Supplementary Data

Title**:** Inflamed and non-inflamed classes of HCC: a revised immunogenomic classification

Authors: **Carla Montironi[1,2*], Florian Castet[1*], Philipp K. Haber[3*],** Roser Pinyol[1], Miguel Torres-Martin[1], , Laura Torrens[1], Agavni Mesropian[1], Huan Wang[4], Marc Puigvehi[3,5], Miho Maeda[3], Leow Wei-Qiang[3,6], Elizabeth Harrod[3,7,8], Patricia Taik[4], Jigjidsuren Chinburen[9], Erdenebileg Taivanbaatar[9], Enkhbold Chinbold[9], Manel Solé[1], Michael Donovan[3], Swan N. Thung[3], Jaclyn Neely[10], Vincenzo Mazzaferro[11], Jeffrey Anderson[10], Sasan Roayaie[12], Myron E Schwartz[3], Augusto Villanueva[3], Scott L. Friedman[3], Andrew Uzilov[4,13], Daniela Sia[3**], Josep M. Llovet[1,3,14**].

* These authors have contributed equally.

**These authors jointly supervised the work.

Affiliations:

[1] Translational Research in Hepatic Oncology, Liver Unit, Institut D'Investigacions Biomèdiques August Pi I Sunyer (IDIBAPS)-Hospital Clínic, Universitat de Barcelona, Catalonia, Spain.

[2] Pathology Department, Hospital Clinic de Barcelona, Barcelona, Catalonia, Spain.

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

[4] Sema4, Stamford, Connecticut, USA.

[5] Hepatology Section, Gastroenterology Department, Parc de Salut Mar, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Catalonia, Spain.

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

[7] Royal Surrey County Hospital, Guildford, UK.

[8] University of Surrey, Guildford, UK.
[9] Hepato-Pancreatico-Biliary Surgery Department, National Cancer Center, Ulaanbaatar, Mongolia.
[10] Bristol-Myers Squibb, Princeton, New Jersey, USA.
[11] Gastrointestinal Surgery and Liver Transplantation Unit, National Cancer Institute, Milan, Italy.
[12] Department of Surgery, White Plains Hospital, White Plains, New York, USA.
[13] Department of Genetics and Genomic Sciences and Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA
[14] Institució Catalana De Recerca I Estudis Avançats, Barcelona, Catalonia, Spain.

Corresponding author:
Josep M. Llovet, MD, PhD
Division of Liver Diseases, Department of Medicine, Tisch Cancer Institute
Liver Cancer Program, Icahn School of Medicine at Mount Sinai.
1425 Madison Avenue, Room 11-70A, 10029, New York, NY, USA.
E-mail: josep.llovet@mssm.edu Tel.: +1-917-488-3207

Daniela Sia, PhD
Division of Liver Diseases, Department of Medicine, Tisch Cancer Institute
Liver Cancer Program, Icahn School of Medicine at Mount Sinai.
1425 Madison Avenue, Room 11-70A, 10029, New York, NY, USA.
E-mail: daniela.sia@ mssm.edu Tel.: +1-212-659-8315; Fax: +1-212-849-2574.

Content

Supplementary Materials and Methods

Patient cohorts

The samples comprising the primary cohort used for this study were collected at several institutions belonging to the HCC Genomic Consortium and the Mongolia National Cancer Center upon approval of the Review Board at each institution.

Additional study cohorts were used to validate our findings. Level 3 preprocessed mutational data, methylation, RNA-seq and copy number data with Log R ratios (LRR) values for the TCGA-LIHC cohort were downloaded from Broad's TCGA team (http://gdac.broadinstitute.org). The Heptromic cohort included a total of 228 surgically resected fresh-frozen samples previously collected by our group (GSE63898)[1]. Microarray profiling and methylation data were available for all 228 HCC samples[2]. For CNV analysis in the Heptromic cohort, we used processed SNP array data, as previously described[3]. The samples comprising the third validation cohort where the liquid-biopsy analysis was performed comprised 71 patients with matched tumor tissue, non-tumor tissue and baseline blood samples in 68 of these patients. Finally, Nanostring data from the cohort of advanced HCC patients treated with ICIs was downloaded from the supplementary materials of the published paper[4].

Histopathological Examination

The diagnosis of HCC was initially confirmed by three independently working expert hepatopathologists blinded to the demographic and genomic findings of the patient cohort.

The histopathological examination of the samples was done by assessing the immune infiltrate, tertiary lymphoid structures (TLS) count and tumor infiltrating lymphocytes (TILs) on hematoxylin and eosin stained slides. In case of disagreement between individual evaluations, the sample was discussed by the pathologists till consensus was reached.

Hematoxylin eosin (H&E) stained sections of 216 patients were available for histological evaluation, 206 of which were assessed for the presence of immune infiltrates as previously described[1]. Briefly, the entire tumor tissue was

evaluated for presence of lymphocytic immune cells, and presence of the immune infiltrates was scored from 0 to 4 as follows: 0 (absence of immune cell infiltration), 1 (minimal infiltration), 2 (mild infiltration), 3 (moderate infiltration), and 4 (severe infiltration). Samples with scores between 0 and 1 were categorized as “low immune infiltration” and scores 2 and 4 were categorized as “high immune infiltration”.

The evaluation of TLS was performed as previously described[1]. Briefly, the entire tumor tissue on the slide was evaluated for tight aggregates of lymphocytic immune cells clearly discernible at a magnification of 100X or with a 10X objective. These aggregates were in close proximity to hepatocytic tumor cells. Aggregates within stromal spaces were not included in the count. The total number of TLS was recorded and ultimately categorized as follows: low (fewer than 5 TLS in the entire tumor tissue), or high (equal to or greater than 5 TLS in the entire tumor tissue).

TILs were evaluated according to the proposed guidelines set out by the International Immunooncology Biomarkers Working Group[5]. In accordance with the guidelines, TILs were evaluated at the center of the tumor (iTILs) and in the stroma (sTILs). The percentage of iTILs is the number of TILs over the tumor area and the percentage of sTILs is over the stomal area. Since there are no pre-defined values for what constitutes a high TILs count, we defined cut off value of more than 10% and 30% to categorize samples for iTILs and sTILs, respectively.

Immunohistochemistry

Immunohistochemical staining was carried out on 5µm-thick FFPE tissue sections after heat-induced antigen retrieval. The following targets of interest were selected: CTLA4, LAG3, TIM3 and TIGIT. Only 163 patients had available FFPE tissue for further analysis. Therefore, immunohistochemistry was performed on these 163 patients (61 HCV, 37 HBV, 65 non-infected) for LAG3 and in 162 (61 HCV, 37 HBV, 64 non-infected) for CTLA4, TIM3 and TIGIT.

The primary antibodies used were anti-CTLA4 monoclonal antibody (Origene, clone UMAB249), anti-LAG3 antibody (LS-Bio, clone 17B4), anti-TIM3 (Cell

Signaling, D5D5R) and anti-TIGIT (Dianova TG1 clone). Signal was captured using diaminobenzidine (DAB) colorimetric reaction. Immunostaining localization, distribution and intensity were assessed by 3 independent expert pathologists blinded to the clinical and genomic features of the cohort.

Scoring of LAG-3, TIM3 and TIGIT expression on intra-tumoral lymphocytes was performed by assessing the entire tumor area on the FFPE slide.

The percentage of LAG-3, TIM3 or TIGIT positive lymphocytes over intra-tumoral lymphocytes was estimated over the entire tumor area of the FFPE slide. Samples with percentages of stained lymphocytes equal to or greater than 1% were considered positive. For CTLA4, we used a previously describe quantification method[6]. In brief, intensity and percentage of stained lymphocytes was scored from 0 to 3, and a final H-score was built by multiplying these two components. The H-score of 20 was used as a threshold to define CTLA4 positive tumors.

For FAK IHC analysis, we performed immunohistochemical staining of 30 representative cases belonging to the Immune (n=10), Intermediate (n=10) and Excluded Class (n=10). Immunohistochemical observations were performed using the primary antibody FAK protein (rabbit polyclonal antibody reference #3285 Cell Signaling Technology). Three-µm-thick histological sections of 10% formalin-fixed, paraffin-embedded materials were cut, mounted on glass slides coated with 3,3-aminopropyltriethoxyslane, and air dried overnight at room temperature. The sections were removed from paraffin with xylene and rehydrated in descending dilutions of ethanol. The endogenous peroxidase activity was blocked by methanol containing 0.3% hydrogen peroxidase for 30 min. To retrieve the antigen, pretreatment with Hier antigen retrieval reagent (10X) for 20 minutes at 99°C in a pressure cooker was performed for FAK and sections were subsequently incubated overnight at 4°C with primary antibodies diluted 1:100. Next, the sections were incubated with the secondary antibody (Dako EnVision anti-rabbit secondary reference #K4003). After rinsing in PBS, the reaction products were visualized by immersing the section in diaminobenzidine tetrahydrochloride as chromogen for 5:30 minutes. No significant staining was observed in the negative controls. Finally, the sections were counterstained with hematoxylin, dehydrated, and coverslipped.

Bile duct and endothelial cells were used as the internal positive controls for FAK in the same sample. Identical reaction times allowed an accurate comparison of all samples. We classified the cases as FAK-positive tumors using two different cut-offs that reflected the percentage of positively stained carcinoma cells, i.e. ≥20% (following the definition used in a previous study)[7] and ≥30% (as it seemed to better reflect protein overexpression in our cohort.

Multiplex immunofluorescence

We performed multispectral immune fluorescence staining for CD8, PD1 and PD-L1 in a subset of the cohort (n=75). FFPE tissue sections (3 µm-thick) underwent automated staining on the Discovery Ultra platform (Roche) for CD8 (rabbit ab405) (Abcam ab4055), PD1 (rabbit clone NAT 105) (Abcam ab52587) or PDL1 (mouse Clone 28-8) (Abcam 205921). Slides were incubated with the primary antibody for 1 hour at 37°C, followed by 30 minutes incubation at 37°C with the secondary antibodies conjugated with Zenon Alexa Fluor Z488, Z594, Z555 and Z647 (Invitrogen Molecular Probe) respectively. Imaging was performed using the Imaging System Nuance 3.0.2 (Perkin Elmer). To ensure representative sampling of the entire tumor, 5 regions of interest (ROI) were obtained. If the invasive margin (IM) was present, 2 out of the 5 ROI were from that region while 3 ROI were from the central part of the tumor (IT). If there was adjacent non-tumor tissue, an additional ROI (#6) was obtained from an area far away from the tumor (NT). Each ROI was captured at a magnification of 200X and contained an average of 2000 cells. Colour intensity thresholds were applied, and artefacts and areas of auto-fluorescence were excluded.

Pixel-based percentage fluorescence intensity of CD8, PD1, and PDL1 was measured for each ROI. For subsequent analyses and correlation with clinicopathological data, we classified patients as positive if the intensity value was greater than or equal to 1% and as negative if lower than 1%.

DNA extraction, whole exome sequencing and data processing

DNA was extracted from tumor and non-tumor tissue using the QIAamp Fast DNA Tissue Kit (Qiagen, Hilden, Germany). Whole exome sequencing (WES) was performed on an Illumina HiSeq 2500 sequencer on 165 samples (64 HBV-infected, 68 HCV-infected and 33 non-infected). A median coverage of 111X was achieved per tumor sample and 65X for non-tumor specimen. Mutations were called comparing the tumor with its paired non-tumoral counterpart. Molecular variant calling was performed by using the Tigris pipeline[8]. The Tigris pipeline (v2.0.1) performs a modified GATK4 best practices (https://software.broadinstitute.org/gatk/)[9] for alignment, BQSR for base quality score recalibration, Picard MarkDuplicates for read deduplication, HaplotypeCaller for germline molecular variant (SNV and small indel) calling and Mutect2 for somatic molecular variant calling. Tigris pipeline uses GATK3 DepthOfCoverage, CallableLoci and Picard to compute depth-based and other NGS library QC. Somatic copy number variants (sCNV) are called using tumor/normal SAAS-CNV workflow that models allele balance to determine balanced versus unbalanced somatic gains and losses, as well as determine somatic copy-neutral loss of heterozygosity[10]. Data is available under accession number (will be provided when publicly available).

Tumor mutational burden (TMB), Tumor Indel burden (TIB) and Neoantigen prediction

TMB and TIB were calculated in specimens by dividing the total number of non-synonymous mutations (TMB) or indels (TIB) in every sample by the total size of the WES library (30MB).

To predict neoantigens, HLA-I typing was conducted by applying the OptiType Tool[11] to the available RNAseq data. Next, non-synonymous or indel mutations called from the WES data were screened to see whether they were transcribed into RNA by applying *isovar*[12]. Having obtained both HLA-typing and the sequence of the putative mutant protein, the NetMHC-4.0[13] computational method was applied to predict and rank the binding of a putative epitope to HLA-I. All peptides with a rank <2% were considered potentially neoantigenic.

Assuming that a high affinity is required to induce the presentation of a specific neoantigen in the HLA-I system and thus induce a T cell activation, we defined “high-affinity neoantigens” as those where the amino acid sequence of HLA-I and the putative protein were within the top 0.5% of complimentary combinations. Neoantigens were reported separately for non-synonymous and indel variants.

For the analysis of neoantigens in the TCGA-LIHC cohort, the list of neoantigens per sample provided by Thorsson et al.[14] was used.

CNV data processing and determination of the CNA level

SAAS-CNV output were further processed using CNApp for somatic CNV analysis, as previously described[3]. Briefly, we first used CNApp[15] with default parameters to refine the copy number segments by adjusting sample-specific CNA thresholds using sample purity values and applying a re-segmentation procedure. Then, we categorized the chromosomal segments as either broad or focal. Broad CNAs were defined as those segments spanning ≥50% of a chromosome arm while the rest of CNAs were considered focal events. Finally, we used CNApp to quantify the individual CNA burdens of each sample[15]. CNApp provides two scores per sample based on the number, amplitude and length of the CNAs. The broad score (BS) reflects the genomic burden of all broad CNAs and the focal score (FS) reflects all focal CNAs. In the event of broad and focal CNAs affecting the same genomic region, they are contemplated separately within BS and FS, respectively.

In order to identify candidate genes involved in determining the immune phenotypes observed, genes present in each copy number gain and loss were extracted. Significant differences were captured using a Fisher exact test. Default CNApp parameters were used to establish the threshold for amplification and depletion (log2 ratio <-0.2 for deletions (1.7 copies) and >0.2 for amplifications (2.3 copies)) unless otherwise stated[15]. To validate the functional role of deletions according to this threshold, we correlated the copy-number aberrations of candidate genes with the expression levels of the corresponding gene. Finally, to further support this analysis, we also analysed WES data by using the GISTIC 2.0 algorithm[16] and observed similar results.

TCR sequencing

In order to investigate the T cell repertoire in HCC samples at baseline, we performed T cell receptor (TCR)-sequencing in a subset of the cohort (n=40). Briefly, DNA was extracted and submitted to Adaptive Biotechnologies for TCR β-chain sequencing. Targeted amplicon libraries were prepared by multiplex PCR targeting all TCR β-chain gene segments and sequenced using the Illumina HiSeq system. Data for individual TCR sequences were obtained from Adaptive Biotechnologies and analysed using the ImmunoSEQ analyser provided by the same manufacturer.

Upon completion of TCR-sequencing, seven samples with less than 1000 productive templates were filtered out and not included in subsequent analyses according to the manufacturer's recommendation. Using the ImmunoSEQ analyzer provided by Adaptive Biotechnologies, we analyzed the following parameters across different immune classes: 1) number of productive rearrangements, which represent the count of unique rearrangements in the sample that are in frame and can produce a functional protein receptor; 2) fraction of T cells/nucleated cells; 3) productive clonality where clonality values range from 0 (polyclonal sample) to 1 (one or few predominant rearrangements); and 4) productive entropy, that was reported as Shannon's entropy of the distribution of read counts for each TCR sequenced. This is a measure of the shape of the distribution and includes information about the number of TCR sequences recovered (more sequences lead to a higher entropy) and clonality (highly clonal samples have low entropy; very diverse samples have high entropy). Samples with higher entropy have a greater diversity of rearrangements. Finally, the diversity and richness of the T cell repertoire was analyzed using several diversity metrics (i.e. Daley-Smith, Efron-Thisted, iChao1).

RNA Extraction, Sequencing and Analysis

RNA was extracted using the RNeasy Plus Mini Kit (Qiagen), following the manufacturer´s protocol. The RNA data was processed by the RAPiD pipeline at the Mount Sinai Genomic Core Facility. Briefly, fastq files were aligned using *STAR*[17] (version 2.5.1b) to hg19 with gencode annotation v19 and were quantified with *featureCounts*[18] (v1.5.2). Normalization for subsequent analysis

was performed using the *edgeR*[19] package with the trimmed mean of M-values (TMM) used for computing the scale factors[20]. The resulting effective library size was used in all downstream analysis unless otherwise indicated. Data is available under accession number (will be provided when publicly available).

In the analysis of the transcriptomic data, positivity for previously reported gene signatures was evaluated using the Nearest Template Prediction[21] module from GenePattern[22]. Significant prediction was defined using an adjusted FDR q-value < 0.05. Differential gene expression analysis was performed by using the *DESeq2* package[23] with default parameters. For this purpose, the original counts for each gene (organized as a matrix of integer values) was used as input. Differentially expressed genes were selected based on an adjusted FDR q-value (Benjamini-Hochberg method) of <0.05 and shrunken log2 fold change of >1.5 (according to adaptive shrinkage estimator from the ashr package[24]). Gene set enrichment analysis of the resulting genes was conducted using the Enrichr web tool[25]. The Gene set enrichment analysis (GSEA) GenePattern module was used to assess enrichment of activated pathways/signatures in each class[26] and the single-sample gene set enrichment analysis (ssGSEA) GenePattern module was used to assess enrichment of activated pathways/signatures in each sample[27]. A list of the gene signatures used for this study can be found in **Supplementary Table 19.** Cytolytic activity was calculated as the geometric mean of the genes granzyme A (GZMA) and perforin-1 (PRF1), as previously described[28]. Richness of the immune and stroma infiltrate in tumor tissue was inferred from expression data through the Immune Score obtained from the ESTIMATE software and by applying the 141 gene signature using the ssGSEA methodology[29].

The CIBERSORTx deconvolution tool was used to impute gene expression profiles and provide an estimation of the relative and absolute abundances of immune cell types in the tumor samples[30]. The xCell method was used as a validation[31].

In the liquid-biopsy cohort comprising 71 patients, RNA extraction was performed as previously reported[2], and RNA profiling was conducted using the Human Genome U219 Array Plate (Affymetrix). The processing of transcriptome data (normalization, background correction and filtering) was carried out as previously

described[2]. Blood samples were centrifugated and the plasma retrieved and submitted for the Immuno-oncology and Inflammation Olink Assay panel (Olink, Sweden), which are a high-throughput, multiplex immunoassay enabling analysis of 92 protein biomarkers.

Generation of the inflamed signature and liquid biopsy-based signature

To identify the immune-like class, we used the Expanded immune gene signature described by Ayers *et al*[32] and applied it using ssGSEA. Subsequently, the immune-like class clustered all those cases that had a score higher than one standard deviation above the mean of the non-immune samples (**Supplementary Figure 2A**). To generate the inflamed signature, we used our previously published immune factor, which was obtained by performing a virtual microdissection through non-negative matrix factorization (NMF) of the expression data[1]. Briefly, an immune-related expression pattern was identified by integrating NMF-identified factors with the immune enrichment score[29]. The top-ranked genes were listed based on their loadings and unsupervised clustering performed to identify the Immune class. Subsequently, differential gene expression analysis between the Immune and non-Immune cases revealed 389 genes with an FDR adjusted q-value <0.05. To generate the inflamed signature, the 17 top-ranked genes based on the value of the t-test statistic that were overexpressed in the Immune class compared to the non-Immune cases were selected. Three genes were additionally incorporated into the signature based on their significant role in cancer-associated inflammation, *CXCL9, GZMB* and *CXCR4*[33–35].

This 20-gene signature was validated in two additional cohorts by applying ssGSEA. The upper tertile was used as a cut-off to differentiate inflamed from non-inflamed tumors. The sensitivity, specificity, positive predictive value, negative predictive value and accuracy were calculated by comparing the inflamed cases as originally defined with the defined cases by the Inflamed signature.

For the analysis of liquid biopsy-based biomarkers, we first performed a univariate binomial logistic regression and all proteins with a p<0.05 were entered

into a multivariate binomial logistic regression model. This allowed us to devise a 13-protein signature with a good performance in capturing the Inflamed class (AUC 0.91, accuracy 85%, sensitivity 92%). Most of these proteins were overexpressed in peripheral blood of Inflamed tumors, and included proteins involved in immune activation (IL18, PDL1), cytolysis (GZMA, GZMH), TNF superfamily (TNFRSF21, TWEAK) or angiogenesis (TIE2, PGF, VEGFR2). This data serves as a proof-of-concept that the Inflamed class can be identified by using blood-based biomarkers, although further validation in additional prospective cohorts will be required in the future.

Statistical analysis

All analyses were conducted using R (version 4.0.2) and SPSS software version 23 (IBM Corporation, Chicago, IL). Intergroup comparisons were performed using Pearson χ^2 test or Fisher's exact test depending on the number of categories for qualitative data. Kruskal-Wallis test with post-hoc Dunn's test or Wilcoxon's rank-sum test was used for intergroup comparisons of continuous data with a non-parametric distribution depending on the number of group categories. ANOVA or T-Student's test was used for intergroup comparisons of variables with a parametric distribution. Correlations between two continuous variables were performed using Spearman correlation *r*. Kaplan-Meier estimates and log-rank test were used to analyze the association of molecular and clinical variables with overall survival and tumor recurrence. Median follow-up was estimated by applying the reverse Kaplan-Meier method. For the multivariate analysis of the impact on survival of the immune classes, all clinical variables were first evaluated in a univariate Cox regression analysis, and those with a p-values of <0.1 were entered in a stepwise backward multivariate Cox regression analysis. A two-sided $p < 0.05$ was considered statistically significant for all analysis.

Supplementary Tables

Supplementary Table 1: Clinicopathologic characteristics of the cohort stratified according to aetiology

	All Patients (n=240)*	HBV (n=75)	HCV (n=90)	Non-Infected (n=75)	p**
Age, years; median (range)	65 (29-91)	61 (29-87)	67 (43-83)	67 (47-91)	**<0.01**
Age ≥ 65 years, n (%)	121 (54)	20 (29.4)	54 (63.5)	47 (66.2)	**<0.01**
Gender, male	172 (77)	53 (78)	59 (69)	60 (85)	0.08
Bilirubin, mg/dL; median (range)	0.8 (0.3-3.8)	0.7 (0.4-3.2)	0.8 (0.4-2.7)	1 (0.3-3.8)	**<0.01**
Albumin, g/dL; median (range)	4.1 (2.2-5.4)	4.1 (2.5-4.9)	4 (2.5-5.2)	4.1 (2.2-5.4)	0.65
Platelets, 109/L; median (range)	160 (27-574)	170 (83-574)	154 (42-379)	147 (27-460)	0.08
AFP, ng/mL; median (range)	13 (0-311190)	15 (0-26628)	29 (1-31692)	8 (1-311190)	0.07
AFP >400ng/mL, n (%)	34 (16.5)	9 (14.5)	16 (20)	9 (14.1)	0.56
AFP >200ng/mL, n (%)	45 (21.7)	14 (22.6)	21 (25.9)	10 (15.6)	0.32
BCLC 0/A, n (%)	179 (83.6)	59 (86.8)	69 (82.1)	51 (82.2)	0.70
Vascular invasion; n (%)	101 (45)	38 (56)	30 (35)	33 (46)	**0.04**
Macrovascular invasion; n (%)	26	9 (13)	11 (13)	6 (9)	0.60
Poor Differentiation, n (%)	47 (24)	18 (27)	17 (25)	12 (20)	0.61
Size, cm; median (range)	4.5 (1-21)	6 (1.5-21)	4 (1-19)	4.4 (1-19)	**0.01**
Satellites; n (%)	50 (31)	11 (30)	22 (34)	17 (27)	0.73
Multiple nodules, n (%)	51 (24)	14 (21)	21 (25)	16 (26)	0.78
Recurrence, n (%)	96 (60)	20 (54)	47 (75)	29 (48)	
Time to recurrence, months; median (range)	27 (21-33)	25 (0-69)	26 (17-35)	27 (2-53)	
Death, n (%)	93 (46)	22 (36)	45 (58)	26 (40)	
OS, months; median (95% CI)	70 (59-80)	73 (67-79)	50 (33-67)	95 (86-104)	

Abbreviations: Alpha-feto protein (AFP), Barcelona Clinic Liver Cancer (BCLC), Overall survival (OS). *Missing data: 7% Age, 7% Gender, 11% BCLC, 19% differentiation, 11% size, 7% vascular invasion, 32% satellites, 11% num. nodules, 9% bilirubin, 9% albumin, 8% platelets, 15% AFP, 33% Recurrence, and 15% Survival. ** Pearson χ^2 for categorical variables, Kruskal Wallis for quantitative variables.

Supplementary Table 2: Univariate and Multivariate Cox regression analysis of Overall survival and Recurrence-free survival.

	Overall survival		Recurrence-free survival	
	Univariate [HR (95% CI), p-value]	Multivariate [HR (95% CI), p-value]	Univariate [HR (95% CI), p-value]	Multivariate [HR (95% CI), p-value]
Age	1.00 (0.98-1.02), p=0.90		**0.97 (0.95-0.99), p=0.009**	0.97 (0.95-1.00), p=0.062
Gender				
Male	Ref.		Ref.	
Female	1.26 (0.80-1.99), p=0.33		1.14 (0.67-1.95), p=0.631	
Etiology				
VHC	Ref.	Ref.	Ref.	
VHB	0.71 (0.42-1.19), p=0.19	0.61 (0.28-1.31), p=0.204	0.67 (0.39-1.14), p=0.138	
Non-viral	**0.51 (0.31-0.83), p=0.007**	0.56 (2.89-1.09), p=0.088	0.7 (0.41-1.19), p=0.188	
Bilirubin	0.89 (0.63-1.26), p=0.50		0.94 (0.58-1.50), p=0.78	
Albumin	0.83 (0.57-1.22), p=0.34		1.06 (0.69-1.64), p=0.79	
Platelets	1.00 (0.998-1.004), p=0.43		1.00 (0.998-1.004), p=0.53	
AFP				
<400ng/mL	Ref.	Ref.	Ref.	
≥400ng/mL	**1.67 (0.97-2.89), p=0.064**	**2.44 (1.16-5.13), p=0.019**	1.51 (0.82-2.80), p=0.19	
BCLC staging				
BCLC 0/A	Ref.		Ref.	Ref.
BCLC B/C	1.46 (0.85-2.51), p=0.17		**2.36 (1.21-4.62), p=0.012**	**3.16 (1.07-9.28), p=0.037)**
Vascular invasion	**1.48 (0.97-2.25), p=0.066**	1.44 (0.82-2.54), p=0.21	1.37 (0.89-2.12), p=0.157	
Size				
<5cm	Ref.	Ref.	Ref.	
≥5cm	**1.62 (1.07-2.47), p=0.023**	0.92 (0.46-1.85), p=0.81	1.25 (0.81-1.94), p=0.315	
Satellitosis	**1.77 (1.10-2.83), p=0.019**	1.16 (0.62-2.15), p=0.65	**1.52 (0.95-2.42), p=0.08**	1.27 (0.70-2.29), p=0.43
Multiple nodules	**1.52 (0.96-2.41), p=0.071**	**3.25 (1.66-6.36), p=0.001**	**2.47 (1.53-3.99), p<0.001**	**2.69 (1.45-4.98), p=0.002**
Degree of differentiation				
Good/moderate	Ref.		Ref.	
Poor	1.02 (0.60-1.73), p=0.95		0.71 (0.40-1.26), p=0.24	
Immune Active	**0.34 (0.13-0.84), p=0.019**	**0.33 (0.13-0.84), p=0.02**	**0.49 (0.24-1.04), p=0.062**	0.72 (0.33-1.56), p=0.401
Immune Exhausted	1.72 (0.78-3.81), p=0.178		**2.8 (1.26-6.22), p=0.011**	**2.84 (1.22-6.64), p=0.016**
Immune-like	1.21 (0.67-2.19), p=0.533		**0.91 (0.47-1.75), p=0.766**	
Intermediate	1.11 (0.68-1.83), p=0.676		**1.07 (0.65-1.77), p=0.784**	
Excluded	1.17 (0.64-2.15), p=0.614		**1.22 (0.66-2.25), p=0.521**	

Supplementary Table 3: Gene list comprising the Wnt-βcatenin activation signature

Provided as an Excel file.

Supplementary Table 4: Clinicopathologic characteristics of the cohort stratified according to the Immune classification

	Immune Active (n=26)	Immune Exhausted (n=12)	Immune-like (n=26)	Non-inflamed (n=107)	p*
Age, years; median (range)	67 (53-82)	57 (41-79)	64 (40-79)	63 (29-82)	0.06
Age ≥ 65 years, n (%)	14 (60.9)	3 (25)	12 (46.2)	42 (44.7)	0.24
Gender, male	19 (82.6)	9 (75)	19 (73.1)	72 (76.6)	0.88
Etiology, n (%) HBV HCV Non-infected	 7 (26.9) 9 (34.6) 10 (38.5)	 6 (50) 4 (33.3) 2 (16.6)	 9 (34.6) 13 (50) 4 (15.4)	 43 (40.2) 45 (42.1) 19 (17.8)	0.29
Bilirubin, mg/dL; median (range)	0.8 (0.4-2.7)	0.8 (0.4-3.2)	0.8 (0.4-2.2)	0.8 (0.4-2.1)	0.96
Albumin, g/dL; median (range)	4 (3.2-4.7)	4 (3.1-4.7)	4.1 (3.2-4.8)	4 (2.5-5.1)	0.74
Platelets, 109/L; median (range)	139 (61-388)	177 (99-574)	164 (82-379)	170 (42-388)	0.48
AFP, ng/mL; median (range)	47 (3-26947)	65 (2-71270)	13 (1-7055)	12 (0-71770)	0.12
AFP >400ng/mL, n (%)	5 (21.7)	4 (33.3)	5 (20)	10 (11.9)	0.22
AFP >200ng/mL, n (%)	5 (21.7)	4 (33.3)	6 (24)	16 (18.8)	0.69
BCLC 0/A, n (%)	22 (95.7)	9 (75)	22 (88)	85 (90.4)	0.29
Vascular invasion; n (%)	7 (30.4)	6 (50)	7 (26.9)	48 (51.1)	0.08
Macrovascular invasion; n (%)	3 (13)	2 (16.7)	3 (11.5)	10 (10.6)	0.93
Poor Differentiation, n (%)	8 (36.4)	2 (22.2)	9 (39.1)	14 (17.7)	0.10
Size, cm; median (range)	4.2 (1.5-18)	5.5 (2-21)	4.8 (1-18)	4.5 (2-20)	0.81
Satellites; n (%)	3 (16.7)	4 (57.1)	4 (23.5)	18 (29)	0.23
Multiple nodules, n (%)	3 (13)	1 (8.3)	6 (23.1)	19 (20.4)	0.61

Abbreviations: Alpha-feto protein (AFP), Barcelona Clinic Liver Cancer (BCLC). Missing data: 9% Age, 9% Gender, 10% BCLC, 22% differentiation, 12% size, 9% vascular invasion, 39% satellites, 10% num. nodules, 11% bilirubin, 10% albumin, 9% platelets, 16% AFP. * Pearson χ^2 for categorical variables, Kruskal Wallis for quantitative variables.

Supplementary Table 5: Integration of Hoshida molecular classes of HCC[36] with the Immune classification

		Immune (n=38)	Immune-like (n=26)	Non-inflamed (n=107)	P-value		
					Immune vs Immune-like	Immune-like vs Non-Inflamed	Immune vs Non-Inflamed
Hoshida Molecular Subclasses	**S1**	26	2	6	$1.6x10^{-6}$	0.65	$3.66x10^{-14}$
	S2	1	5	27	0.04	0.62	$1.5x10^{-3}$
	S3	8	12	50	0.05	1	$6.7x10^{-3}$

Values in blue indicate a two-tailed p-value <0.05

Supplementary Table 6: Integration of Chiang molecular classes of HCC[37] with the Immune classification

		Immune (n=38)	Immune-like (n=26)	Non-inflamed (n=107)	P-value		
					Immune vs Immune-like	Immune-like vs Non-Inflamed	Immune vs Non-Inflamed
Chiang Molecular Subclasses	**CTNNB1**	1	12	28	$2.87x10^{-5}$	0.06	$8.64x10^{-4}$
	Proliferation	10	7	22	1	0.6	0.5
	IFN-related	12	0	9	$9.17x10^{-4}$	0.21	$1.12x10^{-3}$
	Polisomy of Chr 7	3	3	27	0.68	0.29	0.06
	Unnanotated	7	3	10	0.51	0.72	0.15

Values in blue indicate a two-tailed p-value <0.05

Supplementary Table 7: Histological assessment of immune infiltration in the Inflamed class

	Immune (n=29)	Immune-like (n=22)	Non-inflamed (n=95)	P-value		
				Inflamed vs Non-Inflamed	Immune-like vs Non-Inflamed	Immune vs Non-Inflamed
Low Immune Infiltrate (score 0-1)	14	10	65	0.01	0.05	0.08
High Immune Infiltrate (score 2-4)	15	12	30			

Values in blue indicate a two-tailed p-value <0.05

Supplementary Table 8: Histological assessment of TILs and TLS in the Inflamed class

	Immune (n=19)	Immune-like (n=14)	Non-inflamed (n=68)	P-value		
				Inflamed vs Non-Inflamed	Immune-like vs Non-Inflamed	Immune vs Non-Inflamed
High iTILs (≥10%)	5	6	11	0.07	0.06	0.33
High sTILs (≥30%)	7	7	12	0.01	0.02	0.11
High TLS (≥5)	5	4	8	0.09	0.21	0.15

Values in blue indicate a two-tailed p-value <0.05

Supplementary Table 9: Genes comprising the *Inflamed signature*

Gene	Description
CCL5	C-C Motif Chemokine Ligand 5
CD2	CD2 Molecule
CD3D	CD3d Molecule
CD48	CD48 Molecule
CD52	CD52 Molecule
CD53	CD53 Molecule
CXCL9	C-X-C Motif Chemokine Ligand 9
CXCR4	C-X-C Motif Chemokine Receptor 4
FYB	FYN Binding Protein 1
GZMA	Granzyme A
GZMB	Granzyme B
GZMK	Granzyme K
IGHG1	Immunoglobulin Heavy Constant Gamma 1
IGHG3	Immunoglobulin Heavy Constant Gamma 3
LAPTM5	Lysosomal Protein Transmembrane 5
LCP2	Lymphocyte Cytosolic Protein 2
PTPRC	Protein Tyrosine Phosphatase Receptor Type C
SLA	Src Like Adaptor
TRAC	T Cell Receptor Alpha Constant
TRBC2	T Cell Receptor Beta Constant 2

Supplementary Table 10: Performance of the *Inflamed signature* across different datasets

	Discovery Cohort	Heptromic Cohort	TCGA-LIHC Cohort
Accuracy	89%	94%	89%
Sensitivity	80%	86%	80%
Specificity	95%	99%	95%
PPV	91%	97%	91%
NPV	89%	92%	89%

Supplementary Table 11: Distribution of copy number deletions in relevant immune-related genes in the Discovery cohort

	Immune (n=30)	Immune-like (n=22)	Intermediate (n=63)	Excluded (n=28)	P-value		
					Intermediate vs Inflamed	Intermediate vs Excluded	Intermediate vs Rest
16p13.13 (*CIITA*)	5	3	27	5	$2.04x10^{-3}$	0.03	$6.45x10^{-4}$
4q21.1 (*CXCL9, CXCL10, CXCL11*)	6	3	34	6	$4.99x10^{-5}$	0.01	$1.53x10^{-5}$
4q35.1 (*IRF2*)	6	3	35	5	$3.97x10^{-5}$	$1.15x10^{-3}$	$2.74x10^{-6}$

Values in blue indicate a two-tailed p-value <0.05

Supplementary Table 12: Distribution of copy number deletions in relevant immune-related genes in Heptromic cohort

	Immune (n=23)	Immune-like (n=9)	Intermediate (n=49)	Excluded (n=21)	P-value		
					Intermediate vs Inflamed	Intermediate vs Excluded	Intermediate vs Rest
16p13.13 (*CIITA*)	3	1	16	1	0.06	0.01	$6.16x10^{-3}$
4q21.1 (*CXCL9, CXCL10, CXCL11*)	6	2	24	5	0.04	0.07	0.01
4q35.1 (*IRF2*)	4	2	24	4	$9.09x10^{-3}$	0.03	$1.61x10^{-3}$

Values in blue indicate a two-tailed p-value <0.05

Supplementary Table 13: Distribution of copy number deletions in relevant immune-related genes in TCGA-LIHC cohort

	Immune (n=74)	Immune-like (n=37)	Intermediate (n=150)	Excluded (n=81)	P-value		
					Intermediate vs Inflamed	Intermediate vs Excluded	Intermediate vs Rest
16p13.13 (*CIITA*)	14	7	40	9	0.18	6.61×10^{-3}	0.02
4q21.1 (*CXCL9, CXCL10, CXCL11*)	24	12	62	20	0.16	0.01	0.02
4q35.1 (*IRF2*)	28	11	68	26	0.10	0.07	0.03

Values in blue indicate a two-tailed p-value <0.05

Supplementary Table 14: Differential gene expression analysis between Wnt-βcatenin Inflamed and non-inflamed profiles

Provided as an Excel file

Supplementary Table 15: GO biological process enrichment analysis of differentially expressed genes between Wnt-βcatenin Inflamed and non-inflamed profiles

Provided as an Excel file

Supplementary Table 16: Gene set enrichment analysis results of Hallmark gene sets in the Wnt-βcatenin Inflamed and non-Inflamed profiles

Provided as an Excel file

Supplementary Table 17: Gene set enrichment analysis results of WNT-related gene sets in the Wnt-βcatenin Inflamed and non-Inflamed profiles

Provided as an Excel file

Supplementary Table 18: Differentially expressed Wnt-related genes between Inflamed and non-inflamed Wnt-βcatenin activated tumors

Provided as an Excel file

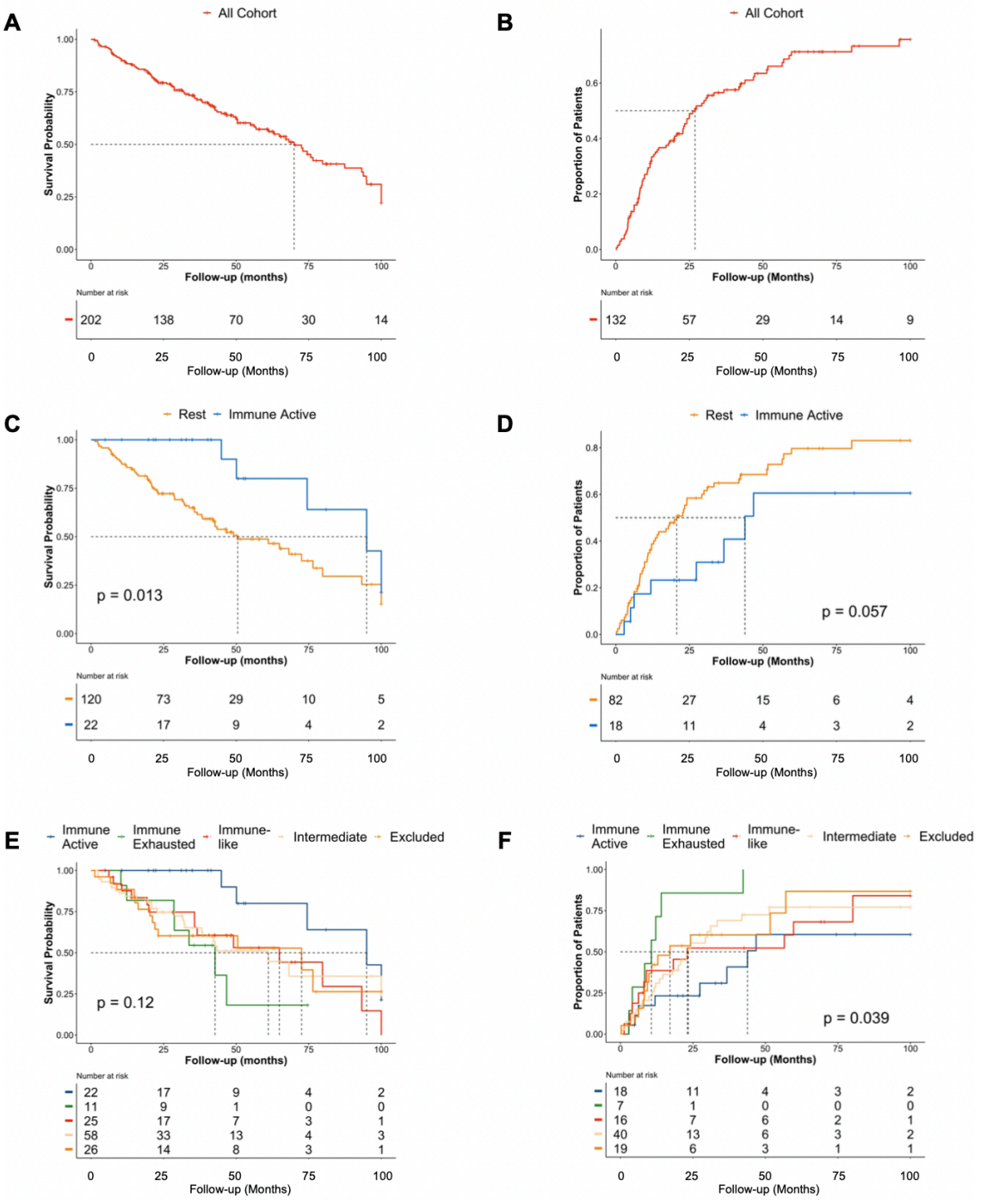
Supplementary Table 19: Publicly available gene signatures and gene sets used in the study

Name		Study	Reference
Immune enrichment score		Yoshihara K, et al. Nat Commun 2013;4:2612	[29]
Stroma enrichment score			
HCC molecular classes (Hoshida)	S1	Hoshida Y, et al. Cancer Res 2009;69:7385-92	[36]
	S2		
	S3		
HCC molecular classes (Chiang)	CTNNB1	Chiang DY, et al. Cancer Res 2008;68:6779-88	[37]
	Interferon		
	Proliferation		
	Poly7		
	Unannotated		
WNT/TGFβ signature		Lachenmayer A, et al. Clin Cancer Res 2012;18:4997-5007	[38]
Wnt-βcatenin activation signature		Lachenmayer A, et al. Clin Cancer Res 2012;18:4997-5007	[38]
Late TGFβ signature		Coulouarn C, et al. Hepatology. 2008 Jun;47(6):2059-67	[39]
Activated stroma		Moffitt RA, et al. Nat Genet 2015;47:1168-78	[40]
KEGG Wnt signalling pathway		MSigDB - GSEA	[41]
Biocarta Wnt pathway			
Boyault liver cancer subclass G6_UP			
Boyault liver cancer subclass G5-6_UP			
GO Canonical Wnt signalling pathway			

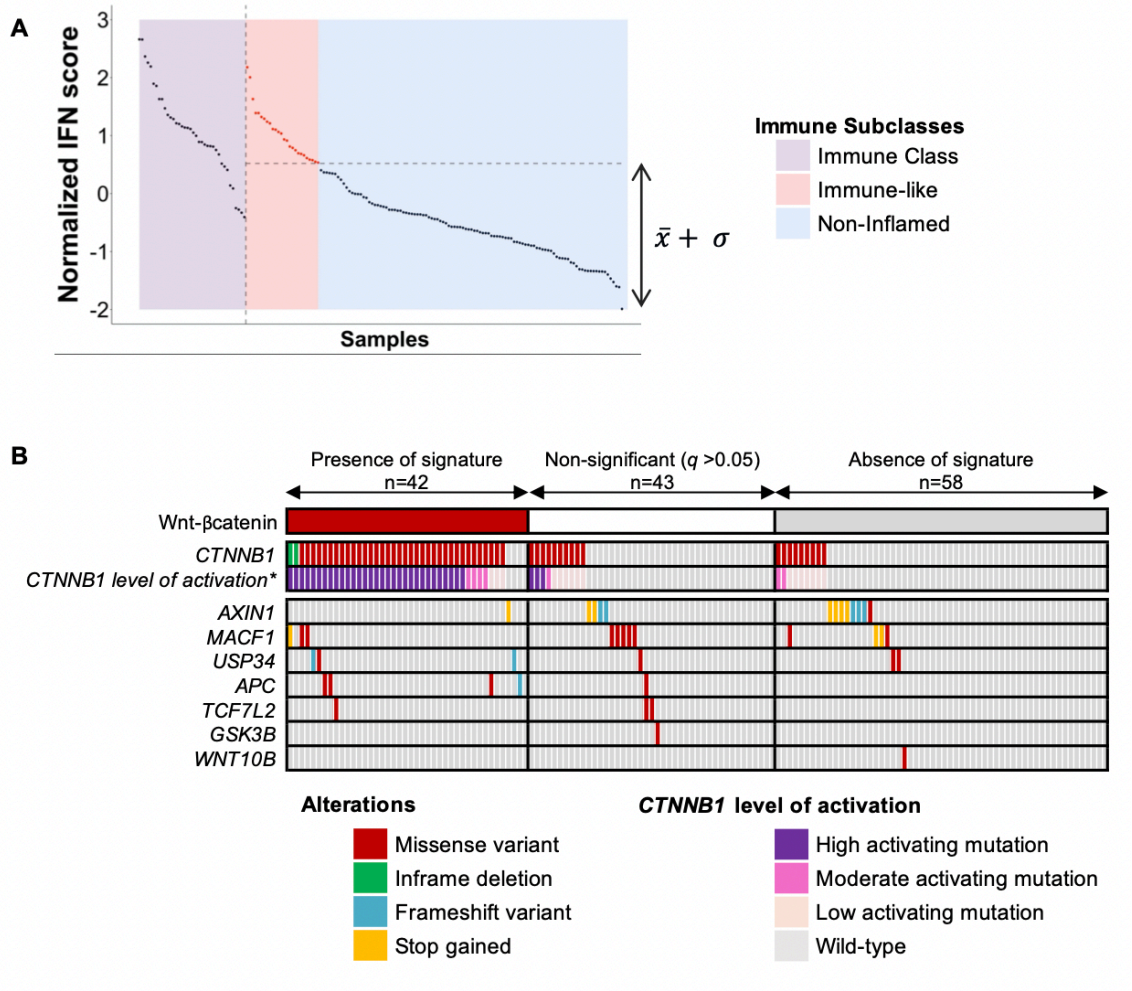
GO Regulation of Wnt signalling pathway		
GO Wnt activated receptor activity		
Hallmark Wnt beta catenin signalling		
PID Wnt canonical pathway		
PID Wnt signalling pathway		
Reactome signalling by Wnt		
Reactome signalling by Wnt in cancer		
WP Wnt signalling		
GO Regulation of non-canonical Wnt pathway		
PID Wnt non-canonical pathway		
GO non-canonical Wnt signalling pathway		
HCC Immune class	Sia D, et al. Gastroenterology. 2017;153:812–826	[1]
IFN signature (18-gene)	Ayers M, et al. J Clin Invest 2017;127:2930–2940	[32]
Inflammatory signature	Sangro B, et al. J Hepatol 2020;73:1460–1469	[42]
Cytolitic activity	Rooney MS, et al. Cell 2015;160:48-61	[28]
INFAP signature	Haber PK, et al. EASL 2021	[43]
Tertiary lymphoid structure (TLS)	Cabrita R, et al. Nature 2020; 577:561–565	[44]
Activated CD8 T cells	Charoentong P, et al. Cell Rep 2017;18:248–262.	[45]
Activated CD4 T cells		
Activated dendritic cells		
Activated B cells		
M1 macrophages	Davoli T, et al. Science 2017;355:eaaf8399.	[46]
M2 macrophages		

Supplementary Figures

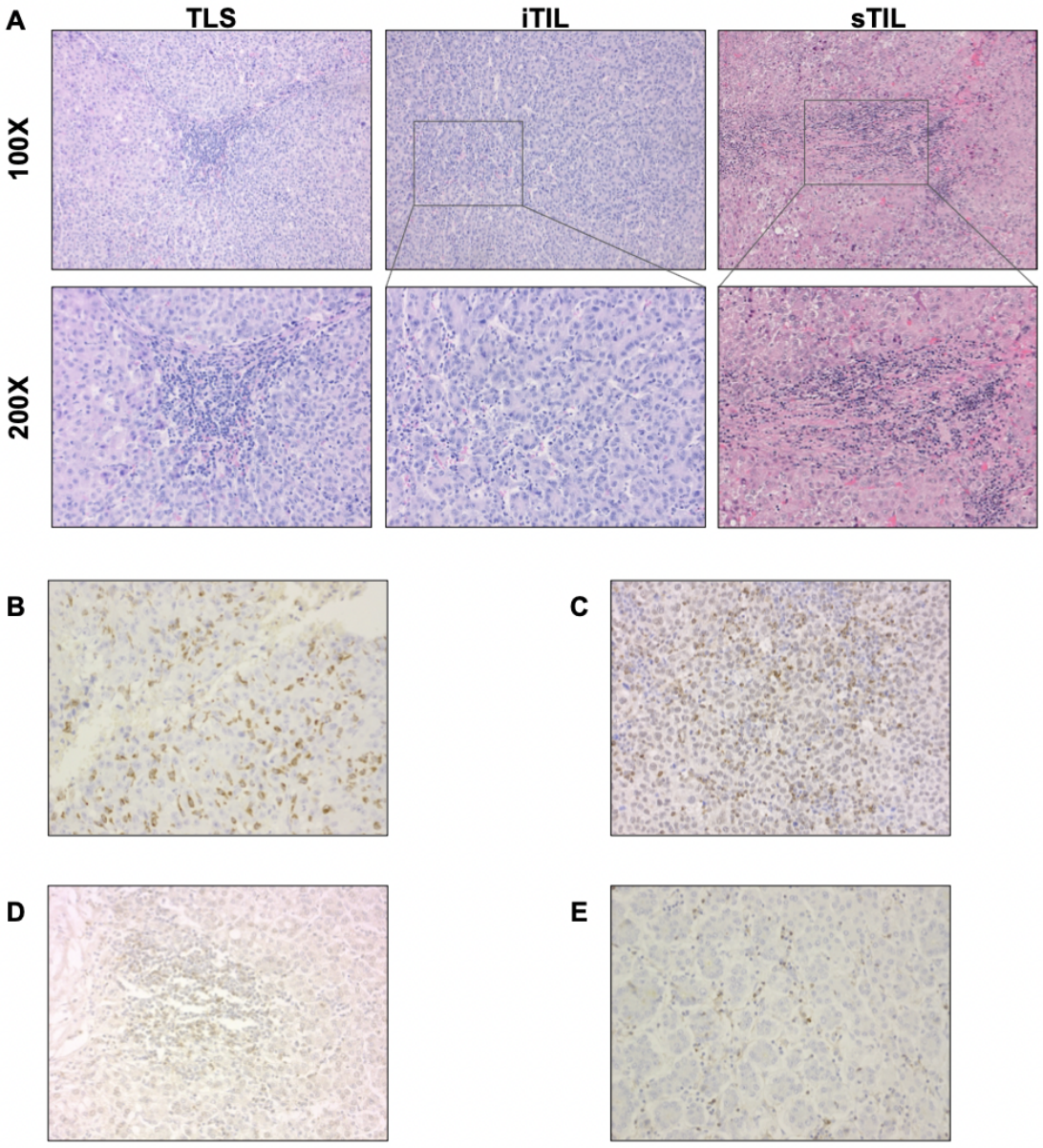
Supplementary Figure 1: Kaplan-Meyer estimates of overall survival and recurrence-free survival for the whole study cohort. Kaplan Meier survival curves for (**A**) Overall survival and (**B**) recurrence-free survival plots of the whole cohort. (**C**) Overall survival and (**D**) recurrence-free survival of the immune Active class compared with the rest of the cohort. (**E**) Overall survival and (**F**) recurrence-free survival of the distinct immune classes. P-value is calculated by Log rank (Mantel-Cox) test.



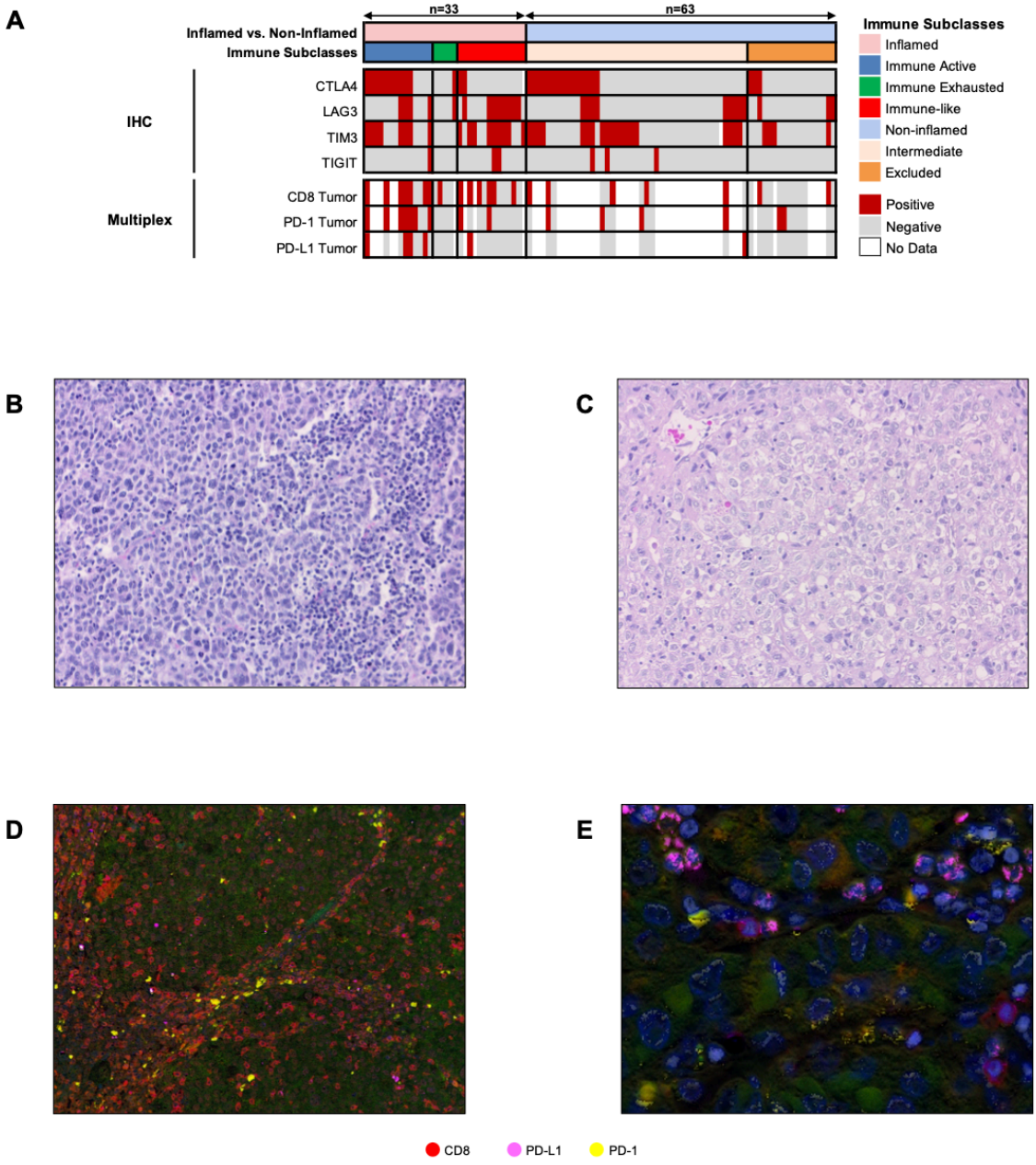
Supplementary Figure 2: Identification of the Immune classes of HCC. (**A**) Scatter plot showing the normalized levels of IFN signaling[32] (y-axis) across all samples included in the cohort (x-axis). Colored rectangles delineate the Immune (purple) and Immune-like (red) classes. (**B**) Heatmap representation of the genomic events captured by the Wnt-βcatenin activation signature.



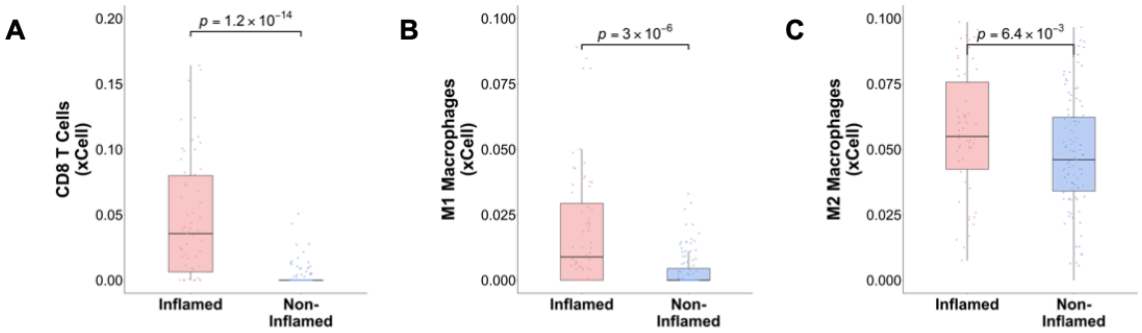
Supplementary Figure 3: Analysis of the immune infiltration in the whole cohort. *Related to Figure 3.* Representative images of (**A**) tertiary lymphoid structures (TLS), intratumoral tumor infiltrating lymphocytes (iTILs) and stromal tumor infiltrating lymphocytes (sTILs). The images were taken at 100 and 200X magnification. Representative images of HCC cases positive for (**B**) CTLA4, (**C**) LAG3, (**D**) TIGIT, (**E**) TIM-3, respectively. All images were taken with a magnification of 200X.



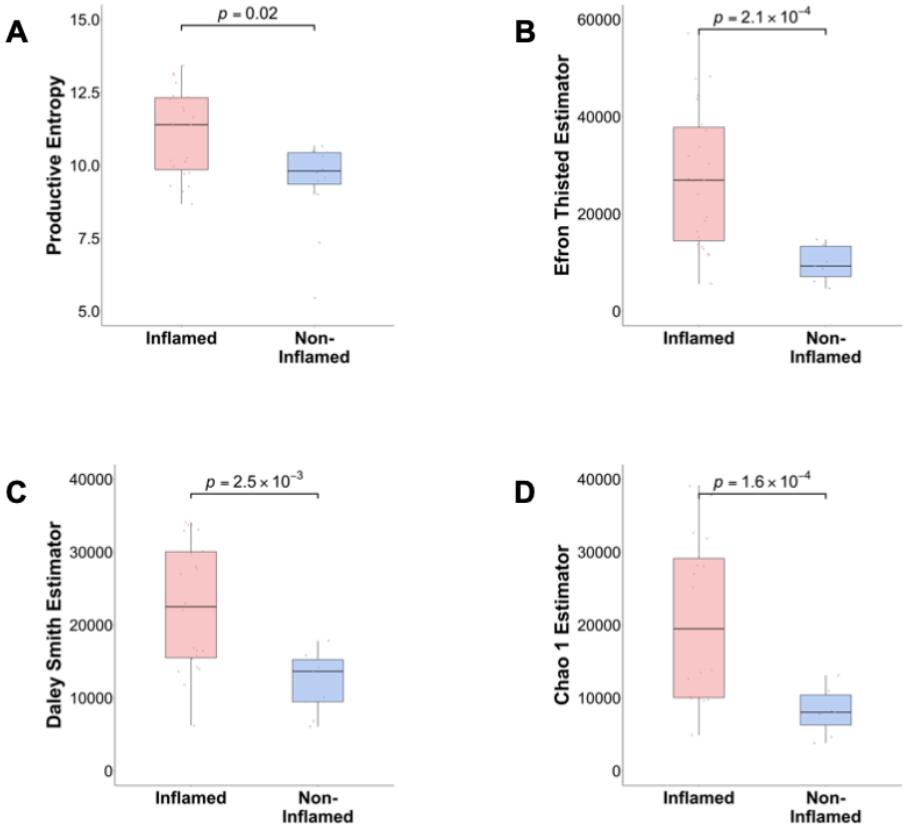
Supplementary Figure 4: Representative images of immune infiltration and multiplex immunofluorescence immunostaining. *Related to Figure 3.* (**A**) Heatmap representation of global IHC and multiplex immunofluorescence findings. (**B**) H&E image of an HCC case with high immune infiltration compared to an HCC case with (**C**) low immune infiltration. Images were taken at 200X magnification. (**D**) Representative image of multiplex immunofluorescence immunostaining. The image is focused on an area with high CD8 immune infiltrate (200X magnification). (**E**) The image represents a zoomed-in tumoral area (1600X magnification). CD8 is seen in red, PD-1 in yellow and PD-L1 is seen in magenta. DAPI counterstains the nuclei of all cells.



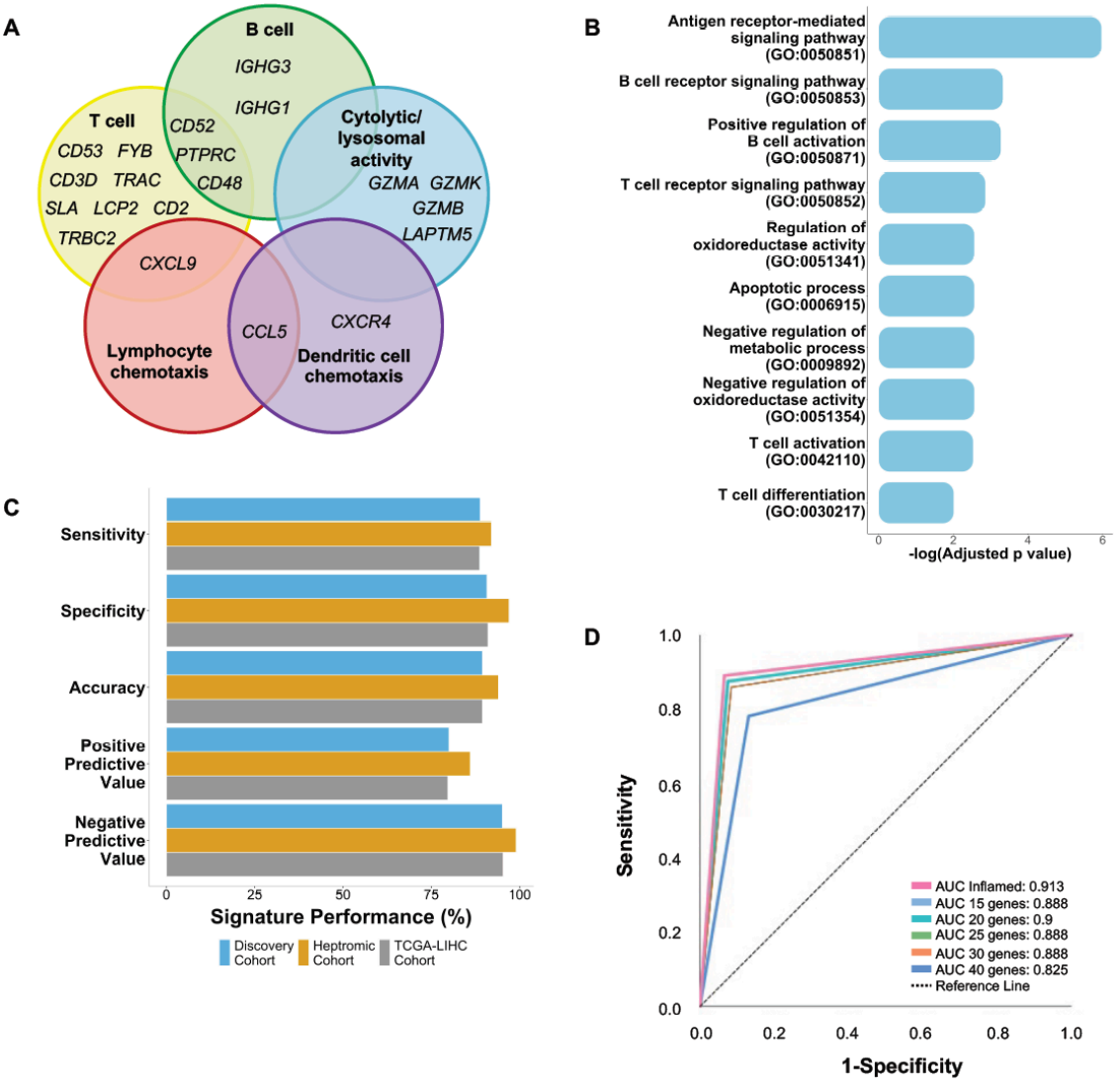
Supplementary Figure 5: Validation of the composition of the immune infiltration by xCell. *Related to Figure 3.* (**A-C**) xCell deconvolution method analyzing (**A**) CD8+ cells, (**B**) M1 macrophages and (**C**) M2 macrophages. P-value is calculated by Wilcoxon rank-sum's test.



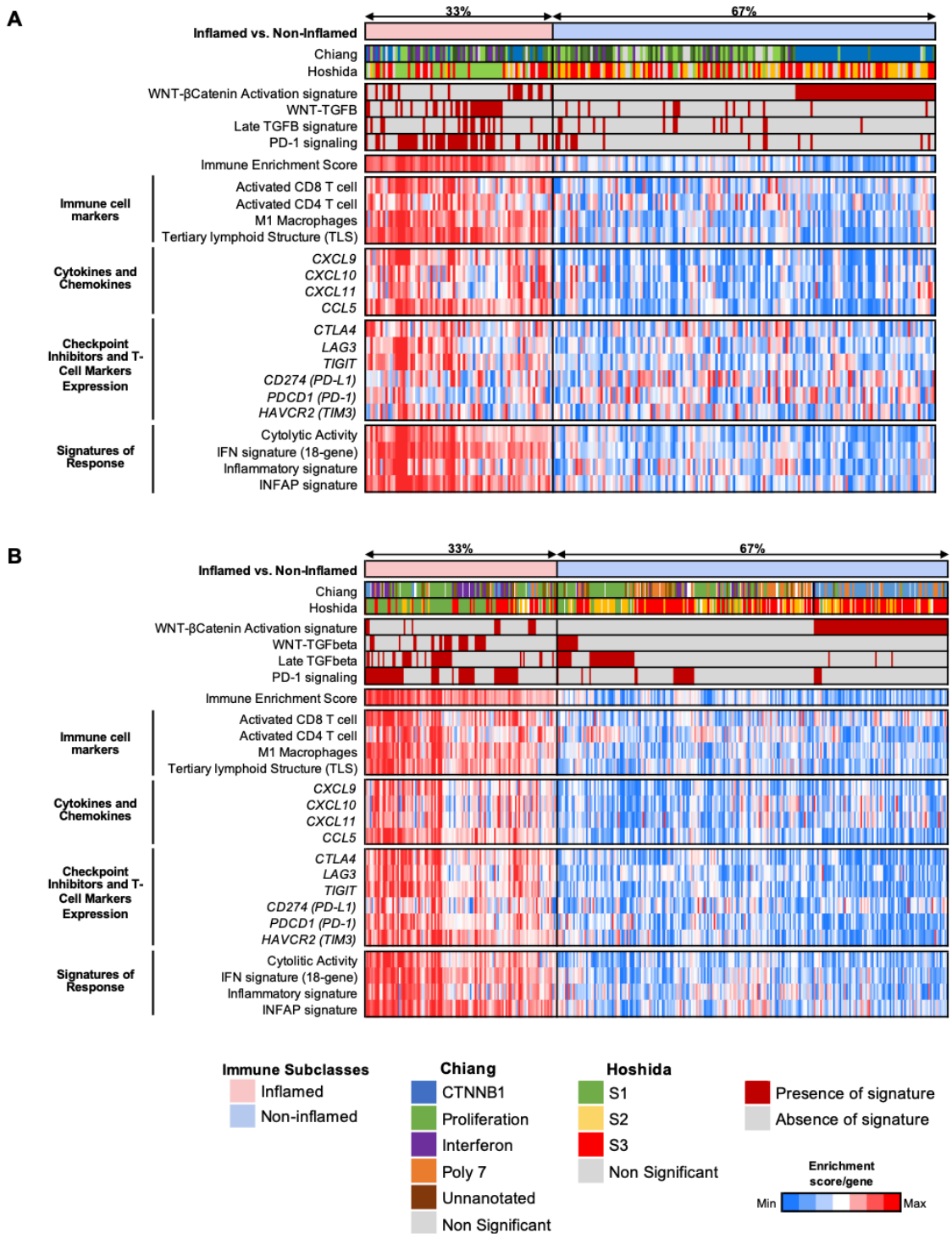
Supplementary Figure 6: TCR-seq analysis confirms a higher immune infiltration in the Immune and Inflamed class and suggests the presence of a more diverse T cell repertoire. *Related to Figure 3.* Boxplots represent the (**A**) productive entropy, (**B**) Efron-Thisted Estimator, (**C**) Daley Smith Estimator and (**D**) Chao1 Estimator. P-value is calculated by Wilcoxon rank-sum's test.



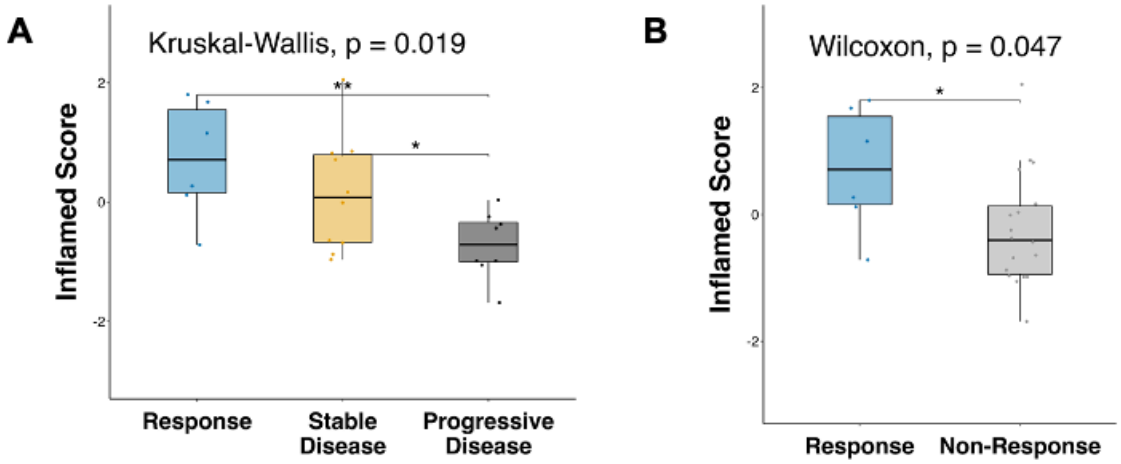
Supplementary Figure 7: Generation and validation of the Inflamed signature across additional datasets. *Related to Figure 3.* (**A**) Circle plot depicting the 20 genes included in the Inflamed signature and their immune-related function. (**B**) Gene Ontology Biological Processes enrichment analysis of the 20 genes. (**C**) Performance of the Inflamed signature in capturing the Inflamed class in all three cohorts. (**D**) Receiver operating characteristic (ROC) curves showing the ability to capture the inflamed class using different gene thresholds.



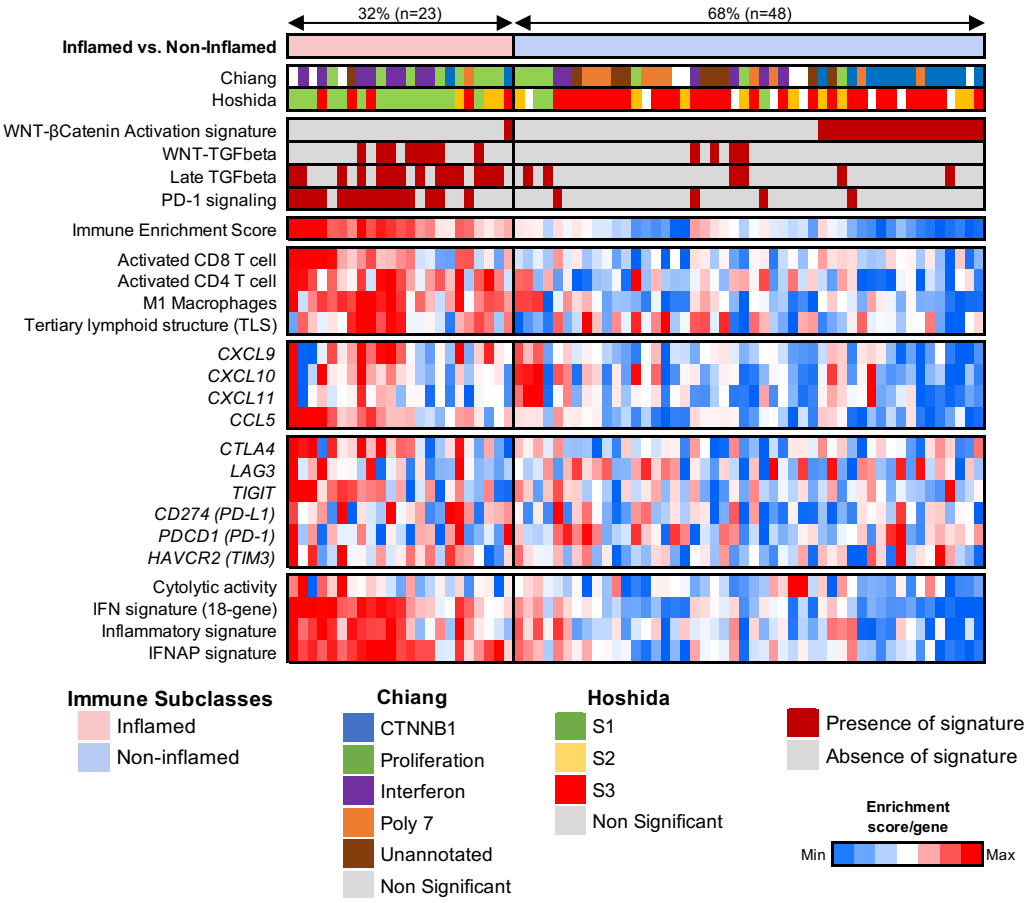
Supplementary Figure 8: Heatmap representation of the main molecular and immune features of the distinct immune-related profiles. *Related to Figure 3.* Heatmap corresponding to (**A**) Heptromic cohort and (**B**) TCGA-LIHC cohort.



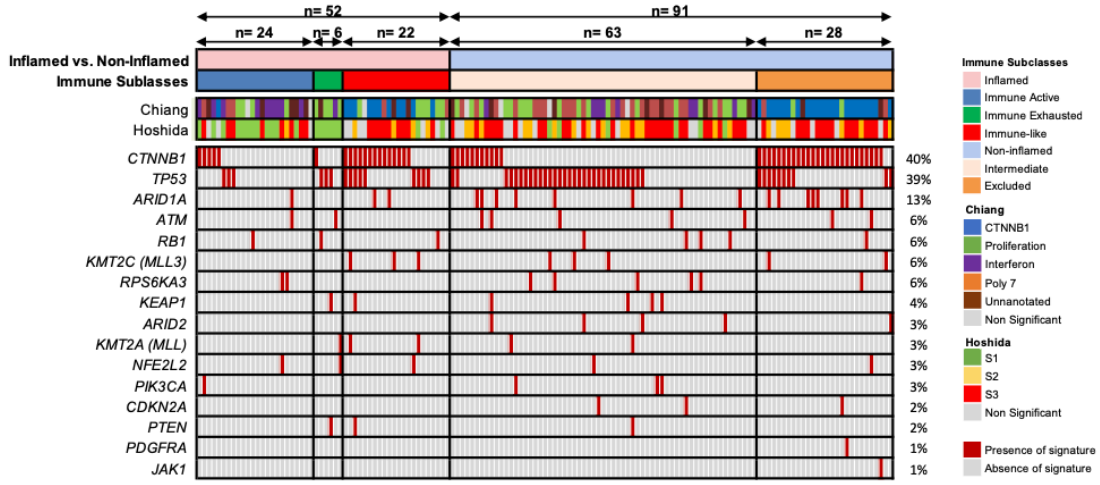
Supplementary Figure 9: Predictive potential of the Inflamed signature in Hsu et al. cohort. Boxplot depicting the inflamed score according to response status (A) Response was categorized in partial, stable disease or progressive disease according to RECIST 1.1 (B) Stable and progressive disease were considered Non-Response. P-value calculated by (A) Kruskal-Wallis test with post-hoc Dunn's test and (B) Wilcoxon rank-sum test. *, p<0.05; **p<0.01.



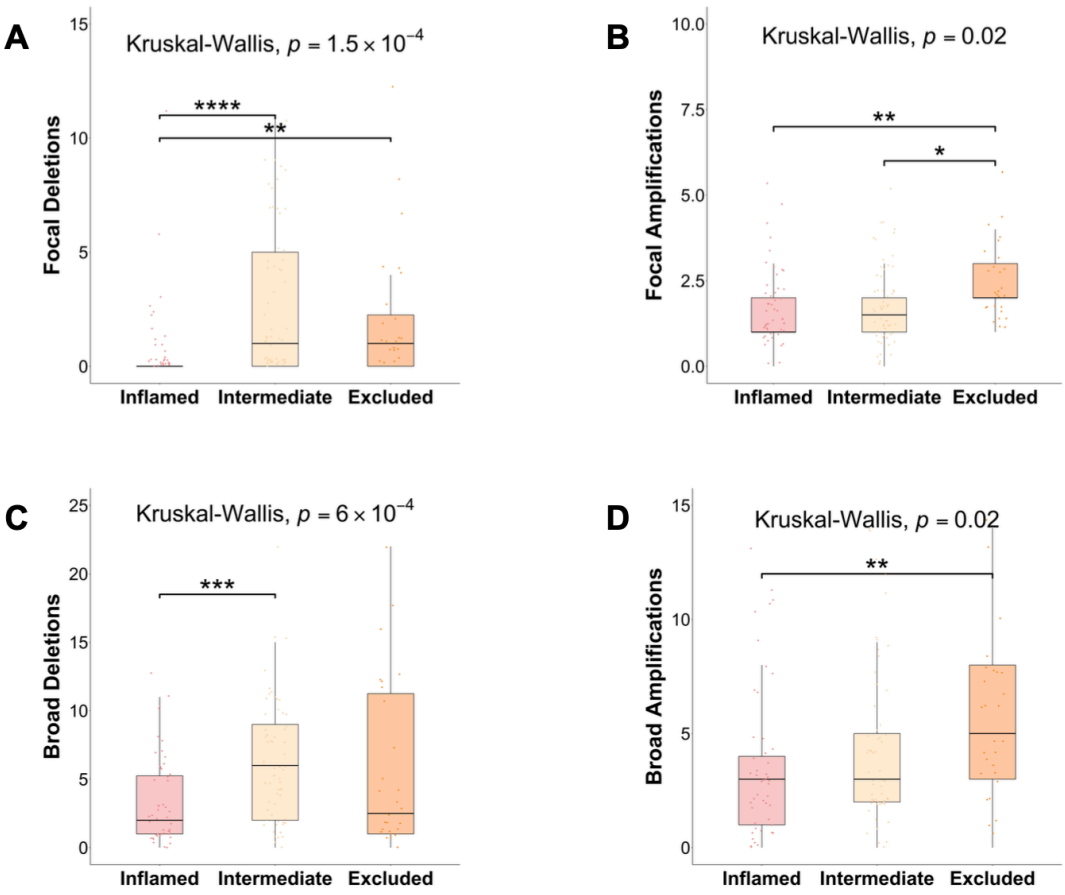
Supplementary Figure 10: Identification of the Inflamed Class by using liquid biopsy-based biomarkers. Heatmap representation of the main molecular and immune features of the distinct immune-related profiles. *p-values shown are calculated by Student's T test for continuous variables or Fisher's exact test for categorical variables and represent differences between the inflamed and non-inflamed classes.



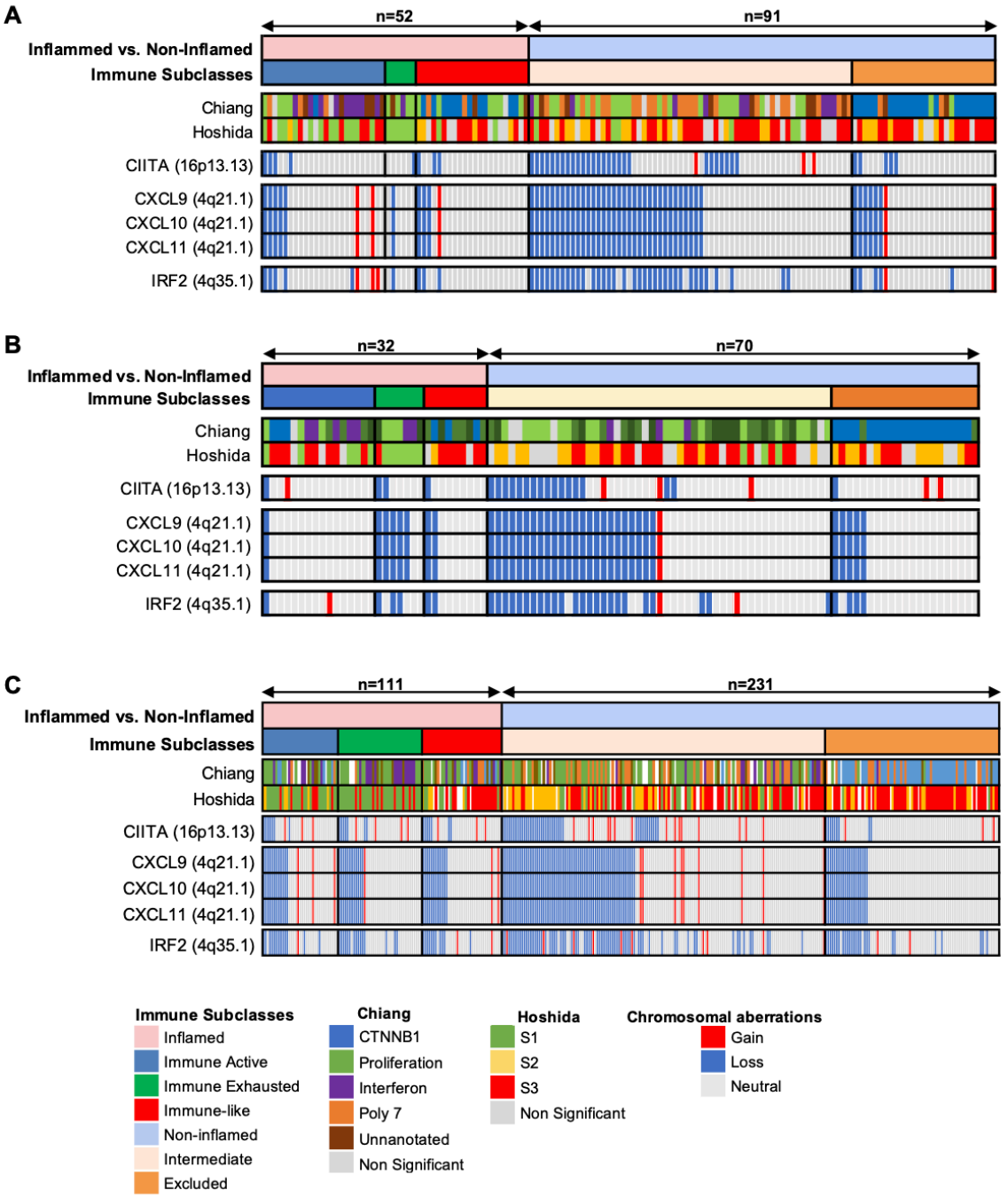
Supplementary Figure 11: Mutational landscape of the distinct immune classes of HCC. *Related to Figure 4.* Heatmap representation of the distribution of mutations in known driver genes of HCC.



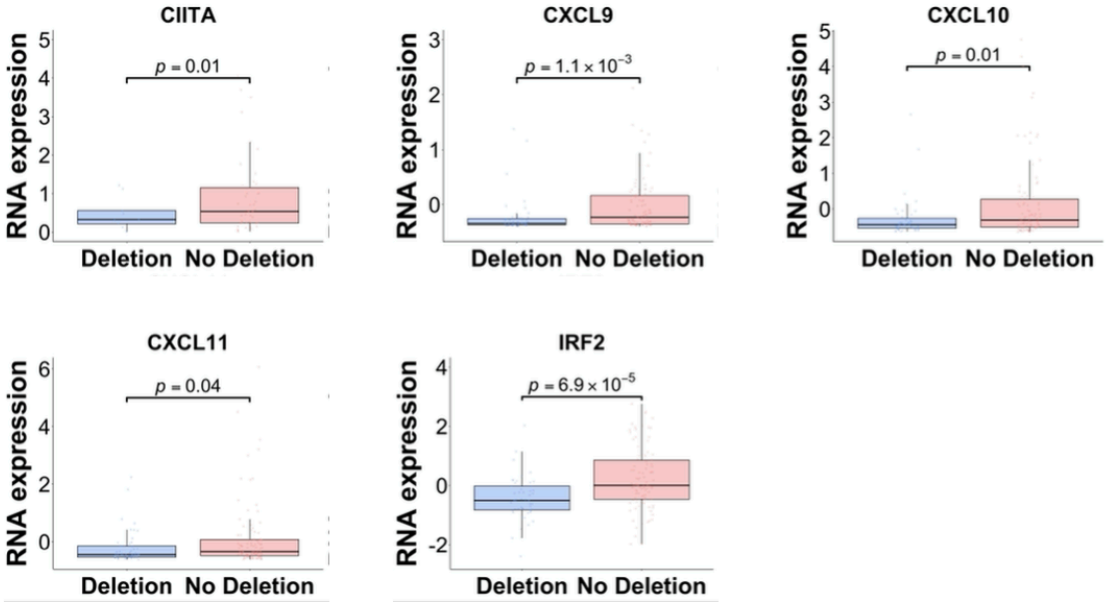
Supplementary Figure 12: Copy number aberrations burden in the whole cohort as determined by the GISTIC2.0 algorithm. *Related to Figure 4.* Box plot representation of the distribution of the chromosomal aberrations: (**A**) focal deletions, (**B**) focal amplifications, (**C**) broad deletions and (**D**) broad amplifications among the distinct Immune classes. P-value is calculated by Kruskal-Wallis test with post-hoc Dunn's test. *, P < 0.05; **, P < 0.01; ***, P < 0.001; ****, P < 0.0001.



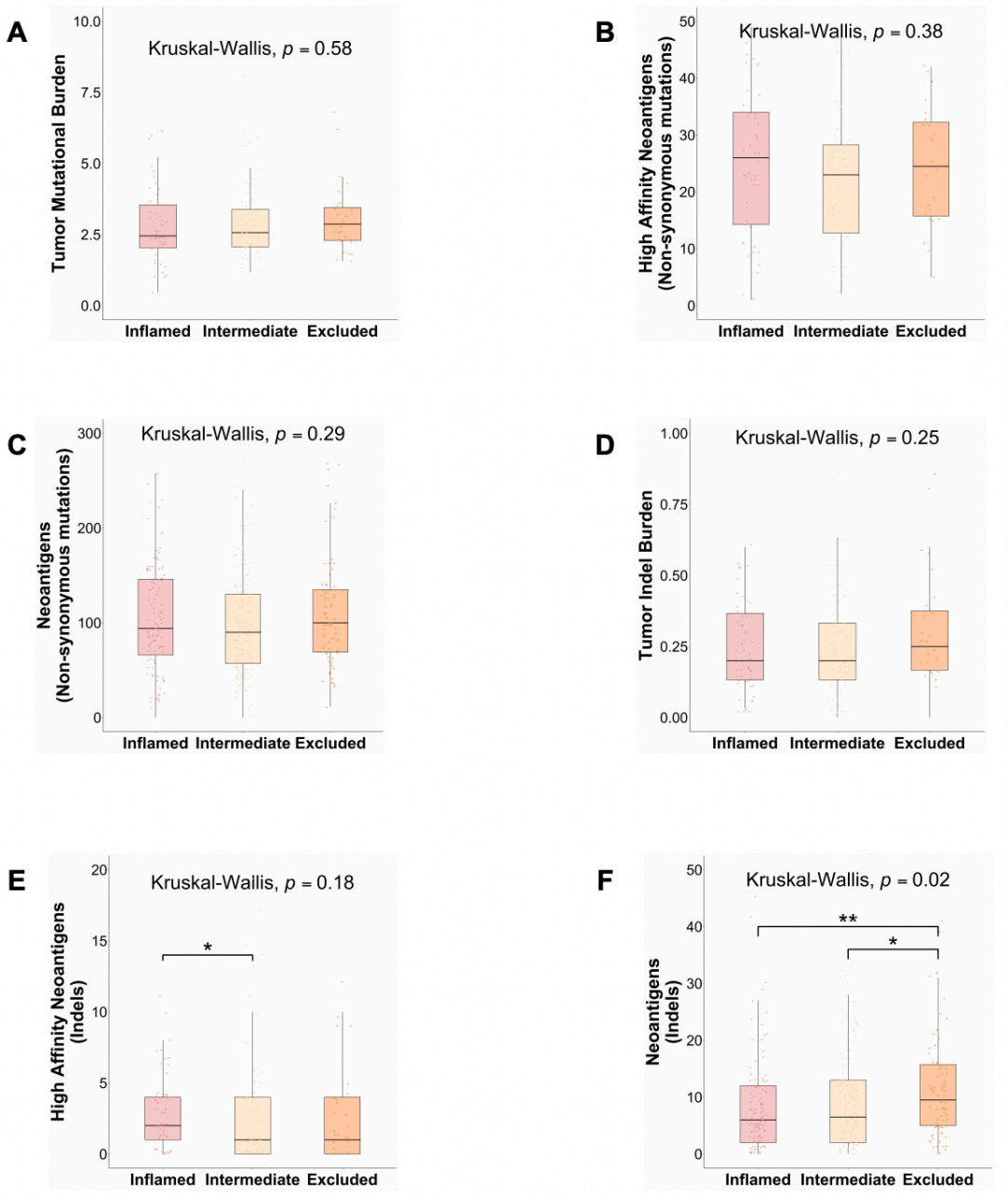
Supplementary Figure 13: Copy number deletions in specific subcytobands harbouring genes related to antigen presentation and interferon signalling. *Related to Figure 4.* Heatmap representation of deletions (blue) and amplifications (red) of subcytobands harbouring specific genes in the (**A**) discovery cohort, (**B**) Heptromic cohort and (**C**) TCGA-LIHC cohort. Threshold was established according to CNApp default parameters.



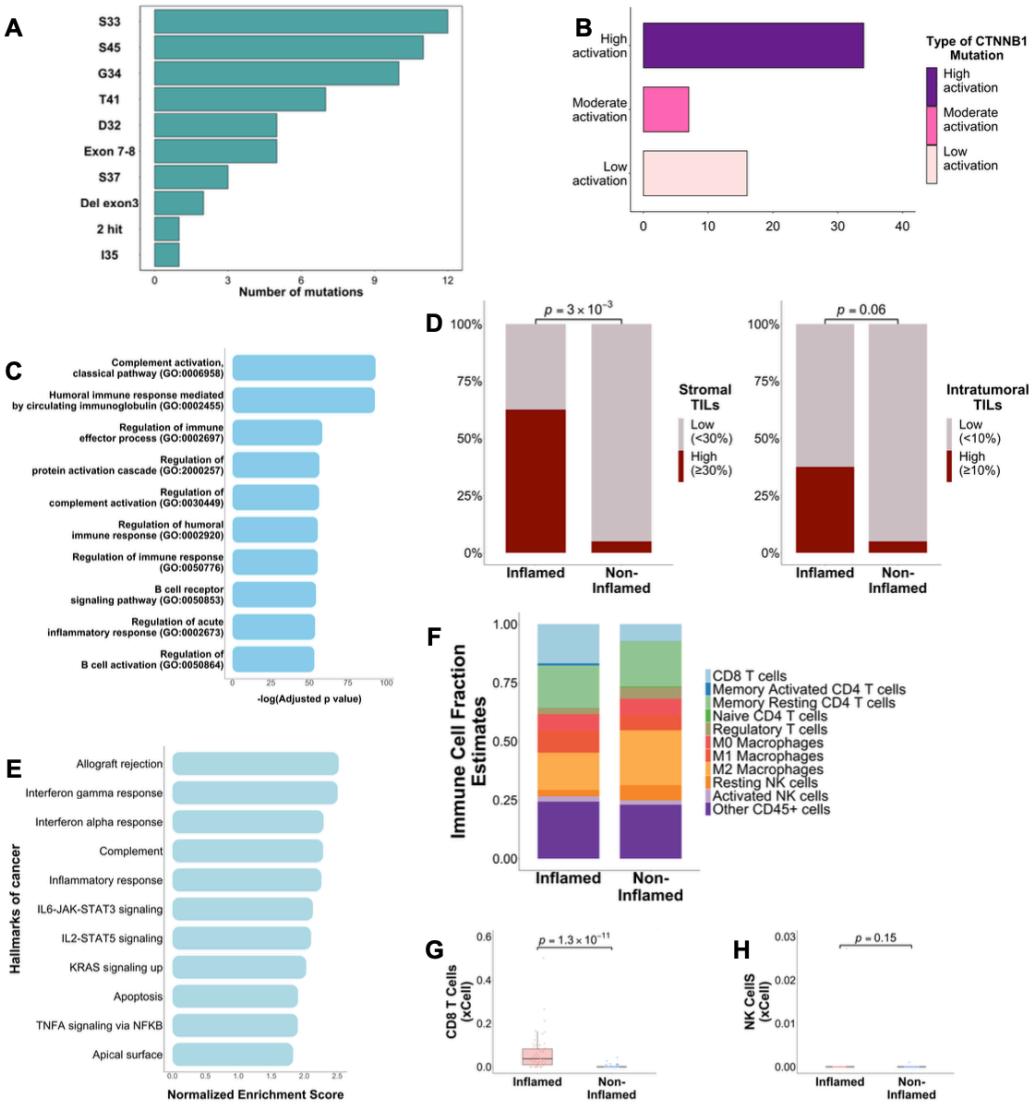
Supplementary Figure 14: Correlation between the mRNA expression and the presence of copy number deletions. *Related to Figure 4.* Boxplot representation of the correlation between the normalized RNA expression and presence of copy number deletions in subcytobands harbouring genes related to antigen presentation and interferon signalling. P-value is calculated by Wilcoxon rank-sum test.



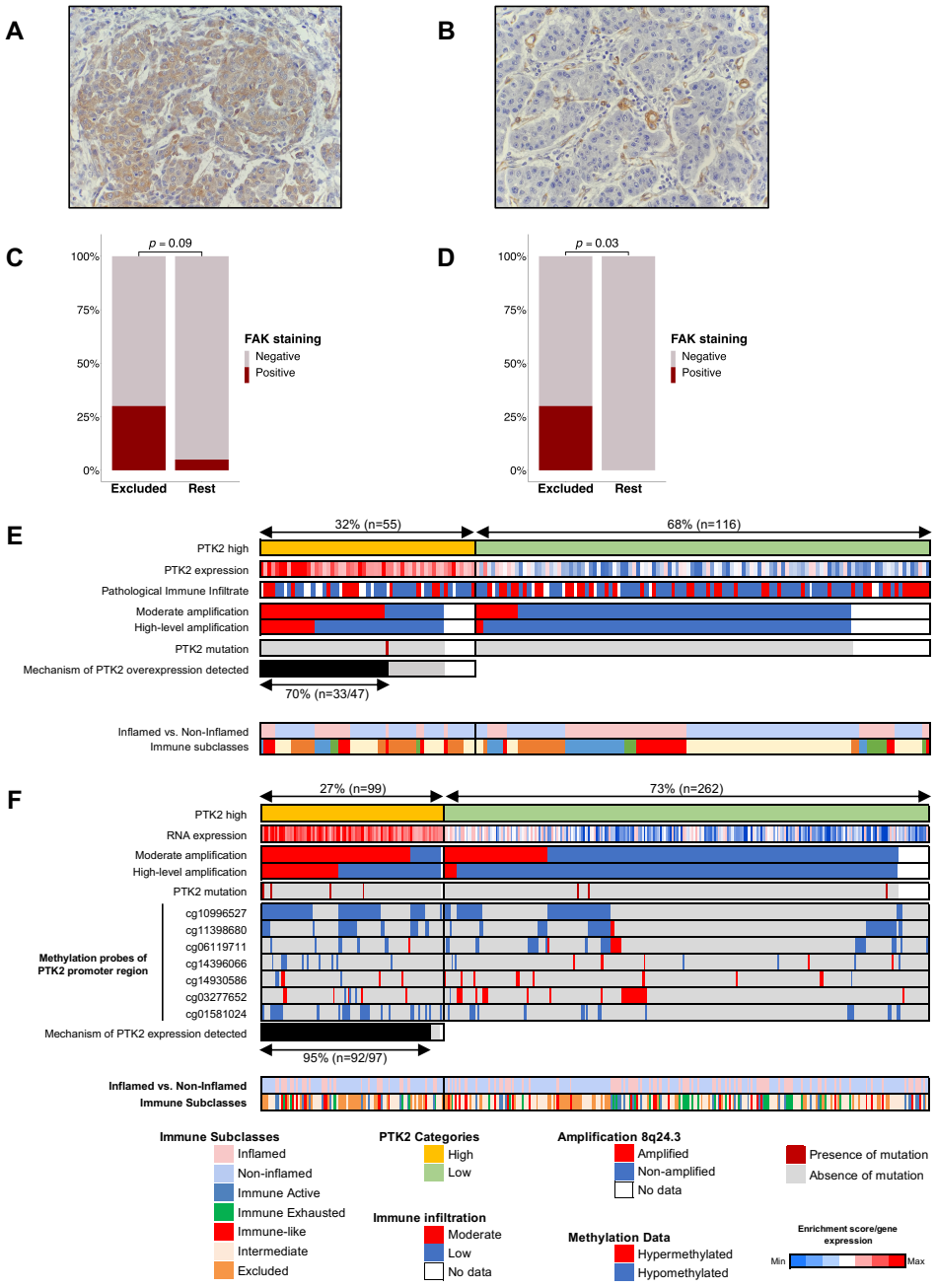
Supplementary Figure 15: Neoantigen burden distribution across the immune classes. *Related to Figure 4.* (**A-C**) Boxplot depicting the differences in (**A**) TMB and (**B**) high-affinity neoantigens from non-synonymous mutations in the discovery cohort and (**C**) neoantigen burden from non-synonimous mutations in the TCGA cohort. (**D-E**) Boxplot depicting the differences in (**D**) TIB and (**E**) high-affinity neoantigens from insertions and deletions in the discovery cohort and (**F**) neoantigen from insertions and deletions in the TCGA cohort. P-value is calculated by Kruskal-Wallis test with post-hoc Dunn's test. *, P < 0.05; **, P < 0.01; ***, P < 0.001; ****, P < 0.0001.



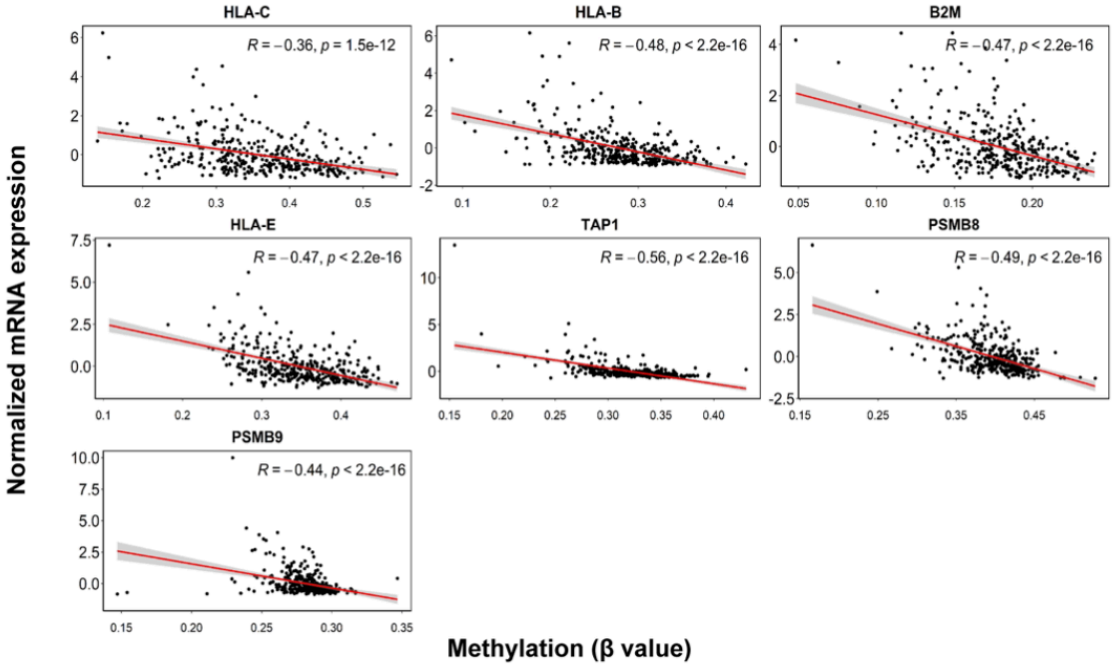
Supplementary Figure 16: Characterization of two distinct profiles of Wnt-βcatenin activated tumors based on immune features. *Related to Figure 5 and 6.* (**A**) Barplot representing the frequency of *CTNNB1* point mutations across the discovery cohort. (**B**) Barplot representing the frequency of *CTNNB1* mutations distributed according to level of activation of the Wnt-βcatenin pathway. (**C**) Barplot showing the Gene Ontology Biological Processes enrichment analysis of the differentially expressed genes in Inflamed and non-Inflamed profiles. (**D**) Stacked barplots showing the frequency of sTILs and iTILs between both profiles. (**E**) Barplot showing the GSEA analysis of Hallmark gene sets between both profiles. (**F**) Stacked barplot depicting the fraction of 22 immune cell types inferred by CIBERSORTx. (**G-H**) xCell deconvolution method analyzing (**G**) $CD8^+$ T cells and (**H**) NK cells in the distinct profiles. P-values are calculated by (**D**) Fisher's exact test and (**G-H**) Wilcoxon's rank-sum test.



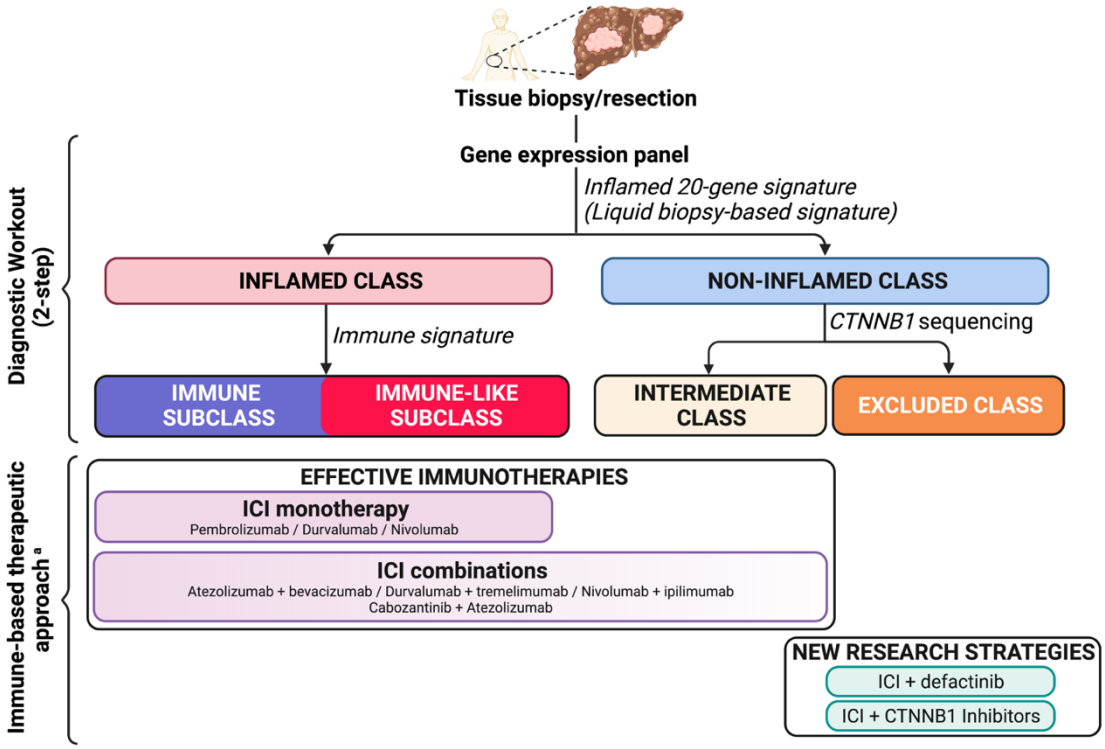
Supplementary Figure 17: Validation of FAK overexpression and mechanisms leading to overexpression of *PTK2*. *Related to Figure 6* **(A-D)** Immunohistochemical staining for focal adhesion kinase (FAK) in a (**A**) positive and (**B**) negative case. (**C-D**) Stacked barplot depicting positive and negative cases in the Excluded class versus the rest of the cohort using a (**C**) 20% or (**D**) 30% cutoff. (**C-D**) P-value calculated by Fisher's exact test. (**E-F**) Heatmap representation showing the correlation with amplification of the 8q24.3 locus and promoter methylation in (**E**) our cohort and (**F**) TCGA-LIHC cohort. The rows indicating moderate and high-level amplifications correspond to more than 3 or 4 copies of the 8q24.3 subcytoband, respectively.



Supplementary Figure 18: Correlation between the level of RNA expression and methylation of antigen type I related genes in the TCGA-LIHC cohort. *Related to Figure 6.* R and p-values are calculated by Spearman's correlation test.



Supplementary Figure 19: Algorithm for the classification or HCC patients into the Inflamed and non-Inflamed classes. *Related to Figure 7.* The diagnostic workout includes a 2-step approach in which the Inflamed signature (or a liquid-based biomarker) will first be applied and followed by the Immune signature in Inflamed tumors and *CTNNB1* sequening in Non-inflamed tumors. Inflamed tumors can be further classified in immune and immune-like subclasses. [a]The proposed treatment algorithm is based on the immune and molecular findings of this study and currently approved therapies in HCC and other tumors. Prospective validation will be required before it can be implemented in routine clinical practice. Figure created using *BioRender.com*.



References

1 Sia D, Jiao Y, Martinez-Quetglas I, *et al.* Identification of an Immune-specific Class of Hepatocellular Carcinoma, Based on Molecular Features. *Gastroenterology* 2017;**153**:812–26. doi:10.1053/j.gastro.2017.06.007

2 Villanueva A, Portela A, Sayols S, *et al.* DNA methylation-based prognosis and epidrivers in hepatocellular carcinoma. *Hepatology* 2015;**61**:1945–56. doi:10.1002/hep.27732

3 Bassaganyas L, Pinyol R, Esteban-Fabró R, *et al.* Copy-Number Alteration Burden Differentially Impacts Immune Profiles and Molecular Features of Hepatocellular Carcinoma. *Clin Cancer Res* 2020;**26**:6350–61. doi:10.1158/1078-0432.CCR-20-1497

4 Hsu CL, Ou DL, Bai LY, *et al.* Exploring Markers of Exhausted CD8 T Cells to Predict Response to Immune Checkpoint Inhibitor Therapy for Hepatocellular Carcinoma. *Liver Cancer* Published Online First: 2021. doi:10.1159/000515305

5 Hendry S, Salgado R, Gevaert T, *et al.* Assessing Tumor-Infiltrating Lymphocytes in Solid Tumors: A Practical Review for Pathologists and Proposal for a Standardized Method from the International Immuno-Oncology Biomarkers Working Group: Part 2: TILs in Melanoma, Gastrointestinal Tract Carcinom. Adv. Anat. Pathol. 2017;**24**:311–35. doi:10.1097/PAP.0000000000000161

6 Salvi S, Fontana V, Boccardo S, *et al.* Evaluation of CTLA-4 expression and relevance as a novel prognostic factor in patients with non-small cell lung cancer. *Cancer Immunol Immunother* 2012;**61**:1463–72. doi:10.1007/s00262-012-1211-y

7 Itoh S, Maeda T, Shimada M, *et al.* Role of Expression of Focal Adhesion Kinase in Progression of Hepatocellular Carcinoma. *Clin Cancer Res* 2004;**10**:2812–7. doi:10.1158/1078-0432.CCR-1046-03

8 Uzilov A V., Taik P, Cheesman KC, *et al.* USP8 and TP53 Drivers are

Associated with CNV in a Corticotroph Adenoma Cohort Enriched for Aggressive Tumors. *J Clin Endocrinol Metab* 2021;**106**:826–42. doi:10.1210/clinem/dgaa853

9 DePristo MA, Banks E, Poplin R, *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;**43**:491–8. doi:10.1038/ng.806

10 Zhang Z, Hao K. SAAS-CNV: A Joint Segmentation Approach on Aggregated and Allele Specific Signals for the Identification of Somatic Copy Number Alterations with Next-Generation Sequencing Data. *PLOS Comput Biol* 2015;**11**:e1004618. doi:10.1371/journal.pcbi.1004618

11 Szolek A, Schubert B, Mohr C, *et al.* OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics* 2014;**30**:3310–6. doi:10.1093/bioinformatics/btu548

12 Rubinsteyn A, Kodysh J, Aksoy B. hammerlab/isovar: Version 0.7.0. 2017.

13 Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* 2016;**32**:511–7. doi:10.1093/bioinformatics/btv639

14 Thorsson VV, Gibbs DL, Brown SD, *et al.* The Immune Landscape of Cancer. *Immunity* 2018;**48**:812-830.e14. doi:10.1016/j.immuni.2018.03.023

15 Franch-Expósito S, Bassaganyas L, Vila-Casadesús M, *et al.* CNApp, a tool for the quantification of copy number alterations and integrative analysis revealing clinical implications. *Elife* 2020;**9**:1–22. doi:10.7554/eLife.50267

16 Mermel CH, Schumacher SE, Hill B, *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* 2011;**12**:R41. doi:10.1186/gb-2011-12-4-r41

17 Dobin A, Davis CA, Schlesinger F, *et al.* STAR: ultrafast universal RNA-

seq aligner. *Bioinformatics* 2013;**29**:15–21. doi:10.1093/bioinformatics/bts635

18 Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;**30**:923–30. doi:10.1093/bioinformatics/btt656

19 Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;**26**:139–40. doi:10.1093/bioinformatics/btp616

20 Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 2010;**11**:R25. doi:10.1186/gb-2010-11-3-r25

21 Hoshida Y, Villanueva A, Kobayashi M, *et al.* Gene Expression in Fixed Tissues and Outcome in Hepatocellular Carcinoma. *N Engl J Med* 2008;**359**:1995–2004. doi:10.1056/NEJMoa0804525

22 Reich M, Liefeld T, Gould J, *et al.* GenePattern 2.0. *Nat Genet* 2006;**38**:500–1. doi:10.1038/ng0506-500

23 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:550. doi:10.1186/s13059-014-0550-8

24 Stephens M. False discovery rates: a new deal. *Biostatistics* 2016;**18**:275–94. doi:10.1093/biostatistics/kxw041

25 Kuleshov M V., Jones MR, Rouillard AD, *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 2016;**44**:W90–7. doi:10.1093/nar/gkw377

26 Subramanian A, Tamayo P, Mootha VK, *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* 2005;**102**:15545–50. doi:10.1073/pnas.0506580102

27 Barbie DA, Tamayo P, Boehm JS, *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*

2009;**462**:108–12. doi:10.1038/nature08460

28 Rooney MS, Shukla SA, Wu CJ, *et al.* Molecular and Genetic Properties of Tumors Associated with Local Immune Cytolytic Activity. *Cell* 2015;**160**:48–61. doi:10.1016/j.cell.2014.12.033

29 Yoshihara K, Shahmoradgoli M, Martínez E, *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 2013;**4**:2612. doi:10.1038/ncomms3612

30 Newman AM, Steen CB, Liu CL, *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* 2019;**37**:773–82. doi:10.1038/s41587-019-0114-2

31 Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol* 2017;**18**:220. doi:10.1186/s13059-017-1349-1

32 Ayers M, Lunceford J, Nebozhyn M, *et al.* IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest* 2017;**127**:2930–40. doi:10.1172/JCI91190

33 Voskoboinik I, Whisstock JC, Trapani JA. Perforin and granzymes: function, dysfunction and human pathology. *Nat Rev Immunol* 2015;**15**:388–400. doi:10.1038/nri3839

34 Dangaj D, Bruand M, Grimm AJ, *et al.* Cooperation between Constitutive and Inducible Chemokines Enables T Cell Engraftment and Immune Attack in Solid Tumors. *Cancer Cell* 2019;**35**:885-900.e10. doi:10.1016/j.ccell.2019.05.004

35 Biasci D, Smoragiewicz M, Connell CM, *et al.* CXCR4 inhibition in human pancreatic and colorectal cancers induces an integrated immune response. *Proc Natl Acad Sci* 2020;**117**:28960–70. doi:10.1073/pnas.2013644117

36 Hoshida Y, Nijman SMB, Kobayashi M, *et al.* Integrative Transcriptome Analysis Reveals Common Molecular Subclasses of Human Hepatocellular Carcinoma. *Cancer Res* 2009;**69**:7385–92.

doi:10.1158/0008-5472.CAN-09-1089

37 Chiang DY, Villanueva A, Hoshida Y, *et al.* Focal Gains of VEGFA and Molecular Classification of Hepatocellular Carcinoma. *Cancer Res* 2008;**68**:6779–88. doi:10.1158/0008-5472.CAN-08-0742

38 Lachenmayer A, Alsinet C, Savic R, *et al.* Wnt-Pathway Activation in Two Molecular Classes of Hepatocellular Carcinoma and Experimental Modulation by Sorafenib. *Clin Cancer Res* 2012;**18**:4997–5007. doi:10.1158/1078-0432.CCR-11-2322

39 Coulouarn C, Factor VM, Thorgeirsson SS. Transforming growth factor-β gene expression signature in mouse hepatocytes predicts clinical outcome in human cancer. *Hepatology* 2008;**47**:2059–67. doi:10.1002/hep.22283

40 Moffitt RA, Marayati R, Flate EL, *et al.* Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat Genet* 2015;**47**:1168–78. doi:10.1038/ng.3398

41 Liberzon A, Subramanian A, Pinchback R, *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011;**27**:1739–40. doi:10.1093/bioinformatics/btr260

42 Sangro B, Melero I, Wadhawan S, *et al.* Association of inflammatory biomarkers with clinical outcomes in nivolumab-treated patients with advanced hepatocellular carcinoma. *J Hepatol* 2020;**73**:1460–9. doi:10.1016/j.jhep.2020.07.026

43 Haber PK, Torres M, Andreu C, *et al.* PO-1420 Molecular markers of response to anti-pd1 therapy in advanced hepatocellular carcinoma. *J Hepatol* 2021;**75**:S521-522.

44 Cabrita R, Lauss M, Sanna A, *et al.* Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* 2020;**577**:561–5. doi:10.1038/s41586-019-1914-8

45 Charoentong P, Finotello F, Angelova M, *et al.* Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype

Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep* 2017;**18**:248–62. doi:10.1016/j.celrep.2016.12.019

46 Davoli T, Uno H, Wooten EC, *et al.* Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science (80-)* 2017;**355**:eaaf8399. doi:10.1126/science.aaf8399