

Development of an immune-related prognostic index associated with hepatocellular carcinoma

Bo Hu¹, Xiao-Bo Yang¹, Xin-Ting Sang¹

¹Department of Liver Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100010, China

Correspondence to: Xin-Ting Sang; **email:** sangxt@pumch.cn

Keywords: hepatocellular carcinoma, oncogenes, immune system, prognosis, precision medicine

Received: December 13, 2019

Accepted: March 2, 2020

Published: March 19, 2020

Copyright: Hu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Liver hepatocellular carcinoma (LIHC), an inflammation-associated cancer induced by a variety of etiological factors, is still one of the most prevalent and lethal cancers in human population. In this study, the expression profiles of immune-related genes (IRGs) were integrated with the overall survival (OS) of 378 LIHC patients based on the Cancer Genome Atlas (TCGA) dataset. Moreover, the differentially expressed and survival related IRGs among LIHC patients were predicted through the computational difference algorithm and COX regression analysis. As a result, 7 genes, including HSPA4, S100A10, FABP6, CACYBP, HDAC1, FCGR2B and SHC1, were retrieved to construct a predictive model associated with the overall survival (OS) of LIHC patients. Typically, the as-constructed model performed moderately in predicting prognosis, which was also correlated with tumor grade. Functional enrichment analysis revealed that the genes of high-risk group were actively involved in mRNA binding and the spliceosome pathway. Intriguingly, the prognostic index established based on IRGs reflected infiltration by multiple types of immunocytes. Our findings screen several IRGs with clinical significance, reveal the drivers of immune repertoire, and illustrate the importance of a personalized, IRG-based immune signature in LIHC recognition, surveillance, and prognosis prediction.

INTRODUCTION

Liver hepatocellular carcinoma (LIHC) is the sixth most prevalent cancer and third leading cause of cancer-related deaths worldwide [1]. It can be managed by surgical treatment and chemotherapy, but the mortality rate remains high [2]. For patients with advanced LIHC, chemotherapy fails to demonstrate a survival benefit [3, 4]. By contrast, therapeutic regimens that counteract the immunosuppressive mechanisms have the potential to dramatically alter the clinical outcomes of LIHC, which lead us to further explore the relationship of the abnormal immune gene expression with LIHC development and prognosis [5]. Cancer immunotherapy is a primary driver of personalized medicine, which makes aggressive efforts to leverage the immune system against tumors and is thus promising for treating human diseases [6, 7]. As an inflammation-associated tumor,

the immunosuppressive microenvironment of LIHC is well evidenced to promote immune tolerance and evasion through various mechanisms, rendering LIHC an attractive candidate for immunotherapy [8]. Recently, immune checkpoint inhibitors (ICIs), including ipilimumab (the CTLA4 inhibitor) and nivolumab (the PD-1 inhibitor), have demonstrated great survival benefits for LIHC [9, 10]. According to a phase I/II study on nivolumab among patients with advanced LIHC, a 19% response rate (including a 5% complete response rate) is achieved, which is expected to be further increased by the combined therapies with different ICIs or the combined treatment of small molecules with ICIs [10]. However, it remains largely unknown about the molecular features of immune checkpoints and their correlations with clinicopathological parameters and LIHC tumor microenvironment (TME). Moreover, the molecular

mechanisms of immunology in LIHC, especially for the immunogenomic effects, are still unclear so far. With the establishment and completion of the large-scale gene expression datasets, cancer researchers are able to identify the responsible biomarkers for tumor monitoring and surveillance in a rapid and accurate way [11–13]. For instance, Yang et al. proposed an immune-related classifier for the prognosis of cutaneous melanoma, which was derived from immune-related genes (IRGs) and demonstrated a powerful predictive ability [14]. Nevertheless, the clinical relevance and prognostic significance of IRGs in LIHC have to be explored.

This study aimed to gain insight into the potential clinical utility of IRGs in prognosis stratification, as well as their implicational potential as biomarkers for targeted LIHC therapy. We are committed to construct a robust immune-related signature to improve prognostic prediction of LIHC via comprehensive genomic data analysis. Results obtained in this study can provide certain foundation for subsequent in-depth immune-related studies, which show great promise for treating LIHC with personalized medicine.

RESULTS

Identification of the differentially expressed IRGs

A total of 7667 differentially expressed genes (DEGs) were identified using the edgeR algorithm, including 7273 up-regulated and 394 down-regulated ones (Figure 1A and 1B). Then, altogether 329 differentially expressed IRGs were extracted from these 7667 DEGs, including 267 up-regulated and 62 down-regulated ones (Figure 1C and 1D).

Identification of the survival-related IRGs and construction of TF regulatory network

After screening, 61 IRGs were identified to show remarkable correlation with OS for TCGA-LIHC patients ($P < 0.01$) (Supplementary Table 1). To explore the potential molecular mechanisms corresponding to the clinical significance of our survival-related IRGs, the regulatory mechanisms of these genes were also investigated. Specifically, the expression patterns of 318 TFs were examined, which suggested that 116 of them were differentially expressed between TCGA-LIHC

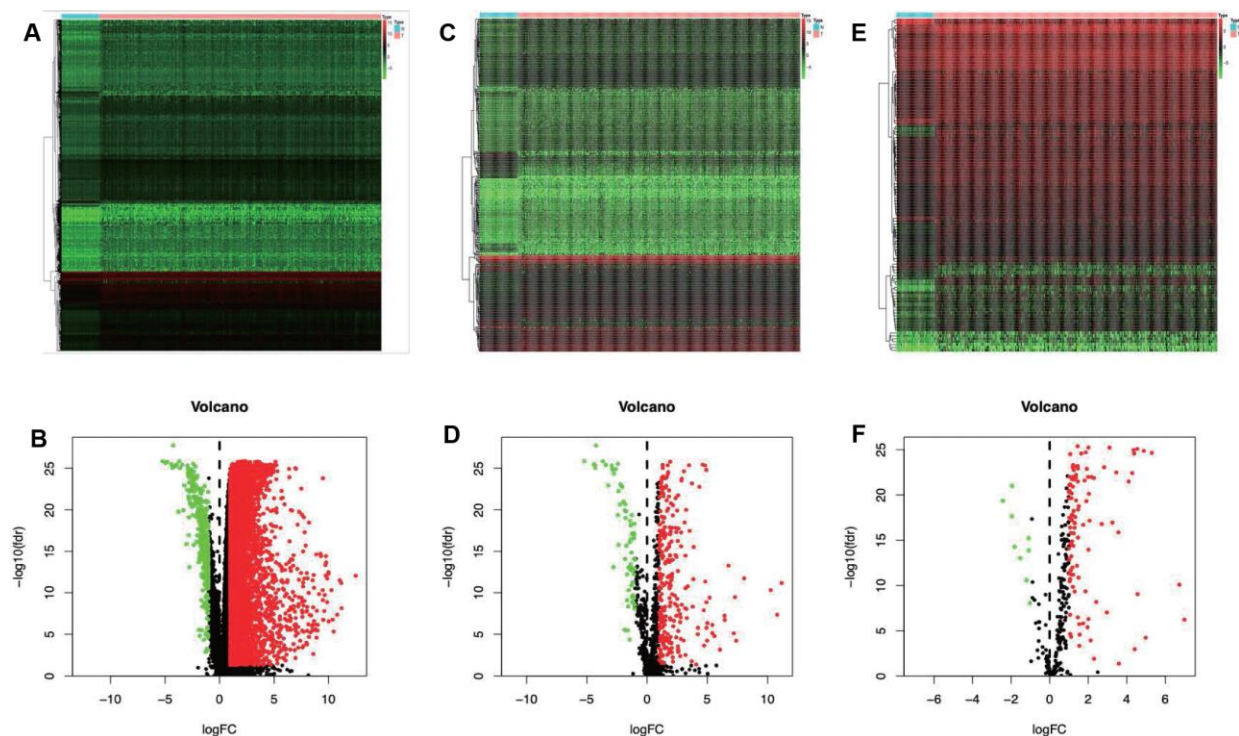


Figure 1. Differentially expressed immune-related genes and transcription factors (TFs). Heatmap (A) and volcano plot (C) demonstrating differentially expressed genes between hepatocellular carcinoma (LIHC) and non-tumor tissues. Differentially expressed immune-related genes (IRGs) are shown in heatmap (B) and volcano plot (D), red and green dots represent differentially expressed genes. Heatmap (E) and volcano plot (F) illustrating differentially TFs between LIHC and non-tumor tissues, red dots represent differentially up-regulated TFs. Red dots represent differentially up-regulated expressed genes, green dots represent differentially down-regulated expressed genes and black dots represent no differentially expressed genes. N, normal tissue. T, tumor.

and non-tumor hepatic samples (Figure 1E and 1F). Afterwards, the associations of these differentially expressed TFs with the prognostic immune genes were also analyzed. Finally, 64 factors were screened to construct the regulatory networks with 58 survival-related immune genes, as selected at the correlation score of > 0.4 and the P-value of < 0.001 . Notably, the TF-based regulatory schematic acutely illustrated the regulatory relationships among these IRGs (Figure 2).

Development and validation of the immune-related prognostic signature in TCGA

The prognostic signature was established based on the LASSO-penalized Cox regression analysis results (Figure 3A and 3B), so as to stratify TCGA-LIHC patients as two groups, namely, the low-risk group and the high-risk group, with discrete clinical outcomes with regard to OS. To be specific, the following formula was

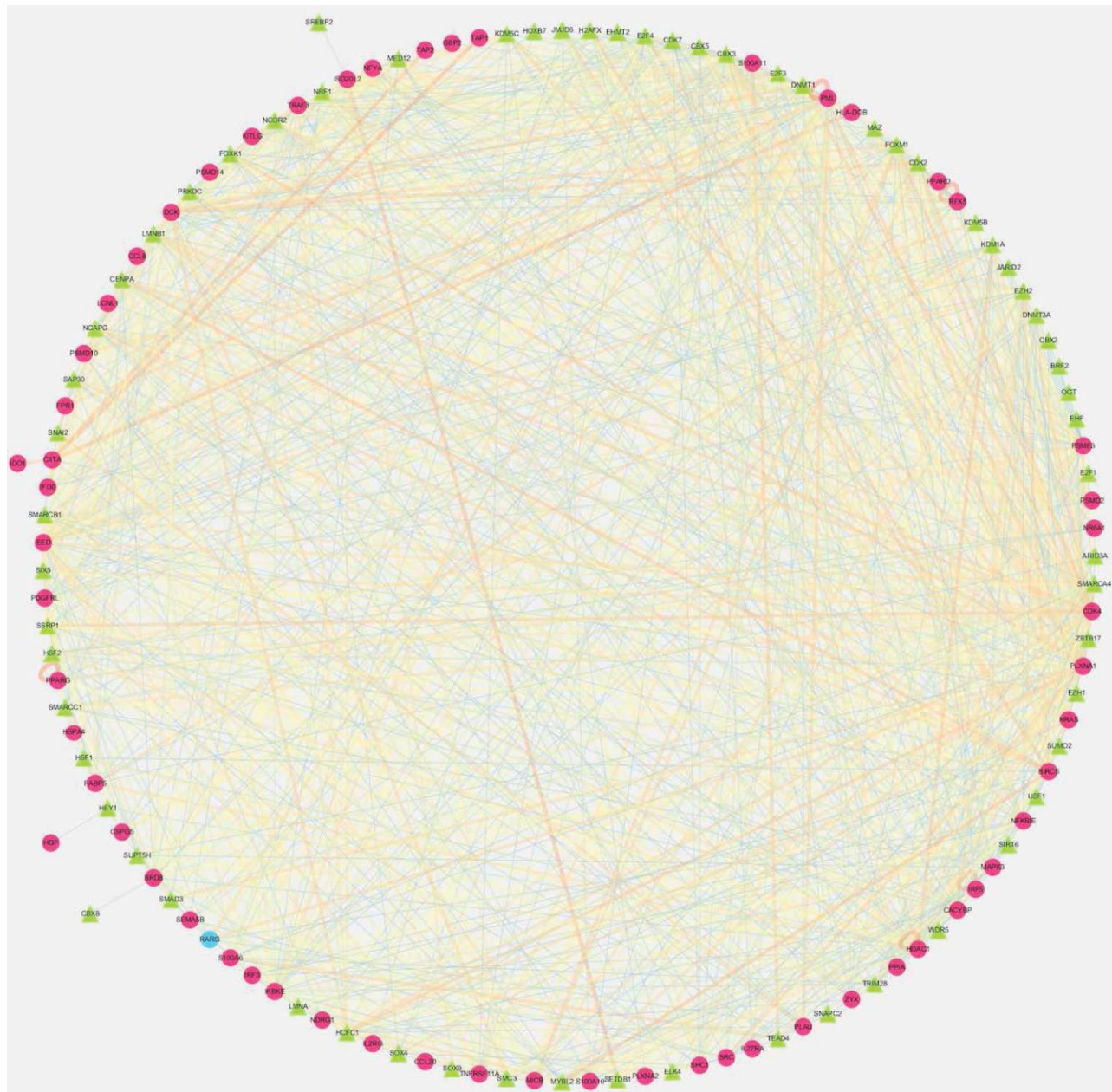


Figure 2. Transcription factors (TFs)- mediated regulatory network. Regulatory network constructed based on survival relevant TFs and IRGs. The red circle represents high-risk genes, the blue circle represents low-risk genes. The green triangle represents transcription factors, and the thickness and brightness of lines between nodes represent the level of relevance (low Cor values to small sizes and dark colors).

adopted for calculation: [Expression level of HSPA4* (0.0361)] + [Expression level of S100A10* (-0.0043)] + [Expression level of CACYBP* (0.0416)] + [Expression level of FABP6 * (0.0797)] + [Expression level of HDAC1* (0.0247)] + [Expression level of FCGR2B* (0.1551)] + [Expression level of SHC1* (0.0091)] (Table 1). Figure 3C shows the comparisons of survival differences between the two groups in training set ($P < 0.001$). Moreover, such findings were further verified in the testing set and the entire set ($P < 0.01$) (Figure 3D and 3E). Additionally, the AUC for 1-year OS were 0.821, 0.833 and 0.828 in the training set, the testing set and the entire set, separately, which indicated moderate potentials for the metabolic gene signature to monitor survival (Figure 3F–3H). Our model achieved the greatest AUC value compared with those of other clinicopathological characteristics, which also reflected its excellent predicting ability. In the entire set, the low-risk group showed markedly superior prognosis to the high-risk group among all ($\leq 65 / > 65$; Figure 4A and 4B), sex (female/male; Figure 4C and 4D), stage (I+II/III+IV; Figure 4E and 4F), grade (G1+G2/G3+G4; Figure 4G and 4H) and T stage

(T1+T2/T3+T4; Figure 4I and 4J) subgroups. Similar results were also observed in patients without distant metastasis (Figure 4K) and lymph node metastasis (Figure 4L). Additionally, the multivariate analysis revealed that the as-constructed immune gene signature could become an independent predictor after other clinicopathological parameters were adjusted (Figure 5). Figure 6 depicts the risk score distribution (A, D, G), survival status (B, E, H) and expression of the metabolic gene signature (C, F, I) in the training set (A–C), the testing set (D–F) and the entire set (G–I). Clearly, their distributions were similar, thus supporting the robust predictive capacity of our immune related gene-based risk score assessment model. Owing to the significant clinical value of the signature genes, we embarked on a comprehensive investigation of their molecular characteristics. Genetic alternations of these genes in LIHC (TCGA-PanCancer Atlas) were explored via cBioPortal, and amplification was found to be the most commonly occurring type of mutation (Supplementary Figure 1). Besides, there were 4 genes (*SHC1*, *FCGR2B*, *S100A10* and *CACYBP*) with a mutation rate $\geq 5\%$.

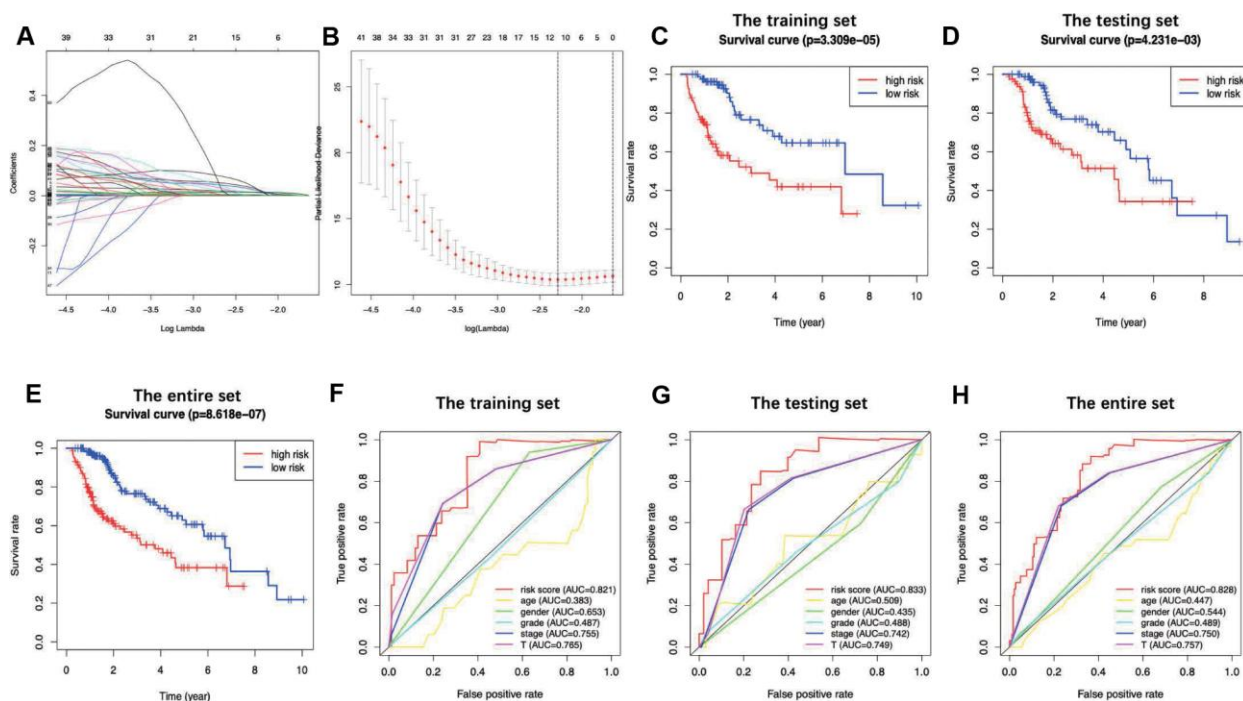


Figure 3. LASSO coefficient profiles of the 7 immune genes are depicted in (A, B) show the selection of the tuning parameter (lambda) in the LASSO model by tenfold cross-validation based on minimum criteria for OS; the lower X axis shows log (lambda), and the upper X axis shows the average number of OS-genes. The Y axis indicates partial likelihood deviance error. Red dots represent average partial likelihood deviances for every model with a given lambda, and vertical bars indicate the upper and lower values of the partial likelihood deviance errors. The vertical black dotted lines define the optimal values of lambda, which provides the best fit. Survival curves of patients in high risk group and low risk group of the training set (C), the testing set (D) and the entire set (E) are shown. Patients in high-risk group suffered shorter overall survival. (F–H) show survival-dependent receiver operating characteristic (ROC) curves validation at 1 – year of prognostic value of the prognostic index in the three sets (the training set, the testing set and the entire set, respectively).

Table 1. Seven immune-related signature genes identified from Cox regression analysis from TCGA.

id	coef	HR	HR.95L	HR.95H	P-value
HSPA4	0.03606573	1.03672399	1.00620270	1.06817108	0.01800376
S100A10	0.00428466	1.00429385	1.00126694	1.00732992	0.00540109
FABP6	0.07972664	1.08299099	1.01358048	1.15715476	0.01831907
CACYBP	0.04157927	1.04245579	0.99974327	1.08699314	0.05142275
HDAC1	0.02472582	1.02503404	1.00243637	1.04814112	0.02971183
FCGR2B	0.15512764	1.16780702	1.07395448	1.26986129	0.00028443
SHC1	0.00914823	1.00919021	1.00025144	1.01820885	0.04386807

Coef, coefficient; HR, hazard ratio; L, low; H, high.

PCA proved the model grouping capacity

PCA was further conducted to examine the difference between low- and high-risk groups based on the immune-related signature (Figure 7A), immune genes (Figure 7B), differently expressed genes (Figure 7C) and the entire gene expression profiles (Figure 7D). The

results obtained based on our model showed that low- and high-risk groups were generally distributed at different directions. Nonetheless, the distributions of high- and low-risk groups displayed in Figure 7B–7D were relatively scattered, which confirmed that our prognosis signature was capable of distinguishing the high-risk group from the low-risk group.

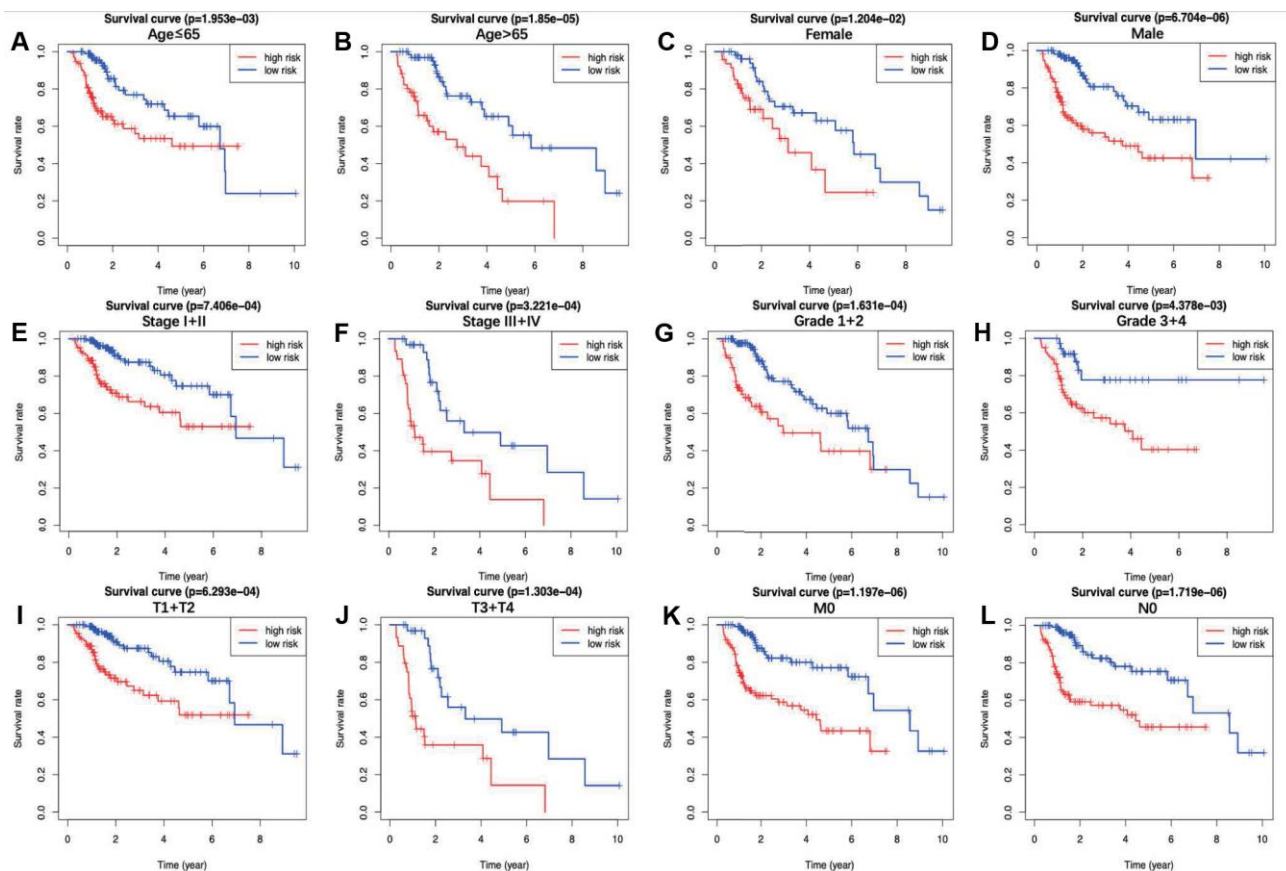


Figure 4. The overall survival differences between the high-risk group and the low-risk group were shown under the conditions of classifying patients by age (A, B), sex (C, D), stage (E, F), grade (G, H) and T stage (I, J). Patients without distant metastasis (K) and lymph node metastasis (L) are also displayed. Detailed notes are described in the main text.

Correlation of the prognostic model with clinicopathological characteristics

Two hundred and twenty-one patients with complete information including gender, age, tumor grade, clinical stage, T stage, lymph node metastasis and distant

metastasis were included in TCGA-LIHC cohort. Among the signature genes we researched, *HSPA4*, *FABP6*, *S100A10*, *CACYBP*, *SHC1* and *HDAC1* were associated with a higher tumor grade (Figure 8A–8F), *CACYBP* was linked with a higher clinical stage as well as T stage (Figure 8G and 8H). Additionally, the

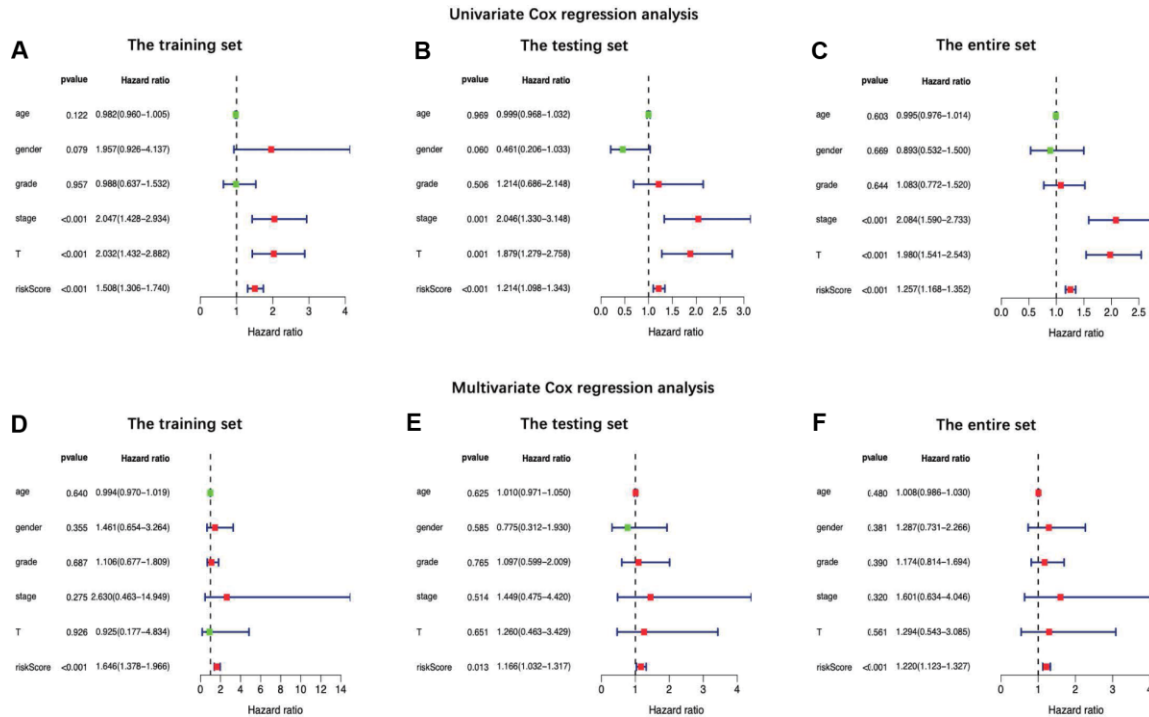


Figure 5. Univariate (A–C) and multiple (D–F) regression analysis of hepatocellular carcinoma and the relationships between the age, gender, grade, stage, T stage, distant metastasis, lymph node metastasis and riskScore in the training set (A and D), the testing set (B and E) and the entire set (C and F). Green dot means hazard ratio (HR) median value is less than 1, red dot means HR median value is greater than 1.

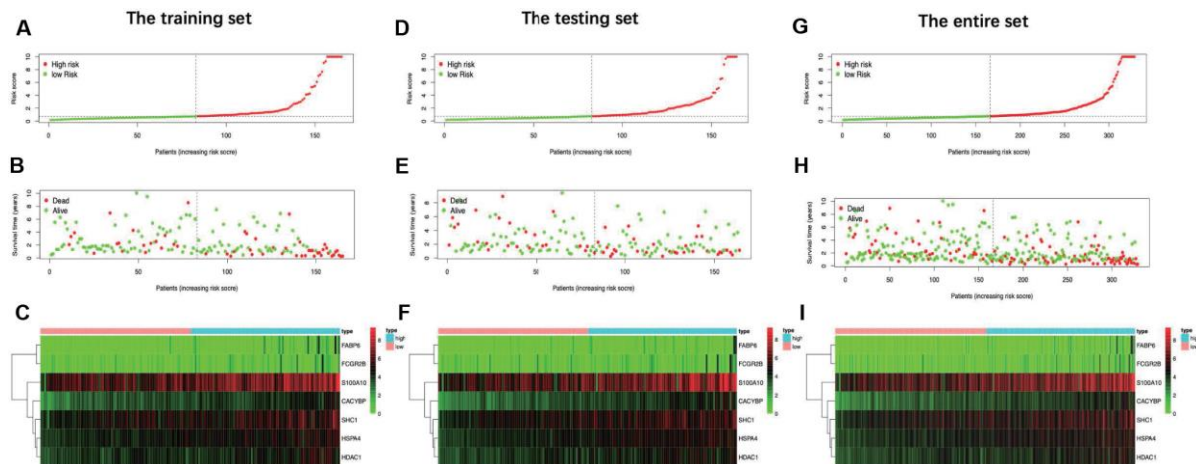


Figure 6. Distribution of risk score, overall survival (OS), gene expression in (A–C) training set, (D–F) testing set and (G–I) entire set. Distribution of risk score, OS and heat map of the expression of 7 signature genes in low-risk and high-risk groups are listed in the picture from top to bottom.

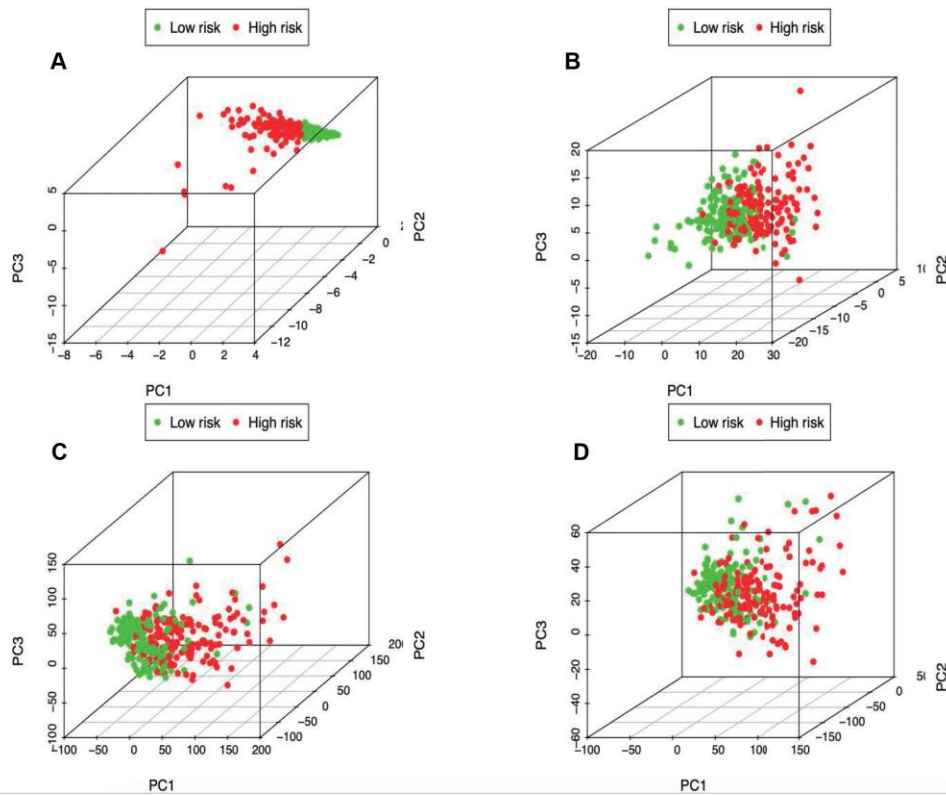


Figure 7. Principal components analysis between low- and high-risk groups based on (A) immune-related signature, (B) immune-related genes, (C) differentially expressed genes and (D) the entire gene expression profiles.

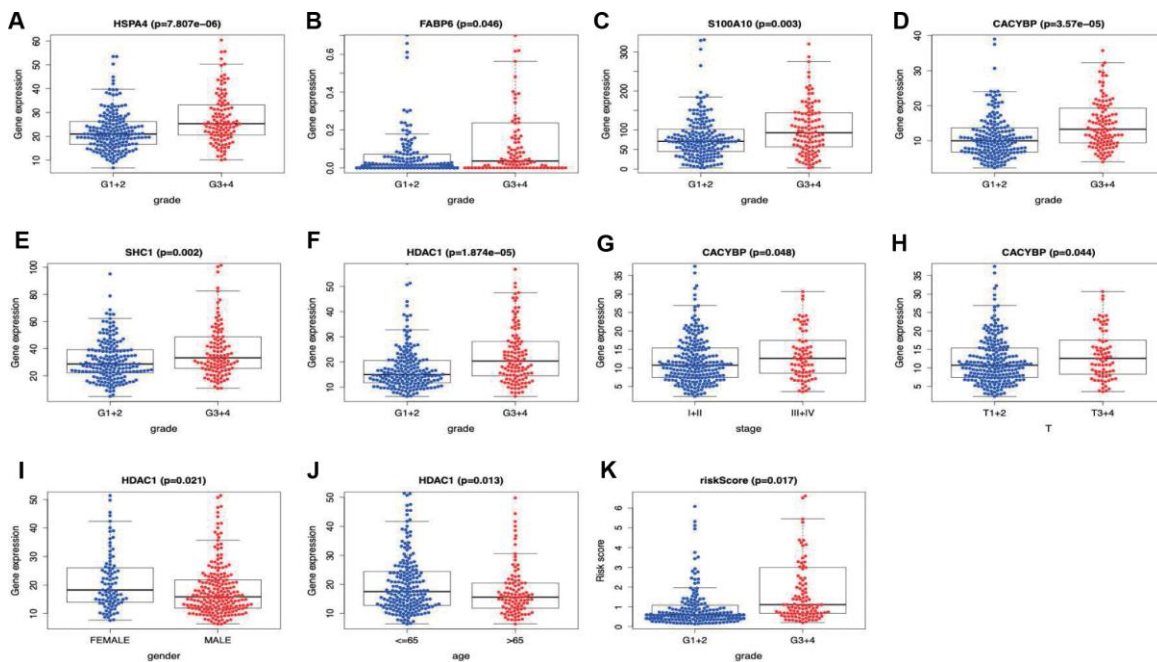


Figure 8. Correlation of the prognostic immune-related signature with clinicopathological characteristics. *HSPA4*, *FABP6*, *S100A10*, *CACYBP*, *SHC1* and *HDAC1* were associated with a higher tumor grade (A–F), *CACYBP* was linked with a higher clinical stage (G) as well as T stage (H). The expression level of *HDAC1* was significantly enhanced in female patients (I) and patients younger than 65 years old (J). Risk score derived from our model was significantly associated with higher tumor grade (K).

expression level of *HDAC1* was significantly enhanced in female patients and patients younger than 65 years old (Figure 8I and 8J). Afterwards, risk score derived from our model was significantly associated with higher tumor grade (Figure 8K).

Functional enrichment analysis revealed different states between high- and low-risk groups

GSEA was performed to further investigate the differences between the high- and low-risk groups. The results revealed that the GO molecular function “mRNA binding” (Figure 9A), biological process “Regulation of

cell cycle phase transition” (Figure 9B) and “Nuclear transport” (Figure 9C) were differentially enriched in high-risk phenotype ($P < 0.01$), while biological process “Organic acid catabolic process” (Figure 9D), molecular function “Steroid hydroxylase activity” (Figure 9E) and biological process “Cellular amino acid catabolic process” (Figure 9F) were closely correlated with the low risk phenotype ($P < 0.01$). In addition, KEGG pathway analysis suggested that the genes in high-risk group were mainly enriched in the “Spliceosome” (Supplementary Figure 2A), “RNA degradation” (Supplementary Figure 2B) and “Oocyte meiosis” (Supplementary Figure 2C) ($P < 0.01$); in

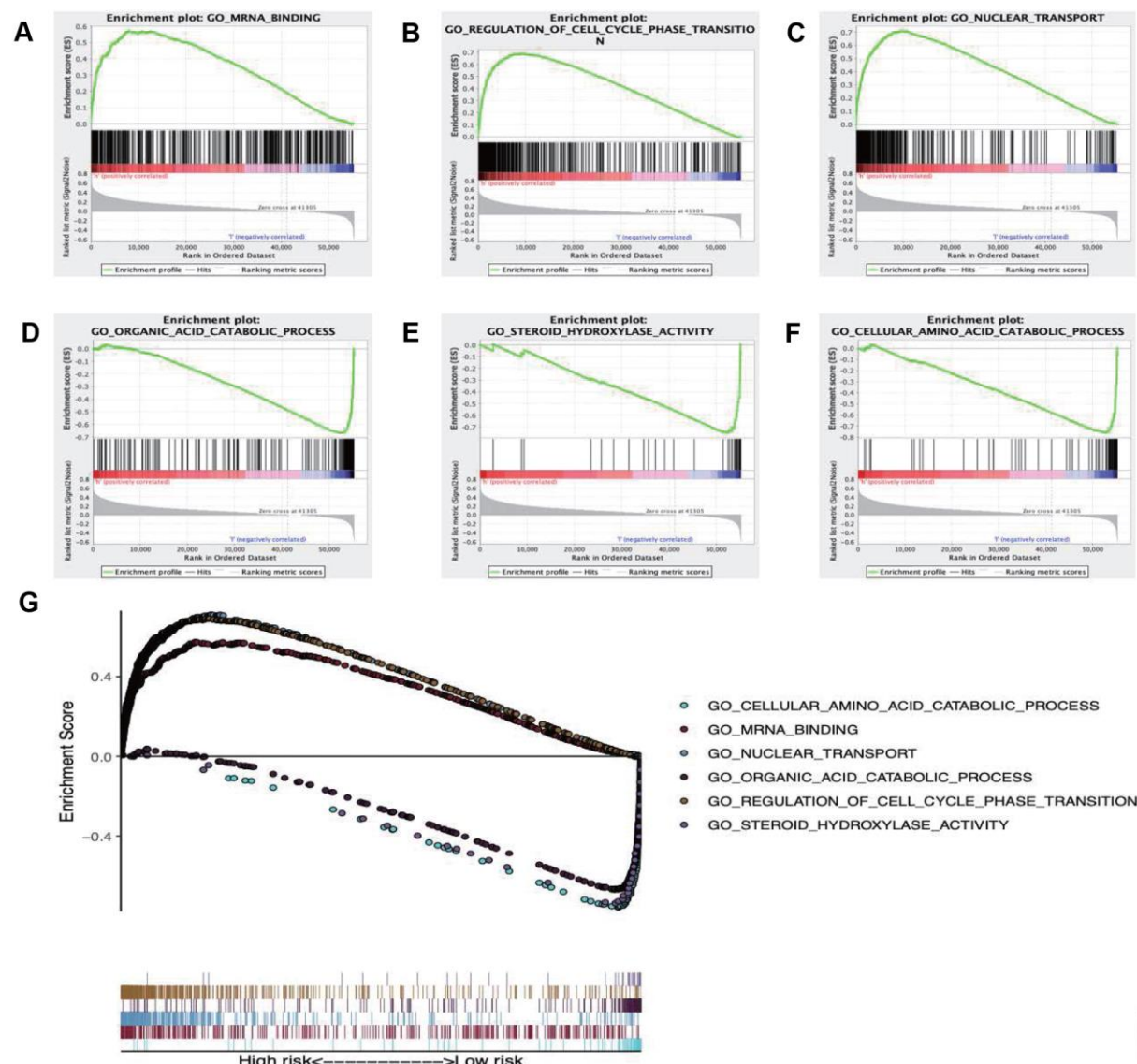


Figure 9. Enrichment plots of Gene Ontology annotation from gene set enrichment analysis (GSEA). GSEA results showing (A) mRNA binding, (B) Regulation of cell cycle phase transition, (C) Nuclear transport were differentially enriched in high risk phenotype, while (D) Organic acid catabolic process, (E) Steroid hydroxylase activity (F) Cellular amino acid catabolic process were closely correlated with the low risk phenotype. (G) Summarizes the above six gene sets.

addition, the “Complement and coagulation cascades” (Supplementary Figure 2D), “Glycine serine and threonine metabolism” (Supplementary Figure 2E) and “Primary bile acid biosynthesis” (Supplementary Figure 2F) were primarily enriched in low risk group ($P < 0.01$). Moreover, the immune status between the low- and high- risk group was also examined via GSEA, and the results suggested that the differentially expressed genes between these two groups were enriched in the immunological signature gene sets (c7. All. V7.0. symbol). According to the normalized enrichment score (NES), the top six immune related gene sets are shown in Table 2. Furthermore, the relationship of the prognostic signature with immune cell infiltration in TCGA-LIHC patients was investigated to examine whether the risk score partially reflected the tumor immune microenvironment status (Figure 10). Our results suggested that, for high risk patients in the entire set, the levels of macrophages (Cor=0.468; $p=7.594e-14$), neutrophils (Cor=0.479; $p=1.475e-14$) and DCs (Cor=0.358; $p=2.447e-08$), significantly increased in tumor microenvironment (TME) (Figure 10A–10C). Besides, CD8⁺ T cells (Cor=0.214; $p=0.001$) (Figure 10D) and B cells (Cor=0.178; $p=0.007$) (Figure 10E) were also showed association with high-risk group.

DISCUSSION

Studies on the revealing of new biomarkers for diagnosis and treatment of LIHC is in full swing [15, 16], and the relationship between immunity and LIHC has always been a research hotspot. Recently, Moeini et al. identified an immune-related gene expression pattern in liver tissues of patients with early-stage LIHC, which is associated with risk of LIHC progression in patients with cirrhosis [17]. The researchers also revealed that administration of nintedanib or aspirin and clopidogrel to mice with chronic liver inflammation caused loss of the above-mentioned gene expression pattern and development of fewer and smaller liver tumors, which further demonstrated the utility and usability of the immune signature. Although the significance of IRGs in LIHC progression and immunotherapy has been explored, there is still an urgent need for comprehensive genome-wide profiling studies to explore their clinical significance and underlying molecular mechanisms. The current study investigated the value of IRGs in LIHC in a holistic and comprehensive way, and our results further revealed their clinical significance and potential molecular characteristics. As suggested by our results, the identified IRGs were tightly involved in LIHC initiation and progression, which might also predict the prognosis for LIHC patients. As a result, these genes might give scope to their potentials as the valuable clinical biomarkers. Also, a regulatory network was

constructed in this study, which added several novel TFs and shed more light on the interactions between factors. Furthermore, an individualized immune prognostic model was proposed based on the selected differentially expressed IRGs, so as to measure immunocyte infiltration and to assess the potential clinical outcomes.

As an inflammation-associated tumor, the immune microenvironment of LIHC is well evidenced to shape tumor progression, invasion and metastasis through establishing a symbiotic relationship with cancer cells [8]. Researchers have gained an insight into the tumor-linked infiltrated inflammatory environments, nonetheless, numerous aspects of LIHC immune-associated molecular mechanisms remain unclear, and the biomarkers for predicting the treatment response are still lacking. A body of studies have uncovered the DEGs between LIHC and non-tumor samples [18–20], which contributes to the fundamental understanding towards the pathogenesis of LIHC at genetic level. For example, Carone et al. detected the expression levels of 579 immune response-related genes in 30 frozen LIHC liver tissue samples and 33 normal tissues, and demonstrated that the longer time to LIHC recurrence (TTR) was associated with up-regulation of immune response- and inflammation-related genes in tumor tissues, whereas down-regulation of these genes in normal tissues [21]. In our study, the immunocyte abundances that served as a means to characterize the immune microenvironment of LIHC, was used in combination with the immuno-genomic profiles and the corresponding clinical significance, which helped to explore the immune landscape from a novel perspective.

The acquisition of the invasive traits in cancer cells depends on a succession of genomic alterations. Therefore, the current investigation was started from the differences in the expression profiles of immune genes between LIHC and adjacent normal liver tissues, so as to uncover the relationships between these profiles and the immune microenvironment, and to illustrate the potential clinical implications. After identifying the prognosis-related immune genes, the TF-mediated network was constructed, for the sake of exposing the vital TFs that regulated the identified IRGs and revealing the underlying molecular mechanisms. Among them, chromobox protein homolog 3 (CBX3), which has the highest number of nodes related to survival-related immune genes, has been illustrated to promote tumor proliferation and predict poor survival in LIHC [22]. Besides, host cell factor C1 (HCFC1), a chromatin-associated transcriptional regulator that may also play an overall regulatory role in survival related immune genes, exerts its function in the cell division cycle during cell culture, embryogenesis as well as adult

Table 2. Immune-related gene sets that associated with high-risk group.

NAME	ES	NES	NOM p-val	FDR q-val
HEALTHY VS HIV AND SIV INFECTED DC UP	0.6998564	2.2639873	0	0
CTRL VS TIV FLU VACCINE PBMC 2008 DN	0.6825686	2.2599075	0	0
NAÏVE VS GC B CELL DN	0.6806816	2.2485397	0	0
CTRL VS POLYIC STIM DC 3H UP	0.6934012	2.2360585	0	0
NAÏVE CD4 TCELL VS DAY5 IL4 CONV TREG DN	0.713608	2.2357495	0	0

ES, enrichment score; NES, normalized enrichment score; NOM, nominal; FDR, false discovery rate.

tissue [23, 24]. In addition, HCFC1 is also proven to play a significant role in liver regeneration and show moderate co-expression with ARIDIA. ARIDIA encodes the BAF250a subunit of the SWI/SNF complex, and it plays an opposite and complex regulatory role at various stages of LIHC initiation and development [25]. Motalebipour et al. demonstrated that HCFC1 possessed the binding sites with the sterol response element-binding protein-1 (SREBP-1), which might be involved in the lipid metabolic reprogramming in LIHC [26]. Noteworthily, both the SWI/SNF chromatin remodeling complex and the aberrant lipid metabolism are demonstrated to be associated with the

immune microenvironment of LIHC, implying a potential genetic link among these three. Previous studies have also shown that, the HCFC1 expression level in mutant β -catenin LIHC is lower than that in non-mutant β -catenin LIHC [27]. Additionally, LMNB1, another TF revealed by our regulatory network, is shown to be markedly up-regulated in LIHC, which is distributed in patient plasma [28]. On the other hand, LMNB1 functions in the nuclear envelope lamina, possesses the transcriptional coregulatory activity, and has an important role in DNA replication, cell aging, and stress responses. Besides, it is positively correlated with tumor stage, tumor size,

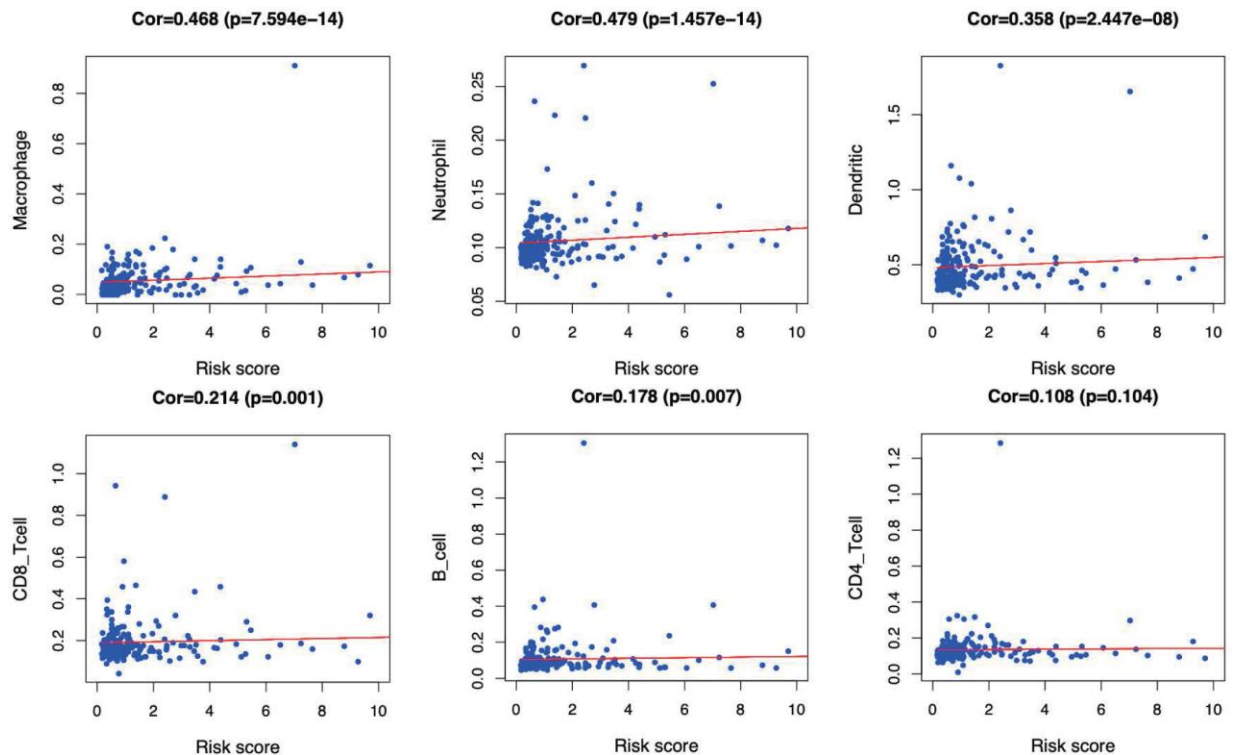


Figure 10. Relationships between the immune-related prognostic index and infiltration abundances of six types of immune cells. The correlation was performed by using Pearson correlation analysis. (A) macrophages; (B) neutrophils; (C) dendritic cells; (D) CD8⁺T cells; (E) B cells; and (F) CD4⁺T cells.

and number of nodules, therefore rendering itself a biomarker for early LIHC in tumor tissues and plasma [29]. In addition, SMARCA4 may enhance the growth and invasion of LIHC cells [30], while NRF1 is involved in stimulating the mitochondrial DNA (mtDNA) replication and transcription within LIHC [31], all of which are revealed in the network. Taken together, our results demonstrate a relatively comprehensive immune gene regulatory network associated with LIHC, yet further in-depth studies are warranted to examine the interactions between immune genes and the effects on liver cancer cells.

In this study, an immune-based prognostic signature was established to develop a simple and convenient protocol for observing the immune status and predicting the clinical outcomes for LIHC patients. This prognostic signature was based on 7 differentially expressed survival related IRGs in LIHC, which had favorable clinical viability. For the training set, LIHC patients in low-risk group had longer OS than those in high-risk group ($P=3.309e-05$), and similar results were also obtained in the testing set ($P=4.231e-03$) and the entire set ($P=8.618e-07$). Furthermore, our data showed that the prognosis signature performed moderately in prognosis prediction. In terms of the clinical utility, there was a significant correlation between the prognostic signature and tumor grade ($P=0.017$), which meant that the risk score calculated by the model we built was significantly higher in advanced grade cases. Besides, the as-constructed prognosis model also displayed the potentials to predict the differential prognosis between high- and low-risk groups, when patients were stratified by age ($\leq 65 / > 65$), clinical stage (I and II/III and IV), grade (G1 and G2/G3 and G4) and T stage (T1 and T2/T3 and T4). It implies that under diverse clinicopathological conditions, our signature is efficacious for predicting the prognosis of patients. Further PCA confirmed that our prognosis signature had sound grouping capacity. Additionally, the as-constructed signature was compared with the other two prognostic immune signatures previously published by Chew et al. to the TCGA dataset, and both of them had reached a lower significance level, with the P-values of 0.0092 and 0.0067, respectively [32, 33]. The model structured in this study exerted its own strengths in initially determining the patient prognosis and rapidly adjusting the treatment plans based on the expression of immune genes and the immunocyte infiltration levels.

Moreover, the GO item “mRNA binding” and the KEGG pathway “Spliceosome” were proved to be most significantly associated with the high-risk group. Some molecules involved in mRNA binding, such as insulin-like growth factor II mRNA-binding protein 3 (IMP3), has been demonstrated to promote tumor invasion and

predicts early recurrence and poor prognosis in LIHC [34]. With regard to the latter, previous literature elaborated that genes in spliceosome pathway, which were shown to be upregulated in tumor tissue compared with normal liver tissue, played a role in progression of LIHC [35, 36]. The two aspects mentioned above might have a bearing on the poor prognosis of high-risk group. GSEA also revealed different immune states in high-risk group compared with low-risk group. Moreover, the relationships between the immune-related prognostic index and immunocyte infiltration were deliberated to reflect the immune microenvironment status of LIHC. Of interest, our analysis indicated that the signature showed positive correlation with the infiltration of 5 kinds of immune cells, especially macrophages, neutrophils and DCs, indicating that the higher infiltration levels of these cells might be observed in the high-risk patients. Consistently, recent studies have expounded that the high densities of tumor-infiltrating macrophages and neutrophils predict the poor prognosis for primary LIHC patients. Intramural neutrophil infiltration is elaborated to be promoted by the CXCL5 and CXCR2-CXCL1 axis; besides, it is remarkably related to the shorter OS and LIHC recurrence, and is also taken as an independent prognostic factor [37, 38]. Therefore, patients with high risk scores that derived from our signature may be candidates for neutrophil targeted therapies. He et al. also illustrated that the co-inhibitory molecule programmed cell death ligand 1 (PD-L1) was observed to be overexpressed on infiltrating neutrophils from patients with LIHC, and further pointed out that the PD-L1+ neutrophils from patients with LIHC effectively suppressed the proliferation and activation of T cells, which could be partially reversed by the blockade of PD-L1 [39]. Thus, we speculated that the high-risk patients in our research may benefit from anti-PD-L1 antibodies. Furthermore, the increased infiltration of tumor-associated macrophages (TAM) (dominated by the M2 macrophages), which may be due to the deletion of the Hippo signaling, has been reported to produce the Wnt/ β -catenin signaling and trigger the elevated intramural FoxP3+ Treg population, while this in turn accelerates LIHC progression [40–42]. Moreover, Li et al. have elaborated that therapeutic blocking of the CCL2/CCR2 axis inhibits the recruitment of inflammatory monocytes, infiltration and M2-polarisation of TAMs, resulting in reversal of the immunosuppression status of the LIHC microenvironment and activation of an anti-tumorous CD8+ T cell response [43], which may be applicable to high-risk patients in our study. In line with this, Chen et al. demonstrated that upregulation of B7-H1 expression is associated with macrophage in LIHC, and anti-inflammatory therapies targeting on TAM or signaling pathways like NF- κ B and STAT3 may downregulate the B7-H1 expression on malignant cells and enhance the

efficacy of immunotherapy based on tumor-specific CD8⁺ T cells [44]. In addition, accumulating evidence reveals a role of DCs as an adverse prognostic factor for LIHC. For instance, Zhou et al. set forth that, the intra-tumoral infiltration by plasmacytoid DCs was a novel indicator of the poor prognosis for LIHC patients, which might be achieved through inducing the immune tolerogenic and inflammatory TME comprising regulatory T and IL-17-producing cells [45]. Such results have underscored the importance of tumor-associated DC cells in predicting the prognosis for LIHC patients. Our results confirmed that immunocytes were essential for LIHC progression, and suggested that the signature we constructed might potentially serve as the predictor for elevated immunocyte infiltration, which coincided with previous reports. Additionally, the immune related signature may potentially provide an instruction for treatment of LIHC. Nevertheless, the immune microenvironment of LIHC is intricate, and the role of immunocytes in LIHC has not been fully illustrated yet, which requires more efforts.

Nonetheless, some limitations should be noted in this study, which should be taken into consideration when interpreting our results. Firstly, it remained unclear about whether pretreatment in LIHC patients, like hepatic resection or transarterial chemoembolization, affected the immune contexture composition, due to the insufficient detailed clinical information. Secondly, transcriptomic analysis only reflected some aspects of the immune status, rather than the global alterations. Thirdly, the immunocyte-specific gene sets applied in this study were limited to 6 major immunocyte types, so the differences in more specialized immunocyte subtypes (like the differently polarized macrophages or myeloid-derived suppressor cells) might not be recognized in this study, while they were known to be mechanistically linked to LIHC progression and stage [46, 47]. Finally, our results were not validated via another independent cohort, which was also a limitation of this study, and the reliability of our molecular results was still challenged by the lack of experiments *in vitro* or *in vivo*.

CONCLUSIONS

To sum up, this study projects the relevance of immune microenvironment for LIHC outcomes and proposes a concise signature for the relationship between immune status and LIHC patient prognosis. The prognostic signature established in this study may be of great clinical significance, and this study can provide new insights in developing new immunotherapies for LIHC. The bioinformatic approach exposed in this study also embodies a straightforward methodology to construct other human malignancies, which will make a crucial difference to guide future clinical studies.

MATERIALS AND METHODS

Clinical samples and data acquisition

The transcriptomic RNA-sequencing data of LIHC samples were downloaded from the TCGA data portal (<https://cancergenome.nih.gov/>). Meanwhile, the clinical data of these patients were also downloaded and extracted. Furthermore, the missing follow-up information was filtered out, and 329 unique LIHC samples, which were randomized as the training group (n=165) and the testing group (n=164) using the R package “caret” [48], were incorporated into subsequent analysis. Of them, the training set was adopted to establish a prognostic immune gene signature, whereas the testing set and the entire set were employed to validate the predictive power of the as-established signature.

A list of IRGs was obtained from the Immunology Database and the Analysis Portal (ImmPort) database [49]. Typically, ImmPort is a database that accurately and timely updates the immunology data, and data shared through ImmPort lay down a powerful foundation for immunologic research. More importantly, this database provides a list of IRGs for cancer research, and these genes are identified to actively participate in the process of immune activity.

Analysis of differentially expressed genes (DEGs)

To select IRGs that were associated with LIHC, the differentially expressed IRGs between LIHC and adjacent non-tumor samples were screened using the limma package of R software (<http://bioconductor.org/packages/limma/>). Afterwards, differential gene analysis was conducted among all transcriptional data, with the false discovery rate (FDR) of < 0.05 and the log₂ |fold change| of > 1 as the cutoff values. Then, the differentially expressed IRGs were extracted from all DEGs. Subsequently, the Gene Ontology (GO) annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway functional enrichment analysis were carried out to probe into the potential molecular mechanisms of the differentially expressed IRGs.

TFs extraction and regulatory network construction

Clinical data downloaded from the TCGA data portal were collected to extract the overall survival time. Later, the survival related IRGs were selected upon univariate COX analysis carried out using the survival package of R software. To explore the interactions between these genes, a regulatory network was constructed. TFs are the important molecules that directly control the gene expression level. The Cistrome Cancer is a data source

that integrates the cancer genomic data from TCGA with over twenty-three thousand ChIP-seq and chromatin accessibility profiles, which can provide the regulatory links between TFs and transcriptomes [50]. Generally, TFs are compared with differential genes of all transcriptional data, so as to identify the differentially expressed TFs and to draw the heatmap and volcano map. Besides, the differential TFs are also connected with the survival-related immune genes and the mapping regulatory network using the Cytoscape software version 3.7.2 [51].

Development and verification of the IRG-based prognostic signature

Through univariate Cox regression analysis and least absolute shrinkage and selection operator (LASSO)-penalized Cox regression analysis [52], we develop the IRG-based prognostic signature. Notably, LASSO is the penalized regression, which employs an L1 penalty to shrink the regression coefficients toward zero, thereby eliminating numerous variables based on the principle that fewer predictors are selected in the presence of a larger penalty. Thereafter, an IRG-related prognostic signature was built to predict patient overall survival (OS). The R package “survival” and “survminer” was used to explore the optimal cut-off of risk score and draw the Kaplan–Meier survival curve. The R package “survivalROC” [53] was used to investigate the time-dependent prognostic value of the gene signature [54]. A two-sided log-rank $P < 0.05$ were considered significant for survival analysis. Afterwards, both univariate and multivariate Cox regression analyses were performed to ascertain whether the signature predicted prognosis independently from the conventional clinical factors (such as age, sex, grade, clinical stage and TNM stage). Besides, the correlation of the signature with the clinicopathological characteristics was also analyzed through the R package “ggpubr” [55]. Also, principal components analysis (PCA) was also conducted as the dimension-reducing procedure to identify a small set of synthetic variables, so as to explore the model grouping capacity. Notably, PCA is a statistical technique to determine the key variables in a multidimensional data set, which explains the observational differences and is utilized to simplify the analysis and visualization of the multidimensional data sets [56]. PCA was implemented through the limma [57] and scatterplot3d [58] packages.

Gene set enrichment analysis (GSEA)

GSEA is a computational approach to determine whether a priori defined gene set shows statistically significant and concordant differences between two biological states [59]. To reveal potential underlying Kyoto Encyclopedia

of Genes and Genomes (KEGG) pathways of the gene signature, GSEA was performed in this study using the JAVA program (<https://www.broadinstitute.org/gsea>), so as to identify enriched terms in TCGA-LIHC cohort. $P < 0.05$ and a false discovery rate $q < 0.25$ were considered statistically significant. Following 1000 permutations, the top 3 pathways in terms of the normalized enrichment score (NES) in each group were employed in multiple GSEA, so as to demonstrate a whole picture of the signaling pathways involved in the metabolic signature in LIHC.

Relationships of immune-related prognostic index with immune cell infiltration

The TIMER online database, which is a web resource to systemically evaluate the clinical impact of various immune cells on diverse cancer types, analyzes and visualizes the abundances of tumor-infiltrating immune cells [60]. It covers 10,009 samples across 23 cancer types from TCGA to estimate the abundance of six tumor-infiltrating immune cell subtypes, including B cells, CD4 T cells, CD8 T cells, macrophages, neutrophils, and dendritic cells (DCs). Therefore, it can be easily used to determine the relationship of immune cells infiltration with other parameters. In this study, the immune infiltrate levels of TCGA-LIHC patients were downloaded, and the associations of the prognostic signature with immune cells infiltration were calculated.

Statistical analysis

The R (v.3.6.1) software was employed for all statistical analyses. Pearson χ^2 test or Fisher’s exact test were used to explore qualitative variables as appropriate. If not specified above, $P < 0.05$ was considered statistically significant.

AUTHOR CONTRIBUTIONS

Xin-Ting Sang and Bo Hu created the idea for the review. Bo Hu performed the selection of literature, drafted the manuscript, and prepared the figures. Xiao-Bo Yang and Xin-Ting Sang revised the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors would like to thank the ImmPort, cBioPortal, TIMER, Cistrome cancer and TCGA databases for the availability of the data.

CONFLICTS OF INTEREST

The authors have no conflicts of interest.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018; 68:394–424. <https://doi.org/10.3322/caac.21492> PMID:30207593
2. Llovet JM, Montal R, Villanueva A. Randomized trials and endpoints in advanced HCC: role of PFS as a surrogate of survival. *J Hepatol.* 2019; 70:1262–77. <https://doi.org/10.1016/j.jhep.2019.01.028> PMID:30943423
3. Abou-Alfa GK, Meyer T, Cheng AL, El-Khoueiry AB, Rimassa L, Ryoo BY, Cicin I, Merle P, Chen Y, Park JW, Blanc JF, Bolondi L, Klumpen HJ, et al. Cabozantinib in patients with advanced and progressing hepatocellular carcinoma. *N Engl J Med.* 2018; 379:54–63. <https://doi.org/10.1056/NEJMoa1717002> PMID:29972759
4. Finn RS, Zhu AX, Farah W, Almasri J, Zaiem F, Prokop LJ, Murad MH, Mohammed K. Therapies for advanced stage hepatocellular carcinoma with macrovascular invasion or metastatic disease: A systematic review and meta-analysis. *Hepatology.* 2018; 67:422–35. <https://doi.org/10.1002/hep.29486> PMID:28881497
5. Khemlina G, Ikeda S, Kurzrock R. The biology of Hepatocellular carcinoma: implications for genomic and immune therapies. *Mol Cancer.* 2017; 16:149. <https://doi.org/10.1186/s12943-017-0712-x> PMID:28854942
6. Brenner MJ, Cho JH, Wong NM, Wong WW. Synthetic biology: immunotherapy by design. *Annu Rev Biomed Eng.* 2018; 20:95–118. <https://doi.org/10.1146/annurev-bioeng-062117-121147> PMID:29345976
7. van Dijk N, Funt SA, Blank CU, Powles T, Rosenberg JE, van der Heijden MS. The cancer immunogram as a framework for personalized immunotherapy in urothelial cancer. *Eur Urol.* 2019; 75:435–44. <https://doi.org/10.1016/j.eururo.2018.09.022> PMID:30274701
8. Fu Y, Liu S, Zeng S, Shen H. From bench to bed: the tumor immune microenvironment and current immunotherapeutic strategies for hepatocellular carcinoma. *J Exp Clin Cancer Res.* 2019; 38:396. <https://doi.org/10.1186/s13046-019-1396-4> PMID:31500650
9. Lee JH, Lee JH, Lim YS, Yeon JE, Song TJ, Yu SJ, Gwak GY, Kim KM, Kim YJ, Lee JW, Yoon JH. Sustained efficacy of adjuvant immunotherapy with cytokine-induced killer cells for hepatocellular carcinoma: an extended 5-year follow-up. *Cancer Immunol Immunother.* 2019; 68:23–32. <https://doi.org/10.1007/s00262-018-2247-4> PMID:30232520
10. Waidmann O. Recent developments with immunotherapy for hepatocellular carcinoma. *Expert Opin Biol Ther.* 2018; 18:905–10. <https://doi.org/10.1080/14712598.2018.1499722> PMID:29995439
11. Blum A, Wang P, Zenklusen JC. SnapShot: TCGA-Analyzed Tumors. *Cell.* 2018; 173:530. <https://doi.org/10.1016/j.cell.2018.03.059> PMID:29625059
12. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, Kovatich AJ, Benz CC, Levine DA, Lee AV, Omberg L, Wolf DM, Shriver CD, Thorsson V; Cancer Genome Atlas Research Network, Hu H. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell.* 2018; 173:400–416.e11. <https://doi.org/10.1016/j.cell.2018.02.052> PMID:29625055
13. Yang H, Zhang X, Cai XY, Wen DY, Ye ZH, Liang L, Zhang L, Wang HL, Chen G, Feng ZB. From big data to diagnosis and prognosis: gene expression signatures in liver hepatocellular carcinoma. *PeerJ.* 2017; 5:e3089. <https://doi.org/10.7717/peerj.3089> PMID:28316892
14. Yang S, Liu T, Nan H, Wang Y, Chen H, Zhang X, Zhang Y, Shen B, Qian P, Xu S, Sui J, Liang G. Comprehensive analysis of prognostic immune-related genes in the tumor microenvironment of cutaneous melanoma. *J Cell Physiol.* 2020; 235:1025–35. <https://doi.org/10.1002/jcp.29018> PMID:31240705
15. Yang M, Zhang X, Liu J. Prognostic value of des-γ-carboxy prothrombin in patients with hepatocellular carcinoma treated with transarterial chemotherapy: A systematic review and meta-analysis. *PLoS One.* 2019; 14:e0225170. <https://doi.org/10.1371/journal.pone.0225170> PMID:31730646
16. Maeda T, Kanzaki H, Chiba T, Ao J, Kanayama K, Maruta S, Kusakabe Y, Saito T, Kobayashi K, Kiyono S, Nakamura M, Ogasawara S, Suzuki E, et al. Serum fibroblast growth factor 19 serves as a potential novel biomarker for hepatocellular carcinoma. *BMC Cancer.* 2019; 19:1088. <https://doi.org/10.1186/s12885-019-6322-9> PMID:31718608
17. Moeini A, Torrecilla S, Tovar V, Montironi C, Andreu-Oller C, Peix J, Higuera M, Pfister D, Ramadori P, Pinyol R, Solé M, Heikenwälder M, Friedman SL, et al. An Immune Gene Expression Signature Associated

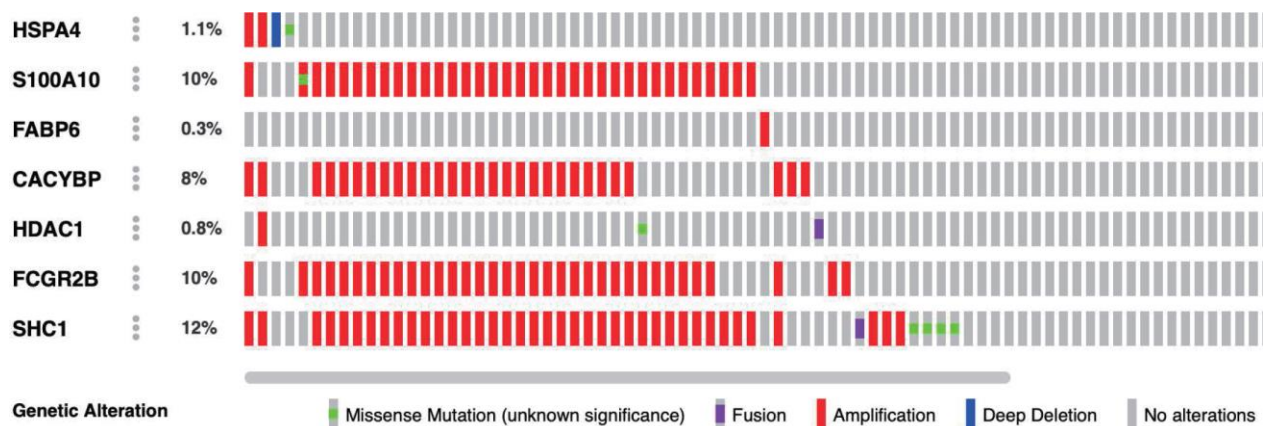
- With Development of Human Hepatocellular Carcinoma Identifies Mice That Respond to Chemopreventive Agents. *Gastroenterology*. 2019; 157:1383–1397.e11.
<https://doi.org/10.1053/j.gastro.2019.07.028>
PMID:[31344396](https://pubmed.ncbi.nlm.nih.gov/31344396/)
18. Selli C, Pearce DA, Sims AH, Tosun M. Differential expression of store-operated calcium- and proliferation-related genes in hepatocellular carcinoma cells following TRPC1 ion channel silencing. *Mol Cell Biochem*. 2016; 420:129–40.
<https://doi.org/10.1007/s11010-016-2776-0>
PMID:[27443843](https://pubmed.ncbi.nlm.nih.gov/27443843/)
 19. Dong LY, Zhou WZ, Ni JW, Xiang W, Hu WH, Yu C, Li HY. Identifying the optimal gene and gene set in hepatocellular carcinoma based on differential expression and differential co-expression algorithm. *Oncol Rep*. 2017; 37:1066–74.
<https://doi.org/10.3892/or.2016.5333>
PMID:[28035405](https://pubmed.ncbi.nlm.nih.gov/28035405/)
 20. Likhitrattanapaisal S, Tipanee J, Janvilisri T. Meta-analysis of gene expression profiles identifies differential biomarkers for hepatocellular carcinoma and cholangiocarcinoma. *Tumour Biol*. 2016; 37:12755–66.
<https://doi.org/10.1007/s13277-016-5186-8>
PMID:[27448818](https://pubmed.ncbi.nlm.nih.gov/27448818/)
 21. Carone C, Olivani A, Dalla Valle R, Manuguerra R, Silini EM, Trenti T, Missale G, Cariani E. Immune gene expression profile in hepatocellular carcinoma and surrounding tissue predicts time to tumor recurrence. *Liver Cancer*. 2018; 7:277–94.
<https://doi.org/10.1159/000486764>
PMID:[30319985](https://pubmed.ncbi.nlm.nih.gov/30319985/)
 22. Zhong X, Kan A, Zhang W, Zhou J, Zhang H, Chen J, Tang S. CBX3/HP1y promotes tumor proliferation and predicts poor survival in hepatocellular carcinoma. *Aging (Albany NY)*. 2019; 11:5483–97.
<https://doi.org/10.18632/aging.102132>
PMID:[31375643](https://pubmed.ncbi.nlm.nih.gov/31375643/)
 23. Quintana AM, Geiger EA, Achilly N, Rosenblatt DS, Maclean KN, Stabler SP, Artinger KB, Appel B, Shaikh TH. Hcfc1b, a zebrafish ortholog of HCFC1, regulates craniofacial development by modulating mmachc expression. *Dev Biol*. 2014; 396:94–106.
<https://doi.org/10.1016/j.ydbio.2014.09.026>
PMID:[25281006](https://pubmed.ncbi.nlm.nih.gov/25281006/)
 24. Minocha S, Villeneuve D, Praz V, Moret C, Lopes M, Pinatel D, Rib L, Guex N, Herr W. Rapid recapitulation of nonalcoholic steatohepatitis upon loss of host cell factor 1 function in mouse hepatocytes. *Mol Cell Biol*. 2019; 39:e00405–18.
<https://doi.org/10.1128/MCB.00127-19>
PMID:[30559308](https://pubmed.ncbi.nlm.nih.gov/30559308/)
 25. Sun X, Wang SC, Wei Y, Luo X, Jia Y, Li L, Gopal P, Zhu M, Nassour I, Chuang JC, Maples T, Celen C, Nguyen LH, et al. Arid1a has context-dependent oncogenic and tumor suppressor functions in liver Cancer. *Cancer Cell*. 2018; 33:151–52.
<https://doi.org/10.1016/j.ccell.2017.12.011>
PMID:[29316428](https://pubmed.ncbi.nlm.nih.gov/29316428/)
 26. Motallebipour M, Enroth S, Punga T, Ameer A, Koch C, Dunham I, Komorowski J, Ericsson J, Wadelius C. Novel genes in cell cycle control and lipid metabolism with dynamically regulated binding sites for sterol regulatory element-binding protein 1 and RNA polymerase II in HepG2 cells detected by chromatin immunoprecipitation with microarray detection. *FEBS J*. 2009; 276:1878–90.
<https://doi.org/10.1111/j.1742-4658.2009.06914.x>
PMID:[19292868](https://pubmed.ncbi.nlm.nih.gov/19292868/)
 27. Cavard C, Colnot S, Audard V, Benhamouche S, Finzi L, Torre C, Grimber G, Godard C, Terris B, Perret C. Wnt/beta-catenin pathway in hepatocellular carcinoma pathogenesis and liver physiology. *Future Oncol*. 2008; 4:647–60.
<https://doi.org/10.2217/14796694.4.5.647>
PMID:[18922122](https://pubmed.ncbi.nlm.nih.gov/18922122/)
 28. Sun S, Xu MZ, Poon RT, Day PJ, Luk JM. Circulating Lamin B1 (LMNB1) biomarker detects early stages of liver cancer in patients. *J Proteome Res*. 2010; 9:70–78.
<https://doi.org/10.1021/pr9002118>
PMID:[19522540](https://pubmed.ncbi.nlm.nih.gov/19522540/)
 29. Bandos AI, Obuchowski NA, and AI. Evaluation of diagnostic accuracy in free-response detection-localization tasks using ROC tools. *Stat Methods Med Res*. 2019; 28:1808–25.
<https://doi.org/10.1177/0962280218776683>
PMID:[29921163](https://pubmed.ncbi.nlm.nih.gov/29921163/)
 30. Chen Z, Lu X, Jia D, Jing Y, Chen D, Wang Q, Zhao F, Li J, Yao M, Cong W, He X. Hepatic SMARCA4 predicts HCC recurrence and promotes tumour cell proliferation by regulating SMAD6 expression. *Cell Death Dis*. 2018; 9:59.
<https://doi.org/10.1038/s41419-017-0090-8>
PMID:[29352111](https://pubmed.ncbi.nlm.nih.gov/29352111/)
 31. Yin PH, Lee HC, Chau GY, Wu YT, Li SH, Lui WY, Wei YH, Liu TY, Chi CW. Alteration of the copy number and deletion of mitochondrial DNA in human hepatocellular carcinoma. *Br J Cancer*. 2004; 90:2390–96.
<https://doi.org/10.1038/sj.bjc.6601838>
PMID:[15150555](https://pubmed.ncbi.nlm.nih.gov/15150555/)
 32. Chew V, Chen J, Lee D, Loh E, Lee J, Lim KH, Weber A, Slankamenac K, Poon RT, Yang H, Ooi LL, Toh HC,

- Heikenwalder M, et al. Chemokine-driven lymphocyte infiltration: an early intratumoural event determining long-term survival in resectable hepatocellular carcinoma. *Gut*. 2012; 61:427–38.
<https://doi.org/10.1136/gutjnl-2011-300509>
PMID:21930732
33. Chew V, Tow C, Teo M, Wong HL, Chan J, Gehring A, Loh M, Bolze A, Quek R, Lee VK, Lee KH, Abastado JP, Toh HC, Nardin A. Inflammatory tumour microenvironment is associated with superior survival in hepatocellular carcinoma patients. *J Hepatol*. 2010; 52:370–79.
<https://doi.org/10.1016/j.jhep.2009.07.013>
PMID:19720422
34. Jeng YM, Chang CC, Hu FC, Chou HY, Kao HL, Wang TH, Hsu HC. RNA-binding protein insulin-like growth factor II mRNA-binding protein 3 expression promotes tumor invasion and predicts early recurrence and poor prognosis in hepatocellular carcinoma. *Hepatology*. 2008; 48:1118–27.
<https://doi.org/10.1002/hep.22459>
PMID:18802962
35. Xu W, Huang H, Yu L, Cao L. Meta-analysis of gene expression profiles indicates genes in spliceosome pathway are up-regulated in hepatocellular carcinoma (HCC). *Med Oncol*. 2015; 32:96.
<https://doi.org/10.1007/s12032-014-0425-6>
PMID:25731616
36. Iguchi T, Komatsu H, Masuda T, Nambara S, Kidogami S, Ogawa Y, Hu Q, Saito T, Hirata H, Sakimura S, Uchi R, Hayashi N, Ito S, et al. Increased copy number of the gene encoding SF3B4 indicates poor prognosis in hepatocellular carcinoma. *Anticancer Res*. 2016; 36:2139–44.
PMID:27127115
37. Li L, Xu L, Yan J, Zhen ZJ, Ji Y, Liu CQ, Lau WY, Zheng L, Xu J. CXCR2-CXCL1 axis is correlated with neutrophil infiltration and predicts a poor prognosis in hepatocellular carcinoma. *J Exp Clin Cancer Res*. 2015; 34:129.
<https://doi.org/10.1186/s13046-015-0247-1>
PMID:26503598
38. Zhou SL, Dai Z, Zhou ZJ, Wang XY, Yang GH, Wang Z, Huang XW, Fan J, Zhou J. Overexpression of CXCL5 mediates neutrophil infiltration and indicates poor prognosis for hepatocellular carcinoma. *Hepatology*. 2012; 56:2242–54.
<https://doi.org/10.1002/hep.25907>
PMID:22711685
39. He G, Zhang H, Zhou J, Wang B, Chen Y, Kong Y, Xie X, Wang X, Fei R, Wei L, Chen H, Zeng H. Peritumoural neutrophils negatively regulate adaptive immunity via the PD-L1/PD-1 signalling pathway in hepatocellular carcinoma. *J Exp Clin Cancer Res*. 2015; 34:141.
<https://doi.org/10.1186/s13046-015-0256-0>
PMID:26581194
40. Ding T, Xu J, Wang F, Shi M, Zhang Y, Li SP, Zheng L. High tumor-infiltrating macrophage density predicts poor prognosis in patients with primary hepatocellular carcinoma after resection. *Hum Pathol*. 2009; 40:381–89.
<https://doi.org/10.1016/j.humpath.2008.08.011>
PMID:18992916
41. Shirabe K, Mano Y, Muto J, Matono R, Motomura T, Toshima T, Takeishi K, Uchiyama H, Yoshizumi T, Taketomi A, Morita M, Tsujitani S, Sakaguchi Y, Maehara Y. Role of tumor-associated macrophages in the progression of hepatocellular carcinoma. *Surg Today*. 2012; 42:1–7.
<https://doi.org/10.1007/s00595-011-0058-8>
PMID:22116397
42. Tian Z, Hou X, Liu W, Han Z, Wei L. Macrophages and hepatocellular carcinoma. *Cell Biosci*. 2019; 9:79.
<https://doi.org/10.1186/s13578-019-0342-7>
PMID:31572568
43. Li X, Yao W, Yuan Y, Chen P, Li B, Li J, Chu R, Song H, Xie D, Jiang X, Wang H. Targeting of tumour-infiltrating macrophages via CCL2/CCR2 signalling as a therapeutic strategy against hepatocellular carcinoma. *Gut*. 2017; 66:157–67.
<https://doi.org/10.1136/gutjnl-2015-310514>
PMID:26452628
44. Chen J, Li G, Meng H, Fan Y, Song Y, Wang S, Zhu F, Guo C, Zhang L, Shi Y. Upregulation of B7-H1 expression is associated with macrophage infiltration in hepatocellular carcinomas. *Cancer Immunol Immunother*. 2012; 61:101–08.
<https://doi.org/10.1007/s00262-011-1094-3>
PMID:21853301
45. Zhou ZJ, Xin HY, Li J, Hu ZQ, Luo CB, Zhou SL. Intratumoral plasmacytoid dendritic cells as a poor prognostic factor for hepatocellular carcinoma following curative resection. *Cancer Immunol Immunother*. 2019; 68:1223–33.
<https://doi.org/10.1007/s00262-019-02355-3>
PMID:31201473
46. Arihara F, Mizukoshi E, Kitahara M, Takata Y, Arai K, Yamashita T, Nakamoto Y, Kaneko S. Increase in CD14+HLA-DR⁻/low myeloid-derived suppressor cells in hepatocellular carcinoma patients and its impact on prognosis. *Cancer Immunol Immunother*. 2013; 62:1421–30.
<https://doi.org/10.1007/s00262-013-1447-1>
PMID:23764929

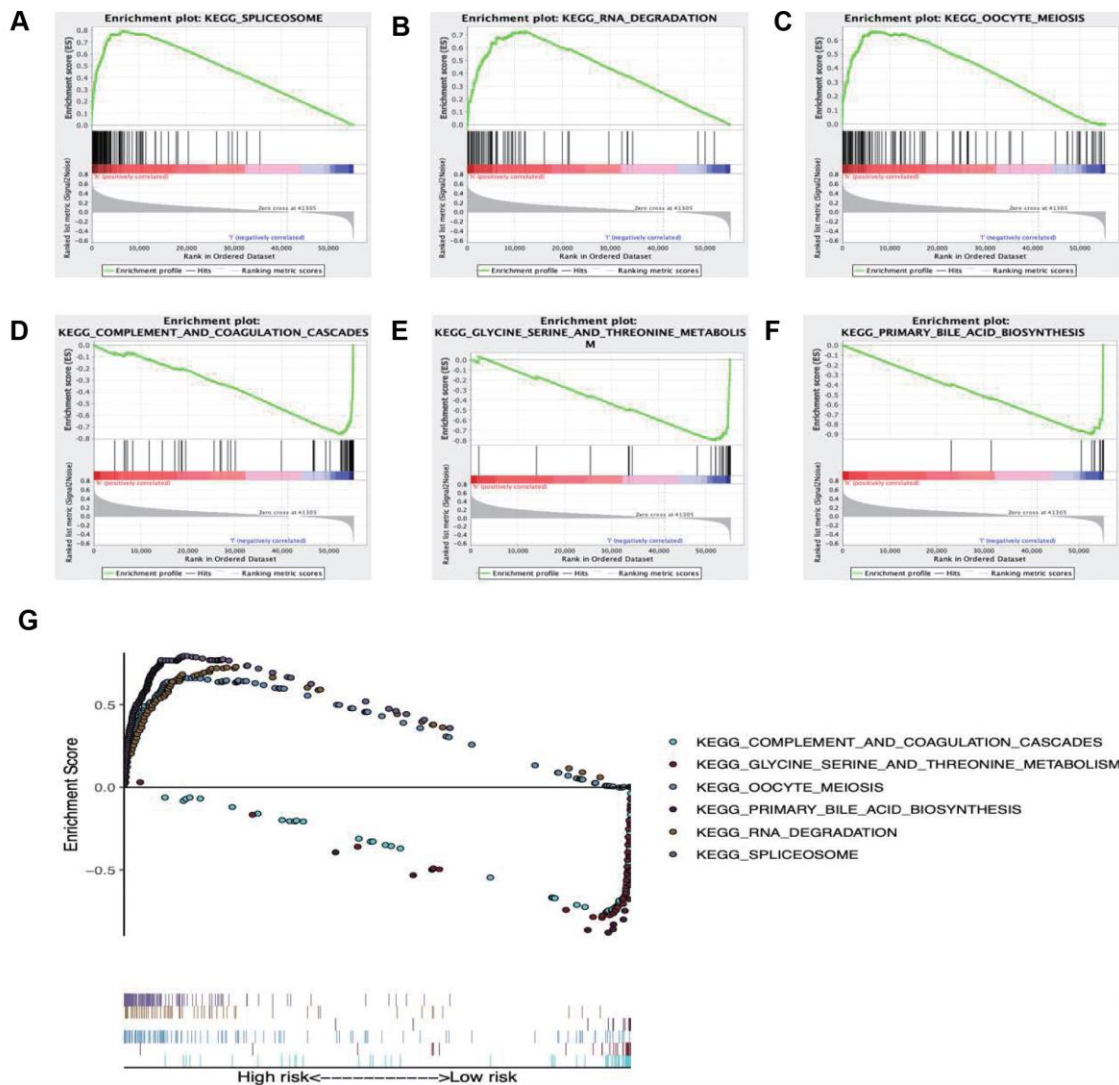
47. Hoehst B, Ormandy LA, Ballmaier M, Lehner F, Krüger C, Manns MP, Greten TF, Korangy F. A new population of myeloid-derived suppressor cells in hepatocellular carcinoma patients induces CD4(+)CD25(+)Foxp3(+) T cells. *Gastroenterology*. 2008; 135:234–43. <https://doi.org/10.1053/j.gastro.2008.03.020> PMID:18485901
48. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw*. 2008; 28:1–26. <https://doi.org/10.18637/jss.v028.i05>
49. Bhattacharya S, Andorf S, Gomes L, Dunn P, Schaefer H, Pontius J, Berger P, Desborough V, Smith T, Campbell J, Thomson E, Monteiro R, Guimaraes P, et al. ImmPort: disseminating data to the public for the future of immunology. *Immunol Res*. 2014; 58:234–39. <https://doi.org/10.1007/s12026-014-8516-1> PMID:24791905
50. Mei S, Meyer CA, Zheng R, Qin Q, Wu Q, Jiang P, Li B, Shi X, Wang B, Fan J, Shih C, Brown M, Zang C, Liu XS. Cistrome cancer: a web resource for integrative gene regulation modeling in cancer. *Cancer Res*. 2017; 77:e19–22. <https://doi.org/10.1158/0008-5472.CAN-17-0327> PMID:29092931
51. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, Wadi L, Meyer M, Wong J, Xu C, Merico D, Bader GD. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc*. 2019; 14:482–517. <https://doi.org/10.1038/s41596-018-0103-9> PMID:30664679
52. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med*. 1997; 16:385–95. [https://doi.org/10.1002/\(SICI\)1097-0258\(19970228\)16:4<385::AID-SIM380>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3) PMID:9044528
53. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*. 2000; 56:337–44. <https://doi.org/10.1111/j.0006-341X.2000.00337.x> PMID:10877287
54. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med*. 2013; 32:5381–97. <https://doi.org/10.1002/sim.5958> PMID:24027076
55. Jo SS, Choi SS. Enrichment of rare alleles within epigenetic chromatin marks in the first intron. *Genomics Inform*. 2019; 17:e9. <https://doi.org/10.5808/GI.2019.17.1.e9> PMID:30929410
56. Raychaudhuri S, Stuart JM, Altman RB. Principal components analysis to summarize microarray experiments: application to sporulation time series. *Pac Symp Biocomput*. 2000:455–66. https://doi.org/10.1142/9789814447331_0043 PMID:10902193
57. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015; 43:e47. <https://doi.org/10.1093/nar/gkv007> PMID:25605792
58. Koutecký P. MorphoTools: a set of R functions for morphometric analysis. *Plant Syst Evol*. 2015; 301:1115–21. <https://doi.org/10.1007/s00606-014-1153-2>
59. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005; 102:15545–50. <https://doi.org/10.1073/pnas.0506580102> PMID:16199517
60. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, Liu XS. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res*. 2017; 77:e108–10. <https://doi.org/10.1158/0008-5472.CAN-17-0307> PMID:29092952

SUPPLEMENTARY MATERIALS

Supplementary Figures



Supplementary Figure 1. Genetic alternations of signature genes in TCGA-LIHC. *SHC1* is the gene with highest mutation frequency. There are 4 genes with a mutation rate $\geq 5\%$.



Supplementary Figure 2. Enrichment plots of Kyoto Encyclopedia of Genes and Genomes analysis from gene set enrichment analysis (GSEA). GSEA results showing (A) Spliceosome, (B) RNA degradation, (C) Oocyte meiosis were differentially enriched in high risk group, while (D) Complement and coagulation cascades, (E) Glycine serine and threonine metabolism, (F) Primary bile acid biosynthesis were primarily enriched in low risk group. (G) Summarizes the above six gene sets.

Supplementary Table

Supplementary Table 1. General characteristics of LIHC-specific survival-related immune genes.

id	HR	HR.95L	HR.95H	P-value
CACYBP	1.0917619	1.05795119	1.12665315	4.51E-08
HDAC1	1.05022551	1.03097483	1.06983565	2.08E-07
PSMD14	1.10778447	1.06262069	1.15486781	1.44E-06
HSPA4	1.05964018	1.03404643	1.08586739	3.42E-06
NRAS	1.08973904	1.05091756	1.12999461	3.43E-06
ISG20L2	1.16874	1.09328287	1.24940509	4.67E-06
TNFRSF11A	2.07200697	1.51253076	2.83843013	5.71E-06
IFI30	3.17933487	1.90457983	5.30729671	9.68E-06
S100A10	1.00566657	1.00308672	1.00825306	1.62E-05
PPIA	1.01599581	1.00862786	1.02341759	1.93E-05
FABP6	1.12300205	1.06476789	1.18442114	1.96E-05
FABP5	1.07739142	1.04069686	1.11537981	2.48E-05
SHC1	1.01663274	1.00869591	1.02463202	3.70E-05
IL2RG	1.02199182	1.01147917	1.03261373	3.73E-05
IKBKE	1.27102829	1.12666386	1.43389077	9.67E-05
HGF	1.1241683	1.05920587	1.19311495	0.00011624
PSMD10	1.0656461	1.0311687	1.10127627	0.00015121
BIRC5	1.0404312	1.01860922	1.06272068	0.0002475
CCL8	1.19891453	1.08519549	1.32455034	0.00035977
GBP2	1.03201284	1.01428214	1.0500535	0.00036553
PDGFRL	1.21816123	1.0896081	1.36188119	0.00052404
IL17D	1.09463878	1.03846349	1.15385284	0.00076792
SPP1	1.00060692	1.00024964	1.00096432	0.00086893
BRD8	1.20845349	1.08066447	1.35135362	0.00089898
TAP2	1.08531931	1.03409177	1.13908459	0.00090377
TAP1	1.01069017	1.00430185	1.01711913	0.00101324
S100A11	1.00167522	1.00067095	1.00268051	0.00107363
PSME3	1.05152035	1.02029916	1.08369692	0.00108801
DCK	1.18182999	1.06736158	1.30857448	0.0013083
ZYX	1.01343573	1.0050076	1.02193455	0.00173453
CCL20	1.00559114	1.00206949	1.00912517	0.00183976
NDRG1	1.00647063	1.00237442	1.01058359	0.00193684
SEMA5B	1.26004029	1.08844097	1.45869329	0.00197114
EED	1.52350585	1.16595225	1.99070766	0.00203529
PSMD2	1.02110969	1.00726704	1.03514258	0.00270242
RFX5	1.07783148	1.0263163	1.13193243	0.00270413
PLXNA2	1.33484634	1.10350362	1.61468864	0.00293744
NFKBIE	1.04940072	1.01632056	1.08355761	0.0031719

CDK4	1.03302289	1.01084017	1.0556924	0.00335215
LCNL1	2.06794791	1.26218984	3.38808668	0.00392231
IL27RA	1.15188728	1.04616157	1.2682977	0.00399336
FGF13	2.65086219	1.36077852	5.16400742	0.00416462
PML	1.18079808	1.05286741	1.32427322	0.00450452
PLXNA1	1.18084436	1.0519579	1.32552205	0.00481816
FCGR2B	1.11560295	1.03356267	1.20415528	0.00500003
IRF5	1.18229283	1.05148895	1.32936854	0.00512223
PPARG	1.10740754	1.03047593	1.19008259	0.00548345
FPR1	1.12723089	1.03554744	1.22703165	0.00565789
CCL14	0.35186891	0.16776124	0.73802346	0.00571367
S100A6	1.00164935	1.00047594	1.00282414	0.00585868
KITLG	1.27115114	1.07133671	1.50823286	0.00596518
HLA-DOB	1.06529841	1.01805969	1.11472904	0.00626831
MAPK3	1.08276978	1.02220825	1.14691934	0.00677033
IDO1	1.07592028	1.0200119	1.13489308	0.00719388
CSPG5	1.67901922	1.14889605	2.45375164	0.00742866
NR6A1	1.37807871	1.08943246	1.74320208	0.00748934
WNT5A	1.19136252	1.04748929	1.35499683	0.00766411
TRAF3	1.28965446	1.06884011	1.55608741	0.0079365
PLAU	1.04381479	1.01096212	1.07773506	0.00858497
MICB	1.15494195	1.03698719	1.28631376	0.00877407
LPA	0.84675749	0.74669709	0.96022638	0.00952743
