ARTICLE IN PRESS

Nature Communications

https://doi.org/10.1038/s41467-026-73212-y

Article in Press

Integrated genomic analyses identify oncogenic pathway interplay in hepatocarcinogenesis defining specific molecular subtypes

Received: 20 February 2025

Accepted: 5 May 2026

Cite this article as: Pan, L., Hirsch, T.Z., Fang, J. *et al.* Integrated genomic analyses identify oncogenic pathway interplay in hepatocarcinogenesis defining specific molecular subtypes. *Nat Commun* (2026). https://doi.org/10.1038/s41467-026-73212-y

Long Pan, Théo Z. Hirsch, Jing Fang, Agnieszka Seretny, Sandrine Imbeaud, Shuoshuo Jin, Olatunji Oluwole Gege, Sandra Rebouissou, Sabrina Sidali, Naïma Ammiche, Julien Calderaro, Giuliana Amaddeo, Jean-Frédéric Blanc, Brigitte Le Bail, Marianne Ziol, Séverine Celton-Morizur, Valérie Paradis, Josep M. Llovet, Darjus Felix Tschaharganeh, Jean-Charles Nault, Chantal Desdouets, Stefano Caruso & Jessica Zucman-Rossi

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

Integrated genomic analyses identify oncogenic pathway interplay in hepatocarcinogenesis defining specific molecular subtypes

Long Pan (潘龙)[1], Théo Z. Hirsch[1], Jing Fang (方静)[2], Agnieszka Seretny[3], Sandrine Imbeaud[1], Shuoshuo Jin (金烁烁)[1], Olatunji Oluwole Gege[3], Sandra Rebouissou[1], Sabrina Sidali[1,4], Naïma Ammiche[1], Julien Calderaro[5,6,7], Giuliana Amaddeo[8], Jean-Frédéric Blanc[9], Brigitte Le Bail[10], Marianne Ziol[1,11], Séverine Celton-Morizur[2], Valérie Paradis[12], Josep M. Llovet[13,14,15], Darjus Felix Tschaharganeh[3], Jean-Charles Nault[1,16], Chantal Desdouets[2], Stefano Caruso[1,5,6,7], Jessica Zucman-Rossi[1,17]

[1] Team « Functional Genomics of Solid Tumors », Cordeliers Research Center, Sorbonne Université, Inserm, Université Paris Cité, équipe Labellisée Ligue Contre le Cancer, Paris, France.

[2] Team « Genomic Instability, Metabolism, Immunity and Liver Tumorigenesis », Cordeliers Research Center, Sorbonne Université, Inserm, Université Paris Cité, équipe Labellisée Ligue Contre le Cancer, Paris, France.

[3] Helmholtz-University Group Cell Plasticity and Epigenetic Remodeling, German Cancer Research Center (DKFZ) and Institute of Pathology University Hospital Heidelberg, Heidelberg, Germany.

[4] Université Paris Cité, Service d'Hépatologie, DMU DIGEST, Hôpital Beaujon, APHP Nord, Clichy, France.

[5] Department of Pathology, Assistance Publique Hôpitaux de Paris, Groupe Hospitalier Henri Mondor, Créteil, France.

[6] Institut Mondor de Recherche Biomédicale, INSERM U955, Université Paris-Est Créteil, Créteil, France.

[7] MINT-HEP, Mondor Integrative Hepatology, Créteil, France.

[8] Department of Hepatology, Assistance Publique-Hôpitaux de Paris, Henri Mondor-Albert Chenevier University Hospital, Créteil, France.

[9] Oncology Digestive Unit, INSERM U1312, University Hospital of Bordeaux, Bordeaux, France.

[10] Department of Pathology, Hôpital Pellegrin, Bordeaux University Hospital, 33076, Bordeaux, France.

[11] Service d'anatomie pathologique et Centre de Ressources Biologiques (BB-0033-00027), Hôpitaux Universitaires Paris-Seine-Saint-Denis, Assistance-Publique Hôpitaux de Paris et Unité de Formation et de Recherche Santé Médecine et Biologie Humaine, Université Paris Nord, Bobigny, France.

[12] Department of pathology, Hôpital Beaujon, Assistance Publique-Hôpitaux de Paris, Clichy, France.

[13] Mount Sinai Liver Cancer Program, Divisions of Liver Diseases, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, USA.

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

[15] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, 08010, Spain

[16] Liver unit, Avicenne Hospital, APHP, Bobigny, France.

[17] Hôpital Européen Georges Pompidou, APHP, Paris, France.

ARTICLE IN PRESS

ARTICLE IN PRESS

Corresponding Author:

Pr Jessica Zucman-Rossi, PUPH Université Paris Cité-HEGP/AP-HP

Director of the Centre de Recherche des Cordeliers

Sorbonne Université-Inserm-Université Paris Cité

15 rue de l'école de Médecine, 75006 Paris, France

Tel: 33 6 01 07 78 75; E-mail: jessica.zucman-rossi@inserm.fr

ARTICLE IN PRESS

ARTICLE IN PRESS

Abstract

Hepatocellular carcinoma (HCC) is a genomically diverse disease, and molecular classification is essential for understanding its biology and improving patient care. Here, integrative analyses of 529 HCC samples across genomic, transcriptomic, epigenomic, and proteomic layers identify nine robust molecular subtypes, validated in 807 external cases. Three common subtypes are defined by alterations in *CTNNB1*^{ex3}*/APC* (25%), *TP53* (21%), and *AXIN1/IRF2* (11%), while four rare subtypes involve *TP53*+*CTNNB1*^{ex3} (6%), *BAP1* (6%), *CCNA2/E1* (6%), and *HNF1A* (1%). Two additional immune-related subtypes, U^{Hot} (17%) and U^{Interm} (8%), have undetermined drivers. Using transcriptomic scores, we demonstrate that *AXIN1* and *TP53* mutations increase β-catenin activity and reduce p53 activity, whereas *BAP1* loss functionally inactivates *TP53*. The rare *TP53*+*CTNNB1*^{ex3} subtype is highly aggressive, associated with poorer outcomes, exhibiting features of both *CTNNB1*$^{ex3-}$ and *TP53*-mutated HCC, with *TP53* mutation occurring prior to *CTNNB1* alteration. The refined molecular classification provides a clinically relevant framework for precision medicine in HCC.

ARTICLE IN PRESS

ARTICLE IN PRESS

Introduction

Liver cancer is a common malignancy and the third leading cause of cancer-related death worldwide.[1] Hepatocellular carcinoma (HCC) is the major type of primary liver cancer, mostly occurring in chronic liver disease and cirrhosis.[2] Despite improvements in surveillance programs, most HCC cases are diagnosed at advanced stages. Treatments of advanced HCC are based on immunotherapy targeting PD-L1 or CTLA4, and anti-angiogenic therapies targeting VEGF or tyrosine kinases. However, only ~ 30% of patients benefit from these treatments[3], due to the high diversity of HCC at the clinical and molecular levels. Thus, classifying HCC into homogeneous subtypes is a crucial step in identifying patients who may benefit from specific treatments.

Over the past decade, cancer driver genes recurrently altered in HCC have been identified, including *TERT promoter*, *CTNNB1*, *TP53*, *AXIN1*, *BAP1*, *ARID1A*, *CCND1-FGF19*, *NFE2L2*, *KEAP1*, *ARID2*, *RPS6KA3, CCNA2/E1, BAP1,* and *HMBS*.[4–12] Moreover, molecular subtyping of HCC has evolved significantly with worldwide efforts. For instance, our group has identified 6 robust transcriptomic subgroups (G1-G6) associated with clinical and genomic characteristics[13]. Hoshida and collaborators identified 3 major transcriptomic subgroups (S1-S3) existing across nine public datasets[14], The Cancer Genome Atlas (TCGA) program determined 3 molecular clusters based on integrative multiple omics data in 196 HCC patients.[7] Additionally, there are several molecular classifications developed based on diverse types of omic data, including methylation profile[15], immune features[16,17] and proteomic data.[8,18] However, most of these molecular classifications were based on expression data, and genomic alterations were not integrated very well. Recent studies have revealed two rare but homogeneous HCC subtypes driven by *CCNA2/E1* alterations or *BAP1* biallelic mutations, highlighting the importance of genomic events in shaping HCC phenotype.[10,12] Besides, current molecular classifications were limited by the small number of included tumors and were underpowered to identify and validate rare subtypes of HCC in integrated analyses.

Here, we aim to identify genotype–phenotype-concordant molecular subtypes of HCC, focusing on genomic alterations. We integrate multi-omics data, including whole-exome sequencing (WES), whole-genome sequencing (WGS), RNA sequencing (RNAseq), methylation array, and reverse-phase protein array (RPPA), from 529 HCC samples from 461 patients primarily treated in France. The findings are further validated in an external cohort of 807 HCC cases. Additionally, we assess the functional impact of major pathways altered in HCC and reconstruct clonal and subclonal evolution in a specific subtype. To make our data more accessible to the liver cancer research community, we develop an interactive web-based tool (https://zucmanlab.com/hccsubtype/) to analyze and visualize the data at the subtype level.

ARTICLE IN PRESS

ARTICLE IN PRESS

Results

Genomic analysis identifies 5 major mutation clusters in HCC

Combined analysis of 2228 genomically independent frozen HCC samples sequenced by WES (n=1,728) or WGS (n=500), including our current LICA-FR French cohort, TCGA from the US, and ICGC-JP from Japan, identified 44 cancer driver genes recurrently altered in more than 1% of HCC (Methods, Supplementary Figure 1a, Supplementary Data 1-2). The effect size of each driver mutation in HCC was quantified (Supplementary Figure 2 and Supplementary Data 3). LICA-FR data included 529 HCC samples from 461 patients treated in France (average age 61 years) with various main causes of chronic liver disease (alcohol intake 34%, HBV 31%, HCV 20%, MASH 19%), and 46% had cirrhosis. Tumor stages varied: 65% were BCLC 0-A, 20% B, and 15% C. In the LICA-FR cohort, which includes 116 WGS and 413 WES samples, we examined genomic alterations in each of the 44 genes, including somatic mutations, copy number alterations (CNAs), structural variations (SVs), gene fusions, and viral insertions (HBV and AAV2) (Supplementary Figure 1b, Supplementary Data 4). The top 10 most frequently altered driver genes were: *TERT* (75%), *TP53* (33%), *CTNNB1* (31%), *ALB* (13%), *AXIN1* (10%), *ARID1A* (9%), *TSC1/2* (9%), *CDKN2A* (8%), *RPS6KA3* (8%), and *KMT2D* (7%). The functional effects of these driver alterations were validated using transcriptomic and proteomic data (Supplementary Figure 1c-e). Protein abundance generally correlated positively with mRNA levels, with an average correlation coefficient of 0.25 (Supplementary Figure 1f,g), consistent with data from another HCC dataset.

Besides the frequent alterations identified in classical cancer driver genes previously described across various tumor types, we found several driver genes involved in liver metabolism with recurring somatic mutations in HCC, including *HMBS* (linked to autosomal dominant acute intermittent porphyria), *HNF1A* (associated with Maturity Onset Diabetes of the Young type 3), and two new HCC drivers, *HPD* and *FOXO1,* which are involved in metabolic liver diseases. *HPD*, which encodes 4-hydroxyphenylpyruvate dioxygenase in the tyrosine catabolic pathway, is the gene responsible for tyrosinemia type 3 and autosomal dominant hawkinsinuria. In this study, *HPD* shows somatic mutations in five of 529 HCC cases (1%), four of which were at codon 241 leading to p.N241S missense mutation, a pathogenic mutation also found in hawkinsinuria. Another mutation hotspot was identified in *FOXO1*, a key transcription factor in insulin signaling, with three HCC cases showing mutations. In these, the mutation chr13:41240285G>C resulting in a p.S22W substitution was similar to the recurrent mosaic mutation seen in alcohol-related (ALD) and metabolic-associated fatty liver (MAFLD) diseases. Interestingly, six of these seven patients also had ALD or MAFLD.

Interaction network analysis of these mutations and alterations in the 489 LICA-FR genomically independent HCC revealed five major clusters of cancer-associated driver gene alterations: (1) *CTNNB1* (30%) with *TERT-ARID2-NFE2L2* or *CDKN2A*, (2) *BAP1* mutations (5%) frequently associated with *PRKACA* amplifications/copy-number gains or *PRKAR2A* losses, (3) *TP53* mutations (33%) often linked with *TSC1/2* mutations or *CCND1/FGF19* amplifications, (4) *AXIN1* mutations (9%) frequently associated with inactivating alterations in ARID1A, RPS6KA3, or IRF2, and (5) *CCNA2/E1* (5%) alterations often tied to *RB1* and *PTEN* inactivating mutations (Figure 1A, Supplementary Data 5). Additionally, *MYC* amplifications/copy-number gains were mutually exclusive with *CCNA2/E1* changes but not linked to specific genomic alterations. Notably, *CTNNB1* mutations were significantly exclusive from the other four main clusters (*TP53*, *BAP1*, *AXIN1,* and *CCNA2/E1*). This exclusivity between *CTNNB1* and *BAP1/AXIN1/TP53* alterations may be due to higher-order antagonistic epistatic effects, while the interactions between *CTNNB1* and *CCNA2/E1*, as well as between *MYC* amplifications/gains and *CCNA2/E1*, remain to be clarified (Supplementary Figure 3, Supplementary Data 6).

Multi-omics integrated analyses identify 9 robust molecular HCC subtypes

We integrated RNAseq transcriptomic profiles of 406 HCC from the LICA-FR cohort using t-distributed stochastic neighbor embedding (t-SNE) with the genomic features. In line with the above, we identified

ARTICLE IN PRESS

homogenous transcriptomic clusters, all validated in external cohorts, driven by either (1) *CTNNB1*^{ex3} mutations or *APC* biallelic mutations, (2) *BAP1* biallelic alterations, (3) *AXIN1* mutations, and (4) *CCNA2/E1* alterations, whereas *TP53*-mutated HCC were interspersed in a cluster that includes non-TP53-mutated tumors (Figure 1B, Supplementary Figure 4). It is worth noting that the *AXIN1* cluster also included HCC with *IRF2* biallelic alterations alone or in combination with *AXIN1* mutations. In addition, we found a fifth HCC cluster related to *HNF1A* biallelic alterations (Figure 1B, Supplementary Figure 4). Among the *CTNNB1*^{ex3}-HCC we identified, a sixth subtype (5% of the patients) consisted of 30 HCC from 27 patients containing *TP53* and *CTNNB1*^{ex3} mutations that are usually exclusive in the vast majority of HCC (Figure 1B).

To characterize the immune profile of HCC, we applied the MCP-counter method[19] (Figure 1C, Supplementary Figure 5a). HCC with high immune infiltrate were distributed along a gradient on the x-axis of the t-SNE, as confirmed with the immune classifications developed by Montironi et al.[16] and Thorsson et al.[20] (Figure 1D and E). Then, we divided HCC without specific genomic alterations (Undetermined HCC) into two subtypes, U^{Hot}-HCC and U^{Interm}-HCC, based on their immune infiltration levels using a centroid method that leveraged MCP-counter readouts (Figure 1B, Methods, Supplementary Figure 6a-e). We further validated this result by applying CosensusTME deconvolution (Supplementary Figure 6f). Also, the t-SNE plot was closely related to the previously described Boyault's (G1-G6)[13], Hoshida's (S1-S3)[14] and Chiang's[21] transcriptomic subgroups, as well as the Meunier's methylation subgroup (M1-M7)[15] (Figure 1F and G, Supplementary Figure 5b). Finally, similar clusters and relationships between different omics data were also observed in TCGA-LIHC/ICGC-JP/CN-Fudan external cohorts, reinforcing the robustness of the analysis using the present method (Figure 1H, Supplementary Figure 5c,f).

Therefore, by integrating all the omics data, we classified HCC into 9 different robust molecular subtypes, all validated in 807 HCC from TCGA, ICGC-JP and CN-Fudan: (1) *CTNNB1*^{ex3}*/APC*-HCC (25%), defined by *CTNNB1* exon 3 mutation (or large in-frame deletion), or *APC* biallelic alterations, were related to G5/G6 transcriptomic, M5 methylation subgroups, *Excluded* and *lymphocyte depleted* immune subtypes, along with high protein expression of glutamine synthetase (*GLUL*), Axin-1 (*AXIN1*), AMPK-alpha (*PRKAA1*) and Bcl-xL (*BCL2L1*), (2) *AXIN1/IRF2*-HCC (11%), defined by inactivating *AXIN1* mutations and/or *IRF2* biallelic alterations, significantly associated with G1 transcriptomic subgroup, M2 methylation subgroup and *Intermediate* immune subtype, along with low protein expression of Axin-1 (*AXIN1*), (3) *TP53+CTNNB1*^{ex3}-HCC (6%) defined by *TP53* associated with *CTNNB1*^{ex3} mutations, significantly associated with G5/G6 transcriptomic subgroups and *Excluded* immune subtype, (4) *BAP1*-HCC (6%), defined by *BAP1* biallelic alterations, significantly associated with G1 transcriptomic subgroup, M2 methylation subgroup, *Inflammatory* and *Wound healing* immune subtypes, along with an increase in the expression of E-cadherin (*CDH1*), PKA C-alpha (*PRKACA*) and TGF-beta I/III (*TGFB1/TGFB3*), (5) *CCNA2/E1*-HCC (6%) defined by *CCNA2* or *CCNE1* genomic alterations by focal amplifications, HBV/AAV2 insertions, fusions or structural rearrangements, significantly associated with G2 transcriptomic subgroup, M7 methylation subgroup and upregulation of NK cell, along with high expression of survivin (*BIRC5*) and Ki67 (*MKI67*) and low expression of cyclin D1 (*CCND1*), (6) *HNF1A*-HCC (1%), defined by *HNF1A* biallelic alterations, related to *lymphocyte depleted* immune subtype, (7) *TP53*-HCC (21%) defined by *TP53* mutations alone without mutations described above, related to G3 transcriptomic subgroup and *Intermediate* and *IFN-γ dominant* immune subtypes, along with an increase of Ki67 (*MKI67*) and pyruvate kinase M1/2 (*PKM*), (8) U^{Hot}-HCC (17%), associated with G4 transcriptomic subgroup, M1 normal-like methylation subgroup, and *immune like*, *immune active* and *inflammatory* subtypes, along with an increase of jun-B (*JUNB*) and *PTEN,* (9) U^{Interm}-HCC (8%), significantly associated with G4 transcriptomic subgroup and *Intermediate* immune subtype (Figure 1I, Supplementary Figure 5d, Supplementary Data 7,8). The enriched biological pathways and mutational signatures for each subtype were shown in Supplementary Figure 5e as we expected (Supplementary Data 7,9).

ARTICLE IN PRESS

To further estimate the impact of subclonal mutations on this molecular classification, we repeated the subtype assignment using a variant allele frequency (VAF) threshold of 5% (previously 10%). Molecular subtyping for most HCC is stable, and only 9 out of 507 HCC (<2%) were reclassified: 8 Unclassified HCC were reclassified as 5 *CTNNB1*^{ex3}*/APC*, 2 *TP53* and 1 *AXIN1/IRF2*, one *TP53*-HCC was reclassified into *AXIN1/IRF2* subtype. Five additional cases of HCC with subclonal *CTNNB1* mutations were identified, suggesting a dynamic evolution of *CTNNB1* mutations in these tumors.

The CNA profiles across the whole genome for each subtype were visualized (Supplementary Figure 7). Generally, 1q copy-number gain was a common event, including in 60% of both U^{Hot}-HCC and U^{Interm}-HCC despite their absence of common driver mutations. *CTNNB1*^{ex3}*/APC*-HCC, *HNF1A*-HCC, U^{Hot}-HCC and U^{Interm}-HCC had fewer CNAs, whereas other subtypes had distinct CNA profiles. Specifically, chromosome 4q, 10q, 13q, 16q and 17p (*TP53* locus) deletions were enriched in *TP53*-HCC, while 5q, 16p/q deletions were enriched in *AXIN1/IRF2*-HCC (adjusted-p<0.0001) (Supplementary Data 10). It is worth noting that the loss of 3p21 caused the deletion of both *BAP1* and *PRKAR2A* in 85% of *BAP1*-HCC.

We further characterized the functional activity of the two major dysfunctional pathways, *TP53* and WNT/β-catenin, accounting for ~70% HCC.

Modulation of the p53 activity in HCC

To quantify the p53 activity, from 6 established p53 transcriptomic signatures based on human expression data, we generated a consensus p53 signature[7,22] (Figure 2A, Methods, Supplementary Figure 8a-f). In our cohort of 406 HCC, in addition to cell cycle signaling, p53 score was also markedly correlated with DNA repair, fatty acid metabolism, hepatic differentiation and apoptosis (Supplementary Figure 8g), consistent with the complex and multiple roles of p53 in human cells.[23] HCC with low p53 score were significantly associated with high chromosome instability, whole genome duplication (WGD), G3 transcriptomic subgroup, HBV infection, Aflatoxin B1 exposure, satellite nodules, advanced stage, high serum AFP, poor differentiation and large tumors (Figure 2A and B). As expected, a low p53 score was strongly correlated with *TP53* mutations (p< 0.0001, Figure 2A). HCC with *TP53* biallelic (n=129) or with monoallelic (n=12) mutations showed similar p53 inactivation; in contrast, HCC with mono-allelic p53 inactivation by loss of heterozygosity (LOH) showed a significantly decreased p53 activity compared to WT tumors (p< 0.001) but exhibited higher score compared to biallelic p53 inactivation, suggesting a dominant negative effect of the monoallelic *TP53* mutations (p< 0.001, Supplementary Figure 9). Within HCC with biallelic *TP53* alterations, no significant relation was observed between the *TP53* mutation type and the activity score (Supplementary Figures 9 and 10). In contrast, *CTNNB1*^{ex3}*/APC*, U^{Hot} and U^{Interm} HCC retained high p53 activity (all the p< 0.05, Figure 2A), as well as HCC with *IL6ST* or *HMBS* mutations (Figure 2A).

We further investigated whether certain driver genes could cooperate to modulate the functional activity of p53 in HCC. We identified that co-occurrences of *TP53* with *AXIN1* or *RB1* mutations in HCC were associated with a lower p53 activity, suggesting that *AXIN1* and *RB1* could cooperate with *TP53* mutations with an additive effect (Figure 2C). In contrast, *CCNA2/E1, PTEN, CCND1-FGF19* and *TSC1/2* alterations co-occurred with *TP53* mutations but without an additive effect on the p53 activity (Figure 2C, Supplementary Figure 11a).

Finally, *BAP1* and *IRF2*-mutated HCC displayed a low p53 activity in the absence of *TP53* mutations in our series and in the external cohorts of 727 HCC, suggesting that these two gene alterations could induce functional p53 inactivation (Figure 2C, Supplementary Figure 11b,c). For *IRF2*, which is known to bind to the p53 inhibitor MDM2, we previously demonstrated that *IRF2* inactivation in the HepaRG cell line led to p53 functional inactivation.[24] Here, in pan-cancer analysis, we confirmed that *BAP1* biallelic alterations were also mutually exclusive with *TP53* mutations (p< 0.0001, Supplementary Figure 11d), with decreased p53 activity in *BAP1*-mutated kidney renal clear (KIRC) and papillary cell carcinoma (KIRP) (Supplementary Figure 11e). Finally, in a liver cancer mouse model induced by *Pten* knockout, we

ARTICLE IN PRESS

identified a strong p53 activity shutdown when *Bap1* was silenced (p= 0.005, Figure 2D, Supplementary Figure 12).

Heterogeneously increased WNT/β-catenin activity associates with *CTNNB1, AXIN1*, *APC* and *TP53* mutations

To evaluate β-catenin activity, we used a set of 23 genes including 9 canonical *CTNNB1* targets, 12 liver-specific positive targets and 2 liver-specific negative targets compiled from Abitbol et al.[25] We confirmed that mutations at amino acid position D32-S37, large in-frame deletions across exon 3, or multiple hits cause high increases in β-catenin activity; in contrast, mutations at amino acid position T41 and S45 were associated with moderate β-catenin activity, and mutations at exon 7/8 exhibited weak activity (Figure 3A, Supplementary Figure 13a)[26] We found that *APC* biallelic alterations and *AXIN1* mutations were also associated with ß-catenin activation with moderate-high or moderate-weak β-catenin activity, respectively (Figure 3A). The highest ß-catenin activation was strongly associated with *CTNNB1*^{ex3}*/APC* and *TP53+CTNNB1*^{ex3} HCC subtypes with significantly higher levels in G6 compared to G5 transcriptomic subgroup (Figure 3B, Supplementary Figure 13b, $p < 0.0001$)[13]. Finally, we found higher β-catenin activities in *TP53*-mutated HCC even without *CTNNB1*, *AXIN1* or *APC* mutations, and similar results were retrieved in the external cohorts of 727 HCC, suggesting that *TP53* inactivation could directly or indirectly activate ß-catenin (Figure 3C).

Co-occurring *TP53* and *CTNNB1*^{ex3} alterations define a rare HCC subtype with poor prognosis

In the rare HCC subtype with both *TP53* and *CTNNB1*^{ex3} mutations (*TP53+CTNNB1*^{ex3} HCC), patients and tumors presented mixed features of *CTNNB1*^{ex3} and *TP53*-mutated HCC (Figure 4A). Indeed, at the clinicopathological and genomic level, *TP53+CTNNB1*^{ex3} HCC were enriched with HBV infection, more CNAs and high proportion of WGD/polyploidy as seen in *TP53*-mutated HCC, with a higher ratio of microvascular invasion (16/21, 76%), cirrhosis (19/27, 70%) and *CCND1-FGF19* ampl (5/30, 17%) (Figure 4A, Supplementary Data 11). However, at the transcriptomic level, *TP53+CTNNB1*^{ex3} HCC resembled *CTNNB1*^{ex3} HCC and mostly belonged to the G5/6 subgroups (Figure 1B, 1D, Figure 4A). Of note, the *TP53+CTNNB1*^{ex3} subtype was associated with poor prognosis in our series (Figure 4B), and was validated in external cohorts (Figure 4C). Furthermore, we established HCC mouse models to mimic these three subtypes using hydrodynamic tail-vein injection in the context of Myc over-expression (Figure 4D, Methods and Supplementary Figure 14a,b).[27] We found *MYC-luc;sg-p53;CTNNB1* mice developed larger HCC with a more aggressive phenotype compared with *MYC-luc;sg-p53 or MYC-luc;CTNNB1* mice, characterized by marked cytologic atypia, necrosis, and microvascular invasion (Figure 4E-G and Supplementary Figure 14c-e).

To understand the evolutionary history of *TP53+CTNNB1*^{ex3} HCC, we used the ability to time mutations relative to WGD events (Supplementary Figure 15a,16). Briefly, 16 of 27 *TP53+CTNNB1*^{ex3} HCC patients had WGD events. In 50% (8/16) of tumors with WGD, the timing difference of *TP53* and *CTNNB1* mutations was determined, and in all 8 cases, *TP53* mutations occurred before WGD while *CTNNB1* mutations occurred after. In the remaining 8 HCC, the timing of *TP53* and *CTNNB1* events was not possible to assess (Supplementary Figure 15a).

To further explore the spatial and longitudinal evolution of these tumors, we constructed phylogenetic trees of 5 patients with *TP53+CTNNB1*^{ex3} HCC with paired primary and/or recurrent/metastatic tumors, highlighting diverse evolution patterns (Figure 5A, Supplementary Figure 15b and Supplementary Data 12). For patient #206, two synchronous tumors were genomically independent and only one tumor had *TP53* and *CTNNB1* mutations arising before and after the WGD, respectively. For patient #4124, two independent tumors harbored different *TP53* mutations arising before the WGD and only one of these tumors had *CTNNB1* mutation arising after the WGD (Figure 5A). For patients #1756 and #2022, *TP53* and *CTNNB1* mutations were shared across the primary tumors and relapses, but patient #1756 had an additional subclonal *CTNNB1* mutation occurring in the primary tumor (Figure 5A and Supplementary Figure 15b).

ARTICLE IN PRESS

For patient #2018, the *TP53* mutation followed by a WGD event was shared in primary tumor and metastasis, then the *CTNNB1* mutation appeared at the metastatic stage (Figure 5A). Interestingly, there was a transition of transcriptomic states from G4 subgroup to G6 subgroup. These results suggest that *TP53* alterations occur early during tumor evolution and prior to *CTNNB1* mutation in most cases. Notably, the temporal order of *TP53* and *CTNNB1* mutations was confirmed by our high-order epistatic analyses in 489 genomically independent HCC (Supplementary Figure 3).

Furthermore, we found *TP53*+*CTNNB1*^{ex3} HCC had a higher level of CINscore and a greater frequency of WGD, and reduced immune infiltrate compared with *TP53* HCC (Figure 5B). Thus, by integrating molecular and clinical features, we propose an evolutionary model for this subtype, where *TP53* mutation is positively selected as the first hit to initiate the onset of HCC, mainly under HBV infection, followed by selection of *CTNNB1* mutation as the second hit during tumor progression. The overall evolution process leads to high genomic instability, shifts in the transcriptomic state to G5/6 subgroup, decreases immune infiltrates and finally causes poor prognosis (Figure 5C).

HNF1A biallelic inactivation reveals a rare homogenous HCC subtype developed on normal liver

In this study we identified *HNF1A* HCC as a rare subtype carrying *HNF1A* biallelic inactivation, including 7 HCC from 4 patients in LICA-FR cohort (Figure 1B) and 4 HCC in TCGA-LIHC cohort (Figure 1E, Supplementary Data 13), associated with a specific transcriptomic profile, particularly in genes controlled by *HNF1A* allelic status *(FABP1*, *UGT2B7, DHRS2* and *ANGPTL7)* (Supplementary Figure 17a). In contrast, *HNF1A* monoallelic mutated HCC did not show any specific pattern (Supplementary Figure 4 and 17a). No difference was found between *HNF1A* biallelic and monoallelic mutated HCC in terms of *HNF1A* mutation type or clonality (Supplementary Figure 17b-d).

At the genomic level, no WGD occurred on these tumors (0/11, 0%, $p< 0.05$, Supplementary Figure 17e) and CNAs across the whole genome were lower compared with other HCC (CINscore: 0.12 vs. 0.46, $p< 0.0001$). At the transcriptomic level, *HNF1A*-inactivated HCC mostly belonged to the G4 subgroup ($p= 0.023$) and all were immune-cold HCC (11/11, 100%, $p= 0.04$). In addition, we found *HNF1A*-inactivated patients displayed uniform clinical characteristics, with enrichment of females (6/8, 75%, $p= 0.0045$), low or no fibrosis in the liver (8/8, 100%, $p= 0.0005$), older age at diagnosis (mean age: 69.6 yr, $p= 0.027$) and an absence of classical HCC risk factors (6/8, 75%, $p= 0.0005$). Besides, all these HCC belonged to clinically AFP-negative HCC (7/7, 100%, $p= 0.015$) and showed a good prognosis compared with other HCC (p= 0.025, Supplementary Figure 17f). Since *HNF1A* mutation is a major driver of hepatocellular adenoma, rare benign liver tumors usually developed in females with normal liver, *HNF1A*-mutated HCC are compatible with malignant transformation. However, only 2 out of the 8 patients had a hepatocellular adenoma lesion.

Characterizing HCC subtypes with comprehensive clinicopathological and molecular features

By analyzing the clinical features associated with each molecular subtype, we identified a specific distribution of the HCC risk factors. HBV infection was highly associated with the *TP53*-driven subtypes (*TP53*, *TP53*+*CTNNB1*^{ex3}, and *AXIN1/IRF2)*. In contrast, *CTNNB1*^{ex3}*/APC* HCC was associated with alcohol consumption, and no specific association was observed with MASH or HCV infection (Figure 6A, Supplementary Figure 18a, Supplementary Data 14). Interestingly, the rare subtypes *HNF1A*, *BAP1,* and *CCNA2/E1* were frequently associated with unusual patient features in HCC, with enrichment in females, low or no fibrotic liver, and younger age for *BAP1* (Figure 6B-D, Supplementary Figure 18a,b). These results show that the natural history of the patient may influence the type of HCC they develop.

AXIN1/IRF2 and *BAP1* subtypes had higher serum AFP levels, while *HNF1A* and *CTNNB1*^{ex3}*/APC* HCC had lower levels (Figure 6D). The tumor size was larger in *CCNA2/E1* and *BAP1* HCC and smaller in U^{Hot} HCC (Figure 6D). At the genomic level, *TP53*+*CTNNB1*^{ex3}, *AXIN1*, *BAP1* and *TP53* HCC had higher CINscore and greater proportions of WGD events, whereas *CTNNB1*^{ex3}, *HNF1A* and U^{Hot} HCC showed

ARTICLE IN PRESS

lower values (Figure 6D). It is worth noting that *CCNA2/E1* HCC had lots of focal copy-number segments although their overall CINscore was not high, consistent with the information from the CNA profile across the genome (Supplementary Figure 7). This is in accordance with the specific signature of replication stress-induced structural variants leading to many breakpoints affecting short segments previously described in *CCNA2/E1* HCC.[12,28] All these associations were mostly validated in external cohorts (Supplementary Figure 19).

Regarding immune profiling across the nine HCC subtypes, *CTNNB1*^{ex3}*/APC*, *TP53+CTNNB1*^{ex3}, and *HNF1A* HCC exhibited fewer immune infiltrates, whereas *BAP1* and U^{Hot} HCC generally showed higher immune infiltrates (Figure 6D). *BAP1* HCC displayed a higher abundance in endothelial cells and overexpressed VEGFR2/3 compared with other HCC subtypes (Supplementary Figure 20a,b). It also presented the second-highest tertiary lymphoid structures (TLS) signature score, following U^{Hot} HCC, consistent with histological analyses that confirmed *BAP1* HCC and U^{Hot} HCC were enriched in TLS (Supplementary Figure 18a). Moreover, *HNF1A* HCC had elevated *LAG3* expression compared with other HCC subtypes, suggesting this subtype might be responsive to anti-*LAG3* therapy (Supplementary Figure 20b).

Finally, we systematically assessed the prognosis of each subtype. The 9 subtypes had different long-term outcomes (Supplementary Figure 18c,d). We confirmed U^{Hot} HCC was also associated with good prognosis in both LICA-FR cohort and external cohorts (Figure 6E and F, disease-specific survival, $p < 0.0006$ and $p < 0.02$), whereas *CCNA2/E1* HCC was associated with poor prognosis (Supplementary Figure 18e, disease-specific survival, $p=0.012$).

ARTICLE IN PRESS

ARTICLE IN PRESS

Discussion

In the present study, we refined the molecular HCC classification in 9 robust subtypes based on cancer driver gene mutations and transcriptomic profiling. These HCC molecular subtypes are strongly related to molecular and clinical features, and they were validated in large external cohorts including Western and Eastern HCC populations. All data can be easily explored interactively at our shiny website https://zucman-lab.shinyapps.io/hccsubtype/. This molecular classification sheds light on the major mechanisms of oncogenesis with their interplay in human HCC, and provides new tools to develop precision care for patients (Figure 7).

In this study, we showed that multi-omic integration in a large set of HCC associated with various etiologies is powerful to better understand the mechanisms of tumorigenesis in the liver. Our comprehensive analysis revealed the functional roles of the two major oncogenic pathways in HCC driven by ß-catenin and p53, usually in an exclusive mode. It has also enlightened how oncogenic specific pathways can cooperate within the same tumor with fundamental consequences on cell differentiation and proliferation and their impact on the natural history of each HCC subtype. In the evolution of the tumors, cooperation between p53 inactivation and ß-catenin activation is particularly demonstrative in humans and mice, leading to more aggressive tumors related to very poor prognosis. Here, we also demonstrated that *AXIN1* loss can induce a weak β-catenin activity and decrease functional p53 activity. Additionally, we observed that monoallelic *TP53* mutations may exert a dominant-negative effect, as they had similar low p53 activity as biallelic *TP53* mutations. Nevertheless, we cannot exclude two alternative explanations: (i) the presence of subclonal biallelic lesions below the detection limit of bulk DNA sequencing (e.g., a fraction of ≤ 20%); and (ii) the contribution of exogenous risk factors, particularly a possible synergistic interaction between HBV infection and AFB1 exposure, in modulating p53 activity. These observations underline the dynamic evolution of the tumor cells identified at the single-cell and spatial levels in HCC.[29–31]

Understanding inter-tumor heterogeneity is important for developing more precise therapies for patients. In the context of curative treatments, since prognosis is highly related to the specific subtype, the prescription of different types of curative treatment or new schemes, including adjuvant systemic therapies, could vary for the extreme molecular profiles associated with the best survival (U^{Hot}, 75% at 5 years) compared to the worst survival (*TP53+CTNNB1*^{ex3}, 25% at 5 years).

A precise molecular classification is indeed required to develop the proof of concept that HCC could be treated tailored to their genomic/transcriptomic features (Figure 7).[32] Overall, our in-depth immune analyses suggested *BAP1* HCC and U^{Hot} HCC were most likely to benefit from current immunotherapy. Additional analyses and results from clinical trials are required to include immune signatures predicting response to immunotherapy in our integrated HCC classification. Besides, several therapeutic strategies have been investigated for treating *BAP1*-mutated tumors, including tazemetostat[33] (an *EZH2* inhibitor) and niraparib[34] (a Poly (ADP-ribose) polymerase (PARP) inhibitor). For other subtypes, alternative strategies may be required. *CTNNB1*^{ex3}*/APC* subtype HCC are difficult to treat since they are frequently immune-cold and less proliferative. Recently, a pre-clinical study from Rialdi et al.[35] found that *CTNNB1*-mutated HCC had a specific vulnerability on EZH2 and identified a novel multi-kinase inhibitor, WNTinib, which inhibits the tumor growth of this HCC subtype in both human and murine models, showing unprecedented therapeutic prospects. In the present work, we showed that p53 exhibited a biallelic inactivation in most *TP53*-mutated HCC. An important therapeutic strategy to consider for this subtype could be based on restoring the WT p53 activity. In this line, Xiao et al.[36] restored WT p53 expression via a CXCR4-targeted *TP53* mRNA nanoparticle in HCC mouse models with p53 deficiency, achieving a good anti-tumor effect. Additionally, *TP53*-inactivating HCC cell lines and mouse models were also sensitive to alisertib[37,38], an Aurora A kinase inhibitor. Moreover, two co-occurring alterations in this *TP53*-subtype, *MET* amplifications and *TSC2* inactivation, displayed hypersensitivity to tepotinib[9,39] (a *MET* inhibitor) and everolimus[32] (a mTOR inhibitor) in clinical trials, respectively. Also, the MEK inhibitor trametinib could be an option in *RPS6KA3* inactivation.[40] In the same line, PARP inhibitors could be proposed for *CCNA2/E1*

ARTICLE IN PRESS

HCC since they exhibit a genome instability signature similar to that of *BRCA1*-mutated breast cancer[12]. Additionally, BLU-222 (a *CDK2* inhibitor) could be an alternative candidate.[41] More studies using cutting-edge techniques, including genome-scale CRISPR screens and mRNA-based cancer therapeutics, are needed to specifically identify the therapeutic targets for each subtype.

The present molecular classification also provides innovative insights and strategies to identify potential biomarkers to improve the diagnosis and follow-up of patients with HCC. Among them, a critical issue to solve is the lack of increased AFP detected in the serum in ~40% of patients with HCC, which is closely related to the lack of AFP gene expression by the tumor according to specific subtypes.[42,43] This is particularly important for the *CTNNB1*^{ex3}*/APC* subtype, which represents 30 to 40% of HCC in Western countries for which serum AFP is usually negative. In this line, we must search for new secreted tumor biomarkers in specific HCC subtypes.

Beyond the classical cancer driver genes altered in HCC and the common HCC subtypes associated with either *CTNNB1*^{ex3}*/APC*, *TP53*, or *AXIN1/IRF2*, we also identified the important role of cancer driver genes, involved frequently (*CTNNB1*, ß-catenin) or rarely (*HPD*, *HMBS*, *HNF1A*, *FOXO1*) in liver metabolic pathways. Overall, HCC with metabolic pathway alterations tended to occur more frequently in females, without cirrhosis, and they were more frequently associated with good prognosis. These metabolic pathways could serve as promising therapeutic targets, following a recent study showing that a loss of HPD promotes HCC development via AMPK/mTOR pathway inhibition in mice.[44]

In comparison with other types of cancers, the use of molecular classifications of liver tumors is currently limited by two main challenges: (1) the availability of the tumor biopsies, and (2) the need to demonstrate therapeutic efficiency guided by the molecular subtypes. To overcome these limitations in the future, we must demonstrate in clinical trials of patients with HCC how homogeneous subgroups of patients will benefit from specific therapies. Several proofs of concept have already been provided in patients resistant to first-line systemic treatments, in particular to target *MET* amplification by tepotinib or *TSC1/2* mutations by mTOR inhibitors.[32,45] Also, the feasibility and safety of liver tumor biopsies and molecular analyses to guide treatment have been proven.[32,46] Molecular analyses to predict response to immunotherapy in clinical trials are also expected to identify patients who will benefit from more precise care.

This study also had several limitations. First, this study attempted to integrate multi-omics data from the genome, epigenome, transcriptome, and proteome to better characterize HCC, and very strong relations were identified across the molecular HCC-subtypes. However, the associations between HCC subtypes and proteomic features should be interpreted with caution, as they have not been validated in an external series due to the limited number of common antibodies used across different cohorts. Second, at the genomic level, structural variations were not fully captured in two-thirds of the tumors sequenced in WES without WGS data. More comprehensive datasets will help us to more precisely recapitulate molecular features in tumors and further reveal tumor heterogeneity in the future. Third, the evolutionary model we proposed for *TP53+CTNNB1*^{ex3} subtype primarily focused on the timing of *TP53* and *CTNNB1* mutations, together with genomic and transcriptomic changes at the bulk level. Single-cell omics could further decipher and validate the clonal evolution of *TP53* and *CTNNB1* mutations.

In conclusion, our study offers the largest homogeneous HCC dataset with comprehensive genomic and clinical annotations to date, which will support clinical implementation and help identify more effective therapeutic targets in HCC.

ARTICLE IN PRESS

Methods

Clinical cohorts

A large series of 529 hepatocellular carcinoma (HCC) samples (LICA-FR cohort) obtained from surgical resections or tumor biopsies were collected from France (n = 479), Spain (n = 9) and Italy (n = 41). The study was approved by institutional review board committees (CCPPRB Paris Saint-Louis IRB00003835) and written informed consents for all patients were acquired in accordance with legislation. Tumors and corresponding non-tumor counterparts were frozen in liquid nitrogen immediately and stored at -80 °C. Clinicopathological data of this series were summarized in Supplementary Data 1.

Besides, we constructed external merged cohorts of 807 HCC from public liver cancer cohorts with available genomic and/or transcriptomic data as the validation cohort, including 365 HCC from The Cancer Genome Atlas (TCGA-LIHC cohort)[7], 283 HCC from International Cancer Genome Consortium Japanese cohort (ICGC-JP cohort)[6,47] and 159 HCC from Chinese Fudan university (CN-Fudan cohort).[8] 727 out of 807 HCC had transcriptomic data (Supplementary Data 1). Mixed hepatocholangiocarcinoma and cholangiocarcinoma were excluded from these cohorts.

Genomic sequencing and Somatic mutation calling

A total of 529 tumors and paired non-tumor tissues were sequenced by WGS (n = 116, comprising 74 previously published and 42 new cases) or WES (n = 413, comprising 353 previously published and 60 new cases), and Sanger or MiSeq sequencing was additionally performed to detect TERT promoter mutations on samples without WGS as previously described[9–12,24,48–50]. MuTect2 was used to call somatic mutations from WGS and WES data by comparing each tumor sample with its matched non-tumor counterpart or a panel of normals (PON) file. To improve specificity in the calling of mutations with low variant allele frequency (VAF), we quantified the number of high quality variant reads in the tumor (mapping quality ≥ 20, base quality ≥ 20) and the number of variant reads in the non-tumor sample with no quality threshold using bamreadcount (https://github.com/genome/bam-readcount). Only variants matching the following criteria were retained: VAF ≥ 2 % in the tumor with ≥ 3 variant reads, VAF ≤ 5% in the non-tumor samples with ≤ 2 variant reads, and a VAF ratio ≥ 5 between the tumor and non-tumor sample.

Pathogenicity annotation for mutations

Only pathogenic variants were retained in the subsequent analysis except driver gene detection. Specifically, truncating mutations (nonsense variants, frameshift variants), splice variants and in-frame variants were kept. For missense mutations, only variants predicted as (probably) damaging variants in at least one of two tools (SIFT and PolyPhen2) were kept. For *TP53* variants, we additionally considered the annotations from UMD_TP53 Mutation Database[51].

Viral insertion screening

For each tumor, HBV and AAV2 insertions were screened as previously described[52,53]. Simply, we integrated the results from viral capture, WGS, WES, or RNA-seq if the data is available. Only clonal insertions were considered using the same definition from Péneau et al.[53]. There is no new AAV2 insertion events found in our series (Supplementary Data 4).

Compilation of liver cancer driver genes

First, we integrated mutation data from our cohort and 6 public cohorts (Supplementary Data 1), with totally 2228 HCC samples.[4–7,54,55] For multiple tumors from the same patient, only genetically independent HCC samples were included. To avoid potential artifacts, noise, different sequencing depth across the large-scale cohorts, only mutations with VAF ≥ 10% were included to analyze. MutSig2CV v3.1[56] was used to detect driver genes which consists of three independent statistical tests including abundance, clustering and conservation. 53 genes with q value ≤ 0.1 were regarded as significantly mutated. For these genes, we

ARTICLE IN PRESS

filtered out genes with mutation frequency in our series < 1% or not expressed in liver tissues (95th TPM percentile < 0.3). Then, if genes are annotated as cancer driver genes in Cancer Gene Census from COSMIC, we considered them as HCC driver genes. For genes not annotated as driver genes in COSMIC, we checked mutational spectrums, coding sequencing length and relative literature for each gene carefully in order to avoid artifacts. Besides, we manually curated 10 genes as driver genes (*CCND1*, *FGF19*, *CCNE1*, *CCNA2*, *MET*, *MYC*, *APC*, *KMT2D* and *PRKACA* and *PRKAR2A*), which have been considered common driver gene in COSMIC cancer gene census or proven to shape HCC phenotype in previous studies[9,10,12,24,27]. Finally, a total of 44 HCC driver genes were determined (Supplementary Data 2). The selection effect sizes of driver gene mutations in HCC were quantified using cancereffectsizeR.[57]

Detection of large in-frame deletions on *CTNNB1* exon 3

CTNNB1 large in-frame deletions were screened in WGS, WES and RNA-seq data using dedicated approaches, as previous described[58]. Specifically, MANTA software[59] was applied to identifying structural variations spanning *CTNNB1* region in WGS and WES data (with the --exome option), with the following parameters to increase sensitivity: minEdgeObservations = 2, minScoredVariantSize = 10. For RNA-seq, we analyzed the *junctions.bed* output files from TopHat2[60] to identify abnormal junctions involving *CTNNB1* exons 3 or 4. We also screened for paired reads with one mate in exon 2 and the other in exon 3/4, and an abnormally long insert size, greater than the mean insert size of the library + 1 standard deviation. Finally, our LICA-FR cohort was thoroughly reviewed and 12 samples were found to have *CTNNB1* large deletions, while raw RNA-seq data was acquired and analyzed for TCGA-LIHC cohort and 4 samples have CTNNB1 large deletions. In addition, Sanger or MiSeq sequencing was additionally performed to detect this event in 353 HCC with available samples, as previously described (Supplementary Data 15).[9,26]

Allele-specific copy number analysis and structural rearrangement analysis

We used FACETS algorithm[61] (version 0.6.1) to infer tumor purity, ploidy and allele-specific absolute copy number profile on WGS and WES using different parameters. Specifically, single nucleotide polymorphisms (SNP) count matrices for tumor and non-tumor tissue were obtained by applying *snp-pileup* on bam files (parameters: -q15; -Q20; -P100; -r20,0). We used *preProcSample* to generate log-R-ratio (LRR) and B Allele Frequency data from SNP matrices (parameters on WGS: cval = 35, snp.nbhd = 200, ndepth = 20; parameters on WES: cval = 25, snp.nbhd = 250, ndepth = 30), *ProcSample* (parameters on WGS: cval = 350, min.nhet = 5; parameters on WES: cval = 150, min.nhet = 5) to estimate the 2-copy state and *emcncf* (min.nhet = 5) to obtain allele-specific copy number profiles. Focal CNA events (segment length ≤ 10 Mb) were determined using the following criteria: if the total copy-number of segments is equal to 0, they are considered homozygous deletions; if the total copy-number of segments is greater than or equal to ploidy plus 5, they are considered focal amplifications. In addition, we also calculated the log ratio of the sequence coverage in each tumor and paired non-tumor liver and then conducted circular binary segmentation algorithm from R package *DNAcopy*[62] to identify high-level CNA events, as previous described.[11] Each focal CNA events involved in driver genes were manually checked in the copy-number profile.

We applied MANTA software[59] to identify somatic SVs in WGS data as previously described.[58] To keep the most reliable events, we only selected SVs supported by ≥ 15 reads and with a VAF ≥ 10% in the tumor and ≤ 1 variant read in the non-tumor counterpart.

Orthogonal validation of tumor purity from copy-number profiles

In this study, we had 61 samples with both WGS and WES data, and we found the ploidy and purity derived from WGS and WES are consistent, suggesting the robustness of Facets (Supplementary Figure 21a,b). Besides, we estimated the tumor purity based on mutations. We reasoned that the VAFs of clonal mutations should reflect the tumor purity and should be compatible with the results derived from copy number solution. We applied TPES tool in our cohort, which used clonal monoallelic single-nucleotide variants (SNVs) within diploid genomic regions to infer the tumor purity.[63] In general, we found the purity obtained from

ARTICLE IN PRESS

Facets and TPES are quite concordance (Supplementary Figure 21c). In addition, we performed an alternative orthogonal validation analysis using targeted sequencing (Miseq) to assess the tumor purity in 33 samples. For each tumor, one clonal driver mutation (*TP53, CTNNB1, AXIN1, BAP1,* or *HNF1A*) was selected according to the clonal peak identified in the WGS/WES data, and sequenced with a minimum depth of 200x and an average depth of 847x. Then, the purity was calculated based on the following formula:

$$\text{purity} = \frac{2 \times \text{VAF}}{\text{VAF} \times (2 - \text{nMut} - \text{nWT}) + \text{nMut}}$$

where nMut denotes the mutant allele copy-number and nWT denotes the non-mutant allele copy-number in the tumor. The results derived between Facets and Miseq are consistent (Supplementary Figure 21d, Supplementary Data 16).

Moreover, we expected that tumor purity would be associated with the abundance of non-malignant components in HCC samples. Consistent with this, we found the purity estimated from Facets was negatively associated with immune infiltration inferred from RNAseq data (Supplementary Figure 21e).

Co-occurrence, mutual exclusivity, and higher-order epistasis analyses

For the analysis of co-occurrence and mutual exclusivity among driver gene alterations, we used one-sided fisher's exact test with Benjamini-Hochberg correction. Then, we applied the algorithm developed by Alfaro-Murillo et al.[64] for higher-order epistasis analysis. Briefly, this method is based on a continuous-time Markov chain model, allowing the estimation of evolutionary trajectories and fluxes using the distribution of somatic genotypes within the cohort. The antagonistic or synergistic epistatic effects were defined as cases in which the presence of a given somatic genotype substantially reduced or increased the evolutionary flux toward acquiring another somatic mutation, compared with the flux from the wild-type genotype. We adopted a conservative approach, considering such effects only if the flux value was greater than the upper confidence interval or less than the lower confidence interval of fluxes from the wild-type genotype.

Clonality analysis and whole genome duplication

Clonality analysis was performed as previously described[48]. Simply, the cancer cell fraction (CCF) for all somatic mutations calculated from VAF was estimated using a binomial distribution and maximum likelihood estimation. Mutations were considered as clonal if the upper boundary of the 95% confidence interval was > 0.95, and subclonal otherwise.

We used the algorithm from Dewhurst et al.[65] to identify WGD status. Simply, a simulation test was conducted to calculate the probability that an observed copy-number profile could be produced from successive and independent copy-number gains and losses. In tumors with WGD, mutations were further timed relative to WGD using a conservative approach adapted from previous studies[66,67]. Specifically, clonal mutations were classified as early if the major allele in a given segment (MCN) ≥ 2 and mutant copy number ≥ 1.75 while late clonal mutations were regarded as those with MCN ≥ 2 and mutant copy number ≤ 1.25. Clonal mutations which could not be timed were classified as untimed including clonal mutations in regions with total copy number = 3 and mutant copy number ≤ 1.25, while all subclonal mutations were considered to have arisen after WGD. The timing of WGD was calculated based on the fraction of mutations occurring after WGD compared to the sum of before and after WGD[68].

Mutational signature and phylogenetic tree reconstruction

We used *Palimpsest*[69] to fit 10 known liver-related single base substitutions (SBS) signatures from Letouzé et al.[48] in our series. For patients with only one sequenced sample, we used function *palimpsest_plotTumorHistories* from *Palimpsest* to construct the natural history of tumors. For patients

ARTICLE IN PRESS

with multiple tumor samples, we reconstructed phylogenetic trees as previously described.[58] Specifically, we used a Bayesian Dirichlet process in multiple dimensions[70] to identify clusters of mutations with a similar CCF distribution across related samples in each patient, which would be assigned to the different branches of the phylogenetic tree. The phylogenetic tree was then manually reconstructed by organizing branches to fit the observed clonal composition across related samples in each patient. Driver mutations were annotated on the trees and the contribution of mutational process on each branch was quantified using Palimpsest.

RNA sequencing and transcriptomic analysis

RNAs were extracted from 406 tumors and 51 non-tumor tissues (including 135 new cases), and sequencing libraries were generated using Illumina TruSeq or Illumina TruSeq Stranded mRNA kit and sequenced by IntegraGen (Evry, France) on an Illumina HiSeq 2000 or 4000 paired-end 75 bp or 100 bp reads as described previously.[12] Fastq files were aligned to the reference human genome GRCh38 using TopHat2.[60] We removed reads mapping to multiple locations and used HTSeq[71] to obtain the number of reads associated with each gene in the Gencode v27 database, restricting reads to protein-coding genes, pseudogenes, antisense and lincRNAs (n = 58288). We used the Bioconductor *DESeq2* package[72] to import raw HTSeq counts for each sample into R statistical software and applied variance stabilizing transformation (VST) to the raw count matrix to obtain an expression matrix without variance-mean dependence. TPM scores (transcripts per kilobase million) were calculated by normalizing the count matrix for the coding length of each gene and the library size. We used the area under the ROC curve (AUC) to identify and remove genes with a significant batch effect (AUC > 0.95 between one sequencing project and others). The same method was also used to remove batch effect for external series across three projects (TCGA-LIHC, ICGC-JP and CN-Fudan).

We applied t-distributed stochastic neighbor embedding (t-SNE) to transcriptomic profiles as previous described.[12] 2000 most variably expressed genes were selected, and 1 minus the weighted Pearson correlation coefficient was used as the distance measure. Pairwise Pearson correlation was calculated using the *wtd.cors* function of the *weights* R package. We used standard deviation substracted by minimum standard deviation of selected 2000 genes as the weight, giving more variable genes greater influence. The resulting distance matrix was used to perform the t-SNE analysis using the R package *Rtsne*[73] with default parameters except the following: theta = 0, is_distance = T, pca = F, max_iter = 2000. Then tumors were annotated with mutation status of 44 driver genes and allelic status was further checked on tumor suppressor genes (TSGs) in t-SNE plot.

To examine whether the batch effect has been minimized, we first checked the t-SNE plots annotated with RNAseq projects before and after batch effect correction (Supplementary Figure 22a,b). We could find that a specific technical cluster driven by project PJ1204037 has been corrected. Furthermore, we performed principal variance component analysis (PVCA), which quantifies the variability of sources (biological, technical or other) in a given dataset, to estimate the batch effects before and after the correction. From the following bar plot, we could find that the weighted average proportion of variance attributed to RNAseq projects decreased from 11.4% to 4.5% (Supplementary Figure 22c,d). In general, the distribution of tumors with WES or WGS were evenly distributed across HCC subtypes shown in the t-SNE and bar plots (Supplementary Figure 22e-g).

Gene fusion detection

Gene fusion detection was performed as previously described[58]. We used TopHat2 to detect fusions (--fusion-search --fusion-min-dist 2000 --fusionanchor-length 13 --fusion-ignore-chromosomes chrM), TopHatFusion-post algorithm to filter and FusionInspector to validate[74]. Only fusions validated by BLAST, with at least 10 split-reads or read pairs spanning the fusion event, and an FFPM (Fusion Fragments Per Million reads) ≥ 0.1 were kept. Fusions identified recurrently in a cohort of 36 normal liver samples were

ARTICLE IN PRESS

removed, involving genes with inconsistent orientations, noncoding genes and putative read-through transcripts. Finally, fusion events involved in driver genes were manually checked.

DNA methylation array

We performed Illumina InfiniumHumanMethylation450 BeadChip arrays to obtain the methylation profile on 259 HCC and 35 non-tumor liver (including 3 new cases) as previously described[15,75]. For most cases in LICA-FR and TCGA-LIHC cohorts, methylation M1-M7 subgroup annotations were obtained from Meunier et al..[15] Annotations for the new included cases were generated using the same pipeline.[15]

Reverse-phase protein array

RPPA was performed as previously described.[38] A total of 126 antibodies targeting total and phosphorylated protein were analyzed on the 182 HCC and 25 non-tumor liver according to the protocol. All primary antibodies used in RPPA have been previously tested by Western Blotting to assess their specificity for the protein of interest. Raw data were normalized for fluorescent background per spot and a total protein stain using Normacurve.[76] Then, each RPPA slide was median centered and scaled (divided by median absolute deviation). Next, we corrected for remaining sample loadings effects on each array by individually adjusting the dependency of the data on the median value of each sample across all the arrays using a linear regression.[38] After quality control, one tumor #734T was excluded from the RPPA analysis due to the low intensity across all the antibodies (Supplementary Data 16).

Immune profiling and identification of immune hot/cold subclass

Microenvironment cell population-counter (MCP-counter) method[19] was performed on RNA-seq data to describe the immune profile of HCC using the gene list after filtering out genes expressed in hepatocytes.[10] Based on the abundance of 5 immune cell types, we classified our LICA-FR cohort into immune-hot and immune-cold classes using centroid method (Supplementary Figures 4a-e). Specifically, we ran consensus K-means clustering using Euclidean distance and Ward.D2 linkage method from the R package *ConsensusClusterPlus* with K values ranging from 2 to 8 (reps=1000, pItem=0.8, pFeature=0.8). After examining the delta area plot, the consensus matrix, and the consensus clustering heatmap under different k-value conditions, the optimal number of clusters was determined to be 5, and subsequently, the centroids of the immune hot and cold were also identified (Supplementary Figure 4b,c). According to the cosine distance to immune hot/cold centroids, we assigned other HCC into immune hot/cold classes (Supplementary Figure 4d). For HCC with qRT-PCR data only, we curated a list of 18 genes so that MCP-counter method can be also performed on qRT-PCR data (Supplementary Data 17). Using the same samples as centroids, we further classified the remaining HCC without RNA-seq data into immune hot/cold classes (Supplementary Figure 4e). Additionally, we also performed ConsensusTME method to characterize the immune profile to validate the results from MCP-counter.[77] Previously established immune-related signatures were collected to characterize the HCC immune profiles (Supplementary Data 18).

Annotations with previously established HCC molecular classifications

MS.liverK R package[78] was conducted to predict HCC transcriptomic subgroups for samples with RNAseq data available and a combination of 16-gene list was used to predict Boyault's G1-G6 classifications for samples with only qRT-PCR data[13]. Immune subgroups from Montironi et al.[16] were also applied in our cohort following the description of original literature. Immune subtypes from Thorsson et al.[20] were performed using *ImmuneSubtypeClassifier* R package.

Integration of multi-omics and definition of nine HCC molecular subtypes

To better reflect the current mutation status and characterize the predominant driver mutations within each tumor, we raised the VAF threshold to 10% for driver mutations used to identify HCC subtypes.

ARTICLE IN PRESS

The omics data from different layers were first analyzed independently. We then used the distribution of tumors in the t-SNE plot derived from RNA-seq data as a surrogate for HCC phenotype (Figure 1B). On this t-SNE plot, we incorporated information from other layers to annotate the tumors, including genome (genomic alterations), epigenome (M1M7 methylation classifications), and immune profiles. By manually examining the relationship between genomic alterations and RNA-seq clusters on the t-SNE plot, together with prior biological knowledge, we found that the phenotypes of seven HCC subtypes were primarily shaped by HCC driver mutations (Supplementary Figure 4). There was a great consistency between genomic alterations and RNA-seq clusters. Therefore, these seven HCC subtypes were defined directly by driver mutations. The remaining U^{Hot} and U^{Interm} HCC were defined according to the level of immune infiltration using the annotation from immune hot or cold subclasses. The detailed definition was shown in the Results section, "Multi-omics integrated analyses identify 9 robust molecular HCC subtypes". For certain tumors harboring multiple driver mutations, we classified them into specific subtypes based on the clustering pattern observed in the RNA-seq t-SNE plot, following the priority order: *HNF1A > CCNA2/E1 > BAP1 > TP53+CTNNB1^{ex3} > CTNNB1^{ex3}/APC > AXIN1/IRF2 > TP53*. Once the nine HCC subtypes were established, other data from additional layers such as methylation and immune classifications were used to assess their enrichment within these subtypes.

Construction and evaluation of wild-type p53 activity signature

To reduce the overfitting to our RNA-seq data, we decided to use established p53 signatures from literature to estimate wild-type (WT) p53 activity. Specifically, we collected 6 p53 transcriptomic signatures developed on human tissues.[7,22,79–82] Then, we checked the overlap of target genes between each signature, evaluated the correlation between different signature scores and took into account the score ranges and the number of targets genes included in each signature. Finally, we selected two signatures from TCGA 2017[7] and Donehower et al.[22] that had low overlap of target genes, similar score range and high correlation to generate our consensus score (Supplementary Figure 6a-f).

TCGA pan-cancer cohorts

Pan-cancer mutation data (mc3.v0.2.8.PUBLIC.maf.gz) was downloaded from https://gdc.cancer.gov/about-data/publications/pancanatlas, allelic-specific copy-number profiles were downloaded from GDC website (https://portal.gdc.cancer.gov) using *TCGAbiolinks* R package[83], and RNAseq data (GDC-PANCAN.htseq_fpkm-uq.tsv.gz) was downloaded from Xena website (https://xenabrowser.net/datapages/). Only samples with available genomic (mutations and copy-number profiles) and transcriptomic data were included. Samples flagged with quality control issues were further excluded according to the file "merged_sample_quality_annotations.tsv". Finally, 9007 patients were used in the analysis after removing patients from TCGA-LIHC cohort.

Mouse model of liver cancer

MYC-luc;sg-p53, *MYC-luc;CTNNB1* and *MYC-luc;sg-p53;CTNNB1* mice: For validation of *TP53+CTNNB1^{ex3}* HCC, we exclusively examined male mice because HCC exhibits a strong male predominance in both human and experimental models, including all three subtypes examined here. C57BL/6J male mice were purchased from Janvier Laboratories at 5 weeks of life and housed in a temperature-controlled environment with 12-hour-light/dark cycles. All animals had free access to water and standard diet (R04-10, Safe), providing 60% carbohydrate, 3% fat and 16% protein in terms of energy. After one week, 6-weeks-old C57BL/6J mice underwent plasmid delivery via hydrodynamic tail-vein injection.[84] *pT3-EF1a-MYC-IRES-luciferase* (*pT3-MYC-luc*), *pCMV-SB13*, *px330-sg-p53* and *pT3-N90-CTNNB1* plasmids were kindly provided by Dr. Amaia Lujambio (Icahn School of Medicine at Mount Sinai, New York, NY).[27,85] *pCMV-SB13*, a vector expressing SB13 transposase, is required to integrate the transposon-based vector into the genomic DNA of hepatocytes. Plasmid productions were made by

ARTICLE IN PRESS

platform in Institut de Psychiatrie et Neurosciences de Paris (IPNP). A volume of 2 mL of saline dilution containing the appropriate plasmid concentration was hydrodynamically injected through the tail vein with a 26-G needle at a rate of 0.4 mL/s. *MYC-luc;sg-p53* mouse received 2.5µg of *pCMV-SB13*, 12µg of *pT3-MYC-luc* and 10µg of *px330-sg-p53*. *MYC-luc;CTNNB1* mouse received 2.5µg of *pCMV-SB13*, 12µg of *pT3-MYC-luc* and 10µg of *pT3-N90-CTNNB1*, *MYC-luc;sg-p53;CTNNB1* mouse received 2.5µg of *pCMV-SB13*, 12µg of *pT3-MYC-luc* and 5µg of *px330-sg-p53* and 5µg of *pT3-N90-CTNNB1*. All animal procedures were performed at the CFE animal facility (Paris, France) in accordance with the guidelines of the French Ministry of Research and the 'Direction départementale de la protection des populations' (DDPP, Paris, France). Experimental protocols were approved by the Institutional Animal Care and Use Committee (IACUC) and the Ethics Committee for Animal Experimentation No. 005 (Comité d'éthique en expérimentation animale Charles Darwin) under 'Autorisation de Projet utilisant des Animaux à des Fins Scientifiques' (APAFIS) authorization #35532-2022022118422179. The maximal tumour burden permitted by the institutional ethics committee was a cumulative tumour volume of 1.2 cm . This threshold was not exceeded in any of the experiments conducted in this study. H&E staining was performed to describe the major histological features across three mouse models. Mice were euthanized if they exhibited severe clinical symptoms, including >20% body weight loss, persistent lethargy, hunched posture, rough hair coat, impaired ambulation, respiratory distress, or high tumor burden as assessed by IVIS bioluminescence imaging, exceeding predefined ethical thresholds.

Pten$^{-/-}$;*Bap1*$^{-/-}$ and *Pten*$^{-/-}$;*Bap1*$^{wt/wt}$ mice: For validation of the role of *BAP1* in functional *TP53* inactivation, we used female mice considering female patients constitute the majority in this HCC subtype. 7-weeks-old female C57BL/6N mice were purchased from Charles River Laboratories. Each group contains five mice. At 8 weeks of age, mice underwent hydrodynamic tail-vein injection with approximately 2 mL of sterile saline solution (equivalent to 10% of body weight) containing the appropriate plasmid DNA concentration. For *Pten* and *Bap1* gene disruption, mice received 20 µg of *px330-sg-Pten* and 20 µg *px330-sg-Bap1*; To achieve *Pten* gene disruption alone, 20 µg of *px330-sg-Pten* was delivered hydrodynamically, with *px330-sg-GFP* replacing *px330-sg-Bap1* as the control plasmid. Mice were palpated twice a week to monitor tumor formation. Mice were sacrificed at humane endpoints: for the *Pten*$^{-/-}$;*Bap1*$^{-/-}$ group, this was between 20-40 weeks post-injection, whereas *Pten*$^{-/-}$;*Bap1*$^{wt/wt}$ mice without sgBap1 were sacrificed at around 1 year, as only sporadic tumor nodules were observed at this time. All animal experiments were conducted in compliance with the regional regulations and with the approval of the regional board (Regierungspräsidium Karlsruhe, Germany), under the permit number: EP-Z160I02, G-178/16-01, G-270/16, G-283/16, A-23/17, G-70/20, and G-81/20. Mice were examined weekly, and daily when any sign of distress appeared. The termination criteria required immediate humane euthanasia if animals showed apathy, motor abnormalities, ascites, significant weight loss (around 20%), or an unphysiological posture. Additional liver-specific criteria included visibly enlarged abdomen, palpable tumor mass associated with pain, ascites, jaundice, constipation, or bloody feces. The maximal tolerated tumor burden was therefore defined by these clinical criteria rather than by absolute size, and in our experiments the permitted tumor burden was not exceeded. The success of CRISPR/Cas9-mediated editing was validated using T7 endonuclease I assay and immunochemistry for the targeted protein with the subsequent RNAseq conducted at the Genomics and Proteomics Core Facility of the DKFZ.[86]

Bioluminescence detection

Luciferase was quantified by bioluminescence at different time points. For this purpose, mice were anesthetized under isofluorane (ISO-VET®, 100% - 1000mg/g) in an induction chamber (TEMSEGA, Minihub® V2.1) and injected with 100µl of D-luciferine i.p. (Thermo Scientific) at a concentration of 30 mg/mL. Mice were placed in the imaging chamber of a IVIS Lumina II (Caliper Life Sciences). A color-

ARTICLE IN PRESS

scale photograph of the animals was acquired, followed by a bioluminescent acquisition. Regions of interest were drawn over the animal image, as well as over regions of no signal, which were used as background readings. Light intensity was quantified using photons/s. The color-scale photograph and data images from all studies were superimposed using Living Image software.

Western blotting

Western blot was performed as previously described.[87] The antibodies used for western blot analysis included, p53 antibody (Cell Signaling Technology, rabbit monoclonal, cat#32532), Beta catenin antibody (BD Biosciences, mouse monoclonal, cat#610154), and HSC-70 antibody (Santa Cruz Biotechnology, mouse monoclonal, cat#7298).

Statistics & Reproducibility

All analyses were performed using R (version 4.1, http://www.R-project.org) or GraphPad Prism (version 5). Generally, wilcoxon's rank-sum test, two-sided fisher's exact test or χ^2 test was applied for intergroup comparisons according to the type of variable. For some variables, we used a trend test or Student's t-test, as specifically mentioned in the text or figure legend. Survival analysis was performed using Kaplan-Meier method and log-rank test. A p value < 0.05 was considered statistically significant. No statistical method was used to predetermine sample size. No data was excluded from the analysis. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

Data availability

The sequencing data of the LICA-FR cohort reported in this paper have been deposited to the EGA (European Genome-phenome Archive) database (WGS accessions: [EGAS00001007957] [https://ega-archive.org/studies/EGAS00001007957], [EGAS00001007694] [https://ega-archive.org/studies/EGAS00001007694], [EGAS00001002408] [https://ega-archive.org/studies/EGAS00001002408], [EGAS00001002888] [https://ega-archive.org/studies/EGAS00001002888], [EGAS00001003837] [https://ega-archive.org/studies/EGAS00001003837], [EGAS00001004629] [https://ega-archive.org/studies/EGAS00001004629], [EGAS00001005629] [https://ega-archive.org/studies/EGAS00001005629], [EGAS00001005986] [https://ega-archive.org/studies/EGAS00001005986]; WES accessions: [EGAS00001007957] [https://ega-archive.org/studies/EGAS00001007957], [EGAS00001007694] [https://ega-archive.org/studies/EGAS00001007694], [EGAS00001000217] [https://ega-archive.org/studies/EGAS00001000217], [EGAS00001000679] [https://ega-archive.org/studies/EGAS00001000679], [EGAS00001001002] [https://ega-archive.org/studies/EGAS00001001002], [EGAS00001003063] [https://ega-archive.org/studies/EGAS00001003063], [EGAS00001003130] [https://ega-archive.org/studies/EGAS00001003130], [EGAS00001003837] [https://ega-archive.org/studies/EGAS00001003837], [EGAS00001004629] [https://ega-archive.org/studies/EGAS00001004629]; RNAseq accession: [EGAS00001007957] [https://ega-archive.org/studies/EGAS00001007957], [EGAS00001007694] [https://ega-archive.org/studies/EGAS00001007694], [EGAS00001001284] [https://ega-archive.org/studies/EGAS00001001284], [EGAS00001002879] [https://ega-archive.org/studies/EGAS00001002879], [EGAS00001003310] [https://ega-archive.org/studies/EGAS00001003310], [EGAS00001003837] [https://ega-archive.org/studies/EGAS00001003837], [EGAS00001004629] [https://ega-archive.org/studies/EGAS00001004629], [EGAS00001005986] [https://ega-archive.org/studies/EGAS00001005986]). The DNA methylation data have been deposited to the Gene

ARTICLE IN PRESS

Expression Omnibus database (accession number: GSE157341 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE157341] and GSE56588 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56588]). RPPA data has been attached in Supplementary Data 8 and was also available on figshare (https://doi.org/10.6084/m9.figshare.30927281). All processed molecular data could be also downloaded from our shiny website (https://zucmanlab.com/hccsubtype/). The ways to download public data were shown in Supplementary Data 1. Source data are provided with this paper.

Code availability

The data was generated as described in the Methods part. All the packages used to generate the results are public and available freely. The code to reproduce the principal figures was archived on Zenodo (https://doi.org/10.5281/zenodo.17236540) under GNU General Public License.

References

ARTICLE IN PRESS

1. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *Ca Cancer J Clin* **71**, 209–249 (2021).

2. Llovet, J. M. *et al.* Hepatocellular carcinoma. *Nat Rev Dis Primers* **7**, 6 (2021).

3. Finn, R. S. *et al.* Atezolizumab plus Bevacizumab in Unresectable Hepatocellular Carcinoma. *New Engl J Med* **382**, 1894–1905 (2020).

4. Totoki, Y. *et al.* Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. *Nat Genet* **46**, 1267–1273 (2014).

5. Ahn, S. *et al.* Genomic portrait of resectable hepatocellular carcinomas: Implications of RB1 and FGF19 aberrations for patient stratification. *Hepatology* **60**, 1972–1982 (2014).

6. Fujimoto, A. *et al.* Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat Genet* **48**, 500–509 (2016).

7. Ally, A. *et al.* Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* **169**, 1327-1341.e23 (2017).

8. Gao, Q. *et al.* Integrated Proteogenomic Characterization of HBV-Related Hepatocellular Carcinoma. *Cell* **179**, 561-577.e22 (2019).

9. Nault, J. *et al.* Clinical Impact of Genomic Diversity From Early to Advanced Hepatocellular Carcinoma. *Hepatology* **71**, 164–182 (2020).

10. Hirsch, T. Z. *et al.* BAP1 mutations define a homogeneous subgroup of hepatocellular carcinoma with fibrolamellar-like features and activated PKA. *J Hepatol* **72**, 924–936 (2020).

ARTICLE IN PRESS

11. Schulze, K. *et al.* Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet* **47**, 505–511 (2015).

12. Bayard, Q. *et al.* Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrangement signature of replication stress. *Nat Commun* **9**, 5235 (2018).

13. Boyault, S. *et al.* Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. *Hepatology* **45**, 42–52 (2007).

14. Hoshida, Y. *et al.* Integrative Transcriptome Analysis Reveals Common Molecular Subclasses of Human Hepatocellular Carcinoma. *Cancer Res* **69**, 7385–7392 (2009).

15. Meunier, L. *et al.* DNA Methylation Signatures Reveal the Diversity of Processes Remodeling Hepatocellular Carcinoma Methylomes. *Hepatology* (2021).

16. Montironi, C. *et al.* Inflamed and non-inflamed classes of HCC: a revised immunogenomic classification. *Gut* gutjnl-2021-325918 (2022).

17. Sia, D. *et al.* Identification of an Immune-specific Class of Hepatocellular Carcinoma, Based on Molecular Features. *Gastroenterology* **153**, 812–826 (2017).

18. Jiang, Y. *et al.* Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature* **567**, 257–261 (2019).

19. Becht, E. *et al.* Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* **17**, 218 (2016).

20. Thorsson, V. *et al.* The Immune Landscape of Cancer. *Immunity* **48**, 812-830.e14 (2018).

21. Chiang, D. Y. *et al.* Focal Gains of VEGFA and Molecular Classification of Hepatocellular Carcinoma. *Cancer Res* **68**, 6779–6788 (2008).

22. Donehower, L. A. *et al.* Integrated Analysis of TP53 Gene and Pathway Alterations in The Cancer Genome Atlas. *Cell Reports* **28**, 1370-1384.e5 (2019).

23. Kastenhuber, E. R. & Lowe, S. W. Putting p53 in Context. *Cell* **170**, 1062–1078 (2017).

24. Guichard, C. *et al.* Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat Genet* **44**, 694–698 (2012).

25. Abitbol, S. *et al.* AXIN deficiency in human and mouse hepatocytes induces hepatocellular carcinoma in the absence of β-catenin activation. *J Hepatol* **68**, 1203–1213 (2018).

26. Rebouissou, S. *et al.* Genotype-phenotype correlation of CTNNB1 mutations reveals different ß-catenin activity associated with liver tumor progression. *Hepatology Baltim Md* **64**, 2047–2061 (2016).

ARTICLE IN PRESS

27. Molina-Sánchez, P. *et al.* Cooperation Between Distinct Cancer Driver Genes Underlies Intertumor Heterogeneity in Hepatocellular Carcinoma. *Gastroenterology* **159**, 2203-2220.e14 (2020).

28. Bayard, Q. *et al.* Structure, Dynamics, and Impact of Replication Stress–Induced Structural Variants in Hepatocellular Carcinoma. *Cancer Res* **82**, OF1–OF12 (2022).

29. Revsine, M. *et al.* Lineage and ecology define liver tumor evolution in response to treatment. *Cell Rep. Med.* 101394 (2024).

30. Liu, X. *et al.* Tumor phylogeography reveals block-shaped spatial heterogeneity and the mode of evolution in Hepatocellular Carcinoma. *Nat. Commun.* **15**, 3169 (2024).

31. Nguyen, P. H. D. *et al.* Trajectory of immune evasion and cancer progression in hepatocellular carcinoma. *Nat. Commun.* **13**, 1441 (2022).

32. Limousin, W. *et al.* Molecular-based targeted therapies in patients with hepatocellular carcinoma and hepato-cholangiocarcinoma refractory to atezolizumab/bevacizumab. *J. Hepatol.* **79**, 1450–1458 (2023).

33. Zauderer, M. G. *et al.* EZH2 inhibitor tazemetostat in patients with relapsed or refractory, BAP1-inactivated malignant pleural mesothelioma: a multicentre, open-label, phase 2 study. *Lancet Oncol.* **23**, 758–767 (2022).

34. George, T. J. *et al.* Results of a phase II trial of the PARP inhibitor, niraparib, in BAP1 and other DNA damage response pathway deficient neoplasms. *J. Clin. Oncol.* **40**, 3122–3122 (2022).

35. Rialdi, A. *et al.* WNTinib is a multi-kinase inhibitor with specificity against β-catenin mutant hepatocellular carcinoma. *Nat. Cancer* 1–19 (2023).

36. Xiao, Y. *et al.* Combining p53 mRNA nanotherapy with immune checkpoint blockade reprograms the immune microenvironment for effective cancer therapy. *Nat Commun* **13**, 758 (2022).

37. Dauch, D. *et al.* A MYC–aurora kinase A protein complex represents an actionable drug target in p53-altered liver cancer. *Nat Med* **22**, 744–753 (2016).

38. Caruso, S. *et al.* Analysis of Liver Cancer Cell Lines Identifies Agents With Likely Efficacy Against Hepatocellular Carcinoma and Markers of Response. *Gastroenterology* **157**, 760–776 (2019).

39. Decaens, T. *et al.* Phase 1b/2 trial of tepotinib in sorafenib pretreated advanced hepatocellular carcinoma with MET overexpression. *Br. J. Cancer* **125**, 190–199 (2021).

ARTICLE IN PRESS

ARTICLE IN PRESS

40. Schaeffer, S. *et al.* RSK2 inactivation cooperates with AXIN1 inactivation or ß-catenin activation to promote hepatocarcinogenesis. *J. Hepatol.* (2023).

41. Brown, V. *et al.* Abstract 2306: BLU-222, an investigational, potent, and selective CDK2 inhibitor, demonstrated robust antitumor activity in CCNE1 -amplified ovarian cancer models. *Cancer Res.* **82**, 2306–2306 (2022).

42. Galle, P. R. *et al.* Biology and significance of alpha-fetoprotein in hepatocellular carcinoma. *Liver Int.* **39**, 2214–2229 (2019).

43. Parikh, N. D., Tayob, N. & Singal, A. G. Blood-based biomarkers for hepatocellular carcinoma screening: Approaching the end of the ultrasound era? *J. Hepatol.* **78**, 207–216 (2023).

44. Tong, M. *et al.* Loss of tyrosine catabolic enzyme HPD promotes glutamine anaplerosis through mTOR signaling in liver cancer. *Cell Reports* **36**, 109617 (2021).

45. Yang, X. *et al.* Precision treatment in advanced hepatocellular carcinoma. *Cancer Cell* **42**, 180–197 (2024).

46. Lehrich, B. M. *et al.* Development of mutated β-catenin gene signature to identify CTNNB1 mutations from whole and spatial transcriptomic data in patients with HCC. *JHEP Rep.* **6**, 101186 (2024).

47. Furuta, M. *et al.* Whole genome sequencing discriminates hepatocellular carcinoma with intrahepatic metastasis from multi-centric tumors. *J Hepatol* **66**, 363–373 (2017).

48. Letouzé, E. *et al.* Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat Commun* **8**, 1315 (2017).

49. Nault, J. C. *et al.* High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions. *Nat Commun* **4**, 2218 (2013).

50. Molina, L. *et al.* Biallelic hydroxymethylbilane synthase inactivation defines a homogenous clinico-molecular subtype of hepatocellular carcinoma. *J Hepatol* (2022).

51. Leroy, B., Anderson, M. & Soussi, T. TP53 Mutations in Human Cancer: Database Reassessment and Prospects for the Next Decade. *Hum. Mutat.* **35**, 672–688 (2014).

52. Bella, T. L. *et al.* Adeno-associated virus in the liver: natural history and consequences in tumour development. *Gut* **69**, 737–747 (2019).

53. Péneau, C. *et al.* Hepatitis B virus integrations promote local and distant oncogenic driver alterations in hepatocellular carcinoma. *Gut* gutjnl-2020-323153 (2021).

ARTICLE IN PRESS

ARTICLE IN PRESS

54. Kan, Z. *et al.* Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res* **23**, 1422–1433 (2013).

55. Ng, C. K. Y. *et al.* Integrative proteogenomic characterization of hepatocellular carcinoma across etiologies and stages. *Nat Commun* **13**, 2436 (2022).

56. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).

57. Cannataro, V. L., Gaffney, S. G. & Townsend, J. P. Effect Sizes of Somatic Mutations in Cancer. *JNCI: J. Natl. Cancer Inst.* **110**, 1171–1177 (2018).

58. Hirsch, T. Z. *et al.* Integrated genomic analysis identifies driver genes and cisplatin-resistant progenitor phenotype in pediatric liver cancer. *Cancer Discov* **11**, candisc.1809.2020 (2021).

59. Chen, X. *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).

60. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36 (2013).

61. Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res* **44**, e131–e131 (2016).

62. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–572 (2004).

63. Locallo, A., Prandi, D., Fedrizzi, T. & Demichelis, F. TPES: Tumor Purity Estimation from SNVs. *Bioinformatics* **35**, 4433–4435 (2019).

64. Alfaro-Murillo, J. A. & Townsend, J. P. Pairwise and higher-order epistatic effects among somatic cancer mutations across oncogenesis. *Math. Biosci.* **366**, 109091 (2023).

65. Dewhurst, S. M. *et al.* Tolerance of Whole-Genome Doubling Propagates Chromosomal Instability and Accelerates Cancer Genome Evolution. *Cancer Discov* **4**, 175–185 (2014).

66. Bielski, C. M. *et al.* Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat Genet* **50**, 1189–1195 (2018).

67. López, S. *et al.* Interplay between whole-genome doubling and the accumulation of deleterious alterations in cancer evolution. *Nat Genet* **52**, 283–293 (2020).

68. Jamal-Hanjani, M. *et al.* Tracking the Evolution of Non–Small-Cell Lung Cancer. *New Engl J Medicine* **376**, 2109–2121 (2017).

ARTICLE IN PRESS

ARTICLE IN PRESS

69. Shinde, J. *et al.* Palimpsest: an R package for studying mutational and structural variant signatures along clonal evolution in cancer. *Bioinformatics* **34**, 3380–3381 (2018).

70. Yates, L. R. *et al.* Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med* **21**, 751–759 (2015).

71. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

72. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

73. Maaten, L. van der & Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* **86**, 2579−2605.

74. Haas, B. J. *et al.* Targeted in silico characterization of fusion transcripts in tumor and normal tissues via FusionInspector. *Cell Rep. Methods* **3**, 100467 (2023).

75. Villanueva, A. *et al.* DNA methylation-based prognosis and epidrivers in hepatocellular carcinoma. *Hepatology* **61**, 1945–1956 (2015).

76. Troncale, S. *et al.* NormaCurve: A SuperCurve-Based Method That Simultaneously Quantifies and Normalizes Reverse Phase Protein Array Data. *PLoS ONE* **7**, e38686 (2012).

77. Jiménez-Sánchez, A., Cast, O. & Miller, M. L. Comprehensive Benchmarking and Integration of Tumor Microenvironment Cell Estimation Methods. *Cancer Res* **79**, 6238–6246 (2019).

78. Petitprez, F. *et al.* MS.liverK: an R package for transcriptome-based computation of molecular subtypes and functional signatures in liver cancer. *Biorxiv* 540005 (2019).

79. Nguyen, T.-A. T. *et al.* Revealing a human p53 universe. *Nucleic Acids Res* **46**, gky720-(2018).

80. Fischer, M. Census and evaluation of p53 target genes. *Oncogene* **36**, 3943–3956 (2017).

81. Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D. & Saez-Rodriguez, J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res* **29**, 1363–1375 (2019).

82. Andrysik, Z. *et al.* Identification of a core TP53 transcriptional program with highly distributed tumor suppressive activity. *Genome Res* **27**, 1645–1657 (2017).

83. Colaprico, A. *et al.* TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* **44**, e71–e71 (2016).

ARTICLE IN PRESS

ARTICLE IN PRESS

84. Chen, X. & Calvisi, D. F. Hydrodynamic Transfection for Generation of Novel Mouse Models for Liver Cancer Research. *Am. J. Pathol.* **184**, 912–923 (2014).

85. Galarreta, M. R. de *et al.* β-Catenin Activation Promotes Immune Escape and Resistance to Anti–PD-1 Therapy in Hepatocellular Carcinoma. *Cancer Discov* **9**, 1124–1141 (2019).

86. Breinig, M. *et al.* Multiplexed orthogonal genome editing and transcriptional activation by Cas12a. *Nat. Methods* **16**, 51–54 (2019).

87. Donne, R. *et al.* Replication stress triggered by nucleotide pool imbalance drives DNA damage and cGAS-STING pathway activation in NAFLD. *Dev. Cell* **57**, 1728-1741.e6 (2022).

ARTICLE IN PRESS

ARTICLE IN PRESS

Acknowledgements

We warmly thank the clinicians, surgeons, pathologists, hepatologists, and oncologists who contributed to the tissue collection and clinical annotations. We thank Eric Letouzé for his advice about mutational signatures and Barkha Gupta for the annotations of pathological features from several HE slides. FunGeST team (Functional Genomics of Solid Tumors) is supported by the Ligue Nationale contre le Cancer (Equipe Labellisée), Labex OncoImmunology (investissement d'avenir, France 2030), the SIRIC CARPEM (INCa), and THRIVE EU Horizon funding Mission Cancer (EU-HORIZON-MISS-2023-CAN-CER-01). L.P., J.F. and S.S.J. were supported by fellowships from China Scholarship Council. J.M.L. is supported by grants from the European Commission (Horizon Europe-Mission Cancer, THRIVE, Ref. 101136622), by an Accelerator Award from Cancer Research UK, Fondazione per la Ricerca sul Cancro (AIRC) and Fundación Científica de la Asociación Española Contra el Cáncer -FAECC (HUNTER, Ref. C9380/A26813), by the NIH (R01-CA273932-01, RO1DK56621 and RO1DK128289), the Samuel Waxman Cancer Research Foundation, by the Spanish National Health Institute (Project PID2022-139365OB-I00), funded by MICIU/AEI/10.13039/501100011033 and FEDER; by the Fundación Científica de la Asociación Española Contra el Cáncer - FAECC (Proyectos Generales: PRYGN223117LLOV); Fundación Científica de la Asociación Española Contra el Cáncer - FAECC- Reto AECC 70% Supervivencia (RETOS245779LLOV), the Generalitat de Catalunya (AGAUR, 2021-SGR 01347), AECC Excellence Program (EPAEC246711CLIN) and from "la Caixa" Foundation under agreement LCF/PR/SP23/52950009. We thank IntegraGen, Macrogen France and the Next Generation Sequencing unit of the Genomics and Proteomics Core Facility, German Cancer Research Center (DKFZ), for providing excellent sequencing services.

Author contributions

L.P.: conceptualization, investigation, visualization, writing-original draft. T.Z.H.: data curation, methodology and writing-review & editing. J.F.: methodology, investigation and visualization. A.S.: investigation and data curation. S.I.: data curation and investigation. S.S.J.: investigation. O.O.G.: investigation, data curation and resources. S.R.: resources and writing-review & editing. S.S.: resources. N.A.: visualization. J.C.: resources. G.A.: resources. J.F.B.: resources. B.L.B: resources. M.Z.: resources. S.C.M.: investigation and writing-review & editing. V.P.: resources. J.M.L.: resources and writing-review & editing. D.F.T.: investigation and resources. J.C.N.: resources and writing-review & editing. C.D.: investigation and writing-review & editing. S.C.: conceptualization, investigation, visualization, supervision and writing-original draft. J.Z.R.: conceptualization, investigation, visualization, funding acquisition, supervision and writing-original draft.

Competing interests:

J.M.L. received research support from Genentech and Roche; consultancy or lecture fees from Eisai Inc., Merck, Roche, Genentech, AstraZeneca, Bayer Pharmaceuticals, Abbvie, Sanofi, Moderna, Glycotest, Exelixis, and Boehringer Ingelheim; and served on a Data Safety Monitoring Board for Bristol Myers Squibb. J.C.N. received grants from Ipsen and Bayer. The remaining authors declare no competing interests.

ARTICLE IN PRESS

ARTICLE IN PRESS

Figure legend:

Figure 1. Definition of 9 molecular subtypes by integrating genomic and transcriptomic data in LICA-FR Cohort. (**A**) Interaction network of HCC driver gene alterations in 489 genomically independent HCC sequenced by WES or WGS. p value was calculated using one-sided fisher's exact test and adjusted using Benjamini-Hochberg method. (**B**) t-SNE plot of 406 HCC RNAseq samples annotated by driver gene mutations. Each dot represents one tumor. (**C**) Representation on the 406 HCC RNAseq t-SNE plot of the total immune infiltrates and individual immune cell population including T cell and NK cell using MCP-counter[19]. (**D-G**) t-SNE plots of 406 HCC RNAseq samples annotated by: (D) immune subgroups (Montironi et al.[16]), (E) immune subtypes (Thorsson et al.[20]), (F) transcriptomic subgroups (Boyault et al.[13]), and (G) methylation subgroups (Meunier et al.[15]). Small grey dots represent HCC without methylation data. (**H**) t-SNE plot depicting distinct transcriptomic clusters shaped by driver gene clusters in TCGA-LIHC/ICGC-JP/CN-Fudan external cohorts (n = 727). Using the same annotations as in (B). (**I**) Molecular characterization of the 9 molecular HCC subtypes. WGS, whole genome sequencing; WES, whole exome sequencing; t-SNE, t-distributed stochastic neighbor embedding. [a], p value was calculated using two-sided wilcoxon's rank-sum test or fisher's exact test, and was adjusted using Benjamini-Hochberg method within each feature (details in Supplementary Data 14). [b], Total immune infiltrates and 6 cell types were assessed by MCP-counter tool[18]. [$], not mutated for *CTNNB1*, *AXIN1*, *CCNA2/E1*, *BAP1*, *HNF1A*, *IRF2*. [$$], biallelic alterations; TIIs, total immune infiltrates. ****, adjusted p< 0.0001; ***, adjusted p< 0.001; **, adjusted p< 0.01; *, adjusted p< 0.05.

Figure 2. **Modulation of functional wild-type p53 activity in HCC.** (**A**) Association of p53 activity score (on the top) with genomic instability, HCC subtypes, driver gene somatic alterations, transcriptomic subgroups[13] and clinical features (two-sided Wilcoxon's rank-sum test). 51 non-tumor liver (NTL) are represented on the left and 406 HCC from LICA-FR cohort on the right. (**B**) Association of p53 activity score with clinical features including BCLC stage, serum AFP level, Edmonson grade, and tumor size using one-sided Jonckheere's trend test. (**C**) Analysis of the p53 activity score according to cancer driver mutations revealed that gene alterations in *AXIN1*, *RB1*, *CCNA2/E1*, *PTEN*, *BAP1* an *IRF2* are associated with low p53 activity score in HCC. Cooperation and additive effect with *TP53* alterations were identified for either *AXIN1* or *RB1* mutations. No additive effect with *TP53* alterations was identified for *CCNA2/E1*, *PTEN*, *BAP1* or *IRF2* alterations. p value was calculated using two-sided Wilcoxon's rank-sum test. (**D**) Validation of the effect of *Bap1* knock-out on p53 activities in *Pten*$^{-/-}$ liver tumor mouse model (two-sided Student's t-test). Each point represents one tumor tissue from one different mouse. Source data are provided as a Source Data file. Boxplots (panels B-C) show the median (center line), 25th, and 75th percentile (lower and upper boundary). The whiskers extend to the minimum and maximum values within 1.5× the interquartile range. CNA, copy-number alteration; WG, whole genome; AFB1, aflatoxin B1; FA, focal amplification; HD, homozygous deletion; SV, structural variation; BCLC, barcelona clinic liver cancer; M, mutated; NM, not mutated; NTL, non tumor liver; ****, p< 0.0001; ***, p< 0.001; **, p< 0.01; *, p< 0.05. ns, not significant.

Figure 3. Functional WNT/ß-catenin activity score is related to *CTNNB1*, *AXIN1*, *APC* and *TP53* in HCC LICA-FR cohort. (**A**) Association of ß-catenin activity score (see Methods) with *CTNNB1* mutations

ARTICLE IN PRESS

at different positions (exon7/8, S45, T41, multiple hits, inframe exon 3 deletion and D32-S37), *AXIN1* and *APC* mutations in RNAseq data of 406 HCC. n, number of samples; ****, p = 6.76e-08; **, p = 0.01. (**B**) Association of ß-catenin activity score (on the top) with HCC subtypes, transcriptomic subgroups[13] and driver gene alterations. 51 non-tumor liver (NTL) are represented on the left and 406 HCC from LICA-FR cohort on the right. #, G6 transcriptomic subgroup versus G5 subgroup. ****, ####, p< 0.0001; **, p< 0.01. (**C**) Upregulation of ß-catenin activity score in *TP53*-mutated HCC in both LICA-FR and external cohorts. Left panel: **, p = 0.007; *, p = 0.016. Right panel: **, p = 0.001. Statistical analysis: p value was calculated using two-sided Wilcoxon's rank-sum test in (A), (B) and (C). Source data are provided as a Source Data file. Boxplots (panels A, C) show the median (center line), 25th, and 75th percentile (lower and upper boundary). The whiskers extend to the minimum and maximum values within 1.5× the interquartile range. $$, biallelic inactivation; M, mutated; NM, not mutated; NTL, non-tumor liver; n, number of samples.

Figure 4. Co-occurring *TP53* and *CTNNB1* exon 3 mutations define a rare subtype of HCC with poor prognosis. (**A**) A summary plot showing the correlation between clinicopathological and molecular features and each HCC subgroup (*TP53+CTNNB1*^{ex3} HCC (n=30), *TP53*^{WT}*CTNNB1*^{ex3} HCC (n=122), *TP53*^{mut}*CTNNB1*^{WT} HCC (n=144)) compared to *TP53*^{WT}*CTNNB1*^{WT} HCC (n=233) in LICA-FR cohort (two-sided Wilcoxon's rank-sum test or fisher's exact test, details in Supplementary Data 20). (**B**) Kaplan-Meier plot showing progression free survival (PFS) or disease specific survival (DSS) according to HCC subgroup (*TP53+CTNNB1*^{ex3} HCC, *TP53*^{WT}*CTNNB1*^{ex3} HCC, *TP53*^{mut}*CTNNB1*^{WT} HCC, *TP53*^{WT}*CTNNB1*^{WT} HCC) in LICA-FR cohort. *TP53+CTNNB1*^{ex3} HCC is associated with poor PFS and DSS. p-value was calculated by log-rank test. (**C**) Associations of *TP53+CTNNB1*^{ex3} HCC with poor PFS and overall survival were validated in TCGA/ICGC-JP/CN-Fudan external cohorts. (**D**) Experimental scheme to reproduce HCC molecular subtypes in mice using hydrodynamic injection with different combinations of plasmids (see Methods, ref.[86]) to inactivate *Trp53* (*px330-sg-p53*), activate β-catenin (*pT3-N90-CTNNB1*) and express *MYC* (*pT3-MYC-luc*). *pCMV-SB13*, a vector expressing SB13 transposase, is required to integrate the transposon-based vector into the genomic DNA of hepatocytes. (**E**) Representative images of tumor development obtained with in vivo imaging system (IVIS) in different HCC models over time. IVIS data were unavailable for mice that have been euthanized due to high tumor burden. (**F**) The mean value and standard deviation of bioluminescence intensity in the liver among different HCC models over time (one-way ANOVA analysis). Source data are provided as a Source Data file. Data are presented as mean values +/- SEM. (**G**) Kaplan-Meier plot showing overall survival of the mice according to the different HCC models. The red line represents *MYC-luc;sg-p53* mice; The blue line represents *MYC-luc;CTNNB1* mice; The slate green line represents *Myc-luc;sg-p53;CTNNB1* mice. #, median survival time. p-value was calculated by log-rank test. *Myc-luc;sg-p53;CTNNB1 vs. MYC-luc;sg-p53, p*= 0.0004; *Myc-luc;sg-p53;CTNNB1 vs. MYC-luc;CTNNB1, p*= 0.0001; n.s., not significant.

Figure 5. **Clonal evolution of *TP53+CTNNB1*^{ex3} HCC.** (**A**) Spatial and longitudinal evolutions were reconstructed for 4 HCC patients with primary tumors and/or relapses/metastases using phylogenetic trees. Basic information of tumors/patients including geographic origin, etiology, metavir fibrosis score (F0-F4), age at sampling, treatment history and transcriptomic subgroup[13] (G1-G6) is shown. Branch lengths are proportional to the total number of mutations, and branches to the left of the dashed lines represent clonal events, whereas those to the right indicate subclonal events, with a color code indicating the contribution of each mutational signature. Driver mutations were assigned to specific branches and were additionally timed if samples have WGD events. (**B**) CINscore, frequency of whole genome duplication, and immune infiltrate score (calculated by MCP-counter[19]) were compared across the three HCC subgroups: *CTNNB1*^{ex3}, *TP53*-mutated and *TP53+CTNNB1*^{ex3}. Median values are indicated on "CINscore" and "Frequency of

ARTICLE IN PRESS

whole genome duplication". ****, p = 1.84e-09; ***, p = 0.0005; **, p = 0.003; *, p = 0.045. One-sided Jonckheere's, One-sided Cochran-Armitage and two-sided wilcoxon's rank-sum tests were used to compare "CINscore", "Frequency of whole genome duplication", and immune infiltrate respectively. Source data are provided as a Source Data file. (**C**) Scheme of evolutionary model for *TP53+CTNNB1*^{ex3} HCC. WGD, whole genome duplication; CINscore, chromosome instability score; cnLOH, copy-neutral loss of heterozygosity; WGS, whole genome sequencing; WES, whole exome sequencing. SBS, single base substitutions. TACE, transarterial chemoembolization.

Figure 6. Correlation between HCC subtypes with clinicopathological and molecular features. (**A**) Distribution of HCC subtypes according to risk factors, (**B**) gender, (**C**) fibrosis score and Edmondson grade. (**D**) Difference of HCC subtypes in age at sampling, serum AFP level (ng/ml), tumor size (cm), total immune infiltrate score calculated by MCP-counter[19], CINscore, number of copy-number segments and frequency of whole genome duplication. The median value of features in each HCC subtype is represented. Source data are provided as a Source Data file. (**E**) Kaplan-Meier plot showing U^{Hot} HCC (pink) compared to all the other subtypes (grey) is associated with good progression-free survival and disease-specific survival in LICA-FR cohort. (**F**) Kaplan-Meier plot showing U^{Hot} HCC is associated with good-progression free survival and overall survival in TCGA-LIHC/ICGC-JP/CN-Fudan external cohorts. Statistic analysis: The Kruskal-Wallis test and two-sided Chi-squared test were used to calculate p value for continuous features and categorical features in (A), (B), (C) and (D), respectively. p value was calculated by log-rank test in (E) and (F). HBV, hepatitis B; HCV, hepatitis C; MASH, metabolic dysfunction-associated steatohepatitis.

Figure 7. Integrated nosology of hepatocellular carcinoma. Only significant associations with clinical and molecular features were shown in the plot. *, transcriptomic subgroups are derived from Boyault et al. (G1-G6)[13], Hoshida et al. (S1-S3)[14] and Chiang et al. (five classes: CTNNB1, Proliferation, Interferon, Poly 7, and Unannotated)[21], and methylation subgroups are derived from Meunier et al. (M1-M7)[15]. Immune subgroups derived from Montironi et al.[16] included Immune active, immune exhausted, Immune-like, Intermediate and Excluded. Immune subtypes derived from Thorsson et al.[20] included wound healing, IFN-γ dominant, inflammatory, lymphocyte depleted, immunologically quiet and TGF-β dominant. Immune infiltrate score was calculated using MCP-counter[19]. Dark red line represents p53 activity score, purple line represents β-catenin activity score and pink line represents immune infiltrate score. #, potential and promising strategies are selected after extensive literature searching regardless of cancer types.

ARTICLE IN PRESS

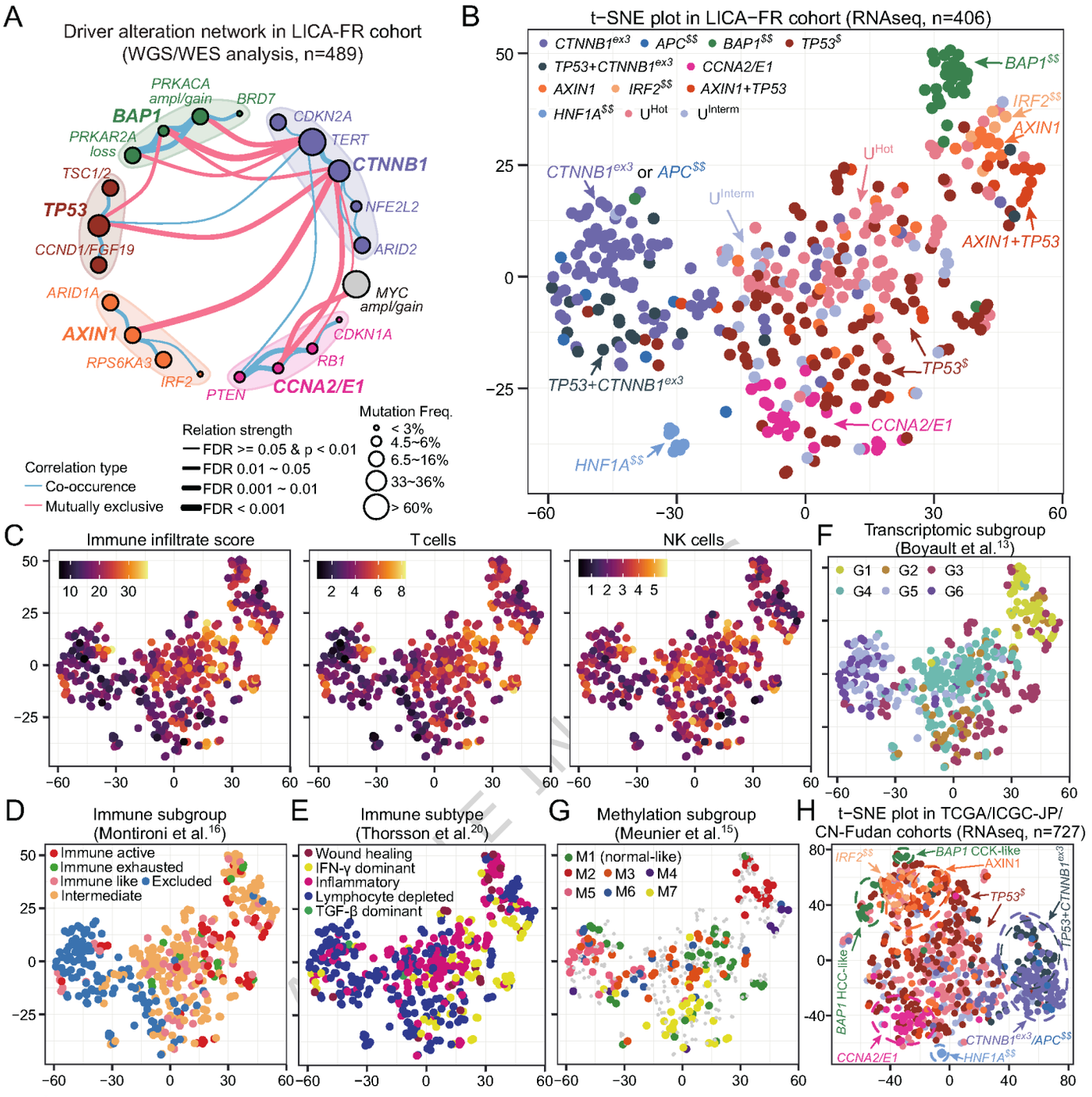
Editor Summary:

Precision oncology for hepatocellular carcinoma (HCC) remains challenging due to the high diversity of HCC at the clinical and molecular levels. Here, the authors employ an integrative multi-omics approach to identify and characterize nine molecular subtypes and explore their biological and clinical significance via p53/WNT scoring, immune infiltration analyses, and in vivo models.

Peer Review Information:

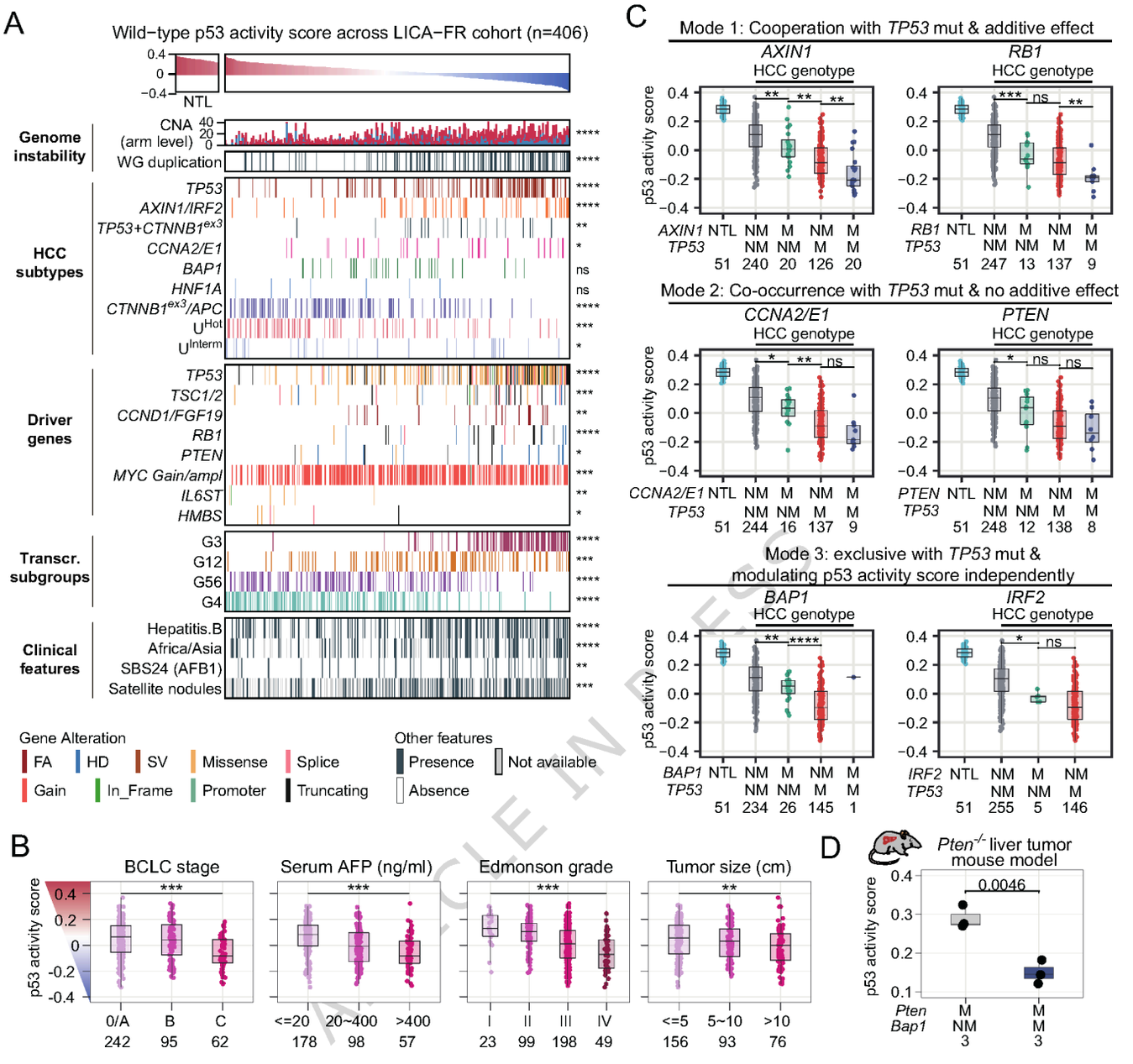
Nature Communications thanks Alexander Scheiter and the other anonymous reviewers for their contribution to the peer review of this work. [A peer review file is available.]

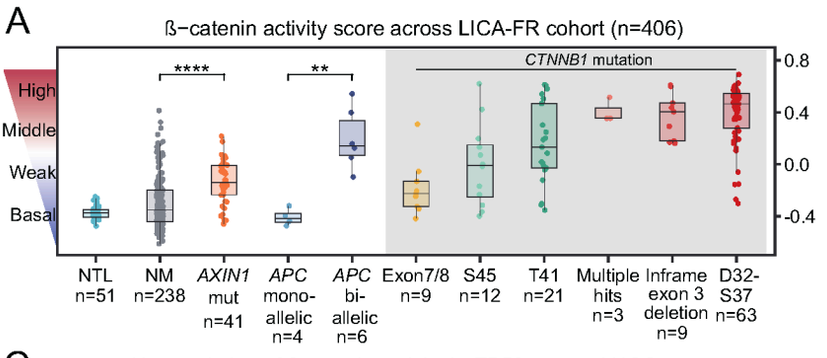
ARTICLE IN PRESS

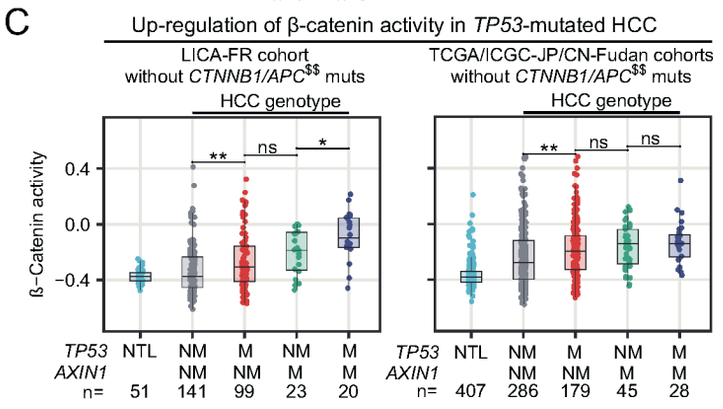


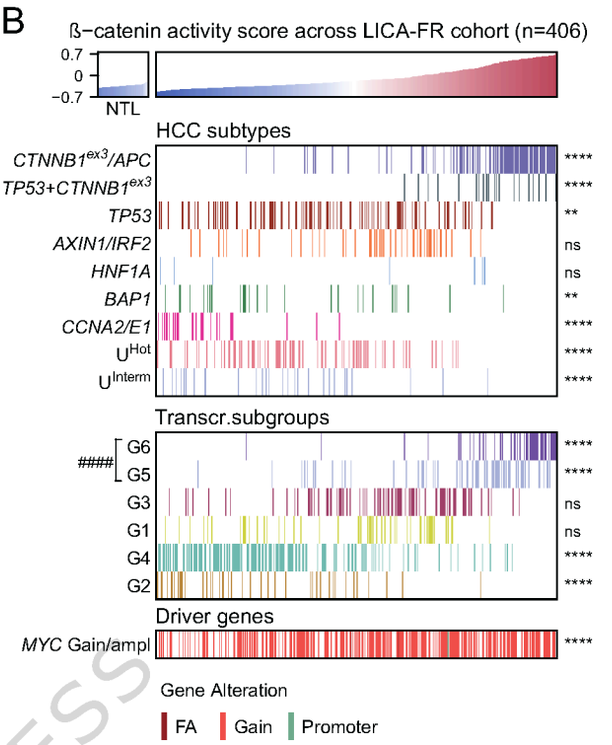
I

	Definition (driver genes)	Freq %	Associated driver mutations[a]	Transcr. subgroup[a,13]	Methyl. subgroup[a,15]	Immune profiles	
						Cell types[a, b]	Immune subtypes[a,16,20]
Common subtypes	*CTNNB1ex3/APC$$* (n=125)	25%	*TERT***, *NFE2L2***, *ARID2**, *HMBS**	G56****	M5****	↓ TIIs/T cell/B cell/myeloid dendritic cell****, Neutrophils***, Cytotoxic lymphocytes*	70% Excluded**** 68% Lymphocyte depleted***
	TP53$ (n=104)	21%	*TSC1/2**, *MET**	G3****	No.sig	No.sig	55% Intermediate* 27% IFN-γ dominant***
	AXIN1/IRF2$$ (n=55)	11%	*RPS6KA3***, *ARID1A***	G1****	M2***	No.sig	67% Intermediate**
Rare subtyps	*TP53+CTNNB1ex3* (n=30)	6%	*CCND1/FGF19**	G56****	No.sig	↓ myeloid dendritic cell**, Neutrophils*	85% Excluded****
	BAP1$$ (n=28)	5.5%	*PRKACA ampl/gain*****, *PRKAR2A loss*****	G1****	M2****	↑ T cell*, myeloid dendritic cell*, Neutrophils*	56% Inflammatory** 22% Wound healing*
	CCNA2/E1 (n=28)	5.5%	*PTEN*****, *RB1**	G2**	M7****	NK cell*	No.sig
	HNF1A$$ (n=7)	1%	No.sig	No.sig	No.sig	↓ TIIs*, T cell**, myeloid dendritic cell*	100% Lymphocyte depleted*
Unclassified	UHot (n=88)	17%	No.sig	G4****	M1***	↑ TIIs/T cell/B cell/myeloid dendritic cell/NK cell/Cytotoxic lymphocytes/Neutrophils****	25% Immune like** 21% Immune active* 65% Inflammatory****
	UInterm (n=42)	8%	No.sig	G4****	No.sig	↓ TIIs**, B cell***, T cell**, myeloid dendritic cell*	83% Intermediate****

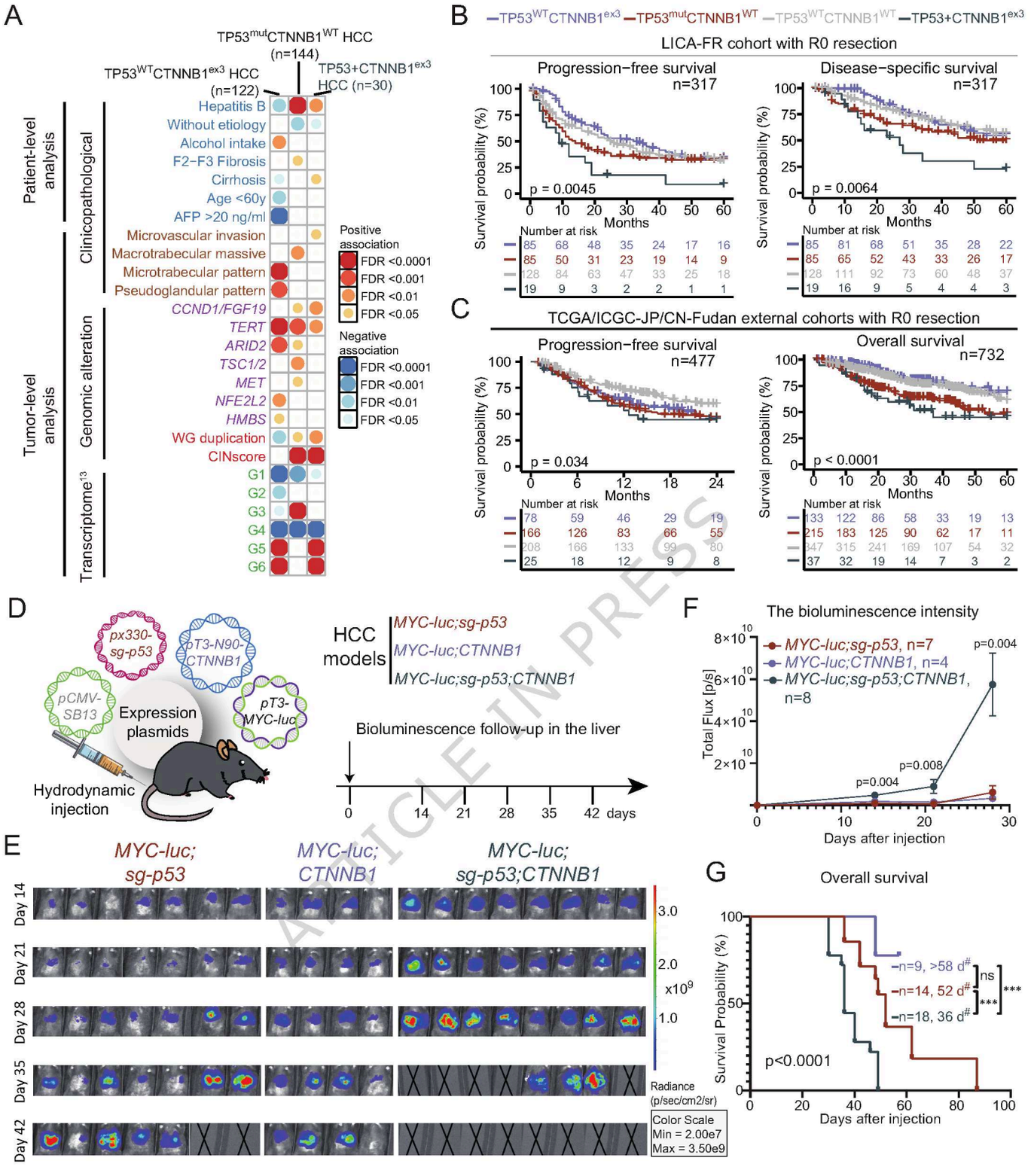

A
Wild−type p53 activity score across LICA−FR cohort (n=406)
NTL
Genome instability
CNA (arm level)
WG duplication
HCC subtypes
TP53
AXIN1/IRF2
TP53+CTNNB1ex3
CCNA2/E1
BAP1
HNF1A
CTNNB1ex3/APC
UHot
UInterm
Driver genes
TP53
TSC1/2
CCND1/FGF19
RB1
PTEN
MYC Gain/ampl
IL6ST
HMBS
Transcr. subgroups
G3
G12
G56
G4
Clinical features
Hepatitis.B
Africa/Asia
SBS24 (AFB1)
Satellite nodules
Gene Alteration
FA
HD
SV
Missense
Splice
Gain
In_Frame
Promoter
Truncating
Other features
Presence
Not available
Absence
C
Mode 1: Cooperation with TP53 mut & additive effect
AXIN1
RB1
HCC genotype
p53 activity score
Mode 2: Co-occurrence with TP53 mut & no additive effect
CCNA2/E1
PTEN
Mode 3: exclusive with TP53 mut & modulating p53 activity score independently
BAP1
IRF2
B
BCLC stage
Serum AFP (ng/ml)
Edmonson grade
Tumor size (cm)
D
Pten-/- liver tumor mouse model
0.0046



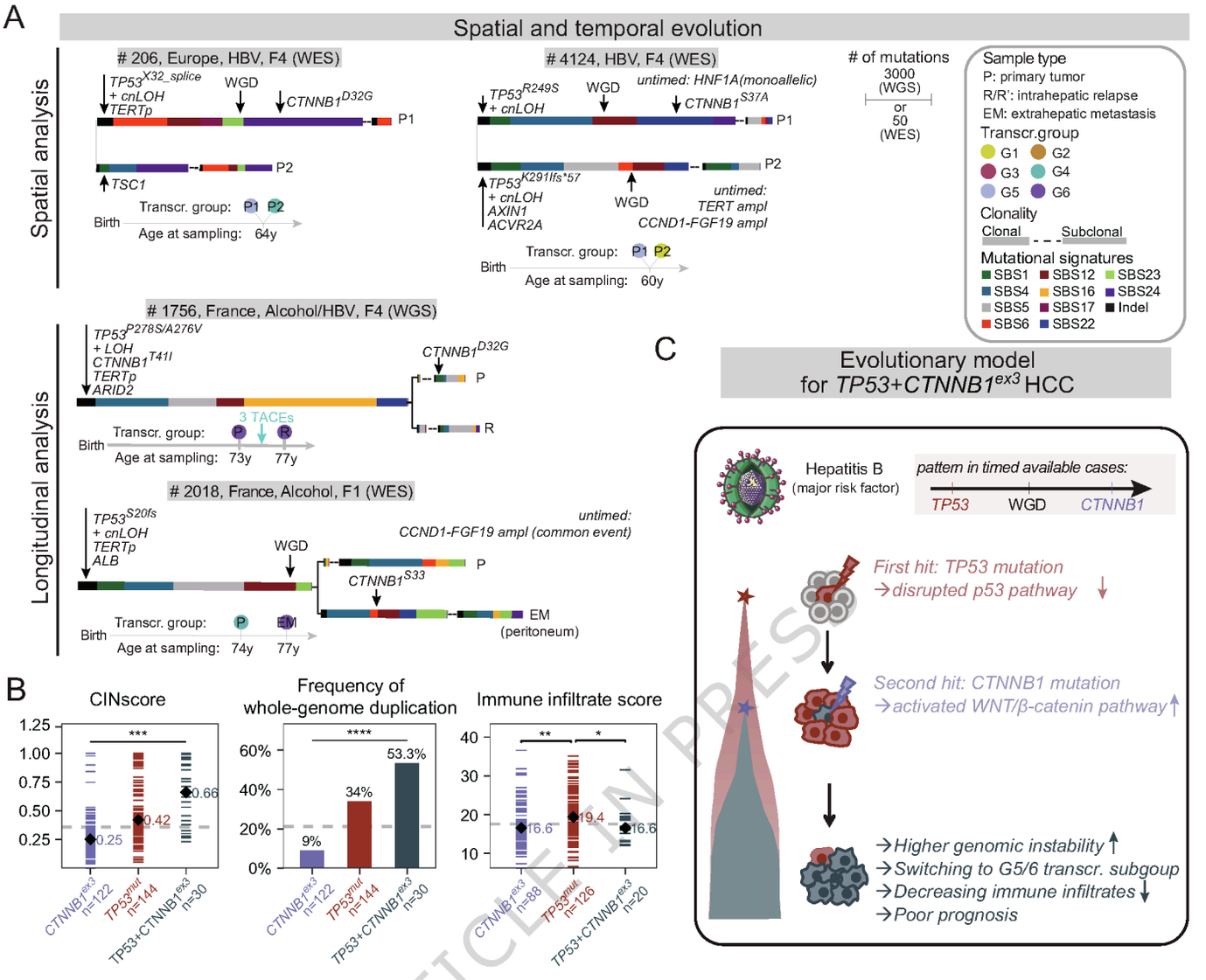




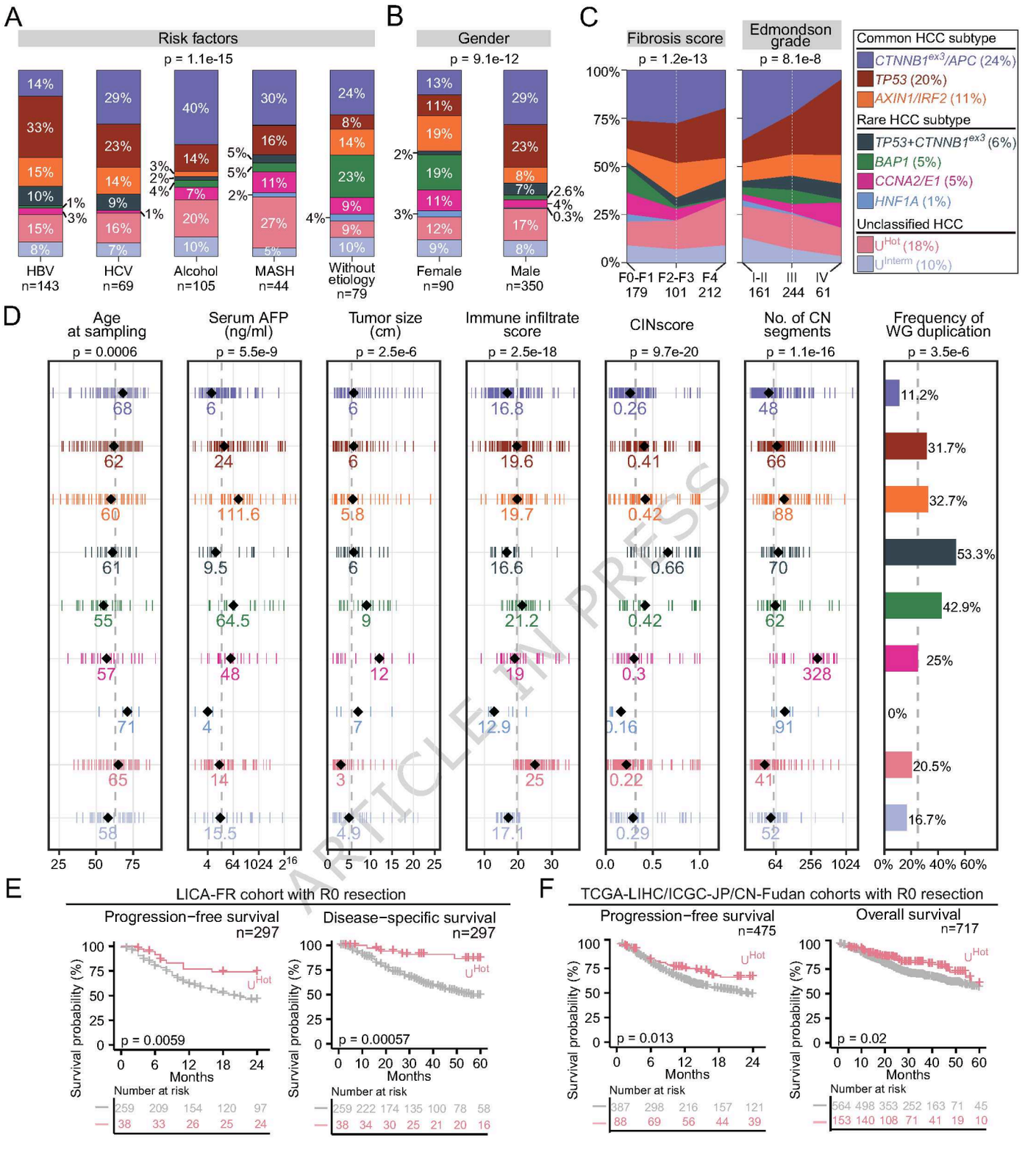
ARTICLE IN PRESS

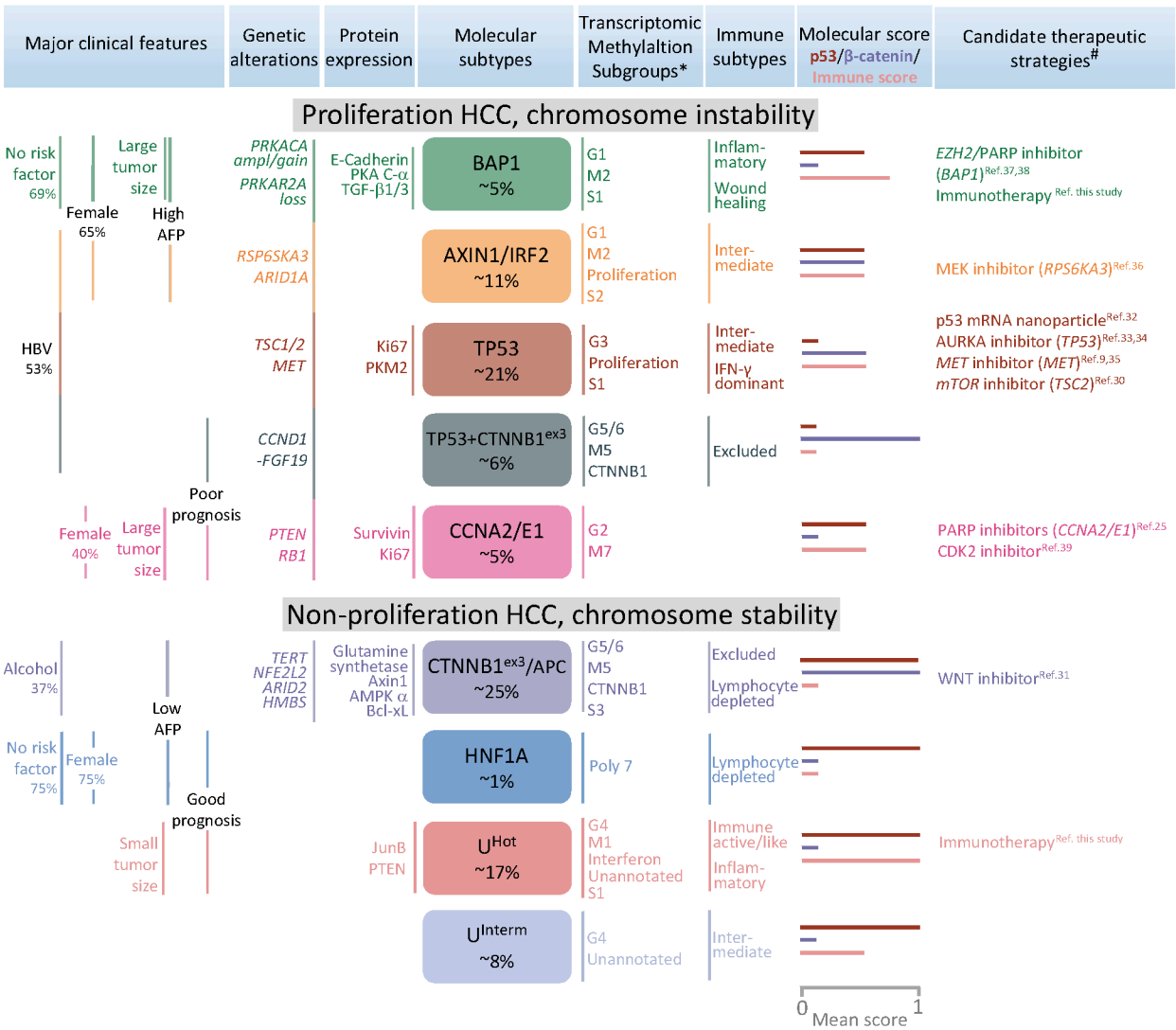


ARTICLE IN PRESS



ARTICLE IN PRESS


A
Risk factors
p = 1.1e-15
HBV n=143
HCV n=69
Alcohol n=105
MASH n=44
Without etiology n=79
B
Gender
p = 9.1e-12
Female n=90
Male n=350
C
Fibrosis score
p = 1.2e-13
F0-F1 179
F2-F3 101
F4 212
Edmondson grade
p = 8.1e-8
I-II 161
III 244
IV 61
Common HCC subtype
CTNNB1ex3/APC (24%)
TP53 (20%)
AXIN1/IRF2 (11%)
Rare HCC subtype
TP53+CTNNB1ex3 (6%)
BAP1 (5%)
CCNA2/E1 (5%)
HNF1A (1%)
Unclassified HCC
UHot (18%)
UInterm (10%)
D
Age at sampling
p = 0.0006
Serum AFP (ng/ml)
p = 5.5e-9
Tumor size (cm)
p = 2.5e-6
Immune infiltrate score
p = 2.5e-18
CINscore
p = 9.7e-20
No. of CN segments
p = 1.1e-16
Frequency of WG duplication
p = 3.5e-6
E
LICA-FR cohort with R0 resection
Progression−free survival
n=297
p = 0.0059
Disease−specific survival
n=297
p = 0.00057
Survival probability (%)
Months
Number at risk
F
TCGA-LIHC/ICGC-JP/CN-Fudan cohorts with R0 resection
Progression−free survival
n=475
p = 0.013
Overall survival
n=717
p = 0.02


Major clinical features
Genetic alterations
Protein expression
Molecular subtypes
Transcriptomic Methylaltion Subgroups*
Immune subtypes
Molecular score p53/β-catenin/ Immune score
Candidate therapeutic strategies#
Proliferation HCC, chromosome instability
No risk factor 69%
Female 65%
Large tumor size
High AFP
HBV 53%
Poor prognosis
Female 40%
Large tumor size
PRKACA ampl/gain
PRKAR2A loss
E-Cadherin
PKA C-α
TGF-β1/3
BAP1 ~5%
G1
M2
S1
Inflam-matory
Wound healing
EZH2/PARP inhibitor (BAP1)Ref.37,38
Immunotherapy Ref. this study
RSP6SKA3
ARID1A
AXIN1/IRF2 ~11%
G1
M2
Proliferation
S2
Inter-mediate
MEK inhibitor (RPS6KA3)Ref.36
TSC1/2
MET
Ki67
PKM2
TP53 ~21%
G3
Proliferation
S1
Inter-mediate
IFN-γ dominant
p53 mRNA nanoparticleRef.32
AURKA inhibitor (TP53)Ref.33,34
MET inhibitor (MET)Ref.9,35
mTOR inhibitor (TSC2)Ref.30
CCND1
-FGF19
TP53+CTNNB1ex3 ~6%
G5/6
M5
CTNNB1
Excluded
PTEN
RB1
Survivin
Ki67
CCNA2/E1 ~5%
G2
M7
PARP inhibitors (CCNA2/E1)Ref.25
CDK2 inhibitorRef.39
Non-proliferation HCC, chromosome stability
Alcohol 37%
No risk factor 75%
Female 75%
Low AFP
Good prognosis
Small tumor size
TERT
NFE2L2
ARID2
HMBS
Glutamine synthetase
Axin1
AMPK α
Bcl-xL
CTNNB1ex3/APC ~25%
G5/6
M5
CTNNB1
S3
Excluded
Lymphocyte depleted
WNT inhibitorRef.31
HNF1A ~1%
Poly 7
Lymphocyte depleted
JunB
PTEN
UHot ~17%
G4
M1
Interferon
Unannotated
S1
Immune active/like
Inflam-matory
ImmunotherapyRef. this study
UInterm ~8%
G4
Unannotated
Inter-mediate
0
1
Mean score