



HAL
open science

A genetic risk score and diabetes predict development of alcohol-related cirrhosis in drinkers

John B Whitfield, Tae-Hwi Schwantes-An, Rebecca Darlay, Guruprasad P Aithal, Stephen R Atkinson, Ramon Bataller, Greg Botwin, Naga P Chalasani, Heather J Cordell, Ann K Daly, et al.

► **To cite this version:**

John B Whitfield, Tae-Hwi Schwantes-An, Rebecca Darlay, Guruprasad P Aithal, Stephen R Atkinson, et al.. A genetic risk score and diabetes predict development of alcohol-related cirrhosis in drinkers. *Journal of Hepatology*, 2022, 76 (2), pp.275-282. 10.1016/j.jhep.2021.10.005 . hal-03418937

HAL Id: hal-03418937

<https://hal.science/hal-03418937>

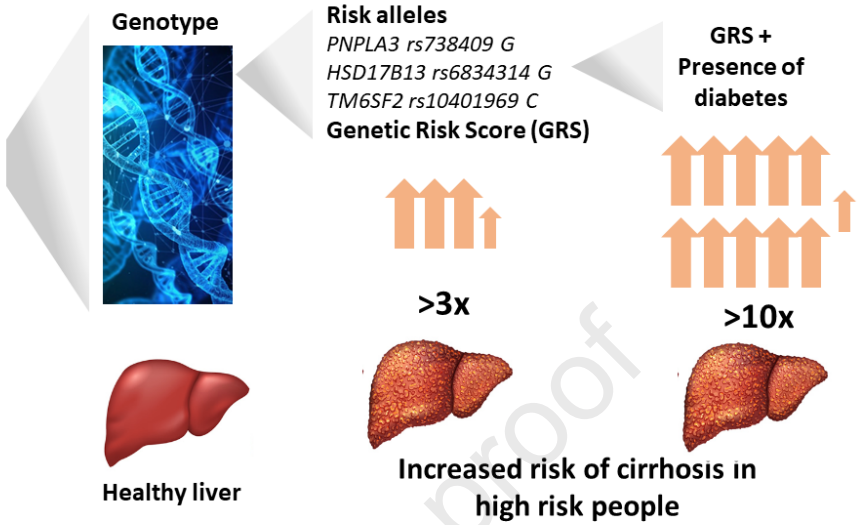
Submitted on 10 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Alcohol-related cirrhosis in carriers of risk alleles



Journal Pre-proof

1 **A genetic risk score and diabetes predict development of alcohol-related**
2 **cirrhosis in drinkers**

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

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

26 Cedars-Sinai Medical Center, Los Angeles, California CA 90048, USA (9) Department of
27 Medicine, Indiana University, Indianapolis, IN 46202-5175, USA, (10) Faculty of Medical
28 Sciences, Newcastle University Medical School, Framlington Place, Newcastle upon Tyne
29 NE2 4HH, United Kingdom, (11) Newcastle University, Framlington Place, Newcastle upon
30 Tyne NE2 4HH, United Kingdom, (12) Division of Clinical Toxicology, Department of
31 Internal Medicine 2, Klinikum rechts der Isar, School of Medicine, Technical University of
32 Munich, Ismaninger Str. 22, 81675 Munich, Germany, (13) Liver Unit, Sheffield Teaching
33 Hospitals, AO Floor Robert Hadfield Building, Northern General Hospital, Sheffield S5 7AU,
34 UK, (14) Laboratory of Neurogenetics, NIAAA, Rockville, MD 20852, USA, (15) Drug Health
35 Services, Royal Prince Alfred Hospital, Missenden Road, Camperdown, NSW 2050, Australia,
36 (16) Faculty of Medicine and Health, the University of Sydney, Sydney, NSW 2006, Australia,
37 (17) Service Addictologie, CHRU Caremeau, 30029 Nîmes, France, (18) Division of
38 Gastroenterology and Hepatology, Department of Medicine, Indiana University and
39 Roudebush Veterans Administration Medical Center, Indianapolis, USA, (19) CHRU de Lille,
40 Hôpital Claude Huriez, Rue M. Polonovski CS 70001, 59 037 Lille Cedex, France, (20) Univ
41 Rennes, INRA, INSERM, CHU Rennes, Institut NUMECAN (Nutrition Metabolisms and
42 Cancer), F-35000 Rennes, France, (21) Molecular Psychiatry Laboratory, Division of
43 Psychiatry, University College London, London WC1E 6DE, UK, (22) CUB Hôpital Erasme,
44 Université Libre de Bruxelles, clinique d'Hépatologie, Brussels, Belgium; Laboratory of
45 Experimental Gastroenterology, Université Libre de Bruxelles, Brussels, Belgium, (23) UCL
46 Institute for Liver & Digestive Health, Division of Medicine, Royal Free Campus, University
47 College London, London NW3 2PF, UK, (24) Department of Internal Medicine, Salem
48 Medical Center and Center for Alcohol Research, University of Heidelberg, Zeppelinstraße 11-
49 33, 69121 Heidelberg, Germany, (25) Department of Gastroenterology and Hepatology,
50 University Hospital Zurich, Rämistrasse 100, CH-8901 Zurich, Switzerland, (26) Lerner

51 Research Institute, 9500 Euclid Avenue, Cleveland, Ohio, OH 44195, USA, (27) Service
52 d'Hépatologie, APHP Hôpital Avicenne et Université Paris 13, Bobigny, France, (28)
53 University Paris 13, Bobigny, France, (29) Inserm U1162 Génomique fonctionnelle des
54 tumeurs solides, Paris, France, (30) DISC, Inserm, 75013 Paris, France, (31) Hôpital Antoine-
55 Bécclère, 157 Rue de la Porte de Trivaux, 92140 Clamart, France, UM1, INSERM U1018 (32)
56 Hôpital Universitaire Caremeau, Place du Pr. Robert Debre, 30029 Nîmes, France, (33) MRC
57 Centre for Drug Safety Science, Liverpool Centre for Alcohol Research, University of
58 Liverpool, The Royal Liverpool and Broadgreen University Hospitals NHS Trust, and
59 Liverpool Health Partners, Liverpool, L69 3GL, UK, (34) Psychiatric Hospital University of
60 Munich, Nussbaumsstr.7, 80336 Munich, Germany and Privatklinik Meiringen, Willigen, CH
61 3860 Meiringen, Switzerland, (35) Health Analytics, Lane Clark & Peacock LLP, London, UK,
62 (36) Department of Medicine, University of California, Irvine, USA, (37) Centenary Institute
63 of Cancer Medicine and Cell Biology, the University of Sydney, Sydney, NSW 2006,
64 Australia.

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

66 *Equal senior authors

67 #Corresponding authors

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

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

72 **KEYWORDS**

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

75 **Word count: 3546**

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

77 **Figures 4.**

78 **CONFLICT OF INTEREST**

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

91

92 **FINANCIAL SUPPORT:** Funding for this study was provided by NIH/NIAAA
93 U01AA018389 and RO1AA018389 for data collection, analysis, interpretation and patient
94 recruitment. SL: U01 grant from the NIAAA and R01 AA025208, U01 AA026917, and

95 1I01CX000361. MT: The UK Medical Research Council Stratified Medicine Award (Ref
96 MR/R014019/1), NIHR Imperial Biomedical Research Centre and NIHR Senior Investigator
97 Award (NIHR 200153).

98

99 **AUTHOR CONTRIBUTIONS**

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

108

109 **ABSTRACT (273 words)**

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

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

120 **Results:** A combination of three single nucleotide polymorphisms (SNPs) (*PNPLA3*:rs738409,
121 *SUGP1-TM6SF2*:rs10401969, *HSD17B13*:rs6834314) and diabetes status best discriminated
122 for cirrhosis risk. The odds ratio (OR) and 95% confidence intervals (CI) for the extreme score
123 quintiles (Q1-Q5) of the 3-SNP score, based on independent allelic effect size estimates, were
124 5.99 (4.18;8.60) (GenomALC-1); 2.81 (2.03;3.89) (GenomALC-2); and 3.10 (2.32;4.14) (UK
125 Biobank). Patients with diabetes and high-risk score, compared to those without diabetes and
126 a low-risk score, had ORs increased to 14.7 (7.69;28.1) (GenomALC-1) and 17.1 (11.3;25.7)
127 (UK Biobank). Patients with cirrhosis and HCC had significantly higher mean risk scores than
128 patients with cirrhosis alone (0.76 ± 0.06 versus 0.61 ± 0.02 , $p=0.007$). Score performance was
129 not significantly enhanced by information on additional genetic risk variants, body mass index
130 or coffee consumption.

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

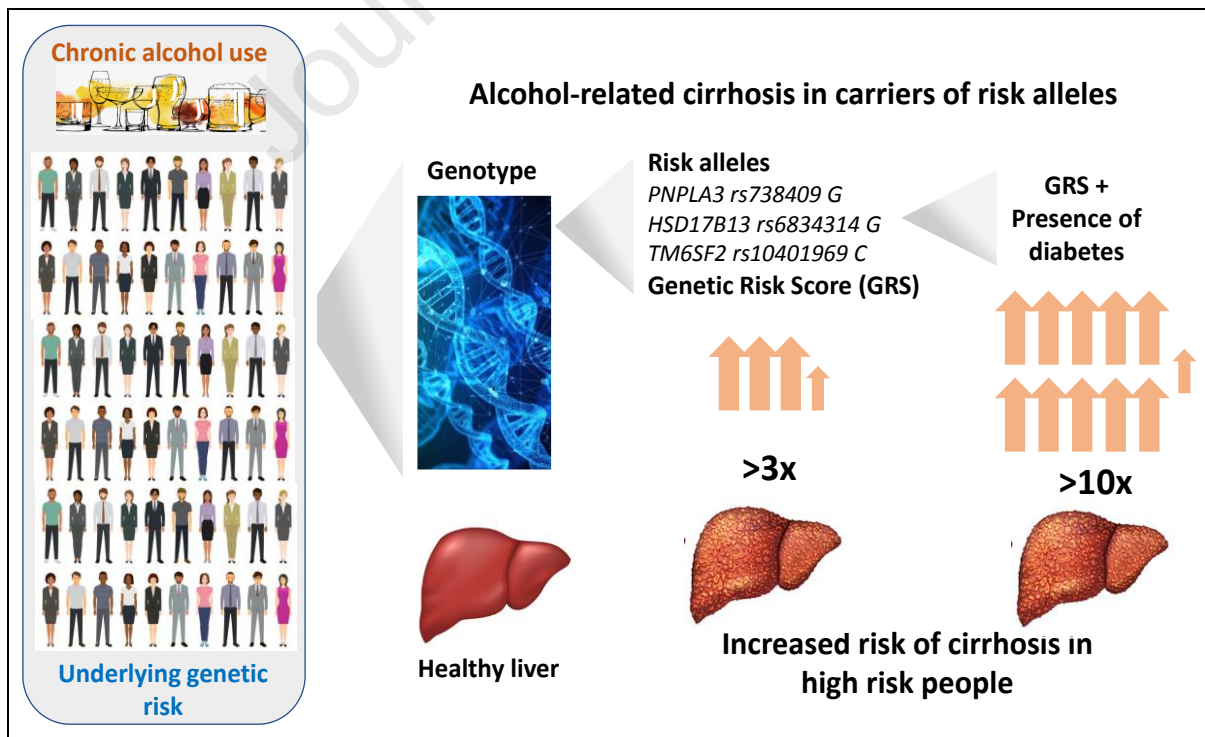
134

135 LAY SUMMARY

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

142

143 GRAPHICAL ABSTRACT



144

145 **INTRODUCTION**

146 Although the risk for developing cirrhosis is positively associated with alcohol consumption,
147 only a minority of people with high-risk alcohol intake develop cirrhosis. The prevalence can
148 vary between 7-16%^{1,2} with some reports suggesting the prevalence to be as low as 2%^{3,4}. The
149 risk threshold for what is considered high-risk intake has changed over time⁵⁻⁷. Long-term
150 consumption of 80 grams per day (g/d) or more is associated with increased risk of cirrhosis^{8,9},
151 but the threshold for liver harm is below this level, especially for women^{10,11}. The 80 g/d (men)
152 and 50 g/d (women) cut-offs were set at a relatively high level to ensure both the cirrhosis and
153 control groups were exposed to a substantial level of alcohol-related risk. We used this
154 threshold to define “heavy drinking” in this study.

155 Primary prevention of alcohol-related liver disease (ALD) would involve decreasing alcohol
156 intake of the whole population but achieving this remains challenging. Focused intervention
157 through the identification of people with high alcohol intake or more specifically through
158 stratification of individuals within this population at risk for developing cirrhosis depends on
159 identification of those at high risk. Evidence from clinical trials¹² suggests that informing
160 excessive drinkers that they have abnormal liver function tests/hepatic fibrosis can motivate
161 them to reduce their alcohol intake. A number of both constitutional¹³⁻¹⁶, and genetic¹⁷⁻²⁰ risk
162 factors for the development of alcoholic-related cirrhosis have been identified, but no attempt
163 appears to have been made, to date, to bring these together to provide an integrated measure of
164 risk. Thus, the aim of this study was to devise risk scores for the stratification of cirrhosis risk
165 and evaluate them in heavy drinkers from three independent cohorts.

166

167

168 **MATERIALS/PATIENTS AND METHODS**

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

174 ***GenomALC-1***

175 The GenomALC-1 cohort was recruited according to a pre-designed protocol between 2012
176 and 2017 in Australia, France, Germany, Switzerland, the UK, and the USA. The recruitment
177 criteria and the data collection protocol were detailed previously²¹. Briefly, all participants had
178 a history of heavy drinking (≥ 80 g/d (men) and ≥ 50 g/d (women) for ≥ 10 years). For cases,
179 cirrhosis had been diagnosed by a combination of clinical criteria, laboratory variables and/or
180 liver elastography (Fibroscan®), with liver biopsy if clinically indicated. Clinical features
181 defining the severity of cirrhosis are shown in Supplementary Table 2. Other liver diseases
182 (hepatitis B or C, haemochromatosis, Wilson's disease, and autoimmune hepatitis) were
183 excluded by laboratory testing or clinical criteria. For controls, liver disease was excluded
184 through a combination of clinical history and measurement of liver function tests (bilirubin,
185 albumin, ALT). For both cases and controls, HIV infection was an exclusion criterion. The
186 study was approved by appropriate Ethics Committees or Institutional Review Boards at each
187 site and conformed to the ethical guidelines of the 1975 Declaration of Helsinki. Participants
188 were provided with explanations of the study and gave written informed consent. Genotyping
189 was performed at Erasmus University Medical Centre, Rotterdam using the Illumina GSA
190 genotyping array, as described²⁰.

191 ***GenomALC-2***

192 The biological samples and data were donated by research groups who had independently
193 collected them for other studies. Some of the GenomALC-2 samples were included in a
194 previous GWAS¹⁷; therefore, for the purposes of this study, overlapping samples were removed
195 from the analysis. Clinical diagnosis of cases and controls was similar to GenomALC-1 criteria
196 but detailed clinical information was limited for this cohort. Patients had given informed
197 consent and the studies were approved by the appropriate Ethics Review Boards. DNA from
198 these participants' samples was also genotyped as outlined above for GenomALC-1.

199 Genotypes in the GenomALC-1 and GenomALC-2 cohorts were²² cleaned using a widely used
200 quality control pipeline, the GWASTools package
201 <https://bioconductor.org/packages/devel/bioc/manuals/GWASTools/man/GWASTools.pdf>
202 and imputed to 1000 Genomes reference using the Michigan Imputation Server (MIS)²²

203 ***UK Biobank***

204 The UK Biobank²³ includes approximately 500,000 volunteers from the UK with a wide range
205 of data including computer-administered questionnaires, physical measurements, laboratory
206 tests, and genotyping. All participants gave informed consent, consistent with the UK Biobank
207 Ethics and Governance Framework. Recruitment and initial assessment occurred between 2006
208 and 2010 when participants were aged 40 to 69 years. Access to the UK Biobank database was
209 obtained (Application 18870) and relevant data (with diagnoses updated to June 2020) were
210 extracted. For cases, information was restricted to assigned clinical diagnosis (Supplementary
211 Table 2) on hospital admissions and diagnoses, and on causes of death in participants who have
212 subsequently died. Information was available on self-reported alcohol intake at the time of
213 assessment (by beverage type, in drinks per week, or for less frequent drinkers per month), and
214 participants also reported whether this was less than, similar to or more than they had been
215 consuming 10 years previously. The amounts were converted to express the alcohol intake in
216 g/d. Participants who had a recorded diagnosis or cause of death of alcohol-related cirrhosis

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

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

227 *Data curation and statistical analysis*

228 Data management and statistical analyses used IBM SPSS Statistics, version 22 (IBM Corp.,
229 New York NY). Binary variables were coded as 0 (absent) or 1 (present). Diabetes status
230 (absent/present), BMI, kg/m²) and coffee consumption (0: not a coffee consumer, 1: coffee
231 consumer) shown in our previous report as associated with cirrhosis¹⁶ were also modelled.
232 Genotype data were coded as single nucleotide polymorphisms (SNPs) minor allele dosages,
233 assuming an additive model for allelic effects.

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

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

241 analysis²⁰. Two significantly associated SNPs have been reported at *SUGP1-TM6SF2* locus¹⁷
242 which are in near-complete linkage disequilibrium ($d'1.00$, r^2 0.955), and rs10401969 was
243 chosen over rs58542926 because of its stronger association with cirrhosis.

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

250 Scores based on three, five, and eight loci were also computed for the GenomALC-2 samples
251 using coefficients of loci with significant association from the published meta-analysis²⁰ or
252 other sources^{17,18} ('3-SNP-M', '5-SNP-M' and '8-SNP-M' scores) (Table 1). The 3-SNP-M
253 score was based on the loci mentioned above, the 5-SNP-M score included above three loci,
254 and *SERPINA1* and *FAF2* identified in our meta-analysis, and the 8-SNP-M score which was
255 derived from the 5-SNP-M score with addition of three reported loci (*MBOAT7*, *MTARCI*
256 [previously *MARCI*], *HNRNPUL1*) significantly associated with alcohol-related
257 cirrhosis^{17,24,25}.

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

261

262 **RESULTS**

263 *Risk stratification by genetic loci-based scores*

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

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

279 The addition of further loci in the 5-SNP-M score (*PNPLA3*:rs2294915, *SUGP1*-
280 *TM6SF2*:rs10401969, *HSD17B13*:rs10433937, *SERPINA1*:rs28929474,
281 *FAF2*:rs11134997)^{17,24,25} and in the 8-SNP-M score, with *MBOAT7*:rs641738,
282 *MTARC1*:rs2642438 and *HNRNPUL1*:rs17251589 in addition to those in the 5-SNP-M score,
283 did not improve the associations between score and outcome or the risk stratification (Table 2).
284 Because the coefficients for *FAF2* and *SERPINA1* were obtained from the meta-analysis of the
285 GenomALC-1, Buch study¹⁷ and UK Biobank data, the 5-SNP-M and 8-SNP-M scores could
286 only be tested in the GenomALC-2 data. To allow a valid comparison between the multi-SNP
287 scores each was based on the coefficients from our meta-analysis of GWAS results. This
288 resulted in an improvement for the meta-analysis-based 3-SNP-M score compared to the 3-

289 SNP score (Q5-Q1 ORs changed from 2.81 [95% CI 2.03,3.89] to 3.65 [2.59,5.15]). There was
290 also a high correlation between the 3-SNP and 3-SNP-M scores in GenomALC-2 ($r = 0.826$, n
291 $= 3037$, $p < 10^{-200}$; Supplementary Figure 1).

292 *Clinical utility of the risk score*

293 Numerical cut-offs that define or quantify risk are needed if the risk score is to have clinical
294 utility. The 3-SNP scores in the GenomALC-1 cases and controls for the lowest and highest
295 quintile boundaries were close to 0 and 1 (0.033 and 0.964, respectively; Figure 1). Division
296 of the scores into three groups at low, intermediate and high cirrhosis risk was based on the
297 3-SNP score distribution (Supplementary Figure 2). The final selected scores were, low: <0 ;
298 intermediate $>0 - 0.7$ and high risk >0.7 . In each study cohort the risk difference between the
299 low- and high-risk groups ranged between 2.5-fold and approximately 5-fold (Table 3). The
300 difference in risk between the high- and low-risk GenomALC-1 groups were similar across
301 the six countries (Figure 2).

302 *Diabetes*

303 Diabetes is known to have a large effect on cirrhosis risk. Inclusion of diabetes status with
304 genetic risks in a combined risk score led to a bimodal distribution and difficulty in defining
305 score quintiles. Thus, to see the effect of genetic risk score in the context of diabetes status, the
306 3-SNP score was subdivided by the diabetes status and is presented separately (Table 4).

307 People with diabetes showed a substantial increase in the risk of cirrhosis in both the
308 GenomALC-1 (OR 3.82, 95% CI 2.67;5.47) and the UK Biobank (OR 5.62, 95% CI 4.33;7.28)
309 cohorts. The genetic score effects were similar for people with and without diabetes, both in
310 the GenomALC-1 (logistic regression coefficients \pm SE, no diabetes: 1.055 ± 0.105 ; diabetes:
311 1.276 ± 0.338) and the UK Biobank data (no diabetes: 0.653 ± 0.093 ; diabetes: 0.735 ± 0.181).
312 Tests for genetic score-diabetes interaction, either by including a (score x diabetes) term in the

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

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

320 The mean values for the 3-SNP score varied across groups defined by alcohol intake and by
321 the diagnostic categories for alcohol-related liver disease for both GenomALC-1 and the UK
322 Biobank cohorts (Supplementary Figure 3). *Post hoc* comparisons showed similar trends of
323 mean 3-SNP risk score increasing with disease severity for the GenomALC-2 cohort that
324 included excessive drinkers with no liver disease and significantly differed between cases with
325 severe alcoholic hepatitis and alcohol-related cirrhosis ($p = 0.011$) (Supplementary Table 4).
326 Mean 3-SNP score increased with severity of liver disease (Supplementary Figure 3), including
327 when comparing cirrhosis with HCC against cirrhosis without HCC, both for GenomALC-1
328 (0.757 ± 0.057 versus 0.613 ± 0.019) and UK Biobank (0.717 ± 0.102 versus 0.396 ± 0.031);
329 see also Supplementary Table 5.

330

331 **DISCUSSION**

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

335 *Development of score for risk stratification*

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

340 Our main measure of genetic risk stratification was to compare people who are in the highest
341 quintile for a score against those in the lowest quintile, providing a more practical measure of
342 stratification success than comparing the most extreme of all possible categories, which will
343 usually contain only a small proportion of people²⁶. Substantial Q5-Q1 risk differences were
344 evident for the simple 3-SNP score in each of the cohorts; approximately six-fold in the
345 GenomALC-1 cohort and three-fold in the other cohorts (Table 2). The greater difference in
346 Q5-Q1 risk for GenomALC-1 is likely to be due to a more refined and pre-defined case-control
347 definition for the recruitment protocol in this cohort.

348 Diabetes status led to a substantial enhancement of the utility of the 3-SNP score, predicting a
349 >10-fold difference in risk between extreme groups (Q5 with diabetes and Q1 non-diabetes).
350 Adding information on further genetic risk variants or BMI and coffee consumption had
351 minimal effect.

352 *Clinical utility of risk-score*

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

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

368 The 3-SNP risk score was also associated with differences across the alcohol-related liver
369 disease spectrum, including HCC. The HCC risk association is consistent with previous
370 information showing that *PNPLA3*, *HSD17B13* and *TM6SF2* polymorphisms²⁷⁻³¹ are
371 associated with a higher risk for this condition compared to advanced cirrhosis, perhaps
372 suggesting a pro-oncogenic role for these variants.

373 *Scope of risk-score*

374 The loci comprising the current risk score are also implicated in the risk for developing
375 cirrhosis of diverse aetiologies. Using similar polygenic risk scores (PRS) in non-alcoholic
376 fatty liver disease (NAFLD) revealed that combining genetic and clinical features refines the
377 predictive utility of the algorithm for identifying those at higher risk of severe liver disease³²⁻
378 ³⁴. Given the many shared genetic and metabolic risks between alcohol-related liver disease
379 and NAFLD, the predictive algorithm defined here may have a wider scope across these
380 diseases for risk stratification of those at higher risk of cirrhosis. Recently Emdin and
381 colleagues³⁰ identified 12 variants, five previously known, including *PNPLA3*, *HSD17B13* and
382 *TM6SF2*, and seven novel, which were associated at genome-wide significance with ‘any cause’
383 cirrhosis, and aggregated these into a PRS. A high PRS, defined as the top quintile of the

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

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

396 The outcome of risk stratification for alcohol-related liver disease can be compared with PRS
397 approaches to other complex diseases, including cardiovascular disease and cancers. A recent
398 study³⁵ showed that for five common diseases (coronary heart disease, type 2 diabetes, atrial
399 fibrillation, breast cancer and prostate cancer), Q1-Q5 differences in PRS were associated with
400 approximately two- to five-fold differences in the cumulative prevalence of diagnosis by age
401 80. Our 3-SNP score performance was equal to or slightly better than these.

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

408 such that the effectiveness of our score would be under-, rather than over-estimated. Finally,
 409 the risk scores were derived from groups of heavy drinkers with cirrhosis or without liver
 410 disease. However, these were validated in case and control groups selected from the
 411 population-based UK Biobank cohort. Application of the risk score to an individual patient
 412 should be performed with an understanding that some patients' outcomes will differ from those
 413 predicted by the score. Prospective studies are needed, both to relate score to progression across
 414 time in patients who present with early stages of liver disease, and to clarify the relationship
 415 between onset of diabetes and of advanced liver disease in patients with excessive alcohol use.
 416 Based on the findings of the present study a 3-SNP score algorithm is proposed for use and
 417 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: $(0.7839 * PNPLA3 \text{ rs738409 G dosage}) + (0.5423 * SUGP1-TM6SF2 \text{ rs10401969 C dosage}) - (0.4463 * HSD17B13 \text{ rs6834314 G dosage})$ 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. | | |

418

419 **Conclusions**

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

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

427

428 **ACKNOWLEDGEMENTS**

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

438

439

440 **REFERENCES**

- 441 [1] Askgaard G, Leon DA, Kjaer MS, Deleuran T, Gerds TA, Tolstrup JS: Risk for alcoholic
442 liver cirrhosis after an initial hospital contact with alcohol problems: A nationwide
443 prospective cohort study. *Hepatology* 2017, 65:929-37. 10.1002/hep.28943.
- 444 [2] Askgaard G, Kjaer MS, Tolstrup JS: Opportunities to Prevent Alcoholic Liver Cirrhosis
445 in High-Risk Populations: A Systematic Review With Meta-Analysis. *Am J*
446 *Gastroenterol* 2019, 114:221-32. 10.1038/s41395-018-0282-6.
- 447 [3] Hasin DS, Stinson FS, Ogburn E, Grant BF: Prevalence, correlates, disability, and
448 comorbidity of DSM-IV alcohol abuse and dependence in the United States: results from
449 the National Epidemiologic Survey on Alcohol and Related Conditions. *Arch Gen*
450 *Psychiatry* 2007, 64:830-42. 10.1001/archpsyc.64.7.830.
- 451 [4] Wong T, Dang K, Ladhani S, Singal AK, Wong RJ: Prevalence of Alcoholic Fatty Liver
452 Disease Among Adults in the United States, 2001-2016. *JAMA* 2019, 321:1723-5.
453 10.1001/jama.2019.2276.
- 454 [5] Pequignot G, Tuyns AJ, Berta JL: Ascitic cirrhosis in relation to alcohol consumption.
455 *IntJEpidemiol* 1978, 7:113-20.
- 456 [6] Rehm J, Taylor B, Mohapatra S, Irving H, Baliunas D, Patra J, et al.: Alcohol as a risk
457 factor for liver cirrhosis: a systematic review and meta-analysis. *Drug Alcohol Rev* 2010,
458 29:437-45. 10.1111/j.1465-3362.2009.00153.x.
- 459 [7] Askgaard G, Gronbaek M, Kjaer MS, Tjonneland A, Tolstrup JS: Alcohol drinking
460 pattern and risk of alcoholic liver cirrhosis: a prospective cohort study. *J Hepatol* 2015,
461 62:1061-7. 10.1016/j.jhep.2014.12.005.
- 462 [8] Corrao G, Bagnardi V, Zambon A, Arico S: Exploring the dose-response relationship
463 between alcohol consumption and the risk of several alcohol-related conditions: a meta-
464 analysis. *Addiction* 1999, 94:1551-73.
- 465 [9] Lelbach WK: Epidemiology of alcoholic liver disease. *ProgLiver Dis* 1976, 5:494-515.

- 466 [10] Roerecke M, Vafaei A, Hasan OSM, Chrystoja BR, Cruz M, Lee R, et al.: Alcohol
467 Consumption and Risk of Liver Cirrhosis: A Systematic Review and Meta-Analysis. *Am*
468 *J Gastroenterol* 2019, 114:1574-86. 10.14309/ajg.0000000000000340.
- 469 [11] Tuyns AJ, Pequignot G: Greater risk of ascitic cirrhosis in females in relation to alcohol
470 consumption. *IntJEpidemiol* 1984, 13:53-7.
- 471 [12] Subhani M, Knight H, Ryder S, Morling JR: Does Advice Based on Biomarkers of Liver
472 Injury or Non-Invasive Tests of Liver Fibrosis Impact High-Risk Drinking Behaviour: A
473 Systematic Review With Meta-analysis. *Alcohol Alcohol* 2021 10.1093/alcalc/agua143.
- 474 [13] Hart CL, Morrison DS, Batty GD, Mitchell RJ, Davey Smith G: Effect of body mass
475 index and alcohol consumption on liver disease: analysis of data from two prospective
476 cohort studies. *BMJ* 2010, 340:c1240. 10.1136/bmj.c1240.
- 477 [14] Liangpunsakul S, Puri P, Shah VH, Kamath P, Sanyal A, Urban T, et al.: Effects of Age,
478 Sex, Body Weight, and Quantity of Alcohol Consumption on Occurrence and Severity of
479 Alcoholic Hepatitis. *Clin Gastroenterol Hepatol* 2016, 14:1831-8 e3.
480 10.1016/j.cgh.2016.05.041.
- 481 [15] Saab S, Mallam D, Cox GA, 2nd, Tong MJ: Impact of coffee on liver diseases: a
482 systematic review. *Liver Int* 2014, 34:495-504. 10.1111/liv.12304.
- 483 [16] Whitfield JB, Masson S, Liangpunsakul S, Mueller S, Aithal GP, Eyer F, et al.: Obesity,
484 Diabetes, Coffee, Tea, and Cannabis Use Alter Risk for Alcohol-Related Cirrhosis in 2
485 Large Cohorts of High-Risk Drinkers. *Am J Gastroenterol* 2020, 116:106-15.
486 10.14309/ajg.0000000000000833.
- 487 [17] Buch S, Stickel F, Trepo E, Way M, Herrmann A, Nischalke HD, et al.: A genome-wide
488 association study confirms PNPLA3 and identifies TM6SF2 and MBOAT7 as risk loci
489 for alcohol-related cirrhosis. *Nat Genet* 2015, 47:1443-8. 10.1038/ng.3417.

- 490 [18] Abul-Husn NS, Cheng X, Li AH, Xin Y, Schurmann C, Stevis P, et al.: A Protein-
491 Truncating HSD17B13 Variant and Protection from Chronic Liver Disease. *N Engl J*
492 *Med* 2018, 378:1096-106. 10.1056/NEJMoa1712191.
- 493 [19] Tian C, Stokowski RP, Kershenovich D, Ballinger DG, Hinds DA: Variant in PNPLA3
494 is associated with alcoholic liver disease. *Nat Genet* 2010, 42:21-3. 10.1038/ng.488
495 ng.488 [pii].
- 496 [20] Schwantes-An TH, Darlay R, Mathurin P, Masson S, Liangpunsakul S, Mueller S, et al.:
497 Genome-wide association study and meta-analysis on alcohol-related liver cirrhosis
498 identifies novel genetic risk factors. *Hepatology* 2021, 73:1920-31. 10.1002/hep.31535.
- 499 [21] Whitfield JB, Rahman K, Haber PS, Day CP, Masson S, Daly AK, et al.: Brief report:
500 genetics of alcoholic cirrhosis-GenomALC multinational study. *Alcohol Clin Exp Res*
501 2015, 39:836-42. 10.1111/acer.12693.
- 502 [22] Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, et al.: Next-generation
503 genotype imputation service and methods. *Nat Genet* 2016, 48:1284-7. 10.1038/ng.3656.
- 504 [23] Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al.: UK biobank: an
505 open access resource for identifying the causes of a wide range of complex diseases of
506 middle and old age. *PLoS Med* 2015, 12:e1001779. 10.1371/journal.pmed.1001779.
- 507 [24] Emdin CA, Haas ME, Khera AV, Aragam K, Chaffin M, Klarin D, et al.: A missense
508 variant in Mitochondrial Amidoxime Reducing Component 1 gene and protection against
509 liver disease. *PLoS Genet* 2020, 16:e1008629. 10.1371/journal.pgen.1008629.
- 510 [25] Innes H, Buch S, Hutchinson S, Guha IN, Morling JR, Barnes E, et al.: Genome-Wide
511 Association Study for Alcohol-Related Cirrhosis Identifies Risk Loci in MARC1 and
512 HNRNPUL1. *Gastroenterology* 2020, 159:1276-89 e7. 10.1053/j.gastro.2020.06.014.
- 513 [26] Gellert-Kristensen H, Richardson TG, Davey Smith G, Nordestgaard BG, Tybjaerg-
514 Hansen A, Stender S: Combined Effect of PNPLA3, TM6SF2, and HSD17B13 Variants

- 515 on Risk of Cirrhosis and Hepatocellular Carcinoma in the General Population.
516 Hepatology 2020, 72:845-56. 10.1002/hep.31238.
- 517 [27] Salameh H, Raff E, Erwin A, Seth D, Nischalke HD, Falletti E, et al.: PNPLA3 Gene
518 Polymorphism Is Associated With Predisposition to and Severity of Alcoholic Liver
519 Disease. Am J Gastroenterol 2015, 110:846-56. 10.1038/ajg.2015.137.
- 520 [28] Yang J, Trepo E, Nahon P, Cao Q, Moreno C, Letouze E, et al.: A 17-Beta-
521 Hydroxysteroid Dehydrogenase 13 Variant Protects From Hepatocellular Carcinoma
522 Development in Alcoholic Liver Disease. Hepatology 2019, 70:231-40.
523 10.1002/hep.30623.
- 524 [29] Tang S, Zhang J, Mei TT, Guo HQ, Wei XH, Zhang WY, et al.: Association of TM6SF2
525 rs58542926 T/C gene polymorphism with hepatocellular carcinoma: a meta-analysis.
526 BMC Cancer 2019, 19:1128. 10.1186/s12885-019-6173-4.
- 527 [30] Trépo E, Nahon P, Bontempi G, Valenti L, Falletti E, Nischalke HD, et al.: Association
528 between the PNPLA3 (rs738409 C>G) variant and hepatocellular carcinoma:
529 Evidence from a meta-analysis of individual participant data. Hepatology 2014, 59 DOI:
530 10.1002/hep.26767.
- 531 [31] Stickel F, Lutz P, Buch S, Nischalke HD, Silva I, Rausch V, et al.: Genetic Variation in
532 HSD17B13 Reduces the Risk of Developing Cirrhosis and Hepatocellular Carcinoma in
533 Alcohol Misusers. Hepatology 2020, 72:88-102. 10.1002/hep.30996.
- 534 [32] Bianco C, Casirati E, Malvestiti F, Valenti L: Genetic predisposition similarities
535 between NASH and ASH: Identification of new therapeutic targets. JHEP Rep 2021,
536 3:100284. 10.1016/j.jhepr.2021.100284.
- 537 [33] Bianco C, Jamialahmadi O, Pelusi S, Baselli G, Dongiovanni P, Zanoni I, et al.: Non-
538 invasive stratification of hepatocellular carcinoma risk in non-alcoholic fatty liver using

- 539 polygenic risk scores. *Journal of Hepatology* 2021, 74:775-82.
540 10.1016/j.jhep.2020.11.024.
- 541 [34] De Vincentis A, Tavaglione F, Jamialahmadi O, Picardi A, Antonelli-Incalzi R, Valenti
542 L, et al.: A Polygenic Risk Score to Refine Risk Stratification and Prediction for Severe
543 Liver Disease by Clinical Fibrosis Scores. *Clin Gastroenterol Hepatol* 2021
544 10.1016/j.cgh.2021.05.056.
- 545 [35] Mars N, Koskela JT, Ripatti P, Kiiskinen TTJ, Havulinna AS, Lindbohm JV, et al.:
546 Polygenic and clinical risk scores and their impact on age at onset and prediction of
547 cardiometabolic diseases and common cancers. *Nat Med* 2020, 26:549-57.
548 10.1038/s41591-020-0800-0.
- 549

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* <i>PNPLA3</i> rs738409 G dosage) + (0.5423* <i>SUGPI-TM6SF2</i> rs10401969 C dosage) – (0.4463* <i>HSD17B13</i> 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* <i>PNPLA3</i> rs2294915 T dosage) + (0.3988* <i>SUGPI</i> rs10401969 C dosage) – (0.2485* <i>HSD17B13</i> 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* <i>SERPINA1</i> rs28929474 T dosage) – (0.2357* <i>FAF2</i> 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* <i>MBOAT7</i> rs641738 T dosage) - (0.2401* <i>MTARCI</i> rs2642438 A dosage) - (0.1304* <i>HNRNPUL1</i> 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 x 10 ⁻²⁸ | 5.99 (4.18 to 8.60) |
| | GenomALC-2 | 0.606 ± 0.014 | 0.669 ± 0.090 | 1.44 x 10 ⁻¹³ | 2.81 (2.03 to 3.89) |
| | UK Biobank | 0.619 ± 0.014 | 0.729 ± 0.080 | 1.06 x 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 x 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 x 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 x 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 x 10 ⁻¹⁹ | 3.37 (2.43 to 4.66) |

ⁱ Coefficients estimated from Buch et al¹⁷ and Abul-Husn et al¹⁸

ⁱⁱ Coefficients estimated from meta-analysis data Schwantes-An et al²⁰

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.

| Risk group | score | Odds Ratios (95% confidence intervals) | | |
|--------------|-------------|--|--|---|
| | | 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).

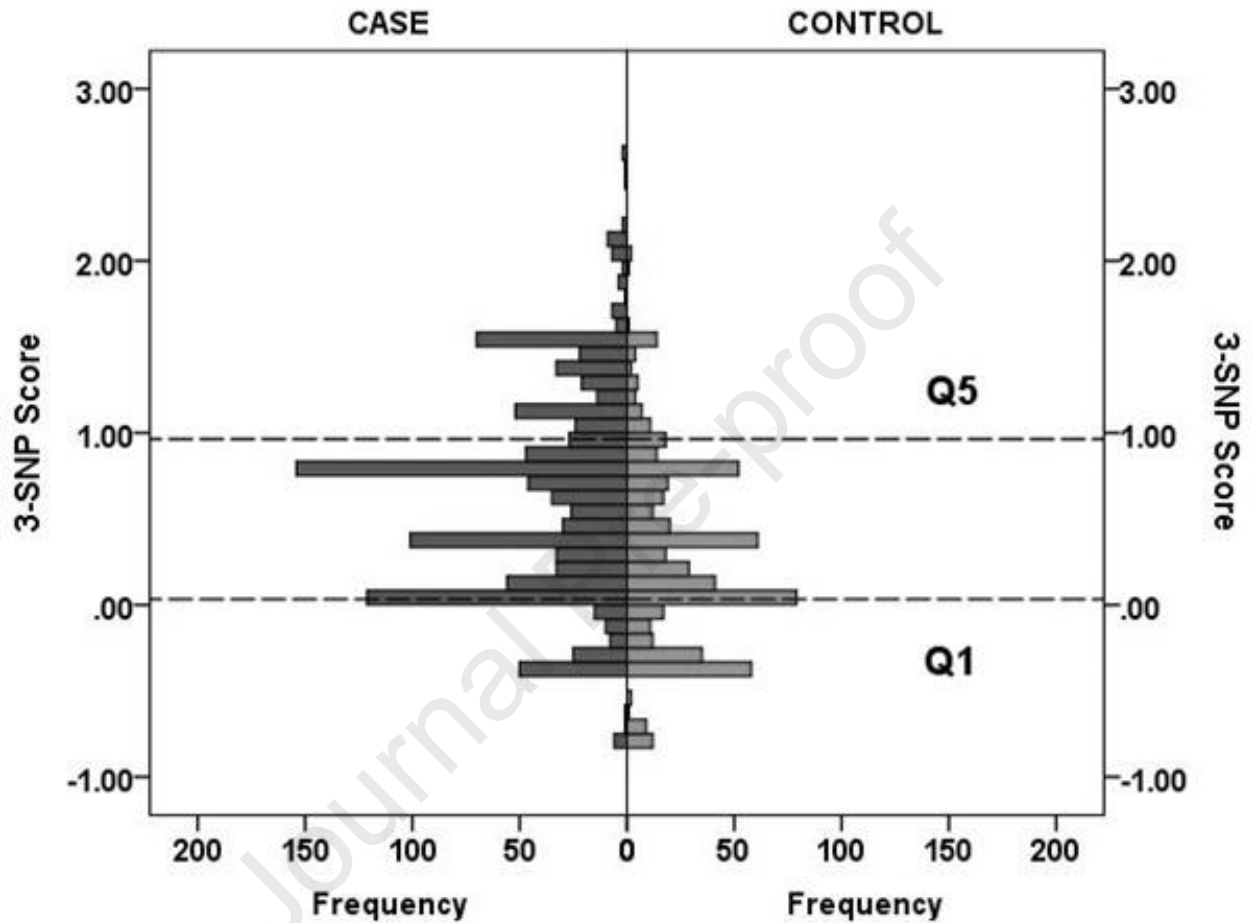
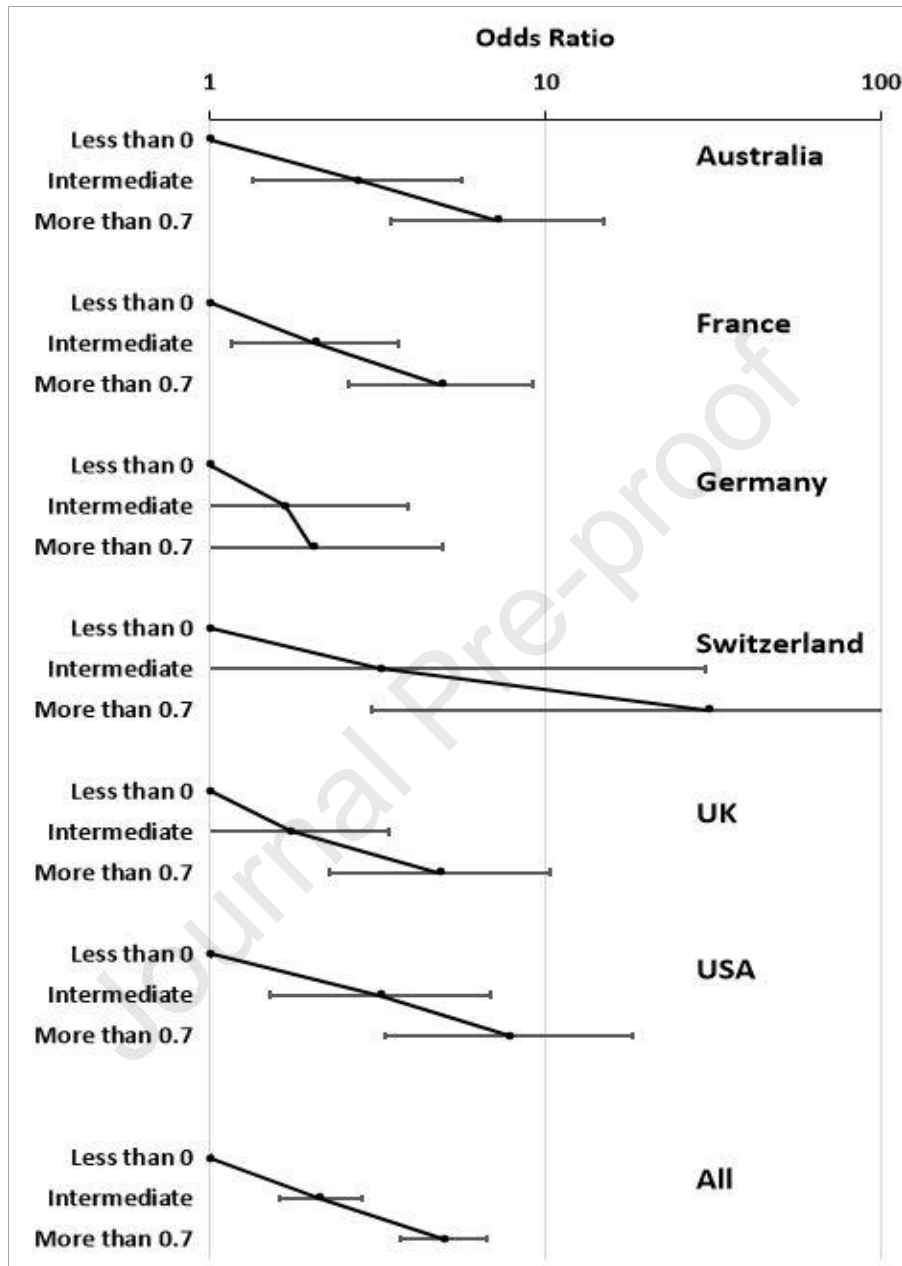


Figure 2. Odds ratios, by country and overall, for the risk of alcohol-related cirrhosis in the GenomALC-1 cohort when results for the 3-SNP score are divided into low (<0), intermediate (0 to 0.7) and high (>0.7) categories. Error bars show 95% confidence intervals.



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.