Title

Molecular characterization of hepatocellular carcinoma in patients with non-alcoholic steatohepatitis

Authors

Roser Pinyol[1]*, Sara Torrecilla[1]*, Huan Wang[2], Carla Montironi[1], Marta Piqué-Gili[1], Miguel Torres-Martin[1,3], Leow Wei-Qiang[3,4], Catherine E. Willoughby[1], Pierluigi Ramadori[5], Carmen Andreu-Oller[1,3], Patricia Taik[2], Youngmin A. Lee[3,6], Agrin Moeini[1], Judit Peix[1], Suzanne Faure-Dupuy[5], Tobias Riedl[5], Svenja Schuehle[5], Claudia P. Oliveira[7], Venancio A. Alves[7], Paolo Boffetta[3,8], Anja Lachenmayer[9], Stephanie Roessler[10], Beatriz Minguez[11], Peter Schirmacher[10], Jean-François Dufour[9], Swan N. Thung[3], Helen L. Reeves[12,13], Flair J. Carrilho[7], Charissa Chang[3], Andrew V. Uzilov[2,14], Mathias Heikenwalder[5], Arun Sanyal[15], Scott L. Friedman[3], Daniela Sia[3], Josep M. Llovet[1,3,16]

*Shared co-first authorship.

Affiliations

[1]Liver Cancer Translational Research Laboratory, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Hospital Clínic, Universitat de Barcelona, Barcelona, Catalonia, Spain.

[2]Sema4, Stamford, Connecticut, USA.

[3]Mount Sinai Liver Cancer Program (Divisions of Liver Diseases, Department of Hematology/Oncology, Department of Medicine, Department of Pathology), Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, USA.

[4]Department of Anatomical Pathology, Singapore General Hospital, Singapore, Singapore.

[5]Division of Chronic Inflammation and Cancer, German Cancer Research Center Heidelberg (DKFZ), Heidelberg, Germany.

[6]Department of Surgical Sciences, Vanderbilt University Medical Center, Nashville, Tennessee, USA

[7]Departments of Gastroenterology and Pathology, University of São Paulo - School of Medicine, Sao Paulo, Brazil.

[8]Department of Medical and Surgical Sciences, University of Bologna.

[9]Department of Visceral Surgery and Medicine, Bern University Hospital, University of Bern, CH-3010, Bern, Switzerland.

[10]Institute of Pathology, University Hospital Heidelberg, Germany.

[11]Liver Unit, Vall d´Hebron Hospital Universitari, Liver Diseases Research Group, Vall d´Hebron Institut of Research (VHIR), Vall d´Hebron Hospital Campus. CIBERehd, Universitat Autonoma de Barcelona, Barcelona, Catalonia, Spain.

[12]Newcastle University Translational and Clinical Research Institute and Newcastle University Centre for Cancer, Medical School, Framlington Place, Newcastle Upon Tyne, NE2 4HH, UK.

[13]Hepatopancreatobiliary Multidisciplinary Team, Newcastle upon Tyne NHS Foundation Trust, Freeman Hospital, Newcastle upon Tyne, UK.

[14]Department of Genetics and Genomic Sciences and Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, New York.

[15]Divisions of Gastroenterology and Hepatology, Virginia Commonwealth University, Richmond, VA, USA.

[16]Institució Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain.

Corresponding author

*Josep M. Llovet, M.D., Ph.D., Mount Sinai Liver Cancer Program, Division of Liver Diseases, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA . E-mail: josep.llovet@mssm.edu

Keywords: liver cancer, obesity, metabolic syndrome, molecular class, mutational signature, animal model.

Electronic word count: 5928

Abstract electronic word count: 271

Number of Figures and Tables: 1 Table and 6 Figures

Availability of data: Whole genome expression data used in the study are available under the accession numbers GSE63898, GSE15654 and GSE164760, and whole exome sequencing data under EGAD00001000131 and EGAS00001005222.

Conflict of interest

J.M.L. receives research support from Bayer HealthCare Pharmaceuticals, Eisai Inc, Bristol-Myers Squibb, Boehringer-Ingelheim and Ipsen, and consulting fees from Eli Lilly, Bayer HealthCare Pharmaceuticals, Bristol-Myers Squibb, Eisai Inc, Celsion Corporation, Exelixis, Merck, Ipsen, Genentech, Roche, Glycotest, Leerink Swann LLC, Fortress Biotech, Nucleix, Can-Fite Biopharma, Sirtex, Mina Alpha Ltd and AstraZeneca. S.L.F consults for the following companies: 89 Bio, Amgen, Axcella

Health, Blade Therapeutics, Bristol Myers Squibb, Can-Fite Biopharma, ChemomAb, Escient Pharmaceuticals, Forbion, Foresite laboratories, Galmed, Gordian Biotechnology, Glycotest, Glympse Bio, Hepgene, In sitro, Morphic Therapeutics, North Sea Therapeutics, Novartis, Ono Pharmaceuticals, Pfizer Pharmaceuticals, Scholar Rock Surrozen. He has stock options in the following companies: Blade Therapeutics, Escient, Galectin, Galmed, Genfit, Glympse, Hepgene, Lifemax, Metacrine, Morphic Therapeutics, Nimbus, North Sea Therapeutics, Scholar Rock, Surrozen. He receives research support from Morphic Therapeutics, Novo Nordisk, and Galmed, and has an SBIR grant with Abalone Bio. A.L. is a consultant for Neuwave and Histosonics. C.P.M.S.O. consults for Bayer, Novartis, Novonordisk, Allergan, Pfizer, Roche and Zambon. P.S. is receiving research grants from BMS, Roche, Incyte, Chugai and consulting fees from BMS, MSD, Incyte, Janssen, Roche, AstraZeneca and Amgen. H.W., P.T., and A.V.U. are or were salaried employees of Sema4 at the time of the study. H.W., P.T., and A.V.U. hold Sema4 stock options. The rest of authors have nothing to disclose. B.M. received consultancy fees from Bayer-Shering Pharma, and speaker fees from Eisai and MSD.

Grant Support

J.M.L. is supported by grants from the European Commission (EC) Horizon 2020 Program (HEPCAR, proposal number 667273-2), the US Department of Defense (CA150272P3), the National Cancer Institute (P30 CA196521), the NIH (RO1DK56621 and RO1DK128289 to SLF), the Samuel Waxman Cancer Research Foundation, the Spanish National Health Institute (MICINN, SAF-2016-76390 and PID2019-105378RB-I00), through a partnership between Cancer Research UK, Fondazione AIRC and

Fundación Científica de la Asociación Española Contra el Cáncer (HUNTER, Ref. C9380/A26813), and by the Generalitat de Catalunya (AGAUR, SGR-1358). S.T. was awarded with a grant from the Spanish National Health Institute (Ref. EEBB-I-17-12316) and with the EASL Andrew K. Burroughs Fellowship. C.M. and C.E.W. are respectively supported by a Rio Hortega fellowship (CM19/00039) and a Sara Borrell fellowship (CD19/00109) from the ISCIII and the European Social Fund. C.A.O. was supported by a Fulbright Fellowship and a "laCaixa" INPhINIT Fellowship (LCF/BQ/IN17/11620024). M.H. is supported by the SFB 179 and an ERCCoG (HeparoMetaboPath). A.L. is supported by the Swiss Transplant Cohort Study. H.L.R. is supported by a Cancer Research UK (CR UK) centre grant C9380/A18084; programme grant C18342/A23390 and Accelerator award C9380/A26813. P.S. and SR were supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 314905040 – SFB/TRR 209 Liver Cancer (B01 to S.R. and Z01, INF to P.S.), the European Union's Horizon 2020 research and innovation programme (grant no. 667273, HEP-CAR), and the Eurostars (grant E! 113707, LiverQR) which was funded by the Bundesministerium für Bildung und Forschung (BMBF). S.R. was supported by the German Cancer Aid (grant no. 70113922). D.S. is supported by the PhD Scientist Innovative Research Award. B. M. received grant support from Instituto de Salud Carlos III (PI18/00961).

Author contributions

ST, RP, DS, JML were involved in the study design; ST, RP, HW, PT, CM, MPG, JP, SFD, TR, SS, LWQ, AM, SNT, AVU and DS performed experiments/analysis; YL, CPO, VAA, AL, SR, BM, PS, PB, JFD, HLR, FJC, CC, MH, AJS, YAL, SLF, JML

provided tissue samples; ST, RP, HW, CM, CEW, CAO, MPG, PR, MTM, AM, AVU, DS, JML provided scientific input; RP, DS and JML wrote the manuscript; all authors were involved in the critical revision of the manuscript.

Abstract

Background and Aims: Non-alcoholic steatohepatitis (NASH)-related hepatocellular carcinoma (HCC) is increasing globally, but its molecular features are not well defined. We aimed to identify unique molecular traits characterizing NASH-HCC compared to other HCC aetiologies.

Methods: We collected 80 NASH-HCC and 125 NASH samples from 5 institutions. Expression array (n=53 NASH-HCC; n=74 NASH) and whole exome sequencing (n=52 NASH-HCC) data were compared to HCCs of other aetiologies (n=184). Three NASH-HCC mouse models were analysed with RNAseq/expression-array (n=20). Activin A Receptor Type 2A (*ACVR2A*) was silenced in HCC cells and proliferation assessed by colorimetric and colony formation assays.

Results: Mutational profiling of NASH-HCC tumours revealed *TERT-promoter* (56%), *CTNNB1* (28%), *TP53* (18%) and *ACVR2A* (10%) as the most-frequently mutated genes. *ACVR2A* mutation rates were higher in NASH-HCC than in other HCC aetiologies (10% *versus* 3%, $p<0.05$). *In vitro*, *ACVR2A* silencing prompted a significant increase in cell proliferation in HCC cells. We identified a novel mutational signature (MutSig-NASH-HCC) significantly associated with NASH-HCC (16% vs 2% in viral/alcohol-HCC, $p=0.03$). Tumour mutational burden (TMB) was higher in non-cirrhotic than in cirrhotic NASH-HCCs (1.45 *versus* 0.94 mutations/Mb; $p<0.0017$). Compared to other aetiologies of HCC, NASH-HCCs were enriched in bile and fatty acid signalling, oxidative stress and inflammation, and presented a higher fraction of Wnt/TGF-β proliferation subclass tumours (42% *versus* 26%, $p=0.01$) and a lower prevalence of the CTNNB1 subclass. Compared to other aetiologies, NASH-HCC showed a significantly higher prevalence of an *immunosuppressive cancer field*. In

three murine models of NASH-HCC, key features of human NASH-HCC were preserved.

Conclusions: NASH-HCCs display unique molecular features including higher rates of *ACVR2A* mutations and the presence of a newly identified mutational signature.

Lay Summary

Hepatocellular carcinoma (HCC) associated with non-alcoholic steatohepatitis (NASH) is increasing globally, but its molecular traits are not well characterized. Our molecular characterization has uncovered higher rates of *ACVR2A* mutations (10%) –a potential tumour suppressor– and the presence of a novel mutational signature (MutSig-NASH-HCC), as well as a more prominent role of a Wnt/TGF-β proliferation subclass in tumours (42%) and immunosuppressive traits in the adjacent non-tumoral tissue.

Highlights

- The most-frequently mutated genes in NASH-HCC were *TERT*, *CTNNB1*, *TP53* and *ACVR2A*.
- Mutations in *ACVR2A* –a potential tumour suppressor gene– were higher in NASH-HCC than in other aetiologies.
- A novel mutational signature significantly associated with NASH-HCC was identified.
- The Wnt/TGF-β proliferation subclass was more prevalent in NASH-HCC than in HCCs of other aetiologies.
- NASH-HCC showed a significantly higher prevalence of an *immunosuppressive pro-carcinogenic cancer field*.

Introduction

Hepatocellular carcinoma (HCC) represents the most common type of liver cancer[1]. Risk factors for HCC development are well defined, and include cirrhosis, hepatitis B (HBV) and C (HCV) virus infection, alcohol abuse and non-alcoholic fatty liver disease (NAFLD)[1,2]. NAFLD is the most common cause of chronic liver disease, with a worldwide prevalence of 25% (ranging from 32% in Middle-East to 14% in Africa, and ~25% in USA and Europe)[3] and is expected to become the leading cause of HCC in developed countries[4,5]. NAFLD occurs in the absence of significant alcohol consumption and it ranges from non-alcoholic fatty liver (NAFL), to non-alcoholic steatohepatitis (NASH), characterized by hepatic triglyceride accumulation, inflammation and hepatocyte injury[4]. It has been recently considered an auto-aggressive disease[6].

Knowledge of HCC's molecular pathogenesis is expanding[1,7]. Genomic analyses have revealed key pathways altered in HCC, including Wnt/β-catenin, PI3K/Ras, and cell-cycle pathways. The most frequent HCC drivers genes (i.e. *TERT*, *CTNNB1*, *TP53*) and mutational signatures associated with risk factors have been identified[7]. However, these studies were conducted mainly in viral-related HCC, whereas tumours with underlying NASH have been underrepresented. Several studies have analysed the relevance of clinical parameters involved in the transition from NASH to NASH-HCC[8,9], but few studies have sought to clarify the molecular drivers of hepatocarcinogenesis in the NASH setting. Such studies have identified a) genetic variants involved in HCC progression in NASH patients (i.e. adiponutrin (PNPLA3), TM6SF2)[4], b) oncogenic factors (i.e. inactivation of T-cell protein tyrosine phosphatase (TCPTP), IL-17A production, overexpression of Squalene Epoxidase (SQLE))[10–12], and c) epigenetic events repressing the transcription of genes

related to bile and fatty acid metabolism[13]. Thus, molecular studies based on large cohorts of NASH-related HCCs are required, particularly in light of recent observations suggesting that these patients benefit less from immune checkpoint inhibitors than patients with viral-HCC[14].

Here, we conducted a comprehensive molecular analysis of a large cohort of histopathologically diagnosed NASH-HCCs and identified: a) significantly higher rates of mutations in the TGF-β-related activin receptor *ACVR2A* (10%) compared to viral/alcohol-HCC (3%); b) a novel mutational signature almost exclusive to NASH-HCCs (MutSig-NASH-HCC); c) enrichment of bile- and fatty acid signalling, oxidative stress and inflammation; and d) lack of molecular differences between adjacent tumoural tissue in HCC associated with NASH livers and cirrhotic NASH livers.

Materials and methods

Study cohorts

We collected 80 NASH-HCCs and 125 NASH formalin fixed paraffin-embedded (FFPE) samples from 5 different institutions, and 20 publicly available fresh frozen (FF) NASH-HCC samples[15] (**Supplementary Table 1**). NASH was diagnosed in FFPE samples by at least two expert pathologists (CM, LWQ and/or SNT) following a described histological algorithm[16] (**Supplementary Data File**). All NASH patients included in the study were HBV- and HCV-negative. Patients reporting alcohol consumption ≥20 g/day for women and ≥30 g/day for men, as well as patients with a known liver disease superimposed to NASH were excluded. **Table 1** details patients' clinico-pathological characteristics.

For comparative purposes, the study included a) the transcriptomic profile of the HEPTROMIC cohort[17] (HBV=48, HCV=103, alcohol=33), b) transcriptomic and mutational data from the HCC-TCGA cohort (n=345)[18], c) whole exome sequencing (WES) data from 45 viral/alcohol-HCCs[15], and d) the mutational information of 624 viral/alcohol-HCCs (alcohol=170, HBV=355, HCV=99)[18–20].

Whole Exome Sequencing and Mutational Signatures

WES data from paired tumour and non-tumour samples (52 NASH-HCC and 45 HCCs of other aetiologies) was used to assess the mutational landscape. A *de novo* mutational signature extraction procedure was conducted using the SNV variants identified in 86 samples (43 NASH-HCC and 43 HCCs of other aetiologies)[15]. The presence of mutational signatures of environmental agents was assessed[22]. Complete details are described in the **Supplementary Data File**.

Methodological details on whole gene expression profiling, histological evaluation, immunohistochemical analyses, NASH-HCC murine models, the assessment of the functional role of ACVR2A and the statistical information are described in the **Supplementary Data File**.

Results

Clinicopathological characteristics of NASH-HCC patients

Our study analysed 80 HCC samples from NASH patients (*NASH-HCC cohort)*, and 125 liver samples from NASH patients without HCC (*NASH cohort).* The NASH-HCC samples were obtained from patients who underwent resection (n=41/80, 51.3%) or transplant (n=38/80, 47.5%), while the NASH samples were obtained from liver biopsy (n=102/125, 81.6%) or liver transplant (n=23/125, 18.4%). In the NASH-HCC cohort, the prevalence of HCC in men was significantly higher than in women (81.3% vs 18.8%, $p<0.001$, **Table 1**)[23]. In addition, most HCC cases were at early/intermediate clinical stages (51% BCLC-0A and 40% BCLC-B), with median size of 2.7 cm; 44% HCCs were multinodular and 41.8% presented satellites. Moreover, the majority were moderately differentiated (G2, 69%) and 41.3% presented microvascular invasion.

Next, we compared the clinical characteristics of NASH-HCC non-cirrhotic patients with those of NASH-HCC cirrhotic patients (**Supplementary Table 2**). Non-cirrhotic patients had more hypertension (95% vs 74%, $p=0.047$), lower body mass index (BMI) (28 vs 31 kg/m^2, $p=0.02$), higher albumin levels (4.1 vs 3.5, $p=0.048$) and serum platelet counts ($206 \cdot 10^3$ vs $85 \cdot 10^3$ platelets/ml), and lower bilirubin (0.6 vs 1.8 mg/dl, $p<0.0001$) and INR (1.0 vs 1.4, $p<0.0001$). Tumours of non-cirrhotic NASH

patients displayed lower rates of multinodular HCCs (22% vs 52%, p=0.029), but more frequent satellite lesions (69% vs 30%, p=0.008; **Supplementary Table 2**).

Median age was significantly lower in NASH compared to NASH-HCC patients (56 vs 65, p<0.00001) and women were more prevalent (58.1% vs 41.9%, p<0.001). Patients with NASH-HCC exhibited higher rates of metabolic syndrome features including hypertension (80.3% vs 52.1%; p<0.001) and diabetes (72.4% vs 50.4%; p=0.003), but similar prevalence of obesity (55.4% vs 61.4%) and hyperlipidaemia (53.7% vs 57.8%). Cirrhosis was also more prevalent among NASH-HCC patients than among NASH patients (70.0% vs 28.8%, p<0.00001; **Supplementary Table 2)**.

Genomic alterations in human NASH-HCC

The mutational landscape of NASH-HCC was assessed in 52 paired samples of HCC tissue and non-tumorous adjacent tissue. For comparative purposes, we analysed 45 additional viral/alcohol-related HCC cases (HCV=12, HBV=16, alcohol=17)[15].

The median sequencing depth of our sequenced was 100x for NASH-HCCs and 35x for non-tumour tissue. The median number of non-silent mutations was 60 (ranging from 6 to 167), corresponding to 1.2 mutations per megabase (TMB, tumour mutational burden), consistent with previous reports[15,24,25]. No significant difference was observed in the number of SNVs between the FFPE and the FF samples (median number of non-silent mutations per sample: 64.0 vs 56.5; p=0.4). In total, we identified 1,653 mutated genes, among which 82 were defined as putative tumour-driver genes[26], with a median of 3 mutated driver genes per sample. After integrating focal copy-number gains and losses there were in aggregate a total of 96 altered genes (**Supplementary Table 3)**. The most frequent identified alterations occurred in the *TERT promoter* (56%), followed by *CTNNB1* (28%), *TP53* (18%) and

ACVR2A (10%) (**Figure 1A**). *TERT promoter* mutations were accompanied by *TERT* overexpression in 28% of the cohort (FC≥1.5; p<0.0001, **Supplementary Table 4**).

When comparing NASH-HCCs with HCCs of other aetiologies, the landscape of mutations was similar, except for *ACVR2A* and *TP53* mutations. Specifically, NASH-HCCs exhibited a trend towards significantly higher rates of *ACVR2A* mutations (10% vs 4.4% in other HCCs) and lower rates of *TP53* mutations (18% vs 31% in other HCCs; **Supplementary Figure 1A**). To confirm these observations, we expanded the other aetiologies HCC cohort through a meta-analysis including 624 samples (HCV=99, HBV=355, alcohol=170). Our results showed a significantly increased frequency of *ACVR2A* mutations (10% vs 3%, p=0.02) and a trend towards lower rates *of TP53* mutations (18% vs 32%, *p*=0.051) in NASH-HCC compared to the viral/alcohol-related HCC cases (**Figure 1B**). Notably, this difference was mainly driven by low rates of *ACVR2A* mutations in HBV-HCC (1% vs 10%, p=0.0037, **Supplementary Figure 1B**), and persisted when comparing the non-cirrhotic NASH-HCC cases with the non-cirrhotic non-NASH-HCC cases.

Correlation analyses between tumour-driver mutations and the patients' clinico-pathological features revealed that, in NASH-HCC: a) *TP53* mutations were associated with multinodular tumours [33% (7/21) vs 0% (0/13) in single HCCs, p= 0.019], b) mutations in *ARID1A* (6%) were related to tumours with vascular invasion [35% (3/8) vs 0% (0/26), p=0.001], and c) *PDGFRA* mutations were significantly more prevalent in female patients [22% (2/9) vs 0% (0/41), p=0.02; **Supplementary Table 5**].

We next analysed the impact of cirrhosis on the mutational landscape of NASH-HCC. Interestingly, the overall burden of mutations was significantly higher in tumours from non-cirrhotic (n=30) than from cirrhotic (n=22) patients (**Figure 1C**), with a median

number of mutations of 72.4 vs 46.9; corresponding to 1.45 and 0.94 mutations/Mb ($p<0.0017$), respectively. This difference was maintained when adjusting for tumour size and tumour differentiation degree (**Supplementary Table 6**). Interestingly, non-cirrhotic non-NASH-HCC tumours also displayed a higher median number of mutations than cirrhotic ones (117.5 vs 66.0; $p=0.00004$).

Finally, we used the WES data to explore the presence of the germline variant rs738409 C>G p.I148M in the *PNPLA3* gene, which is known to be associated with HCC risk[27]. We identified a higher prevalence of the homozygous GG genotype in cirrhotic patients (67% vs 17%, $p=0.001$, **Figure 1C**). At the molecular level, these tumours presented acylglycerol-transacylation and phospholipase-activity signatures, consistent with previous reports, and the poor prognosis TGF-β signature[28] **(Supplementary Figure 1B).** On the other hand, tumours from non-GG homozygous patients were enriched in signatures related to: a) acetyl-CoA metabolism, consistent with PNPLA3 function; b) PPAR transcription factors, naturally activated by fatty acids; and c) oxidative phosphorylation and DNA-damage.

Overall, we identified *TERT (56%)*, *CTNNB1 (28%)*, *TP53* (18%) and *ACVR2A* (10%) as the most frequently altered genes in NASH-HCC, and the incidence of *ACVR2A* mutations was higher in NASH-HCC than in other aetiologies (10% vs 3%).

ACVR2A functions as a tumour suppressor in HCC cell lines

In order to explore the role of *ACVR2A* mutations, we conducted a functional study in cultured HCC cell lines. First, we validated all detected mutations by Sanger sequencing and found *ACVR2A* mutations in five different spots (chr2:148683685, chr2:148676075, chr2:148684650, chr2:148653921 and chr2:148602752) which corresponded to either indels (T>TA and GTCTT>G alterations) or SNVs (T>G, G>T

and A>G; **Figure 1D**; **Supplementary Table 7**). Secondly, we found the expression of *ACVR2A* to be downregulated in 15% of NASH-HCCs (n=8/53; FC<0.5). The decreased expression of *ACVR2A* was significantly associated with mutations in this gene. Specifically, in the TCGA HCC cohort (n=361), *ACVR2A*-mutated tumours presented significantly reduced *ACVR2A* expression when compared to *ACVR2A*-wild type cases (p=0.026, **Supplementary Figure 2C**). No association was seen between *ACVR2A* mutations and cirrhosis (**Supplementary Figure 2A,B**). Based on the above observations, we hypothesized that *ACVR2A* could act as a tumour suppressor in HCC, as it does in colorectal cancer[29]. Hence, we silenced its expression in Hep3B and Huh7 cells using shRNA (**Supplementary Figure 2D**) and performed MTT cell-viability assay and colony formation assays. *ACVR2A* knockdown led to an 8-fold increase in colony formation capacity and a 56% increase in the viability of Hep3B cells, compared to control cells (p=0.03 and p=0.015, respectively) as well as a 41% increase in Huh7 cells viability *versus* controls (p=0.048; **Figure 1E-G and Supplementary Figure 2**). Similar results were obtained when evaluating HCC cell lines that mimic the NASH phenotype (**Supplementary Figure 3**). These data suggest that, in culture, *ACVR2A* functionally acts as a tumour suppressor in HCC, a feature requiring validation *in vivo*.

Mutational signatures underlying the pattern of NASH-HCC mutations

We next aimed to identify the landscape of mutational signatures explaining the SNVs detected by WES. To this end, we submitted the WES profiles of 43 NASH-HCC and 43 HCCs of other aetiologies to a *de novo* extraction process of mutational signatures, which identified three *de novo* signatures. DenovoSig1 and denovoSig3 matched the previously reported liver cancer specific COSMIC v2 signatures 16

(MutSig16) and 24 (MutSig24), respectively (**Supplementary Figure 4A Supplementary Table 8**)[15,30]. The third identified signature (denovoSig2), which was characterized by a higher frequency of C>T and C>A transitions [prevalence of 9% (n=8/86)], did not match any previously reported signature, and was referred to as *MutSig-NASH-HCC.*

The subsequent mutational signature-fitting step generated the spectrum of mutational signatures in both cohorts (**Figure 2A**). Only signatures obtained with a degree of confidence above 90% and able to explain over 20% of the mutations in a sample (exposure > 20%, **Supplementary Figure 4B and Supplementary Table 9**)[31] were considered in the correlation analysis with the clinico-pathological data. MutSig16 was identified as the most prevalent (19%, n=16/86) and was equally distributed among NASH-HCCs and non-NASH-HCCs (**Figure 2B,** and **Supplementary Table 10**). In NASH-HCCs, MutSig16 was associated with *TP53* mutations (p=0.03, **Supplementary Table 10**). The second most prevalent signature was *MutSig-NASH-HCC*, detected in 16% of NASH-HCCs, but only in 2% of viral/alcohol-HCCs (p=0.03, **Supplementary Table 10**). Moreover, female gender rates were significantly enriched in tumours positive for MutSig-NASH-HCC (50% vs. 13% in tumours negative for this signature; p=0.007). Gene Set Enrichment Analysis (GSEA) revealed metabolic and methylation signatures as the two features most significantly associated with these tumours (p<0.0001; FDR = 1; **Supplementary Table 11**). The third most prevalent signature, MutSig24, was identified in 8% of the cases (n=7/86) and was found exclusively in viral/alcohol-related HCCs (0% in NASH-HCC, p=0.006, **Supplementary Table 10**). It was significantly enriched in younger patients (p=0.001) and was related to tumours with vascular invasion

(p=0.012), higher AFP levels (p=0.03) and mutated *TP53* (p=0.001). None of the above reported mutational signatures were associated with differences in survival.

In parallel, we used the WES data to investigate whether exposure to environmental mutagens[22] could explain certain mutational patterns in NASH-HCC. We identified a significantly higher prevalence of the 6-nitrochrysene plus S9 signature[22] in non-cirrhotics vs cirrhotics [41% (n=11/27) vs 6% (n=1/16); p=0.02] and of the diethyl sulphate (DES) signature[22] in cirrhotics [63% (n=10/16) vs 15% (n=4/27); p=0.002]. Summarizing, we detected a new mutational signature (MutSig-NASH-HCC) almost exclusively present in NASH-HCCs (16%) and associated with female gender.

Signalling pathways, and molecular and immune classes in human NASH-HCC

We next sought to identify signalling pathways altered in NASH-HCC. Firstly, we classified the above identified 96 tumour-driver genes according to their pathways. The most commonly altered signalling pathways included telomere maintenance (56%), Wnt/β-catenin (42%) and TP53 (28%), followed by chromatin remodelling (16%), TGF-β (14%), MAPK (12%), PI3K/AKT/MTOR (8%) and oxidative stress (8%, **Figure 3A**).

In addition, NASH-HCCs compared to other aetiologies displayed a significant enrichment of signatures related to: (1) bile acid and fatty acid metabolism (including cholesterol and sterol biosynthesis), (2) oxidative stress and ROS, and (3) inflammation (**Figure 3B**). Of note, specific comparison of NASH-HCC to HCV-HCC revealed higher IFN-α signalling in HCV tumours (**Supplementary Table 12**).

Regarding molecular classes[23], 42% and 15% of NASH-HCCs belonged to Wnt/TGF-β-proliferation (S1) and progenitor cell-proliferation (S2) subclasses[23], respectively, and 36%, to non-proliferation subclass (S3)[23] (**Figure 3C**).

Comparison with HCC of other aetiologies (HCV=103, HBV=48 and alcohol=33)[17] revealed that NASH-HCCs presented significantly higher rates of Wnt/TGF-β proliferation (S1) (42% vs 26%, p=0.01) and a lower prevalence of the CTNNB1 subclass[32] (16% vs 31%, p=0.02; **Supplementary Table 13)**. Further analysis using prognostic and pathway signatures revealed no significant differences between both cohorts[1,33]. Hierarchical clustering analysis further supported this finding **(Supplementary Figure 5A)**.

We next determined the immune profile of the NASH-HCC cohort using reported immune-specific gene signatures (**Supplementary Table 14**). One third of the NASH-HCC cohort (30%, n=16/53) was classified as *Immune Class*[34], with enrichment of signatures related to T cells, cytotoxic cells, and macrophages (**Figure 4A, Supplementary Figure 6**). Among those, 56% were *Immune Active*[34],[38] (n=9/16), and 44% were *Immune Exhausted*[34] (n=7/16) with enrichment for signatures of TGF-β and Active Stroma. No differences were found in the distribution of cirrhotic patients or the number of mutations per sample within the different HCC immune subtypes (**Figure 4A**). Nonetheless, cirrhotic NASH-HCC cases displayed a significant enrichment in features of immune exhaustion (i.e. Tregs, TGF-β) compared with non-cirrhotic NASH (**Supplementary Figure 7A**). In addition, signatures of response to anti-PD1 therapies and overexpression of *CXCL9* were enriched in non-cirrhotic NASH-HCC cases (**Supplementary Figure 7B-C**).

Finally, we sought to analyse the differences between NASH-HCCs developed on cirrhotic livers *versus* non-cirrhotic livers. While non-cirrhotic NASH-HCCs (n=37) were more enriched in pro-proliferative pathways, including the S2 subclass, E2F targets and DNA-damage (FDR<0.005; **Figure 4B, Supplementary Table 15**), NASH-HCCs in cirrhotic livers (n=16) were more associated with signatures of

inflammation, epithelial-to-mesenchymal transition (EMT), angiogenesis, activated stroma and the HCC *Immune Class* (FDR<0.005; **Figure 4C, Supplementary Table 15**).

NASH-related cancer field characterization

We next analysed the transcriptomes of 74 livers from NASH patients without HCC (59 non-cirrhotic and 15 cirrhotic) and found that cirrhotic NASH livers presented marked molecular differences compared to non-cirrhotic NASH livers. In this regard, non-cirrhotic NASH livers presented enrichment of: 1) fatty and bile acid features (including mTOR[36,37]); 2) ROS-related gene sets (i.e. peroxisome, DNA-repair and mitochondria); and 3) insulin signalling (**Figure 5A, Supplementary Figures 8A,B**). Consistently, they displayed a lower immune cancer field (ICF)[38] prevalence compared to cirrhotic NASH (32% vs 93%, $p<0.0001$), where the immunosuppressive subtype (IS-ICF) was the most prevalent form (67%; **Supplementary Figure 9 and Supplementary Table 16**).

Next, we compared NASH livers with NASH-HCC adjacent tissues and found that cirrhotic NASH livers presented molecular similarities with NASH-HCC adjacent tissues (regardless of the cirrhotic status). Non-cirrhotic NASH-HCC and cirrhotic NASH-HCC adjacent tissues were equally characterized by upregulation of inflammatory signatures (IFN, IL17-A, IL6, chemokine signalling or JAK/STAT; $p<0.05$), traits previously linked to NASH pathogenesis[4] (**Figure 5A, Supplementary Figures 7 and 8**). Also, they displayed activation of hepatocarcinogenic pathways including Notch, TGF-β, TP53, and FGF ($p<0.05$; **Supplementary Figure 8**). In terms of immunity, they were both significantly enriched in immune signatures including the HCC *Immune Class* and immune exhaustion features (TGF-β) but no differences in terms of ICF[38] (**Figure 5B, Supplementary Figures 8 and 9**).

When comparing cirrhotic NASH livers with cirrhotic livers of HCV-infected patients from a previous study[39] revealed a higher prevalence of *immunosuppressive*

cancer field in NASH livers (9/15,60% vs 21/216 10% in HCV, p<0.0001)(**Supplementary Table 16**). Finally, we did not identify gatekeeper mutations in the TERT promoter in any of the NASH liver tissue samples.

Altogether, these results suggest that NASH cirrhotic livers (without presence of HCC) present key molecular features that are common with the cancer field traits of adjacent tissue of NASH-HCC patients.

NASH murine models recapitulate features of human NASH-HCC

Several experimental models mimicking metabolic and/or histologic features of NASH have allowed the identification of different molecular mechanisms involved in NASH development and progression to HCC. Here, we compared three well-established NASH-murine models [Western Diet plus Sugar Water (WD+SW), Choline Deficient High Fat Diet (CD-HFD) and Western Diet plus Carbon Tetrachloride (WD+CCl_4)] with human NASH-HCC and viral/alcohol-HCC at the transcriptomic level (for additional details on the pre-clinical models see the **Supplementary Data File**).

Submap analysis revealed that the WD+SW murine HCCs most closely resembled human non-cirrhotic NASH-HCCs (FDR = 0.07, **Figure 6A**). On the other hand, the WD+CCl_4 model appeared equally associated with cirrhotic and non-cirrhotic human NASH-HCCs (FDR=0.45 and FDR=0.35, respectively). With respect to the non-tumour tissue adjacent to murine HCC, the CD-HFD model was the only one associated with human non-cirrhotic NASH-HCC adjacent tissue, while the other models (WD+SW and WD+CCl_4) were associated with both, cirrhotic and non-cirrhotic human NASH-HCC adjacent tissue (**Figure 6A**).

Further analysis revealed that the WD+SW and CD-HFD models significantly recapitulated features observed in the human samples (**Figure 6B**). In terms of HCC molecular classes, HCCs in WD+SW and CH-HFD mice reproduced the heterogeneity of molecular and immune classes observed in human NASH-HCC (**Figure 6C**).

Discussion

Seminal studies have described the molecular pathogenesis of HCC primarily in viral-related tumours[1,7], while NASH aetiology has been underrepresented. Thus, a better understanding of the molecular features characterizing this type of HCCs and their comparison with NASH-HCC pre-clinical models and non-NASH human HCC is a major unmet need. Moreover, since a recent report suggests that NASH-HCC patients benefit less from immune checkpoint inhibitors than those with viral-related HCC, the extensive molecular characterization of this tumour aetiology is critical to optimize therapies boosting checkpoint blockers[14].

Here we report the mutational landscape of NASH-HCC, with *TERT promoter* (56%), *CTNNB1* (28%), *TP53* (18%) and *ACVR2A* (10%), as the most frequently mutated genes, and identified higher rates of *ACVR2A* mutations among NASH-HCCs compared to viral/alcohol-HCCs (10% vs 3%). *ACVR2A* is a cytokine receptor involved in cell differentiation and proliferation, reported as mutated in microsatellite-unstable colorectal cancers and whose downregulation is associated with poor outcomes[40]. Our *in vitro* results indicate that ACVR2A functions as a tumour suppressor, as previously reported in other malignancies[41,42], a feature that warrants validation in HCC animal models.

On the other hand, we hypothesized that NASH-related microenvironment could act as a liver genotoxic and trigger the generation of specific nucleotide substitutions captured as mutational signatures. We identified a pattern of mutations that could be explained by a non-previously described mutational signature, *MutSig-NASH-HCC,* which was present almost exclusively in NASH-HCCs (16% vs 2% in viral/alcohol-HCCs). This signature was associated with tumours developed in females, which

aligns with the fact that C>T transitions have been reported to occur more frequently in female HCC patients[44].

In terms of molecular classes, NASH-HCCs were enriched in the Wnt/TGF-β class and displayed a significantly lower prevalence of the CTNNB1 molecular subclass[32] compared with viral/alcohol-HCCs. These findings correlate with a recent study showing that HCCs in patients with metabolic syndrome were associated with absence of *CTNNB1* mutations[45]. From the signalling pathway perspective, we observed that NASH-HCCs were characterized by signatures related to bile and fatty acid metabolism, oxidative stress or inflammation, all features previously reported in human NASH and in NASH pre-clinical models[3]. Furthermore, NASH-HCCs were enriched in gene sets related to mTOR (involved in lipid biosynthesis[36]) and mitochondria (involved in lipid biosynthesis through the citrate cycle). Finally, higher mitochondrial activities have been reported to produce higher concentrations of ROS, and subsequent DNA damage, two features identified also herein and reported as an initial carcinogenic step.

In our study, the *PNPLA3* pathogenic variant in homozygosis was more prevalent among cirrhotic NASH-HCC patients (67% vs 17%), and its overall prevalence was consistent with the previously reported incidence in Western NAFLD-related HCC patients (29%)[46]. HCCs displaying the homozygous *PNPLA3* I148M variant were strongly associated with signatures of defective DNA repair, reduced TP53 signalling and oxidative stress, which might contribute to the development of liver carcinogenesis in patients with this polymorphism[4,47,48].

Finally, the comparison of our NASH-HCC cohort with three distinct NASH-HCC murine models revealed that they comprehensively recapitulate human NASH-HCC

molecular and immune traits and therefore are suitable for conducting pre-clinical studies.

In summary, our study provides novel insights that help clarify the pathogenesis of NASH-HCC and indicates that tumours arising in NASH are significantly associated with the Wnt/TGF-β subclass, present a higher prevalence of the potential tumour suppressor *ACVR2A*, and are associated with a new mutational signature that may point to unique genotoxic drivers.

Abbreviations

AFP	Alpha-fetoprotein
BMI	Body Mass Index
CD-HFD	Choline Deficient High Fat Diet
CNA	Copy Number Alteration
EMT	Epithelial to Mesenchymal Transition
FC	Fold Change
FDR	False Discovery Rate
FF	Fresh Frozen
FFPE	Formalin Fixed Paraffin-Embedded
FGF	Fibroblast Growth Factor
GSEA	Gene Set Enrichment Analysis
HBV	Hepatitis B virus infection
HCC	Hepatocellular carcinoma
HCV	Hepatitis C virus infection
ICF	Immune cancer field
IFN	Interferon
IGF	Insulin Growth Factor
IGV	Integrative Genomics Viewer
INR	International Normalised Ratio
IS-ICF	Immunosuppressive ICF
Mb	Megabase
MutSig	Mutational signature
NAFL	Non-alcoholic fatty liver
NAFLD	Non-alcoholic fatty liver disease

NASH	Non-alcoholic steatohepatitis
NF-κB	Nuclear Factor kappa B
NK cells	Natural Killer cells
PRO-ICF	Pro-inflammatory ICF
ROS	Reactive Oxygen Species
ssGSEA	single sample Gene Set Enrichment Analysis
TGF-β	Transforming Growth Factor beta
TMB	Tumour Mutational Burden
VAF	Variant Allele Frequency
WD+CCl_4	Western Diet plus carbon tetrachloride
WD+SW	Western Diet plus Sugar Water
WES	Whole Exome Sequencing

Acknowledgements

We thank Prof. Dr. Ferran Torres for assistance in statistical analysis. This study has been developed in part in the *Centre Esther Koplowitz* from IDIBAPS / CERCA Programme / Generalitat de Catalunya. Genomic analyses were run at the Functional Genomics Core Facility of IDIBAPS, at the New York Genome Center and at the Genomics Core Facility from the Icahn School of Medicine at Mount Sinai.

References

Author names in bold designate shared co-first authorship.

[1] Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, et al. Hepatocellular carcinoma. Nat Rev Dis Prim 2021;7:6.

[2] Kanwal F, Singal AG. Surveillance for hepatocellular carcinoma: current best practice and future direction. Gastroenterology 2019;157:54–64.

[3] Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease—Meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology 2016;64:73–84.

[4] Anstee QM, Reeves HL, Kotsiliti E, Govaere O, Heikenwalder M. From NASH to HCC: current concepts and future challenges. Nat Rev Gastroenterol Hepatol 2019;16:411–28.

[5] Villanueva A. Hepatocellular Carcinoma. N Engl J Med 2019;380:1450–62.

[6] Dudek M, Pfister D, Donakonda S, Filpe P, Schneider A, Laschinger M, et al. Auto-aggressive CXCR6+ CD8 T cells cause liver immune pathology in NASH. Nature 2021.

[7] Zucman-Rossi J, Villanueva A, Nault JC, Llovet JM. Genetic landscape and biomarkers of hepatocellular carcinoma. Gastroenterology 2015;149:1226–39.

[8] Fujiwara N, Friedman SL, Goossens N, Hoshida Y. Risk factors and prevention of hepatocellular carcinoma in the era of precision medicine. J Hepatol 2018;68:526–549.

[9] Huang DQ, El-Serag HB, Loomba R. Global epidemiology of NAFLD-related HCC: trends, predictions, risk factors and prevention. Nat Rev Gastroenterol Hepatol 2020;18;223-238.

[10] Grohmann M, Wiede F, Dodd GT, Gurzov EN, Ooi GJ, Butt T, et al. Obesity drives STAT-1-dependent NASH and STAT-3-dependent HCC. Cell 2018;175:1289-1306.e20.

[11] Liu D, Wong CC, Fu L, Chen H, Zhao L, Li C, et al. Squalene epoxidase drives NAFLD-induced hepatocellular carcinoma and is a pharmaceutical target. Sci Transl Med 2018;10:eaap9840.

[12] Gomes AL, Teijeiro A, Burén S, Theurillat J, Tummala KS, Yilmaz M, et al. Metabolic inflammation-associated IL-17A causes non-alcoholic steatohepatitis and hepatocellular carcinoma. Cancer Cell 2016;30:161–175.

[13] **Jühling F**, **Hamdane N**, **Crouchet E**, **Li S**, El Saghire H, Mukherji A, et al. Targeting clinical epigenetic reprogramming for chemoprevention of metabolic and viral hepatocellular carcinoma. Gut 2021;70:157-169.

[14] Pfister D, Núñez NG, Pinyol R, Govaere O, Pinter M, Szydlowska M, et al. NASH limits anti-tumour surveillance in immunotherapy-treated HCC. Nature 2021.

[15] **Schulze K**, **Imbeaud S**, **Letouzé E**, Alexandrov LB, Calderaro J, Rebouissou S, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nat Genet 2015;47:505-511.

[16] Bedossa P, Poitou C, Veyrie N, Bouillot JL, Basdevant A, Paradis V, et al. Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. Hepatology 2012;56:1751–9.

[17] Villanueva A, Portela A, Sayols S, Battiston C, Hoshida Y, Méndez-González J, et al. DNA methylation-based prognosis and epidrivers in hepatocellular carcinoma. Hepatology 2015;61:1–12.

[18] Ally A, Balasundaram M, Carlsen R, Chuah E, Clarke A, Dhalla N, et al. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. Cell 2017;169:1327-1341.e23.

[19] Ahn SM, Jang SJ, Shim JH, Kim D, Hong SM, Sung CO, et al. A genomic portrait of resectable hepatocellular carcinomas: Implications of RB1 and FGF19 aberrations for patient stratification. Hepatology 2014;60:1972-82.

[20] Matano M, Date S, Shimokawa M, Takano A, Fujii M, Ohta Y, et al. Modeling colorectal cancer using CRISPR-Cas9–mediated engineering of human intestinal organoids. Nat Med 2015;21:256–62.

[21] Uzilov AV, Taik P, Cheesman KC, Javanmard P, Ying K, Roehnelt A, et al. USP8 and TP53 drivers are associated with CNV in a corticotroph adenoma cohort enriched for aggressive tumors. J Clin Endocrinol Metab 2021;106:826-842.

[22] Kucab JE, Zou X, Morganella S, Joel M, Nanda AS, Nagy E, et al. A compendium of mutational signatures of environmental agents. Cell 2019;177:821-836.e16.

[23] Llovet JM, Montal R, Sia D, Finn R. Molecular therapies and precision medicine for hepatocellular carcinoma. Nat Rev Clin Oncol 2018;15:599-616.

[24] **Guichard C**, **Amaddeo G**, **Imbeaud S**, Ladeiro Y, Pelletier L, Maad I Ben, et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. Nat Genet 2012;44:694–8.

[25] **Kan Z**, **Zheng H**, **Liu X**, **Li S**, Barber TD, Gong Z, et al. Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. Genome Res 2013;23:1422-33.

[26] **Bailey MH**, **Tokheim C**, **Porta-Pardo E**, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive characterization of cancer driver genes and mutations. Cell 2018;173:371-385.e18.

[27] Trépo E, Nahon P, Bontempi G, Valenti L, Falleti E, Nischalke HD, 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.

[28] Coulouarn C, Factor VM, Thorgeirsson SS. Transforming growth factor-beta gene expression signature in mouse hepatocytes predicts clinical outcome in human cancer. Hepatology 2008;47:2059-67.

[29] Takeda H, Kataoka S, Nakayama M, Ali MAE, Oshima H, Yamamoto D, et al. CRISPR-Cas9-mediated gene knockout in intestinal tumor organoids provides functional validation for colorectal cancer driver genes. Proc Natl Acad Sci U S A 2019;116:15635–44.

[30] **Letouzé E**, **Shinde J**, Renault V, Couchy G, Blanc JF, Tubacher E, et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. Nat Commun 2017;8:1315.

[31] Degasperi A, Amarante TD, Czarnecki J, Shooter S, Zou X, Glodzik D, et al. A practical framework and online tool for mutational signature analyses show intertissue variation and driver dependencies. Nat Cancer 2020;1:249-263.

[32] Chiang DY, Villanueva A, Hoshida Y, Peix J, Newell P, Minguez B, et al. Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. Cancer Res 2008;68:6779–88.

[33] Villanueva A, Hoshida Y, Battiston C, Tovar V, Sia D, Mazzaferro V, et al. Combining clinical, pathology, and gene expression data to predict recurrence of hepatocellular carcinoma. Gastroenterology 2011;140:1501-1512.e2.

[34] Sia D, Jiao Y, Martinez-Quetglas I, Kuchuk O, Villacorta-Martin C, Castro de Moura M, et al. Identification of an immune-specific class of hepatocellular carcinoma, based on molecular features. Gastroenterology 2017;153:812–26.

[35] Pinyol R, Sia D, Llovet JM. Immune exclusion-Wnt/CTNNB1 class predicts resistance to immunotherapies in HCC. Clin Cancer Res 2019;25:2021–3.

[36] Guri Y, Colombi M, Dazert E, Hindupur SK, Roszik J, Moes S, et al. mTORC2 promotes tumorigenesis via lipid synthesis. Cancer Cell 2017;32:807-823.e12.

[37] Li T, Weng J, Zhang Y, Liang K, Fu G, Li Y, et al. mTOR direct crosstalk with STAT5 promotes de novo lipid synthesis and induces hepatocellular carcinoma. Cell Death Dis 2019;10:619.

[38] Moeini A, Torrecilla S, Tovar V, Montironi C, Andreu-Oller C, Peix J, et al. An immune gene expression signature associated with development of human hepatocellular carcinoma identifies mice that respond to chemopreventive agents. Gastroenterology 2019;157:1383-1397.e11.

[39] Hoshida Y, Villanueva A, Sangiovanni A, Sole M, Hur C, Andersson KL, et al. Prognostic gene expression signature for patients with hepatitis C-related early-stage cirrhosis. Gastroenterology 2013;144:1024–30.

[40] Jung B, Doctolero RT, Tajima A, Nguyen AK, Keku T, Sandler RS, et al. Loss of activin receptor type 2 protein expression in microsatellite unstable colon cancers. Gastroenterology 2004;126:64–659.

[41] Jeruss JS, Sturgis CD, Rademaker AW, Woodruff TK. Down-regulation of activin, activin receptors, and Smads in high-grade breast cancer. Cancer Res 2003;62:497–505.

[42] Rossi MR, Ionov Y, Bakin A V., Cowell JK. Truncating mutations in the ACVR2 gene attenuates activin signaling in prostate cancer cells. Cancer Genet Cytogenet 2005;163:123–9.

[43] Li CH, Prokopec SD, Sun RX, Yousif F, Schmitz N, Al-Shahrour F, et al. Sex differences in oncogenic mutational processes. Nat Commun 2020;11:4330.

[44] Shimada S, Mogushi K, Akiyama Y, Furuyama T, Watanabe S, Ogura T, et al. Comprehensive molecular and immunological characterization of hepatocellular carcinoma. EBioMedicine 2019;40:457–70.

[45] Liu YL, Patman GL, Leathart JBS, Piguet AC, Burt AD, Dufour JF, et al. Carriage of the PNPLA3 rs738409 C>G polymorphism confers an increased risk of non-alcoholic fatty liver disease associated hepatocellular carcinoma. J Hepatol 2014;61:75–81.

[46] Teoh NC, Yock YD, Swisshelm K, Lehman S, Wright JH, Haque J, et al. Defective DNA strand break repair causes chromosomal instability and accelerates liver carcinogenesis in mice. Hepatology 2008;47:2078–88.

[47] Limagne E, Cottet V, Cotte AK, Hamza S, Hillon P, Latruffe N, et al. Potential role of oxidative DNA damage in the impact of PNPLA3 variant (rs 738409 C>G) in

hepatocellular carcinoma risk. Hepatology 2014;60:1110–1.

Figure Legends

Fig. 1. Genomic landscape of NASH-HCC and *in vitro* evidences supporting a tumour suppressor role of *ACVR2A* in HCC. (A) Mutations and focal copy-number alterations in driver genes altered in ≥4% of the NASH-HCC cohort. (B) Mutational frequency of the most commonly altered genes in the NASH-HCC cohort (n=50) and in the viral/alcohol-HCC cohort (n=624)[15,18,19]. Statistical test: Fisher. (C) Genomic and clinico-pathological features of NASH-HCC according to cirrhosis. Statistical test: Fisher and Mann-Whitney. (D) *ACVR2A* mutations identified in the NASH-HCC cohort. (E, F) Cell viability rate (E) and colony formation quantification (F) of Hep3B cells stably transfected with *ACVR2A*- or control-shRNA. Error bars represent mean ± SEM of ≥ 3 experiments performed in triplicate. Statistical test: t-test. (G) Representative image of the colony formation assay.

Fig. 2. Mutational signatures in NASH-HCC and in viral/alcohol-HCC. (A) Unsupervised hierarchical clustering of the mutational signatures obtained for 43 NASH-HCCs and 43 viral/alcohol-HCCs. Red asterisks mark samples where *MutSig-NASH-HCC* presented an exposure >20% when setting the confidence at 90%. (B) Heatmap with clinico-pathological data, mutational status of *CTNNB1* and *TP53*, and mutational signatures (confidence >90%, exposure >20%). Statistical test: Fisher and Mann-Whitney.

Fig. 3. Signalling pathways altered in NASH-HCC. (A) Driver genomic alterations identified by WES grouped according to signalling pathways. (B) Heatmap displaying differentially enriched pathways in NASH-HCCs (n=53) compared to viral/alcohol-HCCs (n=184). Statistical test: t-test. (C) Molecular classes and activated signalling

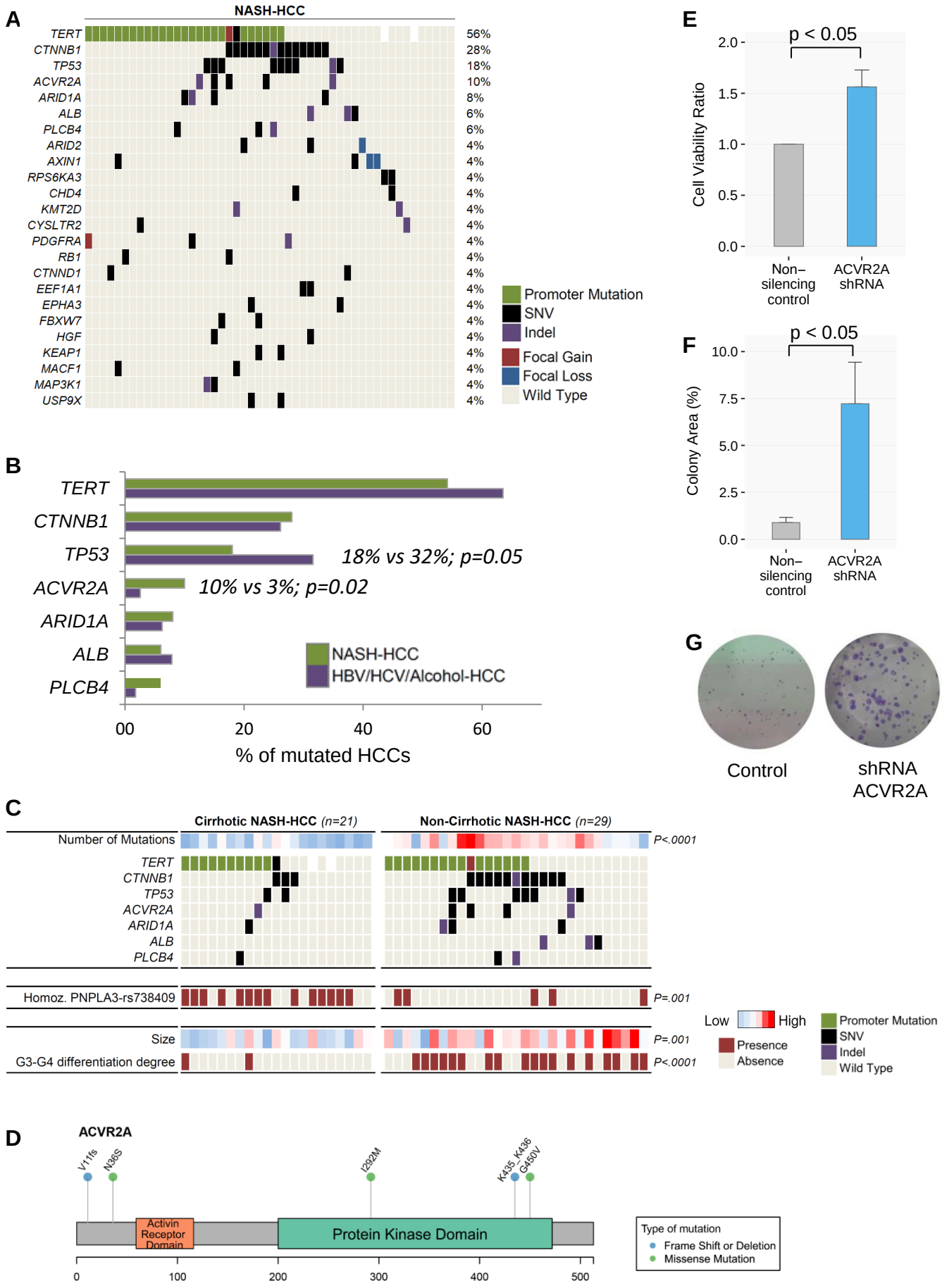
pathways in the NASH-HCC cohort. Samples were classified into proliferative (S1/S2) and non-proliferative tumours (S3). Statistical test: t-test and Fisher. Displayed p values were obtained comparing proliferation and non-proliferation HCCs. Gene signatures were obtained from MSigDB or other sources (see Supplementary Data File).

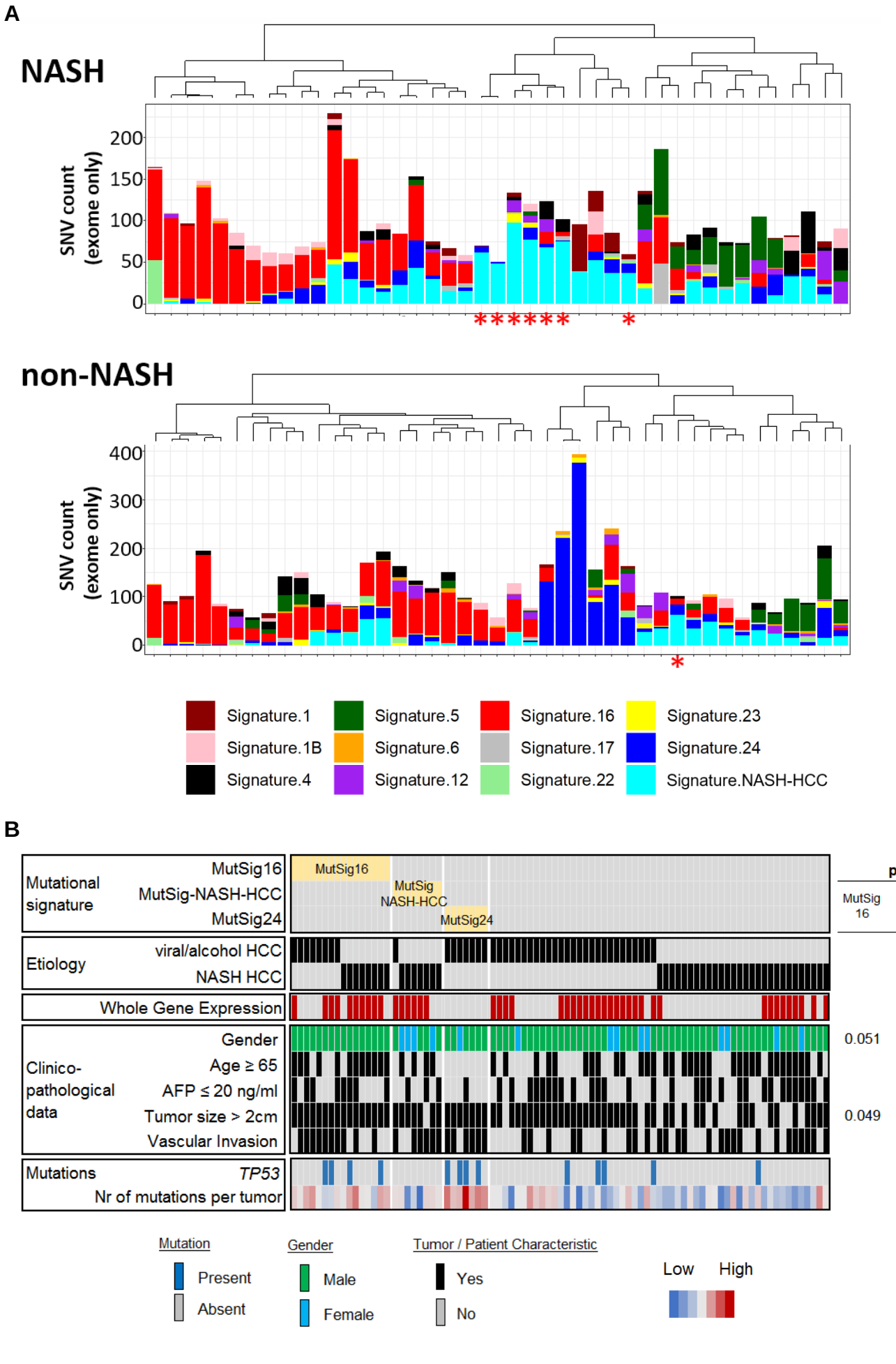
Fig. 4. Characterization of the NASH-HCCs according to HCC immune classes and signalling pathways differentiating cirrhotic from non-cirrhotic NASH-HCC. (A) Heatmap displaying NASH-HCC tumours classified according to the HCC immune classes[34,35]. Gene signatures used are referenced in the Supplementary Data File. Statistical test: t-test. (B, C) Pre-ranked GSEA enrichment plots of representative signalling pathways or molecular classes enriched in non-cirrhotic (B, n=16) and cirrhotic NASH tumours (C, n=37).

Fig. 5. Characterization of the NASH cancer field. (A) Heatmap characterizing the cancer field in NASH livers and NASH-HCC adjacent tissues. Plotted are ssGSEA scores for NASH-related gene sets. T-test p values report differences between cirrhotic and non-cirrhotic samples. Healthy liver (H). Cirrhotic liver (Ci). NASH liver from patients with no HCC (NASH). Non-tumorous tissue adjacent to NASH-HCC (NASH-HCC adjacent). (B) Heatmap displaying ssGSEA scores of immune signatures capturing different immune cell populations. Gene signatures referenced in the Supplementary Data File. Statistical test: t-test.

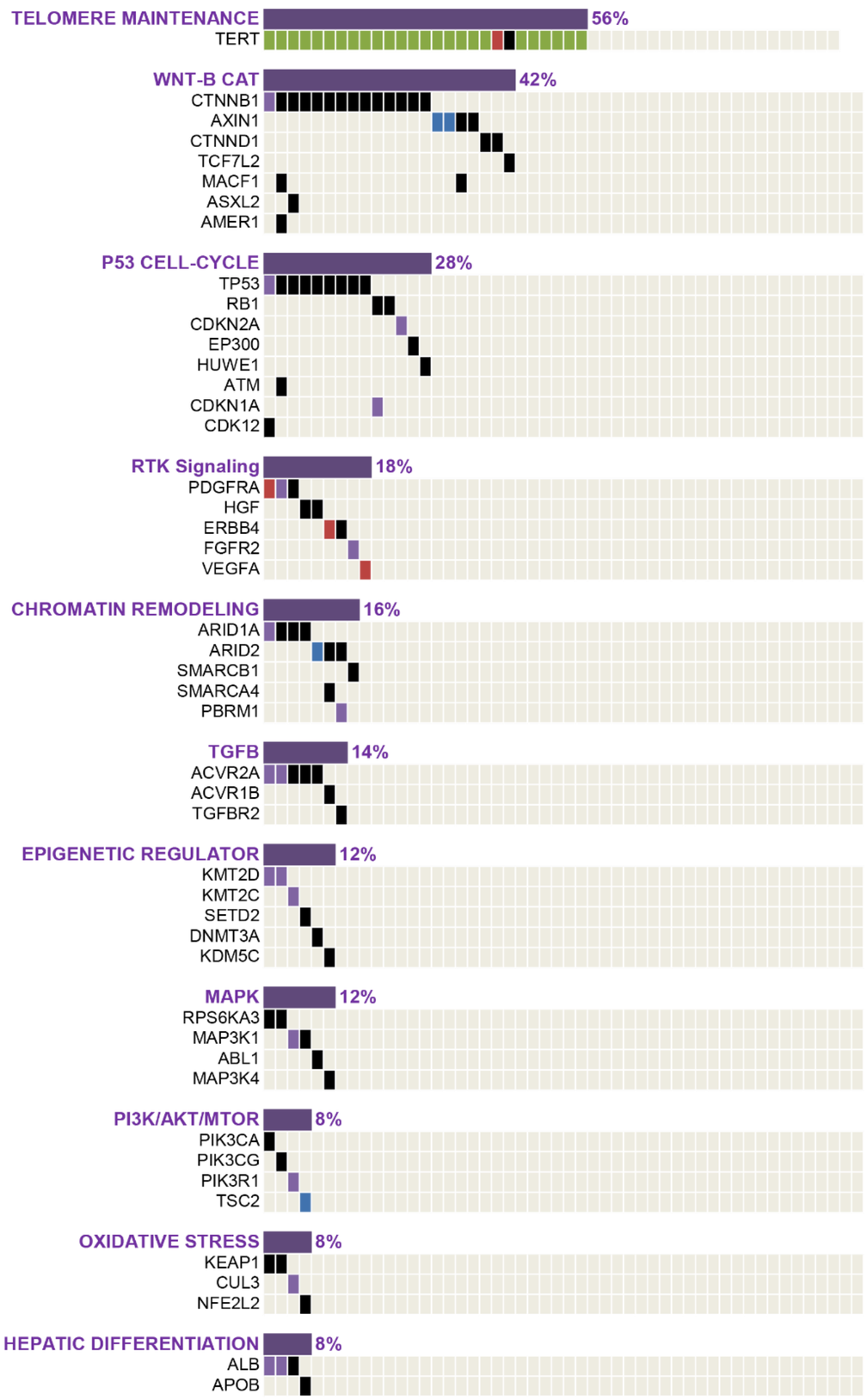
Fig. 6. NASH-HCC murine models recapitulate key molecular and immune features of human NASH-HCC. (A) Submap analysis displaying the molecular

similarity between human and murine NASH-HCC and adjacent tissue samples. Numbers on heatmap indicate FDR values for transcriptome similarity. (B) Heatmap displaying enrichment of fatty and bile acid metabolism, oxidative stress and inflammation-related gene signatures in NASH-HCC vs non-NASH HCC. Statistical test: t-test. (C) NASH-HCC murine and human samples classified according to the HCC molecular and immune classes.

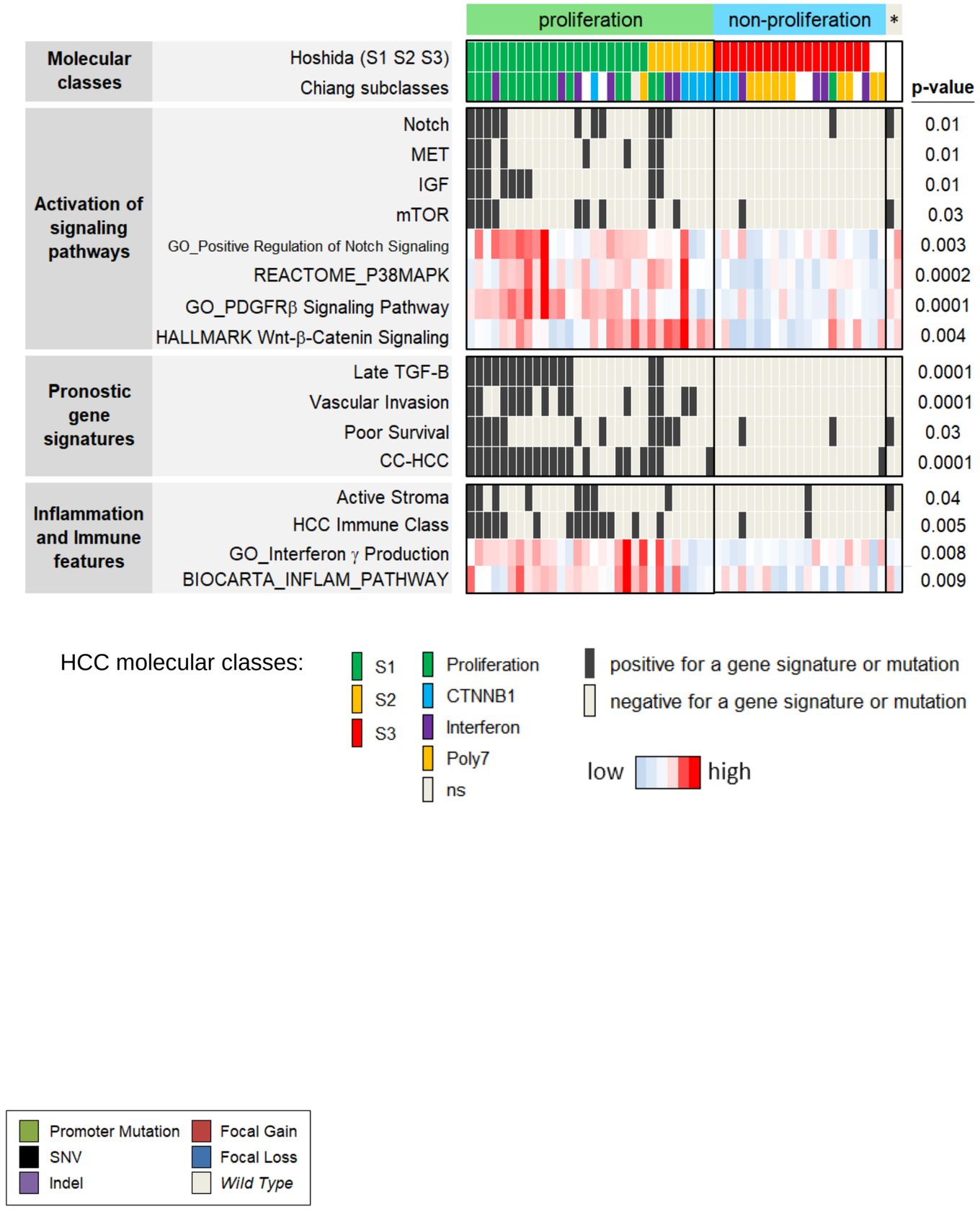

A
NASH-HCC
TERT 56%
CTNNB1 28%
TP53 18%
ACVR2A 10%
ARID1A 8%
ALB 6%
PLCB4 6%
ARID2 4%
AXIN1 4%
RPS6KA3 4%
CHD4 4%
KMT2D 4%
CYSLTR2 4%
PDGFRA 4%
RB1 4%
CTNND1 4%
EEF1A1 4%
EPHA3 4%
FBXW7 4%
HGF 4%
KEAP1 4%
MACF1 4%
MAP3K1 4%
USP9X 4%
Promoter Mutation
SNV
Indel
Focal Gain
Focal Loss
Wild Type
B
TERT
CTNNB1
TP53
ACVR2A
ARID1A
ALB
PLCB4
18% vs 32%; p=0.05
10% vs 3%; p=0.02
NASH-HCC
HBV/HCV/Alcohol-HCC
00
20
40
60
% of mutated HCCs
E
p < 0.05
Cell Viability Ratio
2.0
1.5
1.0
0.5
0.0
Non–silencing control
ACVR2A shRNA
F
p < 0.05
Colony Area (%)
10.0
7.5
5.0
2.5
0.0
Non–silencing control
ACVR2A shRNA
G
Control
shRNA ACVR2A
C
Cirrhotic NASH-HCC (n=21)
Non-Cirrhotic NASH-HCC (n=29)
Number of Mutations
P<.0001
TERT
CTNNB1
TP53
ACVR2A
ARID1A
ALB
PLCB4
Homoz. PNPLA3-rs738409
P=.001
Size
P=.001
G3-G4 differentiation degree
P<.0001
Low
High
Presence
Absence
Promoter Mutation
SNV
Indel
Wild Type
D
ACVR2A
V11fs
N36S
I292M
K435_K436
G450V
Activin Receptor Domain
Protein Kinase Domain
0
100
200
300
400
500
Type of mutation
Frame Shift or Deletion
Missense Mutation


A
NASH
SNV count (exome only)
200
150
100
50
0
******
*
non-NASH
SNV count (exome only)
400
300
200
100
0
*
Signature.1
Signature.1B
Signature.4
Signature.5
Signature.6
Signature.12
Signature.16
Signature.17
Signature.22
Signature.23
Signature.24
Signature.NASH-HCC
B
Mutational signature
MutSig16
MutSig-NASH-HCC
MutSig24
MutSig16
MutSig NASH-HCC
MutSig24
Etiology
viral/alcohol HCC
NASH HCC
Whole Gene Expression
Clinico-pathological data
Gender
Age ≥ 65
AFP ≤ 20 ng/ml
Tumor size > 2cm
Vascular Invasion
Mutations
TP53
Nr of mutations per tumor
p-values
MutSig 16
MutSig-NASH-HCC
MutSig 24
0.03
0.006
0.051
0.007
0.001
0.03
0.049
0.012
0.001
3E-09
Mutation
Present
Absent
Gender
Male
Female
Tumor / Patient Characteristic
Yes
No
Low
High

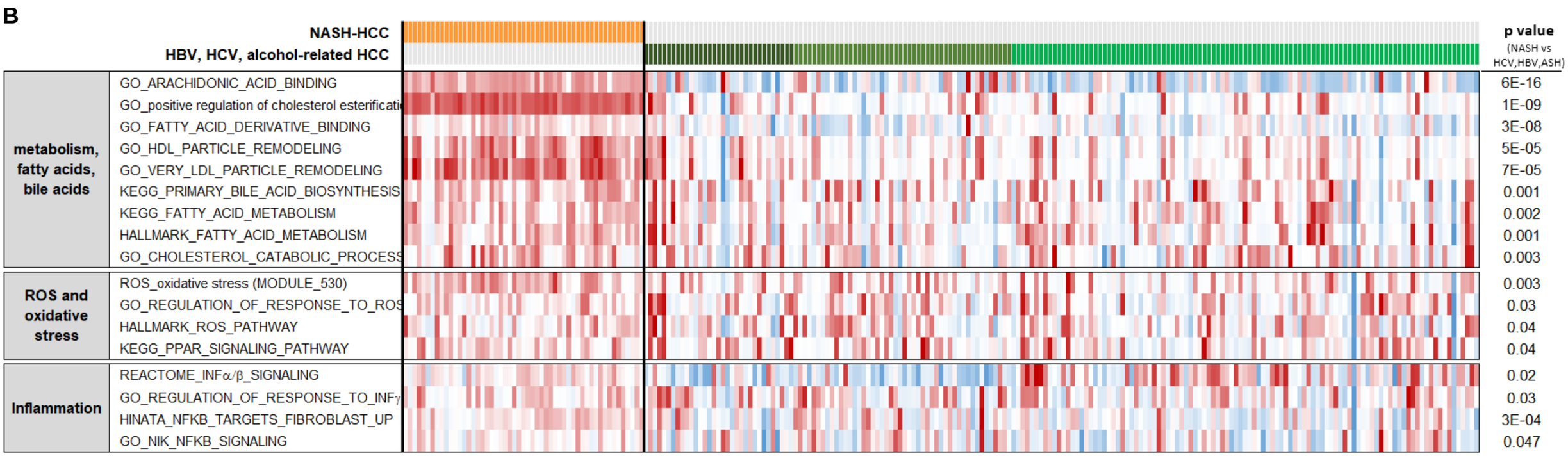
A


TELOMERE MAINTENANCE 56%
TERT
WNT-B CAT 42%
CTNNB1
AXIN1
CTNND1
TCF7L2
MACF1
ASXL2
AMER1
P53 CELL-CYCLE 28%
TP53
RB1
CDKN2A
EP300
HUWE1
ATM
CDKN1A
CDK12
RTK Signaling 18%
PDGFRA
HGF
ERBB4
FGFR2
VEGFA
CHROMATIN REMODELING 16%
ARID1A
ARID2
SMARCB1
SMARCA4
PBRM1
TGFB 14%
ACVR2A
ACVR1B
TGFBR2
EPIGENETIC REGULATOR 12%
KMT2D
KMT2C
SETD2
DNMT3A
KDM5C
MAPK 12%
RPS6KA3
MAP3K1
ABL1
MAP3K4
PI3K/AKT/MTOR 8%
PIK3CA
PIK3CG
PIK3R1
TSC2
OXIDATIVE STRESS 8%
KEAP1
CUL3
NFE2L2
HEPATIC DIFFERENTIATION 8%
ALB
APOB

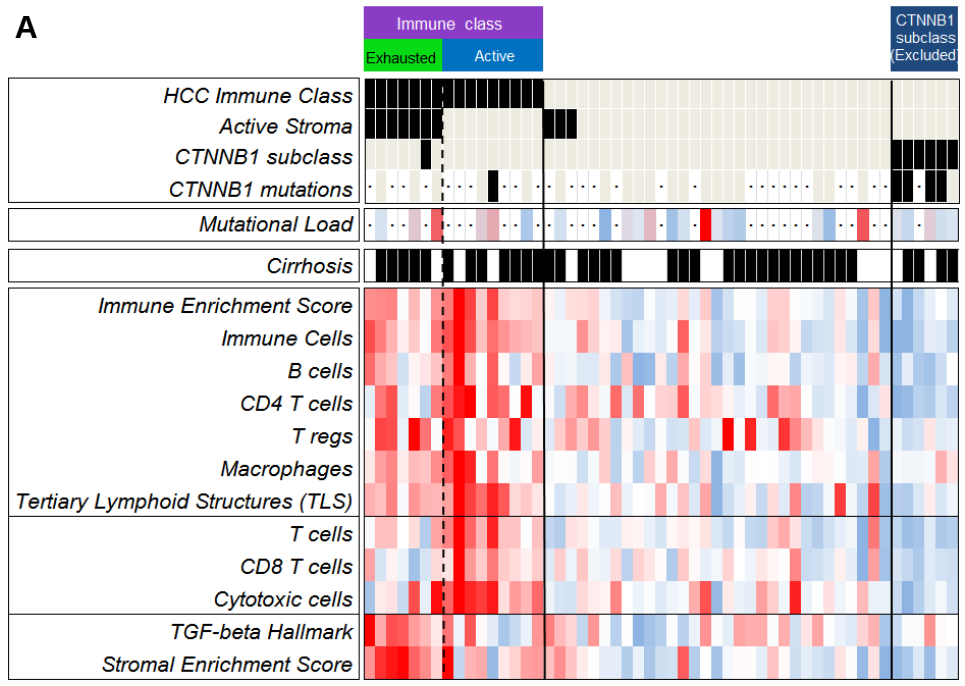
C


proliferation
non-proliferation
*
Molecular classes
Hoshida (S1 S2 S3)
Chiang subclasses
p-value
Activation of signaling pathways
Notch 0.01
MET 0.01
IGF 0.01
mTOR 0.03
GO_Positive Regulation of Notch Signaling 0.003
REACTOME_P38MAPK 0.0002
GO_PDGFRβ Signaling Pathway 0.0001
HALLMARK Wnt-β-Catenin Signaling 0.004
Pronostic gene signatures
Late TGF-B 0.0001
Vascular Invasion 0.0001
Poor Survival 0.03
CC-HCC 0.0001
Inflammation and Immune features
Active Stroma 0.04
HCC Immune Class 0.005
GO_Interferon γ Production 0.008
BIOCARTA_INFLAM_PATHWAY 0.009
HCC molecular classes:
S1
S2
S3
Proliferation
CTNNB1
Interferon
Poly7
ns
positive for a gene signature or mutation
negative for a gene signature or mutation
low
high
Promoter Mutation
SNV
Indel
Focal Gain
Focal Loss
Wild Type

B


NASH-HCC
HBV, HCV, alcohol-related HCC
p value
(NASH vs HCV,HBV,ASH)
metabolism, fatty acids, bile acids
GO_ARACHIDONIC_ACID_BINDING 6E-16
GO_positive regulation of cholesterol esterificati 1E-09
GO_FATTY_ACID_DERIVATIVE_BINDING 3E-08
GO_HDL_PARTICLE_REMODELING 5E-05
GO_VERY_LDL_PARTICLE_REMODELING 7E-05
KEGG_PRIMARY_BILE_ACID_BIOSYNTHESIS 0.001
KEGG_FATTY_ACID_METABOLISM 0.002
HALLMARK_FATTY_ACID_METABOLISM 0.001
GO_CHOLESTEROL_CATABOLIC_PROCESS 0.003
ROS and oxidative stress
ROS_oxidative stress (MODULE_530) 0.003
GO_REGULATION_OF_RESPONSE_TO_ROS 0.03
HALLMARK_ROS_PATHWAY 0.04
KEGG_PPAR_SIGNALING_PATHWAY 0.04
Inflammation
REACTOME_INFα/β_SIGNALING 0.02
GO_REGULATION_OF_RESPONSE_TO_INFγ 0.03
HINATA_NFKB_TARGETS_FIBROBLAST_UP 3E-04
GO_NIK_NFKB_SIGNALING 0.047

A



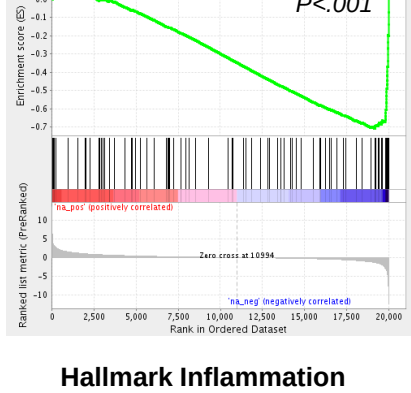
Low — High Enrichment Score / Mutational Load

	p value (Immune vs rest)	p value *(Active vs Exh.)*	p value *(Excl. vs rest)*
Immune Enrichment Score	*<0.001*	*ns*	*0.02*
Immune Cells	*<0.001*	*ns*	*0.002*
B cells	*<0.001*	*ns*	*0.05*
CD4 T cells	*0.04*	*ns*	*<0.001*
T regs	*0.01*	*ns*	*0.07*
Macrophages	*<0.001*	*ns*	*ns*
Tertiary Lymphoid Structures (TLS)	*<0.001*	*ns*	*0.04*
T cells	*<0.001*	*0.03*	*0.01*
CD8 T cells	*0.02*	*0.03*	*0.03*
Cytotoxic cells	*0.005*	*0.03*	*0.002*
TGF-beta Hallmark	*ns*	*0.03*	*ns*
Stromal Enrichment Score	*<0.001*	*0.06*	*0.02*

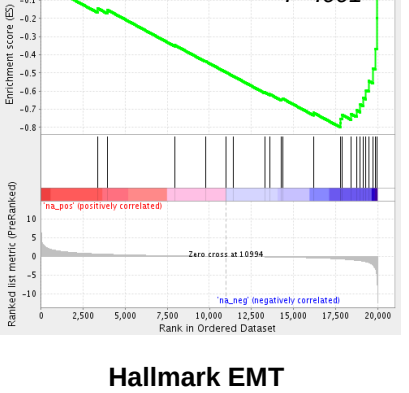
B

Non-cirrhotic NASH-HCC

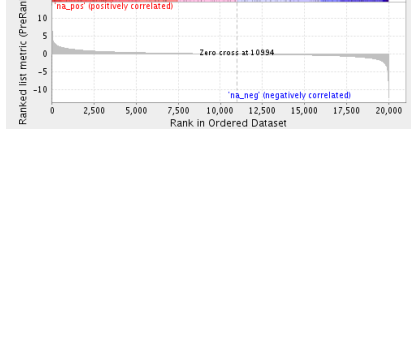
S2 HCC molecular class



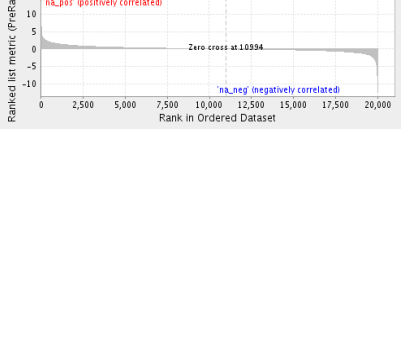
Hallmark E2F Targets

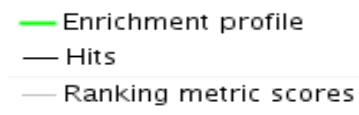


HCC proliferation



Hallmark MYC targets





C

Cirrhotic NASH-HCC

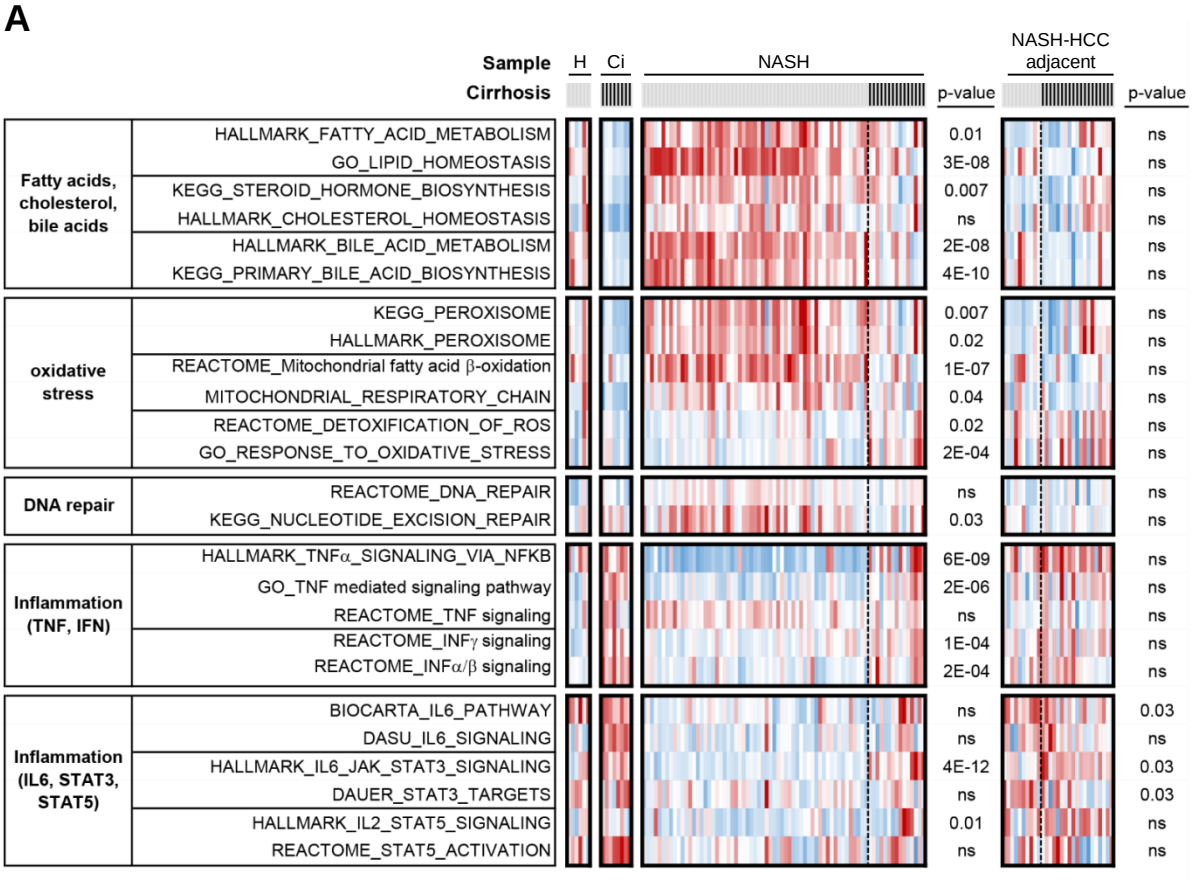
HCC immune class

Activated stroma

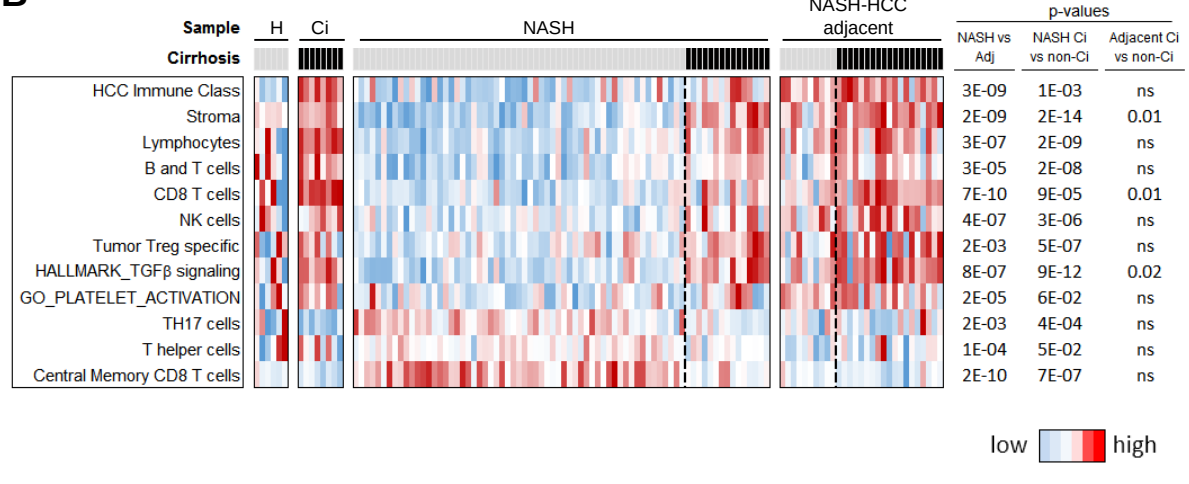
Hallmark Inflammation

Hallmark EMT

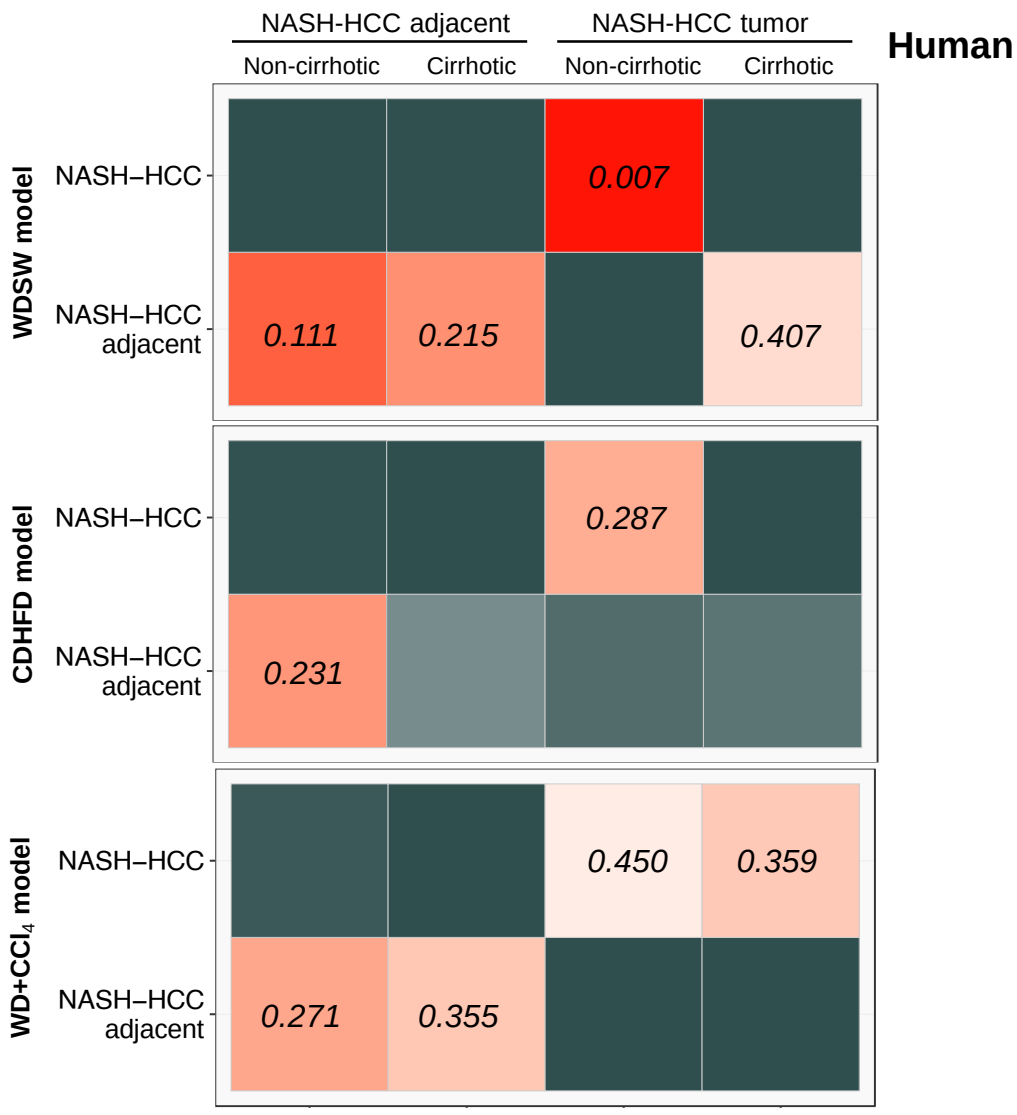
A


Sample
Cirrhosis
H
Ci
NASH
NASH-HCC adjacent
p-value
p-value
Fatty acids, cholesterol, bile acids
HALLMARK_FATTY_ACID_METABOLISM 0.01 ns
GO_LIPID_HOMEOSTASIS 3E-08 ns
KEGG_STEROID_HORMONE_BIOSYNTHESIS 0.007 ns
HALLMARK_CHOLESTEROL_HOMEOSTASIS ns ns
HALLMARK_BILE_ACID_METABOLISM 2E-08 ns
KEGG_PRIMARY_BILE_ACID_BIOSYNTHESIS 4E-10 ns
oxidative stress
KEGG_PEROXISOME 0.007 ns
HALLMARK_PEROXISOME 0.02 ns
REACTOME_Mitochondrial fatty acid β-oxidation 1E-07 ns
MITOCHONDRIAL_RESPIRATORY_CHAIN 0.04 ns
REACTOME_DETOXIFICATION_OF_ROS 0.02 ns
GO_RESPONSE_TO_OXIDATIVE_STRESS 2E-04 ns
DNA repair
REACTOME_DNA_REPAIR ns ns
KEGG_NUCLEOTIDE_EXCISION_REPAIR 0.03 ns
Inflammation (TNF, IFN)
HALLMARK_TNFα_SIGNALING_VIA_NFKB 6E-09 ns
GO_TNF mediated signaling pathway 2E-06 ns
REACTOME_TNF signaling ns ns
REACTOME_INFγ signaling 1E-04 ns
REACTOME_INFα/β signaling 2E-04 ns
Inflammation (IL6, STAT3, STAT5)
BIOCARTA_IL6_PATHWAY ns 0.03
DASU_IL6_SIGNALING ns ns
HALLMARK_IL6_JAK_STAT3_SIGNALING 4E-12 0.03
DAUER_STAT3_TARGETS ns 0.03
HALLMARK_IL2_STAT5_SIGNALING 0.01 ns
REACTOME_STAT5_ACTIVATION ns ns

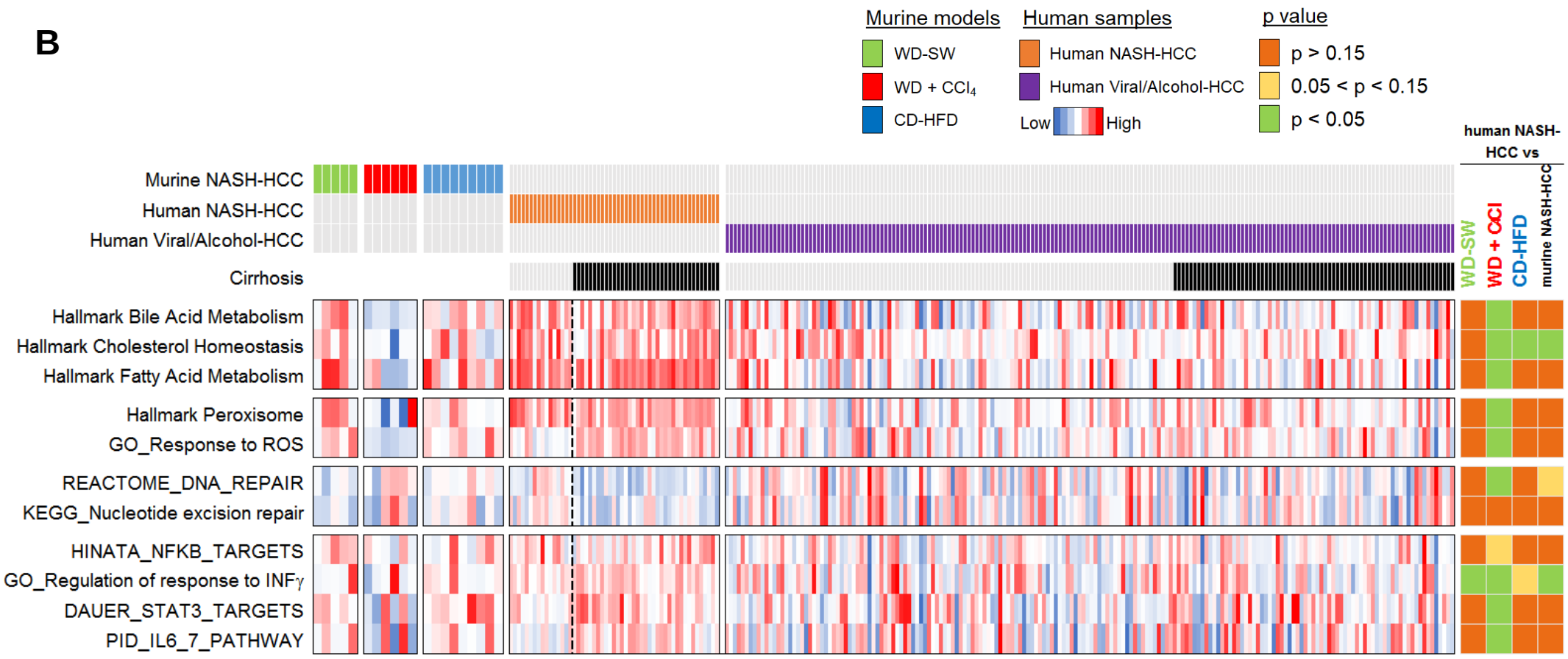
B


Sample
Cirrhosis
H
Ci
NASH
NASH-HCC adjacent
p-values
NASH vs Adj
NASH Ci vs non-Ci
Adjacent Ci vs non-Ci
HCC Immune Class 3E-09 1E-03 ns
Stroma 2E-09 2E-14 0.01
Lymphocytes 3E-07 2E-09 ns
B and T cells 3E-05 2E-08 ns
CD8 T cells 7E-10 9E-05 0.01
NK cells 4E-07 3E-06 ns
Tumor Treg specific 2E-03 5E-07 ns
HALLMARK_TGFβ signaling 8E-07 9E-12 0.02
GO_PLATELET_ACTIVATION 2E-05 6E-02 ns
TH17 cells 2E-03 4E-04 ns
T helper cells 1E-04 5E-02 ns
Central Memory CD8 T cells 2E-10 7E-07 ns
low
high

A


NASH-HCC adjacent
NASH-HCC tumor
Non-cirrhotic
Cirrhotic
Non-cirrhotic
Cirrhotic
Human
FDR
1.00
0.75
0.50
0.25
0.00
WDSW model
NASH–HCC
NASH–HCC adjacent
0.007
0.111
0.215
0.407
CDHFD model
NASH–HCC
NASH–HCC adjacent
0.287
0.231
WD+CCl4 model
NASH–HCC
NASH–HCC adjacent
0.450
0.359
0.271
0.355

B


Murine models
WD-SW
WD + CCl4
CD-HFD
Human samples
Human NASH-HCC
Human Viral/Alcohol-HCC
Low
High
p value
p > 0.15
0.05 < p < 0.15
p < 0.05
human NASH-HCC vs
WD-SW
WD + CCl
CD-HFD
murine NASH-HCC
Murine NASH-HCC
Human NASH-HCC
Human Viral/Alcohol-HCC
Cirrhosis
Hallmark Bile Acid Metabolism
Hallmark Cholesterol Homeostasis
Hallmark Fatty Acid Metabolism
Hallmark Peroxisome
GO_Response to ROS
REACTOME_DNA_REPAIR
KEGG_Nucleotide excision repair
HINATA_NFKB_TARGETS
GO_Regulation of response to INFγ
DAUER_STAT3_TARGETS
PID_IL6_7_PATHWAY

C

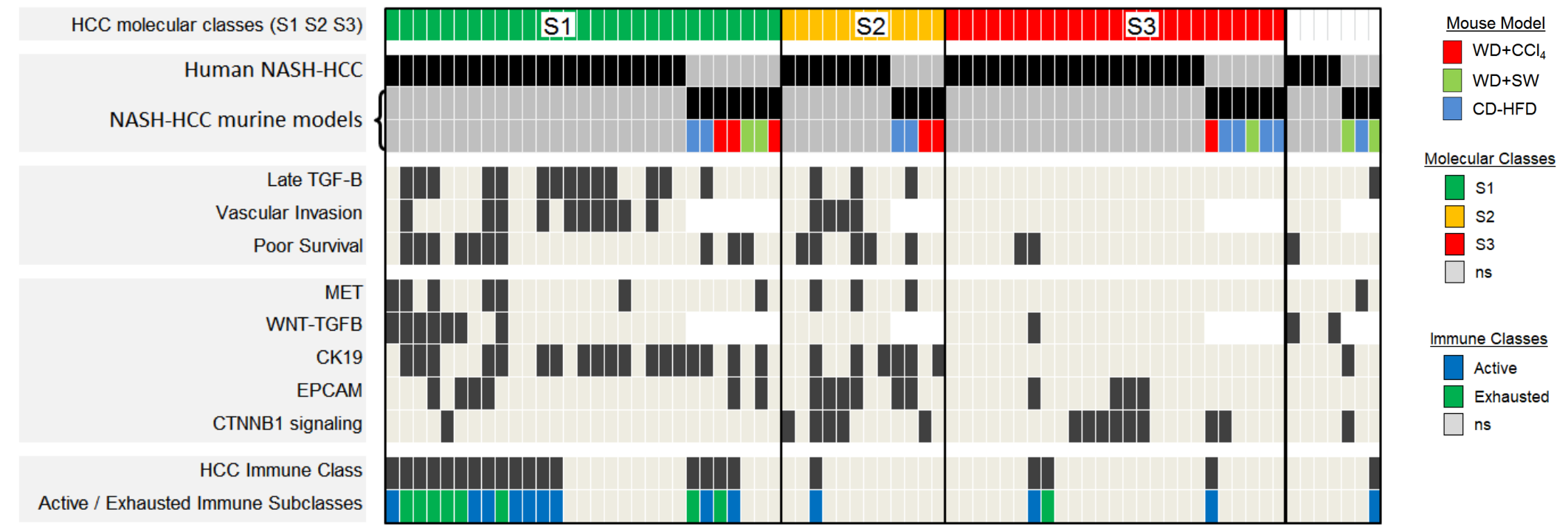

HCC molecular classes (S1 S2 S3)
S1
S2
S3
Human NASH-HCC
NASH-HCC murine models
Late TGF-B
Vascular Invasion
Poor Survival
MET
WNT-TGFB
CK19
EPCAM
CTNNB1 signaling
HCC Immune Class
Active / Exhausted Immune Subclasses
Mouse Model
WD+CCl4
WD+SW
CD-HFD
Molecular Classes
S1
S2
S3
ns
Immune Classes
Active
Exhausted
ns

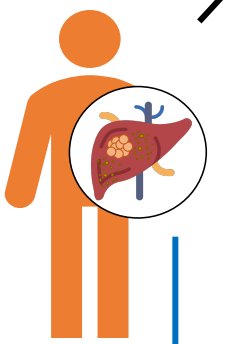
Table 1. Clinicopathological characteristics of NASH-HCC and NASH cohorts

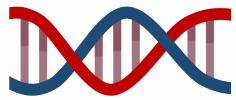
	NASH-HCC (n=80)*	NASH (n=125)*‡	p-value
Age (years), median (range)	65.8 (50.1 - 90.6)	56.1 (18.2 - 81.6)	*< 0.001*
≥ 65, n (%)	47 (58.8)	24 (19.4)	
Gender, n (%)			
Male	65 (81.3)	52 (41.9)	*< 0.001*
Female	15 (18.8)	72 (58.1)	
Race, n (%)			
Asian	6 (7.7)	10 (11.6)	*ns*
White	70 (89.7)	71 (82.6)	*ns*
Other	2 (2.6)	5 (5.8)	*ns*
Hispanic	12 (15.2)	34 (39.1)	*< 0.001*
Etiology, n (%)			
NASH	100%	100%	*ns*
Hypertension, n (%)	61 (80.3)	60 (52.1)	*< 0.001*
Diabetes Mellitus, n (%)	55 (72.4)	58 (50.4)	*0.003*
Body Mass Index (BMI, Kg/m^2), median (range)	29.6 (20.6-41.0)	31.0 (18.4-64.0)	*ns*
Obesity, n (%)	41 (55.4)	70 (61.4)	*ns*
Hyperlipidemia, n (%)	36 (53.7)	67 (57.8)	*ns*
Triglycerides (mg/dl), median (range)	107 (34 - 242)	135 (29 - 433)	*0.008*
Cholesterol (mg/dl), median (range)	153.5 (85 - 252)	189 (40 - 415)	*0.003*
Albumin (g/dl), median (range)	3.7 (2.2 - 5)	4.2 (1.8 - 5.3)	*0.001*
Total bilirubin (mg/dl), median (range)	0.95 (0.3 - 18.3)	0.7 (0.3 - 41.5)	*ns*
Platelet count (10^3/ml), median (range)	111 (24 - 460)	196 (21.5 - 494)	*< 0.001*
< 100,000/ml, n (%)	39 (52.7)	91 (82)	
INR, median (range)	1.2 (0.9 - 3)	1.02 (0.8 - 3.8)	*ns*
Fibrosis¥, n (%)			
F0, F1, F2	12 (15)	74 (59.2)	*< 0.001*
F3, F4	68 (85)	51 (40.8)	
Liver cirrhosis, n (%)	56 (70)	36 (28.8)	*< 0.001*
Degree of steatohepatitis, n (%)			
G1	46 (82.1)	68 (64.8)	*0.03*
G2	6 (10.7)	33 (31.4)	*0.004*
G3	4 (7.1)	4 (3.8)	*ns*
Steatosis, n (%)			
0	12 (15.2)	18 (17.8)	*ns*
1	55 (69.6)	52 (51.5)	*0.02*
2	12 (15.2)	31 (30.7)	*0.02*
Ballooning, n (%)			
0	15 (18.8)	15 (12)	*ns*
1	40 (50)	74 (59.2)	*ns*
2	25 (31.3)	36 (28.8)	*ns*
Lobular inflammation, n (%)			
0	15 (18.8)	14 (11.2)	*ns*
1	50 (62.5)	70 (56)	*ns*
2	10 (12.5)	34 (27.2)	*0.01*
3	5 (6.3)	7 (5.6)	*ns*
Portal inflammation, n (%)			
None	1 (1.3)	15 (12)	*0.006*
Mild	41 (51.3)	68 (54.4)	*ns*
Moderate	35 (43.8)	41 (32.8)	*ns*
Severe	3 (3.8)	1 (0.8)	*ns*
NAFLD Activity Score (NAS), n (%)			
0-2 (non-NASH)	21 (26.3)	19 (15.2)	*ns*
3-4 (borderline NASH)	40 (50)	56 (44.8)	*ns*
5-8 (NASH)	19 (23.8)	50 (40)	*0.02*
Burn out NASH, n (%)	21 (26.3)	20 (16)	*ns*
Child-Pugh, n (%)			
A	22 (47.8)	1 (20)	*ns*
B	18 (39.1)	3 (60)	*ns*
C	6 (13.0)	1 (20)	*ns*
BCLC stage, n (%)			

Very early stage (0)	1 (1.8)	
Early stage (A)	27 (49.1)	
Intermediate stage (B)	22 (40)	
Advanced stage (C)	5 (9.1)	
Number of nodules, n (%)		
1	38 (55.9)	
2	10 (14.7)	
≥ 3	20 (29.4)	
AFP (ng/ml), median (range)	4.5 (0.9 - 516)	
Maximum tumour size (cm), median (range)	2.7 (1 - 19)	
Satellites, n (%)	23 (41.8)	
Tumour differentiation[§], n (%)		
G1	11 (13.8)	
G2	55 (68.8)	
G3	14 (17.5)	
G4	0 (0)	
Vascular Invasion[#], n (%)		
Microvascular invasion	33 (41.3)	
Macrovascular invasion	6 (10.2)	
Events, n (%)		
Recurrence	15 (38.5)	
Death	43 (53)	19 (16.5)
Median Follow-Up, years	5.2	4.2

[¥]Metavir-Score. [§]Edmondson–Steiner Grade. [#]Macro and microvascular invasion were respectively obtained from the pathological report and through the histological examination. In the NASH-HCC cohort (n=80), 51% of the samples were from resected patients and 48% from transplanted patients. The remaining 1% were biopsies. In the NASH cohort (n=125), 82% of the samples were biopsies, while the remaining 18% were from liver transplanted patients. *Not all variables could be collected for all patients. [‡]Two NASH patients developed HCC during follow-up.

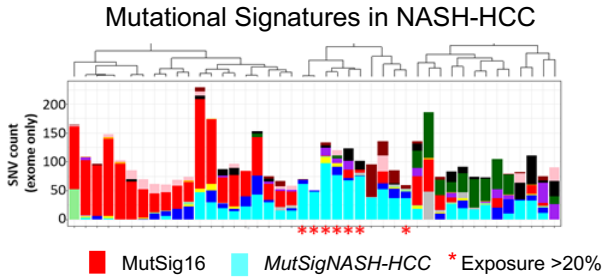
Human NASH-HCC

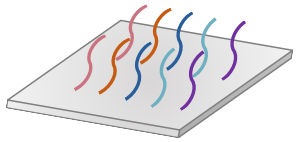




Whole-exome sequencing

↑ *ACVR2A* mutations
↑ *MutSigNASH-HCC*





Microarray-Expression

Signaling pathways

↑ Bile and fatty acid
↑ Oxidative stress
↑ Inflammation

Immune features

↑ *Immunosuppressive cancer field*

Molecular classes

↑ Wnt/TGF-β
↓ *CTNNB1* subclasses

NASH-HCC murine models

- WD+SW
- CD-HFD
- WD+CCl_4

RNA-Seq/Microarray

NASH-HCC
vs
Viral/Alcohol-HCC