Prognostic Value and Regulatory Network of Immune-Related Genes in Hepatocellular Carcinoma

Xiang Zhou
Yan'an University
Keying Zhang
Department of Urology, Xijing Hospital, Fourth Military Medical University, Xi'an
Fa Yang
Department of Urology, Xijing Hospital, Fourth Military Medical University, Xi'an
Chao Xu
Department of Urology, Xijing Hospital, Fourth Military Medical University, Xi'an
Jianhua Jiao
Department of Urology, Xijing Hospital, Fourth Military Medical University, Xi'an
Xiaolong Zhao
Department of Urology, Xijing Hospital, Fourth Military Medical University, Xi'an
Yan Wang
The 987 Hospital of PLA, Baoji
Shaojie Liu
Department of Urology, Xijing Hospital, Fourth Military Medical University, Xi'an
Yao Jiang
Department of Urology, Xijing Hospital, Fourth Military Medical University, Xi'an
Weihong Wen
Northwestern Polytechnical University
Zhaoying Fu (yadxfzy@163.com)
Yan'an University

Research Article

Keywords: Hepatocellular carcinoma, Immune-related genes, TCGA, Prognostic signatures, TFs regulatory mechanism

Posted Date: December 11th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-123214/v1

Abstract

Background: Hepatocellular carcinoma (HCC) is a disease with higher morbidity, mortality, and poor prognosis in the whole world. Understanding the crosslink between HCC and the immune system is essential for people to uncover a few potential and valuable therapeutic strategies. This study aimed to reveal the correlation between HCC and immune-related genes and establish a clinical evaluation model.

Methods: We had analyzed the clinical information consisted of 373 HCC and 49 normal samples from the cancer genome atlas (TCGA). The differentially expressed genes (DEGs) were selected by the Wilcoxon test and the immune-related differentially expressed genes (IRDEGs) in DEGs were identified by matching DEGs with immune-related genes downloaded from the ImmPort database. Furthermore, the univariate Cox regression analysis and multivariate Cox regression analysis were performed to construct a prognostic risk model. Then, twenty-two types of tumor immune-infiltrating cells (TIICs) were downloaded from Tumor Immune Estimation Resource (TIMER) and were used to construct the correlational graphs between the TIICs and risk score by the CIBERSORT. Subsequently, the transcription factors (TFs) were gained in the Cistrome website and the differentially expressed TFs (DETFs) were achieved. Finally, the KEGG pathway analysis and GO analysis were performed to further understand the molecular mechanisms between DETFs and PDIRGs.

Results: In our study, 5839 DEGs, 326 IRDEGs, and 31 prognosis-related IRDEGs (PIRDEGs) were identified. And 8 optimal PIRDEGs were employed to construct a prognostic risk model by multivariate Cox regression analysis. The correlation between risk genes and clinical characterizations and TIICs has verified that the prognostic model was effective in predicting the prognosis of HCC patients. Finally, several important immune-related pathways and molecular functions of the eight PIRDEGs were significantly enriched and there was a distinct association between the risk IRDEGs and TFs.

Conclusion: The prognostic risk model showed a more valuable predicting role for HCC patients, and produced many novel therapeutic targets and strategies for HCC.

Introduction

Hepatocellular carcinoma (HCC) is the most malignant disease which manifested high recurrence, poor prognosis, and resistance to chemotherapy [(1)]. The important curative options such as surgical resection or liver transplantation just limited to early-stage HCC patients. Drug targeted tumor immune microenvironment, chimeric antigen receptor redirected T (CAR-T) cells targeted specific tumor-associated antigens (TAAs), and the inhibitors targeted immune checkpoint (ICI) have been gradually developed in recent years [(2, 3)]. Systemic sorafenib therapy had been proved effective in liver cirrhosis patients who were not suitable for targeting therapy and patients with chronic metastasis HCC, but its effectiveness was still weak [(4)]. The tumor microenvironment of HCC that consists of lots of immune cells and cytokines depends on its genetic and other factors to drive inflammation and immune dysfunction, which

control the malignant progression and poor response to therapy of HCC [5]. Hence, we must pay more attention to the relationship between the immune system and the progression and treatment of HCC.

The Cancer Genome Atlas (TCGA), a landmark cancer genomics program, molecularly characterized over 20,000 primary cancer and matched normal samples spanning 33 cancer types. Recently, advances in high-throughput nucleotide sequencing analysis and single-cell sequencing have enabled us to better understand the difference in gene expression between tumorous and normal tissues. Many studies have implied that the genetic mutation signatures occurred in lots of types of cancer and the survival time of patients were labile, such as breast cancer [6], non-small cell lung cancer [7], renal cell carcinoma[8], and hepatocellular carcinoma [9, 10]. The immune microenvironment of HCC is complicated, with a variety of innate and adaptive immune cells, which would affect the immune evasion, immunotherapy response, and patient survival [3]. The immune checkpoints, including programmed death receptor 1 and its ligand (PD-1/PD-L1), cytotoxic T-lymphocyte-associated protein 4 (CTLA-4), and T cell immunoglobulin mucin-3 (TIM-3), had become the important therapeutic targets of HCC [11–13]. By analyzing the relationship between the liver microenvironment and immune cells or genes, novel and effective therapeutic targets and strategies for HCC will be proposed.

In the present study, we have identified eight prognostic differentially expressed genes incorporated with differential expressed genes and immune-related genes, including *TRAF3*, *NORG1*, *MAPT*, *IL-17D*, *RBP2*, *PSMD14*, *HSPA4*, and *NRAS*. These immune-related genes manifest a pattern of the poor prognosis of HCC patients and revealed the relationship between oncogenes and immune-related genes.

Materials And Methods

Data preparation from TCGA

The 373 sets of HCC data and 49 sets of normal liver tissue data, including RNA transcriptomic sequencing data (RNA-seq FPKM) and clinical characteristics, were downloaded from the TCGA portal (https://portal.gdc.cancer.gov/). Then, those data were integrated according to their ID numbers, and within-array replicate probes were replaced with their average via limma package [14]. Finally, all data were processed and analyzed in R software (https://www.r-project.org/).

Differential expression analysis and IRDEGs identification

Wilcoxon test was adopted to analyze the differentially expressed genes (DEGs) between tumors and normal tissues. The P-value was adjusted with the false discovery rate (FDR), and the filter criteria was FDR < 0.05 and |log2 fold-change [FC]| > 1. After accessing the latest list of immune-related genes (IRGs) in the Immunology Database and Analysis Portal (IMMPORT) website (https://www.immport.org), the immune-related differentially expressed genes (IRDEGs) of HCC were identified by matching the DEGs. Then, univariate Cox regression analysis with a threshold value of P < 0.01 was used to identify possible prognostic IRDEGs (PIRDEGs).

Prognostic signature construction and validation

Initially, to avoid overfitting, PIRDEGs that correlated highly with other genes were deleted. Next, a prognostic risk model constructed by multivariate Cox regression was achieved, and an individualized risk score with coefficients was calculated with the following computational formula [15]:

$$\text{Risk score (individual)} = \sum_{i=1}^{n} \text{coefficient (gene i)} * \text{expression value of (gene i)}$$

Here, "$gene_i$" is the i^{th} selected gene, and "coefficient ($gene_i$)" is the weight of $gene_i$ expression extracted from the risk model. The median risk score was used as the cut-off value to separate the high-risk HCC patients from the low-risk group. Subsequently, survival analysis and receiver operating characteristic (ROC) analysis were performed using the R software, and the area under the curve (AUC) was calculated as reported [16].

Independent prognostic factors analysis

Univariate Cox regression analysis was employed to identify factors associated with the survival of HCC patients and multivariate Cox regression analysis was used to assess which clinical parameters and risk score were independent of the prognosis of HCC patients in the present prognostic model.

Clinical utility of prognostic signature

The TCGA-HCC samples were divided into different groups according to the clinical variables (i.e. age, gender, histological grade, pathological stage, and tumor-node-metastasis (TNM) status). Whereafter, the risk score and the expression level of related risk genes of the prognostic model were compared between groups with a t-test and the box plots were produced with beeswarm package.

Regulatory network between DETFs and PIRDEGs

The reference data of transcription factors (TFs) was acquired on the Cistrome website (https://cistrome.org), and the differentially expressed TF (DETFs) were screened in HCC-TCGA patients by matching the DEGs as shown above. Subsequently, the relationship between DETFs and PIRDEGs was analyzed, and a regulatory network was established by Cytoscape 3.6.0. The parameters of correlation coefficient > 0.1 and $P < 0.05$ was set as the cut-off values.

Enrichment analysis of DETFs and PIRDEGs

After integrating DETFs and PIRDEGs selected above, Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed by clusterProfiler and enrichplot package in R. A network model about the pathways enriched above and related genes was constructed with Cytoscape 3.6.0. $P < 0.05$ was considered a statistically significant difference.

Statistical analysis.

Statistical analysis was performed in R (version 4.0.1). IRDEGs and PIRDEGs between adjacent and HCC tissues were identified by a Wilcoxon Signedrank test. The univariate and multivariate Cox regression analyses were performed to construct the prognostic risk model. The infiltration levels of the immune cells between tumor and paratumor tissues were compared through Pearson's correlation coefficient. $P < 0.05$ was considered a statistically significant difference.

Results

Screening of immune-related differential genes (IRDEGs)

By analyzing the expressed level of 12,881 genes of HCC patients in the TCGA database, 5389 differentially expressed genes (DEGs) were identified, and 326 immune-related DEGs (IRDEGs) that included 267 upregulated and 59 downregulated genes were selected to analyze the relationship between differentially expressed genes of HCC and immune-related genes (Fig. 1A-D; $P < 0.05$, $|\log2 FC| > 1$). Then, the KEGG pathway enrichment analysis of these IRDEGs indicated that the primary pathway of IRDEGs is cytokine-cytokine receptor interaction and MAPK signaling pathway (Fig. 1E), and the results are similar to a previous study [17]. The GO analysis showed that the IRDEGs were primarily associated with signaling receptor activator and ligand activity, immune cells migration, chemotaxis, and cytokines secretion, and the response of immune cells to cytokines (Fig. 1F).

Prognosis signature of IRDEGs

31 prognosis related genes were identified from the IRDEGs, which significantly associated with overall survival (OS) of HCC patients ($P < 0.001$; Fig. 2A). While all of the prognosis related IRDEGs were high-risk genes, which predicting poor prognosis of HCC. Then, we screened 8 optimal genes from the prognosis-related IRDEGs by multivariate Cox regression analysis in R to construct a prognostic risk model, including *HSPA4*, *PSMD14*, *RBP2*, *MAPT*, *TRAF3*, *NDRG1*, *NRAS*, and *IL17D* (Table 1). We also confirmed that these genes had a higher expression in cancerous tissues of the liver than normal tissues through an online website named GIEPA (http://gepia.cancer-pku.cn/; Fig. S1A-H). Subsequently, the risk score for each patient was calculated according to the following computational formula:

Risk score = (0.02225 × expression of *HSPA4*) + (0.042251 × expression of *PSMD14*) + (0.019366 × expression of *RBP2*) + (0.197373 × expression of *MAPT*) + (0.146152 × expression of *TRAF3*) + (0.006049 × expression of *NDRG1*) + (0.027209 × expression of *NRAS*) + (0.074901 × expression of *IL17D*).

Table 1
PIRDEGs selected for prognostic signature construction.

gene	Coefficient	HR	HR.95L	HR.95H	P-value
HSPA4	0.02225	1.0225	1.003363	1.042002	0.020988
PSMD14	0.042251	1.043156	0.99353	1.09526	0.089325
RBP2	0.019366	1.019555	1.008488	1.030743	0.000506
MAPT	0.197373	1.218198	0.986943	1.503641	0.066122
TRAF3	0.146152	1.157372	0.99067	1.352126	0.065499
NDRG1	0.006049	1.006068	1.002659	1.009487	0.000476
NRAS	0.027209	1.027583	0.99719	1.058902	0.075693
IL17D	0.074901	1.077778	1.023002	1.135486	0.004885
HR, hazard ratio; HR.95H and 95L, 95% confidence interval (CI).					

The survival analysis indicated that the high-risk group (n = 184) showed poor prognosis than the low-risk group (n = 185) (Fig. 2B; P < 0.001). To identify the accuracy of the risk score in the prognostic model, we performed a ROC curve and the area under the curve (AUC) was 0.813, indicating that the prognostic model is very accurate in predicting the overall survival (OS) of HCC patients (Fig. 2C). Beyond that, there was a visualized map revealing that patients in the high-risk group died more often than those in the low-risk group (Fig. 2D and E). And a heat map showed that the eight genes selected from the prognostic model were all expressed increasingly in the high-risk group (Fig. 2F).

Independent prognostic analysis Of the prognostic model

As shown in Fig. 3A, tumor stages, T status (primary tumor size), M status (tumor migration), and risk score were significantly associated with the prognosis of HCC patients (Fig. 3A). And a multivariate analysis showed that the risk score was an independent prognostic factor associated with the prognosis of HCC patients (Fig. 3B).

The clinical signatures of the PIRDEGs

By analyzing the expression of eight genes significantly associated with the prognosis of HCC in the prognostic model, we found that the expression level of *HSPA4* and *NDRG1* was positively correlated with the stage, fustat, grade, and T status of HCC (Fig. 4A-H; P < 0.05). Besides, *NRAS* and *PSMD14* were positively correlated with the stage, fustat, and T status of HCC excepted for *NRAS* which showed little correlation with T status. The expression of *IL17D* was only associated with the grade of HCC, and the expression of *TRAF3* was associated with stage and fustat of HCC (Fig. 4I-Q).

The selected eight genes related to the prognostic model of HCC-TCGA patients were performed for survival analysis. And the results showed that the high expression of *TRAF3*, *PSMD14*, *HSPA4*, and *NRAS* were risk factors and predicted poor prognosis of HCC patients (Fig. 5A-D; $P < 0.05$). And we had verified the results on GIEPA online website that the expression of *HSPA4*, *TRAF3*, *MAPT*, *NDRG1*, *NRAS*, and *PSMD14* positively related to overall survival with a statistically significant difference (Fig. S2A-F; $P < 0.05$). But the expression of *RBP2* and *IL17D* showed no statistically significant difference with the survival of HCC patients in the two methods (Fig. 5E-H; Fig. S2G and H).

Correlation analysis between risk score and tumor-infiltrating immune cells (TIIC)

The tumor microenvironment plays an important role in tumor repression and development. Since, we analyzed the correlation between immune score, stromal score, tumor purity, and ESTIMATE score and risk score. But the risk score did not show a significant correlation with the others (Fig. S3A-D). We further analyzed the correlation between risk score and nearly twenty types of TIIC in R. The results indicated that the risk score was positively correlated with M2 macrophages, activated $CD4^+$ T memory cells, follicular helper T cells, and eosinophils (Fig. 6A-D; $P < 0.05$). However, monocytes, activated mast cells, resting mast cells, $CD8^+$ T cells, plasma cells, gamma delta T cells, activated NK cells, resting NK cells, activated DCs, resting dendritic cells (rDCs), naïve B cells, memory B cells, resting $CD4^+$ T memory cells, regulatory T cells (Tregs), neutrophils, M0 and M1 macrophages had a little correlation with the risk score and showed little statistically significant difference (Fig. S4A-P; $P > 0.05$).

The correlation between TFs and PIRDEGs

We further analyzed the DETFs in HCC-TCGA patients. Amongst 318 TFs in our symbols and 118 DETFs were found (Fig. 7A and B; $P < 0.05$, $|\log2\ FC| > 1$). Then, we analyzed the correlation between the 118 DETFs and 31 PIRDEGs. To further comprehend the correlation between DETFs and PIRDEGs, the network structure as shown below (Fig. 7C).

The GO and KEGG pathway enrichment analysis

By the KEGG enrichment analysis, 19 pathways were enriched in the DETFs and PIRDEGs of HCC-TCGA symbols. The most relevant pathways included cell cycle, cellular senescence, Epstein-Barr virus infection, hepatitis B, and HCC (Fig. 8A; $P < 0.05$). And the network map that indicated the correlation between these pathways and selected genes was plotted (Fig. 8B). Then, the GO analysis was performed to classified these selected genes into parts, including cellular component (CC), molecular function (MF), and biological process (BP). The results showed that the DETFs and PIRDEGs of HCC-TCGA patients were associated with 543 BPs, 39 CCs, and 77 MFs ($P < 0.05$; Fig. 8C). Then, we screened some biological processes significantly associated with the immune system to analyze the biological functions of the selected genes ($P < 0.05$; Fig. 8D).

Discussion

HCC is one of the most common malignant tumors and is estimated as the fourth leading cause of death [18]. The liver is a huge immune organ and complex microenvironment under physiologic conditions that could suppress abnormal inflammatory responses and maintain immunotolerance [19]. The development of immuno-oncology has transformed cancer treatment strategies over the past decade. Immune checkpoints are important in the maintenance of self-tolerance under physiologic conditions, including PD-1 and its ligand PD-L1, as well as CTLA-4 [3, 20]. The therapeutic strategies for HCC included surgery and liver transplantation, but it's not suitable for most patients with advanced liver cancer. Accordingly, immunotherapy is a potential method, such as immune checkpoint inhibition (ICI) and chimeric antigen receptor redirected T (CAR-T) cell therapy [2, 3, 20]. To better understand the correlation between the immune system and HCC, we carried out the present work about the IRGs how does it affect the prognosis of HCC patients.

First, we identified eight optimal PIRDEGs from HCC patients which were downloaded in the TCGA database to conduct a precise prognostic model for predicting overall survival (OS) of HCC patients. These eight genes are all tumor protective factors, which prevented regression in HCC patients, including *HSPA4*, *PSMD14*, *RBP2*, *MAPT*, *TRAF3*, *NDRG1*, *NRAS*, *IL17D*. But *IL17D* only showed little significant difference in the survival models. The GO analysis and KEGG pathway analysis were performed in R to discover the important biological functions and cellular molecular pathways of the PIRDEGs. And the correlation between PIRDEGs and TFs was completed with a network diagram. Thereafter, we have sufficient preconditions to accurately predict the prognosis of HCC patients in our prognostic model.

A survival analysis produced via the prognostic model indicated that patients with a high-risk score showed a poorer prognosis than the low-risk score group ($P < 0.05$). Additionally, univariate Cox prognostic analysis indicated that the risk score, pathological stage, T, and M status were independent prognostic factors and the AUC value of ROC was 0.813 which indicated that the risk score was an accurate prognostic factor for predicting the prognosis of HCC patients. And the other clinical signatures were dependent on predicting factors, such as age, gender, histological grade, pathological stage, TNM status. Furthermore, it has been reported that AUC > 0.60 was supposed to be a receivable foretell, and AUC > 0.75 was deemed to have excellent predictive value [21]. As evidence, the risk curve showed that survival time and the expression of eight selected genes were positively proportional to the risk score. Hence, the prognostic model had a precise assessment of HCC patients in many aspects, such as age, gender, pathological stage, and TNM status, for better improving the poor prognosis of patients.

We also analyzed the relationship between the sifted genes and other clinical factors including age, fustat, histological grade, pathological stage, TNM status, and OS. The fustat showed that these genes were significantly expressed higher in dead patients than alive ones. The expression level of *HSPA4* was positively correlated with histological grade, and its expression was higher in stage 2 of HCC patients than stages 1 and 3, and it was similar to the changing pattern of primary tumor size. A previous study also demonstrated that HSPA4 was associated with poor prognosis of liver cancer [22] and breast cancer

[23]. Studies have shown that *NDRG1* can promote the proliferation, progression, and poor prognosis of liver cancer [24, 25]. Our studies further proved that the expression of *NDRG1* was positively correlated with the death, histological grade, pathological stage, and primary tumor size of HCC-TCGA patients. Furthermore, in the stage and grade 2–3, as well as primary tumor size, the expression of *PSMD14* was higher than the stage and grade 1. However, *NRAS*'s influence on staging and tumor size is very weak. The expression of *IL17D* and *TRAF3* was higher in grade 3 and stage 3 respectively. A small number of researchers have used bioinformatics methods to demonstrate the differential expression of IL17D in HCC [26]. And Lv, J., et al. showed that PSMD14 can promote the proliferation and metastasis of liver cancer cells [27]. The survival analysis revealed that high expression of these genes was associated with poor prognosis in HCC patients.

Immune infiltration plays an important role in the genesis and development of the tumor. As reported, tumor-infiltrating cells were preferentially enriched in HCC, such as exhausted $CD8^+$ T cells and Tregs [28]. Increased tumor infiltration of Tregs is associated with $CD8^+$ T cell dysfunction, HCC invasiveness as well as progression, and poor patient outcomes [3, 29]. Macrophages and dendritic cells (DCs) which can have anti-tumor functions may be transformed into tumor-associated macrophages (TAMs) and myeloid-derived suppressor cells (MDSCs) phenotype which have an immunosuppressive and pro-tumor feature. And the high level of TAMs and DCs in HCC was correlated with poor prognosis of HCC patients [3, 30]. In the present study, the results showed that the risk score was positively correlated with M2 macrophages, activated $CD4^+$ T memory cells, follicular helper T cells, and eosinophils. In the previous studies, Junyu Long .et.al had found that HCC patients in the high-risk group exhibited higher fractions of follicular helper T cells than the low-risk group[31]. And Oscar W H Yeung .et.al discovered that M2 macrophages contribute to invasiveness and poor prognosis of HCC[32].

Finally, we conducted GO and KEGG pathway analysis to deeply explored the role of these genes and DETFs selected from our prognostic model, further comprehending the molecular mechanisms of HCC development, deterioration, and poor prognosis. The KEGG pathway analysis showed that the expression of these genes and DETFs was highly associated with cell cycle, cell senescence, Epstein-Barr virus infection, hepatitis B, and hepatocellular carcinoma. And the GO analysis showed that these genes and DETFs were associated with chromatin modification, transcription regulator complex, and DNA-binding transcription activator activity. We selected partial immune-related biological processes associated with the selected TFs and genes, and further analysis revealed that *IRF5*, *POLR3A*, *PRKDC*, *CEBPA*, *SRC*, *PSMD14*, *NRAS*, *IL17D*, *JMJD6*, *MAPT*, *SOX4*, and *TRAF3* were intricately associated with the immune functions as follow: positive regulation of type I interferon production, positive regulation of defense response, macrophage activation, pro-B cell differentiation, regulation of interferon-beta production, lymphoid progenitor cell differentiation and activation of the innate immune response.

In general, the results of our study exhibited a potential direction of factors related to liver cancer prognosis and immunity. Nothing but some deficiencies existed in the present work. On the one hand, our sample size was not large enough to take into account the differences between individual patients and

the heterogeneity of the tumor environment. On the other hand, we have not verified the pathways and mechanisms of these genes by in vitro and in vivo experiments.

Conclusion

We constructed a prognostic model based on the selected IRDEGs of HCC-TCGA patients to precisely predict their prognosis. The risk score of the selected genes could be served as an independent prognostic marker to grasp the survival outcomes of patients. And Some possible mechanisms have been found to correlate with the genes and TFs of our model, but need to be verified by more experiments.

Abbreviations

HCC: Hepatocellular carcinoma

ICI: immune checkpoint inhibitor

CAR-T: chimeric antigen receptor redirected T cell

TAAs: tumor-associated antigens

TCGA: The Cancer Genome Atlas

CTLA-4: cytotoxic T-lymphocyte-associated protein 4

PD-1/PD-L1: programmed death receptor 1 and its ligand

TIM-3: T cell immunoglobulin mucin-3

DEGs: differentially expressed genes

IRGs: immune-related genes

IRDEGs: immune-related differentially expressed genes

PIRDEGs: prognostic immune-related differentially expressed genes

FDR: false discovery rate

ROC: receiver operating characteristic

AUC: the area under the curve

TFs: transcription factors

DETFs: differentially expressed transcription factors

OS: Overall survival

DCs: dendritic cells

TAM: tumor-associated macrophages

MDSCs: myeloid-derived suppressor cells

Declarations

Acknowledgment and funding

This work was supported by grants from the National Natural Science Foundation of China (No.81772734) and the Innovation Capability Support Program of Shanxi (Program No. 2020PT-021).

Author contribution

Xiang Zhou, Ke-Ying Zhang, Fa Yang, Zhao-Ying Fu, and Wei-Hong Wen designed and conceived the experimentations. Xiang Zhou, Ke-Ying Zhang, Fa Yang, Chao Xu, Jian-Hua Jiao, and Shao-Jie Liu contributed to data download and analysis. Xiang Zhou, Yan Wang, Yao Jiang, and Xiao-Long Zhao were responsible for writing the manuscript. All authors read and approved the final manuscript.

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Competing interest

The authors declared that they have no competing interests.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

References

1. McGlynn, K.A., J.L. Petrick, and H.B. El-Serag, *Epidemiology of Hepatocellular Carcinoma.* Hepatology, 2020.
2. Dal Bo, M., et al., *New insights into the pharmacological, immunological, and CAR-T-cell approaches in the treatment of hepatocellular carcinoma.* Drug Resist Updat, 2020. **51**: p. 100702.
3. Rizvi, S., J. Wang, and A.B. El-Khoueiry, *Liver Cancer Immunity.* Hepatology, 2020.
4. Lu, C., et al., *Current perspectives on the immunosuppressive tumor microenvironment in hepatocellular carcinoma: challenges and opportunities.* Mol Cancer, 2019. **18**(1): p. 130.
5. Hou, J., et al., *The immunobiology of hepatocellular carcinoma in humans and mice: Basic concepts and therapeutic implications.* J Hepatol, 2020. **72**(1): p. 167-182.
6. Gao, C., et al., *SNP mutation-related genes in breast cancer for monitoring and prognosis of patients: A study based on the TCGA database.* Cancer Med, 2019. **8**(5): p. 2303-2312.
7. Li, B., et al., *Development and Validation of an Individualized Immune Prognostic Signature in Early-Stage Nonsquamous Non-Small Cell Lung Cancer.* JAMA Oncol, 2017. **3**(11): p. 1529-1537.
8. Tao, H., et al., *Integrative bioinformatics analysis of a prognostic index and immunotherapeutic targets in renal cell carcinoma.* Int Immunopharmacol, 2020. **87**: p. 106832.
9. Hirsch, T.Z., et al., *BAP1 mutations define a homogeneous subgroup of hepatocellular carcinoma with fibrolamellar-like features and activated PKA.* J Hepatol, 2020. **72**(5): p. 924-936.
10. Saillard, C., et al., *Predicting survival after hepatocellular carcinoma resection using deep-learning on histological slides.* Hepatology, 2020.
11. Ma, L.J., et al., *Clinical significance of PD-1/PD-Ls gene amplification and overexpression in patients with hepatocellular carcinoma.* Theranostics, 2018. **8**(20): p. 5690-5702.
12. Xie, K., et al., *OX40 expression in hepatocellular carcinoma is associated with a distinct immune microenvironment, specific mutation signature, and poor prognosis.* Oncoimmunology, 2018. **7**(4): p. e1404214.
13. Tang, X., et al., *Clinical significance of the immune cell landscape in hepatocellular carcinoma patients with different degrees of fibrosis.* Ann Transl Med, 2019. **7**(20): p. 528.
14. Yue, C., H. Ma, and Y. Zhou, *Identification of prognostic gene signature associated with microenvironment of lung adenocarcinoma.* PeerJ, 2019. **7**: p. e8128.
15. Wan, B., et al., *Prognostic value of immune-related genes in clear cell renal cell carcinoma.* Aging (Albany NY), 2019. **11**(23): p. 11474-11489.
16. Qiu, H., et al., *Identification and Validation of an Individualized Prognostic Signature of Bladder Cancer Based on Seven Immune Related Genes.* Front Genet, 2020. **11**: p. 12.
17. Huang, R., et al., *Immune system-associated genes increase malignant progression and can be used to predict clinical outcome in patients with hepatocellular carcinoma.* Int J Oncol, 2020. **56**(5): p. 1199-1211.
18. Fitzmaurice, C., et al., *Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2017: A*

Systematic Analysis for the Global Burden of Disease Study. JAMA Oncol, 2019. **5**(12): p. 1749-68.

19. Ringelhan, M., et al., *The immunology of hepatocellular carcinoma.* Nat Immunol, 2018. **19**(3): p. 222-232.
20. Topalian, S.L., C.G. Drake, and D.M. Pardoll, *Immune checkpoint blockade: a common denominator approach to cancer therapy.* Cancer Cell, 2015. **27**(4): p. 450-61.
21. Cho, S.H., et al., *The AP2M1 gene expression is a promising biomarker for predicting survival of patients with hepatocellular carcinoma.* J Cell Biochem, 2019. **120**(3): p. 4140-4146.
22. Hu, B., X.B. Yang, and X.T. Sang, *Development of an immune-related prognostic index associated with hepatocellular carcinoma.* Aging (Albany NY), 2020. **12**(6): p. 5010-5030.
23. Gu, Y., et al., *Tumor-educated B cells selectively promote breast cancer lymph node metastasis by HSPA4-targeting IgG.* Nat Med, 2019. **25**(2): p. 312-322.
24. Chua, M.S., et al., *Overexpression of NDRG1 is an indicator of poor prognosis in hepatocellular carcinoma.* Mod Pathol, 2007. **20**(1): p. 76-83.
25. Liu, Y., et al., *Long noncoding RNA CCAT2 promotes hepatocellular carcinoma proliferation and metastasis through up-regulation of NDRG1.* Exp Cell Res, 2019. **379**(1): p. 19-29.
26. Shi, W., et al., *Exploration of Prognostic Index based on Immune-related Genes in Patients with Liver Hepatocellular Carcinoma.* Biosci Rep, 2020.
27. Lv, J., et al., *Deubiquitinase PSMD14 enhances hepatocellular carcinoma growth and metastasis by stabilizing GRB2.* Cancer Lett, 2020. **469**: p. 22-34.
28. Zheng, C., et al., *Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing.* Cell, 2017. **169**(7): p. 1342-1356.e16.
29. Fu, J., et al., *Increased regulatory T cells correlate with CD8 T-cell impairment and poor survival in hepatocellular carcinoma patients.* Gastroenterology, 2007. **132**(7): p. 2328-39.
30. Zhang, Q., et al., *Landscape and Dynamics of Single Immune Cells in Hepatocellular Carcinoma.* Cell, 2019. **179**(4): p. 829-845.e20.
31. Long, J., et al., *Development and validation of a TP53-associated immune prognostic model for hepatocellular carcinoma.* EBioMedicine, 2019. **42**: p. 363-374.
32. Yeung, O.W., et al., *Alternatively activated (M2) macrophages promote tumour growth and invasiveness in hepatocellular carcinoma.* J Hepatol, 2015. **62**(3): p. 607-16.

Figures

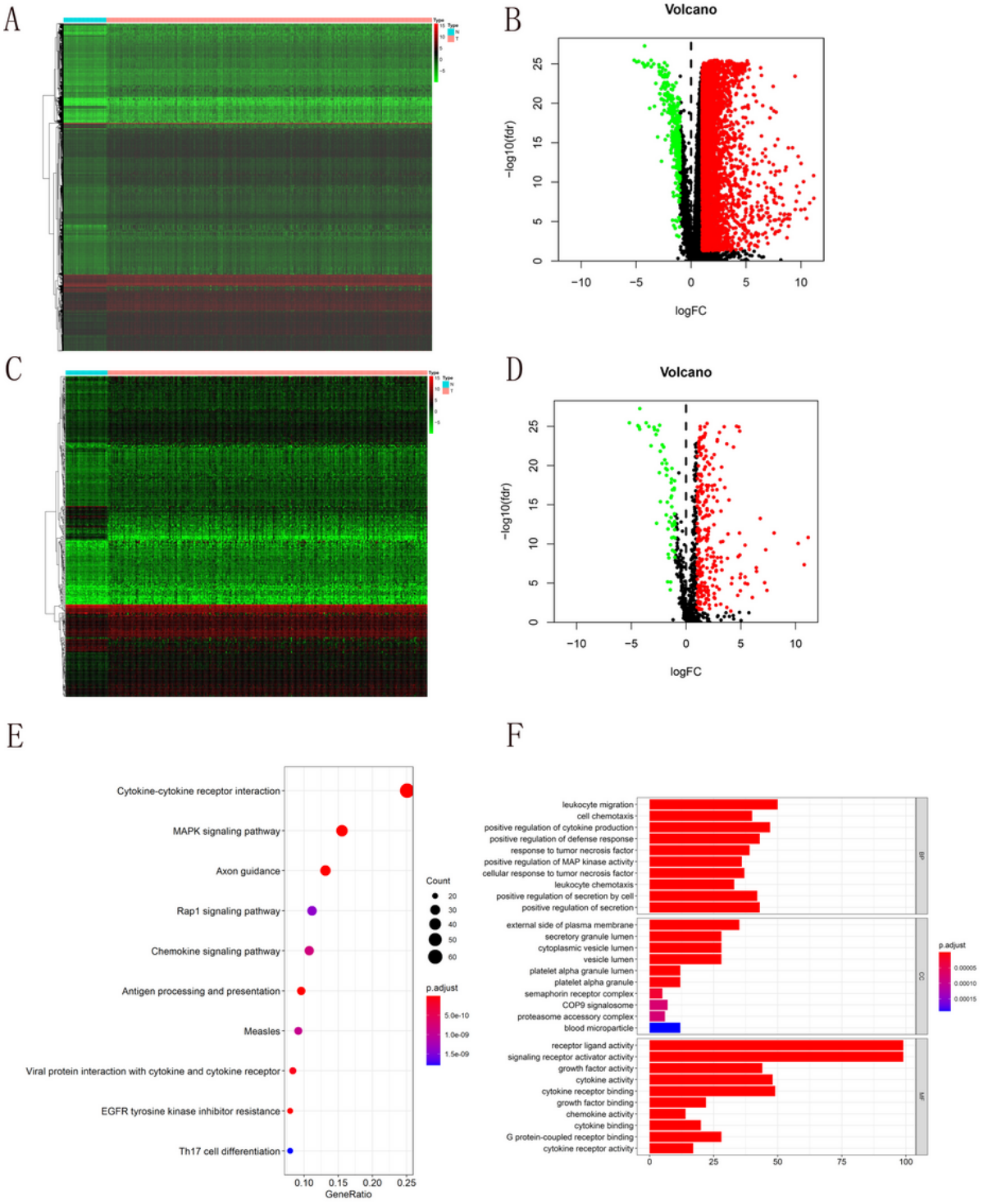

Figure 1

Screening of immune-related differentially expressed genes (IRDEGs). (A-B) Heat map and volcano plot of the differentially expressed genes (DEGs) in patients with HCC derived from the TCGA database. (C-D) Heat map and volcano plot of immune-related DEGs (IRDEGs). The color of red, green, and black repressed up-regulated, down-regulated, and invariable genes respectively. (E) The KEGG pathway analysis of the IRDEGs. (F) The GO analysis of the IRDEGs.

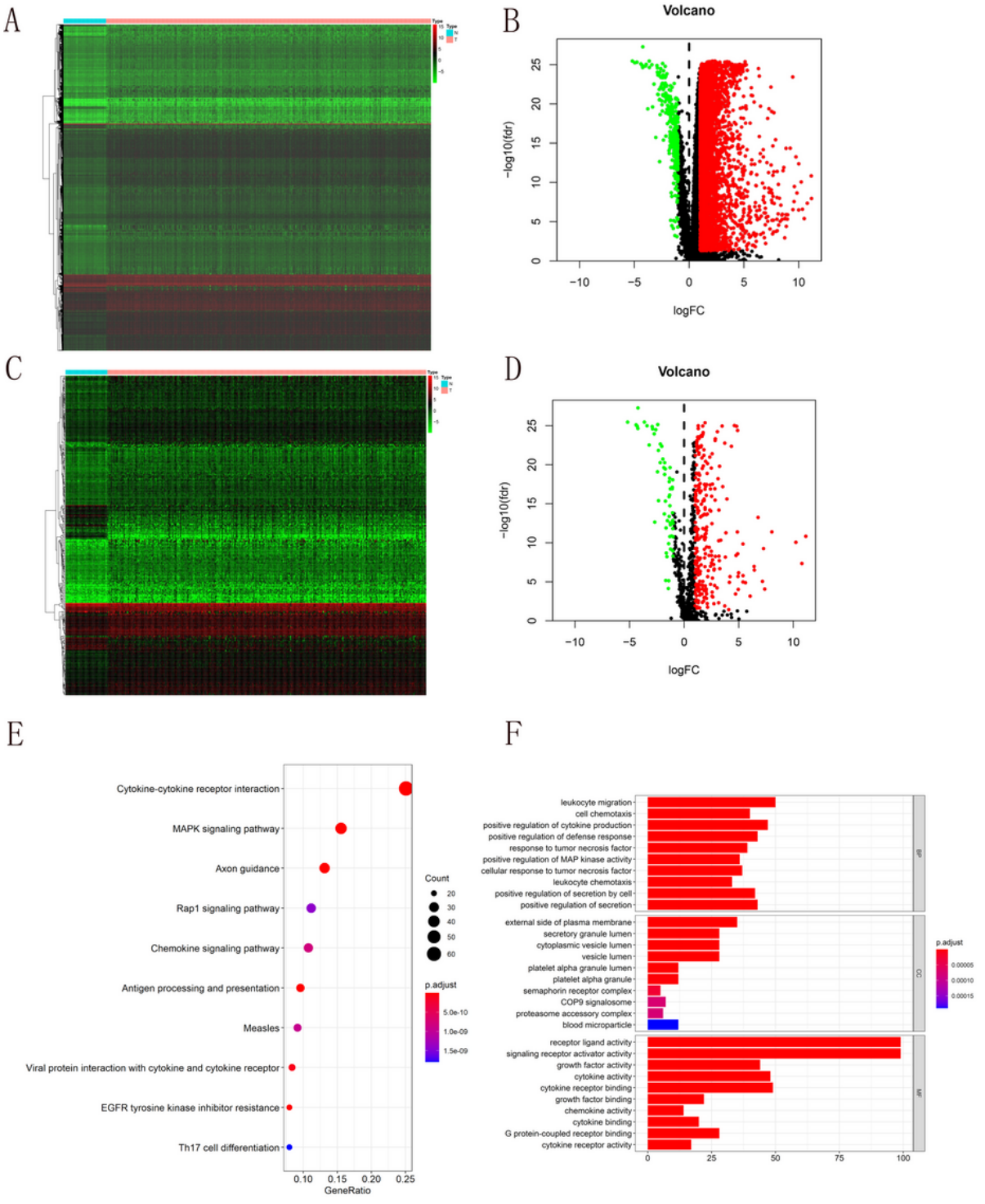

Figure 1

Screening of immune-related differentially expressed genes (IRDEGs). (A-B) Heat map and volcano plot of the differentially expressed genes (DEGs) in patients with HCC derived from the TCGA database. (C-D) Heat map and volcano plot of immune-related DEGs (IRDEGs). The color of red, green, and black repressed up-regulated, down-regulated, and invariable genes respectively. (E) The KEGG pathway analysis of the IRDEGs. (F) The GO analysis of the IRDEGs.

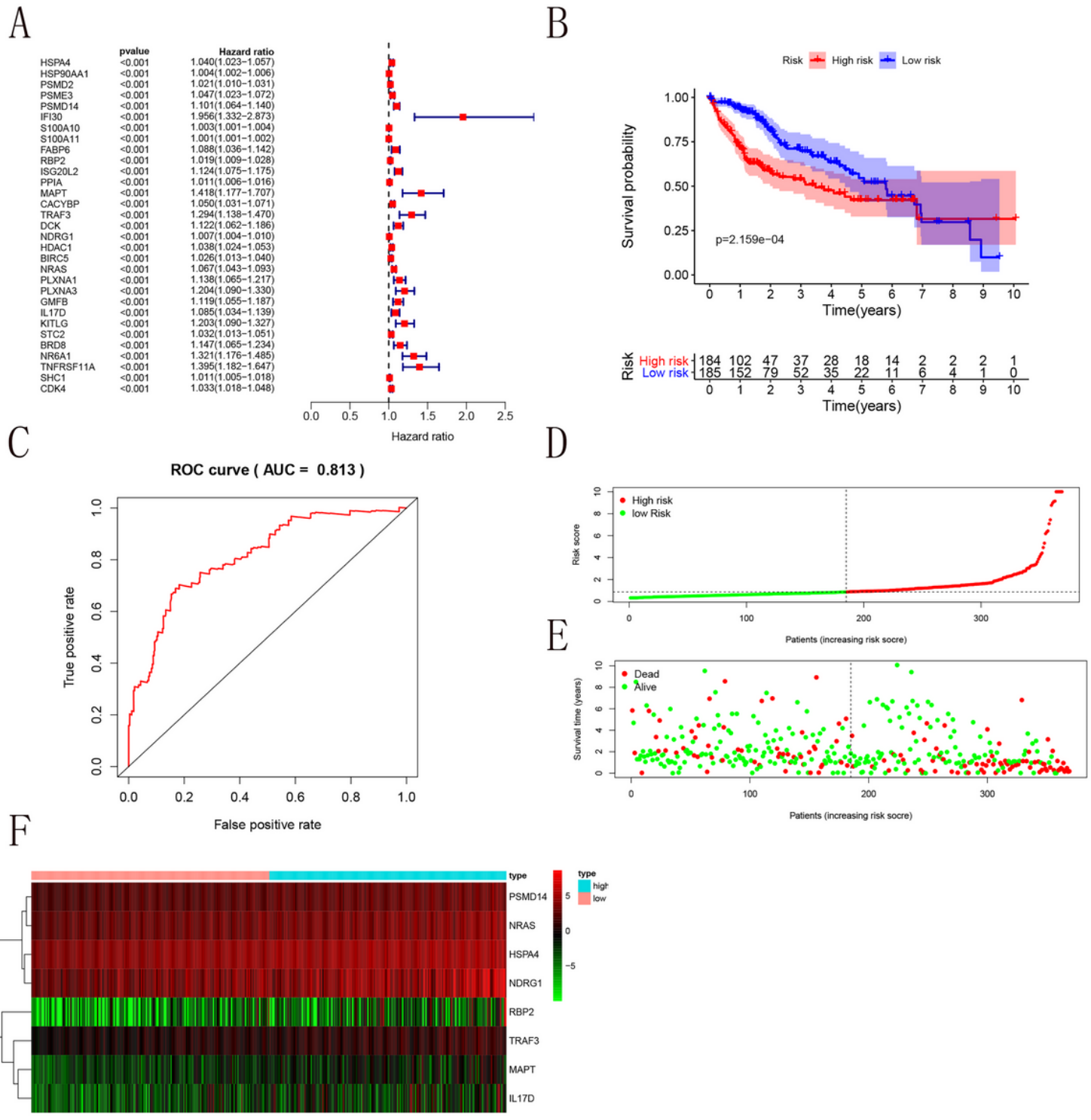

Figure 2

Prognostic Signatures of the IRDEGs. (A) The forest graph of prognosis related IRDEGs. The color of red represented that their hazard ratio > 1. (B) The survival curve contained the high-risk group and low-risk group. (C) ROC curve analysis. (D-E) The distribution diagram consisted of the risk score and the survival status of TCGA-derived HCC patients. Each point represented a sample, and the red and green

represented the dead status and alive status of HCC patients respectively. (F) A heat map showed that the correlation between the expression of eight genes and risk types (high or low).

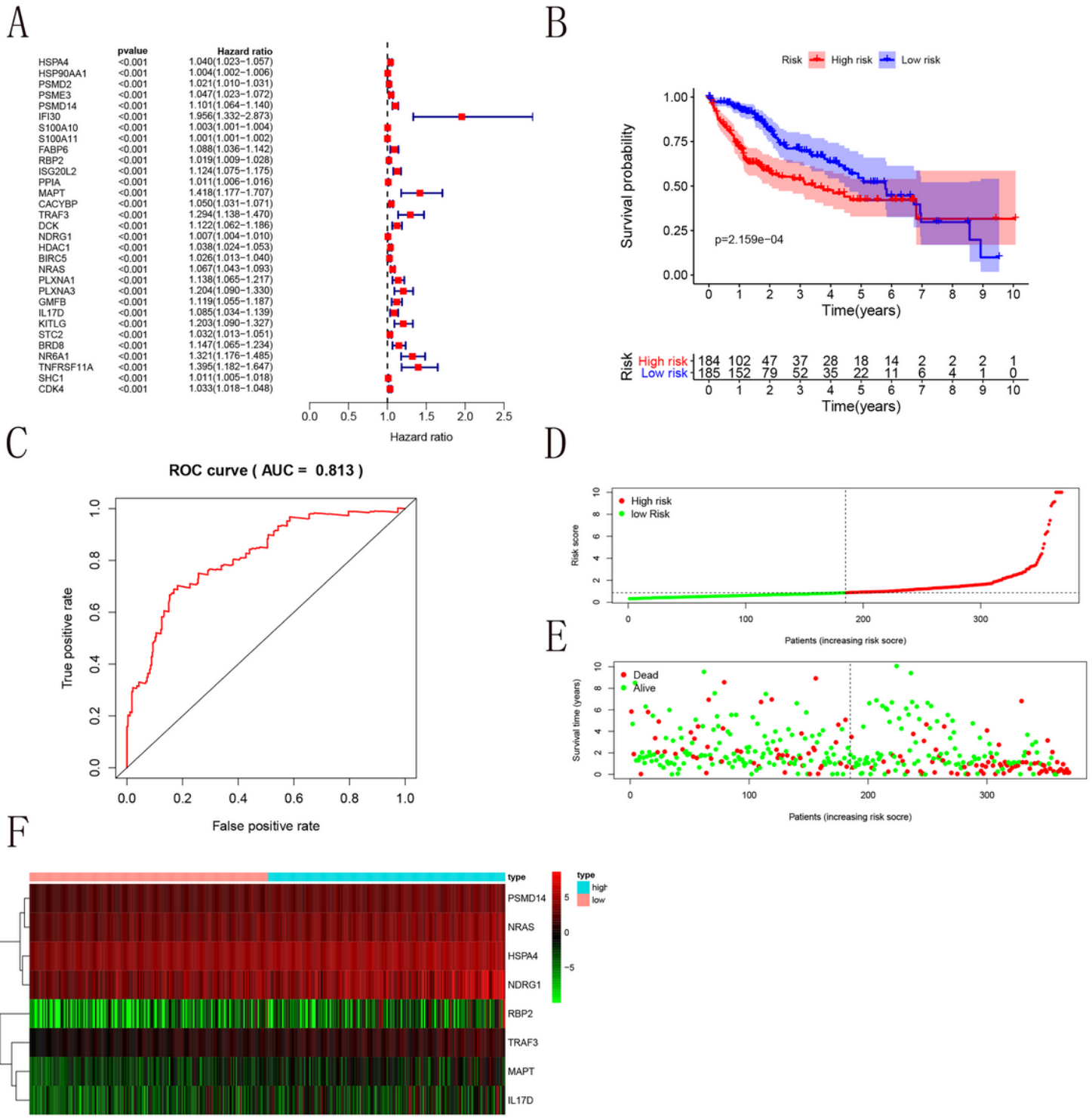

Figure 2

Prognostic Signatures of the IRDEGs. (A) The forest graph of prognosis related IRDEGs. The color of red represented that their hazard ratio > 1. (B) The survival curve contained the high-risk group and low-risk group. (C) ROC curve analysis. (D-E) The distribution diagram consisted of the risk score and the survival

status of TCGA-derived HCC patients. Each point represented a sample, and the red and green represented the dead status and alive status of HCC patients respectively. (F) A heat map showed that the correlation between the expression of eight genes and risk types (high or low).

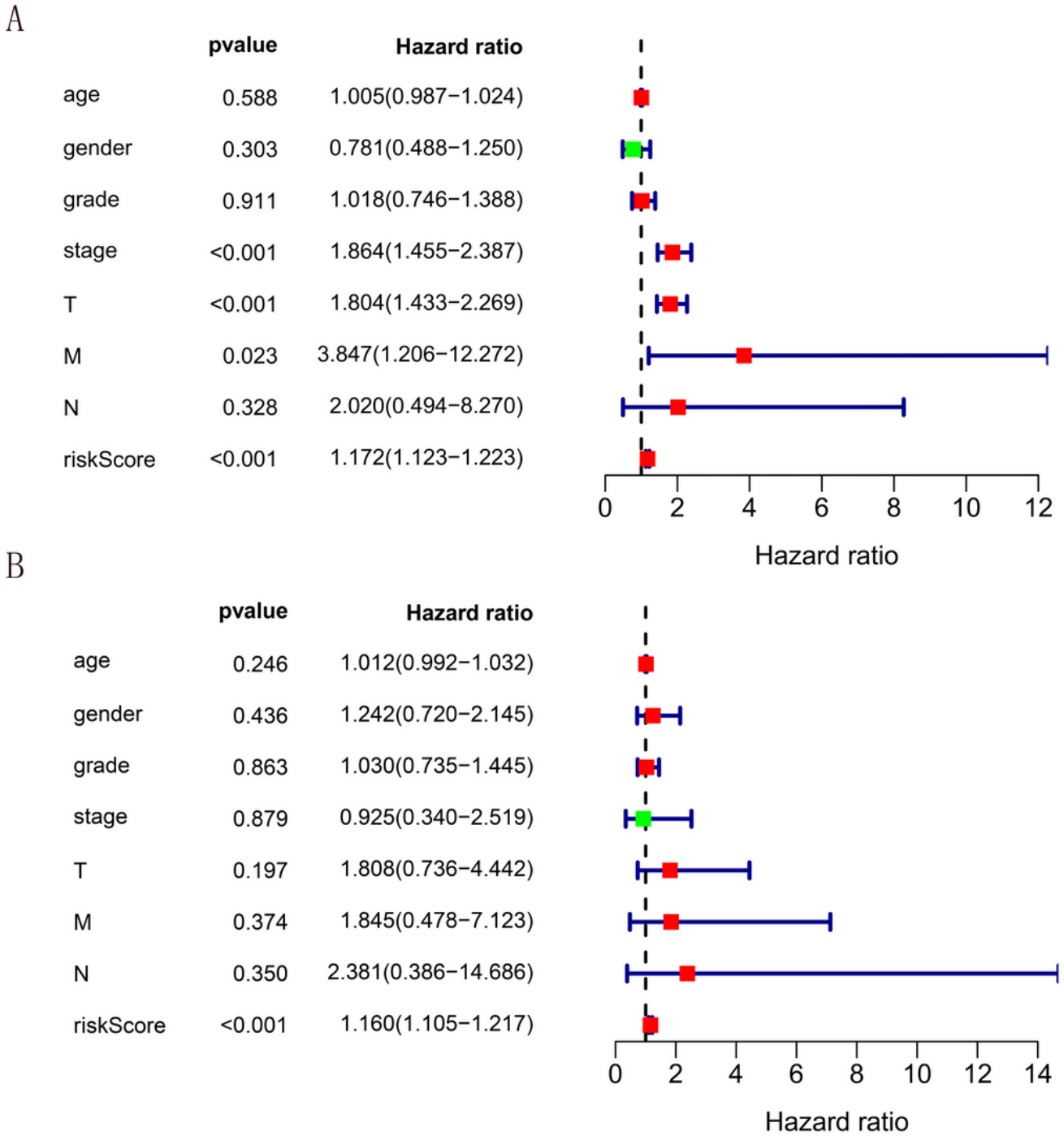

Figure 3

Independent prognostic analysis of the prognostic model. (A) Univariate Cox regression analysis was performed to identify prognosis associated factors. (B) Multivariate Cox regression analysis was used to

assess independent prognostic factors. The red and green spots on behalf of the hazard ratio >1 and <1, respectively.

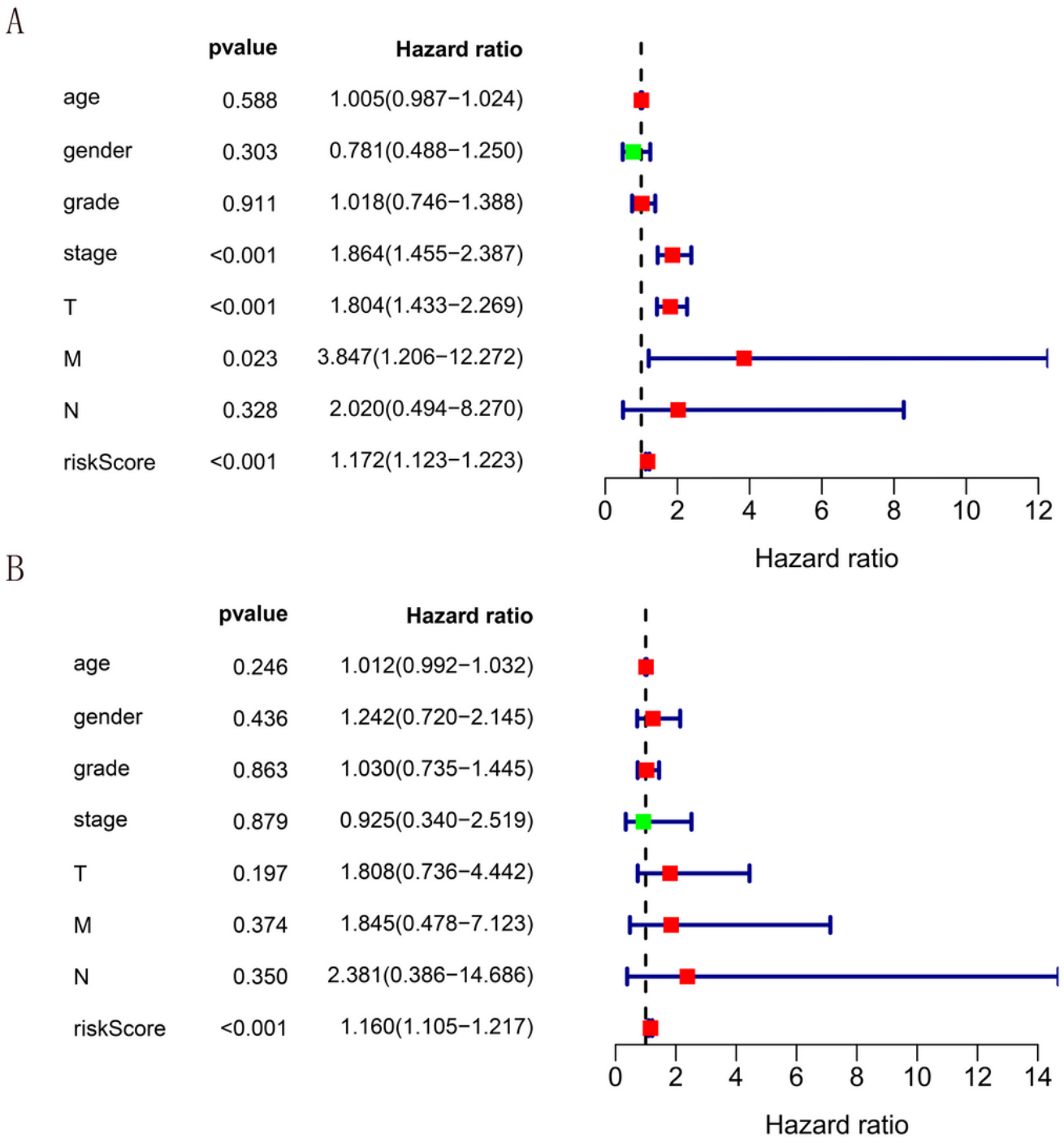

Figure 3

Independent prognostic analysis of the prognostic model. (A) Univariate Cox regression analysis was performed to identify prognosis associated factors. (B) Multivariate Cox regression analysis was used to

assess independent prognostic factors. The red and green spots on behalf of the hazard ratio >1 and <1, respectively.

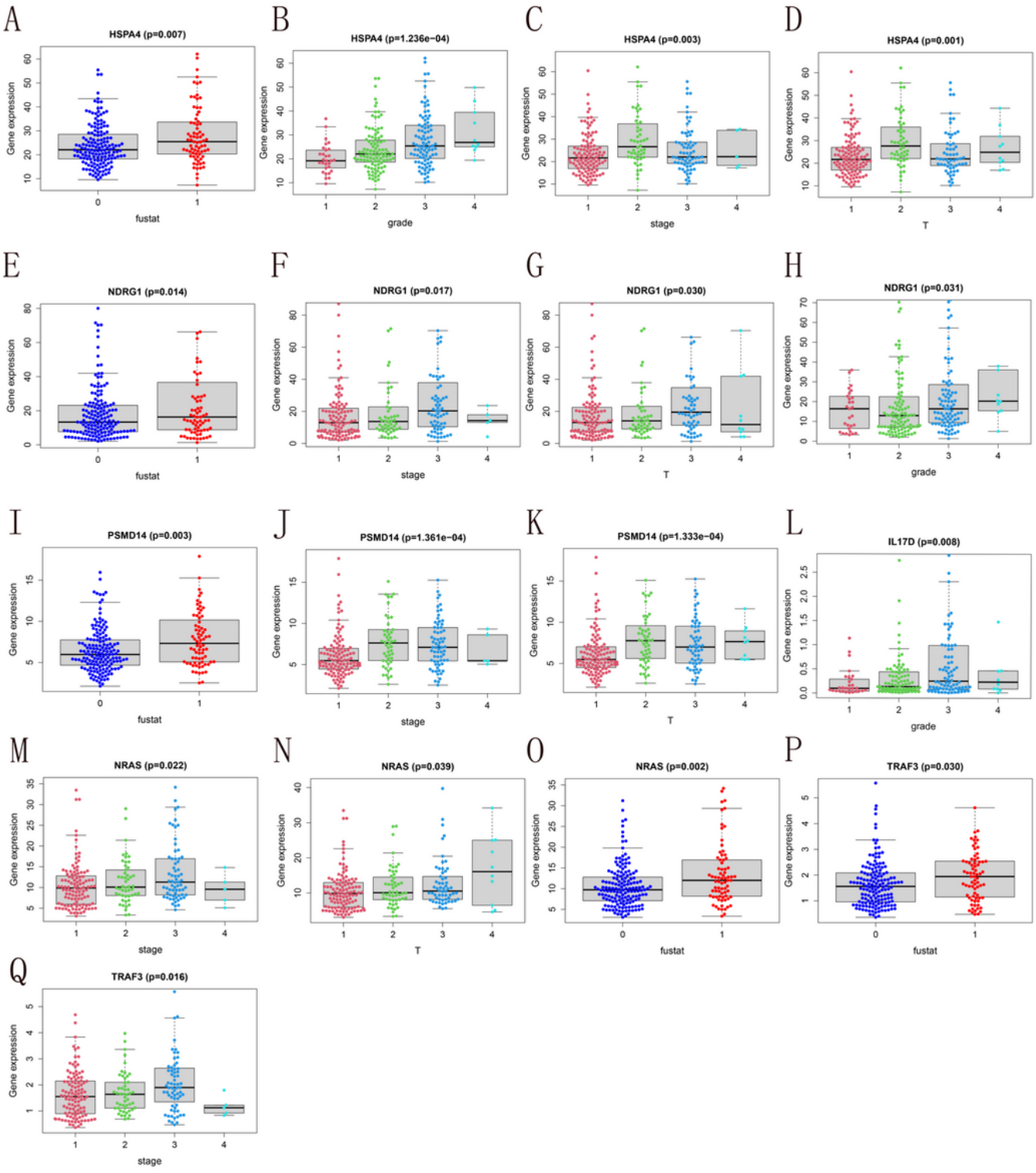

Figure 4

The clinical signatures of the PIRDEGs. (A-D) The correlation between HSPA4 and fustat, grade, stage, and T of HCC-TCGA patients respectively. (E-H) The correlation between NDRG1 and fustat, stage, T, and grade respectively. (I-K) The correlation between PSMD14 and fustat, stage, and T respectively. (L) The

correlation between IL17D and grade. (M-O) The correlation between NRAS and stage, T, and fustat respectively. (P-Q) The correlation between TRAF3 and fustat and stage.

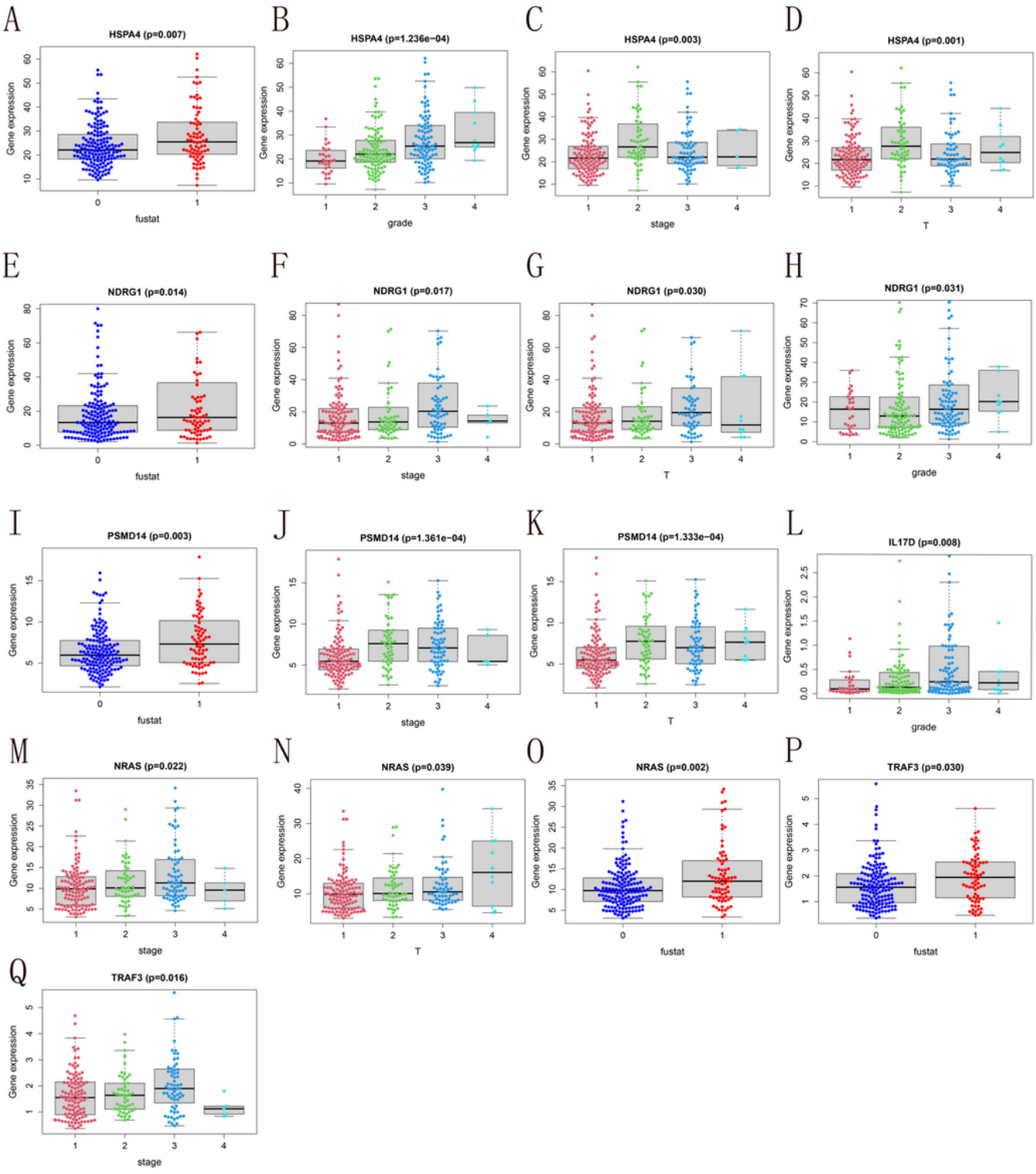

Figure 4

The clinical signatures of the PIRDEGs. (A-D) The correlation between HSPA4 and fustat, grade, stage, and T of HCC-TCGA patients respectively. (E-H) The correlation between NDRG1 and fustat, stage, T, and grade respectively. (I-K) The correlation between PSMD14 and fustat, stage, and T respectively. (L) The

correlation between IL17D and grade. (M-O) The correlation between NRAS and stage, T, and fustat respectively. (P-Q) The correlation between TRAF3 and fustat and stage.

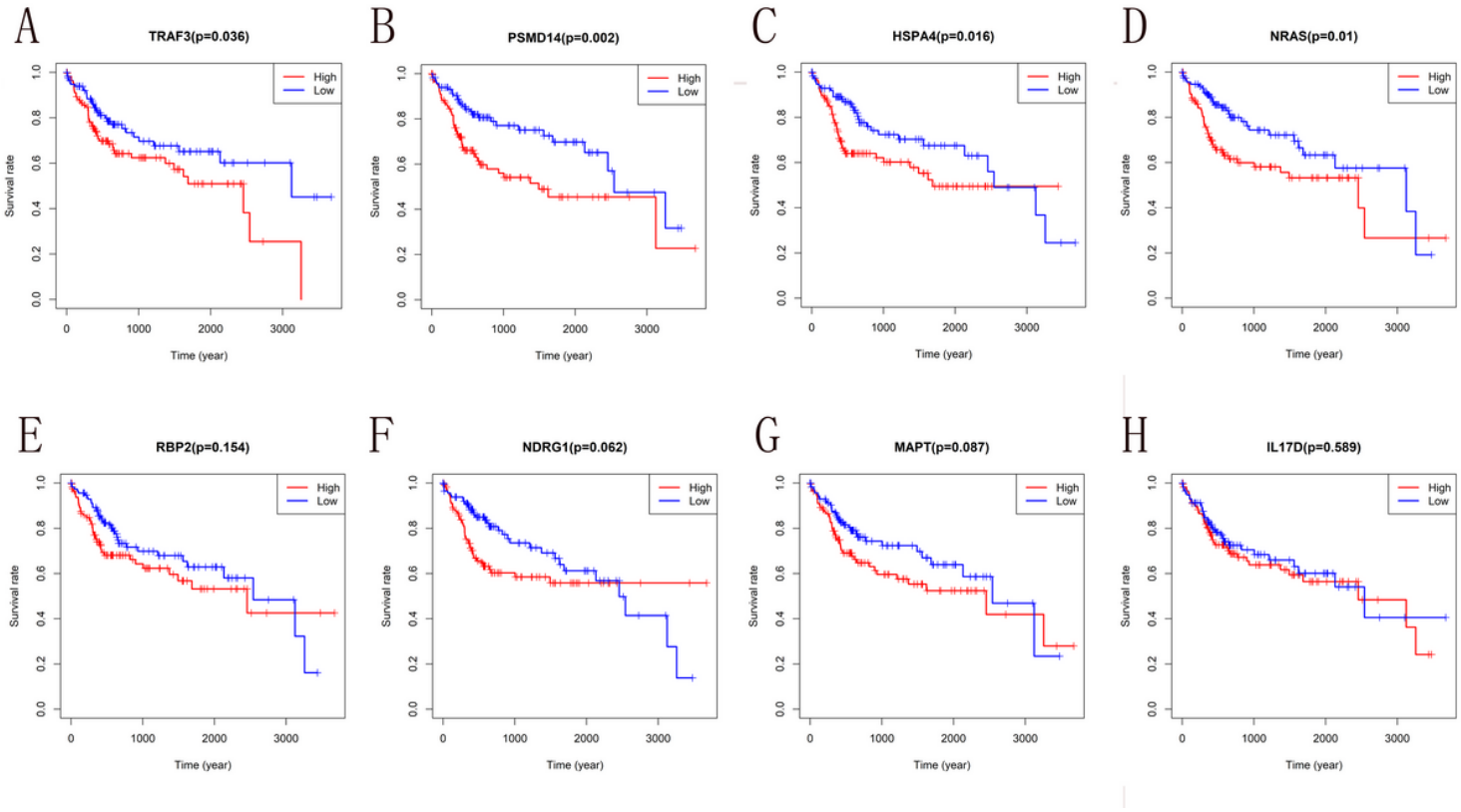

Figure 5

The survival analysis of genes derived from the prognostic model. (A) TRAF3. (B) PAMD14. (C) HSPA4. (D) NRAS. (E) RBP2. (F) NDRG1. (G) MAPT. (H) IL-17D

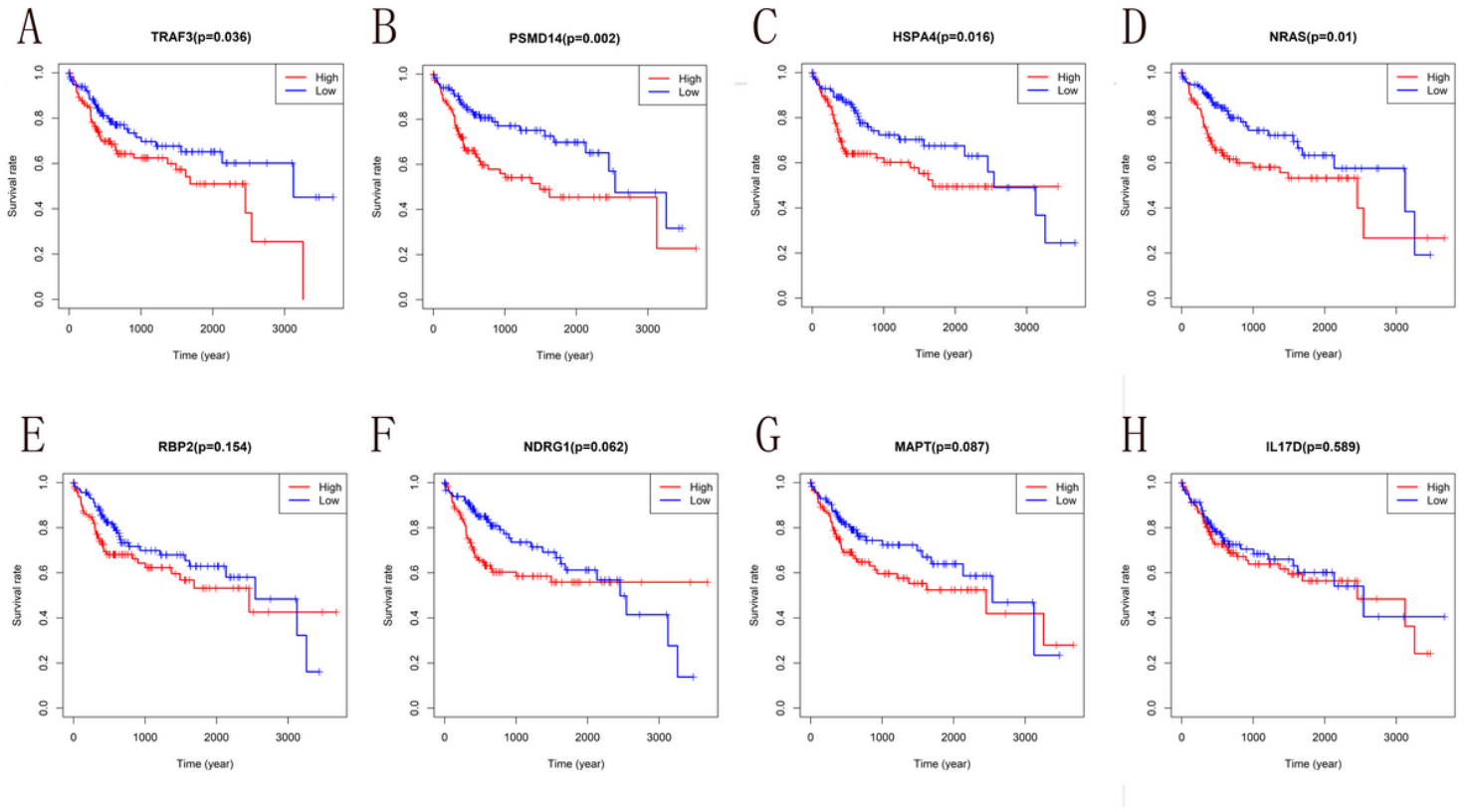

Figure 5

The survival analysis of genes derived from the prognostic model. (A) TRAF3. (B) PAMD14. (C) HSPA4. (D) NRAS. (E) RBP2. (F) NDRG1. (G) MAPT. (H) IL-17D

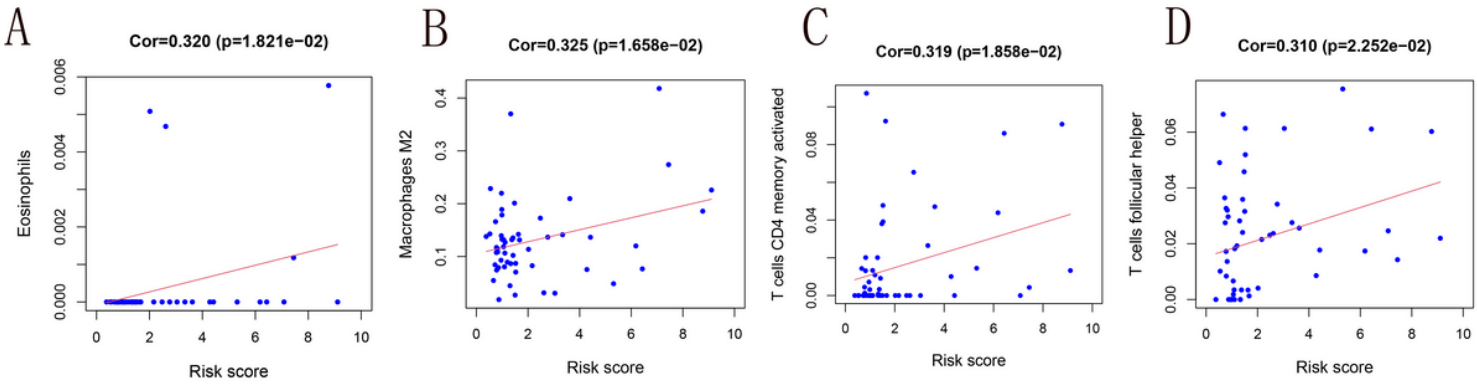

Figure 6

The correlation analysis between risk score and tumor-infiltrating immune cells (TIIC). (A-D) The relationship between TIICs and risk score. (A) Eosinophils. (B) M2 macrophages. (C) activated memory CD4+ T cells. (D) follicular helper T cells.

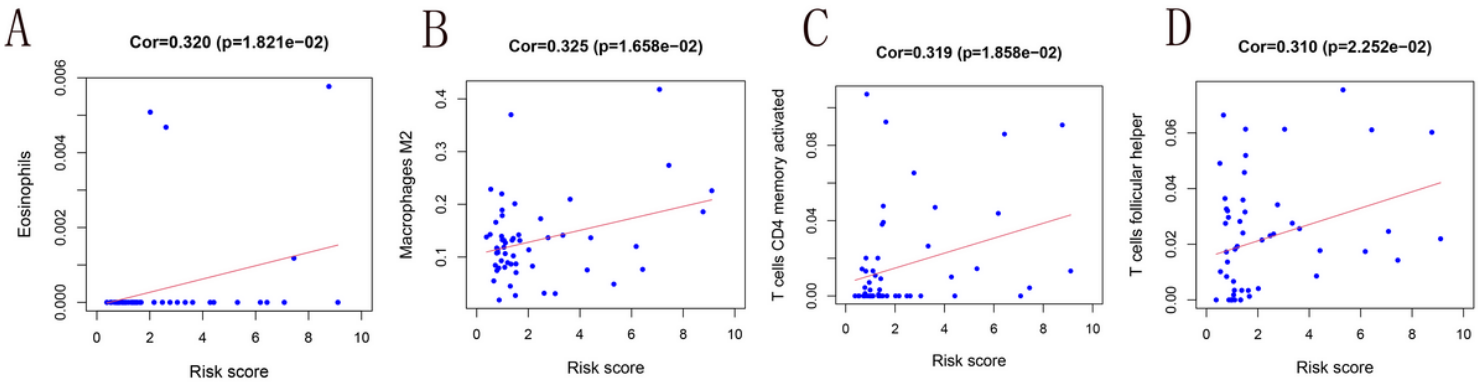

Figure 6

The correlation analysis between risk score and tumor-infiltrating immune cells (TIIC). (A-D) The relationship between TIICs and risk score. (A) Eosinophils. (B) M2 macrophages. (C) activated memory CD4+ T cells. (D) follicular helper T cells.

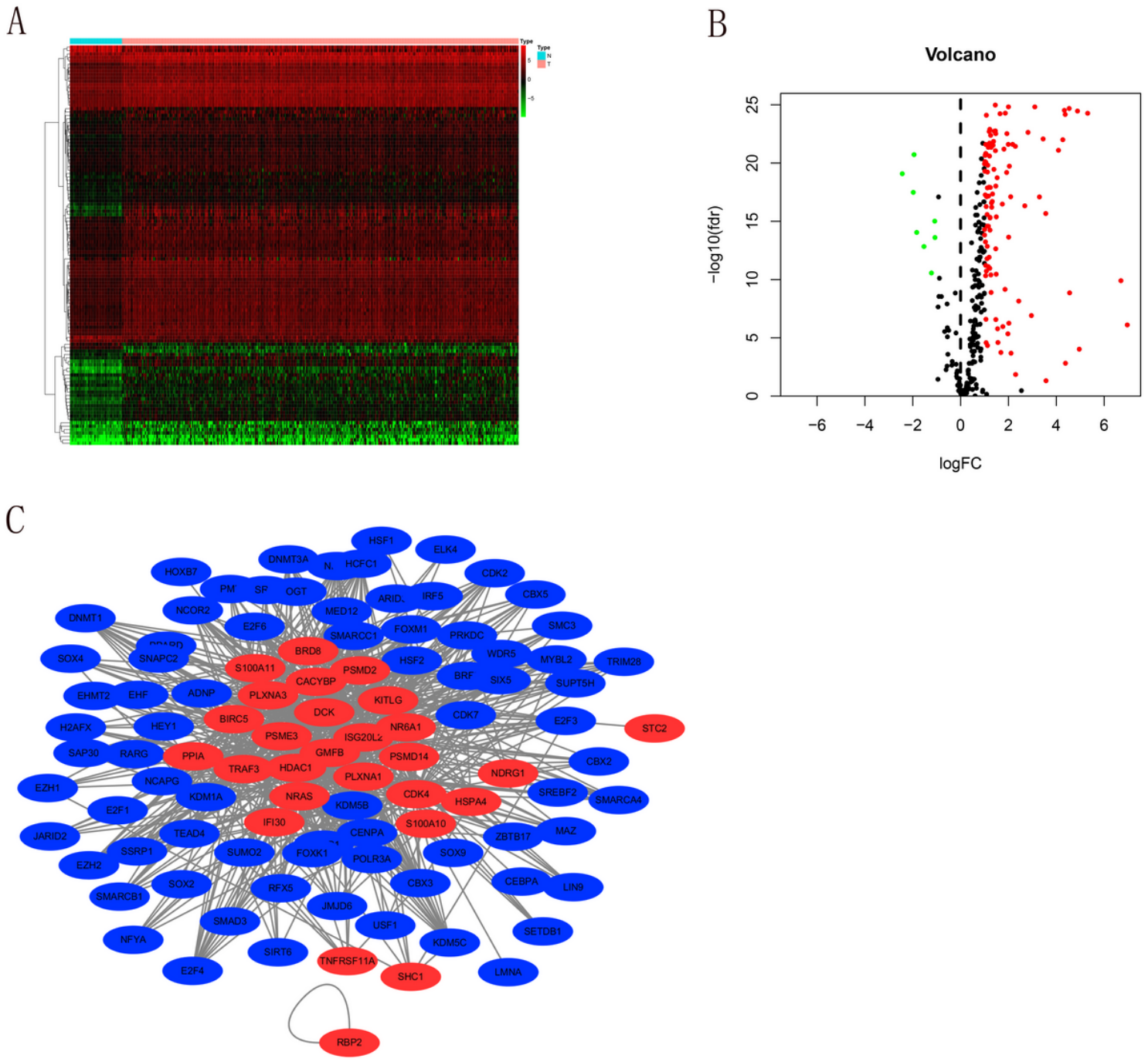

Figure 7

The correlation between TFs and PIRDEGs. (A) Heat map of the differentially expressed TFs (DETFs) of HCC-TCGA patients; (B) Volcano plot of DETFs of HCC-TCGA patients; (C) Regulatory network of DETFs and prognosis related IRDEGs. The color of red represents high-risk score PIRDEGs and the blue represents DETFs.

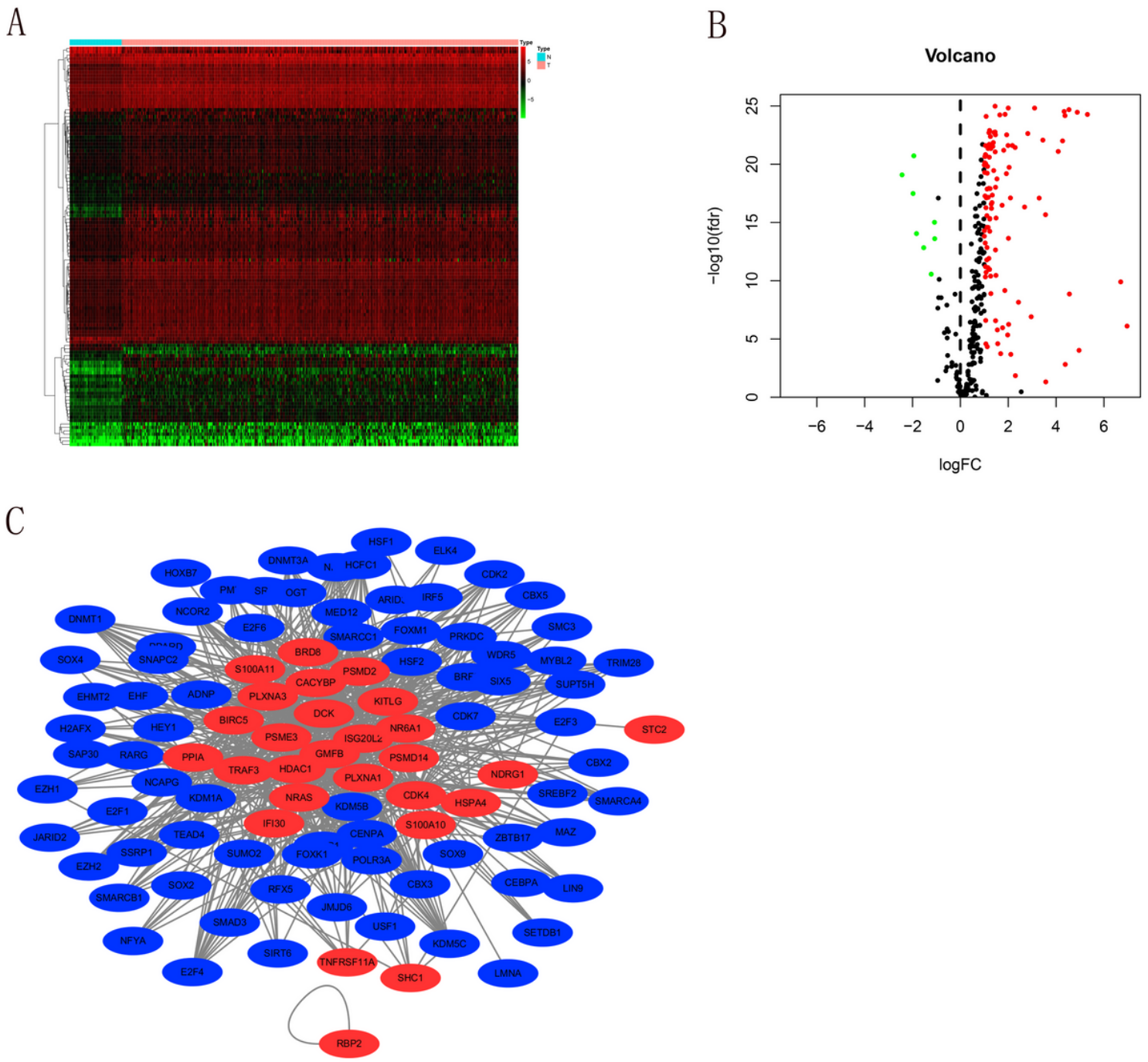

Figure 7

The correlation between TFs and PIRDEGs. (A) Heat map of the differentially expressed TFs (DETFs) of HCC-TCGA patients; (B) Volcano plot of DETFs of HCC-TCGA patients; (C) Regulatory network of DETFs and prognosis related IRDEGs. The color of red represents high-risk score PIRDEGs and the blue represents DETFs.

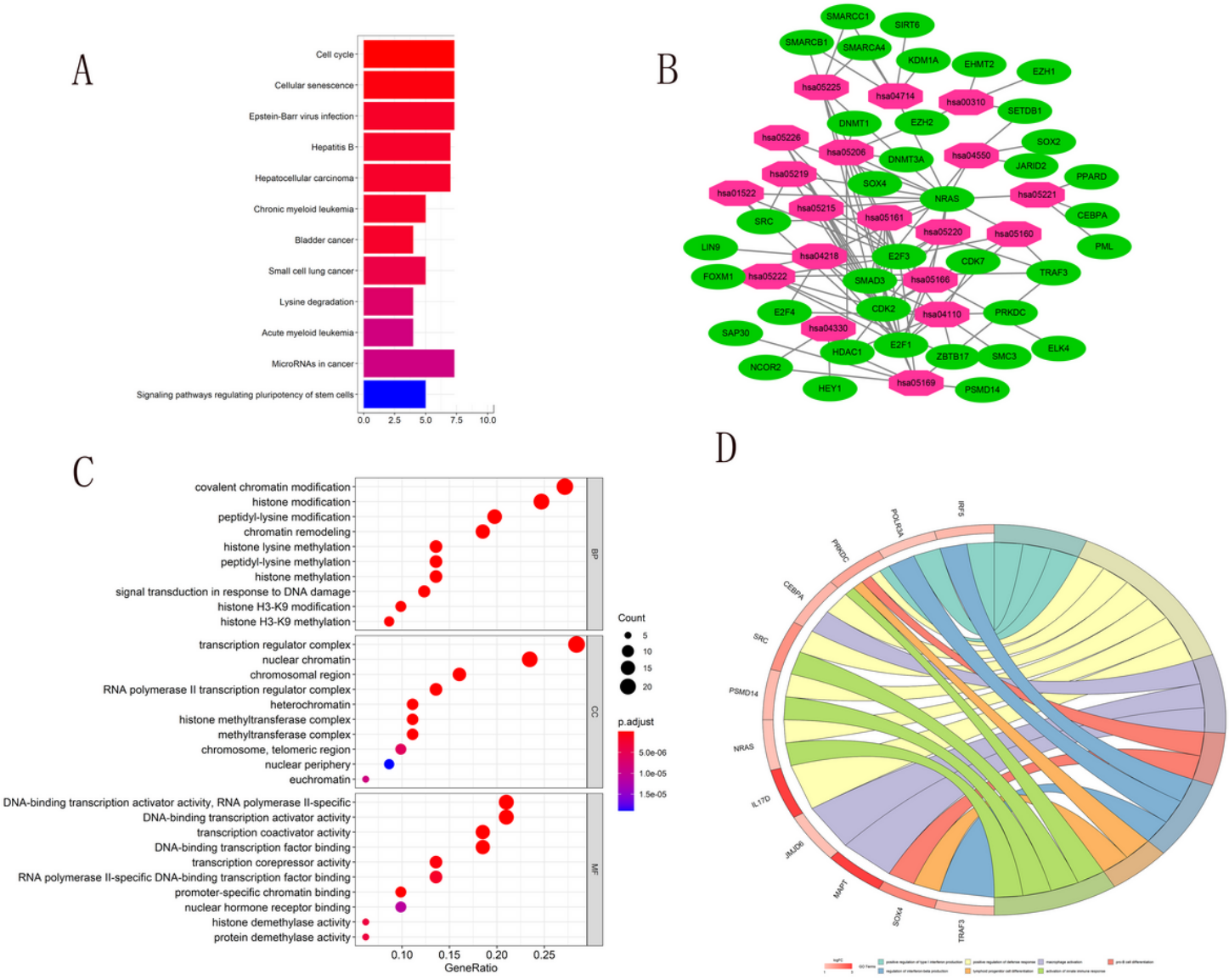

Figure 8

The GO and KEGG enrichment analysis. (A) the KEGG pathway analysis of prognostic IRDEGs and DETFs. (B) network map of screening pathways and genes. hsa04110, Cell cycle; hsa04218, Cellular senescence; hsa05169, Epstein-Barr virus infection; hsa05161, Hepatitis B; hsa05225, Hepatocellular carcinoma; hsa05220, Chronic myeloid leukemia; hsa05219, Bladder cancer; hsa05222, Small cell lung cancer; hsa00310, Lysine degradation; hsa05221, Acute myeloid leukemia; hsa05206, MicroRNAs in cancer; hsa04550, Signaling pathways regulating pluripotency of stem cells; hsa05226, Gastric cancer; hsa05166, Human T-cell leukemia virus 1 infection; hsa05160, Hepatitis C; hsa05215, Prostate cancer; hsa01522, Endocrine resistance; hsa04714, Thermogenesis; hsa04330, Notch signaling pathway; (C) The GO analysis of prognostic IRDEGs and DETFs. (D) immune-related biological processes of selected genes.

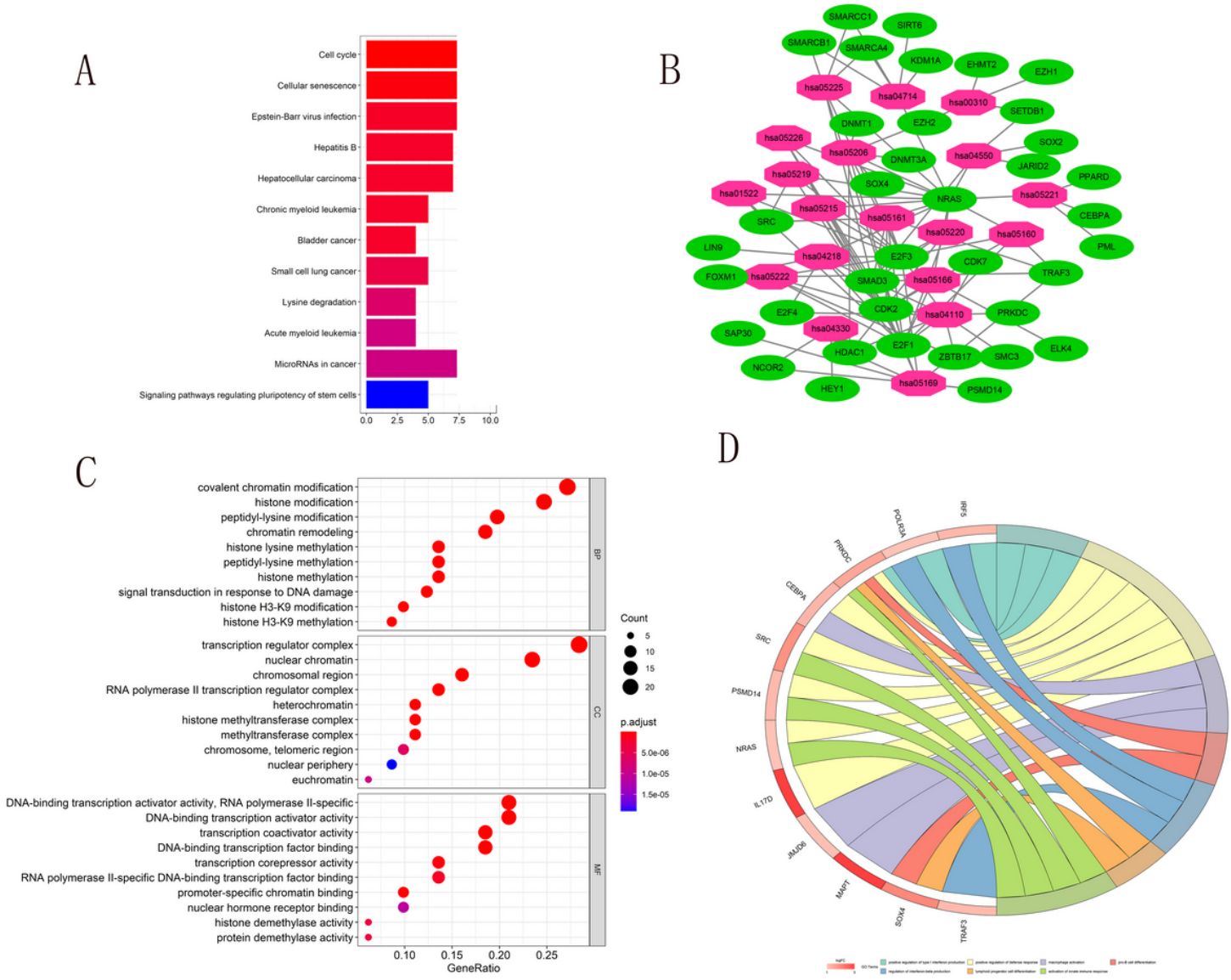

Figure 8

The GO and KEGG enrichment analysis. (A) the KEGG pathway analysis of prognostic IRDEGs and DETFs. (B) network map of screening pathways and genes. hsa04110, Cell cycle; hsa04218, Cellular senescence; hsa05169, Epstein-Barr virus infection; hsa05161, Hepatitis B; hsa05225, Hepatocellular carcinoma; hsa05220, Chronic myeloid leukemia; hsa05219, Bladder cancer; hsa05222, Small cell lung cancer; hsa00310, Lysine degradation; hsa05221, Acute myeloid leukemia; hsa05206, MicroRNAs in cancer; hsa04550, Signaling pathways regulating pluripotency of stem cells; hsa05226, Gastric cancer; hsa05166, Human T-cell leukemia virus 1 infection; hsa05160, Hepatitis C; hsa05215, Prostate cancer; hsa01522, Endocrine resistance; hsa04714, Thermogenesis; hsa04330, Notch signaling pathway; (C) The GO analysis of prognostic IRDEGs and DETFs. (D) immune-related biological processes of selected genes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.tif
- FigureS1.tif
- FigureS2.tif
- FigureS2.tif
- FigureS3.tif
- FigureS3.tif
- FigureS4.tif
- FigureS4.tif
- SupplementaryMaterials.docx
- SupplementaryMaterials.docx