Identification of an Immune-specific Class of Hepatocellular Carcinoma, Based on Molecular Features

Running title: Characterization of the immunological landscape of HCC

Daniela Sia[1], Yang Jiao[1], Iris Martinez-Quetglas[2], Olga Kuchuk[1,3], Carlos Villacorta-Martin[1], Manuel Castro de Moura[4], Juan Putra[1], Genis Camprecios[1], Laia Bassaganyas[2], Nicholas Akers[1,5], Bojan Losic[1,5], Samuel Waxman[1], Swan N. Thung[1], Vincenzo Mazzaferro[3], Manel Esteller[4,6,7], Scott L. Friedman[1], Myron Schwartz[1], Augusto Villanueva[1], Josep M. Llovet[1,2,7]

Affiliations:

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

[2] Liver Cancer Translational Research Laboratory, BCLC, Liver Unit, CIBEREHD, IDIBAPS, Hospital Clinic, University of Barcelona, Catalonia, Spain.

[3] University of Milan and Gastrointestinal Surgery and Liver Transplantation Unit, Fondazione IRCCS, Istituto Nazionale dei Tumori, Milan, Italy

[4] Cancer Epigenetics and Biology Program, IDIBELL, Hospital Universitari Bellvitge, Barcelona, Catalonia, Spain.

[5] Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, USA

[6] Department of Physiological Sciences, School of Medicine and Health Sciences, University of Barcelona, Barcelona, Catalonia, Spain.

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

Grant Support

This project has received funding from the Tisch Cancer Institute at Mount Sinai (P30 CA196521 – Cancer Center Support Grant). Josep M. Llovet is supported by grants from the U.S. Department of Defense (CA150272P3), European Commission Framework Program 7 (HEPTROMIC, proposal number 259744) and Horizon 2020 Program (HEPCAR, proposal number 667273-2), the Asociación Española Contra el Cáncer (AECC), Samuel Waxman Cancer Research Foundation, Spanish National Health Institute (SAF2013-41027) and Grup de Recerca Consolidat –- Recerca Translacional en Oncologia Hepàtica. AGAUR (Generalitat de Catalunya), SGR 1162. Iris Martinez-Quetglas is supported by the European Commission HEPCAR grant.Olga Kuchuk is supported by Onlus Prometeo, Hepato-Oncology Research Project, Istituto Nazionale Tumori (National Cancer Institute) IRCCS Foundation (Milan, Italy). Laia Bassaganyas is supported by the Juan de la Cierva Fellowship.Vincenzo Mazzaferro is supported by grants from AIRC and the Oncology Research Project of the Italian Ministry of Health. Scott L. Friedman is supported by the U.S. Department of Defense (CA150272P3) and National Institutes of Health (R01DK56621). Augusto Villanueva is supported by the U.S. Department of Defense (CA150272P3), The Tisch Cancer Institute, and the American Association for the Study of Liver Diseases Foundation (AASLDF) Alan Hofmann Clinical and Translational Award

Abbreviations

PD-1: programmed death 1

PD-L1: programmed death-ligand 1

TGF-β: transforming growth factor beta

FDA: Food and Drug Administration

CTLA-4: cytotoxic T-Lymphocyte Associated Protein 4

IFN: interferon

CTNNB1: Catenin Beta 1

GEO: Gene Expression Omnibus

NMF: non-negative matrix factorization

TLS: tertiary lymphoid structure

FDR: false discovery rate

NK: natural killer

CCL: Chemokine (C-C motif) ligand

CXCL: chemokine (C-X-C motif) ligand

JAK/STAT: Janus Kinase/Signal Transducer and Activator of Transcription

GSEA: gene set enrichment analysis

VS: versus

EMT: epithelial mesenchymal transition

NTP: nearest template prediction

TBRS: TGF-beta response signature

F-TBRS: Fibroblasts-derived TGF-beta response signature

T-TBRS: T cells-derived TGF-beta response signature

LGALS: lectin, galactose binding, soluble 1

NKG2D: natural killer group 2D

TBX: T-box transcription factor

FF: fresh frozen

FFPE: formalin fixed paraffin embedded

PMEPA1: prostate transmembrane protein, androgen induced 1

PTK2: Protein Tyrosine Kinase 2

WG: whole genome

qRT-PCR: quantitative real time polymerase chain reaction

AFP: Alpha Fetoprotein

RF: random forest

SCNA: somatic copy number aberrations

Address for correspondence:

Josep M. Llovet, MD

Professor of Medicine

Director, Mount Sinai Liver Cancer Program

Division of Liver Diseases

Icahn School of Medicine at Mount Sinai

1425 Madison Avenue, Box 11-23

New York 10029, NY

Phone: 1-2126599503

Fax: 212-849-2574

Email: Josep.Llovet@mssm.edu

Disclosures

The authors declare no potential conflicts of interest.

Author Contributions

D.S. and J.M.L. designed and directed the research; Y.Y., I.M., O.K. and G.C. performed the experiments. J.P. and S.T. performed immunohistochemical analysis. D.S, A.V., C.V.M, M.C.M., L.B., N.A., B.L. and M.E. conducted the in silico analyses. V.M., M.S., S.T., S.W., A.V., S.L.F. and J.M.L. collected the tissues and clinical data. D.S, A. V. and J.M.L interpreted the results and wrote the manuscript. All authors reviewed and approved the manuscript.

GEO accession number: GSE93647, and previously deposited data from our group (GSE63898, GSE20140) and others [GPL1528, GSE1898, GPL2094 a), GPL80b), GPL96 E-TABM-36, GPL5474, GSE10186].

ABSTRACT

BACKGROUND AND AIMS: Agents that induce an immune response against tumors by altering T-cell regulation have increased survival times of patients with advanced-stage tumors, such as melanoma or lung cancer. We aimed to characterize molecular features of immune cells that infiltrate hepatocellular carcinomas (HCCs) to determine whether these types of agents might be effective against liver tumors.

METHODS: We analyzed HCC samples from 956 patients. We separated gene expression profiles from tumor, stromal, and immune cells using a non-negative matrix factorization algorithm. We then analyzed the gene expression pattern of inflammatory cells in HCC tumors samples. We correlated expression patterns with the presence of immune cell infiltrates and immune regulatory molecules, determined by pathology and immunohistochemical analyses, in a training set of 228 HCC samples. We validated the correlation in a validation set of 728 tumor samples. Using data from 190 tumors in the Cancer Genome Atlas, we correlated immune cell gene expression profiles with numbers of chromosomal aberrations (based on single-nucleotide polymorphism array) and mutations (exome sequence data).

RESULTS: We found approximately 25% of HCCs to have markers of an inflammatory response, with high expression levels of the CD274 molecule (PD-L1) and programmed cell death 1 (PD-1), markers of cytolytic activity, and fewer chromosomal aberrations. We called this group of tumors the Immune class. It contained 2 subtypes, characterized by markers of an adaptive T-cell response or exhausted immune response. The exhausted immune response subclass expressed many genes regulated by transforming growth factor beta 1 (TGFB) that mediate immunosuppression. We did not observe any differences in numbers of mutations or expression of tumor antigens between the immune-specific class and other HCCs.

CONCLUSIONS: In an analysis of HCC samples from 956 patients, we found almost 25% to express markers of an inflammatory response. We identified 2 subclasses, characterized by adaptive or exhausted immune responses. These findings indicate that some HCCs might be

susceptible to therapeutic agents designed to block the regulatory pathways in T cells, such as PD-L1, PD-1, or TGFB inhibitors.

SHORT SUMMARY: The study defines the ~25% population with molecular characteristics - including high infiltration of immune cells, expression of PD-1 and CD274 molecule (PD-L1), and active IFN-γ signaling - that highly resemble those of cancers most responsive to immunotherapy.

KEYWORDS: immune checkpoint; virtual microdissection; molecular subgroups; immune regulation

INTRODUCTION

Hepatocellular carcinoma (HCC) is the second leading cause of cancer-related mortality worldwide. The number of HCC deaths (approximately 800,000 per year) overlap with that of new cases, a testament to its high lethality[1, 2]. This malignancy often occurs in the setting of chronic inflammatory liver disease (e.g., cirrhosis) and is associated with well-defined risk factors such as hepatitis B virus (HBV), hepatitis C virus (HCV), alcohol abuse, metabolic syndrome and diabetes[2]. Over the past decade, major advancements have elucidated the molecular pathogenesis of HCC[2, 3], and yet, current therapeutic options remain very limited. Only a minority of HCC patients are diagnosed at early stages when curative approaches, such as surgical resection, transplantation or local ablation, are effective[2]. In patients at advanced stages, the only systemic therapies that increase survival are the multi-tyrosine kinase inhibitors sorafenib (first line)[4] and regorafenib (second line)[5]. Nonetheless, even with the survival benefits provided by these agents, the median life expectancy is of less than 2 years. Therefore, there is a clear need to expand the therapeutic arsenal for advanced HCC.

In recent years, immune checkpoint inhibitors, which unleash the body's own immune response to attack tumors by targeting regulatory pathways in T cells, have shown remarkable efficacy in different solid cancers; this has led to the Food and Drug Administration (FDA) approval of 4 immune-based compounds for the treatment of advanced stage malignancies such as melanoma or lung cancer (i.e., ipilimumab, nivolumab, pembrolizumab, and atezolimumab). These agents include monoclonal antibodies directed against the cytotoxic T-lymphocyte protein 4 (CTLA-4), the programmed cell death protein 1 (PD-1) and its ligand PD-L1[6]. Intriguingly, not all patients have the same likelihood of responding to these regimens[7]. High expression of PD-L1 is currently under investigation as a potential predictor of response to anti-PD1 therapy[8-10]. Emerging experimental data indicate that the presence of a pre-existing intra-tumoral T cell infiltration, interferon (IFN) signaling, checkpoint molecules (PD-1, PD-L1 expression) or high

tumor mutational burden could favor a clinical response[11-13]. Conversely, tumor-intrinsic active ß-catenin (CTNNB1) signaling may result in T cell exclusion and resistance to anti-PD-L1 and anti-CTLA4 antibodies[14]. In HCC, promising responses have been recently reported with nivolumab, a monoclonal antibody directed against PD-1, in a phase I/II trial[15]. Unfortunately, little is known about the immunological profile of HCC tumors and how to leverage this information to maximize response to immune-based therapies.

HCCs comprise a mixture of cell types, including malignant hepatocytes, immune cells and endothelial cells dispersed within the extracellular matrix and supporting stroma. Previous studies have established a set of analytical approaches to virtually dissect the molecular signals deriving from these distinct compartments[16, 17]. Using non-negative matrix factorization (NMF), we have deconvoluted the gene expression data of 956 human HCC samples and isolated the signal released from the inflammatory infiltrates to characterize the immunological landscape of HCC. This has allowed us to identify an immune-specific class of HCC with specific biological traits. Key features of this class include actual presence and activation of immune cells, enhanced cytolytic activity, protein expression of PD-1 and PD-L1, and enrichment of gene signatures predictive of response to immunotherapies. Further dissection of this class has revealed two robust microenvironment-based types with either active or exhausted immune activity.These findings provide a comprehensive understanding of the immunological milieu of HCC and deserve further investigation in HCC patients treated with immunotherapy.

MATERIALS AND METHODS

Patients and samples

For the purpose of the study, gene expression profile from a total of 956 HCC human samples was analyzed (*Flow chart*, **Figure 1**), including a training cohort of 228 surgically resected fresh frozen (FF) samples (Heptromic dataset, GSE63898). All samples of the training set were previously obtained from two institutions of the HCC Genomic Consortium upon IRB approval: IRCCS Istituto Nazionale Tumori (Milan, Italy) and Hospital Clínic (Barcelona, Spain). RNA profiling and methylation data were available for all 228 HCC samples and 168 non-tumor liver adjacent cirrhotic tissues and are published elsewhere[18]. Additional 728 HCC samples of patients with mixed etiology from 7 independent datasets were used for external validation (**Figure 1, Supplementary Table 1**).

Statistical analysis

All analyses were performed using SPSS software version 22. Correlations between molecular classes, histological markers and clinico-pathological variables were analyzed by Fisher's exact test and Wilcoxon rank-sum test for categorical and continuous data, respectively. All signatures used in the study were previously reported (**Supplementary Table 2**).

Additional detailed protocols are provided in the **Supplementary Materials and Methods**.

RESULTS

A novel Immune class of HCC

In order to isolate immune-related genomic signals from bulk gene expression data in HCC tumors, we performed NMF analysis of 228 resected HCC samples (training cohort, **Figure 1**). Clinical characteristics of the training cohort are summarized in **Table 1**. Among the distinct expression patterns identified by NMF, one was attributed to the presence of inflammatory response and immune cells through integration with a previously reported immune enrichment score (**Supplementary Figure 1A)**. Analysis of the top-ranked genes (named exemplar genes) that defined this expression pattern further confirmed immune-related functions and signaling (**Supplementary Figure 1B**). Consensus clustering on exemplar genes (**Supplementary Figure 2)** identified a new molecular subgroup accounting for 24% of the cohort (55/228), referred herein as 'Immune class' (**Figure 2A**). Patients belonging to the Immune class showed significant enrichment of signatures identifying immune cells [i.e. T cells, cytotox, tertiary lymphoid structures (TLS), and macrophages ($p<0.001$)], immune metagenes, IFN gene signatures predictive of response to pembrolizumab in melanoma (28-genes, $p<0.001$) and head and neck squamous cell carcinoma (6-genes , $p<0.001$), and PD-1 signaling (36/55 vs 19/173, $p<0.001$) (**Figure 2A**). Class comparison between the Immune class and remaining samples identified 112 genes significantly deregulated (*Immune Classifier*)*,* including 108 over-expressed immune-related genes such as T cell receptor components and chemo-attractants for Natural Killer (NK) and T cells (*CCL5*, *CXCL9* and *CXCL10*, $p<0.001$, **Supplementary Table 3**). Similarly, GSEA identified enrichment of IFN alfa and gamma signaling, inflammatory response (i.e. lymphocyte activation, T helper 1- cytotoxic module, NK-mediated toxicity, etc.), TGF-ß and JAK/STAT signaling (FDR<0.001, **Supplementary Figure 3** and **Supplementary Table 4**).

We next sought to integrate the Immune class with previously reported HCC molecular classifications. This revealed an enrichment of the IFN-related (18/55 vs 12/173, $p=0.0001$) and

S1 classes (TGF-β/WNT activation) (32/55 vs 15/173, p=0.0001), as well as a significant exclusion of S2 (2/55 vs 46/173, p=0.0001) and CTNNB1 classes (8/55 vs 59/173, p<0.001, **Figure 2A**). All together, these data suggest that we successfully identified an immune-related class of HCC enriched with signatures capturing the presence of immune cells, signatures of response to immune checkpoint therapy and IFN signaling.

Immune class immunophenotype shows enrichment of PD-1/PD-L1 signaling

We performed immunophenotyping to gain further biological insight into the immunological nature of the Immune class. As predicted, patients belonging to this class had significantly higher rates of immune cell infiltration (11/49 vs 14/167, p=0.01, **Figure 2A-B**) and density of TLS (≥ 5 foci, 19/51 vs 34/170, p=0.01, **Figure 2A and Supplementary Figure 4A**) as revealed by the examination of hematoxylin and eosin-stained sections. We then assessed PD-1 and PD-L1 protein expression by immunohistochemistry in a subset of samples of the training cohort (48 within the Immune class and 51 outside, **Figure 2B** and **Supplementary Figure 4B**). Overall, PD-L1 tumoral expression was observed in 16% (16/99) of HCC in accordance with recent reports[19]. PD-1 protein expression was observed in 10% of the cohort (10/99), but no significant correlation was found between high PD-1 and PD-L1 expression, likely due to the small sample size. Nonetheless, tumors with high PD-1 (8/48 within the Immune class vs 2/51 in the rest, p=0.04) and PD-L1 (12/48 within in the Immune class vs 4/51 in the rest, p=0.03) protein expression were significantly enriched in the Immune class. No difference was observed between the Immune class and the rest of the cohort in terms of other clinico-pathological variables (data not shown, p>0.05). In summary, pathological examination revealed that patients belonging to the Immune class showed a high degree of immune infiltration, higher immunohistochemical expression of PD-1/PD-L1, and presence of TLS. These data underscore the performance of the Immune Classifier to capture molecular signals deriving from infiltrating immune cells in HCC.

The Immune class captures two distinct components of the tumor microenvironment: active and exhausted subtypes

The immune system can exert both anti- and pro-tumor activities. Indeed, cross-talk between cancer cells and the tumor microenvironment triggers immune responses which favor cancer progression by supplying growth factors that sustain proliferation and facilitate epithelial-mesenchymal transition (EMT), invasion, and metastasis[20]. To further explore this concept in HCC, we analyzed the type of immune modulation occurring in response to the tumor microenvironment in patients within the Immune class. As depicted in **Figure 3**, 33% of the Immune class (18/55) was characterized by *"activated stroma*" whereas the remaining patients (37/55, 67%) showed lack of such activation, as predicted by nearest template prediction (NTP) analysis using a previously published molecular signature that captures activated inflammatory stromal response. Interestingly, patients with normal or non-active stroma (37/55, 67%) showed significant enrichment of T cells and IFN signatures, including overexpression of adaptive immune response genes (i.e. *T Cells receptor G*, *CD8A*, *IFN-Y, GZMB, etc.*) and IFN signatures predictive of response to pembrolizumab ($p<0.001$). Thus, we named this cluster *Active Immune Response*. Conversely, the presence of activated stroma was significantly associated with a T cell exhaustion signature (10/18 vs 4/37, $p<0.001$), and with immunosuppressive components, such as TGF-ß signaling and M2 macrophages (8/18 vs 1/37, $p=0.0003$). In particular, over-expression of *TGF-ß -1* and *-3* along with enrichment of several signatures reflecting activation of TGF-ß pathway, such as late TGF-ß signature (9/18 vs 6/37, $p=0.02$), S1/TGF-ß signature (16/18 vs 16/37, $p=0.001$), WNT/TGF-ß signaling (15/18 vs 12/37, $p<0.001$), and TGF-beta response signatures (TBRS) of Fibroblasts (F-TBRS) (9/18 vs 6/37, $p=0.02$) and T-Cells (T-TBRS) (10/18 vs 9/37, $p=0.03$), were observed in this subgroup (**Figure 3**). T cell exhaustion and impaired cytotoxic activity in this cluster was supported by the up-regulation of immunosuppressive factors (i.e. *LGALS1, CXCL12*) and myeloid chemo-attractants (*CCL2*). Other essential NK cell activators such as *Granzyme B* (*GZMB*) *IFN-Y*, *NKG2D* and *TBX21*

receptors[21, 22], were strongly down-regulated (**Figure 3**). Based on these features, we named this cluster *Exhausted Immune Response.* GSEA analysis comparing both clusters confirmed the driver role of TGF-ß in the *Exhausted Immune Response,* and enrichment of pathways related to metastasis, EMT, angiogenesis and liver cancer recurrence, suggesting a more aggressive phenotype (**Supplementary Table 5**). Interestingly, we did not observe any significant difference between the Active and Exhausted Immune subtype in terms of immune infiltration, TLS count, PD-L1 and PD-1 expression (**Supplementary Figure 4B-C**).

We further explored the potential prognostic implications of the type of immune response by correlating these clusters with clinico-pathological parameters. Interestingly, patients within the *Active Immune Response cluster* showed lower rates of tumor recurrence after resection compared to the *Exhausted Immune Response cluster* (median time to recurrence 32 *versus* 21 months, p=0.04, **Supplementary Figure 5A-B**); we also observed a trend towards better survival (median survival time of 88 months in the Active Immune vs 63 months in remaining patients, p=0.07) (**Figure 4A, Supplementary Figure 5C**). No differences in other clinico-pathological variables, including HBV and HCV infection, were found between the distinct Immune subtypes (**Supplementary Table 6**). Notably, the Active Immune subtype was retained as independent prognostic factor of overall survival (HR=0.58, CI 0.34-0.98, p=0.04, **Supplementary Table 7**) along with vascular invasion, multinodularity, platelets count, and HCV infection.

Altogether, these data divide the Immune class in two distinct microenvironment-based components: a) *Active Immune Response Subtype* (~65%) characterized by overexpression of adaptive immune response genes **(Figure 3**), and b) *Exhausted Immune Response Subtype* (~35%) characterized by the presence of immunosuppressive signals (i.e. TGF-ß, M2 macrophages).

The Immune class is validated across datasets

The presence of the Immune class was further evaluated in 7 additional datasets (n= 728 HCCs, **Figure 1**) using the 112 gene-expression based Immune Classifier (**Supplementary Table 8**). Firstly, we applied the Immune Classifier to the TCGA dataset, the largest dataset publicly available [n=190 fresh frozen (FF) samples] profiled by RNA-sequencing]. Similar to our training cohort, 42/190 (22%) HCC samples were successfully predicted within the Immune class. Molecular characterization of the Immune class confirmed a significant enrichment of signatures identifying immune cells (i.e. T cells, cytotox, TLS and macrophages, $p<0.001$), signatures predictive of response to immune checkpoint therapy ($p<0.001$) and PD-1 signaling (24/42 vs 31/148, $p<0.001$) (**Figure 4B**). Compared to known HCC molecular classes, we confirmed the enrichment of the IFN-related (13/42 vs 11/148 in the rest, $p<0.001$) and S1 classes (28/42 vs 20/148 in the rest of cohort, $p<0.001$) and the significant exclusion of the CTNNB1 class (2/42 vs 30/148 in the rest of the cohort, $p<0.001$) as previously observed in the training cohort. In addition, half of the TCGA-Immune class showed lack of the activated stroma signature along with over-expression of adaptive immune response genes, recapitulating the *Active Immune Response Subtype* (**Figure 4B**). On the other end, the remaining half of patients showed activated stroma which was associated with TGF-ß signaling (11/21 vs 1/21 in the rest of the Immune class, $p=0.01$) and down-regulation of *NKG2D* and *TBX21* receptors ($p<0.01$), main characteristics of the *Exhausted Immune Response subtype*. Correlation with clinical outcomes confirmed that patients within the Active Immune Response subtype had a better survival (median survival time of 107 months in the Active Immune cluster vs 33 months in the remaining patients, $p=0.03$) (**Figure 4C, Supplementary Figure 5D**).

We next interrogated the Validation cohort previously collected by our group [n=131 formalin-fixed paraffin embedded (FFPE) HCCs] and 5 additional datasets including 4 testing FF tissues (n=289) and 1 of FFPE samples (n=118) (**Figure 1, Supplementary Table 1**). The percentage of patients allocated to the Immune class was consistent across all FF datasets with an average

of 27% of the samples predicted to this class (range 22-28). In the two FFPE datasets (Validation and HCC-V), 37% (48/131) and 30% (35/118) of patients were allocated to the Immune class, respectively (**Supplementary Figures 6 and 7**). The higher percentage could be due to the different genomic platform used [DASL (Illumina) versus Affymetrix] or a different type of tissue material (FFPE versus FF samples). Nonetheless, molecular characteristics of the Immune class and the presence of the two microenvironment-based subtypes were successfully recapitulated in all datasets tested regardless of the platform and type of samples used.

Finally, we tested the capacity of the Immune class to predict response to immunotherapy. The tumoral gene expression derived from two HCC patients treated with nivolumab was analyzed for the presence of the immune classifier rendering a positive result for patient #1 (FDR=0.001) who showed a partial response (**Supplementary Figures 8**) but not for patient #2 (FDR=0.23) who presented with stable disease.

Considering that checkpoint inhibitors are not yet approved for HCC management by regulatory agencies, we compared the gene expression profile of our Immune class with the expression profiles of melanoma patients responding to immunotherapy using a recently published dataset of 32 patients[23]. SubClass mapping analysis revealed that our Immune class, and in particular the Active Immune subtype (**Supplementary Figure 9**), shows similarity to the group of melanoma patients who respond to PD-1 checkpoint inhibitors.

Immune class tumors show lower burden of chromosomal aberrations but no differences in the expression of neo-antigens or tumoral mutational burden

Recent analyses have linked the tumoral genomic landscape with anti-tumor immunity. In particular, it has been proposed that presence of neo-antigens and overall mutational load might drive T cell responses[12, 24, 25] whereas tumor aneuploidy correlates with markers of immune evasion and reduced response to immunotherapy[23, 26]. In order to verify if the burden of somatic copy number aberrations (SCNAs) and mutated neo-antigens may influence local immune

infiltrates in HCC, we used the TCGA dataset. In a recent analysis, the local immune cytolytic activity of several tumors showed strong correlation with cytotoxic T cells and interferon-stimulated chemokines that attract T cells[24]. Interestingly, in HCC patients we observed a strong correlation between the cytolytic activity score and our Immune class ($p<0.0001$, **Figure 4B**). In terms of SCNAs, patients within the Immune class showed lower burden of gains and losses, both broad and focal (**Figure 5A-B** and **Supplementary Figure 10A-B**) with a median of 3 broad gains (range 0-16) and 3.5 broad losses (range 0-20) in the Immune class versus 5 broad gains (range 0-22) and 9 broad losses (range 0-26) in the rest of the cohort ($p=0.046$ and $p=0.01$, respectively). Similarly, we identified a median of 5 focal gains (range 0-18) and 9 focal losses (range 0-25) in the immune class versus 8.5 focal gains (range 0-20) and 13 focal losses (range 0-27) in the rest of the cohort ($p=0.03$ for both comparisons). When analyzing the regions associated with recurrent SCNAs in patients outside the Immune class (low immune infiltrates based on immune signatures), recurrent copy number gain in chromosome 1q and recurrent losses in chromosomes 3p, 17p, and 18p were observed at arm level (**Supplementary Tables 9-10**). In terms of focal high-level amplifications and homozygous deletions, we restricted the analysis to focal structural aberrations involving driver genes previously reported in HCC[27]. As indicated in **Supplementary Table 11**, we only found significant difference for the high level amplification of the locus 11q13 (*CCND1*, *FGF19*, etc.), which was significantly enriched in the Immune class, and particularly in the Active Immune subtype. No significant differences were found regarding loci involving *MYC*, *TERT* and *PTEN*.

We then correlated the Immune class with the overall rate of mutations and rate of predicted neo-antigens, as per previous analysis of the TCGA dataset[24]. There was no association between the Immune class and both features (**Figure 5C and Supplementary Figure 10C**). In particular, the median number of mutations for Immune class compared among the remaining patients was 175 vs 212, respectively ($p=0.1$, **Supplementary Figure 10C**). Similarly, the rate of neo-antigens was not statistically different between the two groups (21 vs 23, respectively,

p=0.28, **Figure 5C**). Nonetheless, when we analyzed both parameters according to the microenvironment-based subtype, the Active Immune subtype showed a trend towards lower neo-epitopes rate (median of 18 versus 33 in Exhausted versus 23 in rest of cohort, p=0.20, **Figure 5D**) and mutations (median of 140 versus 269 in Exhausted subtype versus 212 in rest of cohort, p=0.06, **Supplementary Figure 10C**). Finally, we correlated the Immune class with mutations in known driver genes. With the exception of mutations in the CTNNB1 pathway (12/42 vs 81/148, p=0.003), no other mutations showed differential distribution (**Figure 5E**). All these data show no correlation between neo-antigen load and T cell response, which indicates that additional mechanisms, such as aneuploidy and mutations in specific oncogenic pathways, may impair immune cell recruitment in highly immunogenic tumors.

The Immune class has a unique DNA methylation signature

Considering the profound up-regulation of immune–related genes in the Immune class, we wondered if such deregulation could mirror epigenetic alterations in these tumors. Supervised analysis of whole genome methylation data revealed that 363 CpG sites in 192 immune response gene promoters were differentially methylated in the Immune class compared to the rest of the cohort (FDR<0.05, **Supplementary Figure 11** and **Supplementary Table 12**). Furthermore, among the 192 genes showing differentially methylated CpG sites, 115 showed a significant correlation with gene expression (**Supplementary Table 13**). In particular, the immunosuppressive molecule *LGALS3*, which may play a role in immune escape during tumor progression through the induction of apoptosis of cancer-infiltrating T cells[28] and the regulator of the TGF-ß signaling, *PMEPA1*, were significantly over-expressed in the 2 Immune subtypes (p<0.001**, Supplementary Figure 12**). Overall, these data indicate that the Immune class is characterized by a unique methylation profile. In particular, differential methylation was observed in 192 immune related genes and, in most instances, was associated with altered gene expression.

Specific oncogenic signaling pathways could cooperate to reduce T cell infiltration in the CTNNB1 class of HCC

The integration of the Immune class with previously reported molecular classifications revealed a significant exclusion of the CTNNB1 class in all datasets tested (**Figures 2A and 4B**, **Supplementary Figures 6-7**). The CTNNB1 class of HCC is characterized by over-expression of liver-related Wnt-target-genes, enrichment in nuclear β-catenin staining and CTNNB1-mutations[29]. Exclusion of the CTNNB1 class supports recent reports in melanoma where activation of the pathway is associated with T cell exclusion, through the repression of CCL4 and subsequent failure of T cell priming[14]. In our cohorts, HCC samples within the CTNNB1 class showed significantly lower enrichment score for several immune signatures, in particular T cells, compared to patients within the Immune class or the remaining patients ($p<0.001$, **Supplementary Figure 13A-B**). In addition, in accordance with data in melanoma, patients within the CTNNB1 class showed down-regulation of *CCL4* ($p<0.001$). Further oncogenic pathways have been associated with T cell exclusion, including PTEN[30] and PTK2[31]. Interestingly, *PTK2* was significantly over-expressed in the CTNNB1 class (**Supplementary Figure 13A-C**), suggesting a possible cooperation between PTK2 and CTNNB1 pathways to induce immune cells exclusion in this subgroup**.** In addition, DNA copy number and expression of *PTK2* were highly correlated ($p<0.0001$, **Supplementary Figure 13D-E**). These data suggest that HCC samples within the CTNNB1 class showed lower expression of immune signatures compared to patients of the Immune class and the remaining tumors. Activation of specific oncogenic signaling, such as CTNNB1 and PTK2 signaling –through activating mutations or additional mechanisms- may play a seminal role in influencing the immunological profile of this subgroup.

Compartmentalization of immune signals: immune infiltration in the surrounding tissue does not reflect its tumor counterpart

Finally, in order to assess whether the type of intra-tumoral immune cell infiltrates mirrors its peritumoral counterpart, we correlated the intra-tumoral immune infiltration with the surrounding liver tissue. To do so, we performed a sub-analysis in 167 patients of the training cohort for whom gene expression data were available for both tumor and matched surrounding non-tumoral tissue. Among the 167 cases, 25% (42/167) were positively classified within the Immune class based on the expression profile of the tumor (**Figure 6A**). Interestingly, only a minority of these patients (13/42, 31%) showed a combined positive prediction in both tumor and matched surrounding liver, suggesting that the intra-tumoral immune infiltration does not reflect the profile of the surrounding tissue. Given these observations, we further explored the type of immune infiltration occurring in the non-tumoral liver. Interestingly, patients positively predicted by the Immune Classifier based on the profile of the surrounding tissue showed a strong enrichment of signatures capturing the presence of immune cells (CD8, macrophages, $p<0.001$), activated stroma [31/57 (54%) vs 7/110, (6%), $p=0.0001$], TGF-ß signaling [38/57, (67%) vs 2/110, (2%), p= $p=0.0001$] and additional immunosuppressive components (*LGALS1*, *CXCL12*, etc., $p<0.001$). In addition, exhausted T cells [19/57, (33%) vs 7/110 (6%), $p=0.0001$], and a prognostic 186-gene signature derived from the surrounding liver [43/57 (75%) vs 7/110 (6%), p= 0.0001] were also enriched in this subgroup. In addition, we observed that METAVIR F3-F4 stages [42/45 (93%) vs 66/87 (76%) in rest of cohort, $p=0.02$] and HCV infection [36/54 (67%) vs 39/108 (36%) in rest, $p<0.001$] were significantly associated to a positive Immune Classifier in the surrounding liver. On the other end, HBV infection [7/54 (13%) vs 34/108 (31%), $p=0.01$] and alcohol abuse [2/54 (4%) vs 20/108 (19%), $p=0.008$] were more frequent in patients negative for the immune classifier (**Figure 6A**). Finally, patients positive for the Immune Classifier showed significant worse prognosis with a median survival time of 37 vs 76 months in the rest of the cohort ($p<0.001$, **Figure 6B)**. In essence, these data suggest that the immune profile of the surrounding liver tumor does not reflect the intra-tumoral profile and is mostly characterized by immunosuppressive components associated with survival of HCC patients.

DISCUSSION

Our study represents a comprehensive characterization of the immunological profile of human HCC tumors. The use of virtual separation analytical approaches enabled us to deconvolute the gene expression signals deriving from the intra-tumoral immune infiltrates; this identified a previously unnoticed robust class of HCC (~27% of 956 patients), herein named Immune class. The immune nature of our classifier is supported by the overlap with gene signatures identifying immune cells (i.e. T cells and cytotox), signatures predictive of response to immune checkpoint therapy, presence of high immune cell infiltration, enhanced cytolytic activity and PD-1 and PD-L1 protein expression. Among tumors within the Immune class, we discovered two distinct microenvironment-based immune clusters with either an active or exhausted immune response, ultimately providing a comprehensive description of the intra-tumoral immunologic milieu.

Survival of patients with melanoma or lung cancer has significantly improved since the recent FDA approval of immune checkpoint inhibitors (e.g., nivolumab, pembrolizumab). These compounds elicit durable clinical responses and long-term remissions in a fraction of patients with metastatic disease[7, 32]. Given that these therapies are directed to immune cells rather than tumor cells, they can be effective in a broad range of cancer types, with important activity recently reported in both solid and hematologic malignancies including bladder[33] and colorectal cancer[10, 12]. In HCC, results of the phase II extended clinical trial testing nivolumab indicate an objective response rate of 16%, and median survival of 14 months among the 214 patients treated[15]. In this trial, objective responses (21/145 cases, 15%) were not related to PD-L1 expression on tumor cells[15]. Thus, identification of accurate predictive biomarkers to select ideal candidates for immunotherapy is a major unmet need in HCC. Initial trials, particularly in non-small cell lung cancer, have suggested that patients positive for PD-L1 expression have a greater overall response compared to patients negative for PD-L1[8-10]. Nonetheless, accurate scoring of PD-L1 protein expression is complex due to technical (i.e. affinity, threshold) and

biological pitfalls (i.e. cell type, dynamic expression). Furthermore, responses observed in patients with PD-L1 negative tumors have highlighted the need to investigate more robust biomarkers, such as immune signatures. In this context, a better understanding of the anti-tumor immune responses and the interplay between cancer cells and the microenvironment will be essential to predict reponders to immunotherapies.

Our study identifies a new immune molecular class of HCC, and provides important insights into the immunological profile of this tumor, and how it may be influenced by the interaction with its microenvironment. Close to 25% of HCCs belong to the herein called Immune class, whose molecular characteristics - including high infiltration of immune cells, expression of PD-1 and PD-L1, and active IFN-γ signaling - highly resemble those of cancers most responsive to immunotherapy[11-13]. Indeed, when tested in patients receiving nivolumab, positive prediction of the immune classifier was observed only in the patient achieving objective response to immunotherapy. Accordingly, two immune signatures that predict response to pembrolizumab in melanoma and head and neck squamous cell carcinoma were significantly enriched in patients of our Immune class. These signatures are associated with T cell cytotoxic function and IFN-γ signaling, reinforcing the idea that clinical responses to PD-1 blockade occur in patients with a pre-existing IFN-mediated adaptive immune response[34]. PD-L1 staining was enriched in the Immune class, but failed to capture most of the cases, and thus represents a suboptimal marker. As mentioned before, this is consistent with the lack of predictive capacity observed for PD-L1 expression on tumor cells in the large phase II study with nivolumab for HCC patients[15]. Further investigation in patients receiving immunotherapy is necessary to verify the predictive capacity of the immune classifier. Interestingly, neither the mutational load nor the presence of neo-antigens was associated with the Immune class, suggesting that, unlike melanoma[12] and lung cancer[25], other molecular mechanisms may drive anti-tumor immunity in HCC. A similar lack of association has been described in other tumors with modest mutational burden, such as prostate, ovarian cancer and pancreatic cancer[35]. In these settings, the quality or clonality of

neo-antigens, rather than the quantity, may influence the immune reactivity[25]. In addition, other mutations-independent mechanisms, such as expression of HCC-associated antigens (i.e. AFP, Glypican 3, MAGE, NY-ESO-1) might have an impact on the immune infiltrate. The lack of association between neo-antigenes and immune profile of HCC tumors could also reflect the fact that the immune response is more likely regulated by a combination of both tumor-intrinsic factors, based on the genetic make-up of the tumor (e.g. aneuploidy, activation of specific oncogenic signaling, expression of immune checkpoint molecules, etc), and extrinsic factors present in the microenvironment[36]. Further investigation is needed to fully understand the molecular mechanisms responsible for the different immunogenicity of HCC tumors.

The sole presence of an immune phenotype does not necessarily predict response to immunotherapies. A favorable response to checkpoint inhibitors relies on the intricate and dynamic interactions between tumor cells, immune cells and other immunomodulators present in the microenvironment, which may either dampen or enhance the immune response. In this regard, virtual dissection of the gene expression profile of the Immune class allowed us to elucidate such interactions and identify two clear cut microenvironment-based clusters of samples: 1) *Active Immune Response and 2) Exhausted Immune Response*. Robustness of these subtypes was supported by their successful replication in seven independent datasets across different platforms, ranging from RNA-sequencing to microarray and using distinct types of samples (i.e., fresh frozen and paraffin-embedded tissue). While the A*ctive Immune Response* cluster showed anti-tumor immune features such as enrichment of IFN signatures, overexpression of adaptive immune response genes and better survival, the *Exhausted Immune Response* was characterized by tumor-promoting signals (e.g., activated stroma, T cell exhaustion and immunosuppressive components). In particular, activation of TGF-ß, a potent immunoregulatory cytokine frequently overexpressed in aggressive cancers, was significantly enriched in our *Exhausted Subtype*. TGF-ß regulates tumor-stroma interactions, angiogenesis,

metastasis, and suppresses the host immune response via induction of T cell exhaustion[37, 38] and promotion of M2-type macrophages[39]. Interestingly, we detected differential methylation of 192 immune related genes between the immune clusters, particularly the *Exhausted Immune Response subtype*, which suggests that epigenetic mechanisms could play an important role in influencing the intra-tumoral immune response of HCC patients.

Understanding the interactions between the immune response, oncogenic signaling and the tumor microenvironment is critical to improve the efficacy of current immunotherapies. For example, patients within the *Exhausted Immune Response* subtype could benefit from the combination of TGF-ß inhibition plus immune checkpoint blockade. In this regard, a phase 1b/2 clinical trial testing the novel TGF-ß inhibitor, galunisertib, in combination with nivolumab in advanced solid tumors, including HCC, is currently ongoing in all comers (NCT02423343), with no patient enrichment strategy. Similarly, dissection of the oncogenic mechanisms responsible for T cell exclusion could bring additional combination strategies in patients who otherwise would likely not respond. Recent molecular analyses have revealed a correlation between activation of the CTNNB1 signaling pathway and lower T cell infiltrates in melanoma[14] and other tumors[40]. Consistent with these findings, HCC samples within the CTNNB1 class showed lower immune-cell signature scores. Interestingly, the CTNNB1 class also displayed over-expression of *PTK2*, another oncogenic signal recently reported to drive immune exclusion[31]. While further investigation is required to elucidate the specific role of CTNNB1 and PTK2 signaling and verify the physical absence of T cell infiltrates in the CTNNB1 class, these data suggest that multiple oncogenic pathways could cooperate to modify the immune profile of the tumor.

Finally, we did not observe a correlation between the immune expression profiles of the tumors and the matched surrounding non-tumoral livers. Perhaps more interesting is the fact that among 34% of HCC patients with peritumoral immune profile, most of them contained

immunosuppressive signals, such as TGF-ß activation and T cell exhaustion, associated with shorter survival. This observation is consistent with previous evidence supporting the so-called "field effect" in the damaged liver due to chronic hepatitis and/or cirrhosis[41]. Accordingly, a strong association was observed in the surrounding liver between the Immune Classifier and our previously reported 186-gene signature able to identify HCC patients with poor survival after resection[41], and those HCV cirrhotic patients at higher risk for HCC development[42].

In summary, our study introduces a novel immune-specific class in ~25% of HCC cases which comprises two robust microenvironment-based clusters with either active or exhausted immune responses, who might represent the ideal candidates to receive immunotherapy. Further investigations of this Immune Classifier in a larger cohort of patients receiving immunocheckpoint therapies is needed to determine its potential use as predictive biomarker of response in the design of immune-based clinical trials.

REFERENCES

1. Murray CJ, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380:2197-223.
2. Llovet JM, Zucman-Rossi J, Pikarsky E, et al. Hepatocellular carcinoma. Nat Rev Dis Primers 2016;2:16018.
3. Zucman-Rossi J, Villanueva A, Nault JC, et al. Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. Gastroenterology 2015;149:1226-1239 e4.
4. Llovet JM, Ricci S, Mazzaferro V, et al. Sorafenib in advanced hepatocellular carcinoma. N Engl J Med 2008;359:378-90.
5. Bruix J, Qin S, Merle P, et al. Regorafenib for patients with hepatocellular carcinoma who progressed on sorafenib treatment (RESORCE): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet 2017;389:56-66.
6. Llovet JM, Villanueva A, Lachenmayer A, et al. Advances in targeted therapies for hepatocellular carcinoma in the genomic era. Nat Rev Clin Oncol 2015;12:408-24.
7. Zou W, Wolchok JD, Chen L. PD-L1 (B7-H1) and PD-1 pathway blockade for cancer therapy: Mechanisms, response biomarkers, and combinations. Sci Transl Med 2016;8:328rv4.
8. Herbst RS, Baas P, Kim DW, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet 2016;387:1540-50.
9. Garon EB, Rizvi NA, Hui R, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. N Engl J Med 2015;372:2018-28.
10. Topalian SL, Hodi FS, Brahmer JR, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. N Engl J Med 2012;366:2443-54.
11. Ji RR, Chasalow SD, Wang L, et al. An immune-active tumor microenvironment favors clinical response to ipilimumab. Cancer Immunol Immunother 2012;61:1019-31.
12. Le DT, Uram JN, Wang H, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N Engl J Med 2015;372:2509-20.
13. Bald T, Landsberg J, Lopez-Ramos D, et al. Immune cell-poor melanomas benefit from PD-1 blockade after targeted type I IFN activation. Cancer Discov 2014;4:674-87.
14. Spranger S, Bao R, Gajewski TF. Melanoma-intrinsic beta-catenin signalling prevents anti-tumour immunity. Nature 2015;523:231-5.
15. **El-Khoueiry AB, Sangro B**, Yau T, et al. Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial. Lancet 2017.
16. 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.
17. Yoshihara K, Shahmoradgoli M, Martinez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 2013;4:2612.
18. **Villanueva A, Portela A**, Sayols S, et al. DNA methylation-based prognosis and epidrivers in hepatocellular carcinoma. Hepatology 2015;61:1945-56.
19. Calderaro J, Rousseau B, Amaddeo G, et al. Programmed death ligand 1 expression in hepatocellular carcinoma: Relationship With clinical and pathological features. Hepatology 2016;64:2038-2046.
20. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646-74.
21. Slavuljica I, Krmpotic A, Jonjic S. Manipulation of NKG2D ligands by cytomegaloviruses: impact on innate and adaptive immune response. Front Immunol 2011;2:85.

22. **Viel S, Marcais A, Guimaraes FS**, et al. TGF-beta inhibits the activation and functions of NK cells by repressing the mTOR pathway. Sci Signal 2016;9:ra19.
23. **Roh W, Chen PL, Reuben A,** et al. Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. Sci Transl Med 2017;9.
24. 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.
25. **McGranahan N, Furness AJ, Rosenthal R**, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science 2016;351:1463-9.
26. Davoli T, Uno H, Wooten EC, et al. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. Science 2017;355.
27. **Schulze K, Imbeaud S, Letouze E**, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nat Genet 2015;47:505-11.
28. Fukumori T, Takenaka Y, Yoshii T, et al. CD29 and CD7 mediate galectin-3-induced type II T-cell apoptosis. Cancer Res 2003;63:8302-11.
29. Chiang DY, Villanueva A, Hoshida Y, et al. Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. Cancer Res 2008;68:6779-88.
30. Peng W, Chen JQ, Liu C, et al. Loss of PTEN Promotes Resistance to T Cell-Mediated Immunotherapy. Cancer Discov 2016;6:202-16.
31. Jiang H, Hegde S, Knolhoff BL, et al. Targeting focal adhesion kinase renders pancreatic cancers responsive to checkpoint immunotherapy. Nat Med 2016;22:851-60.
32. **Khalil DN, Smith EL**, Brentjens RJ, et al. The future of cancer treatment: immunomodulation, CARs and combination immunotherapy. Nat Rev Clin Oncol 2016;13:394.
33. Powles T, Eder JP, Fine GD, et al. MPDL3280A (anti-PD-L1) treatment leads to clinical activity in metastatic bladder cancer. Nature 2014;515:558-62.
34. Chow LQM, Mehra R, Haddad RI, et al. Biomarkers and response to pembrolizumab (pembro) in recurrent/metastatic head and neck squamous cell carcinoma (R/M HNSCC). J Clin Oncol 34, (suppl; abstr 6010) 2016.
35. **Balli D, Rech AJ**, Stanger BZ, et al. Immune cytolytic activity stratifies molecular subsets of human pancreatic cancer. Clin Cancer Res 2016.
36. **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;18:248-262.
37. Park BV, Freeman ZT, Ghasemzadeh A, et al. TGFbeta1-Mediated SMAD3 Enhances PD-1 Expression on Antigen-Specific T Cells in Cancer. Cancer Discov 2016;6:1366-1381.
38. Stephen TL, Rutkowski MR, Allegrezza MJ, et al. Transforming growth factor beta-mediated suppression of antitumor T cells requires FoxP1 transcription factor expression. Immunity 2014;41:427-39.
39. Flavell RA, Sanjabi S, Wrzesinski SH, et al. The polarization of immune cells in the tumour environment by TGFbeta. Nat Rev Immunol 2010;10:554-67.
40. Porta-Pardo E, Godzik A. Mutation Drivers of Immunological Responses to Cancer. Cancer Immunol Res 2016;4:789-98.
41. 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.
42. Hoshida Y, Villanueva A, Sangiovanni A, et al. Prognostic gene expression signature for patients with hepatitis C-related early-stage cirrhosis. Gastroenterology 2013;144:1024-30.

Author names in bold designate shared co-first authorship.

Acknowledgments

We thank Prof Jessica Zucman-Rossi for her advice and revision of the paper. We thank Wei Quiang Leow and Agrin Moeini for their support. The results shown here are in part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/.

Table 1. Clinical characteristics of the training (Heptromic) and validation cohorts (Validation and TCGA sets).

Variable†	**Training set (n=225)**	**Validation Set (n=131)**	**TCGA set (n=190)**
Median age (IQR)	66 (61-72)	66 (55-71)	62 (52-70)
Gender, male (%)	178 (79)	96 (73)	123 (65)
Etiology (%)†			
Hepatitis C	101 (46)	64 (50)	N/A
Hepatitis B	48 (21)	39 (30)	N/A
Alcohol	33 (15)	6 (5)	N/A
Others	38 (17)	19 (15)	
Child-Pugh score (%)†			
A	220 (98)	123 (98)	86 (83)
B	3 (1)	2 (2)	17 (17)
Tumor size, cm (%)			
<2	28 (12)	17 (13)	N/A
between 2 and 3	66 (30)	31 (24)	N/A
>3	130 (58)	81 (63)	N/A
Multiple nodules (%)			
Absent	168 (75)	117 (91)	N/A
Present	56 (25)	12 (9)	N/A
Vascular invasion (%)			
Absent	144 (65)	78 (62)	104 (66)
Present	78 (35)	46 (38)	54 (34)
Satellites (%)			
Absent	164 (73)	100 (80)	N/A
Present	60 (27)	25 (20)	N/A
BCLC early stage, 0-A (%)	195 (87)	120 (94)	N/A
Degree of tumor			
Well	33 (15)	31 (26)	31 (17)
Moderately	106 (47)	73 (61)	96 (52)
Poor	44 (20)	16 (13)	58 (31)
Bilirubin, >1 mg/dL (%)	113 (50)	34 (27)	35 (25)
Albumin, <3.5 g/L	26 (12)	13 (11)	42 (31)
Platelet count,	41 (18)	17 (13)	N/A
AFP, >100 mg/dL (%)	51 (23)	38 (31)	43 (33)
Events (%)			
Recurrence	150 (67)	78 (60)	88 (59)
Death	133 (59)	46 (35)	84 (44)
Median follow-up, months	49	51	N/A

†**Missing values Training set:** etiology (n=2); Child-Pugh score (n=2); multiple nodules (n=1); vascular invasion (n=3); BCLC 0-A (n=2); tumor differentiation (n=42); bilirubin (n = 32); AFP (n = 11); albumin and bilirubin (n=4); platelet (n=2); recurrence (n=7). **Missing values Validation set:** etiology (n=3); Child-Pugh score (n=6); tumor size (n=2); multiple nodules (n=2); vascular invasion (n=6); satellites (n=6); BCLC 0-A (n=4); tumor differentiation (n=12); bilirubin (n = 6); AFP (n = 10); albumin (n=9); platelet (n=5); recurrence (n=1). **Missing values TCGA set:** Child-Pugh score (n=87); vascular invasion (n=32); tumor differentiation (n=5); bilirubin (n=50); AFP (n=58); albumin (n=55); inaccurate platelet count; recurrence (n=40); etiology, tumor size and number, satellites and updated follow-up information is not available.

Figure Legends

Figure 1. Flow chart of the study. A total of 956 HCC samples were used in this study. A training cohort (Heptromic) including 228 HCCs was virtually microdissected to identify an Immune class. Validation was then performed in 7 independent datasets.

Figure 2. Identification of the Immune class of HCC. (**A**) Consensus-clustered heatmap of HCC samples (training dataset, n=228) using exemplar genes of the immune expression pattern and refined by Random Forest. In the heatmap, high and low gene set enrichment scores are represented in red and blue, respectively. Positive prediction of signatures is indicated in red and absence in grey. Note: only the 28-gene signature will be shown in following heatmaps. Similar results were obtained with both signatures. (**B**) Representative images of immune cell infiltration, PD-1 and PD-L1 staining in a patient of the Immune class (M515) and a patient outside of the Immune class (B319). Images were captured with 20X.

Figure 3. The Immune class contains two distinct microenvironment-based subtypes. NTP analysis of whole-tumor gene expression data using a molecular signature able to capture activated inflammatory stromal response identified two distinct subtypes of Immune class – the *Active* (blue color bar) and the *Exhausted* (green color bar) *Immune Response Subtypes*. In the heatmap, high and low gene set enrichment scores are represented in red and blue, respectively; same representation is used for high and low gene expression. Positive prediction of signatures as calculated by NTP is indicated in red and absence in grey.

Figure 4. Kaplan-Meier estimates of overall survival according to the immune response type status and robustness of the Immune class. (**A**) Kaplan-Meier estimates of overall survival according to the *Active Immune Response* status in the Heptromic cohort (*Active Immune Response* vs rest plus *Exhausted Immune Response*). (**B**) Kaplan-Meier estimates of

overall survival according to the *Active Immune Response* status in the TCGA cohort. (**C**) External validation of the Immune class was conducted in the publicly available TCGA dataset.

Figure 5. Association of the Immune class with copy number aberrations, presence of neo-antigens and mutations in driver genes. Patients of the Immune class showed significantly lower burden of gains (**A**) and losses (**B**), both broad (left panels) and focal (right panels).(**C**) The rate of mutations predicted to yield a neo-antigen was similar between the Immune class and the rest of the cohort and (**D**) between the two microenvironment-based subtypes. (**E**) Heatmap representation of the distribution of mutations in known driver genes between patients of the Immune class and rest of TGCA cohort.

Figure 6. The intratumoral immune profile does not correspond to the immune infiltration of the surrounding non-tumoral liver. (**A**) Gene expression of the tumor (upper panel) and matched surrounding non-tumoral liver (lower panel) was available for 167 patients of the Heptromic cohort (training dataset). Heatmap represents enrichment scores for immune signatures in the tumors (upper panel) and corresponding surrounding tissues (bottom panel). Multi-nodularity was more frequent in patients positive for the immune classifier [25/55 (45%) vs 24/110 (22%), p=0.01]. (**B**) Kaplan-Meyer estimates of overall survival according to the presence of the Immune Classifier in the surrounding liver.

Figure 1

Figure 1

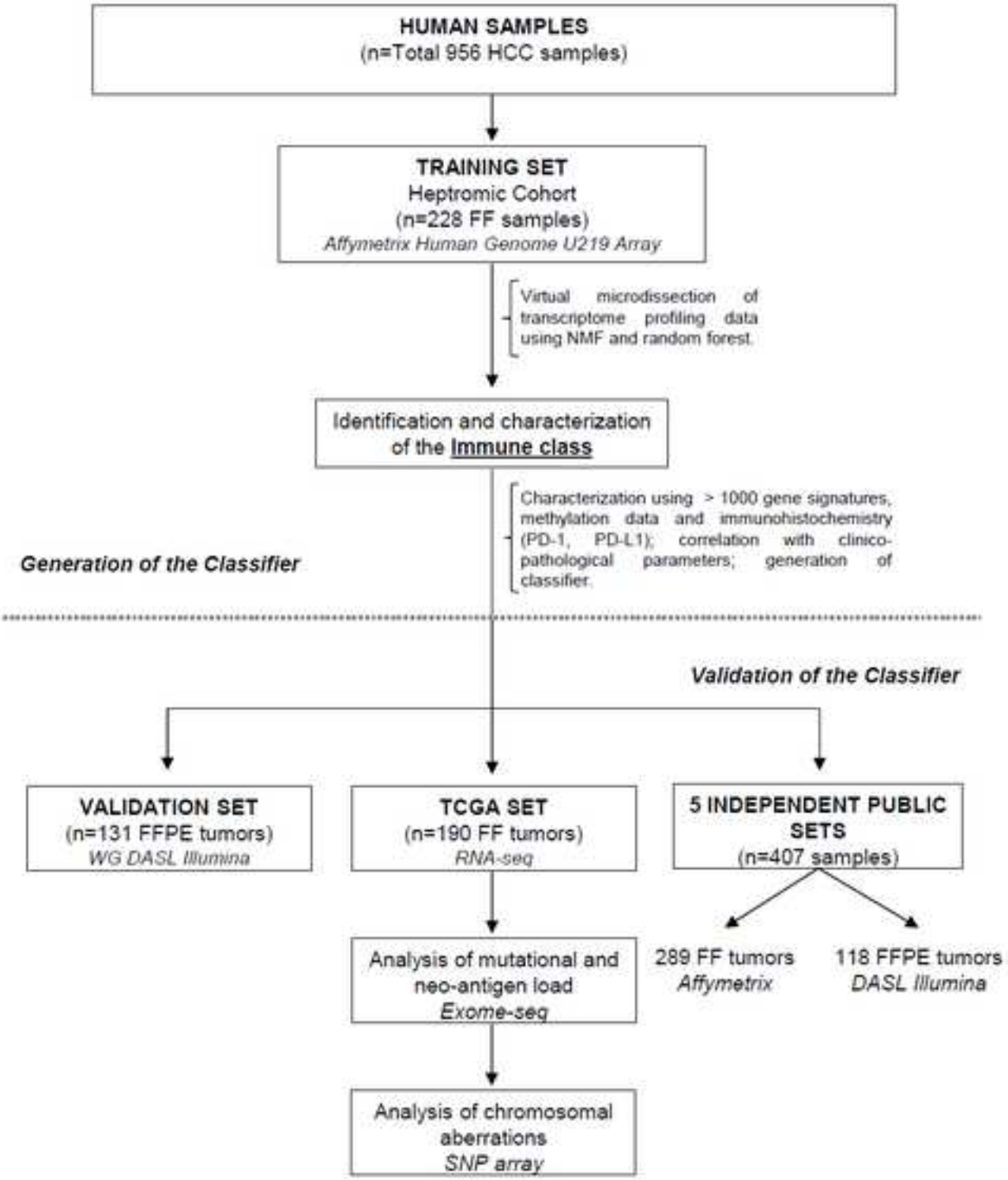
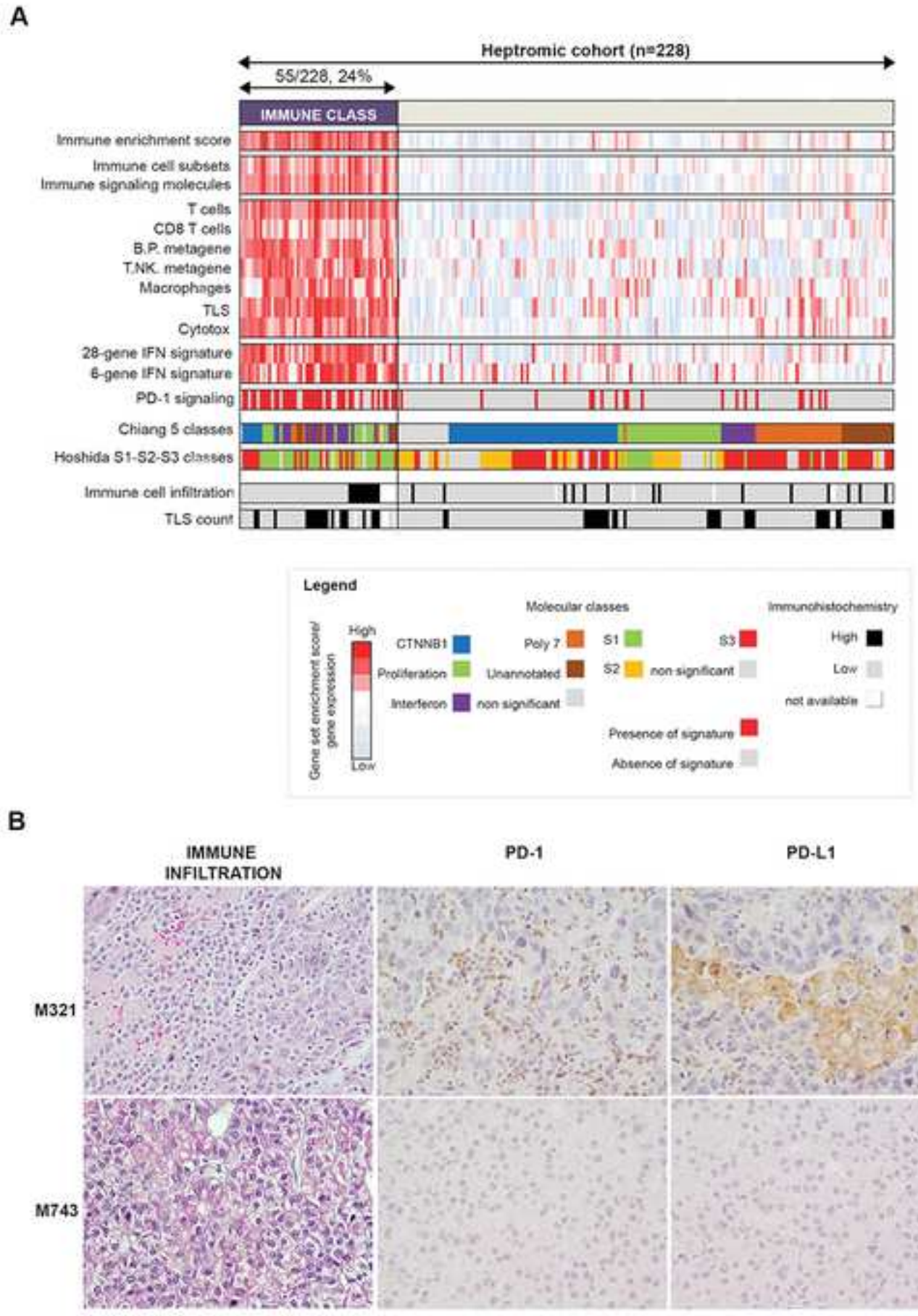


Figure 2

Figure 2



Click here to download Figure Figure 3.tif

Figure 3

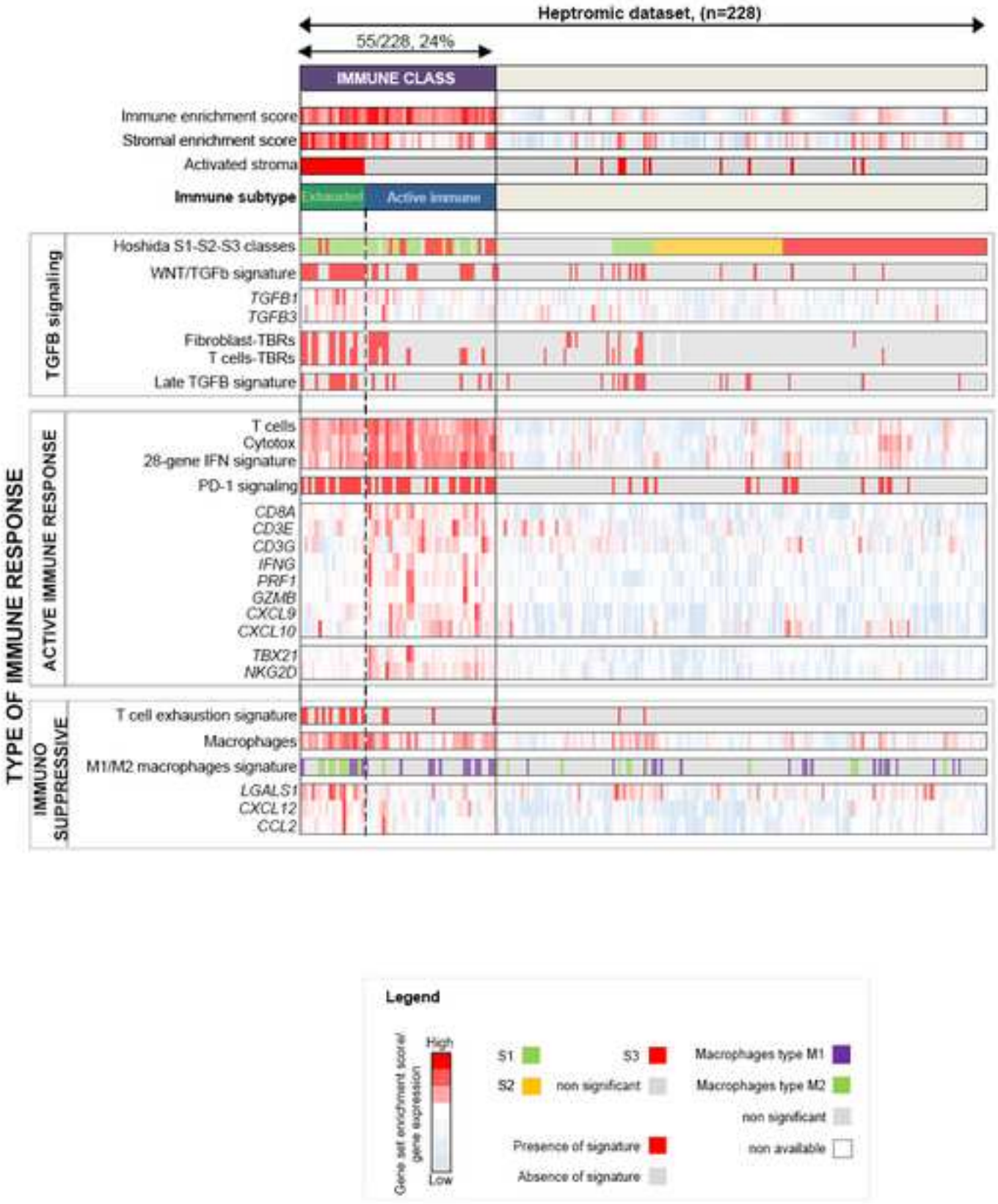


Figure 4

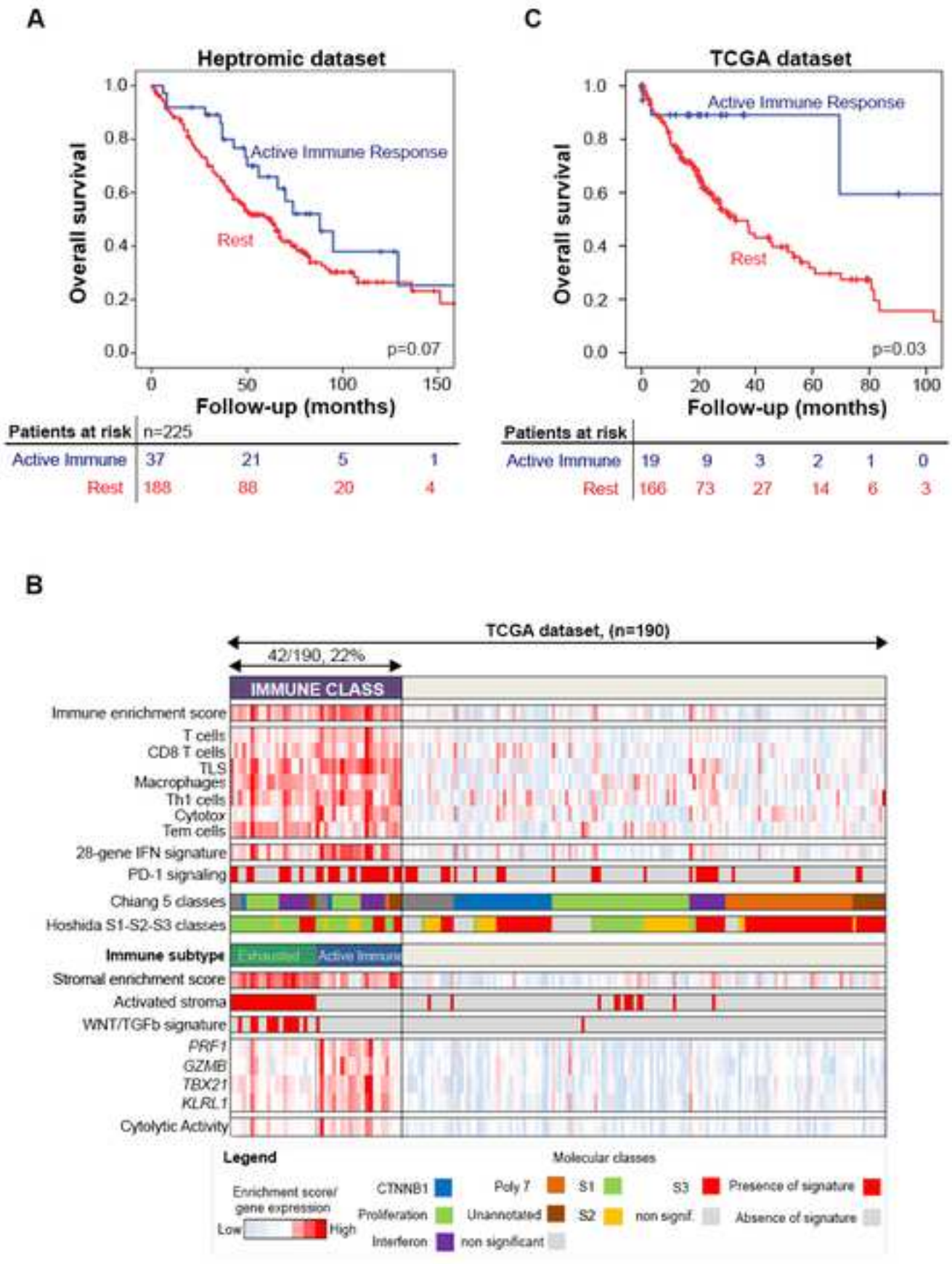


Figure 5

Click here to download Figure Figure 5 - New.tif