

Screening of Key Hub Genes in Hepatocellular Carcinoma and Their Prognostic Value Based on Integrated Bioinformatics Analysis

Can Jin

*School of Life Sciences, Yunnan University, Kunming, China
jincan1@stu.ynu.edu.cn*

Abstract. This study aimed to screen key genes associated with hepatocellular carcinoma (HCC) using integrated bioinformatics methods and evaluate their prognostic significance. Public transcriptome datasets were utilized to conduct differential expression analysis, functional enrichment analysis, immune infiltration analysis, and protein-protein interaction network analysis for hub gene screening. Gene expression characteristics were validated in combination with the TCGA cohort, and a prognostic risk model was constructed to assess its predictive value. Differentially expressed genes were mainly enriched in processes related to cell proliferation and cell cycle. Hub genes screened through network analysis were significantly highly expressed in HCC and correlated with poor prognosis in patients. The risk score model constructed based on these genes exhibited certain efficacy in survival prediction. The HCC-related hub genes identified in this study may be involved in the regulation of tumor progression and possess potential prognostic value, providing a reference for the research of molecular markers in HCC.

Keywords: Hepatocellular carcinoma, Hub genes, Bioinformatics, Prognostic model

1. Introduction

Hepatocellular carcinoma (HCC) is the most common histological type of primary liver cancer and also constitutes a major component of the global cancer mortality burden. Due to the insidious onset of HCC and the lack of specific early clinical manifestations, many patients are diagnosed at intermediate or advanced stages, resulting in limited available treatment options and unsatisfactory long-term survival outcomes. Despite continuous advances in surgical resection, local ablation, interventional therapy, molecular targeted therapy, immunotherapy and other strategies in recent years, there remain substantial differences in treatment response and survival outcomes among different patients. Therefore, in-depth exploration of the molecular basis underlying the occurrence and development of HCC, as well as the identification of biomarkers with diagnostic or prognostic value, is of great significance for optimizing risk assessment and promoting subsequent mechanistic research [1, 2].

The initiation and progression of HCC are not driven by alterations in a single gene or abnormalities in a single signaling pathway, but rather the combined effect of multiple molecular

events. Previous studies have shown that HCC exhibits significant heterogeneity in molecular typing, cellular composition, pathological progression and treatment response, which is also an important reason for its difficulty in early identification, obvious differences in therapeutic efficacy and inconsistent prognosis [3]. With the wide application of high-throughput sequencing technology and the continuous improvement of public database resources, bioinformatics analysis based on transcriptome data has become an important approach to decipher the molecular characteristics of tumors. By systematically comparing gene expression differences between tumor tissues and normal or adjacent non-tumor tissues, combined with functional enrichment analysis, key biological processes, regulatory networks and potential candidate genes associated with tumorigenesis and tumor development can be identified at the overall level.

In addition to intrinsic molecular alterations in tumor cells, the tumor immune microenvironment also participates in the regulation of HCC occurrence, progression and treatment response. The infiltration level and compositional changes of different immune cell subsets in tumor tissues may affect immune escape, inflammatory status, tumor progression rate and immunotherapy sensitivity. Existing studies have suggested [4, 5] that there is a complex immune regulatory network in the HCC microenvironment, and macrophages, T cells, NK cells and other immune cell populations may influence tumor biological behaviors through cytokine secretion, antigen presentation, immune suppression or inflammatory responses. Therefore, the combination of differential expression analysis and immune infiltration evaluation helps to understand the molecular characteristics of HCC from both the intrinsic abnormalities of tumor cells and the remodeling of the external microenvironment [6].

In terms of screening candidate key genes, protein-protein interaction (PPI) network analysis and hub gene identification can further identify molecules that may exert core regulatory roles from the perspective of systems biology. Related studies have shown that hub genes are often located at key nodes of important biological networks, potentially participating in the occurrence, development, disease progression and prognosis of HCC, and possess research value as diagnostic markers, prognostic indicators or potential therapeutic targets. Meanwhile, the establishment of prognostic risk models based on gene expression profiles also provides a feasible approach for survival risk assessment and individualized prognostic stratification of HCC patients [7].

Based on the above background, this study used public transcriptome data to comprehensively perform differential expression analysis, functional enrichment analysis, immune infiltration analysis, PPI network construction and survival analysis, screened key HCC-related genes, and further evaluated their expression patterns, immune-related characteristics and prognostic value. Through the above analyses, this study aims to provide a reference for exploring the molecular mechanism of HCC, screening potential biomarkers and conducting prognostic evaluation research.

2. Materials and methods

2.1. Data source and preprocessing

Transcriptomic data and clinical information used in this study were mainly obtained from the Gene Expression Omnibus (GEO) database and The Cancer Genome Atlas (TCGA) database [8, 9]. Specifically, the GSE202853 dataset, which contains hepatocellular carcinoma (HCC) tissues and their paired adjacent non-tumor tissues, was used for screening differentially expressed genes (DEGs), functional enrichment analysis, and immune-related analysis. The transcriptomic expression matrix and corresponding clinical survival information of the TCGA-LIHC cohort were used for hub gene expression validation, survival analysis, and prognostic model evaluation.

After downloading the expression matrix, gene expression data were organized and normalized. For entries with duplicate gene names, the average expression value was taken as the final expression level of the gene. Genes with excessively low expression or a large amount of missing information were removed, retaining the expression matrix available for subsequent analysis. In the TCGA-LIHC clinical information, samples with missing overall survival time, overall survival status, or key clinical variables were excluded, and the expression matrix was matched with clinical information according to sample IDs.

2.2. Screening of differentially expressed genes

Differential expression analysis was performed using the DESeq2 package in R software [10]. HCC tumor tissues and paired adjacent non-tumor tissues in GSE202853 were used as research objects to construct a differential expression analysis model. For paired samples, individual pairing information was incorporated into the model design to reduce the impact of individual differences on differential expression results. Differential expression results were corrected for multiple testing using the Benjamini-Hochberg method.

In this study, $|\log_2 \text{fold change}| > 1$ and adjusted P value < 0.05 were used as thresholds for screening DEGs. DEGs were classified into upregulated genes and downregulated genes according to the direction of $\log_2$ fold change. The overall distribution of differential expression results was displayed using a volcano plot.

2.3. GO and KEGG functional enrichment analysis

To further explore the biological functions and signaling pathways that DEGs may be involved in, Gene Ontology (GO) functional enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed on DEGs. First, gene symbols of DEGs were converted to ENTREZ IDs, and then enrichment analysis was conducted using the clusterProfiler package in R software [11, 12].

GO functional enrichment analysis included three levels: biological process (BP), cellular component (CC), and molecular function (MF). KEGG enrichment analysis was used to identify metabolic pathways and signaling pathways significantly enriched by DEGs. Adjusted P value < 0.05 was used as the significance screening criterion for enrichment results, which were visualized using bar charts, bubble charts, and other methods.

2.4. Immune-related gene set scoring and immune infiltration analysis

To evaluate differences in immune-related functional status between HCC tumor tissues and adjacent non-tumor tissues, single-sample Gene Set Enrichment Analysis (ssGSEA) was used to calculate immune-related scores for each sample based on a precompiled immune-related gene set [13]. Differences in scores of different immune-related gene sets between tumor tissues and paired adjacent non-tumor tissues were then compared to reflect changes in immune-related pathways or functional status in HCC tissues.

To further analyze differences in immune cell composition, CIBERSORTx was used to estimate the relative proportion of immune cells in samples [14]. CIBERSORTx performs deconvolution analysis on transcriptomic expression matrices based on signature gene sets, which can be used to infer the relative abundance of different immune cell types in samples. For paired samples, paired statistical tests were used to compare differences in immune cell proportions between tumor tissues

and adjacent non-tumor tissues, and visualization was performed according to the trend of immune cell changes in tumor tissues.

2.5. Construction and module analysis of protein-protein interaction network

To further screen candidate genes that may play key regulatory roles in the occurrence and development of HCC, a protein-protein interaction (PPI) network was constructed based on DEGs. DEGs were imported into the STRING database, *Homo sapiens* was selected as the species, and protein-protein interaction relationships were screened according to the confidence threshold recommended by the database [15]. Subsequently, the interaction results were imported into Cytoscape software for network visualization and topological analysis [16]. Functional subnetworks with relatively tight connections were further extracted based on network connectivity, and their main biological processes and pathways were determined in combination with functional enrichment results.

2.6. Hub gene screening

To improve the stability of hub gene screening, network nodes were analyzed based on the topological characteristics of the PPI network. Three centrality metrics, namely Degree, Betweenness, and Closeness, were used to rank network nodes separately, and candidate genes with top rankings in each metric were screened. The intersection of the screening results of the three centrality metrics was then taken as the final hub genes to reduce bias caused by a single metric.

2.7. Hub gene expression validation and survival analysis

To validate the expression characteristics of hub genes in HCC, the TCGA-LIHC cohort was used to compare expression differences of hub genes between tumor tissues and normal tissues. For cases with paired adjacent non-tumor samples, paired expression difference analysis was further performed to enhance the reliability of the results. Expression difference results were displayed using box plots.

Survival analysis was performed based on overall survival time and survival status of the TCGA-LIHC cohort. Patients were divided into high-expression and low-expression groups according to the median expression level of each hub gene. Kaplan-Meier method was used to plot survival curves, and log-rank test was used to compare survival differences between the two groups.

2.8. Prognostic model construction and predictive performance evaluation

To further evaluate the combined prognostic value of candidate hub genes, an Elastic Net Cox regression model was constructed based on the TCGA-LIHC cohort [17]. The expression matrix of candidate hub genes was matched with patient survival information, and cross-validation was used to select model parameters. According to the risk score calculated by the model, patients were divided into high-risk and low-risk groups with the median as the cutoff.

Kaplan-Meier survival analysis was used to compare overall survival differences between high-risk and low-risk groups, and time-dependent receiver operating characteristic (ROC) curves were used to evaluate the model's performance in predicting 1-year, 3-year, and 5-year survival [18]. The area under the curve (AUC) was used to evaluate the predictive performance of the model.

2.9. Statistical analysis

All statistical analyses were performed in R software. Student's t-test was used for intergroup comparison of normally distributed data, and Wilcoxon rank-sum test was used for non-normally distributed data. For paired samples, paired t-test or Wilcoxon signed-rank test was used. Multiple comparisons were corrected using the Benjamini-Hochberg method. Survival analysis was performed using the Kaplan-Meier method and log-rank test. The prognostic risk model was constructed using Elastic Net Cox regression, and predictive performance was evaluated using time-dependent ROC curves. Unless otherwise specified, P value < 0.05 was considered statistically significant.

3. Results

3.1. Screening, functional enrichment and immune-related characteristic analysis of differentially expressed genes

Based on the GSE202853 dataset, this study first performed differential expression analysis on HCC tumor tissues and their paired adjacent non-tumor tissues using DESeq2. With $|\log_2\text{FoldChange}| > 1$ and adjusted P value < 0.05 as screening criteria, a total of 1333 differentially expressed genes were obtained, of which 753 genes were upregulated in tumor tissues and 580 genes were relatively highly expressed in adjacent non-tumor tissues. Volcano plot results showed significant transcriptomic differences between HCC tissues and adjacent non-tumor tissues, suggesting that extensive gene expression changes accompany the occurrence of HCC.

To further clarify the functional characteristics of differentially expressed genes, GO and KEGG enrichment analyses were conducted. GO analysis revealed that differentially expressed genes were mainly enriched in cell cycle- and nuclear division-related processes such as nuclear division, mitotic nuclear division, chromosome segregation, mitotic sister chromatid segregation, and nuclear chromosome segregation. At the cellular component level, they were mainly involved in structures including condensed chromosome, chromosomal region, kinetochore, and spindle. At the molecular function level, they were associated with functions such as ATPase activity, DNA helicase activity, and microtubule binding. KEGG analysis further showed that differentially expressed genes were significantly enriched in pathways including Cell cycle, DNA replication, Oocyte meiosis, p53 signaling pathway, Wnt signaling pathway, and Cholesterol metabolism. Among them, cell cycle- and DNA replication-related pathways were the most prominent, indicating typical tumor biological characteristics such as cell cycle dysregulation and enhanced DNA replication in HCC tissues.

In addition to proliferation-related pathways, this study further analyzed immune-related functions and changes in immune cell composition. ssGSEA results showed that some immune-related gene sets differed between tumor tissues and adjacent non-tumor tissues, among which pathways such as COMPLEMENT, INTERFERON ALPHA RESPONSE, and INFLAMMATORY RESPONSE exhibited significant or nearly significant changes. CIBERSORTx analysis showed that different immune cell types presented distinct changing trends between tumor tissues and adjacent non-tumor tissues, suggesting that immune cell composition may be remodeled to a certain extent in HCC tissues. These results indicated that HCC not only exhibits abnormal activation of cell proliferation-related pathways but is also accompanied by alterations in immune-related functional status.

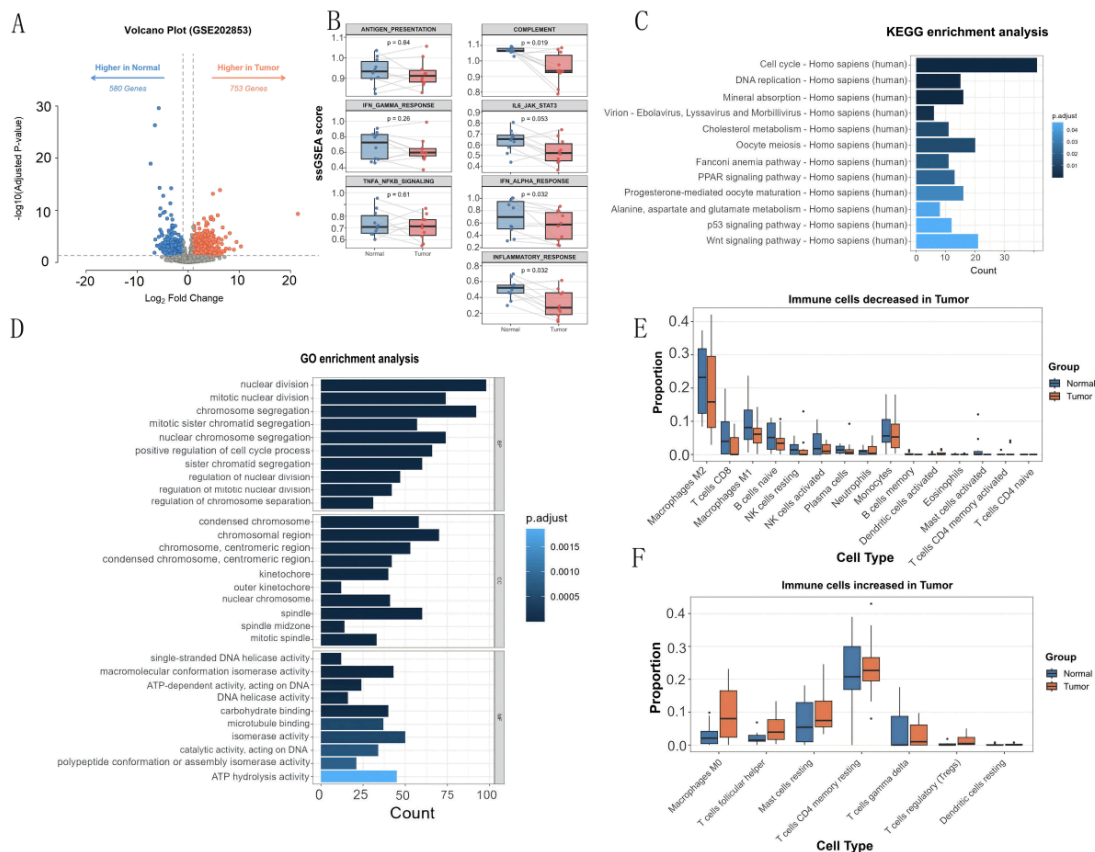


Figure 1. Results of differentially expressed gene screening, functional enrichment and immune-related analysis

Figure 1 comprehensively shows the volcano plot of differentially expressed genes, ssGSEA immune-related scores, KEGG pathway enrichment results, GO functional enrichment results, and changing trends of immune cell composition analyzed by CIBERSORTx in the GSE202853 dataset.

3.2. Hub gene expression validation, network module analysis and key gene screening

To further screen key genes with potential regulatory roles from differentially expressed genes, this study constructed a PPI network based on the STRING database and performed network analysis using Cytoscape. The initial PPI network contained 395 nodes and 5428 edges, and a core network including 387 nodes and 5423 edges was obtained after extracting the maximum connected subgraph. Due to the complex structure of the complete network, this study focused on the analysis of tightly connected functional modules and key nodes within it.

PPI network module analysis showed that some gene modules were closely related to DNA replication, DNA damage repair and cell cycle regulation. Modules represented by BRCA1, TOP2A, MCM family genes, CDC45 and RFC family genes were tightly connected, suggesting that these genes may jointly participate in HCC cell proliferation and genome replication. Another module included genes such as ALB, HGF, ESR1, EZH2, CCL2, CCL23 and CCL25, which may be involved in signal transduction, metabolic regulation and immune-related processes. The module analysis results corroborated the aforementioned GO and KEGG enrichment results, further indicating that abnormal cell cycle and DNA replication are important features of transcriptomic changes in HCC.

Subsequently, three centrality indicators including Degree, Betweenness and Closeness were used to rank PPI network nodes, and the top 20 genes of each indicator were selected for intersection analysis. Results showed that six genes simultaneously appeared in the screening results of the three centrality indicators, namely ASPM, BRCA1, CCNA2, CDK1, PLK1 and TOP2A. Therefore, these six genes were identified as candidate hub genes in this study.

Further validation of hub gene expression levels revealed that ASPM, BRCA1, CCNA2, CDK1, PLK1 and TOP2A all showed significantly high expression trends in HCC tumor tissues. This result was consistent with those of differential expression analysis and functional enrichment analysis, suggesting that these hub genes may be closely related to the enhanced cell proliferation status in HCC tissues.

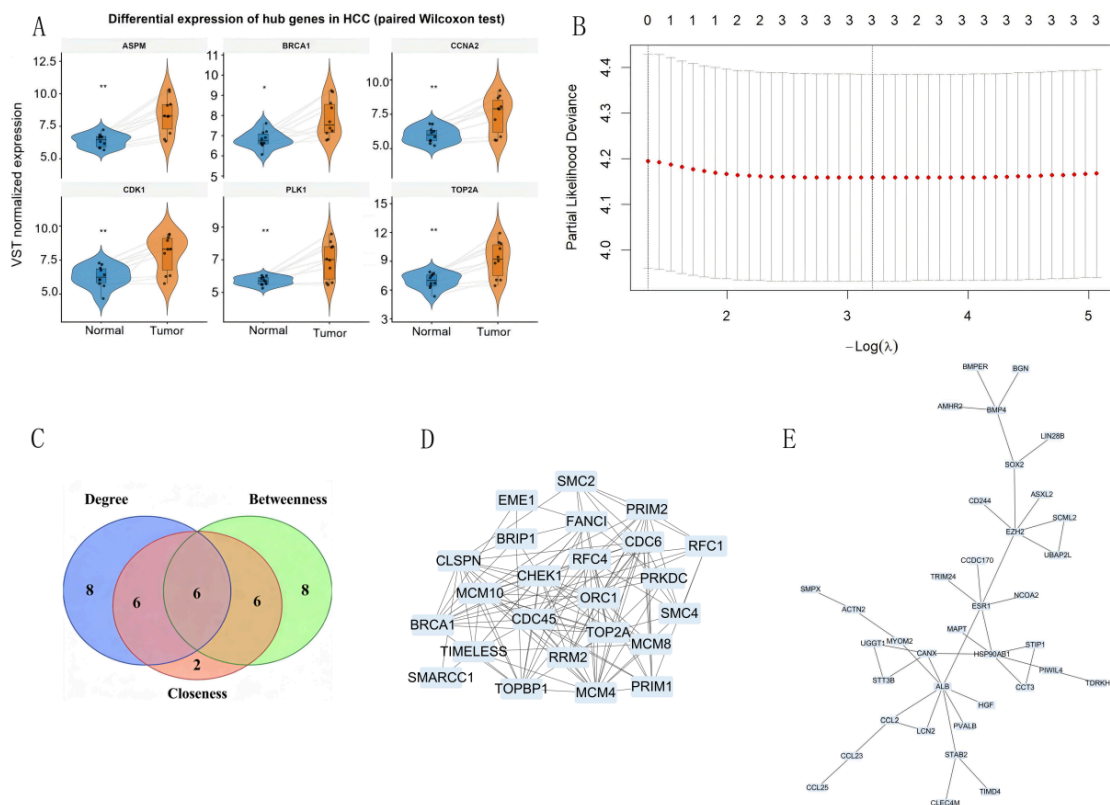


Figure 2. Results of PPI network module analysis, hub gene screening and expression validation

Figure 2 comprehensively shows the main functional modules in the PPI network, the Venn diagram of hub genes screened by three centrality indicators, the expression differences of six hub genes between HCC tissues and adjacent non-tumor tissues, and the cross-validation curve of Elastic Net Cox regression.

3.3. Association between hub genes and survival prognosis of HCC patients

To evaluate the prognostic value of candidate hub genes, Kaplan-Meier survival analysis was performed on hub genes based on the TCGA-LIHC cohort. Patients were divided into high-expression and low-expression groups according to gene expression levels, and differences in overall survival between the two groups were compared.

The results showed that high expression of all six hub genes was associated with poorer overall survival in HCC patients. Specifically, patients in the high-expression groups of ASPM, BRCA1, CCNA2, CDK1, PLK1, and TOP2A exhibited shorter overall survival than those in the low-expression groups, with statistically significant differences. These findings indicate that these hub genes are not only significantly upregulated in HCC tissues but also correlate with poor patient prognosis, suggesting potential prognostic value.

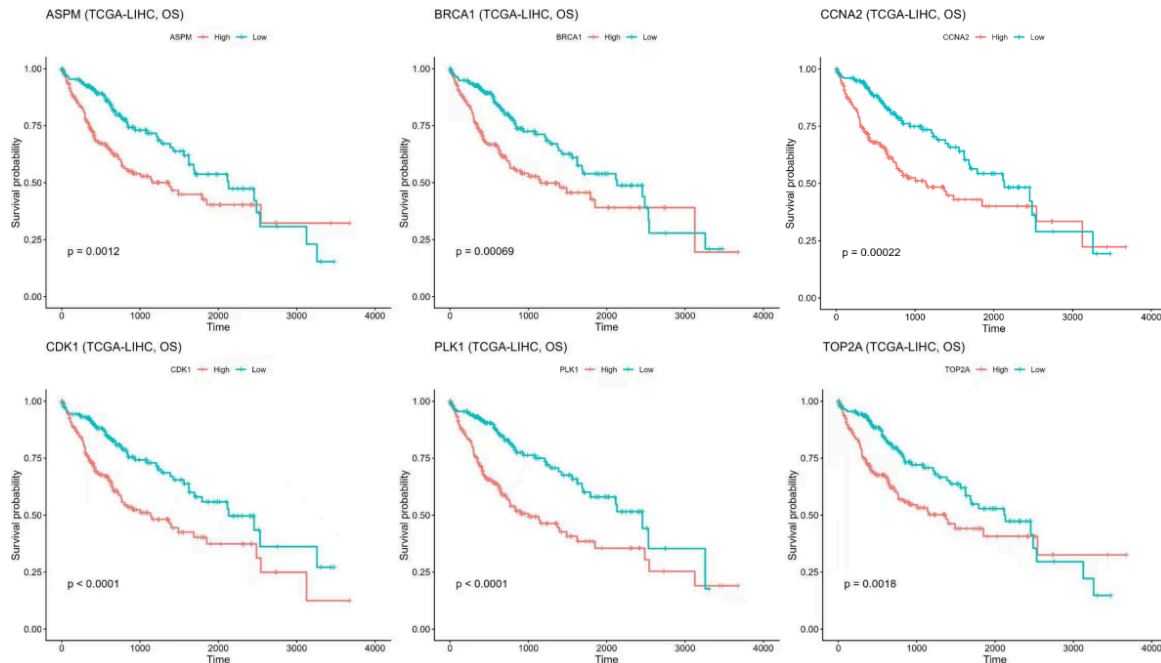


Figure 3. Results of Kaplan-Meier survival analysis for hub genes

In Figure 3, Kaplan-Meier survival analysis was conducted for the six hub genes (ASPM, BRCA1, CCNA2, CDK1, PLK1, and TOP2A) based on the TCGA-LIHC cohort. Patients were stratified into high-expression and low-expression groups by gene expression levels, and overall survival differences were compared. The results demonstrate that high expression of all six hub genes is associated with worse prognosis.

3.4. Construction and predictive performance evaluation of the risk score model

To further assess the combined prognostic value of hub genes, a risk score model was constructed based on candidate hub genes. TCGA-LIHC patients were divided into high-risk and low-risk groups according to the median risk score. Kaplan-Meier survival analysis revealed that the overall survival rate of patients in the high-risk group was significantly lower than that in the low-risk group, indicating that the model can effectively distinguish HCC patients with different prognostic risks to a certain extent.

Time-dependent receiver operating characteristic (ROC) curves were further used to evaluate the model's predictive ability for 1-, 3-, and 5-year overall survival. The results showed that the area under the curve (AUC) values of the model for 1-, 3-, and 5-year survival prediction were 0.731, 0.689, and 0.678, respectively, demonstrating certain predictive efficacy at different time points. The 1-year predictive performance was relatively better, while the 3- and 5-year predictive abilities

decreased slightly. Overall, this risk score model has certain prognostic stratification value but requires further validation in independent cohorts and more clinical samples.

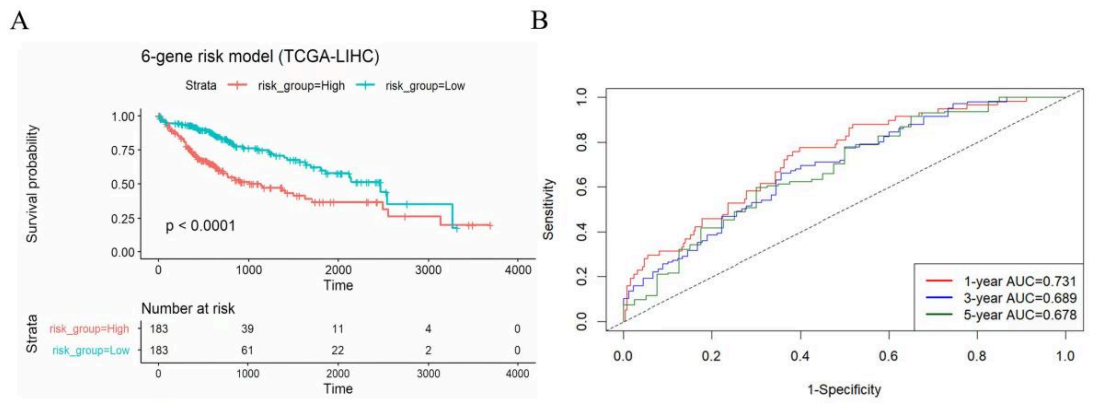


Figure 4. Survival Analysis and time-dependent ROC curves of the risk score model

In Figure 4, patients were divided into high-risk and low-risk groups by the median risk score; patients in the high-risk group had significantly shorter overall survival than those in the low-risk group. ROC curves show that the model has certain predictive ability for 1-, 3-, and 5-year survival.

3.5. Summary of results

In summary, this study identified 1333 differentially expressed genes based on the GSE202853 dataset, including 753 upregulated genes in HCC tissues and 580 genes with relatively high expression in adjacent non-tumor tissues. Functional enrichment analysis revealed that differentially expressed genes were mainly enriched in tumor-related processes such as cell cycle, DNA replication, chromosome segregation, and the p53 signaling pathway. Meanwhile, immune-related scores and immune cell composition also showed certain changes, indicating the coexistence of abnormal proliferation and immune microenvironment remodeling in HCC tissues.

Through PPI network module analysis and network centrality metric screening, six hub genes (ASPM, BRCA1, CCNA2, CDK1, PLK1, and TOP2A) were identified. All these genes were significantly upregulated in HCC tissues, and their high expression was associated with poor patient prognosis. The risk score model constructed based on hub genes effectively distinguished high-risk and low-risk patients and exhibited certain predictive ability for 1-, 3-, and 5-year survival.

4. Discussion

This study systematically analyzed HCC-related differentially expressed genes, functional enrichment patterns, immune microenvironment alterations, core genes in the PPI network, and prognostic models by integrating public databases including GSE202853 and TCGA-LIHC. The results revealed significant transcriptomic differences between HCC tissues and adjacent non-tumor tissues, with the relevant differentially expressed genes primarily involved in biological processes such as cell cycle regulation, DNA replication, chromosome segregation, and mitosis. Six candidate hub genes, namely ASPM, BRCA1, CCNA2, CDK1, PLK1, and TOP2A, were further identified, and high expression of these genes in HCC was associated with poorer overall survival. The risk score model constructed based on these genes also demonstrated certain prognostic stratification ability, suggesting that these molecules may possess potential prognostic value.

Functional enrichment analysis showed that differentially expressed genes were significantly concentrated in processes including nuclear division, mitotic nuclear division, chromosome segregation, Cell cycle, and DNA replication, indicating that abnormal activation of the cell cycle and enhanced DNA replication may be important features of transcriptomic alterations in HCC. Tumor cells typically acquire sustained proliferative advantages by breaking through cell cycle constraints, evading checkpoint regulation, and enhancing replication capacity. As a highly heterogeneous malignant tumor, HCC progression is often accompanied by genomic instability, abnormal chromosome segregation, and enhanced proliferation-related signaling. The results of this study are consistent with these tumor biological characteristics, suggesting the presence of active proliferation and dysregulated cell cycle control in HCC tissues [19].

In addition to proliferation-related pathways, this study also found that differentially expressed genes were involved in pathways such as the p53 signaling pathway, Wnt signaling pathway, and Cholesterol metabolism. Abnormalities in the p53 [20] pathway can impair cell cycle checkpoint and DNA damage response functions, while activation of the Wnt/ β -catenin signaling pathway is closely associated with HCC occurrence, progression, and maintenance of cancer stemness [21]. Meanwhile, cholesterol metabolism and lipid metabolic reprogramming are also considered to be related to HCC cell growth, invasion, and therapeutic response [22]. Therefore, molecular alterations in HCC are not merely manifested as enhanced cell proliferation but simultaneously involve multiple aspects including abnormal signal transduction, altered metabolic status, and disrupted genomic stability.

PPI network analysis further screened six candidate hub genes: ASPM, BRCA1, CCNA2, CDK1, PLK1, and TOP2A. Most of these genes are related to cell cycle progression, DNA replication, mitosis, and maintenance of chromosome stability. Specifically, CDK1 and CCNA2 participate in G2/M phase transition and mitotic regulation; PLK1 plays a crucial role in centrosome maturation, spindle formation, and chromosome segregation; TOP2A is involved in DNA topological structure regulation and chromosome untangling; BRCA1 is mainly associated with DNA damage repair and genomic stability; and ASPM is related to spindle function and cell division processes. The overall high expression trend of these hub genes in HCC tissues further supports the abnormal activation of proliferation- and mitosis-related programs in HCC.

Survival analysis results indicated that high expression of all six hub genes was associated with poorer overall survival in HCC patients. Since these genes are centrally involved in cell cycle, DNA replication, and mitotic regulation, their elevated expression may reflect stronger proliferative capacity and higher malignant progression risk in tumor cells. Therefore, these genes can serve as candidate molecules for HCC prognostic evaluation. However, the expression differences and survival correlations obtained in this study based on public databases remain correlative evidence and cannot directly prove their causal roles in HCC development. Further validation through cell experiments, animal models, and clinical samples is still required.

This study also analyzed alterations in immune-related functional status and immune cell composition. The results showed that immune-related pathways such as COMPLEMENT, INTERFERON ALPHA RESPONSE, and INFLAMMATORY RESPONSE exhibited differences or changing trends between HCC tissues and adjacent non-tumor tissues, and CIBERSORTx analysis also suggested compositional changes in some immune cell types. HCC development is not only affected by intrinsic molecular abnormalities in tumor cells but also closely related to immune microenvironment remodeling [23]. T cells, macrophages, NK cells, and other immune cells can influence tumor progression and immunotherapeutic response through inflammatory responses, immunosuppression, antigen presentation, and cytokine secretion. Therefore, combining differential

expression analysis with immune infiltration analysis helps to understand the biological characteristics of HCC from both the perspectives of tumor cell abnormalities and microenvironment alterations.

Regarding the prognostic model, this study constructed a risk score model based on candidate hub genes and evaluated its predictive performance using time-dependent ROC curves [24]. The results showed that the AUC values of the model for 1-, 3-, and 5-year overall survival prediction were 0.731, 0.689, and 0.678, respectively, indicating certain prognostic discrimination ability. The 1-year predictive effect was relatively better, while the 3- and 5-year AUC values decreased slightly, suggesting that the model may be more suitable for short-term survival risk assessment. The decline in long-term predictive performance may be attributed to factors such as treatment modalities, tumor staging, underlying liver disease status, and the complexity of clinical factors with prolonged follow-up. Therefore, this model is currently more suitable as an exploratory prognostic tool and requires further optimization and validation by incorporating independent cohorts and more clinical variables.

This study has several limitations. First, the research mainly relies on public databases, and potential biases may arise from sample sources, sequencing platforms, and data processing workflows. Second, although GSE202853 includes paired tumor and adjacent non-tumor tissues, which helps reduce the impact of individual differences, the sample size is still relatively limited. Third, hub gene screening is primarily based on expression differences, network topological features, and survival correlations, lacking experimental functional validation. Fourth, immune infiltration analysis is inferred from bulk RNA-seq data and cannot fully replace experimental methods such as single-cell sequencing, flow cytometry, or immunohistochemistry. Finally, although the prognostic model demonstrates certain predictive value, its stability and generalizability still require further evaluation in independent HCC cohorts.

In summary, this study preliminarily revealed alterations in cell cycle, DNA replication, and immune-related functional status in HCC tissues by integrating differential expression, functional enrichment, immune infiltration, PPI network, and survival analyses, and identified potential key genes including ASPM, BRCA1, CCNA2, CDK1, PLK1, and TOP2A. High expression of these hub genes is associated with poor survival prognosis, and the risk score model constructed based on them also exhibits certain prognostic stratification ability. These results provide a reference for understanding the molecular mechanisms of HCC and screening potential prognostic biomarkers, but further experimental validation and external cohort verification are still needed to support the findings.

5. Conclusion

In summary, this study integrated transcriptomic differential expression analysis, functional enrichment analysis, immune infiltration analysis, protein-protein interaction (PPI) network analysis, and survival analysis to screen key genes associated with hepatocellular carcinoma (HCC) and evaluate their potential prognostic value. The results showed significant gene expression differences between HCC tissues and paired adjacent non-tumor tissues. Differentially expressed genes were mainly enriched in biological processes such as cell cycle regulation, DNA replication, and chromosome segregation, suggesting that abnormal cell proliferation and cell cycle dysregulation may be important features of transcriptomic alterations in HCC.

In addition, by constructing a PPI network and screening with topological indicators, this study identified six candidate hub genes: ASPM, BRCA1, CCNA2, CDK1, PLK1, and TOP2A. These genes were significantly upregulated in HCC tissues, and their high expression was correlated with

poorer overall survival. The risk score model constructed based on candidate hub genes also demonstrated certain prognostic stratification ability, indicating that these key genes may have potential value as candidate biomarkers for prognostic evaluation of HCC.

Overall, this study provides a reference for further understanding the molecular mechanisms of HCC and screening potential prognostic biomarkers. However, this study is mainly based on public database analyses, and the relevant results still require further validation in independent clinical cohorts and experimental studies.

References

- [1] Hwang, S. Y., Danpanichkul, P., Agopian, V., Mehta, N., Parikh, N. D., Abou-Alfa, G. K., Singal, A. G., & Yang, J. D. (2025). Hepatocellular carcinoma: Updates on epidemiology, surveillance, diagnosis and treatment. *Clinical and Molecular Hepatology*, 31(Suppl), S228–S254. <https://doi.org/10.3350/cmh.2024.0824>
- [2] Chan, Y. T., Zhang, C., Wu, J., Lu, P., Xu, L., Yuan, H., Feng, Y., Chen, Z. S., & Wang, N. (2024). Biomarkers for diagnosis and therapeutic options in hepatocellular carcinoma. *Molecular Cancer*, 23(1), 189. <https://doi.org/10.1186/s12943-024-02101-z>
- [3] Guo, D. Z., Zhang, X., Zhang, S. Q., Zhang, S. Y., Zhang, X. Y., Yan, J. Y., Dong, S. Y., Zhu, K., Yang, X. R., Fan, J., Zhou, J., & Huang, A. (2024). Single-cell tumor heterogeneity landscape of hepatocellular carcinoma: Unraveling the pro-metastatic subtype and its interaction loop with fibroblasts. *Molecular Cancer*, 23(1), 157. <https://doi.org/10.1186/s12943-024-02062-3>
- [4] Liu, Y., Xun, Z., Ma, K., Liang, S., Li, X., Zhou, S., Sun, L., Liu, Y., Du, Y., Guo, X., Cui, T., Zhou, H., Wang, J., Yin, D., Song, R., Zhang, S., Cai, W., Meng, F., Guo, H., Zhang, B., Yang, D., Bao, R., Hu, Q., Wang, J., Ye, Y., & Liu, L. (2023). Identification of a tumour immune barrier in the HCC microenvironment that determines the efficacy of immunotherapy. *Journal of Hepatology*, 78(4), 770–782. <https://doi.org/10.1016/j.jhep.2023.01.011>
- [5] Cheng, K., Cai, N., Zhu, J., Yang, X., Liang, H., & Zhang, W. (2022). Tumor-associated macrophages in liver cancer: From mechanisms to therapy. *Cancer Communications*, 42(11), 1112–1140. <https://doi.org/10.1002/cac2.12345>
- [6] Hossen, M. A., Reza, M. S., Rana, M. M., Hossen, M. B., Shoaib, M., Mollah, M. N. H., & Han, C. (2024). Identification of most representative hub-genes for diagnosis, prognosis, and therapies of hepatocellular carcinoma. *Chinese Clinical Oncology*, 13(3), 32. <https://doi.org/10.21037/cco-23-151>
- [7] Dai, Q., Liu, T., Gao, Y., Zhou, H., Li, X., & Zhang, W. (2021). Six genes involved in prognosis of hepatocellular carcinoma identified by Cox hazard regression. *BMC Bioinformatics*, 22(1), 167. <https://doi.org/10.1186/s12859-021-04095-7>
- [8] Clough, E., & Barrett, T. (2016). The Gene Expression Omnibus Database. *Methods in Molecular Biology*, 1418, 93–110. https://doi.org/10.1007/978-1-4939-3578-9_5
- [9] Cancer Genome Atlas Research Network. (2017). Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell*, 169(7), 1327–1341.e23. <https://doi.org/10.1016/j.cell.2017.05.046>
- [10] Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550. <https://doi.org/10.1186/s13059-014-0550-8>
- [11] Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS*, 16(5), 284–287. <https://doi.org/10.1089/omi.2011.0118>
- [12] Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., & Ishiguro-Watanabe, M. (2023). KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Research*, 51(D1), D587–D592. <https://doi.org/10.1093/nar/gkac963>
- [13] Hänzelmann, S., Castelo, R., & Guinney, J. (2013). GSVA: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*, 14, 7. <https://doi.org/10.1186/1471-2105-14-7>
- [14] Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., Khodadoust, M. S., Esfahani, M. S., Luca, B. A., Steiner, D., Diehn, M., & Alizadeh, A. A. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nature Biotechnology*, 37(7), 773–782. <https://doi.org/10.1038/s41587-019-0114-2>
- [15] Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., & von Mering, C. (2023). The STRING database in 2023: Protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, 51(D1), D638–D646. <https://doi.org/10.1093/nar/gkac1000>

- [16] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11), 2498–2504. <https://doi.org/10.1101/gr.1239303>
- [17] Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1–22.
- [18] Blanche, P., Dartigues, J. F., & Jacqmin-Gadda, H. (2013). Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Statistics in Medicine*, 32(30), 5381–5397. <https://doi.org/10.1002/sim.5958>
- [19] Li, L., Huang, K., Zhao, H., Chen, B., Ye, Q., & Yue, J. (2020). CDK1-PLK1/SGOL2/ANLN pathway mediating abnormal cell division in cell cycle may be a critical process in hepatocellular carcinoma. *Cell Cycle*, 19(10), 1236–1252. <https://doi.org/10.1080/15384101.2020.1749471>
- [20] Edamoto, Y., Hara, A., Biernat, W., Terracciano, L., Cathomas, G., Riehle, H. M., Matsuda, M., Fujii, H., Scoazec, J. Y., & Ohgaki, H. (2003). Alterations of RB1, p53 and Wnt pathways in hepatocellular carcinomas associated with hepatitis C, hepatitis B and alcoholic liver cirrhosis. *Int J Cancer*, 106(3), 334–341. <https://doi.org/10.1002/ijc.11254>
- [21] Leung, R. W. H., & Lee, T. K. W. (2022). Wnt/ β -Catenin signaling as a driver of stemness and metabolic reprogramming in hepatocellular carcinoma. *Cancers (Basel)*, 14(21), 5468. <https://doi.org/10.3390/cancers14215468>
- [22] Cao, D., & Liu, H. (2023). Dysregulated cholesterol regulatory genes in hepatocellular carcinoma. *Eur J Med Res*, 28(1), 580. <https://doi.org/10.1186/s40001-023-01547-z>
- [23] Zheng, J., Chen, J., Wang, S., Yang, D., & Zhou, P. (2024). Genomic and immune landscape in hepatocellular carcinoma: Implications for personalized therapeutics. *Environ Toxicol*, 39(3), 1601–1616. <https://doi.org/10.1002/tox.24062>
- [24] Dessie, E. Y., Tu, S. J., Chiang, H. S., Tsai, J. J. P., Chang, Y. S., Chang, J. G., & Ng, K. L. (2021). Construction and validation of a prognostic gene-based model for overall survival prediction in hepatocellular carcinoma using an integrated statistical and bioinformatic approach. *Int J Mol Sci*, 22(4), 1632. <https://doi.org/10.3390/ijms22041632>