Journal of Hepatology

A GENETIC RISK SCORE AND DIABETES PREDICTS DEVELOPMENT OF ALCOHOL-RELATED CIRRHOSIS IN DRINKERS

--Manuscript Draft--

Manuscript Number:	JHEPAT-D-21-01108R1
Article Type:	Original Article
Section/Category:	NAFLD and Alcohol-Related Liver Diseases
Keywords:	Hepatocellular carcinoma; risk stratification; chronic alcohol use; genome wide association; single nucleotide polymorphism; coffee
First Author:	John B. Whitfield, PhD FRCPath
Corresponding Author:	Devanshi Seth, PhD MPH MSc BSc (Honours) Royal Prince Alfred Hospital AUSTRALIA
Order of Authors:	John B. Whitfield, PhD FRCPath
	Tae-Hwi Schwantes-An, PhD
	Rebecca Darlay, PhD
	Guruprasad P. Aithal, PhD FRCP
	Stephen R. Atkinson, PhD
	Ramon Bataller, MD PhD
	Gregory Botwin, PhD
	Naga Chalasani, MD
	Heather J. Cordell, DPhil
	Ann K. Daly, PhD
	Christopher P. Day, MD PhD
	Florian Eyer, MD
	Tatiana Foroud, PhD
	Dermot Gleeson, MD FRCP
	David Goldman, MD
	Paul S. Haber, MD PhD
	Jean-Marc Jacquet, MD
	Tiebing Liang, PhD
	Suthat Liangpunsakul, MD
	Steven Masson, FRCP
	Philippe Mathurin, MD PhD
	Romain Moirand, MD PhD
	Andrew McQuillon, PhD
	Christophe Moreno, MD PhD
	Marsha Y. Morgan, MD PhD
	Sebastian Mueller, MD PhD
	Beat Mullhaupt, MD
	Laura E. Nagy, PhD

	Pierre Nahon, MD PhD
	Bertrand Nalpas, MD PhD
	Sylvie Naveau, MD
	Pascal Perney, MD PhD
	Munir Pirmohamed, PhD FRCP
	Helmut K. Seitz, MD
	Michael Soyka, MD
	Felix Stickel, MD PhD
	Andrew Thompson, PhD
	Mark R. Thursz, MD
	Eric Trepo, MD PhD
	Timothy R. Morgan, MD
	Devanshi Seth, PhD MPH MSc BSc (Honours)
Abstract:	Background: Only a minority of excess alcohol drinkers develop cirrhosis. Risk stratification would identify those at highest risk for developing cirrhosis allowing appropriate management including intensive intervention. Methods: Three cohorts (GenomALC-1: n=1690, GenomALC-2: n=3037, UK Biobank: n=502,506) with known alcohol consumption history were included. Cases were participants with alcohol-related cirrhosis while controls had a history of alcohol consumption (≥80 g/day (men), ≥50 g/day (women), for ≥10 years) but with no evidence of liver disease. Risk scores were computed from up to eight genetic risk loci from genome-wide association studies of alcohol-related cirrhosis and three clinical risk factors, and their performance for the diagnosis of alcohol-related cirrhosis assessed and compared. The stratification utility of the risk scores was tested across alcohol-related liver diseases including hepatocellular carcinoma (HCC). Results: A combination of three single nucleotide polymorphisms (SNPs) (PNPLA3:rs738409, SUGP1-TM6SF2:rs10401969, HSD17B13:rs6834314) and diabetes status best discriminated for cirrhosis risk. Based on independent allelic effect size estimates, the odds ratio (95% confidence intervals) for the extreme score quintiles (Q1-Q5) of the 3-SNP score, were 5.99 (4.18;8.60) (GenomALC-1); 2.81 (2.03;3.89) (GenomALC-2); and 3.08 (2.30;4.11) (UK Biobank). Comparing people with diabetes (Q5) versus those without diabetes (Q1), the ORs increased to 26.3 (8.08;85.7) (GenomALC-1) and 23.8 (15.0;37.8) (UK Biobank); diabetes status was unavailable for GenomALC-2. Patients with cirrhosis and HCC had significantly higher mean risk scores than patients with cirrhosis alone (0.76±0.06 versus 0.61±0.02, p=0.007). Score performance was not enhanced by adding additional genetic risk variants, body mass index or coffee consumption. Conclusion : A three genetic variant risk score can provide meaningful risk stratification for cirrhosis and HCC in excess drinkers. Risk stratification is enhanced by inclusion of diabetes status.
Response to Reviewers:	

Dr Paolo Angeli, Editor-in Chief
Dr Vlad Ratizu, Co-editor
Journal of Hepatology

Date: 2 Sep 2021

JHEPAT-D-21-01108
A genetic risk score and diabetes predicts development of alcohol-related cirrhosis in drinkers

Dear Dr Paola Angeli and Dr Vlad Ratizu

We thank the editors and reviewers for their overall positive comments and valuable suggestions. The revised the manuscript has taken into account all the queries and suggestions made by the Reviewers and Editors. A point-by-point response letter is submitted.

We believe the revised manuscript has satisfactorily addressed all the concerns and should be deemed suitable for publication.

Both marked and clean copies of the revised manuscript are attached. Other documents as per checklist have also been submitted.

Please contact me if any other information is required.

We look forward to your positive response.

Sincerely

Devanshi Seth, PhD (on behalf of the authors)

PRINCIPAL SCIENTIST
Edith Collins Centre (Translational Research in Alcohol Drugs and Toxicology)
SYDNEY LOCAL HEALTH DISTRICT

+61 2 9565 6268; +61 2 9515 7201 | d.seth@sydney.edu.au

JHEPAT-D-21-01108_Checklist
A genetic risk score and diabetes predicts development of alcohol-related cirrhosis in drinkers

MANUSCRIPT NUMBER: (given once successful submission has taken place) **JHEPAT-D-21-01108**

√ Letter of submission to the Editor

√ Copyright assignment, Authorship and Responsibility, Financial

√ Disclosure and Institutional Review Board/Animal Care Committee Approval (signed by all authors).

N/A Drug Declaration Form

N/A Consort Guideline Checklist for randomised controlled trials

√ Title Page

√ Short Title of less than 40 spaces: **GRS predicts alcoholic cirrhosis risk**

√ Author(s) and Affiliation(s)

√ Address, telephone and fax numbers and e-mail of corresponding author

√ Electronic Word Count (excluding abstract and references)

√ Article, proper (double-spaced)

√ First author's last name, short title and page number at the top of each page

√ Structured summary of less than 200 words including an electronic word count

√ From three to ten key words

√ Manuscript length: Maximum of 3,000 words **(within limit of 6000 for original article),** excluding abstract, references, figures and tables.

√ Introduction/Background/Aims

√ Materials and methods

√ Results

√ Discussion

√ Acknowledgements

√ References

√ Tables

√ Figure legends

N/A Figures separate from the text of the figure legends (Power Point or JPG files)

√ Permission to reproduce any previously published material and patient permission to publish photographs.

JOURNAL
OF HEPATOLOGY

Journal of Hepatology

CTAT methods

Tables for a "Complete, Transparent, Accurate and Timely account" (CTAT) are now mandatory for all revised submissions. The aim is to enhance the reproducibility of methods.

- Only include the parts relevant to your study
- Refer to the CTAT in the main text as 'Supplementary CTAT Table'
- Do not add subheadings
- Add as many rows as needed to include all information
- Only include one item per row

If the CTAT form is not relevant to your study, please outline the reasons why:

This paper did not use any of the listed items below. The data utilized in this study is available from our ongoing research and is located with the first author Dr Whitfield at the Queensland Institute of Medical Research Berghofer, Queensland, Australia under the institute's secure database and ethics approval (HREC # P1380).

1.1 Antibodies

Name	Citation	Supplier	Cat no.	Clone no.

1.2 Cell lines

Name	Citation	Supplier	Cat no.	Passage no.	Authentication test method

1.3 Organisms

Name	Citation	Supplier	Strain	Sex	Age	Overall n number

1.4 Sequence based reagents

Name	Sequence	Supplier

1.5 Biological samples

Description	Source	Identifier

1.6 Deposited data

Name of repository	Identifier	Link

1.7 Software

Software name	Manufacturer	Version

1.8 Other (e.g. drugs, proteins, vectors etc.)

1.9 Please provide the details of the corresponding methods author for the manuscript:

2.0 Please confirm for randomised controlled trials all versions of the clinical protocol are included in the submission. These will be published online as supplementary information.

Dr Paolo Angeli, Editor-in Chief
Dr Vlad Ratizu, Co-editor
Journal of Hepatology

Date: 2 Sep 2021

JHEPAT-D-21-01108
A genetic risk score and diabetes predicts development of alcohol-related cirrhosis in drinkers

Dear Editorial team,

We thank the editors and reviewers for their overall positive comments and valuable suggestions. We believe the revised manuscript has satisfactorily addressed the Reviewers and Editors queries and suggestions, and should be suitable for publication.

Both marked and clean copies of the revised manuscript are attached.

Point-by-point response to Reviewers
Please see below point-by-point responses. New information added to the revised manuscript is underlined, and location (clean copy page & line #s) is provided here. New supplementary Tables S2 and S3 have been added, and subsequently, previous supplementary tables 2 and 3 are renamed as supplementary tables 4 and 5, respectively.

Reviewer 1
Major comments
- from a clinical point of view, discrimination of a clinical outcome cannot rely on OR 95% ci; for each model Authors must report measures of diagnostic accuracy, including AUROCs, Se/Sp/NPV/PPV/LHRs; AUROCs should be compared by appropriate tests to examine differences in the diagnostic performance

Response: We have now added AUROC values in the revised Table 2. However, the reporting of specificity and sensitivity is complex in this context because presentation of these measures in a Table may lead readers to interpret the genetic score being proposed as a diagnostic test. Therefore, we present this information in the text. Assuming a score cut-off with a sensitivity of 80% is desirable to identify nearly all those who are at high risk, specificity values are in the range 30-40%. We believe that in the context of risk stratification, low Specificity (~30%-40%) is still acceptable identify as many people at high risk as possible, and to accept that many people labelled as high-risk will not develop cirrhosis. We also discussed this in the revised manuscript (pages16-17).

page 13, lines 265-266: "Results in the *three* study cohorts for the 3-SNP score, *AUCs, logistic regressions and the Odds Ratios (ORs), comparing quintiles Q5 and Q1 of the score*, are shown in Table 2."

pages 16-17, lines 360-368: *"For any classification based on a numerical test or score, changing the cut-off point(s) will alter the trade-off between test sensitivity and specificity*

and the optimum cut-offs will depend on the use to be made of the test. The prevalence of the condition is also important because this will affect the predictive value of positive or negative results. The AUCs shown in Table 2 were significant but when the desired test sensitivity was set at 80% (to flag nearly all those at high risk) the test specificities were between 30% and 40%. Thus, a substantial number of false positives must be accepted, making the score suitable for risk stratification but not for prediction of outcome in individual patients."

- along the same line, it would be clinically relevant to test whether the score can predict the risk of cirrhosis prospectively or risk of HCC and/or clinical events in patients with cirrhosis. That is, how would the score be implemented in a heavy drinker undergoing clinical evaluation likely already associated with noninvasive assessment of liver fibrosis? Limitations related to the study design on the possibility to make clinical recommendations at this stage should be better acknowledged.

Response: We agree it would be desirable to study the associations between the risk score and disease progression across time in patients who present with lesser degrees of liver damage. It would also be valuable to follow-up those with high-risk (although genetic risk is constant) across time to see if development of diabetes precedes, follows or exacerbates liver damage, but both are beyond the scope of this paper. Follow-up prospective studies in larger cohorts are required and will be the next logical step forward in this field. We have discussed this as one of the limitations of the study in the revised manuscript as below in the main text.

Page 19, lines 414-416: "*Prospective studies are needed, both to relate score to progression across time in patients who present with early stages of liver disease, and to clarify the relationship between onset of diabetes and of advanced liver disease in patients with excessive alcohol use.*"

- liver damage staging in the clinical cohorts should be reported in much greater details; this is a key methodological issue for the study and cannot be referred to previous publications

Response: As suggested, we have added more detailed information about our criteria for the GenomALC-1 study (page 9). For GenomALC-2 and UK Biobank the available information was restricted to assigned clinical diagnosis. We have now explained this more fully (page 10). In addition, we now have a new Supplementary Table 2 with detailed characteristics on liver injury/disease parameters that were available for each cohort (page 9).

page 9, lines 173-174: "*Cohort characteristics of the cases and controls from each source are described in Supplementary Table 2.*"

Page 9, lines 179-186: *"For cases, cirrhosis had been diagnosed by a combination of clinical criteria and/or liver elastography (Fibroscan®), with liver biopsy if clinically indicated. Other potential causes of liver diseases (hepatitis B or C, haemochromatosis, Wilson's disease, and autoimmune hepatitis) were excluded by laboratory testing or clinical criteria. For controls, liver disease was excluded through a combination of clinical history and*

measurement of liver function tests (bilirubin, albumin, ALT). For both cases and controls, HIV infection was an exclusion criterion."

page 10, lines 196-197: *"Clinical diagnosis of cases and controls were similar to GenomALC-1 criteria but detailed clinical diagnosis information was limited for this cohort."*

Page 10, lines 211-212: "*For cases, information was restricted to assigned clinical diagnosis (Supplementary Table 2)* on hospital admissions….subsequently died)."

- limitations of noninvasive assessment and lack of systematic evaluation in UKBB should be discussed.

Response: We agree UK Biobank cohort may have limitations in accurate case/control classification. It is uncertain how far this may be due to non-invasive assessment and lack of systematic evaluation. However, the diagnoses were based on death certificates or hospital records supplied to UK Biobank and would presumably have been made by a specialist gastroenterologist or hepatologist. In any case, misclassification of cases/controls would lead to poorer stratification so the effectiveness of our score would be under-, rather than over-estimated. This has been clarified in study limitations.

pages 18-19, lines 407-409: *"*.....liver disease, *although it should be recognised that misclassification of some cases as controls would lead to poorer stratification such that the effectiveness of the score would be under-, rather than over-estimated.*

- it seems relevant to note and discuss the possible implications that the selected risk variants are shared among alcoholic cirrhosis - NAFLD - steatohepatitis (PMID: 34027340); Authors may discuss the parallel results from the evaluation of PRSs made up of a very similar panel of variants on the risk of HCC related to NAFLD also conducted in European clinical cohorts and the UKBB that has recently been published in the Journal (PMID: 33248170). These E.g. may these PRS have a larger application?

Response: We have added discussion about the wider applicability of this risk or similar scores, specifically for non-alcohol-related liver diseases.

page 17, lines 376-381: "*Using similar polygenic risk scores in NAFLD revealed that combining genetic and clinical features refines the predictive utility of the algorithm for identifying those at higher risk of severe liver disease (Bianco, J Hepatol 2021, Bianco, J Hepatol Rep 2021, de Vincentis 2021). Given the many shared genetic and metabolic risks between alcohol-related liver disesae and NAFLD, the predictive algorithm defined here may have a wider scope across these diseases for risk stratification of those at higher risk of cirrhosis.*"

- based on the previous results, it would be important to test whether there was a significant interaction between PRS and diabetes in determining the risk of alcoholic cirrhosis.

Response: The original version of this manuscript showed the genetic risk score had similar associations in those with and without diabetes (Results, Diabetes sub-section, page14, lines 310-312). As suggested, we have now also included a test for interaction between diabetes status and the 3- SNP score in predicting case/control status.

pages 14-15, lines 313-315: "*Tests for genetic score-diabetes interaction, either by including a (score x diabetes) term in the logistic regression or by testing for heterogeneity of Odds Ratios between those with and without diabetes, showed no evidence for interaction effects in either cohort (Table 4).*"

Minor comments

- in the abstract, only individuals from UKBB who were actually considered for the analysis should be reported

Response: We have made this suggested change in the Abstract (page 6).
Please also note that the number of UK Biobank participants in the control group has changed from the earlier version of this paper. This is because (i) some excessive-drinking controls with liver diseases not specified as alcohol-related have been removed, and (ii) some participants who met the excessive-drinking criteria were previously included in a separate 'alcohol dependence' category which has been dropped; these people are now included in the excessive-drinking group (please refer to Supplemental Table 1).

page 6, line 114: "Three cohorts … UK Biobank: *relevant n=6898*): with a ….."

- the choice of including a proxy for the TM6SF2 locus, where there is robust evidence of causality for the TM6SF2 p.E167K variant, should be better substantiated

Response: Our choice of TM6SF2 SNPs for the 3-SNP score was based on the study by Buch et al (2015). The results were provided for both rs10401969 and rs58542926 (E167K) at the TM6SF2 locus (their Table 2), which are in strong linkage disequilibrium (d'1.00, R2 0.955). We chose rs10401969 with slightly greater effect size and smaller p-value. We have added a brief explanation of our decision to use the published SNPs and if there was more than one reported at any locus, to use the most significant one.

Page 12, lines 242-244: "*Two significantly associated SNPs have been reported at SUGP1-TM6SF2 locus[17] which are in near-complete linkage disequilibrium (d'1.00, r^2 0.955), and rs10401969 was chosen over rs58542926 because of its stronger association with cirrhosis.*"

Reviewer 2

Major comments

- authors evaluated and compared the different SNP-scores exclusively looking at odds ratios for the association with alcoholic cirrhosis. These measures express the strength of an association, but in fact do not capture the probability of experiencing the disease and are not able to forecast the observed risk. In this reviewer's opinion, the whole

analytic approach should be expanded with the model goodness-of-fit, AUROCs, calibration curves, sensitivity, specificity and predictive values.

Response: As suggested by the Reviewer, we have added new information on AUCs and their standard errors to Table 2. However, the reporting of specificity and sensitivity is complex in this context because presentation of these measures in a Table may lead readers to interpret the genetic score being proposed as a diagnostic test. Therefore, we present this information in the text. Assuming a score cut-off with a sensitivity of 80% is desirable to identify nearly all those who are at high risk, specificity values are in the range 30-40%. We believe that in the context of risk stratification, low Specificity (~30%-40%) is still acceptable to identify as many people at high risk as possible, and to accept that many people labelled as high-risk will not develop cirrhosis. The paragraph below also discusses this in the revised manuscript. [This response is also detailed above to a similar comment by Reviewer 1].

page 13, lines 265-266: "Results in the *three* study cohorts for the 3-SNP score, *AUCs, logistic regressions and the Odds Ratios (ORs), comparing quintiles Q5 and Q1 of the score*, are shown in Table 2."

pages 16-17, lines 360-368: *"For any classification based on a numerical test or score, changing the cut-off point(s) will alter the trade-off between test sensitivity and specificity and the optimum cut-offs will depend on the use to be made of the test. The prevalence of the condition is also important because this will affect the predictive value of positive or negative results. The AUCs shown in Table 2 were significant but when the desired test sensitivity was set at 80% (to flag nearly all those at high risk) the test specificities were between 30% and 40%. Thus, a substantial number of false positives must be accepted, making the score suitable for risk stratification but not for prediction of outcome in individual patients."*

- it is not clear how Authors chose the high and low risk classes and why reported OR for only the fifth vs first quintile. Moreover, Supplementary Figure 2 is not adequately discussed. Authors should report OR for each quintile and clearly define low, intermediate and high-risk classes.

Response: Choice of cut-off values for determining the proportion of people in each (high, intermediate or low risk) group was considered along with balancing sensitivity and specificity. An important factor to consider is the relative undesirability of false positives and false negatives, which depends on the consequences of wrong classification and on the prevalence of the condition in the tested population. Because cut-offs defined by quintiles cannot easily be applied clinically, we defined numerical cut-offs (e.g. 0 and 0.7) as a suitable compromise as mentioned before (page 14, lines 294-295; page 16, lines 357-358) and this is shown in Supplementary Figure 2. The information on odds ratios for each quintile is less relevant in clinical setting and the Q5-Q1 comparisons can mainly be used for comparison across cohorts as in this study and against reports on genetic scores for other diseases.
We have elaborated this in the revised manuscript.

page 16, lines 355-357: "*However, Q5-Q1 comparisons can be useful for comparison across cohorts, such as in our study, and against genetic scores for other diseases.*"

- Authors tried to combine the added value of genetics with that of clinical variables. Indeed, they stratified the analysis by the presence of diabetes mellitus. Could other clinical factors (gender, obese Vs nonobese, other comorbidities?) help in further stratifying the risk? If yes, how did Authors take them into account in their analyses?
- given the different impact of alcohol intake across gender, it would be interesting to see a sensitivity analyses stratified by gender.

Response: We presented information on other co-variates (coffee, BMI) (page 13, lines 270-271), identified in our previous publication Whitfield et al am J Gastro 2020. We have performed stratification by gender and provided this information in new Supplementary Table 3 and added text in the revised manuscript in relation to the other risk factors.

page 13, lines 276-279: "*Coffee data did not improve the risk stratification, and nor did BMI (which was non-significant in the UK Biobank group and not available for GenomALC-2). Similar results were seen in both men and women (Supplementary Table 3).*"

- a general descriptive table reporting the socio-demographic, clinical and biochemical characteristics of the study population would help a meaningful understanding.

Response: As suggested, we have now added a new Supplementary Table 2 with descriptive information that were available for the three cohorts, detailing characteristics on demographics, alcohol use, tobacco and cannabis use, coffee intake, laboratory results and liver injury/disease parameters.

page 9, lines 173-174: "*Cohort characteristics of the cases and controls from each source are shown in Supplementary Table 2.*"

- concerning the UKBB cohort, Authors should more precisely detail the population selection along with the methodology applied to derive the amount of alcohol assumption per day, given the variables available from the UKBB.

Response: We have added more detail on the UK Biobank cohort selection/recruitment procedures, and the calculation of alcohol intake from participants responses to the questions about drinking amounts.

Pages 10-11, lines 213-222: *"Information was available on self-reported alcohol intake at the time of assessment (by beverage type, in drinks per week, or for less frequent drinkers per month), and participants also reported whether this was less than, similar to or more than they had been consuming 10 years previously. The amounts were converted to express the alcohol intake in g/d. Participants who had a recorded diagnosis or cause of death of alcohol-related cirrhosis (ICD-10 K70.3, 'Alcoholic cirrhosis of liver') from hospital records or death certificates were included as cases (n=594), and those with reported drinking above the 80 or 50 g/day limits, with similar or greater consumption 10 years before, but with no*

diagnosed liver disease (either alcohol-related or other causes) were included as controls (n=6304).”

- what were the exclusion criteria for enrolling subjects from the UK Biobank? Were subjects with liver disease from other etiologies included?

Response: Exclusion criteria for UK Biobank subjects were similar to GenomALC-1. This information is now added in cohort description.

page 11, lines 221-222: "*Exclusion criteria for UK Biobank subjects were similar to GenomALC-1.”*

Minor comments
- in the background please report the reference for the sentence "most people who develop alcohol-related cirrhosis report sustained alcohol intake >=80 g/d (men) and >=50 g/d (women) for >=10 years".

Response: The 80 g/d (men) and 50 g/d (women) cut-offs were set at a relatively high level to ensure both the cirrhosis and control groups were exposed to an incidental level of alcohol related risk. We have revised this statement with the following text and added relevant references.

page 8, lines 150-154: *“Long-term consumption of 80 g/d or more is associated with increased risk of cirrhosis (Corrao 1999, Lelbach 1976) but threshold for harm has been below this level, especially for women (Tyuns & Pequignot 1984, Roerecke 2019). The 80 g/d (men) and 50 g/d (women) cut-offs were set at a relatively high level to ensure both the cirrhosis and control groups were exposed to an incidental level of alcohol related risk.”*

- similar works but on non alcoholic fatty liver disease by bianco et al J Hep 2021 and de vincentis et al CGH 2021 may be discussed considering also the overlap in genetic risk factors between NAFLD and ALD

Response: We agree that the potential for using this or a similar score for NAFLD (and perhaps other liver diseases) should be mentioned and have expanded this part of the discussion accordingly. [This response is also detailed above to a similar comment by Reviewer 1].

page 17, lines 376-381: “*Using similar polygenic risk scores in NAFLD revealed that combining genetic and clinical features refines the predictive utility of the algorithm for identifying those at higher risk of severe liver disease (Bianco, J Hepatol 2021, Bianco, J Hepatol Rep 2021, de Vincentis 2021). Given the many shared genetic and metabolic risks between alcohol-related liver disesae and NAFLD, the predictive algorithm defined here may have a wider scope across these diseases for risk stratification of those at higher risk of cirrhosis.*”

A GENETIC RISK SCORE AND DIABETES PREDICTS DEVELOPMENT OF ALCOHOL-RELATED CIRRHOSIS IN DRINKERS

John B. Whitfield[#] (1), Tae-Hwi Schwantes-An (2), Rebecca Darlay (3), Guruprasad P. Aithal (4), Stephen R. Atkinson (5), Ramon Bataller (6), Greg Botwin (7,8), Naga P. Chalasani (9), Heather J. Cordell (3), Ann K. Daly (10), Christopher P. Day (11), Florian Eyer (12), Tatiana Foroud (2), Dermot Gleeson (13), David Goldman (14), Paul S. Haber (15,16), Jean-Marc Jacquet (17), Tiebing Liang (9), Suthat Liangpunsakul (18), Steven Masson (10), Philippe Mathurin (19), Romain Moirand (20), Andrew McQuillin (21), Christophe Moreno (22), Marsha Y. Morgan (23), Sebastian Mueller (24), Beat Müllhaupt (25), Laura E. Nagy (26), Pierre Nahon (27-29), Bertrand Nalpas (17,30), Sylvie Naveau (31), Pascal Perney (32), Munir Pirmohamed (33), Helmut K. Seitz (24), Michael Soyka (34), Felix Stickel (25), Andrew Thompson (33,35), Mark R. Thursz (5), Eric Trépo (22), Timothy R. Morgan* (7,36), Devanshi Seth*[#] (15,16,37), for the GenomALC Consortium.

(1) Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Queensland 4029, Australia, (2) Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis IN, USA, (3) Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, International Centre for Life, Central Parkway, Newcastle upon Tyne NE1 3BZ, United Kingdom, (4) NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals and the University of Nottingham, Nottingham NG7 2UH, United Kingdom, (5) Department of Metabolism, Digestion & Reproduction, Imperial College London, UK, (6) Center for Liver Diseases, University of Pittsburgh Medical Center, 3471 Fifth Avenue, Pittsburgh, PA 15213, USA, (7) Department of Veterans Affairs, VA Long Beach Healthcare System, 5901 East Seventh Street, Long Beach, CA 90822, USA, (8) F. Widjaja Family Foundation Inflammatory Bowel and Immunobiology Research Institute,

Cedars-Sinai Medical Center, Los Angeles, California CA 90048, USA (9) Department of Medicine, Indiana University, Indianapolis, IN 46202-5175, USA, (10) Faculty of Medical Sciences, Newcastle University Medical School, Framlington Place, Newcastle upon Tyne NE2 4HH, United Kingdom, (11) Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, United Kingdom, (12) Division of Clinical Toxicology, Department of Internal Medicine 2, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Ismaninger Str. 22, 81675 Munich, Germany, (13) Liver Unit, Sheffield Teaching Hospitals, AO Floor Robert Hadfield Building, Northern General Hospital, Sheffied S5 7AU, UK, (14) Laboratory of Neurogenetics, NIAAA, Rockville, MD 20852, USA, (15) Drug Health Services, Royal Prince Alfred Hospital, Missenden Road, Camperdown, NSW 2050, Australia, (16) Faculty of Medicine and Health, the University of Sydney, Sydney, NSW 2006, Australia, (17) Service Addictologie, CHRU Caremeau, 30029 Nîmes, France, (18) Division of Gastroenterology and Hepatology, Department of Medicine, Indiana University and Roudebush Veterans Administration Medical Center, Indianapolis, USA, (19) CHRU de Lille, Hôpital Claude Huriez, Rue M. Polonovski CS 70001, 59 037 Lille Cedex, France, (20) Univ Rennes, INRA, INSERM, CHU Rennes, Institut NUMECAN (Nutrition Metabolisms and Cancer), F-35000 Rennes, France, (21) Molecular Psychiatry Laboratory, Division of Psychiatry, University College London, London WC1E 6DE, UK , (22) CUB Hôpital Erasme, Université Libre de Bruxelles, clinique d'Hépatologie, Brussels, Belgium; Laboratory of Experimental Gastroenterology, Université Libre de Bruxelles, Brussels, Belgium, (23) UCL Institute for Liver & Digestive Health, Division of Medicine, Royal Free Campus, University College London, London NW3 2PF, UK, (24) Department of Internal Medicine, Salem Medical Center and Center for Alcohol Research, University of Heidelberg, Zeppelinstraße 11-33, 69121 Heidelberg, Germany, (25) Department of Gastroenterology and Hepatology, University Hospital Zurich, Rämistrasse 100, CH-8901 Zurich, Switzerland, (26) Lerner

Research Institute, 9500 Euclid Avenue, Cleveland, Ohio, OH 44195, USA, (27) Service d'Hépatologie, APHP Hôpital Avicenne et Université Paris 13, Bobigny, France, (28) University Paris 13, Bobigny, France, (29) Inserm U1162 Génomique fonctionnelle des tumeurs solides, Paris, France, (30) DISC, Inserm, 75013 Paris, France, (31) Hôpital Antoine-Béclère, 157 Rue de la Porte de Trivaux, 92140 Clamart, France, UM1, INSERM U1018 (32) Hôpital Universitaire Caremeau, Place du Pr. Robert Debre, 30029 Nîmes, France, (33) MRC Centre for Drug Safety Science, Liverpool Centre for Alcohol Research, University of Liverpool, The Royal Liverpool and Broadgreen University Hospitals NHS Trust, and Liverpool Health Partners, Liverpool, L69 3GL, UK, (34) Psychiatric Hospital University of Munich, Nussbaumsstr.7, 80336 Munich, Germany and Privatklinik Meiringen, Willigen, CH 3860 Meiringen, Switzerland, (35) Health Analytics, Lane Clark & Peacock LLP, London, UK, (36) Department of Medicine, University of California, Irvine, USA, (37) Centenary Institute of Cancer Medicine and Cell Biology, the University of Sydney, Sydney, NSW 2006, Australia.

All authors except first three and last two are in alphabetical order

*Equal senior authors

#Corresponding authors

1. Dr Devanshi Seth, Centenary Institute of Cancer Medicine and Cell Biology, The University of Sydney, Sydney, NSW 2006, Australia. d.seth@sydney.edu.au

2. Dr John B. Whitfield, Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Queensland 4029, Australia. John.Whitfield@qimrberghofer.edu.au

KEYWORDS

Hepatocellular carcinoma; risk stratification; chronic alcohol use; genome wide association; single nucleotide polymorphism; coffee

Word count: 3546

Numbers of figures and tables: Tables 4; Figures 2. Supplemental data: Tables 5, Figures 4.

CONFLICT OF INTEREST

NPC has a number of consulting agreements with and research grants from the pharmaceutical industry but they are not significantly or directly related to this paper. MP receives research funding from various organisations including the MRC and NIHR. He has also received partnership funding for the following: MRC Clinical Pharmacology Training Scheme (co-funded by MRC and Roche, UCB, Eli Lilly and Novartis); a PhD studentship jointly funded by EPSRC and Astra Zeneca; and grant funding from Vistagen Therapeutics. He has also unrestricted educational grant support for the UK Pharmacogenetics and Stratified Medicine Network from Bristol-Myers Squibb and UCB. He has developed an HLA genotyping panel with MC Diagnostics, but does not benefit financially from this. He is part of the IMI Consortium ARDAT (www.ardat.org). None of these funding sources were deployed in the undertaking of this study. TRM has conducted clinical research with AbbVie, Genfit, Gilead, and Merck but none of these are related to this manuscript.

FINANCIAL SUPPORT: Funding for this study was provided by NIH/NIAAA UO1AA018389 and RO1AA018389 for data collection, analysis, interpretation and patient recruitment. SL: U01 grant from the NIAAA and R01 AA025208, U01 AA026917, and

1I01CX000361. MT: The UK Medical Research Council Stratified Medicine Award (Ref MR/R014019/1), NIHR Imperial Biomedical Research Centre and NIHR Senior Investigator Award (NIHR 200153).

AUTHOR CONTRIBUTIONS

DS, CPD, TRM, PM, PSH, HKS, JBW, BN, FS, TF, AKD and HJC conceived and designed the study. Recruitment and data acquisition was done for GenomALC-1 by GPA, FE, DGl, J-MJ, SL, SMa, PM, RM, TRM, SMu, BM, PN, BN, SN, PP, MP, HKS, DS, MS, FS, AT; and for GenomALC-2 by GPA, SA, RB, NPC, AKD, FE, DGl, DGo, SMa, PM, CM, AM, MM, TRM, LEN, DS, FS, AT, MT, ET. Genetic analysis for SNP information was performed by THS-A, HJC, RD, TF. TL facilitated DNA processing for genotyping. JBW and DS led the analyses and writing of the manuscript. All authors read, critically reviewed and approved the final version. DS and TRM are the guarantors.

ABSTRACT (273 words)

Background & Aims: Only a minority of excess alcohol drinkers develop cirrhosis. We developed and evaluated risk stratification scores to identify those at highest risk.

Methods. Three cohorts (GenomALC-1: n=1690, GenomALC-2: n=3037, UK Biobank: relevant n=6898) with a history of heavy alcohol consumption (≥80 g/day (men), ≥50 g/day (women), for ≥10 years) were included. Cases were participants with alcohol-related cirrhosis. Controls had a history of similar alcohol consumption but no evidence of liver disease. Risk scores were computed from up to eight genetic loci identified previously as associated with alcohol-related cirrhosis and three clinical risk factors. Score performance for the stratification of alcohol-related cirrhosis risk was assessed and compared across the alcohol-related liver disease spectrum, including hepatocellular carcinoma (HCC).

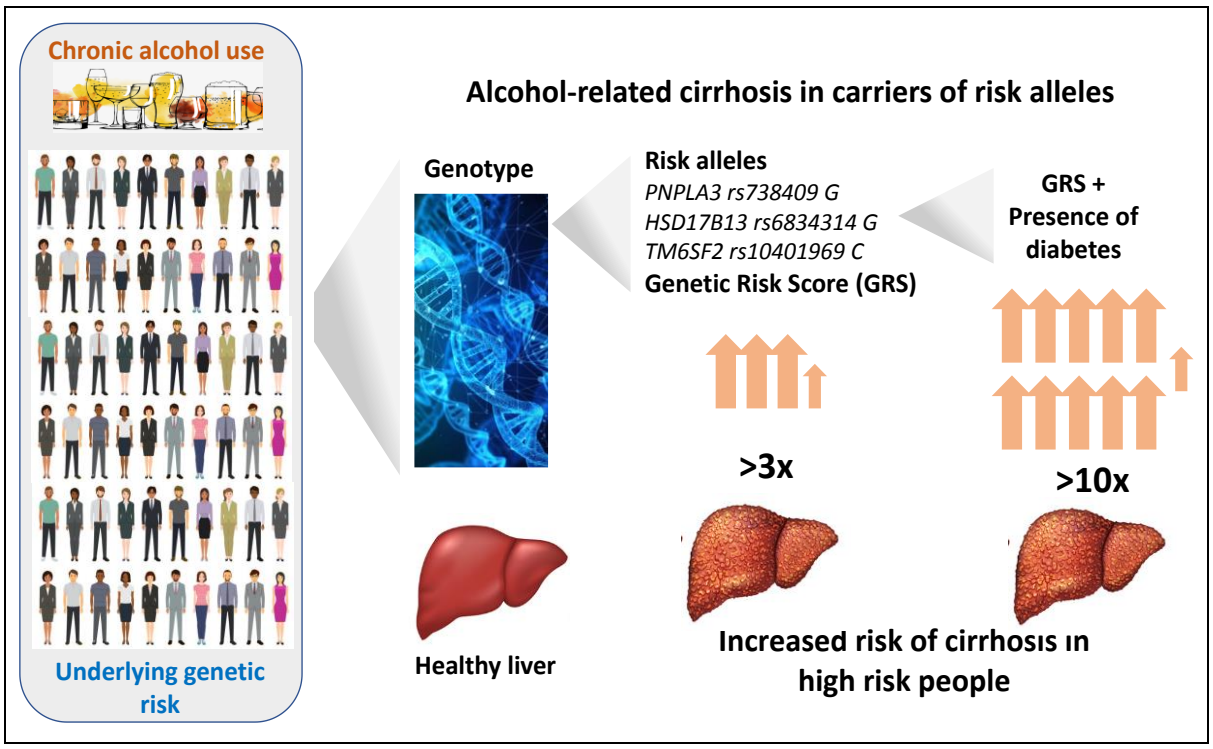
Results: A combination of three single nucleotide polymorphisms (SNPs) (*PNPLA3:*rs738409, *SUGP1-TM6SF2:*rs10401969, *HSD17B13:*rs6834314) and diabetes status best discriminated for cirrhosis risk. The odds ratio (OR) and 95% confidence intervals (CI) for the extreme score quintiles (Q1-Q5) of the 3-SNP score, based on independent allelic effect size estimates, were 5.99 (4.18;8.60) (GenomALC-1); 2.81 (2.03;3.89) (GenomALC-2); and 3.10 (2.32;4.14) (UK Biobank). Patients with diabetes and high-risk score, compared to those without diabetes and a low-risk score, had ORs increased to 14.7 (7.69;28.1) (GenomALC-1) and 17.1 (11.3;25.7) (UK Biobank). Patients with cirrhosis and HCC had significantly higher mean risk scores than patients with cirrhosis alone (0.76±0.06 versus 0.61±0.02, p=0.007). Score performance was not significantly enhanced by information on additional genetic risk variants, body mass index or coffee consumption.

Conclusions: A risk score based on three genetic risk variants and diabetes status can provide meaningful risk stratification for cirrhosis in excess drinkers, allowing earlier prevention planning including intensive intervention.

LAY SUMMARY

Excessive chronic drinking leads to liver cirrhosis in some people, but so far there is no way to identify those at high risk of developing this debilitating disease. Our study has developed a genetic risk score (GRS) test that can identify patients at high risk and shows that the risk of cirrhosis is increased >10-fold with just two risk factors - diabetes and high GRS. Risk assessment using this test has potential for early and personalised management of this disease in high-risk patients.

GRAPHICAL ABSTRACT



INTRODUCTION

Although the risk for developing cirrhosis is positively associated with alcohol consumption, only a minority of people with high-risk alcohol intake develop cirrhosis. The prevalence can vary between 7-16%[1,2] with some reports suggesting the prevalence to be as low as 2%[3,4]. The risk threshold for what is considered high-risk intake has changed over time[5-7]. Long-term consumption of 80 grams per day (g/d) or more is associated with increased risk of cirrhosis[8,9], but the threshold for liver harm is below this level, especially for women[10,11]. The 80 g/d (men) and 50 g/d (women) cut-offs were set at a relatively high level to ensure both the cirrhosis and control groups were exposed to a substantial level of alcohol-related risk. We used this threshold to define "heavy drinking" in this study.

Primary prevention of alcohol-related liver disease (ALD) would involve decreasing alcohol intake of the whole population but achieving this remains challenging. Focused intervention through the identification of people with high alcohol intake or more specifically through stratification of individuals within this population at risk for developing cirrhosis depends on identification of those at high risk. Evidence from clinical trials[12] suggests that informing excessive drinkers that they have abnormal liver function tests/hepatic fibrosis can motivate them to reduce their alcohol intake. A number of both constitutional[13-16], and genetic[17-20] risk factors for the development of alcoholic-related cirrhosis have been identified, but no attempt appears to have been made, to date, to bring these together to provide an integrated measure of risk. Thus, the aim of this study was to devise risk scores for the stratification of cirrhosis risk and evaluate them in heavy drinkers from three independent cohorts.

MATERIALS/PATIENTS AND METHODS

Information on disease status, genotypes and clinical risk factors was available for three cohorts: i) GenomALC-1 and ii) GenomALC-2 from the GenomALC consortium, and iii) the UK Biobank. Details of the recruiting and contributing sites, with numbers of patients by diagnosis and by country are given below and in Supplementary Table 1. Cohort characteristics of the cases and controls from each source are described in Supplementary Table 2.

GenomALC-1

The GenomALC-1 cohort was recruited according to a pre-designed protocol between 2012 and 2017 in Australia, France, Germany, Switzerland, the UK, and the USA. The recruitment criteria and the data collection protocol were detailed previously[21]. Briefly, all participants had a history of heavy drinking (≥80 g/d (men) and ≥50 g/d (women) for ≥10 years). For cases, cirrhosis had been diagnosed by a combination of clinical criteria, laboratory variables and/or liver elastography (Fibroscan®), with liver biopsy if clinically indicated. Clinical features defining the severity of cirrhosis are shown in Supplementary Table 2. Other liver diseases (hepatitis B or C, haemochromatosis, Wilson's disease, and autoimmune hepatitis) were excluded by laboratory testing or clinical criteria. For controls, liver disease was excluded through a combination of clinical history and measurement of liver function tests (bilirubin, albumin, ALT). For both cases and controls, HIV infection was an exclusion criterion. The study was approved by appropriate Ethics Committees or Institutional Review Boards at each site and conformed to the ethical guidelines of the 1975 Declaration of Helsinki. Participants were provided with explanations of the study and gave written informed consent. Genotyping was performed at Erasmus University Medical Centre, Rotterdam using the Illumina GSA genotyping array, as described[20].

GenomALC-2

The biological samples and data were donated by research groups who had independently collected them for other studies. Some of the GenomALC-2 samples were included in a previous GWAS[17]; therefore, for the purposes of this study, overlapping samples were removed from the analysis. Clinical diagnosis of cases and controls was similar to GenomALC-1 criteria but detailed clinical information was limited for this cohort. Patients had given informed consent and the studies were approved by the appropriate Ethics Review Boards. DNA from these participants' samples was also genotyped as outlined above for GenomALC-1.

Genotypes in the GenomALC-1 and GenomALC-2 cohorts were [22] cleaned using a widely used quality control pipeline, the GWASTools package https://bioconductor.org/packages/devel/bioc/manuals/GWASTools/man/GWASTools.pdf and imputed to 1000 Genomes reference using the Michigan Imputation Server (MIS)[22]

UK Biobank

The UK Biobank[23] includes approximately 500,000 volunteers from the UK with a wide range of data including computer-administered questionnaires, physical measurements, laboratory tests, and genotyping. All participants gave informed consent, consistent with the UK Biobank Ethics and Governance Framework. Recruitment and initial assessment occurred between 2006 and 2010 when participants were aged 40 to 69 years. Access to the UK Biobank database was obtained (Application 18870) and relevant data (with diagnoses updated to June 2020) were extracted. For cases, information was restricted to assigned clinical diagnosis (Supplementary Table 2) on hospital admissions and diagnoses, and on causes of death in participants who have subsequently died. Information was available on self-reported alcohol intake at the time of assessment (by beverage type, in drinks per week, or for less frequent drinkers per month), and participants also reported whether this was less than, similar to or more than they had been consuming 10 years previously. The amounts were converted to express the alcohol intake in g/d. Participants who had a recorded diagnosis or cause of death of alcohol-related cirrhosis

(ICD-10 K70.3, 'Alcoholic cirrhosis of liver') from hospital records or death certificates were included as cases (n=594), and those with reported drinking above the 80 or 50 g/day limits, with similar or greater consumption 10 years before, but with no diagnosed liver disease (either alcohol-related or other causes) were included as controls (n=6304). Exclusion criteria for UK Biobank subjects were similar to GenomALC-1.

UK Biobank also included 758 cases within the spectrum of other alcohol-related liver disease diagnoses (Supplementary Table 1). Genotype data for the relevant UK Biobank participants were downloaded from the server and genotypes for the relevant SNPs were extracted. Data on coffee consumption, body mass index (BMI) and diabetes status were recorded (Supplementary Table 2).

Data curation and statistical analysis

Data management and statistical analyses used IBM SPSS Statistics, version 22 (IBM Corp., New York NY). Binary variables were coded as 0 (absent) or 1 (present). Diabetes status (absent/present), BMI, kg/m^2) and coffee consumption (0: not a coffee consumer, 1: coffee consumer) shown in our previous report as associated with cirrhosis[16] were also modelled. Genotype data were coded as single nucleotide polymorphisms (SNPs) minor allele dosages, assuming an additive model for allelic effects.

Calculation of risk scores requires coefficients for the effect sizes associated with each risk factor, and assessment of the performance of the risk scores requires testing in independent cohorts not included in the derivation of these coefficients. The scheme shown in Table 1 sets out the basis for the scores and the data-sets which were used for evaluation.

SNPs with the lowest p-value at three loci (*PNPLA3:*rs738409, *SUGP1-TM6SF2:*rs10401969 and *HSD17B13:*rs6834314) were selected based on previous association with the risk of alcohol-related cirrhosis[17,18], and confirmed at genome-wide significance in our meta-

analysis[20]. Two significantly associated SNPs have been reported at *SUGP1-TM6SF2* locus[17] which are in near-complete linkage disequilibrium (d'1.00, r^2 0.955), and rs10401969 was chosen over rs58542926 because of its stronger association with cirrhosis.

A score based on these three loci ('3-SNP score') was computed for each participant in each of the three cohorts. Minor allele counts ('dosage') were obtained from direct or imputed genotypes for each SNP, multiplied by the beta coefficients for allelic effect sizes (derived from published odds ratios, calculated as beta = $\log_e$(OR)) and summed across SNPs (Table 1). The means for 3-SNP scores were also compared between disease diagnostic groups in the three independent cohorts described in Supplementary Table 1.

Scores based on three, five, and eight loci were also computed for the GenomALC-2 samples using coefficients of loci with significant association from the published meta-analysis[20] or other sources[17,18] ('3-SNP-M', '5-SNP-M' and '8-SNP-M' scores) (Table 1). The 3-SNP-M score was based on the loci mentioned above, the 5-SNP-M score included above three loci, and *SERPINA1 and FAF2* identified in our meta-analysis, and the 8-SNP-M score which was derived from the 5-SNP-M score with addition of three reported loci (*MBOAT7*, *MTARC1* [previously *MARC1*], *HNRNPUL1*) significantly associated with alcohol-related cirrhosis[17,24,25].

Area under the ROC curve (AUC) analysis and logistic regressions (with the score as the predictor variable and case/control status as an outcome) were performed. Odds Ratios (ORs) of the score were compared for extreme quintiles (highest Q5 against lowest Q1).

RESULTS

Risk stratification by genetic loci-based scores

Results in the three study cohorts for the 3-SNP score AUCs, logistic regressions and the ORs comparing quintiles Q5 and Q1 of the score, are shown in Table 2. Each of these measures showed better performance of the score in the GenomALC-1 cohort than in either the GenomALC-2 or UK Biobank cohorts, and there was no significant difference in score between men and women (Supplementary Table 3).

The results of adding two clinical risk factors (BMI and coffee consumption) to the 3-SNP score are shown in Table 2. Because the beta-coefficients for the two clinical risk factors were derived from the GenomALC-1 cohort, and information on these factors was not available for the GenomALC-2 cohort, this score was only evaluated against the UK Biobank data. A moderate, but not significant, improvement in risk stratification was observed following addition of these clinical risk factors; the Q5-Q1 OR estimate increased from 3.10 to 3.37 but the 95% confidence intervals overlapped. Coffee data did not improve the risk stratification, and nor did BMI (which was non-significant in the UK Biobank group and not available for GenomALC-2) (Table 2). Stratification of risk including the clinical factors in the score showed similar results for men and women (Supplementary Table 3).

The addition of further loci in the 5-SNP-M score (*PNPLA3:*rs2294915, *SUGP1-TM6SF2:*rs10401969, *HSD17B13:*rs10433937, *SERPINA1:*rs28929474, *FAF2:*rs11134997)[17,24,25] and in the 8-SNP-M score, with *MBOAT7:rs641738*, *MTARC1:*rs2642438 and *HNRNPUL1:*rs17251589 in addition to those in the 5-SNP-M score, did not improve the associations between score and outcome or the risk stratification (Table 2). Because the coefficients for *FAF2* and *SERPINA1* were obtained from the meta-analysis of the GenomALC-1, Buch study[17] and UK Biobank data, the 5-SNP-M and 8-SNP-M scores could only be tested in the GenomALC-2 data. To allow a valid comparison between the multi-SNP scores each was based on the coefficients from our meta-analysis of GWAS results. This resulted in an improvement for the meta-analysis-based 3-SNP-M score compared to the 3-

SNP score (Q5-Q1 ORs changed from 2.81 [95% CI 2.03,3.89] to 3.65 [2.59,5.15]). There was also a high correlation between the 3-SNP and 3-SNP-M scores in GenomALC-2 ($r = 0.826$, $n = 3037$, $p < 10^{-200}$; Supplementary Figure 1).

Clinical utility of the risk score

Numerical cut-offs that define or quantify risk are needed if the risk score is to have clinical utility. The 3-SNP scores in the GenomALC-1 cases and controls for the lowest and highest quintile boundaries were close to 0 and 1 (0.033 and 0.964, respectively; Figure 1). Division of the scores into three groups at low, intermediate and high cirrhosis risk was based on the 3-SNP score distribution (Supplementary Figure 2). The final selected scores were, low: <0; intermediate >0 - 0.7 and high risk >0.7. In each study cohort the risk difference between the low- and high-risk groups ranged between 2.5-fold and approximately 5-fold (Table 3). The difference in risk between the high- and low-risk GenomALC-1 groups were similar across the six countries (Figure 2).

Diabetes

Diabetes is known to have a large effect on cirrhosis risk. Inclusion of diabetes status with genetic risks in a combined risk score led to a bimodal distribution and difficulty in defining score quintiles. Thus, to see the effect of genetic risk score in the context of diabetes status, the 3-SNP score was subdivided by the diabetes status and is presented separately (Table 4).

People with diabetes showed a substantial increase in the risk of cirrhosis in both the GenomALC-1 (OR 3.82, 95% CI 2.67;5.47) and the UK Biobank (OR 5.62, 95% CI 4.33;7.28) cohorts. The genetic score effects were similar for people with and without diabetes, both in the GenomALC-1 (logistic regression coefficients ± SE, no diabetes: 1.055 ± 0.105; diabetes: 1.276 ± 0.338) and the UK Biobank data (no diabetes: 0.653 ± 0.093; diabetes: 0.735 ± 0.181). Tests for genetic score-diabetes interaction, either by including a (score x diabetes) term in the

logistic regression or by testing for heterogeneity of Odds Ratios between those with and without diabetes, showed no evidence for interaction effects in either cohort (Table 4). The combined effects of having diabetes and a high genetic risk score resulted in a >10-fold increased risk in people with diabetes and a high risk 3-SNP score against people without diabetes and a low-risk score, for both GenomALC-1 (OR 14.7, 95% CI 7.69;28.1) and the UK Biobank (OR 17.1, 95% CI 11.3;25.7) (Table 4).

Genetic loci-based risk scores across alcohol-related liver diseases

The mean values for the 3-SNP score varied across groups defined by alcohol intake and by the diagnostic categories for alcohol-related liver disease for both GenomALC-1 and the UK Biobank cohorts (Supplementary Figure 3). *Post hoc* comparisons showed similar trends of mean 3-SNP risk score increasing with disease severity for the GenomALC-2 cohort that included excessive drinkers with no liver disease and significantly differed between cases with severe alcoholic hepatitis and alcohol-related cirrhosis ($p = 0.011$) (Supplementary Table 4). Mean 3-SNP score increased with severity of liver disease (Supplementary Figure 3), including when comparing cirrhosis with HCC against cirrhosis without HCC, both for GenomALC-1 (0.757 ± 0.057 versus 0.613 ± 0.019) and UK Biobank (0.717 ± 0.102 versus 0.396 ± 0.031); see also Supplementary Table 5.

DISCUSSION

This study shows that a genetic score based on three lead SNPs associated at genome-wide significance with the risk for developing alcohol-related cirrhosis, can risk-stratify people drinking at potentially harmful levels.

Development of score for risk stratification

The performance of 3-SNP score improved considerably when used in conjunction with information on diabetes status, providing a powerful tool for identifying patients at high risk for developing advanced alcohol-related liver diseases. Higher scores were also associated with other severe liver injuries, including alcoholic hepatitis and HCC.

Our main measure of genetic risk stratification was to compare people who are in the highest quintile for a score against those in the lowest quintile, providing a more practical measure of stratification success than comparing the most extreme of all possible categories, which will usually contain only a small proportion of people[26]. Substantial Q5-Q1 risk differences were evident for the simple 3-SNP score in each of the cohorts; approximately six-fold in the GenomALC-1 cohort and three-fold in the other cohorts (Table 2). The greater difference in Q5-Q1 risk for GenomALC-1 is likely to be due to a more refined and pre-defined case-control definition for the recruitment protocol in this cohort.

Diabetes status led to a substantial enhancement of the utility of the 3-SNP score, predicting a >10-fold difference in risk between extreme groups (Q5 with diabetes and Q1 non-diabetes). Adding information on further genetic risk variants or BMI and coffee consumption had minimal effect.

Clinical utility of risk-score

Clinical application of a score requires the definition of decision points in numerical terms rather than by reference to population quintiles. However, Q5-Q1 comparisons can be useful for comparison across cohorts, such as in our study, and against genetic scores for other diseases. For clinical application boundaries of 0 and 0.7 were set for the 3-SNP score that provided a potentially useful stratification of risk in each of the three cohorts. As expected, lowering the high-risk threshold (e.g. from 1.0 to 0.7) identified a higher proportion of the cases as being at high risk but the ORs between the high- and low-risk groups decreased. For any

classification based on a numerical test or score, changing the cut-off point(s) will alter the trade-off between test sensitivity and specificity and the optimum cut-offs will depend on the use to be made of the test. The prevalence of the condition is also important because this will affect the predictive value of positive or negative results. The AUCs shown in Table 2 were significant but when the desired test sensitivity was set at 80% (to flag nearly all those at high risk) the test specificities were between 30% and 40%. Thus, a substantial number of false positives must be accepted, making the score suitable for risk stratification but not for prediction of outcome in individual patients.

The 3-SNP risk score was also associated with differences across the alcohol-related liver disease spectrum, including HCC. The HCC risk association is consistent with previous information showing that *PNPLA3*, *HSD17B13* and *TM6SF2* polymorphisms[27-31] are associated with a higher risk for this condition compared to advanced cirrhosis, perhaps suggesting a pro-oncogenic role for these variants.

Scope of risk-score

The loci comprising the current risk score are also implicated in the risk for developing cirrhosis of diverse aetiologies. Using similar polygenic risk scores (PRS) in non-alcoholic fatty liver disease (NAFLD) revealed that combining genetic and clinical features refines the predictive utility of the algorithm for identifying those at higher risk of severe liver disease[32-34]. Given the many shared genetic and metabolic risks between alcohol-related liver disease and NAFLD, the predictive algorithm defined here may have a wider scope across these diseases for risk stratification of those at higher risk of cirrhosis. Recently Emdin and colleagues[30] identified 12 variants, five previously known, including *PNPLA3*, *HSD17B13* and *TM6SF*, and seven novel, which were associated at genome-wide significance with 'any cause' cirrhosis , and aggregated these into a PRS. A high PRS, defined as the top quintile of the

distribution, was associated with significantly increased risk of cirrhosis compared with the lowest quintile (OR 2.26; P < .001). Our current study indicates that risk stratification for alcohol-related cirrhosis can be achieved as effectively using fewer genetic markers, and with algorithms based on a smaller base of GWAS information, presumably because the genetic architecture of alcohol-related cirrhosis includes a number of common variants with substantial effects on risk.

Preliminary investigation of adding previously reported risk loci over the 3-SNP score did not significantly improve risk stratification. To develop a robust PRS that incorporates many loci for alcohol-related cirrhosis risk would require a larger population based cohort. Another possible extension, again dependent on the availability of more data, would be to incorporate information on patients' alcohol consumption in addition to genotyping for genetic variants associated with cirrhosis risk.

The outcome of risk stratification for alcohol-related liver disease can be compared with PRS approaches to other complex diseases, including cardiovascular disease and cancers. A recent study[35] showed that for five common diseases (coronary heart disease, type 2 diabetes, atrial fibrillation, breast cancer and prostate cancer), Q1-Q5 differences in PRS were associated with approximately two- to five-fold differences in the cumulative prevalence of diagnosis by age 80. Our 3-SNP score performance was equal to or slightly better than these.

The main strengths of this study were that it employed three large independent cohorts and that the case and control definitions were standardised. The study also had its limitations. First, the included populations were of largely European ancestry so that the finding may not be universally applicable. Second, an unknown proportion of the controls, especially in the UK Biobank cohort, may have undiagnosed alcohol-related liver disease, although it should be recognised that misclassification of some cases as controls would lead to poorer stratification

such that the effectiveness of our score would be under-, rather than over-estimated. Finally, the risk scores were derived from groups of heavy drinkers with cirrhosis or without liver disease. However, these were validated in case and control groups selected from the population-based UK Biobank cohort. Application of the risk score to an individual patient should be performed with an understanding that some patients' outcomes will differ from those predicted by the score. Prospective studies are needed, both to relate score to progression across time in patients who present with early stages of liver disease, and to clarify the relationship between onset of diabetes and of advanced liver disease in patients with excessive alcohol use.

Based on the findings of the present study a 3-SNP score algorithm is proposed for use and interpretation of the risk stratification in heavy drinkers (Box 1).

Box 1. ***Use of the 3-SNP risk score for alcohol-related cirrhosis.***

Calculate the risk score as:
(0.7839**PNPLA3* **rs738409 G dosage) + (0.5423****SUGP1-TM6SF2* **rs10401969 C dosage) – (0.4463****HSD17B13* **rs6834314 G dosage)**
Assign the patient to the appropriate stratum of risk, as follows:

	Score less than 0	**Score above 0.7**
	Low risk	High risk
Relative risk if <u>not</u> diabetic	1	3-fold
Relative risk if diabetic	3-fold or more	Over 10-fold

(Patients with scores between 0 and 0.7 are at intermediate risk)

When making use of this risk information, after appropriate explanation, consent and genotyping, be aware that this is a risk stratification scheme rather than providing individual predictions. Some patients whose score places them in the low-risk group will progress to significant liver disease, especially if they continue to drink excessively.

Conclusions

An algorithm for stratifying the risk of developing alcohol-related cirrhosis among heavy drinkers, based on three genetic loci and information on diabetic status, has been developed

and validated. It is intended to identify patients at particularly high risk for developing alcohol-related cirrhosis. In addition to stratifying risk of developing alcohol-related cirrhosis, this algorithm may also stratify risk for developing alcoholic hepatitis and HCC. This risk stratification system could be used to facilitate management of all people at risk for developing significant alcohol-related liver disease.

ACKNOWLEDGEMENTS

We acknowledge Ms Donna Sheedy, Ms Julia Stevens and Prof Jillian Krill for providing access to brain tissue (included in GenomALC-2) from the New South Wales Brain Tissue Resource Centre at the University of Sydney, NSW 2006, Australia (collection of tissues reported in this publication was supported by the National Institute of Alcohol Abuse and Alcoholism of the National Institutes of Health under Award Number R28AA012725. The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health). We are grateful to the four Alcoholic Hepatitis Consortia TREAT, InTEAM, DASH and SCAHC (U01 AA021886) from the USA, for providing DNA contributing towards GenomALC-2.

REFERENCES

1. Askgaard G, Leon DA, Kjaer MS, Deleuran T, Gerds TA, Tolstrup JS. Risk for alcoholic liver cirrhosis after an initial hospital contact with alcohol problems: A nationwide prospective cohort study. *Hepatology* 2017; **65**(3): 929-37.
2. Askgaard G, Kjaer MS, Tolstrup JS. Opportunities to Prevent Alcoholic Liver Cirrhosis in High-Risk Populations: A Systematic Review With Meta-Analysis. *Am J Gastroenterol* 2019; **114**(2): 221-32.
3. Hasin DS, Stinson FS, Ogburn E, Grant BF. Prevalence, correlates, disability, and comorbidity of DSM-IV alcohol abuse and dependence in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Arch Gen Psychiatry* 2007; **64**(7): 830-42.
4. Wong T, Dang K, Ladhani S, Singal AK, Wong RJ. Prevalence of Alcoholic Fatty Liver Disease Among Adults in the United States, 2001-2016. *JAMA* 2019; **321**(17): 1723-5.
5. Pequignot G, Tuyns AJ, Berta JL. Ascitic cirrhosis in relation to alcohol consumption. *IntJEpidemiol* 1978; **7**(2): 113-20.
6. Rehm J, Taylor B, Mohapatra S, et al. Alcohol as a risk factor for liver cirrhosis: a systematic review and meta-analysis. *Drug Alcohol Rev* 2010; **29**(4): 437-45.
7. Askgaard G, Gronbaek M, Kjaer MS, Tjonneland A, Tolstrup JS. Alcohol drinking pattern and risk of alcoholic liver cirrhosis: a prospective cohort study. *J Hepatol* 2015; **62**(5): 1061-7.
8. Corrao G, Bagnardi V, Zambon A, Arico S. Exploring the dose-response relationship between alcohol consumption and the risk of several alcohol-related conditions: a meta-analysis. *Addiction* 1999; **94**(10): 1551-73.
9. Lelbach WK. Epidemiology of alcoholic liver disease. *ProgLiver Dis* 1976; **5**: 494-515.

10. Roerecke M, Vafaei A, Hasan OSM, et al. Alcohol Consumption and Risk of Liver Cirrhosis: A Systematic Review and Meta-Analysis. *Am J Gastroenterol* 2019; **114**(10): 1574-86.
11. Tuyns AJ, Pequignot G. Greater risk of ascitic cirrhosis in females in relation to alcohol consumption. *IntJEpidemiol* 1984; **13**(1): 53-7.
12. Subhani M, Knight H, Ryder S, Morling JR. Does Advice Based on Biomarkers of Liver Injury or Non-Invasive Tests of Liver Fibrosis Impact High-Risk Drinking Behaviour: A Systematic Review With Meta-analysis. *Alcohol Alcohol* 2021.
13. Hart CL, Morrison DS, Batty GD, Mitchell RJ, Davey Smith G. Effect of body mass index and alcohol consumption on liver disease: analysis of data from two prospective cohort studies. *BMJ* 2010; **340**: c1240.
14. Liangpunsakul S, Puri P, Shah VH, et al. Effects of Age, Sex, Body Weight, and Quantity of Alcohol Consumption on Occurrence and Severity of Alcoholic Hepatitis. *Clin Gastroenterol Hepatol* 2016; **14**(12): 1831-8 e3.
15. Saab S, Mallam D, Cox GA, 2nd, Tong MJ. Impact of coffee on liver diseases: a systematic review. *Liver Int* 2014; **34**(4): 495-504.
16. Whitfield JB, Masson S, Liangpunsakul S, et al. Obesity, Diabetes, Coffee, Tea, and Cannabis Use Alter Risk for Alcohol-Related Cirrhosis in 2 Large Cohorts of High-Risk Drinkers. *Am J Gastroenterol* 2020; **116**(1): 106-15.
17. Buch S, Stickel F, Trepo E, et al. A genome-wide association study confirms PNPLA3 and identifies TM6SF2 and MBOAT7 as risk loci for alcohol-related cirrhosis. *Nat Genet* 2015; **47**(12): 1443-8.
18. Abul-Husn NS, Cheng X, Li AH, et al. A Protein-Truncating HSD17B13 Variant and Protection from Chronic Liver Disease. *N Engl J Med* 2018; **378**(12): 1096-106.

19. Tian C, Stokowski RP, Kershenobich D, Ballinger DG, Hinds DA. Variant in PNPLA3 is associated with alcoholic liver disease. *Nat Genet* 2010; **42**(1): 21-3.
20. Schwantes-An TH, Darlay R, Mathurin P, et al. Genome-wide association study and meta-analysis on alcohol-related liver cirrhosis identifies novel genetic risk factors. *Hepatology* 2021; **73**(5): 1920-31.
21. Whitfield JB, Rahman K, Haber PS, et al. Brief report: genetics of alcoholic cirrhosis-GenomALC multinational study. *Alcohol Clin Exp Res* 2015; **39**(5): 836-42.
22. Das S, Forer L, Schonherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet* 2016; **48**(10): 1284-7.
23. Sudlow C, Gallacher J, Allen N, 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**(3): e1001779.
24. Emdin CA, Haas ME, Khera AV, et al. A missense variant in Mitochondrial Amidoxime Reducing Component 1 gene and protection against liver disease. *PLoS Genet* 2020; **16**(4): e1008629.
25. Innes H, Buch S, Hutchinson S, et al. Genome-Wide Association Study for Alcohol-Related Cirrhosis Identifies Risk Loci in MARC1 and HNRNPUL1. *Gastroenterology* 2020; **159**(4): 1276-89 e7.
26. Gellert-Kristensen H, Richardson TG, Davey Smith G, Nordestgaard BG, Tybjaerg-Hansen A, Stender S. Combined Effect of PNPLA3, TM6SF2, and HSD17B13 Variants on Risk of Cirrhosis and Hepatocellular Carcinoma in the General Population. *Hepatology* 2020; **72**(3): 845-56.
27. Salameh H, Raff E, Erwin A, et al. PNPLA3 Gene Polymorphism Is Associated With Predisposition to and Severity of Alcoholic Liver Disease. *Am J Gastroenterol* 2015; **110**(6): 846-56.

28. Yang J, Trepo E, Nahon P, et al. A 17-Beta-Hydroxysteroid Dehydrogenase 13 Variant Protects From Hepatocellular Carcinoma Development in Alcoholic Liver Disease. *Hepatology* 2019; **70**(1): 231-40.

29. Tang S, Zhang J, Mei TT, et al. Association of TM6SF2 rs58542926 T/C gene polymorphism with hepatocellular carcinoma: a meta-analysis. *BMC Cancer* 2019; **19**(1): 1128.

30. Trépo E, Nahon P, Bontempi G, et al. Association between the PNPLA3 (rs738409 C\textgreaterG) variant and hepatocellular carcinoma: Evidence from a meta-analysis of individual participant data. *Hepatology* 2014; **59**(6).

31. Stickel F, Lutz P, Buch S, et al. Genetic Variation in HSD17B13 Reduces the Risk of Developing Cirrhosis and Hepatocellular Carcinoma in Alcohol Misusers. *Hepatology* 2020; **72**(1): 88-102.

32. Bianco C, Casirati E, Malvestiti F, Valenti L. Genetic predisposition similarities between NASH and ASH: Identification of new therapeutic targets. *JHEP Rep* 2021; **3**(3): 100284.

33. Bianco C, Jamialahmadi O, Pelusi S, et al. Non-invasive stratification of hepatocellular carcinoma risk in non-alcoholic fatty liver using polygenic risk scores. *Journal of Hepatology* 2021; **74**(4): 775-82.

34. De Vincentis A, Tavaglione F, Jamialahmadi O, et al. A Polygenic Risk Score to Refine Risk Stratification and Prediction for Severe Liver Disease by Clinical Fibrosis Scores. *Clin Gastroenterol Hepatol* 2021.

35. Mars N, Koskela JT, Ripatti P, et al. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat Med* 2020; **26**(4): 549-57.

Table 1. Score construction and validation plan.

		Cohorts available for independent validation		
		GenomALC-1 (N=1690)	GenomALC-2 (N=3037)	UK Biobank (N=6898)
1	3-SNP score, using SNPs and coefficients from initial reports[17,18] **= (0.7839*_PNPLA3_ rs738409 G dosage) + (0.5423*_SUGP1-TM6SF2_ rs10401969 C dosage) – (0.4463*_HSD17B13_ rs6834314 G dosage)**	Yes	Yes	Yes
2	3-SNP score as in 1 above, with addition of BMI and coffee **= [1] + (0.0709*BMI) – (0.645*Coffee)**	No (BMI and coffee coefficients are derived from this cohort)	No (no information of BMI and coffee)	Yes
3	3-SNP-M score, using SNPs and coefficients from meta-analysis[20] **= (0.7274*_PNPLA3_ rs2294915 T dosage) + (0.3988*_SUGP1_ rs10401969 C dosage) – (0.2485*_HSD17B13_ rs10433937 G dosage)**	No*	Yes	No*
4	5-SNP-M score; as in 3 above but with addition of two GW-significant SNPs from meta-analysis **= [3] + (0.6419*_SERPINA1_ rs28929474 T dosage) – (0.2357*_FAF2_ rs11134997 C dosage)**	No*	Yes	No*
5	8-SNP-M score; as in 4 but with three additional SNPs with genome-wide significant associations with alcohol-related liver disease **= [4] + (0.1446*_MBOAT7_ rs641738 T dosage) - (0.2401*_MTARC1_ rs2642438 A dosage) - (0.1304*_HNRNPUL1_ rs17251589 T dosage)**	No*	Yes	No*

*SNP coefficients are derived from this cohort

Table 2. Results of ROC curve and logistic regression analyses, and estimated odds ratios for cirrhosis between the lowest (Q1) and highest (Q5) quintiles of scores.

		ROC Curve	Logistic regression		Q1-Q5 Odds Ratio (95% CIs)
		AUC	Beta	p-value	
3-SNP score[i]	GenomALC-1	0.665 ± 0.014	1.092 ± 0.099	2.90×10^{-28}	5.99 (4.18 to 8.60)
	GenomALC-2	0.606 ± 0.014	0.669 ± 0.090	1.44×10^{-13}	2.81 (2.03 to 3.89)
	UK Biobank	0.619 ± 0.014	0.729 ± 0.080	1.06×10^{-19}	3.10 (2.32 to 4.14)
3 SNP score[i] + BMI, coffee	GenomALC-1	Not estimated[iii]	Not estimated[iii]		Not estimated[iii]
	GenomALC-2	Not estimated[iii]	Not estimated[iv]		Not estimated[iv]
	UK Biobank	0.636 ± 0.015	0.748 ± 0.073	1.77×10^{-24}	3.37 (2.38 to 4.78)
Comparisons based on coefficients from meta-analysis:					
3-SNP-M score[ii]	GenomALC-2	0.631 ± 0.014	0.909 ± 0.103	1.17×10^{-18}	3.65 (2.59 to 5.15)
5-SNP-M score[ii]	GenomALC-2	0.626 ± 0.014	0.813 ± 0.096	2.96×10^{-17}	3.66 (2.62 to 5.12)
8-SNP-M score[ii]	GenomALC-2	0.633 ± 0.014	0.807 ± 0.091	6.06×10^{-19}	3.37 (2.43 to 4.66)

[i] Coefficients estimated from Buch et al[17] and Abul-Husn et al[18]

[ii] Coefficients estimated from meta-analysis data Schwantes-An et al[20]

[iii] Not estimated because coefficients would be partly based on data for this cohort.

[iv] Not estimated because BMI and coffee data are not available for this cohort.

Table 3. Simplification of scoring system into three groups based on numerical values of the 3-SNP score.

		Odds Ratios (95% confidence intervals)		
Risk group	score	GenomALC-1	GenomALC-2	UK Biobank
Low	≤ 0	1 N = 273 (16.2%)	1 N = 327 (18.5%)	1 N = 3403 (56.1%)
Intermediate	> 0 to 0.70	2.13 (1.61 to 2.83) N = 731 (43.3%)	1.54 (1.18 to 2.00) N = 771 (43.7%)	1.36 (1.04 to 1.77) N = 1207 (19.9%)
High	> 0.70	4.96 (3.67 to 6.71) N = 686 (40.6%)	2.67 (2.02 to 3.53) N = 668 (37.8%)	2.654(2.16 to 3.29) N = 1456 (24.0%)

Table 4. Risk of alcohol-related cirrhosis by diabetes status, and comparison of risk in the low- and high-risk of the 3-SNP score stratified by diabetes status. For GenomALC-1, diabetes status was at time of recruitment and for UK Biobank at the time of (baseline) assessment. Information on diabetes was not available for the GenomALC-2 group. Only those participants with information on diabetes, and a 3-SNP score, are included.

Predictor	Group	Contrast	Odds Ratios (95% Confidence Intervals)	
			GenomALC-1	UK Biobank
Diabetes		Diabetes versus no diabetes	3.82 (2.67 to 5.47)	5.62 (4.33 to 7.28)
3-SNP Score	No diabetes	≤0 versus >0.7 in non-diabetics	4.77 (3.45 to 6.58)	2.37 (1.86 to 3.03)
	Diabetes	≤0 (diabetes) versus >0.7 (diabetes)	5.32 (2.06 to 13.7)[1]	3.74 (2.16 to 6.48)[2]
		≤0 (no-diabetes) versus >0.7 (diabetes)	14.7 (7.69 to 28.1)	17.1 (11.3 to 25.7)

[1] Breslow-Day test for homogeneity of Odds Ratios in Non-Diabetes and Diabetes groups, GenomALC-1 $\chi^2 = 0.05$ p = 0.830.

[2] Breslow-Day test for homogeneity of Odds Ratios in Non-Diabetes and Diabetes groups, UK Biobank $\chi^2 = 2.20$ p = 0.138.

Figure 1. Distribution of 3-SNP scores in cases and controls from the GenomALC-1 data, showing the boundaries of the lowest (Q1) and highest (Q5) quintiles at 0.033 and 0.964, respectively (dotted lines).

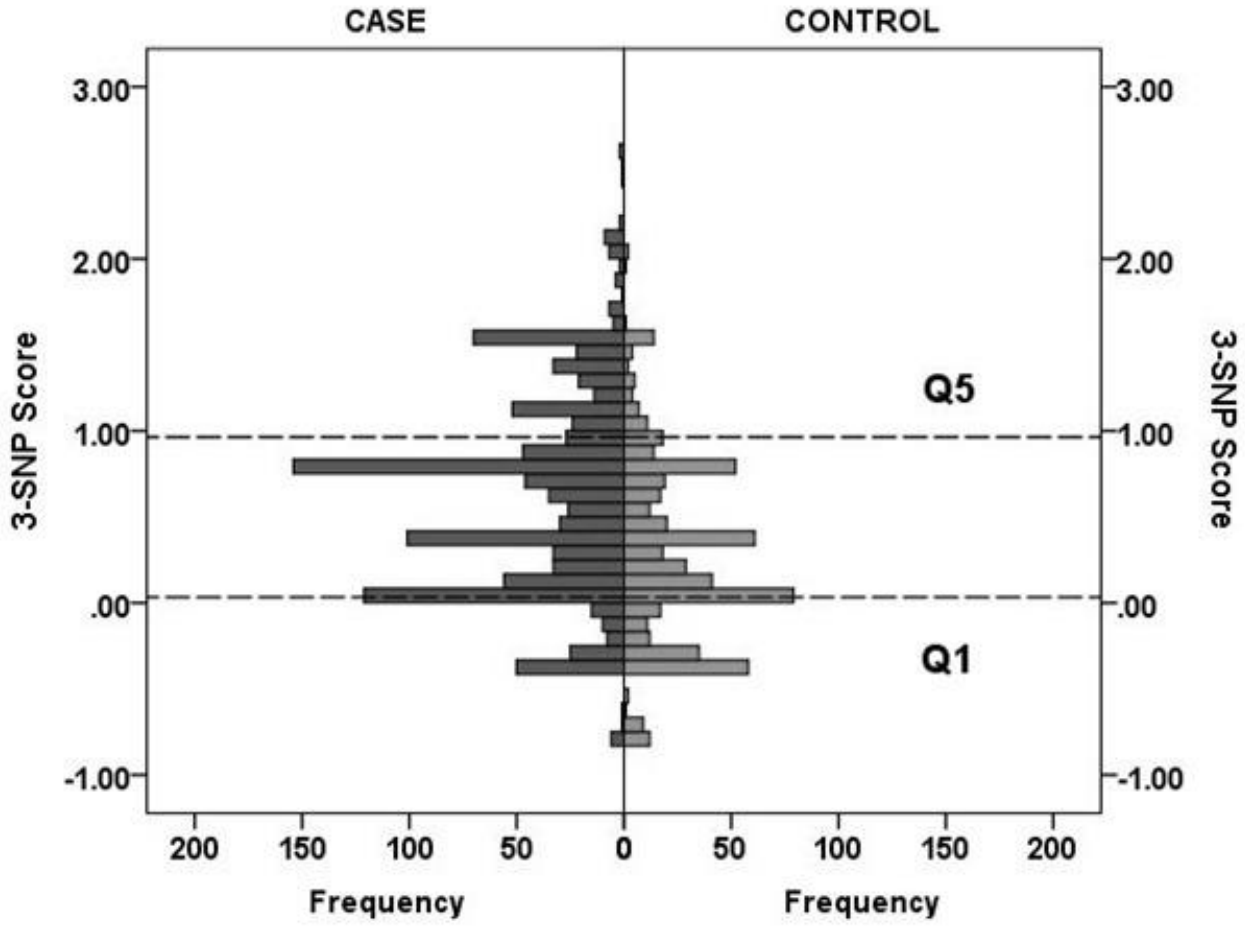
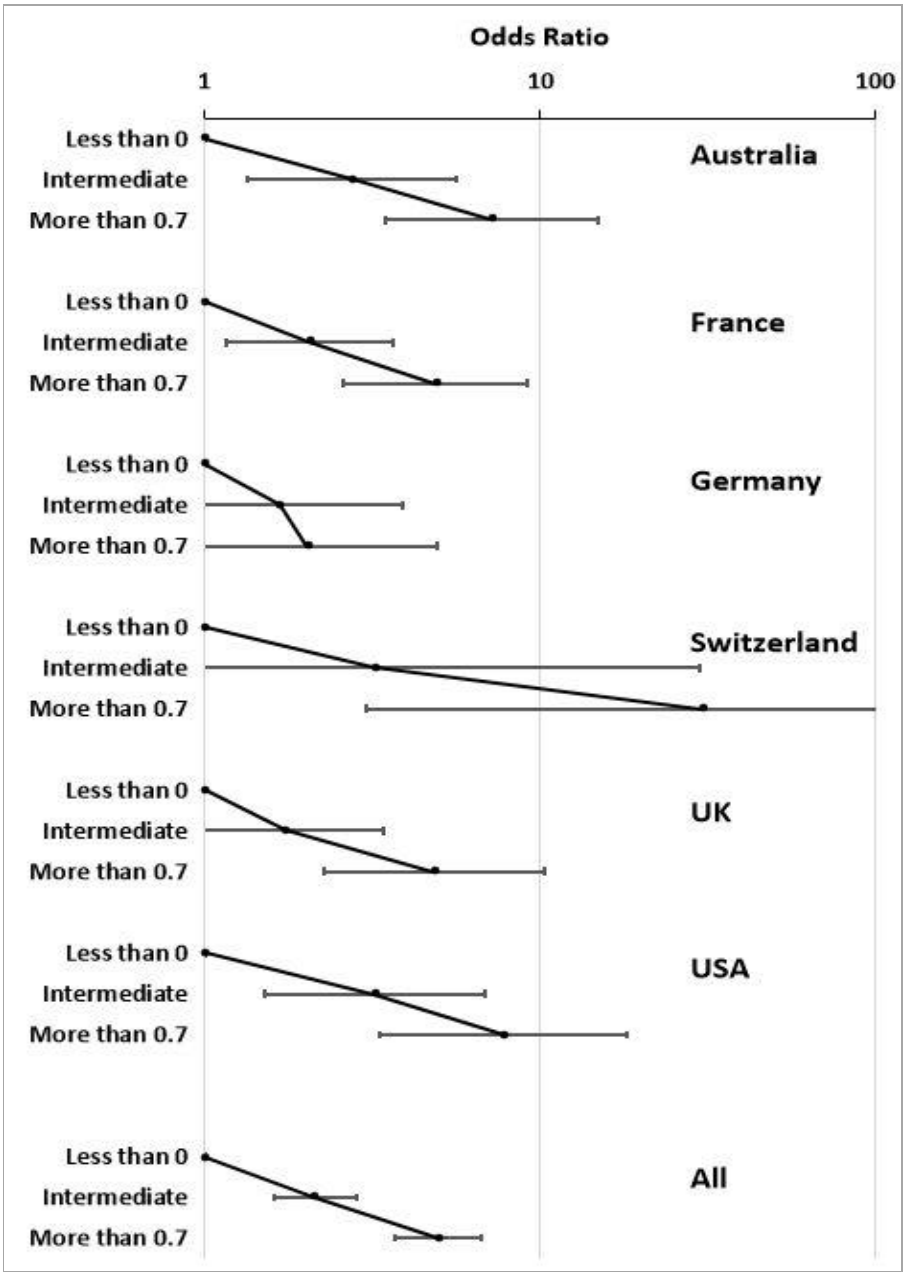


Figure 2. Odds ratios, by country and overall, for the risk of alcohol-related cirrhosis in the GenomALC-1 cohort when results for the 3-SNP score are divided into low (<0), intermediate (0 to 0.7) and high (>0.7) categories. Error bars show 95% confidence intervals.



SUPPLEMENTARY MATERIAL

A GENETIC RISK SCORE AND DIABETES PREDICTS DEVELOPMENT OF ALCOHOL-RELATED CIRRHOSIS IN DRINKERS

Supplementary Table 1. Numbers of participants, and diagnostic categories, by country of recruitment. Restricted to participants with relevant genotyping and clinical risk factor data.

GenomALC-1 samples

	Control	Alcohol-related cirrhosis	Total
Australia	175	129	304
France	103	373	476
Germany	142	75	217
Switzerland	33	28	61
UK	77	257	334
USA	58	240	298
Total	588	1102	1690

GenomALC-2 samples

	Control	Alcoholic hepatitis	Alcohol-related cirrhosis	Total
Australia	68	0	36	104
Belgium	258	223	663	1144
France	58	0	85	143
Germany	31	0	15	46
Switzerland	3	0	0	3
UK	184	771	356	1311
USA	2	277	7	286
Total	604	1271	1162	3037

UK Biobank samples

No relevant diagnosis	494,910	
Excessive drinker, no liver diagnosis	6304	
Alcoholic fatty liver	*95*	
Alcoholic liver disease, unspecified	*428*	
Alcoholic fibrosis and sclerosis	*17*	
Alcoholic hepatitis	*130*	
Alcoholic hepatic failure	*88*	
Alcoholic cirrhosis	594	
Alcoholic cirrhosis without HCC		542
Alcoholic cirrhosis with HCC		52
Total	502,566	

Note for UK Biobank cohort: Total number excludes people who withdrew consent and those with unknown sex or alcohol intake. Numbers in italics are for alcohol-related liver diseases other than cirrhosis.

Supplementary Table 2. Demographic, clinical and substance use characteristics of participants included.

		GenomALC-1 (N = 1390)				GenomALC-2 (N = 1766) [1]				UK Biobank (N= 6898) [2]			
		Cases (N =917)		Controls (N = 473)		Cases (N = 1162)		Controls (N = 604)		Cases (N = 594)		Controls (N = 6304)	
		Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
		Mean ± SD (N)	Mean ± SD (N)	Mean ± SD (N)	Mean ± SD (N)	Mean ± SD (N)	Mean ± SD (N)	Mean ± SD (N)	Mean ± SD (N)	Mean ± SD (N)	Mean ± SD (N)	Mean ± SD (N)	Mean ± SD (N)
Demographics	Age (Mean ± SD, in years)	53.0 ± 8.5 (674)	51.0 ± 8.9 (243)	50.0 ± 9.5 (331)	50.6 ± 9.7 (142)	55.2 ± 10.4 (702)	54.5 ± 10.1 (288)	47.4 ± 10.9 (394)	50.4 ± 10.1 (186)	57.8 ± 6.9 (477)	57.2 ± 7.7 (117)	56.5 ± 7.6 (4836)	54.8 ± 7.7 (1468)
	Years of Education	12.0 ± 3.5 (665)	11.5 ± 303 (240)	12.0 ± 3.9 (326)	12.8 ± 3.9 (141)					11.3 ± 2.2 (357)	11.2 ± 2.2 (83)	11.4 ± 2.1 (3785)	11.7 ± 2.0 (927)
	BMI, kg/m^2	28.0 ± 5.6 (673)	26.1 ± 6.2 (242)	25.8 ± 4.7 (330)	25.4 ± 5.8 (142)					29.1 ± 5.4 (466)	27.4 ± 5.0 (117)	28.4 ± 4.5 (4836)	26.9 ± 4.8 (1460)
	European ethnicity/race (by self-report)	99.6%	99.6%	99.7%	98.6%	100%	100%	99.7%	99.5%	96.2%	94.8%	98.5%	97.4%
Alcohol use	Alcohol intake, g/day	262 ± 431 (674)	189 ± 349 (243)	251 ± 298 (331)	186 ± 100 (142)					49.0 ± 55.0 (477)	22.4 ± 26.8 (117)	104.1 ± 27.7 (4836)	66.3 ± 19.5 (1468)
	Age started XS drinking	26.9 ± 9.8 (674)	31.3 ± 10.6 (242)	25.5 ± 9.0 (329)	29.7 ± 10.8 (141)								
	Years of high-risk drinking	25.0 ± 11.2 (674)	19.1 ± 9.1 (243)	21.6 ± 9.3 (331)	18.4 ± 7.4 (142)								
	Audit Score	10.5 ± 10.6 (673)	11.6 ± 11.5 (241)	26.8 ± 9.4 (330)	26.8 ± 10.1 (140)								
	Lifetime alcohol intake, kg	2310 ± 4034 (674)	1316 ± 2548 (243)	2073 ± 3327 (331)	1244 ± 892 (142)								
Lab results	Haemoglobin (g/L)	117 ± 26 (653)	113 ± 21 (239)	147 ± 14 (315)	135 ± 14 (136)								
	INR (ratio)	1.40 ± 0.43 (609)	1.54 ± 0.58 (223)	0.99 ± 0.17 (272)	0.97 ± 0.11 (111)	1.44 ± 0.52 (51)	1.44 ± 0.48 (134)	1.11 ± 0.41 (51)	1.04 ± 0.11 (16)				
	Albumin (g/L)	34.5 ± 6.8 (623)	34.8 ± 7.5 (227)	43.2 ± 5.1 (314)	43.2 ± 5.7 (132)	35.0 ± 7.3 (336)	33.5 ± 7.1 (128)	40.9 ± 10.7 (100)	40.8 ± 9.7 (45)	42.5 ± 4.7 (407)	43.1 ± 4.4 (101)	45.5 ± 2.7 (4175)	45.5 ± 2.7 (1250)
	Bilirubin (µmol/L)	60.6 ± 100.5 (661)	81± 121 (243)	9.1 ± 5.8 (320)	8.2 ± 5.6 (140)			12.8 ± 5.3 (59)	10.2 ± 4.5 (32)	15.3 ± 12.3 (443)	12.1 ± 7.9 (109)	9.9 ± 4.1 (4505)	8.3 ± 3.3 (1357)
	Creatinine (µmol/L)	94 ± 67 (661)	105 ± 486 (242)	75 ± 17 (324)	62 ± 15 (141)					77 ± 34 (445)	59 ± 15 (110)	77 ± 14 (4536)	62 ± 12 (1365)

	ALT (unit/L)	36.8 ± 40.2 (661)	34.3 ±33.2 (242)	43.3 ± 46.3 (326)	38 ± 38 (141)	54.6 ± 219.2 (342)	34.2 ± 25.7 (132)	45.6 ± 45.2 (240)	34.3 ± 43.9 (122)	45.0 ± 38.7 (444)	38.0 ± 26.3 (110)	31.1 ± 19.1 (4529)	21.9 ± 12.9 (1366)
	AST (unit/L)	60.2 ± 51.0 (654)	64.5 ± 51.3 (238)	43.6 ± 41.0 (320)	41 ± 38 (140)	90.1 ± 307.7 (343)	64.9 ± 51.5 (132)	52.2 ± 56.5 (242)	36.2 ± 22.6 (124)	61.3 ± 46.1 (442)	62.4 ± 51.0 (109)	33.1 ± 19.2 (4504)	26.7 ± 12.6 (1357)
	GGT (unit/L)	225 ± 371 (623)	173 ± 264 (222)	124 ± 255 (315)	122 ± 216 (139)					233 ± 245 (434)	216 ± 239 (108)	86.5 ± 87.8 (4528)	46.8 ± 56.6 (1366)
Liver disease	MELD score	11.2 ± 7.4 (604)	11.1 ± 9.1 (222)	1.9 ± 3.3 (266)	-0.7 ± 3.1 (110)								
	Number with ascites (ever)	520/674 (77%)	189/243, (78%)	0	0	294/572 (51%)	131/241 (54%)	None recorded	None recorded				
	Number with oesophageal varices (ever)	358/669 (53%)	120/239 (49%)	0	0	49/177 (28%)	23/87 (26%)	None recorded	None recorded				
	Number with encephalopathy (ever)	209/651 (31%)	91/240 (37%)	0	0	72/531 (14%)	24/226 (11%)	None recorded	None recorded				
	Number with HCC (ever)	105/674 (16%)	10/243 (4%)	0	0	9/157 (6%)	None recorded	None recorded	None recorded	49/477 (10.3%)	3/117 (2.6%)	0/4721 (0%)	0/1458 (0%)
	Number abstinent for ≥ 60 days	234/674 (35%)	154/243 (63%)	8/332 (2%)	3/142 (2%)	Not recorded	Not recorded	3/394 (0.8%)	1/186 (0.5%)				
	Number (%) with known diabetes	159/674 (24%)	39/243 (16%)	26/332 (8%)	5/142 (4%)					109/468 (23.3%)	16/117 (13.7%)	271/4813 (5.6%)	34/1466 (2.3%)
Cannabis use	Regular use, 5+ years, during period of high alcohol use	67/673 (10%)	8/142 (3%)	90/329 (27%)	24/142 (17%)								
	Years of regular use (if ever regular user)	17.5 ± 14.2 (66)	15.9 ± 11.0 (7)	15.5 ± 10.1 (89)	14.6 ± 10.4 (24)								
	Days per week marijuana	4.6 ± 2.3 (67)	5.1 ± 2.2 (8)	5.3 ± 2.4 (89)	5.7 ± 1.9 (24)								
	Occasions total	4424 ± 4742 (66)	4520 ± 4119 (7)	4374 ± 3795 (89)	4279 ± 3063 (24)								
Smoking history	Regular smoker (ever)	513/674 (76%)	156/243 (64%)	279/331 (84%)	111/142 (78%)					342/473 (72.5%)	78/116 (67.2%)	3589/4825 (74.4%)	1093/1464 (74.7%)
	Pack years (if ever smoker)	35.4 ± 32.3 (370)	24.1 ± 18.4 (96)	35.3 ± 27.8 (174)	33.8 ± 21.3 962)								
Coffee intake	Regular use, 5+ years, during period of high alcohol use	398/674 (59%)	113/243 (47%)	240/331 (73%)	88/142 (62%)								
	Years coffee (if regular user)	32.2 ± 12.9 (347)	28.9 ± 13.3 (100)	25.3 ± 11.8 (225)	24.2 ± 11.3 (81)								
	Cups caffeinated coffee per day (if regular user)	3.7 ± 3.5 (347)	3.7 ± 3.9 (100)	4.1 ± 3.6 (224)	3.4 ± 2.3 (81)					1.87 ± 2.23 (430)	1.80 ± 2.35 (110)	2.23 ± 2.36 (4438)	2.10 ± 2.05 (1378)

Supplementary Table 3. Comparison of score performance measures in men and women.

			ROC Curve	Logistic regression		Q1-Q5 Odds Ratio
			AUC	Beta	p-value	(95% CIs)
3-SNP score	GenomALC-1	Men	0.671 ± 0.016	1.132 ± 0.116	1.41×10^{-22}	6.18 (4.05 to 9.41)
		Women	0.650 ± 0.027	0.974 ± 0.192	3.79×10^{-7}	5.40 (2.67 to 10.92)
	GenomALC-2	Men	0.592 ± 0.017	0.575 ± 0.107	6.76×10^{-8}	2.47 (1.68 to 3.62
		Women	0.635 ± 0.025	0.897 ± 0.172	1.94×10^{-7}	3.81 (2.05 to 7.07)
	UK Biobank	Men	0.635 ± 0.016	0.800 ± 0.088	1.14×10^{-19}	3.44 (2.48 to 4.77)
		Women	0.554 ± 0.036	0.375 ± 0.196	0.056	2.08 (1.11 to 3.89)
3 SNP score + BMI, coffee	UK Biobank	Men	0.645 ± 0.017	0.795 ± 0.082	4.55×10^{-22}	3.50 (2.34 to 5.24)
		Women	0.594 ± 0.034	0.535 ± 0.164	0.0011	2.79 (1.37 to 5.71)
3-SNP-M score[i]	GenomALC-2	Men	0.626 ± 0.017	0.873 ± 0.124	1.94×10^{-12}	3.79 (2.50 to 5.75)
		Women	0.638 ± 0.025	0.978 ± 0.186	$1..42 \times 10^{-7}$	3.27 (1.77 to 6.06)
5-SNP-M score[i]	GenomALC-2	Men	0.622 ±0.017	0.791 ±0.116	7.87×10^{-12}	3.47 (2.31 to 5.23)
		Women	0.632 ±0.025	0.848 ± 0.174	1.09×10^{-6}	3.92 (2.18 to 7.03)
8-SNP-M score	GenomALC-2	Men	0.631 ± 0.017	0.789 ± 0.110	6.37×10^{-13}	3.34 (2.24 to 4.98)
		Women	0.635 ±0.025	0.836 ± 0.162	2.55×10^{-7}	3.33 (1.90 to 5.85)

Supplementary Table 4. Significance (p-values, not adjusted for multiple comparisons) for contrasts between groups of excessive drinkers with and without alcohol-related liver disease diagnoses in UK Biobank and GenomALC-2 cohorts. The dependent variable is the 3-SNP score. Boxes in the UK Biobank table emphasise the significant differences between the control groups and the more severe forms of liver disease, and the lack of significant differences among the more severe categories.

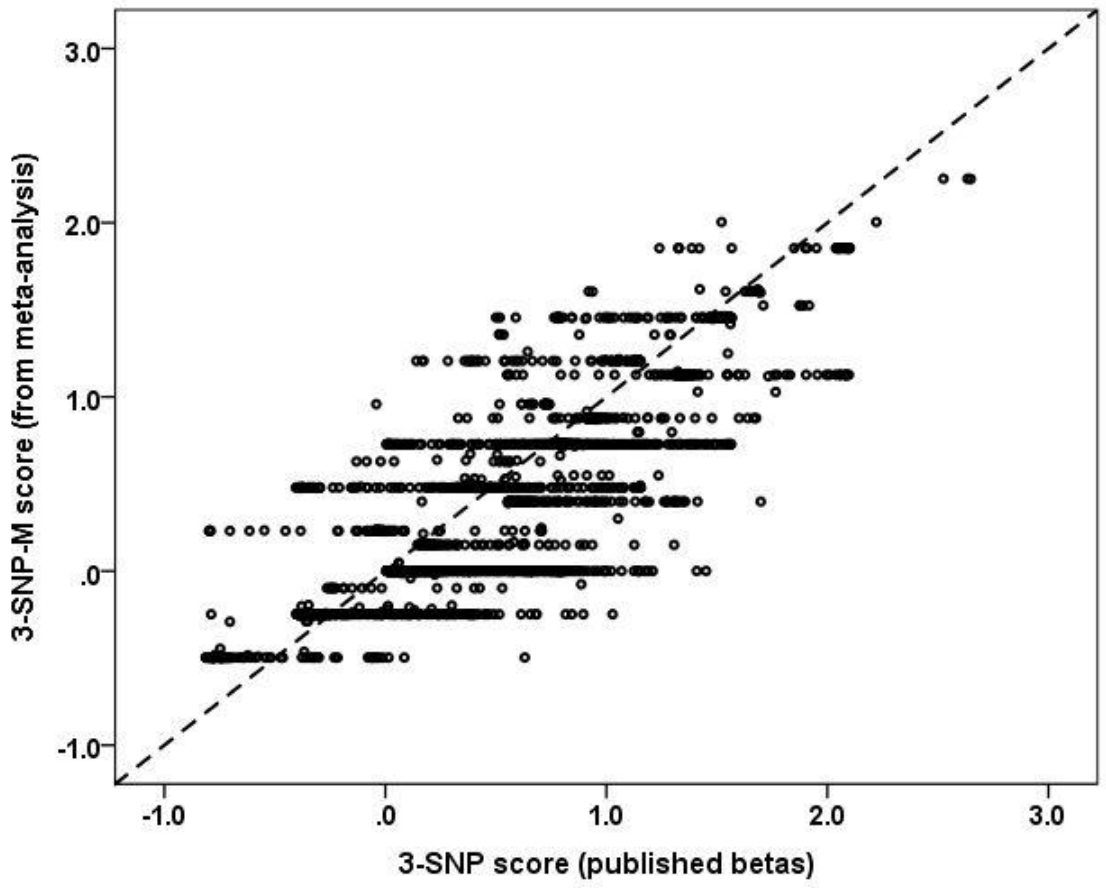
UK Biobank	N	Excessive drinker, no liver diagnosis	Alcoholic fatty liver	ALD unspecified	Alcoholic fibrosis sclerosis	Alcoholic hepatitis	Alcoholic hepatic failure	Alcoholic cirrhosis
Excessive drinker	5618	-						
Fatty liver	71	0.128	-					
ALD unspecified	327	0.0074	0.824	-				
Alcoholic Fibrosis/sclerosis	910	0.222	0.545	0.466	-			
Alcoholic hepatitis	796	7.75×10^{-4}	0.293	0.095	0.904	-		
Alcoholic hepatic failure	52	0.0023	0.191	0.073	0.922	0.667	-	
Alcoholic Cirrhosis	448	2.51×10^{-20}	0. 0332	3.30×10^{-5}	0.8831	0.335	0.814	-

GenomALC-2	N	Control	Alcoholic hepatitis	Alcoholic cirrhosis
Control	604	-		
Alcoholic hepatitis	1271	7.95×10^{-9}	-	
Alcoholic cirrhosis	1162	1.20×10^{-14}	0.011	-

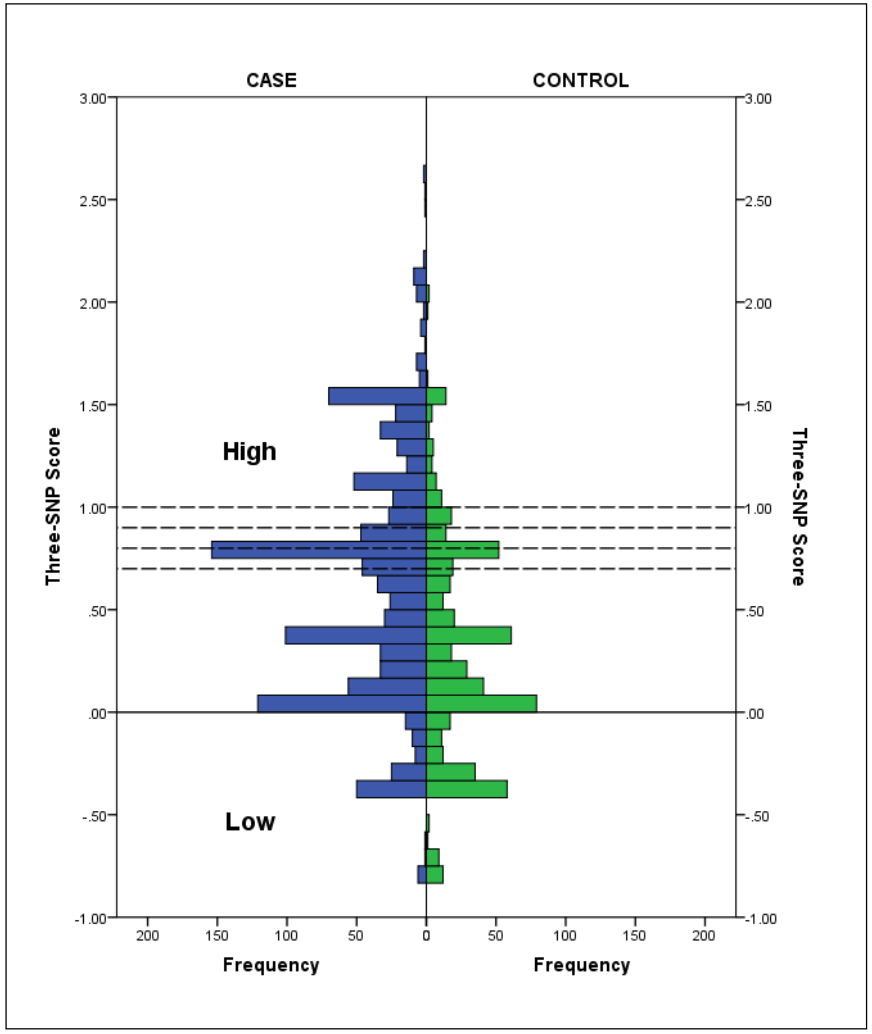
Supplementary Table 5. Effects of 3-SNP score on risk of alcohol-related liver diseases, graded by severity of liver disease, in GenomALC-1 sample and UK Biobank. In each cohort Controls were people who reported daily alcohol intake of ≥80 grams (men)/≥ 50 grams (women), for ≥10 years. UK Biobank participants with 'mild' alcoholic liver disease had ICD-10 diagnoses of K70.0 (alcoholic fatty liver) or K70.9 (alcoholic liver disease, unspecified); there was no comparable group in the GenomALC participants. 'Severe' alcoholic liver disease comprised ICD-10 diagnoses of K70.1 (alcoholic hepatitis), K70.3 (alcoholic cirrhosis) or K70.4 (alcoholic hepatic failure) for UK Biobank, and alcoholic cirrhosis for GenomALC. HCC, hepatocellular carcinoma (ICD-10 C22.0). Coefficients (B) and Odds Ratios (OR) are expressed per unit change in the 3-SNP score.

	GenomALC-1					UK Biobank				
	B	SE	p-value	OR	95% CI	B	SE	p-value	OR	95% CI
Control (excessive drinkers) versus Mild ALD	N/A					0.271	0.089	0.0023	1.31	1.10 to 1.56
Control (excessive drinkers) versus 'Severe' ALD but no HCC	1.020	0.101	3.49×10^{-24}	2.773	2.277 to 3.377	0.652	0.074	8.77×10^{-19}	1.92	1.66 to 2.22
Control (excessive drinkers) versus 'Severe' ALD with HCC	1.327	0.174	2.05×10^{-14}	3.769	2.683 to 5.296	1.264	0.229	3.38×10^{-8}	3.54	2.26 to 5.54
'Mild' ALD versus 'Severe' ALD but no HCC	N/A					0.378	0.111	6.50×10^{-4}	1.46	1.17 to 1.81
'Severe' ALD but no HCC versus 'Severe' ALD with HCC	0.401	0.155	0.010	1.493	1.102 to 2.022	0.631	0.242	0.0091	1.88	1.17 to 3.02

Supplementary Figure 1. Comparison of the 3-SNP and 3-SNP-M scores, for patients in the GenomALC-2 cohort calculated as in Table 1. The diagonal line shows x = y.



Supplementary Figure 2. Effect of varying the 3-SNP score cut-off for classification into the high-risk group on the proportion of cases stratified as high-risk and the Odds Ratios comparing the high-and low-risk groups. Data shown are for GenomALC-1.

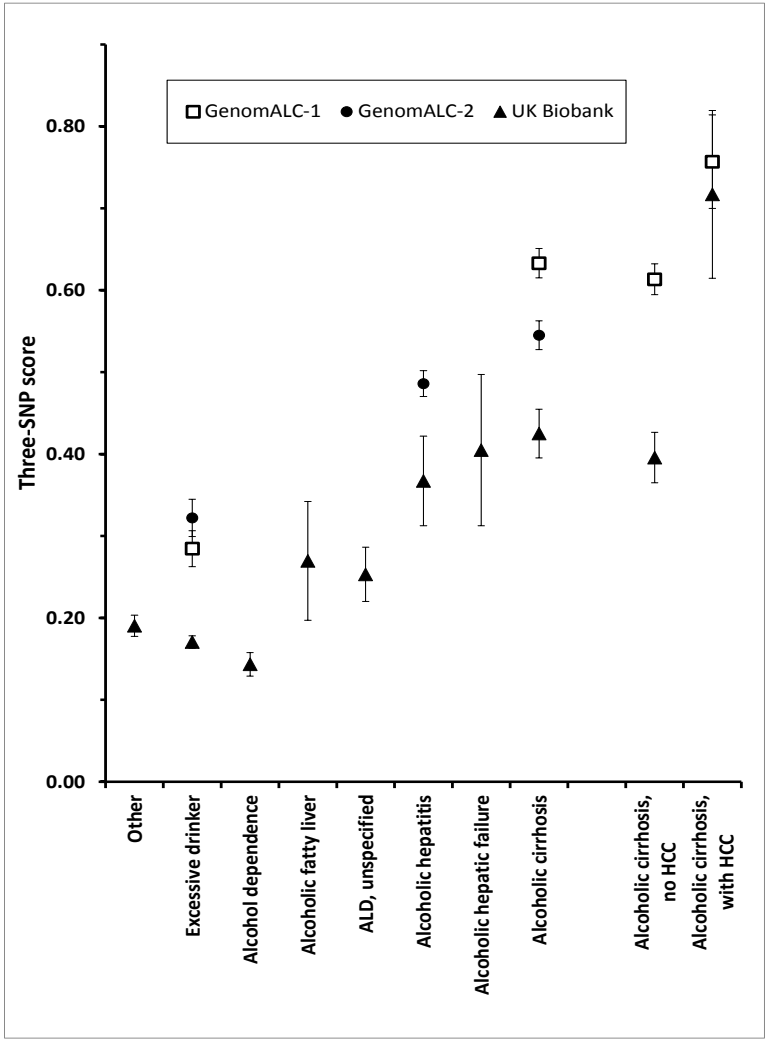


The left-hand panel shows the distribution of scores in Cases and Controls, as in Figure 1; those with scores below 0 (continuous horizontal line) are always considered as the low-risk group while those above the interrupted horizontal lines (at 0.7, 0.8, 0.9 or 1.0) are in the high-risk group. The Table shows results (OR) at each of the evaluated high-risk thresholds.

Cut-off for high-risk group	Proportion of cases above cut-off	Odds Ratio, High versus Low group (95% CI)
1.0	25%	7.35 (5.01 to 10.78)
0.9	28%	5.86 (4.13 to 8.33)
0.8	37%	4.95 (3.60 to 6.81)
0.7	49%	4.96 (3.67 to 6.71)

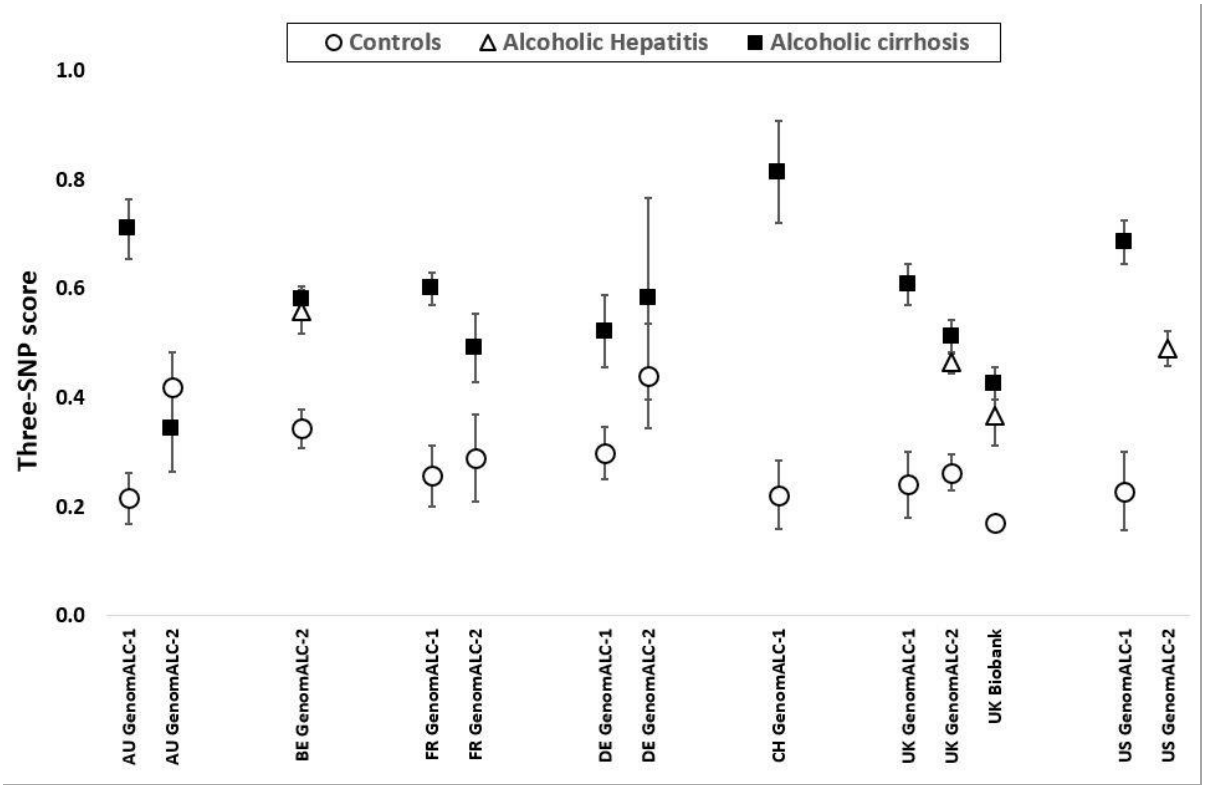
.

Supplemental Figure 3. Comparison of means for 3-SNP scores between diagnostic groups from three independent studies. Points and bars show means and standard errors. *Other:* no excessive drinking; *Excessive drinker:* high-risk drinking by the 50/80 grams/day criterion; *Alcohol dependence:* alcohol dependence [ICD-10 F10.2]; *Categories of alcohol-related liver disease are:* alcoholic fatty liver, ALD unspecified, alcoholic hepatitis, alcoholic hepatic failure and alcoholic cirrhosis (overall, and sub-divided by HCC status).



	GenomALC-1	GenomALC-2	UK Biobank
Other			**0.190 ± 0.013**
Excessive drinker	**0.285 ± 0.022**	**0.322 ± 0.023**	**0.171± 0.008**
Alcohol dependence			**0.143 ± 0.014**
Alcoholic fatty liver			**0.270 ± 0.073**
ALD, unspecified			**0.253 ± 0.033**
Alcoholic hepatitis		**0.486 ± 0.016**	**0.367 ± 0.055**
Alcoholic hepatic failure			**0.405 ± 0.092**
Alcoholic cirrhosis	**0.633 ± 0.018**	**0.545 ± 0.018**	**0.425 ± 0.030**
Alcoholic cirrhosis, no HCC	**0.613 ± 0.019**		**0.396 ± 0.031**
Alcoholic cirrhosis, with HCC	**0.757 ± 0.057**		**0.717 ± 0.102**

Supplementary Figure 4. Means of the 3-SNP score by country (AU Australia, BE Belgium, FR France, DE Germany, CH Switzerland, UK United Kingdom, US United States). Error bars show standard errors for the means.



A GENETIC RISK SCORE AND DIABETES PREDICTS DEVELOPMENT OF ALCOHOL-RELATED CIRRHOSIS IN DRINKERS

John B. Whitfield[#] (1), Tae-Hwi Schwantes-An (2), Rebecca Darlay (3), Guruprasad P. Aithal (4), Stephen R. Atkinson (5), Ramon Bataller (6), Greg Botwin (7,8), Naga P. Chalasani (9), Heather J. Cordell (3), Ann K. Daly (10), Christopher P. Day (11), Florian Eyer (12), Tatiana Foroud (2), Dermot Gleeson (13), David Goldman (14), Paul S. Haber (15,16), Jean-Marc Jacquet (17), Tiebing Liang (9), Suthat Liangpunsakul (18), Steven Masson (10), Philippe Mathurin (19), Romain Moirand (20), Andrew McQuillin (21), Christophe Moreno (22), Marsha Y. Morgan (23), Sebastian Mueller (24), Beat Müllhaupt (25), Laura E. Nagy (26), Pierre Nahon (27-29), Bertrand Nalpas (17,30), Sylvie Naveau (31), Pascal Perney (32), Munir Pirmohamed (33), Helmut K. Seitz (24), Michael Soyka (34), Felix Stickel (25), Andrew Thompson (33,35), Mark R. Thursz (5), Eric Trépo (22), Timothy R. Morgan* (7,36), Devanshi Seth*[#] (15,16,37), for the GenomALC Consortium.

(1) Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Queensland 4029, Australia, (2) Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis IN, USA, (3) Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, International Centre for Life, Central Parkway, Newcastle upon Tyne NE1 3BZ, United Kingdom, (4) NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals and the University of Nottingham, Nottingham NG7 2UH, United Kingdom, (5) Department of Metabolism, Digestion & Reproduction, Imperial College London, UK, (6) Center for Liver Diseases, University of Pittsburgh Medical Center, 3471 Fifth Avenue, Pittsburgh, PA 15213, USA, (7) Department of Veterans Affairs, VA Long Beach Healthcare System, 5901 East Seventh Street, Long Beach, CA 90822, USA, (8) F. Widjaja Family Foundation Inflammatory Bowel and Immunobiology Research Institute,

Cedars-Sinai Medical Center, Los Angeles, California CA 90048, USA (9) Department of Medicine, Indiana University, Indianapolis, IN 46202-5175, USA, (10) Faculty of Medical Sciences, Newcastle University Medical School, Framlington Place, Newcastle upon Tyne NE2 4HH, United Kingdom, (11) Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, United Kingdom, (12) Division of Clinical Toxicology, Department of Internal Medicine 2, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Ismaninger Str. 22, 81675 Munich, Germany, (13) Liver Unit, Sheffield Teaching Hospitals, AO Floor Robert Hadfield Building, Northern General Hospital, Sheffied S5 7AU, UK, (14) Laboratory of Neurogenetics, NIAAA, Rockville, MD 20852, USA, (15) Drug Health Services, Royal Prince Alfred Hospital, Missenden Road, Camperdown, NSW 2050, Australia, (16) Faculty of Medicine and Health, the University of Sydney, Sydney, NSW 2006, Australia, (17) Service Addictologie, CHRU Caremeau, 30029 Nîmes, France, (18) Division of Gastroenterology and Hepatology, Department of Medicine, Indiana University and Roudebush Veterans Administration Medical Center, Indianapolis, USA, (19) CHRU de Lille, Hôpital Claude Huriez, Rue M. Polonovski CS 70001, 59 037 Lille Cedex, France, (20) Univ Rennes, INRA, INSERM, CHU Rennes, Institut NUMECAN (Nutrition Metabolisms and Cancer), F-35000 Rennes, France, (21) Molecular Psychiatry Laboratory, Division of Psychiatry, University College London, London WC1E 6DE, UK , (22) CUB Hôpital Erasme, Université Libre de Bruxelles, clinique d'Hépatologie, Brussels, Belgium; Laboratory of Experimental Gastroenterology, Université Libre de Bruxelles, Brussels, Belgium, (23) UCL Institute for Liver & Digestive Health, Division of Medicine, Royal Free Campus, University College London, London NW3 2PF, UK, (24) Department of Internal Medicine, Salem Medical Center and Center for Alcohol Research, University of Heidelberg, Zeppelinstraße 11-33, 69121 Heidelberg, Germany, (25) Department of Gastroenterology and Hepatology, University Hospital Zurich, Rämistrasse 100, CH-8901 Zurich, Switzerland, (26) Lerner

Research Institute, 9500 Euclid Avenue, Cleveland, Ohio, OH 44195, USA, (27) Service d'Hépatologie, APHP Hôpital Avicenne et Université Paris 13, Bobigny, France, (28) University Paris 13, Bobigny, France, (29) Inserm U1162 Génomique fonctionnelle des tumeurs solides, Paris, France, (30) DISC, Inserm, 75013 Paris, France, (31) Hôpital Antoine-Béclère, 157 Rue de la Porte de Trivaux, 92140 Clamart, France, UM1, INSERM U1018 (32) Hôpital Universitaire Caremeau, Place du Pr. Robert Debre, 30029 Nîmes, France, (33) MRC Centre for Drug Safety Science, Liverpool Centre for Alcohol Research, University of Liverpool, The Royal Liverpool and Broadgreen University Hospitals NHS Trust, and Liverpool Health Partners, Liverpool, L69 3GL, UK, (34) Psychiatric Hospital University of Munich, Nussbaumsstr.7, 80336 Munich, Germany and Privatklinik Meiringen, Willigen, CH 3860 Meiringen, Switzerland, (35) Health Analytics, Lane Clark & Peacock LLP, London, UK, (36) Department of Medicine, University of California, Irvine, USA, (37) Centenary Institute of Cancer Medicine and Cell Biology, the University of Sydney, Sydney, NSW 2006, Australia.

All authors except first three and last two are in alphabetical order

*Equal senior authors

[#]Corresponding authors

1. Dr Devanshi Seth, Centenary Institute of Cancer Medicine and Cell Biology, The University of Sydney, Sydney, NSW 2006, Australia. d.seth@sydney.edu.au

2. Dr John B. Whitfield, Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Queensland 4029, Australia. John.Whitfield@qimrberghofer.edu.au

KEYWORDS

Hepatocellular carcinoma; risk stratification; chronic alcohol use; genome wide association; single nucleotide polymorphism; coffee

Word count: 3546

Numbers of figures and tables: Tables 4; Figures 2. Supplemental data: Tables 5, Figures 4.

CONFLICT OF INTEREST

NPC has a number of consulting agreements with and research grants from the pharmaceutical industry but they are not significantly or directly related to this paper. MP receives research funding from various organisations including the MRC and NIHR. He has also received partnership funding for the following: MRC Clinical Pharmacology Training Scheme (co-funded by MRC and Roche, UCB, Eli Lilly and Novartis); a PhD studentship jointly funded by EPSRC and Astra Zeneca; and grant funding from Vistagen Therapeutics. He has also unrestricted educational grant support for the UK Pharmacogenetics and Stratified Medicine Network from Bristol-Myers Squibb and UCB. He has developed an HLA genotyping panel with MC Diagnostics, but does not benefit financially from this. He is part of the IMI Consortium ARDAT (www.ardat.org). None of these funding sources were deployed in the undertaking of this study. TRM has conducted clinical research with AbbVie, Genfit, Gilead, and Merck but none of these are related to this manuscript.

FINANCIAL SUPPORT: Funding for this study was provided by NIH/NIAAA UO1AA018389 and RO1AA018389 for data collection, analysis, interpretation and patient recruitment. SL: U01 grant from the NIAAA and R01 AA025208, U01 AA026917, and

1I01CX000361. MT: The UK Medical Research Council Stratified Medicine Award (Ref MR/R014019/1), NIHR Imperial Biomedical Research Centre and NIHR Senior Investigator Award (NIHR 200153).

AUTHOR CONTRIBUTIONS

DS, CPD, TRM, PM, PSH, HKS, JBW, BN, FS, TF, AKD and HJC conceived and designed the study. Recruitment and data acquisition was done for GenomALC-1 by GPA, FE, DGl, J-MJ, SL, SMa, PM, RM, TRM, SMu, BM, PN, BN, SN, PP, MP, HKS, DS, MS, FS, AT; and for GenomALC-2 by GPA, SA, RB, NPC, AKD, FE, DGl, DGo, SMa, PM, CM, AM, MM, TRM, LEN, DS, FS, AT, MT, ET. Genetic analysis for SNP information was performed by THS-A, HJC, RD, TF. TL facilitated DNA processing for genotyping. JBW and DS led the analyses and writing of the manuscript. All authors read, critically reviewed and approved the final version. DS and TRM are the guarantors.

ABSTRACT (273 words)

Background & Aims: Only a minority of excess alcohol drinkers develop cirrhosis. We developed and evaluated risk stratification scores to identify those at highest risk.

Methods. Three cohorts (GenomALC-1: n=1690, GenomALC-2: n=3037, UK Biobank: relevant n=6898) with a history of heavy alcohol consumption (≥80 g/day (men), ≥50 g/day (women), for ≥10 years) were included. Cases were participants with alcohol-related cirrhosis. Controls had a history of similar alcohol consumption but no evidence of liver disease. Risk scores were computed from up to eight genetic loci identified previously as associated with alcohol-related cirrhosis and three clinical risk factors. Score performance for the stratification of alcohol-related cirrhosis risk was assessed and compared across the alcohol-related liver disease spectrum, including hepatocellular carcinoma (HCC).

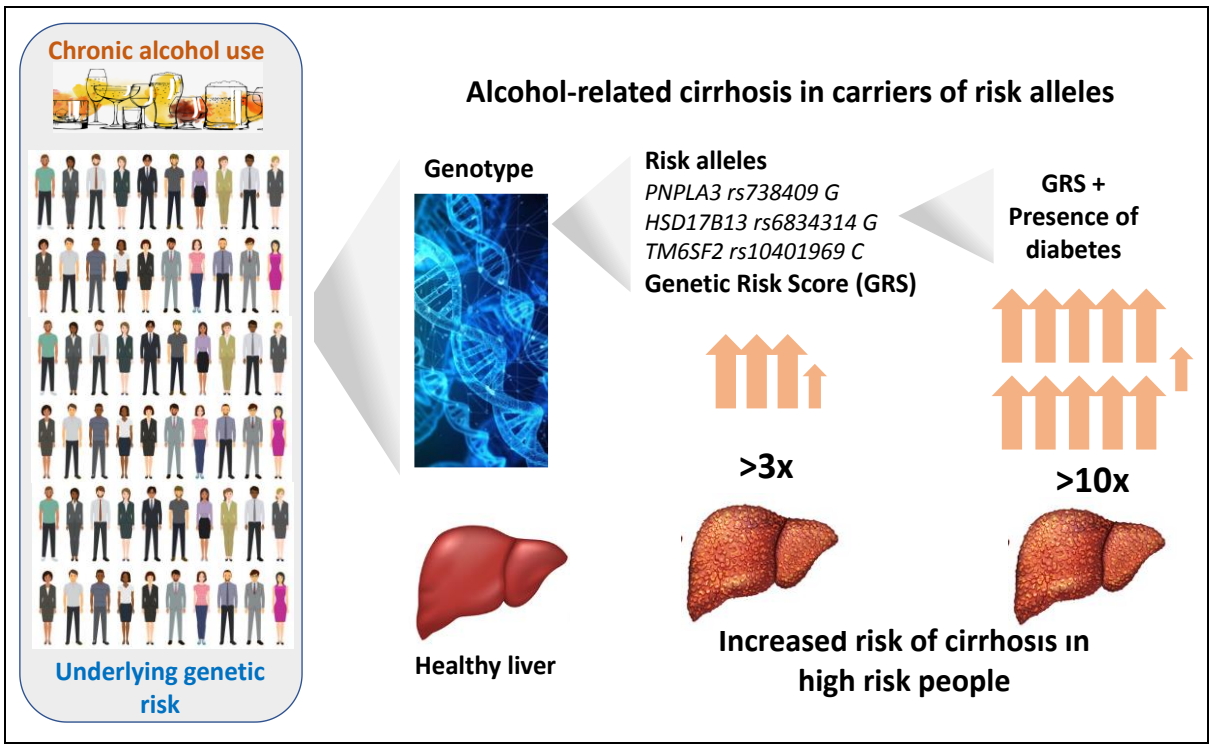
Results: A combination of three single nucleotide polymorphisms (SNPs) (*PNPLA3:*rs738409, *SUGP1-TM6SF2:*rs10401969, *HSD17B13:*rs6834314) and diabetes status best discriminated for cirrhosis risk. The odds ratio (OR) and 95% confidence intervals (CI) for the extreme score quintiles (Q1-Q5) of the 3-SNP score, based on independent allelic effect size estimates, were 5.99 (4.18;8.60) (GenomALC-1); 2.81 (2.03;3.89) (GenomALC-2); and 3.10 (2.32;4.14) (UK Biobank). Patients with diabetes and high-risk score, compared to those without diabetes and a low-risk score, had ORs increased to 14.7 (7.69;28.1) (GenomALC-1) and 17.1 (11.3;25.7) (UK Biobank). Patients with cirrhosis and HCC had significantly higher mean risk scores than patients with cirrhosis alone (0.76±0.06 versus 0.61±0.02, p=0.007). Score performance was not significantly enhanced by information on additional genetic risk variants, body mass index or coffee consumption.

Conclusions: A risk score based on three genetic risk variants and diabetes status can provide meaningful risk stratification for cirrhosis in excess drinkers, allowing earlier prevention planning including intensive intervention.

LAY SUMMARY

Excessive chronic drinking leads to liver cirrhosis in some people, but so far there is no way to identify those at high risk of developing this debilitating disease. Our study has developed a genetic risk score (GRS) test that can identify patients at high risk and shows that the risk of cirrhosis is increased >10-fold with just two risk factors - diabetes and high GRS. Risk assessment using this test has potential for early and personalised management of this disease in high-risk patients.

GRAPHICAL ABSTRACT



INTRODUCTION

Although the risk for developing cirrhosis is positively associated with alcohol consumption, only a minority of people with high-risk alcohol intake develop cirrhosis. The prevalence can vary between 7-16%[1,2] with some reports suggesting the prevalence to be as low as 2%[3,4]. The risk threshold for what is considered high-risk intake has changed over time[5-7]. Long-term consumption of 80 grams per day (g/d) or more is associated with increased risk of cirrhosis[8,9], but the threshold for liver harm is below this level, especially for women[10,11]. The 80 g/d (men) and 50 g/d (women) cut-offs were set at a relatively high level to ensure both the cirrhosis and control groups were exposed to a substantial level of alcohol-related risk. We used this threshold to define "heavy drinking" in this study.

Primary prevention of alcohol-related liver disease (ALD) would involve decreasing alcohol intake of the whole population but achieving this remains challenging. Focused intervention through the identification of people with high alcohol intake or more specifically through stratification of individuals within this population at risk for developing cirrhosis depends on identification of those at high risk. Evidence from clinical trials[12] suggests that informing excessive drinkers that they have abnormal liver function tests/hepatic fibrosis can motivate them to reduce their alcohol intake. A number of both constitutional[13-16], and genetic[17-20] risk factors for the development of alcoholic-related cirrhosis have been identified, but no attempt appears to have been made, to date, to bring these together to provide an integrated measure of risk. Thus, the aim of this study was to devise risk scores for the stratification of cirrhosis risk and evaluate them in heavy drinkers from three independent cohorts.

MATERIALS/PATIENTS AND METHODS

Information on disease status, genotypes and clinical risk factors was available for three cohorts: i) GenomALC-1 and ii) GenomALC-2 from the GenomALC consortium, and iii) the UK Biobank. Details of the recruiting and contributing sites, with numbers of patients by diagnosis and by country are given below and in Supplementary Table 1. Cohort characteristics of the cases and controls from each source are described in Supplementary Table 2.

GenomALC-1

The GenomALC-1 cohort was recruited according to a pre-designed protocol between 2012 and 2017 in Australia, France, Germany, Switzerland, the UK, and the USA. The recruitment criteria and the data collection protocol were detailed previously[21]. Briefly, all participants had a history of heavy drinking (≥80 g/d (men) and ≥50 g/d (women) for ≥10 years). For cases, cirrhosis had been diagnosed by a combination of clinical criteria, laboratory variables and/or liver elastography (Fibroscan®), with liver biopsy if clinically indicated. Clinical features defining the severity of cirrhosis are shown in Supplementary Table 2. Other liver diseases (hepatitis B or C, haemochromatosis, Wilson's disease, and autoimmune hepatitis) were excluded by laboratory testing or clinical criteria. For controls, liver disease was excluded through a combination of clinical history and measurement of liver function tests (bilirubin, albumin, ALT). For both cases and controls, HIV infection was an exclusion criterion. The study was approved by appropriate Ethics Committees or Institutional Review Boards at each site and conformed to the ethical guidelines of the 1975 Declaration of Helsinki. Participants were provided with explanations of the study and gave written informed consent. Genotyping was performed at Erasmus University Medical Centre, Rotterdam using the Illumina GSA genotyping array, as described[20].

GenomALC-2

The biological samples and data were donated by research groups who had independently collected them for other studies. Some of the GenomALC-2 samples were included in a previous GWAS[17]; therefore, for the purposes of this study, overlapping samples were removed from the analysis. Clinical diagnosis of cases and controls was similar to GenomALC-1 criteria but detailed clinical information was limited for this cohort. Patients had given informed consent and the studies were approved by the appropriate Ethics Review Boards. DNA from these participants' samples was also genotyped as outlined above for GenomALC-1.

Genotypes in the GenomALC-1 and GenomALC-2 cohorts were [22] cleaned using a widely used quality control pipeline, the GWASTools package https://bioconductor.org/packages/devel/bioc/manuals/GWASTools/man/GWASTools.pdf and imputed to 1000 Genomes reference using the Michigan Imputation Server (MIS)[22]

UK Biobank

The UK Biobank[23] includes approximately 500,000 volunteers from the UK with a wide range of data including computer-administered questionnaires, physical measurements, laboratory tests, and genotyping. All participants gave informed consent, consistent with the UK Biobank Ethics and Governance Framework. Recruitment and initial assessment occurred between 2006 and 2010 when participants were aged 40 to 69 years. Access to the UK Biobank database was obtained (Application 18870) and relevant data (with diagnoses updated to June 2020) were extracted. For cases, information was restricted to assigned clinical diagnosis (Supplementary Table 2) on hospital admissions and diagnoses, and on causes of death in participants who have subsequently died. Information was available on self-reported alcohol intake at the time of assessment (by beverage type, in drinks per week, or for less frequent drinkers per month), and participants also reported whether this was less than, similar to or more than they had been consuming 10 years previously. The amounts were converted to express the alcohol intake in g/d. Participants who had a recorded diagnosis or cause of death of alcohol-related cirrhosis

(ICD-10 K70.3, 'Alcoholic cirrhosis of liver') from hospital records or death certificates were included as cases (n=594), and those with reported drinking above the 80 or 50 g/day limits, with similar or greater consumption 10 years before, but with no diagnosed liver disease (either alcohol-related or other causes) were included as controls (n=6304). Exclusion criteria for UK Biobank subjects were similar to GenomALC-1.

UK Biobank also included 758 cases within the spectrum of other alcohol-related liver disease diagnoses (Supplementary Table 1). Genotype data for the relevant UK Biobank participants were downloaded from the server and genotypes for the relevant SNPs were extracted. Data on coffee consumption, body mass index (BMI) and diabetes status were recorded (Supplementary Table 2).

Data curation and statistical analysis

Data management and statistical analyses used IBM SPSS Statistics, version 22 (IBM Corp., New York NY). Binary variables were coded as 0 (absent) or 1 (present). Diabetes status (absent/present), BMI, kg/m^2) and coffee consumption (0: not a coffee consumer, 1: coffee consumer) shown in our previous report as associated with cirrhosis[16] were also modelled. Genotype data were coded as single nucleotide polymorphisms (SNPs) minor allele dosages, assuming an additive model for allelic effects.

Calculation of risk scores requires coefficients for the effect sizes associated with each risk factor, and assessment of the performance of the risk scores requires testing in independent cohorts not included in the derivation of these coefficients. The scheme shown in Table 1 sets out the basis for the scores and the data-sets which were used for evaluation.

SNPs with the lowest p-value at three loci (*PNPLA3:*rs738409, *SUGP1-TM6SF2:*rs10401969 and *HSD17B13:*rs6834314) were selected based on previous association with the risk of alcohol-related cirrhosis[17,18], and confirmed at genome-wide significance in our meta-

analysis[20]. Two significantly associated SNPs have been reported at *SUGP1-TM6SF2* locus[17] which are in near-complete linkage disequilibrium (d'1.00, r^2 0.955), and rs10401969 was chosen over rs58542926 because of its stronger association with cirrhosis.

A score based on these three loci ('3-SNP score') was computed for each participant in each of the three cohorts. Minor allele counts ('dosage') were obtained from direct or imputed genotypes for each SNP, multiplied by the beta coefficients for allelic effect sizes (derived from published odds ratios, calculated as beta = $\log_e$(OR)) and summed across SNPs (Table 1). The means for 3-SNP scores were also compared between disease diagnostic groups in the three independent cohorts described in Supplementary Table 1.

Scores based on three, five, and eight loci were also computed for the GenomALC-2 samples using coefficients of loci with significant association from the published meta-analysis[20] or other sources[17,18] ('3-SNP-M', '5-SNP-M' and '8-SNP-M' scores) (Table 1). The 3-SNP-M score was based on the loci mentioned above, the 5-SNP-M score included above three loci, and *SERPINA1 and FAF2* identified in our meta-analysis, and the 8-SNP-M score which was derived from the 5-SNP-M score with addition of three reported loci (*MBOAT7*, *MTARC1* [previously *MARC1*], *HNRNPUL1*) significantly associated with alcohol-related cirrhosis[17,24,25].

Area under the ROC curve (AUC) analysis and logistic regressions (with the score as the predictor variable and case/control status as an outcome) were performed. Odds Ratios (ORs) of the score were compared for extreme quintiles (highest Q5 against lowest Q1).

RESULTS

Risk stratification by genetic loci-based scores

Results in the three study cohorts for the 3-SNP score AUCs, logistic regressions and the ORs comparing quintiles Q5 and Q1 of the score, are shown in Table 2. Each of these measures showed better performance of the score in the GenomALC-1 cohort than in either the GenomALC-2 or UK Biobank cohorts, and there was no significant difference in score between men and women (Supplementary Table 3).

The results of adding two clinical risk factors (BMI and coffee consumption) to the 3-SNP score are shown in Table 2. Because the beta-coefficients for the two clinical risk factors were derived from the GenomALC-1 cohort, and information on these factors was not available for the GenomALC-2 cohort, this score was only evaluated against the UK Biobank data. A moderate, but not significant, improvement in risk stratification was observed following addition of these clinical risk factors; the Q5-Q1 OR estimate increased from 3.10 to 3.37 but the 95% confidence intervals overlapped. Coffee data did not improve the risk stratification, and nor did BMI (which was non-significant in the UK Biobank group and not available for GenomALC-2) (Table 2). Stratification of risk including the clinical factors in the score showed similar results for men and women (Supplementary Table 3).

The addition of further loci in the 5-SNP-M score (*PNPLA3:*rs2294915, *SUGP1-TM6SF2:*rs10401969, *HSD17B13:*rs10433937, *SERPINA1:*rs28929474, *FAF2:*rs11134997)[17,24,25] and in the 8-SNP-M score, with *MBOAT7:rs641738*, *MTARC1:*rs2642438 and *HNRNPUL1:*rs17251589 in addition to those in the 5-SNP-M score, did not improve the associations between score and outcome or the risk stratification (Table 2). Because the coefficients for *FAF2* and *SERPINA1* were obtained from the meta-analysis of the GenomALC-1, Buch study[17] and UK Biobank data, the 5-SNP-M and 8-SNP-M scores could only be tested in the GenomALC-2 data. To allow a valid comparison between the multi-SNP scores each was based on the coefficients from our meta-analysis of GWAS results. This resulted in an improvement for the meta-analysis-based 3-SNP-M score compared to the 3-

SNP score (Q5-Q1 ORs changed from 2.81 [95% CI 2.03,3.89] to 3.65 [2.59,5.15]). There was also a high correlation between the 3-SNP and 3-SNP-M scores in GenomALC-2 ($r = 0.826$, $n = 3037$, $p < 10^{-200}$; Supplementary Figure 1).

Clinical utility of the risk score

Numerical cut-offs that define or quantify risk are needed if the risk score is to have clinical utility. The 3-SNP scores in the GenomALC-1 cases and controls for the lowest and highest quintile boundaries were close to 0 and 1 (0.033 and 0.964, respectively; Figure 1). Division of the scores into three groups at low, intermediate and high cirrhosis risk was based on the 3-SNP score distribution (Supplementary Figure 2). The final selected scores were, low: <0; intermediate >0 - 0.7 and high risk >0.7. In each study cohort the risk difference between the low- and high-risk groups ranged between 2.5-fold and approximately 5-fold (Table 3). The difference in risk between the high- and low-risk GenomALC-1 groups were similar across the six countries (Figure 2).

Diabetes

Diabetes is known to have a large effect on cirrhosis risk. Inclusion of diabetes status with genetic risks in a combined risk score led to a bimodal distribution and difficulty in defining score quintiles. Thus, to see the effect of genetic risk score in the context of diabetes status, the 3-SNP score was subdivided by the diabetes status and is presented separately (Table 4).

People with diabetes showed a substantial increase in the risk of cirrhosis in both the GenomALC-1 (OR 3.82, 95% CI 2.67;5.47) and the UK Biobank (OR 5.62, 95% CI 4.33;7.28) cohorts. The genetic score effects were similar for people with and without diabetes, both in the GenomALC-1 (logistic regression coefficients ± SE, no diabetes: 1.055 ± 0.105; diabetes: 1.276 ± 0.338) and the UK Biobank data (no diabetes: 0.653 ± 0.093; diabetes: 0.735 ± 0.181). Tests for genetic score-diabetes interaction, either by including a (score x diabetes) term in the

logistic regression or by testing for heterogeneity of Odds Ratios between those with and without diabetes, showed no evidence for interaction effects in either cohort (Table 4). The combined effects of having diabetes and a high genetic risk score resulted in a >10-fold increased risk in people with diabetes and a high risk 3-SNP score against people without diabetes and a low-risk score, for both GenomALC-1 (OR 14.7, 95% CI 7.69;28.1) and the UK Biobank (OR 17.1, 95% CI 11.3,25.7) (Table 4).

Genetic loci-based risk scores across alcohol-related liver diseases

The mean values for the 3-SNP score varied across groups defined by alcohol intake and by the diagnostic categories for alcohol-related liver disease for both GenomALC-1 and the UK Biobank cohorts (Supplementary Figure 3). *Post hoc* comparisons showed similar trends of mean 3-SNP risk score increasing with disease severity for the GenomALC-2 cohort that included excessive drinkers with no liver disease and significantly differed between cases with severe alcoholic hepatitis and alcohol-related cirrhosis ($p = 0.011$) (Supplementary Table 4). Mean 3-SNP score increased with severity of liver disease (Supplementary Figure 3), including when comparing cirrhosis with HCC against cirrhosis without HCC, both for GenomALC-1 (0.757 ± 0.057 versus 0.613 ± 0.019) and UK Biobank (0.717 ± 0.102 versus 0.396 ± 0.031); see also Supplementary Table 5.

DISCUSSION

This study shows that a genetic score based on three lead SNPs associated at genome-wide significance with the risk for developing alcohol-related cirrhosis, can risk-stratify people drinking at potentially harmful levels.

Development of score for risk stratification

The performance of 3-SNP score improved considerably when used in conjunction with information on diabetes status, providing a powerful tool for identifying patients at high risk for developing advanced alcohol-related liver diseases. Higher scores were also associated with other severe liver injuries, including alcoholic hepatitis and HCC.

Our main measure of genetic risk stratification was to compare people who are in the highest quintile for a score against those in the lowest quintile, providing a more practical measure of stratification success than comparing the most extreme of all possible categories, which will usually contain only a small proportion of people[26]. Substantial Q5-Q1 risk differences were evident for the simple 3-SNP score in each of the cohorts; approximately six-fold in the GenomALC-1 cohort and three-fold in the other cohorts (Table 2). The greater difference in Q5-Q1 risk for GenomALC-1 is likely to be due to a more refined and pre-defined case-control definition for the recruitment protocol in this cohort.

Diabetes status led to a substantial enhancement of the utility of the 3-SNP score, predicting a >10-fold difference in risk between extreme groups (Q5 with diabetes and Q1 non-diabetes). Adding information on further genetic risk variants or BMI and coffee consumption had minimal effect.

Clinical utility of risk-score

Clinical application of a score requires the definition of decision points in numerical terms rather than by reference to population quintiles. However, Q5-Q1 comparisons can be useful for comparison across cohorts, such as in our study, and against genetic scores for other diseases. For clinical application boundaries of 0 and 0.7 were set for the 3-SNP score that provided a potentially useful stratification of risk in each of the three cohorts. As expected, lowering the high-risk threshold (e.g. from 1.0 to 0.7) identified a higher proportion of the cases as being at high risk but the ORs between the high- and low-risk groups decreased. For any

classification based on a numerical test or score, changing the cut-off point(s) will alter the trade-off between test sensitivity and specificity and the optimum cut-offs will depend on the use to be made of the test. The prevalence of the condition is also important because this will affect the predictive value of positive or negative results. The AUCs shown in Table 2 were significant but when the desired test sensitivity was set at 80% (to flag nearly all those at high risk) the test specificities were between 30% and 40%. Thus, a substantial number of false positives must be accepted, making the score suitable for risk stratification but not for prediction of outcome in individual patients.

The 3-SNP risk score was also associated with differences across the alcohol-related liver disease spectrum, including HCC. The HCC risk association is consistent with previous information showing that *PNPLA3*, *HSD17B13* and *TM6SF2* polymorphisms[27-31] are associated with a higher risk for this condition compared to advanced cirrhosis, perhaps suggesting a pro-oncogenic role for these variants.

Scope of risk-score

The loci comprising the current risk score are also implicated in the risk for developing cirrhosis of diverse aetiologies. Using similar polygenic risk scores (PRS) in non-alcoholic fatty liver disease (NAFLD) revealed that combining genetic and clinical features refines the predictive utility of the algorithm for identifying those at higher risk of severe liver disease[32-34]. Given the many shared genetic and metabolic risks between alcohol-related liver disease and NAFLD, the predictive algorithm defined here may have a wider scope across these diseases for risk stratification of those at higher risk of cirrhosis. Recently Emdin and colleagues[30] identified 12 variants, five previously known, including *PNPLA3, HSD17B13* and *TM6SF,* and seven novel, which were associated at genome-wide significance with 'any cause' cirrhosis , and aggregated these into a PRS. A high PRS, defined as the top quintile of the

distribution, was associated with significantly increased risk of cirrhosis compared with the lowest quintile (OR 2.26; P < .001). Our current study indicates that risk stratification for alcohol-related cirrhosis can be achieved as effectively using fewer genetic markers, and with algorithms based on a smaller base of GWAS information, presumably because the genetic architecture of alcohol-related cirrhosis includes a number of common variants with substantial effects on risk.

Preliminary investigation of adding previously reported risk loci over the 3-SNP score did not significantly improve risk stratification. To develop a robust PRS that incorporates many loci for alcohol-related cirrhosis risk would require a larger population based cohort. Another possible extension, again dependent on the availability of more data, would be to incorporate information on patients' alcohol consumption in addition to genotyping for genetic variants associated with cirrhosis risk.

The outcome of risk stratification for alcohol-related liver disease can be compared with PRS approaches to other complex diseases, including cardiovascular disease and cancers. A recent study[35] showed that for five common diseases (coronary heart disease, type 2 diabetes, atrial fibrillation, breast cancer and prostate cancer), Q1-Q5 differences in PRS were associated with approximately two- to five-fold differences in the cumulative prevalence of diagnosis by age 80. Our 3-SNP score performance was equal to or slightly better than these.

The main strengths of this study were that it employed three large independent cohorts and that the case and control definitions were standardised. The study also had its limitations. First, the included populations were of largely European ancestry so that the finding may not be universally applicable. Second, an unknown proportion of the controls, especially in the UK Biobank cohort, may have undiagnosed alcohol-related liver disease, although it should be recognised that misclassification of some cases as controls would lead to poorer stratification

such that the effectiveness of our score would be under-, rather than over-estimated. Finally, the risk scores were derived from groups of heavy drinkers with cirrhosis or without liver disease. However, these were validated in case and control groups selected from the population-based UK Biobank cohort. Application of the risk score to an individual patient should be performed with an understanding that some patients' outcomes will differ from those predicted by the score. Prospective studies are needed, both to relate score to progression across time in patients who present with early stages of liver disease, and to clarify the relationship between onset of diabetes and of advanced liver disease in patients with excessive alcohol use.

Based on the findings of the present study a 3-SNP score algorithm is proposed for use and interpretation of the risk stratification in heavy drinkers (Box 1).

Box 1. ***Use of the 3-SNP risk score for alcohol-related cirrhosis.***

Calculate the risk score as:
(0.7839**PNPLA3* **rs738409 G dosage) + (0.5423****SUGP1-TM6SF2* **rs10401969 C dosage) – (0.4463****HSD17B13* **rs6834314 G dosage)**
Assign the patient to the appropriate stratum of risk, as follows:

	Score less than 0	**Score above 0.7**
	Low risk	High risk
Relative risk if <u>not</u> diabetic	1	3-fold
Relative risk if diabetic	3-fold or more	Over 10-fold

(Patients with scores between 0 and 0.7 are at intermediate risk)

When making use of this risk information, after appropriate explanation, consent and genotyping, be aware that this is a risk stratification scheme rather than providing individual predictions. Some patients whose score places them in the low-risk group will progress to significant liver disease, especially if they continue to drink excessively.

Conclusions

An algorithm for stratifying the risk of developing alcohol-related cirrhosis among heavy drinkers, based on three genetic loci and information on diabetic status, has been developed

and validated. It is intended to identify patients at particularly high risk for developing alcohol-related cirrhosis. In addition to stratifying risk of developing alcohol-related cirrhosis, this algorithm may also stratify risk for developing alcoholic hepatitis and HCC. This risk stratification system could be used to facilitate management of all people at risk for developing significant alcohol-related liver disease.

ACKNOWLEDGEMENTS

We acknowledge Ms Donna Sheedy, Ms Julia Stevens and Prof Jillian Krill for providing access to brain tissue (included in GenomALC-2) from the New South Wales Brain Tissue Resource Centre at the University of Sydney, NSW 2006, Australia (collection of tissues reported in this publication was supported by the National Institute of Alcohol Abuse and Alcoholism of the National Institutes of Health under Award Number R28AA012725. The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health). We are grateful to the four Alcoholic Hepatitis Consortia TREAT, InTEAM, DASH and SCAHC (U01 AA021886) from the USA, for providing DNA contributing towards GenomALC-2.

REFERENCES

1. Askgaard G, Leon DA, Kjaer MS, Deleuran T, Gerds TA, Tolstrup JS. Risk for alcoholic liver cirrhosis after an initial hospital contact with alcohol problems: A nationwide prospective cohort study. *Hepatology* 2017; **65**(3): 929-37.
2. Askgaard G, Kjaer MS, Tolstrup JS. Opportunities to Prevent Alcoholic Liver Cirrhosis in High-Risk Populations: A Systematic Review With Meta-Analysis. *Am J Gastroenterol* 2019; **114**(2): 221-32.
3. Hasin DS, Stinson FS, Ogburn E, Grant BF. Prevalence, correlates, disability, and comorbidity of DSM-IV alcohol abuse and dependence in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Arch Gen Psychiatry* 2007; **64**(7): 830-42.
4. Wong T, Dang K, Ladhani S, Singal AK, Wong RJ. Prevalence of Alcoholic Fatty Liver Disease Among Adults in the United States, 2001-2016. *JAMA* 2019; **321**(17): 1723-5.
5. Pequignot G, Tuyns AJ, Berta JL. Ascitic cirrhosis in relation to alcohol consumption. *IntJEpidemiol* 1978; **7**(2): 113-20.
6. Rehm J, Taylor B, Mohapatra S, et al. Alcohol as a risk factor for liver cirrhosis: a systematic review and meta-analysis. *Drug Alcohol Rev* 2010; **29**(4): 437-45.
7. Askgaard G, Gronbaek M, Kjaer MS, Tjonneland A, Tolstrup JS. Alcohol drinking pattern and risk of alcoholic liver cirrhosis: a prospective cohort study. *J Hepatol* 2015; **62**(5): 1061-7.
8. Corrao G, Bagnardi V, Zambon A, Arico S. Exploring the dose-response relationship between alcohol consumption and the risk of several alcohol-related conditions: a meta-analysis. *Addiction* 1999; **94**(10): 1551-73.
9. Lelbach WK. Epidemiology of alcoholic liver disease. *ProgLiver Dis* 1976; **5**: 494-515.

10. Roerecke M, Vafaei A, Hasan OSM, et al. Alcohol Consumption and Risk of Liver Cirrhosis: A Systematic Review and Meta-Analysis. *Am J Gastroenterol* 2019; **114**(10): 1574-86.
11. Tuyns AJ, Pequignot G. Greater risk of ascitic cirrhosis in females in relation to alcohol consumption. *IntJEpidemiol* 1984; **13**(1): 53-7.
12. Subhani M, Knight H, Ryder S, Morling JR. Does Advice Based on Biomarkers of Liver Injury or Non-Invasive Tests of Liver Fibrosis Impact High-Risk Drinking Behaviour: A Systematic Review With Meta-analysis. *Alcohol Alcohol* 2021.
13. Hart CL, Morrison DS, Batty GD, Mitchell RJ, Davey Smith G. Effect of body mass index and alcohol consumption on liver disease: analysis of data from two prospective cohort studies. *BMJ* 2010; **340**: c1240.
14. Liangpunsakul S, Puri P, Shah VH, et al. Effects of Age, Sex, Body Weight, and Quantity of Alcohol Consumption on Occurrence and Severity of Alcoholic Hepatitis. *Clin Gastroenterol Hepatol* 2016; **14**(12): 1831-8 e3.
15. Saab S, Mallam D, Cox GA, 2nd, Tong MJ. Impact of coffee on liver diseases: a systematic review. *Liver Int* 2014; **34**(4): 495-504.
16. Whitfield JB, Masson S, Liangpunsakul S, et al. Obesity, Diabetes, Coffee, Tea, and Cannabis Use Alter Risk for Alcohol-Related Cirrhosis in 2 Large Cohorts of High-Risk Drinkers. *Am J Gastroenterol* 2020; **116**(1): 106-15.
17. Buch S, Stickel F, Trepo E, et al. A genome-wide association study confirms PNPLA3 and identifies TM6SF2 and MBOAT7 as risk loci for alcohol-related cirrhosis. *Nat Genet* 2015; **47**(12): 1443-8.
18. Abul-Husn NS, Cheng X, Li AH, et al. A Protein-Truncating HSD17B13 Variant and Protection from Chronic Liver Disease. *N Engl J Med* 2018; **378**(12): 1096-106.

19. Tian C, Stokowski RP, Kershenobich D, Ballinger DG, Hinds DA. Variant in PNPLA3 is associated with alcoholic liver disease. *Nat Genet* 2010; **42**(1): 21-3.
20. Schwantes-An TH, Darlay R, Mathurin P, et al. Genome-wide association study and meta-analysis on alcohol-related liver cirrhosis identifies novel genetic risk factors. *Hepatology* 2021; **73**(5): 1920-31.
21. Whitfield JB, Rahman K, Haber PS, et al. Brief report: genetics of alcoholic cirrhosis-GenomALC multinational study. *Alcohol Clin Exp Res* 2015; **39**(5): 836-42.
22. Das S, Forer L, Schonherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet* 2016; **48**(10): 1284-7.
23. Sudlow C, Gallacher J, Allen N, 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**(3): e1001779.
24. Emdin CA, Haas ME, Khera AV, et al. A missense variant in Mitochondrial Amidoxime Reducing Component 1 gene and protection against liver disease. *PLoS Genet* 2020; **16**(4): e1008629.
25. Innes H, Buch S, Hutchinson S, et al. Genome-Wide Association Study for Alcohol-Related Cirrhosis Identifies Risk Loci in MARC1 and HNRNPUL1. *Gastroenterology* 2020; **159**(4): 1276-89 e7.
26. Gellert-Kristensen H, Richardson TG, Davey Smith G, Nordestgaard BG, Tybjaerg-Hansen A, Stender S. Combined Effect of PNPLA3, TM6SF2, and HSD17B13 Variants on Risk of Cirrhosis and Hepatocellular Carcinoma in the General Population. *Hepatology* 2020; **72**(3): 845-56.
27. Salameh H, Raff E, Erwin A, et al. PNPLA3 Gene Polymorphism Is Associated With Predisposition to and Severity of Alcoholic Liver Disease. *Am J Gastroenterol* 2015; **110**(6): 846-56.

28. Yang J, Trepo E, Nahon P, et al. A 17-Beta-Hydroxysteroid Dehydrogenase 13 Variant Protects From Hepatocellular Carcinoma Development in Alcoholic Liver Disease. *Hepatology* 2019; **70**(1): 231-40.
29. Tang S, Zhang J, Mei TT, et al. Association of TM6SF2 rs58542926 T/C gene polymorphism with hepatocellular carcinoma: a meta-analysis. *BMC Cancer* 2019; **19**(1): 1128.
30. Trépo E, Nahon P, Bontempi G, et al. Association between the PNPLA3 (rs738409 C\textgreaterG) variant and hepatocellular carcinoma: Evidence from a meta-analysis of individual participant data. *Hepatology* 2014; **59**(6).
31. Stickel F, Lutz P, Buch S, et al. Genetic Variation in HSD17B13 Reduces the Risk of Developing Cirrhosis and Hepatocellular Carcinoma in Alcohol Misusers. *Hepatology* 2020; **72**(1): 88-102.
32. Bianco C, Casirati E, Malvestiti F, Valenti L. Genetic predisposition similarities between NASH and ASH: Identification of new therapeutic targets. *JHEP Rep* 2021; **3**(3): 100284.
33. Bianco C, Jamialahmadi O, Pelusi S, et al. Non-invasive stratification of hepatocellular carcinoma risk in non-alcoholic fatty liver using polygenic risk scores. *Journal of Hepatology* 2021; **74**(4): 775-82.
34. De Vincentis A, Tavaglione F, Jamialahmadi O, et al. A Polygenic Risk Score to Refine Risk Stratification and Prediction for Severe Liver Disease by Clinical Fibrosis Scores. *Clin Gastroenterol Hepatol* 2021.
35. Mars N, Koskela JT, Ripatti P, et al. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat Med* 2020; **26**(4): 549-57.

Table 1. Score construction and validation plan.

		Cohorts available for independent validation		
		GenomALC-1 (N=1690)	GenomALC-2 (N=3037)	UK Biobank (N=6898)
1	3-SNP score, using SNPs and coefficients from initial reports[17,18] **= (0.7839****PNPLA3* **rs738409 G dosage) + (0.5423****SUGP1-TM6SF2* **rs10401969 C dosage) – (0.4463****HSD17B13* **rs6834314 G dosage)**	Yes	Yes	Yes
2	3-SNP score as in 1 above, with addition of BMI and coffee **= [1] + (0.0709*BMI) – (0.645*Coffee)**	No (BMI and coffee coefficients are derived from this cohort)	No (no information of BMI and coffee)	Yes
3	3-SNP-M score, using SNPs and coefficients from meta-analysis[20] **= (0.7274****PNPLA3* **rs2294915 T dosage) + (0.3988****SUGP1* **rs10401969 C dosage) – (0.2485****HSD17B13* **rs10433937 G dosage)**	No*	Yes	No*
4	5-SNP-M score; as in 3 above but with addition of two GW-significant SNPs from meta-analysis **= [3] + (0.6419****SERPINA1* **rs28929474 T dosage) – (0.2357****FAF2* **rs11134997 C dosage)**	No*	Yes	No*
5	8-SNP-M score; as in 4 but with three additional SNPs with genome-wide significant associations with alcohol-related liver disease **= [4] + (0.1446****MBOAT7* **rs641738 T dosage) - (0.2401****MTARC1* **rs2642438 A dosage) - (0.1304****HNRNPUL1* **rs17251589 T dosage)**	No*	Yes	No*

*SNP coefficients are derived from this cohort

Table 2. Results of ROC curve and logistic regression analyses, and estimated odds ratios for cirrhosis between the lowest (Q1) and highest (Q5) quintiles of scores.

		ROC Curve	Logistic regression		Q1-Q5 Odds Ratio (95% CIs)
		AUC	Beta	p-value	
3-SNP score[i]	GenomALC-1	0.665 ± 0.014	1.092 ± 0.099	2.90×10^{-28}	5.99 (4.18 to 8.60)
	GenomALC-2	0.606 ± 0.014	0.669 ± 0.090	1.44×10^{-13}	2.81 (2.03 to 3.89)
	UK Biobank	0.619 ± 0.014	0.729 ± 0.080	1.06×10^{-19}	3.10 (2.32 to 4.14)
3 SNP score[i] + BMI, coffee	GenomALC-1	Not estimated[iii]	Not estimated[iii]		Not estimated[iii]
	GenomALC-2	Not estimated[iii]	Not estimated[iv]		Not estimated[iv]
	UK Biobank	0.636 ± 0.015	0.748 ± 0.073	1.77×10^{-24}	3.37 (2.38 to 4.78)
Comparisons based on coefficients from meta-analysis:					
3-SNP-M score[ii]	GenomALC-2	0.631 ± 0.014	0.909 ± 0.103	1.17×10^{-18}	3.65 (2.59 to 5.15)
5-SNP-M score[ii]	GenomALC-2	0.626 ± 0.014	0.813 ± 0.096	2.96×10^{-17}	3.66 (2.62 to 5.12)
8-SNP-M score[ii]	GenomALC-2	0.633 ± 0.014	0.807 ± 0.091	6.06×10^{-19}	3.37 (2.43 to 4.66)

[i] Coefficients estimated from Buch et al[17] and Abul-Husn et al[18]

[ii] Coefficients estimated from meta-analysis data Schwantes-An et al[20]

[iii] Not estimated because coefficients would be partly based on data for this cohort.

[iv] Not estimated because BMI and coffee data are not available for this cohort.

Table 3. Simplification of scoring system into three groups based on numerical values of the 3-SNP score.

		Odds Ratios (95% confidence intervals)		
Risk group	score	GenomALC-1	GenomALC-2	UK Biobank
Low	≤ 0	1 N = 273 (16.2%)	1 N = 327 (18.5%)	1 N = 3403 (56.1%)
Intermediate	> 0 to 0.70	2.13 (1.61 to 2.83) N = 731 (43.3%)	1.54 (1.18 to 2.00) N = 771 (43.7%)	1.36 (1.04 to 1.77) N = 1207 (19.9%)
High	> 0.70	4.96 (3.67 to 6.71) N = 686 (40.6%)	2.67 (2.02 to 3.53) N = 668 (37.8%)	2.654(2.16 to 3.29) N = 1456 (24.0%)

Table 4. Risk of alcohol-related cirrhosis by diabetes status, and comparison of risk in the low- and high-risk of the 3-SNP score stratified by diabetes status. For GenomALC-1, diabetes status was at time of recruitment and for UK Biobank at the time of (baseline) assessment. Information on diabetes was not available for the GenomALC-2 group. Only those participants with information on diabetes, and a 3-SNP score, are included.

Predictor	Group	Contrast	Odds Ratios (95% Confidence Intervals)	
			GenomALC-1	UK Biobank
Diabetes		Diabetes versus no diabetes	3.82 (2.67 to 5.47)	5.62 (4.33 to 7.28)
3-SNP Score	No diabetes	≤0 versus >0.7 in non-diabetics	4.77 (3.45 to 6.58)	2.37 (1.86 to 3.03)
	Diabetes	≤0 (diabetes) versus >0.7 (diabetes)	5.32 (2.06 to 13.7)[1]	3.74 (2.16 to 6.48)[2]
		≤0 (no-diabetes) versus >0.7 (diabetes)	14.7 (7.69 to 28.1)	17.1 (11.3 to 25.7)

[1] Breslow-Day test for homogeneity of Odds Ratios in Non-Diabetes and Diabetes groups, GenomALC-1 $\chi^2 = 0.05$ p = 0.830.

[2] Breslow-Day test for homogeneity of Odds Ratios in Non-Diabetes and Diabetes groups, UK Biobank $\chi^2 = 2.20$ p = 0.138.

Figure 1. Distribution of 3-SNP scores in cases and controls from the GenomALC-1 data, showing the boundaries of the lowest (Q1) and highest (Q5) quintiles at 0.033 and 0.964, respectively (dotted lines).

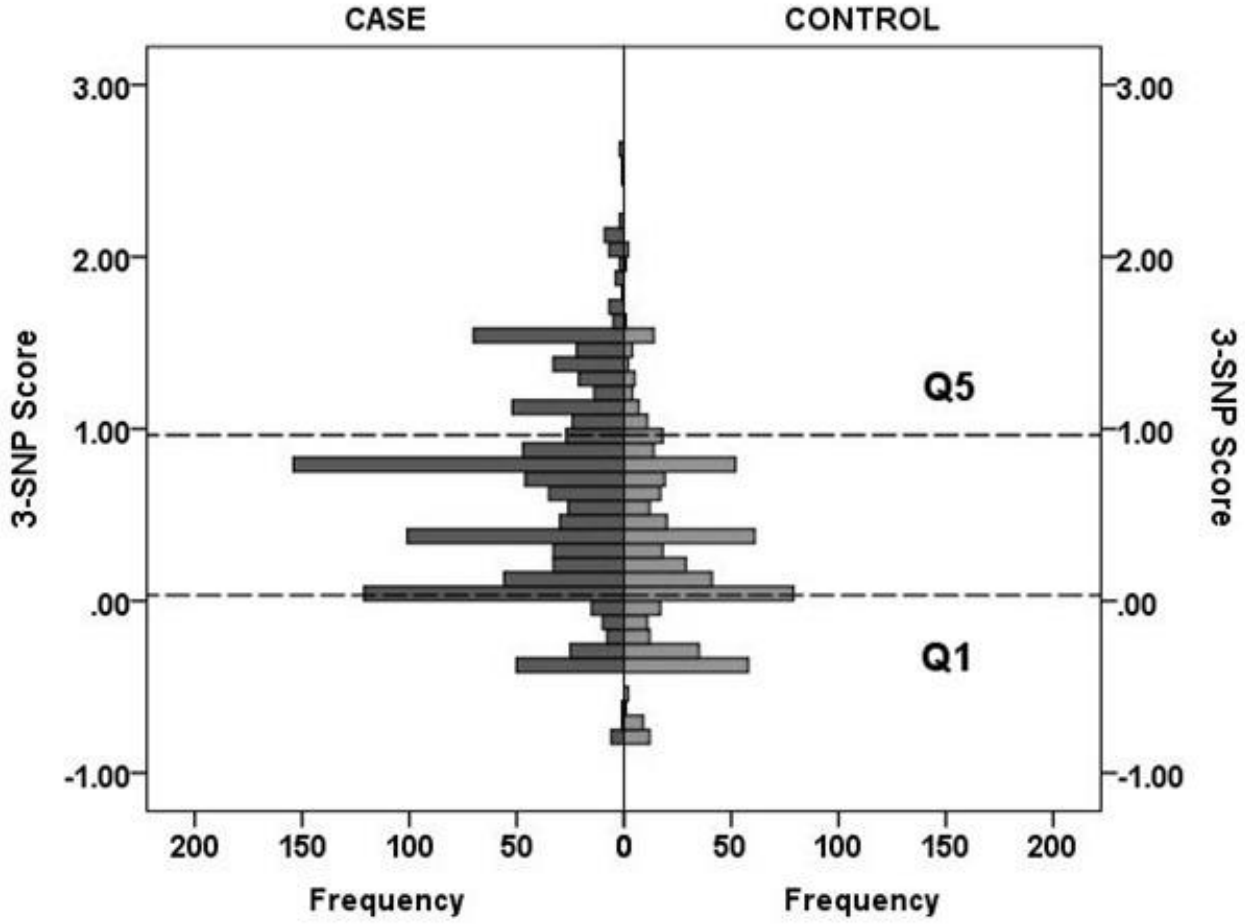
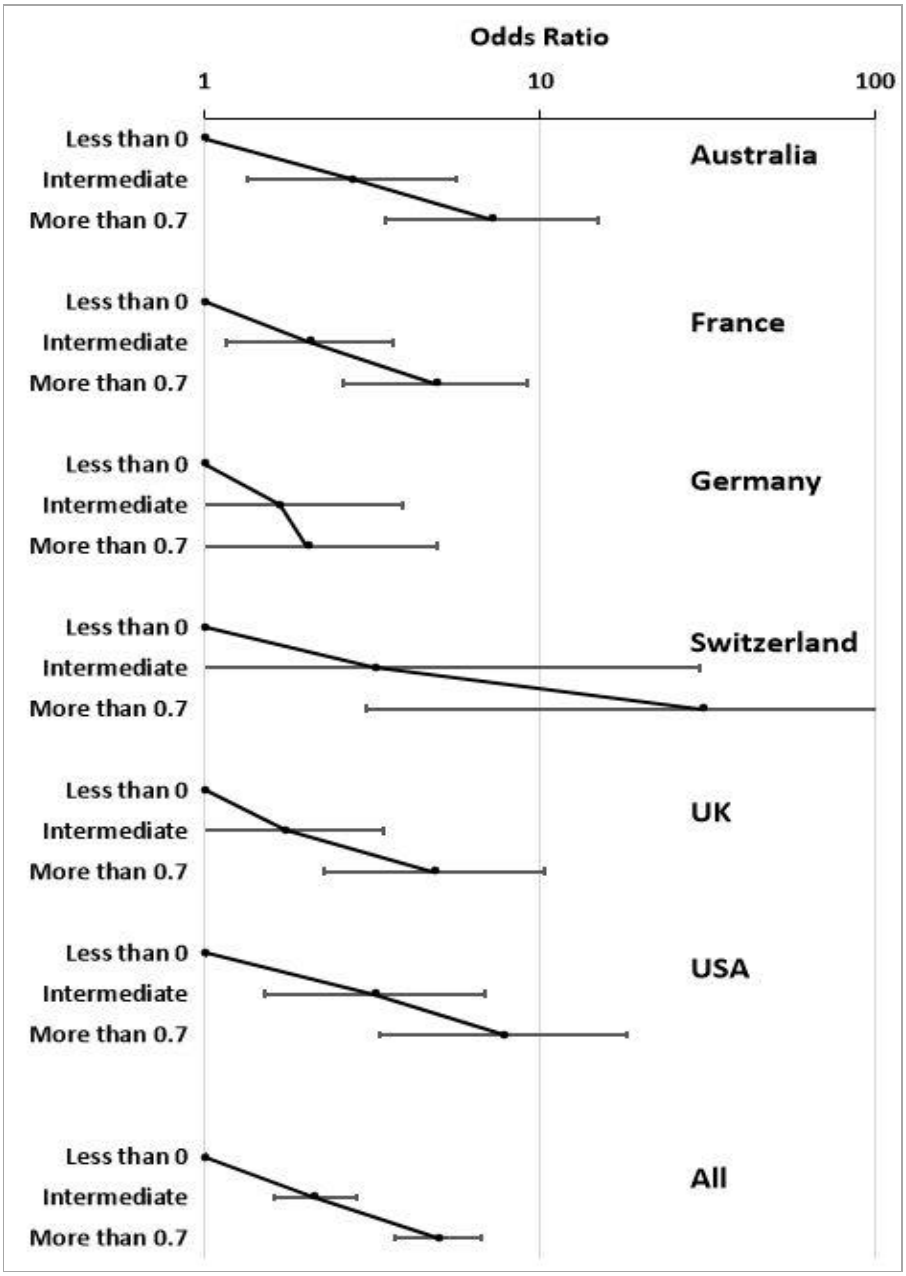


Figure 2. Odds ratios, by country and overall, for the risk of alcohol-related cirrhosis in the GenomALC-1 cohort when results for the 3-SNP score are divided into low (<0), intermediate (0 to 0.7) and high (>0.7) categories. Error bars show 95% confidence intervals.



SUPPLEMENTARY MATERIAL

A GENETIC RISK SCORE AND DIABETES PREDICTS DEVELOPMENT OF ALCOHOL-RELATED CIRRHOSIS IN DRINKERS

Supplementary Table 1. Numbers of participants, and diagnostic categories, by country of recruitment. Restricted to participants with relevant genotyping and clinical risk factor data.

GenomALC-1 samples

	Control	Alcohol-related cirrhosis	Total
Australia	175	129	304
France	103	373	476
Germany	142	75	217
Switzerland	33	28	61
UK	77	257	334
USA	58	240	298
Total	588	1102	1690

GenomALC-2 samples

	Control	Alcoholic hepatitis	Alcohol-related cirrhosis	Total
Australia	68	0	36	104
Belgium	258	223	663	1144
France	58	0	85	143
Germany	31	0	15	46
Switzerland	3	0	0	3
UK	184	771	356	1311
USA	2	277	7	286
Total	604	1271	1162	3037

UK Biobank samples

No relevant diagnosis	494,910	
Excessive drinker, no liver diagnosis	6304	
Alcoholic fatty liver	*95*	
Alcoholic liver disease, unspecified	*428*	
Alcoholic fibrosis and sclerosis	*17*	
Alcoholic hepatitis	*130*	
Alcoholic hepatic failure	*88*	
Alcoholic cirrhosis	594	
Alcoholic cirrhosis without HCC		542
Alcoholic cirrhosis with HCC		52
Total	502,566	

Note for UK Biobank cohort: Total number excludes people who withdrew consent and those with unknown sex or alcohol intake. Numbers in italics are for alcohol-related liver diseases other than cirrhosis.

Supplementary Table 2. Demographic, clinical and substance use characteristics of participants included.

		GenomALC-1 (N = 1390)				GenomALC-2 (N = 1766) [1]				UK Biobank (N= 6898) [2]			
		Cases (N =917)		Controls (N = 473)		Cases (N = 1162)		Controls (N = 604)		Cases (N = 594)		Controls (N = 6304)	
		Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
		Mean ± SD (N)	Mean ± SD (N)	Mean ± SD (N)	Mean ± SD (N)	Mean ± SD (N)	Mean ± SD (N)	Mean ± SD (N)	Mean ± SD (N)	Mean ± SD (N)	Mean ± SD (N)	Mean ± SD (N)	Mean ± SD (N)
Demographics	Age (Mean ± SD, in years)	53.0 ± 8.5 (674)	51.0 ± 8.9 (243)	50.0 ± 9.5 (331)	50.6 ± 9.7 (142)	55.2 ± 10.4 (702)	54.5 ± 10.1 (288)	47.4 ± 10.9 (394)	50.4 ± 10.1 (186)	57.8 ± 6.9 (477)	57.2 ± 7.7 (117)	56.5 ± 7.6 (4836)	54.8 ± 7.7 (1468)
	Years of Education	12.0 ± 3.5 (665)	11.5 ± 303 (240)	12.0 ± 3.9 (326)	12.8 ± 3.9 (141)					11.3 ± 2.2 (357)	11.2 ± 2.2 (83)	11.4 ± 2.1 (3785)	11.7 ± 2.0 (927)
	BMI, kg/m^2	28.0 ± 5.6 (673)	26.1 ± 6.2 (242)	25.8 ± 4.7 (330)	25.4 ± 5.8 (142)					29.1 ± 5.4 (466)	27.4 ± 5.0 (117)	28.4 ± 4.5 (4836)	26.9 ± 4.8 (1460)
	European ethnicity/race (by self-report)	99.6%	99.6%	99.7%	98.6%	100%	100%	99.7%	99.5%	96.2%	94.8%	98.5%	97.4%
Alcohol use	Alcohol intake, g/day	262 ± 431 (674)	189 ± 349 (243)	251 ± 298 (331)	186 ± 100 (142)					49.0 ± 55.0 (477)	22.4 ± 26.8 (117)	104.1 ± 27.7 (4836)	66.3 ± 19.5 (1468)
	Age started XS drinking	26.9 ± 9.8 (674)	31.3 ± 10.6 (242)	25.5 ± 9.0 (329)	29.7 ± 10.8 (141)								
	Years of high-risk drinking	25.0 ± 11.2 (674)	19.1 ± 9.1 (243)	21.6 ± 9.3 (331)	18.4 ± 7.4 (142)								
	Audit Score	10.5 ± 10.6 (673)	11.6 ± 11.5 (241)	26.8 ± 9.4 (330)	26.8 ± 10.1 (140)								
	Lifetime alcohol intake, kg	2310 ± 4034 (674)	1316 ± 2548 (243)	2073 ± 3327 (331)	1244 ± 892 (142)								
Lab results	Haemoglobin (g/L)	117 ± 26 (653)	113 ± 21 (239)	147 ± 14 (315)	135 ± 14 (136)								
	INR (ratio)	1.40 ± 0.43 (609)	1.54 ± 0.58 (223)	0.99 ± 0.17 (272)	0.97 ± 0.11 (111)	1.44 ± 0.52 (51)	1.44 ± 0.48 (134)	1.11 ± 0.41 (51)	1.04 ± 0.11 (16)				
	Albumin (g/L)	34.5 ± 6.8 (623)	34.8 ± 7.5 (227)	43.2 ± 5.1 (314)	43.2 ± 5.7 (132)	35.0 ± 7.3 (336)	33.5 ± 7.1 (128)	40.9 ± 10.7 (100)	40.8 ± 9.7 (45)	42.5 ± 4.7 (407)	43.1 ± 4.4 (101)	45.5 ± 2.7 (4175)	45.5 ± 2.7 (1250)
	Bilirubin (µmol/L)	60.6 ± 100.5 (661)	81± 121 (243)	9.1 ± 5.8 (320)	8.2 ± 5.6 (140)			12.8 ± 5.3 (59)	10.2 ± 4.5 (32)	15.3 ± 12.3 (443)	12.1 ± 7.9 (109)	9.9 ± 4.1 (4505)	8.3 ± 3.3 (1357)
	Creatinine (µmol/L)	94 ± 67 (661)	105 ± 486 (242)	75 ± 17 (324)	62 ± 15 (141)					77 ± 34 (445)	59 ± 15 (110)	77 ± 14 (4536)	62 ± 12 (1365)

	ALT (unit/L)	36.8 ± 40.2 (661)	34.3 ±33.2 (242)	43.3 ± 46.3 (326)	38 ± 38 (141)	54.6 ± 219.2 (342)	34.2 ± 25.7 (132)	45.6 ± 45.2 (240)	34.3 ± 43.9 (122)	45.0 ± 38.7 (444)	38.0 ± 26.3 (110)	31.1 ± 19.1 (4529)	21.9 ± 12.9 (1366)
	AST (unit/L)	60.2 ± 51.0 (654)	64.5 ± 51.3 (238)	43.6 ± 41.0 (320)	41 ± 38 (140)	90.1 ± 307.7 (343)	64.9 ± 51.5 (132)	52.2 ± 56.5 (242)	36.2 ± 22.6 (124)	61.3 ± 46.1 (442)	62.4 ± 51.0 (109)	33.1 ± 19.2 (4504)	26.7 ± 12.6 (1357)
	GGT (unit/L)	225 ± 371 (623)	173 ± 264 (222)	124 ± 255 (315)	122 ± 216 (139)					233 ± 245 (434)	216 ± 239 (108)	86.5 ± 87.8 (4528)	46.8 ± 56.6 (1366)
Liver disease	MELD score	11.2 ± 7.4 (604)	11.1 ± 9.1 (222)	1.9 ± 3.3 (266)	-0.7 ± 3.1 (110)								
	Number with ascites (ever)	520/674 (77%)	189/243, (78%)	0	0	294/572 (51%)	131/241 (54%)	None recorded	None recorded				
	Number with oesophageal varices (ever)	358/669 (53%)	120/239 (49%)	0	0	49/177 (28%)	23/87 (26%)	None recorded	None recorded				
	Number with encephalopathy (ever)	209/651 (31%)	91/240 (37%)	0	0	72/531 (14%)	24/226 (11%)	None recorded	None recorded				
	Number with HCC (ever)	105/674 (16%)	10/243 (4%)	0	0	9/157 (6%)	None recorded	None recorded	None recorded	49/477 (10.3%)	3/117 (2.6%)	0/4721 (0%)	0/1458 (0%)
	Number abstinent for ≥ 60 days	234/674 (35%)	154/243 (63%)	8/332 (2%)	3/142 (2%)	Not recorded	Not recorded	3/394 (0.8%)	1/186 (0.5%)				
	Number (%) with known diabetes	159/674 (24%)	39/243 (16%)	26/332 (8%)	5/142 (4%)					109/468 (23.3%)	16/117 (13.7%)	271/4813 (5.6%)	34/1466 (2.3%)
Cannabis use	Regular use, 5+ years, during period of high alcohol use	67/673 (10%)	8/142 (3%)	90/329 (27%)	24/142 (17%)								
	Years of regular use (if ever regular user)	17.5 ± 14.2 (66)	15.9 ± 11.0 (7)	15.5 ± 10.1 (89)	14.6 ± 10.4 (24)								
	Days per week marijuana	4.6 ± 2.3 (67)	5.1 ± 2.2 (8)	5.3 ± 2.4 (89)	5.7 ± 1.9 (24)								
	Occasions total	4424 ± 4742 (66)	4520 ± 4119 (7)	4374 ± 3795 (89)	4279 ± 3063 (24)								
Smoking history	Regular smoker (ever)	513/674 (76%)	156/243 (64%)	279/331 (84%)	111/142 (78%)					342/473 (72.5%)	78/116 (67.2%)	3589/4825 (74.4%)	1093/1464 (74.7%)
	Pack years (if ever smoker)	35.4 ± 32.3 (370)	24.1 ± 18.4 (96)	35.3 ± 27.8 (174)	33.8 ± 21.3 962)								
	Regular use, 5+ years, during period of high alcohol use	398/674 (59%)	113/243 (47%)	240/331 (73%)	88/142 (62%)								
Coffee intake	Years coffee (if regular user)	32.2 ± 12.9 (347)	28.9 ± 13.3 (100)	25.3 ± 11.8 (225)	24.2 ± 11.3 (81)								
	Cups caffeinated coffee per day (if regular user)	3.7 ± 3.5 (347)	3.7 ± 3.9 (100)	4.1 ± 3.6 (224)	3.4 ± 2.3 (81)					1.87 ± 2.23 (430)	1.80 ± 2.35 (110)	2.23 ± 2.36 (4438)	2.10 ± 2.05 (1378)

Supplementary Table 3. Comparison of score performance measures in men and women.

			ROC Curve	Logistic regression		Q1-Q5 Odds Ratio
			AUC	Beta	p-value	(95% CIs)
3-SNP score	GenomALC-1	Men	0.671 ± 0.016	1.132 ± 0.116	1.41×10^{-22}	6.18 (4.05 to 9.41)
		Women	0.650 ± 0.027	0.974 ± 0.192	3.79×10^{-7}	5.40 (2.67 to 10.92)
	GenomALC-2	Men	0.592 ± 0.017	0.575 ± 0.107	6.76×10^{-8}	2.47 (1.68 to 3.62
		Women	0.635 ± 0.025	0.897 ± 0.172	1.94×10^{-7}	3.81 (2.05 to 7.07)
	UK Biobank	Men	0.635 ± 0.016	0.800 ± 0.088	1.14×10^{-19}	3.44 (2.48 to 4.77)
		Women	0.554 ± 0.036	0.375 ± 0.196	0.056	2.08 (1.11 to 3.89)
3 SNP score + BMI, coffee	UK Biobank	Men	0.645 ± 0.017	0.795 ± 0.082	4.55×10^{-22}	3.50 (2.34 to 5.24)
		Women	0.594 ± 0.034	0.535 ± 0.164	0.0011	2.79 (1.37 to 5.71)
3-SNP-M score[i]	GenomALC-2	Men	0.626 ± 0.017	0.873 ± 0.124	1.94×10^{-12}	3.79 (2.50 to 5.75)
		Women	0.638 ± 0.025	0.978 ± 0.186	$1..42 \times 10^{-7}$	3.27 (1.77 to 6.06)
5-SNP-M score[i]	GenomALC-2	Men	0.622 ±0.017	0.791 ±0.116	7.87×10^{-12}	3.47 (2.31 to 5.23)
		Women	0.632 ±0.025	0.848 ± 0.174	1.09×10^{-6}	3.92 (2.18 to 7.03)
8-SNP-M score	GenomALC-2	Men	0.631 ± 0.017	0.789 ± 0.110	6.37×10^{-13}	3.34 (2.24 to 4.98)
		Women	0.635 ±0.025	0.836 ± 0.162	2.55×10^{-7}	3.33 (1.90 to 5.85)

Supplementary Table 4. Significance (p-values, not adjusted for multiple comparisons) for contrasts between groups of excessive drinkers with and without alcohol-related liver disease diagnoses in UK Biobank and GenomALC-2 cohorts. The dependent variable is the 3-SNP score. Boxes in the UK Biobank table emphasise the significant differences between the control groups and the more severe forms of liver disease, and the lack of significant differences among the more severe categories.

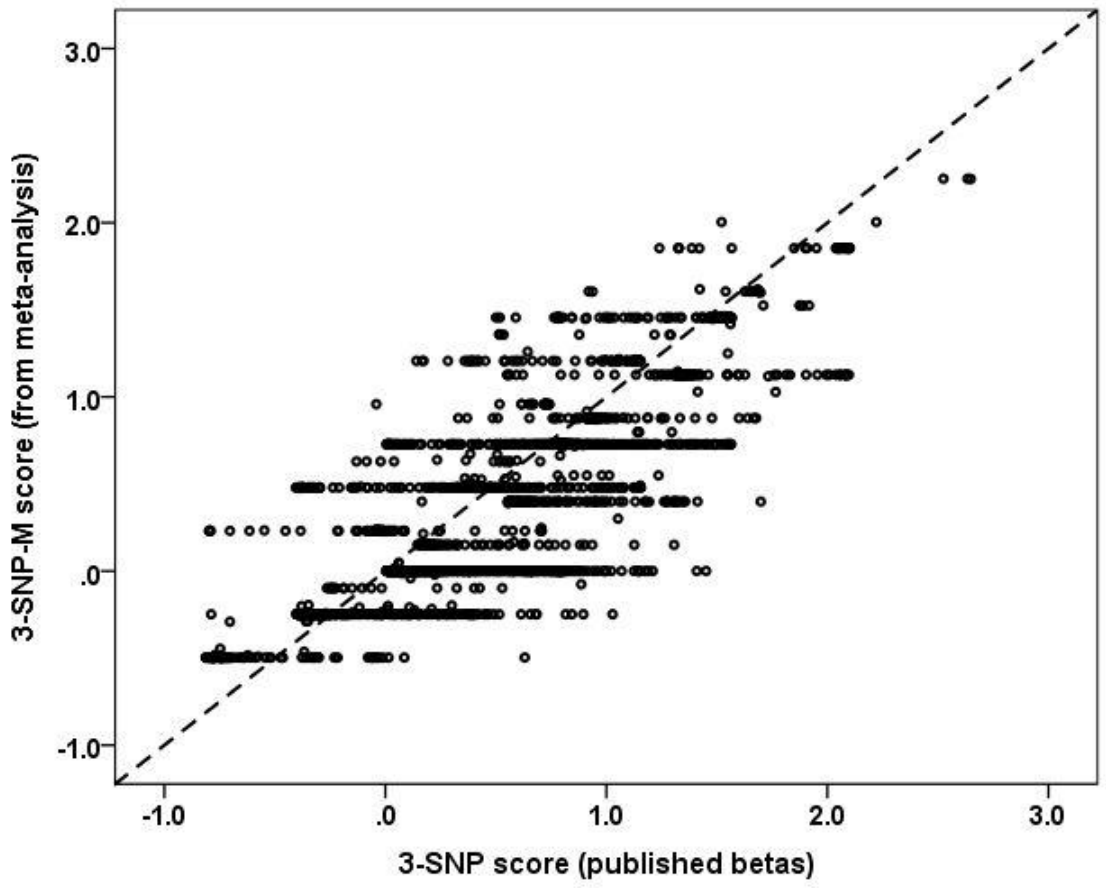
UK Biobank	N	Excessive drinker, no liver diagnosis	Alcoholic fatty liver	ALD unspecified	Alcoholic fibrosis sclerosis	Alcoholic hepatitis	Alcoholic hepatic failure	Alcoholic cirrhosis
Excessive drinker	5618	-						
Fatty liver	71	0.128	-					
ALD unspecified	327	0.0074	0.824	-				
Alcoholic Fibrosis/sclerosis	910	0.222	0.545	0.466	-			
Alcoholic hepatitis	796	7.75×10^{-4}	0.293	0.095	0.904	-		
Alcoholic hepatic failure	52	0.0023	0.191	0.073	0.922	0.667	-	
Alcoholic Cirrhosis	448	2.51×10^{-20}	0. 0332	3.30×10^{-5}	0.8831	0.335	0.814	-

GenomALC-2	N	Control	Alcoholic hepatitis	Alcoholic cirrhosis
Control	604	-		
Alcoholic hepatitis	1271	7.95×10^{-9}	-	
Alcoholic cirrhosis	1162	1.20×10^{-14}	0.011	-

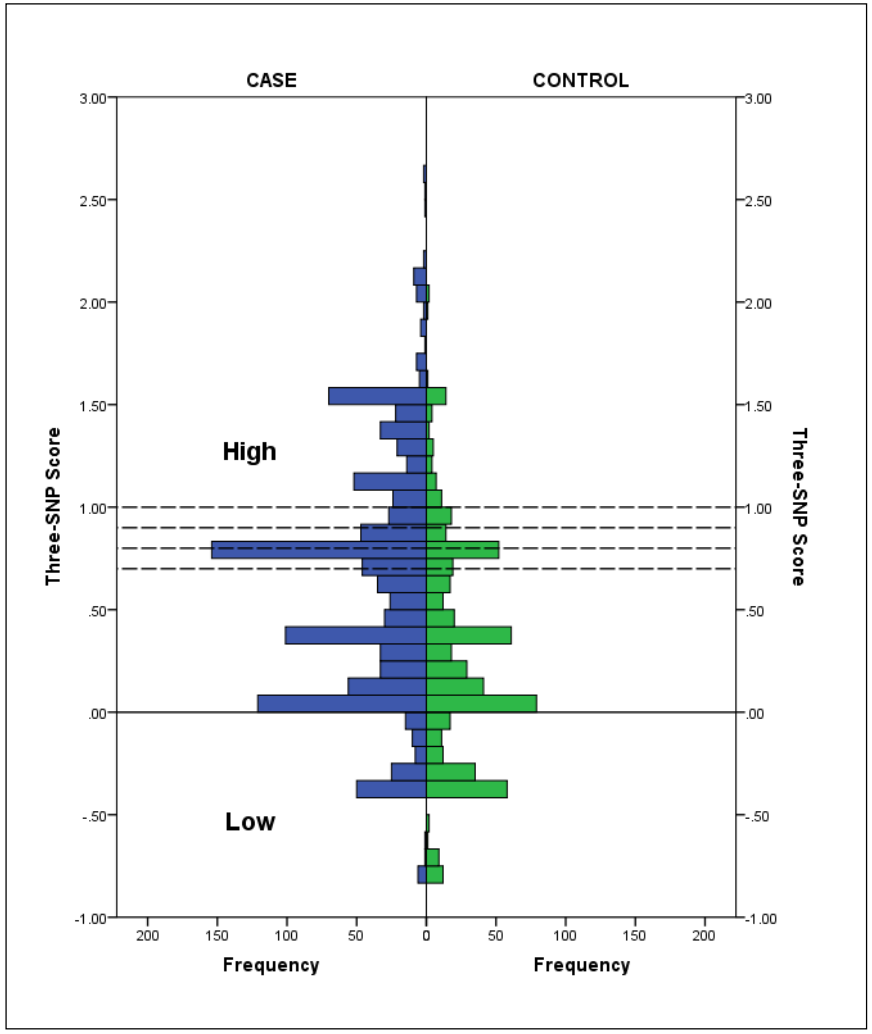
Supplementary Table 5. Effects of 3-SNP score on risk of alcohol-related liver diseases, graded by severity of liver disease, in GenomALC-1 sample and UK Biobank. In each cohort Controls were people who reported daily alcohol intake of ≥80 grams (men)/≥ 50 grams (women), for ≥10 years. UK Biobank participants with 'mild' alcoholic liver disease had ICD-10 diagnoses of K70.0 (alcoholic fatty liver) or K70.9 (alcoholic liver disease, unspecified); there was no comparable group in the GenomALC participants. 'Severe' alcoholic liver disease comprised ICD-10 diagnoses of K70.1 (alcoholic hepatitis), K70.3 (alcoholic cirrhosis) or K70.4 (alcoholic hepatic failure) for UK Biobank, and alcoholic cirrhosis for GenomALC. HCC, hepatocellular carcinoma (ICD-10 C22.0). Coefficients (B) and Odds Ratios (OR) are expressed per unit change in the 3-SNP score.

	GenomALC-1					UK Biobank				
	B	SE	p-value	OR	95% CI	B	SE	p-value	OR	95% CI
Control (excessive drinkers) versus Mild ALD	N/A					0.271	0.089	0.0023	1.31	1.10 to 1.56
Control (excessive drinkers) versus 'Severe' ALD but no HCC	1.020	0.101	3.49×10^{-24}	2.773	2.277 to 3.377	0.652	0.074	8.77×10^{-19}	1.92	1.66 to 2.22
Control (excessive drinkers) versus 'Severe' ALD with HCC	1.327	0.174	2.05×10^{-14}	3.769	2.683 to 5.296	1.264	0.229	3.38×10^{-8}	3.54	2.26 to 5.54
'Mild' ALD versus 'Severe' ALD but no HCC	N/A					0.378	0.111	6.50×10^{-4}	1.46	1.17 to 1.81
'Severe' ALD but no HCC versus 'Severe' ALD with HCC	0.401	0.155	0.010	1.493	1.102 to 2.022	0.631	0.242	0.0091	1.88	1.17 to 3.02

Supplementary Figure 1. Comparison of the 3-SNP and 3-SNP-M scores, for patients in the GenomALC-2 cohort calculated as in Table 1. The diagonal line shows x = y.



Supplementary Figure 2. Effect of varying the 3-SNP score cut-off for classification into the high-risk group on the proportion of cases stratified as high-risk and the Odds Ratios comparing the high-and low-risk groups. Data shown are for GenomALC-1.

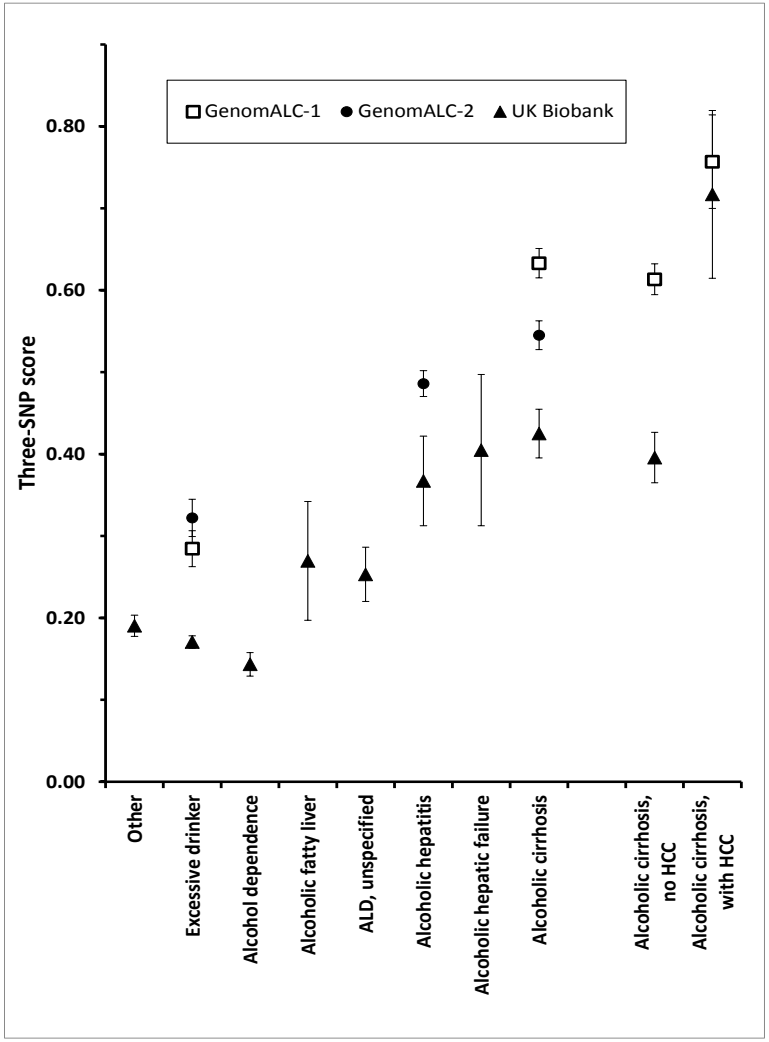


The left-hand panel shows the distribution of scores in Cases and Controls, as in Figure 1; those with scores below 0 (continuous horizontal line) are always considered as the low-risk group while those above the interrupted horizontal lines (at 0.7, 0.8, 0.9 or 1.0) are in the high-risk group. The Table shows results (OR) at each of the evaluated high-risk thresholds.

Cut-off for high-risk group	Proportion of cases above cut-off	Odds Ratio, High versus Low group (95% CI)
1.0	25%	7.35 (5.01 to 10.78)
0.9	28%	5.86 (4.13 to 8.33)
0.8	37%	4.95 (3.60 to 6.81)
0.7	49%	4.96 (3.67 to 6.71)

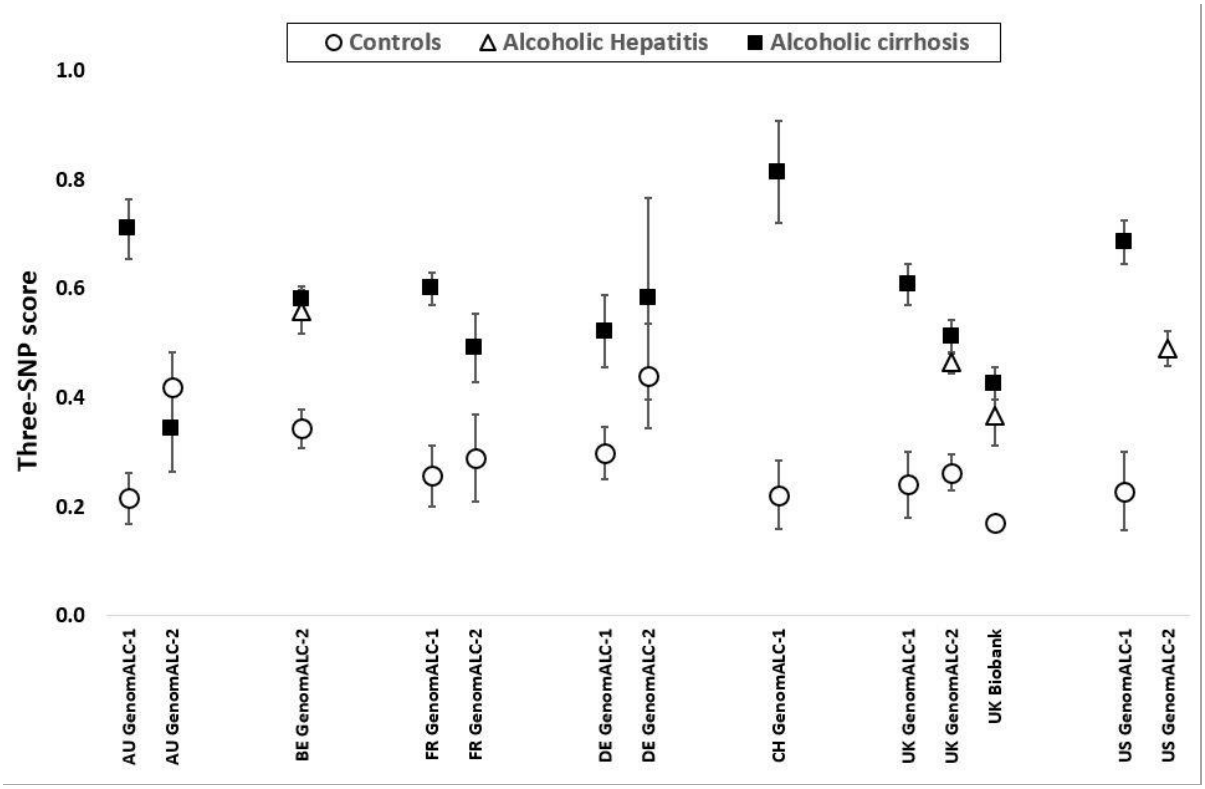
.

Supplemental Figure 3. Comparison of means for 3-SNP scores between diagnostic groups from three independent studies. Points and bars show means and standard errors. *Other:* no excessive drinking; *Excessive drinker:* high-risk drinking by the 50/80 grams/day criterion; *Alcohol dependence:* alcohol dependence [ICD-10 F10.2]; *Categories of alcohol-related liver disease are:* alcoholic fatty liver, ALD unspecified, alcoholic hepatitis, alcoholic hepatic failure and alcoholic cirrhosis (overall, and sub-divided by HCC status).

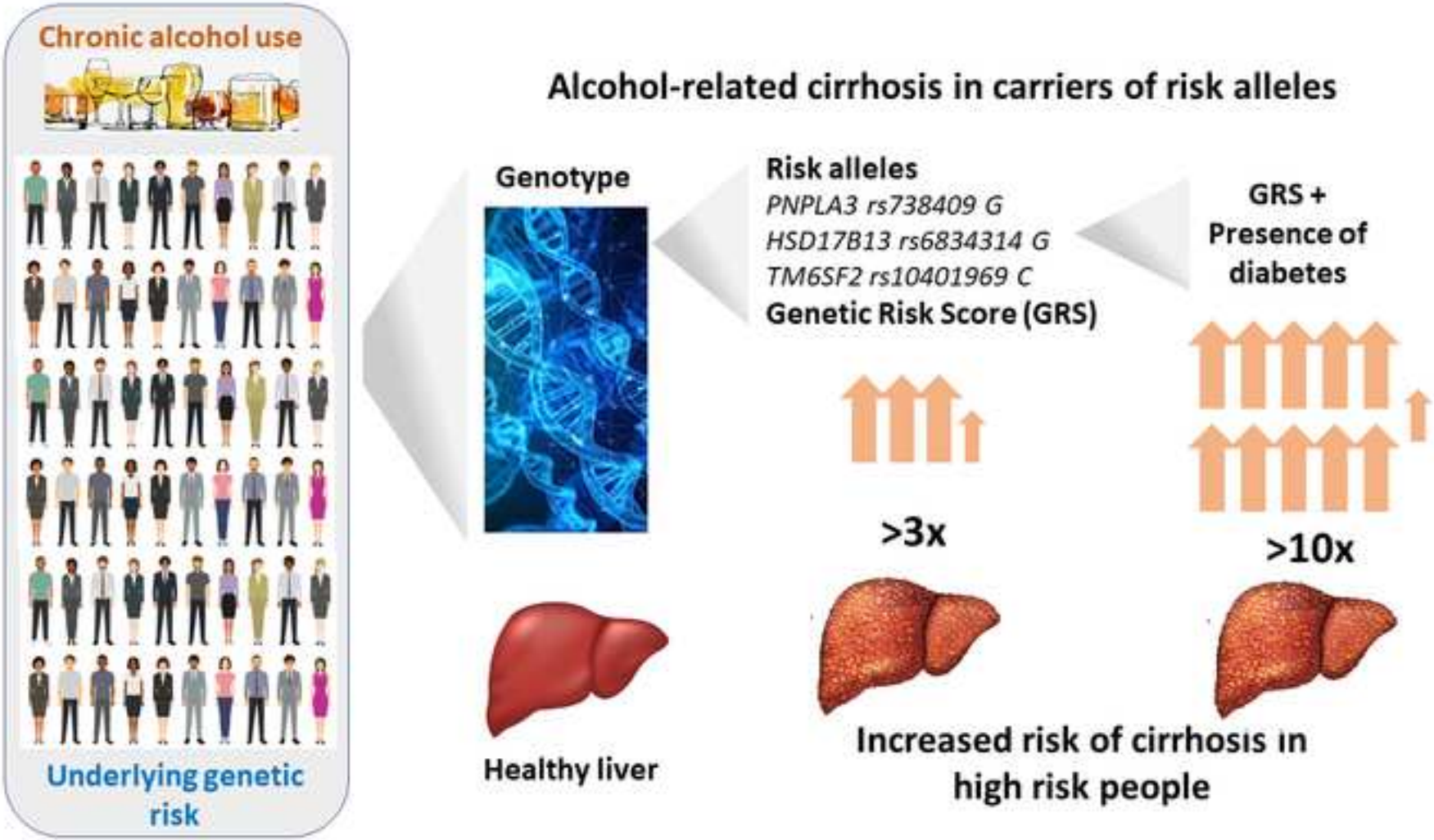


	GenomALC-1	GenomALC-2	UK Biobank
Other			**0.190 ± 0.013**
Excessive drinker	**0.285 ± 0.022**	**0.322 ± 0.023**	**0.171± 0.008**
Alcohol dependence			**0.143 ± 0.014**
Alcoholic fatty liver			**0.270 ± 0.073**
ALD, unspecified			**0.253 ± 0.033**
Alcoholic hepatitis		**0.486 ± 0.016**	**0.367 ± 0.055**
Alcoholic hepatic failure			**0.405 ± 0.092**
Alcoholic cirrhosis	**0.633 ± 0.018**	**0.545 ± 0.018**	**0.425 ± 0.030**
Alcoholic cirrhosis, no HCC	**0.613 ± 0.019**		**0.396 ± 0.031**
Alcoholic cirrhosis, with HCC	**0.757 ± 0.057**		**0.717 ± 0.102**

Supplementary Figure 4. Means of the 3-SNP score by country (AU Australia, BE Belgium, FR France, DE Germany, CH Switzerland, UK United Kingdom, US United States). Error bars show standard errors for the means.



Graphical Abstract



JHEPAT-D-21-01108
A genetic risk score and diabetes predicts development of alcohol-related cirrhosis in drinkers

HIGHLIGHTS

- Currently there is no way to know who amongst alcohol users will develop cirrhosis, but underlying genetic factors (SNPs) are known to be associated with risk of alcohol-related cirrhosis.
- Our 3-SNP Genetic Risk Score (GRS) using PNPLA3:rs73840-G, SUGP1TM6SF2:rs10401969-C and HSD17B13:rs6834314-G risk alleles stratified people at low-/high-risk of alcohol-related cirrhosis.
- High GRS increased relative risk of cirrhosis more than 3-fold in alcohol users.
- Presence of diabetes with high GRS further increased the risk more than 10-fold.
- A GRS based on only three genetic risk variants and diabetes status can provide meaningful risk stratification for cirrhosis in excess drinkers.