Journal Pre-proof

Genome-wide Association Study for Alcohol-related Cirrhosis Identifies Risk Loci in *MARC1* and *HNRNPUL1*.

Hamish Innes, Stephan Buch, Sharon Hutchinson, Indra Neil Guha, Joanne R. Morling, Elleanor Barnes, Will Irving, Ewan Forrest, Vincent Pedergnan, David Goldberg, Esther Aspinall, Stephan Barclay, Peter Hayes, John Dillon, Hans Dieter Nischalke, Philipp Lutz, Ulrich Spengler, Janett Fischer, Thomas Berg, Mario Brosch, Florian Eyer, Christian Datz, Sebastian Mueller, Teresa Peccerella, Pierre Deltenre, Astrid Marot, Michael Soyka, Andrew McQuillin, Marsha Y. Morgan, Jochen Hampe, Felix Stickel

PII: S0016-5085(20)34767-3
DOI: https://doi.org/10.1053/j.gastro.2020.06.014
Reference: YGAST 63537

To appear in: *Gastroenterology*
Accepted Date: 5 June 2020

Please cite this article as: Innes H, Buch S, Hutchinson S, Guha IN, Morling JR, Barnes E, Irving W, Forrest E, Pedergnan V, Goldberg D, Aspinall E, Barclay S, Hayes P, Dillon J, Nischalke HD, Lutz P, Spengler U, Fischer J, Berg T, Brosch M, Eyer F, Datz C, Mueller S, Peccerella T, Deltenre P, Marot A, Soyka M, McQuillin A, Morgan MY, Hampe J, Stickel F, Genome-wide Association Study for Alcohol-related Cirrhosis Identifies Risk Loci in *MARC1* and *HNRNPUL1*., *Gastroenterology* (2020), doi: https://doi.org/10.1053/j.gastro.2020.06.014.

DISCOVERY ANALYSIS (n=35 839)

6.2M variants

50 variants

At risk drinkers in biobank uk

Included in  against GWAS

SERUM MARKERS OF LIVER INJURY

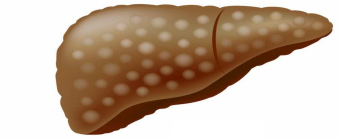
FIB4 APRI Forns

ALT AST

PHASE 1 REPLICATION (n=2 545)

50 variants

9 variants

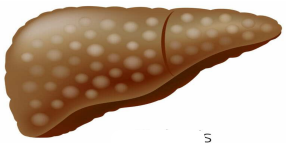


ALCOHOL LIVER CIRRHOSIS STATUS IN **THREE** SEPARATE COHORTS

PHASE 2 REPLICATION (n= 2 068)

9 variants

6 variants



ALCOHOL LIVER CIRRHOSIS STATUS IN **TWO** FURTHER COHORTS

in PNPLA3; HSD17B13; ***MARC1; HNRNPUL1;*** *SERPINA1; TM6SF2*

Genome-wide Association Study for Alcohol-related Cirrhosis Identifies Risk Loci in *MARC1* and *HNRNPUL1*.

Hamish Innes[1,2,3*], Stephan Buch[4*], Sharon Hutchinson[1,3], Indra Neil Guha[5], Joanne R Morling[2,5], Elleanor Barnes[6], Will Irving[5], Ewan Forrest[7], Vincent Pedergnan[8], David Goldberg[1,3], Esther Aspinall[1,3], Stephan Barclay[7], Peter Hayes[9]; John Dillon[10], Hans Dieter Nischalke[11], Philipp Lutz[11], Ulrich Spengler[11], Janett Fischer[12], Thomas Berg[12], Mario Brosch[4], Florian Eyer[13], Christian Datz[14], Sebastian Mueller[15], Teresa Peccerella[15], Pierre Deltenre[16], Astrid Marot[16], Michael Soyka[17], Andrew McQuillin[18], Marsha Y Morgan[19&], Jochen Hampe[4&], Felix Stickel[20&]

[1] School of Health and Life Sciences; Glasgow Caledonian University. Glasgow UK.

[2] Division of Epidemiology and Public Health, University of Nottingham, Nottingham, UK.

[3] Health Protection Scotland, Glasgow, UK.

[4] Medical Department 1, University Hospital Dresden, TU Dresden, Germany;

[5] NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham, Nottingham, UK.

[6] Peter Medawar Building for Pathogen Research, Nuffield Department of Medicine and the Oxford NIHR Biomedical Research Centre, Oxford University, UK.

[7] Glasgow Royal Infirmary, Glasgow, UK.

[8] Laboratoire MIVEGEC (UMR CNRS 5290, UR IRD 224, UM), Montpellier, France.

[9] Royal Infirmary Edinburgh, Edinburgh, UK.

[10] School of Medicine, University of Dundee, Dundee, UK.

[11] Department of Internal Medicine I, University of Bonn, Bonn, Germany;

[12] Division of Hepatology, Clinic and Polyclinic for Gastroenterology, Hepatology, Infectious Diseases and Pneumology, University Clinic Leipzig, Germany

[13] Department of Clinical Toxicology, Klinikum Rechts der Isar, Technical University of Munich, Germany.

[14] Department of Internal Medicine, Hospital Oberndorf, Teaching Hospital of the Paracelsus Private Medical University of Salzburg, Oberndorf, Austria.

[15] Department of Internal Medicine and Center for Alcohol Research, Salem Medical Center University Hospital Heidelberg, Heidelberg, Germany.

[16] Division of Gastroenterology and Hepatology, Centre Hospitalier Universitaire Vaudois, University of Lausanne, Switzerland.

[17] Department of Psychiatry, Ludwig-Maximilian University of Munich & Dept. of Psychiatry, Meiringen Hospital, Switzerland.

[18] Molecular Psychiatry Laboratory, Division of Psychiatry, University College London, United Kingdom

[19] UCL Institute for Liver & Digestive Health, Division of Medicine, Royal Free Campus, University College London, United Kingdom.

[20] Department of Gastroenterology and Hepatology, University Hospital of Zurich, Switzerland

*HI and SB have contributed equally to this work and thus share premier authorship

[&]JH, MYM and FS have contributed equally to this work and therefore share senior authorship

SHORT TITLE**:** GWAS in alcohol-related cirrhosis

MANUSCRIPT INFORMATION: Word count= 7,000 (including abstract, figure legends, table legends and references); Number of Tables= 3; Number supplementary tables =10; Number of figures= 4; Number of supplementary figures=10

CORRESPONDENCE: Address correspondence to: Dr Hamish Innes; Glasgow Caledonian University; Cowcaddens Road; G40BA; Glasgow, UK. Hamish.Innes@gcu.ac.uk

AUTHOR CONTRIBUTIONS: Specific author contributions are outlined in the table below.

Author	**CONTRIBUTIONS**								
	Conceptualisation	Data curation	Formal analysis	Funding acquisition	Study design & Methods	Resources	Supervision	Writing original draft	Writing, review and editing
HI	**X**	**X**	**X**		**X**	**X**		**X**	**X**
SB	**X**	**X**	**X**		**X**	**X**	**X**	**X**	**X**
SH				**X**	**X**	**X**	**X**		**X**
ING					**X**	**X**	**X**		**X**
JRM					**X**	**X**	**X**		**X**
EB				**X**	**X**	**X**	**X**		**X**
WI					**X**	**X**	**X**		**X**
EF					**X**				**X**
VP		**X**			**X**				**X**
DG					**X**				**X**
EA					**X**				**X**
STB					**X**				**X**
PH					**X**				**X**
JD					**X**				**X**
HDN		**X**			**X**				**X**
PL		**X**			**X**				**X**
US		**X**			**X**				**X**
JF		**X**			**X**				**X**
TB		**X**			**X**				**X**
MB			**X**		**X**				**X**
FE		**X**			**X**				**X**
CD		**X**			**X**				**X**
SM		**X**			**X**				**X**
TP		**X**			**X**				**X**
PD		**X**			**X**				**X**
AM		**X**			**X**				**X**
MS		**X**			**X**				**X**
AM		**X**			**X**				**X**
MYM	**X**	**X**			**X**	**X**	**X**	**X**	**X**
JH	**X**	**X**		**X**	**X**	**X**	**X**	**X**	**X**
FS	**X**	**X**		**X**	**X**	**X**	**X**	**X**	**X**

CONFLICTS OF INTEREST: There are no conflicts of interest to disclose.

FUNDING: This work was supported by grants from the Swiss National Funds (SNF no. 310030_169196) and the Swiss Foundation for Alcohol Research (SSA) to FS, and the Liver Systems Medicine (LiSyM) Network funded by the German Federal Ministry for Education and Research (BmBF) to JH. HI's time was supported by: (i) core research funding monies from Glasgow Caledonian University, and (ii) the STOPHCV study, which is led by EB and is supported by a grant from the UK Medical Research Council ((MR/K01532X/1: STOP-HCV). The STOPHCV study

funded access to the Archie West high performance computer platform. The study funders played no role in the: study design, collection of data, analysis of data, or interpretation of data.

ABBREVIATIONS (in order of appearance): United Kingdom Biobank (UKB); Genome-Wide Association Study (GWAS); Genetic Risk Score (GRS); False Discovery Rate (FDR); Body Mass Index (BMI); Mitochondrial Amidoxime Reducing Component 1 *(MARC1)*; Heterogeneous Nuclear RiboNucleoProtein U Like 1 *(HNRNPUL1)*; Single Nucleotide Polymorphism (SNP); Alanine transaminase (ALT); Aspartate transaminase (AST); Aspartate Aminotransferase Platelet Ratio Index (APRI), Fibrosis-4 index (FIB4); FUnctional Mapping and Annotation of genome-wide association studies tool (FUMA); International Classification of Disease (ICD); Odds Ratio (OR); Genotype-Tissue Expression (GTEx); Combined Annotation Dependence Depletion (CADD); Polymorphism Phenotyping version 2 (PolyPhen-2); Messenger RNA (mRNA); Non-alcoholic Fatty Liver Disease (NAFLD); Adjusted Odds Ratio (aOR); Confidence Interval (CI); Hazard Ratio (HR); Transforming Growth Factor Beta 1 *(TGFB1);* Coiled-coil domain containing 97 *(CCDC97);* CEA Cell adhesion molecule 21 *(CEACAM21);* Chromosome 1 open reading frame 115 *(C1orf115)*; Glucocorticoid Response Element *(GRE)*; Hepatocyte Nuclear Factor 1 Homeobox A *(HNF1A);* TRIBbles homolog 1 *(TRIB1).*

ACKNOWLEDGEMENTS:

This research has been conducted using the UK Biobank Resource. Application number: 8764. We would like to also acknowledge the support of the Archie West High Performance Computer team at University of Strathclyde.

ABSTRACT

BACKGROUND & AIMS

Little is known about genetic factors that affect development of alcohol-related cirrhosis. We performed a genome-wide association study (GWAS) of samples from the United Kingdom Biobank (UKB) to identify polymorphisms associated with risk of alcohol-related liver disease.

METHODS

We performed a GWAS of 35,839 participants in the UKB with high intake of alcohol against markers of hepatic fibrosis (FIB-4, APRI and Forns index scores) and hepatocellular injury (levels of aminotransferases). Loci identified in the discovery analysis were tested for their association with alcohol-related cirrhosis in 3 separate European cohorts (phase 1 validation cohort; n=2545). Variants associated with alcohol-related cirrhosis in the validation at a false-discovery rate of less than 20% were then directly genotyped in 2 additional European validation cohorts (phase 2 validation, n=2068).

RESULTS

In the GWAS of the discovery cohort, we identified 50 independent risk loci with genome-wide significance ($P<5 \times 10^{-8}$). Nine of these loci were significantly associated with alcohol-related cirrhosis in the phase 1 validation cohort; 6 of these 9 loci were significantly associated with alcohol-related cirrhosis in phase 2 validation cohort, at a false discovery rate below 5%. The loci included variants in the mitochondrial amidoxime reducing component 1 gene (*MARC1*) and the heterogeneous nuclear ribonucleoprotein U like 1 gene (*HNRNPUL1*). After we adjusted for age, sex, body mass index, and type-2 diabetes in the phase 2 validation cohort, the minor A allele of *MARC1*:rs2642438 was associated with reduced risk of alcohol-related cirrhosis (adjusted odds ratio, 0.76; *P*=.0027);

conversely the minor C allele of *HNRNPUL1*:rs15052 was associated with an increased risk of alcohol-related cirrhosis (adjusted odds ratio, 1.30; *P*=.020).

CONCLUSIONS

In a GWAS of samples from the UKB, we identified and validated (in 5 European cohorts) single-nucleotide polymorphisms that affect risk of alcohol-related cirrhosis in opposite directions: the minor A allele in *MARC1*:rs2642438 decreases risk whereas the minor C allele in *HNRNPUL1*:rs15052 increases risk.

KEYWORDS

biomarker, prognostic factor, SNP, hepatic fibrogenesis

INTRODUCTION

Alcohol-related cirrhosis causes an estimated 350,000 deaths every year from chronic liver failure[1] and is a major risk factor for hepatocellular carcinoma, the 3rd leading cause of cancer mortality worldwide.[2] Current therapies to prevent or retard progression to alcohol-related cirrhosis are limited, and center around reducing alcohol intake, either through behavioral or pharmacological interventions.[3] The transition from a healthy liver to alcohol-related cirrhosis occurs gradually, alongside years of sustained heavy alcohol use and concomitant chronic liver injury. Total volume of alcohol consumed and alcohol drinking patterns are strongly associated with risk of alcohol-related cirrhosis,[4,5] but they do not fully explain why some drinkers develop this outcome (<10%) whereas others do not.[6] Twin studies suggest that there is a heritable component to alcohol-related liver disease,[7,8] yet genome-wide association studies (GWAS) undertaken to date have identified only a handful of specific risk variants, including *PNPLA3*:rs738409; *TM6SF2*:rs58542926; *MBOAT7*:rs641738 and *HSD17B13*:rs72613567.[9-12]

Two factors are likely to have limited the yield of GWAS studies for alcohol-related cirrhosis (and chronic liver disease in general).[9-11] Firstly, the lack of statistical power, and secondly, the limited range of endophenotypes employed in discovery analyses. The development of alcohol-related cirrhosis is strongly underpinned by fibrogenesis, a process that causes substitution of the liver parenchyma with non-functional mesenchymal scar tissue.[13,14] Thiele et al. recently showed that combinations of routine liver blood tests – such as APRI, FIB-4, and Forns index – can differentiate individuals with high alcohol intake and advanced fibrosis from individuals with high alcohol intake without relevant fibrosis with acceptable accuracy (area under the curve of 0.80-0.86).[15] These endophenotypes of alcohol-related cirrhosis have not been leveraged by GWAS studies thus far.

The United Kingdom Biobank (UKB) study integrates host genetic, health behavior, hospital admission records, mortality, and biochemistry data for a cohort of half a million people in the UK aged 40-69 years,[16] and thus provides an unprecedented opportunity to extend existing GWAS data by incorporating a broader set of surrogate phenotype data. Data on fibrosis markers and markers of hepatocellular damage are available for a high proportion of participants.

The primary aim of this study was to identify novel risk variants associated with alcohol-related cirrhosis risk. To this end, we undertook a GWAS in UKB participants reporting high alcohol intake to identify genetic variants associated with surrogate measures of liver fibrosis and conventional markers of hepatocellular injury. Any significantly associated variants were tested for association with alcohol-related cirrhosis across appropriate European cohorts with robust liver-specific phenotypes to validate the findings.

A secondary objective was to assess whether the genetic risk variants identified by this study have any value regarding stratification of "at risk" patients in a community setting. This is highly relevant because chronic liver disease is frequently not diagnosed until decompensated cirrhosis and/or HCC emerges, by which point, liver damage is usually intractable and prognosis bleak.[17,18] Early identification of "at risk" patients is seen as a critical step towards reducing liver mortality; yet existing risk stratification tools are suboptimal.[17,18] Promising Genetic Risk Scores (GRS) have been developed for a variety of diseases including type 2 diabetes, coronary artery disease and inflammatory bowel disease,[19] but the utility of a liver cirrhosis GRS remains unclear at present.

METHODS

DISCOVERY ANALYSIS

At enrolment, UKB participants were asked to report their average alcohol intake per week/month in terms of the number of: glasses of red wine (Field IDs: 1568, 4407), glasses of champagne/white wine (UKB Field IDs: 1578, 4418), pints of beer/cider (Field IDs: 1588, 4429), measures of spirits (Field IDs: 1598, 4440), glasses of fortified wine (Field IDs: 1608, 4451), and glasses of "other" types of

alcoholic drinks (Field IDs: 5364, 4462). This was converted into the average units of alcohol consumed per week, assuming there are 2 units (16g) of pure alcohol in a pint of beer/cider; 1.5 units (12g) in a glass of red wine, champagne, white wine, fortified wine, and "other" alcoholic drink; and 1 unit (8g) in a measure of spirits. These conversions are comparable to those used in the Health Survey for England methods protocol.[20]

The discovery analysis was based on data from the Version 3 release of the UKB imputed genetic dataset (downloaded in May 2019), which provides host genetic information for 487,320 participants.[21] Participants with poor quality genetic data, as indicated by extreme levels of heterozygosity or missing data (see UKB Field ID:22027) were not included in the version 3 imputation file, and thus did not feature in this study. From the imputed dataset, we excluded participants if they were: i) first or second-degree relatives of another participant. This was inferred via the kinship coefficient, generated by the UKB core team for all pairs of participants. Specifically, a second-degree relation or greater was defined as a kinship coefficient ≥0.1;[21] or ii) not of Caucasian British ancestry (defined by UKB according to field ID:22006). However, the small number of UKB participants with gender-sex mismatch or aneuploidy (n<1000; <0.2%), were not excluded from our analysis due to the low level of potential bias they may exert. Of those remaining, women who reported alcohol consumption ≥ 25 units/week (200g) in "an average week", and men who reported alcohol consumption of ≥36 units/week (288 g), were included in the discovery GWAS. These thresholds represent the midpoint between "hazardous" and "harmful" drinking, as set out in UK Government guidelines.[22]

Individual-level data for approximately 6.2M genetic variants were available in the version 3 UKB imputed genetic dataset, after exclusion of variants with: a) minor allele frequency <1%; b) gross deviation from the Hardy Weinberg equilibrium ($p<1.0 \times 10^{-7}$); c) imputation information score <0.8; d) high level missing data (>10%); and e) non-biallelic or duplicate variants. Using PLINK v1.9, we determined the association between each of these 6.2M variants, and five distinct surrogate liver phenotypes. These were: APRI; FIB-4; Forns Index (all defined using standard formulas; see Appendix A); ALT and AST. Phenotypes were log10 transformed to achieve approximate normality and were analysed as continuous variables in a linear regression framework assuming an additive

genetic model (see Supplementary Table 1). In addition, sensitivity analyses were undertaken where we analysed each surrogate phenotype as a categorical variable, comparing participants whose phenotype value was in the top quintile (i.e. top 20%) with participants whose phenotype value was in the bottom quintile (i.e. bottom 20%) via logistic regression. This is tantamount to an "extreme phenotyping" approach. All discovery analyses were adjusted for age, sex, BMI, diagnosis of diabetes, current alcohol consumption and the first five principal components of genetic ancestry. The analyses were performed on the University of Strathclyde's Archie West High Performance Computing platform.

The resultant GWAS summary statistics were then uploaded onto the FUnctional Mapping and Annotation of GWAS tool (FUMA; version 1.3.5),[23] in order to identify a set of independent genomic risk loci for each phenotype. In broad terms, independent genomic risk loci are defined by three main characteristics: a) association with the corresponding phenotype at genome wide significance level ($P<5.0 \times 10^{-8}$); b) a lower p-value than all other variants in the nearby genomic region; c) independence from other independent genomic risk loci for that phenotype (at $r2<0.1$). More detailed specifics around how these loci were selected by FUMA are outlined in Appendix B.

In total, the base-case discovery analysis generated five sets of independent genomic risk loci - one per phenotype - which were then pooled to produce a final combined set of loci. Duplicate variants were removed.

PHASE 1 REPLICATION ANALYSES

The associations between each independent genomic risk locus (identified in our discovery cohort) and the presence of alcohol-related liver cirrhosis, was ascertained in three case-control datasets; as follows:

1) ***UK cohort from the Buch et al. GWAS [9]:*** comprises a) 302 cases with alcohol-related cirrhosis recruited at the Centre for Hepatology, The Royal Free Hospital, London; and b) 346 controls with a history of excess alcohol consumption but without evident liver disease. All participants were of European Caucasian descent. Genotyping was performed using the Illumina BeadChip array.

Cirrhosis was defined through clinical evidence (complications of cirrhosis), imaging results (ascites, hunched liver surface, elastography >19kPa indicating cirrhosis) and/or histology, as described by Buch *et al.*[9] Full details of the criteria used to define cases and controls can be found in appendix C.

2) ***German cohort from the Buch et al. GWAS [9]:*** comprises a) 410 cases diagnosed with alcohol-related cirrhosis recruited from several Gastroenterology and Hepatology hospitals in Germany, Austria and Switzerland; and b) 1080 controls, recruited from psychiatric centres in Germany and Switzerland specialising in addiction medicine, who had a history of excess alcohol use, but were without evident liver disease *(vide supra).* All cases and controls in this cohort are of Caucasian ancestry. Genotyping was again performed using the Illumina BeadChip array. Cirrhosis was again defined according to the diagnostic criteria set out by Buch et al.[9] (see appendix C)
3) ***UK Biobank nested case-control study***: Cases were defined as UKB participants with a) two or more hospital admission for alcohol-related cirrhosis before or after inclusion in the UKB study (defined as ICD 10: K70.3 in any diagnostic position); or b) death from alcohol-related cirrhosis (defined as ICD10: K703 in any cause-of-death position). Controls were defined as all participants who did not indicate that they were lifetime teetotal (UKB field IDs: 3731), and who did not have a hospital admission or death record indicating liver disease (ICD 10: K70-K77). As per our discovery analysis, we excluded cases/controls if they were: i) first or second degree relative with another UKB participant (kinship coefficient≥0.1); or ii) not of Caucasian British ancestry. Furthermore, to ensure that our discovery analysis and phase 1 replication were independent (i.e. non-overlapping), we excluded cases/controls if they were also included in our discovery analysis. In total, 178 cases and 298,248 controls satisfied these criteria, equating to an effective sample size of approximately 712.

Odds ratio (OR) associations were ascertained in the three cohorts, for each independent genomic locus identified in our discovery analysis, under an additive genetic model. All associations were adjusted for differences in age, sex and the first five principal components of genetic ancestry. A fixed-effect meta-analysis of the OR association across the three datasets, was then performed. This

was carried out using METAL's inverse variance function, which weights each effect size estimate by its estimated standard error, in order to provide an overall p-value and effect size.[24] To remove residual linkage disequilibrium created by combining the genomic risk loci of different phenotypes, we performed a standard clumping procedure using an r2 threshold >0.1, and distance parameter of 100KB. For each genomic region, this resulted in retention of the locus with the lowest meta-analysis p-value. Finally, we applied the Benjamini-Hochberg procedure to correct for multiple comparisons.[25] A False Discovery Rate (FDR) of 20% was used to select variants for phase 2 replication.

PHASE 2 REPLICATION ANALYSES

Loci identified in the phase 1 replication analyses were then assessed in two separate replication cohorts, as follows: i) German validation cohort comprising 1,272 cases with alcohol-related cirrhosis and 775 controls without liver disease (effective sample size: 1,926); and ii) Switzerland validation cohort comprising 312 individuals with alcohol-related cirrhosis and 40 individuals without (effective sample size: 142). Analogous to phase 1 replication, all participants in this cohort had a history of heavy alcohol use and were recruited from specialist liver and addiction clinics. Cirrhosis was defined according to the same diagnostic criteria described by Buch et al.[9] (See appendix C). Genotyping for selected loci was undertaken using the TaqMan assay system. The OR association between each locus and case-control status was determined following adjustment for age and sex in the two phase 2 cohorts separately, under an additive allelic effect model. We also performed more extensive adjustment to account for potential differences in BMI and type 2 diabetes, as well as age and sex (albeit with a reduced sample size). Consistent with phase 1 replication, we then performed a fixed-effect meta-analysis of the OR association across the two phase 2 datasets, using METAL's inverse variance function.[24] Meta-analysis p-values were calculated using a conservative two-tailed test, making no assumption about the direction of association. A stringent FDR threshold of <5% was used at this final stage to define statistical significance.

FUNCTIONAL CONSEQUENCES OF PHASE 2 LOCI

BIOINFORMATIC DATABASE INTERROGATION

The ANNOVAR annotation database was used to identify the nearest gene for each locus.[26] However, the nearest gene may not be causal; previous studies show that a locus can sometimes affect the expression and function of genes that are considerably apart on the same chromosome.[27] Thus, in addition to the nearest gene, a broader set of candidate genes were identified for each locus using four approaches *viz*. (i) positional mapping was used to identify all genes within 10KB of each locus, or within 10KB of variants in linkage disequilibrium (r2>0.60) with each locus; (ii)The Gene-Tissue Expression (GTEx) database version 8 [28] was used to identify any genes whose expression is associated with the locus, or genes whose splicing patterns are associated with the locus (at FDR $<1.0 \times 10^{-3}$); (iii) the HiC(GSE87112) liver tissue dataset was used to identify genes that, although physically far apart in terms of their chromosome base-pair positions, may nevertheless interact with the locus – e.g. via chromatin looping (at FDR $<1.0 \times 10^{-6}$);[29] and (iv) the GeneHancer database was used to assess whether the locus lies within a genomic enhancer region, and if so, we enumerated the gene targets of this enhancer.[30]

The Combined Annotation Dependence Depletion (CADD) score [31] and Regulome DB score [32] was also determined for each locus. Finally, for nonsynonymous coding variants, we also determined: (i) the predicted functional impact on the corresponding protein using the Polymorphism Phenotyping version 2 (PolyPhen-2) HumDiv-trained model.[33]; and (ii) if the corresponding amino-acid residue is conserved in mammalian and non-mammalian orthologues. Protein sequences for gene orthologues were obtained from ENSEMBL.org and were aligned using T-Coffee tool kit.[34]

mRNA EXPRESSION ANALYSIS

One hundred and eleven liver tissue samples, collected in a previous study,[35] were used to assess mRNA expression of selected candidate genes in liver tissue, according to identified host genotyping factors. Liver tissue samples were obtained percutaneously for patients undergoing liver biopsy for suspected Non-alcoholic Fatty Liver Disease (NAFLD) (57%, n=63) or intraoperatively during bariatric surgery for assessment of liver histology (43%, n=48). mRNA expression levels were

measured using the Human Gene 1.1 ST Array. The non-parametric Kruskal-Wallis test was used to assess the likelihood of whether differences in mRNA expression according to host genotyping factors were due to sampling error.

CIRRHOSIS GENETIC RISK SCORE

A cirrhosis GRS was generated based on the genetic variants considered in the phase 2 replication analysis. The GRS was calculated as follows:

Cirrhosis GRS $=\sum_{i=1}^{k} wiXi$, where k is the number of risk variants, w_i is effect size (i.e. beta) of each variant estimated from phase 1 validation stage; X_i is the number of risk alleles carried by that individual for genetic variant i. More detail is provided in Appendix D.

Conceptually, this score can be thought of as the number of risk variants each individual carries weighted by their effect size (that is, weighted by the extent to which each variant increases the risk of cirrhosis).

GRS performance was tested on UKB participants at risk of NAFLD. This subgroup is independent of the discovery analysis, and thus is a non-biased group from which to gauge performance of the cirrhosis GRS. Risk factors for NAFLD were defined as: a BMI ≥30 and/or diagnosis of type 2 diabetes, without evidence of any other cause of liver disease including excess alcohol (see Appendix E and Supplementary Table 2 for full details).

The outcome used to assess GRS performance was time to first hospital admission for cirrhosis. A hospital admission for cirrhosis was defined according to Ratib et al's validated algorithm incorporating appropriate ICD discharge codes and OPCS4 hospital procedure codes.[36] We calculated the association between GRS quintile and risk of incident cirrhosis hospitalisation using Cox regression in a survival analysis framework. Follow-up time was commenced at the date of UKB assessment and ended at first date of cirrhosis hospitalisation (if at all), date of mortality (if at all), or date of hospital/mortality registry completion. As well as assessing the association between GRS

quintiles and risk/hazard of incident cirrhosis hospitalisation, we also assessed GRS performance in terms of Harrell's C-statistic. In this context, the C-statistic reflects the probability that an individual with a higher GRS will has a shorter time to cirrhosis hospitalization versus an individual with a lower GRS. A GRS with no ability to differentiate individuals at high versus low risk of cirrhosis hospitalisation would have a C-statistic of 0.50. Conversely, a GRS with perfect ability would have a C-statistic of 1.0.[37]

The base-case GRS included risk variants significant associated with alcohol-related cirrhosis in phase 1 replication analyses at a FDR<20%. In subsequent sensitivity analyses, the number of risk variants incorporated was varied to see if performance was affected using FDRs of <10%, <30%, <40% and <50%, to see if performance was affected.

RESULTS

DISCOVERY ANALYSIS:

Of the UKB participants, 35,839 participants met the inclusion/exclusion criteria for the discovery analysis (see Figure 1). The median (interquartile range [IQR]) age was 58 years (51-63), 63% were male, 3.6% had a diagnosis of type 2 diabetes, whilst the median BMI was 27.3 (24.7-30.1) (Table 1 & Supplementary Table 3). The discovery analysis identified 68 unique genomic risk loci across the five phenotypes (Supplementary Table 4 & Supplementary Figure 1). Detailed information on these loci are provided in Supplementary Table 5. Manhattan plots for each discovery analysis are shown in Figure 2. The genomic inflation factor (λ) varied between 1.03 and 1.05 (see supplementary Figure 2).

PHASE 1 REPLICATION ANALYSIS:

The 68 loci identified in discovery analysis were reduced to 50 independent risk loci, following the phase 1 replication meta-analysis and subsequent clumping (Supplementary Table 6). Of these 50, nine loci were significantly associated with alcohol-related cirrhosis at a FDR of <20%. Four of these *viz PNPLA3:*rs738408 (in complete LD with *PNPLA3:*rs738409), *SUGP1*:rs10401969 (in strong LD

with *TM6SF2:*rs5854926), *SERPINA1*:rs28929474; and *HSD17B13*:rs7694379 - are already known to modulate the risk for alcohol-related cirrhosis, or are in strong LD with known loci. The remaining five loci have not previously been associated with alcohol-related liver disease *viz* *HNF1A*:rs11065384; *ARHGEF3*:rs12485738; rs2954038 (near the *TRIB1* gene); *HNRNPUL1*:rs15052; and *MARC1*:rs2642438. No additional significant loci were identified when using the extreme phenotyping discovery approach.

PHASE 2 REPLICATION ANALYSES:

In the independent replication cohort, six of the nine variants identified in the phase 1 replication were associated with cirrhosis at an FDR of <5% (Table 2 & Supplementary Table 7). Of these six, four are already known, or are in linkage disequilibrium with variants known to modulate cirrhosis risk (rs738408; rs10401969; rs28929474; rs7694379).

The two variants remaining were: i) *MARC1*:rs2642438 on chromosome 1; and ii) *HNRNPUL1:*rs15052 on chromosome 19. The minor A allele of rs2642438 was associated with a reduced risk of cirrhosis in age/sex adjusted (OR:0.76; 95%CI:0.65-0.89; $p=5.37 \times 10^{-4}$) and age/sex/BMI/diabetes adjusted (aOR: 0.76; 95% CI:0.64-0.91; $p=2.7 \times 10^{-3}$) analyses, suggesting a protective effect. Conversely, the minor C allele of rs15052 was associated with an increased risk of alcohol-related cirrhosis in age/sex adjusted (OR:1.24; 95% CI: 1.02-1.52; $p=3.2 \times 10^{-2}$) and age/sex/BMI/diabetes adjusted (aOR:1.30; 95%CI:1.04-1.62; $p=2.0 \times 10^{-2}$) analyses, suggesting a detrimental effect (Figure 3). Regional association plots for *MARC1*:rs2642438 and *HNRNPUL1*:rs15052 are provided in Supplementary Figures 3-7.

FUNCTIONAL CONSEQUENCES OF *MARC1* AND *HNRNPUL1* LOCI

BIOINFORMATIC DATABASE INTERROGATION:

The rs15052 risk variant lies in the 3' untranslated region of the Heterogeneous Nuclear Ribonucleoprotein U Like 1 *(HNRNPUL1;* chromosome 19), It has a CADD score and Regulome DB score of 13.94 and 5, respectively (see Supplementary Table 8). The frequency of the minor rs15052:C allele in the UKB population is 17.8%; the ancestral allele is T, which is conserved within

mammals (see Supplementary Figure 8). Of note, rs15052 lies within a 16.8KB enhancer region (GeneHancer ID: GH19J041294) that regulates expression of 27 genes, including the Transforming Growth Factor Beta 1 *(TGFB1)* gene. GTEx data indicates that rs15052:C is associated with increased expression of *TGFB1* and Coiled-Coil Domain Containing 97 *(CCDC97)* in liver and/or adipose tissue. In addition, chromatin interaction mapping indicates that rs15052 interacts physically with the CEA Cell Adhesion Molecule 21 *(CEACAM21)* gene.

The rs2642438 variant is a missense variant in the Mitochondrial Amidoxime Reducing Component 1 (*MARC1)* gene, resulting in a Alanine [GCC] to Threonine [ACC] substitution at amino acid position 165 of the MARC1 protein (A165T). The PolyPhen-2 algorithm predicts that this substitution has a deleterious impact on MARC1 protein function with a high score of 0.958 (sensitivity: 0.78; specificity: 0.95). The frequency of the minor A allele is 29.7% in the UKB population; the ancestral allele is G, which is conserved among mammals, amphibia (clawed frog; *Xenopus laevis*) and also Actinopterygii (clownfish; *Amphiprion percula*) - see Supplementary Figure 9. GTEx data suggest that rs2642438:A is associated with reduced expression of uncharacterized Chromosome 1 Open Reading Frame 115 *(C1orf115)* gene in cultured fibroblasts cells, and secondly that it is associated with alternative splicing of MARC1 pre-mRNA in adipose tissue (see Supplementary Table 8).

mRNA EXPRESSION ANALYSIS

mRNA expression levels were obtained in 113 liver tissues samples for: (i) *TGFB1*, *CEACAM21*, *CCDC97*, and *HNRNPUL1* with respect to rs15052 genotype; and (ii) *MARC1* and *C1orf115* with respect to rs2642438 genotype. The rs15052:C allele was associated with increased expression of *TGFB1* (p=0.026), *CEACAM21* (p=0.012), but not *CDC97* (P=0.52) or *HNRNPUL1* (p=0.65). The rs2642438 locus was not associated with either *MARC1* expression (p=0.95) or *C1orf115* (p=0.93); see Supplementary Figure 10.

GENETIC RISK SCORE PERFORMANCE:

A total of 107,014 UKB participants met the at-risk criteria for NAFLD. These participants were followed up for 7.9 years on average, during which time 562 incident liver cirrhosis hospitalizations were observed equating to a crude incidence rate of 6.3 events per 10,000 person years of follow-up. There was a clear dose-response relationship between the GRS quintile and the risk of cirrhosis hospitalization. Participants in the highest risk quintile had more than a 3-fold higher risk versus individuals in lowest quintile after adjustment for age, gender, BMI, diabetes and alcohol intake (HR: 3.16; 95% CI: 2.38-4.21) (see Figure 4 and Table 3). The GRS C-statistic was 0.62 (95% CI: 0.59-0.64); but increased to 0.68 (0.65-0.70) when combined with age and sex. In the sensitivity analyses, associations by quintile and the C-statistic was only modestly affected by the FDR selection (see Supplementary Table 9).

DISCUSSION

It is generally agreed that genetic factors play an important role in determining an individual's susceptibility to develop alcohol-related cirrhosis.[7,8] Only a small number of associated risk factors have been identified to date.[9-12, 38] As a complex trait, it is highly likely that additional genetic modifiers exist.[39] Uncovering these variants could help to identify new therapeutic targets for treatment, and also improve patient risk stratification. Against this backdrop, we combined data from the UKB resource with data from clinical cohorts across Europe, to produce the largest, broadest and most comprehensive GWAS on alcohol-related cirrhosis undertaken thus far. Our findings confirm the key roles played by several of the known risk loci *viz PNPLA3*:rs738409; *TM6SF2:*rs58542926; *HSD17B13*:rs72613567 and *SERPINA1*:rs28929474.[9-12, 38] However, we were not able to validate the risk variant rs641738 in *TMC4/MBOAT7* which was detected as a risk locus for alcoholic cirrhosis in our previous GWAS.[9] It was not associated with any of our discovery analysis phenotypes at the requisite genome-wide significance level, but it only narrowly missed this level in relation to ALT and AST with P=8.31 x 10^{-5} and P=3.49 x 10^{-4}, respectively (see supplementary table 10).

As well as confirming most loci known to-date, we have also discovered two additional risk variants for alcohol-related cirrhosis: *MARC1*:rs2642438, which has also been identified independently as a risk factor of cirrhosis regardless of etiology in another very recent publication,[40] and *HNRNPUL1*:rs15052 which has not been described so far. These two loci modify the risk of alcohol-related cirrhosis in opposite directions. Carriage of the minor A allele of *MARC1*:rs2642438 is associated with protection from alcohol-related cirrhosis; while carriage of the minor C allele of *HNRNPUL1:*rs15052 is associated with an increased risk of alcohol-related cirrhosis (see Figure 2). The identification of these additional loci in the present study, but not in others,[9-12] is likely to be due to: (i) high statistical power, gained by combining large discovery (N=35,839) and replication (Effective sample size: 4,599) cohorts; and (ii) the novel inclusion of fibrogenesis endophenotypes in the discovery analysis. For example, (*HNRNPUL1:*rs15052 would not have been identified in this study without including APRI or FIB-4 in the discovery analysis.

The *MARC1* gene is predominantly expressed in liver and subcutaneous adipose tissue, and the corresponding MARC1 protein is located in the outer mitochondrial membrane.[41] Mitochondrial damage is a well-described key feature of alcohol-mediated hepatocellular injury by increasing oxidative stress through the respiratory chain, and interference with beta-oxidation and lipogenesis leading to liver cell apoptosis and steatosis, respectively.[42] MARC1 protein plays an important role in reducing N-hydroxyl compounds, and in this way is involved in detoxification of xenobiotics. The crystal structure of human MARC1, its catalytic mechanism, and its ability to reduce a wide range of N-oxygenated compounds has recently been described.[43] However, its function is still incompletely understood and it is not known whether it may play a role in the metabolism of acetaldehyde, the toxic and mutagenic degradation product of alcohol oxidation. Acetaldehyde is generated from ethanol by cytosolic alcohol dehydrogenase and microsomal cytochrome P450 2E1, and further degraded to acetate in mitochondria through enzymatic conversion by aldehyde dehydrogenase.[44] MARC1 also plays a role in the regulation of nitric oxide production [45], a powerful vasodilator that alters intrahepatic vascular resistance in the liver [46]. Thus, there are many ways in which altered MARC1 function could contribute to the pathophysiology of cirrhosis. At a functional level, the protective

rs2642438:A allele results in an Alanine to Threonine amino-acid substitution at position 165 of the MARC1 protein, which PolyPhen-2 predicts is deleterious to MARC1 protein function. Thus, loss of MARC1 function appears to be beneficial with respect to lowering the risk of alcohol-related cirrhosis. On this basis, therapeutic inhibition/dampening of MARC1 function may be an interesting avenue to explore in future work if the protein can be targeted.

HNRNPUL1 protein has dual DNA and mRNA binding ability, and thus can regulate DNA transcription as well as pre-mRNA processing. In partnership with Bromodomain-containing protein 7, HNRNPUL1 can bind to the DNA glucocorticoid response element (GRE), and activate transcription.[47] The GRE is present in the promoter region of multiple genes that regulate inflammation, and its activation may be among the mechanisms through which corticoid treatments for alcohol hepatitis reduce liver inflammation.[48] Data from GTEx as well as our mRNA expression analysis, demonstrate that rs15052:C is associated with increased expression of *TGFB1* in liver tissue. *TGFB1* is a potent profibrogenic cytokine produced by mesenchymal hepatic stellate cells and portal myofibroblasts, the main effector cells involved in the production of extracellular matrix components including collagens.[49] Increased hepatic expression of *TGFB1 in vivo* leads to the emergence of prominent liver fibrosis.[50] The presence of a GRE in the promoter region of *TGFB1* may suggest that HNRNPUL1 can regulate *TGFB1* directly.[51] However, the functional basis for the association between rs15052 and *TGFB1* expression most probably relates to rs15052's position within a 16.8 Kb enhancer genomic region (enhancer ID: GH19JO41294). This enhancer interacts with 27 distinct genes, including *TGFB1,* and contains binding sites for a wide panel of transcription factors, some of which – e.g. the Aryl Hydrocarbon Receptor – are known regulators of *TGFB1* expression.[52]

We were not able to validate the rs641738 risk variant in *TMC4/MBOAT7* due to it narrowly missing the required genome-wide significance level in our discovery analysis (see supplementary table 10). Thus, although this study represents the largest and most comprehensive GWAS for alcohol cirrhosis to-date, the possibility of "false negative" results nevertheless remains. On this basis, it is likely that some common variants that influence risk of alcohol-related cirrhosis still remain undiscovered. On a related note, some variants identified in our phase 1 replication stage that did not replicate in phase 2

may still warrant further investigation because their failure to replicate may only reflect inadequate statistical power (i.e. a type 2 error) as opposed to a true null association. This includes rs10401969 in *hepatocyte nuclear factor 1 homeobox A (HNF1A)* and the rs2954038 variant near the *tribbles homolog 1* (*TRIB1)* gene. The same *TRIB1* region was identified in a recent GWAS of NAFLD, lending further credibility to the relevance of this locus.[53].

As well as providing insight into the pathobiology of alcohol-related cirrhosis, a deeper understanding of the underlying genetics could, in time, help clinicians differentiate the minority of liver disease patients at high risk of serious liver morbidity from the low risk majority. In principle therefore, host genetic data may help to increase earlier detection of chronic liver disease in high risk patients – thus addressing the issue of frequently delayed diagnosis of chronic liver disease.[17,18] A cirrhosis GRS was, therefore, developed based on variants identified in our phase 1 analysis. To test this score objectively, we were mindful of the need to assess its performance in an independent set of patients (i.e. a different set of participants from those used to develop the score in first place).[37] To that end, we examined how good this score performed at predicting first-time hospitalization for cirrhosis among UKB participants with risk factors for NAFLD, a disease that shows great overlap and multiple similarities (including the underlying host genetics) with alcohol-related liver disease.[54] Although individuals with a risk score in the top quintile had more than 3 times the risk of cirrhosis versus individuals in the lowest quintile, the C-statistic indicated that by itself, this score is unlikely to offer adequate discrimination for effective clinical decision making. Further validation in an independent population of heavy drinkers is clearly warranted.

In summary, in addition to confirming several known genetic risk factors for alcohol-related cirrhosis, this GWAS, the largest and broadest to date, has identified two further risk loci: rs2642438 in *MARC1* and rs15052 in *HNRNPUL1*. These variants, amongst others, warrant functional investigation.

REFERENCES

[1] GBD 2015 Mortality and Causes of Death Collaborators. Global, regional and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016;388:1459-1544.

[2] Ganne-Carrie N, Nahon P. Hepatocellular carcinoma in the setting of alcohol-related liver disease. J Hepatol 2019;70:284-293.

[3] European Association for the Study of the Liver. EASL Clinical Practice Guidelines: Management of alcohol-related liver disease. J Hepatol 2018;69:154-181.

[4] Roerecke M, Vafaei A, Hasan OSM, Chrystoja BR, Cruz M, Lee R, et al. Alcohol consumption and risk of liver cirrhosis: a systematic review and meta-analysis. Am J Gastroenterol 2019;114:1574-1586.

[5] Askgaard G, Gronbaek M, Kjaer MS, Tjonneland A, Tolstrup JS. Alcohol drinking patterns and risk of alcoholic liver cirrhosis: a prospective cohort study. J Hepatol 2015;62:1061-7.

[6] Mathurin P, Bataller R. Trends in the management and burden of alcoholic liver disease. J Hepatol 2015:62:S38-S46.

[7] Hrubec Z, Omenn GS. Evidence of genetic predisposition to alcoholic cirrhosis and psychosis: twin concordances for alcoholism and its biological end points by zygosity among male veterans. Alcohol Clin Exp Res 1981;5:207-215.

[8] Reed T, Page WF, Viken RJ & Christian JC. Genetic predisposition to organ-specific endpoints of alcoholism. Alcohol Clin Exp Res 1996;20:1528-1533.

[9] **Buch S, Stickel F**, Trepo E, Way M, Herrmann A, Nischalke HD, et al. A genome-wide association study confirms PNPLA3 and identifies TM6SF2 and MBOAT7 as risk loci for alcohol-relatred cirrhosis. Nat Genet 2015;47:1443-8.

[10] Abul-Husn NS, Cheng X, LiAH, Xin Y, Schutmann C, Stevis P, et al A protein-truncating HSD17B13 Variant and protection from chronic liver disease. N Engl J Med 2018;378:1096-1106.

[11] Romeo S, Kozlitina J, Xing C, Pertsemlidis A, Cox D, Pennacchio LA, et al. Genetic variation in PNPLA3 confers susceptibility to non-alcoholic fatty liver disease. Nat Genet 2008;40:1461-5.

[12] **Stickel F, Lutz P, Buch S, Nischalke HD**, Silva I, Rausch V, et al. Genetic variation in HSD17B13 reduces risk of developing cirrhosis and hepatocellular carcinoma in alcohol misusers. Hepatology 2019;doi:10.1002/hep.30996.

[13] Lee YA, Wallace MC, Friedman SL. Pathobiology of liver fibrosis: a translational success story. Gut 2015;64:830-41.

[14] Tacke F, Weiskirchen R. An update on the recent advances in antifibrotic therapy. Expert Rev Gastroenterol Hepatol 2018;2:1143-1152.

[15] Thiele M, Madsen BS, Hansen JF, Detlefsen S, Antonsen S, Krag A. Accuracy of the enhanced liver fibrosis test vs fibrotest, elastography and indirect markers in detection of advanced fibrosis in patients with alcohol liver disease. Gastroenterology 2018;154:1369-1379.

[16] Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An open access resource for identifying the causes of a wide range of complex disease of middle and old age. PLoS Med 2015;12:E1001779.

[17] Gines P, Graupera I, Lammert F, Angeli P, Caballeria L, Krag A, et al. Screening for liver fibrosis in the general population: a call for action. Lancet Gastroenterol Hepatol 2016;1:256-260.

[18] William R, Aspinall R, Bellis M, Camps-Walsh G, Cramp M, Dhawan A, et al. Addressing liver disease in the UK: a blueprint for attaining excellence in health care and reducing premature mortality from lifestyle issues of excess consumption of alcohol, obesity and viral hepatitis. Lancet 2014;384:1953-97.

[19] Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet 2018;50:1219-1224.

[20] Osborne B, Cooper V. Health Survey for England. 2017 adult health related behaviours. Version 2. 2019. ISBN: 978-1-78734-255-2.

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

[22] Department of Health. Alcohol Guidelines Review: Report from the guidelines development group to the UK Chief Medical Officers. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/545739/GDG_report-Jan2016.pdf

[23] Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun 2017;8:1826.

[24] Willer CJ, Yun Li, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190-2191.

[25] Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Statist Soc B 1995;57:289-300.

[26] Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.

[27] Deng NA, Shou H, Fan H, Yuan Y. Single nucleotide polymorphisms and cancer susceptibility. Oncotarget 2017;8:110635-110649.

[28] The GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013;45:580-585.

[29] Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoxzy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genomes. Science 2009;326:289-93.

[30] Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Stein TI, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database. 2017;1-17. doi: 10.1093/database/bax028.

[31] Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleoic Acids Res 2018;47:D886-D894.

[32] Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 2012;22:1790-1797.

[33] Adzhubei I, Jordan DM, Sunyaev SR. Predicting function effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet 2013. doi: 10.1002/0471142905.hg0720s76

[34] Zimmermann L, Stephens A, Nam SZ, Rau D, Kubler J, Lozajic M, et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J Mol Biol. 2018;430:2237-2243.

[35] Ahrens M, Ammerpohl O, von Schonfels W, Kolarova J, Bens S, Itzel T, et al. DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. Cell Metab 2013;18:296-302.

[36] Ratib S, West J, Crooks CJ, Flemming KM. Diagnosis of liver cirrhosis in England, a cohort study, 1998-2009: a comparison with cancer. Am J Gastroenterol 2014;109:190-8.

[37] Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariate prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1-73.

[38] Strnad P, Buch S, Hamesch K, Fischer J, Rosendahl J, Schmelz R, et al. Heterozygous carriage of the alpha1-antitrypsin Pi*Z variant increases the risk to develop liver cirrhosis. Gut 2019;68:1099-1107.

[39] Stickel F, Moreno C, Hampe J, Morgan MY. The genetics of alcohol dependence and alcohol-related liver disease. J Hepatol 2017;66:195-211.

[40] Emdin CA, Haas ME, Khera AV, Aragam K, Chaffin M, Klarin D, et al. A missense variant in mitochondrial amidoxime reducing component 1 gene and protection against liver disease.. PLoS Genet 2020;16:e1008629. https://doi.org/10.1371/journal.pgen.1008629

[41] Neve EPA, Kofeler H, Hendriks DFG, Nordling A, Godvadze V, Mkrtchian S, et al. Expression and function of mARC: roles in lipogenesis and metabolic activation of ximelgatran. PLoS One 2015;10:e0138487.

[42] Mansouri A, Gattolliat CH, Asselah T. Mitochondrial dysfunction and signaling in chronic liver diseases. Gastroenterology 2018;155:629-647.

[43] Kubitza C, Bittner F, Ginsel C, Havemeyer A, Clement B, Scheidig AJ. Crystal structure of human mARC1 reveals its exceptional position among eukaryotic molybdenum enzymes. Proc Natl Acad Sci USA 2018;115:11958-11963.

[44] Lu Y, Cederbaum AI. Cytochrome P450s and alcoholic liver disease. Curr Pharm Des 2018;24:1502-1517.

[45] Kotthaus J, Wahl B, Havemeyer A, Kotthaus J, Schade D, Garbe-Schonberg D, et al. Reduction of N-hydroxyl-L-arginine by the mitochondrial amidoxime reducing component (mARC). Biochem J 2011;433:383-391.

[46] Iwakiri Y, Shah V, Rockey DC. Vascular pathobiology in chronic liver disease and cirrhosis - current status and future directions. J Hepatol 2014;61:912-24.

[47] Kzhyshkowska J, Rusch A, Wolf H, Dobner T. Regulation of transcription by the heterogeneous nuclear ribonucleoprotein E1B-AP5 is mediated by complex formation with the novel bromodomain-containing protein BRD7. Biochem J 2003;371:385-93.

[48] Frijters R, Fleuren W, Toonen EJM, Tuckerman JP, Reichardt HM, van der Maaden H, et al. Prednisolone-induced differential gene expression in mouse liver carrying wild type or a dimerization-defective glucocorticoid receptor. BMC Genomics 2010;11:359.

[49] Gressner AM, Weiskirchen R, BreitKopf K, Dooley S. Roles of TGF-beta in hepatic fibrosis. Front Biosci 2002;7:d793-807.

[50] Sanderson N, Factor V, Nagy P, Kopp J, Kondaiah P, Wakefield L, et al. Hepatic expression of mature transforming growth factor beta 1 in transgenic mice results in multiple tissue lesions. Proc Natl Acad Sci U S A 1995;92:2572-2576.

[51] Parrelli JM, Meisler N, Cutroneo KR. Identification of a glucocorticoid response element in the human transforming growth factor beta 1 gene promoter. Int J Biochem Cell Biol 1998;30:623-7.

[52] Peng L, Mayhew CN, Schnekenburger M, Knudsen ES, Puga A. Repression of Ah receptor and induction of transforming growth factor-beta genes in DEN-induced mouse liver tumours. Toxicology 2008;246:242-7.

[53] Namjou B, Lingren T, Huang Y; Parameswaran S, Cobb BL, Stanaway IB, et al. GWAS and enrichment analyses of non-alcoholic fatty liver disease identify new trait-associated genes and pathways across eMERGE network. BMC Med 2019;17:135.

[54] Scott E, Anstee QM. Genetics of alcoholic liver disease and non-alcoholic steatohepatitis. Clin Med (Lond) 2018;18(Suppl 2):s54-s59.

TABLE LEGENDS

Table 1. Summary of the data sources and subgroups used in this study

Table 2. Summary of the discovery analysis and phase 1-2 replication results.

Table 3. Association between quintiles of genetic risk score and incident cirrhosis hospitalisation in UK biobank participants with risk factors for non-alcoholic fatty liver disease.

FIGURE LEGENDS

Figure 1. Derivation of discovery GWAS cohort

Figure 2. Discovery analysis Manhattan Plots. Loci identified in our phase 1 replication analysis are highlighted in red if significant at $P<5.0 \times 10^{-8}$ with the corresponding phenotype

Figure 3. Association of *MARC1*:rs2642438 and *HNRNPUL1*:rs15052 with alcohol-related cirrhosis in phase 2 replication. All associations are adjusted for a minimum of age and sex. "Full adjustment" refers to adjustment for Type 2 diabetes and BMI, as well as age and sex. Phase 1 analysis also includes adjustment for the first five principal components of genetic ancestry. Phase 1 replication analysis is based on data from a previous European GWAS of alcohol-related cirrhosis,[9] plus data from a nested alcohol-related case-control study derived from the UK biobank (total effective sample size:2 546). Phase 2 replication analysis is based on two independent datasets from Germany (effective sample size:1926) and Switzerland (Effective sample size: 142). See main text for full details.

Figure 4. Association between genetic risk score and risk of cirrhosis hospitalization among UK biobank participants at risk of non-alcoholic fatty liver disease. Association is adjusted for age, sex, BMI, diabetes and alcohol consumption.

Table 1: Summary of the data sources/subgroups used in this study

Analysis stage	Data source	Cohorts	Characteristic				
			Number	Median age, yrs (IQR)	Sex (% men)	Median BMI (IQR)	% Type 2 diabetes
Discovery analysis	UK Biobank	Alcohol intake: Women:>25units/week; Men >36 units/week	35 839	58 (51-63)	63	27.3 (24.7-30.1)	3.7
Phase 1 replication	Rep #1. Buch et al UK cohort [9]*	Cases: alcohol-related cirrhosis	302	53 (47-60)	68	24.8 (22.8-26.8)	0.0
		Controls: heavy drinkers without alcohol liver disease	346	49 (42-56)	77	24.6 (22.8-26.6)	0.0
	Rep #2. Buch et al German cohort [9]*	Cases: alcohol-related cirrhosis	410	53 (47-71)	71	26.2 (22.8-29.3)	24.0
		Controls: heavy drinkers without alcohol liver disease	1080	42 (36-48)	100	24.8 (22.7-27.5)	4.0
	Rep #3. UK Biobank†	Cases: alcohol-related cirrhosis	178	60 (53-63)	76	29.0 (25.5-32.8)	21.9
		Controls: non-teetotal participants without evidence of liver disease	298 248	59 (51-64)	45	26.6 (24.1-29.7)	3.8
Phase 2 replication	Rep#4. Germany validation cohort*	Cases: alcohol-related cirrhosis	1272	59 (52-66)	82	27.0 (24.0-30.2)	36.6
		Controls: heavy drinkers, no evidence of liver disease	775	49 (41-55)	78	24.4 (21.8-27.1)	5.0
	Rep#5. Switzerland validation cohort*	Cases: alcohol-related cirrhosis	312	64 (57-71)	85	26.0 (22.8-29.4)	34.5
		Controls: heavy drinkers, no evidence of liver disease	40	63 (58-69)	58	24.8 (22.3-29.0)	13.8
Genetic risk score	UK Biobank	UKB NAFLD risk factor subgroup	107 014	59 (52-64)	43	32.6 (30.9-35.4)	16.9
Functional analysis	Kiel University, Germany[35]	Liver biopsy cohort with mRNA expression data	113	46 (40-60)	43	42.4 (25.8-51.5)	21.6

* Data for BMI and type 2 diabetes is missing for >20% of participants

† excludes individuals included in the UK biobank discovery analysis.

Table 2: Summary of discovery analysis and phase 1 -2 replication results

					Discovery analysis					Phase 1 replication analysis**		Phase 2 replication analysis**			Replication analysis Phase 1 + 2 **		
VARIANT INFORMATION					P_value for association (N=35 839)					Pooled rep#1+#2+#3 (N_eff†=2 545)		Rep#4+#5: N_eff=2068			Pooled rep#1-#5 (N_eff†=4613)		Direction§ (rep#1;#2;#3; #4; #5)
SNP	Ref:Alt allele	Chr	Alt allele freq*	Nearest Gene	FORNS	APRI	FIB4	AST	ALT	Beta	P-value	Beta	P-value	FDR 5%	Beta	P-value	
rs738408	C:T	22	0.216	*PNPLA3*	2.21×10^{-8}	6.77×10^{-63}	8.94×10^{-17}	1.74×10^{-78}	7.51×10^{-82}	0.734	3.54×10^{-24}	0.884ǀ	2.42×10^{-29}	YES	0.803	2.21×10^{-51}	+++++
rs10401969	T:C	19	0.077	*SUGP1*	9.15×10^{-6}	3.67×10^{-8}	2.62×10^{-1}	3.50×10^{-16}	1.13×10^{-18}	0.678	5.74×10^{-10}	0.636¶	3.95×10^{-7}	YES	0.660	1.21×10^{-15}	+++++
rs11065384	C:T	12	0.307	*HNF1A*	1.86×10^{-8}	2.20×10^{-4}	1.86×10^{-3}	4.12×10^{-3}	3.35×10^{-2}	0.275	7.10×10^{-5}	0.108	1.57×10^{-1}	NO	0.199	1.01×10^{-4}	++++-
rs11925835	C:T	3	0.424	*ARHGEF3*	5.88×10^{-22}	1.31×10^{-14}	4.40×10^{-26}	3.21×10^{-1}	1.36×10^{-1}	-0.235	7.32×10^{-4}	-0.031	6.62×10^{-1}	NO	-0.134	6.64×10^{-3}	- - - - +
rs28929474	C:T	14	0.020	*SERPINA1*	3.20×10^{-1}	6.96×10^{-5}	1.11×10^{-1}	1.12×10^{-6}	3.67×10^{-8}	0.561	7.47×10^{-3}	1.029	5.08×10^{-4}	YES	0.717	2.77×10^{-5}	+++++
rs2954038	A:C	8	0.300	*TRIB1*	1.66×10^{-1}	3.52×10^{-3}	2.08×10^{-2}	2.09×10^{-5}	3.77×10^{-20}	0.160	1.29×10^{-2}	0.093¥	2.44×10^{-1}	NO	0.140	8.75×10^{-3}	+++++
rs15052	T:C	19	0.178	*HNRNPUL1*	1.14×10^{-7}	6.87×10^{-10}	6.65×10^{-12}	5.45×10^{-5}	2.25×10^{-1}	0.222	1.34×10^{-2}	0.218	3.20×10^{-2}	YES	0.220	1.06×10^{-3}	-++++
rs2642438	G:A	1	0.297	*MARC1*	7.28×10^{-4}	6.25×10^{-2}	2.13×10^{-2}	2.11×10^{-4}	8.87×10^{-13}	-0.177	1.97×10^{-2}	-0.273	5.38×10^{-4}	YES	-0.223	4.51×10^{-5}	- - - - -
rs72613567	T:TA	4	0.279	*HSD17B13*	1.23×10^{-5}	6.33×10^{-15}	3.00×10^{-5}	1.44×10^{-14}	1.38×10^{-17}	-0.166	2.76×10^{-2}	-0.316	6.61×10^{-5}	YES	-0.237	1.38×10^{-5}	- - - - -

* refers to the allele frequency observed in the UKBiobank Caucasian British subset, excluding related participants.

† effective sample size varies marginally for each SNP

** Data adjusted for age and sex.

§ "+" direction indicates that Alt allele is associated with increased risk of cirrhosis relative to the Ref allele; Vice versa, "-" direction indicates that Alt allele is associated with a reduced risk of cirrhosis relative to the Ref allele

ǀ association is based on rs738409 (r2=1.0) in replication cohort #4; ¶ association is based on rs58542926 (r2=0.91) in replication cohort #4; ; ¥ association is based on rs2980888(r2=1.0) in replication cohort #4.

Grey shaded cells denote statistical significance in discovery analysis at the standard genome-wide significance level ($P<5.0 \times 10^{-8}$)

Abbreviations: SNP – single nucleotide polymorphism; APRI- serum aspartate transaminase /platelet ratio; FIB4 – Fibrosis-4 Index; AST – serum aspartate transaminase; ALT – serum alanine transaminase

Table 3. Association between quintiles of genetic risk score and incident cirrhosis hospitalisation in participants with risk factors for Non-Alcoholic Fatty Liver Disease

Genetic risk score (quintiles)	Person years of follow-up	Events	Incident rate per 10,000 PYs	UNADJUSTED		ADJUSTED*	
				HR (95% CI)	P-value	HR (95% CI)	P-value
1 (lowest risk)	171896	72	4.19	REF (1.00)	\	REF (1.00)	\
2	174261	85	4.88	1.25 (0.90-1.73)	0.180	1.30 (0.93-1.81)	0.169
3	179536	107	5.96	1.44 (1.06-1.98)	0.022	1.44 (1.04-2.00)	0.026
4	155611	103	6.62	1.75 (1.29-2.37)	<0.001	1.77 (1.30-2.42)	<0.001
5 (highest risk)	166357	195	11.72	3.12 (2.37-4.12)	<0.001	3.16 (2.38-4.21)	<0.001

*adjusted for age; gender; BMI; diabetes and units/alcohol consumed per week

Abbreviations: HR- hazards ratio; CI- confidence intervals

Figure 1. Derivation of discovery GWAS cohort

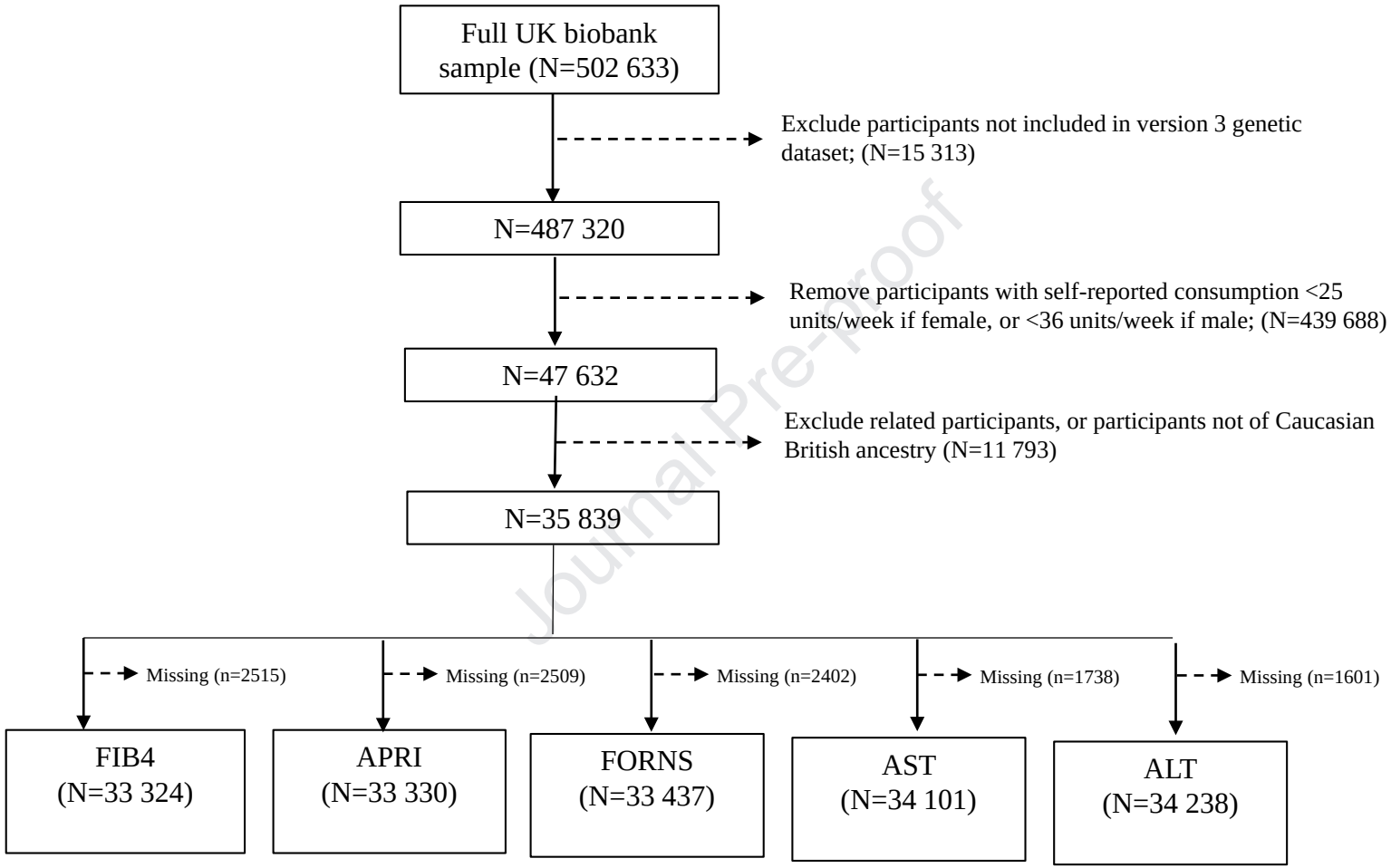
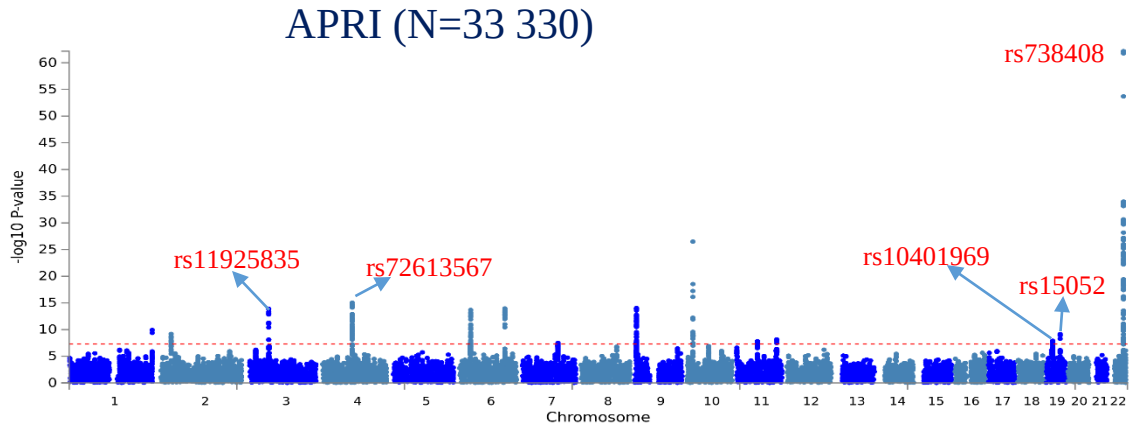
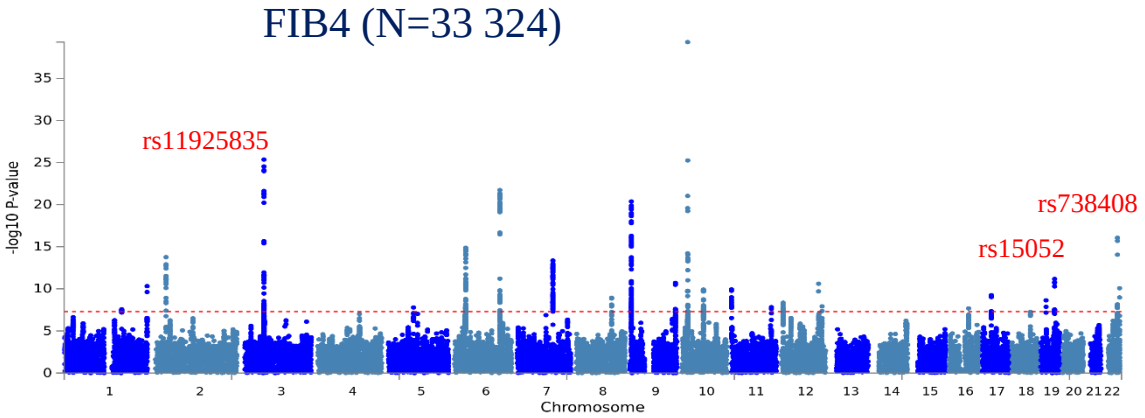
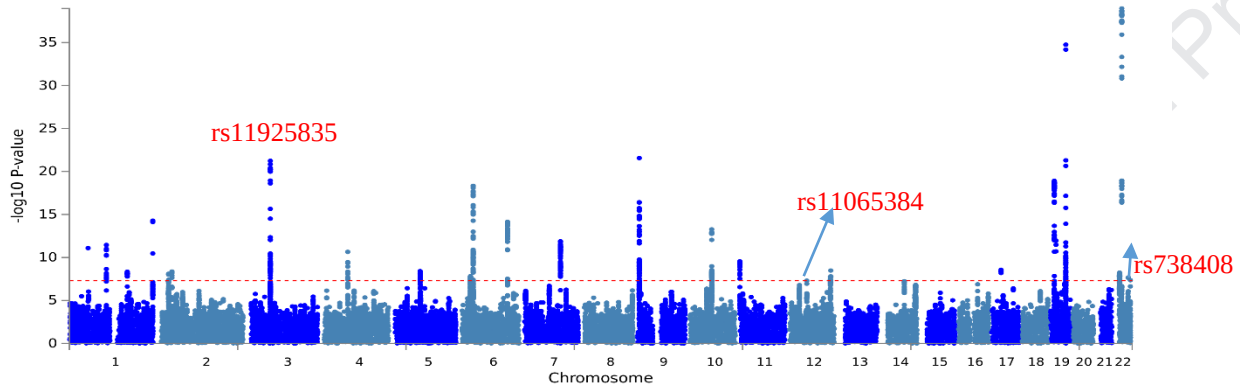
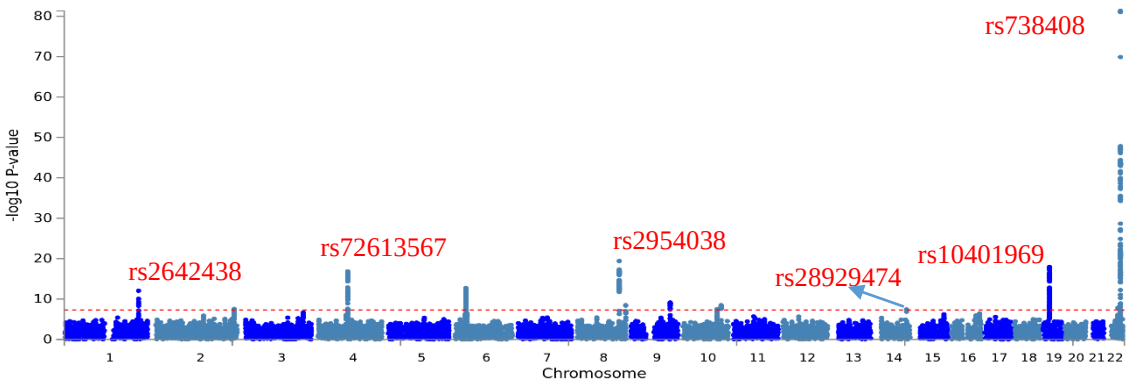


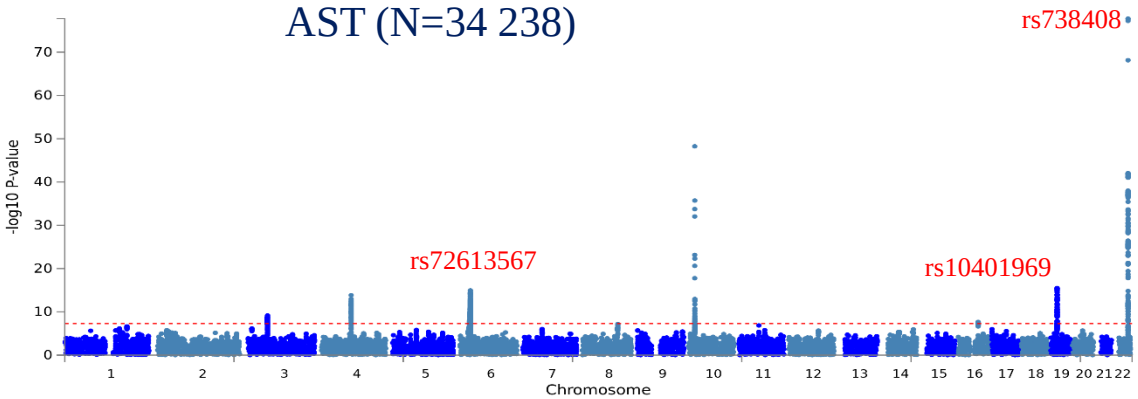
Figure 2. Discovery analysis Manhattan plots*





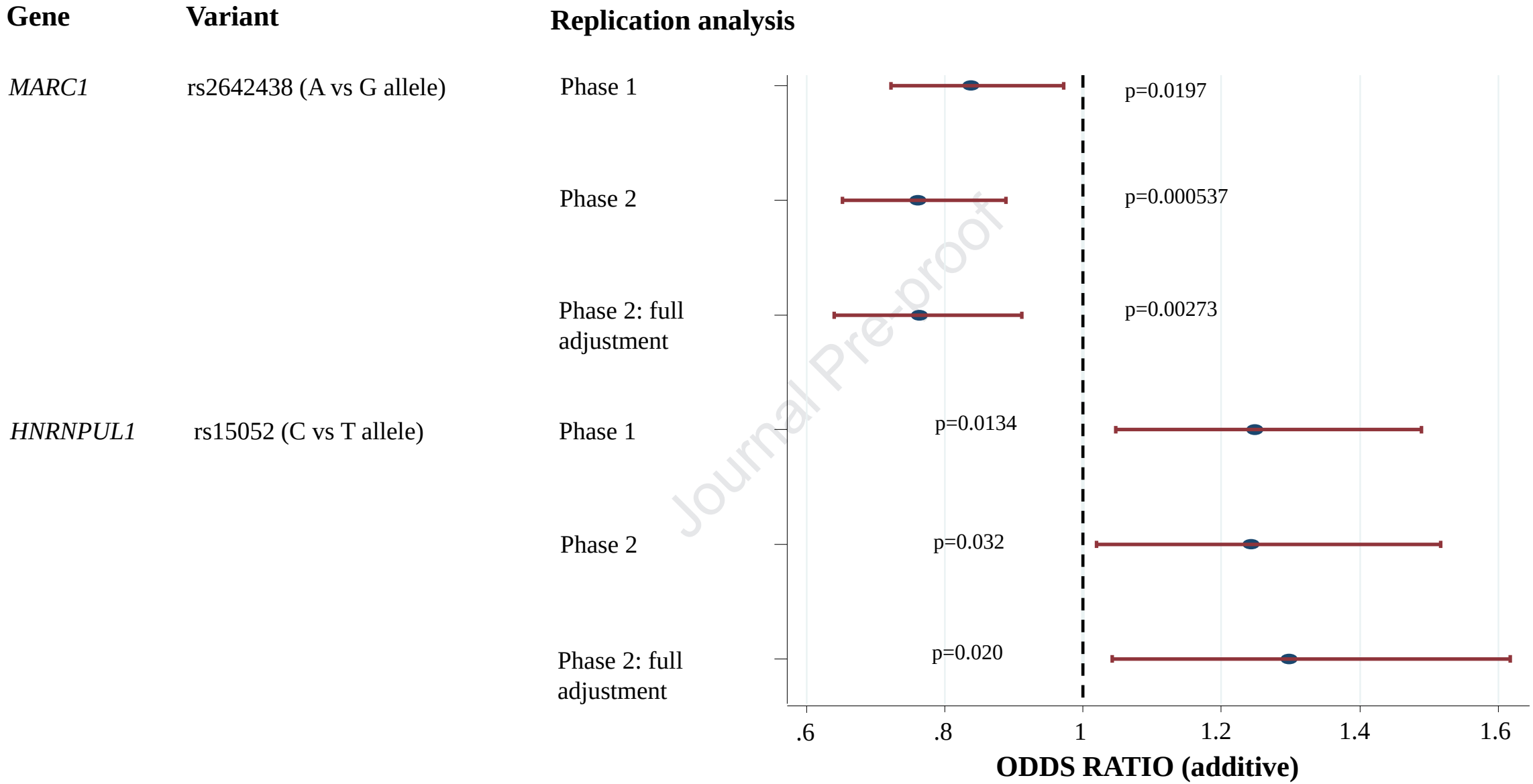






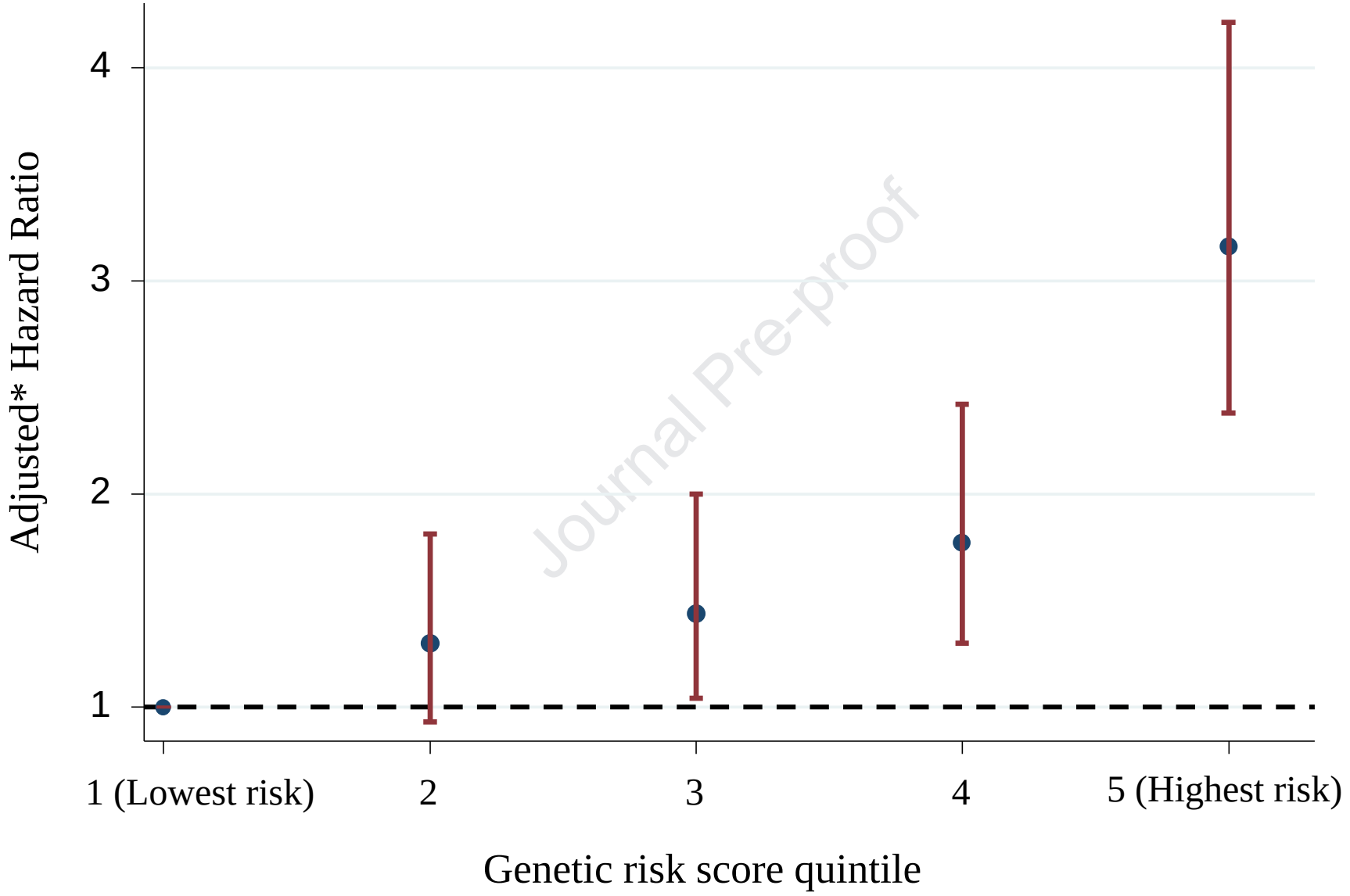
*Loci identified in our phase1 replication analysis are highlighted in red if significant at $P<5.0 \times 10^{-8}$ for corresponding phenotype

Figure 3. Association of *MARC1*:rs2642438 and *HNRNPUL1*:rs15052 with alcohol-related cirrhosis in phase 2 replication



Associations are adjusted for a minimum of age and sex. “Full adjustment” refers to adjustment for Type 2 diabetes and BMI, as well as age and sex. Phase 1 analysis also includes adjustment for the first five principal components of genetic ancestry. Phase 1 replication analysis is based on data from a previous European GWAS of alcohol-related cirrhosis[9], plus data from a nested alcohol-related case-control study derived from the UK biobank (total effective sample size:2 546). Phase 2 replication analysis is based on two validation datasets from Germany (effective sample size:1926) and Switzerland (Effective sample size: 142). See main text for full details.

Figure 4. Association between genetic risk score and risk of cirrhosis hospitalisation, among UK biobank participants at risk of non-alcoholic fatty liver disease.



*adjusted for age, sex, BMI, diabetes and alcohol consumption

WHAT YOU NEED TO KNOW:

Background and Context: Few genetic factors have been associated with development of alcohol-related cirrhosis.

New Findings: In a GWAS of samples from the UKB, the authors identified and validated (in 5 European cohorts) single-nucleotide polymorphisms that affect risk of alcohol-related cirrhosis in opposite directions: the minor A allele in *MARC1*:rs2642438 decreases risk whereas the minor C allele in *HNRNPUL1*:rs15052 increases risk.

Limitations: Studies are needed to determine how variants in these genes might contribute to development of cirrhosis in patients with alcohol use disorders.

Impact: These findings might be used to identify patients at risk for cirrhosis and to determine mechanisms of liver fibrogenesis.

Lay Summary: The authors identify genetic features that increase risk of cirrhosis in persons with high alcohol intake.