Supplementary Data

Title

Copy number alteration burden differentially impacts immune profiles and molecular features of hepatocellular carcinoma

Authors

Laia Bassaganyas[1,#], Roser Pinyol[1,#], Roger Esteban-Fabró[1,#], Laura Torrens[1], Sara Torrecilla[1], Catherine E. Willoughby[1], Sebastià Franch-Expósito[2], Maria Vila-Casadesús[3], Itziar Salaverria[4,5], Robert Montal[1], Vincenzo Mazzaferro[6], Jordi Camps[2], Daniela Sia[7], Josep M Llovet* [1,7,8]

[#]shared first authorship

Affiliations

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

[2]Gastrointestinal and Pancreatic Oncology Group, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Catalonia, Spain

[3]Bioinformatics Unit, CIBERehd, Barcelona, Catalonia, Spain

[4]Lymphoid Neoplasms Program, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Catalonia, Spain

[5]Tumores Hematológicos, Centro de Investigación Biomédica en Red de Cáncer (CIBERonc), Madrid, Spain

[6]Gastrointestinal Surgery and Liver Transplantation Unit, National Cancer Institute, Milan, Italy.

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

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

Corresponding author

*Josep M. Llovet; Liver Cancer Translational Research Laboratory, Liver Unit, IDIBAPS-Hospital Clinic, Rosselló 153, 08036, Barcelona, Catalonia, Spain; Tel. 0034-932.279.155; Email address: jmllovet@clinic.cat

Content:

- **Supplementary Table 3:** Excel sheet presenting the INGENUITY® Pathway Analysis of the discovery cohort according to broad scores (BS).
- **Supplementary Table 4:** Excel sheet with the clinico-pathological data displayed according to BS and FS.
- **Supplementary Table 5:** Molecular and immune characterization of the discovery cohort according to broad scores (BS).
- **Supplementary Table 6:** Molecular and immune characterization of the validation cohort according to broad scores (BS).
- **Supplementary Table 7:** Molecular and immune characterization of the discovery cohort according to focal scores (FS).
- **Supplementary Table 8:** Molecular and immune characterization of the validation cohort according to focal scores (FS).
- **Supplementary Table 9:** Gene groups representing immune clusters as major determinants of immunogenicity.
- **Supplementary Table 10:** Publicly available gene signatures used in the study.
- **References**

Supplementary Methods

Study cohorts

For the purpose of the study, a total of 452 tumor and paired adjacent non-tumor liver tissues were analyzed (Supplementary Figure 1A), including a discovery cohort of 107 surgically-resected fresh frozen (FF) samples (HEPTROMIC)[1] and a validation cohort of 345 Liver Hepatocellular Carcinoma (LIHC) patients publicly available from The Cancer Genome Atlas (TCGA)[2]. Samples from the discovery cohort were retrospectively collected in the setting of the HCC Genomic Consortium upon institutional review board approval and patients' written informed consent, as previously reported[1]. Detailed information of clinical and pathological features of these samples is summarized in Supplementary Table 1. Also, we analyzed one 'in-house' cohort with 25 patients showing low-grade (n=9) and high-grade (n=16) pre-malignant dysplastic nodules (LGDN and HGDN, respectively); and 18 very early HCCs (veHCC) corresponding to tumors of less than 2cm in size from the HEPTROMIC cohort. These two last cohorts had been previously published[1,3]. Finally, a publicly available transcriptome dataset corresponding to pre-treatment samples of 65 cancer patients treated with either nivolumab (43%) or pembrolizumab (57%), both anti-PD-1 immune checkpoint inhibitors[4], was used to run the subclass mapping.

Samples' histological evaluation

Samples were evaluated histopathologically to determine presence of microscopic vascular invasion (0=absent; 1=present), tumor degree of

differentiation (Edmondson-Steiner grades G1 to G4), and tumor immune infiltration (0= absent; 1=mild; 2=mild-moderate; 3=moderate; 4=abundant). Presence of lymphoid aggregates was assessed in the non-tumor liver tissue sections.

Genomic profiling

For DNA extraction of the discovery cohort (HEPTROMIC), we used the Invitrogen ChargeSwitch genomic DNA Mini Tissue kit (Invitrogen). Median sample storage time from collection to DNA extraction was 7 years. Samples (n=107) were analysed with Human OmniExpress Exome8v1 SNParray (Illumina, San Diego, USA). Expression and methylation data of the HEPTROMIC cohort had been previously reported[1], and was available for 102 out of the 107 HCC samples with good quality Single Nucleotide Polymorphism (SNP) array profiles. Data are available under the Gene Expression Omnibus accession number: GSE153338 and GSE63898. Samples from the validation set (TCGA-LIHC, n=345) were analyzed with Affymetrix 6.0 SNP array (Affymetrix, CA, USA). TCGA RNA-sequencing data (Illumina HiSeq 2000) were available through the cBioportal (https://www.cbioportal.org/), and their corresponding methylation data was retrieved from the NCI Genomic Data Commons (GDC) using the R package *TCGAbiolinks*.

CNA data processing and determination of the CNA level

The HEPTROMIC SNP array data were processed using the software Genome Studio Genotyping Module v2.0 (Illumina) to extract Log R Ratio (LRR) and B-Allele Frequency (BAF) values. Segmented copy number data were calculated using the SAAS-CNV software[5] with default parameters. TCGA-LIHC pre-processed level-3 segmented copy number data with LRR values were directly downloaded from the Broad Institute GDAC FireBrowser (http://gdac.broadinstitute.org). The determination of ploidy and aberrant tumor cell fraction (purity) of HEPTROMIC samples was carried out with ASCATv2.4 software[6], whereas TCGA-LIHC ploidy estimation was obtained from COSMIC database[7] and purity values were obtained from published data[8]. First, we used CNApp [9] 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[10,11]. Finally, we used CNApp to quantify the individual CNA burdens of each sample within the HEPTROMIC (n=107) and the TCGA-LIHC (n=345) cohorts. The CNApp provides two scores per sample, one that reflect the genomic burden of all broad CNAs (the so called broad score, BS), and one that reflects all focal CNAs (focal score, FS). These scores are based on the number, amplitude and length of the CNAs. In the event of broad and focal CNAs affecting the same genomic region, they were contemplated separately within BS and FS,

respectively. A more detailed description of the scores can be found in Franch-Exposito et al. [9]. Broad and focal scores for the remaining TCGA samples were also obtained from [9]. Copy number data from the two cohorts of dysplastic nodules were obtained from [3].

Molecular characterization

For the molecular characterization of the HCC tumors according to their different aneuploidy profiles, both broad scores (BS) and focal scores (FS) in the HEPTROMIC and TCGA-LIHC cohorts were dichotomized into low and high. High-BS threshold was set at ≥11 (the upper quartile of both the TCGA pan-cancer cohort and the discovery cohort), and low-BS at ≤4 (coinciding with the lower quartile in the discovery cohort). In terms of FS, low-FS samples were defined as those with FS ≤ 13.5, while high-FS was defined as those with FS ≥ 47, considering the quartile values in the discovery cohort. Tumors with BS and FS values between low and high score thresholds were classified as 'intermediate'. Genes deregulated in low-BS/FS or high-BS/FS were identified through the Comparative Marker Selection module of GenePattern, using a fold change (FC)>1.5 and an FDR<0.05. Gene Set Enrichment Analysis (GSEA)[12], ssGSEA[13] (both GenePattern modules) and INGENUITY® Pathway Analysis software (IPA®) were used to assess enrichment of activated pathways in each group. Sample class predictions to identify molecular classes or enrichment of specific gene signatures (see Supplementary Table 9) were performed using the GenePattern Nearest Template Prediction module as in previous publications

[14]. Subclass mapping (SubMap) analysis (Gene Pattern), a bioinformatic method to quantitatively evaluate the similarity between independent cohorts, was used to measure similarity in gene expression between the HEPTROMIC cohort and a publicly available cohort of cancer patients treated with immune checkpoint inhibitors [4].

Immune characterization

The tumor immunophenotype of samples from both discovery and validation cohorts was assessed with an *in-house* adapted version of the previously published Immunophenoscore[15]. We generated immunophenograms graphically representing the Spearman's correlation between the CNA scores and the expression of the 162 genes or groups of genes described as determinants of tumor immunogenicity (listed in Supplementary Table 8). Cytolytic activity of HEPTROMIC and TCGA-LIHC samples was calculated as the geometric mean of the genes granzyme A (GZMA) and perforin-1 (PRF1), as previously described[16]. Infiltration of immune cells in tumor tissue was inferred from expression data through the Immune Score obtained from the ESTIMATE software[17]. Pan-cancer ESTIMATE values for the 10,635 TCGA samples were directly obtained from the tool developers[17]. For the analysis of mutations and neo-antigens in the TCGA-LIHC cohort, the list of events per sample provided by Rooney et al.[16] was used. In order to identify candidate genes involved in determining the immune phenotypes observed for low-BS and high-BS samples, genes present in each copy number gain and loss were extracted using the

genome coordinates of each CNA segment, and a matrix with the frequency of each gene according to the broad CNA load (low-BS, high-BS, intermediate-BS) was generated. Significant differences were captured using a Fisher exact test, and multiple testing was corrected by FDR. B2M and HLA-DQB1 were identified as significantly different in terms of CNA among low-BS and high-BS simultaneously in the discovery and validation cohorts.

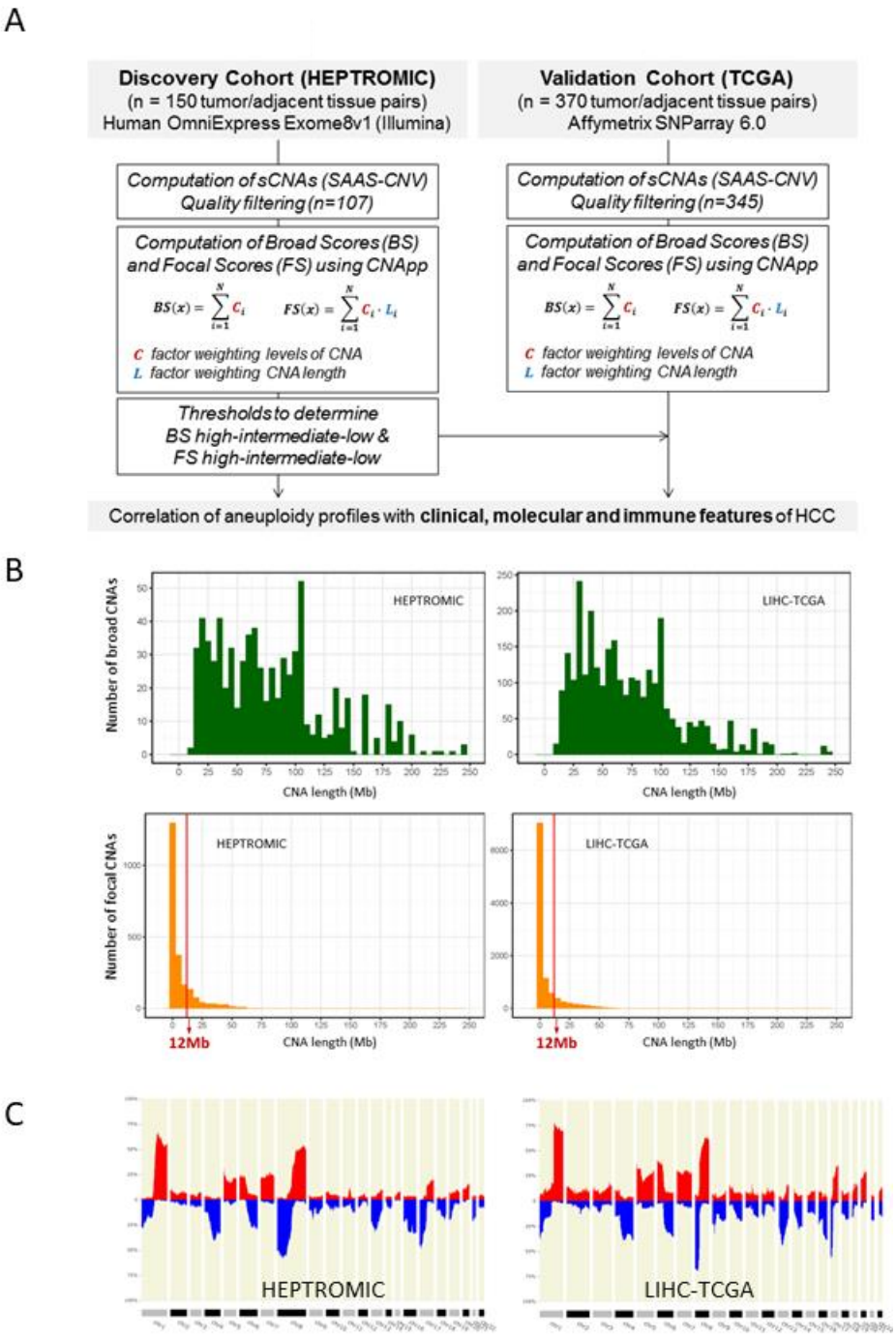
Mutational profiling

The mutational landscape of the TCGA-LIHC cohort was obtained from Rooney et al.[16]. Mutations in *TERT* and *CTNNB1* in the HEPTROMIC cohort were assessed by Sanger sequencing.

Statistical analysis

Statistical analysis was conducted using R (version 3.4.4). Specifically, we performed non-parametric tests for the comparison of distribution of continuous variables (Wilcoxon or Kruskal-Wallis tests) and binary counts (Fisher's exact test). For the assessment of correlations between two continuous variables, we used Spearman's rank correlation coefficient test. IBM-SPSS.v25 software was used for Kaplan-Meier analyses, log-rank tests and Cox regression modelling to evaluate the associations of molecular and clinical variables with overall survival and tumor recurrence. In all analysis, either a p-value <0.05 (two-sided) or an FDR<0.1 were considered to be statistically significant.

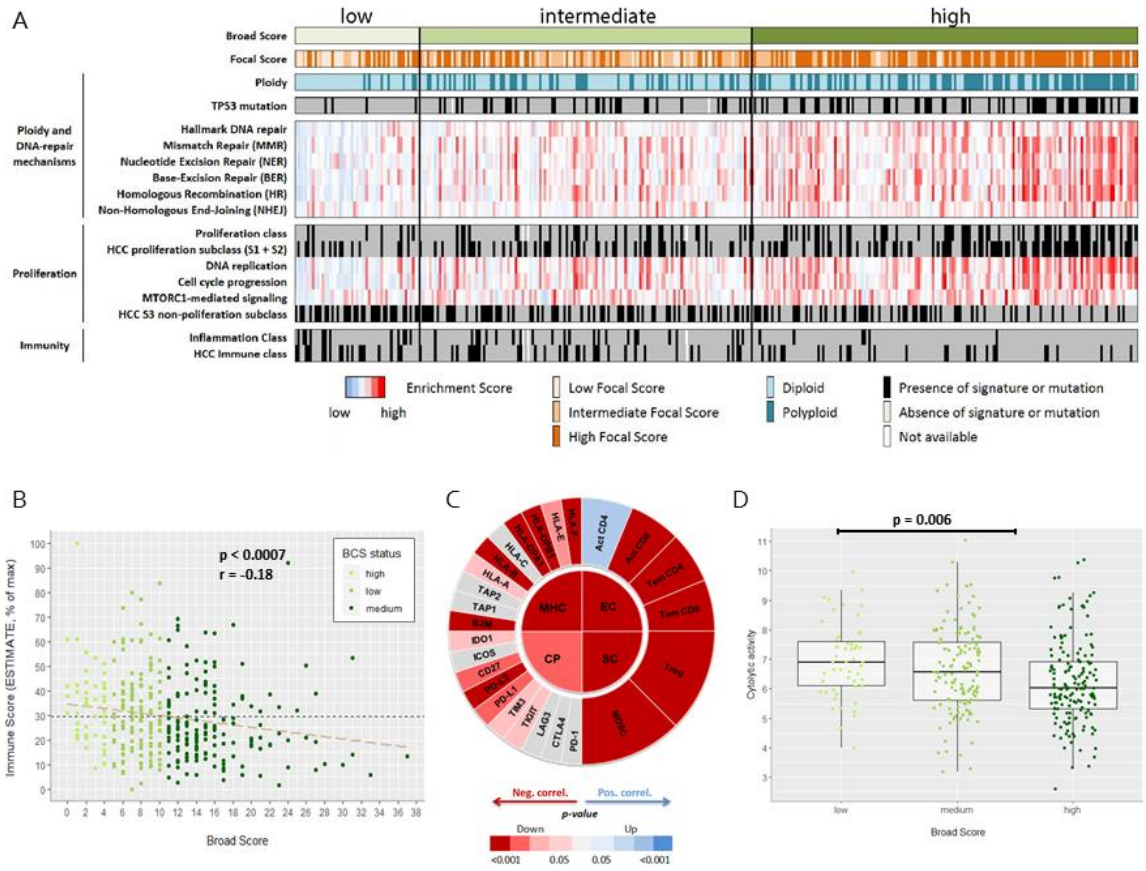
Supplementary Figure 1



Supplementary Figure 1. Flow chart of the study and distribution of broad and focal copy number scores (BS and FS, respectively). (A) A total of 520

HCC samples, and their paired non-tumor adjacent tissue, were analysed in this study. A discovery cohort (HEPTROMIC) consisted of 150 fresh frozen HCCs, and a validation cohort was composed of 370 HCC samples (TCGA). (B) Distribution of copy-number and length of broad (in green) and focal (in orange) CNA segments in the discovery cohort (n=107, HEPTROMIC) and in the validation cohort (n=345, LIHC-TCGA). (C) Genomic frequency distributions of gains (red) and losses (blue) in the discovery and validation cohorts.

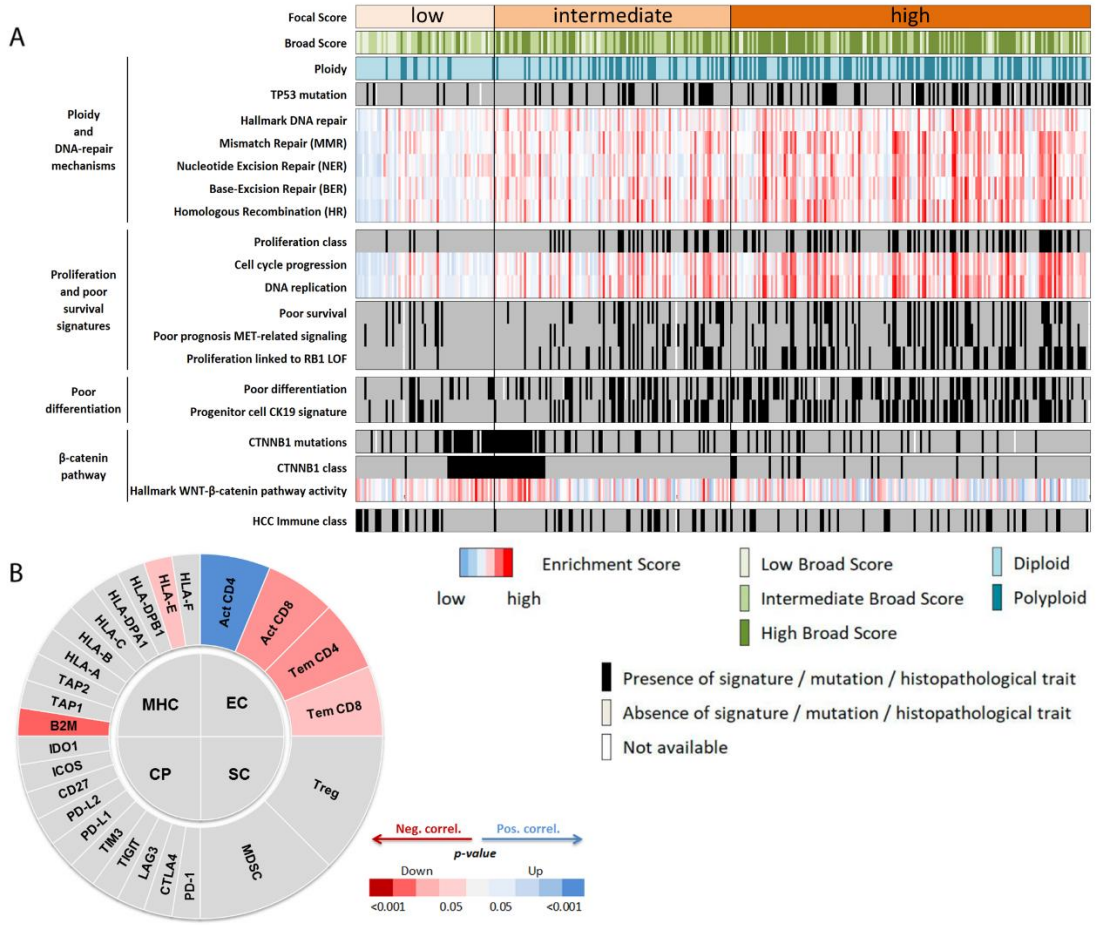
Supplementary Figure 2



Supplementary Figure 2. BS-low and BS-high patterns of CNAs in the validation cohort display also distinct molecular and immune profiles. (A) BS-low tumors were significantly associated with high immune infiltration, diploidy and functional *TP53*. In contrast, BS-high HCCs were polyploid, proliferative and with low immune signaling. (B) Transcriptome-based estimation of immune cell infiltration according to broad chromosomal burdens (validation cohort). (C) Immunophenoscore diagram correlating BS with expression features of effector cells (EC), immune checkpoints (CP), major histocompatibility complex-related

components (MHC) and suppressor cells (SP). (D) BS-low HCCs exhibited higher cytolytic activity compared to the rest of tumors.

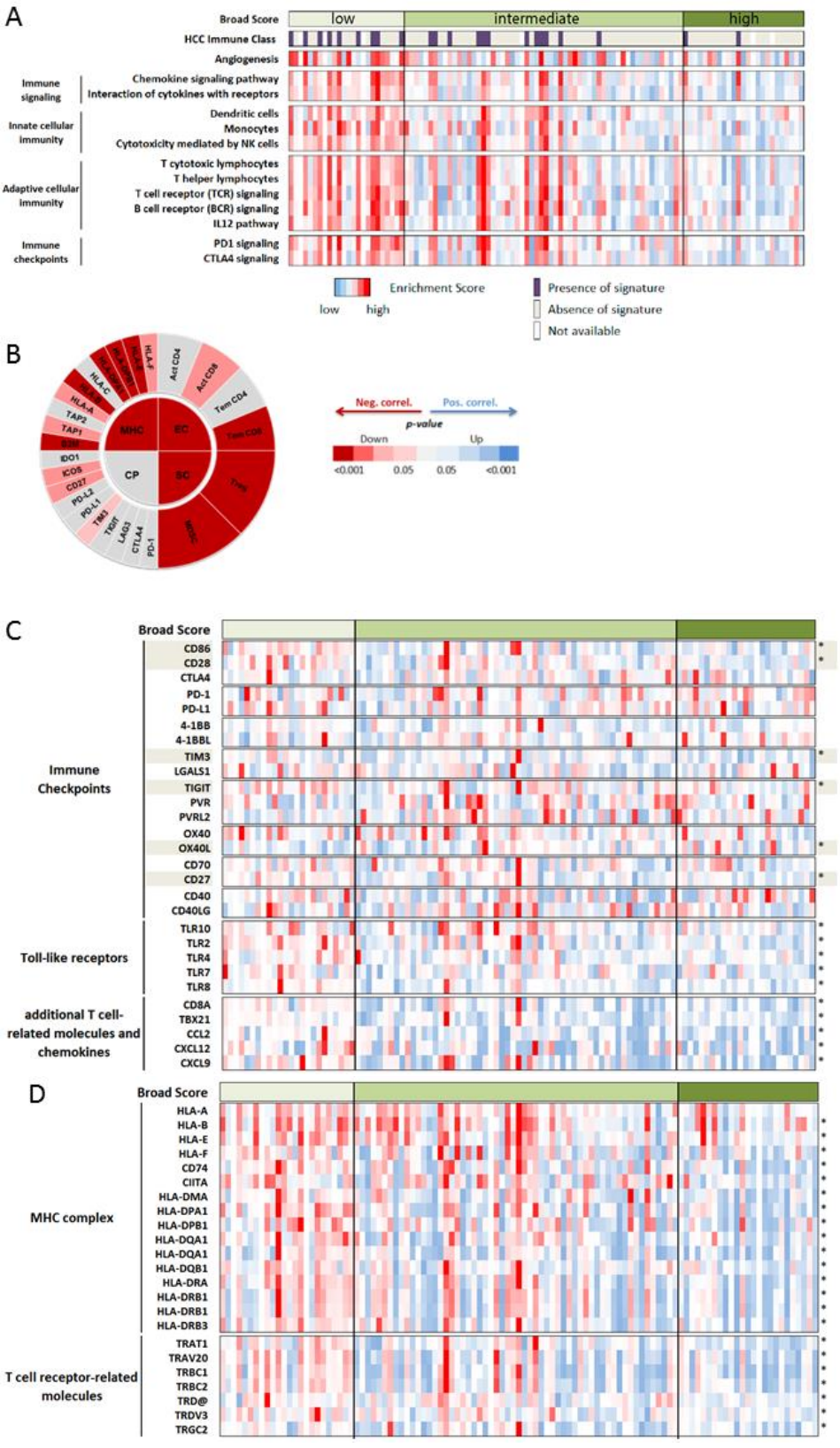
Supplementary Figure 3



Supplementary Figure 3. Burden of focal chromosomal alterations impacts on molecular features but has no effect on the immune profiles also in the validation cohort. (A) Heatmap with data from the LIHC-TCGA cohort, demonstrating that HCC tumors with higher burden of focal events (high-FS) were mainly polyploid, enriched in poor prognosis and proliferation, and presented poor cell differentiation. In contrast, low/intermediate-FS tumors were enriched in Wnt-β-catenin signaling. Overall, focal scores were not associated with immune features. (B) Immunophenoscore diagram displaying the absence of

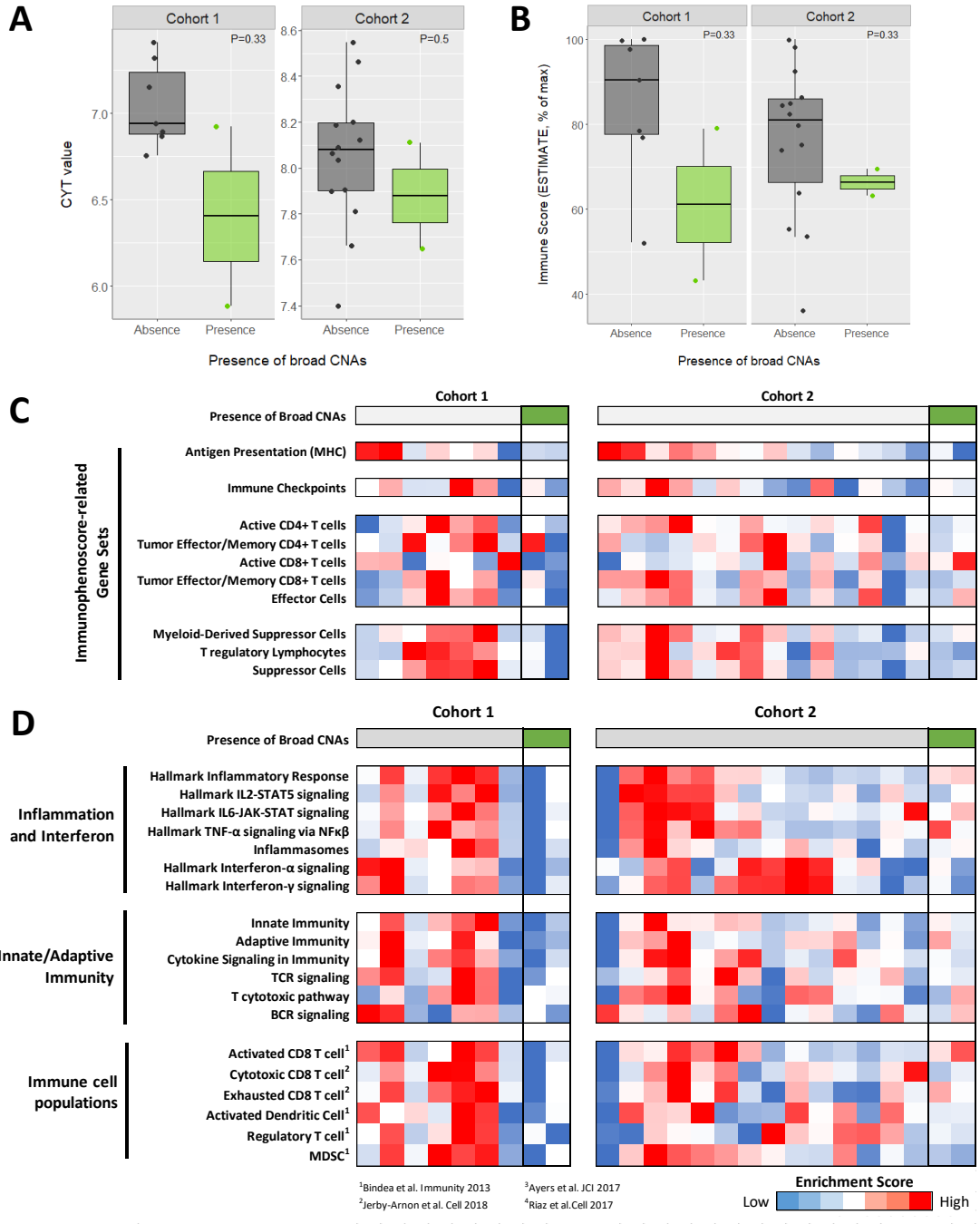
correlation between FS and effector cells (EC), immune checkpoints (CP), major histocompatibility complex-related components (MHC) and suppressor cells (SP).

Supplementary Figure 4



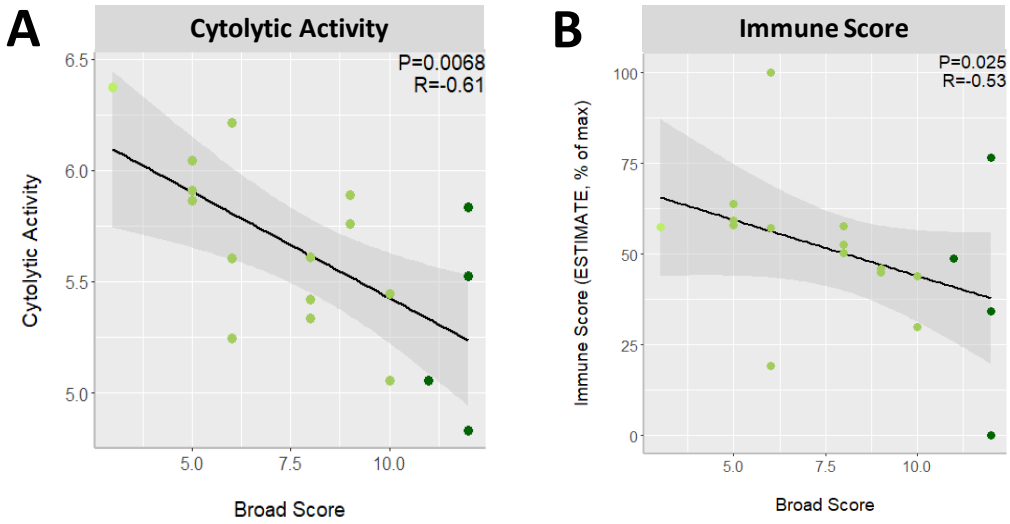
Supplementary Figure 4. BS-low HCC tumors exhibit enrichment of immune traits. (A) Heatmap with ssGSEA signatures related to immunity (discovery cohort), according to the broad score levels. (B) Immunophenoscore diagram correlating BS with expression features from effector cells (EC), immune checkpoints (CP), major histocompatibility complex-related components (MHC) and suppressor cells (SP). High-BSs were negatively correlated with features of EC, MHC and SC. (C) Heatmap displaying gene expression levels of immune checkpoints and immune receptors, according to the broad score levels. (D) Heatmap displaying gene expression levels of MHC-complex and T cell receptor-related molecules. In (C) and (D) gene expression levels are graded from red (high expression) to blue (low expression), and asterisks indicate significant differences in terms of expression between HCC tumors low-BS and high-BS.

Supplementary Figure 5



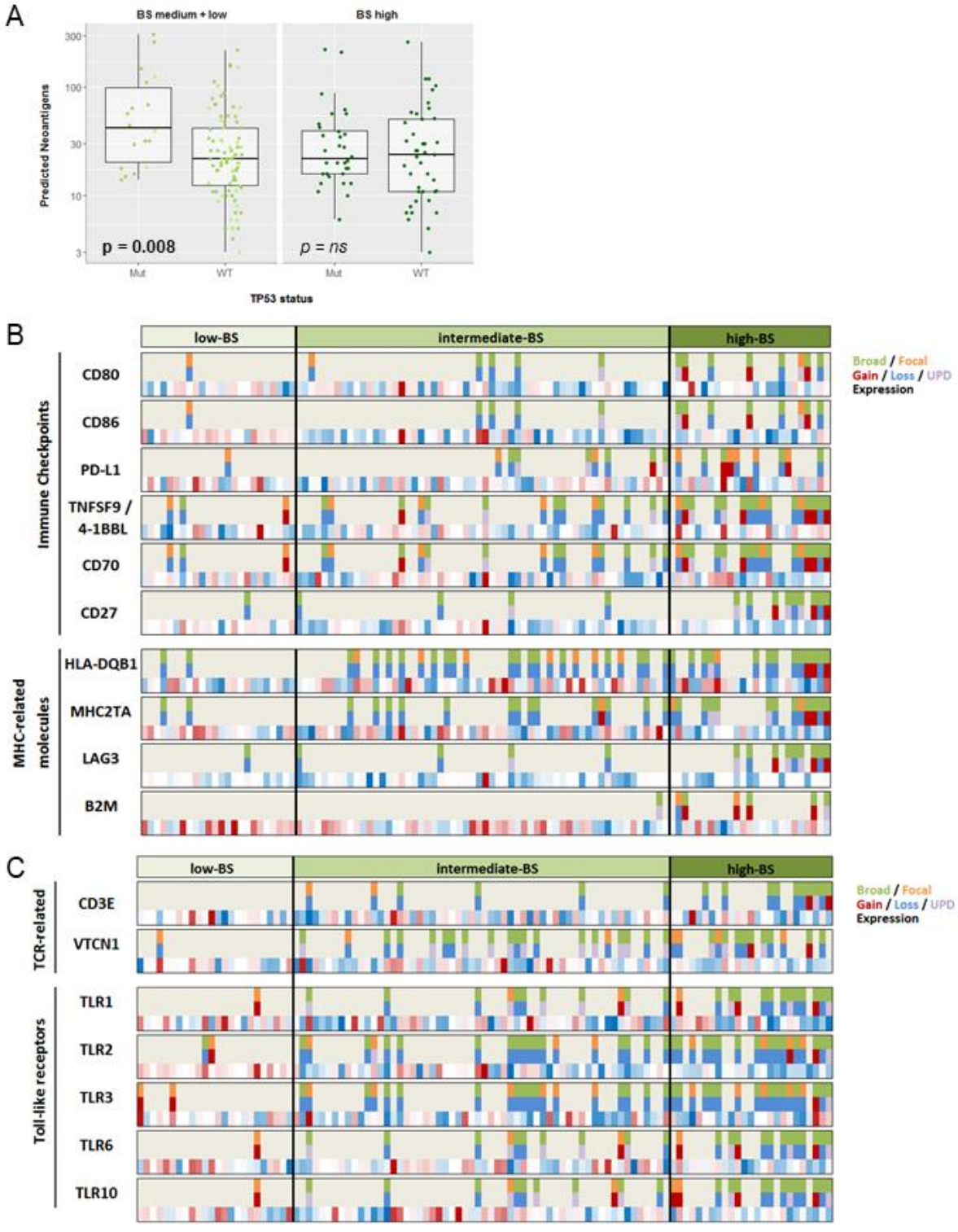
Supplementary Figure 5. Pre-neoplastic lesions classified as high-BS displayed trends towards reduced molecular features of active antitumor immunity. (A,B) The presence of broad CNAs (equivalent to high-BS) in pre-neoplastic lesions (high grade dysplastic nodules and low grade dysplastic nodules) is linked with a trend towards reduced (A) cytolytic activity and (B) expression-inferred immune cell infiltration using the ESTIMATE tool. (C,D) ssGSEA data indicating that broad CNAs in dysplastic nodules were linked to trends towards a reduced enrichment in transcriptomic features of effector cells, immune checkpoints, major histocompatibility complex-related components and suppressor cells (C), as well as trends towards reduced expression of immune-related pathways as observed with HCC tumors (D). P-values have been obtained from a Mann-Whitney-Wilcoxon test.

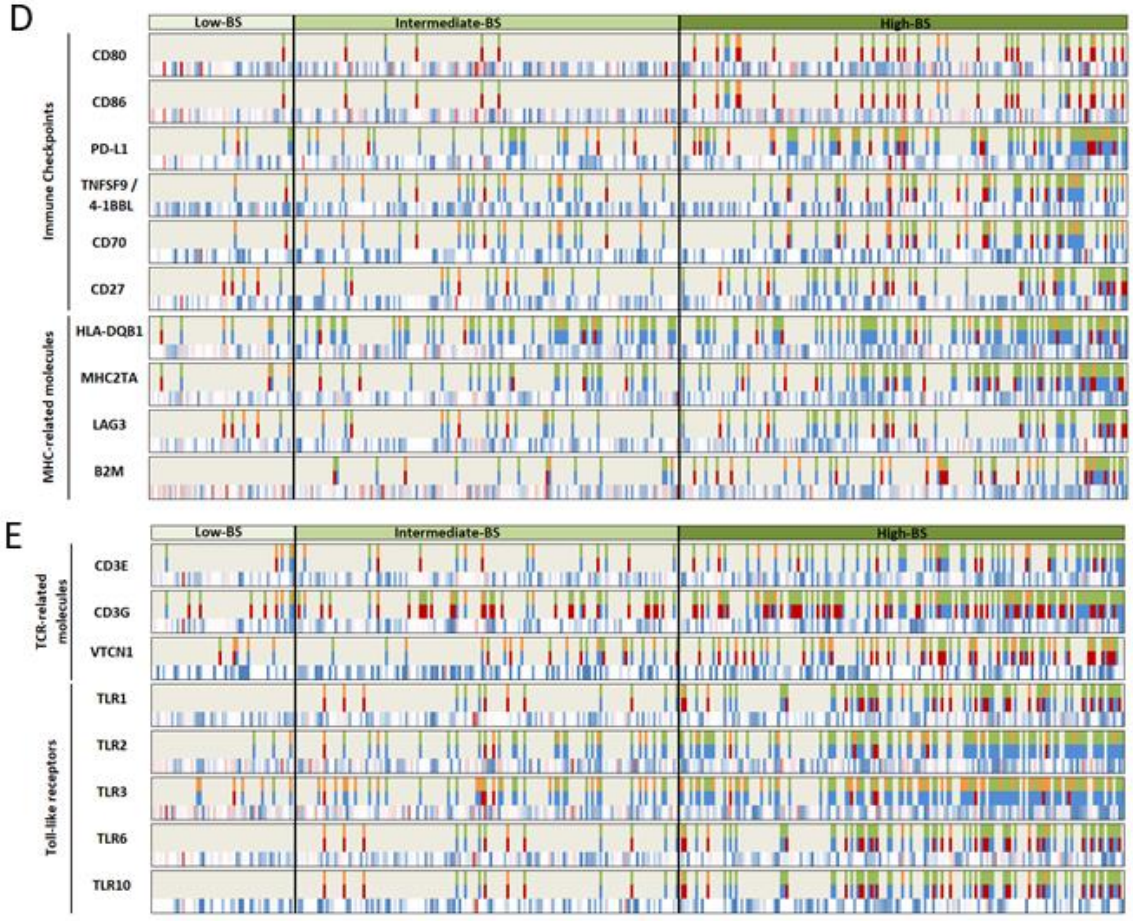
Supplementary Figure 6



Supplementary Figure 6. Correlation between immunity and BS in very early HCCs. The associations between Broad Score (BS) and reduced antitumor immunity features were maintained among the subgroup of small HCCs in Heptromic (n=18). We observed that BS was negatively correlated with (A) Cytolytic activity and (B) expression-inferred immune cell infiltration using the ESTIMATE tool.

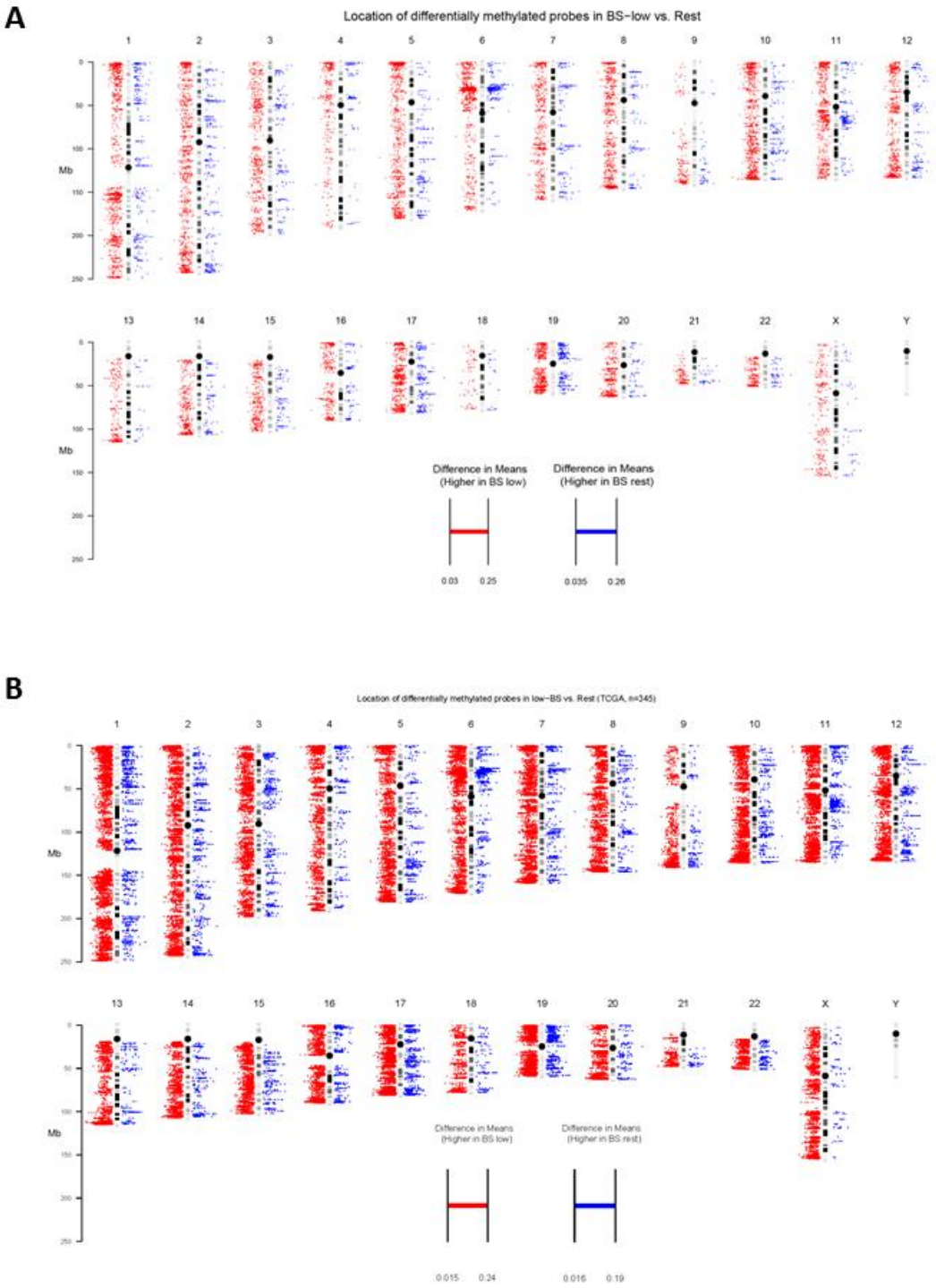
Supplementary Figure 7





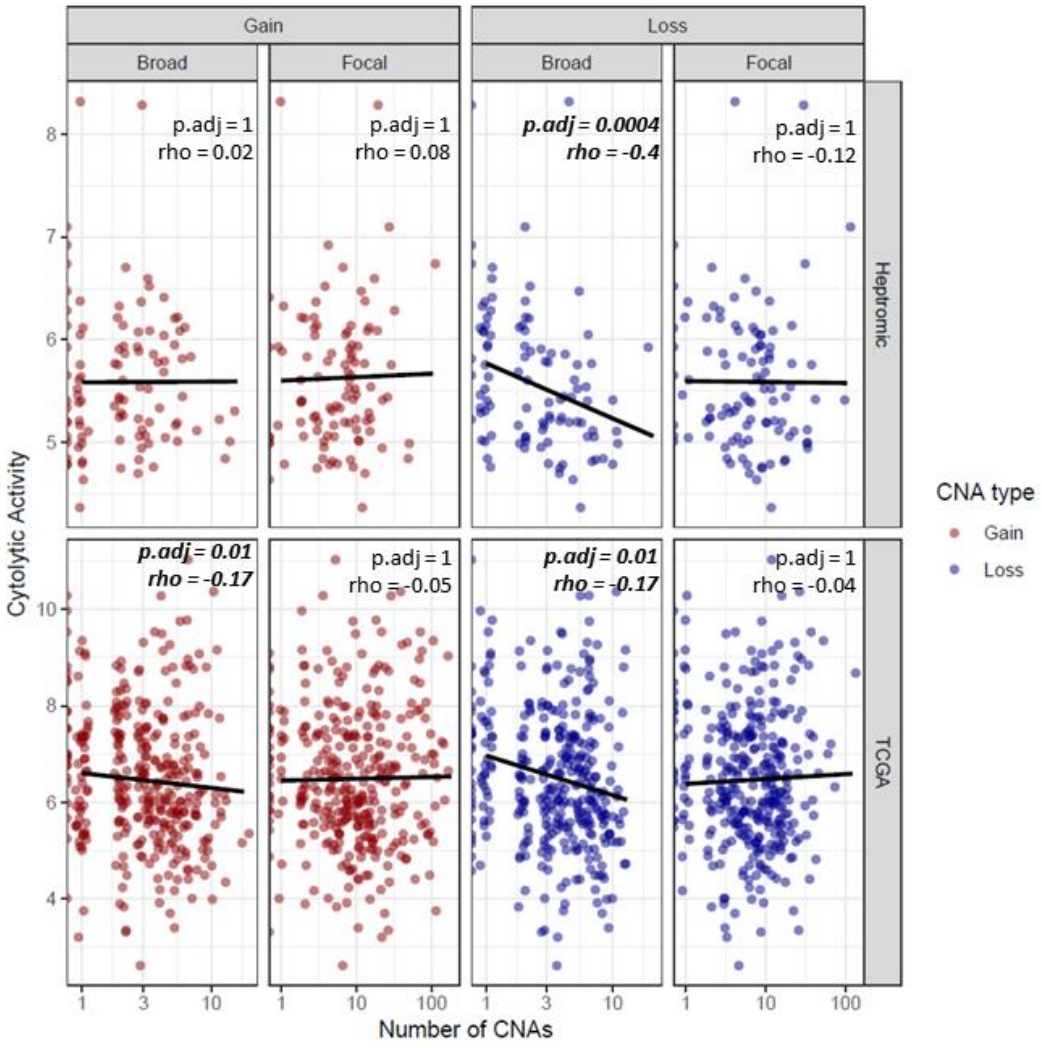
Supplementary Figure 7. Candidate determinants of immune profiles observed for low-BS and high-BS HCC tumors. (A, B) Heatmaps displaying enrichment copy number losses and uniparental disomies (UPDs) −also referred to as copy neutral losses of heterozygosity (CN-LOH) − containing immune-related genes in high-BS tumors in the discovery cohort. Alterations of specific immune-related genes were estimated using the refined copy number segmented data obtained using CNApp. (C, D) Enrichment of copy number losses in immune-related gens in high-BS tumors from the validation cohort.

Supplementary Figure 8



Supplementary Figure 8. High-BS HCCs present widespread demethylation. Genome-wide plot of the discovery (A) and validation (B) cohorts displaying probes differentially methylated in low-BS tumors vs. rest. Probes in red are the ones that were more methylated in low-BS than in non-low-BS tumors. Probes in blue are the ones that were less methylated in low-BS vs rest. The distance to the chromosome represents the difference between the compared means. Plotted were only these probes that were differentially methylated using and FDR<0.05.

Supplementary Figure 9



Supplementary Figure 9. Number of broad losses correlate with low cytolytic activity in HCC. Number of CNAs is represented in a logarithmic scale. Correlation degree was evaluated using the Spearman regression coefficient. Significance was evaluated using the adjusted p value.

Supplementary Table 1: Clinico-pathological features of HCC patients:

Variable	HEPTROMIC (n=107)	LIHC-TCGA (n=345)	p-value
Age (years)			
median (range)	66 (41-83)	61 (52-69)	$2.6 \cdot 10^{-5}$
≥65 years, n (%)	62 (58)	146 (42)	0.005
Gender			
Males, n (%)	83 (78)	235 (68)	ns
Females, n (%)	24 (22)	110 (32)	
Etiology			
Hepatitis B, n (%)	24 (23)	77 (32)	ns
Hepatitis C, n (%)	48 (46)	30 (12)	$3.6 \cdot 10^{-11}$
Alcohol, n (%)	14 (13)	110 (45)	$1.4 \cdot 10^{-8}$
NASH, n (%)	7 (7)	12 (5)	ns
Other, n (%)	12 (11)	15 (6)	ns
Tumor stage			
BCLC			
BCLC 0, n (%)	12 (11)	-	
BCLC A, n (%)	80 (75)	-	
BCLC B, n (%)	8 (8)	-	
BCLC C, n (%)	6 (6)	-	
AJCC			
T1, n (%)	-	169 (49)	
T2, n (%)	-	88 (26)	
T3, n (%)	-	73 (21)	
T4, n (%)	-	12 (4)	
Child-Pugh score			
A, n (%)	104 (98)	206 (91)	0.01
B, n (%)	2 (2)	21 (9)	
Tumor size (cm)			
≤ 2 cm, n (%)	18 (17)	-	
>2 - <3 cm, n (%)	15 (14)	-	
≥ 3 cm, n (%)	73 (69)	-	
Number of nodules			
Single nodule, n (%)	82 (77)	-	
Multiple nodules, n (%)	25 (23)	-	
Vascular invasion			
Yes, n (%)	38 (36)	103 (35)	ns
No, n (%)	67 (64)	193 (65)	
Tumor satellites			
Yes, n (%)	26 (24)	-	
No, n (%)	81 (76)	-	
Degree of tumor differentiation			
Well differentiated, n (%)	13 (16)	43 (13)	ns
Moderately differentiated, n (%)	50 (60)	168 (49)	ns
Poorly differentiated, n (%)	20 (24)	118 (38)	0.02
Bilirubin (mg/dl)			
median (range)	1.1 (0.3 – 3.2)	0.7 (0.1 - 9)	$5.8 \cdot 10^{-8}$
> 1 mg/dl, n (%)	49 (46)	62 (22)	$3 \cdot 10^{-6}$
Albumin (g/l)			
median (range)	4.1 (2.4 - 5.5)	4 (0.2 - 6.9)	0.02
< 3.5 g/l, n (%)	14 (13)	69 (25)	0.02
Platelet count			
median (range)	156 (33 - 373)	200.5 (4 - 608)	$1.7 \cdot 10^{-12}$
< 100,000/mm^3, n (%)	19 (18)	19 (6)	0.0003
AFP (mg/dl)			
median (range)	10.5 (0 – 71,770)	16 (1 – 2,035,400)	ns
> 400 mg/dl, n (%)	16 (15)	63 (24)	

*Number of patients with available data in Heptromic/TCGA: Etiology, 105/244; Tumor stage: 106/342; Child-Pugh score, 106/227; Tumor size, 106/NA; Vascular invasion, 105/296; Degree of differentiation, 83/340; Bilirubin, 106/285; Albumin, 106/281; Platelet count, 106/286; AFP, 106/262

<u>Supplementary Table 2</u>: Descriptive analysis of broad score (left) and focal score (right) distributions in the TCGA pan-cancer cohort (n = 10,635).

Cancer type	Min.	P25	P50	Mean	P75	Max.
LAML	0	0	0	0.9	1	14
THCA	0	0	0	0.7	1	17
THYM	0	0	0	1.5	1	31
PRAD	0	0	1	2.3	3	20
UCEC	0	0	2	4.9	7	31
DLBC	0	1	3	4.5	6.25	20
LGG	0	2	3	3.8	5	31
PAAD	0	0	3	5.0	8	29
KIRC	0	2	4	5.4	7	31
PCPG	0	2	4	4.9	6	33
MESO	0	3	5	7.3	10	29
STAD	0	1	5	6.5	10	34
UVM	0	3	5.5	6.4	9	21
GBM	0	4	6	6.7	8	30
KIRP	0	3	6	7.4	9	35
BRCA	0	4	7	8.8	12	36
BLCA	0	4	8	8.7	13	25
CESC	0	5	8	9.3	12	33
HNSC	0	4	8	8.3	11	31
LIHC	0	5	8	9.5	13	34
LUAD	0	3	8	8.2	12	28
SKCM	0	5	8	9.8	13.3	32
COAD	0	3	9	8.8	13	32
SARC	0	3	9	9.9	15	34
ESCA	0	7	10	10.3	14	26
CHOL	0	6.75	10.5	10.9	15.3	25
LUSC	0	7	11	12.0	16.5	35
OV	0	8	12	13.0	18	40
READ	0	8.75	12	12.2	16	32
UCS	0	10.8	13.5	14.7	19	35
TGCT	0	7	15	13.4	18	32
ACC	0	9.25	18	18.1	27	42
KICH	0	23.3	27	23.7	29	34

Cancer type	Min.	P25	P50	Mean	P75	Max.
LAML	13	35	46	102.6	63	2614
THCA	6	21	27	31.0	34	438
THYM	15	26.8	34.5	40.0	46	243
PRAD	5	43	68	88.7	103	1101
UCEC	9	31	48	158.7	276	1028
DLBC	27	53	77.5	95.2	104	442
LGG	11	31.3	46	60.6	69	637
PAAD	14	33	58.5	72.8	99.3	398
KIRC	6	32	43	60.9	67.8	432
PCPG	15	34	50	91.2	76	1002
MESO	10	44.5	80	88.9	118	219
STAD	6	50.3	127	184.4	262	1520
UVM	23	36	47.5	58.8	65.3	390
GBM	16	79	113	129.5	154	1012
KIRP	7	30	41	57.9	64.8	505
BRCA	9	63	155	206.1	290	2126
BLCA	16	80.3	155	202.8	274	1728
CESC	19	59.5	95	115.3	140	602
HNSC	11	61	106	128.7	165	543
LIHC	12	55.5	92	120.7	155	785
LUAD	16	67.3	120	158.0	226	1037
SKCM	21	58	102	139.0	196	580
COAD	14	54	102	120.7	155	977
SARC	17	102	248	295.6	416	1249
ESCA	23	144	233	251.5	322	733
CHOL	34	56.5	80.5	96.4	127	245
LUSC	14	118	185	225.1	278	2313
OV	29	268	398	450.1	548	2507
READ	25	73	132	143.7	181	696
UCS	40	230	361	363.6	489	865
TGCT	21	59	87	102.5	130	406
ACC	12	68.3	131	183.0	220	936
KICH	15	35.5	51	62.5	64.8	304

Supplementary Table 3: Excel sheet presenting the INGENUITY® Pathway Analysis of the discovery cohort according to broad scores.

Supplementary Table 4: Excel sheet with the clinico-pathological data displayed according to BS and FS.

<u>Supplementary Table 5</u>: Molecular and immune characterization of the discovery cohort according to broad scores (BS). Gene-based signatures were analyzed either by Nearest Template Prediction[14] or ssGSEA[13] Gene Pattern modules. Values in each raw represent the proportion of samples indicated in the raw category that are classified as low-BS or high-BS. E.g. Out of 23 samples classified as positive for the HCC Immune Class, 39% were low-BS and 9% were high-BS.

Variables		Low-BS (n=24)	High-BS (n=25)	p-value		
				Low vs High	Low vs Rest	High vs Rest
HCC Immune Class[1,+]	/ positive	9 (43%)	2 (8%)	0.02	0.02	ns
	/ negative	12 (57%)	22 (92%)			
Active Immune Subtype[1,+]	/ positive	6 (29%)	0 (0%)	0.007	ns	0.01
	/ negative	15 (71%)	24 (100%)			
Exhausted Immune Subtype[1,+]	/ positive	3 (14%)	2 (8%)	0.65	0.151	1
	/ negative	18 (86%)	22 (92%)			
Proliferation S2[2,+]	/ positive	0 (0%)	6 (25%)	0.02	0.006	ns
	/ negative	21 (100%)	18 (75%)			
Ploidy	Diploids	21 (88%)	8 (32%)	<0.001	0.002	0.002
	Polyploids	3 (12%)	17 (68%)			
TLS count (> 5) *		12 (50%)	6 (24%)	0.067	ns	0.089
DNA repair mechanisms (ssGSEA) [+]	p53 pathway activation			0.01	0.02	ns
	Hallmark DNA repair			0.006	0.009	ns
	Mismatch Repair (MMR)			0.03	0.03	ns
	Nucleotide Excision Repair (NER)			0.047	0.03	ns
	Base-Excision Repair (BER)			0.03	0.03	ns
	Chromosome maintenance			0.01	0.001	ns
	Cell cycle progression			0.047	0.01	ns
Proliferation signatures (ssGSEA) [+]	DNA replication			0.03	0.03	ns
	DNA strand elongation			0.047	0.01	ns
	MTORC1-mediated signaling			0.04	0.006	ns

[1] Sia et al. Gastroenterology 2017
[2] Hoshida Y et al. Cancer Res 2009
* histopathological data
** clinical data
[+] Expression data available for 21 BS-low and 24 BS-high

<u>Supplementary Table 6</u>: Molecular and immune characterization of the validation cohort according to the broad scores (BS). Gene-based signatures were analyzed either by Nearest Template Prediction[14] or ssGSEA[13] both Gene Pattern Modules.

Variables		Low-BS (n=51)	High-BS (n=158)	p-value		
				Low vs High	Low vs Rest	High vs Rest
HCC Immune Class[1]	/ positive	19 (31%)	27 (17%)	0.004	0.02	0.01
	/ negative	32 (69%)	131 (83%)			
Active Immune Subtype[1]	/ positive	8 (16%)	15 (9%)	ns	ns	ns
	/ negative	43 (84%)	143 (91%)			
Exhausted Immune Subtype[1]	/ positive	11 (22%)	12 (8%)	0.01	0.033	0.029
	/ negative	40 (88%)	146 (92%)			
Inflammation Class[2]	/ positive	12 (24%)	6 (4%)	<0.0001	0.01	<0.0001
	/ negative	39 (76%)	152 (96%)			
Proliferation Class[2]	/ positive	7 (14%)	61 (39%)	0.001	0.01	0.0002
	/ negative	44 (86%)	97 (61%)			
Proliferation (S1+S2)[3]	/ positive	15 (29%)	82 (52%)	0.008	0.04	0.004
	/ negative	36 (71%)	76 (48%)			
Ploidy	Diploids	7 (14%)	89 (56%)	<0.0001	<0.0001	<0.0001
	Polyploids	44 (86%)	69 (44%)			
TP53 mutations[+]	*TP53*-mut	4 (8%)	60 (38%)	<0.0001	<0.0001	<0.0001
	TP53-wt	47 (92%)	96 (62%)			
DNA repair mechanisms (ssGSEA)	p53 pathway activation			0.02	ns	0.01
	Hallmark DNA repair			<0.0001	<0.0001	<0.0001
	Mismatch Repair (MMR)			<0.0001	<0.0001	<0.0001
	Nucleotide Excision Repair (NER)			<0.0001	0.0001	<0.0001
	Base-Excision Repair (BER)			<0.0001	<0.0001	<0.0001
	Homologous Recombination (HR)			<0.0001	<0.0001	<0.0001
	Non-Homologous End-Joining (NHEJ)			<0.0001	0.0001	<0.0001
Proliferation signatures (ssGSEA)	DNA replication			<0.0001	<0.0001	<0.0001
	Cell cycle progression			<0.0001	<0.0001	<0.0001
	MTORC1-mediated signaling			0.002	0.02	0.002

[1] Sia et al. Gastroenterology 2017
[2] Chiang et al., Cancer Res 2008
[3] Hoshida et al. Cancer Res 2009
[+] 156 Samples available with TP53 mutations in BS-high

<u>Supplementary Table 7</u>: Molecular and immune characterization of the discovery cohort according to focal scores (FS). Gene-based signatures were analyzed either by Nearest Template Prediction (NTP) [14] or ssGSEA[13] Gene Pattern modules.

Variables		High-FS (n=27)	FS Rest (n=80)	p-value
HCC Immune Class [1,+]	/ positive	7 (27%)	16 (21%)	ns
	/ negative	19 (73%)	60 (79%)	
Proliferation Class [2,+]	/ positive	13 (50%)	11 (14%)	0.001
	/ negative	13 (50%)	57 (86%)	
CTNNB1 Class [2,+]	/ positive	2 (8%)	23 (30%)	0.03
	/ negative	24 (32%)	53 (70%)	
Late-TGFβ signalling [3,+]	/ positive	13 (46%)	15 (54%)	0.005
	/ negative	13 (18%)	61 (82%)	
Vascular invasion signature [4,+]	/ positive	13 (50%)	18 (24%)	0.02
	/ negative	13 (50%)	58 (76%)	
Poor survival signature [5,+]	/ positive	13 (50%)	14 (18%)	0.004
	/ negative	13 (50%)	62 (82%)	
Poor prognosis MET-related signature [6,+]	/ positive	6 (23%)	5 (7%)	0.03
	/ negative	20 (77%)	71 (93%)	
Proliferation linked to RB1 LOF [7,+]	/ positive	12 (46%)	12 (16%)	0.003
	/ negative	14 (54%)	64 (84%)	
Progenitor cell CK19 signature [8,+]	/ positive	13 (50%)	17 (22%)	0.01
	/ negative	13 (50%)	59 (78%)	
Fibrosis [++]	F4	6 (16%)	31 (84%)	0.045
	F0-F3	8 (44%)	10 (56%)	
Poor differentiation*	/ positive	11 (41%)	9 (11%)	0.0003
	/ negative	16 (59%)	71 (89%)	

[1] Sia et al. Gastroenterology 2017
[2] Chiang et al. Cancer Res 2008
[3] Coulouarn et al. Hepatology 2008
[4] Mínguez et al. J Hepatology 2011
[5] Lee et al. Hepatology 2004
[6] Kaposi-Novak et al. J Clin Invest 2006
[7] Bollard et al. Gut 2017
[8] Villanueva et al. Gastroenterology 2008
* Histopathological assessment
[+] Expression data available for 102 tumors
[++]Data available for 55 patients

<u>Supplementary Table 8</u>: Molecular and immune characterization of the validation cohort according to focal scores (FS). Gene-based signatures were analyzed either by NTP[14] or ssGSEA[13] Gene Pattern modules.

Variables		High-FS (n=169)	FS Rest (n=176)	p-value
HCC immune class [1,+]	/ positive	34/168 (20%)	46/174 (26%)	ns
	/ negative	134/168 (80%)	128/174 (74%)	
Ploidy	Polyploid	89/169 (53%)	52/176 (30%)	<0.0001
	Diploid	80/169 (47%)	124/176 (70%)	
TP53 mutation [+]	*TP53*-mut	64/167 (38%)	33/174 (19%)	0.0001
	TP53-wt	103/167 (62%)	141/174 (81%)	
CTNNB1 mutation [+]	*CTNNB1*-mut	27/167 (16%)	67/174 (39%)	<0.0001
	CTNNB1-wt	140/167 (84%)	107/174 (61%)	
Proliferation class [2,+]	/ positive	64/168 (38%)	34/174 (20%)	0.0002
	/ negative	104/168 (62%)	140/174 (80%)	
CTNNB1 class [2,+]	/ positive	16/168 (10%)	47/174 (27%)	<0.0001
	/ negative	152/168 (90%)	127/174 (73%)	
Poor survival [3,+]	/ positive	57/168 (34%)	36/174 (21%)	0.009
	/ negative	111/168 (66%)	138/174 (79%)	
Poor prognosis Met-related signature [4,+]	/ positive	50/168 (30%)	34/174 (20%)	0.03
	/ negative	118/168 (70%)	140/174 (80%)	
Proliferation linked to RB1 LOF [5,+]	/ positive	69/168 (41%)	34/174 (20%)	<0.0001
	/ negative	99/168 (59%)	140/174 (80%)	
Progenitor cell CK19 poor prognosis signature [6,+]	/ positive	82/168 (49%)	54/174 (31%)	0.001
	/ negative	86/168 (51%)	120/174 (69%)	
Poor differentiation *[,+]	/ positive	74/167 (44%)	55/173 (32%)	0.02
	/ negative	93/167 (56%)	118/173 (68%)	
DNA damage repair signatures (ssGSEA) [+]	Hallmark DNA repair			0.04
	Mismatch Repair (MMR)			<0.0001
	Nucleotide Excision Repair (NER)			<0.0001
	Base-Excision Repair (BER)			<0.0001
	Homologous Recombination (HR)			<0.0001
Proliferation and β-catenin signatures (ssGSEA) [+]	Cell cycle progression			<0.0001
	DNA replication			<0.0001
	Hallmark Wnt-β-catenin activity			0.002

[1] Sia et al. Gastroenterology 2017
[2] Chiang et al. Cancer Res. 2008
[3] Lee et al. Hepatology 2004
[4] Kaposi-Novak et al. J Clin Invest 2006
[5] Bollard et al. Gut 2017
[6] Villanueva et al. Gastroenterology 2008
* Histopathological assessment
[+] Expression data available for 342 tumors, mutational data available for 341, and histopathological data available for 340.

<u>Supplementary Table 9:</u> Description of the immune clusters. Gene groups representing immune clusters as major determinants of immunogenicity, according to [15] and [16].

Immune clusters		Genes
MHC		B2M, TAP1, TAP2, HLA-A, HLA-B, HLA-C, HLA-DPA1, HLA-DPB1, HLA-E, HLA-F
CP / IM		PDCD1, CTLA4, LAG3, TIGIT, HAVCR2, CD274, PDCD1LG2, CD27, ICOS, IDO1
Effector Cells (ECs)	**Act $CD4^+$**	AIM2, BIRC3, BRIP1, CCL20, CCL4, CCL5, CCNB1, CCR7, DUSP2, ESCO2, ETS1, EXO1, EXOC6, IARS, KIF11, KNTC1, NUF2, PRC1, PSAT1, RGS1, RTKN2, SAMSN1, SELL, TRAT1
	Act $CD8^+$	ADRM1, AHSA1, C1GALT1C1, CCT6B, CD37, CD3D, CD3E, CD3G, CD69, CD8A, CETN3, CSE1L, GEMIN6, GNLY, GPT2, GZMA, GZMH, GZMK, IL2RB, LCK, MPZL1, NKG7, PIK3IP1, PTRH2, TIMM13, ZAP70
	T_{em} $CD4^+$	ATM, CASP3, CASQ1, CD300E, DARS, DOCK9, EXOSC9, EZH2, GDE1, IL34, NCOA4, NEFL, PDGFRL, PTGS1, REPS1, SCG2, SDPR, SIGLEC14, SIGLEC6, TAL1, TFEC, TIPIN, TPK1, UQCRB, USP9Y, WIPF1, ZCRB1
	T_{em} $CD8^+$	ACAP1, APOL3, ARHGAP10, ATP10D, C3AR1, CCR5, CD160, CD55, CFLAR, CMKLR1, DAPP1, FCRL6, FLT3LG, GZMM, HAPLN3, HLA_DMB, HLA_DPA1.1, HLA_DPB1.1, IFI16, LIME1, LTK, NFKBIA, SETD7, SIK1, TRIB2
Suppressor cells (SCs)	**MDSC**	CCR2, CD14, CD2, CD86, CXCR4, FCGR2A, FCGR2B, FCGR3A, FERMT3, GPSM3, IL18BP, IL4R, ITGAL, ITGAM, PARVG, PSAP, PTGER2, PTGES2, S100A8, S100A9
	T-reg	CCL3L1, CD72, CLEC5A, FOXP3, ITGA4, L1CAM, LIPA, LRP1, LRRC42, MARCO, MMP12, MNDA, MRC1, MS4A6A, PELO, PLEK, PRSS23, PTGIR, ST8SIA4, STAB1

Notes: List of genes adapted from [15] and [16]. MHC: Major Histocompatibility Complex; CP/IM: checkpoints / immune-modulators; MDSC: Myeloid-derived suppressive cells; T-reg: Regulatory T cells; Tem: Effector Memory T cells; Act: Activated T cells.

<u>Supplementary Table 10</u>: Publicly-available gene signatures used in the study. Signatures not specifically listed in the table were obtained from the Gene Set Enrichment Analysis (GSEA) Molecular Signatures Database (MSigDB).

Name used to refer to the gene signature	Reference
Proliferation linked to RB1 LOF	Bollard et al. Gut 2017 [18]
Proliferation Class	Chiang et al. Cancer Res 2008 [19]
CTNNB1 Class	
Inflammation Class	
Late-TGFβ signaling	Coulouarn et al. Hepatology 2008 [20]
TGF-β signaling	
HCC proliferation subclass (S1+S2)	Hoshida Y et al. Cancer Res 2009 [21]
Proliferation S2 subclass / Proliferation S2	
HCC S3 non-proliferation subclass	
Poor prognosis MET-related signaling	Kaposi-Novak et al. J Clin Invest 2006 [22]
Poor survival signature	Lee et al. Hepatology 2004 [23]
Vascular invasion signature	Mínguez et al. J Hepatology 2011 [24]
HCC Immune Class	Sia et al. Gastroenterology 2017 [25]
Active Immune Subtype	
Exhausted Immune Subtype	
Progenitor cell CK19 signature	Villanueva et al. Gastroenterology 2008 [26]
Notch signaling	Villanueva et al. Gastroenterology 2012 [27]
EPCAM signaling	Yamashita et al. Cancer Res. 2008 [28]

References

1. Villanueva A, Portela A, Sayols S, Battiston C, Hoshida Y, Méndez-González J, et al. DNA methylation-based prognosis and epidrivers in hepatocellular carcinoma. Hepatology 2015;61:1945–1956.
2. Ally A, Balasundaram M, Carlsen R, Chuah E, Clarke A, Dhalla N, et al. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell 2017;169:1327-1341.e23.
3. Torrecilla S, Sia D, Harrington AN, Zhang Z, Cabellos L, Cornella H, et al. Trunk mutational events present minimal intra- and inter-tumoral heterogeneity in hepatocellular carcinoma. J Hepatol 2017;67:1222–1231.
4. Prat A, Navarro A, Paré L, Reguart N, Galván P, Pascual T, et al. Immune-related gene expression profiling after PD-1 blockade in non–small cell lung carcinoma, head and neck squamous cell carcinoma, and melanoma. Cancer Res 2017;77:3540–3550.
5. 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.
6. Van Loo P, Nordgard SH, Lingjærde OC, Russnes HG, Rye IH, Sun W, et al. Allele-specific copy number analysis of tumors. Proc Natl Acad Sci 2010;107:16910–16915.
7. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res 2018;47:D941–D947.
8. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. Nat Commun 2015;6:8971.

9. Franch-Expósito S, Bassaganyas L, Vila-Casadesús M, Hernández-Illán E, Esteban-Fabró R, Díaz-Gay M, et al. CNApp, a tool for the quantification of copy number alterations and integrative analysis revealing clinical implications. Elife 2020; 9:e50267.
10. Davoli T, Uno H, Wooten EC, Elledge SJ. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. Science 2017;355:eaaf8399.
11. Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, et al. Genomic and Functional Approaches to Understanding Cancer Aneuploidy. Cancer Cell 2018;33:676-689.e3.
12. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–50.
13. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature 2009;462:108–12.
14. Hoshida Y, Villanueva A, Kobayashi M, Peix J, Chiang DY, Camargo A, et al. Gene expression in fixed tissues and outcome in hepatocellular carcinoma. N Engl J Med 2008;359:1995–2004.
15. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep 2017;18:248–262.

16. Rooney MS, Shukla SA, Wu CJJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell 2015;160:48–61.
17. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 2013;4:2612.
18. Bollard J, Miguela V, Ruiz de Galarreta M, Venkatesh A, Bian CB, Roberto MP, et al. No Title. Gut 2017;66:1286–1296.
19. Chiang DY, Villanueva A, Hoshida Y, Peix J, Newell P, Minguez B, et al. Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. Cancer Res 2008;68:6779–88.
20. Coulouarn C, Factor VM, Thorgeirsson SS. Transforming growth factor-beta gene expression signature in mouse hepatocytes predicts clinical outcome in human cancer. Hepatology 2008;47:2059–2067.
21. Hoshida Y, Nijman SMBS, Kobayashi M, Chan J a, Brunet J-P, Chiang DY, et al. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. Cancer Res 2009;69:7385–7392.
22. Kaposi-Novak P, Lee J-S, Gòmez-Quiroz L, Coulouarn C, Factor VM, Thorgeirsson SS. Met-regulated expression signature defines a subset of human hepatocellular carcinomas with poor prognosis and aggressive phenotype. J Clin Invest 2006;116:1582–95.
23. Lee J-SS, Chu I-SS, Heo J, Calvisi DF, Sun Z, Roskams T, et al. Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. Hepatology 2004;40:667–676.

24. Mínguez B, Hoshida Y, Villanueva A, Toffanin S, Cabellos L, Thung S, et al. Gene-expression signature of vascular invasion in hepatocellular carcinoma. J Hepatol 2011;55:1325–1331.
25. Sia D, Jiao Y, Martinez-Quetglas I, Kuchuk O, Villacorta-Martin C, Castro de Moura M, et al. Identification of an Immune-specific Class of Hepatocellular Carcinoma, Based on Molecular Features. Gastroenterology 2017;153:812–826.
26. Villanueva A, Chiang DY, Newell P, Peix J, Thung S, Alsinet C, et al. Pivotal role of mTOR signaling in hepatocellular carcinoma. Gastroenterology 2008;135:1972–83, 1983 e1–11.
27. Villanueva A, Alsinet C, Yanger K, Hoshida Y, Zong Y, Toffanin S, et al. Notch signaling is activated in human hepatocellular carcinoma and induces tumor formation in mice. Gastroenterology 2012;143:1660-1669.e7.
28. Yamashita T, Forgues M, Wang W, Kim JW, Ye Q, Jia H, et al. EpCAM and alpha-Fetoprotein Expression Defines Novel Prognostic Subtypes of Hepatocellular Carcinoma. Cancer Res 2008;68:1451–1461.