

## Common genetic variation in alcohol-related hepatocellular carcinoma: a case-control genome-wide association study

Eric Trépo<sup>1,2,3\*</sup> PhD, Stefano Caruso<sup>1\*</sup> PhD, Jie Yang<sup>1,4\*</sup> PhD, Sandrine Imbeaud<sup>1</sup> PhD, Gabrielle Couchy<sup>1</sup> PhD, Quentin Bayard<sup>1</sup> PhD, Eric Letouzé<sup>1</sup> PhD, Nathalie Ganne-Carrié<sup>1,5</sup> PhD, Christophe Moreno<sup>2,3</sup> PhD, Abderrahim Oussalah<sup>6,7,8</sup> PhD, Cyrille Feray<sup>9</sup> PhD, Jean Frédéric Blanc<sup>10,11</sup> PhD, Bruno Clément<sup>12</sup> PhD, Patrick Hillon<sup>13,14,15</sup> PhD, Jérôme Boursier<sup>16,17</sup> PhD, Valérie Paradis<sup>18</sup> PhD, Julien Calderaro<sup>19,20</sup> PhD, Viviane Gnemmi<sup>21</sup> PhD, Jean-Charles Nault<sup>1,5</sup> PhD, Jean-Louis Guéant<sup>6,7,8,22</sup> PhD, Jacques Devière<sup>2,3</sup> PhD, Isabelle Archambeaud<sup>23</sup> MD, Carole Vitellius<sup>16,17</sup> PhD, Bruno Turlin<sup>12</sup> PhD, Jean-Pierre Bronowicki<sup>8,22</sup> PhD, Thierry Gustot<sup>2,3,24</sup> PhD, Angela Sutton<sup>25,26</sup> PhD, the GENTHEP Consortium<sup>§</sup>, Marianne Ziol<sup>1,27</sup> PhD, Pierre Nahon<sup>1,5</sup> PhD, Jessica Zucman-Rossi PhD<sup>1,28†</sup>

Centre de Recherche des Cordeliers, Sorbonne Université, Université de Paris, INSERM, F-75006, Paris France. (E Trépo PhD, S Caruso PhD, J Yang PhD, S Imbeaud PhD, G Couchy PhD, Q Bayard PhD, E Letouzé PhD, Prof. N Ganne-Carrié PhD, Prof. J-C Nault PhD, Prof. M Ziol PhD, Prof. P Nahon PhD, Prof. J Zucman-Rossi PhD)

Department of Gastroenterology, Hepatopancreatology and Digestive Oncology, C.U.B. Hôpital Erasme, Université Libre de Bruxelles, Brussels, Belgium (E Trépo, Prof. C Moreno PhD, Prof. J Devière PhD, Prof. T Gustot PhD)

Laboratory of Experimental Gastroenterology, Université Libre de Bruxelles, Brussels, Belgium (E Trépo, C Moreno, J Devière, T Gustot)

Department of Radiation Oncology, Peking University Third Hospital, Beijing, China (J Yang)

Service d'hépatologie, Hôpital Avicenne, Hôpitaux Universitaires Paris-Seine-Saint-Denis, Assistance-Publique Hôpitaux de Paris, Bobigny, France (N Ganne-Carrié, JC Nault, M Ziol, P Nahon)

Department of Molecular Medicine, Division of Biochemistry, Molecular Biology, Nutrition, and Metabolism, University Hospital (CHRU) of Nancy, Nancy, France. (Prof. A Oussalah PhD, Prof. JL Guéant PhD)

Reference Centre for Inborn Errors of Metabolism (ORPHA67872), University Hospital (CHRU) of Nancy, Nancy, France. (A Oussalah, JL Guéant)

INSERM UMR\_S 1256, Nutrition, Genetics, and Environmental Risk Exposure (NGERE), Faculty of Medicine of Nancy, University of Lorraine, INSERM, Nancy, France. (A Oussalah, JL Guéant, Prof. JP Bronowicki PhD)

Centre Hépatobiliaire, Université Paris-Saclay, Paul Brousse Hospital, Assistance-Publique Hôpitaux de Paris, Villejuif, France. (Prof. C Feray PhD)

Service Hépatogastroentérologie et Oncologie Digestive, Hôpital Haut-Lévêque, CHU de Bordeaux, F-33000 Bordeaux, France (Prof. JF Blanc PhD)

Research in Translational Oncology, BaRITON, F-33076 Bordeaux, France (JF Blanc)

INSERM U1241, INRAe U1341, Institute of Nutrition, Metabolisms and Cancer, CRB-Santé, The French liver biobank network, Rennes University-Hospital, University of Rennes, Rennes, France (Prof. B Clément PhD, B Turlin PhD)

University of Bourgogne-Franche Comté, F-21000, Dijon, France. (Prof. P. Hillon PhD)

INSERM U1231, Lipids, Nutrition, Cancer, F-21000, Dijon, France. (P. Hillon)

Department of Hepatogastroenterology, University Hospital, F-21000, Dijon, France (P. Hillon)

Service d'Hépatogastroentérologie et Oncologie Digestive, Centre Hospitalier Universitaire d'Angers, Angers, France. (Prof. J Boursier PhD, C Vitellius PhD)

Laboratoire HIFIH, UPRES EA3859, SFR 4208, Université d'Angers, Angers, France. (J Boursier, C Vitellius)

Department of Pathology, Hôpital Beaujon, Assistance-Publique Hôpitaux de Paris, Clichy, France (Prof. V Paradis PhD)

Service d'Anatomopathologie, Hôpital Henri Mondor; Assistance-Publique Hôpitaux de Paris Université Paris Est, Créteil, France (Prof. J Calderaro PhD)

INSERM U955, Team 18, Institut Mondor de Recherche Biomédicale, Créteil, France (J Calderaro)

Univ. Lille, CNRS, Inserm, CHU Lille, Pathology Department, UMR9020-U1277 - CANTHER - Cancer Heterogeneity Plasticity and Resistance to Therapies, F-59000 Lille, France (V Gnemmi PhD)

Department of Hepato-Gastroenterology, Hôpital de Brabois, CHRU de Nancy, University of Lorraine, Nancy, France. (JL Guéant, JP Bronowicki)

Institut des Maladies de l'Appareil Digestif, Hôtel-Dieu, Nantes, France (I Archambeaud MD)  
Centre de Recherche sur l'inflammation (CRI), Inserm UMR S1149, Université de Paris, Paris, France. (T Gustot)

Department of Biochemistry, Hôpital Avicenne, Hôpitaux Universitaires Paris-Seine-Saint-Denis, Assistance-Publique Hôpitaux de Paris, Bobigny, France. (A Sutton PhD)

INSERM U1148 LVTS, UFR SMBH, Université Sorbonne Paris Nord, Bobigny, France. (A Sutton)  
Centre de Ressources Biologiques (BB-0033-00027) Hôpitaux Universitaires Paris-Seine-Saint-Denis, Assistance-Publique Hôpitaux de Paris, Bobigny, France. (M Ziol)

Hôpital Européen Georges Pompidou, Assistance Publique-Hôpitaux de Paris, Paris, France. (J Zucman-Rossi)

\*E.T., S.C. and J.Y. contributed equally to the study and the manuscript

§Authors of the GENTHEP Consortium are listed in the appendix p 3.

‡Corresponding author

Correspondence should be addressed to:

Jessica Zucman-Rossi MD, PhD: Centre de recherche des Cordeliers, Functional genomics of solid tumors, 15 rue de l'Ecole de Médecine, 75006 Paris, France, Tel.: +33 6 01 07 78 75. Email: [jessica.zucman-rossi@inserm.fr](mailto:jessica.zucman-rossi@inserm.fr)

**Word count:** 3414

**Figures:** 3

**Tables:** 2

## Summary

**Background** Hepatocellular carcinoma (HCC) is frequently the consequence of alcohol-related liver disease (ALD) with variable incidence among heavy drinkers. We performed a genome-wide association study (GWAS) to identify common genetic variants for alcohol-related HCC (ALD-HCC).

**Methods** We conducted a two-stage case-control GWAS in a discovery cohort of 2107 unrelated European ALD patients aged 20-92 years recruited between October 22, 1993 and March 12, 2017. Cases were ALD patients with HCC diagnosed by imaging or histology. Controls were ALD patients without HCC. We used an additive logistic regression model adjusted for the first ten principal components to assess genetic variants associated with ALD-HCC. We performed another analysis with adjustment for age, sex and liver fibrosis. New candidate associations ( $p < 1 \times 10^{-6}$ ) and variants previously associated with ALD-HCC were evaluated in a validation cohort of 1933 ALD patients aged 29-92 years recruited between July 21, 1995 and May 2, 2019. A meta-analysis of the two case-control cohorts was performed.

**Findings** The discovery cohort included 775 cases and 1332 controls. Of 7962325 variants assessed, we identified *WNT3A-WNT9A* (rs708113;  $p = 1.11 \times 10^{-8}$ ) and confirmed previously reported regions associated with ALD-HCC risk at *TM6SF2* (rs58542926;  $p = 6.02 \times 10^{-10}$ ), *PNPLA3* (rs738409;  $p = 9.29 \times 10^{-7}$ ) and *HSD17B13* (rs72613567;  $p = 2.49 \times 10^{-4}$ ). The validation cohort gathered 874 cases and 1059 controls and three variants were replicated: *WNT3A-WNT9A* (rs708113;  $p = 1.17 \times 10^{-3}$ ), *TM6SF2* (rs58542926;  $p = 4.06 \times 10^{-5}$ ) and *PNPLA3* (rs738409;  $p = 1.17 \times 10^{-4}$ ). All three variants reached GWAS significance in the meta-analysis: *WNT3A-WNT9A* (odds ratio [OR]=0.73; 95%CI=0.66-0.81;  $p = 3.93 \times 10^{-10}$ ); *TM6SF2* (OR=1.77; 95%CI=1.52-2.07;  $p = 3.84 \times 10^{-13}$ ), *PNPLA3* (OR=1.34; 95%CI=1.22-1.47;  $p = 7.30 \times 10^{-10}$ ). Adjustment for clinical covariates yielded similar results. *WNT3A-WNT9A* rs708113 was not associated with liver fibrosis, an additive effect with other at-risk alleles, on ALD-HCC risk, was observed.

**Interpretation** *WNT3A-WNT9A* is a susceptibility locus for ALD-HCC, enlightening the early role of the Wnt- $\beta$ -catenin pathway in ALD-HCC carcinogenesis.

**Funding** Ligue contre le Cancer, Bpifrance, Inserm.

## Research in context

### Evidence before this study

Only a minority of long-term alcohol misusers will develop alcohol-related hepatocellular carcinoma and a familial clustering has been reported. We searched all peer-reviewed journal publications in English to identify studies assessing the association of common genetic variations with alcohol-related hepatocellular carcinoma using PubMed and Google Scholar, published between Jan 1<sup>st</sup>, 1990, and March 1<sup>st</sup>, 2021. Search terms included “hepatocellular carcinoma”, “liver cancer”, “alcohol”, “candidate gene”, “single nucleotide polymorphism”, and “genome-wide association study”. Only candidate gene studies have been published to date and only three common variants in *PNPLA3*, *TM6SF2* and *HSD17B13* - also predisposing to alcohol-related liver disease and alcohol-related cirrhosis - have been previously associated with alcohol-related hepatocellular carcinoma. Functional exploration of these variants has strongly contributed to a better understanding of the pathobiology of fatty liver disease, however, their precise role in the development of alcohol-related hepatocellular carcinoma remains unknown. Nearly half of alcohol-related HCCs harbor mutations in *CTNNB1* and activation of the Wnt- $\beta$ -catenin signaling pathway, promoting an immune-excluded phenotype. No inherited genetic variations have currently been linked to these somatic alterations in alcohol-related HCC. Therefore, a genome-wide association study (GWAS) may allow a more complete evaluation of the genetic architecture of alcohol-related HCC and signaling pathways driving alcohol-related hepatocarcinogenesis.

### Added value of this study

To our knowledge, our study is the first GWAS of alcohol-related hepatocellular carcinoma. We identified a risk locus on chromosome 1 spanning the *WNT3A* and *WNT9A* genes both involved in the Wnt- $\beta$ -catenin pathway. The *WNT3A-WNT9A* rs708113[T] minor allele mitigated the risk of developing alcohol-related hepatocellular carcinoma in patients of European ancestry. We also confirmed the association with *PNPLA3* and *TM6SF2* at the genome-wide level.

The protective *WNT3A-WNT9A* rs708113[T] allele was also associated with STAT3 activation genes in non-tumor liver tissue from alcohol-related HCC. Finally, rs708113[T] was also associated with an increased prevalence of immune cells and a lower proportion of *CTNNB1* somatic mutations in patients with alcohol-related HCC.

### Implications of all the available evidence

The results of our study underline the importance of the *WNT3A-WNT9A* locus and more generally the Wnt- $\beta$ -catenin signaling pathway in the risk of alcohol-related hepatocellular carcinoma. Future research into the functional significance of this region will help to gain further insights into alcohol-related liver carcinogenesis and may guide the development of genetically-informed drug targets.

## Introduction

Hepatocellular carcinoma (HCC) is the most common primary liver malignancy and the third most common cause of cancer-related death.<sup>1</sup> Alcohol-related liver disease (ALD) contributes to nearly a third of HCC cases worldwide and is the main driver of liver carcinogenesis in North America and many Western countries.<sup>1</sup> Several clinical features have been identified in cohort studies as risk factors for alcohol-related HCC and include older age, male sex, and liver fibrosis with most cases developing at cirrhosis stage.<sup>1</sup> However, many long-term harmful drinkers with these environmental factors never develop HCC and both case-control and cancer databases have highlighted a significant familial clustering.<sup>2,3</sup>

A limited number of genome-wide association studies (GWAS) have specifically evaluated common variants associated with HCC occurrence, solely in the context of viral-related liver disease in individuals of Asian ancestry.<sup>4</sup> In parallel, candidate gene studies have only identified a few single nucleotide polymorphisms (SNPs) associated with alcohol-related HCC in individuals of European ancestry. The most robustly replicated are two missense variations rs738409(C>G) in *PNPLA3* and rs58542926(C>T) in *TM6SF2*. Carriage of rs738409[G] and rs58542926[T] increases the risk of alcohol-related HCC.<sup>5,6</sup> Of note, both variants were initially identified by independent GWAS in the settings of the presence of nonalcoholic fatty liver disease (NAFLD)<sup>7,8</sup> and alcohol-related cirrhosis.<sup>9</sup> Functional exploration of *PNPLA3* rs738409(C>G) and *TM6SF2* rs58542926(C>T) - both involved in lipid turnover - has markedly contributed to a better understanding of fatty liver disease pathobiology.<sup>10</sup> Nevertheless, their precise roles in the accumulation of liver fibrosis and, particularly, alcohol-related hepatocarcinogenesis - through potential direct oncogenic mechanisms - remain unknown.<sup>10</sup> More recently, an association between a variant, rs72613567(T>TA) in *HSD17B13*, and steatosis/cirrhosis in ALD and NAFLD was identified by another GWAS.<sup>11</sup> Follow-up candidate gene studies reported that rs72613567[TA] decreased the risk of alcohol-related HCC but the effect appears to be much milder than variations in *PNPLA3* and *TM6SF2*<sup>12</sup>.

Nearly half of alcohol-related HCCs harbor mutations in *CTNNB1* and activation of the Wnt- $\beta$ -catenin signaling pathway,<sup>13</sup> that correlates with low tumor immune infiltration.<sup>14</sup> No inherited genetic variations have yet been linked to these somatic alterations in alcohol-related HCC.

Here, we performed a GWAS in patients with ALD to better delineate the contribution of common genetic variations to the risk of alcohol-related HCC and signaling pathways driving alcohol-related hepatocarcinogenesis.

## Methods

### *Study design and participants*

In this two-stage case-control GWAS, in patients with ALD, all selected patients were initially referred to tertiary referral centers in France and Belgium for chronic liver disease and/or HCC management. The discovery cohort included 837 patients with HCC (cases) and 1472 without (controls) recruited between October 22, 1993 and March 12, 2017; the validation cohort

included 908 cases and 1077 controls recruited between July 21, 1995 and May 2, 2019 (appendix p7) To allow comparison with other chronic liver disease insults, we also included a total of 1,530 patients with chronic hepatitis C (765 cases and 765 controls), 451 with chronic hepatitis B (220 cases and 231 controls), 1001 with NAFLD (367 cases and 634 controls) between March 9, 1990 and January 16, 2019 (appendix pp 4,7).

All clinical databases were retrospectively reviewed by senior hepatologists. ALD was defined as a history of excessive alcohol consumption of at least 60 g/d for more than 10 years together with increased alanine aminotransferase and aspartate aminotransferase or suspected/confirmed alcoholic cirrhosis.<sup>15</sup> The majority of ALD patients were active drinkers at the moment of inclusion. The presence of concomitant viral hepatitis was excluded in all patients by testing for hepatitis B surface antigen and anti-hepatitis B core antibody and by screening for antibody to hepatitis C. Cirrhosis was diagnosed as a result of liver biopsy using the METAVIR scoring system<sup>16</sup>, and/or at least two of the following criteria in agreement with guidelines from the European Association for the Study of the Liver<sup>15</sup>: 1) unequivocal biological profile (presence of abnormal coagulation, serum albumin concentration, serum bilirubin concentration and platelet count), 2) cirrhosis-related clinical complications including ascites or hepatic encephalopathy 3) typical imaging findings suggestive of cirrhosis on ultrasound and/or computed tomography scan and/or magnetic resonance imaging (dysmorphic liver, ascites or splenomegaly) and 4) presence of esophageal varices on upper gastrointestinal endoscopy. In case of the absence of cirrhosis, liver fibrosis was staged using the METAVIR scoring system<sup>16</sup>, and/or liver stiffness measurement by transient elastography. All HCC cases were identified through multidisciplinary boards registries dedicated to the management of primary liver tumors in each center. Diagnosis of HCC was based on histological analysis or non-invasive criteria in patients with cirrhosis, according to international recommendations<sup>17</sup>. Written informed consent was obtained from all individuals, and the study was approved by the Institutional Review Board of the French Institute of Medical Research and Health (ref:IRB00003888) in France and the Ethics evaluation committee of CUB Hôpital Erasme, Université Libre de Bruxelles (ref:P2016/183, A2020/030) in Belgium.

### *Genotyping and Imputation*

We extracted DNA samples, either from frozen blood or nontumor liver samples, with the Maxwell tissue DNA purification kit (Promega). Genotyping in the discovery cohort was performed with the Global Screening Array, version 1.0 (Illumina), which contains 618,564 variants before quality control. Samples in the discovery cohort that had genetic sex discordance, genotype call rates <95% or outlying heterozygosity for autosomal chromosomes (i.e.  $\pm 3$  standard deviation away from the sample mean), had an estimated identity by descent value of >0.1875 were excluded. Ancestry was identified by principal component analysis, individuals from non-European ancestry were excluded due to their small number (appendix pp 4,17). The resultant 775 cases and 1332 controls were retained for analysis (appendix p 18). SNPs with a genotype call rate lower than 95%, a minor allele

frequency (MAF)<0.1%, different missing genotype rates in cases and controls ( $p<10^{-5}$ ), and failed to meet Hardy-Weinberg equilibrium  $p<10^{-6}$  in controls were filtered. Further details are provided in the appendix p 4.

To maximize genetic coverage, SNP imputation was carried out with the Sanger Imputation Service on genome build GRCh37 using the Haplotype Reference Consortium (HRC) panel,<sup>18</sup> augmented by the Phase 3 1000 Genomes Project panel<sup>19</sup> for variants not present in the HRC panel. Genotyping in the validation cohort was performed using coverage sequencing and analysis of quantitative allele ratios. (appendix p 5).

#### *Expression and somatic mutation analyses*

We assessed gene expression levels by quantitative reverse-transcription polymerase chain reaction in a total of 171 tumor and 94 non-tumor liver tissue samples from patients with alcohol-related HCC. We also evaluated 301 tumors from patients with non-alcohol-related HCC. Total RNA was extracted and reverse transcribed using the High-Capacity Transcription Kit (Life Technologies). Gene expression was assessed using Fluidigm technology and specific TaqMan predesigned assays. Data were normalized with the RNA ribosomal 18S as the calibrator and changes in mRNA expression levels were determined using a comparative CT method. G1-G6 transcriptomic classification was performed for tumor samples as previously described.<sup>20</sup>

Estimation of the relative infiltration by diverse immune cell types and endothelial cells was assessed through microenvironment cell population (MCP) scores computed with the MCP-counter Method, adapted to qPCR data after filtering out genes expressed in hepatocytes as previously described.<sup>21</sup>

Tumor samples from the 171 patients with alcohol-related HCC were screened for *CTNNB1* mutations by Sanger sequencing as previously described.<sup>22</sup>

#### *Statistical analysis*

Association with HCC was assessed using logistic regression, assuming additive allelic effects. In the discovery cohort, a total of 7,962,325 SNPs were tested, we used allele dosages rather than best-guess genotypes to account for imputation uncertainty with PLINK version 1.9 (<https://www.cog-genomics.org/plink2>). The genome-wide additive regression model was adjusted for the first ten principal components to account for potential population stratification (main analysis). We also tested another model including as covariates the first ten principal components, age, sex, and fibrosis stage. We used a threshold of  $p<1\times 10^{-6}$  in the discovery cohort for suggestive genome-wide association<sup>23</sup>, to select candidate SNPs for validation follow-up. More specifically, in each independent region (required to be separated from other loci by at least 500 kb of genomic distance) harboring such variants, the SNP with the strongest signal (i.e. lowest p value) was selected unless the causal SNP was already known. These SNPs were assessed for association with HCC before and after adjustment for age, sex, and fibrosis stage in the validation cohort.

Meta-analysis of SNPs present in both cohorts was conducted using a fixed-effects model (inverse variance method) with METAL (<https://genome.sph.umich.edu/wiki/METAL>)<sup>24</sup>, and the standard genome-wide p values threshold of  $5 \times 10^{-8}$  was used for statistical significance. Bayesian fine-mapping analysis was performed in regions harboring variants reaching genome-wide significance and where causal SNPs are unknown. The population attributable fraction was calculated against ALD control patients for SNPs reaching GWAS significance in the meta-analysis (appendix p 5). We also calculated a genotype score for SNPs reaching genome-wide significance in the meta-analysis. The score was calculated as an unweighted sum of the number of risk alleles carried by each patient in the in the discovery and validation cohorts. Comparison of gene expression levels was performed Kruskal-Wallis test. The association between molecular features and new identified variants was tested using chi-squared test for trend in proportions. All tests were two-tailed and  $p \leq 0.05$  was considered as significant.

#### *Role of the funding source*

The sponsors of the study had no role in study design, data collection, data curation, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

## **Results**

A total of 775 cases and 1,332 controls were analyzed in the discovery cohort; data regarding age, sex, and fibrosis stage are summarized in Table 1. The GWAS main analysis showed minimal genomic control (GC) inflation  $\lambda_{GC} = 1.017$  after adjustment for the top 10 principal components (appendix p 19). The resulting Manhattan plot is shown in Figure 1A. Two loci were associated with alcohol-related HCC with genome-wide significance ( $p < 5 \times 10^{-8}$ ). The strongest signal was at rs8107974, located in the *TM6SF2* region at chromosome 19p13.11 ( $p = 5.05 \times 10^{-10}$ ) which was in high linkage disequilibrium ( $r^2 = 0.99$ ) with rs58542926 ( $p = 6.02 \times 10^{-10}$ ). A total of 17 additional genome-wide significant SNPs mapped to this locus (appendix pp 8,10). A second independent association was observed at chromosome 1q42.13 for rs708113 in the *WNT3A-WNT9A* region ( $p = 1.11 \times 10^{-8}$ ) with two additional genome-wide significant SNPs in the same region (appendix pp 8,10). In addition, rs2294915 in the *PNPLA3* locus at chromosome 22q13.31 was also associated with alcohol-related HCC ( $p = 3.71 \times 10^{-7}$ ) and was in high linkage disequilibrium ( $r^2 = 1$ ) with rs738409 ( $p = 9.29 \times 10^{-7}$ ). The rs72613567(T>TA) variant in *HSD17B13* was associated with HCC without reaching GWAS significance ( $p = 2.49 \times 10^{-4}$ ). Restricting the analyses to patients with F3-F4 fibrosis stage, showed similar results (appendix pp 11,20).

A set of 20 SNPs covering the region of *WNT3A* and *WNT9A* at 1q42.13 were prioritized by Bayesian fine-mapping analysis to identify variants most likely to be causal (Figure 1B and

appendix p 5). Calculation of posterior inclusion probability confirmed rs708113 located in the region spanning *WNT3A* and *WNT9A* genes as the lead signal (appendix p 12). Examination of signal intensity plots of genotyped SNPs showed good discrimination for homozygous and heterozygous calls (appendix p 21).

We performed replication genotyping in a validation cohort, including 874 ALD-HCC cases and 1,059 ALD controls (Table 1). In independent genomic regions with variants reaching the selected replication threshold ( $p < 1 \times 10^{-6}$ ) in the discovery cohort, we tested the leading SNP (i.e. rs708113 in *WNT3A-WNT9A*) or the causal SNP, when it was already known (i.e. rs58542926 in *TM6SF2* and rs738409 in *PNPLA3*). We also included rs72613567 in *HSD17B13* a variant previously associated with a reduced risk of alcohol-related HCC<sup>12</sup> (Table 2). The three selected SNPs in *TM6SF2*, *WNT3A-WNT9A*, and *PNPLA3* were individually associated with alcohol-related HCC in the main analysis and after adjustment for clinical covariates (Table 2). In the meta-analysis, the three SNPs in *TM6SF2*, *WNT3A-WNT9A* and *PNPLA3* achieved conventional genome-wide significance in the main analysis (*TM6SF2*: rs58542926, odds ratio [OR]=1.77, 95% confidence interval [CI]=1.52-2.07,  $p=3.84 \times 10^{-13}$ ; *WNT3A-WNT9A*: rs708113, OR=0.73, 95% CI=0.66-0.81,  $p=3.93 \times 10^{-10}$ ; *PNPLA3*: rs738409, OR=1.34, 95% CI=1.22-1.47,  $p=7.30 \times 10^{-10}$ ; Table 2), and after adjustment for clinical covariates, (*TM6SF2*: rs58542926, OR=1.75, 95% CI=1.47-2.08,  $p=1.98 \times 10^{-10}$ ; *WNT3A-WNT9A*: rs708113, OR=0.72, 95% CI=0.64-0.80,  $p=2.03 \times 10^{-9}$ ; *PNPLA3*: rs738409, OR=1.53, 95% CI=1.37-1.71,  $p=3.07 \times 10^{-14}$ ; Table 2). The population-attributable fraction are shown in the appendix p 13).

We evaluated the association between the sum of risk alleles of *PNPLA3* rs738409[G], *TM6SF2* rs58542926[T], *WNT3A-WNT9A* rs708113[A], and *HSD17B13* rs72613567[T] carried by each patient and alcohol-related HCC in the discovery and validation cohorts. The prevalence of alcohol-related HCC increased gradually with the number of risk alleles carried by ALD patients (Figure 2A, appendix p 22). We observed an additive effect of risk alleles on the presence of alcohol-related HCC (Figure 2B). Additional adjustments for age, sex and fibrosis showed similar results (Figure 2C).

We tested the association between SNPs entering the GWAS validation stage and HCC in 1530 patients with chronic hepatitis C, 451 patients with chronic hepatitis B and 1001 patients with NAFLD (appendix p14). Overall, *WNT3A-WNT9A* rs708113 was not associated with HCC developed on the background of other chronic liver diseases while *TM6SF2* rs58542926, *PNPLA3* rs708409, and *HSD17B13* rs72613567 were significantly associated with NAFLD-related HCC (appendix p 15). Of note, *WNT3A-WNT9A* rs708113 was not associated with age, sex, and, especially fibrosis (appendix p 23); and subgroup analyses, stratified by clinical risk factors, showed that ORs were homogeneous (appendix p 24). Moreover, *WNT3A-WNT9A* rs708113 was not associated with fibrosis in another GWAS with publicly available summary statistics for alcohol-related cirrhosis (appendix p 16).<sup>9</sup>

We investigated *WNT3A* and *WNT9A* mRNA expression levels in 94 non-tumor tissue samples from alcohol-related HCC patients. The expression of *WNT3A* was not detected, while, *WNT3A-WNT9A* rs708113 was not associated with *WNT9A* expression (appendix p 25). Nevertheless, *WNT3A-WNT9A* rs708113[T] protective allele was associated with a significantly higher expression of inflammatory markers, including *Serum Amyloid A2* and *C-reactive protein* involved in the IL6/JAK/STAT pathway ( $p=0.03$ ,  $p=0.02$  respectively); this allele was also associated with a higher prevalence of NK cells ( $p=0.02$ ) (Figure 3A).

We assessed the effect of *WNT3A-WNT9A* rs708113 in 171 alcohol-related and 301 non-alcohol related HCC. *WNT9A* expression also correlated with the presence of immune and endothelial cells infiltration in tumor liver tissue from patients with ALD (Figure 3B). In line with previous findings<sup>13</sup>, we found that alcohol-related HCCs were enriched in *CTNNB1* gene mutations compared to non-alcohol-related tumors ( $p=0.005$ , Figure 3C). However, we observed a significant lower prevalence of *CTNNB1* somatic mutations in alcohol-related HCC patients carrying the *WNT3A-WNT9A* rs708113[T] protective allele ( $p=0.05$ , Figure 3D). This association was validated at protein level in a subset of alcohol-related HCC patients (appendix p 26). Accordingly, the proportion of patients in the G5-G6 transcriptomic groups of HCC was lower among *WNT3A-WNT9A* rs708113[T] carriers, while the presence of this allele was associated with low proliferative HCC, belonging to the G4 HCC transcriptomic group ( $p=0.02$ , Figure 3E).<sup>20</sup> We also observed a lower prevalence of *CTNNB1* somatic mutations in patients harboring the *PNPLA3* rs738409[G] allele, but not in patients with the *TM6SF2* rs58542926[T] allele. However, *PNPLA3* rs738409[G] was not associated with a higher proportion of G4 subgroup (appendix p 27).

Then, we assessed whether the association between the *WNT3A-WNT9A* rs708113 variant and the prevalence of *CTNNB1* mutations is modulated by alcohol consumption. Thus, we tested the interaction between *WNT3A-WNT9A* rs708113 genotypes and *CTNNB1* somatic mutation status in HCC patients with and without ALD. Regression analysis confirmed significant differences in the effect of *WNT3A-WNT9A* rs708113 genotype on the occurrence of *CTNNB1*-mutated HCC only in alcohol-related HCC patients ( $p$ -interaction rs708113xAlcohol intake=0.01, Figure 3F, appendix p 28).

## Discussion

This GWAS of alcohol-related HCC identified *WNT3A-WNT9A* as a susceptibility locus and confirmed the robust association with *PNPLA3*, *TM6SF2* and *HSD17B13* (appendix p 29). Recruitment from expert liver centers led to an unambiguous diagnosis of chronic liver disease and HCC diagnosis based on routine clinical practice and international recommendations. Previous candidate gene studies have established rs738409(C>G) in *PNPLA3* and rs58542926(C>T) in *TM6SF2* as reproducible risk loci for HCC on the background of chronic alcohol consumption.<sup>5,6</sup> This study highlights these associations on a genome-wide level. The functional consequences of these missense variations have already been thoroughly studied. Both favor liver fat accumulation and contribute to chronic liver damage in NAFLD and ALD, however, their exact role in liver carcinogenesis remain mostly unknown.<sup>10</sup>

The rs72613567(T>TA) variant in *HSD17B13* failed to reach genome-wide significance but showed evidence of an association with alcohol-related HCC in the range of a previous candidate gene study.<sup>12</sup>

Here, we identified another association between rs708113(A>T), located on chromosome 1q42.13 in the *WNT3A-WNT9A* locus, and alcohol-related HCC. The rs708113[T] allele mitigated the risk of tumor onset. The effect predominantly concern patients with ALD as this association was not replicated in patients with other chronic liver diseases. Interestingly, this variant was not associated with liver fibrosis in our study and in a previous GWAS specifically designed to capture genetic markers for alcohol-related cirrhosis<sup>9</sup>. Thus, conversely to *PNPLA3* and *TM6SF2* which have been associated with ALD progression, including steatosis and fibrosis accumulation, the *WNT3A-WNT9A* region may have a more direct impact on liver carcinogenesis.

*WNT3A* and *WNT9A* encode for ligands of Wnt- $\beta$ -catenin signaling pathway which plays a critical role in liver development, metabolism, regeneration and carcinogenesis<sup>25</sup>; with a direct role of *WNT3A* in promoting hepatocytes differentiation in chronic liver disease<sup>26</sup>, and fostering HCC growth.<sup>27</sup> Moreover, increasing evidence suggest that the Wnt/ $\beta$ -catenin pathway contributes to alcohol-induced hepatic fibrogenesis and hepatocarcinogenesis.<sup>28</sup> Consistent with these findings, we previously observed that alcohol-related HCC tumors were significantly enriched in *CTNNB1* somatic mutations, the gene coding for  $\beta$ -catenin, usually observed in nearly half of alcohol-related HCC patients.<sup>13</sup> In addition, we showed that products of alcohol metabolism could trigger specific mutational processes that result in activation of the WNT/ $\beta$ -catenin pathway.<sup>29</sup> Interestingly, *WNT3A-WNT9A* rs708113[T] allele was associated with tumors with a lower prevalence of *CTNNB1* somatic mutations and belongs to the G5-G6 subgroup, characterized by the Wnt- $\beta$ -catenin signaling pathway activation.<sup>20</sup> This suggests that the presence of this allele may prevent the formation of Wnt- $\beta$ -catenin addicted tumors.

It is well established that Wnt signaling plays a key role in the modulation of the immune response during inflammation. More specifically, in HCC, several studies reported that  $\beta$ -catenin activation, through *CTNNB1* mutations, correlates with T-cell exclusion, promoting immune escape and resistance to immunotherapy.<sup>14</sup> Here, we showed that the rs708113[T] protective allele was associated with an increased expression in genes involved in immune response in non-tumor liver - mainly the IL6/JAK/STAT signaling pathway - and closely associated with the abundance of immune cell infiltrations, suggesting a potential role in maintaining an immune surveillance program. Accordingly, patients with alcohol-related HCC carrying the protective allele were predominantly of the G4 transcriptomic class, which is known to share a gene expression profile similar to that of non-tumor liver and characterized by the overexpression of CRP, a target gene of JAK/STAT signaling, at the mRNA and protein level. We also observed that the *WNT3A-WNT9A* rs708113[T] variant was associated with a decreased prevalence of *CTNNB1*-mutated HCC specifically in ALD patients. Overall, our results suggest that gene-alcohol interactions may have a significant impact on the development of alcohol-related HCC and that the rs708113[T] allele may promote an

inflammatory environment which prevents the activation of  $\beta$ -catenin to an oncogenic state that could ultimately mitigate HCC development.

Alcohol-related HCC carries a dismal prognosis<sup>1</sup>, therefore, this gene may represent an attractive genetically-informed chemopreventive target. However, whether the rs708113[T] allele has a direct impact on the expression of *WNT9A/WNT3A* in tumor cells or act indirectly on the tumor microenvironment remains to be established. Therefore, further functional studies are needed to clarify the role of *WNT3A-WNT9A* in alcohol-related hepatocarcinogenesis.

Limitations of this study include information bias inherent to the use of retrospective samples and joined databases with limited phenotypic description of the metabolic syndrome in ALD patients, the exact amount of alcohol consumption and duration, and whether alcohol withdrawal could have an impact. In addition, our population was unbalanced towards patients with advanced fibrosis or cirrhosis and future analyses should further validate these results in ALD patients without significant liver fibrosis. Future studies should also ideally evaluate these associations, individually or included in polygenic risk scores, in prospective cohorts as it allows characterization of exposures and risk factors before disease onset, which have fewer potential sources of bias and confounding than case-control studies.<sup>30</sup> Furthermore, our analyses were performed in patients of European descent with ALD and we do not know whether our findings are generalizable to other populations.

In conclusion, *WNT3A-WNT9A*, *PNPLA3*, *TM6SF2*, and *HSD17B13* are important risk loci for alcohol-related HCC. Our data support a role for *WNT3A-WNT9A* in the promotion of alcohol-related HCC in association with immune microenvironment modulation yet independently from the severity of ALD and other clinical risk factors.

### **Data availability**

Summary statistics from the GWAS main analysis in the discovery cohort will be made available to download with publication from the GWAS catalog <https://www.ebi.ac.uk/gwas/>. Summary statistics for appendix p 16 were obtained from [http://gengastro.med.tu-dresden.de/suppl/alc\\_cirrhosis/](http://gengastro.med.tu-dresden.de/suppl/alc_cirrhosis/).

### **Contributors**

ET, MZ, PN and JZR designed the study. ET, PN, JD and JZR obtained funding. ET, NGC, CM, AO, CF, JFB, BC, PH, JB, VP, JC, VG, JCN, JD, JLG, IA, CV, BT, TG JPB, AS, MZ, PN recruited the patients and collected the data. ET, SC and JY accessed, curated and verified the data. JY and GC did the laboratory assays. ET, SC, JY, SI, GC, QB, EL, MZ, PN, JZR analyzed and interpret the data. All authors had access to the raw data reported in the study. ET, SC and JZR drafted the manuscript. All authors critically revised the manuscript and approved the final version to be published. JZR had final responsibility for the decision to submit for publication. Authors of the GENTHEP consortium and their contributions are listed separately in the appendix p 3.

## **Declaration of interests**

ET received research support from Gilead. NGC received honoraria from Roche, Gilead and Ipsen and support for travel and attending meetings from Gilead. CM received research support from Gilead; consulting fees from Gilead, Abbvie, Novartis, Surrozen, Julius Clinical; support for travel and attending meetings from Gilead and Abbvie; honoraria from Bayer and Astellas; and acted as a scientific advisor to Gilead and Novartis. JCN received research support from Bayer and Ipsen. JPB received consulting fees from Bayer, Roche and Ipsen; honoraria from Bayer, Ipsen and Roche; support for travel and attending meetings from Bayer, Ipsen and Roche; and acted as a scientific advisor to Roche. TG received research support from Gilead; honoraria from Abbvie; and acted as a scientific advisor to GoLiver Therapeutics and Promethera Biosciences. PN received research support from BMS, Eisai, and Roche; consulting fees from Astra-Zeneca, Bayer, BMS, Gilead, Eisai, Roche, Ipsen, and Exact Science; honoraria from Astra-Zeneca, Bayer, BMS, Gilead, Eisai, Roche, Ipsen, and Exact Science; support for travel and attending meetings from Ipsen, Roche; and acted as a scientific advisor to Astra-Zeneca, Bayer, BMS, Gilead, Eisai, Roche, Ipsen and Exact Science.

## **Acknowledgements**

This work was supported by the Ligue Nationale contre le Cancer (Equipe Labellisée), an institutional grant from Bpifrance under the HECAM (HEpatocellular Carcinoma Multi-technological) consortium project, INSERM with the « Cancer et Environnement » (plan Cancer), Institut National du Cancer (INCa), the French Association for the Study of the Liver (AFEF), Labex Oncolmmunology (investissement d'avenir), grant IREB, Coup d'Elan de la Fondation Bettencourt-Shueller, the SIRIC CARPEM, FRM prix Rosen, Ligue Contre le Cancer Comité de Paris (prix René et André Duquesne), Fondation Mérieux. Région Ile-de-France and Consortium HETCOLI (ITMO Cancer). The CiRCE study was supported by the French National Research Agency under the program "Investissements d'Avenir" [reference ANR-11-LABX-0021], Conseil Régional de Bourgogne, Fonds Européen de Développement Régional (FEDER), Canceropole Grand-Est, Fondation de France. The SEPT9\_CROSS study was supported by the French Ministry of Health and Solidarity, Direction générale de l'offre de soins (DGOS), French Eastern Interregional Group of Clinical Research and Innovation (GIRCI Est), University Hospital of Nancy/ CRB lorrain, ClinicalTrials.gov Identifier : NCT03311152".

The Fonds Erasme for Medical Research, The Fonds Gaston Ithier for Oncology Research, a Collective Research Initiatives consolidator from the Université Libre de Bruxelles (EXPLORE project) ET was supported by a Marie Skłodowska-Curie Individual Fellowships from the European Union, and is a Research Associate of the F.R.S.-FNRS in Belgium. SC was supported by SIRIC CARPEM, France.

We thank Lucie Del Pozo and Nathalie Barget at the Centre de Ressources Biologiques (BB-0033-00027) Hôpitaux Universitaires Paris-Seine-Saint-Denis, Assistance-Publique Hôpitaux de Paris, Bobigny, France, and Vjola Tafciu at the Laboratory of Experimental Gastroenterology, Université Libre de Bruxelles, for their help with the collection and processing of samples and DNA extraction. We thank the CiRCE and SEPT9\_CROSS studies for

providing DNA samples from patients with liver diseases and associated clinical data, and the Tumorotheque Alliance-Cancer Lille for providing Human's tumor tissue and associated data.

The authors acknowledge the contribution of Sandy Field, PhD, for English language editing of this manuscript. Dr. Field was compensated by CUB Hôpital Erasme, Université Libre de Bruxelles, Brussels, Belgium.

## References

- 1 Llovet JM, Kelley RK, Villanueva A, *et al.* Hepatocellular carcinoma. *Nat Rev Dis Primers* 2021; **7**: 6.
- 2 Yu MW, Chang HC, Liaw YF, *et al.* Familial risk of hepatocellular carcinoma among chronic hepatitis B carriers and their relatives. *Journal of the National Cancer Institute* 2000; **92**: 1159–64.
- 3 Hemminki K, Li X. Familial liver and gall bladder cancer: a nationwide epidemiological study from Sweden. *Gut* 2003; **52**: 592–6.
- 4 Caruso S, O'Brien DR, Cleary SP, Roberts LR, Zucman-Rossi J. Genetics of Hepatocellular Carcinoma: Approaches to Explore Molecular Diversity. *Hepatology* 2021; **73 Suppl 1**: 14–26.
- 5 Trépo E, Nahon P, Bontempi G, *et al.* Association between the PNPLA3 (rs738409 C> G) variant and hepatocellular carcinoma: evidence from a meta-analysis of individual participant data. *Hepatology* 2014; **59**: 2170–7.
- 6 Yang J, Trépo E, Nahon P, *et al.* PNPLA3 and TM6SF2 variants as risk factors of hepatocellular carcinoma across various etiologies and severity of underlying liver diseases. *International Journal of Cancer* 2019; **144**: 533–44.
- 7 Romeo S, Kozlitina J, Xing C, *et al.* Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet* 2008; **40**: 1461–5.
- 8 Kozlitina J, Smagris E, Stender S, *et al.* Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet* 2014; **46**: 352–6.
- 9 Buch S, Stickel F, Trépo 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**: 1443–8.
- 10 Trépo E, Valenti L. Update on NAFLD genetics: From new variants to the clinic. *J Hepatol* 2020; **72**: 1196–209.
- 11 Abul-Husn NS, Cheng X, Li AH, *et al.* A Protein-Truncating HSD17B13 Variant and Protection from Chronic Liver Disease. *The New England journal of medicine* 2018; **378**: 1096–106.
- 12 Yang J, Trépo E, Nahon P, *et al.* A 17-Beta-Hydroxysteroid Dehydrogenase 13 Variant Protects From Hepatocellular Carcinoma Development in Alcoholic Liver Disease. *Hepatology* 2019; **70**: 231–40.
- 13 Schulze K, Imbeaud S, Letouzé E, *et al.* Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet* 2015; **47**: 505–11.

- 14 Ruiz de Galarreta M, Bresnahan E, Molina-Sánchez P, *et al.*  $\beta$ -Catenin Activation Promotes Immune Escape and Resistance to Anti-PD-1 Therapy in Hepatocellular Carcinoma. *Cancer Discov* 2019; **9**: 1124–41.
- 15 Thursz M, Gual A, Lackner C, *et al.* EASL Clinical Practice Guidelines: Management of alcohol-related liver disease. *Journal of Hepatology* 2018; **69**: 154–81.
- 16 Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. *Hepatology* 1996; **24**: 289–93.
- 17 European Association for the Study of the Liver. Electronic address: easloffice@easloffice.eu, European Association for the Study of the Liver. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. *J Hepatol* 2018; **69**: 182–236.
- 18 the Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016; **48**: 1279–83.
- 19 The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015; **526**: 68–74.
- 20 Calderaro J, Couchy G, Imbeaud S, *et al.* Histological subtypes of hepatocellular carcinoma are related to gene mutations and molecular tumour classification. *J Hepatol* 2017; **67**: 727–38.
- 21 Hirsch TZ, Negulescu A, Gupta B, *et al.* BAP1 mutations define a homogeneous subgroup of hepatocellular carcinoma with fibrolamellar-like features and activated PKA. *J Hepatol* 2020; **72**: 924–36.
- 22 Rebouissou S, Franconi A, Calderaro J, *et al.* Genotype-phenotype correlation of *CTNNB1* mutations reveals different  $\beta$ -catenin activity associated with liver tumor progression: *Hepatology*, Vol. XX, No. X, 2016 Rebouissou, Franconi, et al. *Hepatology* 2016; **64**: 2047–61.
- 23 Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet* 2018; **19**: 491–504.
- 24 Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; **26**: 2190–1.
- 25 Parsons MJ, Tammela T, Dow LE. WNT as a Driver and Dependency in Cancer. *Cancer Discov* 2021; **11**: 2413–29.
- 26 Boulter L, Govaere O, Bird TG, *et al.* Macrophage-derived Wnt opposes Notch signaling to specify hepatic progenitor cell fate in chronic liver disease. *Nat Med* 2012; **18**: 572–9.
- 27 Lai J-P, Oseini AM, Moser CD, *et al.* The oncogenic effect of sulfatase 2 in human hepatocellular carcinoma is mediated in part by glypican 3-dependent Wnt activation. *Hepatology* 2010; **52**: 1680–9.

- 28 Mercer KE, Hennings L, Ronis MJJ. Alcohol consumption, Wnt/ $\beta$ -catenin signaling, and hepatocarcinogenesis. *Adv Exp Med Biol* 2015; **815**: 185–95.
- 29 Letouzé E, Shinde J, Renault V, *et al.* Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat Commun* 2017; **8**: 1315.
- 30 Manolio TA, Bailey-Wilson JE, Collins FS. Genes, environment and the value of prospective cohort studies. *Nat Rev Genet* 2006; **7**: 812–20.

## Figure legends

**Figure 1. Genome-wide association study for alcohol-related HCC in 755 cases and 1,332 controls from the discovery cohort.** Panel A shows a Manhattan plot summarizing the results of the association between single-nucleotide polymorphisms (SNPs) and alcohol-related HCC in 755 ALD-HCC cases and 1332 ALD controls from the discovery cohort. SNPs are plotted on the x axis according to their position on each chromosome against the significance of the association (shown as  $-\log_{10}$  p values) on the y axis after adjustment for the first ten principal components. The red solid line indicates the genome-wide significance threshold of  $p=5\times 10^{-8}$ ; SNPs reaching this threshold are colored in red. The blue dashed line indicates the chosen threshold of  $p=1\times 10^{-6}$  for validation follow-up. The nearest gene to the index SNP is indicated above each association peak. Panel B shows a regional linkage disequilibrium plot of the *WNT3A-WNT9A* region at 1q42.13. The  $-\log_{10}$  of p values (left y axis) drawn from the genome-wide association adjusted for the first ten principal components are plotted against single-nucleotide polymorphism (SNP) genomic positions using GRCh37 assembly. Imputed variants are indicated with circles, and genotyped variants are indicated with squares. The linkage disequilibrium values were calculated based on genotypes after imputation. SNPs are colored to reflect correlation with the most significant SNP (purple diamond). Estimated recombination rates shown in centimorgans (cM) per million base pairs (Mb) are plotted in blue (right y axis). Bayesian fine-mapping analysis prioritized 20 variants with greater than 95% certainty (95% Credible Set). The plot shows the names and locations of the genes; the transcribed strand is indicated with an arrow. Genes are represented with intronic and exonic regions.

**Figure 2. Association between the number of risk alleles in *PNPLA3*, *TM6SF2*, *WNT3A-WNT9A* and *HSD17B13* and alcohol-related HCC.** We assessed the association between the unweighted sum of risk alleles - including *PNPLA3* rs738409[G], *TM6SF2* rs58542926[T], *WNT3A-WNT9A* rs708113[A], and *HSD17B13* rs72613567[T] - carried by each patient in the discovery and validation cohorts and alcohol-related HCC. Panel A shows the proportion of patients with alcohol-related HCC plotted against the number of risk alleles. Panel B shows the risk of alcohol-related HCC by the number of risk allele categories. Data are crude odds ratios (OR)  $\pm$  95% confidence intervals (CI), the 0-2 group was the reference. In the discovery cohort, the OR for alcohol-related HCC was 1.84; 95% CI=1.38-2.47 in patients with 3-4 risk alleles, the corresponding OR in those with 5-8 risk alleles was 3.46; 95% CI =2.54-4.74,  $p=6.06\times 10^{-21}$ ). In the validation cohort, the OR for alcohol-related HCC was 1.36; 95% CI=0.98-1.89 in patients with 3-4 risk alleles, the corresponding OR in those with 5-8 risk alleles was 2.29; 95%CI=1.61-3.26,  $p=6.12\times 10^{-8}$ ); Panel C shows the risk of HCC by number of risk allele categories after adjustment for age, sex, and fibrosis stage. In the discovery cohort, the OR for alcohol-related HCC was 1.82; 95% CI=1.32-2.52 in patients with 3-4 risk alleles, the corresponding OR in those with 5-8 risk alleles was 3.44; 95% CI=2.44-4.59,  $p=9.57\times 10^{-18}$ ). In the validation cohort, the OR for alcohol-related HCC was 1.66; 95% CI = 1.14-2.46 in patients with 3-4 risk alleles, the corresponding OR in those with 5-8 risk alleles was 3.08; 95% CI=2.05-

4.67,  $p = 2.17 \times 10^{-10}$ ). Bottom, the table shows the numbers (n) of patients with each risk allele bin in cases and controls with alcohol-related HCC in both cohorts.

**Figure 3. *WNT3A-WNT9A* rs708113 genotype-phenotype analyses in liver tissue from patients with alcohol-related HCC.** Panel A shows the association between *WNT3A-WNT9A* rs708113 and inflammatory genes and NK cells in non-tumor liver tissue. The presence of the rs708113[T] allele was associated with a higher expression of *Serum Amyloid A2 (SAA2)*, *C-reactive protein (CRP)* involved in the involved in the IL6/JAK/STAT pathway ( $p=0.03$ ,  $p=0.02$ , respectively), and a higher prevalence of NK cells ( $p=0.02$ ). Panel B shows the association between *WNT3A-WNT9A* rs708113 and the relative infiltration by diverse immune and endothelial cells of the non-tumor counterpart of patients with alcohol-related HCC. The abundance of immune cells and endothelial cells was significantly higher in carriers of the rs708113[T] allele ( $p=0.02$  and  $p=0.008$  respectively). Panel C shows the enrichment in *CTNNB1* somatic mutations in alcohol-related HCC compared to HCC from other etiologies ( $p=0.005$ ). Panel D shows the association between *WNT3A-WNT9A* rs708113 genotype and the presence of *CTNNB1* somatic mutations in tumor liver tissue from patients with alcohol-related HCC. A significantly lower prevalence of *CTNNB1* somatic mutations was observed in patients harboring the rs708113[T] protective allele ( $p=0.05$ ). Panel E shows the proportion of each transcriptomic groups according to the *WNT3A-WNT9A* rs708113 genotype. The rs708113[T] protective allele was associated with a low prevalence of G5-G6 and a high prevalence of G4, respectively ( $p=0.02$ ). Panel F shows the effect of *WNT3A-WNT9A* rs708113 genotype and alcohol intake in the development of HCC carrying the *CTNNB1* gene mutations ( $p$ -interaction rs708113xAlcohol intake=0.01).

## Tables

Table 1. Study population characteristics

Characteristic	Discovery				Validation			
	n	Cases (n = 775)	Controls (n = 1332)	p	n	Cases (n = 874)	Controls (n = 1059)	p
<b>Age, years: Mean (s.d.)</b>	2107	65 (9)	56 (10)	<0.0001	1932	67 (8)	59 (10)	<0.0001
<b>Sex: n (%)</b>	2107			<0.0001	1933			<0.0001
F		77 (9.9)	370 (28)			70 (8.0)	281 (27)	
M		698 (90)	962 (72)			804 (92)	778 (73)	
<b>Fibrosis: n (%)</b>	2096			<0.0001	1915			<0.0001
F0-F2		154 (20)	80 (6.0)			145 (17)	39 (3.7)	
F3-F4		618 (80)	1244 (94)			717 (83)	1014 (96)	

s.d. standard deviation.

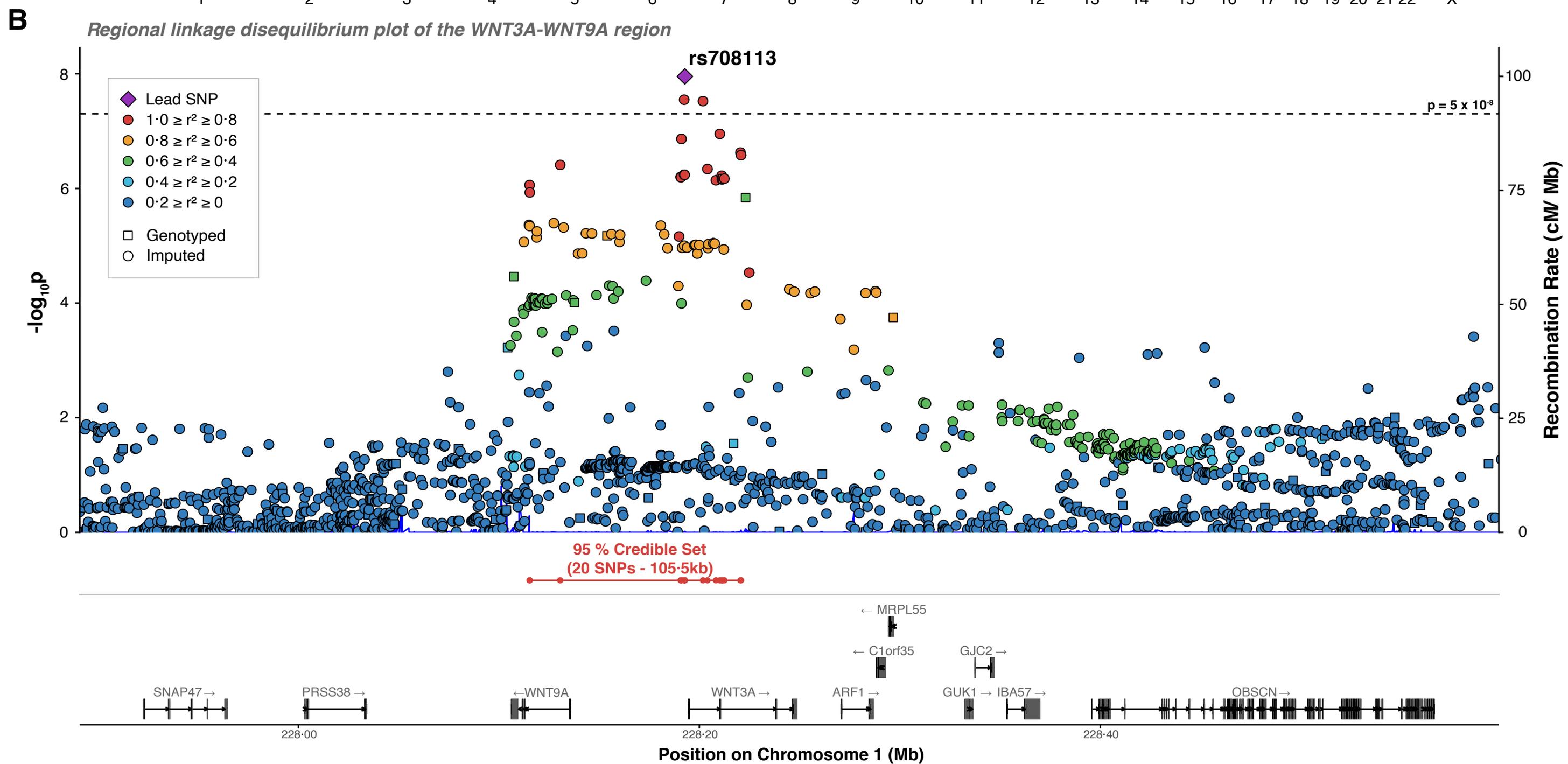
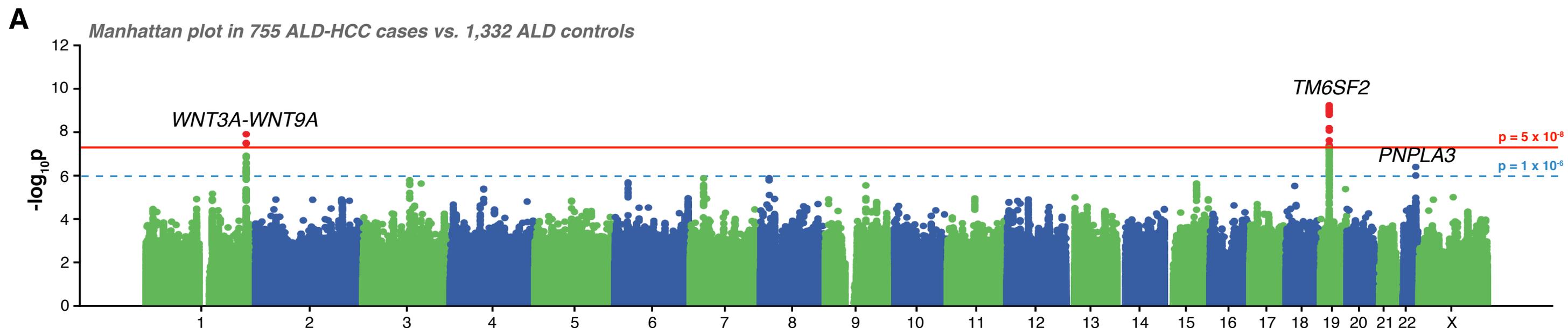
Statistical tests performed: Student's t-test for continuous outcomes and chi-square test of independence for categorical outcomes

**Table 2. Association results with alcohol-related HCC for the lead or known causal SNPs of the regions entering the validation stage of the GWAS**

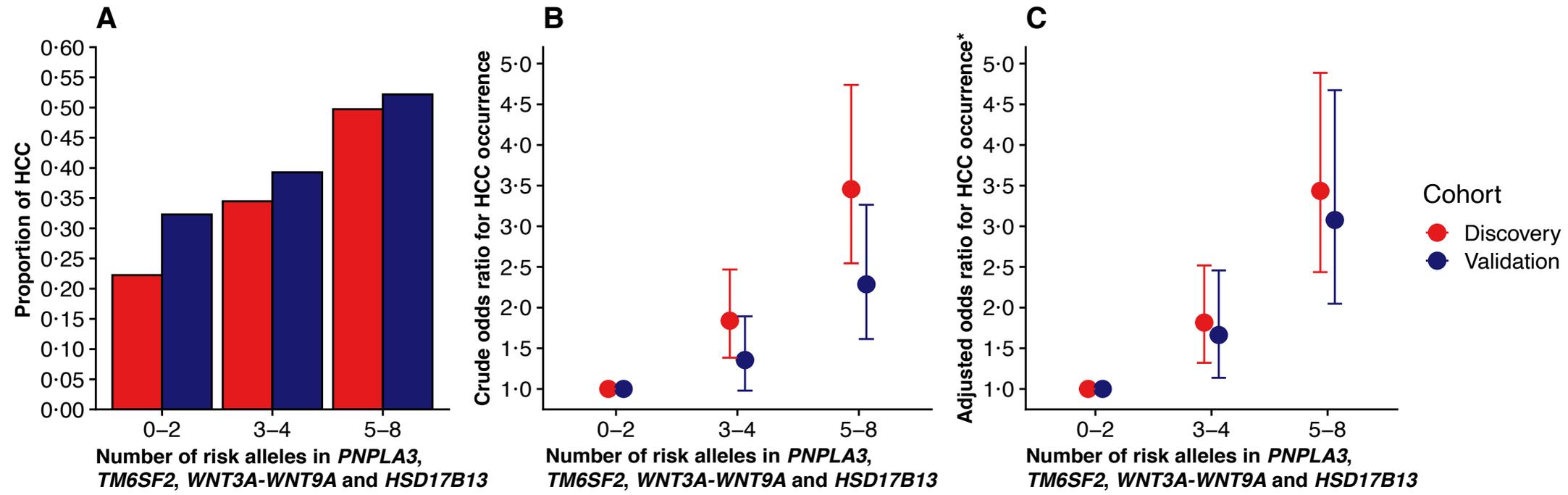
SNP	Chr.	Nearest gene	Minor allele	Analysis	Discovery			Validation			Meta-analysis	
					Allele frequency case/control	OR (95% CI)	p	Allele frequency case / control	OR (95% CI)	p	OR (95% CI)	p
rs58542926	19p13.11	TM6SF2	T	Main <sup>1</sup>	0.13/0.08	2.01 (1.61-2.50)	6.02×10 <sup>-10</sup>	0.14/0.09	1.57 (1.27-1.95)	4.06×10 <sup>-5</sup>	1.77 (1.52-2.07)	3.84×10 <sup>-13</sup>
				Adjusted for clinical covariates <sup>2</sup>	0.13/0.08	2.03 (1.59-2.58)	1.27×10 <sup>-8</sup>	0.14/0.09	1.51 (1.18-1.93)	9.78×10 <sup>-4</sup>	1.75 (1.47-2.08)	1.98×10 <sup>-10</sup>
rs708113	1q42.13	WNT3A-WNT9A	T	Main <sup>1</sup>	0.33/0.41	0.66 (0.57-0.76)	1.11×10 <sup>-8</sup>	0.33/0.38	0.80 (0.70-0.92)	1.17×10 <sup>-3</sup>	0.73 (0.66-0.81)	3.93×10 <sup>-10</sup>
				Adjusted for clinical covariates <sup>2</sup>	0.33/0.41	0.68 (0.58-0.80)	2.67×10 <sup>-6</sup>	0.32/0.38	0.75 (0.64-0.87)	1.39×10 <sup>-4</sup>	0.72 (0.64-0.80)	2.03×10 <sup>-9</sup>
rs738409	22q13.31	PNPLA3	G	Main <sup>1</sup>	0.43/0.33	1.41 (1.23-1.62)	9.29×10 <sup>-7</sup>	0.43/0.37	1.29 (1.13-1.46)	1.17×10 <sup>-4</sup>	1.34 (1.22-1.47)	7.30×10 <sup>-10</sup>
				Adjusted for clinical covariates <sup>2</sup>	0.43/0.33	1.57 (1.35-1.84)	9.45×10 <sup>-9</sup>	0.43/0.36	1.49 (1.29-1.73)	8.78×10 <sup>-8</sup>	1.53 (1.37-1.71)	3.07×10 <sup>-14</sup>
rs72613567	4q22.1	HSD17B13	TA	Main <sup>1</sup>	0.17/0.22	0.72 (0.61-0.86)	2.49×10 <sup>-4</sup>	0.20/0.22	0.87 (0.75-1.02)	0.08	0.80 (0.72-0.90)	1.77×10 <sup>-4</sup>
				Adjusted for clinical covariates <sup>2</sup>	0.17/0.22	0.74 (0.61-0.89)	1.85×10 <sup>-3</sup>	0.20/0.22	0.81 (0.69-0.97)	0.02	0.78 (0.69-0.89)	1.22×10 <sup>-4</sup>

Chr. Chromosome. <sup>1</sup>adjusted for the first ten principal components in the discovery cohort. <sup>2</sup>adjusted for the first ten principal components in the discovery cohort and, age, sex and liver fibrosis stage in both cohort

Figure 1. Genome-wide association study for alcohol-related HCC in 755 cases and 1,332 controls from the discovery cohort.



**Figure 2. Association between the number of risk alleles in *PNPLA3*, *TM6SF2*, *WNT3A-WNT9A* and *HSD17B13* and alcohol-related HCC**

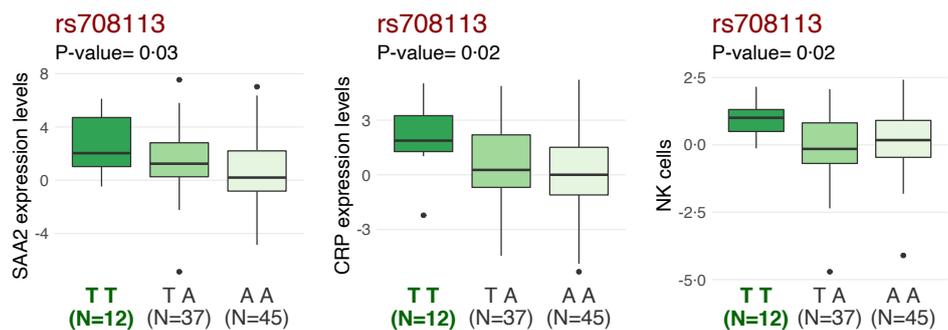


n risk allele	Discovery		Validation	
	n cases	n controls	n cases	n controls
0-2	71	248	63	132
3-4	419	796	348	538
5-8	285	288	239	219

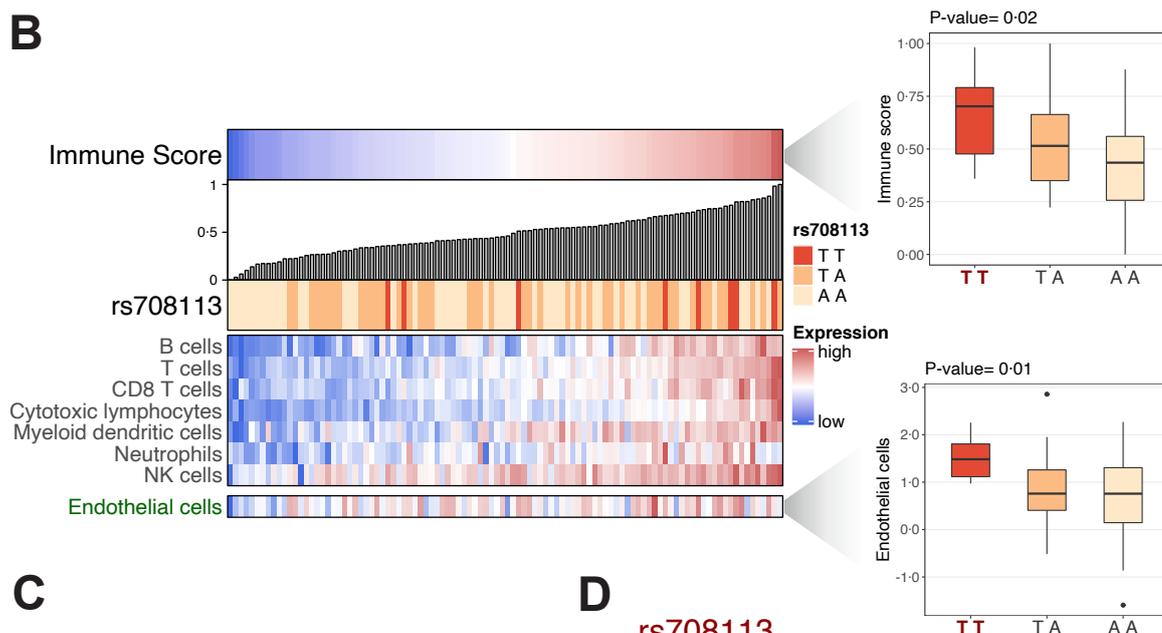
\*adjusted for age, sex and fibrosis stage

**Figure 3. WNT3A-WNT9A rs708113 genotype-phenotype analyses in liver tissue from patients with alcohol-related HCC**

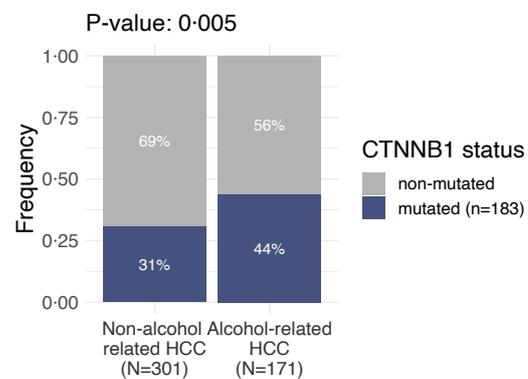
**A**



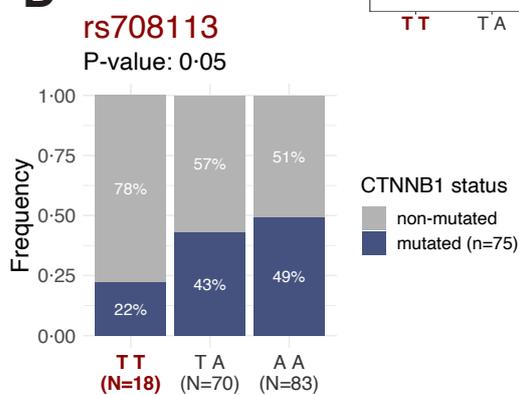
**B**



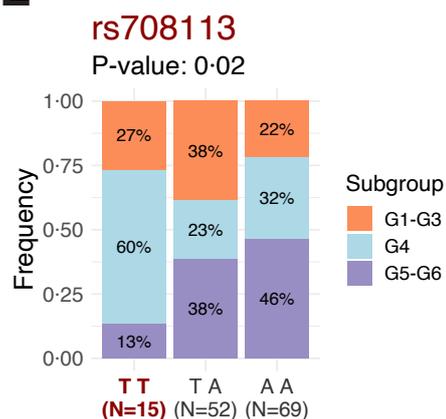
**C**



**D**



**E**



**F**

