ORIGINAL ARTICLE

WILEY

Prognostic evaluation and immune infiltration analysis of five bioinformatic selected genes in hepatocellular carcinoma

Enjiang Lai[1,2] | Yang Tai[1,2] | Jingsun Jiang[2] | Chong Zhao[1,2] | Yang Xiao[1,2] | Xin Quan[1,2] | Hao Wu[2] | Jinhang Gao[1,2] (ORCID)

[1]Lab of Gastroenterology and Hepatology, West China Hospital, Sichuan University, Chengdu, China

[2]Department of Gastroenterology, West China Hospital, Sichuan University, Chengdu, China

Correspondence
Jinhang Gao, Lab of Gastroenterology and Hepatology, West China Hospital, Sichuan University, NO. 37, Guo Xue Xiang, Chengdu 610041, China.
Emails: Gao.jinhang@scu.edu.cn; Gao.jinhang@qq.com

Hao Wu, Department of Gastroenterology, West China Hospital, Sichuan University, NO. 37, Guo Xue Xiang, Chengdu 610041, China.
Email: hxxhwh@163.com

Abstract

Despite the development in hepatocellular carcinoma (HCC) treatment in recent years, the therapeutic outcome of HCC remains unfavourable. This study examines the prognosis of HCC from a genetic level using clinical databases and single-cell data to identify genes with a high prognostic value. Three up-regulated genes (*UBE2S*, *PTTG1*, and *CDC20*) and two down-regulated genes (*SOCS2* and *DNASE1L3*) in HCC tissues were identified. Various analyses confirmed its correlation with tumour stage ($p < 0.01$) and patient survival time (log-rank $p < 0.001$). Immune analysis, single-cell analysis, and gene set enrichment analysis (GSEA) were employed to provide insight on how they affect cancer progression, and we observed a close relation between these genes and tumour immune infiltration. Eventually, we constructed a risk score system that risk score = $(0.0465) \times UBE2S + (0.1851) \times CDC20 + (−0.0461) \times DNASE1L3 + (−0.2279) \times SOCS2$ (5-year area under curve = 0.706). The risk score system may serve as an effective novel prognostic system for HCC patients. This study might provide novel ideas for prognostic or therapeutic biomarkers for HCC.

KEYWORDS
bioinformatics, cell proliferation, hepatocellular carcinoma, immune infiltration, prognosis

1 | INTRODUCTION

Primary liver cancer is the second leading cause of cancer-related death with an increasing global incidence.[1] Hepatocellular carcinoma (HCC) is the most common type of primary hepatic malignancy that accounts for approximately 70%–80% of all primary liver cancers.[1,2] Although early HCC can be treated by liver resection or transplantation, such options may not be available for cases with an advanced stage at the time of diagnosis.[3–5] Even after curative-intent surgical resection, the 5-year survival rate of HCC patients remains poor due to the high recurrence

---

Lai and Tai contributed equally to this study.

rate.[6] Therefore, accurate estimation of the prognosis is crucial for clinical decision-making and personalized treatment. Traditional prognostic prediction for HCC mainly relies on pathological grade and tumour node metastasis (TNM) stage, which is insufficient to predict the outcome of patients. Thus, it is urgent to explore more accurate biomarkers for the early prediction and prognosis evaluation of HCC.

Recently, the development of high-throughput genetic technology has revolutionized the landscape of oncology research, allowing us to study tumour biology at the molecular level.[7] Genetic research has led to substantial advancements in the diagnosis and treatment of various cancers, such as breast cancer, prostate cancer, and colon cancer.[8-11] Identification of key genes in tumours not only reveals the mechanism of tumorigenesis and cancer progression, but also provides therapeutic and prognostic targets for precision and personalized medicine. Previous studies of HCC genetics have mainly focused on the prognostic value of known oncogenic or/and tumour-suppressor genes.[12-14] Some studies have approached this topic by analysing the differentially expressed genes (DEGs) in HCC and normal tissues, but few of them combined the findings with single-cell analysis of the genes to explain the mechanism behind their effect on tumour progression. Increasing evidence indicates that the tumour immune microenvironment plays a pivotal role in the development and prognosis of cancers.[15] However, the genetics behind this process in HCC remains to be fully discovered.

In this study, we explored the prognosis-related DEGs in HCC based on gene expression profiles from multiple databases and analysed their correlation with clinicopathological characteristics, immune infiltration, and patient survival. Eventually, we constructed a prognostic model using these genes to predict clinical outcomes in patients with HCC.

2 | MATERIALS AND METHODS

2.1 | Data acquisition and processing

Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) is an international repository that archives microarray, next-generation sequencing, and other forms of high-throughput functional genomics data. The transcriptome profiles and clinical information of HCC patients were obtained from GEO databases, including 115 cases from GSE76427, 24 cases from GSE101685, and 183 cases from GSE112790.[16-18] Differential expression analysis was performed using the "DESeq2" R package by the standard of adjusted p value < 0.01 and |log2 (fold change)| >1. A Venn diagram was used to generate the overlapped DEGs.

2.2 | Gene Expression Profiling Interactive Analysis (GEPIA)

Gene Expression Profiling Interactive Analysis (http://gepia.cancer-pku.cn/index.html) is a web server for cancer and normal gene expression profiling and interactive analyses based on The Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) data integration.[19] In this study, survival-related genes obtained from GEPIA were used to identify the target genes that overlapped with DEGs from GEO databases. Additionally, gene expression profiles according to cancer types or pathological stages and survival analysis were also performed by GEPIA or GEPIA2 (http://gepia2.cancer-pku.cn/#index). A Sankey diagram was constructed to integrate gene expression, clinicopathological characteristics, and prognosis using the "ggalluval" R package.

2.3 | Survival analysis by Kaplan–Meier Plotter

The prognostic value of the target genes was further validated by an open-access bioinformatic tool Kaplan–Meier Plotter (http://kmplot.com/analysis/),[20] in which 364 HCC cases were classified into the high- or low-expression group according to various quantile expressions of the proposed biomarker. Then, they were compared by a Kaplan–Meier survival plot, and the hazard ratio with 95% confidence intervals and log-rank p value were calculated. A p value < 0.05 was considered statistically significant.

2.4 | Gene interaction analysis by STRING and GeneMANIA

The Spearman's correlation analysis between the target genes was plotted as a heatmap using the "pheatmap" R package. Protein–protein interaction (PPI) analysis of them was further performed using STRING (http://string-db.org/)[21] and GeneMANIA (http://genemania.org/)[22] online tools that predict functional interaction networks based on multiple databases.

2.5 | Gene co-expression and pathway enrichment analysis by LinkedOmics

LinkedOmics (http://www.linkedomics.org/) is a publicly available portal that provides a visual platform for biologists and clinicians to access, analyse, and compare multiomics data from all 32 TCGA cancer types.[23] Gene co-expression analysis with the target genes in HCC was performed using Pearson's correlation coefficient, presenting in scatter plots and heatmaps. Reactome pathway enrichment of the co-expressed genes was then generated from the LinkedOmics database. Additionally, Gene Set Enrichment Analysis (GSEA) was performed in GenePattern using curated gene sets from the Reactome database.

2.6 | Mutation analysis by cBioPortal

The cBioPortal for Cancer Genomics (http://www.cbioportal.org/) is an open-access, open-source resource for interactive exploration of

multidimensional cancer genomics datasets.[24] It integrates data from 126 tumour genome studies, including large tumour research projects such as TCGA and International Cancer Genome Consortium (ICGC). A color-coded map of genetic alterations in the target genes of HCC patients was constructed using OncoPrinter through cBio-Portal TCGA datasets.

2.7 | Immune infiltration analysis by TISIDB and TIMER

TISIDB (http://cis.hku.hk/TISIDB) is a web portal for tumour and immune system interaction, which integrates multiple types of data resources in oncoimmunology.[25] TIMER (http://timer.cistrome.org/) is a comprehensive resource for immune infiltration analysis across diverse cancer types.[26] In this study, the correlation of gene expression with immune features (immunomodulators,

chemokines, and chemokine receptors) and immune cell infiltration levels was evaluated with TISIDB and TIMER, respectively. Moreover, partial Spearman's correlation analysis with the quanTIseq method was also performed for each immune cell subtype to reveal the relationship between infiltrates estimation value and gene expression in HCC samples.

2.8 | Single-cell RNA-sequencing analysis

Human Liver Browser (http://itzkovitzwebapps.weizmann.ac.il/webapps/home/session.html?app=HumanLiverBrowser) and Single-cell Atlas in Liver Cancer (scAtlasLC, https://scatlaslc.ccr.cancer.gov/) are public databases of single-cell transcriptomic profiles for HCC.[27,28] The expression of the target genes in malignant and non-malignant cells in HCC was analysed by Human Liver Browser and scAtlasLC.
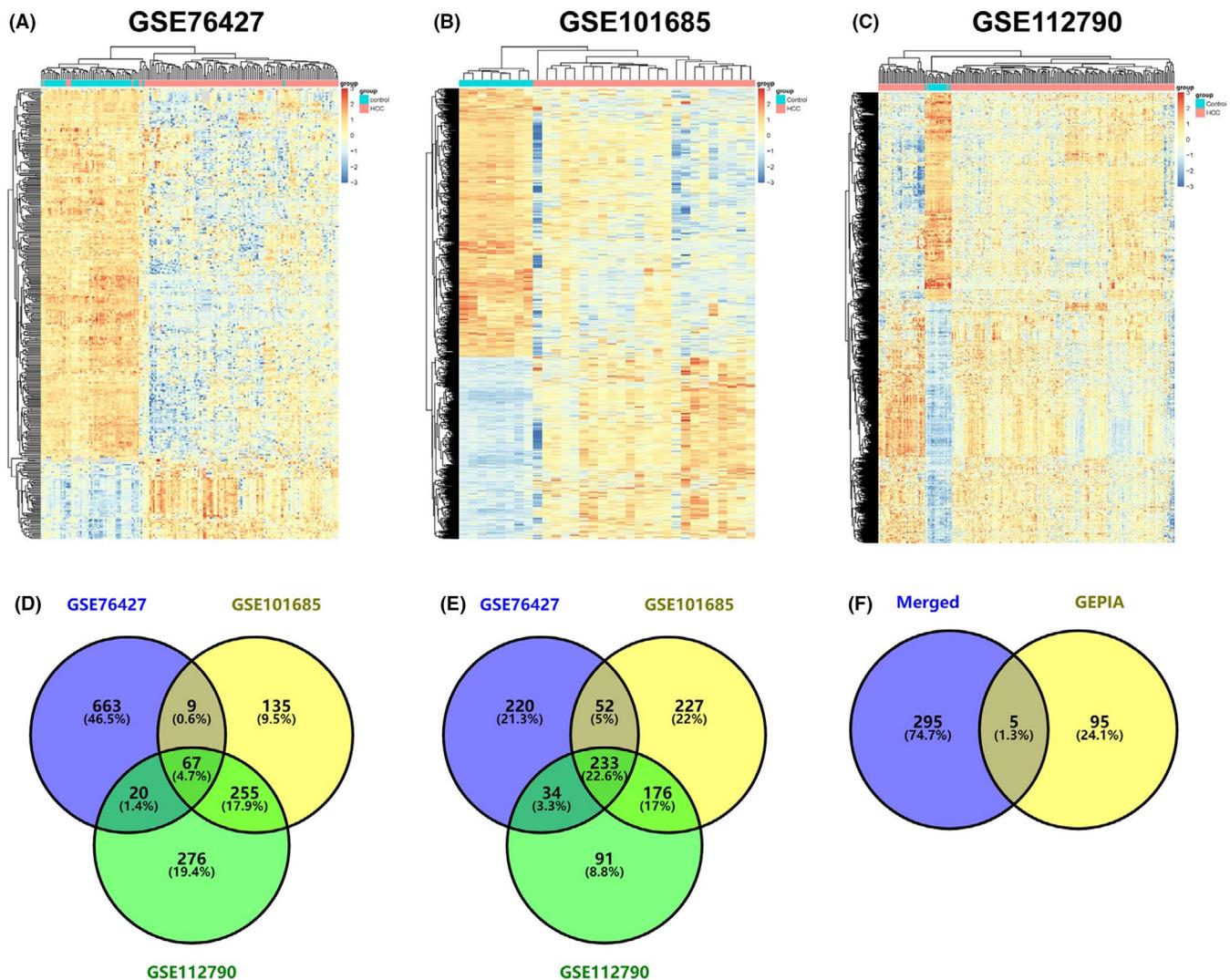
FIGURE 1 Identification of prognosis-related DEGs in HCC. (A–C) Heatmaps of the DEGs associated with HCC in GSE76427 (A), GSE101685 (B), and GSE112790 (C). (D–E) The overlapped genes of the up-regulated DEGs (D) and down-regulated DEGs (E) from the GEO cohorts. (F) The five identified prognosis-related DEGs from the GEO and GEPIA databases

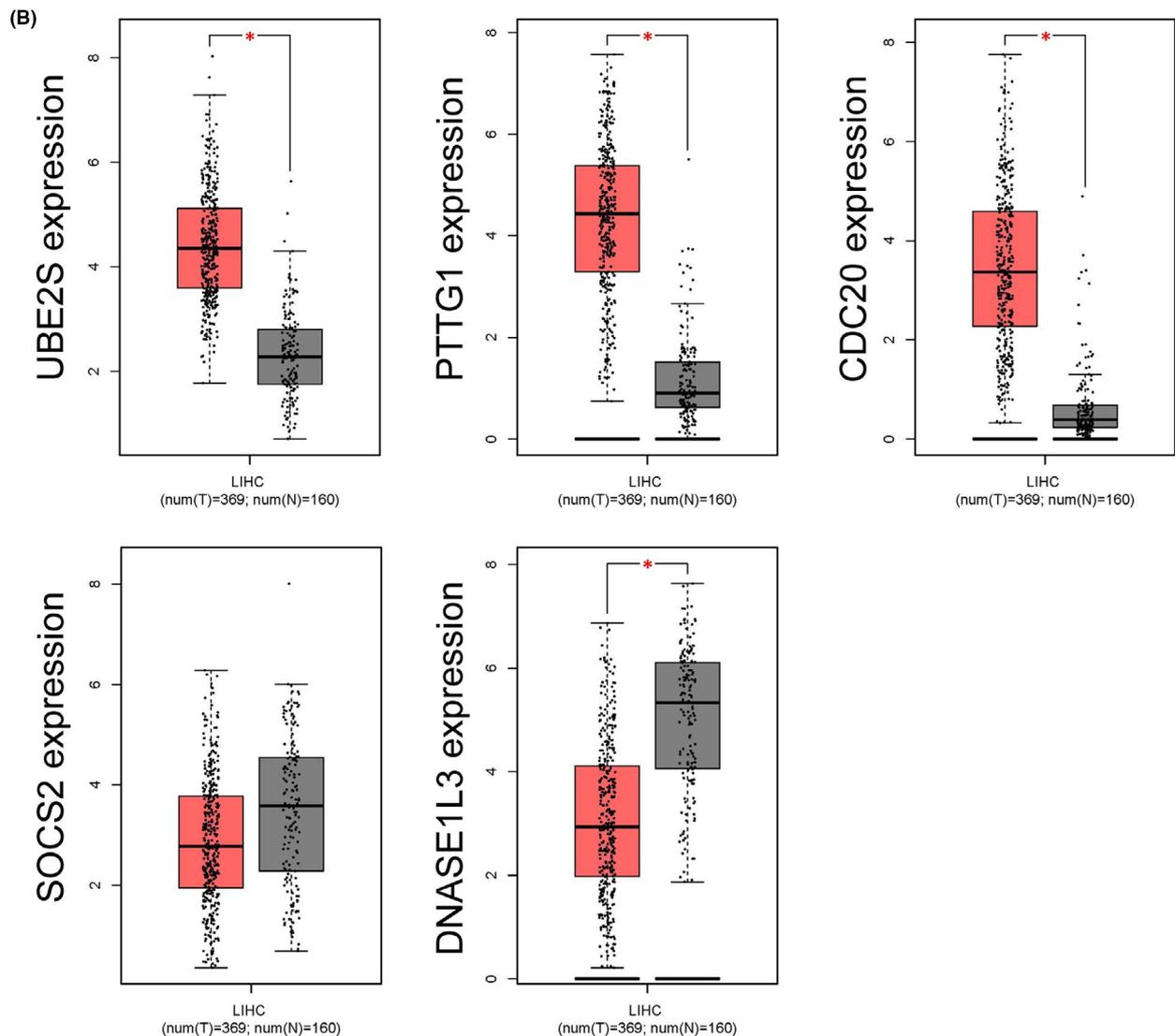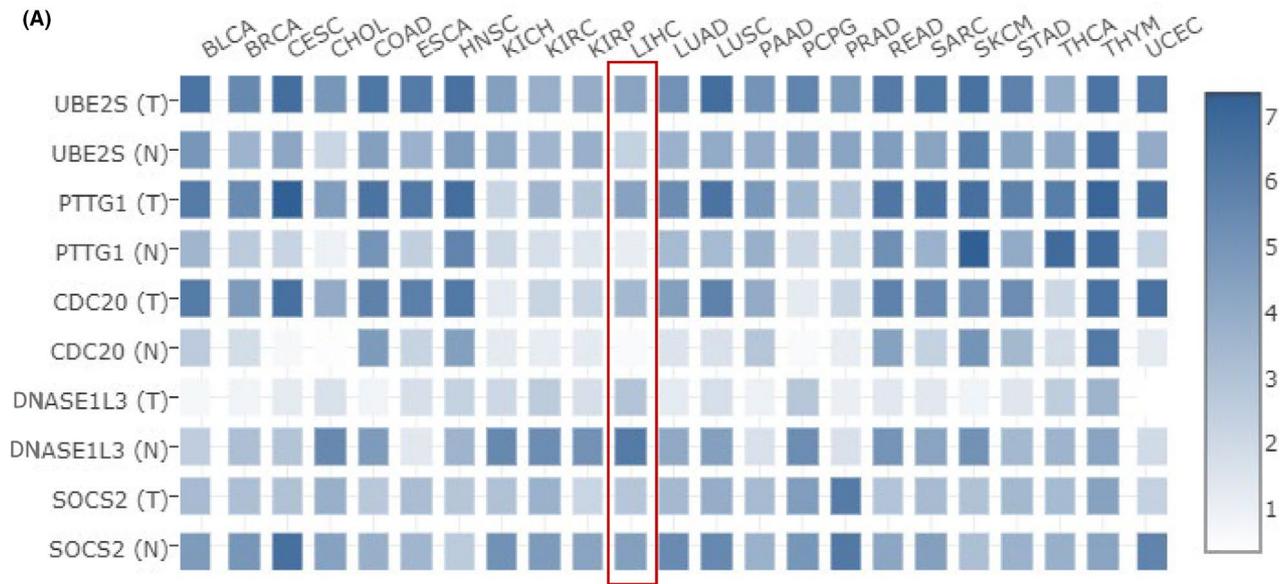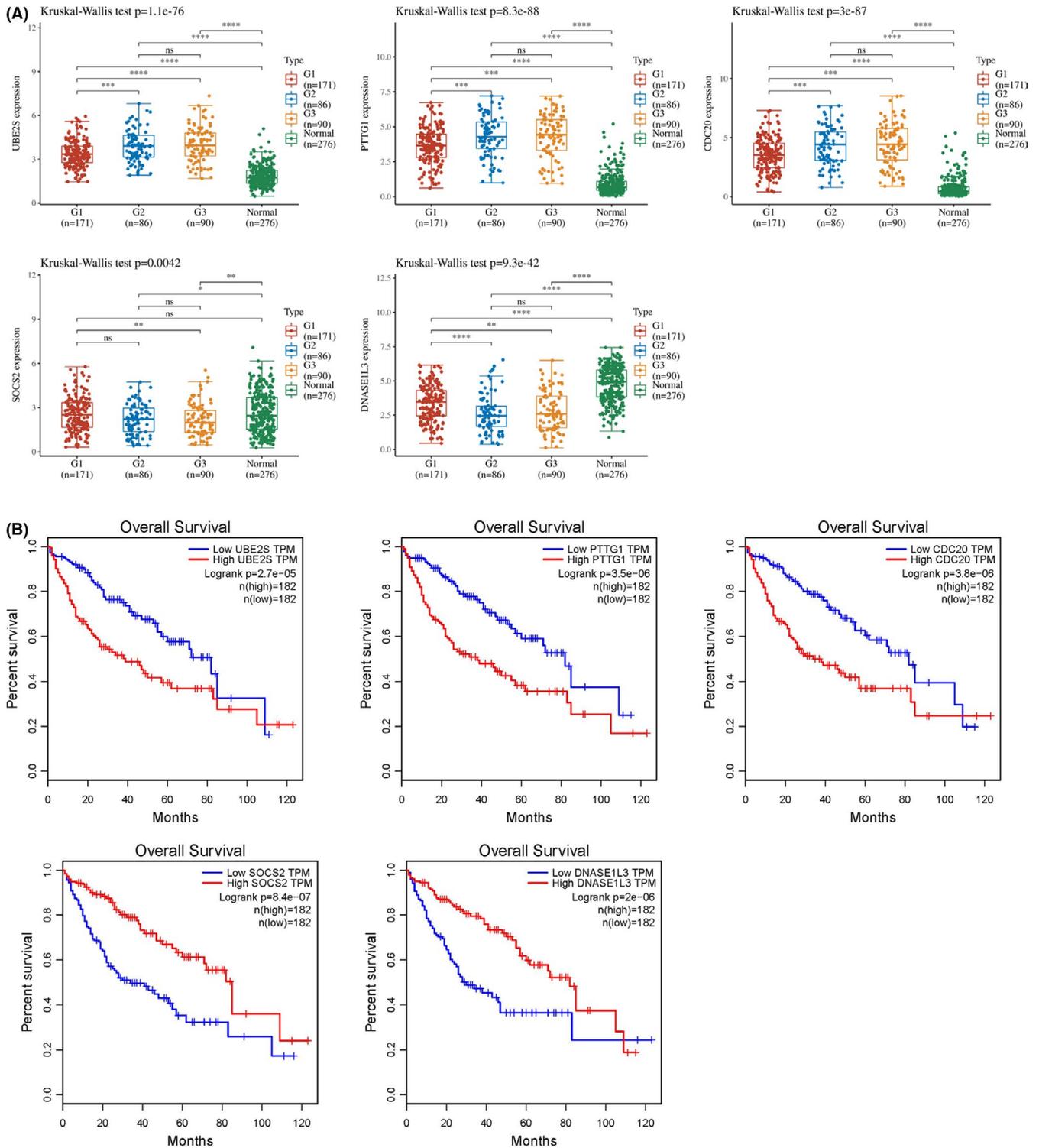**(A)**

(B)

FIGURE 2 Transcriptional expression of the target genes in HCC. (A) Transcription levels of the five target genes in human cancers (GEPIA2). (B) Transcription levels of the five target genes in HCC (GEPIA). *adj. $p < 0.05$

(A)

(B)

FIGURE 3 Correlation of the target genes with pathological stage and patient survival in HCC. (A) Correlation between the target genes expression and pathological stage in HCC (TCGA). (B) Prognostic value of the target genes in HCC (GEPIA). [****]$p < 0.0001$, [***]$p < 0.001$, [**]$p < 0.01$, [*]$p < 0.05$, ns, not significant

2.9 | Construction of prognostic signature and internal validation

Univariate and multivariate Cox regression analyses were performed to identify the target genes related to prognosis in the TCGA cohort. Then, the least absolute shrinkage and selection operator (LASSO) regression model with tenfold cross-validation was performed to identify the most significant survival-related genes. Stepwise multivariate Cox regression analysis was applied to further establish the prognostic signature in HCC. The risk
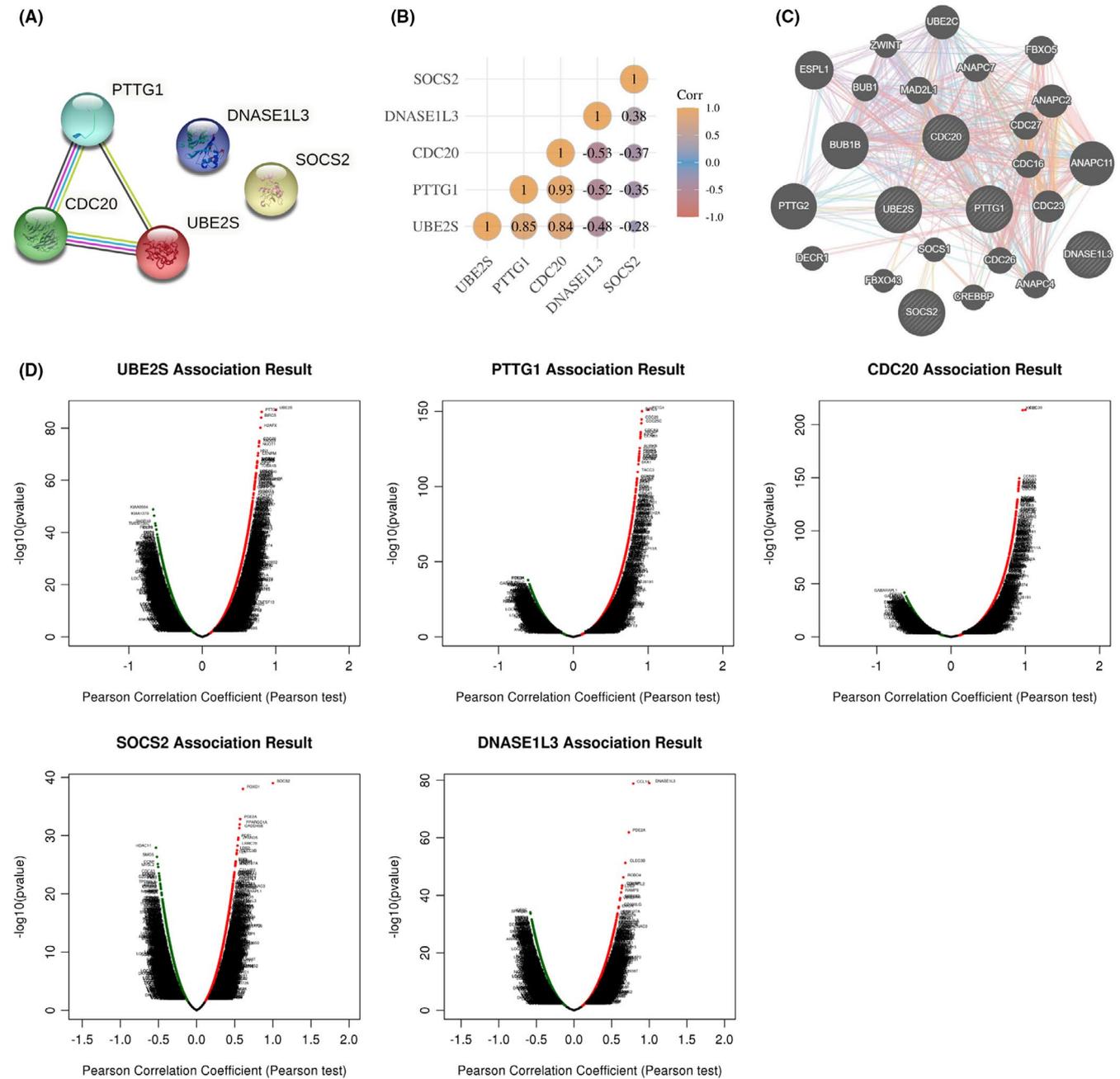
FIGURE 4 Gene interaction and co-expression of the target genes in HCC. (A) Protein–protein interaction network of the five target genes (STRING). (B) Heatmap from the gene-to-gene Spearman's correlation for the five target genes (TCGA). (C) Gene interaction network of the five target genes and their related genes (GeneMANIA). (D) The related genes co-expressed with the five target genes (LinkedOmics)

score was calculated by the following formula: risk score = expression of gene 1 × coefficient 1 + expression of gene 2 × coefficient 2 + ... expression of gene n × coefficient n. To validate the predictive ability, all HCC patients were allocated into high- or low-risk groups according to the median value of the risk score. Kaplan–Meier curve analysis and log-rank test were performed to compare the overall survival difference between the two groups using the "Survival" R package. In addition, the receiver operating characteristic (ROC) model was also utilized to evaluate the predictive power of this prognostic signature.

3 | RESULTS

3.1 | Identification of prognosis-related DEGs in HCC

To identify the candidate genes related to HCC prognosis, the GEO and GEPIA databases were used to screen for DEGs associated with HCC (Figure 1). After taking intersections from different GEO cohorts (GSE76427, GSE101685, and GSE112790, Figure 1A–C), a total of 300 DEGs in HCC samples were identified, with 67 genes up-regulated (Figure 1D) and 233 genes down-regulated (Figure 1E). Then, the top
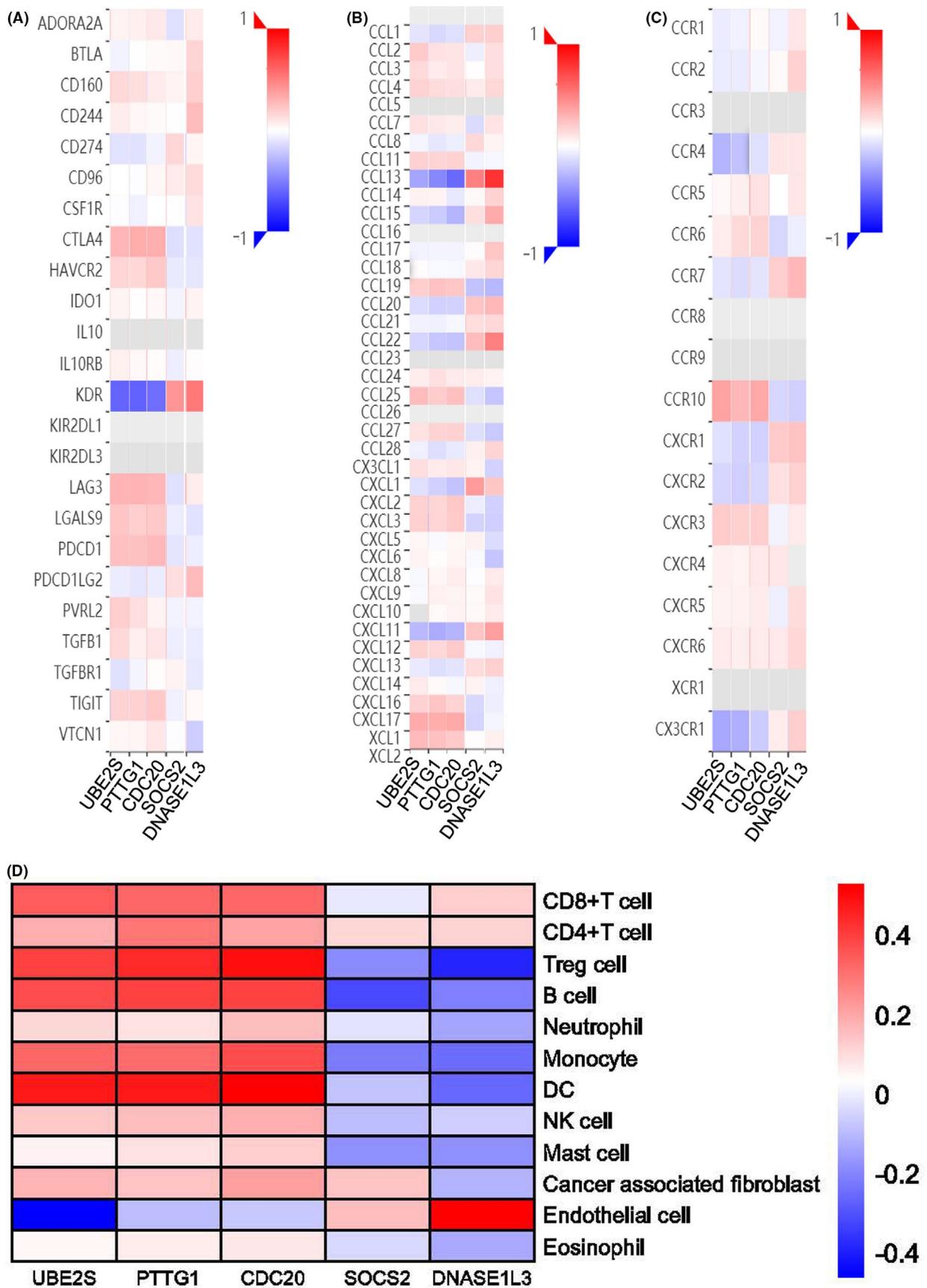
FIGURE 5 Immune infiltration analysis of the target genes in HCC. (A–C) Heatmaps of the correlation between the five target genes and immunomodulators (A), chemokines (B), and chemokine receptors (C) in HCC (TISIDM). (D) Heatmaps of the correlation between the five target genes and immune cells in HCC (TIMER)

100 most significant survival genes were generated from the GEPIA database (Figure 1F). Finally, the overlapped genes with DEGs from GEO, including ubiquitin-conjugating enzyme E2S (*UBE2S*), pituitary tumour-transforming gene 1 (*PTTG1*), cell division cycle 20 (*CDC20*), suppressor of cytokine signalling 2 (*SOCS2*), and deoxyribonuclease 1 like 3 (*DNASE1L3*), were considered as prognosis-related DEGs (target genes) and were subjected to subsequent analyses.

3.2 | Transcriptional expression of the target genes in HCC

The transcriptional expression levels of the target genes in human cancers have been determined using the GEPIA database (Figure 2). As shown in Figure 2A, the expression levels of *UBE2S*, *PTTG1*, and *CDC20* were significantly higher in most cancer tissues than in normal tissues. The increased expression of *UBE2S*, *PTTG1*, and *CDC20* in HCC was also observed compared with that in normal liver ($p < 0.05$, Figure 2A,B). Conversely, *SOCS2* and *DNASE1L3* were down-regulated in most cancer tissues compared to normal tissues (Figure 2A). The decreased expression of *DNASE1L3* was verified in HCC ($p < 0.05$, Figure 2A,B), while *SOCS2* was not significantly reduced in HCC tissues compared to normal livers ($p > 0.05$, Figure 2B).

3.3 | Correlation of the target genes with clinicopathological characteristics and patient survival in HCC

The TCGA database was used to evaluate the relationship between the target genes and the pathological stage of HCC patients. Kruskal–Wallis test showed that the expression levels of the five target genes (*UBE2S*, *PTTG1*, *CDC2*, *SOCS2*, and *DNASE1L3*) were significantly correlated with the pathological stage of HCC ($p < 0.05$, Figure 3A). Next, we used the GEPIA database and Kaplan–Meier plotter to further determine the prognostic values of the target genes in HCC patients. The survival curves revealed that higher levels of *UBE2S*, *PTTG1*, and *CDC20* expression predicted a poor prognosis, while higher expression of *SOCS2* and *DNASE1L3* predicted a better prognosis ($p < 0.05$, Figures 3B and S1). Additionally, Sankey diagrams using TCGA data were generated to better visualize the correlation between gene expression, clinicopathological characteristics, and prognosis in patients with HCC (Figure S2). It is clearly observed in the chart that patients with high levels of *UBE2S*, *PTTG1*, and *CDC20* were more likely to have a higher pathological stage and worse prognosis, while those with high levels of *SOCS2* and *DNASE1L3* were more likely to have the opposite tendency.

3.4 | Gene interaction, co-expression, and pathway enrichment of the target genes in HCC

The correlation and interaction of the target genes were evaluated using the TCGA, STRING, and GeneMANIA databases (Figure 4A–C). When the target genes were mapped into the STRING database for PPI network analysis, the interactions between *UBE2S*, *PTTG1*, and *CDC20* were observed, while *SOCS2* and *DNASE1L3* did not interact with others (Figure 4A). A heatmap from the gene-to-gene correlation for the five target genes was then plotted according to Spearman's correlation analysis (Figure 4B). Furthermore, GeneMANIA network analysis revealed that *UBE2S*, *PTTG1*, *CDC20*, and *SOCS2* could physically interact and co-express with each other (Figure 4C).

Next, LinkedOmics was employed to identify the related genes co-expressed with the target genes and their biological functions. As shown in Figure 4D, the expression of *UBE2S*, *PTTG1*, and *CDC20* was positively correlated with each other. All of the five target genes have close associations with genes regulating cell cycle (*CDC23*, *ANAPC11*, and *ANAPC4*) and cell mitosis (*BUB1B*, *FBXO5*, *MAD2L1*, and *ESPL1*), which was consistent with pathway enrichment results (Figures S3–S7). Moreover, immune-related pathways, such as complement cascade and its regulation, were also significantly enriched in the associated genes (Figures S3–S7).

The genetic alterations of the target genes in HCC patients were determined using the cBioPortal online tool. The results indicated that the target genes were low-frequency mutated genes with the altered rate varying from 0.4% to 1.1% in the queried HCC samples (Figure S8).

3.5 | Role of the target genes in HCC immune infiltration

As the tumour immune microenvironment plays a pivotal role in the tumorigenesis and progression of cancers, the TISIDB and TIMER databases were used to explore the impact of the target genes on immune features and immune cell infiltration in HCC. Firstly, the expression of the target genes was significantly correlated with immunomodulators (Figure 5A), chemokines (Figure 5B), and chemokine receptors (Figure 5C). Next, *UBE2S*, *PTTG1*, and *CDC20* were positively correlated with immune cells in HCC, while *SOCS2* and *DNASE1L3* were negatively correlated with immune cells in HCC (Figure 5D). In addition, the most positive correlation of *UBE2S*, *PTTG1*, and *CDC20*, and the most negative correlation of *SOCS2* and *DNASE1L3* were observed in regulatory T cells (Treg cells), B cells, monocytes, and dendritic cells (DCs). To further

FIGURE 6 Single-cell analysis of the target genes in HCC. (A, B) t-SNE analysis (A) and expression levels (B) of *UBE2S*, *CDC20*, and *PTTG1* in liver cells (Human Liver Browser). (C,D) t-SNE analysis (C) and expression levels (D) of *SOCS2* and *DNASE1L3* in liver cells (Human Liver Browser). CAFs, cancer-associated fibroblasts; cDC, conventional dendritic cell; LSEC, liver sinusoidal endothelial cell; LVEC, lymphatic vascular endothelial cell; LVECm, malignant lymphatic vascular endothelial cell; SAMs, scar-associated macrophages; TM, tissue monocytes; vSMC, vascular smooth muscle cell

(A)

(B)

(C)

(D)
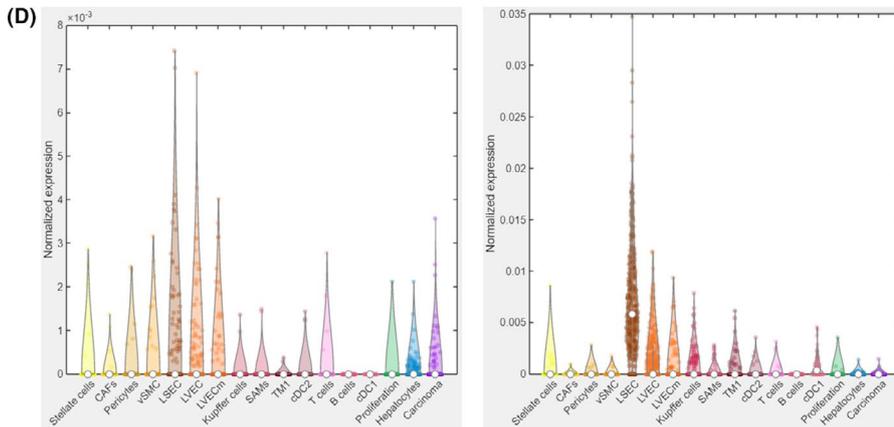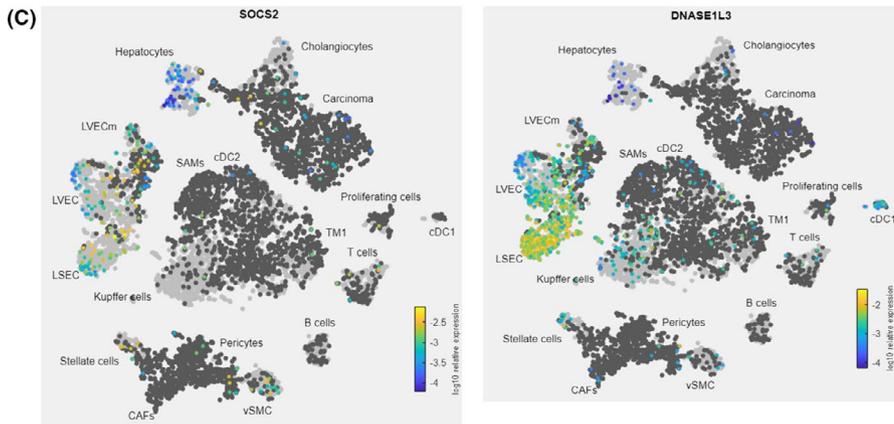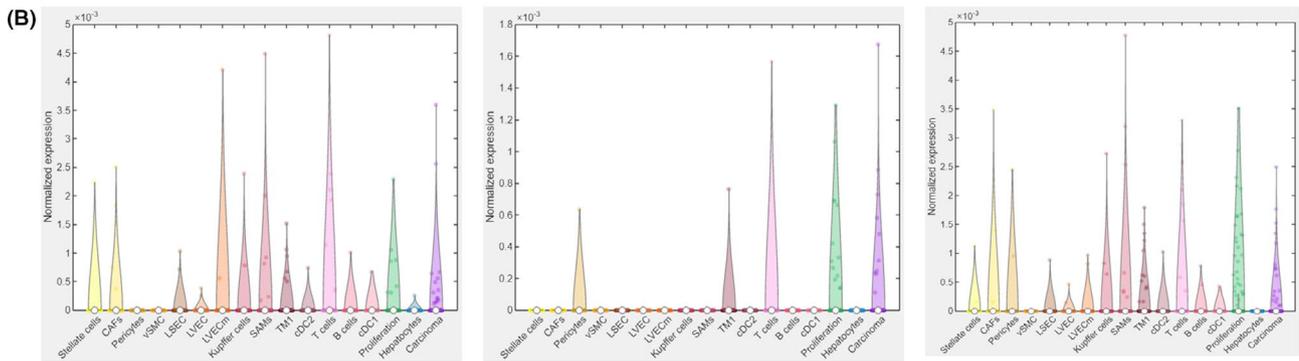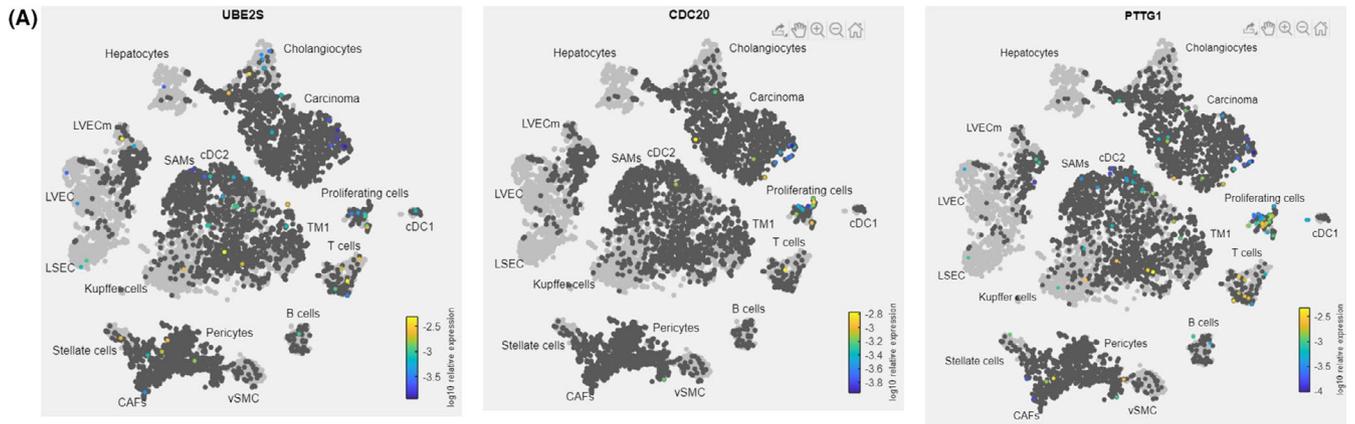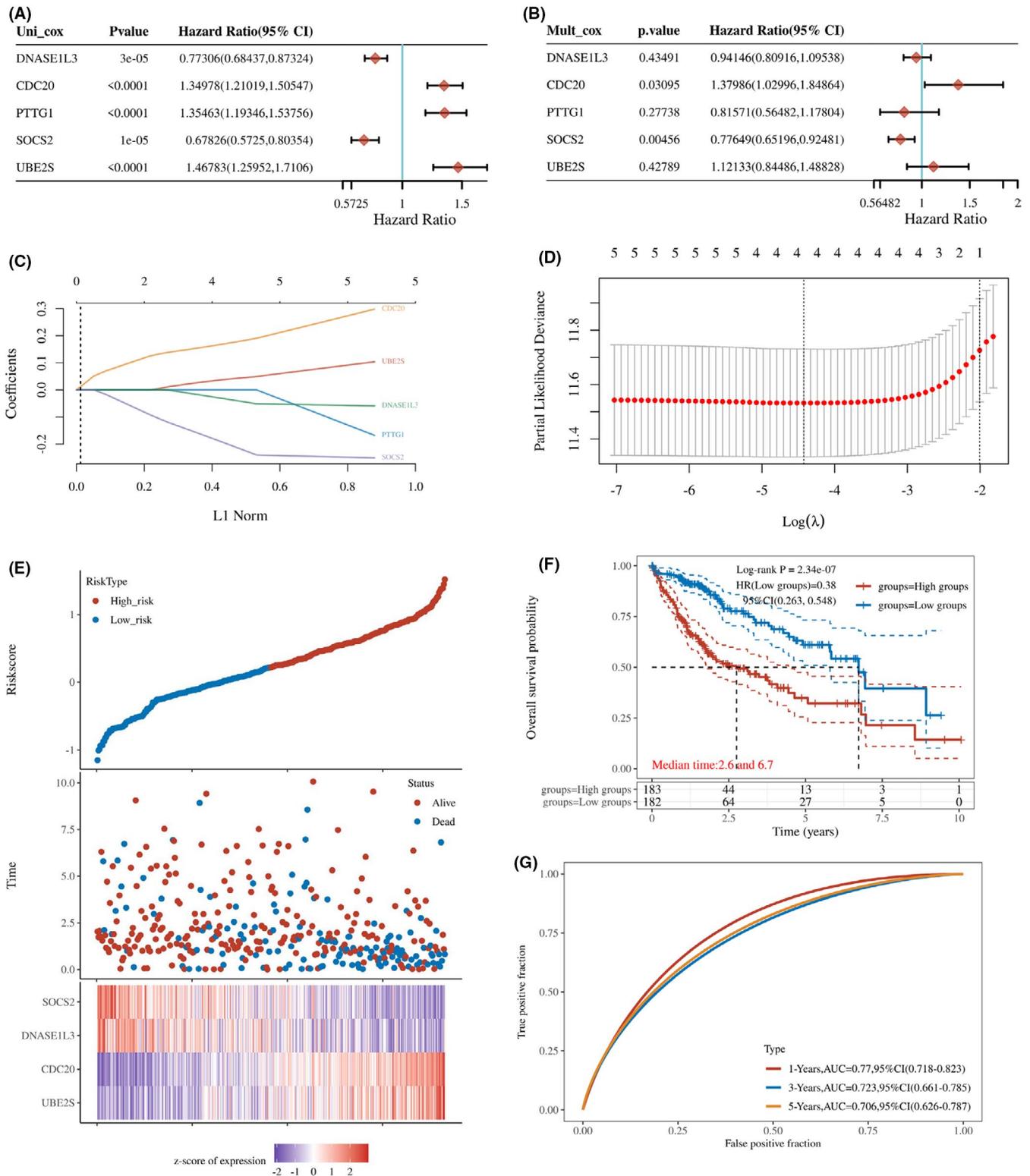
FIGURE 7 Construction of target genes-based prognostic signature and internal validation in HCC. (A, B) Univariate (A) and multivariate (B) Cox regression analyses of the five target genes and clinical prognosis in HCC (TCGA). (C) Selection of the optimal parameter (lambda) in the least absolute shrinkage and selection operator (LASSO) model. (D) LASSO coefficient profiles of the target genes with nonzero coefficients determined by the optimal lambda. (E) Distribution of risk score, survival time, and heatmap of four prognostic genes expression in the TCGA cohort. (F) The Kaplan–Meier survival analysis for HCC patients at the high- or low-risk group in TCGA. (G) ROC curves for predicting 1-, 3-, 5-year overall survival in TCGA

explore the relationship between gene expression and individual immune cells, partial Spearman's correlation analysis was then performed using the quanTIseq method (Figure S9). The results showed a significant positive correlation between *UBE2S*, *PTTG1*, and *CDC20* expression and infiltration levels of CD4$^+$ T cells, CD8$^+$ T cells, Treg cells, B cells, monocytes, and natural killer (NK) cells. In comparison, *SOCS2* showed a significant negative correlation with infiltration levels of DCs and NK cells. Additionally, *DNASE1L3* was found to have a significant negative correlation with infiltration levels of neutrophils, monocytes, and NK cells. These findings indicated a close relationship between the target genes and immune infiltration in HCC.

3.6 | Single-cell analysis of the target genes in HCC

To further explore the expression of the target genes in specific liver cells within HCC, we ran a combined t-distributed stochastic neighbour embedding (t-SNE) analysis from the Human Liver Browser (Figure 6) and scAtlasLC (Figure S10) datasets. It is revealed that *UBE2S* was mainly expressed in T cells, scar-associated macrophages (SAMs), malignant lymphatic vascular endothelial (LVECm) cells, and carcinoma cells (Figure 6A,B). While *CDC20* was highly expressed in carcinoma cells, proliferation cells, T cells, tissue monocytes 1(TM1), and pericytes (Figure 6A,B). The expression of *PTTG1* was more evenly distributed in immune cells, while carcinoma cells showed moderately high levels of *PTTG1* (Figure 6A,B). *SOCS2* was mainly expressed in liver sinusoidal endothelial cells (LSECs) and lymphatic vascular endothelial cells (LVECs), while moderately expressed in carcinoma cells (Figure 6C,D). The expression level of *DNASE1L3* was exceptionally high in LSECs, while it was low in carcinoma cells (Figure 6C,D). The above results indicated that the target genes were differentially expressed in various immune cell types.

3.7 | Construction of target genes-based prognostic signature and internal validation in HCC

Furthermore, the univariate and multivariate Cox regression analyses were conducted to evaluate the target gene as an independent prognostic factor in the TCGA cohort. The univariate Cox analysis demonstrated that all target genes were significantly correlated with clinical prognosis in HCC patients. Among them, *CDC20*, *PTTG1*, and *UBE2S* were high-risk factors (hazard ratio > 1), and *DNASE1L3* and *SOCS2* were protective factors (hazard ratio < 1) (Figure 7A). However, multivariate Cox regression analysis showed that only *CDC20* and *SOCS2* were independent predictors for HCC prognosis (Figure 7B).

Next, LASSO regression analysis with tenfold cross-validation was conducted to select the most predictive genes as prognostic indicators. The coefficients for corresponding genes were generated according to the partial likelihood deviance and determined with its lowest value at a log λ = −4.4 (Figure 7C,D). Eventually, four genes (*UBE2S*, *CDC20*, *DNASE1L3*, and *SOCS2*) were enrolled to construct the prognostic signature using the formula: risk score = (0.0465) × *UBE2S* + (0.1851) × *CDC20* + (−0.0461) × *DNASE1L3* + (−0.2279) × *SOCS2*. The patients were further assigned to the high- or low-risk groups using the median risk score as the cut-off point (Figure 7E). The Kaplan–Meier survival curves revealed a significant difference in overall survival between groups. The high-risk patients showed a worse prognosis compared with the low-risk patients (Figure 7F). Moreover, ROC curve analysis demonstrated the predictive ability of the risk score for 1-, 3- and 5-year overall survival, with areas under the curve (AUCs) of 0.77, 0.723, and 0.706, respectively (Figure 7G).

4 | DISCUSSION

Liver cancer is a highly heterogeneous disease, and the complex mechanism behind it needs more thorough understanding. To date, the prognostic tools used to assess HCC patient risk remain undesirable.[29] Advancements in genetic research have allowed more insights into the mechanism behind this malignant disease and may provide more advanced and accurate ways to evaluate the prognosis of HCC patients. Genetic biomarkers have been identified for cancer detection, risk assessment, and prognosis prediction in multiple types of cancer, including brain cancer, colorectal cancer, and prostate cancer.[30–32] Genetic tools can also help in cancer prevention and treatment by providing precision therapeutic targets, which have been proven to be effective in breast cancer treatment.[33,34] In this study, we identified five target genes (*UBE2S*, *PTTG1*, *CDC20*, *SOCS2*, and *DNASE1L3*) closely correlated with the prognosis of HCC patients through the integration of gene expression profiles from multiple databases. Using these prognostic genes, we eventually constructed a prognostic model for predicting the survival of HCC patients.

Among the five prognostic-related genes, expression of *UBE2S*, *PTTG1*, and *CDC20* was up-regulated, whereas *SOCS2* and *DNASE1L3* were down-regulated in HCC tissues. UBE2S belongs to the ubiquitin-conjugating enzyme (E2) family and is critical in cell cycle regulation, cell differentiation, and DNA repair.[5,35] In HCC, UBE2S has been observed to enhance the ubiquitination of p53 for protein degradation in HCC cells.[36] PTTG1 is a securin protein that inhibits sister–chromatid separation, which is associated with tumorigenesis by promoting cancer cell proliferation, migration, and invasion. PTTG1 also plays an important role in HCC growth and metastasis cascade via activating PI3K/AKT signalling pathway and epithelial–mesenchymal transition-related factors.[37] CDC20 activates the anaphase-promoting complex and, in turn, modulates mitotic exit.[38,39] It has been shown that CDC20 is vital in HCC cells' proliferation by mediating PHD3 ubiquitination and HIF-1α activation.[40] SOCS2 is a member of the suppressor of the cytokine signalling pathways. It is a transcriptional repressor in multiple proliferation-related signalling pathways, and its suppression has

been observed in lung, breast, and ovarian cancers.[41–43] SOCS2 inhibits HCC metastases via negatively regulating JAK/STAT signalling pathway in HCC cells.[44] DNASE1L3 is a member of the deoxyribonuclease I family, which encodes proteins that could digest DNA in chromatin. DNASE1L3 is widely down-regulated in human cancers, and the down-regulation of DNASE1L3 is associated with poor prognosis in various types of cancers.[45,46]

The present study examined the expression of target genes in different cell types through single-cell analysis. We found that the aberrant expression of the genes was mainly present in immune cells, which hinted that the target genes might have affected HCC progression by influencing immune cells. Further analysis revealed that the expression level of the genes has a significant correlation with the infiltration levels of multiple types of immune cells, primarily Treg cells, B cells, monocytes, and DCs, indicating that the target genes might have promoted immune cell infiltration and in turn contributing to cancer proliferation and progression, leading to a worse prognosis.

Notably, the prognostic prediction models used to evaluate HCC patient risk in clinical practice remain undesirable.[29] In this study, we constructed a risk score system to predict the prognosis of HCC patients using the LASSO regression model. This system included four target genes as prognostic parameters (*UBE2S*, *CDC20*, *SOCS2*, and *DNASE1L3*). *UBE2S* and *CDC20* were positively related factors involving in cell mitosis and cell cycle checkpoint pathways. In contrast, *SOCS2* and *DNASE1L3* were negatively related factors, which were associated with cell cycle regulation. Previous studies have pointed out that cell cycle alterations and mitosis signalling pathways are closely associated with cancer progression and affect cancer immune infiltration.[47–49] Recent studies have also suggested that complement cascade may be linked with tumour-promoting inflammation and cancer immune infiltration.[50] Thus, the target genes may contribute to the tumorigenesis and progression of HCC through promoting tumour cell proliferation and immune infiltration. Although our findings showed promising results, additional studies are needed to define the underlying molecular mechanisms.

In conclusion, we constructed a promising gene prognostic signature based on multiple databases for predicting clinical outcomes in patients with HCC. This individualized risk score signature could effectively conduct risk stratification, survival prediction, and immune microenvironment evaluation for HCC patients, which would be conducive to clinical decision-making and personalized treatment.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

All authors declare no conflict of interest.

AUTHOR CONTRIBUTION

Enjiang Lai: Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Writing-original draft (equal); Writing-review & editing (equal). **Yang Tai:** Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (equal); Writing-original draft (equal); Writing-review & editing (equal). **Jingsun Jiang:** Data curation (equal); Project administration (equal); Resources (equal). **Chong Zhao:** Data curation (equal); Formal analysis (equal); Investigation (equal). **Yang Xiao:** Data curation (equal); Investigation (equal). **Xin Quan:** Investigation (equal); Methodology (equal); Validation (equal). **Hao Wu:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal). **Jinhang GAO:** Conceptualization (lead); Funding acquisition (lead); Supervision (lead); Writing-review & editing (lead).

DATA AVAILABILITY STATEMENT

All data are present in the manuscript. All data are available from the corresponding author Jinhang Gao (Gao.jinhang@scu.edu.cn or Gao.jinhang@qq.com) under reasonable request.

ORCID

Jinhang Gao https://orcid.org/0000-0003-0480-7867

REFERENCES

1. McGlynn KA, Petrick JL, London WT. Global epidemiology of hepatocellular carcinoma: an emphasis on demographic and regional variability. *Clin Liver Dis.* 2015;19(2):223-238. 10.1016/j.cld.2015.01.001.
2. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209-249. 10.3322/caac.21660.
3. Couri T, Pillai A. Goals and targets for personalized therapy for HCC. *Hep Int.* 2019;13(2):125-137. 10.1007/s12072-018-9919-1.
4. Kanwal F, Singal AG. Surveillance for hepatocellular carcinoma: current best practice and future direction. *Gastroenterology.* 2019;157(1):54-64. 10.1053/j.gastro.2019.02.049 2021/07/12.
5. Kirkin V, Dikic I. Ubiquitin networks in cancer. *Curr Opin Genet Dev.* 2011;21(1):21-28. 10.1016/j.gde.2010.10.004.
6. Tandon P, Garcia-Tsao G. Prognostic indicators in hepatocellular carcinoma: a systematic review of 72 studies. *Liver Int.* 2009/04/01 2009;29(4):502-510. 10.1111/j.1478-3231.2008.01957.x.
7. Chibon F. Cancer gene expression signatures - The rise and fall? *Eur J Cancer.* 2013;49(8):2000-2009. 10.1016/j.ejca.2013.02.021 2021/07/23.
8. Bull JH, Ellison G, Patel A, et al. Identification of potential diagnostic markers of prostate cancer and prostatic intraepithelial neoplasia using cDNA microarray. *Br J Cancer.* 2001;84(11):1512-1519. 10.1054/bjoc.2001.1816.
9. Latha NR, Rajan A, Nadhan R, et al. Gene expression signatures: A tool for analysis of breast cancer prognosis and therapy. *Crit Rev Oncol Hematol.* 2020;151:102964. 10.1016/j.critrevonc.2020.102964.
10. Saha S, Bardelli A, Buckhaults P, et al. A phosphatase associated with metastasis of colorectal cancer. *Science.* 2001;294(5545):1343-1346. 10.1126/science.1065817.
11. Weeraratna AT. Discovering causes and cures for cancer from gene expression analysis. *Ageing Res Rev.* 2005;4(4):548-563. 10.1016/j.arr.2005.06.004.

12. Zhu Y, Wang R, Chen W, Chen Q, Zhou J. Construction of a prognosis-predicting model based on autophagy-related genes for hepatocellular carcinoma (HCC) patients. *Aging*. 2020;12(14):14582-14592. 10.18632/aging.103507.

13. Zhang F-P, Huang Y-P, Luo W-X, et al. Construction of a risk score prognosis model based on hepatocellular carcinoma microenvironment. *World J Gastroenterol*. 2020;26(2):134-153. 10.3748/wjg.v26.i2.134.

14. Wu H, Liu T, Qi J, Qin C, Zhu Q. Four autophagy-related lncRNAs predict the prognosis of HCC through coexpression and ceRNA mechanism. *Biomed Res Int*. 2020;2020:3801748. 10.1155/2020/3801748.

15. Picard E, Verschoor CP, Ma GW, Pawelec G. Relationships between immune landscapes, genetic subtypes and responses to immunotherapy in colorectal cancer. *Front Immunol*. 2020;11:369. 10.3389/fimmu.2020.00369.

16. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*. 2013;41(D1):D991-D995. 10.1093/nar/gks1193.

17. Grinchuk OV, Yenamandra SP, Iyer R, et al. Tumor-adjacent tissue co-expression profile analysis reveals pro-oncogenic ribosomal gene signature for prognosis of resectable hepatocellular carcinoma. *Mol Oncol*. 2018;12(1):89-113. 10.1002/1878-0261.12153.

18. Shimada S, Mogushi K, Akiyama Y, et al. Comprehensive molecular and immunological characterization of hepatocellular carcinoma. *EBioMedicine*. 2019;40:457-470. 10.1016/j.ebiom.2018.12.058.

19. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017;45:W98-W102. 10.1093/nar/gkx247.

20. Győrffy B. Survival analysis across the entire transcriptome identifies biomarkers with the highest prognostic power in breast cancer. *Comput Struct Biotechnol J*. 2021;19:4101-4109. 10.1016/j.csbj.2021.07.014.

21. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607-D613. 10.1093/nar/gky1131.

22. Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res*. 2010;38(suppl_2):W214-W220. 10.1093/nar/gkq537.

23. Vasaikar SV, Straub P, Wang J, Zhang B. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Res*. 2018;46(D1):D956-D963. 10.1093/nar/gkx1090.

24. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401. 10.1158/2159-8290.CD-12-0095.

25. Ru B, Wong CN, Tong Y, et al. TISIDB: an integrated repository portal for tumor–immune system interactions. *Bioinformatics*. 2019;35(20):4200-4202. 10.1093/bioinformatics/btz210.

26. Li T, Fu J, Zeng Z, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res*. 2020;48(W1):W509-W514. 10.1093/nar/gkaa407.

27. Ma L, Wang L, Khatib SA, et al. Single-cell atlas of tumor cell evolution in response to therapy in hepatocellular carcinoma and intrahepatic cholangiocarcinoma. *J Hepatol*. 2021. 10.1016/j.jhep.2021.06.028.

28. Massalha H, Bahar Halpern K, Abu-Gazala S, et al. A single cell atlas of the human liver tumor microenvironment. *Mol Syst Biol*. 2020;16(12):e9682. 10.15252/msb.20209682.

29. Tellapuri S, Sutphin PD, Beg MS, Singal AG, Kalva SP. Staging systems of hepatocellular carcinoma: a review. *Indian J Gastroenterol*. 2018;37(6):481-491. 10.1007/s12664-018-0915-0.

30. Verma M, Manne U. Genetic and epigenetic biomarkers in cancer diagnosis and identifying high risk populations. *Crit Rev Oncol Hematol*. 2006/10/01 2006;60(1):9-18. 10.1016/j.critrevonc.2006.04.002.

31. Luo X-J, Zhao Q, Liu J, et al. Novel genetic and epigenetic biomarkers of prognostic and predictive significance in stage II/III colorectal cancer. *Mol Ther*. 2021/02/03 2021;29(2):587-596. 10.1016/j.ymthe.2020.12.017.

32. Kavalci E, Onder AU, Brusgaard K, Bostanci A, Selhanoglu MY, Serakinci N. Identification of genetic biomarkers in urine for early detection of prostate cancer. *Curr Probl Cancer*. 2021/02/01 2021;45(1):100616. 10.1016/j.currproblcancer.2020.100616.

33. Byler S, Goldgar S, Heerboth S, et al. Genetic and epigenetic aspects of breast cancer progression and therapy. *Anticancer Res*. 2014;34(3):1071.

34. Odle TG. Precision medicine in breast cancer. *Radiol Technol*. March 1, 2017 2017;88(4):401M-421M.

35. Garnett MJ, Mansfeld J, Godwin C, et al. UBE2S elongates ubiquitin chains on APC/C substrates to promote mitotic exit. *Nat Cell Biol*. 2009;11(11):1363-1369. 10.1038/ncb1983.

36. Pan Y-H, Yang M, Liu L-P, Wu D-C, Li M-Y, Su S-G. UBE2S enhances the ubiquitination of p53 and exerts oncogenic activities in hepatocellular carcinoma. *Biochem Biophys Res Comm*. 2018;503(2):895-902. 10.1016/j.bbrc.2018.06.093.

37. Huang J-L, Cao S-W, Ou Q-S, et al. The long non-coding RNA PTTG3P promotes cell growth and metastasis via up-regulating PTTG1 and activating PI3K/AKT signaling in hepatocellular carcinoma. *Mol Cancer*. 2018;17(1):93. 10.1186/s12943-018-0841-x.

38. Cheng S, Castillo V, Sliva D. CDC20 associated with cancer metastasis and novel mushroom-derived CDC20 inhibitors with antimetastatic activity. *Int J Oncol*. 2019;54(6):2250-2256. 10.3892/ijo.2019.4791.

39. Zhang Q, Huang H, Liu A, et al. Cell division cycle 20 (CDC20) drives prostate cancer progression via stabilization of β-catenin in cancer stem-like cells. *EBioMedicine*. 2019;42:397-407. 10.1016/j.ebiom.2019.03.032.

40. Shi M, Dai W-Q, Jia R-R, et al. APCCDC20-mediated degradation of PHD3 stabilizes HIF-1a and promotes tumorigenesis in hepatocellular carcinoma. *Cancer Lett*. 2021;496:144-155. 10.1016/j.canlet.2020.10.011.

41. Liu Y, Yang Y, Du J, Lin D, Li F. MiR-3613-3p from carcinoma-associated fibroblasts exosomes promoted breast cancer cell proliferation and metastasis by regulating SOCS2 expression. *IUBMB Life*. 2020;72(8):1705-1714. 10.1002/iub.2292.

42. Sutherland KD, Lindeman GJ, Choong DYH, et al. Differential hypermethylation of SOCS genes in ovarian and breast carcinomas. *Oncogene*. 2004;23(46):7726-7733. 10.1038/sj.onc.1207787.

43. Wikman H, Kettunen E, Seppänen JK, et al. Identification of differentially expressed genes in pulmonary adenocarcinoma by using cDNA array. *Oncogene*. 2002;21(37):5804-5813. 10.1038/sj.onc.1205726.

44. Cui M, Sun J, Hou J, et al. The suppressor of cytokine signaling 2 (SOCS2) inhibits tumor metastasis in hepatocellular carcinoma. *Tumor Biol*. 2016;37(10):13521-13531. 10.1007/s13277-016-5215-7.

45. Deng Z, Xiao M, Du D, et al. DNASE1L3 as a prognostic biomarker associated with immune cell infiltration in cancer. *OncoTargets Ther*. 2021;14:2003-2017. 10.2147/OTT.S294332.

46. Wang S, Ma H, Li X, et al. DNASE1L3 as an indicator of favorable survival in hepatocellular carcinoma patients following resection. *Aging*. 2020;12(2):1171-1185. 10.18632/aging.102675.

47. Knarr M, Avelar RA, Sekhar SC, et al. miR-181a initiates and perpetuates oncogenic transformation through the regulation of innate immune signaling. *Nat Commun*. 2020;11(1):3231. 10.1038/s41467-020-17030-w.

48. Mariathasan S, Turley SJ, Nickles D, et al. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature*. 2018;554(7693):544-548. 10.1038/nature25501.

49. Xu T, Zeng Y, Shi L, et al. Targeting NEK2 impairs oncogenesis and radioresistance via inhibiting the Wnt1/β-catenin signaling pathway in cervical cancer. *J Exp Clin Cancer Res*. 2020;39(1):183. 10.1186/s13046-020-01659-y.

50. Bonavita E, Gentile S, Rubino M, et al. PTX3 is an extrinsic oncosuppressor regulating complement-dependent inflammation in cancer. *Cell*. 2015;160(4):700-714. 10.1016/j.cell.2015.01.004.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.