THE LANCET
Oncology

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Trépo E, Caruso S, Yang J, et al. Common genetic variation in alcohol-related hepatocellular carcinoma: a case-control genome-wide association study. *Lancet Oncol* 2021; published online Dec 10. https://doi.org/10.1016/S1470-2045(21)00603-3.

Table of contents

Members of the GENTHEP consortium

Clément Meiller[1], Qian Cao[1], Théo Z. Hirsch[1], Sandra Rebouissou[1], Delphine Degré[2,3], Lukas Otero Sanchez[2,3], Nicolas Rosewick[3], Eric Quertinmont[3], Mireille Desille-Dugast[4], Muriel François-Vié[4], Cécile Moins[4], Emmanuelle Leteurtre[5], Guillaume Lassailly[6], Massih Ningarhari[1,6], Emmanuel Boleslawski[7], Vanessa Cottet[8, 9, 10]

[1]Centre de Recherche des Cordeliers, Sorbonne Université, Université de Paris, INSERM, F-75006, Paris France.
[2]Department of Gastroenterology, Hepatopancreatology and Digestive Oncology, C.U.B. Hôpital Erasme, Université Libre de Bruxelles, Brussels, Belgium
[3]Laboratory of Experimental Gastroenterology, Université Libre de Bruxelles, Brussels, Belgium
[4]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
[5]Univ. Lille, CNRS, Inserm, CHU Lille, Pathology Department, UMR9020-U1277 - CANTHER – Cancer Heterogeneity Plasticity and resistance to therapies, F-59000 Lille, France.
[6]CHU Lille, Hôpital Claude Huriez, Service des Maladies de l'appareil Digestif et Université de Lille, Lille, France
[7]Univ. Lille, INSERM U1189, CHU Lille, Service de Chirurgie Digestive et Transplantations, F-59000 Lille, France
[8]University of Bourgogne, Franche-Comté, Dijon, France
[9]Dijon University Hospital, Dijon, France
[10]INSERM U1231, EPICAD TEAM, CIC 1432, Dijon, France

Contributors

CLM, QC, TZH, SR, LOS, NR, EQ, MN verified and curated the data.

DD, MDD, MFV, CEM, EL, GL, MN, EB, VC recruited the patients and collected the data.

Declaration of interest

EB received honoraria from Ipsen, and support for attending meetings from LEO Pharma.

Supplementary Methods

Patient recruitment with other chronic liver diseases

Patients with other chronic liver diseases (i.e. without ALD) were recruited from tertiary referral centers in France and Belgium between March 9, 1990 and January 16, 2019 (appendix p 7). We enrolled 1) 1,530 patients with chronic hepatitis C, defined by persistent anti-hepatitis C virus (HCV) antibodies for at least 6 months, were HCV-RNA–positive, and had no other signs of any coexisting chronic liver disease shown by histology or imaging or coinfection with hepatitis B virus; 2) 451 patients with chronic hepatitis B, defined by the presence of hepatitis B surface antigen for at least 6 months and had no other signs of any coexisting chronic liver disease shown by histology or imaging; 3) 1001 with nonalcoholic fatty liver disease defined by the presence of steatosis in >5% of hepatocytes according to histological analysis or by magnetic resonance imaging excluding patients with chronic alcohol consumption (>20 g/day and >30 g/day for women and men respectively), and/or other coexisting liver disease (appendix p 14).

Quality Control procedures and genotyping

In the discovery cohort, genetically determined sex was compared to reported sex using PLINK (version 1.9)[1] and samples from individuals with discordant sex values were removed. Samples with genotype call rates less than 95% or outlying heterozygosity for autosomal chromosomes (i.e. ±3 standard deviation away from the sample mean) were also excluded. Relatedness between individuals was assessed among all genotyped samples using an independent linkage disequilibrium-pruned subset of SNPs (leaving no pairs with r^2>0.2, within a window of 50kb).[2] Pairwise percentage identity by descent (IBD) values were calculated using PLINK.[1] One individual (the one showing greater missingness) from each pair with an IBD value of >0.1875 was removed.[3] Principal component analysis (PCA) was performed for identifying large-scale differences in ancestry between individuals using Peddy, a machine learning model trained on individuals from the 1000 Genomes Project reference panel.[4] Due to their small number, individuals from non-European ancestry were excluded (appendix p 17). Following the aforementioned quality control procedures, the discovery cohort included 775 cases and 1,332 controls (appendix p 18).

Genotyping in the discovery cohort was performed with the Global Screening Array, version 1.0 (Illumina Inc., San Diego). A total of 642,824 SNPs were available before quality control. Initial genotype calling was performed with the Illumina GenomeStudio software version 2.0 (https://support.illumina.com/array/array_software/genomestudio/downloads.html). 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 Hardy-Weinberg equilibrium (HWE) $p<10^{-6}$ in controls were filtered and a total of 520,695 SNPs remained for further analysis.

Before performing genotyping in the validation cohort visual examination of intensity cluster plots of candidate SNPs genotyped in the discovery cohort was carried out.[5] Genotyping in the validation cohort was performed using the Midplex Genotyping workflow (Eurofins

Genomics, Konstanz, Germany). Primer pairs for each SNP were designed in the SNP surrounding regions and pooled together. The target regions were amplified in a multiplex approach via PCR, generating sequencing libraries for Illumina sequencing. The sequencing libraries (including negative and positive controls) contained index sequences for demultiplexing and were simultaneously sequenced on the NovaSeq platform (Illumina, San Diego, California). Demultiplexed fastq files were analyzed by Eurofins proprietary Genomic MidPlex analysis software which provides genotype calls for all SNPs in each sample based on quantitative allele ratios.

In addition to selected SNPs from the discovery cohort analysis we also genotyped a panel of 26 ancestry SNPs and performed a PCA to exclude individuals of non-European ancestry.

Genotype imputation

Imputation was carried out with the Sanger Imputation Service (https://imputation.sanger.ac.uk) using EAGLE2 (v2.0.5)[6] for phasing and the positional Burrows-Wheeler transform (PBWT)[7] algorithm on genome build GRCh37 using the Haplotype Reference Consortium (HRC) panel,[8] augmented by the Phase 3 1000 Genomes Project panel[9] for variants not present in HRC. To ensure high quality, the following imputed SNPs were removed 1) imputation $R^2 < 0.6$, 2) MAF < 1% and 3) HWE $P < 10^{-6}$ in controls. After quality control procedures a total of 7,962,325 SNPs were included in the analysis.

Bayesian fine-mapping

In HCC-associated regions of the discovery cohort, including candidate SNPs for validation and where causal SNPs are not already known, we performed a Bayesian fine-mapping analysis to determine a 95% credible set (i.e. the minimum set sets of variants gathering all causal SNPs with a probability ≥ 95%[10]) driving the signal. More specifically, we extracted local linkage disequilibrium from genotypes after imputation in the discovery cohort using PLINK[1] and, for each SNP, we calculated the posterior inclusion probability (the probability that a given SNP should be considered as potentially causative[10]) with FINEMAP (version 1.4),[11] using the option --n-causal-snps 1.

Population attributable fraction

The population attributable fraction (PAF) is used to approximate the effect of removing the genetic risk variant on the overall risk of disease.[12] The PAF was calculated against ALD control patients using the following formula[12]

$$PAF = \frac{2\mathrm{p}(1-\mathrm{p})(OR_{het}-1) + \mathrm{p}^2(OR_{hom}-1)}{1 + 2\mathrm{p}(1-\mathrm{p})(OR_{het}-1) + \mathrm{p}^2(OR_{hom}-1)}$$

where p is the minor allele frequency and OR_{het} and OR_{hom} are the associated odds ratio (OR) for heterozygotes and homozygotes carriers respectively. The OR was used as an approximation for the estimated relative risk

Supplementary Table 1. Patient recruitment at participating centers.

Center	Principal investigator	Patients recruited
Hôpital Avicenne[1,2]	Prof. Nathalie Ganne-Carrié Prof. Pierre Nahon,	2175
CUB Hôpital Erasme[3]	Prof. Jacques Devière	1330
Centre de Recherche des Cordeliers[4,5]	Prof. Jessica Zucman-Rossi	1160
CHRU de Nancy[6]	Prof. Jean-Louis Guéant Prof. Patrick Hillon Prof. Abderrahim Oussalah	1036
CHU Hôtel-Dieu[7]	Prof. Cyrille Feray	653
CHU de Bordeaux[8]	Prof. Jean-Fréderic Blanc	438
CHU d'Angers[9]	Prof. Jérôme Boursier	302
Rennes University-Hospital[4,10]	Prof. Bruno Clément	136
Hôpital Beaujon[11]	Prof. Valérie Paradis	40
Hôpital Henri Mondor[12]	Prof. Julien Calderaro	33
CHU de Lille[13]	Prof. Emmanuelle Leteurtre	30

[1]Hôpital Avicenne, Hôpitaux Universitaires Paris-Seine-Saint-Denis, Assistance-Publique Hôpitaux de Paris, Bobigny, France

[2]Centre de Ressources Biologiques (BB-0033-00027) Hôpitaux Universitaires Paris-Seine-Saint-Denis, Assistance-Publique Hôpitaux de Paris, Bobigny, France.

[3]C.U.B. Hôpital Erasme, Université Libre de Bruxelles, Brussels, Belgium

[4]The French Liver cancer biobanks network – INCa - BB-0033-00085

[5]Centre de Recherche des Cordeliers, Sorbonne Université, Université de Paris, INSERM, F-75006, Paris France.

[6]Hôpital de Brabois, CHRU de Nancy, University of Lorraine, Nancy, France (CiRCE study ref[13], and SEPT9_CROSS study, ClinicalTrials.gov Identifier : NCT03311152).

[7]CHU-Hôtel-Dieu, Nantes, France (CDCHC study, ref[14])

[8]Hôpital Haut-Lévêque, CHU de Bordeaux, F-33000 Bordeaux, France

[9]Centre Hospitalier Universitaire d'Angers, Angers, France

[10]Centre de Ressources Biologiques (CRB) Santé of Rennes BB-0033-00056

[11]Hôpital Beaujon, Assistance-Publique Hôpitaux de Paris, Clichy, France

[12]Hôpital Henri Mondor; Assistance-Publique Hôpitaux de Paris Université Paris Est, Créteil, France

[13]CHU Lille, Tumorothèque ALLIANCE-CANCER, Lille, France

Supplementary Table 2. Association results for SNPs that reached suggestive genome-wide significance ($p<1x10^{-6}$) for the main analysis in the discovery cohort. Variants selected for validation genotyping (rs58542926, rs708113 and rs738409) with a $p<1x10^{-6}$ are highlighted in bold. SNPs are ranked by p-value. Chr. chromosome

SNP ID	Minor allele	Chr.	Nearest gene	OR (CI 95%)	p	Consequence
rs8107974	T	19	*SUGP1*	2·01 (1·61-2·50)	$5·05x10^{-10}$	intron
rs58542926	**T**	**19**	***TM6SF2***	**2·01 (1·61-2·50)**	**$6·02x10^{-10}$**	**missense**
rs10401969	C	19	*SUGP1*	1·97 (1·59-2·45)	$7·25x10^{-10}$	intron
rs150641967	T	19	*HAPLN4*	2·02 (1·61-2·53)	$8·82x10^{-10}$	intron
rs739846	A	19	*SUGP1*	1·97 (1·59-2·45)	$9·84x10^{-10}$	intron
rs56255430	C	19	*GATAD2A*	1·94 (1·57-2·40)	$1·13x10^{-9}$	intron
rs200210321	AG	19	*SUGP1*	1·98 (1·59-2·47)	$1·41x10^{-9}$	intron
rs73001065	C	19	*MAU2*	1·99 (1·58-2·50)	$5·81x10^{-9}$	intron
rs73002956	G	19	*GATAD2A*	1·87 (1·51-2·31)	$7·50x10^{-9}$	intron
rs708113	**T**	**1**	***WNT3A***	**0·66 (0·57-0·76)**	**$1·11x10^{-8}$**	**upstream**
rs3794991	T	19	*GATAD2A*	1·83 (1·48-2·26)	$2·17x10^{-8}$	intron
rs697762	G	1	*WNT3A*	0·67 (0·58-0·77)	$2·84x10^{-8}$	upstream
rs708118	C	1	*WNT3A*	0·67 (0·58-0·77)	$3·01x10^{-8}$	intron
rs73004962	T	19	*PBX4*	1·82 (1·47-2·26)	$3·61x10^{-8}$	intron
rs73004959	T	19	*PBX4*	1·82 (1·47-2·25)	$3·72x10^{-8}$	intron
rs73004926	T	19	*PBX4*	1·82 (1·47-2·26)	$3·85x10^{-8}$	downstream
rs73004951	T	19	*PBX4*	1·82 (1·47-2·25)	$4·01x10^{-8}$	intron
rs73004933	T	19	*PBX4*	1·82 (1·47-2·25)	$4·02x10^{-8}$	intron
rs12608729	T	19	*PBX4*	1·81 (1·47-2·25)	$4·14x10^{-8}$	intron
rs138295924	G	19	*SUGP1*	2·24 (1·68-2·99)	$4·40x10^{-8}$	intron
rs150824230	A	19	*PBX4*	1·81 (1·47-2·25)	$4·59x10^{-8}$	downstream
rs58489806	T	19	*MAU2*	1·80 [1·45-2·22]	$5·13x10^{-8}$	intron
rs73004966	T	19	*PBX4*	1·81 [1·46-2·24]	$5·30x10^{-8}$	intron
rs140868651	A	19	*SUGP1*	1·94 [1·53-2·46]	$5·42x10^{-8}$	upstream
rs141756246	GT	19	*PBX4*	1·82 [1·46-2·25]	$5·51x10^{-8}$	intron
rs57504626	T	19	*PBX4*	1·81 [1·46-2·23]	$5·75x10^{-8}$	intron
rs16996185	G	19	*PBX4*	1·81 [1·46-2·23]	$5·75x10^{-8}$	intron
rs73004975	G	19	*PBX4*	1·80 [1·45-2·23]	$6·25x10^{-8}$	intron
rs10500212	T	19	*PBX4*	1·80 [1·45-2·23]	$6·92x10^{-8}$	intron
rs58847337	A	19	*PBX4*	1·80 [1·45-2·23]	$6·92x10^{-8}$	intron
rs12610185	A	19	*PBX4*	1·79 [1·45-2·22]	$7·84x10^{-8}$	intron
rs12610191	T	19	*PBX4*	1·79 [1·45-2·22]	$7·84x10^{-8}$	intron
rs2285626	T	19	*MAU2*	1·68 [1·39-2·04]	$9·94x10^{-8}$	3_prime_UTR

Supplementary Table 2. Association results for SNPs that reached suggestive genome-wide significance ($p<1x10^{-6}$) for the main analysis in the discovery cohort (continuation)

SNP ID	Minor allele	Chr.	Nearest gene	OR (CI 95%)	p	Consequence
rs16996148	T	19	*CILP2*	1·78 [1·44-2·19]	$1·04x10^{-7}$	downstream
rs17216588	T	19	*CILP2*	1·78 [1·44-2·21]	$1·08x10^{-7}$	intergenic
rs150268548	A	19	*GATAD2A*	1·89 [1·50-2·40]	$1·09x10^{-7}$	upstream
rs3094911	A	1	*WNT3A*	0·68 [0·59-0·78]	$1·12x10^{-7}$	intron
rs17216525	T	19	*CILP2*	1·78 [1·44-2·20]	$1·19x10^{-7}$	downstream
rs143988316	T	19	*CILP2*	1·77 [1·43-2·20]	$1·34x10^{-7}$	intergenic
rs10442629	T	1	*WNT3A*	0·68 [0·59-0·78]	$1·37x10^{-7}$	upstream
rs72999033	T	19	*NCAN*	1·94 [1·51-2·49]	$1·71x10^{-7}$	downstream
rs11672355	C	19	*MAU2*	1·69 [1·38-2·05]	$2·06x10^{-7}$	intron
rs708123	A	1	*WNT3A*	0·69 [0·59-0·79]	$2·37x10^{-7}$	intron
rs11668104	A	19	*SUGP1*	1·68 [1·38-2·04]	$2·42x10^{-7}$	intron
rs708124	T	1	*WNT3A*	0·69 [0·59-0·79]	$2·63x10^{-7}$	intron
rs57962361	T	19	*SUGP1*	1·68 [1·38-2·04]	$2·77x10^{-7}$	intron
rs111234557	G	19	*SUGP1*	1·68 [1·38-2·04]	$2·88x10^{-7}$	upstream
rs11411903	TA	19	*MAU2*	1·67 [1·37-2·03]	$3·42x10^{-7}$	intron
rs73004967	G	19	*PBX4*	1·86 [1·47-2·36]	$3·49x10^{-7}$	intron
rs17217098	A	19	*PBX4*	1·86 [1·46-2·36]	$3·53x10^{-7}$	intron
rs2294915	T	22	*PNPLA3*	1·43 [1·25-1·64]	$3·71x10^{-7}$	intron
rs638877	C	1	*WNT9A*	0·70 [0·61-0·80]	$3·89x10^{-7}$	intron
rs1636196	A	1	*WNT3A*	0·69 [0·60-0·80]	$4·59x10^{-7}$	intron
rs57009615	G	19	*GATAD2A*	1·61 [1·34-1·94]	$4·81x10^{-7}$	intron
rs12052117	T	19	*GATAD2A*	1·64 [1·35-1·98]	$5·31x10^{-7}$	intergenic
rs8182472	C	19	*GATAD2A*	1·63 [1·35-1·97]	$5·68x10^{-7}$	intron
rs697761	G	1	*WNT3A*	0·69 [0·60-0·80]	$5·77x10^{-7}$	upstream
rs697763	G	1	*WNT3A*	0·69 [0·60-0·80]	$5·81x10^{-7}$	upstream
rs1774757	C	1	*WNT3A*	0·69 [0·60-0·80]	$6·01x10^{-7}$	intron
rs34324111	G	19	*GATAD2A*	1·63 [1·34-1·98]	$6·31x10^{-7}$	intron
rs35629458	G	19	*GATAD2A*	1·63 [1·34-1·98]	$6·31x10^{-7}$	intron
rs113460678	G	19	*GATAD2A*	1·63 [1·34-1·98]	$6·31x10^{-7}$	intron
rs708109	C	1	*WNT3A*	0·69 [0·60-0·80]	$6·32x10^{-7}$	upstream
rs708110	T	1	*WNT3A*	0·69 [0·60-0·80]	$6·40x10^{-7}$	upstream
rs10408875	C	19	*GATAD2A*	1·63 [1·34-1·97]	$6·47x10^{-7}$	intron
rs56241616	T	19	*GATAD2A*	1·63 [1·34-1·97]	$6·48x10^{-7}$	intron
rs10415849	T	19	*GATAD2A*	1·63 [1·34-1·97]	$6·53x10^{-7}$	intron
rs1745413	A	1	*WNT3A*	0·69 [0·60-0·80]	$6·75x10^{-7}$	intron
rs1636195	T	1	*WNT3A*	0·69 [0·60-0·80]	$6·76x10^{-7}$	intron
rs3094910	G	1	*WNT3A*	0·69 [0·60-0·80]	$6·96x10^{-7}$	intron
rs10408596	T	19	*GATAD2A*	1·62 [1·34-1·96]	$7·11x10^{-7}$	intron

Supplementary Table 2. Association results for SNPs that reached suggestive genome-wide significance ($p<1x10^{-6}$) for the main analysis in the discovery cohort (continuation)

SNP ID	Minor allele	Chr.	Nearest gene	OR (CI 95%)	p	Consequence
rs708121	A	1	*WNT3A*	0·69 [0·60-0·80]	$7·20x10^{-7}$	intron
rs73002960	T	19	*GATAD2A*	1·62 [1·34-1·96]	$7·28x10^{-7}$	intron
rs28720066	T	19	*GATAD2A*	1·62 [1·34-1·96]	$8·03x10^{-7}$	intron
rs59148799	G	19	*GATAD2A*	1·62 [1·34-1·96]	$8·11x10^{-7}$	intergenic
rs1774754	G	1	*WNT9A*	0·70 [0·61-0·81]	$8·74x10^{-7}$	intron
rs738409	**G**	**22**	***PNPLA3***	**1·41 [1·23-1·62]**	**$9·29x10^{-7}$**	**missense**

Supplementary Table 3. Association results between SNPs reaching genome-wide significance or previously associated with alcohol-related HCC in patients with F3-F4 fibrosis stage in the discovery cohort.

SNP	Chr.	Nearest gene	Minor allele	Analysis	Allele frequency case/control	OR (95% CI)	p
rs58542926	19p13.11	*TM6SF2*	T	Main[1]	0·14/0·08	2·07 (1·64-2·61)	$9·48\times10^{-10}$
				Adjusted for clinical covariates[2]	0·14/0·08	2·04 (1·59-2·63)	$2·28\times10^{-8}$
rs708113	1q42.13	*WNT3A-WNT9A*	T	Main[1]	0·33/0·40	0·67 (0·58-0·79)	$4·89\times10^{-7}$
				Adjusted for clinical covariates[2]	0·33/0·40	0·68 (0·58-0·81)	$7·84\times10^{-6}$
rs738409	22q13.31	*PNPLA3*	G	Main[1]	0·46/0·35	1·45 (1·25-1·68)	$6·49\times10^{-7}$
				Adjusted for clinical covariates[2]	0·46/0·35	1·56 (1·33-1·84)	$5·00\times10^{-8}$
rs72613567	*4q22.1*	*HSD17B13*	*TA*	Main[1]	0·17/0·22	0·75 (0·62-0·90)	$2·28\times10^{-3}$
				Adjusted for clinical covariates[2]	0·17/0·22	0·74 (0·61-0·91)	$3·33\times10^{-3}$

Chr. Chromosome. [1]adjusted for the first ten principal components in the discovery cohort. [2]adjusted for the first ten principal components in the discovery cohort and age and sex in both cohorts.

Supplementary Table 4. Variants in 95% fine-mapped credible set at *WNT3A*-*WNT9A* locus.

SNP ID	Posterior inclusion probability	consequence
rs708113	0·379	upstream
rs697762	0·162	upstream
rs708118	0·153	intron
rs3094911	0·047	intron
rs10442629	0·039	upstream
rs708123	0·024	intron
rs708124	0·022	intron
rs638877	0·015	intron
rs1636196	0·013	intron
rs697761	0·011	upstream
rs697763	0·011	upstream
rs1774757	0·010	intron
rs708109	0·010	upstream
rs708110	0·010	upstream
rs1745413	0·009	intron
rs1636195	0·009	intron
rs3094910	0·009	intron
rs708121	0·009	intron
rs1774754	0·007	intron
rs1636193	0·006	intron

Supplementary Table 5. Population attributable fraction for the variants reaching genome-wide significance in the meta-analysis.

SNP	Nearest gene	Analysis	Discovery	Validation
rs58542926	*TM6SF2*	Main[1]	0·224	0·152
		Adjusted for clinical covariates[2]	0·250	0·133
rs708113	*WNT3A-WNT9A*	Main[1]	-0·224	-0·162
		Adjusted for clinical covariates[2]	-0·171	-0·196
rs738409	*PNPLA3*	Main[1]	0·297	0·187
		Adjusted for clinical covariates[2]	0·358	0·293

[1]adjusted for the first ten principal components in the discovery cohort.
[2]adjusted for the first ten principal components in the discovery cohort and, age, sex and liver fibrosis stage in both cohorts.

Supplementary Table 6. Characteristic of patients with other chronic liver diseases.

Characteristic	Chronic hepatitis C n = 1530	Chronic hepatitis B n = 451	NAFLD n = 1001
HCC			
Cases	765 (50%)	220 (49%)	367 (37%)
Controls	765 (50%)	231 (51%)	634 (63%)
Age, years: Mean (s.d.)	60 (13)	55 (15)	63 (12)
Gender: n (%)			
F	485 (32%)	83 (18%)	307 (31%)
M	1045 (68%)	368 (82%)	694 (69%)
Fibrosis: n (%)			
F0-F2	406 (27%)	192 (44%)	593 (61%)
F3-F4	1088 (73%)	245 (56%)	383 (39%)

NAFLD: nonalcoholic fatty liver disease, s.d. standard deviation

Supplementary Table 7. Association results between *PNPLA3* rs738409, *TM6SF2* rs58542926, *WNT3A*-*WNT9A* rs708113, *HSD17B13* rs72613567 and HCC in patients with other chronic liver diseases.

Model	Chronic hepatitis C		Chronic hepatitis B		NAFLD	
	Odds ratio (95% CI)	p	Odds ratio (95% CI)	p	Odds ratio (95% CI)	p
***PNPLA3* rs738409**						
unadjusted	1·03 (0·88-1·21)	0·72	0·91 (0·69-1·21)	0·53	1·28 (1·07-1·54)	0.007
Adjusted[1]	1·07 (0·90-1·27)	0·45	0·78 (0·57-1·08)	0·14	1·29 (1·04-1·60)	0.02
***TM6SF2* rs58542926**						
unadjusted	1·08 (0·82-1·43)	0·57	0·93 (0·54-1·58)	0·79	1·51 (1·13-2·00)	0.005
Adjusted[1]	1·13 (0·83-1·53)	0·43	0·94 (0·52-1·70)	0·84	1·32 (0·94-1·86)	0·11
***WNT3A*-*WNT9A* rs708113**						
unadjusted	0·96 (0·82-1·11)	0·57	0·90 (0·68-1·18)	0·44	0·86 (0·71-1·05)	0·13
Adjusted[1]	0·95 (0·81-1·12)	0·58	0·91 (0·67-1·24)	0·54	0·85 (0·67-1·07)	0·16
***HSD17B13* rs72613567**						
unadjusted	1·05 (0·89-1·25)	0·55	1·09 (0·78-1·51)	0·62	0·80 (0·63-1·00)	0·05
Adjusted[1]	1·01 (0·84-1·21)	0·95	0·98 (0·67-1·42)	0·90	0·81 (0·62-1·07)	0·14

NAFLD, nonalcoholic fatty liver disease. [1]Adjusted for age, sex and fibrosis stage

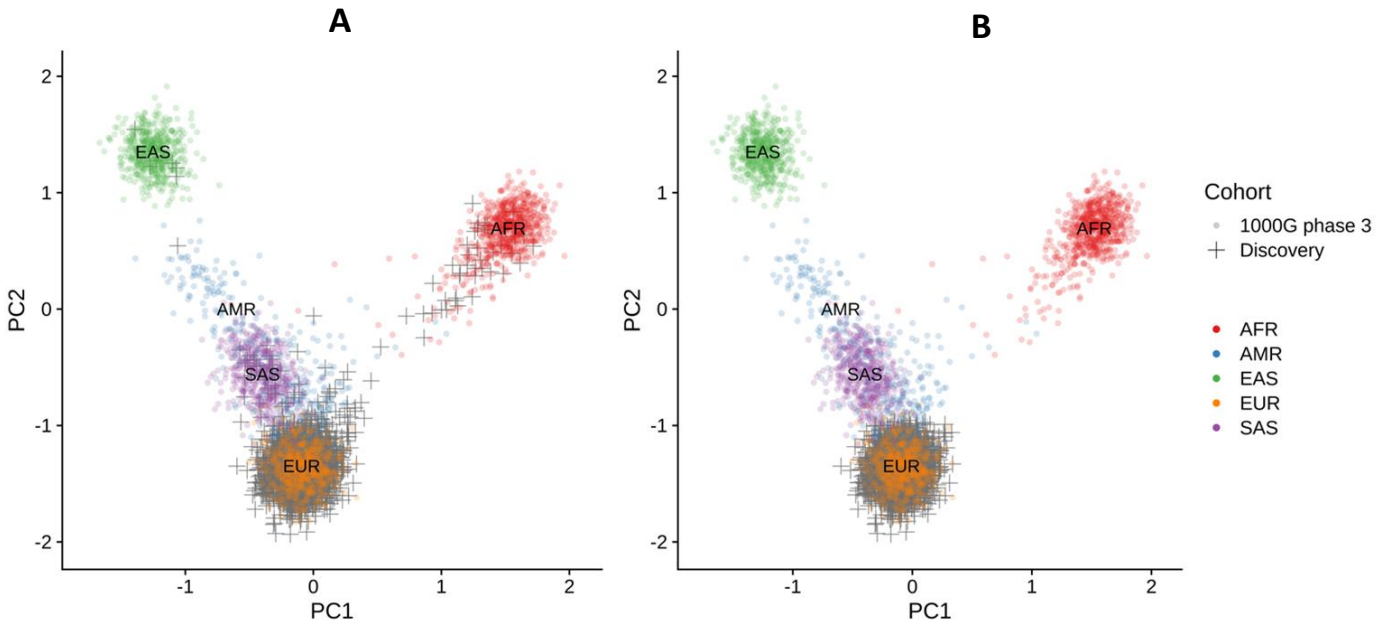
Supplementary Table 8. Association between WNT3A-WNT9A rs708113 and alcohol-related cirrhosis. The population of this published GWAS included 712 cases with alcohol-related cirrhosis and 1,466 controls who were heavy drinkers without evidence of liver damage[15]. Summary statistics were obtained from http://gengastro.med.tu-dresden.de/suppl/alc_cirrhosis/

SNP	Chr.	Nearest gene	Minor allele	MAF	OR (95% CI)	p
rs708113	1q42.13	*WNT3A-WNT9A*	T	0·37	0·91 (0·79-1·05)	0·22

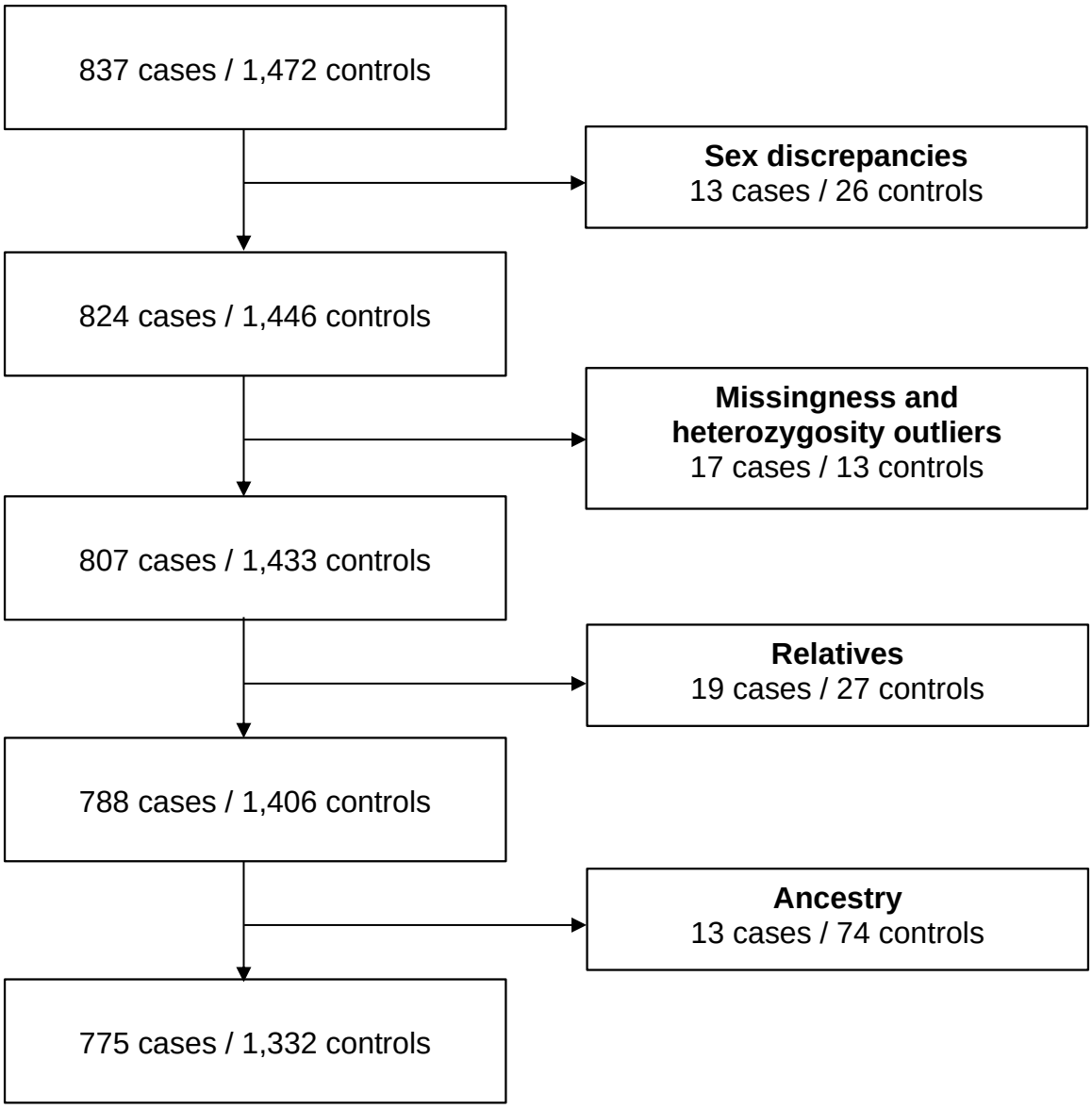
Chr. Chromosome, MAF: minor allele frequency, OR: odds ratio, CI: confidence interval.

Supplementary Figures

Supplementary Figure 1. Principal components analysis of HCC cases and controls. Principal component analysis (PCA) was performed for identifying large-scale differences in ancestry between individuals (appendix p 4). The grey crosses represent HCC cases and controls. The colored points represent the five super populations retrieved form the 1,000 Genomes data.[9] The two panels show the results of the PCA before (A) and after (B) exclusion of non-European ancestry outliers.

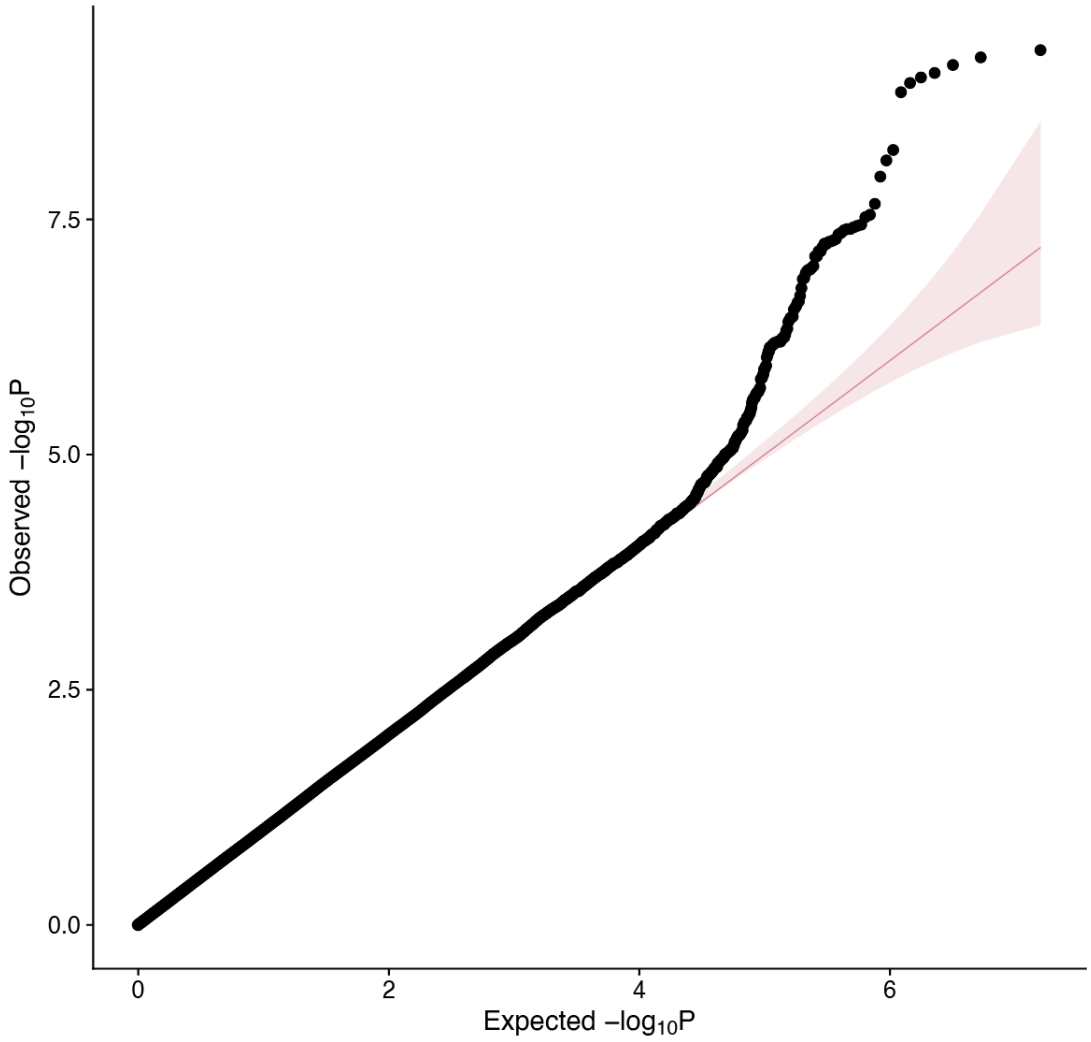


Supplementary Figure 2. Flow charts of genotyped cases and controls in the discovery cohort with reasons for exclusion.

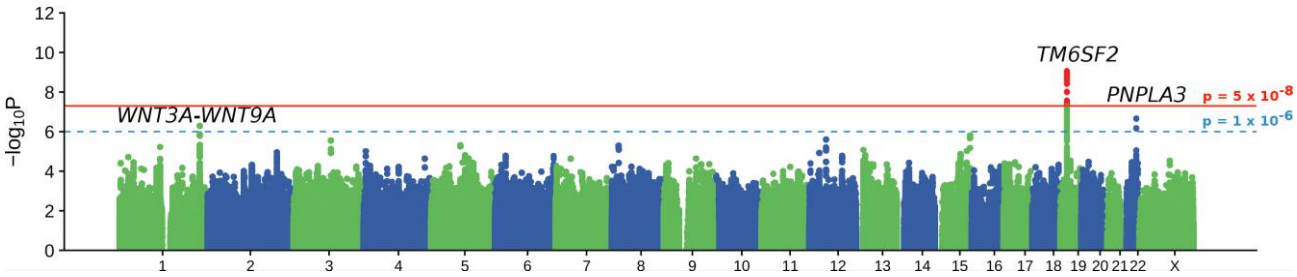


Supplementary Figure 3. Quantile-Quantile (Q-Q) plot for the genome-wide association study for alcohol-related HCC (main analysis) from the discovery cohort.

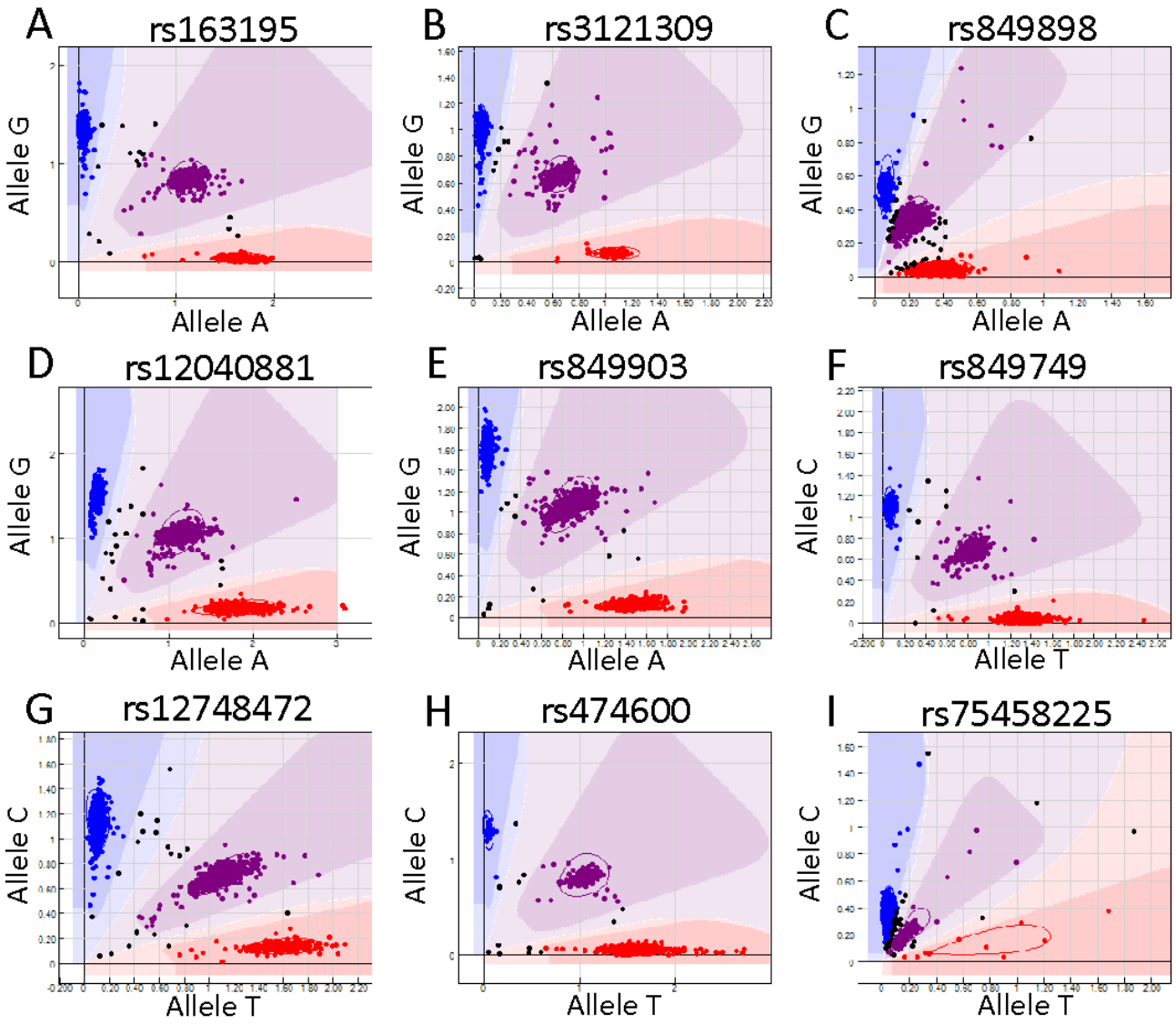
Observed p-values are plotted against expected p-values for all 7,956,264 SNPs after quality control procedures. The 2·5th and 97·5th centiles of the distribution under random sampling and the null hypothesis (red line) form the 95% confidence interval (red band). Black dots show the p-values corrected for the first ten principal components. The genomic inflation factor lambda (λ) was 1·017 suggesting no obvious evidence of population stratification.[16]



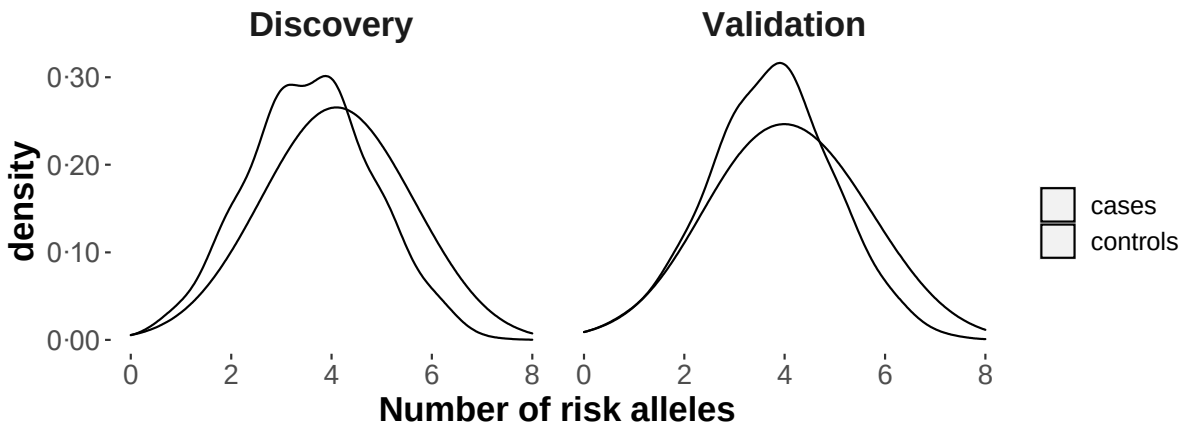
Supplementary Figure 4. Genome-wide association study for alcohol-related HCC in patients with F3-F4 fibrosis from the discovery cohort. Panel A shows a Manhattan plot summarizing the results of the association between, single-nucleotide polymorphisms (SNPs) and alcohol-related HCC in 618 cases and 1244 controls with F3-F4 fibrosis 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 –log10 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 $p = 5\times10^{-8}$; SNPs reaching this threshold are colored in red. The blue dashed line indicates the chosen threshold of $p = 1\times10^{-6}$ for validation follow-up. The nearest gene to the index SNP is indicated above each association peak.



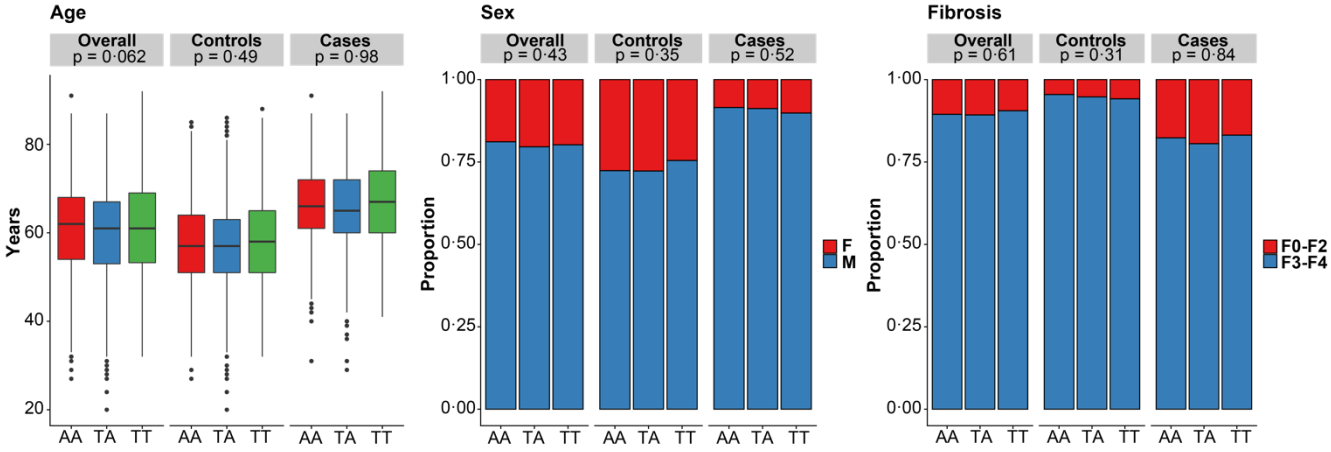
Supplementary Figure 5. Genotype intensity cluster plots in the *WNT3A-WNT9A* region. Genotype intensity cluster plots at 1q42.13 of the nine genotyped SNPs most strongly associated with HCC (ordered by P values) from Genomestudio software are shown. Genotypes depicted in black were automatically set to missing during initial genotype calling (prior to GWAS quality control). All intensity cluster plots demonstrate high quality, well separated clusters for homozygous (red and blue) and heterozygous (purple) genotypes.



Supplementary Figure 6. Distribution of the sum of the number of risk alleles in *PNPLA3* rs738409, *TM6SF2* rs58542926, *WNT3A*-*WNT9A* rs708113 and *HSD17B13* rs72613567.
Density plot showing the distribution of the sum of number of risk alleles in cases and controls in the discovery and validation cohorts.

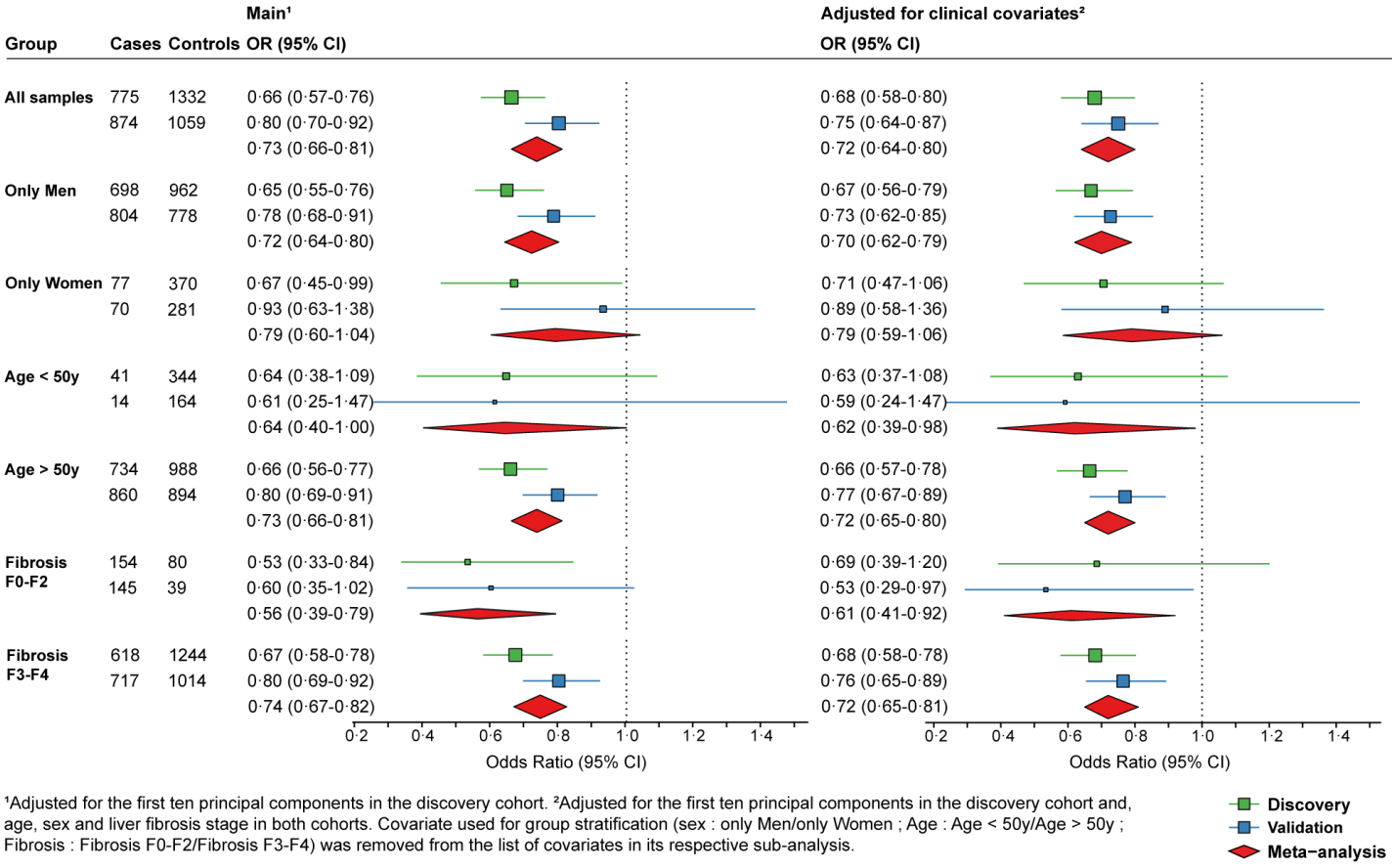


Supplementary Figure 7. Association between *WNT3A*-*WNT9A* rs708113 and clinical characteristics.



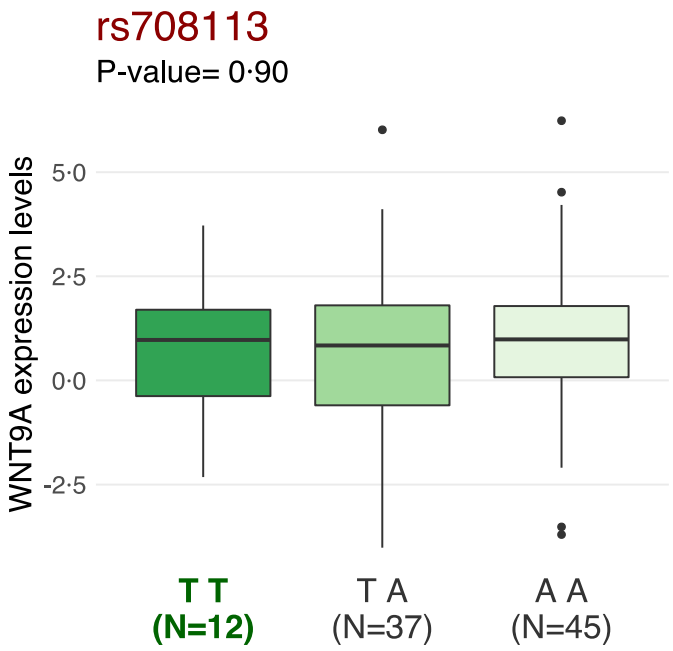
Supplementary Figure 8. Association between *WNT3A-WNT9A* rs708113 and alcohol-related HCC after stratification for clinical risk factors.

Subgroup analyses for age (age < 50 years; age > 50 years), sex (only Men; only Women) and fibrosis stage fibrosis (F0-F2; F3-F4) are tested in the discovery and validation cohorts and after meta-analysis (inverse-variance method). The squares and lines represent odds ratio and the associated 95% confidence interval. The size of the square is proportional to the size of the subgroup cohort size. The diamonds represent the odds ratios after meta-analysis with the width representing the 95% confidence interval.



[1]Adjusted for the first ten principal components in the discovery cohort. [2]Adjusted for the first ten principal components in the discovery cohort and, age, sex and liver fibrosis stage in both cohorts. Covariate used for group stratification (sex : only Men/only Women ; Age : Age < 50y/Age > 50y ; Fibrosis : Fibrosis F0-F2/Fibrosis F3-F4) was removed from the list of covariates in its respective sub-analysis.

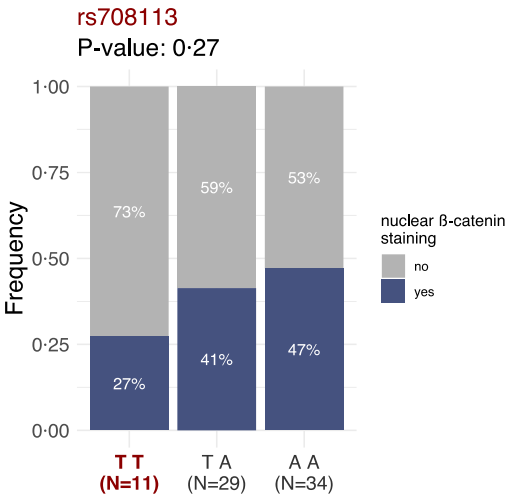
Supplementary Figure 9. Association between *WNT3A-WNT9A* rs708113 and *WNT9A* expression levels in non-tumor liver tissue from patients with alcohol-related HCC.



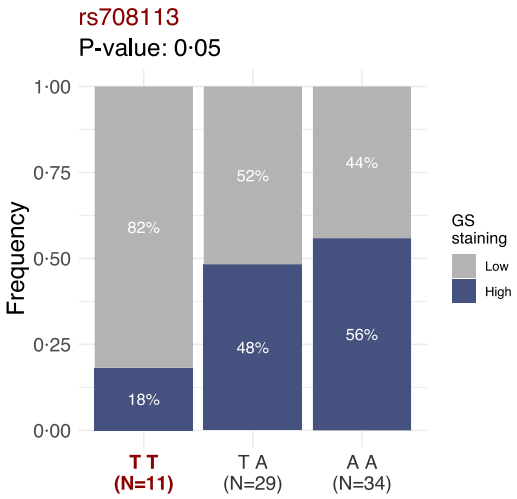
Supplementary Figure 10. Association between *WNT3A-WNT9A* rs708113 genotypes and nuclear ß-catenin and glutamine synthetase immunohistochemistry in a subset of patients with alcohol-related HCC.

This barplot shows the proportion of samples positive for nuclear ß-catenin (Panel A) and glutamine synthetase (GS)[Panel B] staining in a subset of patients with alcohol-related HCC.

A

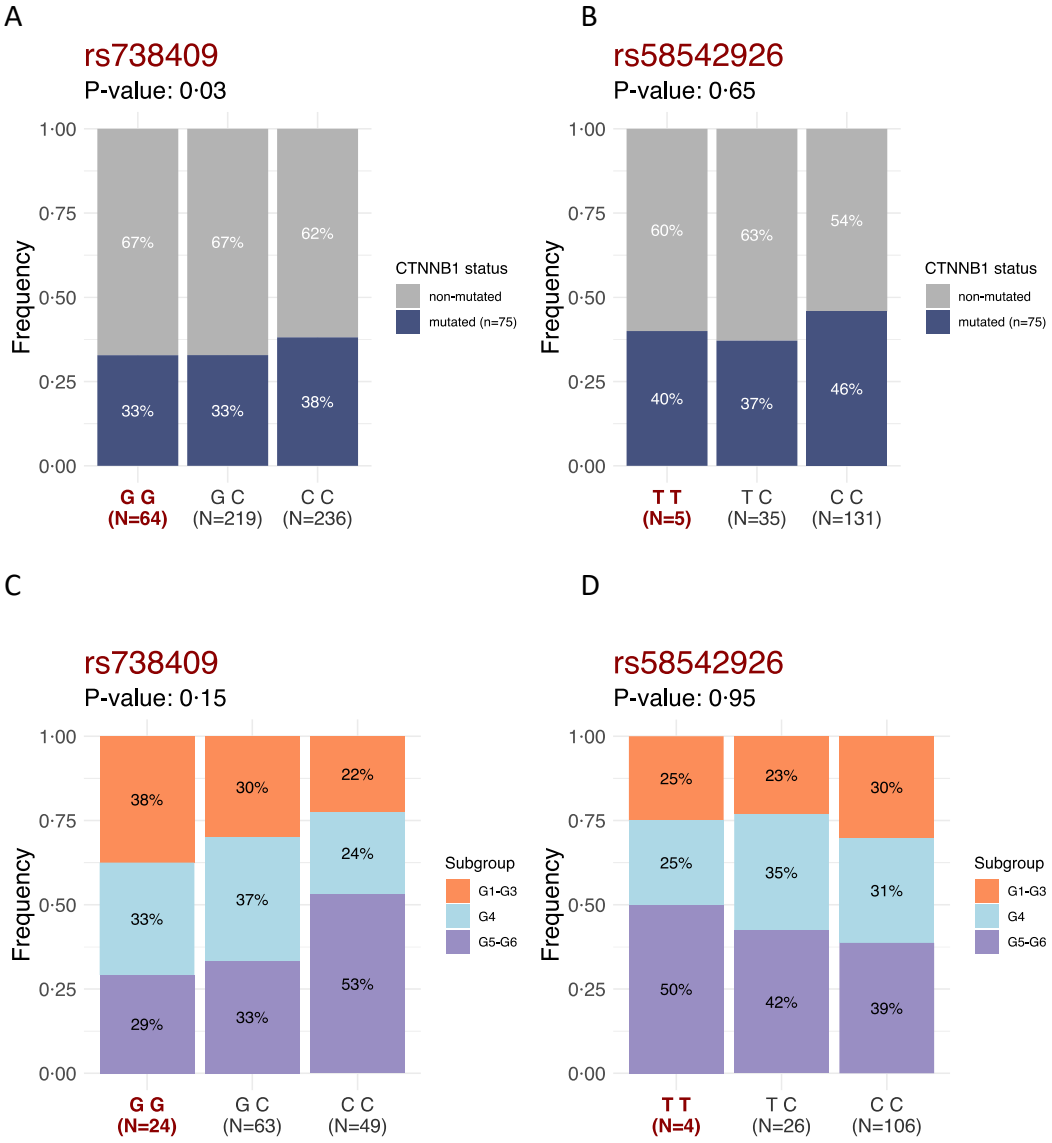


B

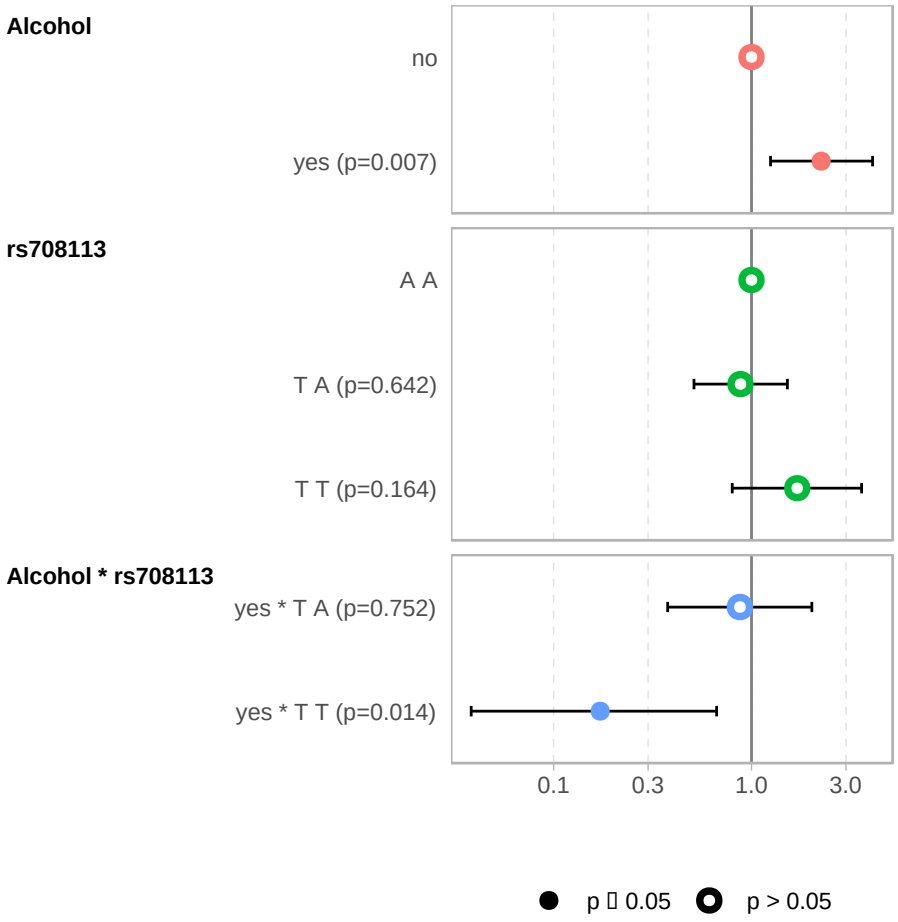


Supplementary Figure 11. Association between *PNPLA3* rs738409, *TM6SF2* rs58542926 genotypes and *CTNNB1* somatic mutations.

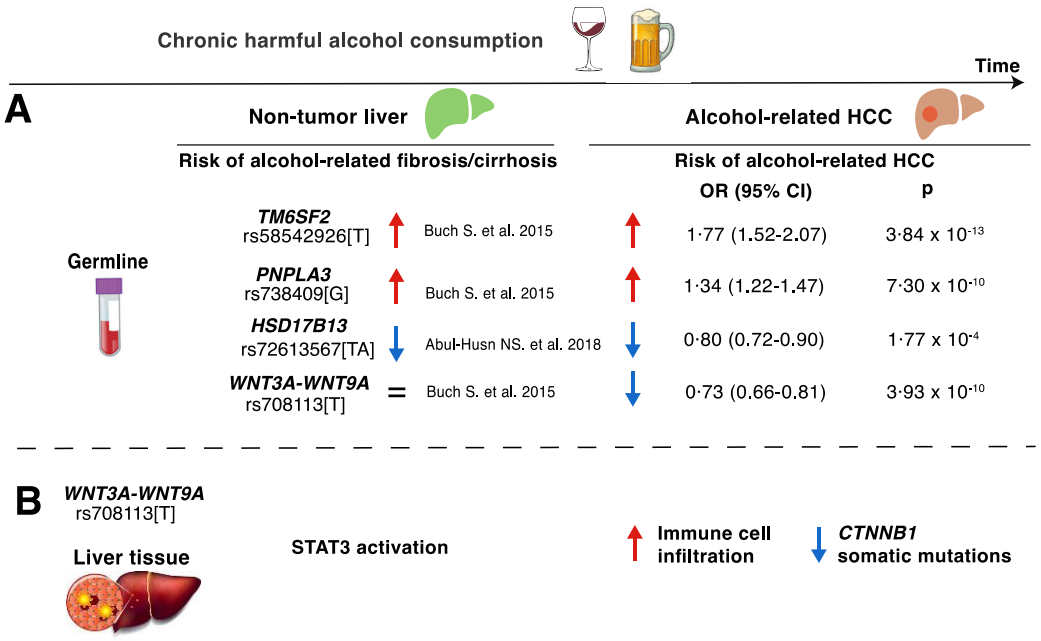
Association between *PNPLA3* rs738409 and *TM6SF2* rs58542926 genotypes and the presence of *CTNNB1* somatic mutations in tumor liver tissue from patients with alcohol-related HCC. A significantly lower proportion of *CTNNB1* somatic mutations was observed in patients harboring the *PNPLA3* rs738409[G] allele (Panel A) but not in patients with the *TM6SF2* rs58542926[T] allele (Panel B). Panel C and D show the proportion of each transcriptomic group according to the *PNPLA3* rs738409 and *TM6SF2* rs58542926 genotypes.



Supplementary Figure 12. Effect of alcohol consumption and *WNT3A-WNT9A* rs708113 variant on CTNNB1-mutated HCC prevalence. Logistic regression model showed a significant interaction effect between *WNT3A-WNT9A* rs708113 and alcohol intake on the prevalence of CTNNB1-mutated HCC**.**



Supplementary Figure 13. Common inherited genetic variants associated with alcohol-related HCC and their effect on ALD progression.



Panel A summarizes inherited genetic associated with alcohol-related HCC. The left part displays the effect of these variants on the risk of alcohol-related fibrosis/cirrhosis based on the results of previous GWAS. The red and blue arrows indicate an increased and decreased risk, respectively. The *TM6SF2* rs58542926[T] and *PNPLA3* rs738409[G] alleles increase the risk of alcohol-related fibrosis/cirrhosis while *HSD17B13* rs72613567[TA] has a protective effect. Conversely to other variants, *WNT3A-WNT9A* rs708113[T] did not show any significant impact on the risk of alcohol-related fibrosis/cirrhosis. The right part displays the effect of the same variants on the risk of alcohol-related HCC. Data are odds ratios (OR) and 95% confidence intervals (CI) from the meta-analysis as observed in this study. Similarly, the red and blue arrows indicate an increased and decreased risk, respectively. The *TM6SF2* rs58542926[T] and *PNPLA3* rs738409[G] allele increase the risk of alcohol-related HCC while *HSD17B13* rs72613567[TA] and *WNT3A-WNT9A* rs708113[T] mitigated that risk. Buch S. et al. 2015[15]; Abul-Husn NS et al. 2018[17]. Panel B shows the effect of *WNT3A-WNT9A* rs708113[T] protective allele in liver tissue of patients with alcohol-related HCC. In non-tumor liver tissue (left), rs708113[T] was associated with STAT3 activation. In tumor tissue (right), rs708113[T] was associated with an increased immune and stromal cell infiltration and low frequency of *CTNNB1* somatic mutations.

References

1 Purcell S, Neale B, Todd-Brown K, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–75.

2 Price AL, Weale ME, Patterson N, *et al.* Long-Range LD Can Confound Genome Scans in Admixed Populations. *The American Journal of Human Genetics* 2008; **83**: 132–5.

3 Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc* 2010; **5**: 1564–73.

4 Pedersen BS, Quinlan AR. Who's Who? Detecting and Resolving Sample Anomalies in Human DNA Sequencing Studies with Peddy. *The American Journal of Human Genetics* 2017; **100**: 406–13.

5 NCI-NHGRI Working Group on Replication in Association Studies. Replicating genotype–phenotype associations. *Nature* 2007; **447**: 655–60.

6 Loh P-R, Danecek P, Palamara PF, *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* 2016; **48**: 1443–8.

7 Durbin R. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). *Bioinformatics* 2014; **30**: 1266–72.

8 the Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016; **48**: 1279–83.

9 The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015; **526**: 68–74.

10 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.

11 Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* 2016; **32**: 1493–501.

12 Witte JS, Visscher PM, Wray NR. The contribution of genetic variants to disease depends on the ruler. *Nat Rev Genet* 2014; **15**: 765–76.

13 for the CiRCE Study Group, Oussalah A, Avogbe PH, *et al. BRIP1* coding variants are associated with a high risk of hepatocellular carcinoma occurrence in patients with HCV- or HBV-related liver disease. *Oncotarget* 2017; **8**: 62842–57.

14 Archambeaud I, Auble H, Nahon P, *et al.* Risk factors for hepatocellular carcinoma in Caucasian patients with non-viral cirrhosis: the importance of prior obesity. *Liver Int* 2015; **35**: 1872–6.

15 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.

16 Devlin B, Roeder K. Genomic Control for Association Studies. *Biometrics* 1999; **55**: 997–1004.

17 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.