OPEN ACCESS

University of BRISTOL

Gellert-Kristensen, H., Richardson, T. G., Davey Smith, G., Nordestgaard, B. G., Tybjaerg-Hansen, A., & Stender, S. (2020). Combined effect of PNPLA3, TM6SF2, and HSD17B13 variants on risk of cirrhosis and hepatocellular carcinoma in the general population. *Hepatology*, [hep.31238]. https://doi.org/10.1002/hep.31238

Peer reviewed version

Link to published version (if available):
10.1002/hep.31238

Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research
General rights

Combined effect of *PNPLA3*, *TM6SF2*, and *HSD17B13* variants on risk of cirrhosis and hepatocellular carcinoma in the general population

Helene Gellert-Kristensen[1], Tom G. Richardson[2], George Davey Smith[2], Børge G. Nordestgaard[3,4,5], Anne Tybjærg-Hansen[1,4,5], Stefan Stender[1]

[1]Department of Clinical Biochemistry, Rigshospitalet; [2]Medical Research Council Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Barley House, Oakfield Grove, Bristol, BS8 2BN, United Kingdom; [3]Department of Clinical Biochemistry, and [4]The Copenhagen General Population Study, Herlev and Gentofte Hospital; [5]The Copenhagen City Heart Study, Bispebjerg and Frederiksberg Hospital; [1, 3-5]Copenhagen University Hospitals and Faculty of Health and Medical Sciences, University of Copenhagen, Denmark.,

Word count: Abstract: 261, main text: 4415.

Tables and Figures: 1 Table, 4 Figures; 9 Supplementary Tables and 12 Supplementary Figures.

Address for correspondence

Stefan Stender MD, PhD

Department of Clinical Biochemistry KB3011, Section for Molecular Genetics

Rigshospitalet, Copenhagen University Hospital

Blegdamsvej 9, DK-2100 Copenhagen, Denmark

Phone: +45 3545 4159; Fax: +45 3545 4160; E-mail: stefan.stender@regionh.dk

Abbreviations

ALT: Alanine transaminase

BMI: Body mass index

CCHS: Copenhagen City Heart Study

CGPS: Copenhagen General Population Study

CI: Confidence interval

HSD17B13: Hydroxysteroid 17-beta dehydrogenase 13

PNPLA3: Patatin-like phospholipase domain-containing protein 3

TM6SF2: Transmembrane 6 superfamily 2

following: design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript.

Abstract

We hypothesized that a genetic risk score for fatty liver disease influences the risk of cirrhosis and hepatocellular carcinoma. Three genetic variants (*PNPLA3* p.I148M, *TM6SF2* p.E167K, and *HSD17B13* rs72613567) were combined into a risk score, ranging from zero to six risk-increasing alleles. We examined association of the risk score with plasma markers of liver disease, and with cirrhosis and hepatocellular carcinoma in 110,903 individuals from Copenhagen, Denmark, and in 334,886 individuals from the UK Biobank. The frequencies of risk score zero, one, two, three, four, and five to six were 5%, 25%, 41%, 23%, 5.5%, and 0.5%, respectively. A higher genetic risk score was associated with increases in plasma alanine transaminase of up to xx% in those with score five to six versus zero. In meta-analysis of the Copenhagen Studies and UK Biobank, individuals with scores one, two, three, four, and five to six had odds ratios for cirrhosis of 1.6 (95% confidence interval, 1.3 to 1.9), 2.0 (95% CI, 1.8 to 2.2), 3.1 (95% CI, 2.7 to 3.5), 5.2 (95% CI, 4.2 to 6.4), and 12 (95% CI, 7.7 to 19), respectively, as compared to those with score zero. The corresponding odds ratios for hepatocellular carcinoma were 1.2 (95% CI, 0.9 to 1.7), 1.0 (95% CI, 0.7 to 1.3), 2.4 (95% CI, 1.9 to 3.0), 3.3 (95% CI, 2.2 to 5.5), and 29 (95% CI, 17 to 51).

Conclusions: A genetic risk score for fatty liver disease confers up to 12-fold higher risk of cirrhosis and up to 29-fold higher risk of hepatocellular carcinoma in individuals from the general population.

Fatty liver disease has become the most common gastrointestinal disease, affecting up to 40% of the population in Westernized societies(1, 2). The disorder encompasses a spectrum of hepatic abnormalities ranging from an excess of triglycerides in hepatocytes (steatosis), to inflammation (steatohepatitis) and scarring of the liver (fibrosis and cirrhosis), to hepatic failure(1). Ultimately, fatty liver disease can progress to hepatocellular carcinoma, a cancer with a five-year survival rate of less than 25%(3, 4).

Obesity and a habitually high alcohol intake are the main causal drivers of fatty liver disease(1, 2). Most individuals with these risk factors have hepatic steatosis, the first stage of the fatty liver disease spectrum. However, only a subset of those with hepatic steatosis go on to develop severe liver disease. Identifying these individuals early might allow for behavioral or pharmacological interventions aimed at forestalling disease progression. Unfortunately, we lack methods to accurately predict the course of fatty liver disease, despite an intense hunt for prognostic biomarkers(5, 6).

Genetics account for approximately half of the interindividual variation in risk of fatty liver disease(7, 8). So far, three sequence variations with robust effects on multiple stages of the disorder have been identified in genome-wide association studies: *PNPLA3* p.I148M(9), *TM6SF2* p.E167K(10), and a splice variant (rs72613567) in *HSD17B13*(11). For each of the three variants, individuals carrying two risk alleles have a two to four-fold higher risk of developing fibrosis, cirrhosis, and hepatocellular carcinoma compared to those carrying zero risk alleles(11-17). These relatively large effect sizes have spurred an interest into using the variants prognostically in fatty liver disease(18, 19).

We hypothesized that a genetic risk score for fatty liver disease influences risk of cirrhosis and hepatocellular carcinoma. To test this, we combined the three variants in *PNPLA3*,

TM6SF2 and *HSD17B13* into a score ranging from zero to six risk-increasing alleles. We then determined the effects of a higher genetic risk score on biochemical markers of liver disease and on the risk of cirrhosis and hepatocellular carcinoma in 110,903 individuals from the Danish general population, in 334,886 individuals from the UK Biobank, and in meta-analysis of the two studies combined. Finally, we estimated absolute ten-year risks of cirrhosis and hepatocellular carcinoma, stratified by genetic risk score, alcohol intake, obesity, age, and gender.

Methods

Participants

Copenhagen Studies

We combined two studies of the Danish general population, the Copenhagen General Population Study (CGPS, n=100,836) and the Copenhagen City Heart Study (CCHS, n=10,067), into one cohort, referred to here as the Copenhagen Studies(10, 14, 20). The CGPS and CCHS are prospective studies of the Danish general population initiated, respectively, in 2003-2015 and in 1976–1978, with follow-up examinations for CCHS in 1981-83, 1991-94, and 2001-03. Individuals were selected based on the national Danish Civil Registration System to reflect the adult population aged 20 to 100+ and of white, Danish descent. Each Danish citizen gets assigned a unique personal identification number at birth. All participants in the Copenhagen Studies had a unique personal identification number, ensuring no participant overlap between the studies. Data were obtained from a self-administered questionnaire reviewed together with an investigator on the day of attendance, a physical examination and blood samples. Blood samples for DNA extraction were drawn on the day of enrollment in the CGPS (2003–2015) and at the CCHS examinations in 1991–1994 and 2001–2003. Studies were approved by institutional review boards and Danish ethical committees and were conducted according to the Declaration of Helsinki. Written informed consent was obtained from participants.

UK Biobank

The UK Biobank is a cohort of about half a million individuals aged 40 to 69 from across the United Kingdom(21). All individuals were identified through the National Health Service patient

registries and went through a baseline assessment including a questionnaire, measurement of anthropometrics and blood sampling for subsequent analysis. National Health Service records on in-hospital admissions and causes of death until March 2019 have been linked to the participants. Genotype data were available for approximately 490,000 individuals enrolled in the study. The quality control pipeline for these data are explained in detail elsewhere(22). Briefly, the pipeline included both marker and sample-based quality control steps, including checks for population substructure, missing rates, heterozygosity frequencies, and sex mismatch. Individuals with withdrawn consent, evidence of genetic relatedness or who were not of white European ancestry were excluded from analysis. After filtering, we included 334,886 non-related individuals of self-reported British descent from the UK Biobank.

Cirrhosis and hepatocellular carcinoma

In the Copenhagen Studies, diagnoses of cirrhosis were collected from the national Danish Patient Registry and the national Danish Causes of Death Registry from January 1st, 1977 to April 18th, 2018. The national Danish Patient Registry has information on all patient contacts with all clinical hospital departments in Denmark. From 1994 and onwards, this includes emergency wards and outpatient clinics. The national Danish Causes of Death Registry contains data on the causes of all deaths in Denmark, as reported by hospitals and general practitioners. Cirrhosis cases were defined as individuals with International Classification of Diseases, tenth edition (ICD-10) codes: K70.3 (alcoholic cirrhosis, n=246), and/or K74.6 (unspecified cirrhosis of liver, n=165), and/or with ICD-8 codes 57109 (alcoholic cirrhosis, n=48), 57192 (unspecific cirrhosis, n=24), and/or 57199 (cirrhosis of non-alcoholic causes, n=13). We also divided cirrhosis cases into "alcoholic cirrhosis" (K70.3 or 57109) and "other cirrhosis" (K74.6, 57912, or 57199).

Hepatocellular carcinoma cases were defined as individuals with an ICD-10 code of C22.0 (liver cell carcinoma, n=80) or C22.9 (unspecified liver cancer, n=18) in the Danish Cancer Registry. The C22.9 diagnosis likely captures hepatocellular carcinoma cases without a verifying histology(23, 24). The Danish Cancer Registry contains information on cancer events in Denmark since 1943. Of the cancer events in the registry, nearly 100% have been validated histologically(25). We included cases of hepatocellular carcinoma recorded up to December 31st, 2016 (our last update of the registry). From January 1st, 2017 to March 1st, 2018 hepatocellular carcinoma was collected from the national Danish Patient Registry, using the same ICD-codes (C22.0, n=13 and C22.9, n=2).

In the UK Biobank, we used the same ICD-10 codes described above to define cirrhosis and hepatocellular carcinoma cases, based on data from in-hospital records and causes of death registries (UK Biobank data fields 41202, 41204, 40001, and 40002). The National Health Services initiated the use of ICD-10 codes in April 1995.

Individuals identified as cirrhosis or hepatocellular carcinoma cases that had a concurrent diagnosis of chronic (viral) hepatitis (ICD-10 B18 and/or ICD-8 57193), or acute hepatitis C (ICD-10 B17.1) were excluded from the studies (18 and 24 cirrhosis cases were excluded from the Copenhagen Studies and UK Biobank, respectively; the corresponding numbers for hepatocellular carcinoma were 10 and eight). Of the cirrhosis cases in the Copenhagen Studies, 58 received their diagnosis before the discovery of hepatitis C in 1989. Some of these cases may in theory have been caused by hepatitis C. However, more than 80% of these cases survived for more than a decade following 1989; if they had hepatitis C, it would likely have been diagnosed during follow-up. See Supplemental Figure S1 for an overview of included and excluded events in each cohort.

Genotyping

PNPLA3 p.I148M (rs738409), *TM6SF2* p.E167K (rs58542926), *HSD17B13* (rs72613567) and

GCKR P.446L (rs1260326) were genotyped using Taqman assays (ABI PRISM 7900HT Sequence

Detection System, Applied Biosystems) in the CGPS and CCHS(10, 14, 20), and by GWAS-chip in

the UK Biobank (Affymetrix UK BiLEVE and UK Biobank Axiom arrays). Genotype call

clustering in UK Biobank was assessed using the Scattershot website(26) (Supplemental Figure S2).

Genetic risk score

PNPLA3 p.I148M, *TM6SF2* p.E167K, and *HSD17B13* rs72613567 T/TA, genotypes were coded 0,

1, and 2 for non-carriers, heterozygous, and homozygous carriers of the risk-increasing allele. For

PNPLA3 p.I148M and *TM6SF2* p.E167K, this was the minor allele (the M-allele and K-allele,

respectively). For *HSD17B13* rs72613567 T/TA, we considered the major T-allele a risk-increasing

allele since the minor TA-allele of this variant associates with protection from chronic liver

disease(11). For each participant, a combined genetic risk score was calculated as the sum of these

risk-increasing alleles (range: zero to six). Due to very few individuals with score six, we combined

group five and six into one group in all main analyses. The distribution of the risk score in

populations of White Europeans, African American, Hispanic American, and East Asian ancestry

was estimated based on allele frequencies in the 1000 Genomes Project, available via Ensembl. In a

sensitivity analysis, we created a score where each variant was weighted by the ALT-effect reported

by Abul-Husn et al(11). We also tested a score that included *GCKR* p.P446L, coded 0, 1, and 2 for

PP-homozygotes, PL-heterozygotes and LL-homozygotes, respectively.

Biochemistry

Plasma levels of ALT, albumin, aspartate transaminase, alkaline phosphatase, gamma glutamyltransferase, bilirubin, and high-sensitivity C-reactive protein were measured using standard hospital assays (Konelab, Helsinki, Finland, ACL-Top, Instrumentation Laboratory, Kirchheim, Germany, and Boehringer Mannheim, Mannheim, Germany in the Copenhagen Studies, and Beckman Coulter, High Wycombe, United Kingdom in the UK Biobank).

Other covariates

Body mass index (BMI) was calculated as weight in kilograms divided by measured height in meters squared. Alcohol consumption was self-reported current intake of alcohol in units per week (1 unit = 12g alcohol). Diabetes mellitus, including both type 1 and type 2, was defined as individuals with one or more of the following ICD-codes in the national Danish Patient Registry: ICD-10 E10, E11, E13, or E14 and/or ICD-8 249 or 250 for Copenhagen Studies and identical ICD-10 codes in one of the national registries for individuals in the UK Biobank. In individuals with cirrhosis in the Copenhagen Studies we calculated CirCom score, a comorbidity index developed specifically for cirrhosis patients(27). For those with hepatocellular carcinoma in the Copenhagen Studies, we calculated the Charlson Comorbidity Index (excluding liver disease), an index of comorbidities affecting one-year survival in critically ill patients(28). Both indices were based on ICD-codes from the National Danish Patient Registry received before a diagnosis of cirrhosis or hepatocellular carcinoma (see Supplemental Tables S1 and S2 for details).

Statistical analyses

All analyses were performed using R statistical software version 3.6.0. Differences in baseline characteristics were tested with a χ^2-test for categorical traits and a Kruskal-Wallis rank sum test for continuous traits. Associations between the genetic risk score (encoded 0, 1, 2, 3, 4, and 5 and entered as a continuous variable) and biochemical markers or liver disease outcomes (prevalent and incident events combined) were tested with linear and logistic regression. The assumptions of linearity between exposure and outcome in the linear and logistic regression models were checked by visual inspection of plots of fitted values versus residuals and of predictors versus the logit of the outcome. No evidence of nonlinearity was detected. In the logistic regression analyses, we estimated 95% confidence intervals (CIs) for each group of the risk score (including the reference group) that corresponded to the amount of information underlying each group calculated with the Quasi Variances for Factor Effects in Statistical Models (qvcalc) package in R(29). This method avoids potential biases owing to low numbers in the reference group, and allows comparison of confidence intervals between any two groups(29). Results from the logistic regressions were meta-analyzed across the two cohorts using a fixed effects model. All logistic and linear regressions were adjusted for age, sex, and BMI, and in the UK Biobank additionally for the first ten genetic principal components (to account for potential population stratification). Due to skewed distributions, all plasma measurements (except albumin) were natural log-transformed before entering linear regressions when calculating the *P* for trend value. We tested for interaction between the genetic risk score and BMI, alcohol consumption, and diabetes on plasma ALT or the risk of cirrhosis or hepatocellular carcinoma by the inclusion of an interaction term between the genetic risk score and each of the covariates mentioned (one at a time) in the linear or logistic regression models. The genetic risk score and each covariate was entered as continuous variables (i.e., all tests for interaction are 1 degree of freedom).

Survival analyses were used to test the risk of incident hepatocellular carcinoma after a diagnosis of cirrhosis or all-cause mortality after a diagnosis of cirrhosis or hepatocellular carcinoma. Follow-up began at the date of cirrhosis or hepatocellular carcinoma diagnosis and ended at the date of hepatocellular carcinoma (for the cirrhosis to hepatocellular carcinoma analysis), death, or last update of the registries, whichever occurred first. The underlying timescale was years since diagnosis. Hazard ratios and 95% confidence intervals (CIs) were calculated by Cox proportional hazards regression models adjusted for age at diagnosis, sex and diabetes. Missing variable were imputed using 'predicted mean matching'(30). For each incomplete case, we identified five complete cases with predicted values closest to that of the predicted value of the incomplete case. The incomplete case's missing value is then replaced with one of the observed values from the complete cases chosen at random. The proportional hazards assumption was checked visually by plotting Schoenfeld residuals. In the cirrhosis to hepatocellular carcinoma and cirrhosis to death analyses, patients diagnosed with hepatocellular carcinoma prior to or at the time of their cirrhosis diagnosis were excluded. Cumulative incidences were calculated using the Aalen-Johansen estimator, with all-cause mortality entered as a competing event in the cirrhosis to hepatocellular carcinoma model. In a sensitivity analysis, we excluded individuals with cirrhosis or hepatocellular carcinoma diagnosed prior to baseline (ie. prevalent cases). To address potential collider bias we checked for associations between the genetic risk score and covariates in the cirrhosis and hepatocellular cases, and by adjusting the analyses for potential confounders(31).

Results

Baseline characteristics and genotyping

The baseline characteristics of the 110,761 individuals from the Copenhagen Studies and the 334,691 UK Biobank participants stratified by disease status are shown in Table 1. There were 478 with cirrhosis and 103 with hepatocellular carcinoma in the Copenhagen Studies. The corresponding numbers in the UK Biobank were 339 and 101. *PNPLA3*, *TM6SF2*, and *HSD17B13* genotypes were in Hardy-Weinberg equilibrium in both cohorts (Supplemental Figure S3 and S4, all P>0.07). Sex, age, BMI, diabetes, and alcohol consumption did not differ by genetic risk score in either cohort (Supplemental Table S3 and S4). The frequencies of risk score zero, one, two, three, four, and five to six were 5%, 25%, 41%, 23%, 5.5%, and 0.5%, respectively. The predicted distributions of the risk score in other ethnicities are shown in Supplemental Table S5.

Biochemical markers of liver disease

Higher genetic risk score was associated with higher plasma ALT in both the Copenhagen Studies and the UK Biobank (Figure 1). The association appeared curvilinear, with negligible effects in the range below a score of three, and larger effects seen for scores of three and above. Individuals with risk score zero, one, or two had similar medians and interquartile ranges of ALT. In contrast, the medians and 75th percentiles of ALT increased from score three through five to six. The absolute difference in median ALT between individuals with score zero and five to six was 5 U/L (relative difference: 26%) in both cohorts. A higher risk score was also associated with higher aspartate transaminase, gamma glutamyltransferase and bilirubin, and with lower alkaline phosphatase (Supplemental Table S6 and S7). The *PNPLA3*, *TM6SF2* and *HSD17B13* variants (but not *GCKR* p.P446L) were individually associated with ALT in both cohorts (Supplemental Figures S3, S4, and

S5). A risk score in which the alleles were weighted by their effect on ALT performed similarly to the unweighted score (Supplemental Figure S6). Adding *GCKR* p.P446L to the risk score did not improve its association with ALT (Supplemental Figure S7).

Interaction with adiposity, alcohol intake, and diabetes

Previous studies have reported synergistic relationships between the individual risk variants in *PNPLA3*, *TM6SF2*, and *HSD17B13* and adiposity on hepatic steatosis and biochemical markers of liver disease(14, 20). We wondered whether similar synergistic relationships modify the ALT-increasing effect of the three-gene risk score in the present study. In both the Copenhagen Studies and the UK Biobank, the association of the risk score with increased ALT was amplified with increasing BMI, alcohol intake, and in individuals with diabetes (Figure 2, all P for interaction <0.001). For example, among lean individuals with a BMI below 25 kg/m^2 in the Copenhagen Studies, median plasma ALT was 17 U/L in those with score zero and 19 U/L in those with score five to six (absolute difference: 2 U/L, relative difference: 12%). Among the most obese individuals with a BMI above 35 kg/m^2, the corresponding values were 23 U/L and 37 U/L (absolute difference: 14 U/L, relative difference: 61%).

Risk of cirrhosis and hepatocellular carcinoma

A higher genetic risk score was associated with cirrhosis and hepatocellular carcinoma in both the Copenhagen Studies and UK Biobank (Figure 2). Compared to individuals with score zero, those with score five to six had odds ratios for cirrhosis of 7.6 (95% CI: 4.2 - 14) and 21 (95% CI: 11 - 40) in the Copenhagen Studies and UK Biobank, respectively. The corresponding odds ratios for hepatocellular carcinoma were 47 (95% CI: 25 - 87) and 6.5 (95% CI: 2.1 - 20). In meta-analysis including the Copenhagen Studies and UK Biobank, the corresponding odds ratios were 12 (95%

CI: 7.7 - 19) for cirrhosis and 29 (95% CI: 17 - 51) for hepatocellular carcinoma. The associations

for the score with groups five and six viewed individually are shown in Supplemental Figure S5.

The risk conferred by the score was comparable for alcohol-related cirrhosis and for cirrhosis due to

other causes (odds ratios of 1.33 [95% CI, 1.19-1.50] and 1.47 [95% CI, 1.27-1.70], respectively,

for a one-unit higher score in the Copenhagen Studies). The associations were comparable for the

ALT-weighted risk score (Supplemental Figure S8), and for a score that included *GCKR* p.P446L

(Supplemental Figure S9).

We examined whether adiposity, alcohol intake, and diabetes influenced the effects of

the risk score on cirrhosis and hepatocellular carcinoma (Supplemental Figure S10). Although

power was limited in these analyses, the effect of the risk on cirrhosis score was amplified

amplified by increasing BMI, and by diabetes in the UK Biobank (P-interaction<0.05).

We wondered how the associations of the risk score would compare to those seen for

elevated plasma ALT, a routinely used biochemical marker of liver disease. In the Copenhagen

Studies, 3% of the participants had baseline ALT above the upper limit of normal (45 U/L for

women and 70 U/L for men). These individuals had hazard ratios for incident cirrhosis and

hepatocellular carcinoma of 10 (95% CI, 7.8 - 14) and 5.6 (95% CI, 3.1 - 10) as compared to those

with ALT below the upper limit of normal.

Disease progression after a diagnosis of cirrhosis or hepatocellular carcinoma

Baseline characteristics of patients with cirrhosis and hepatocellular carcinoma stratified by genetic

risk score are shown in Supplemental Tables S8 and S9. Potential confounders did not differ by risk

score, apart from diabetes which was more common among patients with cirrhosis and a high

genetic risk score. As expected, patients with a higher risk score had higher levels of plasma ALT

and gamma glutamyl-transferase, biochemical markers of liver cell damage. Of the 478 cases with

cirrhosis in the Copenhagen Studies, 38 received their diagnosis post-mortem or after a diagnosis of

hepatocellular carcinoma, leaving 440 for the prospective analyses. Of these, 44% received their

diagnosis prior to baseline assessment (Supplemental Figure S11, Panel A). Among the 103 cases

with hepatocellular carcinoma, ten were prevalent at baseline. Of the 440 individuals who received

a cirrhosis diagnosis, 27 (6%) subsequently developed hepatocellular carcinoma and 252 (57%)

died during a median follow-up of 5.1 years (Supplemental Figure S11, panel B).

The cumulative fraction of individuals with hepatocellular carcinoma as well as all-

cause mortality as a function of years post cirrhosis diagnosis, and all-cause mortality as a function

of years post hepatocellular carcinoma diagnosis all increased as a function of genetic risk score

(Figure 4, A-C). Compared to patients with cirrhosis and risk score zero to two, those with score

three to four and five to six had hazard ratios for incident hepatocellular carcinoma of 8.9 (95% CI:

3.3 - 25) and 19 (95% CI: 4.3 - 86), respectively (Figure 4A). The corresponding hazard ratios for

all-cause mortality were 1.3 (95% CI: 1.1 - 1.7) and 1.6 (95% CI: 0.7 - 3.8) (Figure 4B). A higher

genetic score also predicted all-cause mortality in patients with hepatocellular carcinoma (Figure

4C). Compared to those with score zero to two, those with score three to four and five to six had

hazard ratios for all-cause mortality of 1.8 (95% CI: 1.1 - 2.9) and 1.5 (95% CI: 0.7 - 3.2),

respectively.

The associations remained after multifactorial adjustment for age at diagnosis, sex,

diabetes, BMI, alcohol consumption, international normalized ratio, albumin, bilirubin, ALT, high

sensitivity C-reactive protein and disease-specific comorbidity indices (all P for trend <0.03). In a

sensitivity analysis, we excluded individuals with a diagnosis of cirrhosis (n=194) or hepatocellular

carcinoma (n=10) prior to baseline (see Supplemental Figure S11 for details). After these

exclusions, cirrhosis patients with risk score three to four had hazard ratios for hepatocellular

carcinoma and all-cause mortality of 8.7 (95% CI, 2.3 - 32.3) and 1.4 (1.0 - 1.9), respectively, as

compared to those with score zero to two (Supplemental Figure S12). The corresponding hazard

ratio for all-cause mortality after a diagnosis of hepatocellular carcinoma was 1.6 (95% CI, 1.0-2.7).

Discussion

The main finding of this study is that a genetic risk score comprising three common variants in *PNPLA3*, *TM6SF2*, and *HSD17B13* is associated with up to 12-fold higher risk of cirrhosis, and up to 29-fold higher risk of hepatocellular carcinoma. Moreover, a high genetic risk score conferred up to 19-fold increased risk of progressing from cirrhosis to hepatocellular carcinoma, and up to 1.6-fold higher rate of all-cause mortality after a diagnosis of cirrhosis. These results raise the question whether a similar genetic risk score might be used in a clinical setting to predict the onset and progression of chronic liver disease. The genetic risk score might be used together with established risk factors to identify individuals at high risk of developing chronic liver disease. For example, individuals with obesity, a high alcohol intake, and a high genetic risk score could be offered regular clinical surveillance, aiming to detect the development of steatohepatitis or fibrosis at an early, treatable stage.

While the highest genetic risk scores conferred a markedly higher risk of liver disease, it is important to note that 94% of the participants in our study had a score of three or lower, a range in which risk of liver disease only increased moderately. For example, participants with score three had an approximately two-fold higher risk of cirrhosis compared to those with score zero to one. Larger effect sizes are likely required for effective risk discrimination. We speculate that focusing on those in the extreme tails of the score (in combination with existing risk factors) is most likely to yield clinically useful risk discrimination.

We constructed the genetic risk score by simply counting the number of risk alleles in three variants with known effects on fatty liver disease and/or cirrhosis and hepatocellular carcinoma. This may seem crude compared to the complex methods and hundreds to thousands of variants used in recent studies to construct genetic risk scores for other traits and diseases(32, 33).

However, the three variants used here are the only common variants that have been consistently associated with fatty liver disease. The L-variant of *GCKR* p.P446L has been associated with a moderately increased risk of hepatic steatosis, but its effect on the more advanced stages of fatty liver disease is less clear(20, 34). Adding *GCKR* p.P446L to the score did not improve its performance in our study. The simplicity of the three-gene score can also be viewed as a strength, because it is easy and inexpensive to replicate and to use in a clinical setting. Another strength of the score is that the biology of the three implicated variants and genes is relatively well-known, facilitating mechanistic interpretations (*PNPLA3* and *HSD17B13* play a role in the metabolism of hepatic lipid droplet content, and *TM6SF2* is implicated in the efflux of triglycerides from the liver to the circulation(10, 35-39)).

Our study has limitations that should be considered. Despite the large sample size, the number of cases was modest, resulting in wide confidence intervals for some of the estimates (especially for those pertaining to score five to six). We only included European ancestry individuals of the Danish and British general population. The findings may therefore not necessarily be generalizable to other ethnicities. To begin to address this issue, we examined the predicted distribution of the genetic risk score in other ethnicities. The score was predicted to have a flatter distribution in East Asian populations, with relatively more individuals in the extreme tails of the score compared to other ethnicities. For example, the predicted frequencies of risk scores zero and five/six were 0.03 and 0.003, respectively, in white Europeans, and 0.04 and 0.01 in East Asians, suggesting that the risk score might be particularly useful for risk discrimination in East Asian populations. There are also limitations relating to the covariables and clinical endpoints in the study. We did not have detailed information on the etiology underlying the cirrhosis and hepatocellular carcinoma cases. Self-reported alcohol intake at baseline is an imperfect measure of lifelong alcohol consumption and may not reflect future changes in drinking patterns. Disease definitions based on

ICD-codes inevitably suffer from some degree of misclassification. For example, some individuals with compensated (asymptomatic) cirrhosis may lack an ICD-code for cirrhosis in our study. That said, most individuals with compensated cirrhosis eventually progress to symptomatic disease(40). Those individuals that did receive an ICD-code for cirrhosis in our study were unlikely to have been misclassified, because cirrhosis is a hard endpoint with well-defined diagnostic criteria. In support of this, a study from 1997 found that 85% of individuals with an ICD-code for cirrhosis in the national Danish Patient Registry had biopsy-proven cirrhosis or fulfilled the standard clinical diagnostic criteria(41). Because the methods to diagnose cirrhosis with have improved since 1997, we would expect the specificity of the cirrhosis ICD-codes to be even higher in our study. Further supporting the validity of the cirrhosis endpoint is that each of the variants in *PNPLA3*, *TM6SF2* and *HSD17B13* was strongly associated with this endpoint, with effect sizes comparable to other studies that used histology or imaging. Misclassification was likely negligible for the hepatocellular carcinoma cases in the Copenhagen Studies, because these were extracted from the Danish Cancer Registry, a registry which mainly includes histologically verified cancers. In any case, misclassification of cirrhosis and hepatocellular carcinoma most likely is nondifferential to genotypes and would only bias results toward the null hypothesis, and thus cannot explain the positive findings of the present study.

In conclusion, we found that a genetic risk score comprising three common variants influenced risk of cirrhosis by up to 12-fold and hepatocellular carcinoma by up to 29-fold in individuals from the general population.

Acknowledgements

References

1.   Younossi ZM. Non-alcoholic fatty liver disease - A global public health perspective. J Hepatol 2019;70:531-544.

2.   Crabb DW, Im GY, Szabo G, Mellinger JL, Lucey MR. Diagnosis and Treatment of Alcohol-Related Liver Diseases: 2019 Practice Guidance from the American Association for the Study of Liver Diseases. Hepatology 2019.

3.   Hassan I, Gane E. Improving survival in patients with hepatocellular carcinoma related to chronic hepatitis C and B but not in those related to non-alcoholic steatohepatitis or alcoholic liver disease: a 20-year experience from a national programme. Intern Med J 2019.

4.   Bengtsson B, Stal P, Wahlin S, Bjorkstrom NK, Hagstrom H. Characteristics and outcome of hepatocellular carcinoma in patients with NAFLD without cirrhosis. Liver Int 2019;39:1098-1108.

5.   Castera L, Friedrich-Rust M, Loomba R. Noninvasive assessment of liver disease in patients with nonalcoholic fatty liver disease. Gastroenterology 2019;156:1264-1281 e1264.

6.   Wong VW, Adams LA, de Ledinghen V, Wong GL, Sookoian S. Noninvasive biomarkers in NAFLD and NASH - current progress and future promise. Nat Rev Gastroenterol Hepatol 2018;15:461-478.

7.   Loomba R, Schork N, Chen CH, Bettencourt R, Bhatt A, Ang B, Nguyen P, et al. Heritability of hepatic fibrosis and steatosis based on a prospective twin study. Gastroenterology 2015;149:1784-1793.

8.   Schwimmer JB, Celedon MA, Lavine JE, Salem R, Campbell N, Schork NJ, Shiehmorteza M, et al. Heritability of nonalcoholic fatty liver disease. Gastroenterology 2009;136:1585-1592.

9.   Romeo S, Kozlitina J, Xing C, Pertsemlidis A, Cox D, Pennacchio LA, Boerwinkle E, et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. Nat Genet 2008;40:1461-1465.

10.   Kozlitina J, Smagris E, Stender S, Nordestgaard BG, Zhou HH, Tybjaerg-Hansen A, Vogt TF, et al. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. Nat Genet 2014;46:352-356.

11.   Abul-Husn NS, Cheng X, Li AH, Xin Y, Schurmann C, Stevis P, Liu Y, et al. A protein-truncating HSD17B13 variant and protection from chronic liver disease. N Engl J Med 2018;378:1096-1106.

12.   Yang J, Trepo E, Nahon P, Cao Q, Moreno C, Letouze E, Imbeaud S, et al. PNPLA3 and TM6SF2 variants as risk factors of hepatocellular carcinoma across various etiologies and severity of underlying liver diseases. Int J Cancer 2019;144:533-544.

13.   Yang J, Trepo E, Nahon P, Cao Q, Moreno C, Letouze E, Imbeaud S, et al. A 17-beta-hydroxysteroid dehydrogenase 13 variant protects from hepatocellular carcinoma development in alcoholic liver disease. Hepatology 2019.

14.   Gellert-Kristensen H, Nordestgaard BG, Tybjaerg-Hansen A, Stender S. High risk of fatty liver disease amplifies the alanine transaminase-lowering effect of a HSD17B13 variant. Hepatology 2019.

15.   Stickel F, Buch S, Nischalke HD, Weiss KH, Gotthardt D, Fischer J, Rosendahl J, et al. Genetic variants in PNPLA3 and TM6SF2 predispose to the development of hepatocellular carcinoma in individuals with alcohol-related cirrhosis. Am J Gastroenterol 2018;113:1475-1483.

16.   Liu YL, Reeves HL, Burt AD, Tiniakos D, McPherson S, Leathart JB, Allison ME, et al. TM6SF2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease. Nat Commun 2014;5:4309.

17.   Stickel F, Buch S, Lau K, Meyer zu Schwabedissen H, Berg T, Ridinger M, Rietschel M, et al. Genetic variation in the PNPLA3 gene is associated with alcoholic liver injury in caucasians. Hepatology 2011;53:86-95.

18.   Mancina RM, Ferri F, Farcomeni A, Molinaro A, Maffongelli A, Mischitelli M, Poli E, et al. A two gene-based risk score predicts alcoholic cirrhosis development in males with at-risk alcohol consumption. Appl Clin Genet 2019;12:1-10.

19.		Vespasiani-Gentilucci U, Gallo P, Dell'Unto C, Volpentesta M, Antonelli-Incalzi R, Picardi A. Promoting genetics in non-alcoholic fatty liver disease: Combined risk score through polymorphisms and clinical variables. World J Gastroenterol 2018;24:4835-4845.

20.		Stender S, Kozlitina J, Nordestgaard BG, Tybjaerg-Hansen A, Hobbs HH, Cohen JC. Adiposity amplifies the genetic risk of fatty liver disease conferred by multiple loci. Nat Genet 2017;49:842-847.

21.		Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015;12:e1001779.

22.		Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018;562:203-209.

23.		Hallager S, Ladelund S, Christensen PB, Kjaer M, Thorup Roege B, Gronbaek KE, Belard E, et al. Liver-related morbidity and mortality in patients with chronic hepatitis C and cirrhosis with and without sustained virologic response. Clin Epidemiol 2017;9:501-516.

24.		Omland LH, Jepsen P, Krarup H, Christensen PB, Weis N, Nielsen L, Obel N, et al. Liver cancer and non-Hodgkin lymphoma in hepatitis C virus-infected patients: results from the DANVIR cohort study. Int J Cancer 2012;130:2310-2317.

25.		Gjerstorff ML. The Danish Cancer Registry. Scand J Public Health 2011;39:42-45.

26.		http://mccarthy.well.ox.ac.uk/static/software/scattershot/. In.

27.		Jepsen P, Vilstrup H, Lash TL. Development and validation of a comorbidity scoring system for patients with cirrhosis. Gastroenterology 2014;146:147-156; quiz e115-146.

28.		Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, Januel JM, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. Am J Epidemiol 2011;173:676-682.

29.		Firth D. Overcoming the reference category problem in the presentation of statistical models. Sociological Methodology 2003;23:1-18.

30.		Groothuis-Oudshoorn SvBK. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software 2011;Volume 45, Issue 3:1-67.

31.		Paternoster L, Tilling K, Davey Smith G. Genetic epidemiology and Mendelian randomization for informing disease therapeutics: Conceptual and methodological challenges. PLoS Genet 2017;13:e1006944.

32.		Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, Tyrer JP, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. Am J Hum Genet 2019;104:21-34.

33.		Inouye M, Abraham G, Nelson CP, Wood AM, Sweeting MJ, Dudbridge F, Lai FY, et al. Genomic risk prediction of coronary artery disease in 480,000 adults: Implications for primary prevention. J Am Coll Cardiol 2018;72:1883-1893.

34.		Speliotes EK, Yerges-Armstrong LM, Wu J, Hernaez R, Kim LJ, Palmer CD, Gudnason V, et al. Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. PLoS Genet 2011;7:e1001324.

35.		Wang Y, Kory N, BasuRay S, Cohen JC, Hobbs HH. PNPLA3, CGI-58, and inhibition of hepatic triglyceride hydrolysis in mice. Hepatology 2019;69:2427-2441.

36.		BasuRay S, Wang Y, Smagris E, Cohen JC, Hobbs HH. Accumulation of PNPLA3 on lipid droplets is the basis of associated hepatic steatosis. Proc Natl Acad Sci U S A 2019;116:9521-9526.

37.		Smagris E, Gilyard S, BasuRay S, Cohen JC, Hobbs HH. Inactivation of TM6SF2, a gene defective in fatty liver disease, impairs lipidation but not secretion of very low density lipoproteins. J Biol Chem 2016;291:10659-10676.

38.		Smagris E, BasuRay S, Li J, Huang Y, Lai KM, Gromada J, Cohen JC, et al. Pnpla3I148M knockin mice accumulate PNPLA3 on lipid droplets and develop hepatic steatosis. Hepatology 2015;61:108-118.

39.		Su W, Wang Y, Jia X, Wu W, Li L, Tian X, Li S, et al. Comparative proteomic study reveals 17beta-HSD13 as a pathogenic protein in nonalcoholic fatty liver disease. Proc Natl Acad Sci U S A 2014;111:11437-11442.

40.   Fleming KM, Aithal GP, Card TR, West J. The rate of decompensation and clinical progression of disease in people with cirrhosis: a cohort study. Aliment Pharmacol Ther 2010;32:1343-1350.
41.   Vestberg K, Thulstrup AM, Sorensen HT, Ottesen P, Sabroe S, Vilstrup H. Data quality of administratively collected hospital discharge data for liver cirrhosis epidemiology. J Med Syst 1997;21:11-20.

Author names in bold designate shared co-first authorship.

Figure legends

Figure 1. Plasma alanine transaminase by genetic risk score in the Copenhagen Studies and in the UK Biobank. The box plots depict medians and interquartile ranges, and the whiskers extend to the 5th and 95th percentiles. P-value by linear regression.

Figure 2. Plasma ALT as a function of body mass index, alcohol intake, or diabetes, stratified by genetic risk score. The lines in the BMI and alcohol-panels depict regression lines, and the light shading show the 95% confidence intervals. The box plots in the diabetes plots depict medians and interquartile ranges, and the whiskers extend to the 5th and 95th percentiles. The ALT-increasing effect of a higher genetic risk score was amplified by increasing adiposity, alcohol intake, and diabetes mellitus in the Copenhagen Studies and UK Biobank (all P values for interaction <0.001).

Figure 3. Risk of liver cirrhosis or hepatocellular carcinoma by genetic risk score in the Copenhagen Studies and in the UK Biobank, and in the combined cohorts. Odds ratios were calculated by logistic regression. Results from the Copenhagen Studies and UK Biobank were meta-analyzed using a fixed-effects model. Error bars are 95% confidence intervals.

Figure 4. Cumulative incidence of hepatocellular carcinoma (A) or mortality (B) in patients with cirrhosis, and of mortality in patients with hepatocellular carcinoma (C), stratified by genetic risk score. A) Individuals were followed prospectively from the time of cirrhosis diagnosis until development of hepatocellular carcinoma, death, or end of follow-up. Death due to other causes was entered as a competing risk in the model. B) Participants were followed prospectively

from the time of cirrhosis diagnosis until death or end of follow-up. C) Individuals were followed prospectively from the time of hepatocellular carcinoma until death or end of follow-up. Cumulative incidences are Aalen-Johansen estimates. Hazard ratios were calculated by Cox regression, adjusted for sex, age at diagnosis, and diabetes and with time since diagnosis as the underlying time scale.

Table. Baseline characteristics.

Copenhagen Studies	Controls	Cirrhosis	HCC
N	110,219	478	103
Male (%)	49,287 (45)	312 (65)**	72 (70)**
Age, years	58 (48 – 68)	63 (56 – 70)**	67 (61 - 73)**
Body mass index, kg/m²	26 (23 – 28)	27 (24 -30)**	29 (25 – 33)**
Diabetes mellitus (%)	6278 (5.7)	119 (24.9)**	40 (38.8)**
Alcohol intake, grams per week	96 (36 – 180)	168 (48 – 372)**	180 (60 – 324)**
UK Biobank			
N	334,276	339	101
Male (%)	154,277 (46)	254 (75)**	71 (70)**
Age, years	58 (51 – 63)	61 (54 – 64)**	64 (59 – 66)**
Body mass index, kg/m²	27 (24 – 30)	29 (25 – 33)**	29 (25 – 31)*
Diabetes mellitus (%)	10,006 (3.0)	89 (26.3)**	26 (25.7)**
Alcohol intake, grams per week	84 (36 – 168)	48 (0 – 228)**	108 (36 – 240)

Values are numbers and (percentage) for categorical traits, or medians and (interquartile ranges) for continuous traits. P-values are by χ^2- test for categorial traits and by Kruskal-Wallis Rank Sum test for continuous traits both with controls as reference. Alcohol consumer is defined as at least once monthly self-reported alcohol intake. *P < 0.05 **P < 0.001. HCC: Hepatocellular carcinoma.
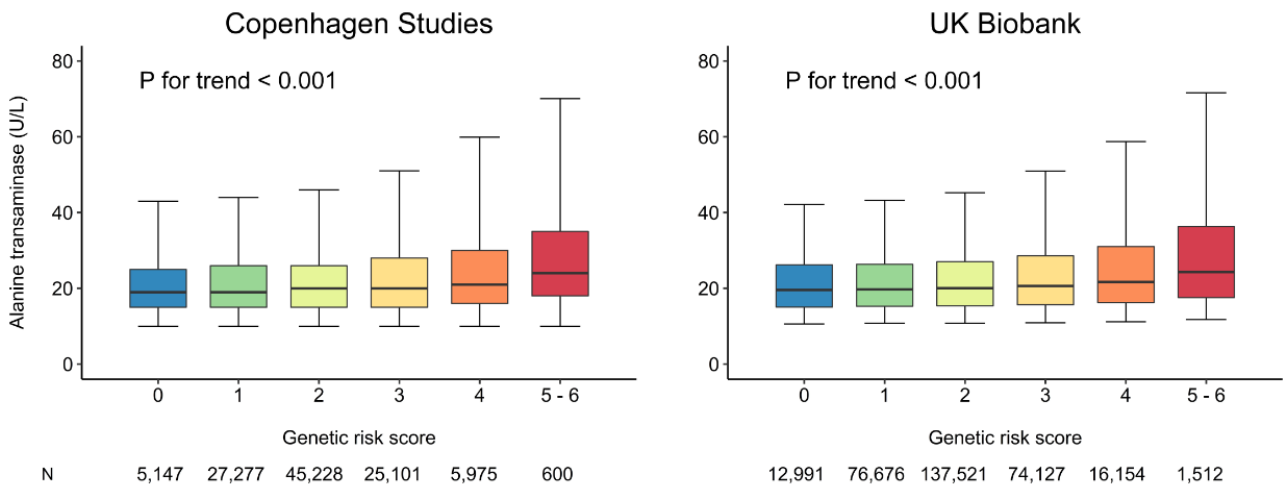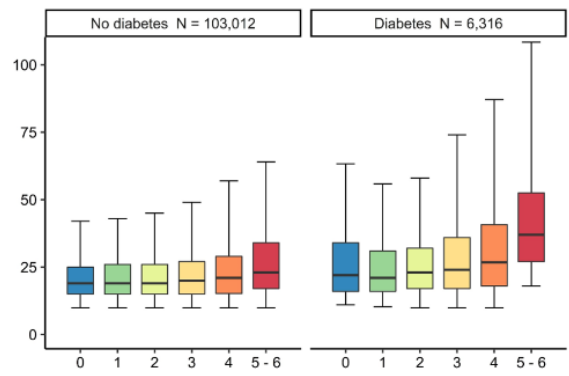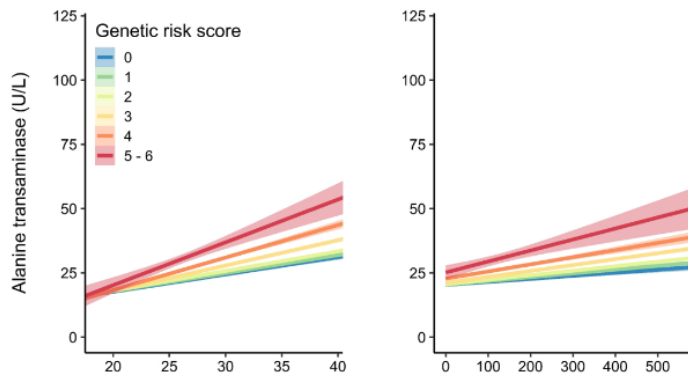
Figure 1.
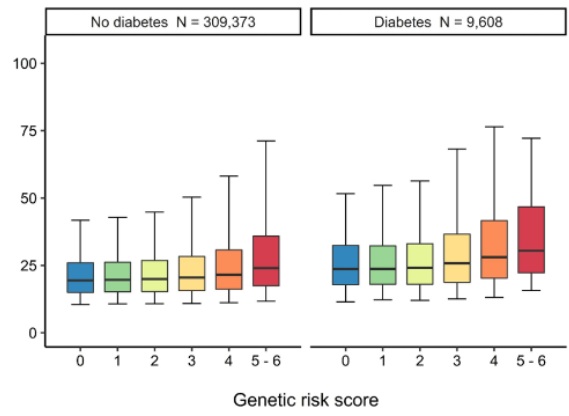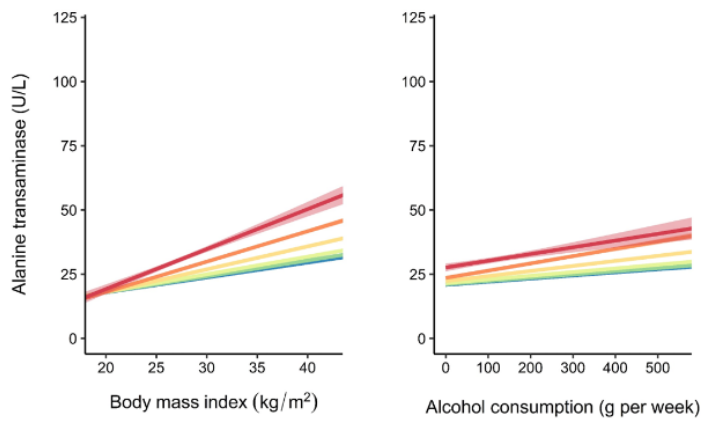
Figure 2.

Figure 3.

Figure 4.