

Multi-omics analysis of pyroptosis regulation patterns and characterization of tumor microenvironment in patients with hepatocellular carcinoma (#80824)

1

First submission

Guidance from your Editor

Please submit by **22 Jan 2023** for the benefit of the authors (and your token reward) .



Structure and Criteria

Please read the 'Structure and Criteria' page for general guidance.



Custom checks

Make sure you include the custom checks shown below, in your review.



Raw data check

Review the raw data.



Image check

Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files

Download and review all files from the [materials page](#).

12 Figure file(s)

5 Table file(s)

1 Other file(s)

! Custom checks

Human participant/human tissue checks



Have you checked the authors [ethical approval statement](#)?



Does the study meet our [article requirements](#)?



Has identifiable info been removed from all files?



Were the experiments necessary and ethical?



Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. BASIC REPORTING
2. EXPERIMENTAL DESIGN
3. VALIDITY OF THE FINDINGS
4. General comments
5. Confidential notes to the editor

 You can also annotate this PDF and upload it as part of your review

When ready [submit online](#).

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your [guidance page](#).

BASIC REPORTING

-  Clear, unambiguous, professional English language used throughout.
-  Intro & background to show context. Literature well referenced & relevant.
-  Structure conforms to [PeerJ standards](#), discipline norm, or improved for clarity.
-  Figures are relevant, high quality, well labelled & described.
-  Raw data supplied (see [PeerJ policy](#)).

EXPERIMENTAL DESIGN

-  Original primary research within [Scope of the journal](#).
-  Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
-  Rigorous investigation performed to a high technical & ethical standard.
-  Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

-  Impact and novelty not assessed. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
-  All underlying data have been provided; they are robust, statistically sound, & controlled.
-  Conclusions are well stated, linked to original research question & limited to supporting results.



The best reviewers use these techniques

Tip

Example

Support criticisms with evidence from the text or from other sources

Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.

Give specific suggestions on how to improve the manuscript

Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).

Comment on language and grammar issues

The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult. I suggest you have a colleague who is proficient in English and familiar with the subject matter review your manuscript, or contact a professional editing service.

Organize by importance of the issues, and number your points

1. Your most important issue
2. The next most important item
3. ...
4. The least important points

Please provide constructive criticism, and avoid personal opinions

I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC

Comment on strengths (as well as weaknesses) of the manuscript

I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.

Multi-omics analysis of pyroptosis regulation patterns and characterization of tumor microenvironment in patients with hepatocellular carcinoma


Bingbing Shang¹, Ruohan Wang¹, Haiyan Qiao², Xixi Zhao¹, Liang Wang^{Corresp., 2}, Shaoguang Sui¹

¹ Emergency Department, The Second Affiliated Hospital, Dalian Medical University, Dalian, China

² Research and teaching department of comparative medicine, Dalian Medical University, Dalian, China

Corresponding Author: Liang Wang
Email address: wangliang-dy@dmu.edu.cn

Background. Hepatocellular carcinoma (HCC) is a very dangerous tumor to human health. Pyroptosis is a novel cell program that is involved in several diseases including cancer. However, the functional role of pyroptosis in HCC remains unclear. The purpose of this study was to establish a prediction model, explored the biological role of differentially expressed genes, found hub genes and provided targets for clinical treatment. **Methods.** The gene data and clinical related information of patients with Liver Hepatocellular Carcinoma (LIHC) were collected from The Cancer Genome Atlas (TCGA) database. After obtaining the differentially expressed genes (DEGs), they intersected with the genes related to pyroptosis. The patients were divided into high-risk and low-risk groups by risk score using univariate Cox regression analysis, and a risk prediction model was established to predict the overall survival (OS). Then we analyzed the biological characteristics of DEGs by drug sensitivity analysis, Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Set Enrichment Analysis (GSEA), Gene Set Variation Analysis (GSVA). Different immune cell infiltration and related pathways were analyzed, and obtained hub genes by protein-protein interaction (PPI). Finally, the expression of hub genes was verified by real-time quantitative PCR (qRT-PCR) and immunohistochemistry. **Results.** We identified a total of 8958 differentially expressed genes and obtained 37 differentially expressed genes related to pyroptosis after intersection with pyroptosis. We found that the risk score is an independent factor in predicting the prognosis of LIHC patients. Moreover, we constructed the prognosis model of OS and found that it has good prediction ability. We obtained the differences of biological function, drug sensitivity and immune microenvironment between high-risk group and low-risk group, through enrichment analysis, we found that the differentially expressed genes are different in a variety of biological processes, but this paper mainly discusses their role in the process of digestive system, the maintenance of epithelial cells and cell cycle, then ten hub genes

were obtained by PPI. **Conclusion.** We have established a relatively stable and reliable prediction model and identified hub genes to predict the prognosis of patients, which provides a direction for clinical research and treatment. 

Multi-omics Analysis of Pyroptosis Regulation Patterns and Characterization of Tumor Microenvironment in Patients with hepatocellular carcinoma

Bingbing Shang^{1*}, Ruohan Wang^{1*}, Haiyan Qiao^{2*}, Xixi Zhao¹, Liang Wang^{2#}, Shaoguang Sui^{1#}

¹Emergency Department, The Second Affiliated Hospital, Dalian Medical University, Dalian, 116027, Liaoning Province, China

²Research and teaching department of comparative medicine, Dalian Medical University, Dalian 116044, Dalian City, Liaoning province, China

* Bingbing Shang, Ruohan Wang and Haiyan Qiao equal contributions to this article. They are co-first authors.

#Corresponding Author: Shaoguang Sui, suishao Guang@163.com and Liang Wang, wangliang-dy@dmu.edu.cn

Keywords: hepatocellular carcinoma, multi-omics analysis, pyroptosis, immune cells infiltration

Abbreviation: Hepatocellular carcinoma (HCC), Liver Hepatocellular Carcinoma (LIHC), overall survival (OS), Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Set Enrichment Analysis (GSEA), Gene Set Variation Analysis (GSVA), protein-protein interaction (PPI), real-time quantitative PCR (qRT-PCR), hepatitis B virus (HBV), differential pyroptosis related genes (PRG), The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO), differentially expressed genes (DEGs), least absolute shrinkage and selection operator (LASSO), overall survival (OS), Molecular Signatures Database (MSigDB), Midkine (MDK).

Abstract

Background. Hepatocellular carcinoma (HCC) is a very dangerous tumor to human health. Pyroptosis is a novel cell program that is involved in several diseases including cancer. However, the functional role of pyroptosis in HCC remains unclear. The purpose of this study was to establish a prediction model, explored the biological role of differentially expressed genes, found hub genes and provided targets for clinical treatment.

Methods. The gene data and clinical related information of patients with Liver Hepatocellular Carcinoma (LIHC) were collected from The Cancer Genome Atlas (TCGA) database. After obtaining the differentially expressed genes (DEGs), they intersected with the genes related to pyroptosis. The patients were divided into high-risk and low-risk groups by risk score using univariate Cox regression analysis, and a risk prediction model was established to predict the overall survival (OS). Then we analyzed the biological characteristics of DEGs by drug sensitivity analysis, Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Set Enrichment Analysis (GSEA), Gene Set Variation Analysis (GSVA). Different immune cell infiltration and related pathways were analyzed, and obtained hub genes by protein-protein interaction (PPI). Finally, the expression of hub genes was verified by real-time quantitative PCR (qRT-PCR) and immunohistochemistry.

Results. We identified a total of 8958 differentially expressed genes and obtained 37 differentially expressed genes related to pyroptosis after intersection with pyroptosis. We found that the risk score is an independent factor in predicting the prognosis of LIHC patients. Moreover, we constructed the prognosis model of OS and found that it has good prediction ability. We obtained the differences of biological function, drug sensitivity and immune microenvironment between high-risk group and low-risk group, through enrichment analysis, we found that the differentially expressed genes are different in a variety of biological processes, but this paper mainly discusses their role in the process of digestive system, the maintenance of epithelial cells and cell cycle, then ten hub genes were obtained by PPI.

Conclusion. We have established a relatively stable and reliable prediction model and identified

hub genes to predict the prognosis of patients, which provides a direction for clinical research and treatment.

1. Introduction

Primary liver cancer is the third leading cause of cancer-related deaths and one of the most serious malignant tumors worldwide [1]. Genome sequencing has revealed that primary liver cancer is heterogeneous in both etiology and biology. As a heterogeneous disease, primary liver cancer comprises many molecular subtypes, which presents an imposing challenge for clinicians and researchers in the precision treatment of liver cancers. As the predominant type of primary liver cancer, HCC is a typical inflammation-associated cancer [2] and accounts for approximately 75%–85% of cases, followed by intrahepatic cholangiocarcinoma (10%–15%). HCC development is a complex pathological process involving chronic inflammation, liver injury, hepatocyte proliferation, and fibrosis [3]. Currently, the treatment of HCC is limited, and its prognosis is poor [4, 5]. Curative ablative therapies are only suitable for patients with early-stage HCC. Once at an advanced stage, patients with HCC are resistant to conventional chemotherapy and radiotherapy. In addition to tumor heterogeneity, the tumor microenvironment also plays an important role in patient staging and treatment evaluation. The clinicopathological characteristics of HCC are the main prognostic factors. Owing to the complex heterogeneous character of HCC, a single biomarker combined with clinicopathological staging is not sufficient to predict the prognosis of patients with HCC. Although HCC is considered as a typical immunogenic cancer, immunotherapy has been shown to have limited effectiveness [6, 7]. To date, there is no credible biomarker to predict the efficacy of immunotherapy for HCC, including PD-L1, which cannot predict the response to nivolumab and pembrolizumab [8, 9]. Therefore, it is particularly important to explore a combination of biomarkers and reliable prognostic models to optimize the medical decision-making. Pyrocytosis is an inflammatory process of cell death characterized by rapid plasma membrane rupture, DNA fragmentation, and the release of proinflammatory intracellular contents [10]. Pyroptosis is regulated by inflammatory caspases (caspase-1/4/5/11) and is closely associated

with inflammasome activation. It is mediated by the gasdermin superfamily, including GSDMA, GSDMB, GSDMC, GSDMD, and GSDME (or DFNA5) [11]. Among these, GSDMD can enhance antitumor immunity by activating pyroptosis [12]. Pyroptosis is involved in various pathological processes, including inflammatory disorders and multiple types of cancer [6, 13], and has been reported to play an important role in both the initiation and progression of HCC [14, 15]. Till date, not many studies have been conducted to investigate the effect of pyroptosis-related genes on HCC prognosis. Therefore, an in-depth and systematic study of pyroptosis-related genes and their correlation with the occurrence and development of HCC may help to elucidate the mechanism of HCC and provide novel insights for specific diagnosis, clinical prevention, and effective therapy. There have been many bioinformatics studies on liver cancer. For example, Xiao et al. [16] obtained prognostic genes of HCC via analysis of the immune and biological functions of differentially expressed pyroptosis-related genes and verified them using qRT-PCR *in vitro* and *in vivo*; however, the sample size was small. Nevertheless, the study provided a new research idea connecting genes related to pyroptosis with those related to liver cancer. Zhang et al. [17] selected biomarkers using TCGA, Gene Expression Omnibus (GEO), and logistic regression analysis; however, their analysis lacked verifiable survival information and clinical data. To bridge these gaps, we linked the genes related to pyroptosis with those related to liver cancer using an expanded sample size and comprehensively analyzed the associated differential genes and performed a more reliable gene analysis. In this study, we identified 37 differentially expressed pyroptosis-related genes by taking the intersection of DEGs and pyroptosis-related genes and explored their biological functions. Then, we constructed a risk prognostic model of HCC to explore its predictive ability, which was divided into low-risk and high-risk groups based on the calculated risk scores. To further analyze the biological function, drug sensitivity, and immune characteristics of DEGs, we explored the reasons affecting the grouping of different risks, and finally constructed PPI network to analyze the 10 key genes leading to the grouping difference.



2. Materia and methods

2.1 Data downloaded

Gene expression profile data and clinical information for HCC were downloaded from TCGA. The clinical data included age, sex, and survival status. We obtained 369 tumor and 50 normal tissue samples. The copy number variation data for HCC were downloaded simultaneously. In addition, GSE76426 [18] from the GEO database was analyzed as a validation dataset and included 115 cases of tumor tissue and 52 cases of normal tissue. The dataset was obtained using the GPL 10558 sequencing platform. All the gene sets were derived from *Homo sapiens*.

2.2 Differentially expressed genes of HCC

To analyze the influence of gene expression values on the occurrence and development of HCC, differential gene analysis was performed on normal and tumor samples from TCGA- LIHC dataset using the DESeq2 R package [19]. To define the threshold of DEGs, we applied the absolute value of $\log_2\text{Fold-change} > 2$ and $P_{\text{adj}} < 0.05$; cut-off values of $\log\text{FC} > 2$ and $P_{\text{adj}} < 0.05$ and $\log\text{FC} < -2$ and $P_{\text{adj}} < 0.05$ were used for upregulated and downregulated genes, respectively. To uncover the influence of pyroptosis-related genes on biological functions, these genes were extracted from the GOBP_PYROPTOSIS dataset of the Genecards [20] database and REACTOME_PYROPTOSIS dataset of the MSigDB [21] database. We then examined the intersection of pyroptosis-related genes and DEGs in HCC; the results were represented in the form of Venn diagrams. A volcano plot was constructed to show HCC-related and pyroptosis-related DEGs.

2.3 HCC risk model rebuilding

To analyze the effect of pyroptosis-related DEGs on the prognosis of patients with HCC, the expression and survival data of patients from TCGA-LIHC database were analyzed. Prognostic risk genes for HCC were evaluated using univariate Cox proportional hazards regression analysis. After the prognostic risk genes for HCC were included in the model, we used the least absolute

shrinkage and selection operator (LASSO) algorithm for dimension reduction and obtained prognostic feature genes. Each normalized gene expression value was weighted by penalty coefficients using LASSO-Cox analysis, and a risk score formula was established. Patients were then divided into high- and low-risk groups based on the mean risk score calculated as follows:

$$riskScore = \sum_i Coefficient (risk\ gene_i) * mRNA\ Expression (risk\ gene_i).$$

To verify the accuracy of the risk score, we used the GSE76427 dataset, which included patient gene expression and clinical data, to group patients and statistically analyze the differences in survival.

2.4 Clinical prediction model based on risk model

We used the risk score combined with clinicopathological characteristics to assess the personalized prognosis of patients. Univariate and multivariate Cox regression analyses of risk scores and clinicopathological characteristics were used to assess the predictive power of OS. Finally, the risk score model and important clinicopathological parameters were included to construct a clinical predictive nomogram model. To quantify the discriminative performance, we compared the nomogram-predicted probability with the observed actual survival rate and used a calibration curve to evaluate the nomogram performance.

2.5 Differentially expressed genes between the high-risk and low-risk groups

To analyze the effect of the risk score on HCC, DEG analysis of the samples that were divided into high- and low-risk groups from TCGA-LIHC dataset was performed using the R package DESeq2 [19]. Subsequently, the significant differential genes were screened. The threshold values to determine the DEGs were as follows: the absolute values of the log₂fold change (log₂FC) > 2 and Padj < 0.05 were defined as upregulated DEGs and log₂FC < -2 and Padj < 0.05 were defined as downregulated DEGs. The results were then represented as a volcano plot.

We obtained the dataset of HCC cell line-drug effects from the GDSC [22] database. The gene

expression data of high- and low-risk patients from TCGA-LIHC database were subjected to drug sensitivity analysis using the R package OncoPredict [23]. Subsequently, the sensitivity differences of the therapeutic drugs for HCC were compared between the high- and low-risk patients. In addition, to determine if the genes in the high- and low-risk patients with HCC showed copy number variation, GISTIC2.0 in Genepattern (<https://cloud.genepattern.org/>) [24] was used to analyze the copy number variation of patients with HCC from TCGA database.

2.6 Assessment of biological characteristics

We used GO, KEGG [25, 26], GSEA, and GSVA[27] to analyze the biological characteristics of patients of different risk groups. GO annotation analysis and KEGG pathway enrichment analysis were performed on the DEGs using the clusterProfiler package ($FDR < 0.05$) [28]. To investigate whether there were biological process differences between these two patient groups, we downloaded “c5.go.v7.4.entrez.gmt” and “c2.cp.kegg.v7.4.entrez.gmt” as reference gene sets from the Molecular Signatures Database (MSigDB) based on the gene expression profile data sets of patients with HCC. GSEA was performed, and the results were visualized using the clusterProfiler R package ($P < 0.05$). To study the variation in the biological processes of high-risk samples compared to that in those of low-risk samples, datasets of gene expression profiles based on TCGA-LIHC were used to perform GSVA using the GSVA package in R. The reference gene set (“h.all.v7.4. symbols.gmt”) was downloaded from MSigDB. We then calculated an enrichment score for each sample in each pathway and the association between the enrichment score for patients and the risk score ($P < 0.05$).

2.7 Identification and correlation analysis of tumor-infiltrating immune cells

Analysis of immune cell infiltrates in tissues plays an important role in the study of diseases and disease prognosis prediction. The stromal and immune cell contents of high- and low-risk samples from TCGA-LIHC dataset were estimated using the R package “estimate” [29]. The correlations between the risk score and ESTIMATE score were calculated. We evaluated the proportion of 22

immune cells in the immune microenvironment of high- and low-risk samples using the CIBERSORT algorithm [30]. We set the number of permutations to 1000 and used the Wilcoxon test to calculate the differences in proportions of immune cells between the high- and low-risk groups ($P < 0.05$).

2.8 Construction of Protein-protein interaction (PPI) network

We used the STRING [31] dataset to construct PPIs of differentially expressed pyroptosis-related genes with a combined score of >400 . Cytoscape (v3.7.0) was used for visualization of the PPI network, and Clue GO [32] was used for functional annotation. Next, we used the cytoHubba plug-in to obtain hub genes [33] using the MCC method in the PPI network.

2.9 RNA extraction and real-time quantitative PCR assay

Total RNA was extracted from liver cancer and adjacent tissues (from the Second Affiliated Hospital of Dalian Medical University) using the M5 HiPer Universal RNA Mini Kit (Mei5 Biotechnology Co. Ltd, Beijing, China). We then used the PerfectStart® Uni RT&qPCR Kit (TransGen Biotech, Beijing, China) for reverse transcription into cDNA and performed qRT-PCR. The primers were as follows: midkine (MDK) forward: 5'-GCAACTGGAAGAAGGAGT-3', and MDK reverse: 5'-TGGCACTGAGCATTGTAG-3'.

2.11 Immunohistochemistry

We performed dewaxing with xylene, gradient dehydration with alcohol, and the inactivation of endogenous peroxidase with hydrogen peroxide. The sections were then blocked and treated with primary antibodies against MDK (anti-Midkine, Abcam, Cambridge, UK). The antibody was diluted with PBS at a ratio of 1:600, and then incubated for 12 h at 4 °C. To each film, 100 µL of working fluid (PV-9000, Beijing Zhongshan Golgen Bridge Biological Technology Co. Ltd, Beijing, China) was added and then incubated at 37 °C for 20 min. The solution was dehydrated again after dyeing, and the film was sealed.

2.12 Statistic analysis

All data were processed and analyzed using R software (version 4.1.1). Continuous variables were compared between the two groups. For normally distributed variables, we used independent *t*-tests; for non-normally distributed variables, we used the Mann–Whitney rank-sum test. Correlations between different genes were analyzed using the Pearson correlation coefficient calculation. Survival analysis was performed using the R survival package. Survival differences are shown using Kaplan–Meier survival curves. The significance of the difference in survival between the two groups was assessed using the log-rank test. Univariate and multivariate Cox regression analyses were used to identify independent prognostic factors. All statistical *P*-values were two-sided, and $P < 0.05$ was considered statistically significant.

3.Results

3.1 Differentially expressed pyroptosis-regulated factors and the effect on their biological processes

To reveal the biological differences between HCC and normal samples from transcriptomes, we first performed differential gene expression analysis in HCC and normal samples. After screening, 8958 genes were defined as significant DEGs, comprising 7030 upregulated and 1928 downregulated genes (Figure 2A). In total, 6866 pyroptosis-related genes were obtained from the GeneCards database. The intersection of DEGs and pyroptosis-related genes revealed 37 differentially expressed pyroptosis-related genes (22 upregulated and 16 downregulated; Figure 2B, 2C).

Functional enrichment was performed for differentially expressed pyroptosis-related genes (Figure 2D). The results showed that the differential genes were related to biological processes (Figure 2E) including digestion, maintenance of gastrointestinal epithelium, digestive system process, epithelial structure maintenance, and O-glycan processing; cellular components (Figure 2F)

including extracellular space, extracellular region, endosome membrane, and plasma membrane; and molecular functions including interleukin-1 receptor binding, lipoteichoic acid binding, G-protein coupled adenosine receptor activity, lipopolysaccharide binding, and identical protein binding (Figure 2G). Furthermore, KEGG pathway enrichment analysis indicated that the differentially expressed pyroptosis-related genes were significantly enriched in the pertussis, *Salmonella* infection, toll-like receptor signaling, and tuberculosis pathways (Figure 2H).

3.2 Construction of risk models and prognostic analysis

To analyze the effects of differentially expressed pyroptosis-related genes on the prognosis of patients with HCC, we used univariate Cox regression analysis and identified 45 prognostic risk genes of HCC. The prognosis-associated risk genes of HCC were included in the LASSO-Cox analysis to choose 12 genes with the best prognostic value. We subsequently performed a correlation analysis of the expression of the 12 characteristic genes, and the results showed a high level of correlation between these genes (Figure 3A). The penalty coefficients of characteristic genes were calculated based on LASSO-Cox analysis (ADRA1A: -0.023, GPRC6A: 0.037, MAGEA8: 0.034, MDK: 0.024, MMP9: 0.019, NETO2: 0.027, OLFM3: 0.031, PYCR1: 0.020, RET: 0.029, and TNFRSF4: 0.065). The risk score was calculated by multiplying the expression level of a gene with its corresponding coefficient and then adding them together. Patients with HCC were then divided into high- and low-risk groups based on their mean risk scores. In addition, principal component analysis revealed that the 12 characteristic genes could better distinguish patients in both TCGA-LIHC and GSE76427 datasets (Figure 3B, 3C). Kaplan–Meier analysis revealed worse OS scores in patients with high risk scores (Log-rank $P < 0.0001$, Figure 3D). The risk score of each patient with HCC in GSE76427 was then computed, and the Kaplan–Meier analysis revealed that patients with a high-risk score showed worse OS (Log-rank $P < 0.05$, Figure 3E).

3.3 Construction of clinical prediction model based on risk score

We evaluated the prognostic impact of the risk score in patients with HCC. Univariate and multivariate Cox analyses revealed that the risk score was an independent prognostic factor for patients with HCC (Figure 4G, 4H). Different clinicopathological features were combined into a risk score to generate a prediction model nomogram for predicting the OS of patients with HCC (Figure 4A). The calibration curve showed a good prediction of 1-, 2-, and 3-year survival for patients using the model (Figure 4B–D). Concurrently, the results of time-dependent receiver operating characteristic curve analysis showed that the predicted percentages of 1-, 3-, and 5-year survival were 77.8%, 79.8%, and 82.4%, respectively (Figure 4E). The decision curve analysis supported the model, which could offer greater prognostic gains for patients (Figure 4F).

3.4 Divergence analysis between high-risk group and low-risk group

Owing to the clear distinction in survival between the high- and low-risk groups, we performed differential expression analysis of the expressed genes in the high- and low-risk groups. In total, 349 DEGs were identified. Of these, 51 genes were upregulated and 298 were downregulated (Figure 5B). Concurrently, the DEGs were subjected to dimensionality reduction using principal component analysis. The results showed a significant difference between the high- and low-risk groups (Figure 5A). We then analyzed the effect of differentially expressed mRNA on the biological functions associated with the high- and low-risk groups. GO biological function annotation analysis of the DEGs showed that these genes were mainly enriched in digestion, maintenance of gastrointestinal epithelium, digestive system process, epithelial structure maintenance, O-glycan processing, and other biological processes (Figure 5C); catenin complex, GABAergic synapse, Golgi lumen, dendrite membrane, apical part of cell, and other cellular components (Figure 5D); and ligand-gated anion channel activity, chloride channel activity, hormone activity, anion channel activity, and other molecular functions (Figure 5E). Additionally, KEGG pathway enrichment analysis of the DEGs showed that these genes were mainly enriched in pathways associated with neuroactive ligand-receptor interaction, taste transduction, pancreatic secretion, taurine and hypotaurine metabolism, and the cAMP signaling pathway (Figure 5F). We

also analyzed the copy number differences between the high- and low-risk groups. The results revealed that the copy number amplification of patients assigned to the high-risk group was significantly lower than that of those assigned to the low-risk group, and the high-risk group included more chromosomal deletions (Figure 6).

To analyze the influence of risk score on the treatment of HCC, we integrated the gene expression data of high- and low-risk patients in TCGA-LIHC dataset with drug sensitivity for HCC from the GDSC database. We identified drugs that were more sensitive in high- and low-risk patients and had lower IC_{50} values. Among the drugs currently used for the treatment of HCC, IGFR_3801, tanespimycin, docetax, and lestaurtinib were more sensitive in high-risk patients (Figure 7A), whereas belinostat, ispinesib mesylate, shikonin, and midostaurin were more sensitive in low-risk patients (Figure 7B). Among the currently used drugs for other cancers, trametinib, mitomycin-C, rapamycin and obatoclast mesylate were more sensitive in high-risk patients (Figure 7C), whereas pevonedistat_1529, methotrexate_1008, foretinib_308, and AZD8055_1059 were more sensitive in low-risk patients (Figure 7D).

Next, we performed GSEA of the gene expression data of the high- and low-risk groups. We found that the patients in the high- and low-risk groups showed significant differences mainly in the following biological processes: inhibitory extracellular ligand-gated ion channel activity, limb bud formation, condensed chromosome outer kinetochore, spindle elongation, regulation of chloride transport, and other processes were inhibited, whereas very low-density lipoprotein particle remodeling, triglyceride-rich lipoprotein particle remodeling, complement activation lectin pathway, platelet dense granule lumen, alcohol dehydrogenase were activated (Figure 8A, 8C). Simultaneously, we found that primary bile acid biosynthesis, fatty acid metabolism, complement and coagulation cascades, glycine serine and threonine metabolism, renin angiotensin system, and other pathways were activated, whereas some pathways, such as cell cycle, ECM receptor interaction, neuroactive ligand receptor interaction, gap junction, and axon guidance, remained inactivated (Figure 8B, D).

We further used GSVA to calculate the DEGs of high- and low-risk patients from TCGA-LIHC database to determine their effect on biological characteristics and oncogenic signaling pathways. The results revealed that the risk score was closely correlated with the activity of multiple oncogenic signaling pathways (Figure 8E). Among them, there was a significant negative correlation between risk score and Hallmark KRAS Signaling DN, Hallmark Pancreas Beta Cells, Hallmark Coagulation, Hallmark Xenobiotic Metabolism, and Hallmark Bile Acid Metabolism, whereas there was a significant positive correlation between risk score and Hallmark G2M Checkpoint, Hallmark E2F Targets, Hallmark MYC Targets V2, Hallmark DNA Repair, and Hallmark MYC Targets V1.

3.5 Immune characterize differences in high-risk and low-risk groups

We analyzed the differences in immune cell content in the high- and low-risk groups (Figure 9A), and the high- and low-risk groups showed significant differences in multiple immune cells (Figure 9B). The risk score was positively correlated with B cell memory, dendritic cell resting, non-activated macrophages, and regulatory T cells (Tregs), and negatively correlated with resting mast cells, monocytes, and gamma delta T cells ($P < 0.05$, Figure 9C). Additionally, we calculated the correlation between immune cells and patients in the high- and low-risk groups. We found that there was a positive correlation between memory CD4 T cells and resting dendritic cells in patients in the high-risk group, while there was a negative correlation between memory B cell, resting dendritic cells, neutrophils, Tregs, and non-activated macrophages in patients in the low-risk group (Figure 9E).

We also analyzed the correlations between the risk score of HCC and the immune score, stromal score, tumor purity, and ESTIMATE score, among which there was a significant positive correlation between the risk score and stromal score ($P < 0.05$, Figure 10A). We constructed a PPI network of DEGs in patients in the high- and low-risk groups (Figure 10B). Visualization of the PPI network was carried out with Cytoscape, which included 159 DEGs and 272 PPI pairs in total. The top five genes that interacted closely with other DEGs were *MUC5AC* (interacted with 16

DEGs), *MUC5B* (interacted with 16 DEGs), *MUC1* (interacted with 15 DEGs), *CFTR* (interacted with 14 DEGs), and *CHGA* (interacted with 13 DEGs). In addition, DEGs related to the PPI network were mainly enriched in functions related to calcium-ion regulated exocytosis, digestive system processes, and cochlear development (Figure 10C). The hub genes of the PPI network were screened using the cytoHubba plugin, and these genes are more likely to be the key genes causing differences between the high- and low-risk groups (Figure 10D).




3.6 High expression of MDK in hepatocellular carcinoma



From the RT-qPCR results (Figure 11), it was clear that the expression of *MDK* in HCC is much higher than that in adjacent tissues. The immunohistochemical analysis (Figure 12) also showed that MDK was highly expressed in HCC tissues compared to that in normal liver tissues.

Discussion

Recently, the number of patients with liver cancer has increased. HCC has gradually become one of the main causes of cancer-related deaths. In China alone, the 5-year survival rate of HCC patients is less than 13% [34]. The pathogenesis of HCC and its specific mechanism are still under study [35], and the early diagnosis and prognosis of liver cancer are also widely studied. The diagnosis and prognosis of patients with HCC are mainly judged from the aspects of pathology and the Barcelona Clinic Liver Cancer staging [36]; however, these diagnostic and prognostic methods still lack sensitivity. Accurately predicting the OS of patients with HCC is important for clinical decision-making. However, there are currently no effective and reliable prognostic biomarkers for patients with HCC. Thus, establishing a robust prediction model and identifying effective biomarkers to predict the outcomes of patients with HCC are urgently needed. Therefore, we screened the main markers, such as MDK, using bioinformatics analysis and verified the degree of expression of related genes in HCC using qRT-PCR and immunohistochemistry; with this method, research on the relationship between pyroptosis and HCC occurrence can guide clinical treatment.

We obtained the genes with the best prognostic value using single-factor Cox regression analysis and LASSO Cox analysis, among which we conducted a more in-depth study on the *MDK* gene. *MDK*, also known as neurite growth-promoting factor 2 (NEGF2), is a small heparin-binding growth factor with a molecular weight of approximately 13 kDa. The human *MDK* gene is located on chromosome 11p11.2 (57) [37, 38]. *MDK* is involved in nervous system development during embryonic development. In addition, *MDK* is involved in inflammatory reactions, which are related to autoimmune diseases and cancer. An increasing number of studies have shown that *MDK* plays important roles in many tumors, such as proliferation, promotion of vascular growth, and anti-apoptosis [39-41]. It has been reported that *MDK* also plays a role in the occurrence, development, metastasis, and prognosis of liver cancer [41]. We showed by PCR that *MDK* is highly expressed in HCC cells, which suggests that *MDK* has a potential to become a new therapeutic target in the near future, providing a new method of clinical treatment. Alpha-fetoprotein (AFP) plays a key role in the early diagnosis of HCC, but studies have found that the sensitivity of AFP in serum is less than 70% and the specificity in HCC is more than 90% [42]. However, some studies using meta-analysis have found that the sensitivity of *MDK* for the diagnosis of HCC could reach 85% through meta-analysis [43], and other studies have shown that the sensitivity can reach more than 80% in the early stage of liver cancer and AFP-negative liver cancer. In addition, many liver cancers are associated with HBV and HCV infections, and *MDK* has been suggested to have an effect on hepatitis-associated HCC [44]. In addition to its role in early diagnosis, high *MDK* expression can indicate poor prognosis and the possibility of recurrence [45]. Not only can *MDK* be used as a new biomarker of HCC and improve its early diagnosis and diagnostic accuracy, but it can also serve as a target for clinical treatment. Through this study, we found that the OS of the high-risk group was relatively worse than that of the low-risk group. To further explore the differences between the high- and low-risk groups, we analyzed the DEGs between these groups. We obtained hub genes by constructing PPIs of differentially expressed pyroptosis-related genes. These hub genes play a key role in the risk grouping of liver cancer. The hub genes were *SST*, *KRT19*, *CRHR1*, *GRM5*, *KRT20*, *AGR2*, *MUC6*,

TFF1, *TFF2*, *CLCA1*, *CFTR*, *CHGA*, *MUC5A*, *TAC1*, *MUC1*, *MUC5B*, *GALNT12*, *FOXJ1*, and *SCGB1A1*.  the PPI of differentially expressed pyroptosis-related genes, we found that MUC5AC interacted the most with other DEGs. MUC5AC is a mucin, which is a product of secretory cells. *MUC5AC* is located on the chromosome 11 (11p15) [46]. It is the main component of airway mucus, is also expressed in intestinal epithelial cells, and plays a protective role. In recent years, it has been found to be abnormally expressed in various diseases, such as asthma and malignant mucinous lung tumor cells. It can also be used as an indicator of the prognosis of colorectal cancer [47-49]. In the past, researchers believed that HCC does not produce mucin [50]. Similarly, many scholars have shown that *MUC5AC* is expressed in cholangiocarcinoma (CCA), is related to its prognosis, and can be used as an indicator of CCA in early screening [51]. Currently, it is known that HCC also produces MUC5AC, which confirms that it is expressed in all tumors [51-53]. Our study also proves that *MUC5AC* plays an important role in HCC risk, but its potential mechanism requires further research. In the future, further exploring the biological mechanism of *MUC5AC* in HCC is likely to provide a potential target for clinical treatment and prognosis judgment.  *MUC1* is also a widely recognized oncogene. MUC1 is a one-way type I transmembrane protein that is mostly expressed in epithelial cells such as those in the stomach and colorectal and respiratory tracts. It has the same protective role as MUC5AC [54]. MUC1 plays a key role in a variety of epithelial cancers through abnormal glycosylation and overexpression. MUC1 has different functions in normal and abnormal cells [55] and is overexpressed in HCC. Its mechanism is to promote the occurrence and development of tumors via the JNK/TGF signaling pathway [52], which is also consistent with our conclusion. , suggesting that MUC1 may be a potential prognostic marker and a therapeutic target for HCC.

We performed GO and KEGG analyses and GSEA to analyze the DEGs.  We found that they are mainly enriched in biological processes such as digestive system processes, epithelial structure maintenance, and O-glycan processing. TFF1 is a member of the TFF family of proteins.  The TFF family gene is located on the chromosomes 21q22.3. The TFF family includes *TFF1*, *TFF2*, and *TFF3* [56]. TFF1 is mainly expressed in digestive tract epithelial cells [57], and is mainly involved

in the protection of the gastrointestinal tract via the repair of epithelial cells; it is highly expressed in gastric and colorectal cancer [50]. Yosuke et al. found that TFF1 has a tumor inhibitory effect. It could inhibit the proliferation of liver cancer cells and induce apoptosis and inhibit the cell cycle by negatively regulating the β -catenin signaling pathway [54, 58]. At present, research on the involvement of TFF1 in HCC progression is limited, and the specific mechanism needs to be further studied. TFF2 is mainly expressed in the mucous cells of the digestive tract and plays a protective role in promoting intestinal epithelial cell migration. TFF2 is expressed at low levels in gastric cancer and is related to methylation. Moreover, TFF2 plays an inhibitory role in breast and pancreatic cancer [59-61]. Overall, our study provides novel clinical ideas for the future treatment of HCC.

This study had some limitations. First, our data were collected from TCGA, which lacks relevant clinical data and its analysis. Moreover, we applied multiple datasets in this study, and there may have been batch differences. Therefore, the statistical power may be low. Second, based on bioinformatics analysis, we identified the hub genes that are involved in inducing the difference between high- and low-risk groups of patients with HCC; however, we still need experimental evidence, such as western blotting and immunohistochemistry, to clarify the role of hub genes at the molecular and tissue levels.

This study established a risk prediction model by exploring the role of differentially expressed pyroptosis-related genes in the prognosis of HCC. Combined with clinical information, we analyzed the biological and immune characteristics and PPI of the DEGs to comprehensively explore the pathogenesis of HCC. This study improves our understanding of the molecular mechanisms of HCC, and the hub genes are potential targets for HCC treatment and provide new ideas for clinical treatment. Nevertheless, the specific pathogenesis and molecular targets of HCC require further verification by prospective studies.

Conclusions

The workflow of this study is shown in Figure 1. Via the analysis of DEGs related to pyroptosis

and LASSO Cox analysis, we calculated the risk score and established an independent prognosis model related to the risk score and clinicopathological features. This model can more accurately predict and judge the 1-, 3-, and 5-year survival rates of patients with HCC and confirm its reliability. These findings are helpful for determining the prognosis of patients with HCC. At the same time, further research on predictive biomarkers would help to explore the pathogenesis of HCC. Using experiments, we verified the high expression of MDK, a key gene in liver cancer, in liver cancer cells, thereby, providing a biomarker for the treatment of liver cancer.

Acknowledgments

None.

Ethics statement

All patient samples were approved by the Ethics Committee of the Second Affiliated Hospital of Dalian Medical University. No.134

Declarations of interest

None.

Authors'contributions

Bingbing Shang and Ruohan Wang jointly wrote the manuscript, Haiyan Qiao completes the experiment, Xixi Zhao finished proofreading the manuscript, Liang Wang and Shaoguang Sui provided ideas for the article.

References

1. Sung, H., 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): p. 209-249.
2. Zheng, X., et al., *IL-21 receptor signaling is essential for control of hepatocellular carcinoma growth and immunological memory for tumor challenge*. Oncoimmunology, 2018. **7**(12): p. e1500673.

3. Krishna-Subramanian, S., et al., *RIPK1 and death receptor signaling drive biliary damage and early liver tumorigenesis in mice with chronic hepatobiliary injury*. Cell Death Differ, 2019. **26**(12): p. 2710-2726.
4. Bollard, J., et al., *Palbociclib (PD-0332991), a selective CDK4/6 inhibitor, restricts tumour growth in preclinical models of hepatocellular carcinoma*. Gut, 2017. **66**(7): p. 1286-1296.
5. Ardelt, M.A., et al., *Inhibition of Cyclin-Dependent Kinase 5: A Strategy to Improve Sorafenib Response in Hepatocellular Carcinoma Therapy*. Hepatology, 2019. **69**(1): p. 376-393.
6. Yu, S.J., et al., *Targeting the crosstalk between cytokine-induced killer cells and myeloid-derived suppressor cells in hepatocellular carcinoma*. J Hepatol, 2019. **70**(3): p. 449-457.
7. Zhao, Y., et al., *Development of a new patient-derived xenograft humanised mouse model to study human-specific tumour microenvironment and immunotherapy*. Gut, 2018. **67**(10): p. 1845-1854.
8. Pinter, M., B. Scheiner, and M. Peck-Radosavljevic, *Immunotherapy for advanced hepatocellular carcinoma: a focus on special subgroups*. Gut, 2021. **70**(1): p. 204-214.
9. Yau, T., et al., *Nivolumab versus sorafenib in advanced hepatocellular carcinoma (CheckMate 459): a randomised, multicentre, open-label, phase 3 trial*. Lancet Oncol, 2022. **23**(1): p. 77-90.
10. Valderrama, J.A., et al., *Group A streptococcal M protein activates the NLRP3 inflammasome*. Nat Microbiol, 2017. **2**(10): p. 1425-1434.
11. Rogers, C., et al., *Gasdermin pores permeabilize mitochondria to augment caspase-3 activation during apoptosis and inflammasome activation*. Nat Commun, 2019. **10**(1): p. 1689.
12. Zhang, Z., et al., *Gasdermin E suppresses tumour growth by activating anti-tumour immunity*. Nature, 2020. **579**(7799): p. 415-420.
13. Orning, P., et al., *Pathogen blockade of TAK1 triggers caspase-8-dependent cleavage of gasdermin D and cell death*. Science, 2018. **362**(6418): p. 1064-1069.
14. Gaul, S., et al., *Hepatocyte pyroptosis and release of inflammasome particles induce stellate cell activation and liver fibrosis*. J Hepatol, 2021. **74**(1): p. 156-167.
15. Huo, J., et al., *A Ferroptosis and Pyroptosis Molecular Subtype-Related Signature Applicable for Prognosis and Immune Microenvironment Estimation in Hepatocellular Carcinoma*. Front Cell Dev Biol, 2021. **9**: p. 761839.
16. Fu, X.W. and C.Q. Song, *Identification and Validation of Pyroptosis-Related Gene Signature to Predict Prognosis and Reveal Immune Infiltration in Hepatocellular Carcinoma*. Front Cell Dev Biol, 2021. **9**: p. 748039.
17. Zhang, R.K. and J.L. Liu, *Screening the genome for HCC-specific CpG methylation signatures as biomarkers for diagnosis and prognosis evaluation*. BMC Med Genomics, 2021. **14**(1): p. 163.
18. Grinchuk, O.V., 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): p. 89-113.
19. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2*. Genome Biol, 2014. **15**(12): p. 550.
20. Safran, M., et al., *GeneCards Version 3: the human gene integrator*. Database (Oxford), 2010. **2010**: p. baq020.
21. Liberzon, A., et al., *The Molecular Signatures Database (MSigDB) hallmark gene set collection*. Cell Syst, 2015. **1**(6): p. 417-425.
22. Yang, W., et al., *Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells*. Nucleic Acids Res, 2013. **41**(Database issue): p. D955-61.
23. Maeser, D., R.F. Gruener, and R.S. Huang, *oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data*. Brief Bioinform, 2021. **22**(6).
24. Reich, M., et al., *GenePattern 2.0*. Nat Genet, 2006. **38**(5): p. 500-1.
25. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium*. Nat Genet, 2000. **25**(1): p. 25-9.
26. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes*. Nucleic Acids Res, 2000. **28**(1): p. 27-30.
27. Hänzelmann, S., R. Castelo, and J. Guinney, *GSVA: gene set variation analysis for microarray and RNA-seq data*. BMC Bioinformatics, 2013. **14**: p. 7.
28. Yu, G., et al., *clusterProfiler: an R package for comparing biological themes among gene clusters*. Omics, 2012. **16**(5): p. 284-7.
29. Yoshihara, K., et al., *Inferring tumour purity and stromal and immune cell admixture from expression data*. Nat Commun, 2013. **4**: p. 2612.
30. Newman, A.M., et al., *Determining cell type abundance and expression from bulk tissues with digital cytometry*. Nat Biotechnol, 2019. **37**(7): p. 773-782.
31. Szklarczyk, 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): p. D607-d613.
32. Bindea, G., et al., *ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks*. Bioinformatics, 2009. **25**(8): p. 1091-3.
33. Chin, C.H., et al., *cytoHubba: identifying hub objects and sub-networks from complex interactome*. BMC Syst Biol, 2014. **8 Suppl 4**(Suppl 4): p. S11.
34. Wang, W. and C. Wei, *Advances in the early diagnosis of hepatocellular carcinoma*. Genes Dis, 2020. **7**(3): p. 308-319.
35. Ye, J., et al., *Eukaryotic Initiation Factor 4A-3: A Review of Its Physiological Role and Involvement in Oncogenesis*. Front Oncol, 2021. **11**: p. 712045.
36. Bruix, J., M. Reig, and M. Sherman, *Evidence-Based Diagnosis, Staging, and Treatment of Patients With Hepatocellular Carcinoma*. Gastroenterology, 2016. **150**(4): p. 835-53.
37. Zhu, W.W., et al., *Evaluation of midkine as a diagnostic serum biomarker in hepatocellular carcinoma*. Clin Cancer Res, 2013. **19**(14): p. 3944-54.

38. Gowhari Shabgah, A., et al., *Shedding more light on the role of Midkine in hepatocellular carcinoma: New perspectives on diagnosis and therapy*. IUBMB Life, 2021. **73**(4): p. 659-669.
39. Du, Q., et al., *Dietary cholesterol promotes AOM-induced colorectal cancer through activating the NLRP3 inflammasome*. Biochem Pharmacol, 2016. **105**: p. 42-54.
40. Kato, M., et al., *Increased midkine expression in hepatocellular carcinoma*. Arch Pathol Lab Med, 2000. **124**(6): p. 848-52.
41. Muramatsu, T., *Midkine, a heparin-binding cytokine with multiple roles in development, repair and diseases*. Proc Jpn Acad Ser B Phys Biol Sci, 2010. **86**(4): p. 410-25.
42. Daniele, B., et al., *Alpha-fetoprotein and ultrasonography screening for hepatocellular carcinoma*. Gastroenterology, 2004. **127**(5 Suppl 1): p. S108-12.
43. Zhang, B.H., et al., *Diagnostic accuracy of midkine on hepatocellular carcinoma: A meta-analysis*. PLoS One, 2019. **14**(10): p. e0223514.
44. Lu, Q., et al., *Comparison of diagnostic accuracy of Midkine and AFP for detecting hepatocellular carcinoma: a systematic review and meta-analysis*. Biosci Rep, 2020. **40**(3).
45. Hodeib, H., et al., *Serum Midkine and Osteopontin Levels as Diagnostic Biomarkers of Hepatocellular Carcinoma*. Electron Physician, 2017. **9**(1): p. 3492-3498.
46. <gendler2001.pdf>.
47. Bonser, L.R. and D.J. Erle, *Airway Mucus and Asthma: The Role of MUC5AC and MUC5B*. J Clin Med, 2017. **6**(12).
48. Lin, Y.H., et al., *Resveratrol inhibits MUC5AC expression by regulating SPDEF in lung cancer cells*. Phytomedicine, 2021. **89**: p. 153601.
49. Hazgui, M., et al., *MUC1 and MUC5AC implication in Tunisian colorectal cancer patients*. Turk J Med Sci, 2021. **51**(1): p. 309-318.
50. Vocka, M., et al., *Trefoil factor family (TFF) proteins as potential serum biomarkers in patients with metastatic colorectal cancer*. Neoplasma, 2015. **62**(3): p. 470-7.
51. Pabalan, N., et al., *Expression and Serum Levels of Mucin 5AC (MUC5AC) as a Biomarker for Cholangiocarcinoma: a Meta-analysis*. J Gastrointest Cancer, 2019. **50**(1): p. 54-61.
52. Wang, J., et al., *Targeting MUC1 and JNK by RNA interference and inhibitor inhibit the development of hepatocellular carcinoma*. Cancer Sci, 2017. **108**(3): p. 504-511.
53. Xuan, J., et al., *The diagnostic performance of serum MUC5AC for cholangiocarcinoma: A systematic review and meta-analysis*. Medicine (Baltimore), 2016. **95**(24): p. e3513.
54. Ochiai, Y., et al., *Trefoil Factor Family 1 Inhibits the Development of Hepatocellular Carcinoma by Regulating beta-Catenin Activation*. Hepatology, 2020. **72**(2): p. 503-517.
55. Nath, S. and P. Mukherjee, *MUC1: a multifaceted oncoprotein with a key role in cancer progression*. Trends Mol Med, 2014. **20**(6): p. 332-42.
56. Chinery, R., J. Williamson, and R. Poulson, *The gene encoding human intestinal trefoil factor (TFF3) is located on chromosome 21q22.3 clustered with other members of the trefoil peptide family*. Genomics, 1996. **32**(2): p. 281-4.
57. Zhang, Y., et al., *The expression and role of trefoil factors in human tumors*. Transl Cancer

- Res, 2019. **8**(4): p. 1609-1617.
58. Hoffmann, W., *Trefoil Factor Family (TFF) Peptides and Chemokine Receptors: A Promising Relationship*. Journal of Medicinal Chemistry, 2009. **52**(21): p. 6505-6510.
59. Ge, Y., et al., *MUC1 is associated with TFF2 methylation in gastric cancer*. Clin Epigenetics, 2020. **12**(1): p. 37.
60. Yamaguchi, J., et al., *Loss of Trefoil Factor 2 From Pancreatic Duct Glands Promotes Formation of Intraductal Papillary Mucinous Neoplasms in Mice*. Gastroenterology, 2016. **151**(6): p. 1232-1244.e10.
61. Ishibashi, Y., et al., *Serum TFF1 and TFF3 but not TFF2 are higher in women with breast cancer than in women without breast cancer*. Sci Rep, 2017. **7**(1): p. 4846.

Figure 1

Work flow chart.

We took the intersection between 8958 DEGs and 197 pyroptosis-related genes and got 37 pyroptosis-DEGs. The best prognostic gene was selected and the risk score was calculated, The best prognostic gene was selected and the risk score was calculated, then it is verified by GSE76427 and divided into low-risk group and high-risk group, 349 differentially expressed genes were identified from the two groups, the differential genes were functional annotation and immune analysis, finally the prognosis model was established.

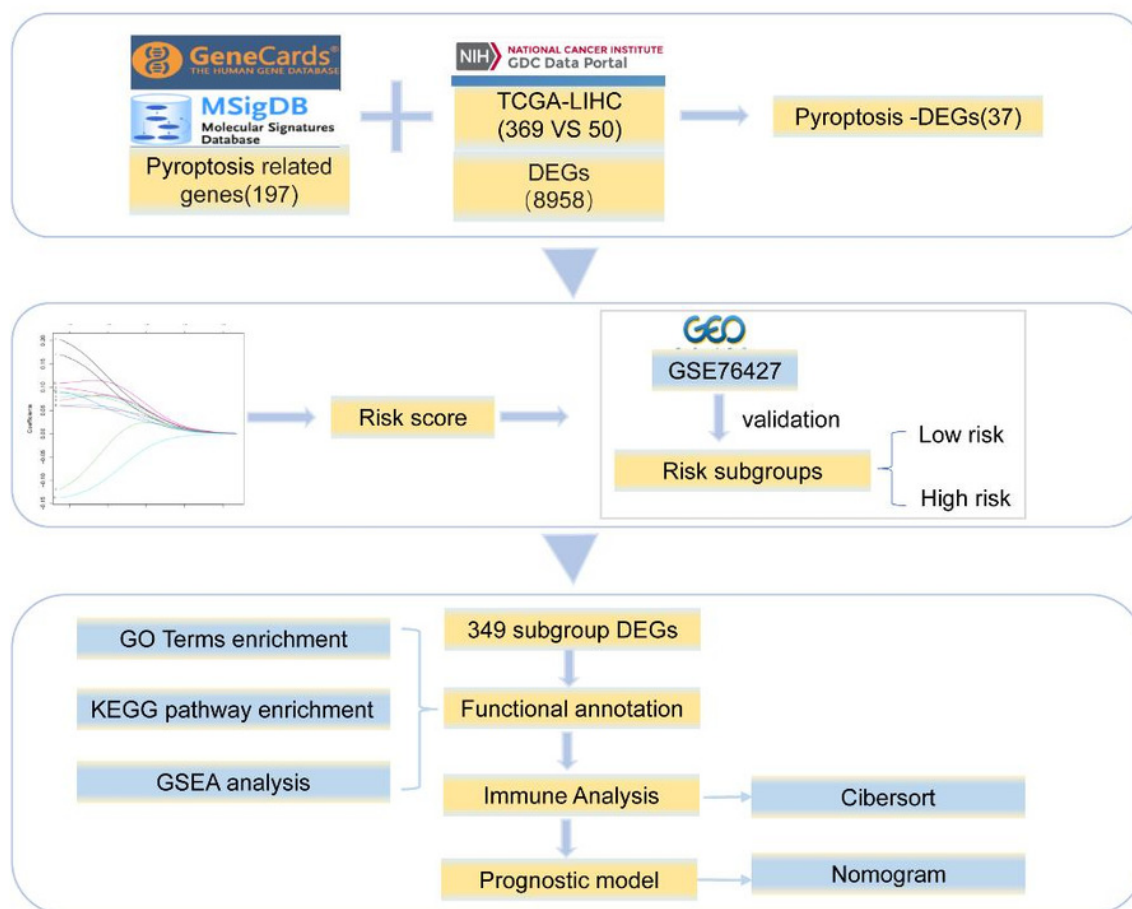


Figure 2

Differentially expressed genes and their biological functions.

Fig.A-B: The volcanic map shows differentially expressed genes and differentially expressed pyroptosis-related genes. The abscissa is $\log_2\text{FoldChange}$ and the ordinate is $-\log_{10}(\text{Adjust P-value})$. The up-regulated differentially expressed genes are represented by red nodes, the down-regulated differentially expressed genes are represented by blue nodes, and the genes with insignificant differentially expressed genes are represented by black nodes. FigC: A venn diagram of differentially expressed genes and pyroptosis-related genes. The blue circle indicates the pyroptosis gene, and the orange circle indicates the differentially expressed genes in the TCGA data set. FigD: The figure shows the Go of differentially expressed pyroptosis-related genes. The color of the bar graph indicates the zscore of GO term, and the size of zscore indicates the activation or suppression of GO term. FigE-G: These photos represent the results of BP, CC and MF in the GO of differentially expressed pyroptosis-related genes. FigH: The results of KEGG pathway enrichment analysis of differentially expressed pyroptosis-related genes are shown in the figure.

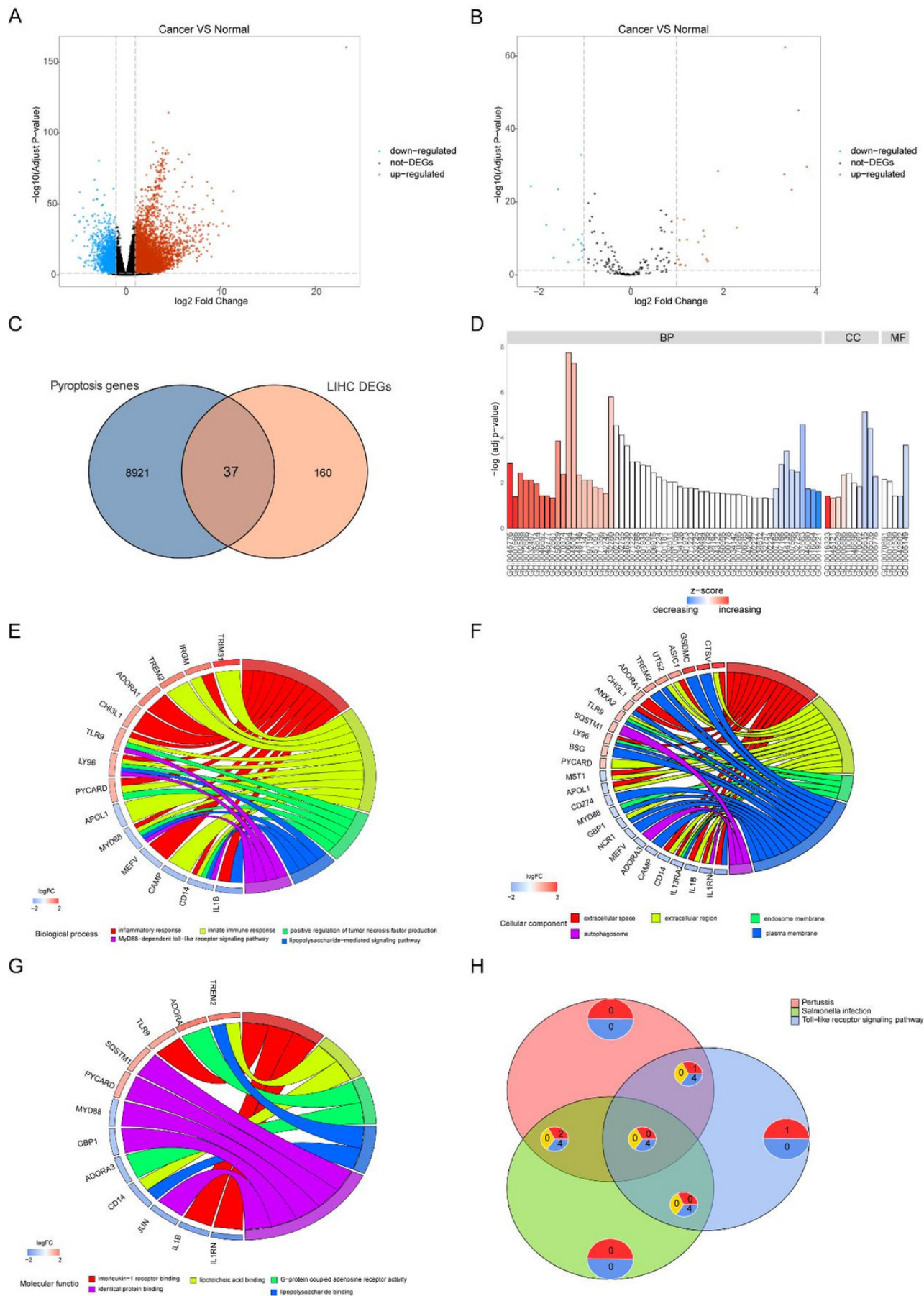


Figure 3

Construction of risk scoring model.

FigA: This figure shows a correlation analysis of characteristic genes in LIHC. FigB-C: The photos represent the PCA analysis of characteristic genes in both TCGA-LIHC and GSE76427 datasets. Red indicates low-risk groups and blue indicates high-risk groups. FigD-E: Kaplan-Meier analysis revealed the effect of risk score on the overall survival of patients with hepatocellular carcinoma in TCGA-LIHC and GSE76427. Red indicates the low-risk group and blue indicates the high-risk group.

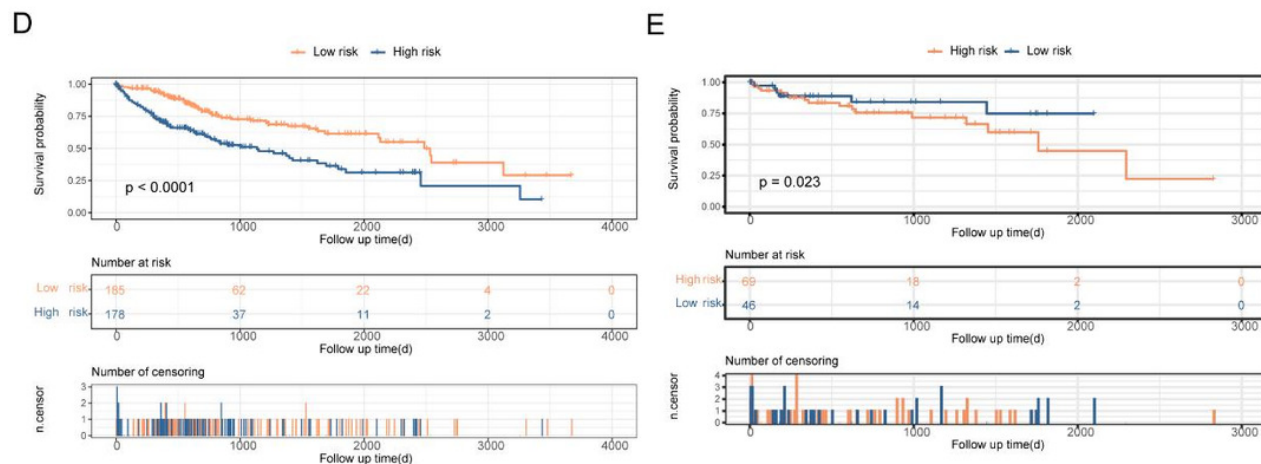
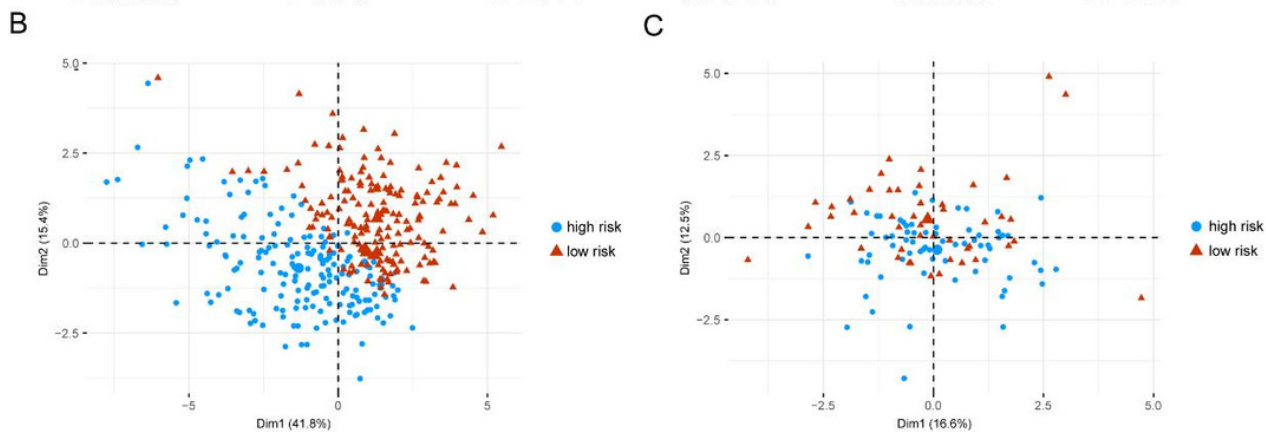
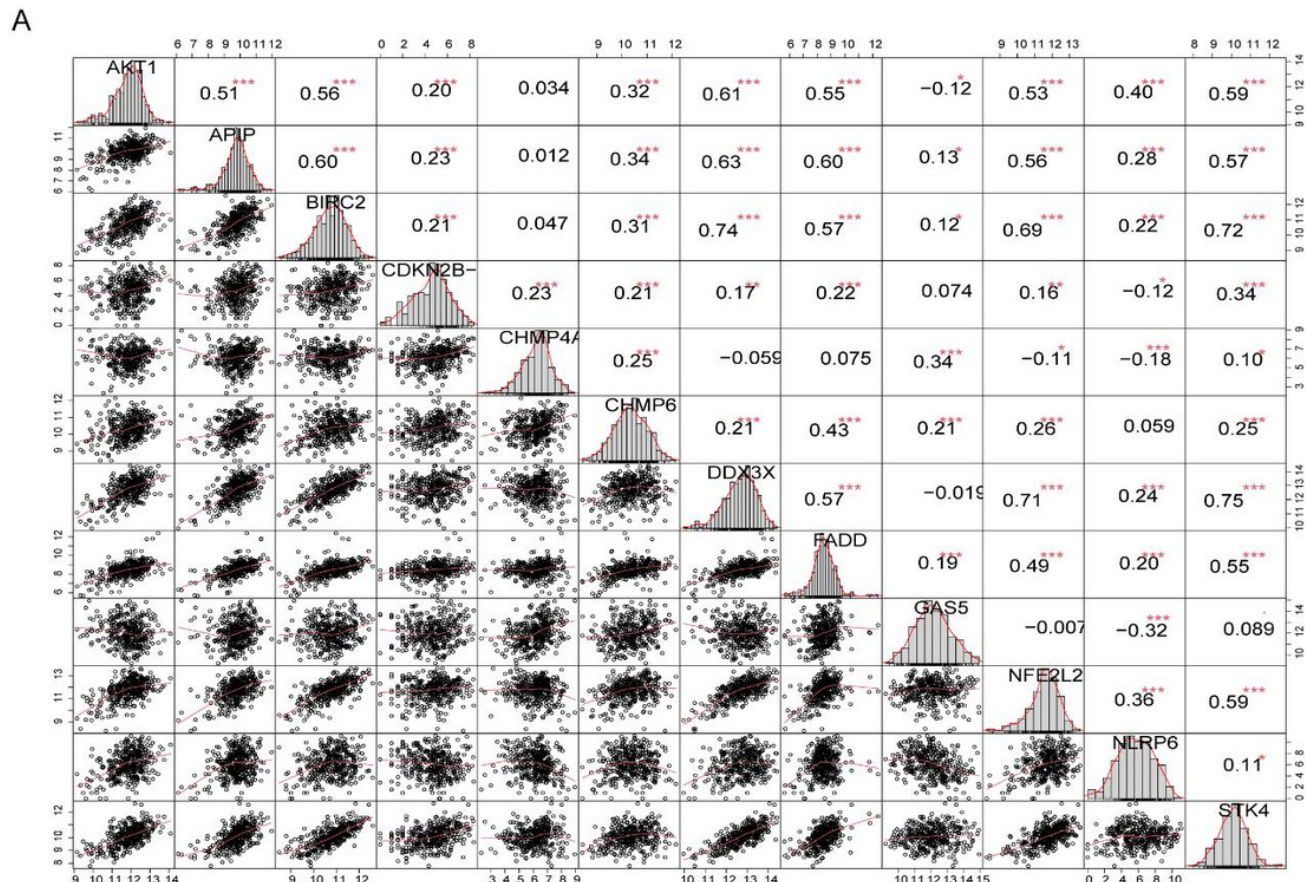


Figure 4

Analysis of the predictive ability of risk score on the prognosis of LIHC patients.

FigA: The figure shows the prediction model nomogram. FigB-D: These pictures are the calibration curves of the prediction model nomogram. The abscissa is the survival predicted by the nomogram, and the ordinate is the actual observed survival. The curve shows the prognosis prediction of the model for patients with hepatocellular carcinoma at 1, 3 and 5 years. FigE: Time-ROC showed the predicted potencies of one-year, three-year and five-year survival. FigF: This is the calibration curve of nomogram model for predicting 1-year survival, 3-year survival and 5-year survival of patients with hepatocellular carcinoma. FigG: Univariate Cox analysis. FigH: Multivariate Cox analysis.

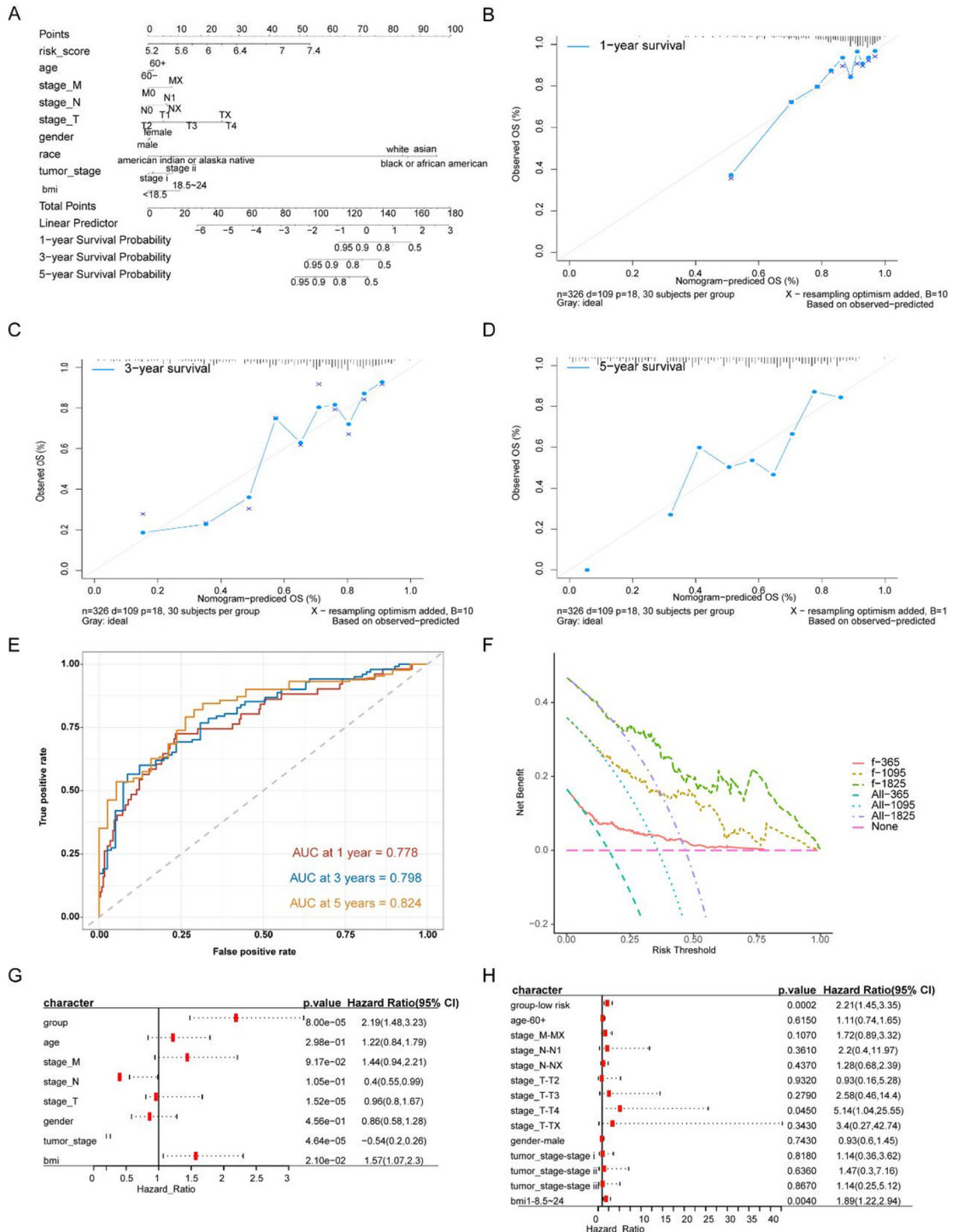


Figure 5

Differentially expressed genes between high-risk group and low-risk group.

FigA: The figure shows the PCA analysis of differentially expressed genes between high-risk group and low-risk group. FigB: The figure shows the volcanic map of differentially expressed genes between high-risk group and low-risk group. The abscissa is $\log_2\text{FoldChange}$ and the ordinate is $-\log_{10}(\text{Adjust P-value})$. The red node indicates the up-regulated differentially expressed genes, the blue node indicates the down-regulated differentially expressed genes, and the black node indicates the genes that are not significantly differentially expressed. FigC-E: These pictures respectively represent the results of BP, CC and MF in the GO biological function annotation analysis of the differentially expressed genes. FigF: This figure represents the result of KEGG pathway enrichment analysis of differentially expressed genes.

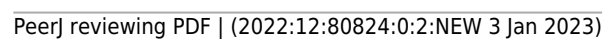


Figure 6

Copy number differences between high-risk and low-risk groups.

FigA-D: These graphs represent genes with significant amplification and deletion. Error detection rate (Q value) and GISTIC2.0 of the change score (x-axis) corresponds to the genomic position (Y-axis). The dotted line indicates the centromere. The green line indicates the 0.25 Q cut-off point for determining significance. These figures represent the copy number amplification of patients in the high-risk group, copy number deletion in high-risk group, copy number amplification in low-risk group and copy number deletion in patients in the low-risk group.

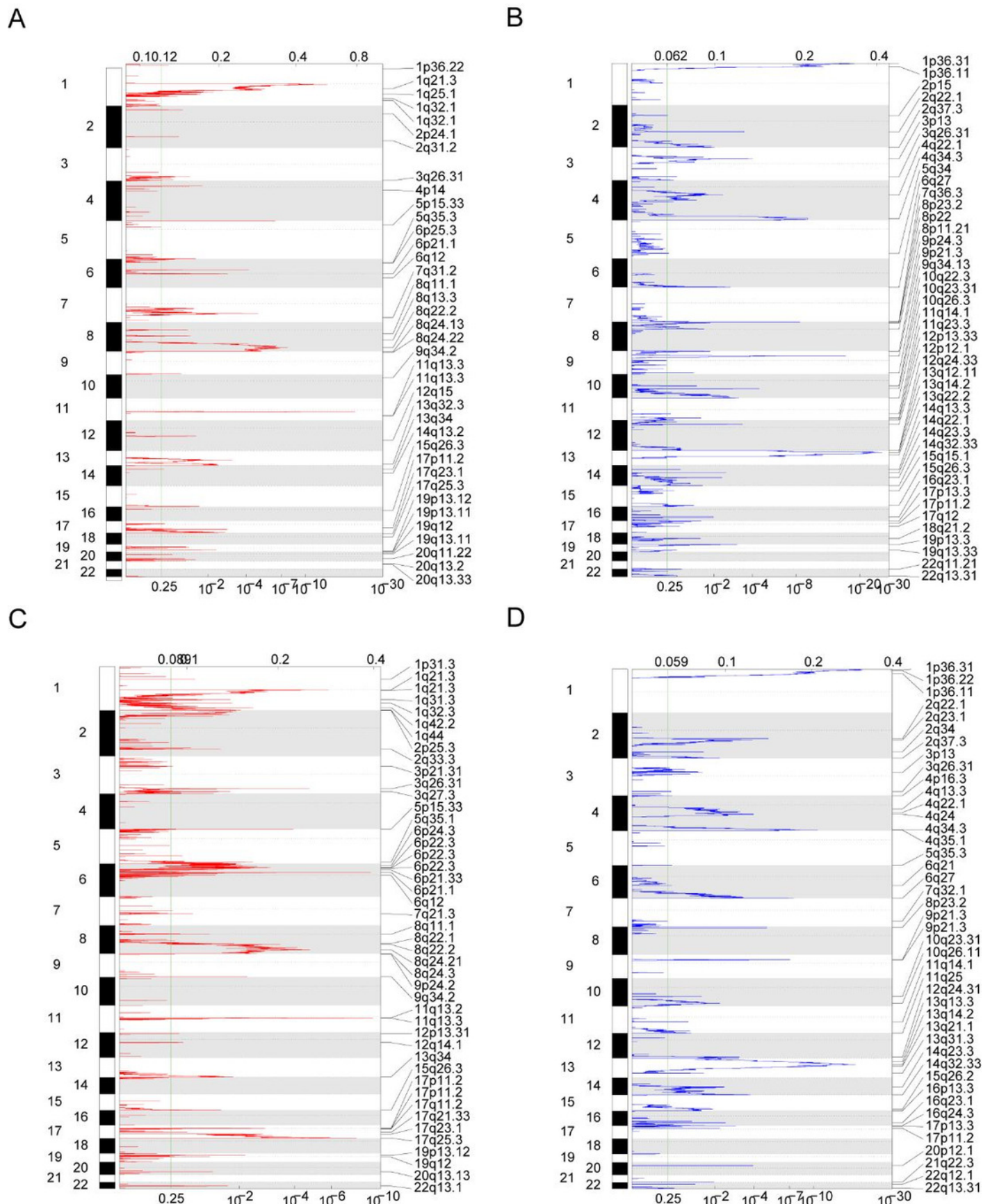


Figure 7

Drug sensitivity analysis of patients in high-risk and low-risk groups.

FigA-D: The horizontal axis of these pictures represents the patient grouping, and the vertical axis represents the \log_{10} (IC₅₀) of the drug. These pictures in turn show the more sensitive hepatocellular carcinoma drugs in the high-risk group, the more sensitive hepatocellular carcinoma drugs in the low-risk group, the more sensitive other cancer drugs in the high-risk group and the more sensitive other cancer drugs in the low-risk group.

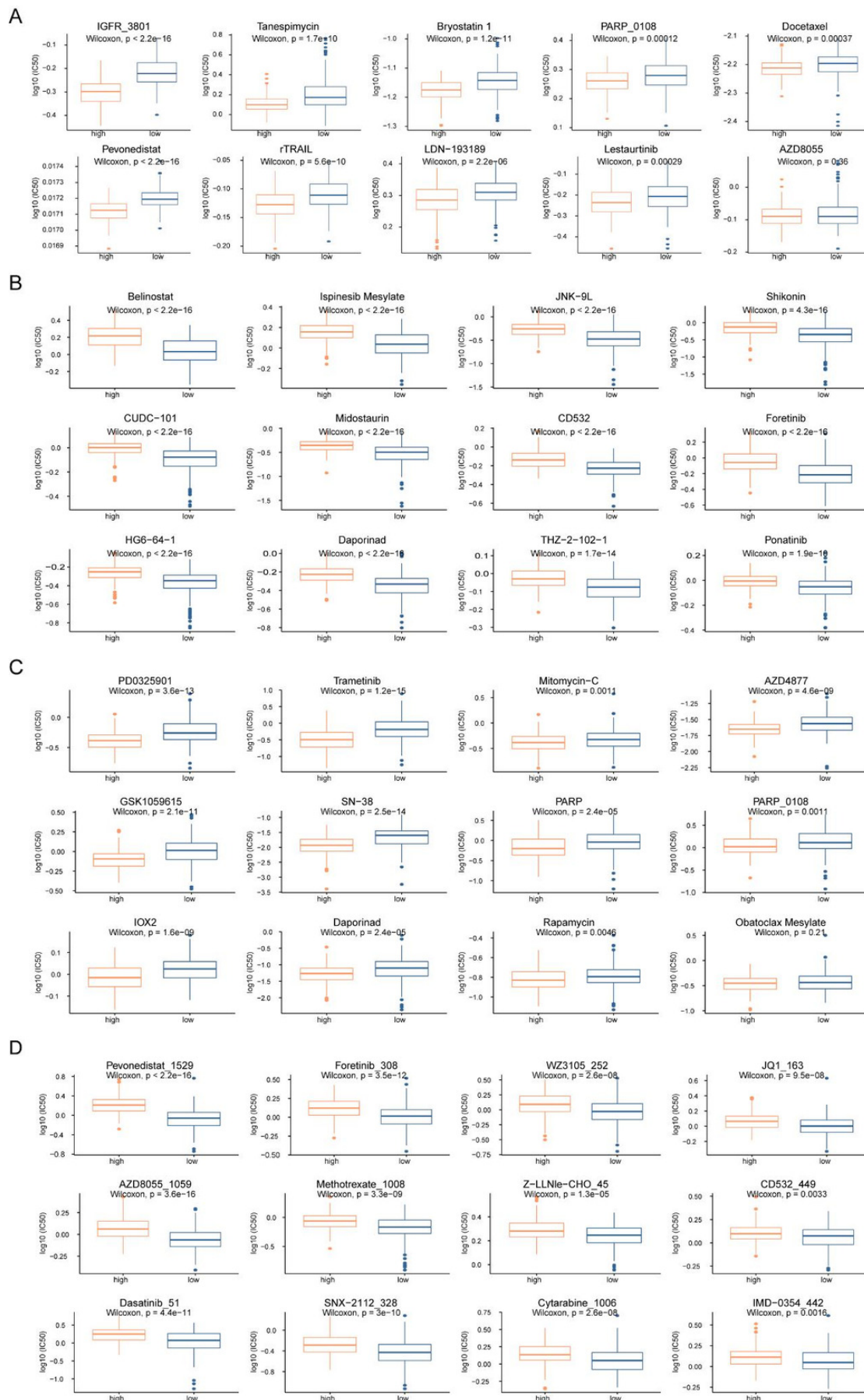


Figure 8

GSEA analysis of gene expression data of high- and low- risk groups.

FigA: The figure shows the GSEA-GO analysis of TCGA-LIHC data set. The abscissa is gene ratio, the ordinate is GO terms, and the color represents $-\log_{10}$ (pvalue). FigB: The figure shows the GSEA-KEGG analysis of TCGA-LIHC data set. The abscissa is gene ratio, the ordinate is GO terms, the node size represents the number of genes enriched in GO terms, and the node color represents $-\log_{10}$ (pvalue). FigC: The first three items of GSEA-GO analysis of TCGA-LIHC data set are shown in the figure. FigD: The first three items of TCGA-LIHC analysis of TCGA-LIHC data set are shown in the figure. FigE: The heat map shows the GSVA analysis of high-risk and low-risk groups. The horizontal axis is the patient ID and the vertical axis is the hallmark gene set.

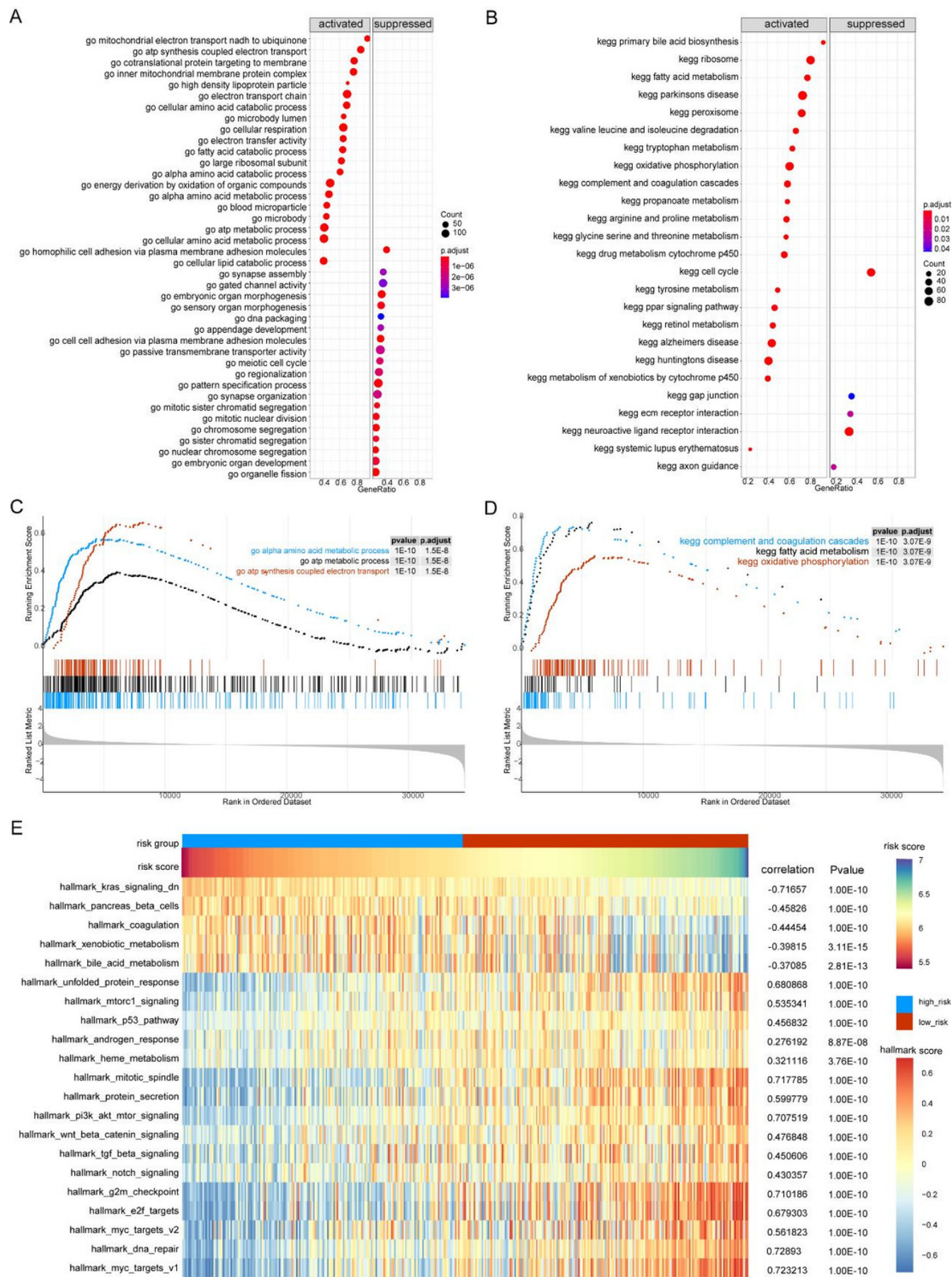


Figure 9

Analysis of immune cell infiltration in high-risk and low-risk groups.

FigA: The heat map shows the distribution of immune cells in different risk groups. The horizontal axis is the patient ID and the vertical axis is the proportion of immune cells. FigB: The results showed the difference of immune cell content among different risk groups. The horizontal axis is immune cells, the vertical axis is immune cell content, blue indicates patients in low-risk group, and orange indicates patients in high-risk group. FigC: The results showed the correlation analysis between the prognostic risk score and the content of immune cells. The horizontal axis is the immune cells significantly related to the prognostic risk score, the vertical axis is the correlation score, and the color of the bar graph indicated the significance of the correlation, $P < 0.05$ means statistically significant. FigD-E: This is the heat map of immune cell correlation between high-risk and low-risk groups. Blue indicates positive correlation and red indicates negative correlation.

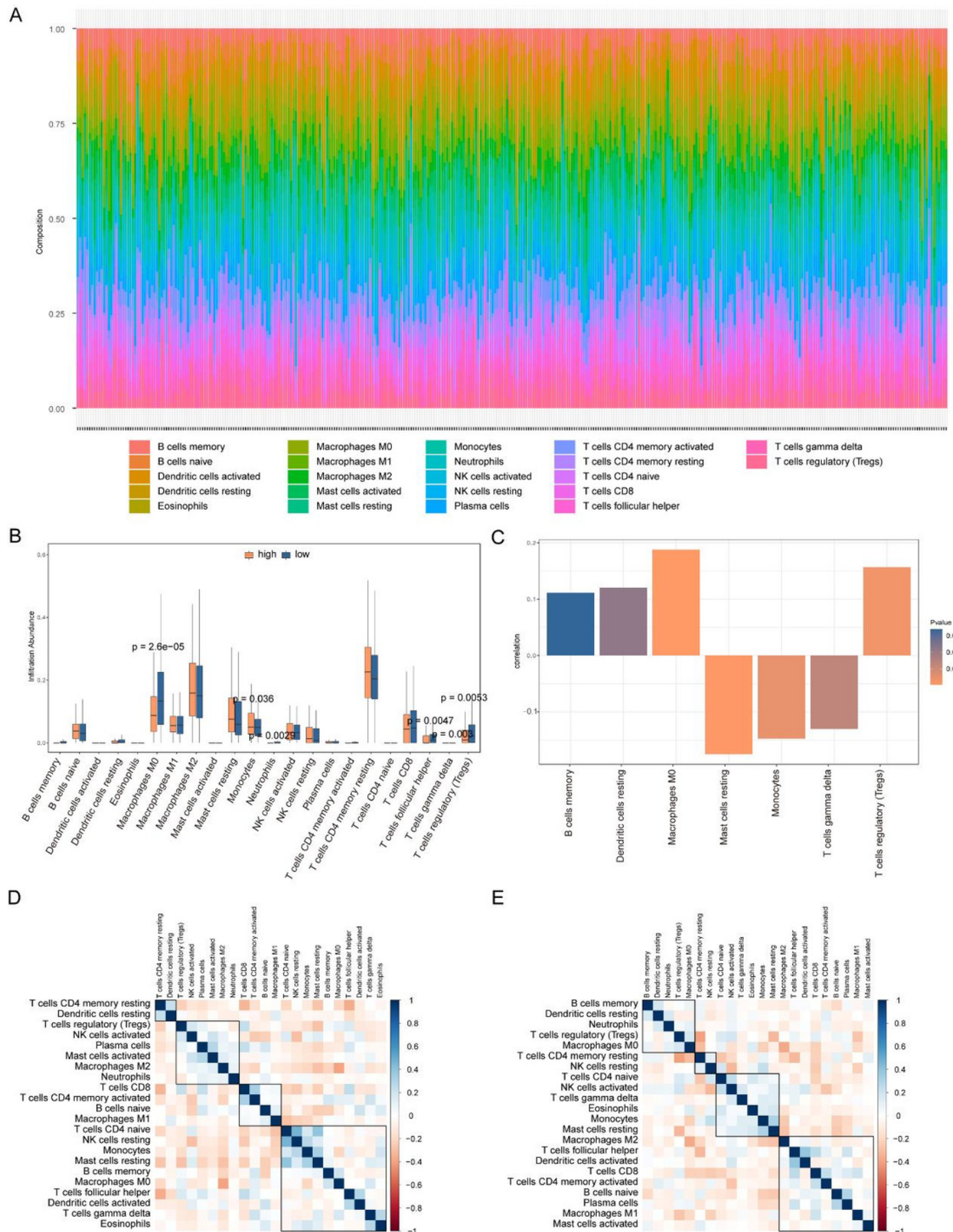


Figure 10

PPI network.

FigA: Violin chart shows the difference of stromal score between high-risk and low-risk groups. Blue indicates patients in low-risk group, orange indicates patients in high-risk group, $P < 0.05$ means statistically significant. FigB: The figure shows a protein-protein interaction network of differentially expressed genes in patients of high- and low-risk groups. FigC: This figure shows the results of clueGO enrichment analysis in the protein-protein interaction network related to differentially expressed genes. FigD: This figure shows the hub genes in the protein-protein interaction network related to differentially expressed genes.

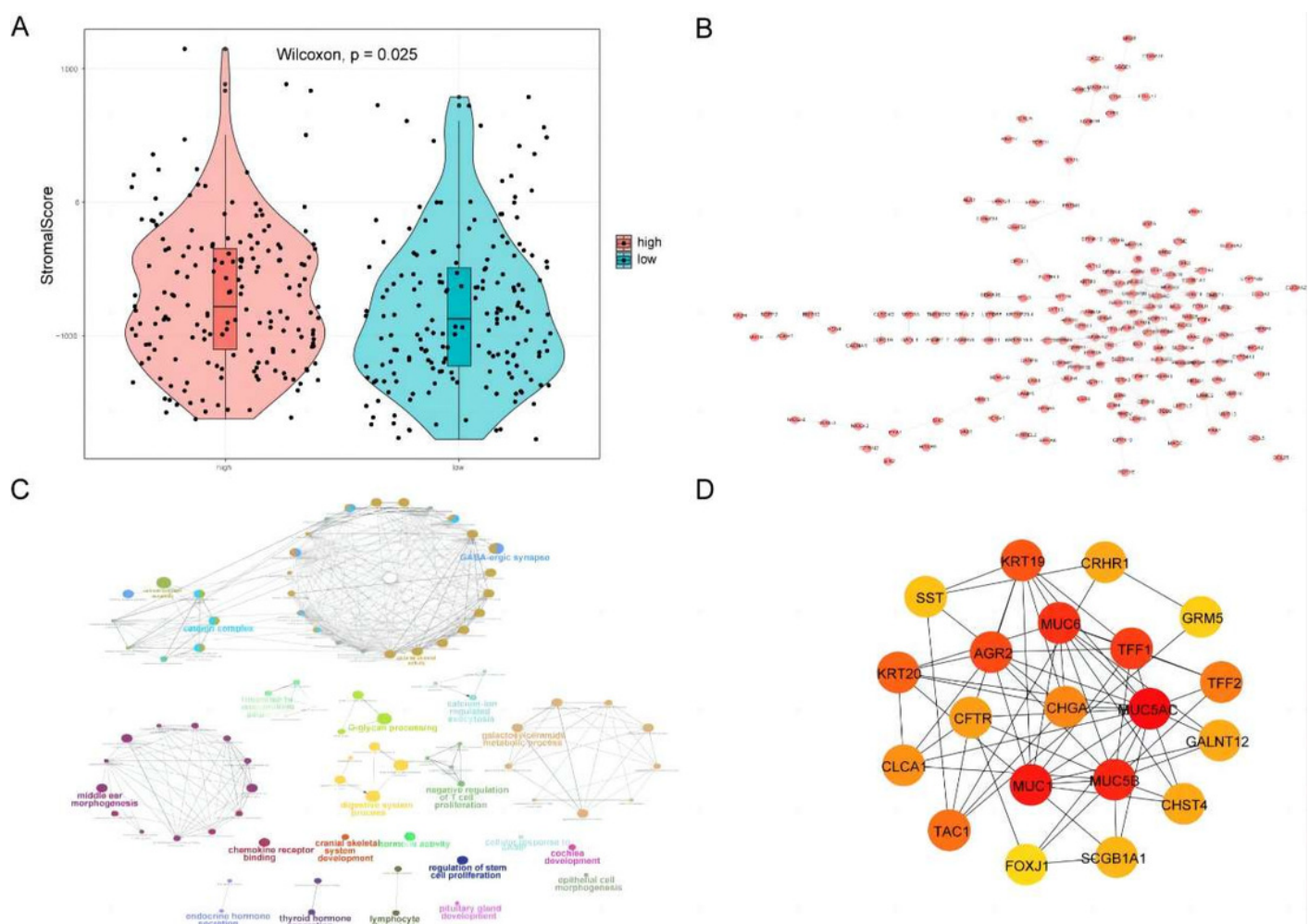


Figure 11

qRT- PCR results.

Detection of MDK expression in hepatocellular carcinoma and adjacent tissues by qRT- PCR.

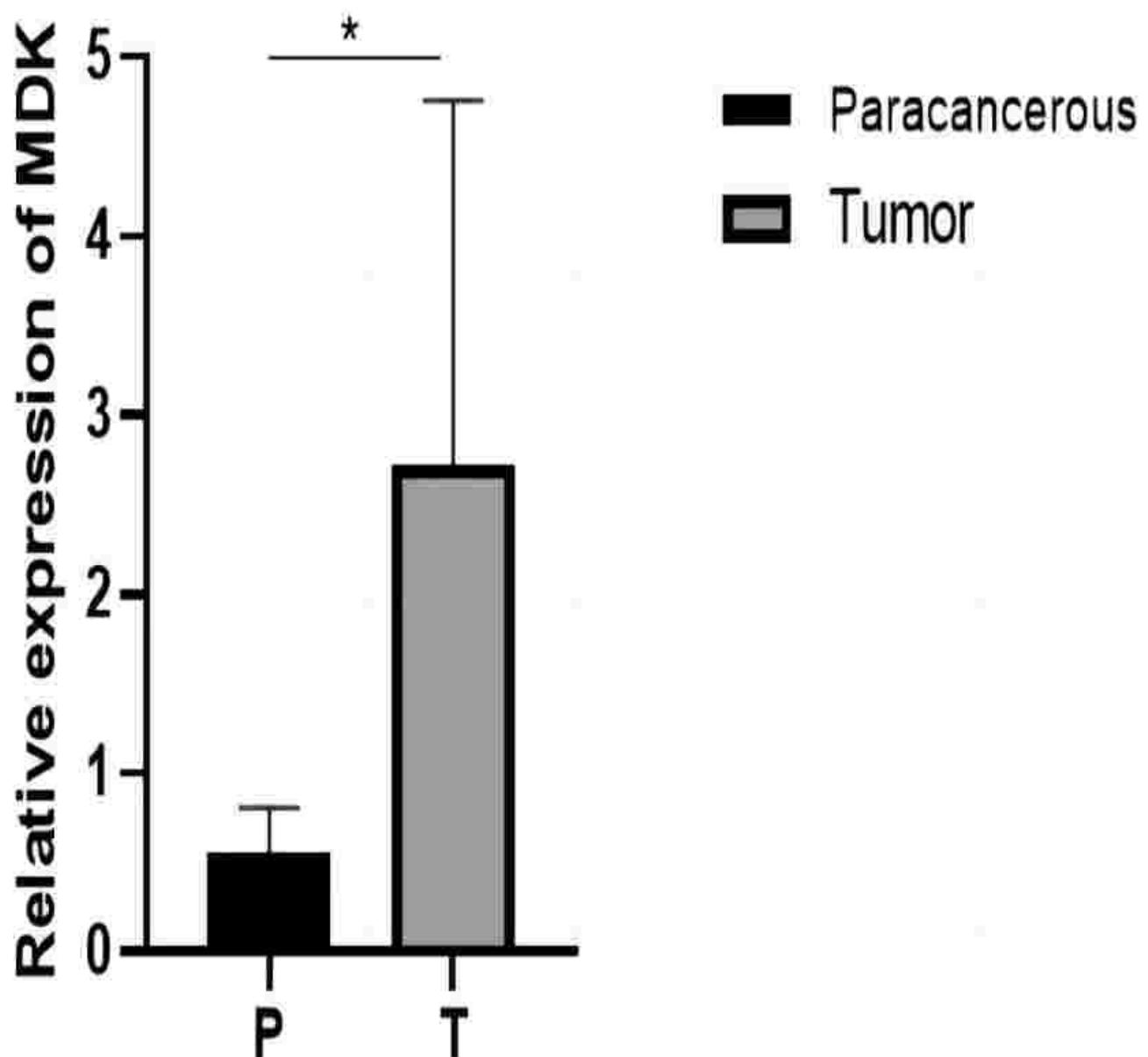


Figure 12

Immunohistochemical results.

The three pictures on the left are cancer tissue after immunohistochemical staining with MDK antibody, and the three pictures on the right are cancer tissue after immunohistochemical staining with MDK antibody.

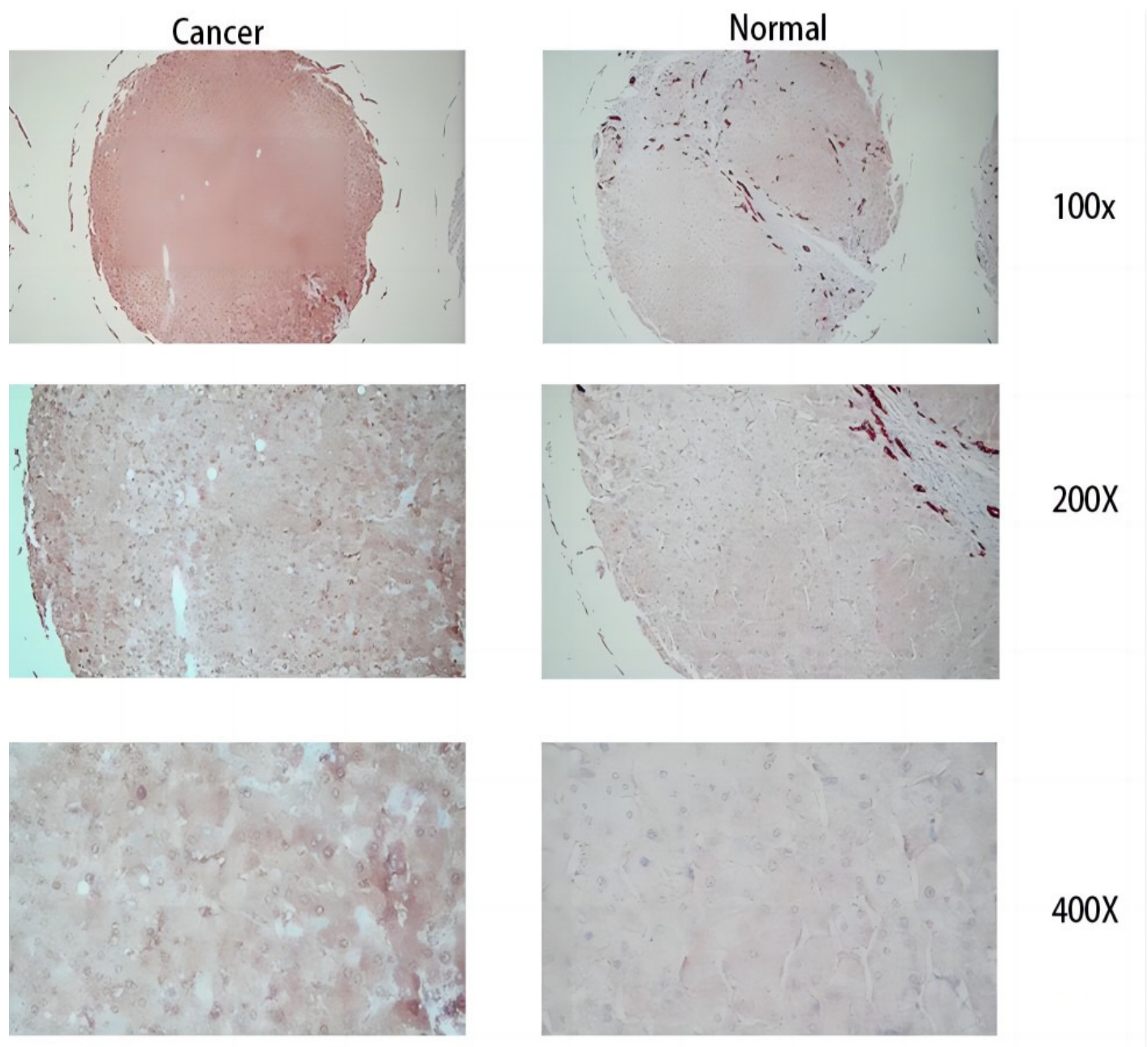


Table 1 (on next page)

TCGA-LIHC patient baseline table.

1 Table 1: TCGA-LIHC patient baseline table.

	high risk	low risk	p	test
n	163	163		
group = low risk (%)	0 (0.0)	163 (100.0)	<0.001	exact
age = 60+ (%)	83 (50.9)	90 (55.2)	0.506	exact
stage_M = MX (%)	36 (22.1)	39 (23.9)	0.793	exact
stage_N (%)		0.234	exact	
N0	120 (73.6)	113 (69.3)		
N1	0 (0.0)	3 (1.8)		
NX	43 (26.4)	47 (28.8)		
stage_T (%)		0.009	exact	
T1	99 (60.7)	71 (43.6)		
T2	34 (20.9)	45 (27.6)		
T3	27 (16.6)	40 (24.5)		
T4	2 (1.2)	7 (4.3)		
TX	1 (0.6)	0 (0.0)		
gender = male (%)	117 (71.8)	107 (65.6)	0.282	exact
race (%)			0.262	exact
american indian or alaska native	1 (0.6)	0 (0.0)		
asian	75 (46.0)	75 (46.0)		
black or african american	7 (4.3)	6 (3.7)		
not reported	8 (4.9)	2 (1.2)		
white	72 (44.2)	80 (49.1)		
tumor_stage (%)		0.008	exact	
not reported	7 (4.3)	10 (6.1)		
stage i	95 (58.3)	65 (39.9)		

stage ii	33 (20.2)	41 (25.2)		
----------	-----------	-----------	--	--

2
3

Table 2 (on next page)

KEGG enrichment analysis of DEGs and Go enrichment analysis of differentially expressed genes between high-risk and low-risk groups.

1 **Table 2-1: Go enrichment analysis of differentially expressed genes between high-risk and low-risk**
 2 **groups.**

Categor y	ID	Description	pvalue
BP	GO:0007586	digestion	2.52E-09
BP	GO:0030277	maintenance of gastrointestinal epithelium	2.83E-06
BP	GO:0022600	digestive system process	1E-05
BP	GO:0010669	epithelial structure maintenance	1.2E-05
BP	GO:0016266	O-glycan processing	4.38E-05
BP		cell-cell adhesion via plasma-membrane	
	GO:0098742	adhesion molecules	0.000159
BP	GO:0042471	ear morphogenesis	0.000167
BP	GO:0046879	hormone secretion	0.000378
BP	GO:0048568	embryonic organ development	0.000451
BP	GO:0001894	tissue homeostasis	0.000455
MF	GO:0099095	ligand-gated anion channel activity	0.000898
MF	GO:0005254	chloride channel activity	0.00123
MF	GO:0005179	hormone activity	0.002106
MF	GO:0005253	anion channel activity	0.002396
MF	GO:0048018	receptor ligand activity	0.002464
MF	GO:0030246	carbohydrate binding	0.002533
MF	GO:0030546	signaling receptor activator activity	0.002738
MF	GO:0022836	gated channel activity	0.003724
MF	GO:0015276	ligand-gated ion channel activity	0.004027
CC	GO:0016342	catenin complex	0.000397
CC	GO:0098982	GABA-ergic synapse	0.000681

CC	GO:0005796	Golgi lumen	0.000909
CC	GO:0032590	dendrite membrane	0.00124
CC	GO:0045177	apical part of cell	0.002132
CC	GO:0034707	chloride channel complex	0.00218
CC	GO:0099056	integral component of presynaptic membrane	0.003304
CC	GO:0032589	neuron projection membrane	0.003991
CC	GO:0099055	integral component of postsynaptic membrane	0.004901

Table2-2: KEGG enrichment analysis of DEGs of differentially expressed genes between high-risk and low-risk groups.

Category	ID	Description	pvalue
KEGG_PATHW		Neuroactive ligand-receptor	
AY	hsa04080	interaction	5.63E-05
KEGG_PATHW			
AY	hsa04742	Taste transduction	0.001914
KEGG_PATHW			
AY	hsa04972	Pancreatic secretion	0.004025
KEGG_PATHW			
AY	hsa00430	Taurine and hypotaurine metabolism	0.005494
KEGG_PATHW			
AY	hsa04024	cAMP signaling pathway	0.007434
KEGG_PATHW			
AY	hsa04540	Gap junction	0.012952
KEGG_PATHW			
AY	hsa00600	Sphingolipid metabolism	0.014045
KEGG_PATHW			
AY	hsa04657	IL-17 signaling pathway	0.016168
KEGG_PATHW	hsa04514	Cell adhesion molecules	0.01897

AY

KEGG_PATHW

AY

hsa04530

Tight junction

0.030624

16

17

Table 3(on next page)

GSEA analysis between high-risk and low-risk groups (promoted biological functions).

1 **Table 3: GSEA analysis between high-risk and low-risk groups (promoted biological functions).**

ID	ES	NES	pvalue
go_very_low_density_lipoprotein_particle_re modeling	0.887434	2.390189	1.98E-06
go_triglyceride_rich_lipoprotein_particle_re odeling	0.885009	2.464035	3.42E-07
go_complement_activation_lectin_pathway	0.873119	2.389054	2.67E-06
go_platelet_dense_granule_lumen	0.871294	2.42585	9.9E-07
go_alcohol_dehydrogenase_nad_p_plus_activ ity	0.86067	2.2559	2.65E-05
go_chylomicron	0.847338	2.359152	6.56E-06
go_glyoxylate_metabolic_process	0.844408	2.213275	6.14E-05
go_blood_coagulation_intrinsic_pathway	0.837189	2.530543	2E-07
go_fatty_acid_beta_oxidation_using_acyl_coa _dehydrogenase	0.83296	2.183269	0.000118
go_urea_cycle	0.83096	2.238084	0.000121
kegg_primary_bile_acid_biosynthesis	0.847215	2.524414	2.89E-07
kegg_fatty_acid_metabolism	0.765637	2.896549	1E-10
kegg_complement_and_coagulation_cascades	0.738595	3.04783	1E-10
kegg_glycine_serine_and_threonine_metaboli sm	0.711765	2.544408	2.27E-07
kegg_renin_angiotensin_system	0.669718	2.062439	0.001817
kegg_valine_leucine_and_isoleucine_degradat ion	0.667422	2.542512	8.42E-08
kegg_peroxisome	0.639025	2.697852	1E-10
kegg_histidine_metabolism	0.625501	2.165818	0.000229
kegg_parkinsons_disease	0.608925	2.673643	1E-10

kegg_ribosome	0.604297	2.580318	1E-10
---------------	----------	----------	-------

2
 3

Table 4(on next page)

GSEA analysis between high-risk and low-risk groups (inhibited biological function).

Table 3: GSEA analysis between high-risk and low-risk groups (inhibited biological function).

ID	ES	NES	pvalue
go_inhibitory_extracellular_ligand_gated_ion_channel_activity	-0.85224	-1.93587	9.71E-05
go_limb_bud_formation	-0.83795	-1.9034	0.000239
go_condensed_chromosome_outer_kinetochore	-0.83151	-1.97688	5.58E-05
go_spindle_elongation	-0.82814	-1.83236	0.000416
go_regulation_of_chloride_transport	-0.80699	-1.78556	0.001196
go_mitotic_dna_replication	-0.79769	-1.9836	9.96E-05
go_synaptic_vesicle_docking	-0.79619	-1.80854	0.001997
go_chromatoid_body	-0.78929	-1.8765	0.000609
go_negative_regulation_of_transcription_by_competitive_promoter_binding	-0.78684	-1.74097	0.002793
go_galactolipid_metabolic_process	-0.77716	-1.71956	0.003771
kegg_cell_cycle	-0.50829	-1.75203	3.25E-05
kegg_ecm_receptor_interaction	-0.48663	-1.61201	0.004043
kegg_neuroactive_ligand_receptor_interaction	-0.46544	-1.67554	2.41E-06
kegg_gap_junction	-0.46196	-1.53597	0.007855
kegg_axon_guidance	-0.44686	-1.54182	0.003792
kegg_huntingtons_disease	0.379055	1.8326	5.28E-06
kegg_steroid_hormone_biosynthesis	0.390165	1.552235	0.009584
kegg_alzheimers_disease	0.400588	1.910789	1.57E-06
kegg_arachidonic_acid_metabolism	0.402226	1.612863	0.008925
kegg_ppar_signaling_pathway	0.449864	1.845516	0.000117