Whole-Exome Sequencing Study of Extreme Phenotypes of NAFLD

Sarah E. Kleinstein,[1] Matthew Rein,[2] Manal F. Abdelmalek,[3] Cynthia D. Guy,[4] David B. Goldstein,[5] Anna Mae Diehl,[3] and Cynthia A. Moylan[3,6]

Nonalcoholic fatty liver disease (NAFLD) is a heterogeneous disease with highly variable outcomes. Patients with simple steatosis typically experience a benign course, whereas those with more advanced liver injury, nonalcoholic steatohepatitis (NASH), and advanced stage fibrosis suffer increased risk for complications such as cirrhosis, hepatic decompensation, and liver cancer. Genetic variants in patatin-like phospholipase domain-containing 3 (*PNPLA3*) and transmembrane 6 superfamily member 2 (*TM6SF2*) and clinical factors including diabetes, obesity, and older age increase a patient's risk for NASH, advanced fibrosis, and worse outcomes. Despite substantial investigation and identification of some common variants associated with NAFLD and advanced fibrosis, the genetics and functional mechanisms remain poorly understood. This study aimed to identify genetic variants by whole-exome sequencing of NAFLD phenotypes to provide novel insights into mechanisms behind NAFLD pathogenesis and variability. We sequenced 82 patients with liver biopsy–confirmed NAFLD and 4455 population controls. NAFLD patients were divided into extreme phenotypes based on liver fibrosis stage and clinical risk factors to investigate rare variants that might predispose to or protect from advanced NAFLD fibrosis. We compared NAFLD extremes to each other and individually to population controls, exploring genetic variation at both the single-variant and gene-based level. We replicated known associations with *PNPLA3* and *TM6SF2* and advanced fibrosis, despite sample-size limitations. We also observed enrichment of variation in distinct genes for progressor or protective NAFLD phenotypes, although these genes did not reach statistical significance. *Conclusion:* We report the first whole-exome sequencing study of genetic variation in liver biopsy–confirmed NAFLD susceptibility and severity, using a small cohort of extreme NAFLD phenotypes and a large cohort of population controls. (*hepatology communications* 2018;2:1021-1029)

NAFLD is a significant and increasing cause of morbidity and mortality worldwide, with global prevalence estimated at 25%.[1] NAFLD consists of a spectrum of histology, ranging from benign liver fat accumulation to NASH, characterized by steatosis, necroinflammation, and fibrosis.[2] NASH increases fibrosis progression risk, and advanced fibrosis predisposes to poor outcomes including decompensated cirrhosis, liver transplantation, and liver cancer.[2] Several clinical factors (diabetes mellitus,

Abbreviations: BMI, body mass index; CDKN1A, cyclin-dependent kinase inhibitor 1A; GWAS, genome-wide association studies; HSD17B13, hydroxysteroid 17-beta dehydrogenase 13; IL, interleukin; IRAK, interleukin receptor associated kinase; MPO, myeloperoxidase; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; ORM1, orosomucoid 1; PNPLA3, patatin-like phospholipase domain-containing 3; SLAM, signaling lymphocytic activation molecule; T2DM, type 2 diabetes mellitus; TM6SF2, transmembrane 6 superfamily member 2; TNF-α, tumor necrosis factor α; and WES, whole-exome sequencing.

Received March 23, 2018; accepted June 11, 2018.

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep4.1227/full.

Supported by a Duke University Chancellor's Award and an American Recovery and Reinvestment Act (ARRA) grant from the NIAAA (5RC2 AA019399 [Anna Mae Diehl, Principal Investigator]). Drs. Diehl and Abdelmalek received funding support from NIH/NIDDK (U01-DK57149). Dr. Abdelmalek was supported by a NIH/NIDDK K23 Career Development Award (K23-DK062116). The collection of control samples and data was funded in part by Biogen Inc., Gilead Sciences Inc., Bryan ADRC NIA (P30AG028377); B57 SAIC-Fredrick Inc. (M11-074); National Institute of Neurological Disorders and Stroke (RC2NS070344, U01NS077303, U01NS053998, U54NS078059); National Institute of Mental Health (RC2MH089915, K01MH098126, R01MH097971); National Institute of Allergy and Infectious Diseases Center for HIV/AIDS Vaccine Immunology (U19-AI067854); National Institute of Allergy and Infectious Diseases Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery (UM1-AI100645); Bill and Melinda Gates Foundation; the Ellison Medical Foundation New Scholar Award (AG-NS-0441-08); the Murdock Study Community Registry and Biorepository; New York Presbyterian Hospital; Columbia University College Physicians and Surgeons; and Columbia University Medical Center.

obesity, male gender, and older age) are associated with hepatic fibrosis risk.[3] However, not all NAFLD patients with risk factors develop advanced liver disease, suggesting a genetic contribution.[4,5] Genome-wide association studies (GWAS) have reproducibly identified patatin-like phospholipase domain-containing 3 (*PNPLA3*) I148M as associated with NAFLD susceptibility and severity.[6–9] Although several additional genes have been implicated, including hydroxysteroid 17-beta dehydrogenase 13 (*HSD17B13*) most recently,[10] currently the only other independently reproducible association is with transmembrane 6 superfamily member 2 (*TM6SF2*),[8,9,11] and few signals have been tracked to causal variants.

In contrast to GWAS, next-generation sequencing interrogates rare variation and can often directly pinpoint causal variants including for complex diseases.[12,13] Understanding the full spectrum of NAFLD genetic variation predisposing to or protecting from advanced fibrosis may facilitate biomarker discovery or assist with novel treatment development. We used whole-exome sequencing (WES) to examine the entire coding portion of the genome to identify potential causal variants for NAFLD fibrosis progression and protection. To accurately define different risk categories within the NAFLD spectrum, we used gold standard liver biopsy for NAFLD fibrosis staging and common clinical measurements related to NASH and advanced fibrosis. We sampled NAFLD phenotypic distribution extremes to enrich for selection of rare causal variants with potentially larger effect sizes.[14] We defined two extreme NAFLD fibrosis phenotypes: "protective" and "progressor." We hypothesized that protective patients (i.e., those without advanced fibrosis despite being high risk [older, obese, and diabetic]) might harbor genetic variants that protect them from fibrosis progression, whereas progressor patients (i.e., those with advanced fibrosis despite lacking this clinical risk profile) might carry genetic variants enhancing their fibrosis vulnerability. Here we report a comprehensive WES study that investigates genetic variation underlying NAFLD fibrosis risk and progression.

Materials and Methods

PATIENT SELECTION

We selected two cohorts of NAFLD patients from the Duke University Health System NAFLD Biorepository using an extreme phenotype design.[14] The NAFLD Biorepository, details of which have been published previously, contains specimens and clinical data from NAFLD patients who underwent diagnostic liver biopsy to grade and stage NAFLD severity as standard of care.[15] The Biorepository has the Duke Institutional Review Board's approval, and patients consented to genomic analyses.

We defined two extreme phenotypes of NAFLD: protective and progressor, based on the development of advanced liver fibrosis (fibrosis stage, F3-F4). The protective phenotype included NAFLD patients expected to have significant liver injury and fibrosis based on clinical risk factors (age >50 years, body mass index [BMI] ≥30 kg/m^2, and type 2 diabetes mellitus

View this article online at wileyonlinelibrary.com.
DOI 10.1002/hep4.1227

Potential conflicts of interest: Dr. Diehl consults for Pfizer, Novartis, Boehringer Ingelheim, Lumena, and Celgene; she received grants from Immuron, Gilead, Metabolomics, Conatus, Galmed, NGM, Madrigal, Exalenz, Shire, Intercept, and Genfit. Dr. Guy received grants from Madrigal and NGM. Dr. Moylan consults for NuSirt; she received grants from Conatus, Galmed, NGM, Taiwan J, Madrigal, Gilead, Galectin, Genfit, Intercept, and Allergan.

ARTICLE INFORMATION:

From the [1]Institute for Genomic Medicine, Columbia University, New York, NY; [2]Durham Gastroenterology Consultants, Duke University Health System, Durham, NC; [3]Division of Gastroenterology, Duke University Health System, Durham, NC; [4]Department of Pathology, Duke University Health System, Durham, NC; [5]Institute for Genomic Medicine, Hammer Health Sciences, New York, NY; [6]Department of Medicine, Durham Veterans Affairs Health System, Durham, NC.

ADDRESS CORRESPONDENCE AND REPRINT REQUESTS TO:

Cynthia A. Moylan, M.D., M.H.Sc.
Division of Gastroenterology
Duke University Health System
905 South LaSalle Street, GSRB1
Durham, NC 27710
E-mail: cynthia.moylan@duke.edu
Tel: +1-919-668-0401

[T2DM]) but with no or little fibrosis on liver biopsy (fibrosis stage, F0-F1). At the other extreme, the progressor phenotype included NAFLD patients expected to have little fibrosis based on a lack of clinical risk factors (age <55 years, BMI <35 kg/m^2, no T2DM), but biopsy showed advanced liver fibrosis or cirrhosis.

Final analyses included Caucasian individuals, the majority of both NAFLD cohorts. We compared the two extreme NAFLD phenotypes as well as each extreme phenotype to previously sequenced Caucasian population controls from unrelated Duke University studies with available consent for use.

SEQUENCING AND QUALITY CONTROL

WES was performed on 103 NAFLD patients with available stored genomic DNA. Standard protocols and Illumina HiSeq platforms were used to sequence NAFLD patients and population controls (Supporting Information). Following quality control, 82 Caucasian NAFLD samples (54 protective, 28 progressor) and 4455 Caucasian population controls were available for analysis.

DATA ANALYSIS AND ASSOCIATION TESTING

We performed downstream statistical analyses using in-house pipeline ATAV version 5.8 software (https://github.com/igm-team/atav/).[13] Three primary comparisons were conducted: (1) NAFLD progressor versus NAFLD protective; (2) NAFLD protective versus population controls; and (3) NAFLD progressor versus population controls. For each comparison, we performed single-variant (Fisher's exact test) and gene-based collapsing analyses (progressor versus protective λ = 0.50-0.99, protective versus controls λ = 1.16-2.23, progressor versus controls λ = 1.16-2.50; representative quantile–quantile plots in Supporting Figs. S1-S3). Statistical significance was based on Bonferroni-corrected thresholds for the number of variants or genes tested, respectively (Supporting Table S1).

Results

Eighty-two Caucasian NAFLD patients were included for analysis. Given the extreme phenotype design, all protective phenotype patients were age 50 or older with T2DM and obesity (median BMI = 41 kg/m^2), whereas progressors were age 55 or younger without T2DM and lower median BMI (32 kg/m^2). Most of the progressors were male and had a NAFLD activity score greater than or equal to 4 (Supporting Table S2).

NAFLD PROGRESSOR VERSUS NAFLD PROTECTIVE

No variants or genes reached genome-wide statistical significance after quality control and multiple testing correction in the progressor versus protective extreme phenotypes comparison. However, there was nonsignificant enrichment of the known *PNPLA3* I148M (rs738409, *P* = 8.42E-05) and *TM6SF2* E167K (rs58542926, *P* = 4.10E-03) polymorphisms under single-variant allelic models among the NAFLD progressors.[5–9,11] An adjacent synonymous variant in *PNPLA3*, P149 (rs738408, *P* = 8.42E-05), in perfect linkage disequilibrium (r^2 = 1) with I148M, also neared significance. *PNPLA3* I148M is in a common haplotype block and, although variants in nearby genes *PARVB* W37R and *SAMM50* G453 neared the genome-wide significance threshold, adjustment for *PNPLA3* I148M completely eliminated any association signal in this region (data not shown), consistent with previous literature.[16]

Among the top nonsignificant variants in our analysis, several were enriched in genes that differed between the NAFLD cohorts. These associations highlight genes and pathways that may either promote or protect against fibrosis progression; however, with current available evidence, their involvement remains uncertain. Nonsignificant variants enriched among progressors in biologically plausible, although not previously implicated, NAFLD genes included several immune-related findings (Table 1 and Supporting Tables S3 and S4). Common variants with nonsignificant enrichment under allelic models included S31R (rs1801270, *P* = 1.30E-03) in cyclin-dependent kinase inhibitor 1A (*CDKN1A*) and L439V (rs11465927, *P* = 6.70E-03) in interleukin 1 (IL1) receptor associated kinase 2 (*IRAK2)*. *CDKN1A* encodes p21, a senescence marker involved in innate immunity, whose hepatocyte expression has been associated with NAFLD fibrosis stage.[17] A previous candidate gene study linked several *CDKN1A* variants with NAFLD fibrosis development, including S31R, although S31R was not associated with fibrosis and only borderline associated with steatohepatitis.[17] IRAK2 is part of the innate immune response, acting in IL1R-mediated and IL1-mediated signaling. Under recessive models,

TABLE 1. NAFLD PROGRESSOR (N = 28) VERSUS PROTECTIVE (N = 54) COMPARISON: TOP ASSOCIATED VARIANTS IN BIOLOGICALLY RELEVANT GENES

Gene	Variant	rs#	NAFLD Progressors With Variant, N (MAF)	NAFLD Protective With Variant, N (MAF)	PolyPhen Prediction	ExAC Global MAF	Genetic Model	*P* Value	Enrichment
PNPLA3	I148M	rs738409	24 (0.61)	24 (0.28)	Probably damaging	0.26	SV allelic	8.42E-05	Progressor
PNPLA3	P149	rs738408	24 (0.61)	24 (0.28)	NA	0.26	SV allelic	8.42E-05	Progressor
SAMM50	G453	rs7587	1 (0.04)	26 (0.27)	NA	0.22	SV allelic	4.23E-04	Protective
PTX4	G36C	rs1040499	11 (0.20)	39 (0.47)	Unknown	0.41	SV allelic	6.26E-04	Protective
PTX4	R276K	rs2745098	12 (0.21)	37 (0.46)	Benign	0.41	SV allelic	2.10E-03	Protective
CDKN1A	S31R	rs1801270	9 (0.18)	3 (0.03)	Benign	0.15	SV allelic	1.30E-03	Progressor
GBP1	A409G	rs1048443	9 (0.16)	36 (0.41)	Probably/possibly damaging	0.30	SV allelic	1.40E-03	Protective
TM6SF2	E167K	rs58542926	5 (0.09)	0 (0)	Probably/possibly damaging	0.02	SV allelic	4.10E-03	Progressor
IRAK2	L439V	rs11465927	5 (0.11)	1 (0.01)	Benign	0.03	SV allelic	6.70E-03	Progressor
PARVB	W37R	rs1007863	39 (0.44)	39 (0.44)	Benign	0.44	SV recessive	2.28E-04	Progressor
VRK2	I167V	rs1051061	21 (0.57)	36 (0.37)	Probably damaging	0.36	SV recessive	7.63E-04	Progressor
SEC31B	L1071V	NA	1	0	Benign	0	GB dominant	0.006	Progressor
	V504M	rs41290542	6	3	Benign	0.03			Progressor
	G86R	NA	1	0	Benign	0			Progressor

Abbreviations: ExAC, exome aggregation consortium; GB, gene-based; MAF, minor allele frequency; NA, not available; SV, single-variant.

I167V (rs1051061, *P* = 7.63E-04) was enriched in vaccinia-related kinase 2 (*VRK2*), an IL1-mediated signaling effector.

Among the protective phenotype, in which variants might reduce fibrosis progression risk, several immune genes were among the top results, although not in the IL1 pathway. We saw nonsignificant enrichment in allelic models of G36C (rs1040499, *P* = 6.26E-04) and R276K (rs2745098, *P* = 2.10E-03) in long pentraxin 4 (*PTX4*), which are in high linkage disequilibrium (r^2 = 0.93) as well as A409G (rs1048443, *P* = 1.40E-03) in interferon (IFN)-inducible guanylate binding protein 1 (*GBP1*). PTX4 is a potential functional antibody ancestor that acts in innate immunity, whereas GBP1 is a guanosine triphosphatase that regulates IL2 secretion and acts in cytokine, IFN-γ, and T-cell receptor–mediated signaling pathways.

NAFLD PROTECTIVE VERSUS POPULATION CONTROLS

The protective versus population control comparison investigated susceptibility to this NAFLD phenotype, as well as possible protective variants against advanced fibrosis. We did not discover any significant, high-quality variants. Several nonsignificant rare variants in genes involved in immune-, liver-, lipid-, or fibrosis-related pathways were identified, although their NAFLD role remains unconfirmed (Table 2 and Supporting Tables S5 and S6).

Three extremely rare variants in immune-related genes were nonsignficantly enriched among the NAFLD protective phenotype relative to population controls under single-variant allelic models: I13V (*P* = 1.42E-04) in SLAM (signaling lymphocytic activation molecule) family member 7 (*SLAMF7*), C161 (*P* = 1.42E-04) in IL32, and T151 (*P* = 4.24E-04) in orosomucoid 1 (*ORM1*). Under a dominant gene-based model, D162V (rs2069860, *P* = 0.004) in IL6 was also enriched. IL6 influences inflammation-associated disease states, including metabolic syndrome diseases such as diabetes. IL6 also upregulates IL32, which induces tumor necrosis factor-α (TNF-α) macrophage production and acts in oxidative damage response. Interestingly, IL32 can attenuate alcohol-induced liver injury, as well as lipid accumulation in mice on a high-fat diet.[18,19] Also within the IL6 pathway, ORM1 is an acute phase plasma reactant protein involved in immunosuppression, including negative regulation of IL6 and TNF-α, with a potential role in alcoholic liver cirrhosis.[20] SLAMF7 activates natural killer cells, inhibits pro-inflammatory

TABLE 2. NAFLD PROTECTIVE (N = 54) VERSUS POPULATION CONTROLS (N = 4455) COMPARISON: TOP ASSOCIATED VARIANTS IN BIOLOGICALLY RELEVANT GENES

Gene	Variant	rs#	NAFLD Protective With Variant, N (MAF)	Controls With Variant, N (MAF)	PolyPhen Prediction	ExAC Global MAF	Genetic Model	*P* Value
OIT3	Y60	NA	2 (0.02)	0 (0)	NA	0	SV allelic	1.42E-04
ABCA8	P1396L	rs148226092	2 (0.02)	0 (0)	Probably damaging	8.68E-05	SV allelic	1.42E-04
SLAMF7	I13V	NA	2 (0.02)	0 (0)	Benign	7.15E-05	SV allelic	1.42E-04
PINK1	L288	NA	2 (0.02)	0 (0)	NA	3.95E-05	SV allelic	1.42E-04
PINK1	P289T	NA	2 (0.02)	0 (0)	Probably damaging	3.95E-05	SV allelic	1.42E-04
IL32	C161	NA	2 (0.02)	0 (0)	NA	1.61E-05	SV allelic	1.42E-04
SMEK2	T723A/ T808A	rs76512669	4 (0.04)	20 (0.002)	Benign	0.001	SV allelic	1.72E-04
ORM1	T151	NA	2 (0.02)	1 (0.0001)	NA	5.54E-05	SV allelic	4.24E-04
HNF1A	P379A	NA	2 (0.02)	1 (0.0001)	Probably damaging	2.10E-04	SV allelic	4.30E-04
PKD2L1	L138*	NA	1 (0.02)	2 (0.0002)	NA	1.90E-04	SV allelic	8.39E-04
IL6	D162V	rs2069860	5	75	Benign	0.006	GB dominant	0.004
THEM5	P246L	NA	1	0	Benign	8.00E-06	GB dominant	0.006
	E168K	NA	1	0	Probably/possibly damaging	2.00E-04		
CYP26B1	V456L	NA	1	0	Benign	0	GB recessive	0.012

Abbreviations: ExAC, exome aggregation consortium; GB, gene-based; MAF, minor allele frequency; NA, not available; SV, single-variant.

cytokines including TNF-α, and has been associated with several autoimmune diseases including diabetes.[21]

NAFLD PROGRESSOR VERSUS POPULATION CONTROLS

In the progressor versus population control comparison, *PNPLA3* variants I148M and P149 reached statistical significance (both P = 2.10E-09 allelic) despite small progressor sample size. Although there were no other robust, significant associations, we observed nonsignificant enrichment in several NAFLD-associated genes: *TM6SF2* E167K (P = 8.88E-04 allelic), *PARVB*, and *SAMM50*. The enrichment of *PNPLA3* I148M and *TM6SF2* E167K in both the progressor versus protective and progressor versus control comparisons, but not in the protective versus control comparison, provides further evidence that these variants are important for NAFLD fibrosis progression independent of T2DM and obesity.

Among the top nonsignificant gene-based associations, several genes were associated with metabolic syndrome and lipids (Table 3 and Supporting Tables S7 and S8). Under a dominant gene-based model, we observed nonsignificant enrichment of rare variation in the major histocompatibility complex I immune response molecule alpha-2-glycoprotein 1 zinc-binding (*AZGP1*; P = 0.007, H214Q and A46V), which regulates fatty acid synthesis and cell adhesion. AZGP1 is implicated in metabolic syndrome and insulin sensitivity, is elevated in kidney injury, and is a lipid catabolism biomarker. Additionally, under recessive single-variant and gene-based models, a single nonsynonymous variant, I717V (rs2759, P = 7.76E-04 single-variant; P = 7.75E-04 gene-based) was enriched in myeloperoxidase (*MPO*), a major component of neutrophil granules that acts in low-density lipoprotein remodeling.[22]

TABLE 3. NAFLD PROGRESSOR (N = 28) VERSUS POPULATION CONTROLS (N = 4455) COMPARISON: TOP ASSOCIATED VARIANTS IN BIOLOGICALLY RELEVANT GENES

Gene	Variant	rs#	NAFLD Progressors With Variant, N (MAF)	Controls With Variant, N (MAF)	PolyPhen Prediction	ExAC Global MAF	Genetic Model	*P* Value
PNPLA3	*P149*	*rs738408*	*24 (0.61)*	*1826 (0.23)*	*NA*	*0.26*	*SV allelic (and recessive)*	***2.09E-09***
PNPLA3	*I148M*	*rs738409*	*24 (0.61)*	*1826 (0.23)*	*Probably damaging*	*0.26*	*SV allelic (and recessive)*	***2.10E-09***
PARVB	W37R	rs1007863	23 (0.70)	2691 (0.40)	Benign	0.44	SV allelic (and recessive)	8.20E-06
SAMM50	D110G	rs3761472	17 (0.38)	1329 (0.16)	Benign	0.21	SV allelic	1.45E-04
TM6SF2	E167K	rs58542926	9 (0.20)	549 (0.06)	Probably/possibly damaging	0.07	SV allelic	8.88E-04
MPO	I717V	rs2759	4 (0.11)	237 (0.03)	Benign	0.02	SV recessive	7.76E-04
HIST1H2BC	FS (chr6: 26124019 insT)	NA	1	1	NA	0	GB dominant	0.002
	A22S	NA	1	1	Unknown	2.00E-05		
AZGP1	H214Q	NA	1	1	Benign	2.00E-05	GB dominant	0.007
	A46V	rs142669146	1	0	Benign	1.00E-04		
MRGPRX1	Q307R	rs138752944	1	0	Benign	3.00E-04	GB recessive	7.75E-04
	Q307*	rs140371088	1	0	NA	3.00E-04		
	Y272C	NA	1	0	Probably damaging	1.00E-04		
CYP26B1	A420G and R191H	rs7568553 and rs76025186	1	0	Benign and probably damaging	0.005 and 0.001	GB compound-heterozygous	0.006
EFCAB13	K244* and T577R/T481R	NA and rs142664574	1	0	NA and possibly damaging/benign	2.00E-05 and 0.005	GB compound-heterozygous	0.007

Note: Italicized and bolded variants reached statistical significance.
Abbreviations: ExAC, exome aggregation consortium; GB, gene-based; MAF, minor allele frequency; NA, not available; SV, single-variant

MPO has been linked to various metabolic syndrome diseases and is involved in stimuli responses including to food and lipopolysaccharides.[22] Further, a *MPO* promoter polymorphism (–463G > A) was previously implicated in fibrosis severity in women with hepatitis C,[23] consistent with its enrichment among advanced fibrosis observed here. *MPO* was also enriched among progressors in the gene-based recessive model for the progressor versus protective analysis, potentially implying a role in advanced fibrosis, if confirmed.

Discussion

A major goal in caring for patients with NAFLD is determining who is at risk for fibrosis and thus poor outcomes. We aimed to determine the contribution of genetic variants to this risk by using WES in patients at NAFLD extremes. We confirmed the significance of *PNPLA3* I148M in NAFLD fibrosis progression as identified in previous studies.[8,9] Importantly, we were able to do this in a small NAFLD cohort through accurate phenotyping of biopsy-proven NAFLD patients. Our results support a role for *PNPLA3* directly in NAFLD fibrosis severity, as previously implicated,[7] and the recent suggestion that *PNPLA3* potentiates the pro-fibrogenic features of hepatic stellate cells.[24] Similarly, the enrichment of *TM6SF2* E167K observed among progressors is consistent with previous associations between E167K and hepatic fibrosis progression in NAFLD,[11] although it did not reach significance in this study.

Although only *PNPLA3* reached statistical significance, and only in the larger progressor versus control comparison, we observed several suggestive associations in biologically relevant genes broadly in line with a role for a pro-inflammatory state in NAFLD development. These findings are important for future hypothesis-driven research but require replication in independent NAFLD cohorts, as immune genes compose a substantial fraction of the human genome, enabling considerable narrative potential. Among the top associations in the NAFLD progressor versus protective comparison, several affected the immune genes, with distinct genes and biological processes observed for each phenotype. Among progressors, two variants were enriched in IL1 signaling pathway genes (*IRAK2* and *VRK2*), which could conceivably reflect disruption of this innate immune and tissue regeneration pathway in advanced fibrosis progression.[25]

In contrast, several variants enriched among high-risk, low-fibrotic NAFLD patients were observed in IL6-related genes (*IL6*, *IL32*, *ORM1,* and *SLAMF7*) in the NAFLD susceptibility investigation of protective versus population controls, suggesting a potential role for a pro-inflammatory immune response. This finding is supported by recently published NASH Clinical Research Network data implicating pro-inflammatory pathways, including common variants in *IL1B* and *IL6*, with NAFLD fibrosis risk and ballooning.[25] However, as protective individuals were obese with T2DM as part of the study design, the IL6-related genes may also reflect T2DM risk.

Our findings are important for several reasons. First, we used gold standard liver biopsy–confirmed NAFLD to ensure accurate histologic phenotyping of patients. Second, we report the first WES investigation of genetic variation in NAFLD fibrosis. Third, the extreme phenotype design assisted in overcoming difficulties of accurate phenotyping in a highly heterogeneous disease such as NAFLD. This has been a major challenge for GWAS, as variable clinical factors contribute to diverse molecular pathogenesis even with similar histologic phenotypes. We selected extreme phenotypes to control clinical factors associated with risk or protection from advanced fibrosis including obesity, age, and T2DM. Due to the feasibility of having sufficient specimen availability meeting criteria, the histologic extremes could not be completely separated. This may have reduced our ability to detect fibrosis-associated variants specific to NASH and represents a study limitation. Additionally, as most patients had stages 0, 1, and 3 fibrosis, it was not possible to identify extremes with just isolated steatosis or cirrhosis. Although we included population controls from studies unrelated to viral hepatitis or liver disease to increase power and generalizability, detailed phenotyping information about NAFLD and risk factors, such as obesity and diabetes, was unavailable. As NAFLD is relatively common in the general population,[1] another limitation is the potential for misclassification among controls, leaving our results conservative. Indeed, we did not observe any association with *HSD17B13,* likely due to a combination of sample-size constraints and differences in investigated phenotypes, as our study focused on histological fibrosis, whereas *HSD17B13* initially associated with serum alanine and aspartate aminotransferases.[10] Finally, as individuals with the protective phenotype were older, obese and diabetic, the protective phenotype comparison with population

controls may have reflected an underlying genetic risk for NAFLD or diabetes susceptibility, whereas the progressor versus control comparison uniquely interrogated risks for advanced fibrosis among NAFLD patients without diabetes.

Although many of our results were not significant, we hope that this investigation will pave the way for further next-generation-sequencing studies. We have highlighted several biologically plausible genes among our top associations, although we currently lack the evidence necessary to determine their NAFLD fibrosis involvement. However, if confirmed in future studies, they may warrant further investigation in the search for pathogenic and therapeutic targets in NASH and liver fibrosis. This study suggests that "extreme" NAFLD phenotypes may represent distinct disease subtypes, perhaps accounting for the nonlinear nature of fibrosis progression. Improved delineation of these subtypes and genetic risks will require large, well-phenotyped follow-up studies. Ideally, the goal is to develop personalized variant profiles for NAFLD based on the risk of progression to fibrosis and fatal liver outcomes, enabling not only personalized treatment options for NAFLD but early identification and prevention of this extremely common disease.

Acknowledgments: We thank Dr. Patrick R. Shea, Dr. Julia Wattacheril, Dr. Xiaolin Zhu, and Dr. K. Melodi McSweeney for their comments on the drafts of this manuscript and Dr. Thomas J. Urban for his contributions to the project study concept and design. We also acknowledge the following individuals or groups for the contributions of control samples: K. Welsh-Bomer, C. Hulette, J. Burke, F. McMahon, N. Akula, D. Valle, J. Hoover-Fong, N. Sobriera, A. Poduri, R. Buckley, N. Calakos, D. Murdock, R. Ottman, J. McEvoy, C. Depondt, S. Sisodiya, G. Cavalleri, N. Delanty, S. Schuman, E. Nading, W. Lowe, D. Marchuk, D. Levy, M. Carrington, C. Woods, Croasdaile Village, K. Schmader, S. McDonald, M. Yanamadala, H. White, P. Lugar, G. Nestadt, S. Palmer, E. Pras, D. Lancet, Z. Farfel, Y. Jiang, T. Young, K. Whisenhunt, D. Daskalakis, M. Winn, R. Gbadegesin, M. Hauser, A. Holden, E. Behr, M. Connors, L. Morris, V. Shashi, K. Schoch, M. Harms, T. Miller, A. Pestronk, R. Bedlack, E. Cirulli, K. Grace, R. Brown, N. Shneider, S. Gibson, J. Ravits, A. Gitler, J. Glass, F. Baas, S. Appel, E. Simpson, G. Rouleau, the ALS Sequencing Consortium, the National Institute of Allergy and Infectious Diseases Center for HIV/AIDS Vaccine Immunology, the National Institute of Allergy and Infectious Diseases Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery, and the Epi4K Consortium and Epilepsy Phenome/Genome Project.

REFERENCES

1) Younossi ZM, Blissett D, Blissett R, Henry L, Stepanova M, Younossi Y, et al. The economic and clinical burden of non-alcoholic fatty liver disease (NAFLD) in the United States and Europe. Hepatol Baltim Md 2016;64:1577-1586.
2) Angulo P, Kleiner DE, Dam-Larsen S, Adams LA, Bjornsson ES, Charatcharoenwitthaya P, et al. Liver fibrosis, but no other histologic features, is associated with long-term outcomes of patients with nonalcoholic fatty liver disease. Gastroenterology 2015;149:389-397,e10.
3) Hossain N, Afendy A, Stepanova M, Nader F, Srishord M, Rafiq N, et al. Independent predictors of fibrosis in patients with nonalcoholic fatty liver disease. Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc 2009;7:1224-1229.e2.
4) Loomba R, Schork N, Chen C-H, Bettencourt R, Bhatt A, Ang B, et al. Heritability of hepatic fibrosis and steatosis based on a prospective twin study. Gastroenterology 2015;149:1784-1793.
5) **Speliotes EK, Yerges-Armstrong LM, Wu J, Hernaez R,** Kim LJ, Palmer CD, et al. Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. PLoS Genet 2011;7:e1001324.
6) 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;40:1461-1465.
7) Valenti L, Al-Serri A, Daly AK, Galmozzi E, Rametta R, Dongiovanni P, et al. Homozygosity for the patatin-like phospholipase-3/adiponutrin I148M polymorphism influences liver fibrosis in patients with nonalcoholic fatty liver disease. Hepatol Baltim Md 2010;51:1209-1217.
8) Anstee QM, Seth D, Day CP. Genetic factors that affect risk of alcoholic and nonalcoholic fatty liver disease. Gastroenterology 2016;150:1728-1744,e7.
9) Severson TJ, Besur S, Bonkovsky HL. Genetic factors that affect nonalcoholic fatty liver disease: a systematic clinical review. World J Gastroenterol 2016;22:6742-6756.
10) Abul-Husn NS, Cheng X, Li AH, Xin Y, Schurmann C, Stevis P, et al. A protein-truncating *HSD17B13* variant and protection from chronic liver disease. N Engl J Med 2018;378:1096-1106.
11) Liu Y-L, 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 [Internet] 2014 [cited 2016 Aug 19];5. Available from: https://www.nature.com/doifinder/10.1038/ncomms5309.
12) Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet 2010;11:415-425.
13) **Cirulli ET, Lasseigne BN,** Petrovski S, Sapp PC, Dion PA, Leblond CS, et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. Science 2015;347:1436-1441.
14) Barnett IJ, Lee S, Lin X. Detecting rare variant effects using extreme phenotype sampling in sequencing association studies. Genet Epidemiol 2013;37:142-151.
15) Moylan CA, Pang H, Dellinger A, Suzuki A, Garrett ME, Guy CD, et al. Hepatic gene expression profiles differentiate

presymptomatic patients with mild versus severe nonalcoholic fatty liver disease. Hepatol Baltim Md 2014;59:471-482.
16) 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;132:783-792.
17) **Aravinthan A, Mells G,** Allison M, Leathart J, Kotronen A, Yki-Jarvinen H, et al. Gene polymorphisms of cellular senescence marker p21 and disease progression in non-alcohol-related fatty liver disease. Cell Cycle Georget Tex 2014;13:1489-1494.
18) Lee DH, Kim DH, Hwang CJ, Song S, Han SB, Kim Y, et al. Interleukin-32γ attenuates ethanol-induced liver injury by the inhibition of cytochrome P450 2E1 expression and inflammatory responses. Clin Sci Lond Engl. 1979;2015:695-706.
19) Lee DH, Hong JE, Yun H-M, Hwang CJ, Park JH, Han SB, et al. Interleukin-32β ameliorates metabolic disorder and liver damage in mice fed high-fat diet. Obes Silver Spring Md 2015;23:615-622.
20) Mandal G, Yagi H, Kato K, Chatterjee BP. Structural heterogeneity of glycoform of alpha-1 Acid glycoprotein in alcoholic cirrhosis patients. Adv Exp Med Biol. 2015;842:389-401.
21) DIAbetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium, Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium, South Asian Type 2 Diabetes (SAT2D) Consortium, Mexican American Type 2 Diabetes (MAT2D) Consortium, Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples (T2D-GENES) Consortium, Mahajan A, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet 2014;46:234-244.
22) Podrez EA, Febbraio M, Sheibani N, Schmitt D, Silverstein RL, Hajjar DP, et al. Macrophage scavenger receptor CD36 is the major receptor for LDL modified by monocyte-generated reactive nitrogen species. J Clin Invest 2000;105:1095-1108.
23) do Carmo RF, Vasconcelos LR, Mendonça TF, deMendonça Cavalcanti Mdo S, Pereira LM, et al. Myeloperoxidase gene polymorphism predicts fibrosis severity in women with hepatitis C. Hum Immunol 2014;75:766-770.
24) Bruschi FV, Claudel T, Tardelli M, Caligiuri A, Stulnig TM, Marra F, et al. The PNPLA3 I148M variant modulates the fibrogenic phenotype of human hepatic stellate cells. Hepatol Baltim Md 2017;65:1875-1890.
25) Nelson JE, Handa P, Aouizerat B, Wilson L, Vemulakonda LA, Yeh MM et al. Increased parenchymal damage and steatohepatitis in Caucasian non-alcoholic fatty liver disease patients with common IL1B and IL6 polymorphisms. Aliment Pharmacol Ther 2016;44:1253-1264.

Author names highlighted in bold designate co-first authorship.

Supporting Information

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep4.1227/full.