UNIVERSITY OF CALIFORNIA SAN DIEGO

Socioeconomic conditions, and not genetics alone, predict progression of nonalcoholic fatty liver disease

A thesis submitted in partial satisfaction of the requirements for the degree Master of Arts

in

Anthropology

by

Maria Christine Rieman-Klingler

Committee in charge:

Professor Amy Non, Chair
Professor Pascal Gagneux
Professor Bronwyn Kaiser

2022

The Thesis of Maria Christine Rieman-Klingler is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my advisor and professor, Dr. Amy L. Non, for her continued support, guidance, and mentorship over the course of this research project and throughout my time in graduate school. I would also like to express my appreciation of, and gratitude for, my professors and fellow Committee members, Drs. Pascal Gagneux and Bonnie Kaiser, who have broadened my knowledge and understanding of the world through both their teachings and characters.

I would also like to thank and acknowledge the lab of Rohit Loomba at the University of California, San Diego NAFLD Research Center for providing me with the opportunity to undertake this project, as well as each patient who graciously volunteered to be part of this research.

ABSTRACT OF THE THESIS

Socioeconomic conditions, and not genetics alone,
predict progression of nonalcoholic fatty liver disease

by

Maria Christine Rieman-Klingler

Master of Arts in Anthropology

University of California San Diego, 2022

Professor Amy Non, Chair

Nonalcoholic fatty liver disease (NAFLD) is the leading cause of chronic liver disease globally, with an estimated prevalence exceeding 25%. Variants in the *PNPLA3* and *HSD17B13* genes have been a focus of investigations surrounding the etiology of NAFLD and are believed to contribute to the greater burden of disease experienced by Hispanic Americans. However, little is known about socioeconomic factors influencing NAFLD progression or its increased prevalence among Hispanics. We cross-sectionally analyzed 264 patients to assess the role of

genetic and socioeconomic variables in the development of advanced liver fibrosis in individuals at risk for NAFLD. Adjusting for age, sex, body mass index (BMI), and *PNPLA3* genotype, lacking a college degree was associated with 3.3 times higher odds of advanced fibrosis (95% CI: 1.21-8.76, p=0.019), an effect comparable to that of possessing the *PNPLA3* risk variant. Notably, the effect of *PNPLA3* genotype on advanced fibrosis was attenuated to non-significance following adjustment for education and other socioeconomic markers. The effect of the protective *HSD17B13* variant, moreover, diminished after adjustment for education (OR: 0.39 [95% CI: 0.13-1.16, p=0.092]), while lower education continued to predict advanced fibrosis following multivariable adjustment with an odds ratio of 8.0 (95% CI: 1.91-33.86, p=0.005). Adjusting for education attenuated the effect of both genotype and Hispanic ethnicity on liver fibrosis, suggesting that social factors—rather than genes or ethnicity alone—may be driving disease severity within this population. This study reveals the importance of including socioeconomic variables when considering the role of genetics or ethnicity in complex disease.

INTRODUCTION

Nonalcoholic fatty liver disease (NAFLD) is the most common cause of chronic liver disease globally, with an estimated prevalence of over 25% (Younossi et al., 2016). Characterized by the accumulation of excessive hepatic fat in the absence of significant alcohol consumption, NAFLD is recognized as an umbrella category encompassing a spectrum of diseases that range in severity from simple steatosis, also known as nonalcoholic fatty liver (NAFL), to nonalcoholic steatohepatitis (NASH). While both NAFL and NASH involve the initial deposition of surplus fat within liver hepatocytes, NAFL may be a relatively benign condition, whereas NASH is a more serious and aggressive form of disease that involves liver inflammation and hepatocellular injury subsequent to the process of fatty change. This inflammation and injury can cause fibrosis, or scarring and stiffening, of the liver, which may eventually progress to cirrhosis (advanced scarring), liver failure, and hepatocellular carcinoma (Brunt et al., 2015; Friedman et al., 2018; Huang et al., 2021; Rinella, 2015). While once a relatively obscure and unrecognized disorder (Ludwig et al., 1980; Pais et al., 2016), NAFLD has increased dramatically in prevalence over the past three decades (Ge et al., 2020; Pais et al., 2016; Younossi et al., 2018, 2019) and now constitutes the fastest growing indication for liver transplantation in Western nations (Burra et al., 2020). With predictive models forecasting an exponential increase in the mortality and burden of disease caused by NAFLD (Estes et al., 2018), elucidating the precise causes of the disorder has become a major effort in biomedical research (Farrell & Larter, 2006; Sanyal et al., 2016; Zhu et al., 2020) and an endeavor with particular relevance to global public health.

Although the precise etiology of NAFLD is unknown, the disease is highly associated with features of metabolic syndrome, such as obesity, insulin resistance and type 2 diabetes,

elevated serum triglyceride levels, and low serum levels of HDL cholesterol (Angulo, 2002). Beyond these metabolic abnormalities, risk factors for the development and progression of NAFLD include belonging to particular racial and ethnic groups. Within the United States, recent estimates have placed the overall prevalence of NAFLD above 30% (Ciardullo & Perseghin, 2021; Zou et al., 2020), but significant racial and ethnic disparities—particularly characterized by a disproportionately high rate of disease among Hispanics—are present. In a systematic review and meta-analysis of literature published through 2016, for example, Rich and colleagues (2018) analyzed 34 studies characterizing racial/ethnic disparities in NAFLD prevalence, severity, or prognosis in the United States, and found pooled disease prevalence to be highest in Hispanic individuals (23%), intermediate in non-Hispanic Whites (14%), and lowest in non-Hispanic Blacks (13%). Among patients with NAFLD, moreover, the study found a higher pooled relative risk of NASH in Hispanic Americans than in either non-Hispanic White or non-Hispanic Black Americans.[*] These general race-based trends are supported by the findings of many studies (Lim et al., 2019; Patel et al., 2020; Zou et al., 2020), and both NAFLD and NASH are thus well-established to exhibit heterogeneous distributions across racial and ethnic groups in the United States, with Hispanics bearing a greater burden of disease.

The strong relationship between NAFLD and metabolic conditions, particularly those of obesity and insulin resistance, has resulted in a general recognition of the disease as the hepatic manifestation of metabolic syndrome (Katsiki et al., 2018; Kim & Younossi, 2008). However, some incongruities in these metabolic relationships render a simple explanation of metabolic

[*] Of note, Rich et al. (2018) report that between-race differences in NAFLD prevalence were greater in population-based cohorts (ranging from 13% to 23%) than in high-risk cohorts (ranging from 47.6% to 55.5%) composed primarily of patients with obesity or other metabolic risk factors. While among high-risk patients, for example, Blacks maintained a lower risk of NAFLD than Whites, the risk of NAFLD in high-risk Hispanics was not significantly different from that in high-risk Whites.

difference unlikely to fully account for the observed racial disparities in disease outcomes. For example, although roughly 95% of individuals with NAFLD in the United States are overweight or obese (Zou et al., 2020), this proportion is reduced to roughly 80% worldwide (Ye et al., 2020). Thus, while some evidence has suggested that elevated body mass index (BMI) may be independently associated with NAFLD and increase the risk of fatty liver in a dose-dependent fashion (Fan et al., 2018), a subset of NAFLD does occur in individuals of normal weight, suggesting that factors beyond obesity are involved in its development. In the United States, the higher prevalence of obesity in Hispanic relative to White Americans—with recent estimates of 45.6% and 41.4%, respectively—is itself insufficient to explain racial disparities in NAFLD prevalence, as Black Americans have a higher rate of obesity than both Hispanics and Whites, at 49.4% (Stierman et al., 2021), yet exhibit the lowest prevalence of NAFLD of the three racial subgroups. NAFLD and hepatic fibrosis are additionally highly associated with insulin resistance (Angulo, 2002; Angulo et al., 2004; Bugianesi et al., 2005; Jinjuvadia et al., 2017; Marchesini et al., 1999, 2001; Marušić et al., 2021), and although Hispanic Americans have a higher prevalence of type 2 diabetes than Whites (22.2% vs. 12.0%), this proportion is not significantly different from that of other major racial subgroups, such as Asian (18.1%) or Black (18.8%) Americans (Stierman et al., 2021). Paradoxically, Black Americans experience a drastically lower degree of NAFLD in comparison to Hispanics, despite having a similar prevalence of type

2 diabetes and a higher average BMI.[†] These discrepancies suggest that disparate rates of individual metabolic conditions alone are unlikely to explain the racial and ethnic disparities observed in the distribution and progression of NAFLD.

The complexity of these associations, in conjunction with the characteristically high frequency of NAFLD within Hispanic populations, has garnered the attention of genetic researchers. Gene variants in multiple chromosomal regions are believed to modify an individual's susceptibility to NAFLD, the most well-characterized of which is a non-synonymous variant (rs738409 C>G p.I148M) in the patatin-like phospholipase domain-containing protein 3 (*PNPLA3)* gene, which codes for adiponutrin, a protein found in fat and liver cells (Trépo et al., 2016). The risk variant at this locus, characterized by the substitution of cytosine (C) with guanine (G), is robustly associated with NAFLD across different geographical and ethnic contexts and predicts a higher risk of hepatic steatosis (Romeo et al., 2008), NASH (Sookoian & Pirola, 2011; Zain et al., 2012), advanced fibrosis (Ajmera et al., 2021), cirrhosis (Emdin et al., 2021), and hepatocellular carcinoma (Huang et al., 2021; Liu et al., 2014; Trépo et al., 2014). Notably, this G risk variant is more prevalent in Hispanic than in Black or White Americans, with allele frequencies previously reported as 0.49, 0.17, and 0.23, respectively (Romeo et al., 2008).

[†] This apparent contradiction may be partially resolved in light of the significantly lower frequencies of high triglyceride and low HDL cholesterol levels observed in Black relative to White Americans (Park et al., 2003; Yu et al., 2012) or Mexican Americans (Park et al., 2003), as these are two features of metabolic syndrome highly associated with NAFLD (Cortez-Pinto et al., 1999; Fukuda et al., 2016). Some data have actually suggested that serum triglyceride level alone may be a stronger predictor of fatty liver than obesity, as well as an independent commonality among patients with NAFLD irrespective of BMI (Bonacini et al., 2021; Fan et al., 2019; Wagenknecht et al., 2009). NAFLD also appears to relate more strongly to visceral adipose tissue than to overall obesity or BMI, and Blacks display less visceral adiposity, on average, than both Hispanics and Whites, despite having higher BMI (Carroll et al., 2008; Wagenknecht et al., 2009). Still, even in subgroup analyses restricted to individuals with insulin sensitivity or with low visceral adiposity, Wagenknecht et al. (2009) have shown that the lower prevalence of hepatic steatosis in Blacks compared to Hispanics remains persistent.

In contrast, an additional variant of particular interest in NAFLD is a protective protein-truncating splice variant (rs72613567:TA) in *HSD17B13 (Abul-Husn et al., 2018)*, a gene encoding a hepatic lipid droplet protein, 17β-hydroxysteroid dehydrogenase-13, involved in lipogenesis (Su et al., 2014). Individuals with this *HSD17B13* T to TA insertion variant have reduced serum levels of alanine aminotransferase (ALT) and aspartate aminotransferase (AST)—two markers of hepatocellular injury—and a decreased risk of NASH, hepatic fibrosis, and other forms of fatty liver disease (Abul-Husn et al., 2018; Wang et al., 2020). This protective TA allele, furthermore, is less common in Hispanic than in non-Hispanic individuals, with minor allele frequencies previously reported as 0.09 and 0.23, respectively (Ajmera et al., 2021).

Given the higher prevalence and progression of NAFLD in Hispanics, and the relative frequencies of the *PNPLA3* G risk variant and *HSD17B13* TA protective variant among Hispanics, scientists have speculated that the *PNPLA3* and *HSD17B13* genes may causally contribute to both the development of NAFLD as well as to its increased incidence and severity among Hispanics (Krawczyk et al., 2020). While there has been explosive interest in determining the genetic and metabolic underpinnings of the disease, however, few studies have approached these disparities with a social, rather than a molecular, lens. Fibrosis stage, which describes the presence and extent of extracellular protein buildup and hepatic scarring in liver disease, is the only histologic marker consistently associated with long-term outcomes and mortality in patients with NAFLD (Angulo et al., 2015; Ekstedt et al., 2015). Yet, at the time of writing, remarkably few studies could be found that analyze the risk of advanced fibrosis according to axes of social inequality, such as income, education level, occupation, or other social-environmental characteristics. Available analyses of social factors have primarily focused on food and nutritional access, and have revealed an important association between food insecurity and the

presence of advanced fibrosis among low-income, U.S. adults (Golovaty et al., 2020). While some research has addressed other environmental and lifestyle risk factors for advanced fibrosis, these are typically limited to those involving individual choices and behavior, such as patterns of diet and exercise (Heredia et al., 2022; Kistler et al., 2011). Although understanding these factors is highly relevant, understanding only these factors may foster a belief that personal choices, irrespective of systemic and structural issues, are alone responsible for between-group differences in health outcomes.

In addition, studies have reported the prevalence of NAFLD and liver disease outcomes according to racial and ethnic group, but without determining the driving factors contributing to these disparities. Investigations of NAFLD that attempt to assess effects of ethnicity have typically controlled, if at all, for biochemical markers or metabolic conditions alone, thus omitting potential mediatory environmental or socioeconomic variables. Evidence from multiple studies has revealed an association between NAFLD and Hispanic ethnicity, but much of this research has consequently attributed the association a priori to genetic or molecular causes, without adjusting for social or environmental circumstances that could confound the relationship. Since ethnic groups may share particular genetic variants, but are likely to share particular social and environmental circumstances as well, additional research is needed that incorporates genetic, racial/ethnic, and social variables in order to disentangle the relationships between predictive genetic markers of disease and disease outcomes themselves. To our knowledge, no studies of advanced fibrosis in NAFLD have simultaneously controlled for genetic, social, and racial/ethnic variables.

The purpose of the current study is to explore the relationship between social variables, genetics, and disease progression in NAFLD. Examining a well-described cohort of patients at

risk for NAFLD, we comparatively assessed the value of social-environmental markers and genetic variants in *PNPLA3* and *HSD17B13* in predicting the development of advanced liver fibrosis, the primary indicator of progression and prognosis in fatty liver disease. While a prior study of this same cohort demonstrated strong associations between these gene variants and advanced fibrosis, adjustment for variables beyond age, sex, and BMI was absent from this research. Here we hypothesized that education level and other social markers would associate with advanced liver fibrosis measured by magnetic resonance elastography (MRE), and that associations would persist following adjustments for genotype.

METHODS

Study Design

A cross-sectional analysis was performed on a well-characterized prospective cohort of 264 patients who had visited the University of California, San Diego (UCSD) NAFLD Research Center between 2011-2017 for a standardized research assessment. The assessment included a clinical physical exam and collection of anthropometric measurements; documentation of past medical history and social history; performance of biochemical laboratory tests, MRE assessment of liver stiffness, and *PNPLA3* and *HSD17B13* genotyping; and collection of socioeconomic and demographic data via patient questionnaire. Written, informed consent was obtained from all participants prior to their enrollment in the study, and the study was approved by the UCSD Institutional Review Board.

Participants

All patients in the study were 18 years of age or older and had provided written informed consent prior to their enrollment. Patients were ineligible for the study if they met one or more of the following criteria: alcohol intake ≥ 7 drinks/week for women or ≥ 14 drinks/week for men within the previous two years; evidence of current substance use; clinical- or laboratory-based evidence of other causes or chronic conditions associated with fatty liver disease, such as nutritional disorders, HIV infection, or use of steatosis-inducing medications (such as amiodarone, glucocorticoids, methotrexate, I-asparaginase, and valproic acid); presence of underlying liver disease aside from NAFLD, including viral hepatitis (determined using serum hepatitis B surface antigen and hepatitis C RNA assays), autoimmune hepatitis, alpha-1 antitrypsin deficiency, Wilson's disease, glycogen storage disease, hemochromatosis, and

vascular or cholestatic liver disease; decompensated liver disease (cirrhosis), defined as having a Child-Pugh score above 7 points; major systemic diseases; ineligibility for magnetic resonance imaging (MRI) assessment, based on the presence of metallic implants, claustrophobia, or a body circumference greater than the capacity of the MRI chamber; current pregnancy or active attempts to become pregnant; and any other conditions believed by the principal investigator to impact a patient's ability to complete the study. A total of 264 patients met inclusion and exclusion criteria and were included in the study.

Biomedical Variables

Each patient was seen at the research center for a standardized clinical assessment, which included a detailed history and a physical examination (comprising vital signs, height, weight, and other anthropometric measurements) conducted by a trained clinical investigator. BMI was calculated using the standard formula kg/m^2 (body weight [in kilograms] divided by height [in meters] squared).

***PNPLA3* and *HSD17B13* Genotyping**

Patient DNA was extracted from whole-blood samples obtained during the research visit using Qiagen DNeasy® Blood & Tissue Kits. *PNPLA3* and *HSD17B13* genotyping was conducted in triplicate using Applied Biosystems SNP assays and were analyzed using Quant Studio.

Socioeconomic Variables

Patients completed a registration form and questionnaire during the research visit that

asked them to report their highest education level (selecting one of the following eight options: “never attended school”; “kindergarten or below”; “grade 1-5”; “grade 6-8”; “grade 9-11”; “completed high school”; “some college”; “bachelor’s degree or higher”); marital status (selecting one of the following four options: “single/never married”; “married/marriage-like relationship”; “separated/divorced/annulled”; “widowed”); annual household income (selecting one of the following four options: “less than $15,000/year”; “$15,000-$29,999/year”; “$30,000-$49,999/year”; “$50,000/year or more”); and race (selecting one of the following six options: “American Indian or Alaska Native”; “Asian”; “Black or African American or Negro or Haitian”; “Caucasian”; “Hispanic or Latino or Latina”; “Native Hawaiian or other Pacific Islander”).

Education data were binned as a binary categorical variable prior to statistical analysis. Given the high level of education within this cohort (40% of patients reported having obtained a bachelor’s degree or higher), responses from the bottom seven education levels (“never attended school” through “some college”) were recombined into a single category, “less than college degree.” Data from the top (highest) education level, “college degree or higher,” served as the reference category in all analyses. This method of data reclassification was informed by previous research suggesting that degree attainment may be a stronger predictor of health than overall quantity or years of schooling. In other words, individuals possessing a degree may experience additional benefits to health relative to those who lack official credentials but possess an equivalent number of years of education—a phenomenon known as the “sheepskin” effect (Liu et al., 2011, 2013). For these reasons, we chose attainment of a college degree as our threshold for patient categorization, separating those possessing degrees from those reporting some college completion.

Data on annual household income (AHI), consisting of four possible responses from the patient questionnaire, were additionally binned as a binary categorical variable. Responses in the upper two income levels of the patient questionnaire were regrouped into a single income category, "$30,000/year or more." Responses in the lower two income levels of the patient questionnaire were regrouped into one income category, "less than $30,000/year."

Alcohol intake was reported outside of clinical appointments but was confirmed in the clinic with the Alcohol Use Disorders Identification Test (AUDIT; Saunders et al., 1993)) and the Lifetime Drinking History (LDH; Skinner & Sheu, 1982). Secondary causes of liver disease, such as alcohol overconsumption, were systematically excluded based on patient history and results from clinical laboratory testing.

Magnetic Resonance Elastography (MRE) Assessment of Liver Fibrosis

Magnetic resonance imaging (MRI) of the abdomen was obtained using a single 3 Tesla MR scanner (GE Signa EXCITE HDxt 3.0T; GE Healthcare) at the UCSD MR3T Research Laboratory. Liver stiffness was calculated in kilopascals (kPa) using two-dimensional MRE—the most accurate non-invasive method to quantitatively assess liver stiffness as a proxy for liver fibrosis—as previously described (Imajo et al., 2016; Park et al., 2017). In brief, an acoustic passive driver was secured to the body wall over the liver and paired with an acoustic active driver located outside of the exam room. Continuous 60 Hertz shear waves, propagated by the passive driver and directed into the liver, were imaged and processed to create transverse liver stiffness maps, known as elastograms. Four transverse slices were obtained; co-localized anatomical regions of interest were specified manually, as described previously (Dulai et al., 2016).

Definition of Advanced Fibrosis

Patients were classified as having advanced fibrosis if their estimated liver stiffness upon MRE assessment was at least 3.63 kPa. Previous studies have indicated that a threshold of ≥3.63 kPa on MRE provides a diagnostic accuracy of 0.92 for the detection of advanced fibrosis in individuals with biopsy-proven NAFLD, measured as the area under the ROC curve (AUROC) for discriminating advanced fibrosis (Imajo et al., 2016; Loomba et al., 2014; Park et al., 2017).

Statistical Analysis

In comparing patient descriptive characteristics stratified by the presence or absence of advanced fibrosis, significance was assessed using Mann-Whitney U test for continuous variables and Fisher's exact test for categorical variables (Table 1). Significance was assessed analogously in comparisons of the prevalence of advanced fibrosis stratified by patient characteristics (Figure 1). Direction of effect for characteristics involving more than two possible subcategories (e.g., to test for the presence of an ordered association between advanced fibrosis and the three possible *PNPLA3* genotypes) was confirmed using the Cochran-Armitage test for trend. Logistic regression analyses were performed to determine the odds of advanced fibrosis (outcome measure) associated with particular patient characteristics. All logistic regression models were multivariable-adjusted for patient age, sex, and BMI, with additional adjustments made, where noted, in individual analyses. Linear regression analyses were also conducted to assess the effect of various patient characteristics on the outcome of liver stiffness as a continuous measure, with adjustment for the same covariates as those in the logistic models. P-values below 0.05 were considered statistically significant.

Regression models were first tested to replicate the effects of genotype on advanced

fibrosis, with adjustments for patient age, sex, and BMI (as in Ajmera et al., 2021). For each series of analyses, we then conducted a comparative model that controlled for patient age, sex, BMI, and education level in place of genotype. We next sequentially added education level and other socioeconomic variables (marital status, income, ethnicity) to the initial "genotype-only" model, assessing effects on the outcome measure of advanced fibrosis.

All logistic regression and linear regression analyses were performed in Python programming language version 3.8.12 (Python Software Foundation). Logistic regression analyses used the logistic regression function from Statsmodels software package version 0.13.0, and linear regression analyses used the linear regression function from Scikit-learn software package version 1.0.2.

Figures were generated in Python programming language version 3.8.12, in combination with Matplotlib version 3.5.1 and Seaborn version 0.11.2.

RESULTS

I. Patient Characteristics

Two hundred sixty-four genotyped patients with liver stiffness assessed by MRE were included in this study. The median age of all patients was 58 years (IQR 42-65), with 63.3% of individuals identifying as female. The median participant BMI was 28.4 kg/m^2 (IQR 24.7-33.6) (Table 1).

Patients were stratified according to their presentation with or without advanced fibrosis on MRE assessment (≥3.63 kPa or <3.63 kPa, respectively; Table 1). Thirty-four patients (13%) presented with advanced fibrosis. Patients with advanced fibrosis, relative to those without, displayed significantly higher age (median age of 66 years vs. median age of 57 years in patients with no advanced fibrosis, $p < 0.01$), lower attainment of a four-year college degree (23% college graduates vs. 43% college graduates in patients with no advanced fibrosis, $p < 0.05$), greater Hispanic ethnicity (53% Hispanic vs. 28% Hispanic in patients with no advanced fibrosis, $p < 0.01$), and greater prevalence of the *PNPLA3* G risk variant (15% C/C, 41% C/G, and 44% G/G vs. 42% C/C, 38% C/G, and 20% G/G in patients with no advanced fibrosis, $p < 0.001$).

Select patient characteristics from Table 1 are displayed in inverse form in Figure 1, revealing the percentage of patients with advanced fibrosis stratified by patient characteristic. There was a statistically significant difference in the prevalence of advanced fibrosis according to patient *PNPLA3* genotype (with 5%, 14%, and 25% prevalence among patients of genotypes C/C, C/G, and G/G, respectively, $p < 0.001$). The Cochran-Armitage test for trend confirmed the presence of an ordered association between advanced fibrosis and *PNPLA3* genotype ($p < 0.001$), indicating that advanced fibrosis was positively associated with the number of copies of

the G risk allele. The prevalence of advanced fibrosis was additionally significantly lower among college graduates than among those without a college degree (7% vs. 16%, $p = 0.033$), and significantly higher in Hispanic than in non-Hispanic individuals (22% vs. 9% $p = 0.005$).

Table 1: Patient characteristics stratified by the presence or absence of advanced fibrosis. Age and BMI data are displayed as sample median (IQR); all other data are displayed as sample number (proportion). *P*-values comparing difference in characteristic between patients with and patients without advanced fibrosis were calculated using Mann-Whitney U test for continuous variables and Fisher's exact test for categorical variables. Bold font signifies statistical significance at $p < 0.05$.

Characteristic	All patients	Patients with advanced fibrosis (≥3.63 kPa) on MRE (n=34)	Patients with no advanced fibrosis on MRE (n=230)	p-value
Age (years)	n=263 58 (42-65)	n=34 66 (61-71)	n=229 57 (37-64)	**0.00001**
Sex	n=264	n=34	n=230	0.851
Female	167 (63.3%)	21 (61.8%)	146 (63.5%)	
Male	97 (36.7%)	13 (38.2%)	84 (36.5%)	
BMI (kg/m^2)	n=256 28.4 (24.7-33.6)	n=33 30.6 (26.4-34.9)	n=223 28.1 (24.4-33.5)	0.103
BMI (kg/m^2) class	n=256	n=33	n=223	0.249
<25 (normal or below)	67 (26.2%)	5 (15.2%)	62 (27.8%)	
25-29.9 (overweight)	84 (32.8%)	11 (33.3%)	73 (32.7%)	
≥30 (obese)	105 (41.0%)	17 (51.5%)	88 (39.5%)	
Type 2 diabetes	n=264	n=34	n=230	0.067
Yes	116 (43.9%)	20 (58.8%)	96 (41.7%)	
No	148 (56.1%)	14 (41.2%)	134 (58.3%)	
***PNPLA3* genotype**	n=264	n=34	n=230	**0.0009**
C/C	102 (38.6%)	5 (14.7%)	97 (42.2%)	
C/G	102 (38.6%)	14 (41.2%)	88 (38.3%)	
G/G	60 (22.7%)	15 (44.1%)	45 (19.6%)	
***HSD17B13* genotype**	n=242	n=30	n=212	0.086
T/T	158 (65.3%)	25 (83.3%)	133 (62.7%)	
T/TA	79 (32.6%)	5 (16.7%)	74 (34.9%)	
TA/TA	5 (2.1%)	0 (0%)	5 (2.4%)	
Ethnicity	n=259	n=34	n=225	**0.005**
Hispanic	81 (31.3%)	18 (52.9%)	63 (28.0%)	
Non-Hispanic	178 (68.7%)	16 (47.0%)	162 (72.0%)	
Education	n=250	n=31	n=219	**0.033**
College degree	101 (40.4%)	7 (22.6%)	94 (42.9%)	
No college degree	149 (59.6%)	24 (77.4%)	125 (57.1%)	
Annual household income	n=238	n=30	n=208	0.460
<$15,000	40 (16.8%)	6 (20%)	34 (16.3%)	
$15,000-$29,999	38 (19.2%)	3 (10%)	35 (16.8%)	
$30,000-$49,999	43 (26.9%)	8 (26.7%)	35 (16.8%)	
≥$50,000	117 (49.2%)	13 (43.3%)	104 (50.0%)	
Marital status	n=250	n=32	n=218	0.570
Married	133 (53.2%)	19 (59.4%)	114 (52.3%)	
Not married	117 (46.8%)	13 (40.6%)	104 (47.7%)	

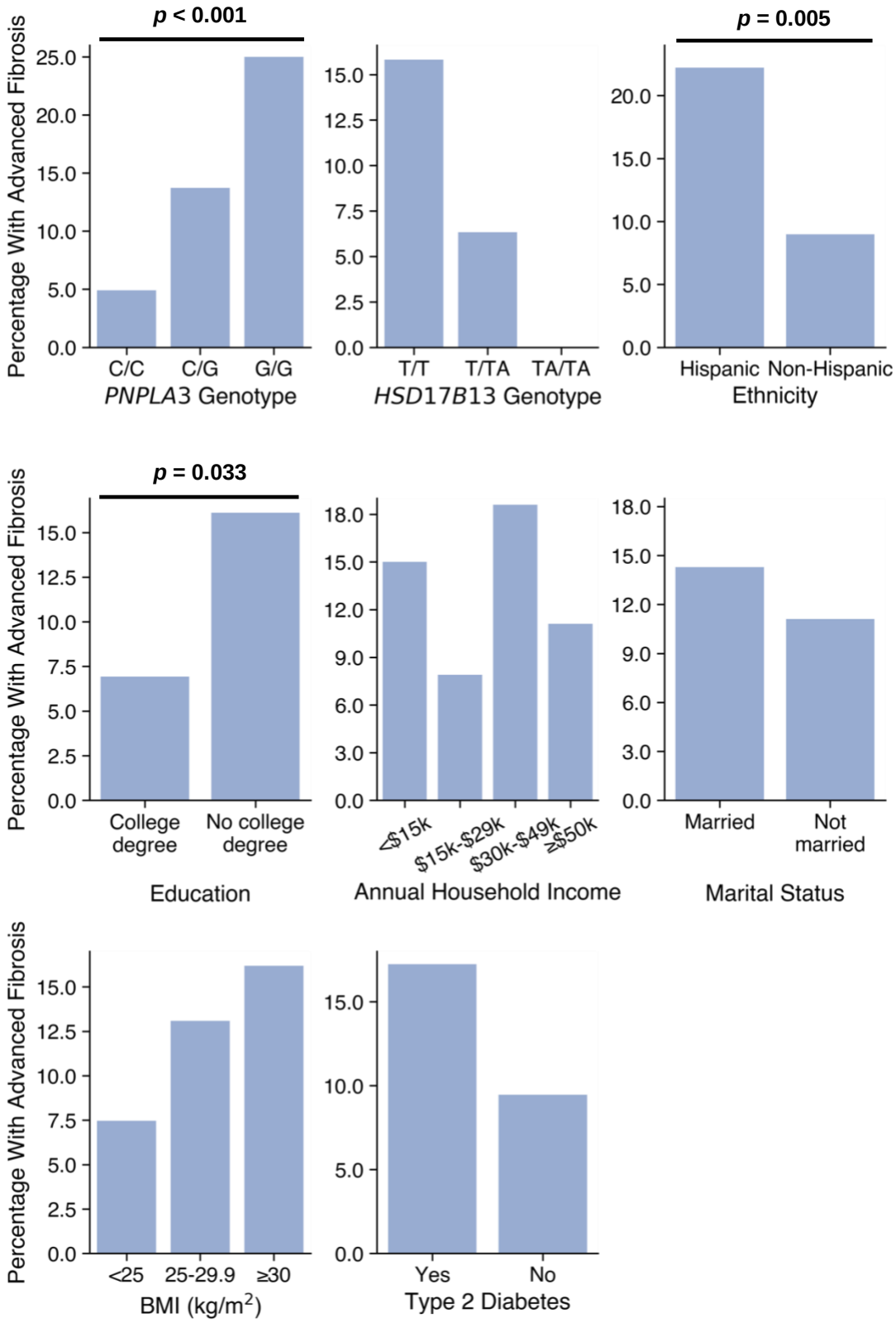

Figure 1: Presence of advanced fibrosis stratified by patient characteristics.

A significantly higher proportion of Hispanics relative to non-Hispanics possessed the *PNPLA3* G risk variant (85.2% of Hispanics vs. 49.4% of non-Hispanics, $p < 0.0001$). Conversely, a significantly lower proportion of Hispanics than non-Hispanics possessed the protective *HSD17B13* TA variant (18.7% of Hispanics vs. 42% of non-Hispanics; $p < 0.001$) (Figure 2). The *PNPLA3* G risk allele frequency was 0.42 and the *HSD17B13* TA protective allele frequency was 0.18 in our sample overall. Hispanic ethnicity was associated with a greater number of copies of the G risk allele (G allele frequency 0.66 compared to 0.30 in non-Hispanics, $p < 0.00001$), and with a lesser number of copies of the TA protective allele (TA allele frequency 0.09 compared to 0.23 in non-Hispanics, $p < 0.001$; Table 2).

Table 2: *PNPLA3* and *HSD17B13* genotypes and allele frequencies stratified by Hispanic ethnicity.

Genotype and allele frequency	**All patients**	**Hispanic**	**Non-Hispanic**	**Unknown**
	N=264	*n*=81	*n*=178	*n*=5
***PNPLA3* genotype, *n* (%)**				
C/C	102 (38.6%)	12 (14.8%)	90 (50.6%)	0 (0%)
C/G	102 (38.6%)	31 (38.3%)	69 (38.8%)	2 (40%)
G/G	60 (22.7%)	38 (46.9%)	19 (10.7%)	3 (60%)
***n* (%) with G allele**	162 (61.4%)	69 (85.2%)	88 (49.4%)	5 (100%)
G allele frequency	0.42	0.66	0.30	0.80
	N=242	*n*=75	*n*=162	*n*=5
***HSD17B13* genotype, *n* (%)**				
T/T	158 (65.3%)	61 (81.3%)	94 (58.0%)	3 (60%)
T/TA	79 (32.6%)	14 (18.7%)	63 (38.9%)	2 (40%)
TA/TA	5 (2.1%)	0 (0%)	5 (3.1%)	0 (0%)
***n* (%) with TA allele**	84 (34.7%)	14 (18.7%)	68 (42.0%)	2 (40%)
TA allele frequency	0.18	0.09	0.23	0.20

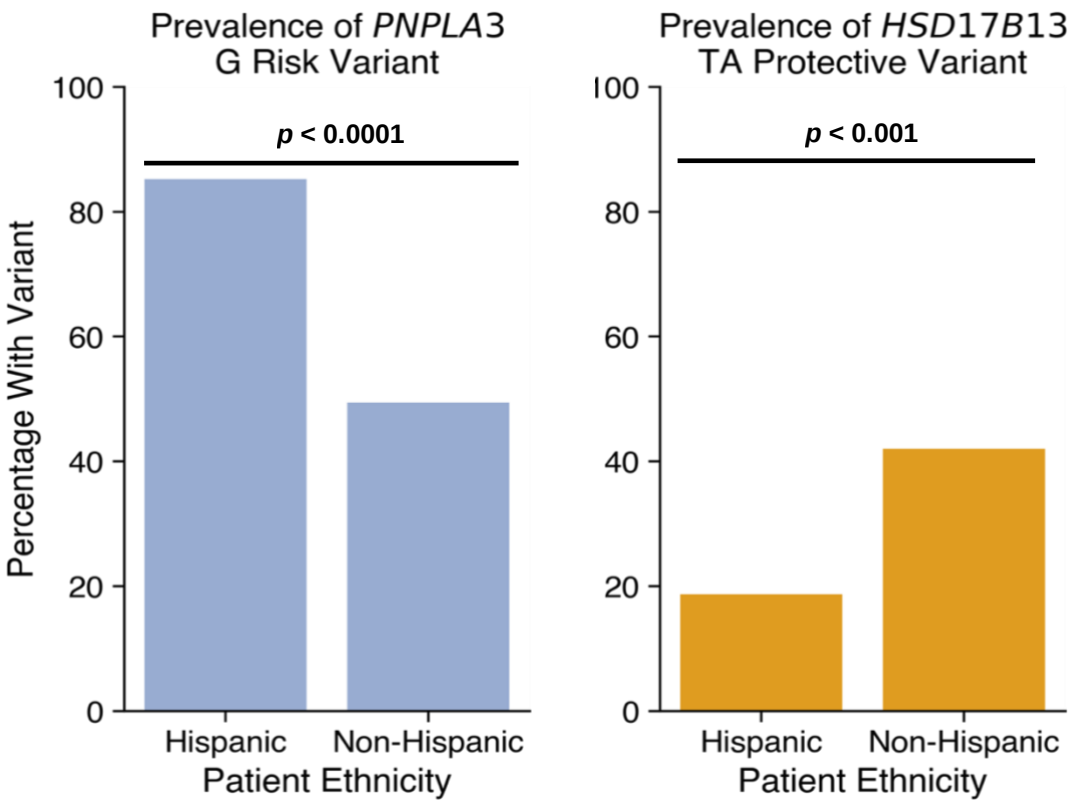

Figure 2: Proportion of patients possessing the *PNPLA3* G risk allele and *HSD17B13* TA protective allele, stratified by Hispanic ethnicity.

The education levels and annual household incomes of participants were skewed in the direction of the higher subcategories of both variables. Over 40% of patients reported having obtained at least a bachelor's degree, and an additional 39% reported having some college education. Only 21% of patients fell within the bottom six subcategories (which encompassed "never attended school" through "completed high school"). With regard to annual household income (AHI), almost half (49%) of study participants reported having an AHI within the highest income level of $50,000 or more (Figure 3).

Mean patient liver stiffness as a function of patient education level and patient annual household income is shown in Figure 4. Plotting mean liver stiffness against patient annual

household income suggested an overall negative trend toward decreasing average liver stiffness with increasing AHI, but this relationship was not consistently linear, given the reversal of the trend within the middle two income categories. Plotting mean liver stiffness by patient education level revealed a consistently negative, linear association—as patient education level increased, mean liver stiffness tended to decrease (Figure 4).

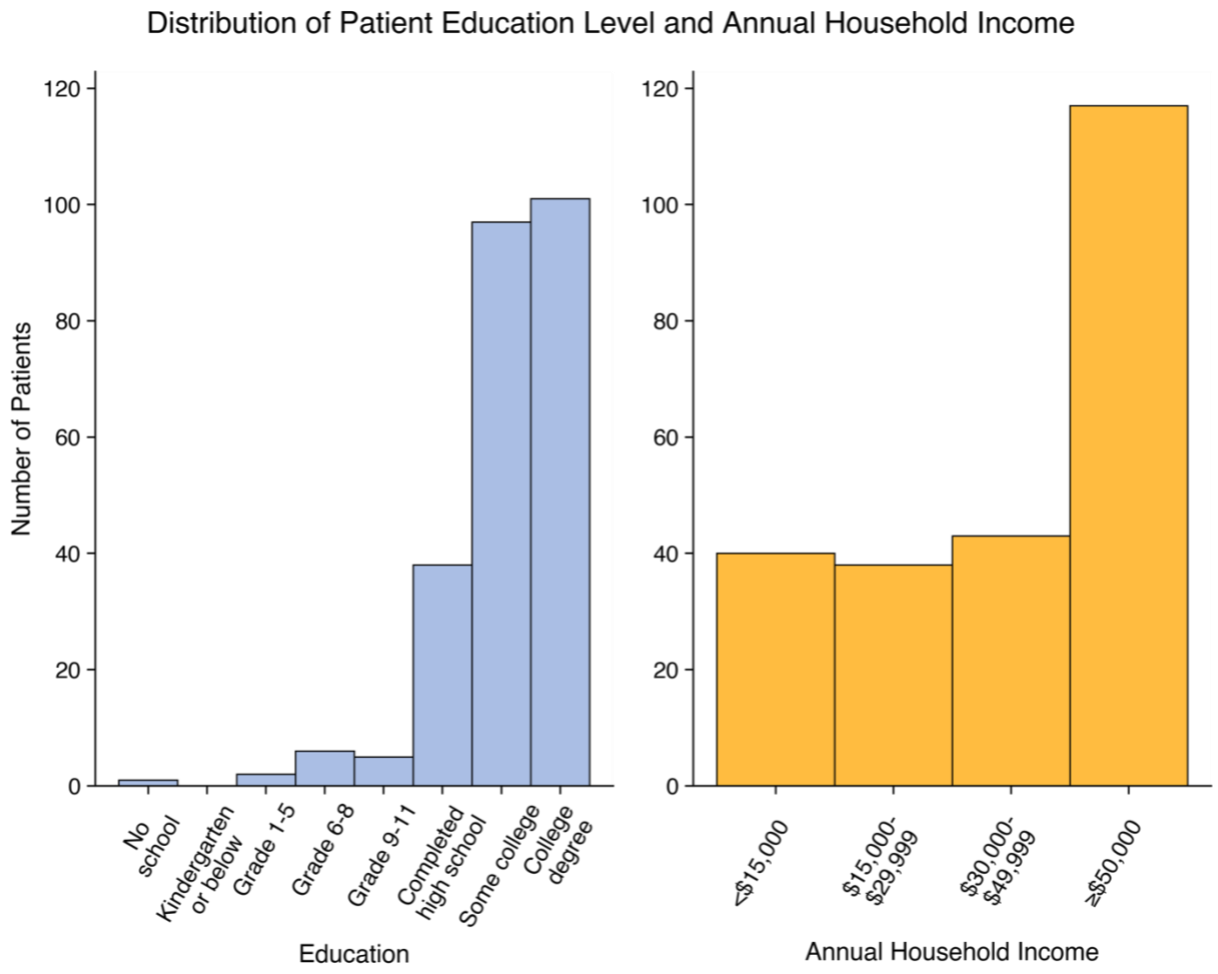

Figure 3: Distribution of patient education level and annual household income.

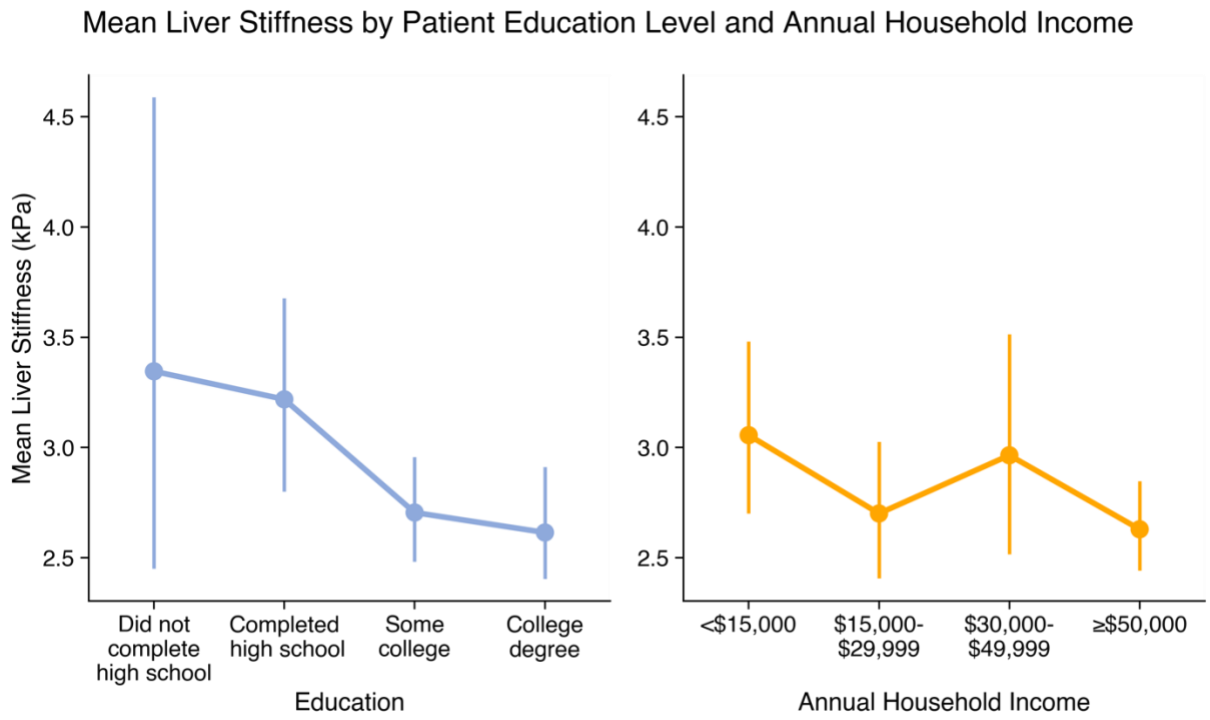

Figure 4: Mean liver stiffness (in kPa) as a function of patient education level and annual household income. Vertical bars represent 95% confidence intervals for the means.

II. Factors Associated with the Presence of Advanced Fibrosis

A. *Logistic Regression Analyses with PNPLA3 Risk Variant*

As previously demonstrated in Ajmera et al. (2021), the presence of at least one copy of the *PNPLA3* G risk variant (i.e., genotypes C/G and G/G), when controlling only for age, sex, and BMI, was associated with 4.3 times higher odds of advanced fibrosis compared to the odds of advanced fibrosis in patients of genotype C/C (95% CI: 1.6-11.8, $p < 0.01$) (Table 3a, Model 0). In a new model adjusting for patient age, sex, BMI, and education level in place of *PNPLA3* genotype, not having a college degree was associated with 3.7 times higher odds of advanced

fibrosis (95% CI: 1.4–9.9, $p < 0.01$) compared to the odds of advanced fibrosis in college graduates (Table 3a, Model 1).

Table 3a: Effect of *PNPLA3* genotype and education level on the presence of advanced liver fibrosis. Progressive multivariable-adjusted logistic regression models; the reference category for genotype is C/C; bold font signifies statistical significance at $p < 0.05$; BMI = body mass index.

	Model 0: Genotype only		**Model 1: Education only**	
Variable	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**
Age (per year)	1.067 (1.032-1.103)	**<0.001**	1.066 (2.029-1.103)	**<0.001**
Sex (female)	0.923 (0.403-2.118)	0.851	0.819 (0.349-1.921)	0.646
BMI (per kg/m^2)	1.030 (0.975-1.087)	0.291	1.032 (0.974-1.094)	0.281
Genotype (C/G or G/G)	4.275 (1.550-11.792)	**0.005**	--	--
Education (no college degree)	--	--	3.731 (1.411-9.867)	**0.008**

We then tested the effect of adding a series of socioeconomic variables, in succession, to the initial genotype-only model (Table 3b, Models 1a-1c). The strength of effect of *PNPLA3* genotype on advanced fibrosis progressively decreased after cumulative adjustment for education (Table 3b, Model 1a) and marital status (Table 3b, Model 1b), and was eventually attenuated to non-significance following adjustment for annual household income in Model 1c. In this final model, advanced fibrosis was associated with higher patient age (with each additional year of age predicting 1.1 times higher odds of advanced fibrosis [95% CI: 1.03-1.11, $p < 0.001$]) and with lack of a college degree, which was associated with 5.2 times higher odds of advanced fibrosis compared to the odds for college graduates (95% CI: 1.7-15.9, $p = 0.004$). In all of these logistic regression models, the effect of lack of a college degree on the outcome measure of advanced fibrosis was comparable to that of possessing the *PNPLA3* G risk allele.

A comparison of the effect of all variables in models with and without inclusion of socioeconomic data is depicted in a forest plot in Figure 5. The *PNPLA3* risk allele, lack of a

college degree, never being married, higher age, and higher BMI all increased the odds of advanced fibrosis. The largest effect was seen with lack of a college degree, which was the only variable other than age to retain statistical significance in the model including all variables.

Table 3b: *PNPLA3* genotype and socioeconomic variables associated with advanced liver fibrosis. Progressive multivariable-adjusted logistic regression models; the reference category for genotype is C/C; bold font signifies statistical significance at $p < 0.05$; BMI = body mass index; AHI = annual household income.

	Model 1a		Model 1b		Model 1c	
Variable	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**
Age (per year)	1.064 (1.028-1.101)	**<0.001**	1.066 (1.029-1.104)	**<0.001**	1.072 (1.032-1.113)	**0.003**
Sex (female)	0.778 (0.324-1.867)	0.574	0.778 (0.324-1.867)	0.574	0.668 (0.268-1.664)	0.386
BMI (per kg/m^2)	1.026 (0.967-1.089)	0.392	1.027 (0.968-1.090)	0.375	1.036 (0.973-1.103)	0.264
Genotype (C/G or G/G)	3.348 (1.187-9.443)	**0.022**	3.401 (1.201-9.629)	**0.021**	2.790 (0.952-8.176)	0.061
Education (no college degree)	3.260 (1.213-8.761)	**0.019**	3.295 (1.224-8.869)	**0.018**	5.226 (1.722-15.862)	**0.004**
Marital Status (never married)	--	--	1.237 (0.408-3.753)	0.706	1.633 (0.443-6.022)	0.461
AHI (<$30,000)	--	--	--	--	0.356 (0.119-1.058)	0.063

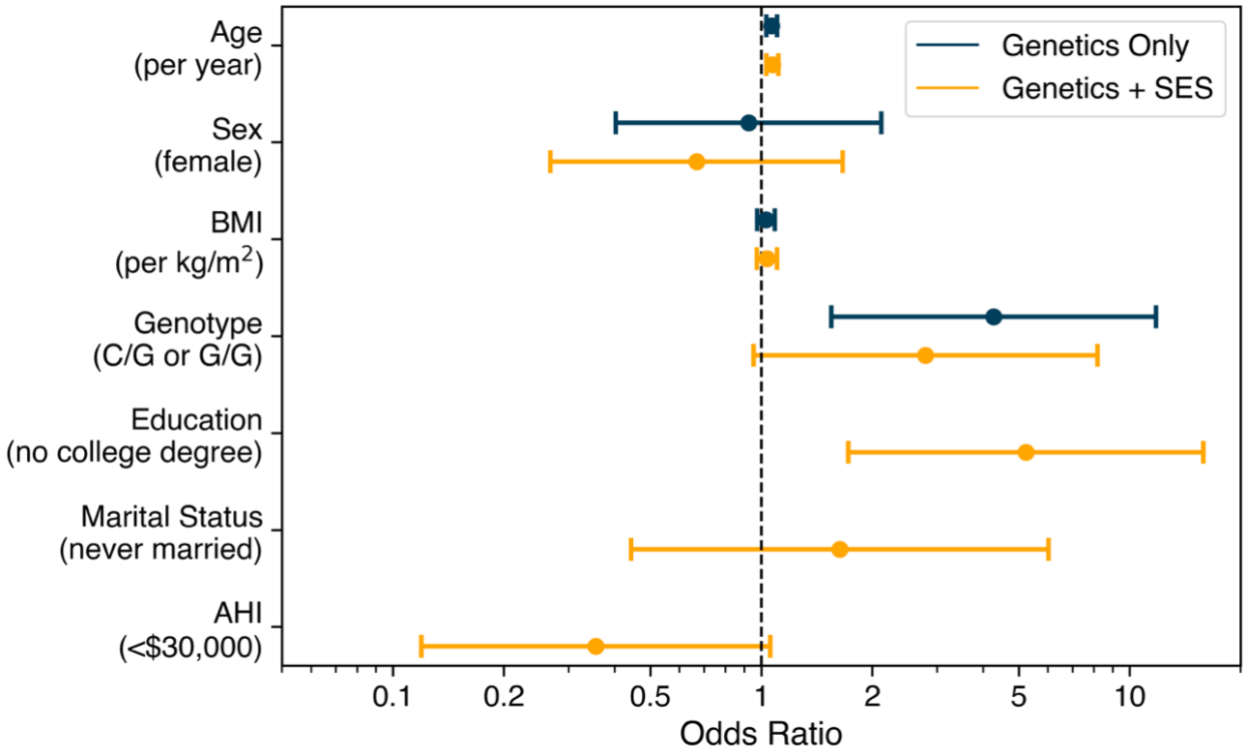

Figure 5: Forest plot of factors predicting advanced fibrosis, comparing odds ratios and 95% confidence intervals of *PNPLA3*-only model (Table 3a, Model 0) and model incorporating *PNPLA3* and socioeconomic variables (Table 3b, Model 1c).

Models 1d and 1e, presented in Table 4, display the effect of adding Hispanic ethnicity to the previous analysis. Because education level was the most significant predictor of advanced fibrosis of the socioeconomic variables modeled above, we incorporated Hispanic ethnicity *with* and *without* adjustment for education to explore whether adjustment for education might attenuate an otherwise-present association between Hispanic ethnicity and advanced fibrosis. Controlling for age, sex, BMI, marital status, income, and the presence of the *PNPLA3* G risk allele, Hispanic ethnicity was associated with 2.9 times higher odds of advanced fibrosis compared to the odds of advanced fibrosis in non-Hispanics (Table 4, Model 1d*;* 95% CI: 1.14-7.22, $p = 0.025$). In this model, age and ethnicity were the only variables that retained statistical significance at $p < 0.05$. When education level was added in Model 1e, the odds ratio for advanced fibrosis in those of Hispanic ethnicity was attenuated to non-significance, while lacking a college degree was associated with a fourfold increased odds of advanced fibrosis (Table 4, Model 1e; 95% CI: 1.27-13.17, $p = 0.018$).

Table 4: Multivariable-adjusted logistic regression models for advanced fibrosis incorporating *PNPLA3* genotype, socioeconomic variables, and Hispanic ethnicity. The reference category for genotype is C/C; bold font signifies statistical significance at $p < 0.05$; BMI = body mass index. AHI = annual household income.

	Model 1d: Inclusion of ethnicity without adjustment for education		**Model 1e: Inclusion of ethnicity with adjustment for education**	
Variable	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**
Age (per year)	1.065 (1.029-1.103)	**<0.001**	1.070 (1.031-1.111)	**<0.001**
Sex (female)	0.811 (0.331-1.990)	0.648	0.654 (0.260-1.650)	0.369
BMI (per kg/m^2)	1.036 (0.972-1.104)	0.274	1.033 (0.968-1.102)	0.330
Genotype (C/G or G/G)	2.663 (0.903-7.853)	0.076	2.381 (0.776-7.311)	0.130
Education (no college degree)	--	--	4.092 (1.272-13.166)	**0.018**
Marital Status (never married)	1.507 (0.418-5.439)	0.531	1.754 (0.469-6.563)	0.404
AHI (<$30,000)	0.496 (0.176-1.400)	0.185	0.328 (0.108-0.989)	**0.048**
Ethnicity (Hispanic)	2.871 (1.142-7.218)	**0.025**	1.827 (0.679-4.913)	0.233

B. ***Logistic Regression Analyses with HSD17B13 Protective Variant***

We then performed an analogous series of logistic regression analyses with *HSD17B13* genotype in place of that of *PNPLA3*. Adjusting for patient age, sex, and BMI (Model 0; replicated analysis of Ajmera et al., 2021), the TA allele in *HSD17B13* was associated with lower odds of advanced fibrosis (OR: 0.33 [95% CI: 0.12-0.94, $p = 0.04$]). Adding education to this model attenuated the protective effect of the TA allele to non-significance (Table 5, Model 1a), while lower education was associated with 5.2 times higher odds of advanced fibrosis (95% CI: 1.61-16.75, $p = 0.006$). In all subsequent analyses that included both the TA variant and education level as variables, education level, age, and income were the only characteristics significantly associated with advanced fibrosis (Table 5, Models 1a-1c).

Table 5: *HSD17B13* genotype and socioeconomic variables associated with advanced liver fibrosis. Progressive multivariable-adjusted logistic regression models; the reference category for genotype is T/T; bold font signifies statistical significance at $p < 0.05$; BMI = body mass index; AHI = annual household income.

	Model 0: Genetics only		**Model 1: Education only**		**Model 1a**		**Model 1b**		**Model 1c**	
Variable	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**
Age (per year)	1.084 (1.041-1.129)	**<0.001**	1.066 (2.029-1.103)	**<0.001**	1.084 (1.039-1.131)	**<0.001**	1.085 (1.038-1.133)	**<0.001**	1.096 (1.045-1.15)	**<0.001**
Sex (female)	1.265 (0.517-3.098)	0.607	0.819 (0.349-1.921)	0.646	0.958 (0.364-2.525)	0.931	0.955 (0.362-2.515)	0.925	0.838 (0.301-2.333)	0.735
BMI (per kg/m^2)	1.047 (0.986-1.112)	0.133	1.032 (0.974-1.094)	0.281	1.044 (0.978-1.114)	0.193	1.044 (0.979-1.114)	0.190	1.046 (0.975-1.122)	0.211
Genotype (T/TA or TA/TA)	0.329 (0.116-0.938)	**0.038**	--	--	0.394 (0.133-1.163)	0.092	0.396 (0.134-1.17)	0.094	0.34 (0.108-1.066)	0.064
Education (no college degree)	--	--	3.731 (1.411-9.867)	**0.008**	5.186 (1.606-16.745)	**0.006**	5.194 (1.609-16.763)	**0.006**	10.484 (2.671-41.157)	**<0.001**
Marital Status (never married)	--	--	--	--	--	--	1.119 (0.326-3.84)	0.858	1.421 (0.322-6.278)	0.643
AHI (<$30,000)	--	--	--	--	--	--	--	--	0.27 (0.082-0.883)	**0.030**

A comparison of the effect of all variables in models with and without inclusion of socioeconomic data is depicted in a forest plot in Figure 6. In the final model incorporating *HSD17B13* genotype and socioeconomic variables (Table 5, Model 1c), lack of a college degree and higher age were the only variables significantly associated with increased odds of advanced fibrosis on MRE.

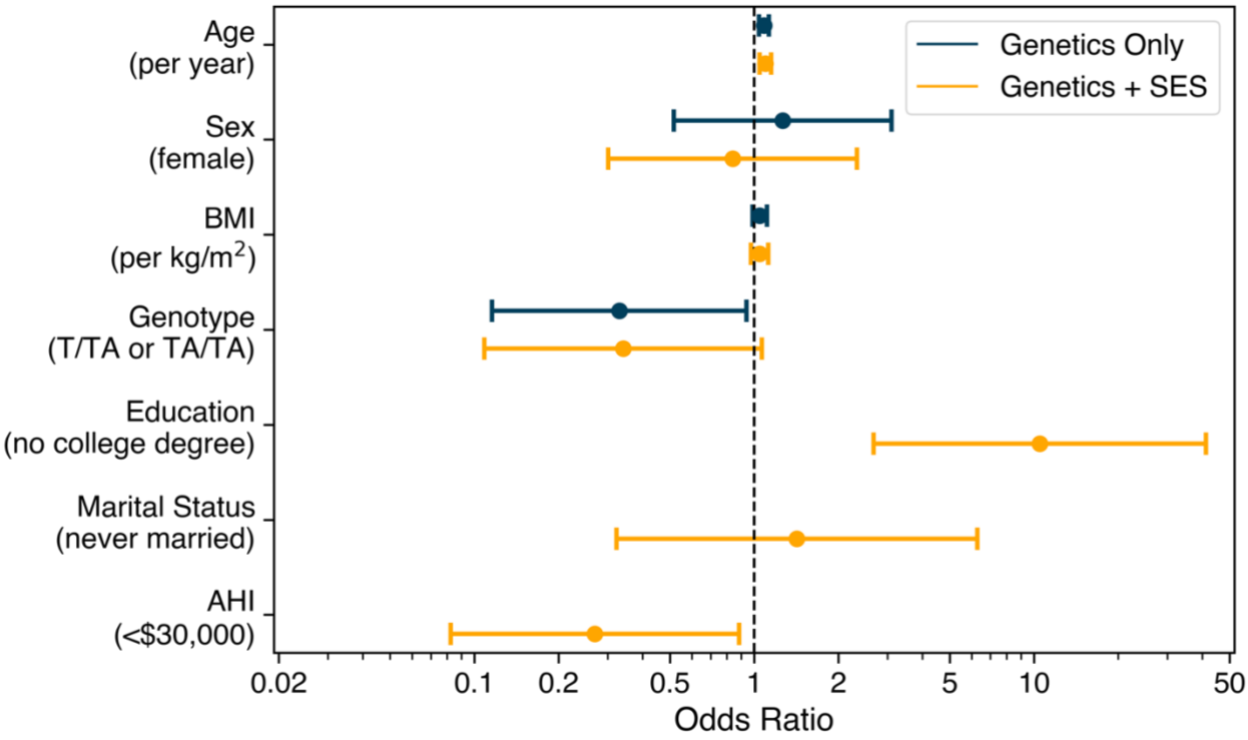

Figure 6. Forest plot of factors predicting advanced fibrosis, comparing odds ratios and 95% confidence intervals of *HSD17B13*-only model (Table 5, Model 0) and model incorporating *HSD17B13* and socioeconomic variables (Table 5, Model 1c).

Models 1d and 1e, presented in Table 6, show the effect of adding Hispanic ethnicity before and after adjustment for education. Controlling for age, sex, BMI, marital status, income, and the presence of the *HSD17B13* protective TA allele, Hispanic ethnicity was associated with 3.5 times higher odds of advanced fibrosis relative to non-Hispanic ethnicity (Table 6, Model 1d*;* 95% CI: 1.36-9.09, $p = 0.010$). In this model, age and ethnicity were the only variables that

retained statistical significance at $p < 0.05$. When education level was added in Model 1e, the effect of Hispanic ethnicity was attenuated to non-significance, while lacking a college degree was associated with an eightfold increased odds of advanced fibrosis (Table 6, Model 1e; 95% CI: 1.91-33.86, $p = 0.005$).

Table 6: Multivariable-adjusted logistic regression models for advanced fibrosis incorporating *HSD17B13* genotype, socioeconomic variables, and Hispanic ethnicity. The reference category for genotype is T/T; bold font signifies statistical significance at $p < 0.05$; BMI = body mass index; AHI = annual household income.

	Model 1d: Inclusion of ethnicity without adjustment for education		**Model 1e: Inclusion of ethnicity with adjustment for education**	
Variable	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**
Age (per year)	1.085 (1.039-1.133)	**<0.001**	1.094 (1.044-1.147)	**<0.001**
Sex (female)	1.188 (0.443-3.188)	0.732	0.818 (0.289-2.318)	0.705
BMI (per kg/m^2)	1.046 (0.976-1.122)	0.201	1.043 (0.970-1.121)	0.258
Genotype (T/TA or TA/TA)	0.385 (0.126-1.177)	0.094	0.376 (0.118-1.202)	0.099
Education (no college degree)	--	--	8.035 (1.907-33.858)	**0.005**
Marital Status (never married)	1.173 (0.280-4.916)	0.828	1.450 (0.325-6.470)	0.626
AHI (<$30,000)	0.446 (0.149-1.338)	0.150	0.259 (0.078-0.858)	**0.027**
Ethnicity (Hispanic)	3.515 (1.359-9.087)	**0.010**	1.794 (0.634-5.073)	0.270

III. Factors Associated with Advanced Fibrosis as a Continuous Measure

A. *Linear Regression Analyses with PNPLA3 Risk Variant*

Linear regression analyses were performed to assess the effects of patient characteristics on liver stiffness on MRE as a continuous measure After adjusting for age, sex, and BMI, possession of the *PNPLA3* G risk variant was associated with a 0.55 kPa increase in liver stiffness on MRE (Table 7, Model 0; 95% CI: 0.23-0.88, $p < 0.01$). In a subsequent, analogous model adjusting for age, sex, BMI, and education level in place of *PNPLA3* G risk variant (Model 1), lacking a college degree was associated with a 0.38 kPa increase in liver stiffness on

MRE (95% CI: 0.05-0.71, $p = 0.024$). After adjusting for education and the presence of the G risk variant (Model 1a), as well as cumulatively controlling for education, genotype, and marital status (Model 1b), the association between college education and increased liver stiffness fell to marginal statistical significance ($p = 0.053$ and $p = 0.051$, respectively). Surprisingly, adding adjustments for income (Table 7, Model 1c) and Hispanic ethnicity (Table 8, Model 1e) revealed new associations between liver stiffness as a continuous measure and lack of a college degree, the effect sizes of which were similar in magnitude to those of *PNPLA3* genotype (Table 7; Table 8). Consistent with previous analyses, a significant association observed between Hispanic ethnicity and increased liver stiffness was attenuated to non-significance once education level was added to the models (Table 8, Models 1d-1e).

Table 7: *PNPLA3* genotype and socioeconomic variables associated with liver stiffness on MRE as a continuous measure. Multivariable-adjusted linear regression models; the reference category for genotype is C/C; bold font signifies statistical significance at $p < 0.05$; BMI = body mass index; AHI = annual household income.

	Model 0: Genetics only		**Model 1: Education only**		**Model 1a**		**Model 1b**		**Model 1c**	
	Beta coefficient (95% CI)	**p-value**	**Beta coefficient (95% CI)**	**p-value**	**Beta coefficient (95% CI)**	**p-value**	**Beta coefficient (95% CI)**	**p-value**	**Beta coefficient (95% CI)**	**p-value**
Age (per year)	0.022 (0.013-0.032)	**<0.001**	0.022 (0.012-0.031)	**<0.001**	0.021 (0.012-0.031)	**<0.001**	0.025 (0.015-0.036)	**<0.001**	0.023 (0.014-0.032)	**<0.001**
Sex (female)	-0.168 (-0.499-0.163)	0.318	-0.175 (-0.510-0.160)	0.304	-0.196 (-0.526-0.134)	0.242	-0.191 (-0.519-0.137)	0.252	-0.248 (-0.540-0.043)	0.094
BMI (per kg/m^2)	0.021 (-0.002-0.044)	0.072	0.024 (-0.000-0.048)	0.054	0.021 (-0.002-0.045)	0.078	0.024 (0.000-0.048)	**0.048**	0.020 (-0.002-0.042)	0.069
Genotype (C/G or G/G)	0.554 (0.228-0.880)	**0.001**	--	--	0.482 (0.157-0.807)	**0.004**	0.505 (0.181-0.829)	**0.002**	0.460 (0.154-0.766)	**0.007**
Education (no college degree)	--	--	0.379 (0.051-0.707)	**0.024**	0.321 (-0.004-0.646)	0.053	0.323 (-0.001-0.646)	0.051	0.405 (0.112-0.699)	**0.003**
Marital Status (never married)	--	--	--	--	--	--	0.386 (0.057-0.012)	0.057	0.229 (-0.157-0.614)	0.243
AHI (<$30,000)	--	--	--	--	--	--	--	--	-0.076 (-0.413-0.262)	0.659

Table 8: Multivariable-adjusted linear regression models incorporating *PNPLA3* genotype, socioeconomic variables, and Hispanic ethnicity. The reference category for genotype is C/C; bold font signifies statistical significance at $p < 0.05$; BMI = body mass index. AHI = annual household income.

	Model 1d: Inclusion of ethnicity without adjustment for education		**Model 1e: Inclusion of ethnicity with adjustment for education**	
Variable	**Beta coefficient (95% CI)**	**p-value**	**Beta coefficient (95% CI)**	**p-value**
Age (per year)	0.024 (0.014-0.034)	**<0.001**	0.025 (0.015-0.035)	**<0.001**
Sex (female)	-0.192 (-0.518-0.134)	0.248	-0.233 (-0.547-0.081)	0.146
BMI (per kg/m^2)	0.022 (-0.002-0.046)	0.070	0.020 (-0.003-0.043)	0.090
Genotype (C/G or G/G)	-0.243 (-0.578-0.091)	0.153	-0.220 (-0.545-0.104)	0.182
Education (no college degree)	--	--	0.484 (0.147-0.821)	**0.005**
Marital Status (never married)	0.193 (-0.235-0.621)	0.375	0.274 (-0.142-0.689)	0.196
AHI (<$30,000)	-0.007 (-0.359-0.344)	0.968	-0.189 (-0.541-0.163)	0.291
Ethnicity (Hispanic)	0.481 (0.126-0.835)	**0.008**	0.312 (-0.047-0.671)	0.089

B. ***Linear Regression Analyses with HSD17B13 Protective Variant***

After adjusting for age, sex, and BMI, possession of the *HSD17B13* TA protective variant was associated with a 0.41 kPa decrease in liver stiffness on MRE (Table 9, Model 0; 95% CI: -0.77--0.05, $p < 0.05$). After adjusting for education level and the presence of the TA protective variant together (Model 1a), the effect of the variant on decreased liver stiffness fell to marginal statistical significance (Table 9, Model 1a; ß = -0.35 [95% CI: -0.71-0.004], $p = 0.053$), and decreased further following adjustment for marital status (Model 1b), income (Model 1c), and Hispanic ethnicity (Table 10). After cumulative adjustment for all variables in Table 10 (Model 1e), lacking a college degree was associated with a 0.32 kPa increase in liver stiffness (95% CI: 0.01-0.63, $p = 0.016$), and retained statistical significance in all previous models.

Table 9: *HSD17B13* genotype and socioeconomic variables associated with liver stiffness on MRE as a continuous measure. Multivariable-adjusted linear regression models; bold font signifies statistical significance at $p < 0.05$. BMI = body mass index. AHI = annual household income.

	Model 0: Genetics only		**Model 1: Education only**		**Model 1a**		**Model 1b**		**Model 1c**	
	Beta coefficient (95% CI)	**p-value**	**Beta coefficient (95% CI)**	**p-value**	**Beta coefficient (95% CI)**	**p-value**	**Beta coefficient (95% CI)**	**p-value**	**Beta coefficient (95% CI)**	**p-value**
Age (per year)	0.024 (0.014-0.034)	**<0.001**	0.022 (0.012-0.031)	**<0.001**	0.023 (0.013-0.033)	**<0.001**	0.027 (0.016-0.038)	**<0.001**	0.024 (0.015-0.034)	**<0.001**
Sex (female)	-0.095 (-0.451-0.261)	0.600	-0.175 (-0.510-0.160)	0.304	-0.131 (-0.487-0.224)	0.468	-0.141 (-0.495-0.213)	0.433	-0.203 (-0.514-0.108)	0.199
BMI (per kg/m^2)	0.024 (0.000-0.048)	**0.047**	0.024 (-0.000-0.048)	0.054	0.025 (0.000-0.050)	**0.049**	0.028 (0.003-0.054)	**0.028**	0.023 (0.000-0.046)	**0.046**
Genotype (T/TA or TA/TA)	-0.407 (-0.765--0.048)	**0.026**	--	--	-0.353 (-0.710-0.004)	0.053	-0.339 (-0.695-0.016)	0.061	-0.285 (-0.599-0.030)	0.076
Education (no college degree)	--	--	0.379 (0.051-0.707)	**0.024**	0.372 (0.023-0.720)	**0.037**	0.386 (0.039-0.734)	**0.029**	0.563 (0.241-0.884)	**0.001**
Marital Status (never married)	--	--	--	--	--	--	0.390 (-0.041-0.820)	0.076	0.234 (-0.173-0.640)	0.258
AHI (<$30,000)	--	--	--	--	--	--	--	--	-0.165 (-0.514-0.184)	0.353

Table 10: Multivariable-adjusted linear regression models incorporating *HSD17B13* genotype, socioeconomic variables, and Hispanic ethnicity. Bold font signifies statistical significance at $p < 0.05$. BMI = body mass index. AHI = annual household income.

	Model 1d: Inclusion of ethnicity without adjustment for education		**Model 1e: Inclusion of ethnicity with adjustment for education**	
Variable	**Beta coefficient (95% CI)**	**p-value**	**Beta coefficient (95% CI)**	**p-value**
Age (per year)	0.023 (0.013-0.033)	**<0.001**	0.024 (0.014-0.033)	**<0.001**
Sex (female)	-0.254 (-0.557-0.050)	0.101	-0.277 (-0.572-0.019)	0.066
BMI (per kg/m^2)	0.020 (-0.003-0.043)	0.081	0.018 (-0.004-0.040)	0.113
Genotype (T/TA or TA/TA)	0.374 (0.059-0.690)	**0.020**	0.390 (0.072-0.708)	**0.040**
Education (no college degree)	--	--	0.324 (0.014-0.633)	**0.016**
Marital Status (never married)	0.227 (-0.176-0.630)	0.269	0.270 (-0.123-0.663)	0.177
AHI (<$30,000)	0.045(-0.290-0.381)	0.791	-0.117 (-0.458-0.225)	0.501
Ethnicity (Hispanic)	0.432 (0.085-0.778)	**0.015**	0.307 (-0.045-0.658)	0.087

DISCUSSION

Main Findings

Analyzing a phenotypically diverse, well-described cohort of 264 patients at risk for NAFLD, we demonstrate that patient education level is significantly associated with the presence of advanced liver fibrosis with an effect comparable to that of possessing the *PNPLA3* gene risk variant (OR: 3.3; controlling for age, sex, BMI, and *PNPLA3* genotype). The association between *PNPLA3* genotype and advanced fibrosis was attenuated with the addition of each social variable to regression analysis, eventually losing statistical significance following cumulative adjustment for education, marital status, and annual household income. Lacking a college degree, in contrast, continued to predict advanced fibrosis in the cumulatively adjusted model with an odds ratio of 5.2. In logistic regression analyses incorporating *HSD17B13* genotype, the protective effect of the TA allele was no longer significant following adjustment for education, while lower education predicted advanced fibrosis across *HSD17B13* models, with an odds ratio of 10.5 following cumulative adjustment for education, marital status, and annual household income.

Modeling liver stiffness continuously to test for linear trends, continuous measure of fibrosis in NAFLD, we additionally found that lacking a college degree was significantly associated with a 0.41 kPa increase in liver stiffness after controlling for age, sex, BMI, marital status, income, and *PNPLA3* genotype. This increase, which was similar in magnitude to that of possessing the *PNPLA3* risk allele, decreased to 0.32 kPa ($p = 0.016$) following adjustment for Hispanic ethnicity. This reduced effect of education was still roughly comparable to that of *PNPLA3* genotype (ß = 0.39 kPa, $p = 0.040$). In linear models incorporating *HSD17B13* in place of *PNPLA3*, the protective effect of the TA allele on liver stiffness fell to marginal statistical

significance (ß = -0.35 kPa, $p = 0.053$) after controlling for age, sex, BMI, and education level, and this effect progressively lessened following further adjustment for marital status, income, and Hispanic ethnicity. In contrast, lacking a college degree was significantly associated with a 0.48 kPa increase in liver stiffness after adjustment for *HSD17B13* and all social variables, and remained positively and significantly associated with liver stiffness across all linear models.

Notably, in all regression analyses incorporating social variables and either *PNPLA3* or *HSD17B13* genotype, both genotype and Hispanic ethnicity were significantly associated with advanced fibrosis only in the absence of education level from the analysis. In other words, adding education to the model resulted in a decreased significance and strength of effect for both genotype and ethnicity, suggesting that a patient's level of education may partially mediate the observed relationships between genotype or ethnicity and advanced fibrosis.

In Dialogue With Published Research

Of the social variables included in our study, the strongest and most consistent association with advanced fibrosis was observed for education. Educational attainment has long been recognized as varying inversely with mortality (Kitagawa & Hauser, 1973; Kunst & Mackenbach, 1994), and numerous studies in the United States and abroad have reported a higher life expectancy of several years among college-educated individuals in comparison to high school graduates (Kitagawa & Hauser, 1973; Meara et al., 2008; Richards & Barry, 1998). Health itself is known to vary with education, with the more educated in the United States reporting better health, fewer health conditions, and less frequent engagement in high-risk behaviors such as smoking (Cutler & Lleras-Muney, 2010, 2012). Molecular indicators of health correlate with education as well, and less educated individuals have been shown to exhibit a

greater prevalence of high-risk values in biomarkers of inflammatory, metabolic, and cardiovascular disease (Seeman et al., 2008). In apparent agreement with these trends, the attainment of a college degree among U.S. adults is associated with a lower risk of NAFLD (Vilar-Gomez et al., 2022). While minimal research has assessed the relationship between education and advanced fibrosis specifically, the inverse association observed within our own study was therefore not unexpected.

Our analyses incorporating education data and ethnicity produced findings similar to those of other chronic disease studies that have explored the effects of adding education level to models including only biological or genetic data. Non et al. (2012) have reported that adding education level to a model of blood pressure revealed a 0.51 mm Hg reduction in systolic blood pressure with each increasing year of education, and additionally attenuated the association between blood pressure and African genetic ancestry. Interestingly, the observed effects of education were stronger among Black than among White Americans, revealing an interaction between race and education that may highlight the importance of considering potential social-environmental influences on producing observed social disparities in disease.

The mechanisms mediating this general association between education and health are not entirely understood, but a factor that appears to be universally involved is that social support and socioeconomic position—which is determined, in part, by educational attainment (Galobardes, 2006)—affect access to resources that help people avoid risk factors for death and disease (Herd et al., 2007; Link & Phelan, 1995). Thus, while one may speculate as to the precise, individual causes mediating the relationship between education and fibrosis—perhaps college graduates, for example, are able to attain jobs with higher pay, greater prestige, and health insurance benefits—it may be more helpful to view educational attainment as reflective of a general societal position

corresponding to a complex of social, emotional, behavioral, and structural forces that coalesce to produce better health. To commit to pinpointing a single risk factor, or determining to identify its precise causal mechanism, can sometimes miss the point of the larger picture. The "resources" that help us to avoid risk factors may be material goods, such as healthier food or immunizations; or they may be abstract concepts, like the accruement of knowledge, loving relationships with others, or freedom from the stress of experiencing racism. In this light, strategies aimed at identifying and ameliorating individual risk factors or proximal causes of illness are unlikely to be as ultimately effective as efforts addressing fundamental social and internal experience, as it is what places individuals "at risk of risks" (Link & Phelan, 1995), and less an individually-based risk factor itself, that is likely to affect numerous disease pathways simultaneously and impact health in a multitude of ways.

Education is thought by some to be a superior variable to income in predicting health (Seeman et al., 2008; Smith, 2004; Winkleby et al., 1992). Although there is a lack of consensus on this issue, one benefit of utilizing education data is that income is typically an acute measure of a moment in time, which can sometimes fail to represent an individual's true wealth, generational assets, or overall access to resources. In this way, income may often, but not always, accurately reflect a person's social class, which is a core element underlying the concept of socioeconomic status (SES) (Saegert et al., 2006). Still, income is a traditional component in assessments of SES and one that often correlates highly with health and disease (Herd et al., 2007; Sabanayagam & Shankar, 2012). Interestingly, we did not find any consistently significant associations between advanced fibrosis and annual household income in our study. Unexpectedly, in the three instances in which a significant association was present (Table 4, Model 1d; Table 5, Model 1c; Table 6, Model 1d), this relationship was positive in direction:

lower AHI (below $30,000) predicted *decreased* odds of advanced fibrosis. The reason for this effect is unclear. One possibility is that the positive relationship was merely a random artifact of our relatively small sample size of under 300 individuals. It is also possible that the four response categories constructed for the collection of income data, of which the highest had a minimum threshold of only $50,000 per year, may not have provided enough range or resolution to detect a meaningful or valid direction of effect for income and fibrosis in our sample. However, we also acknowledge the possibility that the positive association seen between income and advanced fibrosis is a real effect produced by some underlying, as-of-yet undetermined, variable. For example, one very speculative possibility could relate to alcohol intake. Studies have found that individuals in higher income groups consume alcohol significantly more frequently than those in lower income groups (Huckle et al., 2010). An international analysis of moderate- to high-income countries, moreover, revealed that alcohol may constitute a normal good, that is, a product for which demand increases following an increase in income (Cutler & Lleras-Muney, 2012). (This same study also revealed that higher education increased the odds of drinking in almost every country examined.) Although the precise level of alcohol consumption required to cause fatty liver is unknown, some evidence has suggested that the greater prevalence of NAFLD observed in men than in women may be related to the higher prevalence of alcohol consumption among men (Weston et al., 2005). It may therefore warrant investigating whether the ingestion of alcohol, even in moderate amounts, could influence the onset or progression of disease. More research assessing individual alcohol use, alcohol consumption as a function of income, and the effect of moderate alcohol intake on the progression of NAFLD is needed.

Strengths and Limitations

This study evaluated a phenotypically diverse sample of patients at risk for NAFLD, 44% of whom identified as non-White. Detailed biological data, genotyping, and socioeconomic information were collected from participants, including a measurement of liver stiffness assessed by MRE—the most accurate non-invasive, continuous measure of fibrosis in NAFLD. Statistical analyses included adjustment for multiple biological, genetic, and social variables, strengthening the reliability of observed associations. Still, certain limitations merit acknowledgment. First, we employed a cross-sectional design, which limits our ability to sequence events or definitively determine the factors influencing, or being influenced by, our patients' health status. Longitudinal, prospective studies are needed to better assess whether and how social conditions may causally mediate the development of higher-risk NAFLD. Moreover, this was a single-center study with a relatively small sample size, which likely limited our ability to detect all potentially relevant associations; employing a greater number of patients will increase the power of future studies and strengthen the ability to detect meaningful existing relationships. Finally, and perhaps most importantly, our assessment of socioeconomic position was limited in granularity, restricted to pre-determined and approximate subcategories for variables such as "race" and "annual household income." Future research will benefit from offering more nuanced and individualized characterizations of the social, economic, and environmental circumstances of participants, and we hope that such social determinants of health soon become as critical a focus in mainstream medical literature as genetic and biological data are now.

NAFLD in the Context of Genetic Research

Evaluating over 250 patients with descriptive genetic and socioeconomic data in addition

to outcome data quantifying liver fibrosis, this study employed an intersectional approach that is frequently absent in published genetic literature. While previous research has demonstrated an association between variants in *PNPLA3* and *HSD17B13* and liver stiffness on MRE in NAFLD (Ajmera et al., 2021), this prior research controlled only for biological variables, limiting its ability to derive precise and meaningful associations between genes and disease. As observed through initial analysis in the present study, lacking a college degree—while adjusting only for age, sex, and BMI—predicted odds of advanced fibrosis similar to those predicted by the presence of the *PNPLA3* risk variant in an analogous, genotype-only model previously presented by Ajmera et al. (2021). While the current study is by no means exhaustive in its identification of potential factors related to disease progression in NAFLD, these findings may help to illustrate the intricacies of correlation in studies of complex liver disease.

Compelling genetic evidence has suggested a role for heritability in NAFLD, but the disease has also been repeatedly attributed to genetic causes by a striking proportion of studies that, in fact, present no genetic evidence. An article in *Gastroenterology*, for example, the most prominent journal in the field of gastrointestinal disease, explains its methodology by stating that "a familial aggregation study was performed to test the hypothesis that NAFLD is highly heritable" (Schwimmer et al., 2009, p. 1). Although the authors explicitly note that "heritability estimates consider the fraction of the total variation shown by a phenotype that can be attributed to *genetic* [emphasis added] factors" (p. 2), they proceed to state that they "confidently can conclude that the heritability [of NAFLD] is high" (p. 6) based on their finding that fatty liver is more common in parents and siblings of children with NAFLD than in parents and siblings of children without NAFLD. This reasoning, which disregards the fact that families generally share significant aspects of their environmental circumstances as well as their genes, renders their

estimates of genetic heritability unfounded. Other studies within the leading, peer-reviewed *American Journal of Gastroenterology* indicate similar disregard, as exemplified by an article that asserts, in the context of NASH, that "a positive family history is frequent, suggesting a genetic etiology for NASH" (Willner et al., 2001, p. 2957). Assumptive genetic conclusions such as these, made in the absence of controlled evidence, are remarkably common in studies of NAFLD and saturate well-respected journals still today. A recent publication in *Nature* states that "the importance of genetic . . . risk factors in the development and severity of NASH have been underscored by studies identifying a higher risk of fibrosis among family members of those diagnosed with NASH" (Friedman et al., 2018, p. 908). Another, within a prominent journal of medicine and pharmacology, argues, "the strong genetic underpinnings of NAFLD are clearly demonstrated by familial clustering of advanced liver disease and racial/ethnic predisposition to disease" (Ajmera et al., 2021, p. 2). While it may well be true that genetics play a significant causal role in the development of NAFLD, or in any other disease outcome, concluding that a disease has strong genetic underpinnings based upon the existence of familial clustering or racial and ethnic predisposition to disease is not warranted. Indeed, results from the present study, in which variables beyond the biological were incorporated, may suggest that some of the variation in NAFLD progression typically attributed to genetics could, in fact, be effects of environment.

EPILOGUE

The turn of the 21st-century marked the dawn of a new era in genetic research. With the Human Genome Project declared formally complete in 2003, scientists had, for the first time, successfully determined the order of nucleotides comprising nearly all gene-encoding regions of human DNA (Collins et al., 2003). Although the first, full-length reference sequence of a human genome—roughly three billion base pairs long—would not be finalized, gap-free, until 2022 (Nurk et al., 2022; Wang et al., 2022), this initial achievement signified a critical step in the progression of modern genomic research. The release of the near-finished sequence was soon followed by the widespread development of modern, high-throughput DNA sequencing processes. In combination with the achievements of the Human Genome Project, these powerful technologies, collectively known as next-generation sequencing, revolutionized molecular biology and the future of genomic research. With the majority of human genes mapped and identified, and a new potential to link sequence to function, scientists were positioned to more efficiently examine the full spectrum of human genomic variation. In medical research specifically, the recognition of individual differences within a shared genetic code compelled increasing exploration of new associations between genes and disease. The conduction of genome-wide association studies (GWAS), for example, has enabled the discovery of thousands of specific genetic variants linked to particular traits; as of May 17, 2022, the world's leading GWAS repository contained nearly 377,000 such associations, and grew to house more than 400,000 by the end of July (Buniello et al., 2019). The widespread application of these findings has followed, with increasing research and resources devoted to the development of novel therapeutic drugs targeting diseases as diverse as schizophrenia (Allen & Bishop, 2019) to type 2

diabetes (Imamura et al., 2016); to be sure, the pharmaceutical industry invested $83 billion dollars in research and development in 2019, an amount that is, adjusting for inflation, roughly ten times that of the industry's spending per year in the 1980s (Austin & Hayford, 2021).

There is little doubt that advances in genetic technology have enhanced our understanding and utilization of genomic information immensely. The improved efficiency and reduced cost of DNA sequencing have brought these technologies to the individual sphere, with personal genome sequencing becoming increasingly available to those wishing to understand their own inherited traits or their risk of passing these on to potential offspring. And at the societal level, GWAS have identified small genetic variations between groups that may be able to explain why certain populations experience more, or less, relative disease burden than others. Genomic research is appealing to medical audiences largely for its implicit promise to identify disease-influencing mutations, the possibility of which inspires us to envision more straightforward ways of treating and preventing disease. The power in such a capability is clear, and it is no mystery that thousands of articles reporting gene-trait associations have been published within the last decade (Buniello et al., 2019; Mills & Rahal, 2019), many hoping to find a more straightforward, biological explanation for a difficult-to-understand, and oftentimes devastating, difference in disease outcomes between individuals and populations.

But as is true with all sources of power, care should be taken to recognize the limits of these capabilities. Although GWAS are named precisely for their ability to identify *associations*, these associations are often prematurely assumed to reflect causal relationships between genes and disease—and often by scientists themselves. Recent research amidst the COVID-19 pandemic, for example, discovered a region on chromosome 3 that is significantly associated, at the genome-wide level, with increased illness severity and hospitalization in patients infected by

SARS-CoV-2 (COVID-19 Host Genetics Initiative, 2020; Ellinghaus et al., 2020). In a subsequent article published in *Nature*, Zeberg and Pääbo (2020) report their discovery that this genomic region is inherited from Neanderthals, and also note that the highest carrier frequency for the risk haplotype is found in Bangladesh, where 63% of the population carry at least one copy. Without noting whether and which social factors or comorbidities were controlled for, the authors state that "the Neanderthal haplotype may thus be a substantial contributor to COVID-19 risk in some populations . . . In apparent agreement with this, individuals of Bangladeshi origin in the UK have an about two times higher risk of dying from COVID-19 than the general population" (Zeberg & Pääbo, 2020, p. 611). This assertion, in the absence of sufficient controls for potential confounding variables that could alternatively explain the increased risk of death for Bangladeshis in the UK, is similarly problematic. Individuals within an ethnic group may share particular genetic variants, but they are also likely to share particular environmental and social circumstances as well. Without adequately controlling for these factors, the conclusion that a genetic difference plays a causal role in its association with an observed disease outcome is not justified. Zeberg and Pääbo conclude their piece by stating, "with respect to the current pandemic, it is clear that gene flow from Neanderthals has tragic consequences" (2020, p. 612). The prematurity of such a binary (and, perhaps, a bit sensationalist) statement becomes clear in light of a subsequent publication by the same authors, in which they report a different haplotype on chromosome 12, also inherited from Neanderthals, that is associated with a 22% decrease in relative risk of severe illness from COVID-19 following infection by SARS-CoV-2 (Zeberg & Pääbo, 2021).

Seeking genetic explanations for disparities in health outcomes is particularly tempting when these disparities fall along socially-defined racial lines. It is well-documented, for example,

that Black Americans experience higher rates of prematurity and low birth weight than do White Americans (Costa, 2004; Osterman et al., 2021), and the potential contribution of genetics to this disparity has been a focal point of conversation among investigators for decades (Hoffman & Ward, 1999; Menon, 2008; Naylor & Myrianthopoulos, 1967). Despite compelling evidence suggesting the lack of a meaningful role for genetics in producing these racial differences in birth outcomes (David & Collins, 2007)—including a 1997 study revealing similar birth weight distributions among infants born to U.S.-born White women and those born to African-born Black women living in the United States (David & Collins, 1997)—scores of studies have still been published proposing the likely involvement of genetic polymorphisms in generating the higher incidence of prematurity observed in Blacks (Anum et al., 2009; Dizon-Townson, 2001; Green et al., 2005; Plunkett & Muglia, 2008; Varner & Esplin, 2005; see also Fiscella, 2005). Given that Black Americans experience higher infant mortality, lower life expectancy at birth, and poorer outcomes with respect to most categories of health than White Americans (Cunningham et al., 2017; Haines, 2003; Harper et al., 2021), the incessant search for so many of these "genetic causes" is not only somewhat illogical, but also somewhat reminiscent of aspects of nineteenth- and twentieth-century race science, in which the racial health disparities revealed in statistical data were used as evidence that Blacks were inherently predisposed to death and disease and thus, "biologically inferior" (Krieger, 1987; Williams & Sternthal, 2010). Of course, respected scientists today do not work in promotion of racist ideologies, but the increasing emphasis placed on identifying the genetic causes of observed health disparities may often, at best, be distracting, as it suggests that poor health in America is not primarily a result of environmental and lifestyle factors, but rather an effect of inherent differences in predisposition to disease.

Science is revered for its commitment to objectivity in the face of phenomenological uncertainty. The fields of medicine and genetic research may sit at the heart of scientific inquiry, but they are also situated within the same cultural and social contexts that impact us all and render us susceptible to bias. Just recognition of this alone will enable even further progression of medicine and health than we have currently.

REFERENCES

Abul-Husn, N. S., Cheng, X., Li, A. H., Xin, Y., Schurmann, C., Stevis, P., Liu, Y., Kozlitina, J., Stender, S., Wood, G. C., Stepanchick, A. N., Still, M. D., McCarthy, S., O'Dushlaine, C., Packer, J. S., Balasubramanian, S., Gosalia, N., Esopi, D., Kim, S. Y., Mukherjee, S., Lopez, A. E., Fuller, E. D., Penn, J., Chu, X., Luo, J. Z., Mirshahi, U. L., Carey, D. J., Still, C. D., Feldman, M. D., Small, A., Damrauer, S. M., Rader, D. J., Zambrowicz, B., Olson, W., Murphy, A. J., Borecki, I. B., Shuldiner, A. R., Reid, J. G., Overton, J. D., Yancopoulos, G. D., Hobbs, H. H., Cohen, J. C., Gottesman, O., Teslovich, T. M., Baras, A., Mirshahi, T., Gromada, J., & Dewey, F. E. (2018). A Protein-Truncating HSD17B13 Variant and Protection from Chronic Liver Disease. *New England Journal of Medicine*, *378*(12), 1096–1106. https://doi.org/10.1056/NEJMoa1712191

Ajmera, V., Liu, A., Bettencourt, R., Dhar, D., Richards, L., & Loomba, R. (2021). The impact of genetic risk on liver fibrosis in non-alcoholic fatty liver disease as assessed by magnetic resonance elastography. *Alimentary Pharmacology & Therapeutics*, *54*(1), 68–77. https://doi.org/10.1111/apt.16392

Allen, J. D., & Bishop, J. R. (2019). A systematic review of genome-wide association studies of antipsychotic response. *Pharmacogenomics*, *20*(4), 291–306. https://doi.org/10.2217/pgs-2018-0163

Angulo, P. (2002). Nonalcoholic Fatty Liver Disease. *New England Journal of Medicine*, *346*(16), 1221–1231. https://doi.org/10.1056/NEJMra011775

Angulo, P., Alba, L. M., Petrovic, L. M., Adams, L. A., Lindor, K. D., & Jensen, M. D. (2004). Leptin, insulin resistance, and liver fibrosis in human nonalcoholic fatty liver disease. *Journal of Hepatology*, *41*(6), 943–949. https://doi.org/10.1016/j.jhep.2004.08.020

Angulo, P., Kleiner, D. E., Dam-Larsen, S., Adams, L. A., Bjornsson, E. S., Charatcharoenwitthaya, P., Mills, P. R., Keach, J. C., Lafferty, H. D., Stahler, A., Haflidadottir, S., & Bendtsen, F. (2015). Liver Fibrosis, but No Other Histologic Features, Is Associated With Long-term Outcomes of Patients With Nonalcoholic Fatty Liver Disease. *Gastroenterology*, *149*(2), 389-397.e10. https://doi.org/10.1053/j.gastro.2015.04.043

Anum, E. A., Springel, E. H., Shriver, M. D., & Strauss, J. F. (2009). Genetic Contributions to Disparities in Preterm Birth. *Pediatric Research*, *65*(1), 1–9. https://doi.org/10.1203/PDR.0b013e31818912e7

Austin, D., & Hayford, T. (2021). *Research and Development in the Pharmaceutical Industry*. United States Congressional Budget Office. https://www.cbo.gov/publication/57025

Bonacini, M., Kassamali, F., Kari, S., Lopez Barrera, N., & Kohla, M. (2021). Racial differences in prevalence and severity of non-alcoholic fatty liver disease. *World Journal of Hepatology*, *13*(7), 763–773. https://doi.org/10.4254/wjh.v13.i7.763

Brunt, E. M., Wong, V. W.-S., Nobili, V., Day, C. P., Sookoian, S., Maher, J. J., Bugianesi, E., Sirlin, C. B., Neuschwander-Tetri, B. A., & Rinella, M. E. (2015). Nonalcoholic fatty liver disease. *Nature Reviews Disease Primers*, *1*(1), 1–22. https://doi.org/10.1038/nrdp.2015.80

Bugianesi, E., Gastaldelli, A., Vanni, E., Gambino, R., Cassader, M., Baldi, S., Ponti, V., Pagano, G., Ferrannini, E., & Rizzetto, M. (2005). Insulin resistance in non-diabetic patients with non-alcoholic fatty liver disease: sites and mechanisms. *Diabetologia*, *48*(4), 634–642. https://doi.org/10.1007/s00125-005-1682-x

Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., Suveges, D., Vrousgou, O., Whetzel, P. L., Amode, R., Guillen, J. A., Riat, H. S., Trevanion, S. J., Hall, P., Junkins, H., Flicek, P., Burdett, T., Hindorff, L. A., Cunningham, F., & Parkinson, H. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, *47*(Database issue), D1005–D1012. https://doi.org/10.1093/nar/gky1120

Burra, P., Becchetti, C., & Germani, G. (2020). NAFLD and liver transplantation: Disease burden, current management and future challenges. *JHEP Reports*, *2*(6), 100192. https://doi.org/10.1016/j.jhepr.2020.100192

Carroll, J. F., Chiapa, A. L., Rodriquez, M., Phelps, D. R., Cardarelli, K. M., Vishwanatha, J. K., Bae, S., & Cardarelli, R. (2008). Visceral Fat, Waist Circumference, and BMI: Impact of Race/ethnicity. *Obesity*, *16*(3), 600–607. https://doi.org/10.1038/oby.2007.92

Ciardullo, S., & Perseghin, G. (2021). Prevalence of NAFLD, MAFLD and associated advanced fibrosis in the contemporary United States population. *Liver International*, *41*(6), 1290–1293. https://doi.org/10.1111/liv.14828

Collins, F. S., Morgan, M., & Patrinos, A. (2003). The Human Genome Project: Lessons from Large-Scale Biology. *Science*, *300*(5617), 286–290. https://doi.org/10.1126/science.1084564

Costa, D. L. (2004). Race and Pregnancy Outcomes in the Twentieth Century: A Long-Term Comparison. *The Journal of Economic History*, *64*(4), 1056–1086.

COVID-19 Host Genetics Initiative. (2020). The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *European Journal of Human Genetics*, *28*(6), 715–718. https://doi.org/10.1038/s41431-020-0636-6

Cunningham, T. J., Croft, J. B., Liu, Y., Lu, H., Eke, P. I., & Giles, W. H. (2017). Vital Signs: Racial Disparities in Age-Specific Mortality Among Blacks or African Americans — United States, 1999–2015. *Morbidity and Mortality Weekly Report*, *66*(17), 444–456. https://doi.org/10.15585/mmwr.mm6617e1

Cutler, D., & Lleras-Muney, A. (2010). Understanding Differences in Health Behaviors by Education. *Journal of Health Economics*, *29*(1), 1–28. https://doi.org/10.1016/j.jhealeco.2009.10.003

Cutler, D., & Lleras-Muney, A. (2012). *Education and Health: Insights from International Comparisons* (No. w17738; p. w17738). National Bureau of Economic Research. https://doi.org/10.3386/w17738

David, R., & Collins, J. (2007). Disparities in Infant Mortality: What's Genetics Got to Do With It? *American Journal of Public Health*, *97*(7), 1191–1197. https://doi.org/10.2105/AJPH.2005.068387

David, R. J., & Collins, J. W. (1997). Differing Birth Weight among Infants of U.S.-Born Blacks, African-Born Blacks, and U.S.-Born Whites. *New England Journal of Medicine*, *337*(17), 1209–1214. https://doi.org/10.1056/NEJM199710233371706

Dizon-Townson, D. S. (2001). Preterm labour and delivery: a genetic predisposition. *Paediatric and Perinatal Epidemiology*, *15*(s2), 57–62. https://doi.org/10.1046/j.1365-3016.2001.00008.x

Dulai, P. S., Sirlin, C. B., & Loomba, R. (2016). MRI and MRE for non-invasive quantitative assessment of hepatic steatosis and fibrosis in NAFLD and NASH: Clinical trials to clinical practice. *Journal of Hepatology*, *65*(5), 1006–1016. https://doi.org/10.1016/j.jhep.2016.06.005

Ekstedt, M., Hagström, H., Nasr, P., Fredrikson, M., Stål, P., Kechagias, S., & Hultcrantz, R. (2015). Fibrosis stage is the strongest predictor for disease-specific mortality in NAFLD

after up to 33 years of follow-up. *Hepatology*, *61*(5), 1547–1554. https://doi.org/10.1002/hep.27368

Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., Fernández, J., Prati, D., Baselli, G., Asselta, R., Grimsrud, M. M., Milani, C., Aziz, F., Kässens, J., May, S., Wendorff, M., Wienbrandt, L., Uellendahl-Werth, F., Zheng, T., Yi, X., de Pablo, R., Chercoles, A. G., Palom, A., Garcia-Fernandez, A.-E., Rodriguez-Frias, F., Zanella, A., Bandera, A., Protti, A., Aghemo, A., Lleo, A., Biondi, A., Caballero-Garralda, A., Gori, A., Tanck, A., Carreras Nolla, A., Latiano, A., Fracanzani, A. L., Peschuck, A., Julià, A., Pesenti, A., Voza, A., Jiménez, D., Mateos, B., Nafria Jimenez, B., Quereda, C., Paccapelo, C., Gassner, C., Angelini, C., Cea, C., Solier, A., Pestaña, D., Muñiz-Diaz, E., Sandoval, E., Paraboschi, E. M., Navas, E., García Sánchez, F., Ceriotti, F., Martinelli-Boneschi, F., Peyvandi, F., Blasi, F., Téllez, L., Blanco-Grau, A., Hemmrich-Stanisak, G., Grasselli, G., Costantino, G., Cardamone, G., Foti, G., Aneli, S., Kurihara, H., ElAbd, H., My, I., Galván-Femenia, I., Martín, J., Erdmann, J., Ferrusquía-Acosta, J., Garcia-Etxebarria, K., Izquierdo-Sanchez, L., Bettini, L. R., Sumoy, L., Terranova, L., Moreira, L., Santoro, L., Scudeller, L., Mesonero, F., Roade, L., Rühlemann, M. C., Schaefer, M., Carrabba, M., Riveiro-Barciela, M., Figuera Basso, M. E., Valsecchi, M. G., Hernandez-Tejero, M., Acosta-Herrera, M., D'Angiò, M., Baldini, M., Cazzaniga, M., Schulzky, M., Cecconi, M., Wittig, M., Ciccarelli, M., Rodríguez-Gandía, M., Bocciolone, M., Miozzo, M., Montano, N., Braun, N., Sacchi, N., Martínez, N., Özer, O., Palmieri, O., Faverio, P., Preatoni, P., Bonfanti, P., Omodei, P., Tentorio, P., Castro, P., Rodrigues, P. M., Blandino Ortiz, A., de Cid, R., Ferrer, R., Gualtierotti, R., Nieto, R., Goerg, S., Badalamenti, S., Marsal, S., Matullo, G., Pelusi, S., Juzenas, S., Aliberti, S., Monzani, V., Moreno, V., Wesse, T., Lenz, T. L., Pumarola, T., Rimoldi, V., Bosari, S., Albrecht, W., Peter, W., Romero-Gómez, M., D'Amato, M., Duga, S., Banales, J. M., Hov, J. R., Folseraas, T., Valenti, L., Franke, A., & Karlsen, T. H. (2020). Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *The New England Journal of Medicine*, NEJMoa2020283. https://doi.org/10.1056/NEJMoa2020283

Emdin, C. A., Haas, M., Ajmera, V., Simon, T. G., Homburger, J., Neben, C., Jiang, L., Wei, W.-Q., Feng, Q., Zhou, A., Denny, J., Corey, K., Loomba, R., Kathiresan, S., & Khera, A. V. (2021). Association of Genetic Variation With Cirrhosis: A Multi-Trait Genome-Wide Association and Gene-Environment Interaction Study. *Gastroenterology*, *160*(5), 1620-1633.e13. https://doi.org/10.1053/j.gastro.2020.12.011

Estes, C., Razavi, H., Loomba, R., Younossi, Z., & Sanyal, A. J. (2018). Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. *Hepatology*, *67*(1), 123–133. https://doi.org/10.1002/hep.29466

Fan, N., Peng, L., Xia, Z., Zhang, L., Song, Z., Wang, Y., & Peng, Y. (2019). Triglycerides to high-density lipoprotein cholesterol ratio as a surrogate for nonalcoholic fatty liver

disease: a cross-sectional study. *Lipids in Health and Disease*, *18*(1), 39. https://doi.org/10.1186/s12944-019-0986-7

Fan, R., Wang, J., & Du, J. (2018). Association between body mass index and fatty liver risk: A dose-response analysis. *Scientific Reports*, *8*(1), 15273. https://doi.org/10.1038/s41598-018-33419-6

Farrell, G. C., & Larter, C. Z. (2006). Nonalcoholic fatty liver disease: From steatosis to cirrhosis. *Hepatology*, *43*(S1), S99–S112. https://doi.org/10.1002/hep.20973

Fiscella, K. (2005). Race, genes and preterm delivery. *Journal of the National Medical Association*, *97*(11), 1516–1526.

Friedman, S. L., Neuschwander-Tetri, B. A., Rinella, M., & Sanyal, A. J. (2018). Mechanisms of NAFLD development and therapeutic strategies. *Nature Medicine*, *24*(7), 908–922. https://doi.org/10.1038/s41591-018-0104-9

Galobardes, B. (2006). Indicators of socioeconomic position (part 1). *Journal of Epidemiology & Community Health*, *60*(1), 7–12. https://doi.org/10.1136/jech.2004.023531

Ge, X., Zheng, L., Wang, M., Du, Y., & Jiang, J. (2020). Prevalence trends in non-alcoholic fatty liver disease at the global, regional and national levels, 1990–2017: a population-based observational study. *BMJ Open*, *10*(8), e036663. https://doi.org/10.1136/bmjopen-2019-036663

Golovaty, I., Tien, P. C., Price, J. C., Sheira, L., Seligman, H., & Weiser, S. D. (2020). Food Insecurity May Be an Independent Risk Factor Associated with Nonalcoholic Fatty Liver Disease among Low-Income Adults in the United States. *The Journal of Nutrition*, *150*(1), 91–98. https://doi.org/10.1093/jn/nxz212

Green, N. S., Damus, K., Simpson, J. L., Iams, J., Reece, E. A., Hobel, C. J., Merkatz, I. R., Greene, M. F., Schwarz, R. H., & the March of Dimes Scientific Advisory Committee on Prematurity. (2005). Research agenda for preterm birth: Recommendations from the March of Dimes. *American Journal of Obstetrics and Gynecology*, *193*(3), 626–635. https://doi.org/10.1016/j.ajog.2005.02.106

Haines, M. R. (2003). Ethnic Differences in Demographic Behavior in the United States Has There Been Convergence? *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *36*(4), 157–195. https://doi.org/10.1080/01615440309604818

Harper, S., Riddell, C. A., & King, N. B. (2021). Declining Life Expectancy in the United States: Missing the Trees for the Forest. *Annual Review of Public Health*, *42*(1), 381–403. https://doi.org/10.1146/annurev-publhealth-082619-104231

Herd, P., Goesling, B., & House, J. S. (2007). Socioeconomic Position and Health: The Differential Effects of Education versus Income on the Onset versus Progression of Health Problems. *Journal of Health and Social Behavior*, *48*(3), 223–238. https://doi.org/10.1177/002214650704800302

Heredia, N. I., Zhang, X., Balakrishnan, M., Hwang, J. P., & Thrift, A. P. (2022). Association of lifestyle behaviors with non-alcoholic fatty liver disease and advanced fibrosis detected by transient elastography among Hispanic/Latinos adults in the U.S. *Ethnicity & Health*, *0*(0), 1–14. https://doi.org/10.1080/13557858.2022.2027883

Hoffman, J. D., & Ward, K. (1999). Genetic Factors in Preterm Delivery. *Obstetrical & Gynecological Survey*, *54*(3), 203–210.

Huang, D. Q., El-Serag, H. B., & Loomba, R. (2021). Global epidemiology of NAFLD-related HCC: trends, predictions, risk factors and prevention. *Nature Reviews Gastroenterology & Hepatology*, *18*(4), 223–238. https://doi.org/10.1038/s41575-020-00381-6

Huckle, T., You, R. Q., & Casswell, S. (2010). Socio-economic status predicts drinking patterns but not alcohol-related consequences independently. *Addiction*, *105*(7), 1192–1202. https://doi.org/10.1111/j.1360-0443.2010.02931.x

Imajo, K., Kessoku, T., Honda, Y., Tomeno, W., Ogawa, Y., Mawatari, H., Fujita, K., Yoneda, M., Taguri, M., Hyogo, H., Sumida, Y., Ono, M., Eguchi, Y., Inoue, T., Yamanaka, T., Wada, K., Saito, S., & Nakajima, A. (2016). Magnetic Resonance Imaging More Accurately Classifies Steatosis and Fibrosis in Patients With Nonalcoholic Fatty Liver Disease Than Transient Elastography. *Gastroenterology*, *150*(3), 626-637.e7. https://doi.org/10.1053/j.gastro.2015.11.048

Imamura, M., Takahashi, A., Yamauchi, T., Hara, K., Yasuda, K., Grarup, N., Zhao, W., Wang, X., Huerta-Chagoya, A., Hu, C., Moon, S., Long, J., Kwak, S. H., Rasheed, A., Saxena, R., Ma, R. C. W., Okada, Y., Iwata, M., Hosoe, J., Shojima, N., Iwasaki, M., Fujita, H., Suzuki, K., Danesh, J., Jørgensen, T., Jørgensen, M. E., Witte, D. R., Brandslund, I., Christensen, C., Hansen, T., Mercader, J. M., Flannick, J., Moreno-Macías, H., Burtt, N. P., Zhang, R., Kim, Y. J., Zheng, W., Singh, J. R., Tam, C. H. T., Hirose, H., Maegawa, H., Ito, C., Kaku, K., Watada, H., Tanaka, Y., Tobe, K., Kawamori, R., Kubo, M., Cho, Y. S., Chan, J. C. N., Sanghera, D., Frossard, P., Park, K. S., Shu, X.-O., Kim, B.-J., Florez, J. C., Tusié-Luna, T., Jia, W., Tai, E. S., Pedersen, O., Saleheen, D., Maeda, S., &

Kadowaki, T. (2016). Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes. *Nature Communications*, *7*(1), 10531. https://doi.org/10.1038/ncomms10531

Jinjuvadia, R., Antaki, F., Lohia, P., & Liangpunsakul, S. (2017). The Association between Nonalcoholic Fatty Liver Disease and Metabolic Abnormalities in United States Population. *Journal of Clinical Gastroenterology*, *51*(2), 160–166. https://doi.org/10.1097/MCG.0000000000000666

Katsiki, N., Perez-Martinez, P., Anagnostis, P., Mikhailidis, D. P., & Karagiannis, A. (2018). Is Nonalcoholic Fatty Liver Disease Indeed the Hepatic Manifestation of Metabolic Syndrome? *Current Vascular Pharmacology*, *16*(3), 219–227. https://doi.org/10.2174/1570161115666170621075619

Kim, C. H., & Younossi, Z. M. (2008). Nonalcoholic fatty liver disease: a manifestation of the metabolic syndrome. *Cleveland Clinic Journal of Medicine*, *75*(10), 721–728. https://doi.org/10.3949/ccjm.75.10.721

Kistler, K. D., Brunt, E. M., Clark, J. M., Diehl, A. M., Sallis, J. F., & Schwimmer, J. B. (2011). Physical Activity Recommendations, Exercise Intensity, and Histological Severity of Nonalcoholic Fatty Liver Disease. *The American Journal of Gastroenterology*, *106*(3), 460–468. https://doi.org/10.1038/ajg.2010.488

Kitagawa, E. M., & Hauser, P. M. (1973). *Differential mortality in the United States : a study in socioeconomic epidemiology*. Harvard Univiversity Press.

Krawczyk, M., Liebe, R., & Lammert, F. (2020). Toward Genetic Prediction of Nonalcoholic Fatty Liver Disease Trajectories: PNPLA3 and Beyond. *Gastroenterology*, *158*(7), 1865-1880.e1. https://doi.org/10.1053/j.gastro.2020.01.053

Krieger, N. (1987). Shades of Difference: Theoretical Underpinnings of the Medical Controversy on Black/White Differences in the United States, 1830–1870. *International Journal of Health Services*, *17*(2), 259–278. https://doi.org/10.2190/DBY6-VDQ8-HME8-ME3R

Kunst, A. E., & Mackenbach, J. P. (1994). The size of mortality differences associated with educational level in nine industrialized countries. *American Journal of Public Health*, *84*(6), 932–937. https://doi.org/10.2105/AJPH.84.6.932

Lim, U., Monroe, K. R., Buchthal, S., Fan, B., Cheng, I., Kristal, B. S., Lampe, J. W., Hullar, M. A., Franke, A. A., Stram, D. O., Wilkens, L. R., Shepherd, J., Ernst, T., & Le Marchand,

L. (2019). Propensity for Intra-abdominal and Hepatic Adiposity Varies Among Ethnic Groups. *Gastroenterology*, *156*(4), 966-975.e10. https://doi.org/10.1053/j.gastro.2018.11.021

Link, B. G., & Phelan, J. (1995). Social conditions as fundamental causes of disease. *Journal of Health and Social Behavior*, *Spec No*, 80–94.

Liu, S. Y., Buka, S. L., Kubzansky, L. D., Kawachi, I., Gilman, S. E., & Loucks, E. B. (2013). Sheepskin effects of education in the 10-year Framingham risk of coronary heart disease. *Social Science & Medicine*, *80*, 31–36. https://doi.org/10.1016/j.socscimed.2012.12.026

Liu, S. Y., Buka, S. L., Linkletter, C. D., Kawachi, I., Kubzansky, L., & Loucks, E. B. (2011). The Association Between Blood Pressure and Years of Schooling Versus Educational Credentials: Test of the Sheepskin Effect. *Annals of Epidemiology*, *21*(2), 128–138. https://doi.org/10.1016/j.annepidem.2010.11.004

Liu, Y.-L., Patman, G. L., Leathart, J. B. S., Piguet, A.-C., Burt, A. D., Dufour, J.-F., Day, C. P., Daly, A. K., Reeves, H. L., & Anstee, Q. M. (2014). Carriage of the PNPLA3 rs738409 C >G polymorphism confers an increased risk of non-alcoholic fatty liver disease associated hepatocellular carcinoma. *Journal of Hepatology*, *61*(1), 75–81. https://doi.org/10.1016/j.jhep.2014.02.030

Loomba, R., Wolfson, T., Ang, B., Hooker, J., Behling, C., Peterson, M., Valasek, M., Lin, G., Brenner, D., Gamst, A., Ehman, R., & Sirlin, C. (2014). Magnetic Resonance Elastography Predicts Advanced Fibrosis in Patients With Nonalcoholic Fatty Liver Disease: A Prospective Study. *Hepatology (Baltimore, Md.)*, *60*(6), 1920–1928. https://doi.org/10.1002/hep.27362

Ludwig, J., Viggiano, T. R., McGill, D. B., & Oh, B. J. (1980). Nonalcoholic steatohepatitis: Mayo Clinic experiences with a hitherto unnamed disease. *Mayo Clinic Proceedings*, *55*(7), 434–438.

Marchesini, G., Brizi, M., Bianchi, G., Tomassetti, S., Bugianesi, E., Lenzi, M., McCullough, A. J., Natale, S., Forlani, G., & Melchionda, N. (2001). Nonalcoholic Fatty Liver Disease: A Feature of the Metabolic Syndrome. *Diabetes*, *50*(8), 1844–1850. https://doi.org/10.2337/diabetes.50.8.1844

Marchesini, G., Brizi, M., Morselli-Labate, A. M., Bianchi, G., Bugianesi, E., McCullough, A. J., Forlani, G., & Melchionda, N. (1999). Association of nonalcoholic fatty liver disease with insulin resistance. *The American Journal of Medicine*, *107*(5), 450–455. https://doi.org/10.1016/s0002-9343(99)00271-5

Marušić, M., Paić, M., Knobloch, M., & Liberati Pršo, A.-M. (2021). NAFLD, Insulin Resistance, and Diabetes Mellitus Type 2. *Canadian Journal of Gastroenterology and Hepatology, 2021*, e6613827. https://doi.org/10.1155/2021/6613827

Meara, E. R., Richards, S., & Cutler, D. M. (2008). The Gap Gets Bigger: Changes In Mortality And Life Expectancy, By Education, 1981–2000. *Health Affairs*, *27*(2), 350–360. https://doi.org/10.1377/hlthaff.27.2.350

Menon, R. (2008). Spontaneous preterm birth, a clinical dilemma: Etiologic, pathophysiologic and genetic heterogeneities and racial disparity. *Acta Obstetricia et Gynecologica Scandinavica*, *87*(6), 590–600. https://doi.org/10.1080/00016340802005126

Mills, M. C., & Rahal, C. (2019). A scientometric review of genome-wide association studies. *Communications Biology*, *2*(1), 9. https://doi.org/10.1038/s42003-018-0261-x

Naylor, A. F., & Myrianthopoulos, N. C. (1967). The relation of ethnic and selected socio-economic factors to human birth-weight*. *Annals of Human Genetics*, *31*(1), 71–83. https://doi.org/10.1111/j.1469-1809.1967.tb01255.x

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., Vollger, M. R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S. J., Diekhans, M., Logsdon, G. A., Alonge, M., Antonarakis, S. E., Borchers, M., Bouffard, G. G., Brooks, S. Y., Caldas, G. V., Chen, N.-C., Cheng, H., Chin, C.-S., Chow, W., de Lima, L. G., Dishuck, P. C., Durbin, R., Dvorkina, T., Fiddes, I. T., Formenti, G., Fulton, R. S., Fungtammasan, A., Garrison, E., Grady, P. G. S., Graves-Lindsay, T. A., Hall, I. M., Hansen, N. F., Hartley, G. A., Haukness, M., Howe, K., Hunkapiller, M. W., Jain, C., Jain, M., Jarvis, E. D., Kerpedjiev, P., Kirsche, M., Kolmogorov, M., Korlach, J., Kremitzki, M., Li, H., Maduro, V. V., Marschall, T., McCartney, A. M., McDaniel, J., Miller, D. E., Mullikin, J. C., Myers, E. W., Olson, N. D., Paten, B., Peluso, P., Pevzner, P. A., Porubsky, D., Potapova, T., Rogaev, E. I., Rosenfeld, J. A., Salzberg, S. L., Schneider, V. A., Sedlazeck, F. J., Shafin, K., Shew, C. J., Shumate, A., Sims, Y., Smit, A. F. A., Soto, D. C., Sović, I., Storer, J. M., Streets, A., Sullivan, B. A., Thibaud-Nissen, F., Torrance, J., Wagner, J., Walenz, B. P., Wenger, A., Wood, J. M. D., Xiao, C., Yan, S. M., Young, A. C., Zarate, S., Surti, U., McCoy, R. C., Dennis, M. Y., Alexandrov, I. A., Gerton, J. L., O'Neill, R. J., Timp, W., Zook, J. M., Schatz, M. C., Eichler, E. E., Miga, K. H., & Phillippy, A. M. (2022). The complete sequence of a human genome. *Science*, *376*(6588), 44–53. https://doi.org/10.1126/science.abj6987

Osterman, M., Hamilton, B., Martin, J., Driscoll, A., & Valenzuela, C. (2021). *Births: Final Data for 2020*. National Center for Health Statistics (U.S.). https://doi.org/10.15620/cdc:112078

Pais, R., Barritt, A. S., Calmus, Y., Scatton, O., Runge, T., Lebray, P., Poynard, T., Ratziu, V., & Conti, F. (2016). NAFLD and liver transplantation: Current burden and expected challenges. *Journal of Hepatology*, *65*(6), 1245–1257. https://doi.org/10.1016/j.jhep.2016.07.033

Park, C. C., Nguyen, P., Hernandez, C., Bettencourt, R., Ramirez, K., Fortney, L., Hooker, J., Sy, E., Alquiraish, M. H., Valasek, M. A., Rizo, E., Richards, L., Brenner, D., Sirlin, C. B., & Loomba, R. (2017). Magnetic Resonance Elastography vs Transient Elastography in Detection of Fibrosis and Noninvasive Measurement of Steatosis in Patients with Biopsy-proven Nonalcoholic Fatty Liver Disease. *Gastroenterology*, *152*(3), 598-607.e2. https://doi.org/10.1053/j.gastro.2016.10.026

Park, Y.-W., Zhu, S., Palaniappan, L., Heshka, S., Carnethon, M. R., & Heymsfield, S. B. (2003). The Metabolic Syndrome: Prevalence and Associated Risk Factor Findings in the US Population From the Third National Health and Nutrition Examination Survey, 1988-1994. *Archives of Internal Medicine*, *163*(4), 427–436. https://doi.org/10.1001/archinte.163.4.427

Patel, P., Muller, C., & Paul, S. (2020). Racial disparities in nonalcoholic fatty liver disease clinical trial enrollment: A systematic review and meta-analysis. *World Journal of Hepatology*, *12*(8), 506–518. https://doi.org/10.4254/wjh.v12.i8.506

Plunkett, J., & Muglia, L. J. (2008). Genetic contributions to preterm birth: Implications from epidemiological and genetic association studies. *Annals of Medicine*, *40*(3), 167–179. https://doi.org/10.1080/07853890701806181

Richards, H., & Barry, R. (1998). U.S. Life Tables For 1990 By Sex, Race, And Education. *Journal of Forensic Economics*, *11*(1), 9–26.

Rinella, M. E. (2015). Nonalcoholic Fatty Liver Disease: A Systematic Review. *JAMA*, *313*(22), 2263–2273. https://doi.org/10.1001/jama.2015.5370

Romeo, S., Kozlitina, J., Xing, C., Pertsemlidis, A., Cox, D., Pennacchio, L. A., Boerwinkle, E., Cohen, J. C., & Hobbs, H. H. (2008). Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nature Genetics*, *40*(12), 1461–1465. https://doi.org/10.1038/ng.257

Sabanayagam, C., & Shankar, A. (2012). Income Is a Stronger Predictor of Mortality than Education in a National Sample of US Adults. *Journal of Health, Population, and Nutrition*, *30*(1), 82–86.

Saegert, S. C., Adler, N. E., Bullock, H. E., Cauce, A. M., Liu, W. M., & Wyche, K. F. (2006). *APA TASK FORCE ON SOCIOECONOMIC STATUS (SES)* (pp. 1–111). American Psychological Association.

Sanyal, A. J., Mathurin, P., & Nagy, L. A. (2016). Commonalities and Distinctions Between Alcoholic and Nonalcoholic Fatty Liver Disease. *Gastroenterology*, *150*(8), 1695–1697. https://doi.org/10.1053/j.gastro.2016.04.038

Saunders, J. B., Aasland, O. G., Babor, T. F., de la Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. *Addiction (Abingdon, England)*, *88*(6), 791–804. https://doi.org/10.1111/j.1360-0443.1993.tb02093.x

Schwimmer, J. B., Celedon, M. A., Lavine, J. E., Salem, R., Campbell, N., Schork, N. J., Shiehmorteza, M., Yokoo, T., Chavez, A., Middleton, M. S., & Sirlin, C. B. (2009). *Heritability of Nonalcoholic Fatty Liver Disease*. *136*(5), 8.

Seeman, T., Merkin, S. S., Crimmins, E., Koretz, B., Charette, S., & Karlamangla, A. (2008). Education, income and ethnic differences in cumulative biological risk profiles in a national sample of US adults: NHANES III (1988–1994). *Social Science & Medicine*, *66*(1), 72–87. https://doi.org/10.1016/j.socscimed.2007.08.027

Skinner, H. A., & Sheu, W. J. (1982). Reliability of alcohol use indices. The Lifetime Drinking History and the MAST. *Journal of Studies on Alcohol*, *43*(11), 1157–1170.

Smith, J. P. (2004). Unraveling the SES: Health Connection. *Population and Development Review*, *30*, 108–132.

Sookoian, S., & Pirola, C. J. (2011). Meta-analysis of the influence of I148M variant of patatin-like phospholipase domain containing 3 gene (PNPLA3) on the susceptibility and histological severity of nonalcoholic fatty liver disease. *Hepatology (Baltimore, Md.)*, *53*(6), 1883–1894. https://doi.org/10.1002/hep.24283

Stierman, B., Afful, J., Carroll, M., Te-Ching, C., Davy, O., Fink, S., Fryar, C., Gu, Q., Hales, C., Hughes, J., Ostchega, Y., Storandt, R., & Akinbami, L. (2021). *National Health and Nutrition Examination Survey 2017–March 2020 Prepandemic Data Files—Development of Files and Prevalence Estimates for Selected Health Outcomes* (No. 158; National Health Statistics Reports). National Center for Health Statistics. https://stacks.cdc.gov/view/cdc/106273

Su, W., Wang, Y., Jia, X., Wu, W., Li, L., Tian, X., Li, S., Wang, C., Xu, H., Cao, J., Han, Q., Xu, S., Chen, Y., Zhong, Y., Zhang, X., Liu, P., Gustafsson, J.-Å., & Guan, Y. (2014). Comparative proteomic study reveals 17β-HSD13 as a pathogenic protein in nonalcoholic fatty liver disease. *Proceedings of the National Academy of Sciences*, *111*(31), 11437–11442. https://doi.org/10.1073/pnas.1410741111

Trépo, E., Nahon, P., Bontempi, G., Valenti, L., Falleti, E., Nischalke, H.-D., Hamza, S., Corradini, S. G., Burza, M. A., Guyot, E., Donati, B., Spengler, U., Hillon, P., Toniutto, P., Henrion, J., Franchimont, D., Devière, J., Mathurin, P., Moreno, C., Romeo, S., & Deltenre, P. (2014). Association between the PNPLA3 (rs738409 C>G) variant and hepatocellular carcinoma: Evidence from a meta-analysis of individual participant data. *Hepatology*, *59*(6), 2170–2177. https://doi.org/10.1002/hep.26767

Trépo, E., Romeo, S., Zucman-Rossi, J., & Nahon, P. (2016). PNPLA3 gene in liver diseases. *Journal of Hepatology*, *65*(2), 399–412. https://doi.org/10.1016/j.jhep.2016.03.011

Varner, M. W., & Esplin, M. S. (2005). Current understanding of genetic factors in preterm birth. *BJOG: An International Journal of Obstetrics & Gynaecology*, *112*(s1), 28–31. https://doi.org/10.1111/j.1471-0528.2005.00581.x

Vilar-Gomez, E., Nephew, L. D., Vuppalanchi, R., Gawrieh, S., Mladenovic, A., Pike, F., Samala, N., & Chalasani, N. (2022). High-quality diet, physical activity, and college education are associated with low risk of NAFLD among the US population. *Hepatology*, *75*(6), 1491–1506. https://doi.org/10.1002/hep.32207

Wagenknecht, L. E., Scherzinger, A. L., Stamm, E. R., Hanley, A. J. G., Norris, J. M., Chen, Y.-D. I., Bryer-Ash, M., Haffner, S. M., & Rotter, J. I. (2009). Correlates and Heritability of Nonalcoholic Fatty Liver Disease in a Minority Cohort. *Obesity*, *17*(6), 1240–1246. https://doi.org/10.1038/oby.2009.4

Wang, P., Wu, C.-X., Li, Y., & Shen, N. (2020). HSD17B13 rs72613567 protects against liver diseases and histological progression of nonalcoholic fatty liver disease: a systematic review and meta-analysis. *European Review for Medical and Pharmacological Sciences*, *24*, 8997–9007.

Wang, T., Antonacci-Fulton, L., Howe, K., Lawson, H. A., Lucas, J. K., Phillippy, A. M., Popejoy, A. B., Asri, M., Carson, C., Chaisson, M. J. P., Chang, X., Cook-Deegan, R., Felsenfeld, A. L., Fulton, R. S., Garrison, E. P., Garrison, N. A., Graves-Lindsay, T. A., Ji, H., Kenny, E. E., Koenig, B. A., Li, D., Marschall, T., McMichael, J. F., Novak, A. M., Purushotham, D., Schneider, V. A., Schultz, B. I., Smith, M. W., Sofia, H. J., Weissman, T., Flicek, P., Li, H., Miga, K. H., Paten, B., Jarvis, E. D., Hall, I. M., Eichler,

E. E., & Haussler, D. (2022). The Human Pangenome Project: a global resource to map genomic diversity. *Nature*, *604*(7906), 437–446. https://doi.org/10.1038/s41586-022-04601-8

Weston, S. R., Leyden, W., Murphy, R., Bass, N. M., Bell, B. P., Manos, M. M., & Terrault, N. A. (2005). Racial and ethnic distribution of nonalcoholic fatty liver in persons with newly diagnosed chronic liver disease. *Hepatology*, *41*(2), 372–379. https://doi.org/10.1002/hep.20554

Williams, D. R., & Sternthal, M. (2010). Understanding Racial/ethnic Disparities in Health: Sociological Contributions. *Journal of Health and Social Behavior*, *51*(Suppl), S15–S27. https://doi.org/10.1177/0022146510383838

Willner, I. R., Waters, B., Patil, S. R., Reuben, A., Morelli, J., & Riely, C. A. (2001). Ninety Patients With Nonalcoholic Steatohepatitis: Insulin Resistance, Familial Tendency, and Severity of Disease. *The American Journal of Gastroenterology*, *96*(10), 2957–2961.

Winkleby, M. A., Jatulis, D. E., Frank, E., & Fortmann, S. P. (1992). Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease. *American Journal of Public Health*, *82*(6), 816–820. https://doi.org/10.2105/AJPH.82.6.816

Ye, Q., Zou, B., Yeo, Y. H., Li, J., Huang, D. Q., Wu, Y., Yang, H., Liu, C., Kam, L. Y., Tan, X. X. E., Chien, N., Trinh, S., Henry, L., Stave, C. D., Hosaka, T., Cheung, R. C., & Nguyen, M. H. (2020). Global prevalence, incidence, and outcomes of non-obese or lean non-alcoholic fatty liver disease: a systematic review and meta-analysis. *The Lancet. Gastroenterology & Hepatology*, *5*(8), 739–752. https://doi.org/10.1016/S2468-1253(20)30077-7

Younossi, Z., Anstee, Q. M., Marietti, M., Hardy, T., Henry, L., Eslam, M., George, J., & Bugianesi, E. (2018). Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nature Reviews Gastroenterology & Hepatology*, *15*(1), 11–20. https://doi.org/10.1038/nrgastro.2017.109

Younossi, Z. M., Koenig, A. B., Abdelatif, D., Fazel, Y., Henry, L., & Wymer, M. (2016). Global epidemiology of nonalcoholic fatty liver disease—Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology*, *64*(1), 73–84. https://doi.org/10.1002/hep.28431

Younossi, Z., Tacke, F., Arrese, M., Chander Sharma, B., Mostafa, I., Bugianesi, E., Wai-Sun Wong, V., Yilmaz, Y., George, J., Fan, J., & Vos, M. B. (2019). Global Perspectives on

Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis. *Hepatology*, *69*(6), 2672–2682. https://doi.org/10.1002/hep.30251

Yu, S. S. K., Castillo, D. C., Courville, A. B., & Sumner, A. E. (2012). The Triglyceride Paradox in People of African Descent. *Metabolic Syndrome and Related Disorders*, *10*(2), 77–82. https://doi.org/10.1089/met.2011.0108

Zain, S. M., Mohamed, R., Mahadeva, S., Cheah, P. L., Rampal, S., Basu, R. C., & Mohamed, Z. (2012). A multi-ethnic study of a PNPLA3 gene variant and its association with disease severity in non-alcoholic fatty liver disease. *Human Genetics*, *131*(7), 1145–1152. https://doi.org/10.1007/s00439-012-1141-y

Zeberg, H., & Pääbo, S. (2020). The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature*, *587*(7835), 610–612. https://doi.org/10.1038/s41586-020-2818-3

Zeberg, H., & Pääbo, S. (2021). A genomic region associated with protection against severe COVID-19 is inherited from Neandertals. *Proceedings of the National Academy of Sciences*, *118*(9), e2026309118. https://doi.org/10.1073/pnas.2026309118

Zhu, X.-Y., Xia, H.-G., Wang, Z.-H., Li, B., Jiang, H.-Y., Li, D.-L., Jin, R., & Jin, Y. (2020). In vitro and in vivo approaches for identifying the role of aryl hydrocarbon receptor in the development of nonalcoholic fatty liver disease. *Toxicology Letters*, *319*, 85–94. https://doi.org/10.1016/j.toxlet.2019.10.010

Zou, B., Yeo, Y. H., Nguyen, V. H., Cheung, R., Ingelsson, E., & Nguyen, M. H. (2020). Prevalence, characteristics and mortality outcomes of obese, nonobese and lean NAFLD in the United States, 1999–2016. *Journal of Internal Medicine*, *288*(1), 139–151. https://doi.org/10.1111/joim.13069