Genome-Wide Associations Related to Hepatic Histology in Nonalcoholic Fatty Liver Disease in Hispanic Boys

Julia Wattacheril, MD, MPH[1], Joel E. Lavine, MD, PhD[2], Naga P. Chalasani, MD[3], Xiuqing Guo, PhD[4], Soonil Kwon, PhD[4], Jeffrey Schwimmer, MD[5], Jean P. Molleston, MD[6], Rohit Loomba, MD[7], Elizabeth M. Brunt, MD[8], Yii-Der Ida Chen, PhD[4], Mark O. Goodarzi, MD, PhD[9], Kent D. Taylor, PhD[4], Katherine P. Yates, ScM[10], James Tonascia, PhD[10], and Jerome I. Rotter, PhD[4]

Objective To identify genetic loci associated with features of histologic severity of nonalcoholic fatty liver disease in a cohort of Hispanic boys.

Study design There were 234 eligible Hispanic boys age 2-17 years with clinical, laboratory, and histologic data enrolled in the Nonalcoholic Steatohepatitis Clinical Research Network included in the analysis of 624 297 single nucleotide polymorphisms (SNPs). After the elimination of 4 outliers and 22 boys with cryptic relatedness, association analyses were performed on 208 DNA samples with corresponding liver histology. Logistic regression analyses were carried out for qualitative traits and linear regression analyses were applied for quantitative traits.

Results The median age and body mass index *z*-score were 12.0 years (IQR, 11.0-14.0) and 2.4 (IQR, 2.1-2.6), respectively. The nonalcoholic fatty liver disease activity score (scores 1-4 vs 5-8) was associated with SNP rs11166927 on chromosome 8 in the TRAPPC9 region ($P = 8.7^{-07}$). Fibrosis stage was associated with SNP rs6128907 on chromosome 20, near actin related protein 5 homolog ($p = 9.9^{-07}$). In comparing our results in Hispanic boys with those of previously reported SNPs in adult nonalcoholic steatohepatitis, 2 of 26 susceptibility loci were associated with nonalcoholic fatty liver disease activity score and 2 were associated with fibrosis stage.

Conclusions In this discovery genome-wide association study, we found significant novel gene effects on histologic traits associated with nonalcoholic fatty liver disease activity score and fibrosis that are distinct from those previously recognized by adult nonalcoholic fatty liver disease genome-wide association studies. *(J Pediatr 2017;■■:■■-■■).*

Nonalcoholic fatty liver disease (NAFLD) is the most common chronic liver disease in children and adults in the developed world. The histologic expression of the disease ranges from steatosis to steatohepatitis, to marked fibrosis and cirrhosis. Multiple factors contribute to disease progression, including environmental influences such as dietary intake, physical activity, and genetic predisposition. Genome-wide association studies (GWAS) have shown that the heritability of complex traits and complex disorders, such as NAFLD, may be due to multiple genes of small effect size.[1] The genetic susceptibility in NAFLD merits investigation given the implications for disease progression associated with unique pathogenic variants.

The basis for a genetic component in NAFLD is well-founded. Familial clustering, epidemiologic data, and twin studies demonstrate that inherited factors influence the likelihood and severity of pediatric nonalcoholic steatohepatitis (NASH).[2] The heritability of fat fraction as a continuous trait and the heritability of steatosis and fibrosis have been demonstrated using noninvasive means.[3,4] Furthermore, striking ethnic variability found in the prevalence of NAFLD provides further evidence for significant genetic factors impacting pathogenesis.[5,6]

Several large-scale GWAS have sought to identify common genetic variants associated with NAFLD susceptibility and disease progression. These studies have consistently reported strong associations between the nonsynonymous amino acid substitution I148M in the patatin-like phospholipase domain containing 3 gene (PNPLA3), that encodes a triacylglycerol lipase expressed in adipocytes, with the

BMI	Body mass index
CRN	Clinical Research Network
GWAS	Genome-wide association study
NAFLD	Nonalcoholic fatty liver disease
NAS	Nonalcoholic fatty liver disease activity score
NASH	Nonalcoholic steatohepatitis
SNP	Single nucleotide polymorphism

From the [1]Medicine; [2]Pediatrics, Columbia University, New York, NY; [3]Medicine, Indiana University, Indianapolis, IN; [4]Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute and Department of Pediatrics, Harbor-UCLA Medical Center, Torrance; [5]Pediatrics, University of California, San Diego, San Diego, CA; [6]Pediatrics, Indiana University, Indianapolis, IN; [7]Medicine, University of California, San Diego, San Diego, CA; [8]Pathology, Washington University, St. Louis, MO; [9]Medicine, Division of Endocrinology, Diabetes and Metabolism, Cedars-Sinai Medical Center, Los Angeles, CA; and [10]School of Public Health, Johns Hopkins University, Baltimore, MD

Supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) (U01DK061734, U01DK061718, U01DK061728, U01DK061731, U01DK061732, U01DK061737, U01DK061738, U01DK061730, U01DK061713) and National Institute of Child Health and Human Development (NICHD). Also supported by National Institutes of Health Clinical and Translational Science Awards (NIH CTSA) (UL1 TR000040, UL1RR024989, UL1RR025761, M01RR00188, UL1RR024131, UL1RR025014, UL1RR031990, UL1RR025741, UL1RR029887, UL1RR24156, UL1RR025055, UL1RR031980). Genotyping supported in part by UL1 TR000124 and NIDDK DRC DK063491. The authors declare no conflicts of interest.

http://dx.doi.org10.1016/j.jpeds.2017.08.004

risk of steatosis and steatohepatitis.[7,8] Additionally, single nucleotide polymorphisms (SNPs) near TM6SF2 and PPP1R3B have been associated with hepatic steatosis and variants within or near NCAN, GCKR, and LYPLAL1, in addition to PNPLA3, emerged in a study relating these variants to severity of lobular inflammation and fibrosis.[9] Smaller scale GWAS of similar sample size in adults provided additional evidence that genetic variants seen in adults with NAFLD were associated with key histologic features.[10] The I148M variant of PNPLA3, the major genetic risk factor for NASH identified in adults, also is associated with liver enzyme elevations and degree of steatosis in obese children.[11-14]

Epidemiologic data not only points to a genetic component of the disease but also highlights the influence of sex hormones. NAFLD is consistently more prevalent in boys than in girls, which suggests that sex hormones are associated with the predilection for pediatric NAFLD.[15] Given the aforementioned effects from sex hormones and ethnicity, we aimed to minimize noise from heterogeneity in ethnicity and sex by focusing on Hispanic boys.

Methods

All Hispanic boys enrolled in the NASH Clinical Research Network (CRN) in the NAFLD Database I Study (n = 234) were included in this discovery cohort. This multicenter, prospective, longitudinal cohort was established in 2002 by the National Institute of Diabetes and Digestive and Kidney Diseases and contains >4400 subjects. Clinical and histologic features of database participants have been described by Patton et al.[16] These subjects met exclusion criteria for any other potential contributors to fatty liver disease. From the pediatric cohort, all Hispanic boys (2-17 years of age) with liver biopsies were included. Biopsy specimens were reviewed and scored centrally by the NASH CRN Pathology Committee. Specimens were scored according to the histology scoring system established by the NASH CRN.[17] The Institutional Review Boards at each participating center approved the protocols in addition to the NASH CRN Steering Committee. All parents provided consent for their child's participation and all children >7 years of age provided assent.

Genotyping was performed at the Medical Genetics Institute at Cedars–Sinai Medical Center using Illumina OmniExpress chips technology (HumanCNV370-Quadv3 BeadChips; Illumina, San Diego, California[18,19]). Genotypes were determined based on clustering of the raw intensity data for the 2 dyes using Illumina BeadStudio software. Two samples performed in duplicate yielded 100% concordance. Quality controls were performed on the 234 samples and 624 297 SNPs using PLINK.[20] For SNPs on the 22 autosomal chromosomes, we applied the following filter criteria: genotype missing rate >0.02, minor allele frequency <0.05, Hardy-Weinberg equilibrium $P < 10^{-6}$, and heterozygosity >0.53. We also performed quality control at an individual level to check for missing rate and cryptic relatedness ($\hat{\pi}$). Cryptic relatedness refers to unknown more recent relatedness (as opposed to distant relatedness) and includes family relationships such as grandparent–grandchild and full sibling pairs. Population-based association studies assume independent (unrelated) individuals.

We observed no sample with a missing rate of >0.02, but found 22 pairs of samples with $\hat{\pi} \geq 0.25$. Principal component analysis was then carried out using EIGENSTRAT (Harvard, Boston, MA)[21] to examine potential population stratification among our study samples. Four samples were identified as population outliers (spurious component analysis). We thereafter excluded 26 samples (22 cryptic relatedness and 4 principal component analysis outliers) from further association analysis. To adjust for potential population stratification, we included the first 2 principal components as covariates in the model of association analysis. The final dataset for the association analysis after quality controls had 208 samples.

We evaluated the ability to detect an association between an SNP and nonalcoholic fatty liver disease activity score (NAS [NADH CRN, Bethesda, MD]) by power calculation implemented in QUANTO version 1.2.4. Based on the mean and SD of NAS (4.27 ± 1.69) in a preliminary sample of patients enrolled into the NAFLD Database Study, we assessed the power using 208 independent individuals under an additive genetic model. For detectable effect size of >0.8, a sample size of 208 will have enough power (>0.83) to identify the association under additive model with minor allele frequency of >0.1.

Statistical Analyses

The GWAS was performed to identify genetic factors associated with specific features or combinations of liver histology. End points of interest in this study were NAS, definite NASH as defined by prespecified histologic criteria, and fibrosis stage. NAS ranges in score from 1 to 8. In our association analysis, NAS was analyzed as a qualitative (binary) trait by comparing a NAS of ≤4 with a NAS of >4. This cutpoint was chosen based on clinical because a NAS of >4 has been shown to correlate with the presence of NASH.[22] Fibrosis stage was analyzed as a quantitative trait based on stage, which ranged from 0 to 4. The association between the endpoints of interest and each SNP was evaluated in the model with the first 2 PCs as covariates. For each SNP, association analyses were run both under an additive and a dominant genetic model using the PLINK software. Logistic regression analysis was carried out for NAS and definite NASH and linear regression analysis was run for fibrosis grade. We further examined SNPs with $P < 10^{-5}$ from the GWAS (referred below as "top SNPs"). We used SCAN (SNP and CNV Annotation Database; available from: http://www.scandb.org/newinterface/about.html) to annotate genes for the top SNPs.

We also examined previously identified genetic loci associated with NAFLD in adult populations to determine if they generalize to this group of Hispanic boys. We studied the 26 susceptibility loci that were reported in the previously published GWAS.[10] Proxy SNPs found by SNAP[23] were used for SNPs not included in our current study. SNPs with a P value of <.05 in our analysis provided the degree of confirmation for the association with NAS and/or fibrosis grade.

FLA 5.5.0 DTD ■ YMPD9370_proof ■ September 14, 2017