의학박사 학위논문

간세포암의 예후 결정인자로서 유전체 변이 및 면역분자적 특성의 중요성

Mutational and Immunomolecular Profiles as Determinants of Clinical Outcomes of Hepatocellular Carcinoma

울 산 대 학 교 대 학 원

의 학 과

황 희 상

Mutational and Immunomolecular Profiles as Determinants of Clinical Outcomes of Hepatocellular Carcinoma

지도교수 성 창 옥

이 논문을 의학박사학위 논문으로 제출함

2018년 12월

울 산 대 학 교 대 학 원

의 학 과

황 희 상

황희상의 의학박사학위 논문을 인준함

심사위원 박 찬 식 인

심사위원 김 지 훈 인

심사위원 천 성 민 인

심사위원 이 다 근 인

심사위원 성 창 옥 인

울 산 대 학 교 대 학 원

2018년 12월

감사의 글

이 논문의 전반적인 내용을 기획하시고 수많은 결과들을 일목요연하게 엮어 주신 유은실 교수님, 심주현 교수님과 성창옥 교수님께 깊은 감사를 드립니다. 또한 실험 대상 환자군의 임상의학적 특성과 실험실 부분에서 많은 도움을 주신 안지현 선생님, 조직병리 및 면역조직화학 염색 판독 부분에 큰 도움을 주신 강효정 선생님, 그리고 생물정보학 분석에 있어서 바쁘신 와중에도 조언을 아끼지 않으신 김덕훈 선생님께도 깊은 감사를 드립니다.

Abstract

Hepatocellular carcinoma (HCC) is prevalent worldwide and the third leading cause of cancer-related deaths. We used high-resolution transcriptomic and genomic analysis of 206 resected HCC samples to identify biological alterations in tumors and immune systems influencing clinical outcomes. We identified five discrete subtypes of HCC: AMC-C1, well-differentiated tumors; AMC-C2, tumors displaying high levels of epithelial-mesenchymal transition (EMT); AMC-C3, those harboring *CTNNB1* mutations; AMC-C4, those with stem cell features; and AMC-C5, finally a subtype expressing cancer testis antigen (CTA). The subtype with stem cell features (AMC-C4) was associated with the worst prognosis both in our dataset and other datasets. An immunogenomic analysis showed that one third of the tumors had an immunogenic subtype characterized by enhanced cytolytic activity, higher levels of cytotoxic lymphocytes (CD8+ to Treg cell ratio) and higher expression of immunomodulatory genes, and patients with these tumors had a better prognosis than those with the alternative immune-desert subtype. Multiple control networks involving microRNAs and copy number alterations were implicated in tumor-immune interplay and the resultant outcomes. We also discovered and validated a promising trackable biomolecule and actionable target, S100P, associated with early recurrence after HCC resection. We conclude that the bioclinical phenotypes of HCC are based on collaboration between genetic and immunomolecular architectures.

Key words: Hepatocellular carcinoma; Molecular subgroup; Tumor immune response; Treatment outcome.

Contents

Lists of tables and figures

INTRODUCTION

Hepatocellular carcinoma (HCC) is now the third leading cause of cancer mortality worldwide.[1)] Exceedingly high rates of HCC have been consistently observed in East and South Asia, and furthermore its incidence is increasing in Western countries.[2)] In the United States, the incidence has nearly tripled since the early 1980s, making it the fastest-rising cause of cancer-related deaths. Mortality rates due to HCC have also increased in many European countries over recent decades, especially in Northern and Central Europe.[2)]

Unsatisfactory outcomes due to high rates of recurrence and progression despite curative treatment, and therapeutic resistance at later stages remain major concerns in HCC management.[3, 4)] Since the evolution of HCC is a multi-step process involving diverse genetic events,[5)] the precise factors contributing to HCC development are only partially understood. Due to the lack of preventive and therapeutic targets among the clinical phenotypes of HCC, understanding its pathogenic origin and molecular hallmarks is crucial to improving patient outcomes.

In this context, several microarray-based classifications of HCC tumors have been developed linking genomic information and clinico-histopathological factors.[6-8)] The recent identification of a distinct immune-specific group of HCC suggests that such a classification could be helpful in guiding patient selection for immunotherapeutic interventions.[9)] However, the individual properties of molecular pathways have provided limited insight into the biologic underpinnings of this disease, as they mostly focus on a single genetic structure.[6, 7, 10, 11)] The bigger picture involving complex networks of biological and immunological systems remains elusive.

Here, we describe an integrated multi-platform analysis of mRNA and microRNA (miRNA) expression, copy number variation (CNVs), and mutational signatures, combined with identification of predefined genetic alterations by whole exome sequencing,[12)] in resectable HCC cases. We comprehensively characterized the mutational and transcriptional landscape

of HCC and identified functional cross-talk between genomic instabilities, miRNA-mediated regulation, and immune systems in the tumor microenvironment that could determine early and late outcomes after resection.

MATERIALS AND METHODS

RNA sequencing

RNA was extracted from fresh-frozen hepatocellular carcinoma (HCC) tissue, and sequenced. Briefly, 1ug of total RNA was converted into an mRNA library using an Illumina mRNA TruSeq kit, and the library was sequenced using the HiSeq2000 platform (Illumina Inc., San Diego, CA). Total RNA containing microRNA was extracted using mirVana miRNA Isolation kit (Life Technologies). Total RNA integrity was checked using an Agilent Technologies 2100 Bioanalyzer with an RNA Integrity Number (RIN) value greater than 8. Small RNA sequencing libraries were prepared according to the manufacturer's instructions (Illumina Small RNA Prep kit). Total RNA (1μg per sample) was ligated with RNA3' and RNA5' RNA adapters. Reverse transcription followed by PCR is used to create cDNA constructs based on the small RNA ligated 3' and 5' adapters. This process selectively enriches those fragments that have adapter molecules on both ends. PCR is performed with two primers that anneal to the ends of the adapters. cDNA was purified with Sage Science's the Pippin prep electrophoresis platform. The quality of the libraries was verified by capillary electrophoresis (Bioanalyzer, Agilent).

After qPCR using SYBR Green PCR Master Mix (Applied Biosystems), we combined libraries that index tagged in equimolar amounts in the pool. Cluster generation occurred in the flow cell on the cBot automated cluster generation system (Illumina). And then the flow cell loaded on HiSeq 2500 sequencing system (Illumina). RNA-seq and miRNA-seq data analyses were carried out with the Cancer Genome Atlas (TCGA) RNAseq version 2 pipeline, with mapping by MapSplice v2.2.1.[13] GRCh37 served as the human reference

genome. Gene expression was estimated using RSEM v1.3.0[14] and normalized within-sample to a fixed upper quartile. Final normalized expression profiles were used for downstream analyses by direct $\log_2$-transformation.

Discovery of HCC clusters

Unsupervised consensus clustering of $\log_2$-transformed RNA-seq expression data was carried out by the non-negative matrix factorization (NMF) algorithm implemented in R package *NMF*.[15] Genes belonging to the top 10 percent of the variance were used as input. After 200 iterations, cophenetic statistics according to number of clusters were compared, to identify the optimal number of clusters. For the TCGA Liver Hepatocellular Carcinoma (TCGA-LIHC) extended dataset,[5] the previously normalized RNA-seq expression data for 360 pathologically-confirmed HCCs was downloaded from cBioPortal (http://www.cbioportal.org), and consensus clustering was performed in the same manner. Genes differentially expressed between clusters were identified using empirical Bayes-moderated t-statistics in R package *limma*.[16]

Calculation of single sample gene set scores

The list of MSigDB version 6.0 gene set collections was downloaded from the gene set enrichment analysis web site (http://software.broadinstitute.org/gsea). For the individual gene sets, single-sample scores for each gene set were calculated from the RNA-seq expression data by the gene set variation analysis method implemented in R package *GSVA*.[17] Comparisons of the gene set scores between the patient groups were carried out using empirical Bayes-moderated t-statistics.

Gene set enrichment analysis

Pathway enrichment in the postulated groups was evaluated by gene set enrichment analysis

(GSEA) with the java-based program available on the GSEA web site (http://software.broadinstitute.org/gsea). Like the single sample gene set score, MSigDB version 6.0 gene set collections were used for the analysis.

Estimation of immune cell infiltration and immune-associated parameters

Stromal and immune scores for each HCC were calculated using R package *ESTIMATE*.[18] For the TCGA-LIHC dataset, pre-calculated stromal and immune scores were downloaded from the ESTIMATE website (http://bioinformatics.mdanderson.org/estimate/). The relative amounts of the immune cell subtypes were calculated using the absolute value mode of the CIBERSORT web-based tool (http://cibersort.stanford.edu) with the default settings.[19]

Several parameters such as "CD8+ T-cell to regulatory T-cell (Treg) ratio", "cytolytic signature" and "pro- to anti-inflammatory cytokine ratio" were calculated from the $\log_2$-transformed AMC and TCGA-LIHC extended datasets following the methods of Davoli et al.[20]

Immune classes and their subclasses (immune-exhausted and immune-activated), as defined by Sia et al,[9] were also applied to the AMC dataset using NearestTemplatePrediction module version 3 implemented in the GenePattern website (https://genepattern.broadinstitute.org), by the author-defined method. The classifier genes defined in Sia et al. and Moffitt et al.,[21] were used to classify immune subtypes.

Processing of somatic mutation data

Mutation annotation format (MAF) files for the somatic mutation statuses of HCCs in the AMC [12] and TCGA-LIHC datasets were downloaded from cBioPortal (http://www.cbioportal.org). The mutation data were visualized using R package *maftools*.[22] Mutational signatures previously defined in COSMIC[23] were calculated from the MAF files using R package *deconstructSigs*.[24]

Processing of somatic copy number aberration data

Copy number variation (CNV) data based on CytoScan HD array (Affymetrix, Santa Clara, CA) were analyzed with GISTIC 2.0 software for each HCC in the AMC dataset[12] (**Supplementary Fig. 1**).[25] The thresholds for confidence levels and absolute $\log_2$ ratios were defined as 0.99 and 0.4, respectively. Additionally, the germline CNV-corrected Genome-Wide SNP array 6.0 data file of the TCGA-LIHC dataset was downloaded from GDAC firehose (http://gdac.broadinstitute.org) and processed as above. We defined the CNV fraction as the ratio between the length of genome segments more than 0.4 of the absolute log2 ratio value and the length of hg19 genome and it was calculated for the AMC dataset.

Identification of known HCC subgroups using gene expression data

The assignments of each HCC in the AMC dataset to the Hoshida[6] and TCGA iCluster[5] subgroups were predicted using previously annotated gene expression data. The SOFT-formatted family file of GSE10186 was downloaded and processed using R package *GEOquery*.[26] After adjusting for the batch effect using the *ComBat* function in R package *sva*,[27] a multinomial elastic net-regularized logistic regression model for the Hoshida subgroups was fitted by R package *glmnet*.[28] The optimal statistical model was selected by leave-one-out cross validation. Using the optimized model, the probabilities of each Hoshida subclass were calculated from the AMC expression dataset. Likewise, the probabilities of each TCGA iCluster subclass were calculated from the AMC expression data using a multinomial logistic regression model established from the TCGA gene expression dataset. Conversely, AMC clusters defined by unsupervised NMF clustering were assigned to the various HCC gene expression datasets namely the TCGA-LIHC, RIKEN (downloaded from ICGC data portal: https://dcc.icgc.org) and GSE14520 datasets (downloaded from https://www.ncbi.nlm.nih.gov/geo/).

Immunohistochemical staining and evaluation

Immunohistochemical staining was carried out for PD-1 (rabbit monoclonal ab137132 1/1,000, AbCam, Cambridge, MA, USA) and for PD-L1 (rabbit monoclonal E1L3N, 1/100; Cell Signaling, Danvers, MA, USA) in serial 4-μm-thick whole slide sections from key formalin-fixed paraffin-embedded (FFPE) blocks, for CTLA-4 (mouse monoclonal ab134090, 1/500; Abcam, Cambridge, UK) antibody in tissue microarray (TMA) blocks, and for S100P (goat polyclonal AF2957, 1/200; R&D systems, Minneapolis, MN, USA) antibody in whole section sand TMAs. The slides were processed on a Bench Mark XT automatic immunostaining device with an OpticView DAB IHC Detection Kit (Ventana Medical Systems) according to the manufacturer's instructions. The results of immunostaining were independently evaluated by two pathologists (E.Y and H.J.K) who were blinded to clinical outcomes. Membranous PD-L1 expression in tumor cells and tumor-infiltrating mononuclear cells (TIMCs), and membranous PD-1 expression in TIMCs, were scored by the Immunoreactivity Scoring System (IRS) based on the percentage of stained cells and staining intensities (**Supplementary Table 1**). TIMCs in each cellular fraction with IRS scores for PD-1 or PD-L1 ≥ 1, and tumor cells with IRS scores ≥ 3 for PD-L1, were defined as positive. The percentages of TIMCs with membranous CTLA-4 expression and tumor cells with cytoplasmic and/or nuclear S100P expression were determined, and cases in each fraction with ≥ 1% of CTLA-4-expressing TIMCs and with ≥ 1% of S100P-expressing tumor cells, respectively, were defined as positive ,.

Statistical analysis

Comparisons of continuous and categorical variables between pairs of independent groups were carried out using Student's t-tests and chi-square tests, respectively. For comparisons of continuous variables between three or more groups, one-way analysis of variance was used.

Kruskal-Walls and Fisher's exact tests were used for variables that deviated from a normal distribution and that were expected to occur too rarely (less than 5 occurrences), respectively. Receiver-operating characteristics (ROC) curves with area-under-curve (AUC) values were calculated using R package *pROC*.[29] AUC values for the same response variables between *AFP* and *S100P* were compared using the DeLong test.[30]

As we wished to estimate cumulative incidence rates of HCC-specific death, and to compare rates between clusters, we estimated the cumulative incidence rates nonparametrically after considering deaths from other causes as competing risks. We also carried out competing risks multivariable regression analyses to evaluate the prognostic effects of components of our molecular classification, adjusting for covariates. The competing risk analyses were performed using R packages *cmprsk*.[31, 32]

Cell culture and reagents

Human HCC cell lines SNU182, SNU449, and SNU475 were purchased from the Korean Cell Line Bank (KCLB; Seoul, Korea) and grown in RPMI-1640 containing 10% fetal bovine serum (GIBCO, Waltham, MA). The cells were trypsinized and 3×10^5 cells/plate were seeded in 100 mm dishes and allowed to grow overnight at 37 °C in a humidified incubator with 5% CO_2 for 24 h They were then transfected with S100P-specific siRNAs in Opti-MEM (Life Technologies, Gaithersburg, MD, USA). Transfection was carried out with Lipofectamine 2000 reagent (Invitrogen, Carlsbad, CA, USA).

siRNA preparation

S100P siRNA and negative control siRNA were designed and synthesized by Bioneer (Seoul, Korea) and used to transfect SNU475 cells (S100P siRNA sense sequence: GGUGGGUCUGAAUCUAGCA dTdT, antisense sequence: UGCUAGAUU CAGACCCACC dTdT and negative control siRNA sense sequence:

GAUAUAGGAAGUCC AUACU dTdT, antisense sequence: AGUAUGGACUUCCUAUAUC dTdT).

Cell viability assay

Cell viability was measured using the CellTiter-Glo luminescence assay (Promega, Madison, WI). Briefly, 3×10^3 cells in 100 µl were plated in triplicate wells in microtiter plates. After 72 h, 100 µL of CellTiter-Glo reagent was added, and the cells were incubated for 10 min. at room temperature. Luminescence was measured with a Wallac 1420 (PerkinElmer, Boston, MA).

Clonogenicity assay

After transfection for 72 hours, cells were seeded in six-well plates at 1000 cells per well and incubated in a 37 °C incubator for 7–10 days. Colonies were washed twice with PBS, fixed with 95% ethanol for 10 minutes, and stained with 0.1% crystal violet for 20 minutes.

Cell cycle analysis

Cell cycle distributions were analyzed by flow cytometer following PI staining. After transfection for 72 hours, cells were harvested, resuspended, centrifuged at 4 °C at 1,200 rpm for 10 minutes and washed three times with cold PBS. Ethanol (75%) was added dropwise, and the cells were fixed at 4 °C overnight. The ethanol was removed by centrifugation, and the cell pellets were resuspended in 100 µg/mL DNase-free RNase for 30 minutes at room temperature in the dark. Finally, PI solution (100 µg/mL) was added and the cells were analyzed by flow cytometer (FACS Calibur™, BD Biosciences). Each sample was measured in duplicate, and the results are presented as averages of three independent assays.

Invasion assays

Invasion assays were performed using an insert chamber (8 μm pore size; Falcon). 1×10^5 cells were collected and suspended in 750 μl of serum-free RPMI and transferred into the upper chamber coated with matrigel. In the lower chamber, 10% FBS-containing medium served as chemoattractant. After 72 h, the cells were fixed in 3.7% formaldehyde for 30 minutes and methanol for 15 minutes, stained with 0.25% crystal violet, imaged, and counted under an inverted microscope.

RESULTS

Study samples and flow design

We used samples of 206 HCCs and paired non-tumors obtained with informed consent after surgical resection from the Bio-Resource Center at the Asan Medical Center (AMC), Seoul, Republic of Korea (http://brc.amc.seoul.kr). The Asan Medical Center Institutional Review Board approved the study protocol. All the cases were estimated by two pathologists (H.J.K and E.Y) to have a tumor cellularity of more than 70%, and complete baseline and long-term follow-up data were available for them. We generated mRNA and miRNA deep sequencing data for all 206 samples and also used the raw whole exome and copy number dataset from our published study currently contained in the cBio Cancer Genomics Portal (http://www.cbioportal.org/),[12] in which all samples analyzed in this study are included (**Materials and Methods**). This cohort contained 203 early stage (98.5%: American Joint Committee on Cancer [AJCC] stage IA, IB, or II) and 3 more advanced stage HCC cases (1.5%: AJCC stage IIIA).[33] Their clinical backgrounds are shown in **Supplementary Table 2**. Based on these data, systematic molecular and immunogenomic profiling was performed using computational genomic tools; in addition, transcriptome markers for early recurrence of HCC were discovered and confirmed.

Five distinct subgroups of HCC revealed by transcriptome profiling

To determine the extent of heterogeneity in mRNA expression across our cohort, we performed non-negative matrix factorization (NMF) clustering[15] with the mRNA expression profiles of the 206 primary samples (**Supplementary Fig. 2**). This identified five robust subgroups [referred as to as AMC-C1 (n = 59), C2 (n = 19), C3 (n = 39), C4 (n = 53), and C5 (n = 36)] that were characterized by distinct mRNA expression patterns, enriched for different cytogenetic and molecular alterations, and correlated with different clinical features and outcomes (**Fig. 1** and **Supplementary Fig. 3**). The subgroups were also integrated with data on differences in miRNA expression, somatic mutations and CNV (**Fig. 1a, Fig. 2, Supplementary Fig. 4,** and **Supplementary Fig. 5**).[12]

AMC-C1 consists mostly of "well-differentiated HCC", with strong expression of liver-specific metabolic enzymes and low levels of mitosis-related genes. miR-10b, miR-34c, and miR-200b (which act as tumor suppressors in several cancers) are also down-regulated in AMC-C1 tumors.[34, 35] Using gene expression and gene-set-enrichment analysis (GSEA), AMC-C2 was characterized by up-regulation of the TGF-β, NOTCH, and VEGF signaling pathways and of gene sets involved in epithelial-mesenchymal transition (EMT) ("high EMT" group) (**Supplementary Fig. 6**). The molecular associations in AMC-C2 are supported by prior experimental findings: 1) TGF-β is a key element of EMT leading to HCC progression,[36] and 2) EMT is associated with the production of pro-angiogenic factors such as VEGF.[37] Elevation of miR-10a, which promotes EMT in glioma,[38] and down-regulation of miR-194-1 (TGF-β inhibitor), were also observed in AMC-C2.[39] Interestingly, AMC-C2 had the lowest fraction of CNVs among the subtypes ($P = 2.5 \times 10^{-5}$). AMC-C3 ("*CTNNB1*-mutated HCC") had the highest frequency of *CTNNB1* mutations (82%; $P < 2.2 \times 10^{-16}$), with elevation of Wnt/β-catenin-related signals including miR-34a.[40] Like those of subtype 1, subtype 3 tumors exhibited hepatocyte-associated gene expression. Intriguingly, miR-122, which is liver-specific and the most abundant miRNA in liver,[41] was significantly upregulated in the AMC-C3 samples. In contrast to the other subtypes, AMC-C4 and AMC-

C5 were comprised mostly of tumors in which cell-proliferation- related genes were up regulated, accompanied by tumors exhibiting poor differentiation and microvascular invasion.

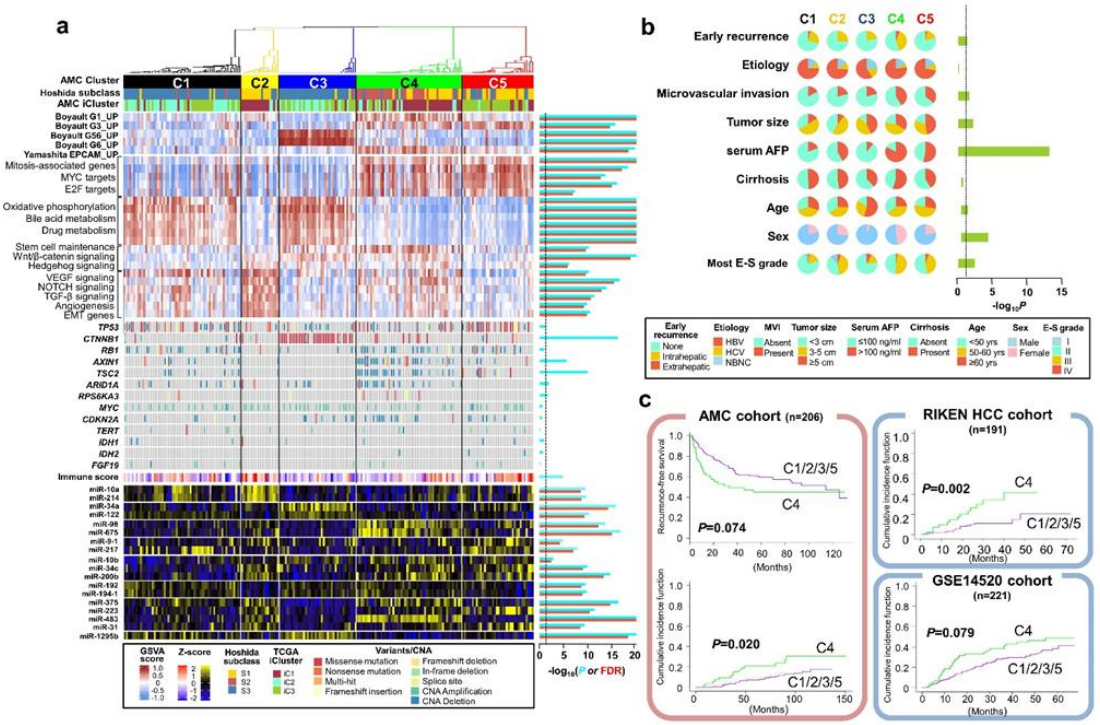


Fig 1. Delineation of the five molecular classes (AMC subgroups) in resectable HCCs

(a) Unsupervised clustering of RNA-seq identifies 5 HCC subgroups. Subgroups significantly enriched for biological functions by gene ontology (GO) analysis are shown to the right. From top to bottom, rows are comparisons of gene expression data, and selected genetic alterations in the pre-defined classes of HCC (b) Clinico-pathological; and (c) prognostic associations of the molecular subtypes. The results of the competing risks analysis indicate that the AMC-C4 group is significantly linked to high AFP in serum and has a worse prognosis than the other subgroups. The poorer outcomes of subgroup 4 were confirmed in the RIKEN (Japanese samples) and GSE 14520 cohorts (Chinese samples). CNA, copy number alteration; AFP, alpha-fetoprotein; E-S grade, Edmonson-Steiner grade; MVI, microvascular invasion; HBV, hepatitis B virus; HCV, hepatitis C virus; NBNC, non-hepatitis B non-hepatitis C hepatocellular carcinoma.

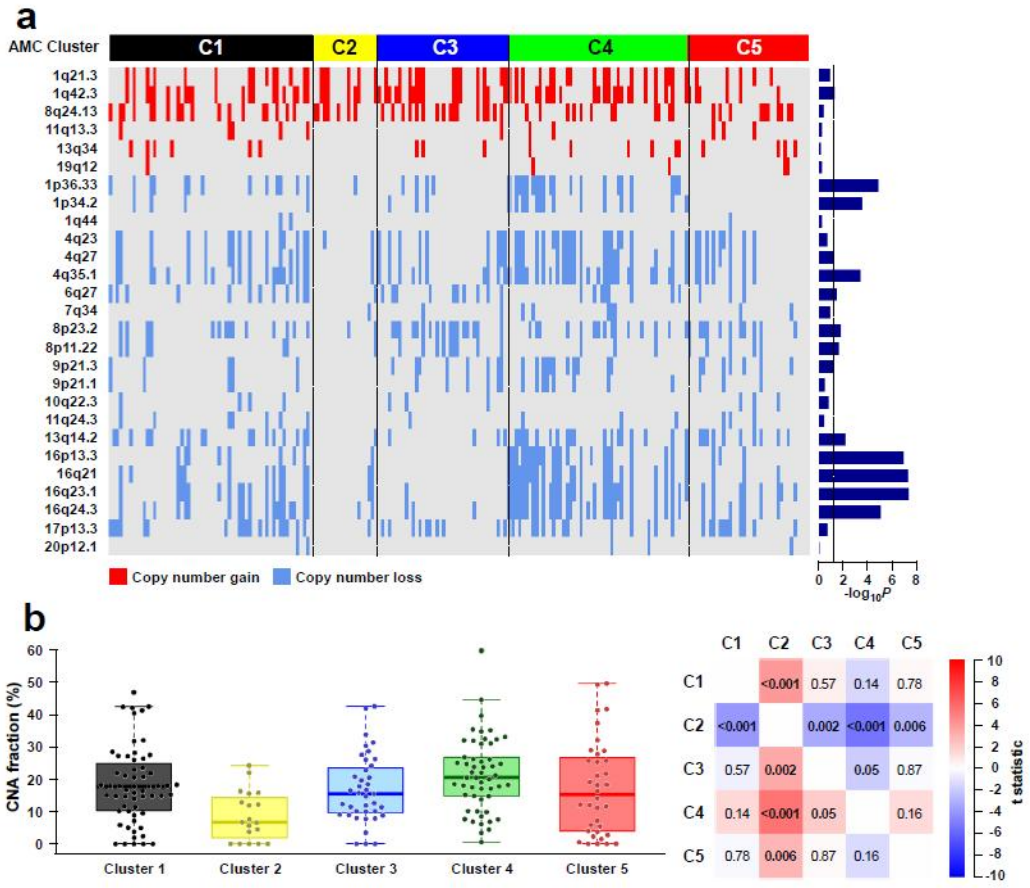


Fig 2. GISTIC analysis of foci of recurrent amplification and deletion across the 5 AMC subgroups of HCC. (a) CNV in focal regions detected by GISTIC 2.0. Regions of recurrent focal amplification (red) and focal deletion (blue) are shown. (b) Distributions of frequencies of somatic copy number alterations in complete genomes in the subgroups. The statistical significances of pairwise comparisons are shown. AMC-C2 tumors had significantly less CNV than the other clusters. CNV, copy number variation.

AMC-C4 (the "stem cell feature" subgroup) exhibited enrichment of oncofetal genes (*GPC3, SALL4*, and *AFP*) and frequent multiple losses of copies of chromosomal regions including chromosome regions 1p, 4q, 13q and 16q, which encompass *ARID1A*, *CXCL13*, *RB1*, *AXIN1* and *TSC2*, respectively. Interestingly, decreased expression of *ARID1A, RB1,* and *TSC2* has been found to be associated with poor prognoses in several next-generation sequencing-based studies.[12, 42-44] The miRNA analysis also revealed over-expression of miR-675, which has been associated with down-regulation of *RB1* in AFP-secreting HCC.[45] The GSVA score[17] for stem cell maintenance (GO:2000036) is high in the AMC-C4 subgroup, which also contains many AFP-expressing (≥ 100 ng/ml) tumors with more aggressive clinical phenotypes (83%; $P = 6.69 \times 10^{-14}$). AMC-C5 (the "CTA" group) is distinguished by over-expression of cancer-testis antigen (CTA) genes, has been never documented in predefined classifications, up-regulation of miR-9-1 (a well-known cancer-related miRNA)[46], and frequent *TP53* gene alterations (56%). In terms of mutational signature, AMC-C2 was the subgroup most highly correlated with COSMIC signature 24, which suggests exposure to aflatoxin (**Supplementary Fig. 8**).[23, 47] The smoking-related signature (signature 4) was least common in the AMC-C2 group.[23]

We next sought to find interrelationships between our subgroups and previously reported HCC molecular classifications. When we assigned each of our patients to the subclasses of Hoshida (S1-S3)[6] and TCGA (iCluster 1-3)[5], there was good correspondence between our AMC subgroups and these classification systems (**Supplementary Fig. 8** and **Supplementary Fig. 9**). Tumors considered to belong to the Hoshida S3 subclass were present across the AMC-C1 and C3 subgroups, and representatives of the Hoshida S1 subclass were found in AMC-C2, C4, and C5. Similarly, in terms of the RT-PCR-based classification described by Boyault et al.,[7] G1 and G3 tumor signatures were enriched exclusively as AMC-C4 and C5, respectively (**Fig. 1a**), the G5 and G6 signatures were enriched in AMC-C3, and G2 and G4 were not represented in our series of mainly early

cancers.

Survival patterns and actionable targets according to the AMC clusters

Clinically, competing risks analysis, which can correctly estimate cancer deaths of interest in the presence of preceding competing events, indicated that patients in AMC-C4 had a significantly higher proportion of cancer-related deaths than the other patients ($P = 0.02$, log-rank test; **Fig. 1c** and **Supplementary Fig. 10**). We also showed that this cluster-associated survival difference persisted, together with liver cirrhosis, after competing risks multivariable regression analysis (**Supplementary Table 3**). To determine whether our expression-based clusters could be reproduced, we took advantage of other public datasets with disease stage distributions different from ours. These consisted of the extended TCGA-LIHC (AJCC stage I, 46.4%; stage II, 22.5%; stage III, 23.3%; stage IV, 1.1%; and others without information on tumor stage, 6.7%),[5] the RIKEN (Liver Cancer Study Group of Japan [LCSGJ] stage I, 16.2%; stage II, 48.2%; stage III, 29.3%; and stage IV, 6.3%),[48] and GSE14520 cohorts (Barcelona Clinic Liver Cancer [BCLC] stage 0, 9.0%; stage A, 67.0%; stage B, 10.0%; stage C, 13.1%; and others without information on tumor stage, 0.9%)[49] from the U.S., Japan, and China, respectively. When we classified them with the elastic net-regularized logistic regression model, all five AMC subgroups were identified in these three external cohorts (**Supplementary Fig. 3** and **Supplementary Fig. 8**). Similar survival trends were also observed in the RIKEN and GSE14520 cohorts (**Fig. 1c** and **Supplementary Fig. 10**).

To investigate the clinical relevance of the molecular characterization of HCC, we matched subgroup-specific pathways with potential standard or investigative treatments. TGF-β signaling-enriched AMC-C2 tumors may benefit from Galunisertib (a TGF-βR1 inhibitor), which was recently reported to have favorable treatment efficacy in advanced HCC (NCT01246986).[50, 51] Galunisertib also reversed EMT and stemness-derived aggressiveness

in multiple preclinical experiments.[52, 53] A NOTCH inhibitor (LY3039478) whose, anti-tumor activity was proven in a murine HCC model, could be of potential use in the AMC-C2 subgroup in which NOTCH pathways are activated.[54] The upregulation of Wnt/β-catenin signaling characteristic of AMC-C3 patients may perhaps also be treatable since the specific alteration involved corresponds to the target for a treatment (PRI-724) that showed clinical efficacy in pancreatic cancer in a phase-I trial.[55] Lastly, CTA-expressing AMC-C5 tumors are good candidates for immunotherapy, since CTA genes are exclusively expressed in cancer cells and thus can be selectively surveilled by the host immune system.[56, 57] Taken together, our observations suggest that the AMC-subtypes of HCC provide unique targets treatable by targeted therapy or immunotherapy.

Comprehensive assessment of the immunologic microenvironments of HCC

We next focused on immune gene expression and infiltration of specific immune gene subsets into the HCC. Nearest template prediction classification using exemplar genes (a 112 gene expression-based immune classifier) described by Sia et al.,[9] identified "immunogenic" (80 of 206, 38.8%) and "immune-desert" (126 of 206, 61.2%) subgroups in the AMC cohorts (**Fig. 3a**). Similarly, 160 of 360 HCC samples (44.4%) from the extended TCGA-LIHC cohorts were predicted to belong to the immunogenic group.[5] The subjects in the immunogenic group were enriched relative to the immune-desert group in IFN-α and IFN-γ responses, in inflammatory responses, TGF-β signaling pathways and immune scores, as well as in up-regulation of immunomodulatory genes (*PDCD1, CD274,* and *CTLA4),* adaptive immune response genes (*GZMB, CD8A, CD3E,* and *CD3G*), NK cell activators (*GZMB, TBX21*) and *CXCL13* genes. *CTNNB1*-mutated HCCs, which have reduced immune cell infiltration in melanomas and HCC, were also discovered to be significantly related to the immune-desert group (P = 0.0009), with similar trends to the AMC-C3 subgroup (P = 0.0004) (**Fig 3a**).[8, 9, 58]

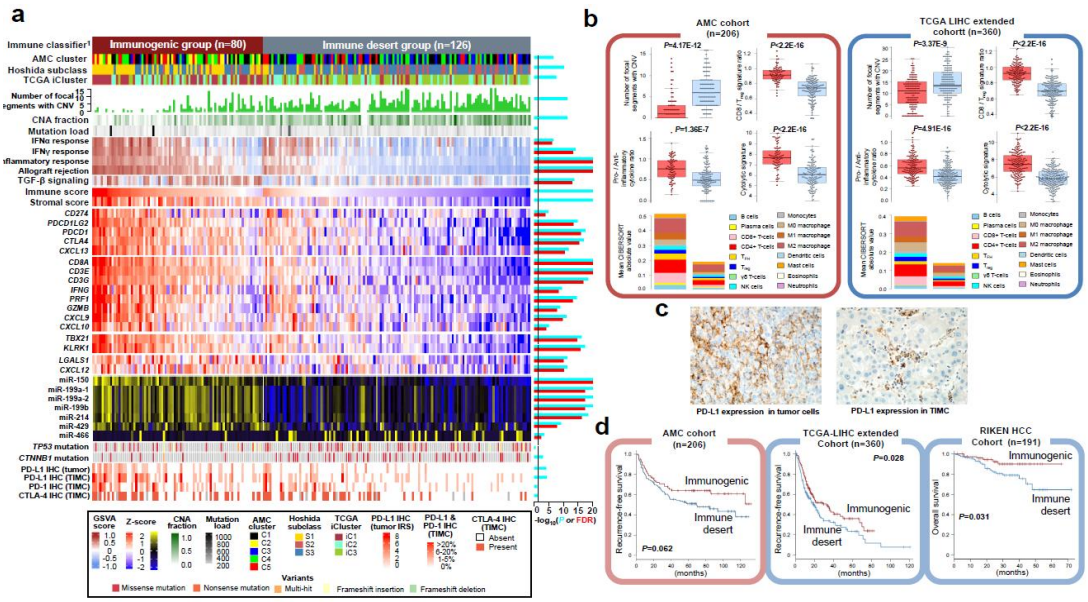


Fig 3. Mutational, histomolecular, and clinical characteristics according to immune HCC subtype. (a) Heatmaps identify an immunogenic subtype enriched in immune cells and immunophenotypic gene signatures. CNV and microRNA profiles matched with immune-related signatures are also shown in this expression heatmap. (b) Boxplot of the number of focal segments with CNV and immune-related signatures stratified by immune subtype. (c) Representative images of PD-L1 IHC staining of tumor cells and TIMC from a patient in the immunogenic group. (d) Kaplan-Meier analysis of overall survival of patients with immunogenic versus immune-desert tumors in our series and in an external validation cohort. TIMC, tumor infiltrating mononuclear cells; IHC, immunohistochemistry.

Recent experiments have demonstrated that inhibition of the Wnt/β-catenin pathways can enhance intratumoral T-cell infiltration.[58] As treatments against immune checkpoint inhibitors are showing promise in liver cancer,[59] further studies investigating combination treatments of HCC with Wnt/β-catenin inhibitors and immune checkpoint inhibitors are to be expected.

When we estimated immune cell compositions using CIBERSORT deconvolution approaches,[60] immune cell types, except for eosinophils, neutrophils, and γδ-type T-cells, were more frequent in the immunogenic group (**Fig. 3b** and **Supplementary Fig. 11**). As expected, overexpression of immune regulatory molecules such as PD-L1 in both tumor cells and tumor-infiltrating mononuclear cells revealed by immunohistochemical (IHC) methods was also observed in the immunogenic group (**Fig. 3c**). It has been proposed that tumor aneuploidy influences local immune infiltrates in several types of cancers.[20, 61] In agreement with this view, tumors of the immunogenic group had a lower burden of segments showing CNV (mean fraction in whole genome, 11.1% *vs.* 21.9%; $P = 4.07 \times 10^{-12}$). Thus AMC-C2 tumors, which have the lowest fraction of CNVs among the subgroups, are characterized by increased immune cell infiltration (**Supplementary Fig. 12**). We found that the ratios of mean mRNA levels of CD8+ T-cell–specific genes to Treg cell-specific genes, and pro-inflammatory molecules to immunosuppressive (anti-inflammatory) molecules, were significantly reduced in immune-desert tumors ($P < 2.2 \times 10^{-16}$ and $P = 1.36 \times 10^{-7}$, respectively, **Fig. 3b**). However, the tumor mutation burdens of the two groups were similar, with median numbers of mutations per tumor of 71 and 61.5 in the immunogenic and immune-desert samples, respectively ($P = 0.914$), which is not in agreement with previous observations.[62, 63] In terms of expression of miRNAs as modulators of tumor immune responses, the immune-related subtype is characterized by activation of the miR-150, miR-214 and miR-199 families and attenuation of miR-466 expression (**Fig. 3a**). miR-150 is expressed selectively in mature B- and T-cells, and its expression in the immune-related

subtype is perhaps indicative of a role in shaping the tumor micro-environment.[64, 65] In addition, numbers of NK cells, which are another key component of the innate immune response, have been reported to be determined by the miR-150 expression level.[66] miR-199a is thought to be a negative regulator of tumorigenesis and progression in epithelial ovarian cancer,[67] and miR-214 secretion from tumor cells is thought to result in the expansion of Treg.[68] This observation prompted us to examine the effect of the immune landscape of HCC on outcomes (**Fig. 3d**). Immunogenic clusters showed a clinical tendency toward longer recurrence-free survival in the AMC cohort (P = 0.062), as well as in the TCGA-LIHC extended cohorts (44.4% of patients with the immune phenotype, P = 0.028), and the same was true for overall survival among the RIKEN patients (38.7% patients belonging to the immunogenic group, P = 0.031). However, patient outcomes were not related to the prior defined subclassification[9] within our immunogenic group (i.e., adaptive *vs.* exhausted) in the AMC and extended TCGA-LIHC cohorts, which thus did not parallel the favorable outcomes of patients with an adaptive immune response in the Heptromic and core TCGA cohorts (**Supplementary Fig. 13**).[9] This inconsistency may be explained by the dynamic nature of the immunologic makeup of tumor microenvironments, with heterogeneous T-cell populations retaining the ability to mount potent effector responses.[69]

We then compared the immune landscapes of patients with [61 (29.6%)] and without recurrence of HCC [145 (70.4%)] within 2 years of resection. Signature ratios including pro- to anti-inflammatory cytokines, and M1 to M2 macrophages, were significantly lower in the subjects with early recurrence (P = 0.049 and P = 0.034 respectively, **Supplementary Fig. 14**). This suggests that the immune microenvironment of the tumors with early recurrence is more pro-tumorigenic and immunosuppressive, as also seen in other cancers.[70, 71] CIBERSORT[60] deconvolution analysis suggested that patients who did not experience early recurrence were enriched in follicular T helper cells, which is a $CD4^+$ T-cell subset believed to participate in the co-operative interplay between B-cells and T-cells that maintains

effective and sustainable antitumor immunity,[72] and in M1 macrophages, which are considered 'good' macrophages as they produce pro-inflammatory cytokines and reactive oxygen/nitrogen species crucial for host defense and tumor cell killing (**Supplementary Fig. 14**).[60, 73]

Identification of traceable genomarkers associated with early recurrence of HCC

Differential expression analysis using Wilcoxon tests identified the following genes that were differentially expressed in patients with and without early recurrence after HCC resection: *S100P, GLP1D1, OSTBETA, CPN1,* and *CXCL13* (**Fig. 4a**). Of note, *S100P* was the most overexpressed gene in patients with early recurrence, and this result was robustly validated in the GSE14520 cohort (**Fig. 4b**), whereas *CXCL13*, which is known to be mainly secreted by follicular helper T-cells and is dominant in our immunogenic subtype,[72] was the least expressed in those patients (**Fig. 4a** and **Supplementary Fig. 15**). S100P is a protein belong to the S100 protein family member implicated in several malignancies including pancreatic, breast, colorectal, prostate, lung and bile duct cancers and has important roles in tumorigenesis and cancer progression by affecting proliferation, apoptosis and metastasis.[74, 75] The area under the ROC curve (AUC) of the *S100P* mRNA for predicting early recurrence confirmed that it had prognostic utility comparable to that of serum AFP, which is used in clinical practice (0.675 *vs.* 0.638, $P = 0.435$, **Fig. 4c**). Multivariable competing risks analysis also confirmed an independent effect of *S100P* on cancer mortality (adjusted $P = 0.042$; **Supplementary Table 4**), which makes it likely to be an appealing drug target for HCC. To strengthen the potential of this marker in clinical practice, we performed an IHC analysis of *S100P* using our RNA-seq samples and found a significant correlation between its histological and transcriptomic values (**Fig. 4d** and **Fig. 4e**). By IHC, 87 patients (42.2%) were positive for S100P, which was significantly associated with early recurrence of HCC ($P = 0.037$) as was S*100P* mRNA (**Fig. 4e** and **Supplementary Fig. 16**). *S100P* was up-regulated

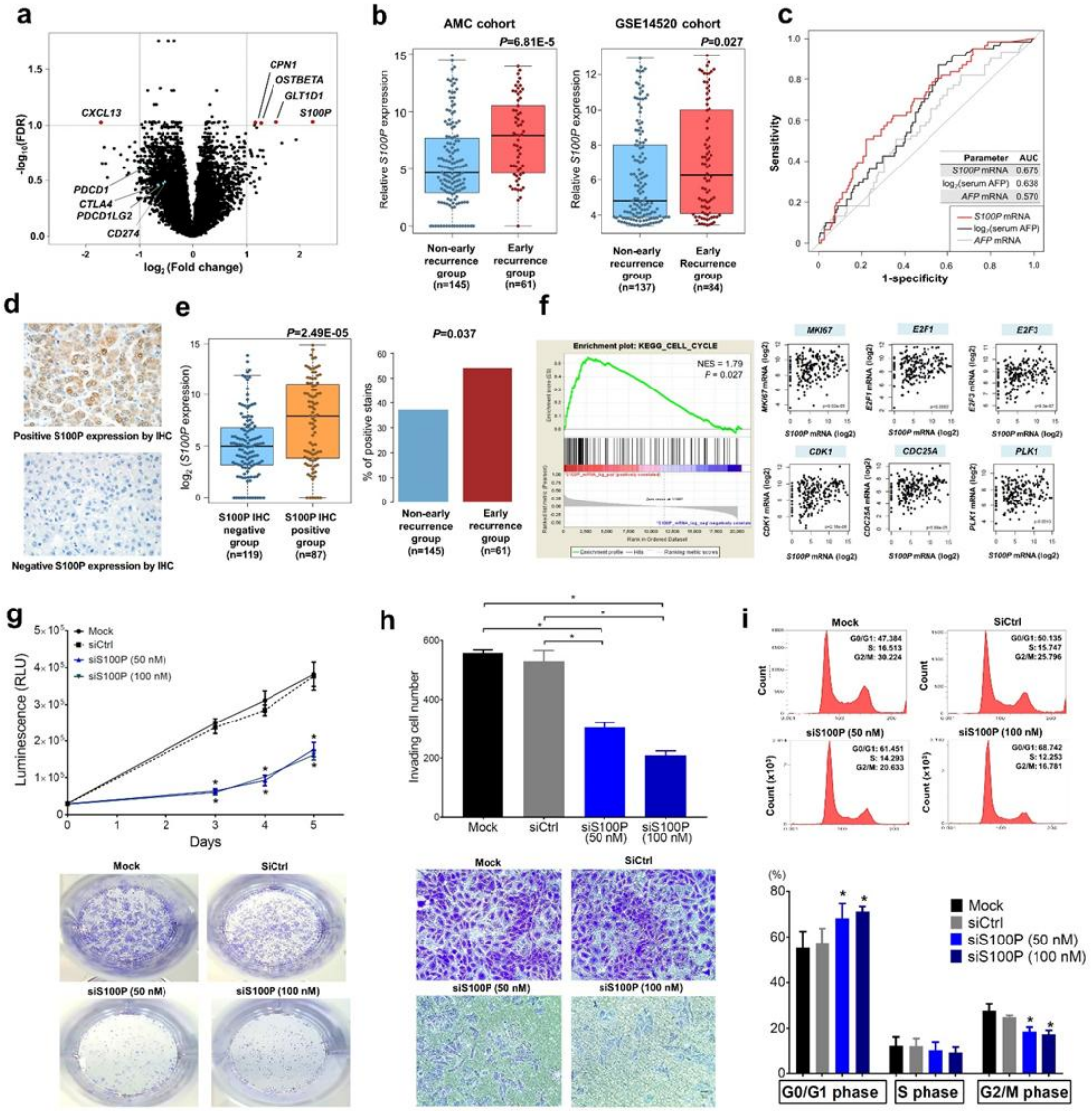


Fig 4. Molecular alterations related to early recurrence of HCC, and experimental validation of the *S100P* gene. (a) The volcano plot shows the differentially expressed genes associated with early recurrence, *S100P*, *CXCL13*, *GLT1D1*, *OSTBETA*, and *CPN1* (b) Tumors recurring early after resection are associated with a significantly higher expression of *S100P*. The GSE14520 cohort (n = 221) showed similar trend. (c) Receiver-operating characteristics plot of the predictive ability of *S100P* gene expression value for 2-year recurrence, compared to thoseof serum AFP level and tumoral *AFP* mRNA expressions. (d) Representative photos of positive and negative S100P IHC staining in the AMC dataset.

Scale bars, 400 μm. (e) Positive cases of S100P immunostaining have significantly higher levels of *S100P* mRNA expression and give rise to more early recurrence than negative cases. (f) Tumors with elevated *S100P* mRNA levels have elevated cell cycle-related signaling levels as evaluated by gene set enrichment analysis (GSEA). Cell cycle-related genes were also correlated with tumoral *S100P* expression. (g) In cell proliferation assays using cellTiter-Glo and the Crystal violet staining, SNU475 cells transfected with siRNAs against *S100P* grew more slowly than controls transfected with scrambled siRNA (siCtrl) or receiving no treatment (mock). (h) HCC cells transfected with *S100P* siRNA were less invasive than the others in invasion assays using matrigel-Boyden chambers. The quantification below shows that the siS100P-transfected HCC cells generate significantly fewer invasive cells than control cells. (i) Cell cycle analysis by flow cytometry demonstrates G1 arrest in SNU475 cells 72 h post-transfection with *S100P* siRNA. *P-value < 0.05.

mainly in the AMC-C4 subgroup, which has poor outcomes (P = 0.0002, **Supplementary Fig. 17**). The association between histological *S100P* expression (49 out of 171, 28.7%) and post-surgical early recurrence (82 out of 171, 48.0%) was reproduced in an AMC validation cohort consisting of 88.3% early (AJCC stage IA, IB or II) and 11.7% later stage (AJCC stage IIIA, IIIB or IV) tumors (P = 0.042; **Supplementary Fig. 18** and **Supplementary Table 5**).

Biological significance of *S100P* expression in HCC

We next explored whether the *S100P* expression in HCC has biological significance. To gain insight into the action of *S100P,* we investigated the potential consequence of *S100P* expression on regulatory pathways by enrichment analyses. The results showed that the samples with S100P transcriptional upregulation had well-regulated cell cycle-related pathways (**Fig. 4f**). S100P over-expressing tumors also had increased levels of cell cycle-related genes including *MKI67, E2F1, E2F3, CDK1, PLK1,* and *CDC25A*. To examine the potential function of *S100P* expression, we performed an *in vitro* knockdown experiments with a small interfering RNA (siRNA), siS100P, in cells of the cell line SNU475, which expresses *S100P* more highly than other SNU HCC cell lines (**Supplementary Fig. 19**). SiRNA-mediated targeting of *S100P* was associated with inhibition of cell proliferation, as reported in other gastrointestinal cancer cell lines (**Fig. 4g**).[74, 76] siS100P also attenuated the invasive properties of SNU475 cells (**Fig. 4h**). Moreover, examination of the cell cycle distribution by flow cytometry revealed that G0/G1 cells increased in response to *S100P* siRNA (68.1% for siS100P 50 nM, and 71.1% for siS100P 100 nM) *vs.* mock (55.0%: $Ps < 0.05$) and scrambled siRNA (57.4%: $Ps < 0.05$). (**Fig. 4i**). Taken together, our results indicate that *S100P* expression contributes to HCC tumor growth, and that inhibiting it can suppress HCC cell growth, invasion, and tumorigenesis; this may explain its prognostic value in our clinical cohort and others.

DISCUSSION

In the present study we applied multi-omic approaches to a large set of early HCC surgical samples to generate an immunomolecular classification scheme. The combinatorial interplay of genomic, transcriptomic, and immunomic architectures revealed by our results may underlie the observed biological and clinical differences between patients with HCC. It also provides signposts for identifying actionable targets, as well as reliable biomarkers for predicting early recurrence and related mortality. Subsequently, histomolecular and functional validation of the relevant transcriptomic characterization was achieved using both internal and external tumor sources.

Although previous initiatives have yielded genomic landscapes of driver events,[12, 44, 47] histopathologic or molecular gene expression-based stratification systems for HCC are not yet being taken into account in clinical management of the disease. Our new taxonomy of the HCC, based on profound molecular and clinicopathologic features, comprises five subgroups, as follows: C1 (well-differentiated HCC), C2 (high EMT with immune predominance), C3 (*CTNNB1*-mutated HCC), C4 (stem cell features with poor prognosis), and C5 (CTA group) (**Fig. 5**). The genomic alterations and biological differences between these subtypes could be used in clinical decision-making based on prognostic factors and rational therapeutic opportunities. We emphasize that our findings require further validation.

Recent genomic and transcriptomic analyses have revealed that subtypes characterized by extensive immune infiltration are present in several types of cancer.[9, 77-79] Approximately one third of our HCC were assigned to the "immunogenic" subtype, and we showed that this immune infiltration was correlated with a more favorable prognosis, preferentially in AMC-C4 tumors. In addition, the presence of specific immune cell types (*i.e.*, follicular helper T-cell and M1 macrophages) was associated with reduced recurrence, potentially reflecting the survival advantage of immune activation.[60, 70] It is of note that our combined CNV and miRNA analysis suggests that chromosomal instability and regulatory factors, not mutational

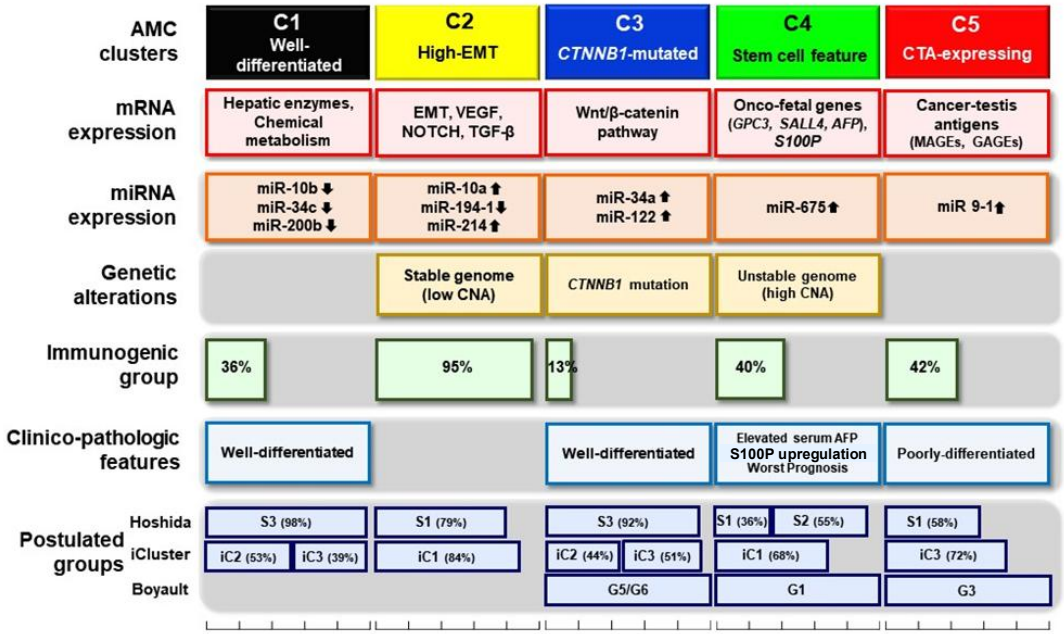


Fig 5. Proposed taxonomy of HCC reflecting significant differences in biological, immunological, and clinical patterns in the gene expression-based molecular classification.

load, shape anti-cancer immune environments. Indeed, recent pan-cancer data indicate that major somatic CNV correlates with poor immune infiltration and poor response to immune-checkpoint inhibition.[20] A growing body of literature has demonstrated the importance of immune controls involving epigenetic and miRNA mechanisms in the tumor microenvironment,[80, 81] as do our findings. Although we have identified broad aspects of these controls, the nature of the specific functional networks acting in tumor, immune, and stromal cell compartments requires further investigation.

There are few clinically useful cancer biomarkers for predicting early recurrence after hepatectomy in patients with HCC. Based on our RNA-seq and immunohistochemical results we showed that *S100P*, which is preferentially expressed in the AMC-C4 subgroup with a poor prognosis, may be a predictor of unfavorable events, comparable to the traditional serum oncomarker, AFP. Aberrant expression of *S100P* has been found in various types of cancer, and its overexpression is thought to be associated with a poor clinical outcome.[74, 75] There is also evidence that S100 proteins contribute to tumorigenic processes such as cell proliferation, disturbance of cell cycle control, and metastasis.[74, 75, 82] In addition, a small peptide antagonist of receptor for advanced glycation end products (RAGE) appears to disrupt the S100P–RAGE interaction, which activates key signaling pathways such as the ERK, NF-kB, and the JAK/STAT pathways, eventually reducing tumor growth and inhibiting metastasis.[76, 83] Our nosological catalog of molecular alterations in HCC provides an opportunity to identify novel prognostic and therapeutic candidates such as *S100P* for future functional and mechanistic investigations.

CONCLUSION

This study provides a comprehensive overview of the global orchestration of immunomolecular and mutational events that influence tumor behavior and patient outcomes in HCC. This pathobiologic insight may serve as the basis for more targeted treatment of HCC.

REFERENCES

1. Jemal A, Ward EM, Johnson CJ, Cronin KA, Ma J, Ryerson B, et al. Annual Report to the Nation on the Status of Cancer, 1975-2014, Featuring Survival. J Natl Cancer Inst 2017;109(9).
2. Bertuccio P, Turati F, Carioli G, Rodriguez T, La Vecchia C, Malvezzi M, et al. Global trends and predictions in hepatocellular carcinoma mortality. J Hepatol 2017;67(2):302-9.
3. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. Lancet 2018;391(10127):1301-14.
4. Hasegawa K, Kokudo N, Makuuchi M, Izumi N, Ichida T, Kudo M, et al. Comparison of resection and ablation for hepatocellular carcinoma: a cohort study based on a Japanese nationwide survey. J Hepatol 2013;58(4):724-9.
5. Cancer Genome Atlas Research Network. Electronic address wbe, Cancer Genome Atlas Research N. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell 2017;169(7):1327-41 e23.
6. Hoshida Y, Nijman SM, Kobayashi M, Chan JA, Brunet JP, Chiang DY, et al. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. Cancer Res 2009;69(18):7385-92.
7. Boyault S, Rickman DS, de Reynies A, Balabaud C, Rebouissou S, Jeannot E, et al. Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. Hepatology 2007;45(1):42-52.
8. Calderaro J, Couchy G, Imbeaud S, Amaddeo G, Letouze E, Blanc JF, et al. Histological subtypes of hepatocellular carcinoma are related to gene mutations and molecular tumour classification. J Hepatol 2017;67(4):727-38.
9. 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(3):812-26.

10. Roessler S, Jia HL, Budhu A, Forgues M, Ye QH, Lee JS, et al. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. Cancer Res 2010;70(24):10202-12.

11. Yamashita T, Ji J, Budhu A, Forgues M, Yang W, Wang HY, et al. EpCAM-positive hepatocellular carcinoma cells are tumor-initiating cells with stem/progenitor cell features. Gastroenterology 2009;136(3):1012-24.

12. Ahn SM, Jang SJ, Shim JH, Kim D, Hong SM, Sung CO, et al. Genomic portrait of resectable hepatocellular carcinomas: implications of RB1 and FGF19 aberrations for patient stratification. Hepatology 2014;60(6):1972-82.

13. Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL, et al. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res 2010;38(18):e178.

14. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 2011;12:323.

15. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics 2010;11:367.

16. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43(7):e47.

17. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 2013;14:7.

18. Yoshihara K, Shahmoradgoli M, Martinez 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.

19. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust

enumeration of cell subsets from tissue expression profiles. Nat Methods 2015;12(5):453-7.

20. 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(6322).

21. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SG, Hoadley KA, et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat Genet 2015;47(10):1168-78.

22. Mayakonda A, Koeffler HP. Maftools: Efficient analysis, visualization and summarization of MAF files from large-scale cohort based cancer studies. bioRxiv 2016:052662.

23. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature 2013;500(7463):415-21.

24. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol 2016;17:31.

25. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol 2011;12(4):R41.

26. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics 2007;23(14):1846-7.

27. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007;8(1):118-27.

28. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 2010;33(1):1-22.

29. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12:77.

30. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44(3):837-45.

31. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association 1999;94(446):496-509.

32. Gray RJ. A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk. Annals of Statistics 1988;16(3):1141-54.

33. Amin MB, American Joint Committee on Cancer., American Cancer Society. AJCC cancer staging manual. (Eight edition / editor-in-chief, Mahul B. Amin, MD, FCAP ; editors, Stephen B. Edge, MD, FACS and 16 others ; Donna M. Gress, RHIT, CTR - Technical editor ; Laura R. Meyer, CAPM - Managing editor. ed). Chicago IL: American Joint Committee on Cancer, Springer; (2017).

34. Tang H, Deng M, Tang Y, Xie X, Guo J, Kong Y, et al. miR-200b and miR-200c as prognostic factors and mediators of gastric cancer cell progression. Clin Cancer Res 2013;19(20):5602-12.

35. Yu Y, Wu J, Guan L, Qi L, Tang Y, Ma B, et al. Kindlin 2 promotes breast cancer invasion via epigenetic silencing of the microRNA200 gene family. Int J Cancer 2013;133(6):1368-79.

36. David CJ, Huang YH, Chen M, Su J, Zou Y, Bardeesy N, et al. TGF-beta Tumor Suppression through a Lethal EMT. Cell 2016;164(5):1015-30.

37. Fu J, Chen Y, Cao J, Luo T, Qian YW, Yang W, et al. p28GANK overexpression accelerates hepatocellular carcinoma invasiveness and metastasis via

phosphoinositol 3-kinase/AKT/hypoxia-inducible factor-1alpha pathways. Hepatology 2011;53(1):181-92.

38. Yan Y, Wang Q, Yan XL, Zhang Y, Li W, Tang F, et al. miR-10a controls glioma migration and invasion through regulating epithelial-mesenchymal transition via EphA8. FEBS Lett 2015;589(6):756-65.

39. Wu X, Liu T, Fang O, Leach LJ, Hu X, Luo Z. miR-194 suppresses metastasis of non-small cell lung cancer through regulating expression of BMP1 and p27(kip1). Oncogene 2014;33(12):1506-14.

40. Gougelet A, Sartor C, Bachelot L, Godard C, Marchiol C, Renault G, et al. Antitumour activity of an inhibitor of miR-34a in liver cancer with beta-catenin-mutations. Gut 2016;65(6):1024-34.

41. Girard M, Jacquemin E, Munnich A, Lyonnet S, Henrion-Caude A. miR-122, a paradigm for the role of microRNAs in the liver. J Hepatol 2008;48(4):648-56.

42. Ho DWH, Chan LK, Chiu YT, Xu IMJ, Poon RTP, Cheung TT, et al. TSC1/2 mutations define a molecular subset of HCC with aggressive behaviour and treatment implication. Gut 2017;66(8):1496-506.

43. He F, Li J, Xu J, Zhang S, Xu Y, Zhao W, et al. Decreased expression of ARID1A associates with poor prognosis and promotes metastases of hepatocellular carcinoma. J Exp Clin Cancer Res 2015;34:47.

44. Huang J, Deng Q, Wang Q, Li KY, Dai JH, Li N, et al. Exome sequencing of hepatitis B virus-associated hepatocellular carcinoma. Nat Genet 2012;44(10):1117-21.

45. Hernandez JM, Elahi A, Clark CW, Wang J, Humphries LA, Centeno B, et al. miR-675 mediates downregulation of Twist1 and Rb in AFP-secreting hepatocellular carcinoma. Ann Surg Oncol 2013;20 Suppl 3:S625-35.

46. Lujambio A, Calin GA, Villanueva A, Ropero S, Sanchez-Cespedes M, Blanco D, et

al. A microRNA DNA methylation signature for human cancer metastasis. Proc Natl Acad Sci U S A 2008;105(36):13556-61.

47. Schulze K, Imbeaud S, Letouze E, Alexandrov LB, Calderaro J, Rebouissou S, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nat Genet 2015;47(5):505-11.

48. Fujimoto A, Furuta M, Totoki Y, Tsunoda T, Kato M, Shiraishi Y, et al. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. Nat Genet 2016;48(5):500-9.

49. Roessler S, Long EL, Budhu A, Chen Y, Zhao X, Ji J, et al. Integrative genomic identification of genes on 8p associated with hepatocellular carcinoma progression and patient survival. Gastroenterology 2012;142(4):957-66 e12.

50. Fransvea E, Mazzocca A, Santamato A, Azzariti A, Antonaci S, Giannelli G. Kinase activation profile associated with TGF-beta-dependent migration of HCC cells: a preclinical study. Cancer Chemother Pharmacol 2011;68(1):79-86.

51. Herbertz S, Sawyer JS, Stauber AJ, Gueorguieva I, Driscoll KE, Estrem ST, et al. Clinical development of galunisertib (LY2157299 monohydrate), a small molecule inhibitor of transforming growth factor-beta signaling pathway. Drug Des Devel Ther 2015;9:4479-99.

52. Rani B, Malfettone A, Dituri F, Soukupova J, Lupo L, Mancarella S, et al. Galunisertib suppresses the staminal phenotype in hepatocellular carcinoma by modulating CD44 expression. Cell Death Dis 2018;9(3):373.

53. Yingling JM, McMillen WT, Yan L, Huang H, Sawyer JS, Graff J, et al. Preclinical assessment of galunisertib (LY2157299 monohydrate), a first-in-class transforming growth factor-beta receptor type I inhibitor. Oncotarget 2018;9(6):6659-77.

54. Tijeras-Raballand A, Hobeika C, Martinet M, Paven E, Bonin P, Eveno C, et al. Abstract B144: LY3039478, a novel Notch inhibitor, potentiates the antitumor

effects of sorafenib in hepatocellular carcinoma (HCC). Molecular Cancer Therapeutics 2018;17(1):Supplement.

55. Ko AH, Chiorean EG, Kwak EL, Lenz H-J, Nadler PI, Wood DL, et al. Final results of a phase Ib dose-escalation study of PRI-724, a CBP/beta-catenin modulator, plus gemcitabine (GEM) in patients with advanced pancreatic adenocarcinoma (APC) as second-line therapy after FOLFIRINOX or FOLFOX. Journal of Clinical Oncology 2016;34(16):suppl.

56. Karpf AR. A potential role for epigenetic modulatory drugs in the enhancement of cancer/germ-line antigen vaccine efficacy. Epigenetics 2006;1(3):116-20.

57. Jones PA, Issa JP, Baylin S. Targeting the cancer epigenome for therapy. Nat Rev Genet 2016;17(10):630-41.

58. Spranger S, Bao R, Gajewski TF. Melanoma-intrinsic beta-catenin signalling prevents anti-tumour immunity. Nature 2015;523(7559):231-5.

59. El-Khoueiry AB, Sangro B, Yau T, Crocenzi TS, Kudo M, Hsu C, 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;389(10088):2492-502.

60. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med 2015;21(8):938-45.

61. 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(4):676-89 e3.

62. Ock CY, Hwang JE, Keam B, Kim SB, Shim JJ, Jang HJ, et al. Genomic landscape associated with potential response to anti-CTLA-4 treatment in cancers. Nat Commun 2017;8(1):1050.

63. Hellmann MD, Callahan MK, Awad MM, Calvo E, Ascierto PA, Atmaca A, et al. Tumor Mutational Burden and Efficacy of Nivolumab Monotherapy and in Combination with Ipilimumab in Small-Cell Lung Cancer. Cancer Cell 2018;33(5):853-61 e4.

64. Xiao C, Calado DP, Galler G, Thai TH, Patterson HC, Wang J, et al. MiR-150 controls B cell differentiation by targeting the transcription factor c-Myb. Cell 2007;131(1):146-59.

65. Zhou B, Wang S, Mayr C, Bartel DP, Lodish HF. miR-150, a microRNA expressed in mature B and T cells, blocks early B cell development when expressed prematurely. Proc Natl Acad Sci U S A 2007;104(17):7080-5.

66. Bezman NA, Chakraborty T, Bender T, Lanier LL. miR-150 regulates the development of NK and iNKT cells. J Exp Med 2011;208(13):2717-31.

67. Li SD, Zhang JR, Wang YQ, Wan XP. The role of microRNAs in ovarian cancer initiation and progression. J Cell Mol Med 2010;14(9):2240-9.

68. Yin Y, Cai X, Chen X, Liang H, Zhang Y, Li J, et al. Tumor-secreted miR-214 induces regulatory T cells: a major link between immune evasion and tumor growth. Cell Res 2014;24(10):1164-80.

69. Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. Nature 2017;541(7637):321-30.

70. Fridman WH, Pages F, Sautes-Fridman C, Galon J. The immune contexture in human tumours: impact on clinical outcome. Nat Rev Cancer 2012;12(4):298-306.

71. Becht E, de Reynies A, Giraldo NA, Pilati C, Buttard B, Lacroix L, et al. Immune and Stromal Classification of Colorectal Cancer Is Associated with Molecular Subtypes and Relevant for Precision Immunotherapy. Clin Cancer Res 2016;22(16):4057-66.

72. Crotty S. T follicular helper cell differentiation, function, and roles in disease.

Immunity 2014;41(4):529-42.

73. Sica A, Schioppa T, Mantovani A, Allavena P. Tumour-associated macrophages are a distinct M2 polarised population promoting tumour progression: potential targets of anti-cancer therapy. Eur J Cancer 2006;42(6):717-27.

74. Arumugam T, Simeone DM, Van Golen K, Logsdon CD. S100P promotes pancreatic cancer growth, survival, and invasion. Clin Cancer Res 2005;11(15):5356-64.

75. Prica F, Radon T, Cheng Y, Crnogorac-Jurcevic T. The life and works of S100P - from conception to cancer. Am J Cancer Res 2016;6(2):562-76.

76. Mercado-Pimentel ME, Onyeagucha BC, Li Q, Pimentel AC, Jandova J, Nelson MA. The S100P/RAGE signaling pathway regulates expression of microRNA-21 in colon cancer cells. FEBS Lett 2015;589(18):2388-93.

77. Bijlsma MF, Sadanandam A, Tan P, Vermeulen L. Molecular subtypes in cancers of the gastrointestinal tract. Nat Rev Gastroenterol Hepatol 2017;14(6):333-42.

78. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The Immune Landscape of Cancer. Immunity 2018;48(4):812-30 e14.

79. Miller LD, Chou JA, Black MA, Print C, Chifman J, Alistar A, et al. Immunogenic Subtypes of Breast Cancer Delineated by Gene Classifiers of Immune Responsiveness. Cancer Immunol Res 2016;4(7):600-10.

80. Chiappinelli KB, Zahnow CA, Ahuja N, Baylin SB. Combining Epigenetic and Immunotherapy to Combat Cancer. Cancer Res 2016;76(7):1683-9.

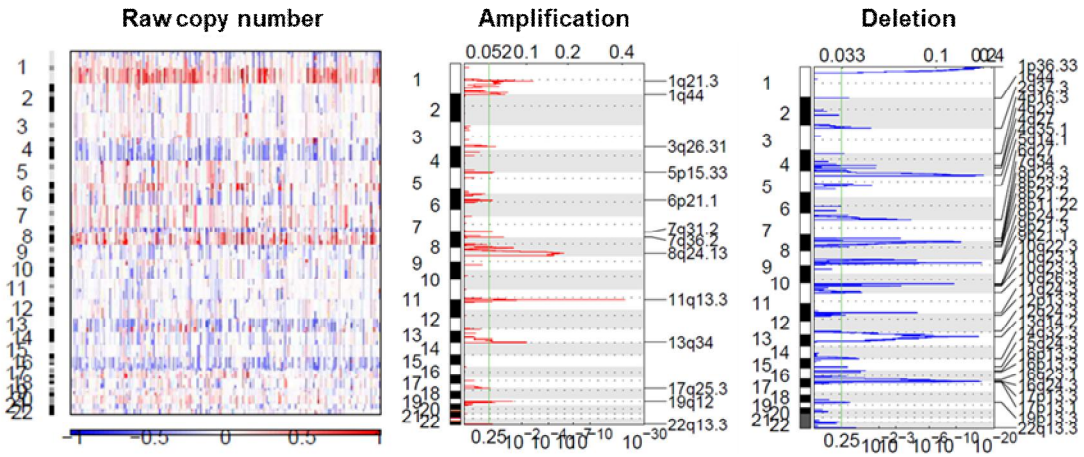
81. O'Connell RM, Rao DS, Chaudhuri AA, Baltimore D. Physiological and pathological roles for microRNAs in the immune system. Nat Rev Immunol 2010;10(2):111-22.

82. Wu Z, Boonmars T, Nagano I, Boonjaraspinyo S, Srinontong P, Ratasuwan P, et al. Significance of S100P as a biomarker in diagnosis, prognosis and therapy of

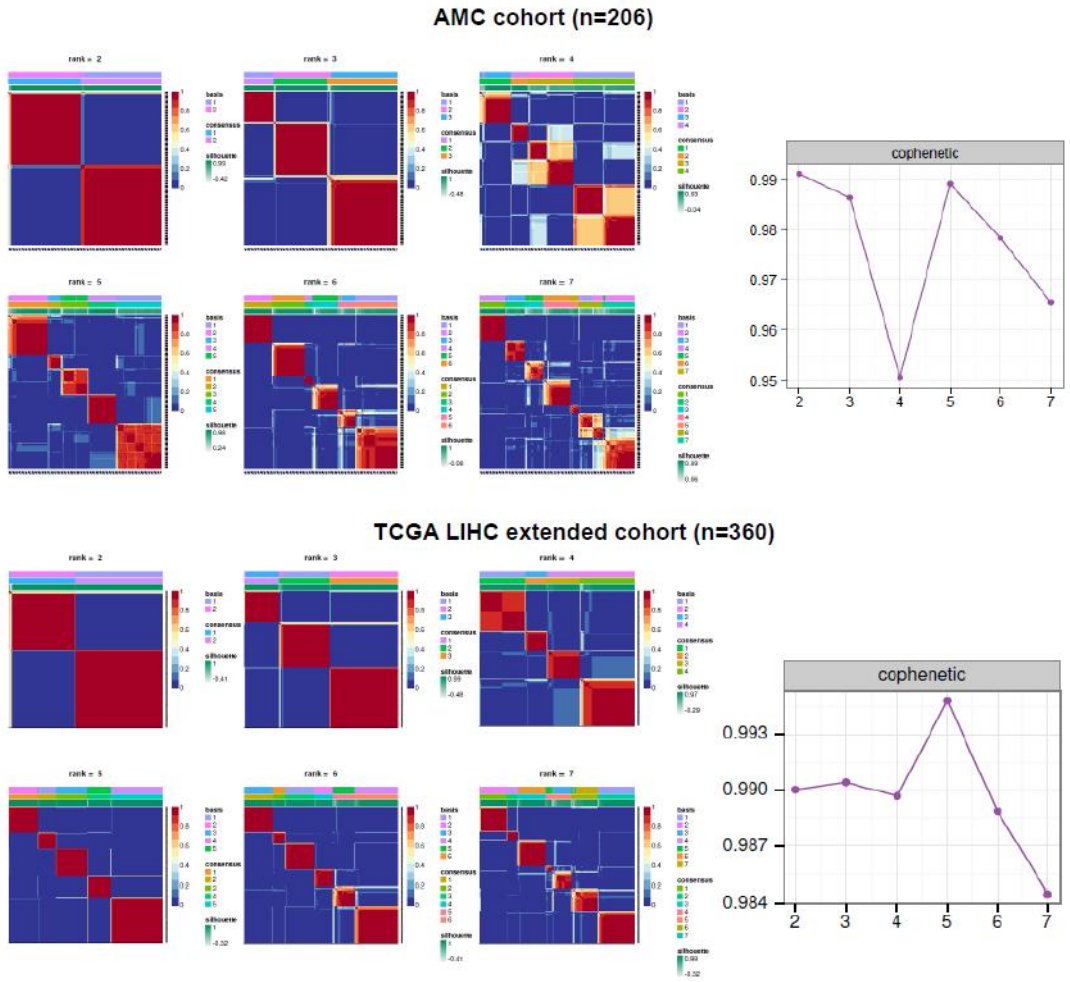
opisthorchiasis-associated cholangiocarcinoma. Int J Cancer 2016;138(2):396-408.

83. Arumugam T, Ramachandran V, Logsdon CD. Effect of cromolyn on S100P interactions with RAGE and pancreatic cancer growth and invasion in mouse models. J Natl Cancer Inst 2006;98(24):1806-18.

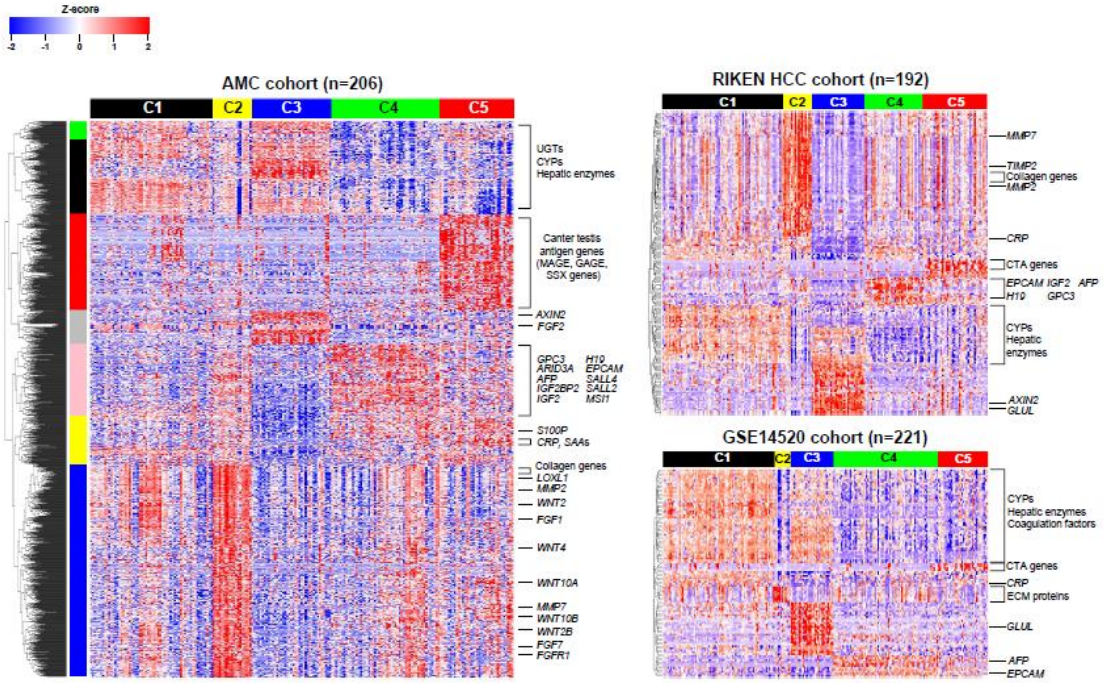
SUPPLEMENTARY FIGURES



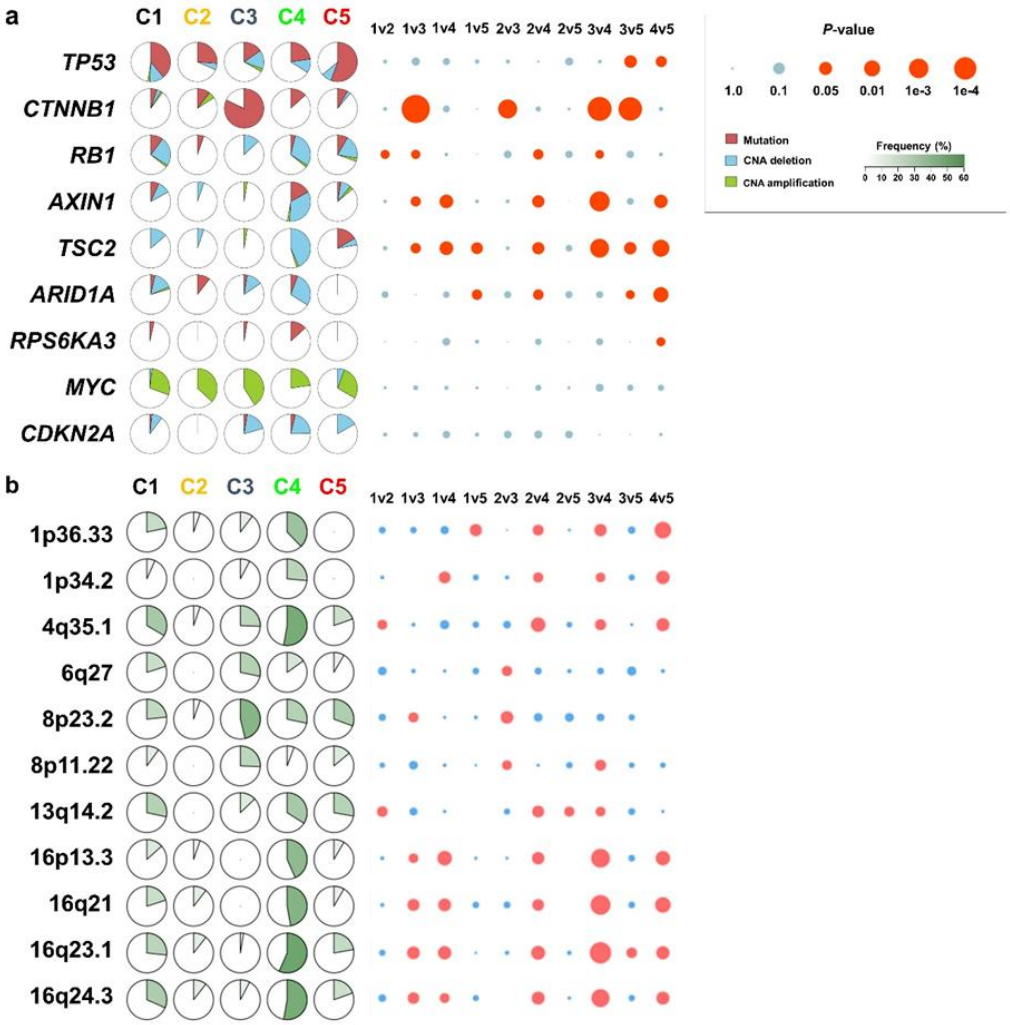
Supplementary Figure 1. GISTIC2 analysis results for the copy number alterations of AMC dataset (n=206). Plots for raw copy number heatmap, amplification and deletion peaks are presented. Confidence interval of 0.99 and absolute log2 ratio of 0.4 were used as cutoffs during analysis.



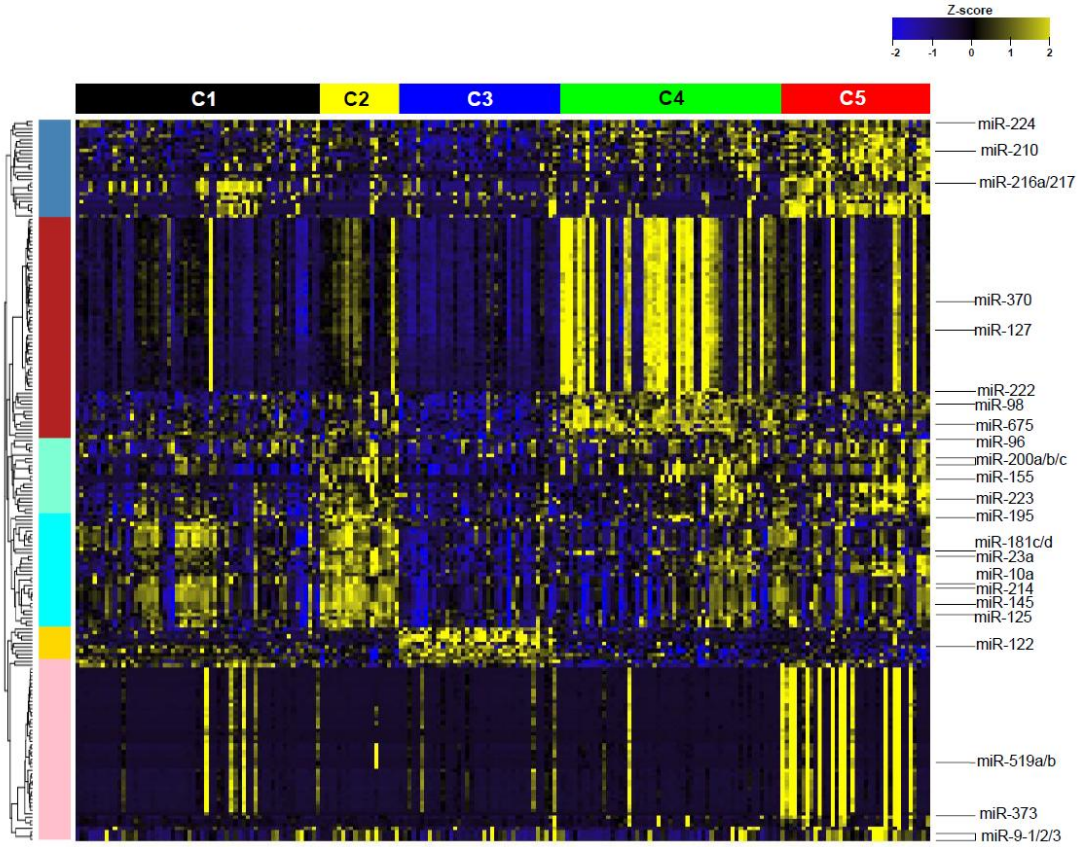
Supplementary Figure 2. Results of unsupervised consensus clustering of the RNA-seq gene expression datasets of the AMC (n = 206) and TCGA-LIHC extended (n = 360) cohorts. Blue/red heatmaps for the consensus matrix of each rank (k = 2 to 7) and the corresponding cophenetic values are shown. In the AMC dataset, the cophenetic values in ranks 2, 3 and 5 are higher than in the other ranks. In the TCGA-LIHC dataset, the cophenetic value in rank=5 is higher than in the other ranks. These results suggest that rank 5 is the optimal number for both cohorts.



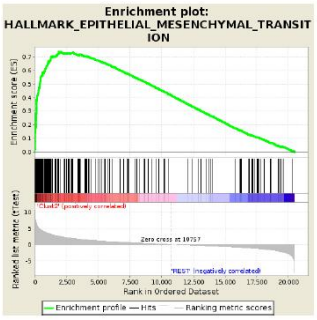
Supplementary Figure 3. Clustering analysis of the AMC and public validation cohorts. Heatmaps of differentially expressed genes (DEGs) with a false discovery rate < 0.05 in the five AMC clusters in the three different gene expression datasets. DEGs with an absolute $\log_2$-fold change (logFC) > 1.5 were retrieved for the heatmaps of the AMC dataset. For the other datasets, AMC clusters were predicted using a multinomial logistic regression model with elastic net regularization (see **Methods**), and DEGs with an absolute logFC > 2 were selected for the heat maps. AMC-C1 and C3 exhibit significant upregulation of genes for hepatocyte enzymes, AMC-C2 exhibits a significant enrichment of epithelial-mesenchymal transition (EMT)-related genes, and AMC-C3 shows increased expression of β-catenin signaling regulators (*AXIN2*, *FGF2*). AMC-C4 exhibits upregulation of oncofetal genes and *EPCAM,* and AMC-C5 has elevated expression of cancer-testis antigen genes. These characteristic features of the AMC clusters were also evident in separate gene expression validation datasets including the RIKEN HCC and GSE14520 cohorts.



Supplementary Figure 4. Pie charts of (a) somatic mutations and (b) focal copy number deletions in the AMC subgroups.

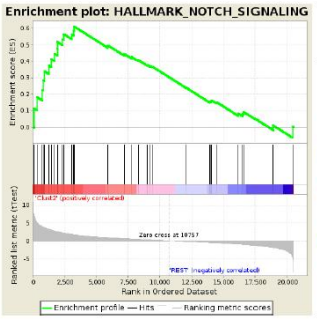


Supplementary Figure 5. Heatmap of differentially expressed microRNAs (miRNAs) with an absolute logFC > 1 in the five AMC subgroups. MicroRNAs associated with EMT regulation (miR-10a, miR-214, miR-125, miR-145, and miR-223), Wnt/β-catenin regulator miR-122 and oncofetal miRNAs miR-98 and miR-675 are elevated in subsets of the patients in AMC-C2, C4, and C5, respectively.



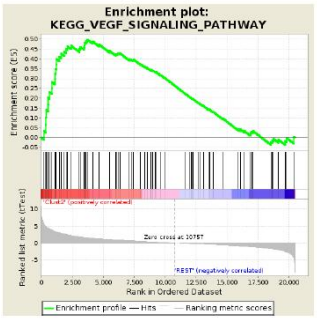
Normalized ES	2.064
Nominal *P*-value	<0.001
FDR q-value	0.046

	Rank in gene list	Rank metric score
LOXL1	6	9.832
MMP2	17	8.705
CDH11	34	8.099
COL1A1	69	7.118
COL1A2	90	6.894
CD44	420	4.794
PDGFRB	544	4.445
TGFB1	928	3.704
SNAI2	1144	3.426
VIM	1307	3.229



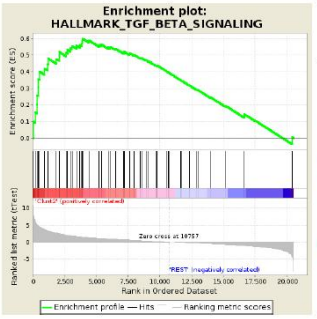
Normalized ES	1.797
Nominal *P*-value	0.002
FDR q-value	0.043

	Rank in gene list	Rank metric score
HEYL	59	7.379
NOTCH3	314	5.161
DLL1	714	4.098
LFNG	803	3.924
JAG1	884	3.783
FZD1	1250	3.300
WNT2	1441	3.065
FZD7	1912	2.652
NOTCH1	2432	2.259
HES1	3213	1.810



Normalized ES	1.775
Nominal *P*-value	<0.001
FDR q-value	0.043

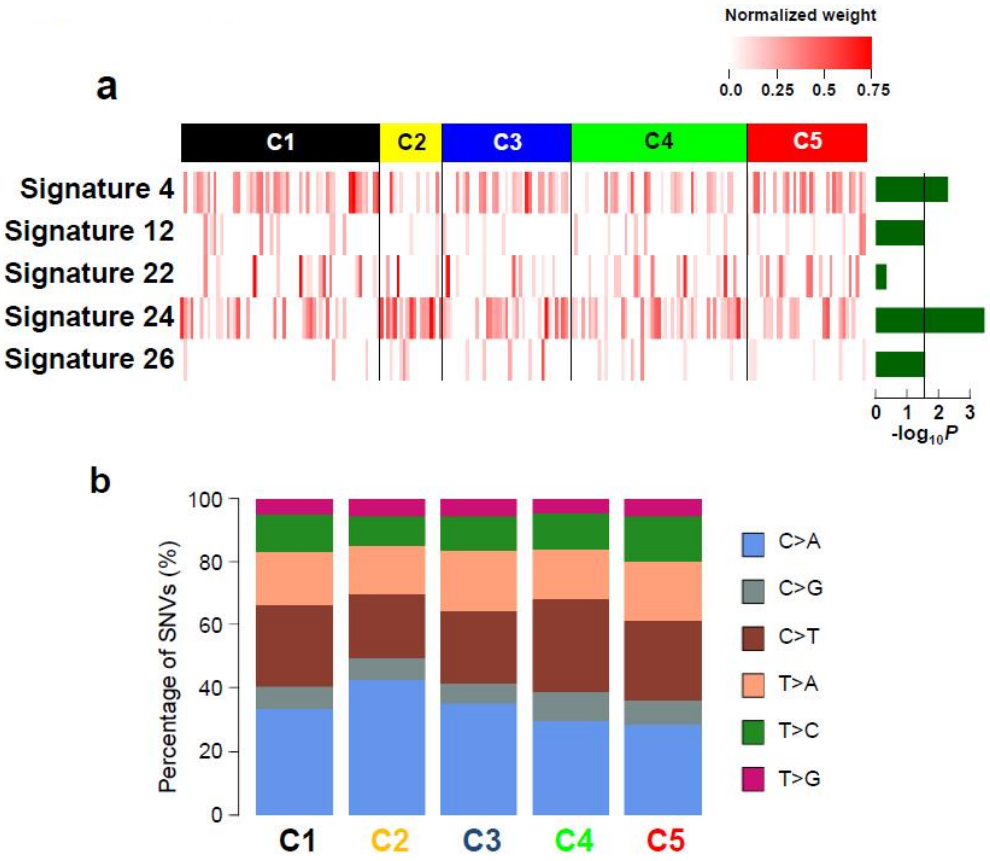
	Rank in gene list	Rank metric score
NFATC1	227	5.662
NOS3	378	4.923
PLA2G5	398	4.870
MAPK11	554	4.420
AKT3	861	3.817
MAPK13	1233	3.321
MAPK12	1237	3.312
SPHK1	1488	3.024
KDR	1622	2.899
SRC	2411	2.273



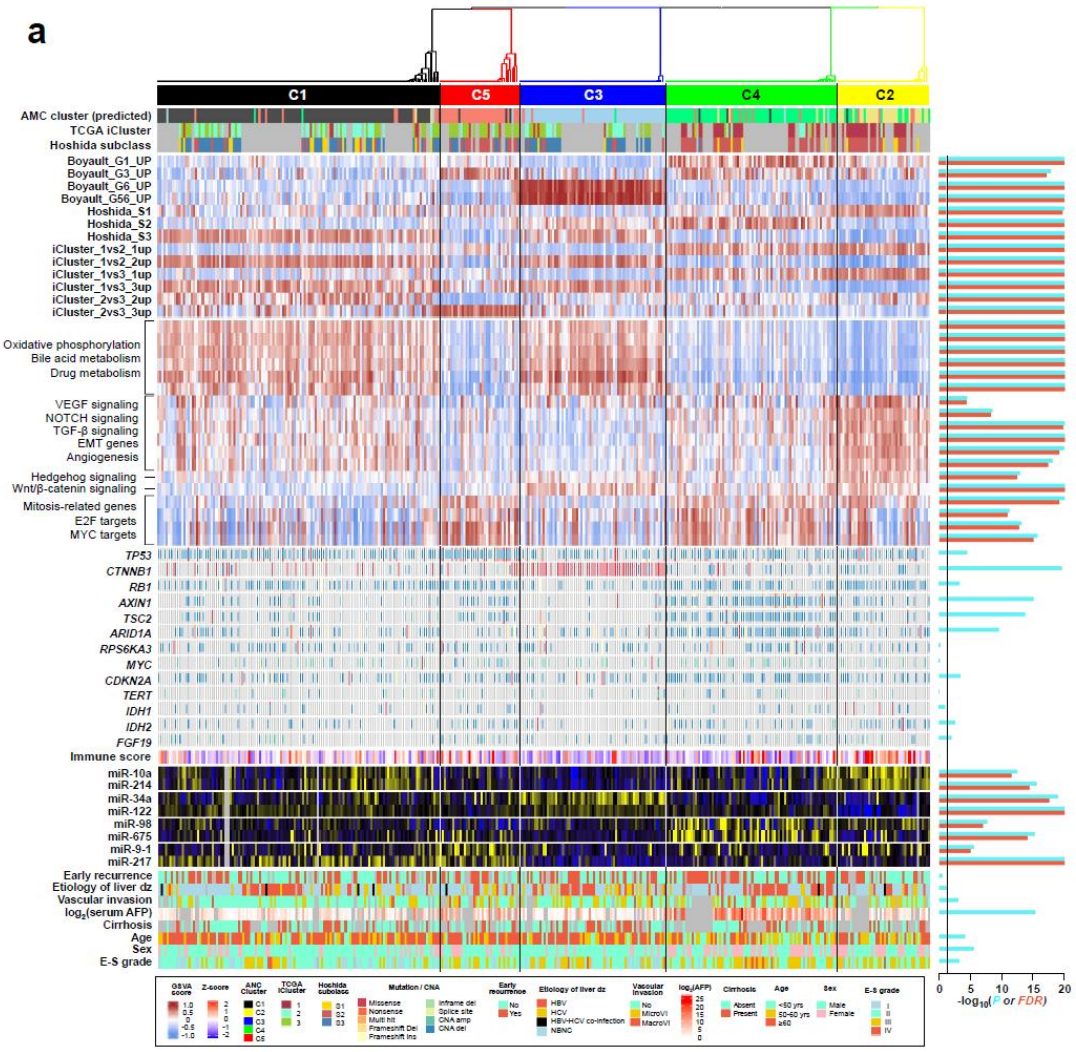
Normalized ES	1.762
Nominal *P*-value	0.002
FDR q-value	0.045

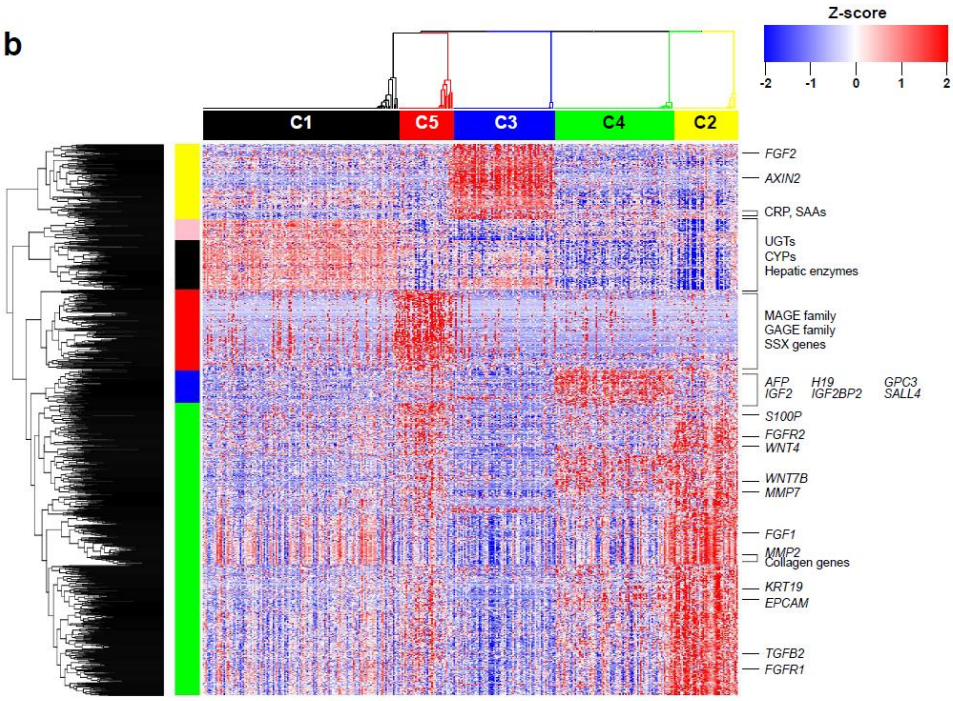
	Rank in gene list	Rank metric score
LTBP2	9	9.416
THBS1	181	5.949
ID3	336	5.081
ENG	413	4.815
RAB31	505	4.542
TGFB1	928	3.704
ID1	1133	3.437
JUNB	2069	2.513
BMPR2	3690	1.591
TGFBR1	3856	1.530

Supplementary Figure 6. Gene set enrichment analysis (GSEA) in AMC-C2 HCCs versus the others. Gene sets associated with EMT, NOTCH, VEGF, and TGF-β signaling pathways are enriched in the AMC-C2 tumors. The genes listed are the top 10 highly enriched genes in the corresponding gene sets.

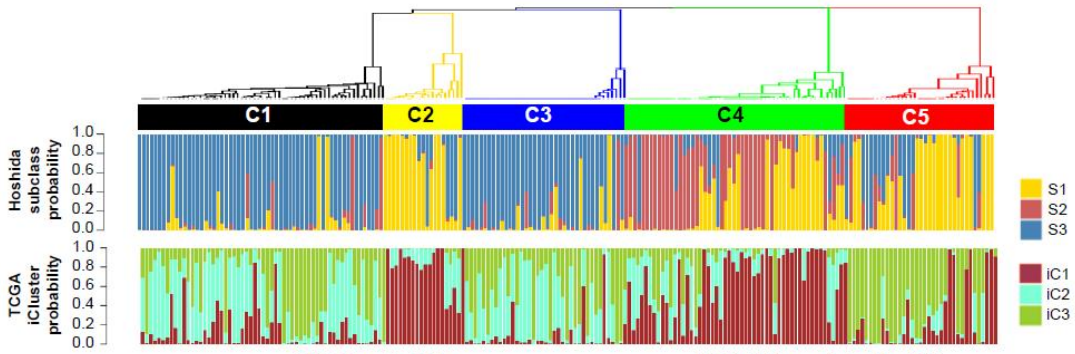


Supplementary Figure 7. Patterns of mutational signatures in the AMC dataset. (a) Distributions of mutational signatures according to the AMC subgroups. The normalized weight of signature 24, known to be related to aflatoxin-associated liver cancer, is significantly elevated in AMC-C2 (P = 0.0007). (B) Stacked bar plots showing the mean proportions of single nucleotide transitions and transversions in the AMC clusters. There is no significant difference in the proportions across the AMC clusters. SNV, single nucleotide variations.


a
C1
C5
C3
C4
C2
AMC cluster (predicted)
TCGA iCluster
Hoshida subclass
Boyault_G1_UP
Boyault_G3_UP
Boyault_G6_UP
Boyault_G56_UP
Hoshida_S1
Hoshida_S2
Hoshida_S3
iCluster_1vs2_1up
iCluster_1vs2_2up
iCluster_1vs3_1up
iCluster_1vs3_3up
iCluster_2vs3_2up
iCluster_2vs3_3up
Oxidative phosphorylation
Bile acid metabolism
Drug metabolism
VEGF signaling
NOTCH signaling
TGF-β signaling
EMT genes
Angiogenesis
Hedgehog signaling
Wnt/β-catenin signaling
Mitosis-related genes
E2F targets
MYC targets
TP53
CTNNB1
RB1
AXIN1
TSC2
ARID1A
RPS6KA3
MYC
CDKN2A
TERT
IDH1
IDH2
FGF19
Immune score
miR-10a
miR-214
miR-34a
miR-122
miR-98
miR-675
miR-9-1
miR-217
Early recurrence
Etiology of liver dz
Vascular invasion
log2(serum AFP)
Cirrhosis
Age
Sex
E-S grade
0
5
10
15
20
-log10(P or FDR)



Supplementary Figure 8. Identification of the five molecular tumor classes (AMC subgroups) in the TCGA-LIHC extended cohorts (n = 360). (a) Heatmap showing the gene set scores of the HCC subgroups and the relevant signaling pathways. The cluster labeling located most upper portion denotes unsupervised consensus clustering using NMF algorithm. The strip plots below the cluster labeling show the AMC clusters predicted by a multinomial penalized logistic regression model based on the gene expression data. The clusters separated by unsupervised consensus clustering are closely correlated with the predicted AMC subgroups and exhibit similar characteristics to those of the AMC subgroups in terms of gene set enrichment, gene mutations, copy number reductions and miRNA expression. (b) Heatmap depicting the DEGs between the five-tiered consensus clusters in the TCGA LIHC dataset. DEGs with an absolute logFC value > 1.5 and a false discovery rate (FDR) < 0.05 were used for the heatmaps.



AMC dataset (n=206)

	AMC-C1	AMC-C2	AMC-C3	AMC-C4	AMC-C5	Total
Hoshida S1	3	15	2	19	21	60
Hoshida S2	4	0	1	29	5	39
Hoshida S3	52	4	36	5	10	107
Total	59	19	39	53	36	206

Chi-square *P*<2.2E-16

	AMC-C1	AMC-C2	AMC-C3	AMC-C4	AMC-C5	Total
iCluster 1	5	16	2	36	8	67
iCluster 2	31	3	17	10	2	63
iCluster 3	23	0	20	7	26	76
Total	59	19	39	53	36	206

Chi-square *P*<2.2E-16

TCGA LIHC core dataset (n=191)

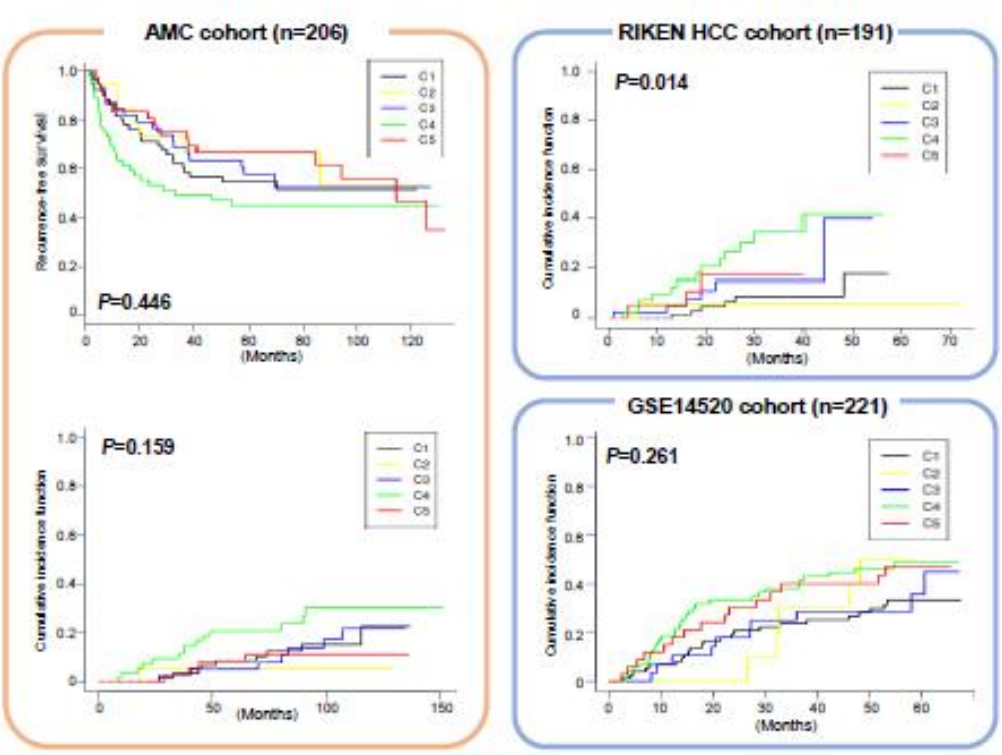
	AMC-C1	AMC-C2	AMC-C3	AMC-C4	AMC-C5	Total
Hoshida S1	15	5	2	4	0	26
Hoshida S2	11	20	6	32	17	86
Hoshida S3	45	0	25	1	8	79
Total	71	25	33	37	25	191

Chi-square *P*<2.2E-16

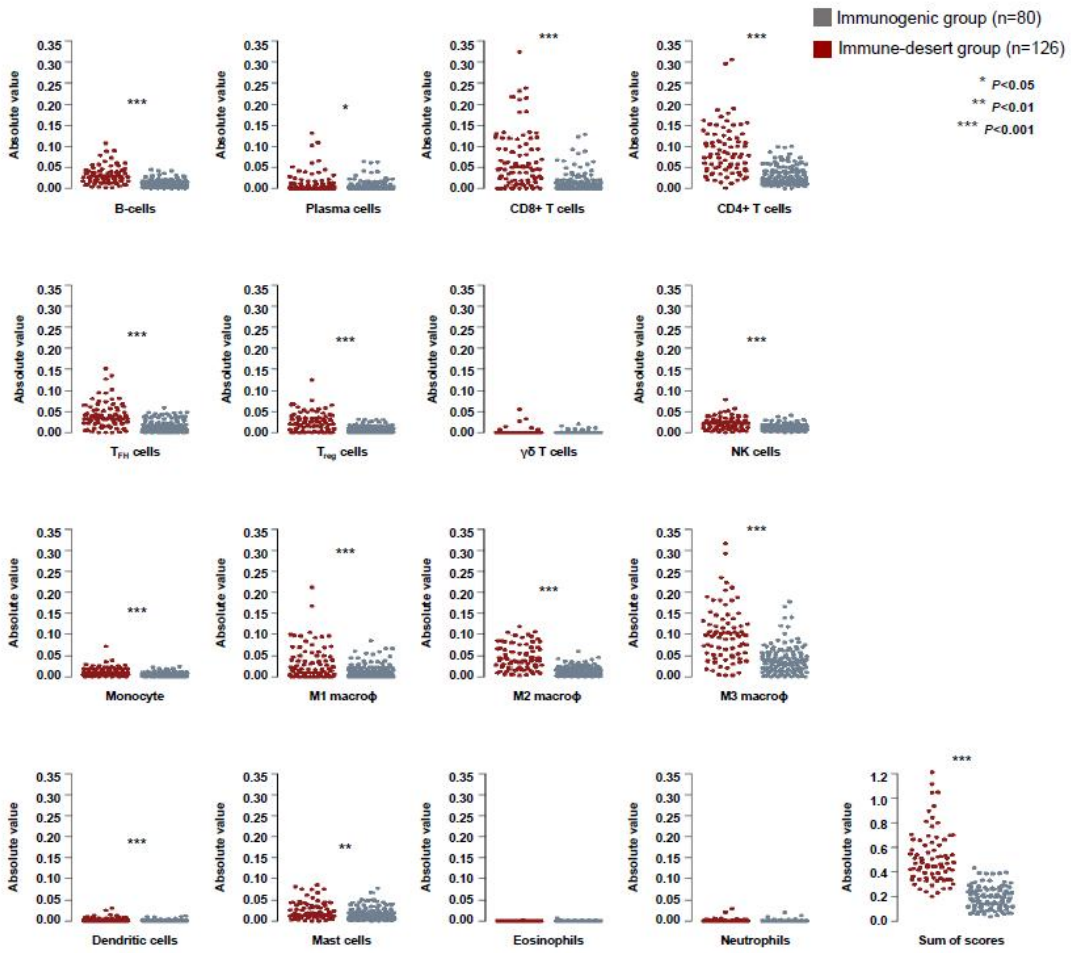
	AMC-C1	AMC-C2	AMC-C3	AMC-C4	AMC-C5	Total
iCluster 1	7	23	3	27	3	63
iCluster 2	33	1	14	5	2	55
iCluster 3	25	0	15	3	20	63
Total	71	25	33	37	25	181*

Chi-square *P*<2.2E-16
* iCluster classification was not available in 10 cases .

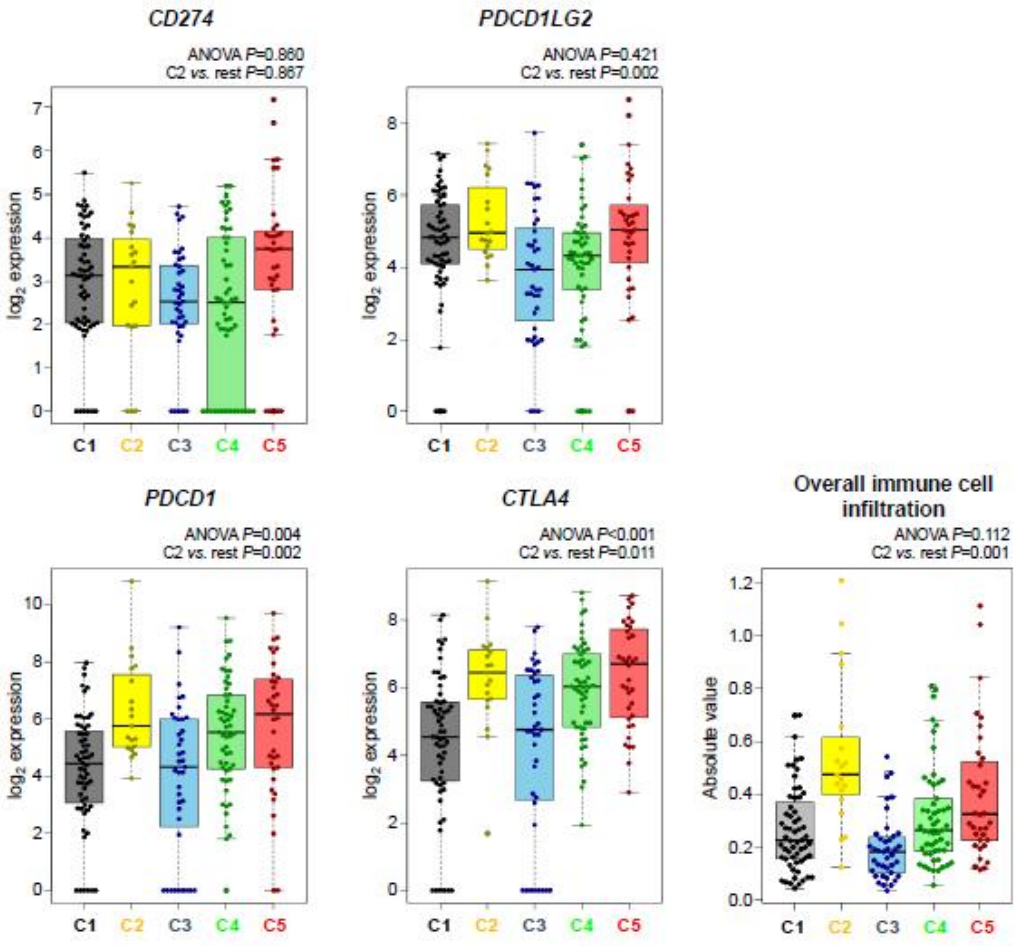
Supplementary Figure 9. Bar plots depicting the estimated probabilities of the Hoshida[6] and TCGA iCluster[5] assignments of the AMC dataset. The lower table presents the relationships between the AMC subgroups, and the Hoshida and TCGA iCluster classes in the AMC (n = 206) and TCGA-LIHC core cohorts (n = 191).



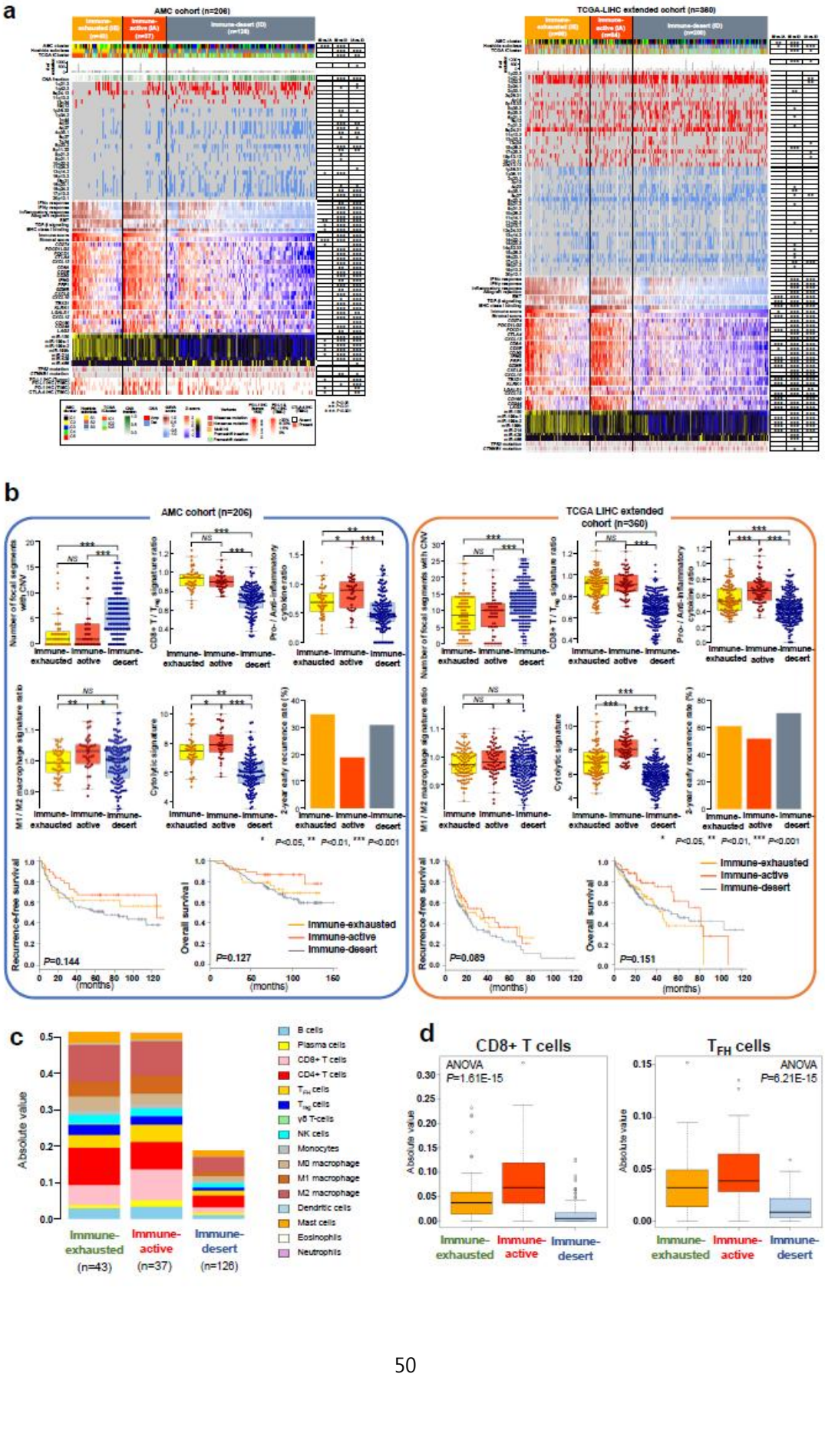
Supplementary Figure 10. Outcome analyses based on the AMC clustering in the AMC and validation datasets. Recurrence-free survival by the Kaplan-Meier method and cancer-specific mortality by competing risks analysis of the AMC datasets. The trends revealed by the competing risks analysis were also observed in the RIKEN HCC and GSE 14520 cohorts.



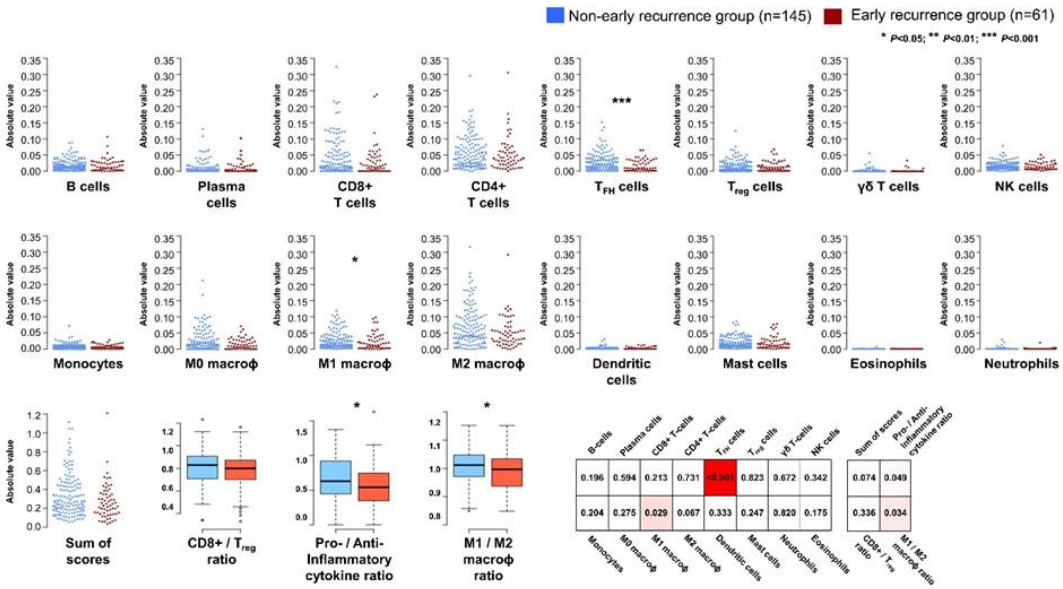
Supplementary Figure 11. Comparisons of the absolute numbers of immune cell types in the immunogenic and immune-desert groups in the AMC dataset. The absolute values of the individual immune cell types were calculated using CIBERSORT, and statistical significance was evaluated by the Wilcoxon rank-sum test. The absolute numbers of most of the immune cell types were significantly increased, the exceptions being eosinophils, neutrophils, and γδ-type T-cells.



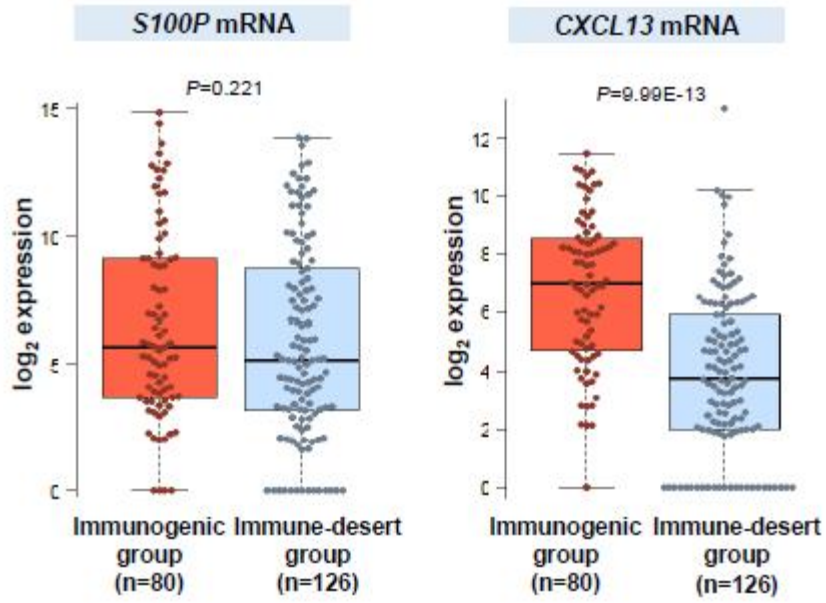
Supplementary Figure 12. Absolute immune cell infiltration numbers and levels of expression of immune-related genes according to the AMC subgroups. Overall immune cell infiltration and levels of expression of *PDCD1LG2*, *PDCD1* and *CTLA4* were highest in AMC-C2 tumors, as estimated by CIBERSORT.


a
AMC cohort (n=206)
TCGA-LIHC extended cohort (n=360)
b
AMC cohort (n=206)
TCGA LIHC extended cohort (n=360)
Number of focal segments with CNV
CD8+ T / T_reg signature ratio
Pro-/ Anti-inflammatory cytokine ratio
M1 / M2 macrophage signature ratio
Cytolytic signature
2-year early recurrence rate (%)
Immune-exhausted
Immune-active
Immune-desert
NS
* P<0.05, ** P<0.01, *** P<0.001
Recurrence-free survival
Overall survival
P=0.144
P=0.127
P=0.089
P=0.151
(months)
c
Absolute value
B cells
Plasma cells
CD8+ T cells
CD4+ T cells
T_FH cells
T_reg cells
γδ T-cells
NK cells
Monocytes
M0 macrophage
M1 macrophage
M2 macrophage
Dendritic cells
Mast cells
Eosinophils
Neutrophils
Immune-exhausted (n=43)
Immune-active (n=37)
Immune-desert (n=126)
d
CD8+ T cells
ANOVA P=1.61E-15
T_FH cells
ANOVA P=6.21E-15

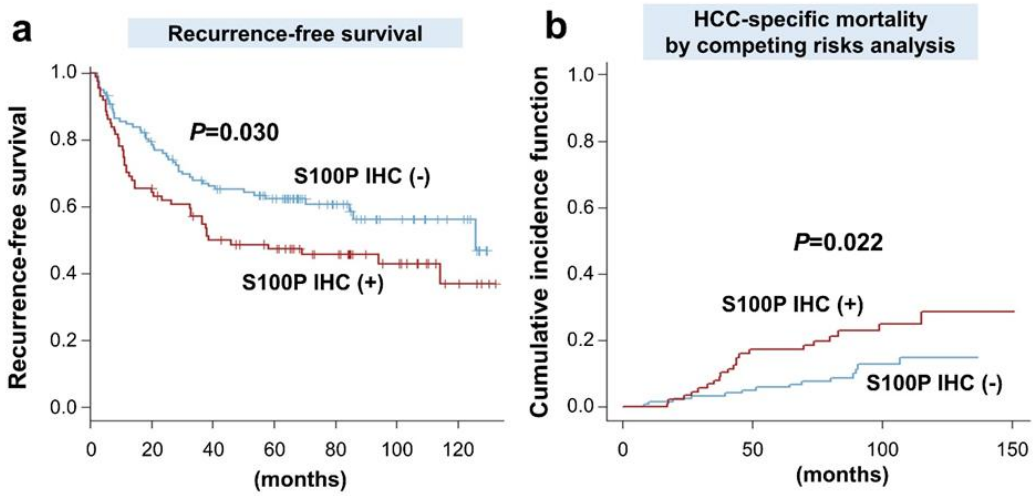
Supplementary Figure 13. Classification and characteristics based on prior-defined immune subtypes (i.e., immune-active *vs.* immune-exhausted *vs.* immune-desert) in the AMC and extended TCGA LIHC cohorts. (a) Heatmaps showing the differences of CNV profile, immune signatures, and immune-associated genes between the immune-exhausted, -active, and -desert groups. There were no significant differences in either CNV burden or most of the immune-related signatures and genes between the immune-exhausted and active tumors in the AMC cohort (left). However, the immune-active group in the TCGA LIHC extended cohort showed over-expression of adaptive immune response genes (e.g., *CD8A, Granzyme B*) and down-regulation of stromal scores and TGF-β signals (right). (b) Boxplots representing the CNV and immune-related signatures across the prior-defined immune subtypes. The immune-active group had a higher pro-inflammatory /anti-inflammatory ratio and higher cytolytic activities than the immune exhausted group in both the AMC (left) and TCGA LIHC extended cohorts (right). However, there were no significant differences in outcomes including overall survival, recurrence-free survival, and rates of early recurrence between the three immune subclasses. (c) Stacked bar plots displaying mean immune cell type-specific absolute values obtained by CIBERSORT according to the prior defined immune subtypes. (d) Numbers of CD8+ T-cells (left) and follicular helper T-cells (right) were significantly higher in the immune-active tumors than in those displaying immune exhaustion.



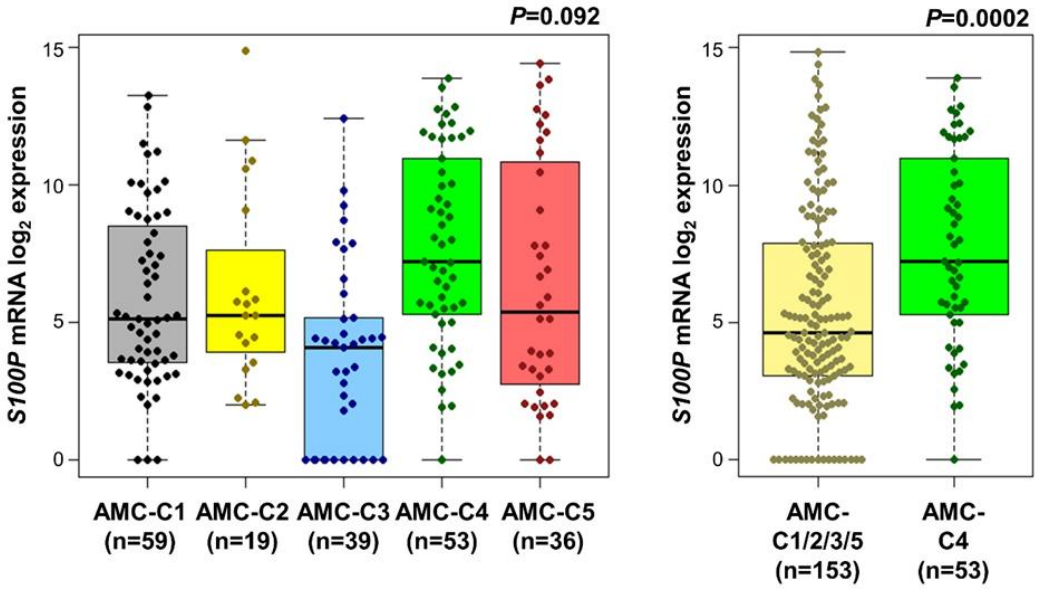
Supplementary Figure 14. Plots of the distributions of immune cell type-specific absolute values in patients with early recurrence (within 2 years) and those without recurrence, in the AMC cohorts. Absolute values of individual immune cell types were calculated using CIBERSORT. *P*-values given by Wilcoxon rank-sum tests are also shown. Numbers of follicular helper T-cells and M1 macrophages were significantly decreased in the early recurrence group. Differences in gene expression signature ratios between pro- and anti-inflammatory cytokines and between M1 and M2 macrophages were also statistically significant.



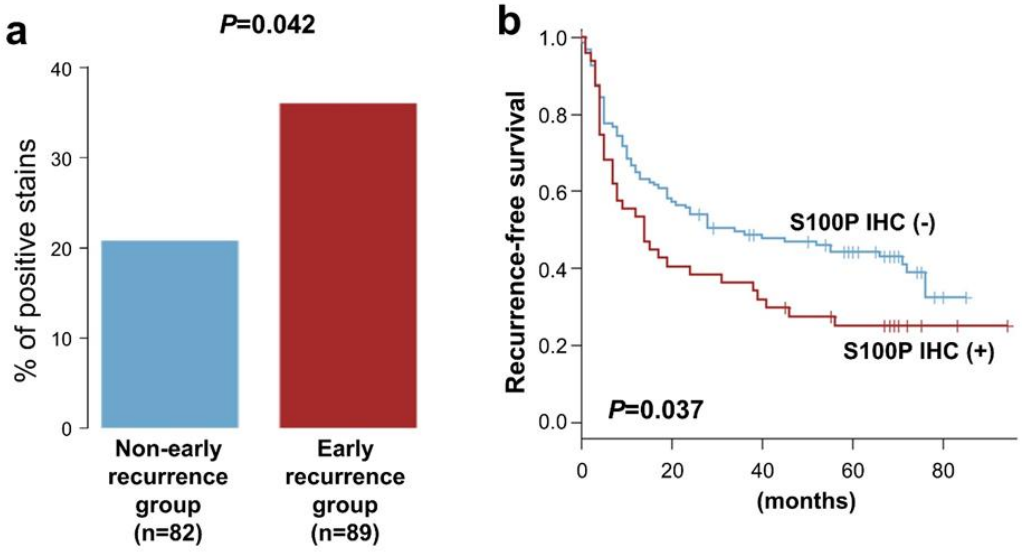
Supplementary Figure 15. Comparison of the expression levels of *S100P* and *CXCL13* in the immunogenic and immune-desert group. While the levels of *S100P* mRNA were not significantly different as a function of immune status, *CXCL13* mRNA was clearly elevated in the immunogenic group.



Supplementary Figure 16. The association between immunohistochemical S100P expression and outcomes in original AMC cohort (n = 206). S100P IHC-positive patients were associated with a (a) worse recurrence-free survival by Kaplan-Meier methods and (b) HCC specific mortality by competing risks analysis.



Supplementary Figure 17. Comparisons of *S100P* mRNA expression across the AMC subgroups in the AMC cohorts. Mean *S100P* expression was significantly higher in AMC-C4 than in the other subgroups.



Supplementary Figure 18. Association between histologic S100P expression and outcomes in the independent AMC validation cohort (n = 171). (a) Positive S100P IHC staining was more frequent in the patients with early recurrence. (b) S100P IHC-positivity was associated with poorer recurrence-free survival.

Supplementary Figure 19. (a) Electrophoretic analysis of RT-PCR products reveals *S100P* mRNA expression in the indicated HCC cell lines. (b) Silencing of endogenous *S100P* mRNA in SNU475 cells by transfection with the indicated concentration of siRNA against *S100P* but not by scrambled siRNA (siCtrl).

SUPPLEMENTARY TABLES

Supplementary Table 1. The scoring systems for PD-L1 and PD-1 immunohistochemistry

PD-L1			**PD-1**
Tumor cells Immunoreactivity Scoring System (IRS): A+B		**TIMC**	**TIMC**
PD-L1 quantity score (A) 0: 0% 1: ≤1% 2: 2–10% 3: 11–50% 4: > 50%	PD-L1 intensity score (B) 0: negative 1: weak 2: moderate 3: strong	Percentage of PD-L1 positive cells 0: 0% 1: 1–5% 2: 6–20% 3: > 20%	Percentage of PD-1 positive cells 0: 0% 1: 1–5% 2: 6–20% 3: > 20%

TIMC, tumor-infiltrating mononuclear cells.

Supplementary Table 2. Patient and tumor characteristics of the main AMC cohort

Variable	Main AMC cohort (n = 206)
Age at surgery (years)	55 (48-61)
< 60 years	139 (67.5%)
≥ 60 years	67 (32.5%)
Gender	
Male	154 (74.8%)
Female	52 (25.2%)
Etiology of the liver disease	
Hepatitis B virus infection	149 (72.3%)
Hepatitis C virus infection	20 (9.7%)
Others	37 (18.0%)
Serum AFP level (ng/ml)	40.2 (4.9-371.1)
< 100 ng/ml	120 (58.3%)
≥ 100 ng/ml	86 (41.7%)
Tumor numbers	
Single	194 (94.2%)
Multiple	12 (5.8%)
Tumor size (cm)	3.8 (2.9-5.5)
< 5 cm	143 (69.4%)
≥ 5 cm	63 (30.6%)
Microvascular invasion	
Yes	57 (27.7%)
No	149 (72.3%)
Edmondson-Steiner grade	

I-II	132 (64.1%)
III-IV	74 (35.9%)
AJCC stage	
Stage IA	18 (8.7%)
Stage IB	176 (85.4%)
Stage II	9 (4.4%)
Stage IIIA	3 (1.5%)

Data are presented as median (interquartile range) or number (%).

AFP, alpha-fetoprotein; AJCC, American Joint Committee on Cancer.

Supplementary Table 3. Unadjusted and adjusted hazard ratios from a competing risks regression model representing the risk of cancer-specific mortality in the AMC cohort (n = 206)

	Univariate analysis			Multivariate analysis		
	HR	**95% CI**	***P* value**	**HR**	**95% CI**	***P* value**
AMC-C4 *vs.* the others	1.49	1.06-2.09	0.021	**1.47**	**1.05-2.05**	**0.025**
Age	0.99	0.95-1.02	0.460			
Male Sex	1.63	0.81-3.28	0.170			
HBV infection	0.99	0.48-2.04	0.980			
Cirrhosis	1.79	0.91-3.54	0.091	1.73	0.87-3.44	0.120
Edmondson-Steiner grade, III or IV	1.14	0.81-1.59	0.460			
Microvascular invasion	1.27	0.61-2.64	0.520			
Capsular invasion	1.02	0.47-2.25	0.950			
Serum AFP > 20 ng/ml	1.49	0.76-2.94	0.250			
Tumor size > 5 cm	1.72	0.88-3.38	0.110			
Multiple tumors	1.65	0.50-5.51	0.410			

AFP, alpha-fetoprotein; HBV, hepatitis B virus.

Supplementary Table 4. Factors related to early recurrence (within 2 years) and cancer-specific mortality in the AMC cohort (n = 206)

Early recurrence (within 2 years)						
	Univariate analysis			Multivariate analysis		
	OR	**95% CI**	***P* value**	**HR**	**95% CI**	***P* value**
S100P mRNA expression	1.17	1.08-1.28	0.0002	**1.15**	**1.05-1.25**	**0.002**
Age	1.02	0.99-1.05	0.239			
Male Sex	1.05	0.53-2.14	0.889			
HBV infection	0.88	0.46-1.73	0.702			
Cirrhosis	1.49	0.82-2.73	0.192			
Edmondson-Steiner grade, III or IV	1.11	0.81-1.51	0.507			
Microvascular invasion	1.27	0.65-2.44	0.470			
Capsular invasion	1.30	0.62-2.67	0.478			
Serum AFP > 20 ng/ml	2.43	1.31-4.66	0.006	1.76	0.90-3.48	0.100
Tumor size > 5 cm	1.96	1.04-3.67	0.037	1.73	0.89-3.37	0.104
Multiple tumors	1.20	0.31-3.98	0.771			
Cancer-specific mortality by competing risks analysis						
	Univariate analysis			Multivariate analysis		
	HR	**95% CI**	***P* value**	**HR**	**95% CI**	***P* value**
S100P mRNA expression	1.09	1.00-1.19	0.053	**1.09**	**1.00-1.19**	**0.042**
Age	0.99	0.95-1.02	0.460			
Male Sex	1.63	0.81-3.28	0.170			
HBV infection	1.01	0.49-2.09	0.980			
Cirrhosis	1.79	0.91-3.54	0.091	1.86	0.94-3.68	0.074

Edmondson-Steiner grade, III or IV	1.14	0.81-1.59	0.460
Microvascular invasion	1.27	0.61-2.64	0.520
Capsular invasion	1.02	0.47-2.25	0.950
Serum AFP > 20 ng/ml	1.49	0.75-2.94	0.250
Tumor size > 5 cm	1.72	0.88-3.38	0.110
Multiple tumors	1.65	0.50-5.51	0.410

AFP, alpha-fetoprotein; HBV, hepatitis B virus.

Supplementary Table 5. Patient and tumor characteristics of the independent AMC validation cohort (n = 171)

Variable	independent AMC validation cohort (n=171)
Age at surgery (years)	55 (48-63)
< 60 years	109 (63.7%)
≥ 60 years	62 (36.3%)
Gender	
Male	135 (78.9%)
Female	36 (21.1%)
Etiology of the liver disease	
HBV infection	102 (59.6%)
Others	69 (40.4%)
Liver cirrhosis	96 (56.1%)
Serum AFP level (ng/ml)	27.6 (4.2-624.0)
< 100 ng/ml	105 (61.4%)
≥ 100 ng/ml	66 (38.6%)
Tumor numbers	
Single	144 (84.2%)
Multiple	27 (15.8%)
Tumor size (cm)	4.0 (2.5-8.0)
< 5 cm	104 (60.8%)
≥ 5 cm	67 (39.2%)
Microvascular invasion	
Yes	32 (18.7%)
No	139 (81.3%)

Edmondson-Steiner grade	
I-II	107 (62.6%)
III-IV	64 (37.4%)
AJCC stage	
Stage IA	21 (12.3%)
Stage IB	106 (62.0%)
Stage II	24 (14.0%)
Stage IIIA	9 (5.3%)
Stage IIIB	10 (5.8%)
Stage IVA	1 (0.6%)

Data are presented as median (interquartile range) or number (%).

AFP, alpha-fetoprotein; AJCC, American Joint Committee on Cancer; HBV, hepatitis B virus.

국문요약

간세포암은 암으로 인한 사망의 세 번째 순위를 차지하는 암종으로, 최근 이에 대한 분자유전학적 특성이 연구되고 있으나 실제 진료에 적용하기에는 아직 요원한 실정이다. 본 연구는 206례의 수술적으로 절제된 간세포암 조직에 대해 전사체 및 유전체 분석을 실시하였고, 이를 통해 임상 예후에 영향을 미치는 암종 및 면역반응의 생물학적 변이를 확인하고자 하였다. 유전자 발현량 분석 결과 간세포암은 다섯 가지의 소분류로 구분할 수 있었다. AMC-C1 군은 고분화 암종으로 생각되었고, AMC-C2 군은 높은 수준의 상피-중간엽 전이를 보였으며, AMC-C3 군은 *CTNNB1* 유전자의 변이를 특징으로 보였고, AMC-C4 군은 줄기세포와 유사한 특징을 보였으며, AMC-C5 군은 암종-정소 항원의 발현을 특징으로 보였다. 이들 중 AMC-C4 군은 다른 군에 비해 좋지 않은 예후를 보였고, 다른 유전자 발현 공개 데이터에서도 유사한 결과를 보였다. 또한 면역유전체 분석을 거쳐 간세포암의 약 1/3이 세포융해 작용 및 세포독성 T 세포의 양이 증가되어 있는 면역유발성 소분류에 속해 있는 것을 확인할 수 있었고, 이 분류와 유전자 복제수 변이 및 microRNA 발현 사이의 유의미한 상관관계를 확인할 수 있었다. 또한 면역유발성 소분류에 속한 증례는 그렇지 않은 증례에 비해 더 좋은 예후를 보이는 것을 알 수 있었다. 마지막으로 수술 후 조기 재발과 연관을 보이는 생체 표지자로 S100P를 발견할 수 있었고, 이 생체 표지자는 독립적인 검증 군에서도 유의미한 결과를 보였다. 이 연구를 통해 간세포암의 생물학적, 임상적 특성이 실제 환자 진료에서도 확인 가능한 유전체 및 면역분자적 특성의 차이와 관련이 있음을 확인할 수 있었다.

Key words: Hepatocellular carcinoma; Molecular subgroup; Tumor immune response; Treatment outcome.