



OPEN ACCESS

EDITED BY

Zhigang Ren,
First Affiliated Hospital of Zhengzhou
University, China

REVIEWED BY

Frankie Chi Fat Ko,
The University of Hong Kong, Hong Kong
SAR, China
Christian Cotsoglou,
Ospedale di Vimercate - ASST Brianza, Italy
Jiang Chen,
Zhejiang University, China

*CORRESPONDENCE

Ren P. Xia

✉ rpxia819@163.com

RECEIVED 31 March 2023

ACCEPTED 16 May 2023

PUBLISHED 26 June 2023

CITATION

Zhao Y, Xu X, Wang Y, Wu LD, Luo RL and
Xia RP (2023) Tumor purity–associated
genes influence hepatocellular carcinoma
prognosis and tumor microenvironment.
Front. Oncol. 13:1197898.
doi: 10.3389/fonc.2023.1197898

COPYRIGHT

© 2023 Zhao, Xu, Wang, Wu, Luo and Xia.
This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Tumor purity–associated genes influence hepatocellular carcinoma prognosis and tumor microenvironment

Yan Zhao¹, Xu Xu², Yue Wang¹, Lin D. Wu¹, Rui L. Luo²
and Ren P. Xia^{1*}

¹Department of Organ Transplantation, Kunming Medical University First Affiliated Hospital, Kunming, China, ²Department of Urology, The First Affiliated Hospital of Chengdu Medical College, Chengdu, China

Introduction: Tumor purity takes on critical significance to the progression of solid tumors. The aim of this study was at exploring potential prognostic genes correlated with tumor purity in hepatocellular carcinoma (HCC) by bioinformatics analysis.

Methods: The ESTIMATE algorithm was applied for determining the tumor purity of HCC samples from The Cancer Genome Atlas (TCGA). The tumor purity–associated genes with differential expression (DEGs) were identified based on overlap analysis, weighted gene co-expression network analysis (WGCNA), and differential expression analysis. The prognostic genes were identified in terms of the prognostic model construction based on the Kaplan–Meier (K–M) survival analysis and Least Absolute Shrinkage and Selection Operator (LASSO) regression analyses. The expression of the above-described genes was further validated by the GSE105130 dataset from the Gene Expression Omnibus (GEO) database. We also characterized the clinical and immunophenotypes of prognostic genes. Gene set enrichment analysis (GSEA) was carried out for exploring the biological signaling pathway.

Results: A total of 26 tumor purity–associated DEGs were identified, which were involved in biological processes such as immune/inflammatory responses and fatty acid elongation. Ultimately, we identified ADCK3, HK3, and PPT1 as the prognostic genes for HCC. Moreover, HCC patients exhibiting higher ADCK3 expression and lower HK3 and PPT1 expressions had a better prognosis. Furthermore, high HK3 and PPT1 expressions and low ADCK3 expression resulted in high tumor purity, high immune score, high stromal score, and high ESTIMATE score. GSEA showed that the abovementioned prognostic genes showed a significant correlation with immune-inflammatory response, tumor growth, and fatty acid production/degradation.

Discussion: In conclusion, this study identified novel predictive biomarkers (ADCK3, HK3, and PPT1) and studied the underlying molecular mechanisms of HCC pathology initially.

KEYWORDS

hepatocellular carcinoma, tumor microenvironment, tumor purity, prognostic gene, bioinformatics analysis

1 Introduction

Hepatocellular carcinoma (HCC) has been confirmed as the most frequently reported primary liver cancer, that is, the sixth most common cancer worldwide (1, 2). It is estimated that about 80% of liver cancer cases are HCC (3–5). HCC was a polygenic disease caused by the interaction of a variety of tumor-promoting and tumor-suppressing genes with the tumor microenvironment (TME). However, its molecular mechanism remains unclear (6, 7). HCC is subjected to poor prognosis, and the major reason for this drawback arises from its late presentation and limited therapeutic options (8). Currently, surgery and chemotherapy are the mainstays of treatment for HCC, whereas the incidence of HCC continues to rise and is increasing more rapidly than any other cancer in men and women. HCC ranks among the top 10 cancers for morbidity and mortality, as reported by the systematic analysis of the Global Burden of Disease Study (GBD) (9). Since most HCC patients are diagnosed late, they too weak to resist the risk of surgery. Accordingly, it is urgent to find new diagnostic and prognostic markers for HCC, which may help in early diagnosis and guide treatment decisions for improving patient survival and quality of life. Tumor purity has been defined as the proportion of tumor cells in the TME (10), TME is a cell population composed of a wide variety of cells (e.g., stromal cells, fibroblasts, endothelial cells, and immune cells), and this cell population takes on critical significance to tumors' occurrence and development (11). Moreover, cellular and molecular components in the TME may exert certain effects on treatment outcomes (12). Existing research has suggested that tumor purity may affect co-expression networks, cluster-based classification of tumor sub-types or molecular classification, and identification of differentially expressed genes (13). However, tumor purity-associated markers in HCC have been rarely investigated.

In clinical practice, generally obtained solid tumors constantly comprise multiple clonal populations of cancer cells and adjacent normal tissue, stroma, and infiltrating immune cells, with a high degree of heterogeneity (14). This heterogeneous structure is capable of complicating or biasing the analysis of various genomic data. ESTIMATE has been adopted for the assessment of tumor purity in existing studies using gene expression signatures to determine the proportion of stromal and immune cells in a tumor sample (15). Furthermore, the algorithm has been reported as a robust tumor purity prediction algorithm. A total of 10 tumor purity prognosis-associated genes were extracted from this study for the investigation of tumor purity (16).

However, there is no immune-associated prognostic analysis of tumor purity in liver cancer. This study aims to investigate the correlation and mechanism between tumor purity and clinical prognosis in HCC, which devotes to better clarify prognostic prediction and therapeutic strategies. The analytical method in this study draws much inspiration from existing research (17).

2 Materials and methods

2.1 Data source

The Cancer Genome Atlas (TCGA) database provided RNA-sequencing data (i.e., 50 normal samples and 374 HCC samples). A

total of 342 HCC samples were obtained after excluding samples from patients with incomplete clinical and survival information (Supplementary Table S1). The GSE105130 dataset originated from the GEO database (18) (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE105130>), containing transcriptomic data from paired 25 HCC patients with tumors and adjacent non-tumors. Moreover, the Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) database provided GSE36376 dataset with 240 HCC patients and 193 adjacent non-tumor samples. Data were downloaded from the publicly available database; hence, it was not applicable for additional ethical approval.

2.2 Differential expression analysis

In accordance with the RNA-sequencing data of 50 normal and 374 HCC samples in the TCGA database, differential expression analysis was carried out with the use of the R package limma. In accordance with the HCC versus normal comparison method, genes satisfying $|\log_2\text{fold change (FC)}| > 0.5$ and $P < 0.05$ were considered differentially expressed genes (DEGs). A volcano plot and a heat map were employed for demonstrating the distribution and expression patterns of DEGs.

2.3 Tumor purity and immune landscape

RNA-sequencing data of 374 HCC samples from the TCGA database were extracted as the basis for this step of the analysis. The immune score, stroma score, and ESTIMATE score of all TCGA-HCC samples were assessed using the ESTIMATE algorithm (R package ESTIMATE) (16). Equation 1 was adopted to estimate the respective HCC patient's tumor purity:

$$\begin{aligned} \text{Tumor purity} &= \cos(0.6049872018 + 0.0001467884 \\ &\times \text{ESTIMATE Score}) \end{aligned} \quad (1)$$

The single-sample gene set enrichment analysis (ssGSEA) was further used to assess the abundance of 24 immune cell species in the TCGA-HCC ($n = 374$) cohort, and the result was visualized into a stacked plot.

2.4 Weighted Gene Co-expression Network Analysis

Weighted Gene Co-expression Network Analysis (WGCNA) refers to an analytical method for the analysis of gene expression patterns of multiple samples, allowing clustering of genes with similar expression patterns and analysis of correlations between modules and specific traits or phenotypes (19). For filtering 374 HCC samples, goodSamplesGenes function in R package WGCNA was employed for the first time, and the samples with TURE results were introduced into subsequent analyses. Tumor purity and

immune infiltrating cell abundance were the specific traits of interest, and they are listed in [Supplementary Table S2](#), with the aim of identifying tumor purity-associated genes.

Afterward, a scale-free co-expression network was built based on a soft-threshold parameter β (β was a soft-threshold parameter that could enhance strong correlations between genes and penalize weak correlations) using the adjacency matrix (20).

Subsequently, in accordance with the standard of hybrid dynamic tree cutting algorithm, the minimum number of genes in the respective gene module was set to 50, and MEDissThres was set to 0.2 to merge similar modules analyzed using the dynamic cut tree algorithm.

To identify the gene modules that are immune-associated to tumor purity in HCC, the tumor purity in HCC was defined as the trait was, and the relationship of the respective module and tumor purity was analyzed. The module with the maximum absolute value of the immune correlation coefficient with tumor purity was defined as a critical module for subsequent analysis, with $p < 0.05$ as the threshold with statistical significance, and genes in the vital module were considered hub genes.

2.5 Identification and enrichment analyses of tumor purity-associated genes with differential expression

Common genes between DEGs and hub genes were obtained through the overlap analysis, which were defined as tumor purity-associated DEGs. To explore whether there is an interaction between the tumor purity-associated DEGs, we used the STRING (<https://string-db.org>) website to construct a PPI network for them with a confidence level of 0.4. Subsequently, functional enrichment analyses were performed on tumor purity-associated DEGs by the R package clusterProfiler for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG), and $p < 0.05$ was regarded as notably enriched.

2.6 Survival analysis

For the assessment of whether tumor purity-associated DEGs were correlated with patient survival, we performed a K–M analysis by R-package survival. In brief, patients were divided into high- and low-expression groups based on whether the expression of the tumor purity-associated DEGs was greater than the median expression level of each tumor purity-associated DEG in the TCGA-HCC cohort. The K–M curve of each expression group of each tumor purity-associated DEGs was plotted and compared. The survival difference between high- and low-expression groups with $P < 0.05$ was considered as a gene notably correlated with OS in HCC patients.

2.7 Least Absolute Shrinkage and Selection Operator algorithm

Based on the 342 TCGA-HCC cases with complete survival and clinical information, gene notably correlated with OS obtained in

the previous step were introduced into Least Absolute Shrinkage and Selection Operator (LASSO) algorithm with the parameters “famil” to “binomial” and “type.measure” to “class,” to select strong correlation features and obtain prognostic genes when the cross-validation error was the lowest.

To evaluate the prognostic value of the risk model, the risk score of each HCC sample was firstly calculated according to the expressions of prognostic genes and the risk coefficient obtained by LASSO with the formula: risk score. Then, the HCC samples were divided into high- and low-risk groups based on the optimal threshold of the risk score calculated by the surv cutpoint function of the R package survminer. Subsequently, a K–M survival difference analysis was employed on the high- and low-risk groups.

2.8 Establishment of a nomogram

The prognostic genes were further enrolled in developing a nomogram using rms and survival packages to predict the survival rate of patients with HCC, and riskRegression and survival packages were applied to plot calibration curve and quantified data of each risk group.

2.9 Clinical and immunological phenotypes of prognostic genes

First, clinical characteristics, covering T-phase (T1–T4), N-phase (N0 and N1), M-phase (M0 and M1), and PHASE (phase I–phase IV) were extracted from the 342 HCC samples. Then, the expression of each model gene was compared between subgroups of each clinical characteristic. To analyze the correlation between the expression of prognostic genes and the clinical characteristics of HCC, we compared the differences in the expression of three prognostic genes according to the clinical characteristics of different groups.

Moreover, based on the median expression value of the respective model gene, the HCC samples were separated into high- and low-expression groups of the respective model gene. Subsequently, CD8.T.cells, stromal scores, ESTIMATE scores, and tumor purity were compared between high and low expression groups of each model gene.

Moreover, the Pearson correlation of each model gene with the immune score, stromal score, ESTIMATE score, CD8.T.cells, and tumor purity was determined.

2.10 Single-gene gene set enrichment analysis

To investigate the relevant pathways and biological functions of the prognostic genes, the prognostic genes served as target genes, and the correlation coefficients between all genes in HCC samples and the expression levels of target genes were considered the sorting standards in terms of GSEA, which was performed by clusterProfiler with the selection standards of adjusted $p < 0.05$.

2.11 Validation of expression levels of prognostic genes

To further verify the expression levels of the prognostic genes, the expression of each model gene was compared between normal and HCC samples in TCGA cohort and GSE105130 dataset respectively. Furthermore, human hepatic stellate cell line LX-2, as well as three human HCC cell lines HepG2, SK-HeP-1, and Huh-1 were purchased from Procell Life Science and Technology Co., Ltd. (Wuhan, China). Total RNA from the above four cell lines (logarithmic phase) was segmented *via* the TRIzol Reagent based on the producer's indications (Ambion, TX, USA). For the next processes, the inverse transcription of RNA into cDNA was done *via* the SweScript-First-strand-cDNA-synthesis-kit (Servicebio, Wuhan, China) and qPCR was completed through the 2× SYBR Green qPCR Master Mix depending on the manuals' indications (Servicebio, Wuhan, China). The detailed information of primers synthesized by Beijing Tsingke Biotech Co., Ltd. (Beijing, China) is listed in Table 1. The relative expression of the respective prognostic gene was uniformized by GAPDH. The student's *t*-test was carried out to compute the *P*-values between two groups. The *P*-value < 0.05 (two-tailed) was delimited as statistically significant.

2.12 Statistical analysis

All statistical analyses were carried out using R software (version 3.5.2) and the relevant software packages. The specific statistical methods were stated in the relevant sections. Without special remarks, *P* < 0.05 was considered with statistical significance.

3 Results

3.1 Identification of the hepatocellular carcinoma-associated genes with differential expression

Differential expression analysis was conducted for 50 normal and 374 HCC samples in the TCGA database using the R package limma. Based on $|\log_2 FC| > 0.5$ and *P* < 0.05, a total of 6251 DEGs

were identified, of which 848 were upregulated and 5,403 were downregulated in HCC samples in comparison with normal samples (Figure 1; Supplementary Table S3).

3.2 Analysis of tumor purity and immune infiltrating cell co-expression network

The expression profiles of 374 HCC samples from the TCGA database were selected as the basis for WGCNA. The tumor purity and immune-infiltrating cell content of all TCGA-HCC samples obtained were available in Supplementary Figure S1, and they would serve as the phenotype of interest for this study. The similarity of TCGA-HCC samples was detected using the goodSamplesGenes function (Figure 2A). A total of nine modules were obtained by WGCNA (Figure 2B). Subsequently, we examined the correlation of tumor purity and immune-infiltrating cells with the respective module. As indicated by the results, the light-yellow module was the highest correlated with tumor purity (*cor* = 0.89, *P* = 2e-147) and also exhibited a moderate to the strong relationship with a variety of immune infiltrating cells (e.g., T cells, macrophages, Th 1 cells, aDC, cytotoxic cells, etc.) (Figure 2C), and the 420 genes in the light-yellow module were regarded as hub genes (Figure 2D).

3.3 Identification and enrichment analysis of the tumor purity-associated genes with differential expression

Overlap analysis revealed a total of 26 common genes (Supplementary Table S4) were found between DEGs and hub genes as shown in Figure 3A, and they were defined as the tumor purity-associated DEGs. Moreover, the 26 tumor purity-associated DEGs enriched 10 GO terms and nine KEGG pathways. For instance, in the BP group, they were mainly involved in neutrophil activation, degranulation, and their mediated immunity, and granulocyte migration; in the CC category, the "tertiary granule," "ficolin-1-rich granule," "ficolin-1-rich granule membrane," and "tertiary granule membrane" were notably enriched; in the MF group, the above-described genes were closely correlated with "RAGE receptor binding" (Figure 3B). According to KEGG pathway enrichment analysis (Figure 3C), the abovementioned tumor purity-associated DEGs were involved in immune/inflammatory responses ("NF-kappa B signaling pathway," "Primary immunodeficiency," and "Cytokine-cytokine receptor interaction") and metabolism ("Galactose metabolism," "Fructose and mannose metabolism," and "starch and sucrose metabolism")-associated pathways, "salmonella infection," and "biosynthesis of nucleotide sugars"; moreover, surprisingly, "fatty acid elongation" pathway also appeared notably enriched. Recently, it was reported that fatty acid extension from C16 to C18 can promote hepatic lipid accumulation and inflammation thereby promoting liver disease (21–23). The above evidence suggested that the above-described tumor purity-associated DEGs may affect the immune and inflammatory responses of patients by

TABLE 1 The detailed information of primers for qPCR.

Primer name	Primer sequence
PPT1 For	GAGGACGTGGAGAACAGCTT
PPT Rev	GGCATCGAGGGAGTCCAAAA
HK3 For	TGAGGTTGGGCTAGTTGTAGACAC
HK3 Rev	TGAGCACCAGGATTCAGGGA
ADCK3 For	CAGCCAGGAGATTCGGAACG
ADCK3 Rev	TATGGATTCGCCGCACA
GAPDH For	CGAAGGTGGAGTCAACGGATTT
GAPDH Rev	ATGGGTGGAATCATATTGGAAC

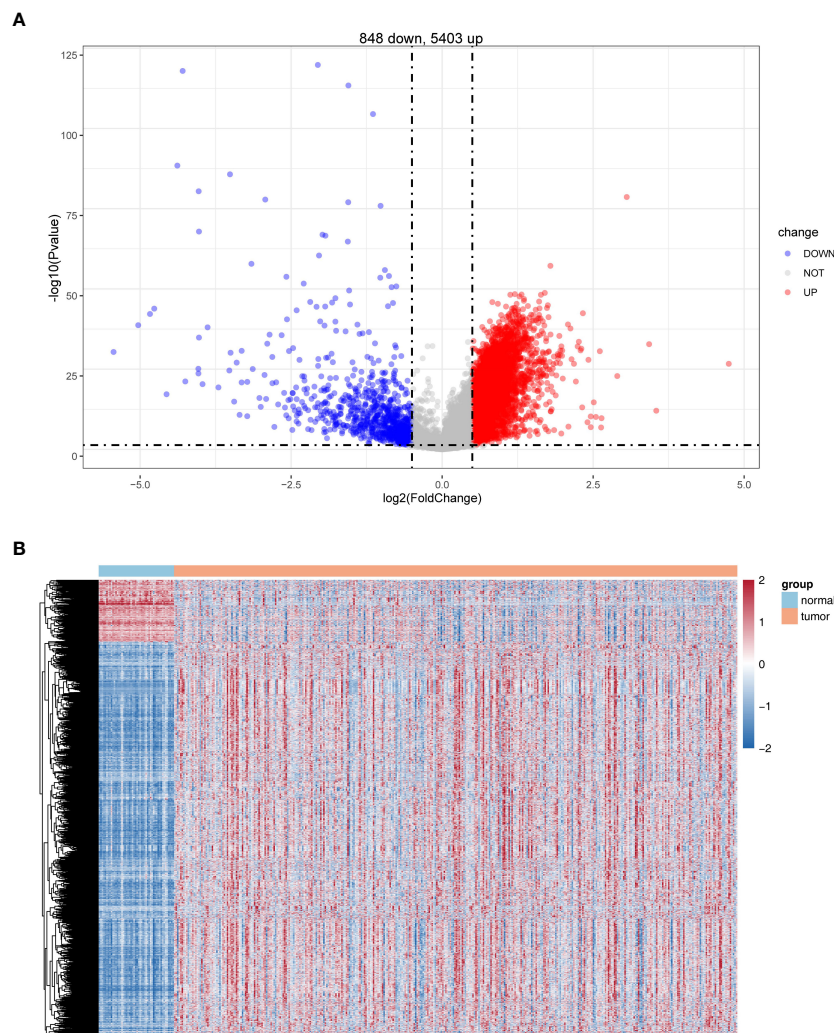


FIGURE 1 Identification of differentially expressed genes (DEGs). (A) Volcano plot of DEG (HCC vs. normal). (B) Heatmaps of DEGs (HCC vs. normal).

regulating fatty acid elongation and thus play an important role in HCC progression.

3.4 Identification of tumor purity-associated prognostic genes

We performed a K-M survival analysis designed to assess the correlation of changes in expression of 26 tumor purity-associated DEGs with OS in TCGA-HCC patients ($n = 342$). As indicated by the results, the expression of CD300A, FPR1, HK3, PPT1, and RGS10 was inversely correlated with patient survival, and high expression of the above-described genes was notably related to short survival time; patients with high ADCK3 and DCAF8 expression had notably better OS than those in the low-expression group (Figure 4A). The expression changes of the remaining 19 genes could not notably differentiate the clinical outcomes of HCC patients (Supplementary Figure S3). Subsequently, we included the seven genes mentioned above that were notably correlated with HCC prognosis to the LASSO regression analysis.

Ultimately, ADCK3, HK3, and PPT1 were identified as prognostic genes based on $\lambda \text{ min} = 0.0373$ (Figure 4B). In addition, based on Equation 2:

$$\begin{aligned} \text{Risk Score} = & (-0.01588) \times \text{expression ADCK3} \\ & + 0.183293 \times \text{expression PPT1} + 0.09556 \\ & \times \text{expression HK3} \end{aligned} \quad (2)$$

The optimal threshold of 5.65, and the cut point of maximally selected rank statistics = 0.73, the HCC samples were divided into high- ($n = 178$) and low-risk ($n = 187$) groups. Furthermore, the survival analysis result illustrated that the OS of low-risk group was notably higher than high-risk group ($p = 0.00011$) (Figure 4C).

Next, the expression levels of three key prognostic genes were compared between HCC samples and controls in the TCGA-HCC cohort and GEO150130. In both datasets, HCC tissues had lower HK3 expression and higher PPT1 and ADCK3 expressions than paired adjacent non-tumor tissues (Figures 4D–E). To further validate the prognostic gene expression trends experimentally, we utilized cell lines and qRT-PCR technique. As indicated by the

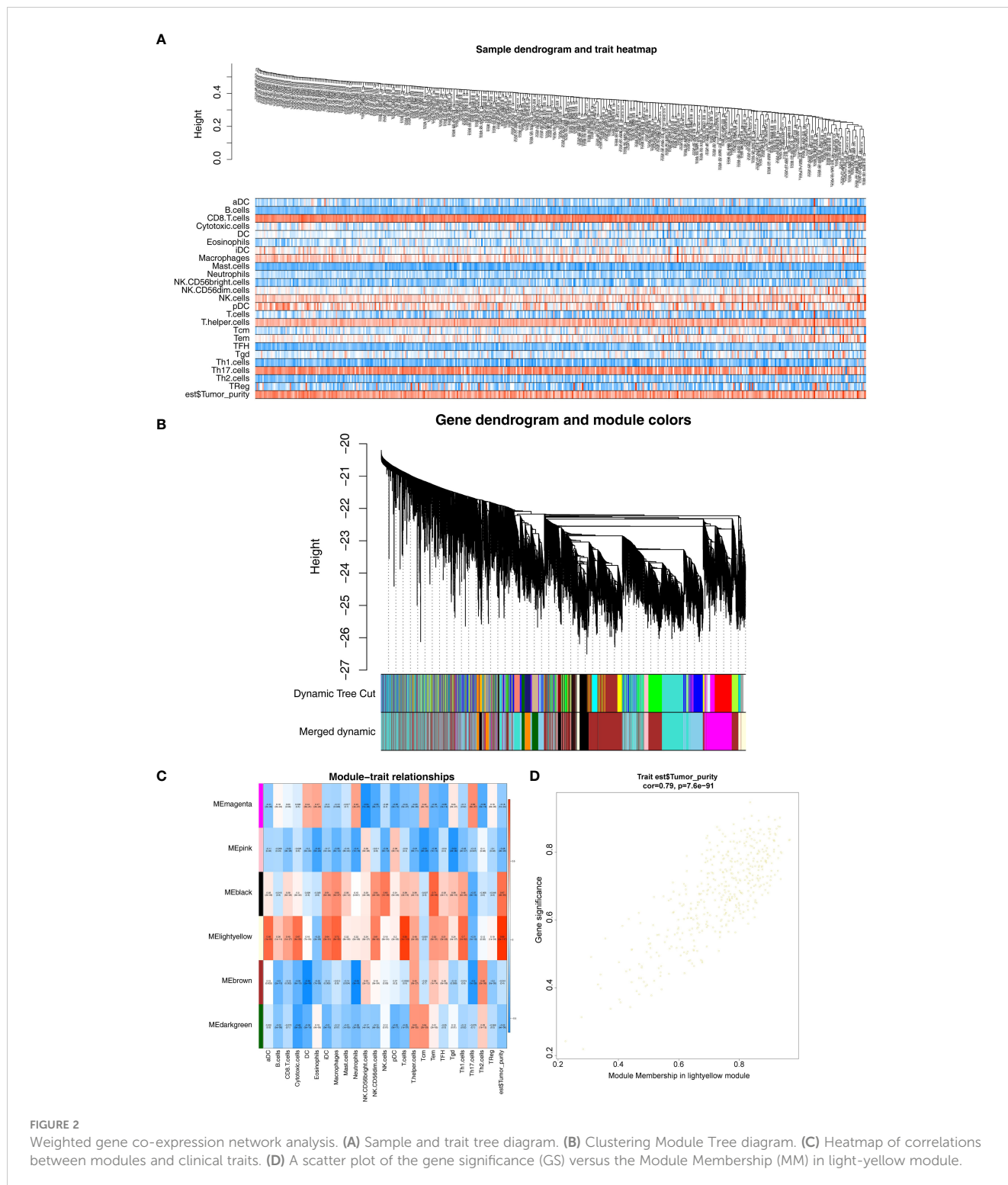


FIGURE 2 Weighted gene co-expression network analysis. **(A)** Sample and trait tree diagram. **(B)** Clustering Module Tree diagram. **(C)** Heatmap of correlations between modules and clinical traits. **(D)** A scatter plot of the gene significance (GS) versus the Module Membership (MM) in light-yellow module.

results, in agreement with the results of the public database tissue, PPT1 and ADCK3 expression was upregulated in three hepatocellular carcinoma cell lines (Huh-1, HepG2, and SK-HeP-1) in comparison with normal hepatocytes LX-2 (Supplementary Figure S4). However, probably due to differences in cell lines and tissues, the expression trend of HK3 in cell lines was opposite to that in tissues (Supplementary Figure S4). The expression levels of three key prognostic genes were also validated in the GSE36376 dataset

and achieved a consistent result (Supplementary Figure S5A). The AUC values of the prognostic model at 1, 2, 3, 4, and 5 years reached 0.68, 0.64, 0.60, 0.60, and 0.59, respectively (Supplementary Figure S5B).

Furthermore, we analyzed the correlation of the three prognostic genes exhibiting the clinical characteristics (TNM phase and phase) of TCGA-HCC samples. As indicated by the results, only PPT1 was notably correlated with the phase ($P =$

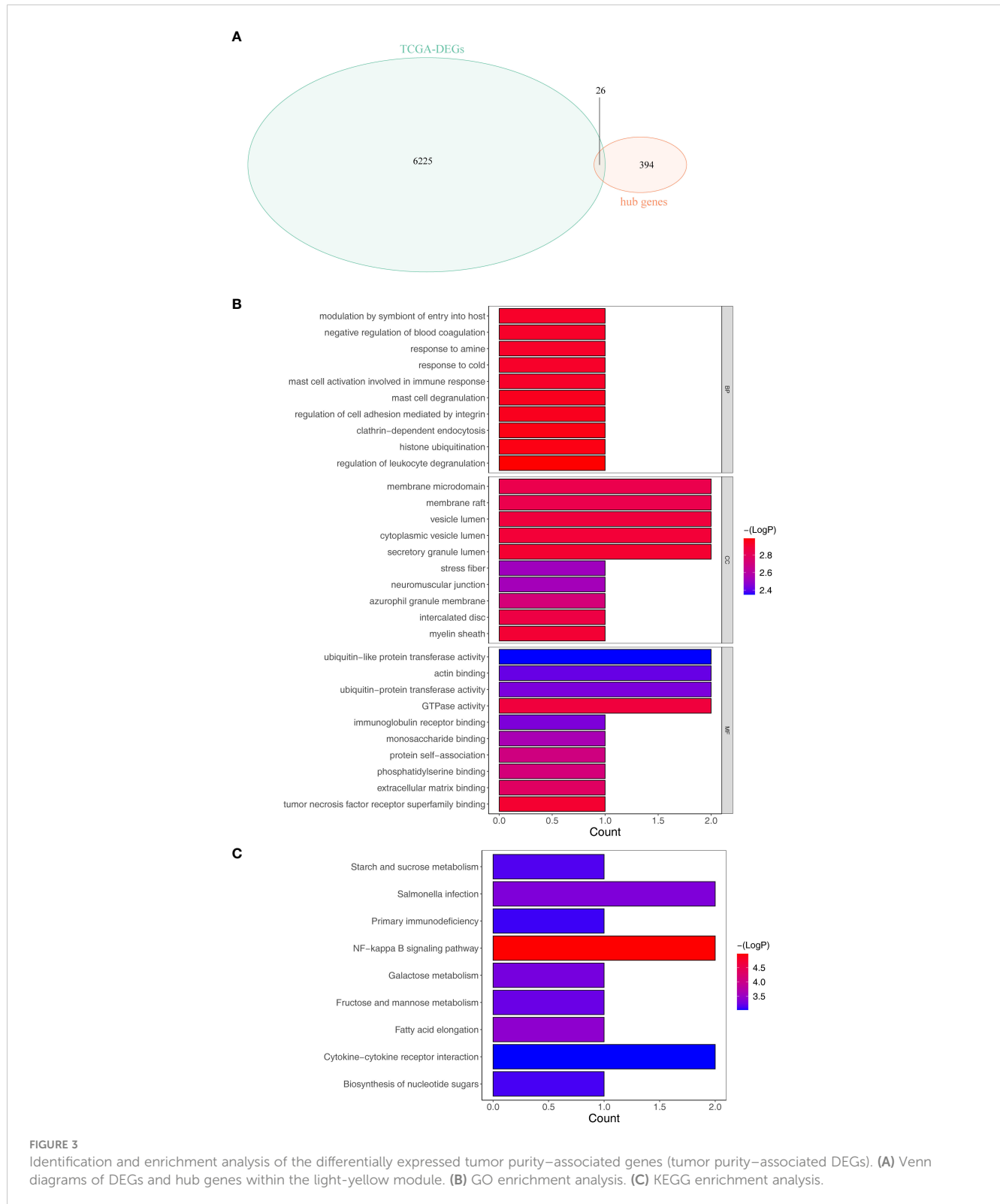


FIGURE 3 Identification and enrichment analysis of the differentially expressed tumor purity-associated genes (tumor purity-associated DEGs). **(A)** Venn diagrams of DEGs and hub genes within the light-yellow module. **(B)** GO enrichment analysis. **(C)** KEGG enrichment analysis.

0.0033) and T-phase ($P = 0.0014$). In the phase subgroup, the expression level of PPT1 gradually increased in patients at phase I to phase III but decreased in patients at phase IV. In the T-phase subgroup, PPT1 expression increased with increasing the tumor size and depth of infiltration (Supplementary Figure S6).

3.5 Development of a nomogram

To evaluate the role played by the prognostic model, a nomogram was developed with an optimal concordance index (C-index, 0.61353) for the prediction of the survival time of HCC patients at 1, 3, and 5 years (Figure 5A). The calibration diagram

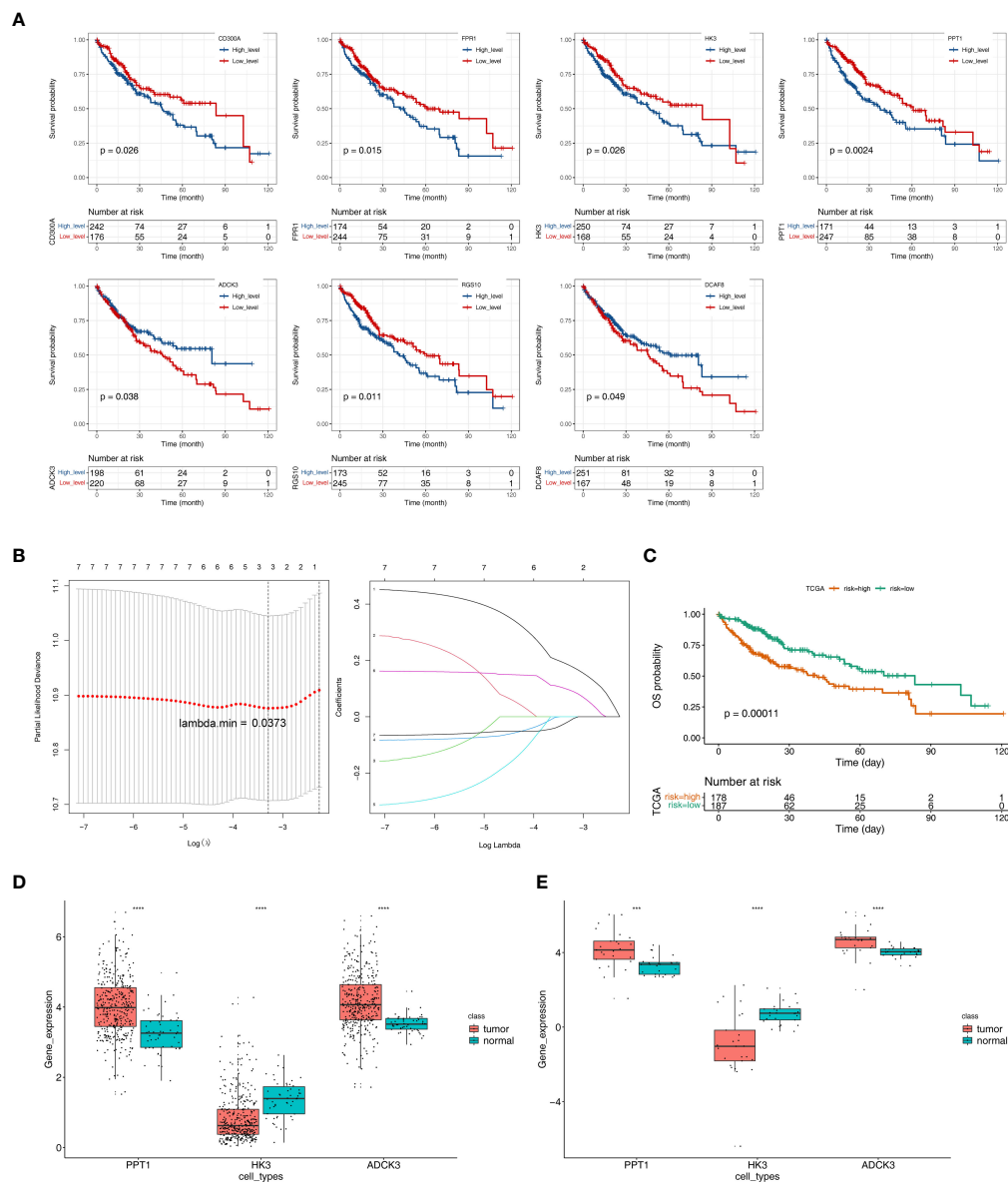


FIGURE 4 Identification of tumor purity-associated prognostic genes. **(A)** K–M survival analysis of 26 tumor purity-associated DEGs with OS in TCGA-HCC patients. **(B)** LASSO regression analysis. **(C)** Survival analysis of high- and low-risk groups. **(D)** The expression levels of three key prognostic genes in the TCGA-HCC cohort. **(E)** The expression levels of three key prognostic genes in GSE105130 dataset. *** $p < 0.001$, **** $p < 0.0001$.

was plotted, closer to the ideal curve, suggesting the perfect stability of the nomogram (Figure 5B). The above results demonstrated that the nomogram based on three prognostic genes could supply a high value for predicting the prognosis of patients with HCC.

3.6 Correlation of prognostic genes with immunophenotypes

We evaluated the immunophenotype of three prognostic genes. As indicated by the results, ADCK3, HK3, and PPT1 were notably correlated with tumor purity, immune score, stromal score, and ESTIMATE score (Figure 6). High HK3 and PPT1 expression and low ADCK3 expression resulted in high tumor purity, high immune

score (Figure 6A), high immune score (Figure 6B), high stromal score (Figure 6C), and high ESTIMATE score (Figure 6D). Scatter plots of the correlations between the abovementioned prognostic genes and tumor environment scores were shown in Supplementary Figure S7.

3.7 Exploration of the potential molecular mechanisms for prognostic genes

To reveal the molecular mechanisms of prognostic genes, we performed a single-gene GSEA on three prognostic genes. ADCK3 was enriched for a total of 2,172 GO terms (1659 BP terms, 245 CC terms, and 268 MF terms; Supplementary Table S5, sheet1); HK3 was enriched for a total of 1,669 GO terms (1,374 BP terms, 124 CC

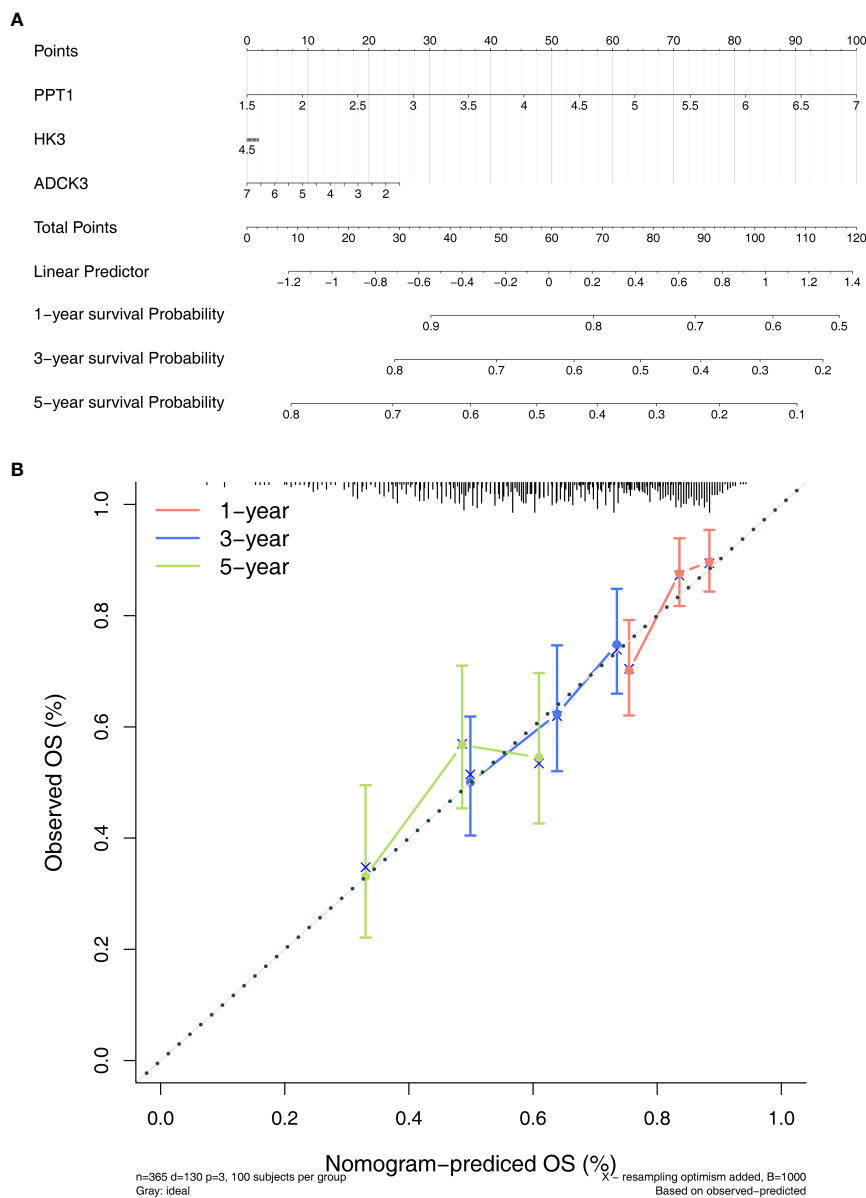


FIGURE 5 Establishment of a nomogram. **(A)** Nomogram for predicting the survival rate of patients with HCC based on three prognostic genes. **(B)** Calibration curve plotted to evaluate consistency of predicted and actual observations.

terms, and 171 MF terms; [Supplementary Table S5](#), sheet2); PPT1 was enriched with a total of 1,341 GO terms (964 BP terms, 183 GO terms, and 194 MF terms; [Supplementary Table S5](#), sheet3). [Figure 7A](#) showed the top 10 terms of each prognostic gene in the GO Overall; all three prognostic genes were closely correlated with immune cell physiological processes (e.g., differentiation, activation, proliferation, adhesion, and related regulatory signals), immune-inflammatory response, cell cycle, vascular growth, and fatty acid production. The top 10 KEGG pathways of the three prognostic genes were displayed in [Figure 7B](#), and they were notably correlated with immune-inflammatory response-associated pathways, also involved in Fatty acid degradation, Tyrosine metabolism, and Primary bile acid biosynthesis. Exhaustive

KEGG enrichment results for ADCK3, HK3, and PPT1 could be reviewed in [Supplementary Table S6](#).

3.8 Identification of differential immune cells and correlation analysis

To clarify the correlation between the prognostic model and immune infiltration, we recognized the differential immune cells between high- and low-risk subgroups. [Figure 8A](#) revealed the proportion of 14 immune cells [eosinophils, aDC, iDC, macrophages, CD56 bright natural killer (NK) cells, CD56 dim NK cells, T cells, T-helper cells, Tcm, Tem, TFH, TH1 cells, TH2

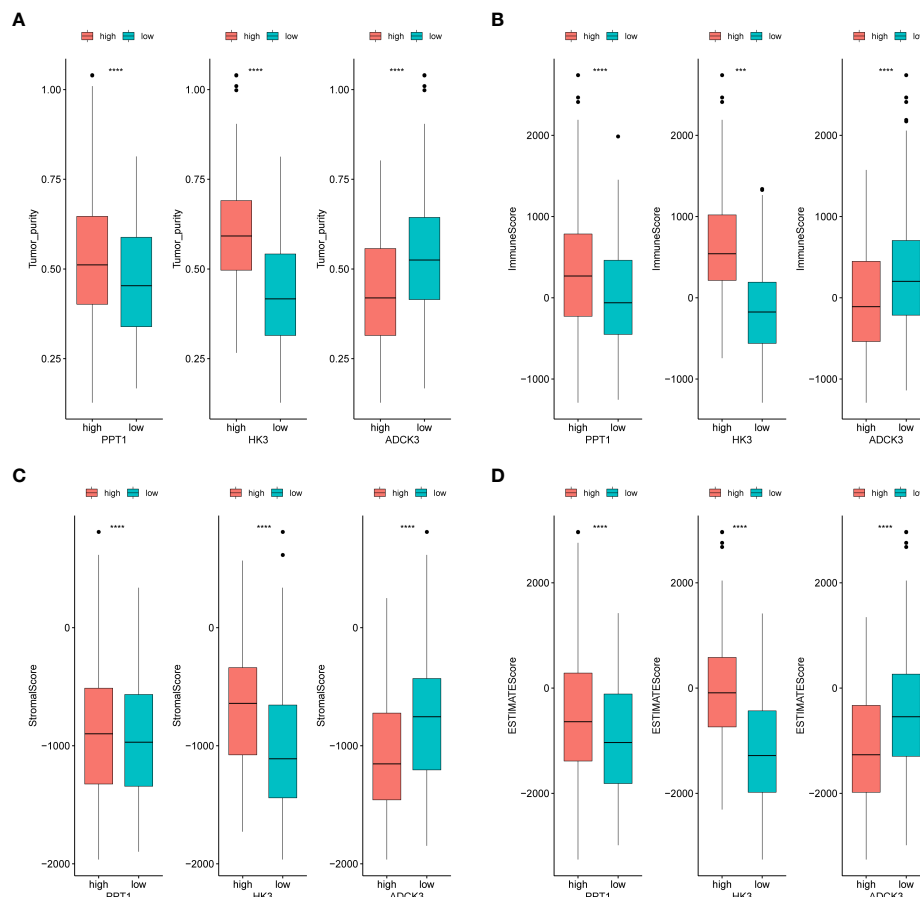


FIGURE 6 Identification of tumor purity-associated prognostic genes. **(A)** Differences in tumor purity between high- and low-expression groups of prognostic genes. **(B)** Differences in immune score between high- and low-expression groups of prognostic genes. **(C)** Differences in stromal score between high- and low-expression groups of prognostic genes. **(D)** Differences in ESTIMATE score between high- and low-expression groups of prognostic genes. *** $p < 0.001$, **** $p < 0.0001$.

cells, and Th17 cells] was notably different between the high- and low-risk groups. In addition, most of the differential immune cells were positively correlated with PPT1 and HK3 but negatively correlated with ADCK3 (Figure 8B). Thus, differential immune cells might be linked with occurrence and development of HCC.

4 Discussion

Hepatocellular carcinoma (HCC) is the most common primary liver cancer with poor prognosis. The incidence of HCC and HCC-associated deaths have increased over the past several decades (24). Consequently, there is an urgent need to find prognostic markers for HCC. Previous studies have shown that tumor purity is correlated with patient prognosis (11, 25). Tumor purity refers to the proportion of cancer cells in the tumor tissue. Several computational methods that can determine tumor purity have been introduced with the advance of genomics, which made the measurement of tumor purity more objective and accurate. In accordance with the ESTIMATE algorithm, tumor purity was estimated based on immune score and stromal score. Tumor immune score is an important factor affecting tumor progression

and immunotherapy outcomes (26). In this study, ESTIMATE was used to calculate tumor purity of each HCC sample in TCGA-LIHC cohort. Through the ssGSEA algorithm, the immune activity of the respective sample can be accurately obtained.

In this study, we screened out three key prognostic genes; they are PPT1, HK3, and ADCK3. Palmitoyl-protein thioesterase 1 (PPT1) is an enzyme that cleaves thioester-linked palmitate from S-acylated proteins in lysosomes (27), Palmitoyl-protein thioesterase 1 (PPT1) was transported to lysosomes through the mannose-6-phosphate receptor-mediated pathway, and it participates in the lysosomal degradation (28). In addition, PPT1 is known to be widely and notably overexpressed in a variety of cancers, including breast, thyroid, and gastric cancers. In addition, higher expression levels of PPT1 in tumors are correlated with shorter overall survival for a variety of cancers. In our study, PPT1 expression was highly expressed in HCC tissues compared with normal tissue. Meanwhile, PPT1 expression in clinical samples is correlated with worse prognosis. Survival analyses for TCGA cancer patients also demonstrated that tumor expression of PPT1 was correlated with shorter overall survival in HCC (29). Meanwhile, PPT1 expression in clinical samples is correlated with worse prognosis. Therefore, we may infer that HCC patients exhibiting

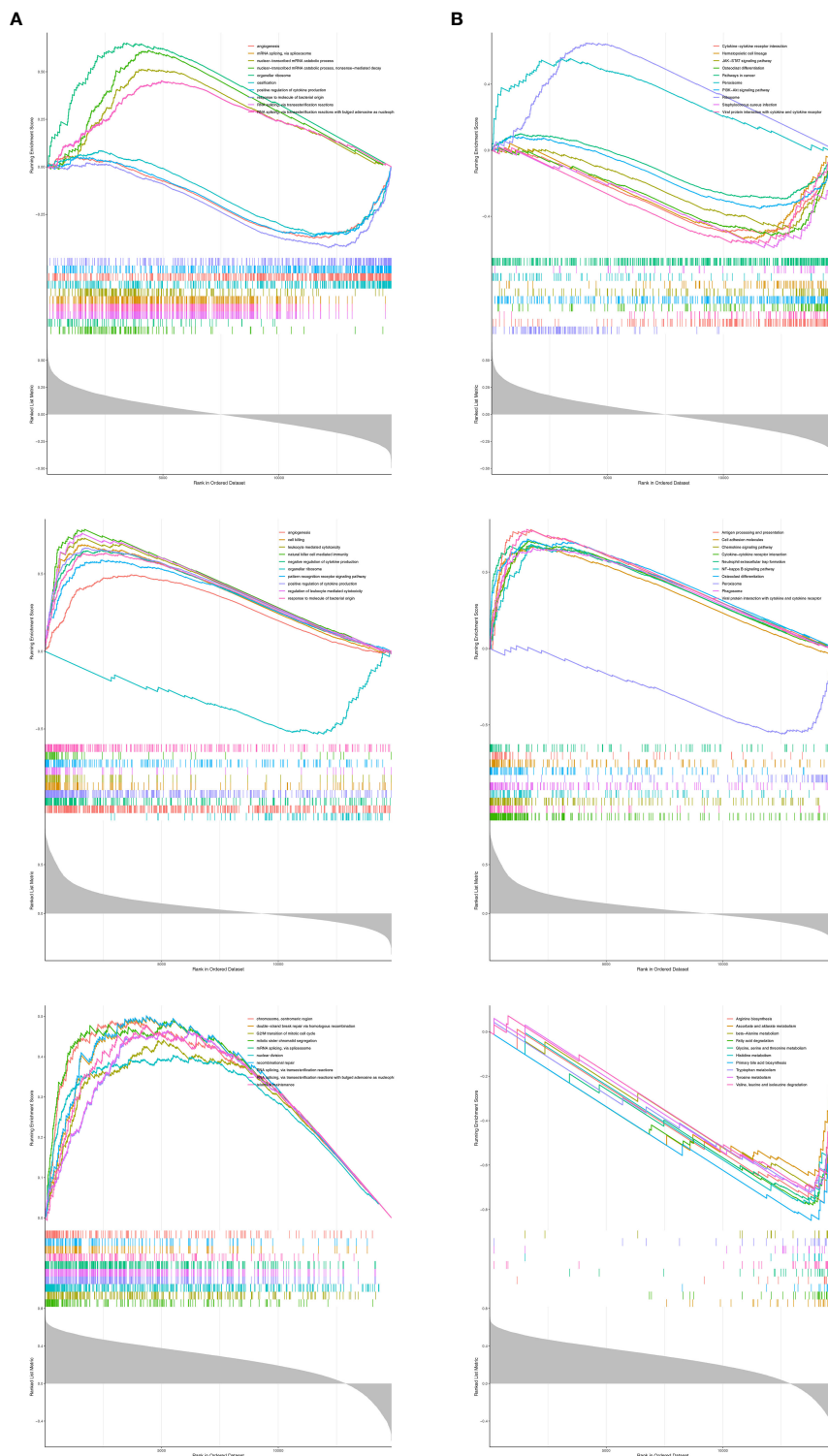


FIGURE 7 The top 10 terms of each prognostic gene in GSEA analysis. (A) GO analysis. (B) KEGG analysis.

higher expression of PPT1 had a worse prognosis. In addition, PPT1 was notably enriched in pathways such as fat metabolism. There are studies that have indicated that, in variety of cancers, fat uptake, storage, and fat production are upregulated, which in turn promotes the rapid growth, invasion, and migration of tumors (30).The above-described results suggest that PPT1 gene may affect

the occurrence and development of HCC through fat metabolism. Also, multiple substance synthesis pathways promotes the progression and malignant behaviors of cancers (31).

Hexokinase-3 (HK3), that is, a member of the hexokinase family, is involved in the first step of glucose metabolism, and its coding gene is located on the human chromosomal 5q35.2 segment.

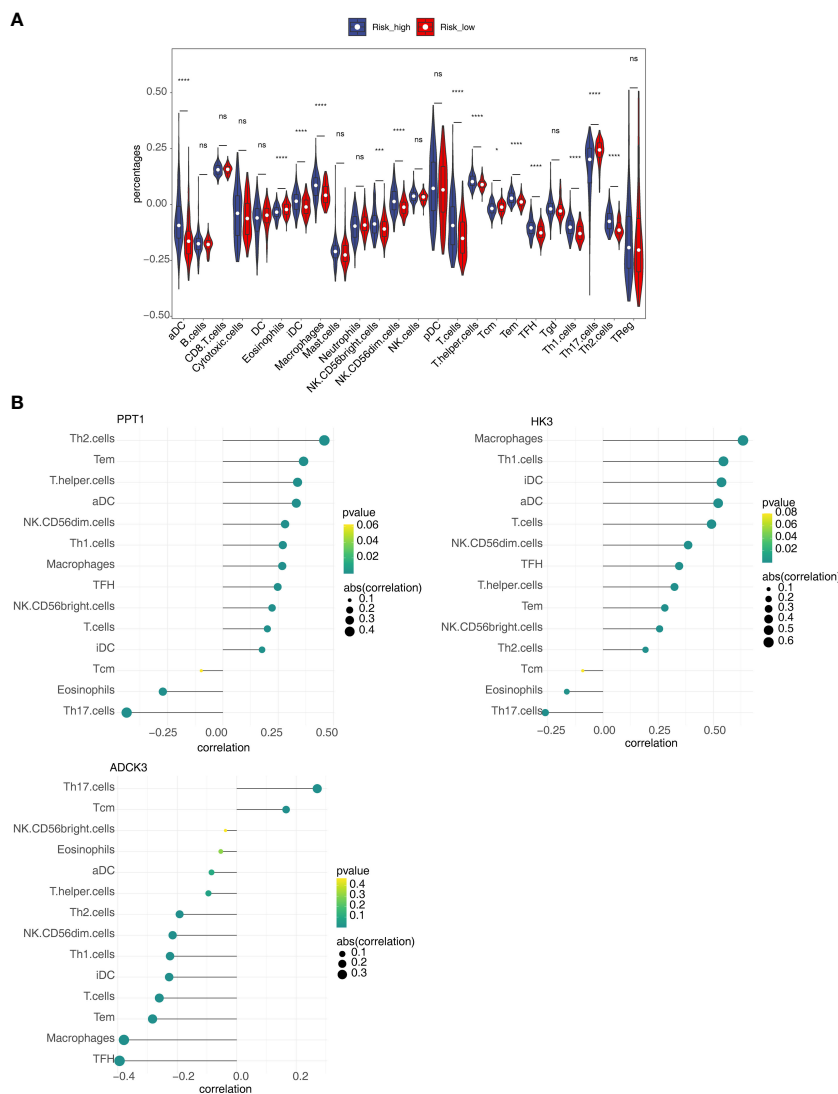


FIGURE 8 Relevance of three prognostic genes and 14 differential immune cells. **(A)** The proportion of 24 immune cells in high- and low-risk groups. **(B)** The relevance of three prognostic genes (PPT1, HK3, and ADCK3) to 14 differential immune cells. ns, not significant; * $p < 0.05$, *** $p < 0.001$, and **** $p < 0.0001$.

The inactivation of HK3 notably affects the activation of cancer cells glycolysis, and then activates the downstream signaling pathways, such as apoptosis and endoplasmic reticulum stress. In cancer cells, which plays a vital role in the progression and development of cancers (32). Also, in numerous studies, HK3 has been identified as a potential marker for regulating the tumor metabolic microenvironment and malignant progression, with predictive efficacy for tumor progression and prognosis. In addition, low expression of HK3 were usually malignant entities and were shown to be obvious genomic aberrations of driver oncogenes poor (33). In this study, from enrichment analysis of GSEA, the HK3 gene was notably enriched in cancer pathways, cellular communication factor, PI3K-Akt Pathway, and virus response. Cancer pathway takes on critical significance to cancer progression and cancer-associated genes' expression, regulating

tumor progression and prognosis (34). Furthermore, the validation of expression level in the GEO external dataset confirmed that the HK3 gene was upregulated in HCC compared with normal groups, consistent with previous studies, suggesting higher expression of HK3 in HCC patients with poor prognosis.

ADCK3 has been confirmed as a mitochondrial protein homologous to the yeast COQ8 and the bacterial UbiB proteins, required for CoQ biosynthesis. Amount of experiments strongly suggested that ADCK3 is also involved in CoQ10 biosynthesis (35). CoQ10 may play a certain role in regulating apoptosis (36, 37). CoQ10, an energy transfer molecule, occurs in high levels in the liver. As reported by some research, a possible inverse correlation exists between blood CoQ 10 levels and cancer (38). In the above research, similar to the HK3 gene, the ADCK3 gene also enriched in cancer pathways, cellular communication factor, PI3K-Akt Pathway

and virus response. The PI3K-Akt Pathway is closely related to the malignant behavior of tumor cells, and it manifested its compelling influence on multiple cellular process in different cancers, which are closely related with tumorigenesis, proliferation, growth, apoptosis, invasion, metastasis, epithelial–mesenchymal transition, stem-like phenotype, immune microenvironment and drug resistance of cancer cells (39). Moreover, the PI3K/Akt pathway plays an important role in tumor formation and metastasis. Various analysis identified, PI3K/Akt mutation status can be used as a novel predictor of cancer patients (40). Our research in GEO external validation dataset also identified ADCK3 showed high-expression level in cancer tissue in comparison with normal tissue. Thus, we can speculate that better prognosis of HCC patients correlated with high ADCK3 expression.

In summary, our study identified the abovementioned TME-associated prognostic markers in HCC. PPT1 and HK3 were overexpressed in tumor and its high expression was correlated with poor prognosis, while high ADCK3 expression was correlated with better survival. However, the prognostic value of the three genes warrants further validation by more clinical data. Importantly, the HK3 and ADCK3 genes were not only highly expressed in HCC but also correlated with PI3K/Akt Pathway and cancer pathways. It holds a great potential as a candidate for targeted immunotherapy of HCC.

5 Conclusions

In conclusion, this study identified three novel predictive biomarkers (ADCK3, HK3, and PPT1) correlated with tumor purity for HCC by bioinformatics methods. In addition, the nomogram was established based on three biomarkers to predict the survival time of HCC patients at 1, 3, and 5 years. The current study might contribute to the prognostic workup of HCC and played a complementary role in determining tumor purity phenotypes.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

YZ and RX contributed equally to this work. YZ and LW conducted data collection, procession and analysis. RX and YZ, XX conducted result summaries, visualizations, and validation. All the authors contributed to the manuscript's development.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2023.1197898/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Distribution of immune infiltrating cells in all samples. The horizontal axis represents the respective sample, the vertical axis represents the size of the immune cell score, and different colors represent different immune cells.

SUPPLEMENTARY FIGURE 2

Weighted gene co-expression network analysis. (A) Screening of scale-free soft thresholds. The horizontal axis of the above graphs all represent the weight parameter power value, and the vertical axis of the left figure represents the square of the correlation coefficient of $\log(k)$ and $\log(p(k))$ in the relevant network, that is, signedR2, the vertical axis of the right graph represents the mean of the adjacency functions of all genes in the relevant gene module. (B) Clustering of module Eigen genes.

SUPPLEMENTARY FIGURE 3

K–M survival analysis of the 19 differentially expressed tumor purity-associated genes (tumor purity-associated DEGs).

SUPPLEMENTARY FIGURE 4

PPT1 and ADCK3 expression was upregulated in three hepatocellular carcinoma cell lines (Huh-1, HepG2, and SK-HeP-1) in comparison with normal hepatocytes LX-2.

SUPPLEMENTARY FIGURE 5

Validation of the expression levels of prognostic genes and the prognostic model. (A) The expression levels of prognostic genes in the GSE36376 dataset. **** $p < 0.0001$. (B) The ROC curve of the prognostic model.

SUPPLEMENTARY FIGURE 6

Difference analysis of three model genes in different clinical groups (TNM phase and phase) of TCGA-HCC samples.

SUPPLEMENTARY FIGURE 7

The correlations between prognostic genes and tumor environment scores, (A) tumor purity, (B) immune score, (C) stromal score, (D) ESTIMATE score.

SUPPLEMENTARY TABLE 1

The clinical and survival information of samples in TCGA dataset.

SUPPLEMENTARY TABLE 2

Tumor purity and immune infiltrating cell abundance in 374 HCC samples.

SUPPLEMENTARY TABLE 3

Differentially expressed genes (DEGs) between HCC samples and normal samples.

SUPPLEMENTARY TABLE 4

The tumor purity-associated DEGs in HCC.

SUPPLEMENTARY TABLE 5

The GO enrichment results of ADCK3, HK3, and PPT1.

SUPPLEMENTARY TABLE 6

The KEGG enrichment results of ADCK3, HK3, and PPT1.

References

- Wang HD, Naghavi M, Allen C, Barber RM, Bhutta ZA, Carter A, et al. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the global burden of disease study 2015. *Lancet* (2016) 388(10053):1459–544. doi: 10.1016/S0140-6736(16)31012-1
- Villanueva A. Hepatocellular carcinoma. *New Engl J Med* (2019) 380(15):1450–62. doi: 10.1056/NEJMra1713263
- Marquardt JU, Andersen JB, Thorgeirsson SS. Functional and genetic deconstruction of the cellular origin in liver cancer. *Nat Rev Cancer* (2015) 15(11):653–67. doi: 10.1038/nrc4017
- Jemal A, Ward EM, Johnson CJ, Cronin KA, Ma J, Ryerson B, et al. Annual report to the nation on the status of cancer, 1975–2014, featuring survival. *J Natl Cancer Institute* (2017) 109(9):dix030. doi: 10.1093/jnci/djx030
- Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA: Cancer J Clin* (2015) 65(2):87–108. doi: 10.3322/caac.21262
- Chen F, Zhang K, Huang Y, Luo F, Hu KP, Cai Q. SPC25 may promote proliferation and metastasis of hepatocellular carcinoma via p53. *FEBS Open Bio* (2020) 10(7):1261–75. doi: 10.1002/22111-5463.12872
- Hao X, Sun G, Zhang Y, Kong X, Rong D, Song J, et al. Targeting immune cells in the tumor microenvironment of HCC: new opportunities and challenges. *Front Cell Dev Biol* (2021) 9:775462. doi: 10.3389/fcell.2021.775462
- Ikeda M, Morizane C, Ueno M, Okusaka T, Ishii H, Furue J. Chemotherapy for hepatocellular carcinoma: current status and future perspectives. *Japanese J Clin Oncol* (2018) 48(2):103–14. doi: 10.1093/jjco/hyx180
- Laursen L. A preventable cancer. *Nature* (2014) 516(7529):S2–3. doi: 10.1038/516S2a
- Gong Z, Zhang J, Guo W. Tumor purity as a prognosis and immunotherapy relevant feature in gastric cancer. *Cancer Med* (2020) 9(23):9052–63. doi: 10.1002/cam4.3505
- Zeng Z, Li J, Zhang J, Li Y, Liu X, Chen J, et al. Immune and stromal scoring system associated with tumor microenvironment and prognosis: a gene-based multi-cancer analysis. *J Trans Med* (2021) 19(1):330. doi: 10.1186/s12967-021-03002-1
- Cheng YQ, Wang SB, Liu JH, Jin L, Liu Y, Li CY, et al. Modifying the tumour microenvironment and reverting tumour cells: new strategies for treating malignant tumours. *Cell proliferation* (2020) 53(8):e12865. doi: 10.1111/cpr.12865
- Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nat Commun* (2015) 6:8971. doi: 10.1038/ncomms9971
- Wang F, Zhang N, Wang J, Wu H, Zheng X. Tumor purity and differential methylation in cancer epigenomics. *Briefings Funct Genomics* (2016) 15(6):408–19. doi: 10.1093/bfpg/elw016
- Yoshihara K, Shahmoradgol M, Martínez E, Vegesna R, Kim H, Torres-García W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* (2013) 4:2612. doi: 10.1038/ncomms3612
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43(7):e47. doi: 10.1093/nar/gkv007
- Ming B, Qi P, Chen S. Tumor purity coexpressed genes related to immune microenvironment and clinical outcomes of lung adenocarcinoma. *J Oncol* (2021) 14(2021):9548648. doi: 10.1155/2021/9548648
- Jin Y, Lee WY, Toh ST, Tennakoon C, Toh HC, Chow PK, et al. Comprehensive analysis of transcriptome profiles in hepatocellular carcinoma. *J Trans Med* (2019) 17(1):273. doi: 10.1186/s12967-019-2025-x
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf* (2008) 9:559. doi: 10.1186/1471-2105-9-559
- Chen L, Yuan L, Qian K, Qian G, Zhu Y, Wu C, et al. Identification of biomarkers associated with pathological stage and prognosis of clear cell renal cell carcinoma by Co-expression network analysis. *Front Physiol* (2018) 9:399. doi: 10.3389/fphys.2018.00399
- Kessler SM, Simon Y, Gempferlein K, Gianmoena K, Cadenas C, Zimmer V, et al. Fatty acid elongation in non-alcoholic steatohepatitis and hepatocellular carcinoma. *Int J Mol Sci* (2014) 15(4):5762–73. doi: 10.3390/ijms15045762
- Matsuzaka T, Atsumi A, Matsumori R, Nie T, Shinozaki H, Suzuki-Kemuriyama N, et al. Elov6 promotes nonalcoholic steatohepatitis. *Hepatology* (2012) 56(6):2199–208. doi: 10.1002/hep.25932
- Muir K, Hazim A, He Y, Peyressatre M, Kim D, Song X, et al. Proteomic and lipidomic signatures of lipid metabolism in NASH-associated hepatocellular carcinoma. *Cancer Res* (2013) 73(15):4722–31. doi: 10.1158/0008-5472.CAN-12-3797
- Kim DW, Talati C, Kim R. Hepatocellular carcinoma (HCC): beyond sorafenib-chemotherapy. *J Gastrointestinal Oncol* (2017) 8(2):256–65. doi: 10.21037/jgo.2016.09.07
- Mao Y, Feng Q, Zheng P, Yang L, Liu T, Xu Y, et al. Low tumor purity is associated with poor prognosis, heavy mutation burden, and intense immune phenotype in colon cancer. *Cancer Manage Res* (2018) 10:3569–77. doi: 10.2147/CMAR.S171855
- Ge P, Wang W, Li L, Gao Z, Tang Z, Dang X, et al. Profiles of immune cell infiltration and immune-related genes in the tumor microenvironment of colorectal cancer. *Biomedicine pharmacotherapy = Biomedicine pharmacotherapie* (2019) 118:109228. doi: 10.1016/j.biopha.2019.109228
- Yun HR, Jo YH, Kim J, Nguyen NNY, Shin Y, Kim SS, et al. Palmitoyl protein thioesterase 1 is essential for myogenic autophagy of C2C12 skeletal myoblast. *Front Physiol* (2020) 11:569221. doi: 10.3389/fphys.2020.569221
- Zeidman R, Jackson CS, Magee AI. Protein acyl thioesterases (Review). *Mol Membr Biol* (2009) 26(1):32–41. doi: 10.1080/09687680802629329
- Rebecca VW, Nicastrì MC, Fennelly C, Chude CI, Barber-Rotenberg JS, Ronghe A, et al. PPT1 promotes tumor growth and is the molecular target of chloroquine derivatives in cancer. *Cancer Discovery* (2019) 9, no. 2:220–9. doi: 10.1158/2159-8290.CD-18-0706
- Cheng H, Wang M, Su J, Li Y, Long J, Chu J, et al. Lipid metabolism and cancer. *Life* (2022) 12(6):784. doi: 10.3390/life12060784
- Miranda-Gonçalves V, Lameirinhas A, Henrique R, Baltazar F, Jerónimo C. The metabolic landscape of urological cancers: new therapeutic perspectives. *Cancer Lett* (2020) 477:76–87. doi: 10.1016/j.canlet.2020.02.034
- Xu W, Liu WR, Xu Y, Tian X, Anwaier A, Su J, et al. Hexokinase 3 dysfunction promotes tumorigenesis and immune escape by upregulating monocyte/macrophage infiltration into the clear cell renal cell carcinoma microenvironment. *Int J Biol Sci* (2021) 17(9):2205–22. doi: 10.7150/ijbs.58295
- Tuo Z, Zheng X, Zong Y, Li J, Zou C, Lv Y, et al. HK3 is correlated with immune infiltrates and predicts response to immunotherapy in non-small cell lung cancer. *Clin Trans Med* (2020) 10(1):319–30. doi: 10.1002/ctm2.6
- Xue C, Li G, Lu J, Li L. Crosstalk between circRNAs and the PI3K/AKT signaling pathway in cancer progression. *Signal transduction targeted Ther* (2021) 6(1):400. doi: 10.1038/s41392-021-00788-w
- Lagier-Tourenne C, Tazir M, López LC, Quinzii CM, Assoum M, Drouot N, et al. ADCK3, an ancestral kinase, is mutated in a form of recessive ataxia associated with coenzyme Q10 deficiency. *Am J Hum Genet* (2008) 82(3):661–72. doi: 10.1016/j.ajhg.2007.12.024
- Walter L, Miyoshi H, Leverve X, Bernard P, Fontaine E. Regulation of the mitochondrial permeability transition pore by ubiquinone analogs. a progress report. *Free Radical Res* (2002) 36(4):405–12. doi: 10.1080/10715760290021252
- Devun F, Walter L, Belliere J, Cottet-Rousselle C, Leverve X, Fontaine E, et al. Ubiquinone analogs: a mitochondrial permeability transition pore-dependent pathway to selective cell death. *PLoS One* (2010) 5(7):e11792. doi: 10.1371/journal.pone.0011792
- Testai L, Martelli A, Flori L, Cicero AFG, Colletti A. Coenzyme Q(10): clinical applications beyond cardiovascular diseases. *Nutrients* (2021) 13(5):1697. doi: 10.3390/nu13051697
- Jiang N, Dai Q, Su X, Fu J, Feng X, Peng J, et al. Role of PI3K/AKT pathway in cancer: the framework of malignant behavior. *Mol Biol Rep* (2020) 47(6):4587–629. doi: 10.1007/s11033-020-05435-1
- Lin A, Gu T, Hu X, Zhang J, Luo P. Comprehensive analysis identifies PI3K/Akt pathway alternations as an immune-related prognostic biomarker in colon adenocarcinoma patients receiving immune checkpoint inhibitor treatment. *J Immunol Res* (2022) 2022:8179799. doi: 10.1155/2022/8179799