Molecular Therapy
Oncolytics
Original Article

Ubiquitin-related molecular classification and risk stratification of hepatocellular carcinoma

Si Yang,[1,2,5] Bowen Yao,[3,5] Liming Wu,[4] Yuanxing Liu,[4] Kang Liu,[3] Peng Xu,[1,2] Yi Zheng,[1,2] Yujiao Deng,[1,2] Zhen Zhai,[1,2] Ying Wu,[1,2] Na Li,[1,2] Dai Zhang,[1,2] Huafeng Kang,[2] and Zhijun Dai[1,2]

[1]Department of Breast Surgery, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou 310003, China; [2]Department of Oncology, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710004, China; [3]Department of Hepatobiliary Surgery, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, China; [4]Department of Hepatobiliary and Pancreatic Surgery, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou 310003, China

The roles of ubiquitin-related genes in hepatocellular carcinoma (HCC) have not been thoroughly investigated. This study aimed to systematically examine ubiquitin-related genes and identify subtypes and stratify prognosis of HCC by using ubiquitin-related signatures. Survival, biological processes, tumor microenvironment (TME), and genomic alterations of the HCC subtypes were investigated. Patients with HCC were classified into two subtypes (clusters 1 and 2) with distinct survival outcomes, pathways, and genomic alterations. Cluster 2 had better prognosis than did cluster 1. Hepatitis B, hepatitis C, Janus tyrosine kinase (JAK)-signal transducer and activator of transcription (STAT) pathway, and natural killer cell-mediated cytotoxicity were enriched in cluster 1. Moreover, cluster 2 had a higher immune score and immune cell infiltrations, whereas cluster 1 had a lower immune score and immune infiltrations. Additionally, mutations, amplifications, and deletions among the phosphatidylinositol 3-kinase (PI3K)-AKT, p53, and receptor tyrosine kinase (RTK)-RAS pathways more frequently occurred in cluster 1, while those among the Hippo, MYC, and Notch signaling pathways were found in cluster 2. Finally, a prognostic signature, consisting of eight ubiquitin-related genes, was established and validated. In brief, our study established a new classification and developed a prognostic signature for HCC.

INTRODUCTION

Hepatocellular carcinoma (HCC) is one of the most common malignancies and the second leading cause of cancer-related deaths worldwide.[1] In spite of the advances in treatment and diagnostic methods, the clinical outcome of patients with HCC remains poor.[2] Therefore, screening populations at high risk of developing HCC, discovering new therapeutics targets, and improving prognosis are urgently needed.

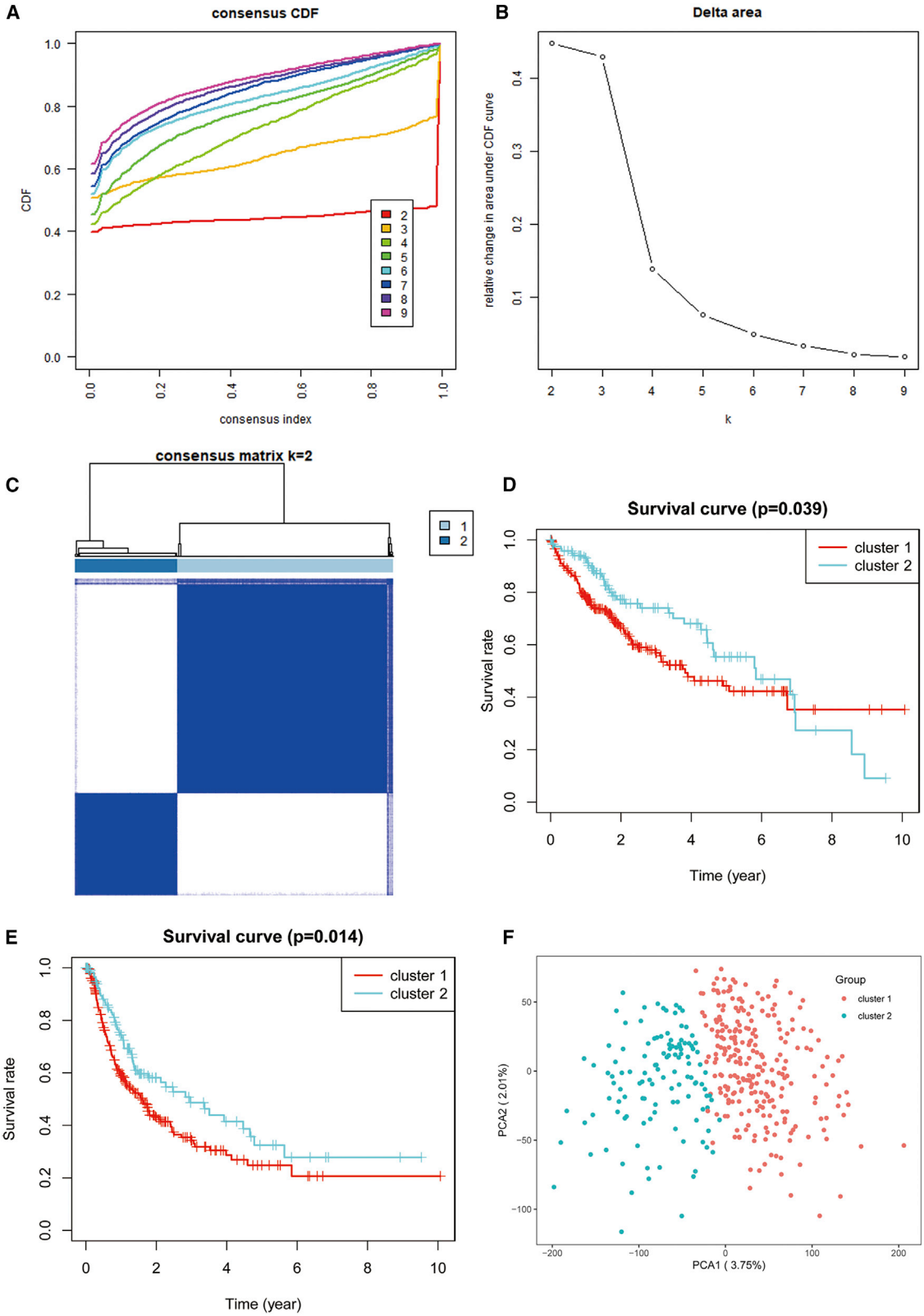
Ubiquitin is a small protein and serves as a post-translational protein modifier by marking proteins for degradation.[3] Ubiquitin transfer cascades (known as ubiquitination), including ubiquitin-activating enzymes (E1s), ubiquitin-conjugating enzymes (E2s), and ubiquitin-protein ligases (E3s), constitute a complex network to modify protein substrates.[4] This ubiquitination process can be reversed by deubiquitinating enzymes (DUBs), that is, removing ubiquitin from modified proteins.[5] In addition, many proteins with ubiquitin-binding domains (UBDs) or ubiquitin-like domains (ULDs) also play important roles in the ubiquitin systems.[5] These ubiquitin system-related genes regulate a number of biological processes such as protein degradation, DNA damage repair, signal transduction, and cell cycle.[6–8] Ubiquitin chain dysregulation and ubiquitin-related protein malfunctions are involved in the development of various diseases such as cancers, metabolic diseases, and neurodegenerative diseases.[9,10]

Over the past few decades, multiple studies have revealed that ubiquitin-related genes are aberrantly expressed in cancers and regulate many cancer-related genes such as tumor suppressor genes (TSGs: *VHL*, *PTEN*, *p27*, *p53*, and *RB*) and oncogenes (*EGFR* and *MYC*).[9–12] These deregulated ubiquitin-related genes can cause aberrant activation or inactivation of cancer-associated pathways and play important roles in carcinogenesis.[9,10] In HCC, several ubiquitin-related proteins interplay with cancer-related proteins and act as oncogenic proteins. For instance, the E3 ligase TRIM25 promotes HCC cell survival by targeting Keap1 for ubiquitination and degradation, activating Nrf2 signaling and reducing reactive oxygen species levels during endoplasmic reticulum stress.[13] TRAF6, another E3 ligase, was reported to interact with histone deacetylase 3 to increase gene expression levels and the protein stability of MYC.[14] Targeting ubiquitin-related genes has been a promising strategy for anticancer drug development.[15] However, the roles of ubiquitin-related genes in HCC have not been thoroughly investigated. A better understanding of the mechanisms and roles of ubiquitin-related genes could provide

Received 26 December 2020; accepted 6 April 2021;
https://doi.org/10.1016/j.omto.2021.04.003.

[5]These authors contributed equally

Correspondence: Zhijun Dai, Department of Breast Surgery, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou 310003, China.
E-mail: dzj0911@zju.edu.cn
Correspondence: Huafeng Kang, Department of Oncology, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710004, China.
E-mail: kanghuafeng1973@126.com



(legend on next page)

Table 1. The demographic and clinicopathological characteristics of two clusters in the TCGA HCC cohort

Variables	Group	Cluster 1 (n = 250)	Cluster 2 (n = 121)	p value	Method
Age (mean ± SE)		59.6 ± 0.9	59.2 ± 1.1	0.8296	t-test
Sex	female	87 (34.8%)	34 (28.1%)	0.1968	$\chi 2$ test
	male	163 (65.2%)	87 (71.9%)		
Clinical stage	I	100 (40.0%)	71 (58.7%)	0.0015	Fisher's exact test
	II	60 (24.0%)	26 (21.5%)		
	III	67 (26.8%)	18 (14.9%)		
	IV	2 (0.8%)	3 (2.5%)		
	NA	21 (8.4%)	3 (2.5%)		
T stage	T1	110 (44.0%)	71 (58.7%)	0.0471	Fisher's exact test
	T2	68 (27.2%)	26 (21.5%)		
	T3	61 (24.4%)	19 (15.7%)		
	T4	10 (4.0%)	3 (2.5%)		
	TX	1 (0.4%)	2 (1.7%)		
N stage	N0	163 (65.2%)	89 (73.6%)	0.1371	Fisher's exact test
	N1	4 (1.6%)	0 (0.0%)		
	NX/NA	83 (33.2%)	32 (26.4%)		
M stage	M0	172 (68.8%)	94 (77.7%)	0.0206	Fisher's exact test
	M1	1 (0.4%)	3 (2.5%)		
	MX	77 (30.8%)	24 (19.8%)		
HBV/HCV/HBV+HCV	yes	122 (48.8%)	42 (34.7%)	0.0104	$\chi 2$ test
	none	128 (51.2%)	79 (65.3%)		
Histological grade	G1	36 (14.4%)	19 (15.7%)	0.8140	Fisher's exact test
	G2	117 (46.8%)	60 (49.6%)		
	G3	86 (34.4%)	36 (29.8%)		
	G4	7 (2.8%)	5 (4.1%)		
	NA	4 (1.6%)	1 (0.8%)		

NA, not available; TX, unknown T stage; MX, unknown M stage; NX, unknown N stage; HBV, hepatitis B virus; HCV, hepatitis C virus.

a basis for understanding HCC and the development of novel therapeutics. Therefore, this study aimed to systematically study ubiquitin-related genes in HCC. Given the high heterogeneity in HCC, our study also intended to identify HCC subtypes from tumor samples and stratify the risk and prognosis of patients with HCC by using ubiquitin-related signatures.

RESULTS

Identification of HCC subtypes based on ubiquitin-related genes

The expression profiles of 777 ubiquitin-related genes from The Cancer Genome Atlas (TCGA) were used for consensus clustering analysis of HCC. The optimal cluster was obtained when the *k* value was 2 according to the cumulative distribution function (CDF) curves (Figures 1A and 1B). The 371 HCC patients were classified into two subtypes as follows: cluster 1 (n = 250) and cluster 2 (n = 121; Figure 1C). A survival analysis revealed that the overall survival (OS) time and progression-free survival (PFS) time in cluster 1 were shorter than those in cluster 2 ($p < 0.05$; Figures 1D and 1E). The principal component analysis (PCA) revealed that the samples from cluster 1/2 were well separated from each other (Figure 1F). The results suggested significant differences between cluster 1 and cluster 2.

Clinicopathological features, biological processes, and pathways in clusters 1 and 2

The expression levels of the ubiquitin-related genes in clusters 1 and 2 are shown in the heatmap (Figure S1). The clinicopathological features of the two subtypes were compared. We found that the proportion of patients with lower tumor stage (stage I/II) was significantly higher in cluster 2 ($p = 0.0015$; Table 1). Moreover, the proportions of T1 and T2 were significantly higher in cluster 2 ($p = 0.0471$; Table 1), while HCC samples with advanced tumor stages (stage III/IV) and higher tumor sizes (T3/T4) were included in cluster 1. There were significantly more patients with hepatitis B and C infections ($p = 0.0104$) in cluster 1. No significant differences in the distribution of age, sex, histological grade, and N stage were found between the two subtypes ($p > 0.05$; Table 1).

A gene set enrichment analysis (GSEA) was performed to identify pathways enriched in the two subtypes. Under the threshold of an adjusted p value of <0.05, there are 18 significantly enriched pathways in cluster 1 and there were no significant pathways involved in cluster 2. In detail, the results showed that the cytosolic DNA-sensing pathway, hepatitis B, hepatitis C, the Janus tyrosine kinase (JAK)-signal transducer and activator of transcription (STAT) pathway, natural killer cell-mediated cytotoxicity, the neuroactive ligand-receptor interaction, the non-obese diabetic (NOD)-like receptor pathway, pentose and glucuronate interconversions, the retinoic acid-inducible gene (RIG)-I-like receptor pathway, and the Toll-like receptor pathway were highly enriched in cluster 1 (normalized enrichment score > 1, adjusted $p < 0.05$; Figure S2).

Tumor microenvironment (TME) characterization in clusters 1 and 2

GSEA results indicated that some immune-related pathways, such as natural killer cell-mediated cytotoxicity, the NOD-like/RIG-I-like/Toll-like receptor pathways, and JAK-STAT pathway were significantly activated in cluster 1. Subsequently, the TME (tumor purity and infiltrating stromal/immune cells in tumor samples) in clusters 1 and 2 were examined. We found that the immune score

Figure 1. Identification of two subtypes of HCC in TCGA cohort

(A) Consensus clustering cumulative distribution function (CDF) for $k = 2–9$. (B) Relative change in area under the CDF curve for $k = 2–9$. (C) The consensus score matrix of HCC samples when $k = 2$ (1, cluster 1; 2, cluster 2). (D) Kaplan-Meier curves of overall survival (OS) of clusters 1 and 2. (E) Kaplan-Meier curves of progression-free survival of clusters 1 and 2. (F) Principal component analysis of the gene-expression profiles in the TGGA HCC cohort.

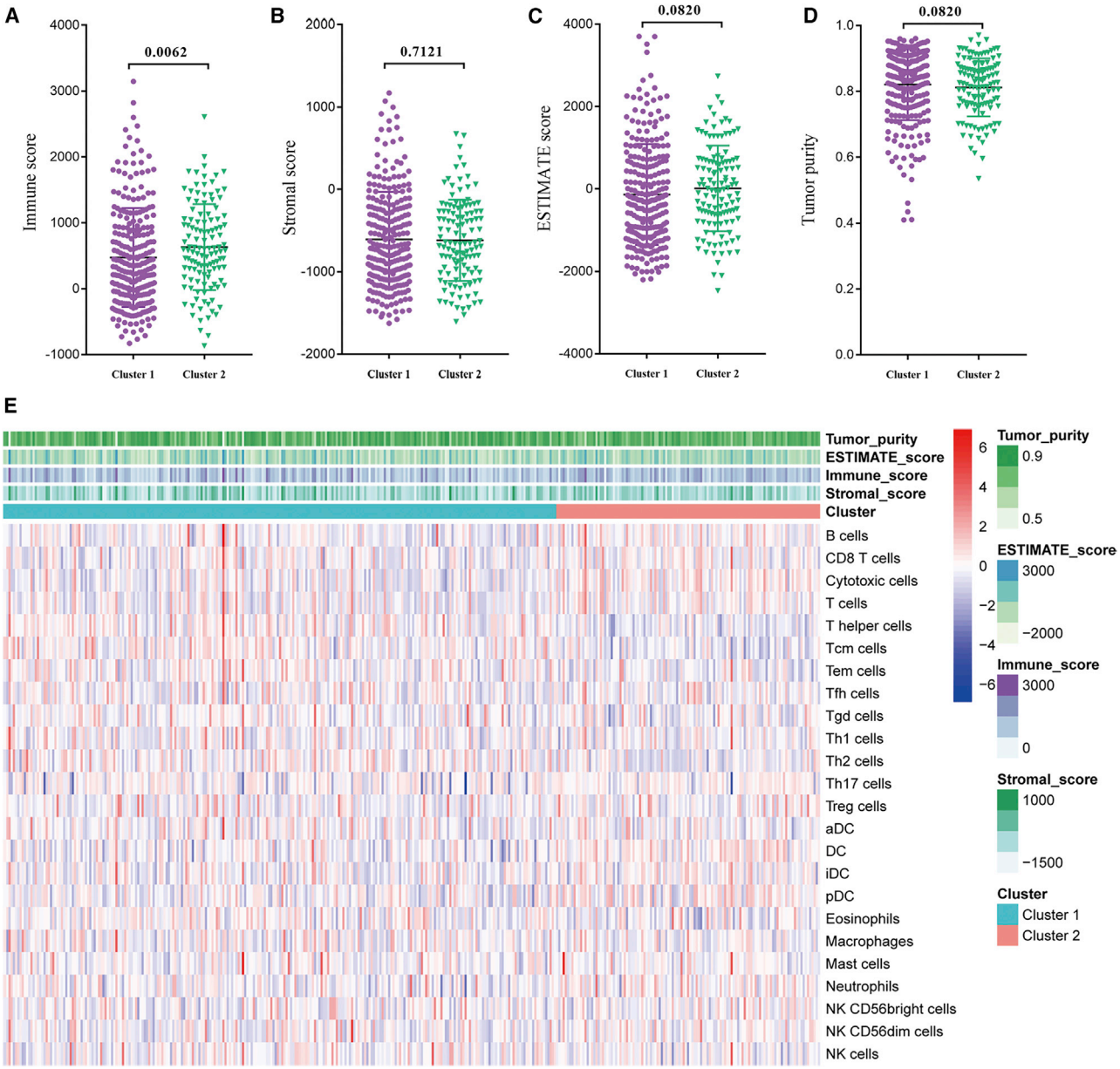


Figure 2. Tumor microenvironment characterization in clusters 1 and 2

(A–D) Comparison of immune score, stromal score, ESTIMATE score, and tumor purity in the two subtypes. (E) Heatmap describing the abundance of 24 immune cell types in the two subtypes.

was significantly higher in cluster 2 ($p < 0.05$; Figure 2A). However, we found no significant differences in stromal score, ESTIMATE score, and tumor purity between the two subtypes (Figures 2B–2D). Given the significant differences in immune score between the two subtypes, immune cell infiltration was further investigated to characterize the immune landscapes of clusters 1 and 2 (Figure 2E). Cluster 2 had a significantly higher abundance of B cells, cytotoxic cells, T cells, dendritic cells, and neutrophils ($p < 0.05$; Figure 3A). The expression levels of eight immune checkpoint genes (ICGs: *PD1*, *PD-L1*, *PD-L2*, *CTLA4*, *VTCN1*, *TIM3*, *LAG3*, and *TIGIT*) in the two subtypes were further investigated. The results indicated that cluster 1 exhibited higher expression levels for the eight ICGs (except for the *LAG3* gene) than cluster 2 ($p < 0.05$; Figure 3B).

Comparison of mutations and CNVs between clusters 1 and 2

To identify potential drug targets to reverse the poorer survival in cluster 1, we explored the significant mutations and copy number variations

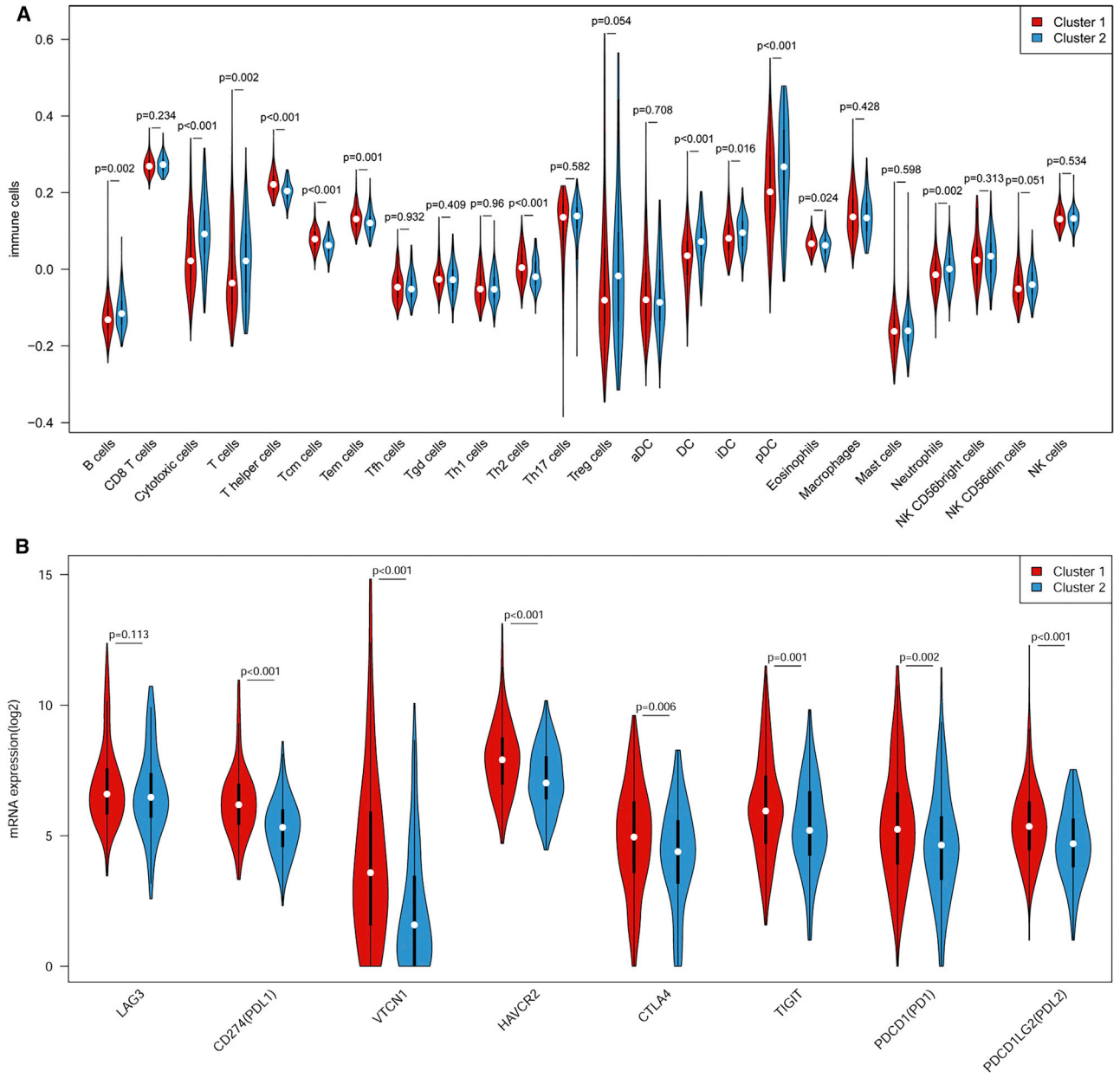
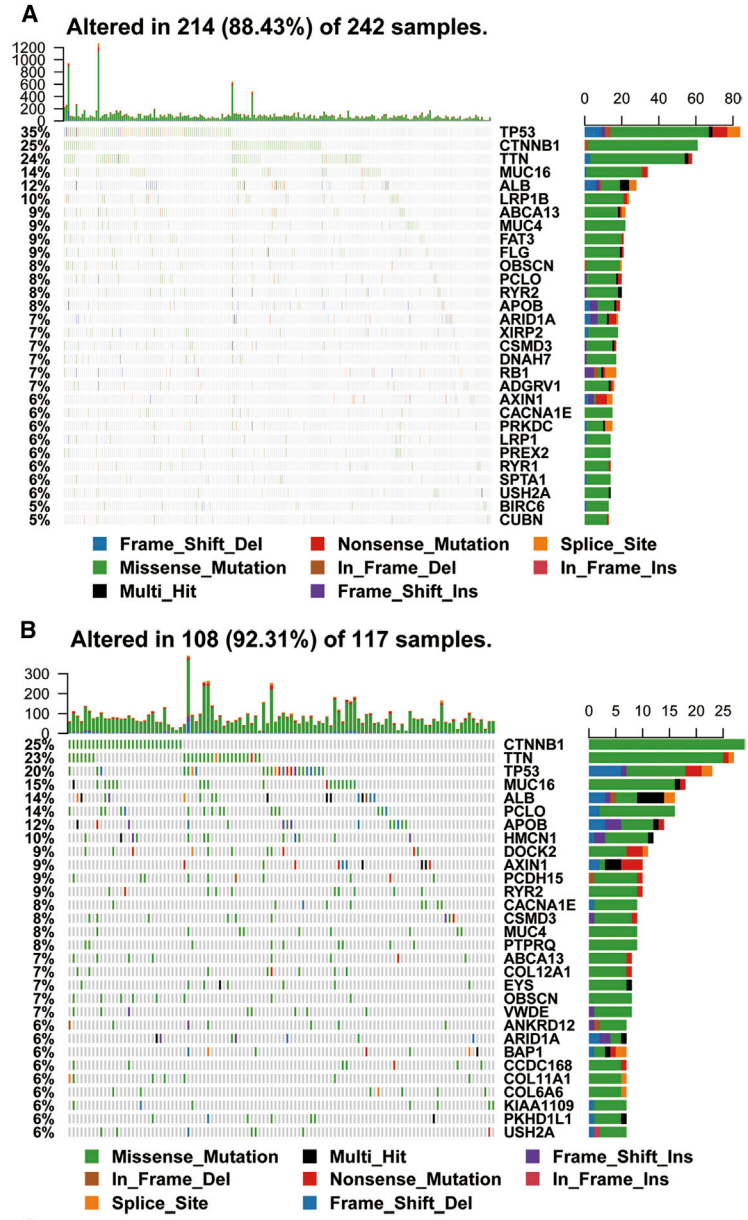


Figure 3. Immune cells abundance and expression levels of immune checkpoint genes in clusters 1 and 2

(A) Comparison of the abundance of 24 immune cell types in clusters 1 and 2. (B) Expression levels of eight immune checkpoint genes in clusters 1 and 2. Level of gene expression is reported as log2-transformed count.

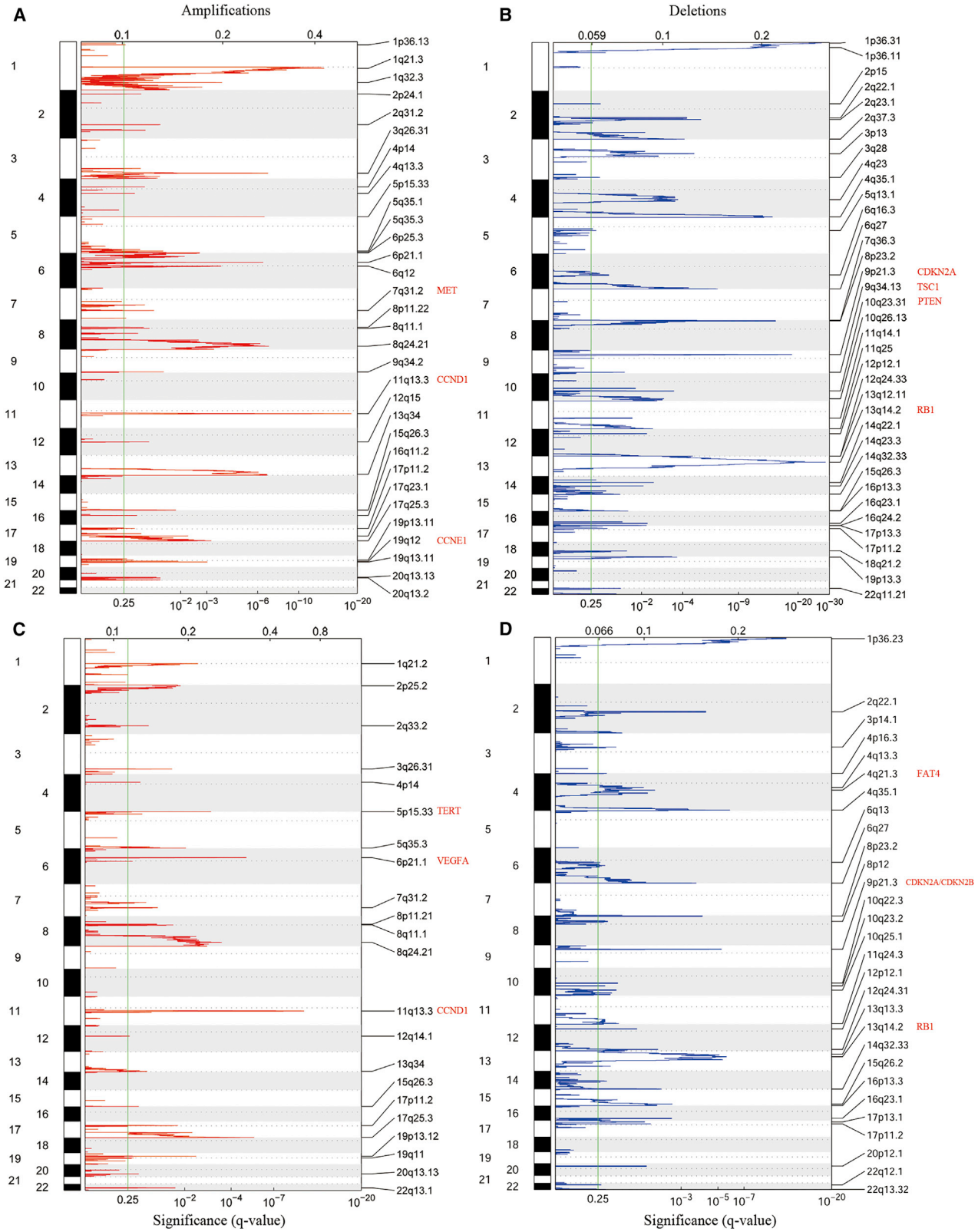
(CNVs) between clusters 1 and 2. The top 30 most frequently mutated genes in clusters 1 and 2 are presented in Figures 4A and 4B. Cluster 1 had a significantly higher mutation frequency of *TP53* (TSG in the p53 pathway) than did cluster 2 (35% versus 20%). The mutation frequency of *CTNNB1* (oncogene in the Wnt pathway) was equal in the two subtypes (both 25%). The mutation frequencies of 10 critical oncogenic pathways in clusters 1 and 2 were summarized (Figure 4C). Mutations in the phosphatidylinositol 3-kinase (PI3K)-AKT, p53, and receptor tyrosine kinase (RTK)-RAS pathways more frequently occurred in cluster 1 (Figure 4C). The Hippo, MYC, and Notch pathways displayed higher mutation frequencies in cluster 2 (Figure 4C). The mutation frequency of the cell-cycle pathway was predominant in both subtypes (Figure 4C). In addition, differences in somatic CNV in clusters 1 and 2 were evaluated using GISTIC 2.0. The CNV analysis revealed that amplifications of 11q13.3 (*CCND1* [oncogene in the cell-cycle pathway]), 19q12 (*CCNE1* [oncogene in the cell-cycle pathway]), and 7q31.2 (*MET*



	Cell cycle	Hippo	MYC	Notch	NRF2	PI3K-AKT	RTK RAS	P53	TGF-β	Wnt
Cluster 1	70.3%	10.6%	16.5%	25.0%	6.8%	28.8%	22.0%	41.5%	7.6%	41.1%
Cluster 2	66.1%	16.1%	25.0%	27.7%	6.3%	17.0%	20.5%	28.6%	5.4%	22.9%

Figure 4. Comparison of mutations in clusters 1 and 2

(A and B) The top 30 most frequently mutated genes in two HCC subtypes. (C) The mutation frequencies of ten critical oncogenic pathways in two HCC subtypes.



(legend on next page)

[oncogene in the RTK-RAS pathway]), and deletions of 13q14.2 (*RB1* [TSG in the cell-cycle pathway]), 9p21.3 (*CDKN2A* [TSG in the cell-cycle pathway]), 10q23.31 (*PTEN* [TSG in the PI3K-AKT pathway]), and 9q34.13 (*TSC1* [TSG in the PI3K-AKT pathway]) were identified in cluster 1 (Figures 5A and 5B). Moreover, amplifications of 11q13.3 (*CCND1* [oncogene in the cell-cycle pathway]), 6p21.1 (*VEGFA* [oncogene]), and 5p15.33 (*TERT* [oncogene]), and deletions of 9p21.3 (*CDKN2A* and *CDKN2B* [TSGs in the cell-cycle pathway]), 13q14.2 (*RB1* [TSG in the cell-cycle pathway]), and 4q21.3 (*FAT4* [TSG in the Hippo pathway]) were identified in cluster 2 (Figures 5C and 5D).

Establishment and evaluation of a ubiquitin-related signature

From 777 ubiquitin-related genes, 336 genes with prognostic value ($p < 0.01$) were identified by univariate analyses. From the 336 genes, 22 genes were selected by the least absolute shrinkage and selection operator (LASSO) regression analysis (Figure S3A). Finally, we identified eight genes to establish the prognostic signature by the stepwise multivariate regression analysis (Table S1; Figure S3B). The risk score was calculated using the following formula: risk score = $(0.1553 \times \mathrm{Exp}_{\mathrm{UBE2S}}) + (-0.3136 \times \mathrm{Exp}_{\mathrm{SOCS2}}) + (0.2454 \times \mathrm{Exp}_{\mathrm{RNF2}}) + (0.3879 \times \mathrm{Exp}_{\mathrm{HECTD3}}) + (0.2298 \times \mathrm{Exp}_{\mathrm{ATG10}}) + (0.0634 \times \mathrm{Exp}_{\mathrm{BRSK2}}) + (0.1460 \times \mathrm{Exp}_{\mathrm{RNF133}}) + (0.3940 \times \mathrm{Exp}_{\mathrm{TRIM6\text{-}TRIM34}})$. According to the median risk score, the patients in the TCGA HCC cohort (training cohort) were assigned to a high- or low-risk subgroup. The survival analysis result indicated that patients in the high-risk subgroup displayed worse OS ($p < 0.0001$; Figure 6A). The area under the receiver-operating characteristic (ROC) curve (AUC) for OS was 0.826 at 1 year and 0.748 at 3 years (Figure 6B). The robustness and effectiveness of the ubiquitin-related signature were evaluated on the validation cohort. Similarly, the patients in the high-risk subgroup had poorer OS than did those in the low-risk group ($p = 0.041$; Figure 6C). The 1- and 3-year AUCs were 0.701 and 0.616, respectively, in the International Cancer Genomics Consortium (ICGC) cohort (Figure 6D). We further investigated the correlation between the two risk subgroups and clinicopathological features in the TCGA cohort. Significant differences were found between the high- and low-risk subgroups, which were marked with * in the heatmap (Figure 6E). The high-risk subgroup correlated with advanced T stage ($p < 0.01$), higher tumor stage ($p < 0.001$), and higher histological grade ($p < 0.01$).

Establishment of a nomogram

Univariate analyses were performed to examine the prognostic values of several clinicopathological features. Consequently, tumor stage, T stage, and M stage correlated with OS ($p < 0.05$; Figure S4A). Tumor and T stages also correlated with PFS ($p < 0.05$; Figure S4B). The multivariate regression analyses suggested that the risk score was an independent prognostic indicator of OS and PFS in the TCGA cohort ($p < 0.001$; Figures S4C and S4D). In addition, the ROC analysis revealed that the sensitivity and specificity of the ubiquitin-related signature were greater than those of the other clinicopathological features (Figure 7A).

A nomogram was constructed, which included risk score, tumor stage, and T/M stage, to predict the probability of OS in HCC patients from the TCGA cohort (Figure 7B). The calibration curves at 1 and 3 years showed good consistency between the actual OS and predicted OS (Figures 7C and 7D).

DISCUSSION

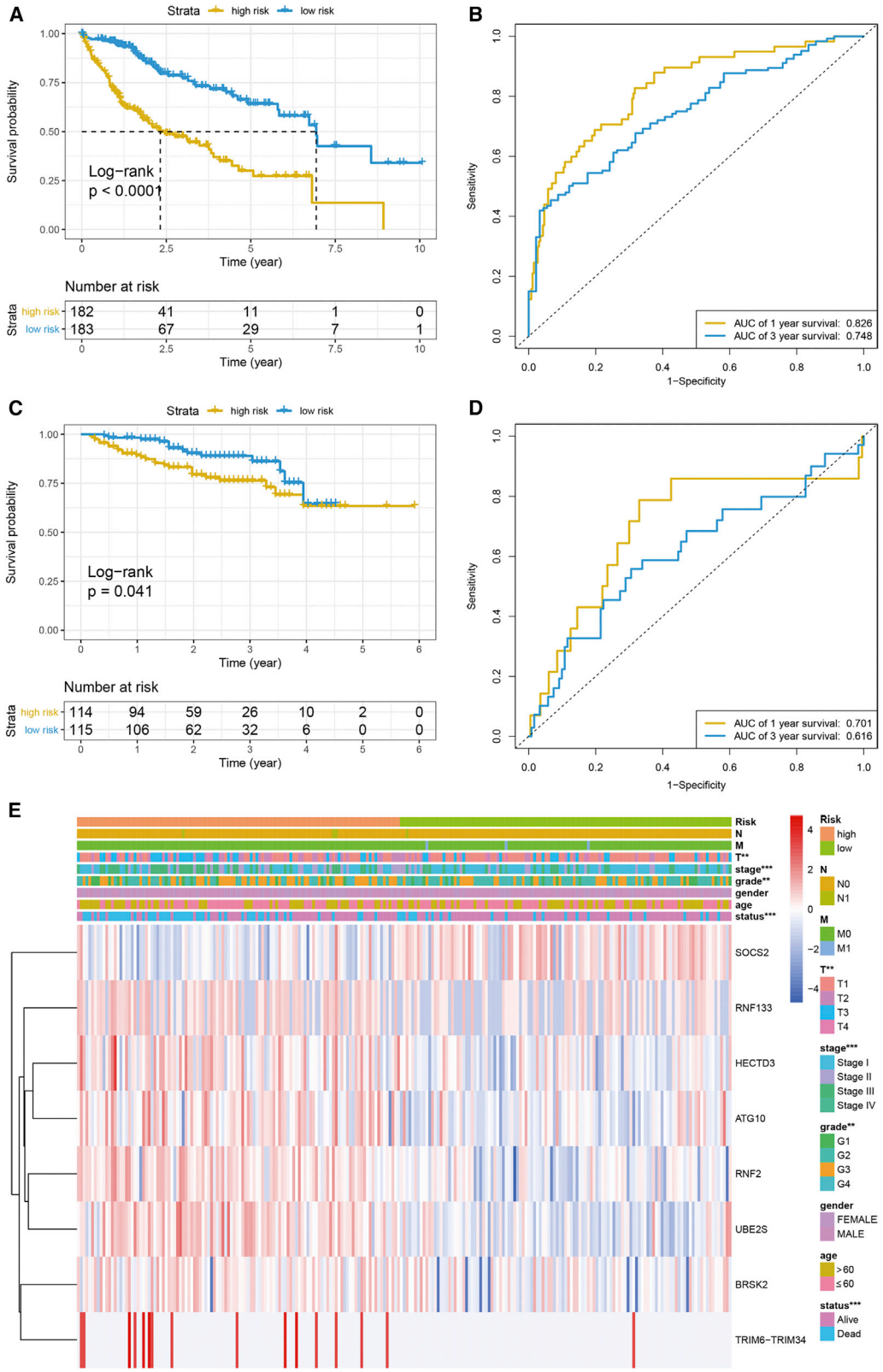
HCC is an aggressive and heterogeneous tumor with a high incidence and a short survival time. Subtype characterization and identification is critical for risk and treatment stratification of HCC patients. In the present study, ubiquitin-related genes were investigated to identify different molecular classifications in HCC samples and to stratify the risk levels of patients with HCC.

Subtype characterization and identification

On the basis of the expression profiles of the ubiquitin-related genes, patients with HCC were classified into two clusters with distinct survival outcomes, pathological features, pathways, TME, and genomic alterations. In detail, cluster 2 had a significantly longer survival time, including OS and PFS, and a higher proportion of patients with early-stage tumors. Hepatitis B and C virus infections are the most common risk factors in the progression of HCC. Of all HCC cases, 75% were associated with hepatitis infection.[16] In our study, we found that cluster 1 with worse prognosis had a higher proportion of patients with hepatitis B and C virus infections, which is consistent with the GSEA results. The GSEA revealed that hepatitis B and hepatitis C virus infection-related pathways were highly enriched in cluster 1. Moreover, the other pathways identified in the GSEA, such as the JAK-STAT pathway and NOD-like/RIG-I-like/Toll-like receptor pathways were frequently reported to play key roles in hepatitis infection or immune processes in HCC.[17–22] For instance, hepatitis B virus infection upregulates Toll-like receptor 2 to promote the invasion of hepatitis B virus-related HCC cells.[19] Therefore, we supposed that clusters 1 and 2 had distinct immune microenvironments. Furthermore, TME characterization in clusters 1 and 2 were investigated to confirm our hypothesis. As a result, cluster 2 had a significantly higher immune score and greater abundance of B cells, cytotoxic cells, T cells, dendritic cells, and neutrophils, while cluster 1 had relatively lower immune cell infiltration and lower immune score. Briefly, cluster 1 had lower immune infiltration and a poor prognosis, while cluster 2 had higher immune infiltration and a favorable prognosis. These findings were consistent with those of a previous study that showed that abundant immune infiltration was associated with better survival in HCC.[23] In addition, cluster 1 exhibited higher expression of ICGs, such as *PD-L1*, *CTLA4*, and *PD-1* than did cluster 2. In a latest meta-analysis, high

Figure 5. Comparison of CNVs in clusters 1 and 2

(A–D) GISTIC 2.0 amplifications and deletions in clusters 1 (A and B) and 2 (C and D). Chromosomal locations of peaks of significantly recurring focal amplifications (red) and deletions (blue) are displayed. The q values, representing the statistical significance, are displayed along the bottom. Regions with q values < 0.25 (green lines) were considered significantly altered. The locations of the peak regions of maximal copy-number change and the known cancer-related genes within those peaks are indicated to the right of each panel.



(legend on next page)

PD-L1 expression level in tumor tissue was reported to correlate with shorter OS, poor tumor differentiation, hepatitis, and tumor-infiltrating lymphocytes in HCC.[24] Higher immune checkpoint expression could decrease immune cell infiltration and inhibit the immune response in cancer.[25] Our results (a negative association between immune checkpoint expression and immune cell infiltration) were consistent with these previous studies. The combined effects of lower immune infiltration, high expression levels of ICGs, advanced tumor stage, large tumor size, and hepatitis B and C virus infections may be responsible for the worse survival in cluster 1.

We further investigated the genomic alterations and intended to detect potential drug targets. At the gene level, cluster 1 had a higher mutation frequency of *TP53*, which leads to p53 pathway activation, than did cluster 2. The mutation frequency of *CTNNB1*, which leads to Wnt pathway activation, was equal in two subtypes. At the pathway level, the mutation frequency of the cell-cycle pathway is predominant in both subtypes. Mutations in the PI3K-AKT, p53, and RTK-RAS pathways more frequently occurred in cluster 1. The Hippo, MYC, and Notch pathways displayed higher mutation frequencies in cluster 2. The amplifications of oncogenes such as *CCND1*, *CCNE1*, and *MET* and the deletions of TSGs such as *RB1*, *CDKN2A*, and *PTEN* were identified in cluster 1. Therefore, we speculated that the hyperactivated p53, cell-cycle, RTK-RAS, or PI3K-AKT pathway might be responsible for the worse survival in cluster 1. In addition, the amplifications of oncogenes such as *CCND1* and *VEGFA* were identified in cluster 2. All in all, our study demonstrated that certain genomic alterations in these critical pathways may be related to prognosis in clusters 1 and 2. These genes with aberrant mutations, amplifications, or deletions could be considered as effective therapeutics targets for different HCC subtypes. In fact, some corresponding inhibitors that target these key pathways or genes showed promising outcomes in trials, such as *MET*.[26] Targeting molecular agents that control multiple signaling pathways are also under development.[27]

Risk stratification

A prognostic signature was developed using eight ubiquitin-related genes, which could significantly distinguish the high- and low-risk patients from TCGA and the ICGC cohorts. Among the eight genes, *SOCS2* was a protective factor and the rest (*UBE2S*, *RNF2*, *HECTD3*, *ATG10*, *BRSK2*, *RNF133*, and *TRIM6-TRIM34*) were risk factors. Some of these genes have been reported in HCC. A study demonstrated that *UBE2S*, as a member of the E2s, is overexpressed in HCC and promotes the progression of HCC cells by enhancing the ubiquitination of p53.[28] *RNF2*, as an E3 ligase, promotes HCC cells growth and metastasis by targeting *SIK1* for degradation.[29] Another study showed that *ATG10* rs10514231 might affect the expression of *ATG10* and was significantly associated with HCC susceptibility.[30] *SOCS2* are reported to inhibit the migration, invasion, and metastasis of HCC cells.[31] These genes may represent promising therapeutic strategies for HCC treatment. The other four genes have not been reported in HCC but in other tumors. Future studies are needed to examine the functional impact of these four genes in the carcinogenesis and development of HCC.

Finally, we integrated risk score, tumor stage, and T/M stage to construct a nomogram, and we found good agreement between the actual OS and predicted OS at 1 and 3 years. These results indicated that the ubiquitin-related signature combined with tumor and T/M stages might be a promising prognostic tool for HCC patients.

Strength and limitations

To the best of my knowledge, our study first identified HCC subtypes in tumor samples and stratified the risk and survival of HCC patients according to ubiquitin-related signatures. The differences in signaling pathways, molecular mechanisms, TME, and genomic alterations between the two subtypes identified suggest that the therapy sensitivity of the two subtypes will be distinct and should be targeted under specific therapeutic strategies. Moreover, a new prognostic signature consisting of eight ubiquitin-related genes was identified and validated. This signature can be used as a screening tool for patients at high risk of developing HCC. The identification of high-risk patients is crucial for early intervention and survival.

This study has a few limitations. First, this was a retrospective study, and all HCC patients were collected from public databases. Second, our HCC samples were smaller than those used in other studies that have integrated multiple datasets and may generate more comprehensive outcomes. Third, a large sample size of patients with HCC from our own hospital is needed for further prospective external validation, and future functional studies are essential to elucidate the precise roles of ubiquitin-related genes in the development of HCC.

Conclusions

Our study established a new classification for HCC based on the expression profiles of ubiquitin-related genes. In addition, we developed a prognostic signature using eight ubiquitin-related genes to screen patients at high risk of developing HCC.

MATERIALS AND METHODS

Data acquisition and processing

In total, 600 patients with HCC were selected from two datasets, including 371 American patients (a training cohort) from TCGA and 229 Japanese patients (an external validation cohort) from the ICGC. The ICGC Liver Cancer-RIKEN, Japan (LIRI-JP) dataset, including transcriptome data and clinical information, was downloaded from the ICGC website (https://dcc.icgc.org). Multi-omics data from the TCGA HCC cohort, including transcriptome data, somatic mutation data, CNV

Figure 6. Construction and evaluation of a ubiquitin-related signature

(A and C) Kaplan-Meier OS curves for patients assigned to high- and low-risk groups based on the risk score in the TCGA (A) and ICGC (C) cohorts. (B and D) Time-dependent ROC curves were performed in the TCGA (B) and ICGC (D) cohorts. (E) The heatmap of the eight ubiquitin-related genes in low- and high-risk groups. The distribution of clinicopathological features was compared between the low- and high-risk groups. *$p < 0.05$ and ***$p < 0.001$.

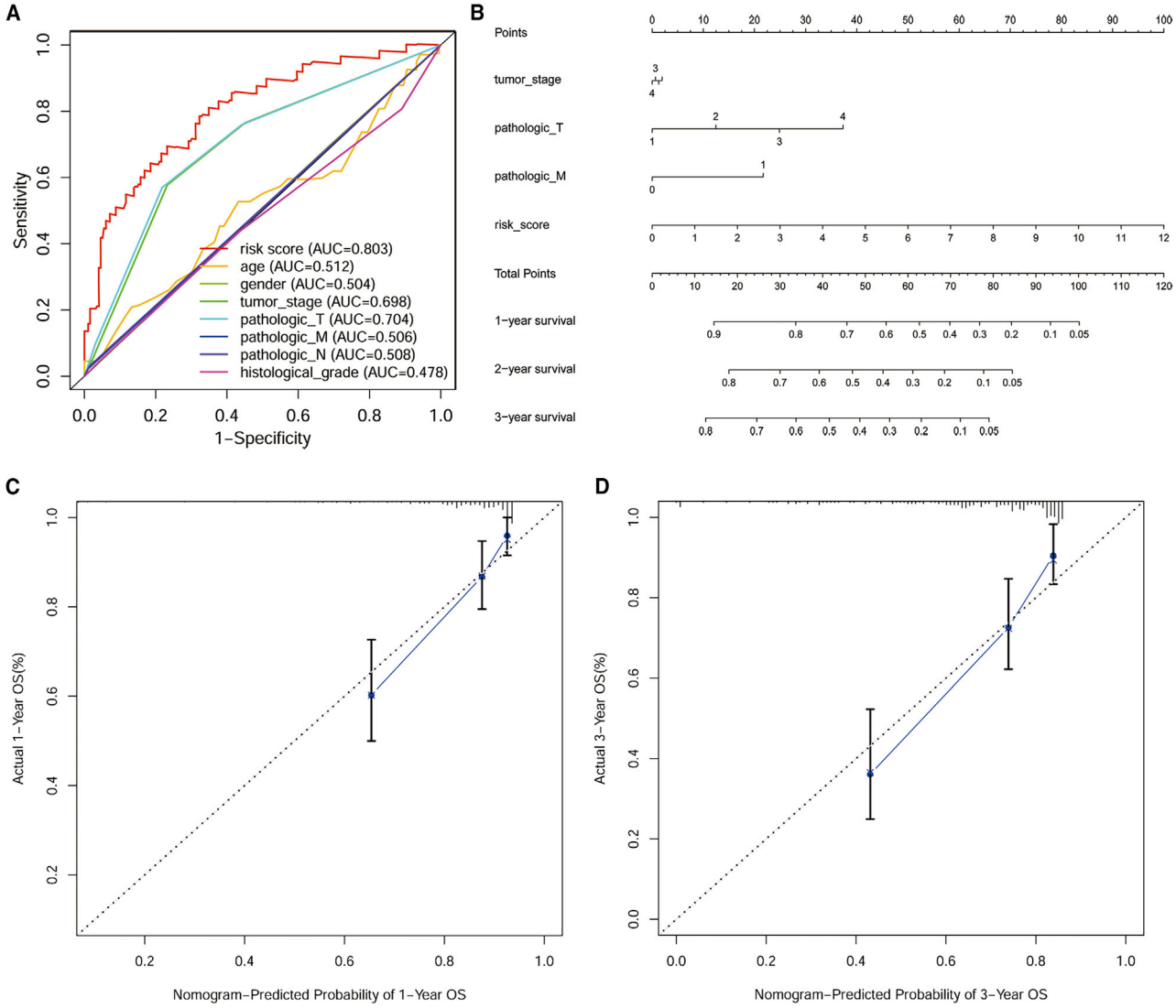


Figure 7. Integration of the ubiquitin-related signature and clinicopathologic features

(A) ROC curves show the sensitivity and specificity of the ubiquitin-related signature and clinicopathologic features in predicting the OS of HCC patients. (B) Nomogram constructed to predict the 1-, 2-, and 3-year OS in the TCGA cohort. (C and D) Calibration curves of the nomogram for predicting the probability of OS at 1 and 3 years.

data, and clinical information, was downloaded from the TCGA website (https://portal.gdc.cancer.gov/). By using the maftools R package, somatic mutation data were analyzed, summarized, visualized.[32] For CNV data, GISTIC 2.0 was performed to determine genes with significant amplification or deletion, as based on q values <0.25.[33]

List of ubiquitin-related genes

A list of 807 ubiquitin-related human genes was collected from the iUUCD 2.0 database (http://iuucd.biocuckoo.org/), including 8 E1s, 38 E2s, 501 E3s, 97 DUBs, 121 UBDs, and 42 ULDs.[34] We extracted 777 ubiquitin-related genes with available mRNA expression profiles from the TCGA HCC dataset.

Screening for HCC subtypes

On the basis of the expression profiles of ubiquitin-related genes, molecular subtypes were clustered and identified using the ConsensusClusterPlus R package.[35] The clustering was performed using the following settings: 50 iterations, 80% resampling rate, and Euclidean distance. The optimal cluster number was determined by constructing CDF curves.[36] A PCA was performed to compare the gene-expression patterns among different HCC subtypes.[37] A GSEA using the clusterprofiler R package was performed to investigate the pathways correlated with the different HCC subtypes.[38] The Kyoto Encyclopedia of Genes and Genomes (KEGG) gene sets retrieved from the Molecular Signatures Database were used for the GSEA, and the significance threshold was set at an adjusted p value of <0.05.

TME analysis

To evaluate the heterogeneity in the TME among different HCC subtypes, we inferred the immune and stromal cell components. Immune and stromal scores that can represent the infiltration of TME cells were calculated using ESTIMATE algorithm.[39] Specially, stromal score represents the infiltration of stroma cells in neoplastic samples, immune score represents the presence of immune cells in neoplastic samples, and ESTIMATE score infers tumor purity, which is the proportion of cancer cells in the admixture. In addition, the immune cell infiltration was estimated by performing a single-sample GSEA (ssGSEA) using the GSVA R package. Marker genes for 24 types of immune cells, including 11 innate and 13 adaptive immune cells, were acquired from a published study.[40] Based on the gene-expression profiles and marker genes, infiltrating immune cells were quantified by ssGSEA in individual tumor samples.[41]

Establishment and evaluation of a ubiquitin-related signature and nomogram

Univariate Cox, LASSO-penalized, and stepwise multivariate Cox regression analyses were performed sequentially to establish a ubiquitin-related prognostic signature.[42] Univariate regression analyses were used to identify ubiquitin-related genes with prognostic value. If $p < 0.01$, the corresponding genes were considered as prognostic genes. LASSO-penalized regression analysis was then performed to screen prognostic genes. Finally, stepwise multivariate regression analysis was conducted to further screen prognostic genes by using the lowest Akaike information criterions (AIC) value and establish the ubiquitin-related signature. The risk score was calculated as follows: risk score = $\Sigma(Ci \times Expi)$, where "C" is the coefficient of gene derived from the multivariate Cox analysis and "Exp" is the gene expression level.[43] We drew ROC curves to compare the prediction efficiency of the signature and several clinicopathological features.[44] Univariate and multivariate Cox regression analyses were performed to investigate prognostic values for the signature and several clinicopathological features. A nomogram was constructed to predict the probability of OS in HCC patients using the rms R package. Calibration curves were drawn to evaluate the effectiveness of the nomogram using the rms R package.

Statistical analyses

A chi-square test was performed to compare the distribution of clinicopathological features, including age, sex, T/N/M pathological stage, clinical stage, and histological grade, between the different groups. Differences between two groups were tested using the Wilcoxon rank test for non-normally distributed variables, and the unpaired t test for normally distributed variables. The patients were categorized into low- and high-risk groups according to median risk score. Survival analyses were performed to compare the survival of the patients in the different subtypes or in the high- and low-risk groups. All statistical analyses were performed using the R software (version 3.5.2) and GraphPad Prism software programs (version 7). All statistical results with a p value of <0.05 were considered significant.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omto.2021.04.003.

ACKNOWLEDGMENTS

We thank the other members of our research team for their assistance. The LIRI-JP dataset, including transcriptome data and clinical information, was acquired from the ICGC portal (https://dcc.icgc.org). The TCGA HCC dataset, including transcriptome data, somatic mutation data, CNV data, and clinical information, was downloaded from the data portal (https://portal.gdc.cancer.gov/). A list of 807 ubiquitin-related human genes was collected from the iUUCD 2.0 database (http://iuucd.biocuckoo.org/).

AUTHOR CONTRIBUTIONS

S.Y. and B.Y. collected and analyzed the data and wrote the manuscript. L.W., Y.L., and K.L. analyzed the data and reviewed the manuscript. P.X., Y.Z., and Y.D. participated in analyzing the data. Z.Z., Y.W., N.L., and D.Z. participated in preparation of the figures and tables. Z.D. and H.K. designed the study and revised the manuscript. All the authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Siegel, R.L., Miller, K.D., and Jemal, A. (2020). Cancer statistics, 2020. CA Cancer J. Clin. *70*, 7–30.
2. Llovet, J.M., Zucman-Rossi, J., Pikarsky, E., Sangro, B., Schwartz, M., Sherman, M., and Gores, G. (2016). Hepatocellular carcinoma. Nat. Rev. Dis. Primers *2*, 16018.
3. Ernst, A., Avvakumov, G., Tong, J., Fan, Y., Zhao, Y., Alberts, P., Persaud, A., Walker, J.R., Neculai, A.-M., Neculai, D., et al. (2013). A strategy for modulation of enzymes in the ubiquitin system. Science *339*, 590–595.
4. Scheffner, M., Nuber, U., and Huibregtse, J.M. (1995). Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. Nature *373*, 81–83.
5. Heride, C., Urbé, S., and Clague, M.J. (2014). Ubiquitin code assembly and disassembly. Curr. Biol. *24*, R215–R220.
6. Kerscher, O., Felberbaum, R., and Hochstrasser, M. (2006). Modification of proteins by ubiquitin and ubiquitin-like proteins. Annu. Rev. Cell Dev. Biol. *22*, 159–180.
7. Hochstrasser, M. (2009). Origin and function of ubiquitin-like proteins. Nature *458*, 422–429.
8. Swatek, K.N., and Komander, D. (2016). Ubiquitin modifications. Cell Res. *26*, 399–422.
9. Popovic, D., Vucic, D., and Dikic, I. (2014). Ubiquitination in disease pathogenesis and treatment. Nat. Med. *20*, 1242–1253.
10. Senft, D., Qi, J., and Ronai, Z.A. (2018). Ubiquitin ligases in oncogenic transformation and cancer therapy. Nat. Rev. Cancer *18*, 69–88.
11. Wade, M., Li, Y.-C., and Wahl, G.M. (2013). MDM2, MDMX and p53 in oncogenesis and cancer therapy. Nat. Rev. Cancer *13*, 83–96.
12. Qi, J., and Ronai, Z.A. (2015). Dysregulation of ubiquitin ligases in cancer. Drug Resist. Updat. *23*, 1–11.
13. Liu, Y., Tao, S., Liao, L., Li, Y., Li, H., Li, Z., Lin, L., Wan, X., Yang, X., and Chen, L. (2020). TRIM25 promotes the cell survival and growth of hepatocellular carcinoma through targeting Keap1-Nrf2 pathway. Nat. Commun. *11*, 348.
14. Wu, H., Yang, T.-Y., Li, Y., Ye, W.-L., Liu, F., He, X.-S., Wang, J.-R., Gan, W.-J., Li, X.-M., Zhang, S., et al. (2020). Tumor Necrosis Factor Receptor-Associated Factor 6

Promotes Hepatocarcinogenesis by Interacting With Histone Deacetylase 3 to Enhance c-Myc Gene Expression and Protein Stability. Hepatology *71*, 148–163.

15. Pal, A., Young, M.A., and Donato, N.J. (2014). Emerging potential of therapeutic targeting of ubiquitin-specific proteases in the treatment of cancer. Cancer Res. *74*, 4955–4966.
16. Dimri, M., and Satyanarayana, A. (2020). Molecular Signaling Pathways and Therapeutic Targets in Hepatocellular Carcinoma. Cancers (Basel) *12*, 491.
17. Heim, M.H. (2013). Innate immunity and HCV. J. Hepatol. *58*, 564–574.
18. Qian, F., Bolen, C.R., Jing, C., Wang, X., Zheng, W., Zhao, H., Fikrig, E., Bruce, R.D., Kleinstein, S.H., and Montgomery, R.R. (2013). Impaired toll-like receptor 3-mediated immune responses from macrophages of patients chronically infected with hepatitis C virus. Clin. Vaccine Immunol. *20*, 146–155.
19. Cheng, S., Zhang, B., Du, J.-Y., Jin, Y.-H., Lang, H.-Y., and Zeng, L.-H. (2017). Hepatitis B Surface Antigen Promotes the Invasion of Hepatitis B Virus-Related Hepatocellular Carcinoma Cells by Upregulation of Toll-Like Receptor 2. Viral Immunol. *30*, 232–239.
20. Lokau, J., Schoeder, V., Haybaeck, J., and Garbers, C. (2019). Jak-Stat Signaling Induced by Interleukin-6 Family Cytokines in Hepatocellular Carcinoma. Cancers (Basel) *11*, 1704.
21. Israelow, B., Narbus, C.M., Sourisseau, M., and Evans, M.J. (2014). HepG2 cells mount an effective antiviral interferon-lambda based innate immune response to hepatitis C virus infection. Hepatology *60*, 1170–1179.
22. Hu, B., Ding, G.-Y., Fu, P.-Y., Zhu, X.-D., Ji, Y., Shi, G.-M., Shen, Y.-H., Cai, J.-B., Yang, Z., Zhou, J., et al. (2018). NOD-like receptor X1 functions as a tumor suppressor by inhibiting epithelial-mesenchymal transition and inducing aging in hepatocellular carcinoma cells. J. Hematol. Oncol. *11*, 28.
23. Kurebayashi, Y., Ojima, H., Tsujikawa, H., Kubota, N., Maehara, J., Abe, Y., Kitago, M., Shinoda, M., Kitagawa, Y., and Sakamoto, M. (2018). Landscape of immune microenvironment in hepatocellular carcinoma and its additional impact on histological and molecular classification. Hepatology *68*, 1025–1041.
24. Li, X.-S., Li, J.-W., Li, H., and Jiang, T. (2020). Prognostic value of programmed cell death ligand 1 (PD-L1) for hepatocellular carcinoma: a meta-analysis. Biosci. Rep. *40*, BSR20200459.
25. Dong, H., Strome, S.E., Salomao, D.R., Tamura, H., Hirano, F., Flies, D.B., Roche, P.C., Lu, J., Zhu, G., Tamada, K., et al. (2002). Tumor-associated B7-H1 promotes T-cell apoptosis: a potential mechanism of immune evasion. Nat. Med. *8*, 793–800.
26. Santoro, A., Rimassa, L., Borbath, I., Daniele, B., Salvagni, S., Van Laethem, J.L., Van Vlierberghe, H., Trojan, J., Kolligs, F.T., Weiss, A., et al. (2013). Tivantinib for second-line treatment of advanced hepatocellular carcinoma: a randomised, placebo-controlled phase 2 study. Lancet Oncol. *14*, 55–63.
27. Abou-Alfa, G.K., Meyer, T., Cheng, A.-L., El-Khoueiry, A.B., Rimassa, L., Ryoo, B.-Y., Cicin, I., Merle, P., Chen, Y., Park, J.-W., et al. (2018). Cabozantinib in Patients with Advanced and Progressing Hepatocellular Carcinoma. N. Engl. J. Med. *379*, 54–63.
28. Pan, Y.-H., Yang, M., Liu, L.-P., Wu, D.-C., Li, M.-Y., and Su, S.-G. (2018). UBE2S enhances the ubiquitination of p53 and exerts oncogenic activities in hepatocellular carcinoma. Biochem. Biophys. Res. Commun. *503*, 895–902.
29. Qu, C., and Qu, Y. (2017). Down-regulation of salt-inducible kinase 1 (SIK1) is mediated by RNF2 in hepatocarcinogenesis. Oncotarget *8*, 3144–3155.
30. Shen, M., and Lin, L. (2019). Functional variants of autophagy-related genes are associated with the development of hepatocellular carcinoma. Life Sci. *235*, 116675.
31. Cui, M., Sun, J., Hou, J., Fang, T., Wang, X., Ge, C., Zhao, F., Chen, T., Xie, H., Cui, Y., et al. (2016). The suppressor of cytokine signaling 2 (SOCS2) inhibits tumor metastasis in hepatocellular carcinoma. Tumour Biol. *37*, 13521–13531.
32. Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C., and Koeffler, H.P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. *28*, 1747–1756.
33. Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. *12*, R41.
34. Zhou, J., Xu, Y., Lin, S., Guo, Y., Deng, W., Zhang, Y., Guo, A., and Xue, Y. (2018). iUUCD 2.0: an update with rich annotations for ubiquitin and ubiquitin-like conjugations. Nucleic Acids Res. *46* (D1), D447–D453.
35. Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics *26*, 1572–1573.
36. Li, W., Wang, H., Ma, Z., Zhang, J., Ou-Yang, W., Qi, Y., and Liu, J. (2019). Multi-omics Analysis of Microenvironment Characteristics and Immune Escape Mechanisms of Hepatocellular Carcinoma. Front. Oncol. *9*, 1019.
37. Raychaudhuri, S., Stuart, J.M., and Altman, R.B. (2000). Principal components analysis to summarize microarray experiments: application to sporulation time series. Pac. Symp. Biocomput. *2000*, 455–466.
38. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287.
39. Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. Nat. Commun. *4*, 2612.
40. Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A.C., Angell, H., Fredriksen, T., Lafontaine, L., Berger, A., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity *39*, 782–795.
41. Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell *160*, 48–61.
42. Wang, H., Lengerich, B.J., Aragam, B., and Xing, E.P. (2019). Precision Lasso: accounting for correlations and linear dependencies in high-dimensional genomic data. Bioinformatics *35*, 1181–1187.
43. Yang, S., Wu, Y., Deng, Y., Zhou, L., Yang, P., Zheng, Y., Zhang, D., Zhai, Z., Li, N., Hao, Q., et al. (2019). Identification of a prognostic immune signature for cervical cancer to predict survival and response to immune checkpoint inhibitors. OncoImmunology *8*, e1659094.
44. Lorent, M., Giral, M., and Foucher, Y. (2014). Net time-dependent ROC curves: a solution for evaluating the accuracy of a marker to predict disease-related mortality. Stat. Med. *33*, 2379–2389.