UC San Diego

UC San Diego Previously Published Works

Title

Nonalcoholic fatty liver disease risk and histologic severity are associated with genetic polymorphisms in children

Permalink

https://escholarship.org/uc/item/6kx269n9

Journal

Hepatology, 77(1)

ISSN

0270-9139

Authors

Goyal, Nidhi P

Rosenthal, Sara B

Nasamran, Chanod

et al.

Publication Date

2023

DOI

10.1002/hep.32570

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Hepatology. Author manuscript; available in PMC 2024 January 01.

Published in final edited form as:

Hepatology. 2023 January 01; 77(1): 197–212. doi:10.1002/hep.32570.

Nonalcoholic fatty liver disease risk and histologic severity are associated with genetic polymorphisms in children

Nidhi P. Goyal[1,2], **Sara B. Rosenthal**[3], **Chanod Nasamran**[3], **Cynthia A. Behling**[4], **Jorge E. Angeles**[1], **Mark H. Fishbein**[5], **Kathryn E. Harlow**[6], **Ajay K. Jain**[7], **Jean P. Molleston**[8], **Kimberly P. Newton**[1,2], **Patricia Ugalde-Nicalo**[1], **Stavra A. Xanthankos**[9], **Katherine Yates**[10], **Nicholas J. Schork**[11,12], **Kathleen M. Fisch**[3], **Jeffrey B. Schwimmer**[1,2], **for the NASH Clinical Research Network**

[1]Division of Gastroenterology, Hepatology, and Nutrition, Department of Pediatrics, University of California, San Diego School of Medicine, San Diego, California, USA

[2]Department of Gastroenterology, Rady Children's Hospital San Diego, San Diego, California, USA

[3]Center for Computational Biology and Bioinformatics, University of California, San Diego, La Jolla, California, USA

[4]Department of Pathology, Sharp Memorial Hospital; Division of Gastroenterology, Hepatology, and Nutrition, Department of Pediatrics, University of California, San Diego, California, USA

[5]Department of Pediatrics, Feinberg Medical School of Northwestern University, Chicago, Illinois, USA

[6]Riley Hospital for Children At Indiana University Health, Indianapolis, Indiana, USA

[7]Division of Gastroenterology, Hepatology and Nutrition, Department of Pediatrics, St. Louis University, St. Louis, Missouri, USA

[8]Section of Pediatric Gastroenterology, Hepatology and Nutrition, Riley Hospital for Children, Indiana University School of Medicine, Indiana University, Indianapolis, Indiana, USA

Correspondence: Jeffrey B. Schwimmer, Department of Pediatrics, UC San Diego and Rady Children's Hospital, 3020 Children's Way, MC 5030 San Diego, CA 92123, USA. jschwimmer@ucsd.edu.

AUTHOR CONTRIBUTIONS

Study concept and design: Schork and Schwimmer. Data acquisition: Behling, Angeles, Fishbein, Harlow, Jain, Molleston, Newton, Ugalde-Nicalo, Xanthakos, Yates, Schwimmer. Data analysis and interpretation: Rosenthal, Nasamran, Fisch. Manuscript draft: Goyal, Rosenthal, Angeles, Schwimmer. Critical review and revisions: Goyal, Rosenthal, Behling, Fishbein, Harlow, Jain, Molleston, Newton, Ugalde-Nicalo, Xanthankos, Yates, Schork, Fisch, and Schwimmer. All authors read and approved the final manuscript.

[Correction statement added June 25, 2022 after first online publication: A correction was made to the author name "Jean P. Molleston." Text was mistakenly entered as an Acknowledgements statement and has since been removed from the article. We apologize to the author and our readers for these errors.]

CONFLICTS OF INTEREST

Ajay K. Jain advises Mirum Pharma. He consults for and received grants from Camp 4. Jeffrey B. Schwimmer received grants from Intercept, Genfit, and Seraphina. Jean P. Molleston received grants from Gilead, AbbVie, Albireo, and Shire. Stavra A. Xanthankos received grants from Target RWE.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[9]Division of Gastroenterology, Hepatology and Nutrition, Cincinnati Children's Hospital Medical Center, Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA

[10]Department of Epidemiology and Biostatistics, Johns Hopkins University, Baltimore, Maryland, USA

[11]The Translational Genomics Research Institute (TGen), Phoenix, Arizona, USA

[12]Department of Molecular and Cell Biology, The City of Hope National Medical Center, Duarte, California, USA

Abstract

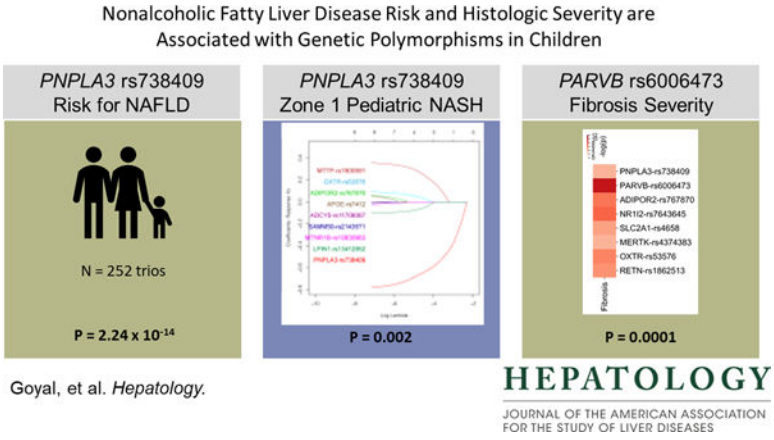
Background and Aims: NAFLD is the most common chronic liver disease in children. Large pediatric studies identifying single nucleotide polymorphisms (SNPs) associated with risk and histologic severity of NAFLD are limited. Study aims included investigating SNPs associated with risk for NAFLD using family trios and association of candidate alleles with histologic severity.

Approach and Results: Children with biopsy-confirmed NAFLD were enrolled from the NASH Clinical Research Network. The Expert Pathology Committee reviewed liver histology. Genotyping was conducted with allele-specific primers for 60 candidate SNPs. Parents were enrolled for trio analysis. To assess risk for NAFLD, the transmission disequilibrium test was conducted in trios. Among cases, regression analysis assessed associations with histologic severity. A total of 822 children with NAFLD had mean age 13.2 years (SD 2.7) and mean ALT 101 U/L (SD 90). *PNPLA3* (rs738409) demonstrated the strongest risk ($p = 2.24 \times 10^{-14}$) for NAFLD. Among children with NAFLD, stratifying by *PNPLA3* s738409 genotype, the variant genotype associated with steatosis ($p = 0.005$), lobular ($p = 0.03$) and portal inflammation ($p = 0.002$). Steatosis grade associated with *TM6SF2* ($p = 0.0009$), *GCKR* ($p = 0.0032$), *PNPLA3* rs738409 ($p = 0.0053$), and *MTTP* ($p = 0.0051$). Fibrosis stage associated with *PARVB* rs6006473 ($p = 0.0001$), *NR1I2* ($p = 0.0021$), *ADIPOR2* ($p = 0.0038$), and *OXTR* ($p = 0.0065$). *PNPLA3* rs738409 ($p = 0.0002$) associated with borderline zone 1 NASH.

Conclusions: This study demonstrated disease-associated SNPs in children with NAFLD. In particular, rs6006473 was highly associated with severity of fibrosis. These hypothesis-generating results support future mechanistic studies of development of adverse outcomes such as fibrosis and generation of therapeutic targets for NAFLD in children.

Graphical Abstract

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



INTRODUCTION

NAFLD is the most common chronic liver disease in children and the incidence is increasing.[1] NAFLD is strongly associated with obesity, type 2 diabetes, and dyslipidemia. However, even among those with risk factors, there remains wide variability regarding susceptibility, and more importantly, progression to NASH, and cirrhosis. Though obesity is commonly present in NAFLD, only about 25% of those with obesity have NAFLD.[2] The prevalence also varies by race and ethnicity. The highest reported rates are in Hispanic children and the lowest reported rates are in Black children.[3] Additionally, 20% of children with NAFLD had a normal BMI. Furthermore, there is wide interindividual variability with disease progression with rapid advancement to severe fibrosis and cirrhosis occurring within a few years of diagnosis.[4] Previous studies have shown the familial nature of NAFLD[5] with increased likelihood of NAFLD in 1st and 2nd degree relatives of affected children. Approximately 40% of the person-to-person variance in liver fat content has been attributed to heritability.[6] These observations allude to a likely role of genetic contributions to NAFLD.

Studies in adults have identified several single nucleotide polymorphisms (SNPs) associated with greater risk of NAFLD as well as more severe liver histology. Studies of the genetics of pediatric NAFLD are more limited. There are also distinct differences in the pathophysiology and histology of NAFLD that hinder extrapolation of results from adults to children. For example, borderline zone 1 NASH has been described as a predominantly pediatric pattern of disease with portal predominance and lack of adult histologic characteristics such as ballooning hepatocytes.[7] These differences raise the likelihood for unique genetic risk alleles in children. Another reason for discrepancies is imaging including ultrasound or magnetic resonance based techniques are often preferred in the pediatric population for disease phenotyping, due to their ease and safety, but they do not allow for the assessment of histopathology. Larger studies in children with liver histology are needed to study genetic disease associations.

We designed the Genetics of Obesity Associated Liver Steatosis (GOALS) study to explore both risk for NAFLD and severity of NAFLD. The overarching study hypothesis was that

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

NAFLD is a polygenic disorder of variable severity and that the genetic variants contributing to the genetic risk for having NAFLD may differ from the variants that are associated with disease severity. The study included two overlapping studies. The first was a family trio study with the aim to investigate candidate alleles influencing the risk for NAFLD in children. The second was performed in children with NAFLD to investigate the association of SNPs with histologic severity. This was done both for *PNPLA3* polymorphism (rs738409) individually (aim 2) and for a panel of selected candidate alleles more broadly (aim 3).

PATIENTS AND METHODS

Study participants

Children with biopsy-confirmed NAFLD were enrolled in the NASH Clinical Research Network (NASH CRN).[8] Each child underwent detailed phenotyping including demographic, clinical, and histological evaluation. Liver pathology was reviewed centrally in consensus by the Pathology Committee. Participants were recruited from the following NASH CRN studies: longitudinal cohort studies of Database and Database 2 (NCT01061684), randomized controlled trials of TONIC (NCT00063635), and CyNCh (NCT01529268). DNA samples were extracted from blood samples stored at the NIDDK biorepository. This study was approved by the institutional review board at each participating center. Inclusion criteria were age <18 years, biopsy-proven NAFLD, consent for DNA testing, and available blood sample. We screened for alcohol using the Alcohol Use Disorders Identification Test and excluded those children with more than sporadic alcohol use. Both biologic parents consented for inclusion in the trio analysis; if only one parent was available the child participated in Aims 2 and 3 only. Written consent was obtained from a parent or guardian, and written assent was obtained from children ≥8 years.

Phenotyping of cohort

Demographic data were obtained via structured interview. Weight and height were measured in duplicate. BMI was calculated as weight (kg) divided by height (m) squared. BMI Z score was used to compare BMI among different ages and sexes. Phlebotomy was done after a 12 h overnight fast for glucose (mg/dl), insulin (mIU/L), glycosylated hemoglobin (HgbA1C, %), total cholesterol (mg/dl), high-density lipoprotein (HDL, mg/dl), low-density lipoprotein (LDL, mg/dl), triglycerides (mg/dl), alanine aminotransferase (ALT, U/L), aspartate aminotransferase (AST, U/L), and gamma glutamyl-transferase (GGT, U/L).

NAFLD diagnosis

Diagnosis was based on the appropriate clinical history, laboratory studies, and liver histology with ≥5% of hepatocytes containing macrovesicular steatosis. The Pathology Committee was masked to demographic and clinical data and reviewed liver biopsy specimens centrally using the NASH CRN scoring system.[9] Biopsies were scored for the degree of steatosis present in hepatocytes as follows: grade 0, <5% steatosis; grade 1, 5–33%; grade 2, 34–66%; and grade 3, >66%. Liver biopsies were diagnosed as definite NASH, borderline zone 1 or zone 3 NASH, or NAFLD not NASH based on the aggregate presence and degree of individual features of NAFLD. A typical set of minimum criteria to diagnose NASH included >5% macrovesicular steatosis, lobular inflammation,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and hepatocyte injury as manifest by ballooning degeneration. Cases determined to be NAFLD not NASH showed >5% steatosis with no/minimal inflammation. This assignment of NASH, borderline NASH, or NAFLD was made as consensus agreement of the Pathology Committee.

Information collected from parents

To evaluate genetic risk for NAFLD in the absence of a control group, a subsample of these children was recruited with their parents for trio analysis. Families who chose to participate were included in this analysis. One approach to determine the risk of NAFLD is to perform a case–control study. There are challenges in the acquisition of verified cases because of the complexity of diagnosis and there are challenges in accurately determining true disease-free status. In lieu of controls, a family trio design can be used to implicate polymorphisms related to disease susceptibility by examining whether an allele is preferentially transmitted to children with NASH from heterozygous parents. Therefore, trios were recruited and basic demographic data including age, sex, race, and ethnicity were collected from parents. No laboratory or histologic data were obtained. The parents were evaluated at the same time as their affected child with NAFLD in order to compare the genetic variants in the genes of the child with those in the parents.

Candidate genes

The genes of interest evaluated in this study were chosen based on a literature review on PubMed with the key words “fatty liver” and “polymorphism.” Preference was given to genes with demonstrated associations with NAFLD in children especially those studies including liver histology. The most popular gene candidates were *PNPLA3* and *TM6SF2*.[10] SNPs associated with NAFLD comorbidities including type 2 diabetes, obesity, cardiovascular disease, and dyslipidemia were also sought (Table S1).

Genotyping methods

Whole-blood samples were collected from each participant in EDTA and stored (−80°C) until DNA extraction. Genomic DNA was extracted from whole-blood samples using Chemagic STAR DNA Blood Kits (Baesweiler, Germany). DNA concentrations and purity were measured using Trinean DropSense/Unchained Lunatic (Pleasanton, CA). A custom panel of amplification and sequencing primers was designed on the web-based Fluidigm D3 Assay Design tool (San Francisco, CA), using tagged, allele-specific PCR primers and a common reverse primer per manufacturer’s protocol. Personnel blinded to participant clinical status and histology performed genotyping of selected SNP variants on the Biomark HD system per manufacturer’s instructions (Fluidigm, San Francisco, CA). PCR conditions were: initial or thermal mix phase consisted of 30 m at 70°C, followed by 10 m at 25°C, and finally 5 m at 95°C; followed by a touchdown period of 4 cycles for 15S at 95°C, followed by 45 s from 64.0 to 61.0°C, dropping 1°C per cycle and 15 s at 72°C; finally, 34 cycles of 15 s at 95°C, 45 s at 60°C and 15 s at 72°C. For quality control, SNPs were excluded from analysis if the call rate was <95%, Hardy–Weinberg equilibrium was $<10^{-4}$, or minor allele frequency was ≤2%.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Statistical analysis

The demographic and clinical characteristics were reported using standard descriptive statistics. Means and standard deviation were reported for continuous variables and counts, and percentages were given for categorical variables. The R packages MASS[11] and nnet[12] were used for all statistical analyses.

Transmission disequilibrium test (TDT)

To assess the risk for NAFLD, the TDT[13] was conducted for 60 SNPs in sets of parents and affected child trios. In TDT analysis, if a particular SNP is not associated with disease risk, one would expect the parents' transmission of alleles at the relevant locus to be random, or each of their alleles being transmitted 50% of the time. If, however, an allele at the SNP locus is associated with diseased risk, there will be overtransmission of that allele (or under-transmission if there is a lower risk). Because TDT analyzes allelic transmissions from heterozygous parents to children, the number of heterozygous parents varied for each SNP. For this reason, parents who were homozygous for a particular SNP were excluded from the analysis. For each test, high-risk allele transmissions were recorded as 1 and low-risk allele transmissions were recorded as 0. The X^2 test was used to assess the association between allelic transmission and NAFLD. Lastly, *p* values were adjusted for multiple comparisons using the Benjamini-Hochberg method. The false discovery rate (FDR) was set at ≤10% for multiple comparisons.

Regression analysis

To assess the association of SNPs with histologic disease severity, regression analysis was conducted on 60 SNPs and various histological response variables, including steatosis, lobular inflammation, portal inflammation, ballooning, fibrosis, and NASH. As a substantial proportion of children were of Hispanic ethnicity, analyses were also conducted separately in the Hispanic only population and the non-Hispanic population. Because the relationships between NASH categories are nonordinal, NASH was modeled using multinomial logistic regression, which does not assume any relationship between categories. The other response variables did not have ambiguous categories and were modeled with ordinal regression.[14] The R functions 'polr' and 'nnet' were used for ordinal and multinomial logistic regression, respectively. As some genes are near *PNPLA3* and in partial linkage disequilibrium (LD) (notably *SAMM50* and *PARVB*), we controlled for the effects of *PNPLA3* on the associations by using *PNPLA3* as a covariate when assessing the histologic association of these SNPs in the regression analysis. SNPs that did not maintain significance when controlling for *PNPLA3* were excluded from further analysis (rs5764455, rs3761472, rs738491, rs6006460). SNPs that maintained significance in at least one phenotype were carried forward in the analysis (rs2143571, rs6006473). The conditioned results are included in Table S2.

An additive genetic model was used for analysis with the assumption that one minor allele has an intermediate effect to two minor alleles. Individuals with two copies of the minor allele at a locus were assigned a score of 2, individuals with one copy of the minor allele were assigned a score of 1, and individuals with zero copies of the minor allele were assigned a score of 0. Potential confounding variables were included as covariates in the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

model. Three separate sets of covariates were tested, as follows: 1: no covariates, 2: age, sex, height, weight, ethnicity, 3: age, sex, BMI Z score, ethnicity. The Benjamini-Hochberg method of controlling FDR was applied to the regression results, to correct for 56 SNPs tested.[15]

Least absolute shrinkage and selection operator (LASSO) regression[16] was used to assess the relative feature importance among the nominally significant results, using functionality from the R packages 'glmnet'[17] and 'polr'.[18] LASSO regression was utilized to improve the prediction accuracy and interpretation of the regression analyses by selecting the SNP most informative to a histologic outcome of interest. LASSO regression enables analysis of data where variables are not fully independent. By introducing a shrinkage penalty, the LASSO model encourages models with fewer independent variables. As the shrinkage penalty is increased, independent variables that contribute less unique signal to the model will be driven to zero. In this application, some SNPs are in LD and thus are correlated. SNPs that are driven to zero earlier are likely not contributing a lot to the model apart from the other included SNPs. We measured the coefficients as the penalty term in the regression increased and recorded the relative weights. The LASSO method was applied to all nominally significant SNPs in the fibrosis and NASH models.

Multi-SNP predictive model

A multi-SNP predictive model was constructed using all SNPs that were nominally significant in the fibrosis model, to find the weights for each SNP. The model was used to discriminate cases with absent-to-mild fibrosis from those with moderate-to-severe fibrosis. All nominally significant individual SNPs were included along with ALT and control variables of age, sex, height, weight, and ethnicity. Five-fold cross validation was used to evaluate predictive power of the model on fibrosis severity. The model was fit on 4 folds of training data and used for prediction on the held-out test dataset fold. This procedure was repeated for each of the 5 folds of data, to generate predictions of moderate-to-severe fibrosis, which were compared to the true values from the held-out data.

RESULTS

Study population

The study included 822 children with biopsy-proven NAFLD (Table 1). Of these, 69.5% (571/822) were male with mean age 13.2 years (SD 2.7). A majority of the study population were of Hispanic ethnicity; 63.4% (521/822). The mean BMI was 32.4 kg/m^2 (SD 6.5) with mean ALT 101 U/L (SD 90). Of the 822 children, 47 (5.7%) had pre-existing type 2 diabetes. The participants encompassed the full range of the histologic spectrum of NAFLD (Table 1). Of note, 62% (508/822) had grade 2–3 steatosis and 27% (223/822) had fibrosis stage 2–4.

Family Trios - Risk for Having NAFLD

To assess the genetic risk of transmission of NAFLD, the trio study included 756 participants comprising 252 complete trios with a child with biopsy-proven NAFLD and their mother and father. There was no significant difference in severity of liver histology

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

between children in the larger study population and the subset of children in the trio analysis (Table 1).

In TDT analysis, with FDR ≤10% to correct for multiple comparisons, 10 SNPs in a total of 8 genes were significantly associated with the risk of having NAFLD (Table 2). SNPs in *PNPLA3* (s738409), *SAMM50* (rs3761472), *PARVB* (rs6006473), and *SLC2A1* (rs4658) were associated with increased risk of NAFLD as the variant alleles were enriched in our population, with rs738409 in *PNPLA3* having the strongest risk association ($p = 2.24 \times 10^{-14}$). Whereas SNPs in *SLC27A5* (rs56225452), *PNPLA3* (rs6006460), *LYPLAL1* (rs12137855), *APOE* (rs74120), and *TNF* (rs361525) conferred reduced risk of NAFLD.

Analysis of *PNPLA3* rs738409 genotype in children with NAFLD

In this study, the evaluation of *PNPLA3* rs738409 was a stated aim, therefore the characteristics of children with NAFLD were stratified by *PNPLA3* rs738409 genotype (Table 3). Overall, 49% were homozygous variant (GG) and the G allele frequency was 0.66. Children with the GG genotype were significantly younger than children with the CC genotype; 12.9 years (SD 2.7) vs 13.8 years (SD 3.0, $p < 0.001$). The G allele frequency in those of Hispanic ethnicity was 0.78 compared to 0.45 in non-Hispanics ($p < 0.0001$). Children with the GG genotype had a significantly lower mean weight than children with the CC genotype; 79 kg (SD 25) vs 94 kg (SD 28, $p < 0.0001$). Furthermore, children with variant GG genotype also had significantly lower triglycerides than children with the CC genotype; mean 139 mg/dl (SD 81) vs 164 mg/dl (SD 98) ($p = 0.009$). However, despite lower weight and triglycerides, those with the GG genotype had a significantly worse steatosis grade ($p = 0.005$). In addition to steatosis, GG genotype compared to the CC genotype was associated with significantly greater lobular ($p = 0.03$) and portal inflammation ($p = 0.002$). Notably, there was a much higher proportion of children with borderline zone 1 (type 1b) NASH in those with the GG genotype (39%) than in those with the CC genotype (9%). Of those participants with borderline zone 1 NASH, 70% ($n = 143/204$) had the GG genotype. Children with the heterozygous CG genotype tended to be intermediate between the CC and GG cohorts with respect to mean age, serum triglyceride levels and frequency of borderline zone 1 NASH, which demonstrates an additive effect of each G allele (Table 3).

Separate analyses were done by ethnicity (Hispanic or non-Hispanic) and are presented in Tables S3 and S4. In the Hispanic ethnicity cohort 63% were homozygous variant (GG). Hispanic children with the GG genotype also had more severe steatosis ($p = 0.006$), NASH ($p = 0.01$), lobular ($p = 0.05$), and portal inflammation ($p = 0.0004$). Of the Hispanic children with borderline zone 1 NASH, 77% had the GG genotype, reflecting a pattern similar to the study population as a whole.

Liver histology: Individual features grade and stage

We evaluated the strength of the association between candidate SNPs and individual histologic features including grades of steatosis, lobular inflammation, portal inflammation, and ballooning degeneration, as well as fibrosis stage (Table 4). The results from ordinal regression analysis for disease-associated SNPs were combined and summarized

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

in heatmaps, showing both the regression *p* values and effect sizes (Figure 1). Heatmaps separated by ethnicity are presented in Figures S1 and S2, respectively. SNPs in partial LD with *PNPLA3* (*SAMM50* and *PARVB*) were only included after controlling for the effects of *PNPLA3* on the histologic associations in the regression analysis.

Steatosis

Four SNPs in *TM6SF2* ($p = 0.0009$), *GCKR* ($p = 0.0032$), *PNPLA3* rs738409 ($p = 0.0053$), and *MTTP* ($p = 0.0051$) were significantly associated with severity of steatosis after correcting for false discovery. *TM6SF2* demonstrated the most significant association with steatosis. LASSO regression analysis was performed to determine which SNPs contributed most to the association with each histologic feature (Figure 1). The SNPs for *GCKR*, *TM6SF2*, and *PNPLA3* remained significant with the highest lambda value in concordance with the ordinal regression analysis. When stratified by ethnicity, *GCKR* and *TM6SF2* were nominally associated with steatosis grade in Hispanic children and *PNPLA3* was nominally associated with steatosis in non-Hispanic children (Tables S5 and S6). SNPs not associated with the larger cohort emerged when data were stratified by ethnicity. In the Hispanic cohort, rs11932595 in *CLOCK* ($p = 0.0398$) and in the non-Hispanic group rs1862513 in *RETN* ($p = 0.0108$) and rs11708067 in *ADCY5* ($p = 0.0117$) were associated with steatosis.

Lobular inflammation

PNPLA3 rs738409 ($p = 0.0004$) and *UCP2* ($p = 0.0110$) were associated with lobular inflammation and *PNPLA3* remained significant after FDR correction.

Portal inflammation

Four SNPs in *PNPLA3* rs738409 ($p = 0.0023$), *ENPP1* ($p = 0.00943$), *ADIPOR2* ($p = 0.0172$), and *SAMM50* ($p = 0.0381$) were nominally associated with portal inflammation. *PNPLA3* was the most significant and *ADIPOR2* had the largest effect size, although none were significant after FDR correction. Additional SNPs that emerged in the analyses by ethnicity were *PPARa* rs1800206 ($p = 0.032$) for Hispanic children and *LEPR* rs1137101 ($p = 0.0325$) for non-Hispanic children.

Ballooning

Two SNPs were nominally associated with ballooning: *TNF* ($p = 0.0310$) and *HHEX-IDE* ($p = 0.0494$). Neither remained significant after FDR correction.

Fibrosis

Eight SNPs were associated with fibrosis, of which four remained significant after FDR correction. These SNPs were found in *PARVB* rs6006473 ($p = 0.0001$), *NR1I2* ($p = 0.0021$), *ADIPOR2* ($p = 0.0038$), and *OXTR* ($p = 0.0065$). *PARVB* was the most significant and *ADIPOR2* had the largest effect size. Because rs6006473 is in LD with variants in the *SAMM50-PNPLA3* locus, the relationship of rs6006473 with fibrosis, is depicted separately for each *PNPLA3* genotype in Figure S3. Notably, *PNPLA3* rs738409, was nominally associated with fibrosis, but did not survive FDR correction. In LASSO regression analysis, SNPs in *ADIPOR2* and *PARVB* contributed most to the model for fibrosis. Similarly, in

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Hispanic children alone, *PARVB* ($p = 0.001$) and *NR1I2* ($p = 0.036$) were associated with fibrosis and survived FDR correction. Both *PARVB* ($p = 0.0244$) and *NR1I2* ($p = 0.0158$) were also associated with fibrosis in non-Hispanic children. An additional SNP, rs12495941 in *ADIPOQ* ($p = 0.0296$), was associated in non-Hispanic children.

Diagnostic categories of NASH

The association between candidate SNPs and NASH diagnosis is shown as heatmaps in Figure 2. In total, nine SNPs were associated with a category of NASH, however only *PNPLA3* rs738409 survived FDR correction. *PNPLA3* rs738409 ($p = 0.0002$) was strongly associated with borderline zone 1 NASH (type 1b NASH) and remained significant after FDR correction. It also had the largest effect size for any category of NASH. This was also true for Hispanic children, *PNPLA3* rs738409 was strongly associated with borderline zone 1 NASH ($p = 0.0002$) (Table S5). *PNPLA3* was not associated with borderline zone 3 (type 1a) or definite NASH (type 2). *MTTP* was most significantly associated with definite NASH ($p = 0.0026$), and had the largest effect size, however it did not survive FDR correction. Two SNPs were nominally associated with borderline zone 3 NASH: *LPIN1* ($p = 0.0197$) and *ADCY5* ($p = 0.0403$). In LASSO regression analysis, *PNPLA3* contributed most to borderline zone 1 NASH, and *MTTP* contributed most to definite NASH, paralleling the regression analysis. NASH was also examined as an ordinal variable and when borderline zone 1 NASH is collapsed with borderline zone 3 NASH, *MTTP* rs1800591 was the most associated with NASH category (Table S7). When borderline zone 1 NASH was compared to NAFLD (no NASH), *PNPLA3* rs739409 was the most associated with borderline zone 1 NASH (Table S8). These results reflected the above categorical regression results.

Multi-SNP predictive model

A multi-SNP predictive model was constructed to determine fibrosis severity using all the SNPs nominally associated with fibrosis (rs6006473, rs767870, rs53576, rs1862513, rs4658, rs3761472, rs4374383, rs738409). The model including genetics and clinical variables had an AUC of 0.74 (Figure 3).

DISCUSSION

We studied the genetic associations of biopsy-proven NAFLD in a large, multicenter, well-characterized population of children. This included a family trio study to assess the genetic risks for having NAFLD as well as regression analyses to evaluate the association of candidate SNPs with histologic severity of NAFLD. This study contributes to the understanding of genetics of NAFLD in children as we have demonstrated associations not previously found in children. In the trio analysis, the SNP most significantly associated with transmission of NAFLD was rs738409 in the *PNPLA3* gene. This present study demonstrated an association of the SNPs in *SLC27A5*, *SAMM50*, *PARVB* (rs6006473), *LYPLAL1*, *PNPLA3* (rs6006460), *TNF*, *APOE* and *SLC2A1* with a risk for NAFLD in children. There were also associations demonstrated with disease severity including rs6006473 in *PARVB* and *ADIPOR2* with fibrosis, *TM6SF2* and *GCKR* with histologic steatosis, and *PNPLA3* with borderline zone 1 NASH. The associations of *SAMM50* and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

rs6006473 in *PARVB* were only included after controlling for the effects of *PNPLA3* on the histologic associations in the regression analysis as these SNPs are in partial linkage.

An important finding in the present study is the association of rs6006473 with fibrosis stage in pediatric NAFLD. This SNP demonstrated the most significant association in the whole cohort as well in the Hispanic cohort alone. This SNP is in the intergenic region near the *PARVB* gene that encodes parvin-β, which forms integrin-linked kinase-pinch-parvin complex. Integrin receptors bind to extracellular matrix components, thus they are thought to play a crucial role in the evolution and progression of fibrosis.[19] Only one study to date has reported on rs6006473 in association with liver histology in NAFLD. In a GWAS of over 200,000 SNPs, this particular SNP demonstrated the strongest association with the severity of fibrosis in 392 Japanese adults with NAFLD.[20]

PNPLA3 rs738409 remains the best studied NAFLD associated SNP and demonstrates the strongest association with NAFLD in both children and adults. *PNPLA3* appears to be a triglyceride hydrolase and the hydrolase activity is reduced with the variant protein. In a mechanistic study, it was demonstrated that steatosis associated with the variant SNP is caused by accumulation of *PNPLA3* on lipid droplets.[21] The G allele frequency in our NAFLD cohort was 0.66 overall and 0.78 in those of Hispanic ethnicity. The allele frequency in the general population is 0.23.[22] Published data in children with obesity have demonstrated the G allele frequency is estimated at 0.32 in Caucasian and 0.48 in Hispanic children.[23] The allele frequency in our population was likely higher as it included a larger proportion of children with Hispanic ethnicity and only children with histologic NAFLD. Several prior studies have evaluated the risk *PNPLA3* variant allele confers on children for NAFLD. There are two studies in children evaluating the association of the *PNPLA3* allele with MRI measured liver fat. In a study assessing hepatic steatosis as measured by MRI measured hepatic fat fraction (HFF) in 85 adolescents with obesity, the frequency of the variant *PNPLA3* allele was significantly higher in Caucasian and African American children with hepatic steatosis, but not in those of Hispanic ethnicity.[23] In another study evaluating MRI HFF in 188 Hispanic children, those with the *PNPLA3* homozygous variant allele demonstrated MRI HFF of 11% versus 5% ($p < 0.0001$) in those with the homozygous wildtype allele.[24] The present study confirms the prior associations of *PNPLA3* and risk of NAFLD in children. The lesser-known variant of *PNPLA3* (rs6006460) has been associated with lower hepatic fat content and is the more frequent allele in African Americans.[25] In the present study, the wildtype allele was overrepresented, indicating that the variant allele is protective, corroborating the prior study.

In addition to the risk of NAFLD, the present study demonstrated an association of *PNPLA3* with the severity of NAFLD in children, where prior studies in children had conflicting results. In this study, the *PNPLA3* variant allele was associated with both lobular and portal inflammation. More importantly, of the 204 children with borderline zone 1 NASH, which is primarily seen in pediatrics, 70% were homozygous for the variant *PNPLA3* allele (GG). This indicates that *PNPLA3* may confer a different disease phenotype in children with NAFLD and that different histologic patterns may be relevant. For example, in a therapeutic trial with cysteamine for pediatric NAFLD, children with borderline zone 1 NASH pattern had 4 times the odds of having histologic improvement with cysteamine compared to

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

placebo.[26] In a GWAS with 1483 European adults with NAFLD and 17,781 controls, *PNPLA3* (rs738409) was significantly associated with steatosis, fibrosis and NAS score.[27] In a study of 223 children with histologic NAFLD, there was no association between *PNPLA3* (rs738409) and histologic severity.[28] In a study of 149 Italian children ages 6–13 years with histologic NAFLD, the rs738409 G allele was associated with the severity of steatosis ($p < 0.0001$), the presence of NASH ($p < 0.0001$), hepatocellular ballooning ($p < 0.0001$), lobular inflammation ($p < 0.0001$), and the presence of fibrosis ($p = 0.01$).[29] In a study from Germany of 70 adolescents with biopsy-proven NAFLD compared to 200 controls, the G allele was associated with a portal pattern of steatosis, inflammation and fibrosis.[30] Our study confirms the findings that the variant *PNPLA3* allele is associated with the severity of NAFLD and that each variant allele has an additive effect with the homozygous variant children with GG having worse disease than heterozygous carrier. In this study, this SNP is predominantly associated with borderline zone 1 NASH in children. In contrast to prior studies, this study, after FDR correction, did not support an association with *PNPLA3* and fibrosis stage in pediatric NAFLD. The prior studies were in cohorts of children from Europe principally of white race and non-Hispanic ethnicity, which may contribute to some of these phenotypic differences.

In this study, the *TM6SF2* gene polymorphism demonstrated the most significant association with steatosis severity. *TM6SF2* is responsible for regulating lipid metabolism in the liver and prior studies demonstrated that *TM6SF2* overexpression is associated with reduced liver cell steatosis.[31] To date, disease associations of *TM6SF2* with pediatric NAFLD have been demonstrated in imaging studies,[32–34] and our study demonstrated an association with the risk and severity of histologic NAFLD in children. In histologic studies with European adults, this SNP was associated with NASH, severity of steatosis, and severity of fibrosis.[27,35,36] In contrast, in GWAS of 1483 European adults and 17,781 controls, *TM6SF2* was associated with the risk of NAFLD, but did not reach significance when evaluated for steatosis, NASH, or fibrosis.[27]

This was a large multicenter study of well-characterized patients. All children had a complete diagnostic evaluation including histology, which was evaluated in standardized fashion by expert liver pathologists. A broad spectrum of disease severity was captured as nearly one third of children had stage 2 fibrosis or higher. There were two components to this study including a trio analysis to determine risk for NAFLD and a case only assessment of disease severity. Although the trio design assessing the risk for NAFLD did not have controls as would a classic case–control study, the family trio design uses transmission disequilibrium to overcome the lack of a control group. This statistical approach can accommodate linkage and direct association in addition to being unaffected by population stratification. For the severity of disease analysis, a major strength was the sample size of 822 children with biopsy-proven NAFLD. In this analysis, a healthy control group would not be relevant. Our study included a broad panel of 60 SNPs thought to contribute to the pathogenesis of NAFLD. However, the associations found may not be attributable to the SNPs alone and may not be associated with a change in the gene associated with the SNP. The patient population had a large representation of Hispanic children; thus, results may not be generalizable in other populations, however we adjusted for ethnicity in the analyses. All significant findings were considered risk associations and not causative. We recognize

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

that the field continues to evolve, and as such there were some SNPs not included in this study notably those in the genes *MARC1, MBOAT7*, and *HSD17B13*. This study included 60 disease specific SNPs but was not exhaustive.

Pediatric NAFLD is a heterogenous disease with a polygenic basis underlying development of the disease as well as the severity of liver histology. Using an approach targeted to SNPs of interest, the present study has demonstrated several that have not been previously identified as disease-associated polymorphisms in children. Using TDT analysis, the *PNPLA3* (s738409) variant SNP was confirmed to be highly associated with the risk for NAFLD. Additionally, this *PNPLA3* SNP was significantly enriched in children with borderline zone 1 NASH, which is a unique histologic pattern in children with NAFLD. To date, pediatric clinical trials have not stratified treatment for genotype, but this may be relevant and portend differential response to therapy. Additionally, the genetic associations may differ based on ethnicity of the children, and this deserves further study. Furthermore, *PARVB* (rs6006473) was very significantly associated with fibrosis, and this is a disease severity association in children with NAFLD. This reinforces that NAFLD in children may be a distinct disease compared to adults. Recognizing these SNPs in children and determining how they differentially affect histologic severity and fibrosis may guide treatment of NAFLD in children. Using these genetic predispositions as guides, the field can begin to stratify disease risk, prognosis, and response to treatment in children. The hypothesis-generating results of this study support future mechanistic studies to evaluate the development of adverse outcomes such as fibrosis. By identifying SNPs associated with fibrosis, such as rs6006473, therapeutic interventions may target pathways known to be associated with those particular genes in an attempt to curtail fibrosis. Elucidating such genetic associations can be pivotal in generating future therapeutic targets in NAFLD, which continues to lack effective treatments in children.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

FUNDING INFORMATION

The TONIC trial was conducted by the NASH CRN and supported in part by the Intramural Research Program of the National Cancer Institute and the Eunice Kennedy Shriver National Institute of Child Health and Human Development. The vitamin E and matching placebo were provided by Pharmavite through a Clinical Trial Agreement with the NIH. The CyNCh trial was conducted by the NASH CRN and supported in part by the Intramural Research Program of the National Cancer Institute and by a Collaborative Research and Development Agreement (CRADA) between NIDDK and Raptor Pharmaceuticals. The project was also supported by the Rady Children's Hospital Academic Enrichment Fund, the UC San Diego Altman Clinical and Translational Research Institute supported by NIH Grants UL1TR000100, UL1TR001442, the UC San Diego Center for Computational Biology & Bioinformatics, and the San Diego Digestive Diseases Research Center supported by NIDDK grant P30 DK120515.

National Center for Advancing Translational Sciences, Grant/Award Number: UL1TR000004, UL1TR000006, UL1TR000040, UL1TR000077, UL1TR000100, UL1TR000150, UL1TR000423, UL1TR000424, UL1TR000448 and UL1TR000454; National Institute of Diabetes and Digestive and Kidney Diseases, Grant/Award Number: U01DK061713, U01DK061718, U01DK061728, U01DK061730, U01DK061731, U01DK061732, U01DK061734, U01DK061737, U01DK061738 and U24DK061730

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

DATA AVAILABILITY STATEMENT

Deidentified data for NASH CRN studies are made available by the NIDDK Central Repository at https://repository.niddk.nih.gov/home/.

APPENDIX

Members of the Nonalcoholic Steatohepatitis Clinical Research Network, Pediatric Clinical Centers: Baylor College of Medicine, Houston, TX: Donna Garner, CPNP; Paula Hertel, MD; Alicia Lawson, BS; Yen Pham, MD; Nicole Triggs, CPNP; Cincinnati Children's Hospital Medical Center, Cincinnati, OH:Kristin Bramlage, MD; April Carr, BS, CCRP; Kim Cecil, PhD; Meghan McNeill, MS; Marialena Mouzaki, MD; Andrew Trout, MD; Stavra Xanthakos, MD; Emory University, Atlanta, GA:Adina Alazraki, MD; Rebecca Cleeton, MPH, CCRP; Maria Cordero, CCRP; Saul Karpen, MD, PhD; Miriam Vos, MD, MSPH, FAHA; Indiana University School of Medicine/Riley Hospital for Children, Indianapolis, IN:Molly Bozic, MD; Laura Carr, RN; Oscar W. Cummings, MD; Kathryn Harlow, MD; Ann Klipsch, RN; Jean P. Molleston, MD; Wendy Morlan, RN; Emily Ragozzino; Girish Rao, MD; Cindy Sawyers; Northwestern University Feinberg School of Medicine/Ann & Robert H. Lurie Children's Hospital of Chicago:Angela Anthony, CRC; Mark H. Fishbein, MD; Saeed Mohammad, MD; Saint Louis University, St Louis, MO:Danielle Carpenter, MD; Theresa Cattoor, RN; Janet Freebersyser, RN; Ajay K Jain, MD; Susan Torretta; University of California San Diego, San Diego, CA:Cynthia Behling, MD, PhD; Janis Durelle; Nidhi P. Goyal, MD, MPH; Kimberly P. Newton, MD; Jeffrey B. Schwimmer, MD; Jaret Skonieczny; Patricia Ugalde-Nicalo, MD; Andrew Wang, MD; University of Washington Medical Center and Seattle Children's Hospital, Seattle, WA:Niviann Blondet, MD; Kara Cooper; Randolph Otto, MD; Matthew Yeh, MD, PhD; Melissa Young; Washington University, St. Louis, MO:Elizabeth M. Brunt, MD (2008–2015); *Resource Centers*: National Cancer Institute, Bethesda, MD:David E. Kleiner, MD, PhD; National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD:Edward C. Doo, MD; Sherry Hall, MS; Jay H. Hoofnagle, MD; Averell H. Sherker, MD; Rebecca Torrance, RN, MS; Patricia R. Robuck, PhD, MPH (2002–2011); Data Coordinating Center, Johns Hopkins University, Bloomberg School of Public Health, Baltimore, MD:Peggy Adamo, BS; Patricia Belt, BS; Jeanne M. Clark, MD, MPH; Jill Meinert; Laura Miriel, BS; Carrie Shade; Emily P. Sharkey, MPH, MBA; Jacqueline Smith, AA; Michael Smith, BS; Alice Sternberg, ScM; James Tonascia, PhD; Mark L. Van Natta, MHS; Annette Wagoner; Laura A. Wilson, ScM; Tinsay Woreta, MD, MPH; Katherine P. Yates, ScM

Abbreviations:

ALT	alanine aminotransferase
AST	aspartate aminotransferase
AUC	area under the curve
BMI	body mass index

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

CRN	clinical research network
GGT	gamma glutamyltransferase
HDL	high-density lipoprotein
HgbA1c	glycosylated hemoglobin
LASSO	least absolute shrinkage and selection operator
LDL	low-density lipoprotein
MRI	magnetic resonance imaging
ROC	receiver operating curve
SNP	single nucleotide polymorphism
TDT	transmission disequilibrium test

REFERENCES

1. Sahota AK, Shapiro WL, Newton KP, Kim ST, Chung J, Schwimmer JB. Incidence of nonalcoholic fatty liver disease in children: 2009–2018. Pediatrics 2020;146(6):e20200771. [PubMed: 33214329]
2. Yu EL, Golshan S, Harlow KE, Angeles JE, Durelle J, Goyal NP, et al. Prevalence of nonalcoholic fatty liver disease in children with obesity. J Pediatr 2019;207:64–70. [PubMed: 30559024]
3. Schwimmer JB, Deutsch R, Kahen T, Lavine JE, Stanley C, Behling C. Prevalence of fatty liver in children and adolescents. Pediatrics 2006;118(4):1388–93. [PubMed: 17015527]
4. Molleston JP, White F, Teckman J, Fitzgerald JF. Obese children with steatohepatitis can develop cirrhosis in childhood. Am J Gastroenterology 2002;97(9):2460–2.
5. Loomba R, Schork N, Chen CH, Bettencourt R, Bhatt A, Ang B, et al. Heritability of hepatic fibrosis and steatosis based on a prospective twin study. Gastroenterology 2015;149(7):1784–93. [PubMed: 26299412]
6. Schwimmer JB, Celedon MA, Lavine JE, Salem R, Campbell N, Schork NJ, et al. Heritability of nonalcoholic fatty liver disease. Gastroenterology 2009;136(5):1585–92. [PubMed: 19208353]
7. Schwimmer JB, Behling C, Newbury R, Deutsch R, Nievergelt C, Schork NJ, et al. Histopathology of pediatric nonalcoholic fatty liver disease. Hepatology 2005;42(3):641–9. [PubMed: 16116629]
8. Patton H, Lavine JE, Van Natta ML, Schwimmer JB, Kleiner D, Molleston J, et al. Clinical correlates of histopathology in pediatric nonalcoholic steatohepatitis (NASH). Gastroenterology 2008;135(6):1961–71.e2. [PubMed: 19013463]
9. Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 2005;41(6):1313–21. [PubMed: 15915461]
10. Goyal NP, Schwimmer JB. The genetics of pediatric nonalcoholic fatty liver disease. Clin Liver Dis 2018;22(1):59–71. [PubMed: 29128061]
11. Venables WN, Ripley BD. Modern applied statistics with S 4th ed. New York: Springer; 2002.
12. Venables WN, Ripley B. Feed-forward neural networks and multinomial log-linear models Berlin Heidelberg: Springer; 2011.
13. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 1993 Mar;52(3):506–16. [PubMed: 8447318]
14. Agresti A, Kateri M. Categorical data analysis. In: Lovric M, editor. International encyclopedia of statistical science Berlin Heidelberg: Springer; 2011. p. 206–8.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

15. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 1995;57(1):289–300.
16. Tibshiranit R Regression shrinkage and selection via the lasso. J R Stat Soc B 1996;58:267–88.
17. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33(1):1–22. [PubMed: 20808728]
18. Ripley B polr: ordered logistic or probit regression. MASS v7.3–51.4 Published 2019. Available from: https://www.rdocumentation.org/packages/MASS/versions/7.3–51.4/topics/polr
19. Patsenker E, Stickel F. Role of integrins in fibrosing liver diseases. Am J Physiol Gastrointest Liver Physiol 2011;301(3):G425–34. [PubMed: 21659620]
20. Kitamoto T, Kitamoto A, Yoneda M, Hyogo H, Ochi H, Nakamura T, et al. Genome-wide scan revealed that polymorphisms in the PNPLA3, SAMM50, and PARVB genes are associated with development and progression of nonalcoholic fatty liver disease in Japan. Hum Genet 2013;132(7):783–92. [PubMed: 23535911]
21. BasuRay S, Wang Y, Smagris E, Cohen JC, Hobbs HH. Accumulation of PNPLA3 on lipid droplets is the basis of associated hepatic steatosis. Proc Natl Acad Sci U S A 2019;116(19):9521–6. [PubMed: 31019090]
22. NIH National Library of Medicine. ALFA allele frequency 2021. Available from: https://www.ncbi.nlm.nih.gov/snp/rs738409#frequency_tab
23. Santoro N, Kursawe R, D'Adamo E, Dykas DJ, Zhang CK, Bale AE, et al. A common variant in the patatin-like phospholipase 3 gene (PNPLA3) is associated with fatty liver disease in obese children and adolescents. Hepatology 2010;52(4):1281–90. [PubMed: 20803499]
24. Goran MI, Walker R, Le KA, Mahurkar S, Vikman S, Davis JN, et al. Effects of PNPLA3 on liver fat and metabolic profile in Hispanic children and adolescents. Diabetes 2010;59(12):3127–30. [PubMed: 20852027]
25. Romeo S, Kozlitina J, Xing C, Pertsemlidis A, Cox D, Pennacchio LA, et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. Nat Genet 2008;40(12):1461–5. [PubMed: 18820647]
26. Schwimmer JB, Lavine JE, Wilson LA, Neuschwander-Tetri BA, Xanthakos SA, Kohli R, et al. In children with nonalcoholic fatty liver disease, cysteamine bitartrate delayed release improves liver enzymes but does not reduce disease activity scores. Gastroenterology 2016;151(6):1141–54.e9. [PubMed: 27569726]
27. Anstee QM, Darlay R, Cockell S, Meroni M, Govaere O, Tiniakos D, et al. Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically characterised cohort. J Hepatol 2020;73(3):505–15. [PubMed: 32298765]
28. Rotman Y, Koh C, Zmuda JM, Kleiner DE, Liang TJ, the NASH CRN. The association of genetic variability in PNPLA3 with histological severity of non-alcoholic fatty liver disease. Hepatology 2010;52(3):894–903. [PubMed: 20684021]
29. Valenti L, Alisi A, Galmozzi E, Bartuli A, del Menico B, Alterio A, et al. I148M patatin-like phospholipase domain-containing 3 gene variant and severity of pediatric nonalcoholic fatty liver disease. Hepatology 2010;52(4):1274–80. [PubMed: 20648474]
30. Hudert CA, Selinski S, Rudolph B, Bläker H, Loddenkemper C, Thielhorn R, et al. Genetic determinants of steatosis and fibrosis progression in paediatric non-alcoholic fatty liver disease. Liver Int 2019;39(3):540–56. [PubMed: 30444569]
31. Mahdessian H, Taxiarchis A, Popov S, Silveira A, Franco-Cereceda A, Hamsten A, et al. TM6SF2 is a regulator of liver fat metabolism influencing triglyceride secretion and hepatic lipid droplet content. Proc Natl Acad Sci U S A 2014;111(24):8913–8. [PubMed: 24927523]
32. Goffredo M, Caprio S, Feldstein AE, D'Adamo E, Shaw MM, Pierpont B, et al. Role of TM6SF2 rs58542926 in the pathogenesis of nonalcoholic pediatric fatty liver disease: a multiethnic study. Hepatology 2016;63(1):117–25. [PubMed: 26457389]
33. Grandone A, Cozzolino D, Marzuillo P, Cirillo G, di Sessa A, Ruggiero L, et al. TM6SF2 Glu167Lys polymorphism is associated with low levels of LDL-cholesterol and increased liver injury in obese children. Pediatr Obes 2016;11(2):115–9. [PubMed: 25893821]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

34. Di Costanzo A, Pacifico L, Chiesa C, Perla FM, Ceci F, Angeloni A, et al. Genetic and metabolic predictors of hepatic fat content in a cohort of Italian children with obesity. Pediatr Res 2019;85(5):671–7. [PubMed: 30710115]

35. Dongiovanni P, Anstee QM, Valenti L. Genetic predisposition in NAFLD and NASH: impact on severity of liver disease and response to treatment. Curr Pharm Des 2013;19(29):5219–38. [PubMed: 23394097]

36. Liu YL, Reeves HL, Burt AD, Tiniakos D, McPherson S, Leathart JBS, et al. TM6SF2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease. Nat Commun 2014;5:4309. [PubMed: 24978903]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

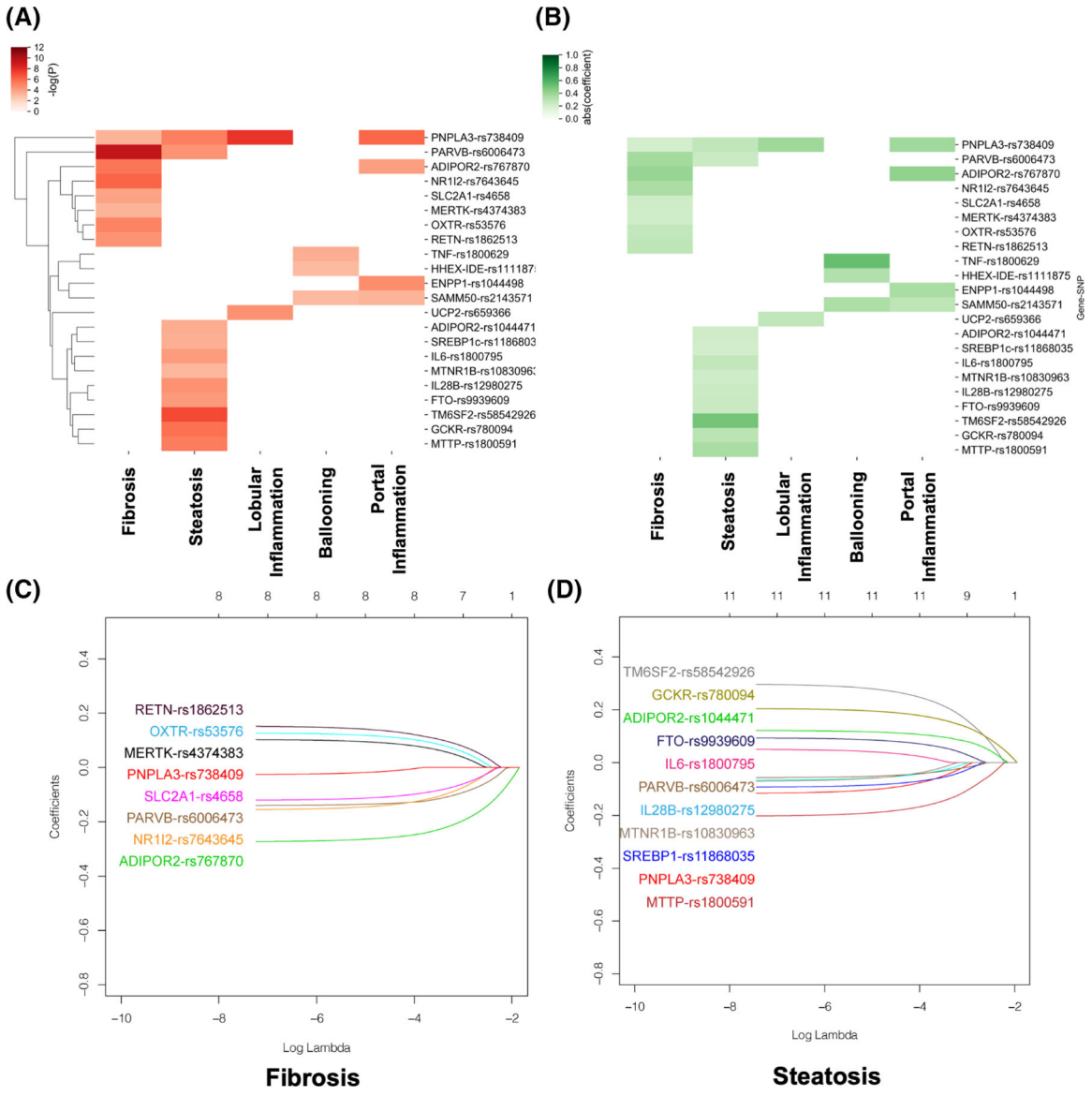


FIGURE 1.

Summary of results from regression analysis. (A) Heatmap shows where there are nominally significant associations between response variables and SNPs of interest (non-white squares). (B) Heatmap shows the absolute value of effect sizes for nominally significant associations (non-white squares). Regression models adjust for Age, Sex, Height, Weight, and Ethnicity. The results from each response variable tested are shown in the heatmap columns, whereas the results from each SNP tested are shown in the rows. Rows and columns are ordered by hierarchical clustering. (C,D) Results from the lasso regression on nominally significant associations, showing relative feature importance, for fibrosis (C), and steatosis (D).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

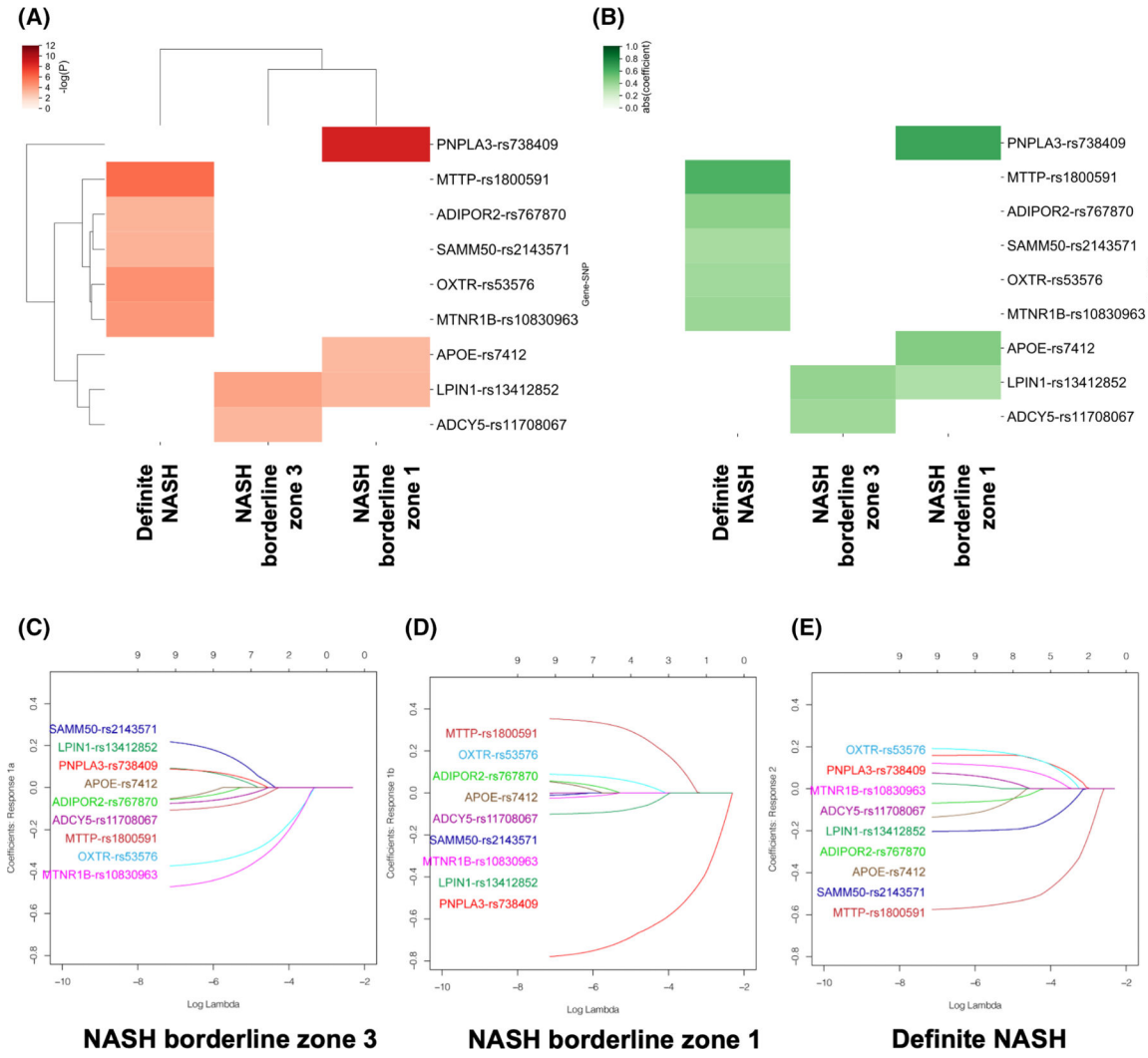


FIGURE 2.

Summary of results from regression analysis. (A) Heatmap shows where there are nominally significant associations between response variables and SNPs of interest (non-white squares). (B) Heatmap shows the absolute value of effect sizes for nominally significant associations (non-white squares). Regression models adjust for Age, Sex, Height, Weight, and Ethnicity. The results from each response variable tested are shown in the heatmap columns, whereas the results from each SNP tested are shown in the rows. Rows and columns are ordered by hierarchical clustering. (C–E) Results from the lasso regression on nominally significant associations, showing relative feature importance, for borderline zone 3 NASH (C), borderline zone 1 NASH (D), and definite NASH (E).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

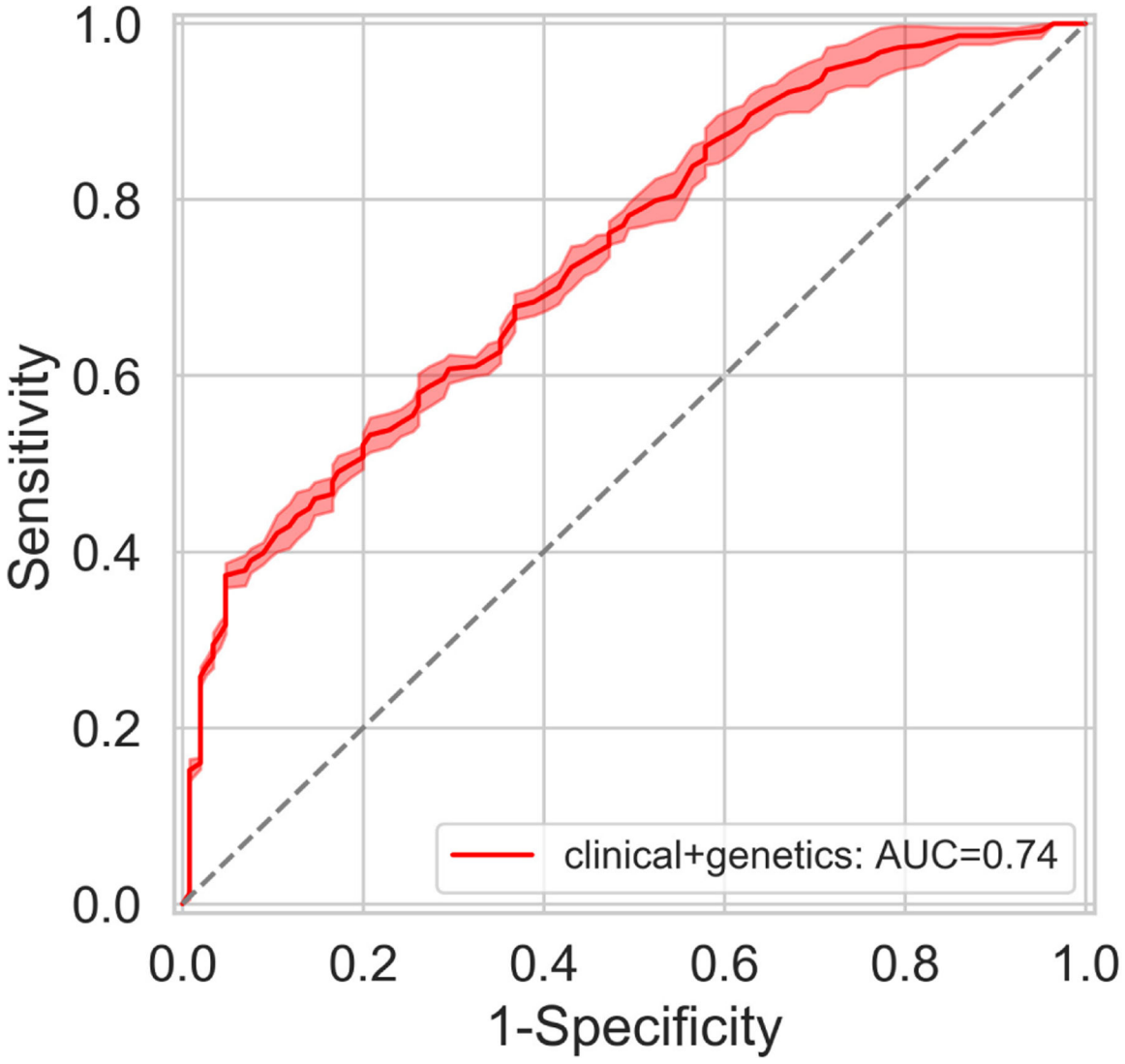


FIGURE 3.

A multi-SNP predictive model constructed using all SNPs that were nominally significant in the fibrosis model. The model discriminates cases with absent-to-mild fibrosis from those with moderate-to-severe fibrosis. All nominally significant individual SNPs were included along with ALT and genetics and control variables of age, sex, height, weight, and ethnicity. The mean ± 1 SD are shown for the area under the curve.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

TABLE 1

Characteristics of the study population

Characteristic	All children (*n* = 822)	Children in trios (*n* = 252)	Parents in trios (*n* = 504)
Age, mean (SD)	13.2 (2.7)	12.5 (2.6)	42.5 (7.3)
Sex, *N*(%)			
Male	571 (69.5%)	189 (75%)	252 (50%)
Female	251 (30.5%)	63 (25%)	252 (50%)
Self-reported race, *N*(%)			
White	552 (67.2%)	158 (62.7%)	261 (51.8%)
Black	25 (3.0%)	2 (0.8%)	8 (1.6%)
Asian	16 (1.9%)	12 (4.8%)	21 (4.2%)
Indian	54 (6.6%)	15 (6.0%)	22 (4.4%)
Native	4 (0.5%)	2 (0.8%)	4 (0.8%)
Multiracial	28 (3.4%)	10 (4.0%)	11 (2.2%)
Other	143 (17.4%)	53 (20.6%)	177 (35.1%)
Self-reported ethnicity, *N*(%)			
Hispanic	521 (63.4%)	184 (73.0%)	364 (72.2%)
Non-Hispanic	301 (36.6%)	68 (67.0%)	140 (27.8%)
Mean weight (SD), kg	84.4 (26.2)	78.7 (26.5)	
Mean height (SD), m	1.6 (0.1)	1.6 (0.1)	
Mean BMI (SD), kg/m^2	32.4 (6.5)	31.1 (6.7)	
BMI Z score	2.3 (0.5)	2.2 (0.4)	
ALT (SD), U/L	101 (90)	111 (109)	
AST (SD), U/L	62 (49)	68 (59)	
GGT (SD), U/L	45 (37)	45 (36)	
Glucose (SD), mg/dl	88 (19)	87 (13)	
Insulin (SD), mIU/L	32 (32)	33 (43)	
HgbA1c (SD), %	5.5 (0.8)	5.5 (0.7)	
LDL (SD), mg/dl	99 (31)	100 (32)	
HDL (SD), mg/dl	39 (10)	40 (10)	
Triglycerides (SD), mg/dl	148 (86)	145 (83)	
Histology Steatosis grade, *N*(%)[a]			
1	200 (28.2%)	65 (27.7%)	
2	210 (29.7%)	69 (29.4%)	
3	298 (42.1%)	101 (42.9%)	
NASH, *N*(%)[b]			
NAFLD not NASH	203 (26.9%)	64 (26.9%)	
Borderline Zone 3	155 (20.5%)	30 (12.6%)	
Borderline Zone 1	211 (27.9%)	86 (36.1%)	
Definite NASH	187 (24.7%)	58 (24.4%)	
Fibrosis stage, *N*(%)[c]			

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Characteristic	All children (*n* = 822)	Children in trios (*n* = 252)	Parents in trios (*n* = 504)
0	228 (30.8%)	71 (29.0%)	
1a	52 (7.0%)	15 (6.1%)	
1b	34 (4.6%)	5 (2.0%)	
1c	204 (27.5%)	79 (32.2%)	
2	117 (15.8%)	34 (13.9%)	
3	93 (12.6%)	36 (14.7%)	
4	13 (1.8%)	5 (2.0%)	
Ballooning[d]			
0	242 (61.3%)	70 (61.4%)	
1	101 (25.6%)	30 (26.3%)	
2	52 (13.2%)	14 (12.3%)	
Lobular Inflammation[e]			
0	3 (0.4%)	0 (0%)	
1	410 (55.3%)	131 (53.5%)	
2	278 (37.5%)	102 (41.6%)	
3	51 (6.9%)	12 (4.9%)	
Portal Inflammation[f]			
0	71 (9.6%)	22 (9.0%)	
1	548 (74.0%)	185 (75.5%)	
2	122 (16.5%)	38 (15.5%)	

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; GGT, gamma glutamyl transferase; HbA1c, glycosylated hemoglobin; HDL, high-density lipoprotein; LDL, low-density lipoprotein.

[a] $p = 0.98$.

[b] $p = 0.017$.

[c] $p = 0.42$.

[d] $p = 0.96$.

[e] $p = 0.56$.

[f] $p = 0.89$.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 2

Trio transmission disequilibrium test. With false discovery rate (FDR) set at ≤10% for testing multiple comparisons, there were 10 SNPs in a total of 8 genes that were significantly associated with the risk of having NAFLD. The probability is the probability of transmission of the variant allele

SNP	Gene	Variant	Wild type	Children with Var	Children with WT	Probability of Var	*p* value	(i/m)Q
Variant alleles associated with increased risk for NAFLD								
rs738409	PNPLA3	G	C	166	53	0.76	2.24E-14	0.002
rs3761472	SAMM50	G	A	140	79	0.64	3.76E-05	0.007
rs738491	SAMM50	T	C	142	91	0.61	0.0008	0.008
rs6006473	PARVB	T	C	148	96	0.61	0.0009	0.010
rs4658	SLC2A1	G	C	57	32	0.64	0.0080	0.015
Variant alleles associated with decreased risk for NAFLD								
rs56225452	SLC27A5	T	C	6	36	0.14	3.67E-06	0.003
rs6006460	PNPLA3	T	G	8	39	0.17	6.13E-06	0.005
rs12137855	LYPLAL1	T	C	43	77	0.36	0.0019	0.012
rs7412	APOE	T	C	18	42	0.30	0.0019	0.013
rs361525	TNF	A	G	13	29	0.31	0.0136	0.017

Note: (i/m)Q – Benjamini-Hochberg correction.

Abbreviations: WT, wild type; Var, variant.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

TABLE 3

Characteristics of children with NAFLD by *PNPLA3* genotype

	PNPLA3 rs438709 Genotype			
Characteristic (*n* = 810)	**CC (*n* = 129)**	**CG (*n* = 287)**	**GG (*n* = 394)**	***p* value**
Age, mean (SD)	13.8 (3.0)	13.4 (2.6)	12.9 (2.7)	<0.001
Sex, *N* (%)				0.712
Male	89 (69%)	195 (33.1%)	279 (70.8%)	
Female	40 (31%)	92 (66.9%)	115 (29.2%)	
Race, *N* (%)				<0.0001
Hispanic	34 (26.4%)	152 (53%)	326 (83%)	
Non-Hispanic	95 (73.6%)	134 (47%)	67 (17%)	
Mean weight (SD), kg	94.0 (27.7)	87.5 (25.1)	79.1 (25.3)	<0.0001
Mean height (SD), m	163 (17)	161 (13)	158 (13)	<0.0001
BMI (SD), kg/m^2	34.5 (6.7)	33.2 (6.5)	31.1 (6.2)	<0.0001
BMI Z score	2.36 (0.60)	2.29 (0.37)	2.18 (0.44)	0.0001
ALT (SD), U/L	91 (83)	95 (89)	109 (94)	0.053
AST (SD), U/L	60 (50)	58 (47)	65 (51)	0.225
GGT (SD), U/L	43 (32)	48 (42)	44 (35)	0.351
Glucose (SD), mg/dl	87 (19)	89 (21)	88 (18)	0.486
Insulin (SD), mIU/L	34 (26)	33 (28)	32 (37)	0.718
HgbA1c (SD), %	5.53 (0.74)	5.50 (0.87)	5.45 (0.74)	0.537
LDL (SD), mg/dl	102 (33)	99 (29)	98 (31)	0.494
HDL (SD), mg/dl	38 (10)	39 (10)	40 (10)	0.151
Triglycerides (SD), mg/dl	164 (98)	152 (85)	139 (81)	0.009
Histology				
Steatosis grade, *N* (%)				0.005
1	32 (32%)	69 (29.7%)	93 (25.5%)	
2	35 (35%)	65 (28.0%)	107 (29.4%)	
3	33 (33%)	98 (42.2%)	164 (45.1%)	
NASH, *N* (%)				0.129
NAFLD not NASH	29 (28.2%)	74 (31.4%)	96 (26.1%)	
Borderline Zone 3	29 (28.2%)	37 (15.7%)	51 (13.9%)	
Borderline Zone 1	9 (8.7%)	52 (22.0%)	143 (38.9%)	
Definite NASH	36 (35.0%)	73 (31.0%)	78 (21.2%)	
Fibrosis stage, *N* (%)				0.019
0	35 (32.4%)	85 (34.0%)	104 (28.0%)	
1a	9 (8.3%)	24 (9.6%)	18 (4.9%)	
1b	13 (12.0%)	12 (4.8%)	9 (2.4%)	
1c	20 (18.5%)	63 (25.2%)	117 (31.5%)	
2	23 (21.3%)	36 (14.4%)	56 (15.1%)	
3	8 (7.4%)	24 (9.6%)	60 (16.2%)	
4	0 (0%)	6 (2.4%)	7 (1.9%)	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

	PNPLA3 rs438709 Genotype			
Characteristic (*n* = 810)	**CC (*n* = 129)**	**CG (*n* = 287)**	**GG (*n* = 394)**	***p* value**
Ballooning, *N*(%)				0.881
0	42 (59.1%)	95 (63.0%)	99 (60.4%)	
1	19 (26.8%)	36 (23.8%)	44 (26.8%)	
2	10 (14.1%)	20 (13.2%)	21 (12.8%)	
Lobular inflammation, *N*(%)				0.03
0	1 (0.9%)	1 (0.4%)	0 (0%)	
1	72 (66.7%)	146 (58.4%)	187 (50.3%)	
2	30 (27.8%)	88 (35.2%)	155 (41.7%)	
3	5 (4.6%)	15 (6.0%)	30 (8.1%)	
Portal inflammation, *N*(%)				0.002
0	19 (17.6%)	30 (12.0%)	22 (5.9%)	
1	75 (69.4%)	184 (73.6%)	281 (75.7%)	
2	14 (13.0%)	36 (14.4%)	68 (18.3%)	

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

TABLE 4

Outcomes table for SNPs associated with histologic features of NAFLD with FDR set at ≤10%

Histologic feature	SNP	Gene	*p* value	BH FDR	Effect size
Steatosis	rs58542926	TM6SF2	0.0009	0.0489	0.4971
	rs780094	GCKR	0.0032	0.0737	0.2929
	rs1800591	MTTP	0.0051	0.0737	0.3442
	rs738409	PNPLA3	0.0053	0.0737	0.2702
	rs12980275	IL28B	0.0106	0.1082	0.2422
	rs6006473	PARVB	0.0119	0.1082	0.2390
	rs9939609	FTO	0.0150	0.1082	0.2480
	rs1800795	IL6	0.0155	0.1082	0.2661
	rs1044471	ADIPOR2	0.0288	0.1783	0.2181
	rs11868035	SREBP1c	0.0318	0.1783	0.2062
	rs10830963	MTNR1B	0.0421	0.2142	0.2180
Lobular inflammation	rs738409	PNPLA3	0.0004	0.0231	0.3811
	rs659366	UCP2	0.0110	0.3072	0.2752
Portal inflammation	rs738409	PNPLA3	0.0023	0.1271	0.3755
	rs1044498	ENPP1	0.0094	0.2642	0.3471
	rs767870	ADIPOR2	0.0172	0.3216	0.4207
	rs2143571	SAMM50	0.0381	0.5340	0.2832
Ballooning	rs1800629	TNF	0.0310	0.8581	0.5283
	rs1111875	HHEX-IDE	0.0494	0.8581	0.3180
Fibrosis	rs6006473	PARVB	0.0001	0.0048	0.3667
	rs7643645	NR1I2	0.0021	0.0581	0.3411
	rs767870	ADIPOR2	0.0038	0.0706	0.4006
	rs53576	OXTR	0.0065	0.0916	0.2647
	rs1862513	RETN	0.0114	0.1282	0.2820
	rs4658	SLC2A1	0.0199	0.1854	0.2233
	rs4374383	MERTK	0.0383	0.2756	0.2110
	rs738409	PNPLA3	0.0394	0.2756	0.2012
Borderline zone 3 NASH	rs13412852	LPIN1	0.0197	0.6818	0.4140
	rs11708067	ADCY5	0.0403	0.6818	0.3795
Borderline zone 1 NASH	rs738409	PNPLA3	0.0002	0.0100	0.6509
	rs13412852	LPIN1	0.0405	0.7463	0.3300
	rs7412	APOE	0.0456	0.7463	0.4533
Definite NASH	rs1800591	MTTP	0.0026	0.1481	0.5976
	rs53576	OXTR	0.0104	0.2559	0.3772
	rs10830963	MTNR1B	0.0137	0.2559	0.3919
	rs2143571	SAMM50	0.0346	0.4229	0.3588
	rs767870	ADIPOR2	0.0378	0.4229	0.4320

Abbreviation: BH FDR, Benjamini-Hochberg false discovery rate.