DOI: 10.14188/j.1671-8852.2024.0045
文章编号 1671-8852(2025)05-0644-11

基于生物信息学分析筛选肝癌核心基因并构建诊断预后模型

陈伊伊 荣 媛
武汉大学中南医院检验科 湖北 武汉 430071

摘要 目的：通过生物信息学多维分析，筛选肝细胞癌（HCC）特异性生物标志物，探索其诊断及预后预测价值。方法：从GEO数据库中下载HCC数据集GSE136247和TCGA数据库中肝癌相关基因表达数据。利用R包edgeR分别对两个队列数据进行差异表达基因（DEGs）的分析并取交集。随后基于GO、KEGG、Reactive和WikiPathway数据库对DEGs进行富集和功能注释。利用STRING数据库构建基于DEGs的蛋白质相互作用（PPI）网络，并通过Cytoscape软件获取PPI网络中的关键基因。利用GEPIA数据库、HPA数据库及GEO额外数据队列验证关键基因表达，使用逐步回归法筛选关键基因，构建肝癌预后Cox模型，并通过二元逻辑回归构建诊断模型。通过R包rms绘制Nomogram图，R包IOBR的Timer算法探索核心基因与免疫浸润的关联；R包clusterProfiler进行核心基因的GSEA分析；Spearman相关性分析核心基因与免疫检查点基因的关联；R包oncoPredict探索核心基因的药物相关性。结果：在肝细胞癌和癌旁组织中共识别出432个DEGs和15个关键基因，通过一系列生物信息学分析方法筛选出了两个高风险分子*CCT3*和*KPNA2*，并构建双分子联合Cox模型。双分子联合Cox模型分数和肿瘤分期是独立的临床风险评估指标，该风险模型用于预测患者1、3、5、7年预后具有价值。同时，Nomogram图对患者预后预测与实际吻合良好。此外，基于双分子逻辑回归诊断模型可有效区分癌与非癌（$AUC>0.95$）。免疫相关性分析表明，*CCT3*和*KPNA2*表达与免疫细胞浸润水平有显著相关性。GSEA分析显示，*CCT3*和*KPNA2*富集通路主要在E2F-TARGETS和G2M-CHECKPOINT；免疫检查点相关性分析显示，*CCT3*和*KPNA2*都与CD276相关性最高；药物相关性分析发现*CCT3*和*KPNA2*都对西地尼布敏感。结论：该研究成功筛选出HCC特异性生物标志物*CCT3*和*KPNA2*，基于双分子的逻辑回归诊断模型、Cox风险模型和Nomogram图，有效用于HCC诊断及预后预测，*CCT3*和*KPNA2*可作为潜在的治疗靶点。

关键词 肝细胞癌；生物信息学；生物标志物；诊断及风险模型
中图分类号 R735.7
文献标识码 A
开放科学（资源服务）标识码（OSID）：

Unveiling hepatocellular carcinoma through bioinformatics: Screening core genes and constructing diagnosticand prognostic models

CHEN Yiyi, RONG Yuan
Dept. of Clinical Laboratory, Zhongnan Hospital of Wuhan University, Wuhan 430071, Hubei, China

Abstract

Objective: To screen the specific biomarkers of hepatocellular carcinoma (HCC) through multi-dimensional analysis of bioinformatics and explore their diagnostic and prognostic value. **Methods:** The HCC dataset GSE136247 from the GEO database and the gene expression data related to liver cancer from the TCGA database were downloaded. The R package edgeR was used to ana-

作者简介：陈伊伊，女，1996—，医学硕士生，主要从事肝癌生物标志物研究，E-mail：2021283030211@whu.edu.cn
通信作者：荣媛，女，1980—，医学博士，副教授，硕士生导师，主要从事肿瘤免疫学研究，E-mail：whushine@163.com

lyze the differential expression genes (DEGs) of the data from the two cohorts separately and the intersection was taken. DEGs were subsequently enriched and functionally annotated based on GO, KEGG, Reactive, and WikiPathway databases. The protein-protein interaction (PPI) network based on DEGs was constructed by the STRING database, and the key genes in the PPI network were obtained by Cytoscape software. GEPIA database, HPA database, and the GEO additional data cohort were used to verify the expression of key genes, and the stepwise regression method was used to screen the key genes to construct the Cox model of liver cancer prognosis, and binary logistic regression was used to construct the diagnostic model. The Nomogram was drawn by the R package rms, and the Timer algorithm of the R package IOBR was used to explore the association between core genes and immune infiltration. The R package clusterProfiler was used for the GSEA analysis of core genes. Spearman correlation analysis was used to analyze the association between core genes and immune checkpoint genes. The R package oncoPredict explores the drug associations of core genes. **Results:** A total of 432 DEGs and 15 key genes were identified in HCC and adjacent tissues. Two high-risk molecules, *CCT3* and *KPNA2*, were screened out by a series of bioinformatics analysis methods, and a bi-molecular Cox model was constructed. The bimolecular Cox model score and tumor stage are independent clinical risk assessment indicators, and the risk model has value in predicting the 1-, 3-, 5-, and 7-year prognosis of patients. At the same time, the prognosis predicted by the Nomogram was in good agreement with the actual prognosis. In addition, the bimolecular logistic regression model could effectively distinguish cancer from non-cancer ($AUC>0.95$). Immune correlation analysis showed that the expression of *CCT3* and *KPNA2* was significantly correlated with the level of immune cell infiltration. GSEA analysis showed that *CCT3* and *KPNA2* were mainly enriched in E2F-TARGETS and G2M-checkpoints. Immune checkpoint correlation analysis showed that *CCT3* and *KPNA2* were the most correlated with CD276. **Conclusion:** *CCT3* and *KPNA2* are HCC-specific biomarkers successfully screened in this study. The bi-molecular logistic regression diagnostic model, Cox risk model, and Nomogram can be effectively used for HCC diagnosis and prognosis prediction.

Key Words　Hepatocellular Carcinoma; Bioinformatics; Biomarkers; Diagnostic and Risk Models

原发性肝癌是全球肿瘤的第三大死亡原因，在中国其病死率位居第二[1]，是影响建设健康中国的重大难题。肝细胞癌（hepatocellular carcinoma，HCC）是原发性肝癌中最常见的类型[2]，发病隐匿、早期诊断困难、生长速度快、复发转移率高，出现转移的肝癌患者5年生存率仅为18%[3]。常见的治疗方法包括手术切除、放化疗法和肝移植等[4]。除早期肝癌可通过手术切除痊愈外，目前对于晚期HCC仍缺乏有效治疗方法[5]。因此，筛选HCC发展的特异性生物标志物，有利于早期肝癌的诊断和指导生物标志物匹配的分子靶向治疗。随着生物信息学分析方法和高通量测序技术的不断发展，这使得快速筛选关键基因，并深入研究这些基因在肝癌诊断和预后中的作用成为了可能[6]。既往研究[7]已发现多个基因和分子参与肝细胞癌的进展，但其具体的机制尚未完全阐明。

本研究对2个数据库中与HCC相关的数据集进行生物信息学多维分析，筛选出参与肝癌的关键基因，构建逻辑回归诊断模型、Cox风险模型并绘制诺莫图用于HCC诊断及预后预测。首先，在基因表达数据库（gene expression omnibus，GEO）中筛选并下载肝癌数据集GSE136247并与TCGA数据库取交集，比较鉴定人HCC组织与非癌性肝组织之间的差异表达基因（differentially expressed genes，DEGs），并对这些DEGs进行功能注释和通路富集分析。通过可视化蛋白质相互作用（protein-protein interaction，PPI）网络最终筛选得到15个关键基因进行生物信息学分析。最后获得*CCT3*和*KPNA2*两个高风险分子，并对双分子进行通路、免疫相关分析以及药物敏感性分析，结果表明双分子模型可

用于肝癌诊断和预后预测,并作为潜在的免疫治疗靶点。

1 材料与方法

1.1 数据来源 从GEO(http://www.ncbi.nlm.nih.gov/geo)数据库中筛选并下载HCC相关数据集GSE136247中69例样本,其中39例肿瘤,30例癌旁组织。从UCSC-xena数据库(https://xenabrowser.net/datapages/)下载TCGA-LIHC的RNA-*seq*原始count矩阵数据,共424例样本,其中包括374例肿瘤和50例癌旁样本。两个数据库共包括413例肿瘤组织和80例癌旁组织样本。

1.2 表达差异分析 利用R包edgeR进行差异基因分析。导入RNA-*seq*原始count计数矩阵,基于CPM(counts per million)去除低表达基因,TMM(trimmed mean of M-values)将数据标准化以校正样本间的技术误差。使用广义线性模型拟合数据并评估基因在不同分组的差异(Wald检验),最后通过BH(Benjamini-Hochberg)法进行*P*值校正,其中错误发现率(FDR)<0.05的基因为显著差异表达基因。

1.3 通路富集分析 采用g:Profiler数据库(Https://biit.cs.ut.ee/gprofiler/gost)进行基因功能富集分析。背景数据库选择GO、KEGG、Reactome(REAC)和Wikipathway(WP),使用BH法进行*P*值校正。

1.4 PPI网络构建和核心基因的筛选 利用NetworkAnalyst(Https://www.NetworkAnalyst.ca)构建肝脏特异性PPI网络,选择STRING背景数据库,置信分数设为900,纳入实验验证的蛋白互作信息。以Degree>50为截断值筛选出排名前15(Top15)的核心基因。

1.5 核心基因的表达和ROC曲线分析 使用GEPIA工具(http://gepia.cancer-pku.cn/)基于TCGA-LIHC样本对top15核心基因进行表达分析和生存分析,表达分析中|$\log_2$Fold Change,FC|>1,*P*<0.01的基因被认为具有显著差异,生存分析中按基因中位表达对患者进行分组,log rank *P*<0.05的基因被认为与生存显著相关。通过R包pROC绘制ROC曲线评估不同基因的诊断性能。

1.6 核心基因在蛋白质水平的验证 通过人类蛋白质图谱(human protein atlas,HPA)数据库(www.proteinatlas.org)获取正常肝组织和肝癌组织的免疫组化染色标本,验证核心基因在肝癌和正常组织中的蛋白质表达水平。

1.7 双分子模型构建和验证 使用R包My.stepwise和R包survival基于核心基因进行逐步回归筛选,获取用于预测患者生存的最佳Cox风险模型,其中两个分子*CCT3*和*KPNA2*作为最终的模型构建分子,并基于以下公式计算双分子风险分数:$Risk\ score=0.446\ 5\times Exp_{KPNA2}+0.217\ 5\times Exp_{CCT3}$。

本研究纳入患者年龄、性别、肿瘤分期、分级等临床特征与双分子风险分数进行单因素和多因素Cox分析,评估变量的风险关联及独立预后预测能力。使用KM曲线评估风险分数的生存预测能力,其中R包tdROC用于进行时间依赖的ROC曲线分析,评估风险分数对患者1、3、5、7年的生存预测效能。使用二元逻辑回归构建基于双分子的肝癌诊断模型,并使用ROC曲线在两个独立的队列评估其对癌和癌旁的区分效能。同时,采用R包rms绘制Nomogram图评估和预测肝癌患者的预后。

1.8 *CCT3*与*KPNA2*的基因通路分析与免疫相关性分析 基于TCGA-LIHC RNA-*seq*原始count矩阵数据,依据*CCT3*和*KPNA2*中位表达对样本分组,利用edgeR包分析差异基因,按$\log_2$FC排序作为基因富集分析(Gene Set Enrichment Analysis,GSEA)输入。背景基因集选用MSigDB的H:hallmark癌症通路数据集。显著富集的通路要求FDR(BH校正)<0.05,NES值的正或负分别暗示通路的激活或抑制状态。利用R包IOBR的Timer法推断样本免疫细胞浸润分数,通过Spearman相关分析*CCT3*、*KPNA2*与免疫细胞浸润评分及免疫检查点的相关性。基于癌症细胞药物敏感性(Genomics of Drug Sensitivity in Cancer,GDSC)数据库,使用R包oncoPredict初步估算了198种药物在TCGA-LIHC样本中的50%抑制浓度(IC_{50}),通过Spearman相关分析基因表达与每个样本中推断的药物IC_{50}值相关性,保留*P*<0.05的药物,通过R包ggplot2可视化正、负相关排名前15的药物。负相关表明高表达CCT3/KPNA2的患者可能对该药物敏感,正相关则耐受。

2 结果

2.1 肝癌组织和非癌组织之间差异基因的鉴定 从GSE136247与TCGA数据库中获取HCC与非癌性肝组织的基因表达谱,然后计算两数据集各自的

差异基因，分别获得 653、5 756 个 DEGs（如图 1A、B），然后交叉筛选两组数据集的差异分析结果，共筛选出 432 个重叠的 DEGs（图 1C）。

2.2　DEGs 的 GO、KEGG、REAC 和 WP 富集分析　如图 1D 所示，GO 分析包括差异基因的生物过程（biological process，BP）、分子功能（molecular function，MF）及细胞组分（cellular component，CC）等内容。差异基因参与的生物过程包括肌酸代谢过程、羧酸代谢过程和氧酸代谢过程。差异基因的分子功能主要包括催化活性、氧化还原酶活性。差异基因的细胞组分主要富集到细胞外空间、细胞外区域和细胞外囊泡等。在 KEGG 分析中，差异基因富集的信号通路主要在代谢途径、药物代谢-细胞色素 P450、补体和凝血级联、脂肪酸降解等方面。在 REAC 分析中，差异基因主要富集在新陈代谢、生物氧化反应、脂质代谢方面。在 WP 分析中，差异基因富集通路主要在补体系统、核受体通路、原途径生物转化Ⅰ期和Ⅱ期。这些结果表明肝癌相关差异基因与肝脏的代谢功能密切相关。

2.3　DEGs 的 PPI 网络分析和关键基因的筛选　基于 STRING 平台建立差异表达基因 PPI 网络（图 2），Cytoscape 软件用于可视化 SRTING 结果，使用 Cytoscape 中的 CytoHubba 插件计算 Degree 值并按大小进行排序，以 Degree＞50 为标准，从显著的差异基因中选出居前 15 位的关键基因，分别为 *ESR1*、*HSP90AB1*、*AR*、*GABARAPL1*、*ECT2*、*FOS*、*PCK1*、*ANXA2*、*KPNA2*、*SPP1*、*CCT3*、*GADD45G*、*SHBG*、*NR0B2* 和 *PPARGC1A*。

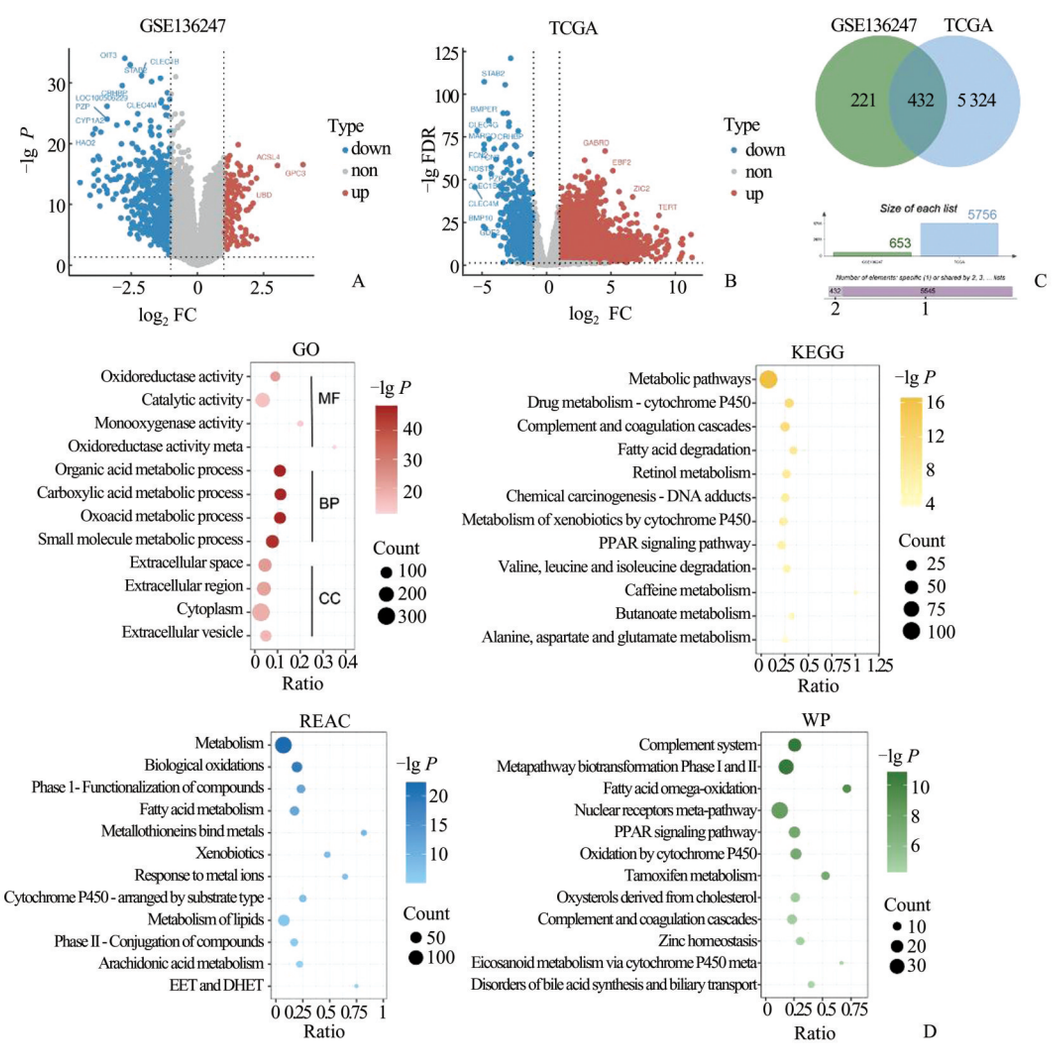


图 1　GSE136247、TCGA 数据集差异基因火山图、差异基因的韦恩图、富集分析

A. GSE136247 数据集差异基因火山图；B. TCGA 数据集差异基因火山图；C. 2 个高通量数据集差异基因的韦恩图；D. GO、KEGG、REAC 和 WP 富集分析图

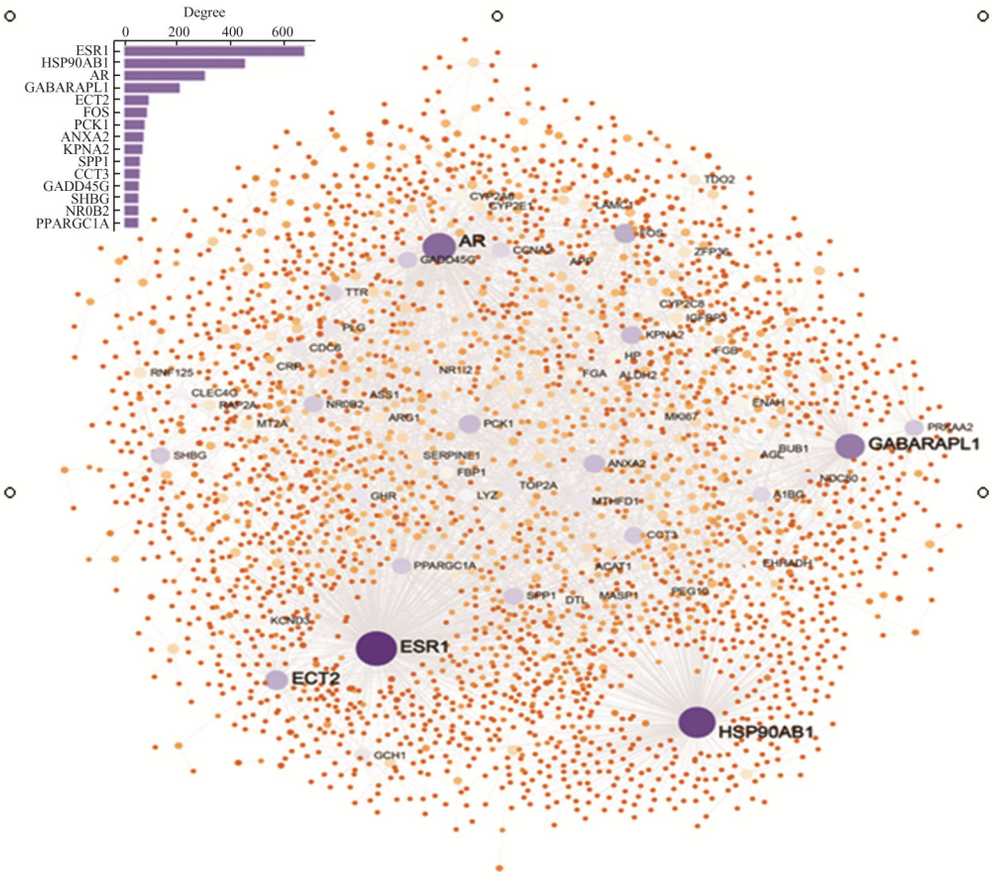


图2 肝细胞癌差异基因的PPI网络分析图(Degree>50,Top15基因)

2.4 关键基因的差异表达分析、生存分析和ROC曲线分析 在GEO和TCGA数据库中验证15个基因在肿瘤中的mRNA水平。结果表明(图3),两个数据库中15个基因的表达一致,6个基因(*CCT3*、*ANXA2*、*KPNA2*、*SPP1*、*HSP90AB1*和*ECT2*)表达水平上调,9个基因(*ESR1*、*GADD45G*、*SHBG*、*NR0B2*、*AR*、*FOS*、*PCK1*、*PPARGC1A*、*GABARAPL1*)表达水平下调。以$P<0.01$为标准的有9个基因,分别为*CCT3*、*ANXA2*、*KPNA2*、*HSP90AB1*、*ECT2*、*SPP1*、*PCK1*、*ESR1*、*FOS*。随后,使用GEPIA数据库对15个基因进行生存分析,以log rank $P<0.05$为标准,在前期筛选的9个基因中,除*FOS*基因,其余基因的生存预后与其在mRNA表达水平相一致(图4)。利用R包pROC对15个关键基因进行ROC曲线分析。结果显示(图5),有13个基因的ROC曲线下面积(*AUC*)均大于90%,在前期筛选的9个关键基因中,其*SPP1*基因的ROC曲线下面积小于90%。综合以上分析结果,我们筛选出了7个核心基因,分别是*CCT3*、*ANXA2*、*KPNA2*、*HSP90AB1*、*ECT2*、*PCK1*、*ESR1*。

2.5 核心基因在HPA数据库中蛋白质表达水平验证 使用HPA数据库中核心基因的免疫组化结果。由于*ECT2*在HPA数据库中不存在免疫组织化学信息,我们显示了*CCT3*、*KPNA2*、*HSP90AB1*、*ANXA2*、*ESR1*、*PCK1*的结果(图6)。与mRNA表达类似的是,*CCT3*、*KPNA2*、*HSP90AB1*、*ANXA2*在正常组织中的蛋白表达水平较低,而在肝癌细胞中这些基因的蛋白表达水平则较高。*ESR1*、*PCK1*在肝癌组织中的蛋白水平相对正常组织较低。

2.6 基于*CCT3*和*KPNA2*构建模型及验证 逐步回归法进一步分析7个核心基因得到高危风险分子*CCT3*和*KPNA2*,并基于此构建双分子联合Cox模型(图7A)。单因素Cox分析(图7B1)与多因素Cox分析独立风险评估(图7B2)显示,双分子联合Cox模型分数和肿瘤不同的分期是独立的临床风险评估指标($P<0.000\ 1$)(图7C~7E)。该风险模型用于预测患者1、3、5、7年的有较好的准确性(图7F),并对患者无疾病生存,疾病无进展生存和疾病特异性生存同样具有较高预测价值($P<0.01$)

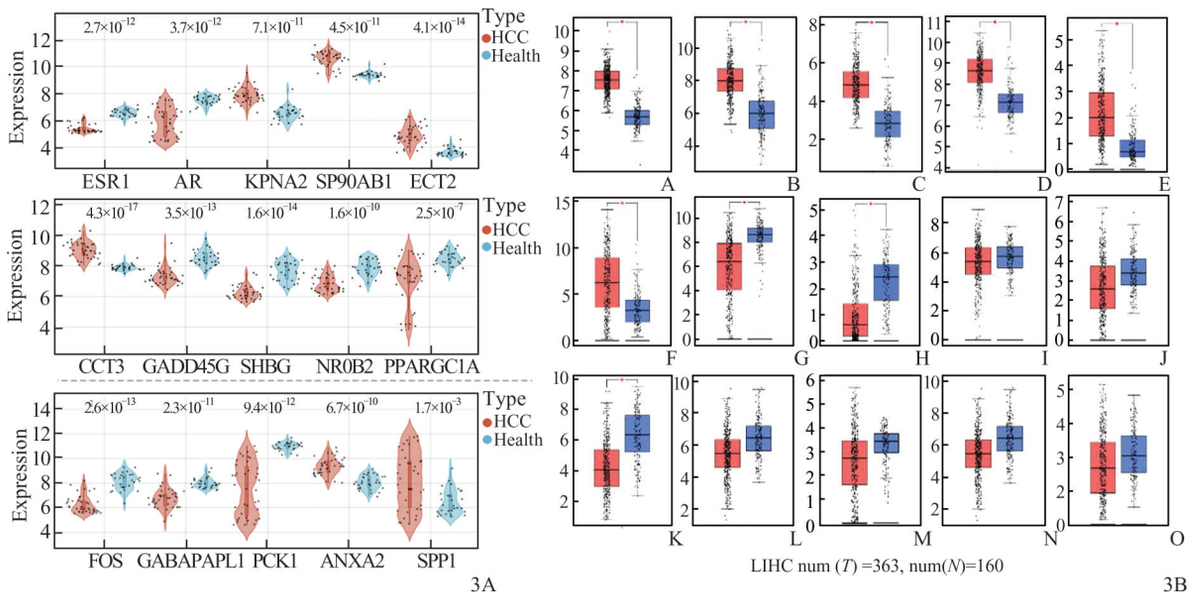


图 3　不同数据库中验证 15 个关键基因的表达水平

3A. GEO 数据库；3B. TCGA 数据库；3B 中 A～O 分别显示 *CCT3*、*ANXA2*、*KPNA2*、*HSP90AB1*、*ECT2*、*SPP1*、*PCK1*、*ESR1*、*NR0B2*、*PPARGC1A*、*FOS*、*GABARAPL1*、*AR*、*GADD45G*、*SHBG* 的表达水平

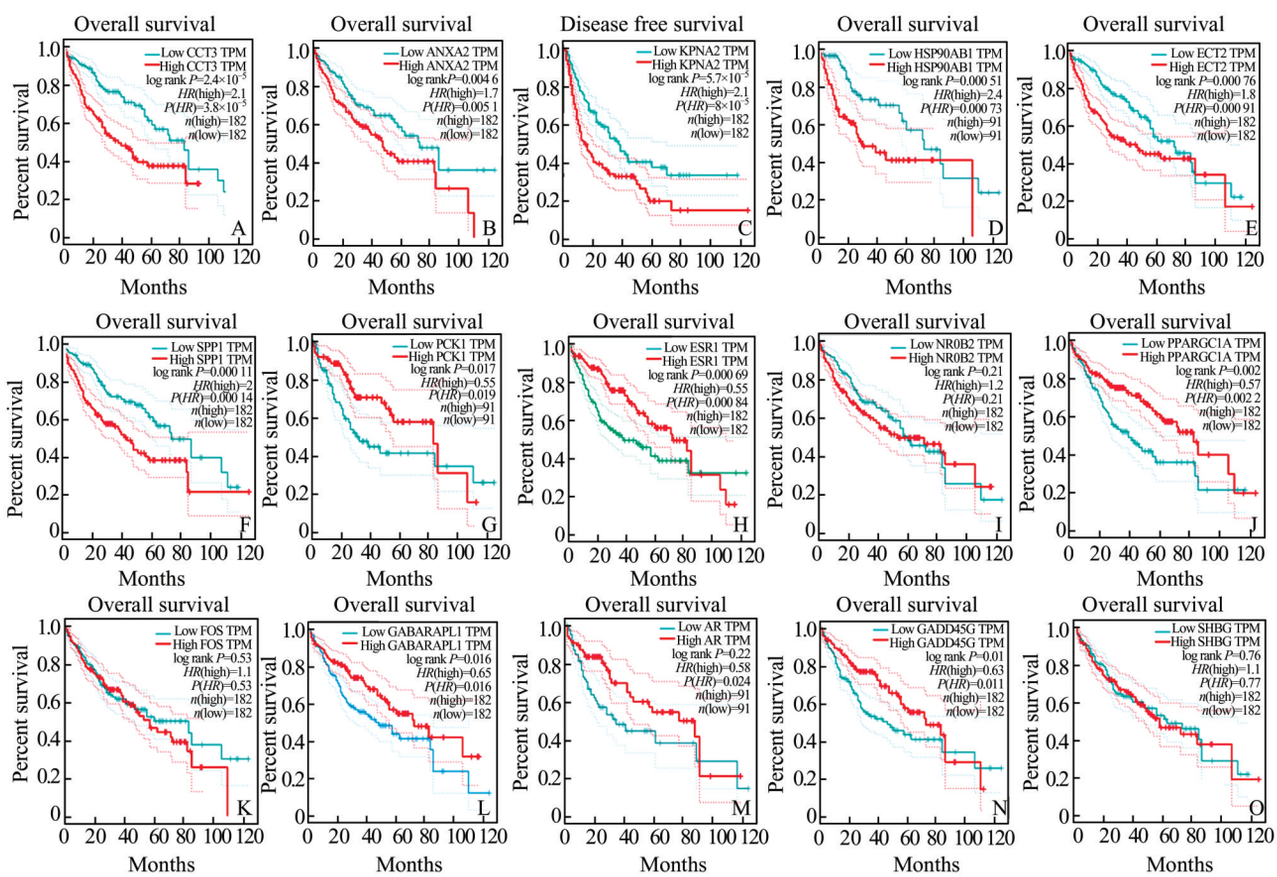


图 4　15 个关键基因的 mRNA 表达水平与生存预后一致性分析

A～O 分别显示 *CCT3*、*ANXA2*、*KPNA2*、*HSP90AB1*、*ECT2*、*SPP1*、*PCK1*、*ESR1*、*NR0B2*、*PPARGC1A*、*FOS*、*GABARAPL1*、*AR*、*GADD45G*、*SHBG* 的生存曲线

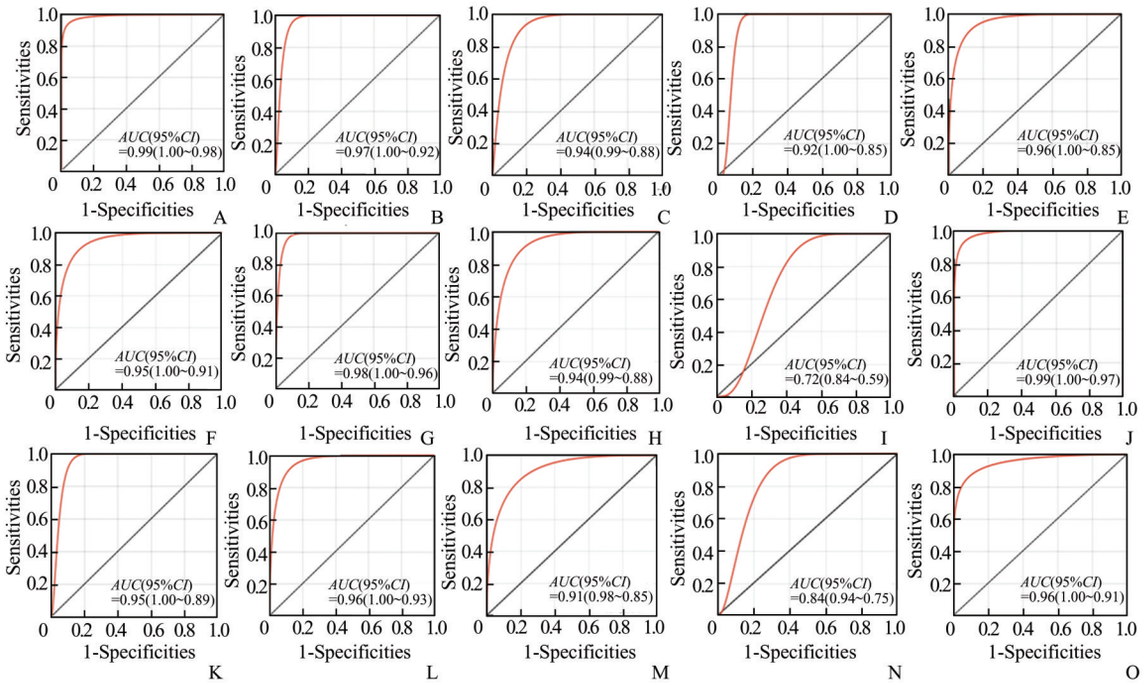


图5 15个关键基因的ROC曲线分析

A～O分别显示*ESR1*、*HSP90AB1*、*AR*、*GABARAPL1*、*ECT2*、*FOS*、*PCK1*、*ANXA2*、*SPP1*、*CCT3*、*GADD45G*、*SHBG*、*NR0B2*、*PPARGC1A*、*KPNA2*的ROC曲线

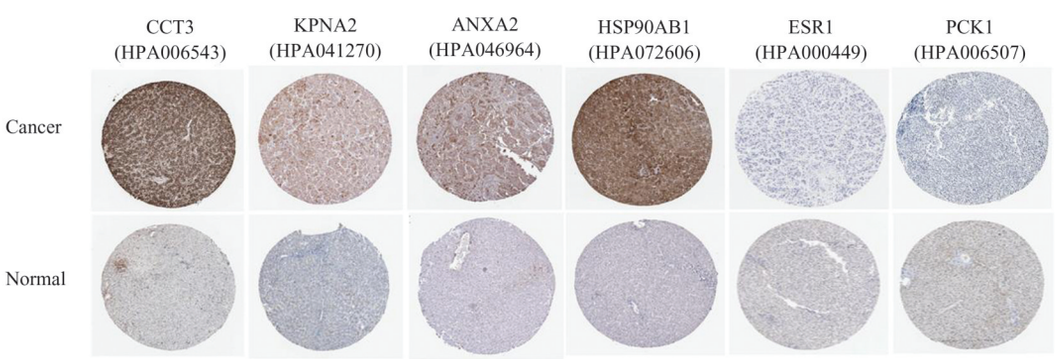


图6 HCC患者CCT3、KPNA2、ANXA2、HSP90AB1、ESR1、PCK1的蛋白质表达水平

棕色代表蛋白高表达,蓝色代表蛋白低表达

(图7G1～G3),此外该诊断模型可有效区分癌与非癌(*AUC*>0.95)(图7H),在用于肝癌早期诊断方面,表现出优秀的诊断效能。同时,基于*CCT3*和*KPNA2*与临床特征构建诺莫图,综合Nomogram图中的总体得分,预测肝癌患者的生存概率(图8A),该Nomogram图的C-index=0.731,具有较好的预测能力。校准分析显示:Nomogram图预测肝癌患者1、3、5和7年生存概率与实际观察结果吻合良好(图8B～8E)。

2.7 核心基因的肿瘤免疫相关分析及通路和药物敏感性探索 Timer分析结果显示,*CCT3*表达与$CD4^+$ T细胞、B细胞、巨噬细胞、中性粒细胞和树突状细胞的浸润丰度呈正相关,与$CD8^+$ T细胞浸润丰度呈负相关(图9A)。*KPNA2*表达与$CD4^+$ T细胞、B细胞、巨噬细胞、中性粒细胞和树突状细胞的浸润丰度正相关性,但与$CD8^+$T细胞无显著相关性(图9B)。免疫检查点相关性分析结果显示,大多数基因均与*CCT3*和*KPNA2*表达呈正相关(P<0.01),其中最相关的免疫检查点分子是*CD276*(图9C)。GSEA分析表明,*CCT3*和*KPNA2*的富集通路主要在E2F-TARGETS、G2M-CHECKPOINT(NES>2,P<0.05)(图9D),表明*CCT3*和*KPNA2*与细胞增殖具有密切关联。药物敏感性分析结果显示,*CCT3*和*KPNA2*都与cediranib药物高度负相关

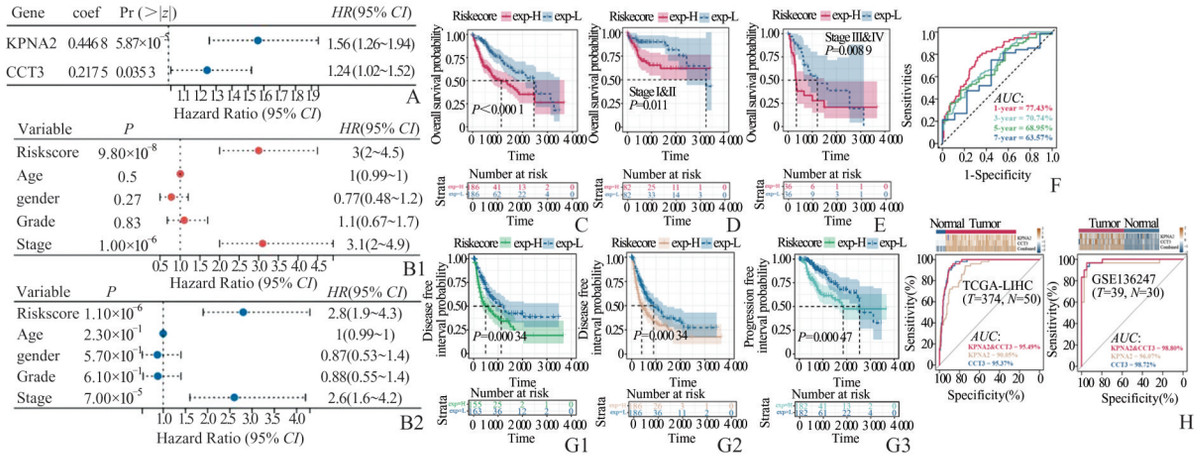


图 7　构建双分子联合 Cox 模型

A. 逐步回归模型；B1、B2. 年龄、性别、肿瘤分期、分级与双分子风险分数的单因素与多因素 Cox 回归分析；C～E. 双分子 Cox 模型分数和肿瘤分期临床风险评估；F. 1、3、5、7 年时间依赖 ROC 曲线；G1～G3. 患者无疾病生存、疾病无进展生存和疾病特异性生存预测；H. 2 个数据集验证模型区分癌与非癌

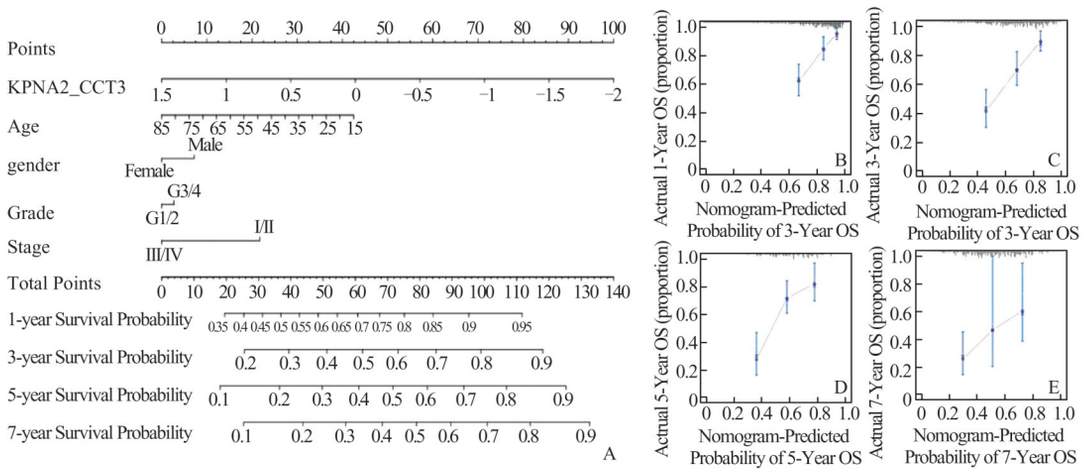


图 8　基于 *CCT3* 和 *KPNA2* 的 Nomogram 图和 Calibration 图

A. 预测肝癌患者的生存概率；B～E. 校准分析

（图 9E），表明当患者高表达 CCT3 和 KPNA2 时，使用该药物可能会获益更多。

3　讨论

肝细胞癌的发生涉及多种基因通过不同途径的共同作用，尽管肝细胞癌的诊断和治疗水平有所提高，但其早期诊断和预后治疗的效果仍未达到理想水平[8]。生物信息学分析具有处理大规模数据、整合多维数据、预测生物标志物和靶点，以及数据可视化和解释等优势。它在生物学研究、疾病诊断和药物开发中发挥重要作用，并为个性化医学提供关键支持。因此，有必要通过生物信息学分析筛选和鉴定在肝癌发生和发展过程中起关键作用的特异性基因，并构建相应的诊断和预后模型，以提升肝癌的诊断准确性和预后评估能力。

在本研究中，我们筛选了 432 个 DEGs。通过多种通路富集分析发现 DEGs 主要参与了肝癌细胞的代谢调控、氧化还原平衡以及细胞间通信等关键过程。例如，糖酵解代谢是肝癌发展的重要机制之一[9]，并且癌细胞代谢的改变可能诱导活性氧的增加，导致正常细胞损伤，促进肿瘤的发展[10]。研究[11]表明，细胞间通信通过多种信号分子和通路促使肝癌细胞的增殖、侵袭和转移，如细胞外泌体通过携带 miRNA、蛋白质和其他生物分子传递信息，促进肝癌细胞的增殖、侵袭和转移。

本研究通过构建 DEGs 的 PPI 网络，以 Degree＞50 标准，筛选了前 15 个关键基因，用 GEPIA 数据库和 GEO 额外数据队列对 15 个关键基因

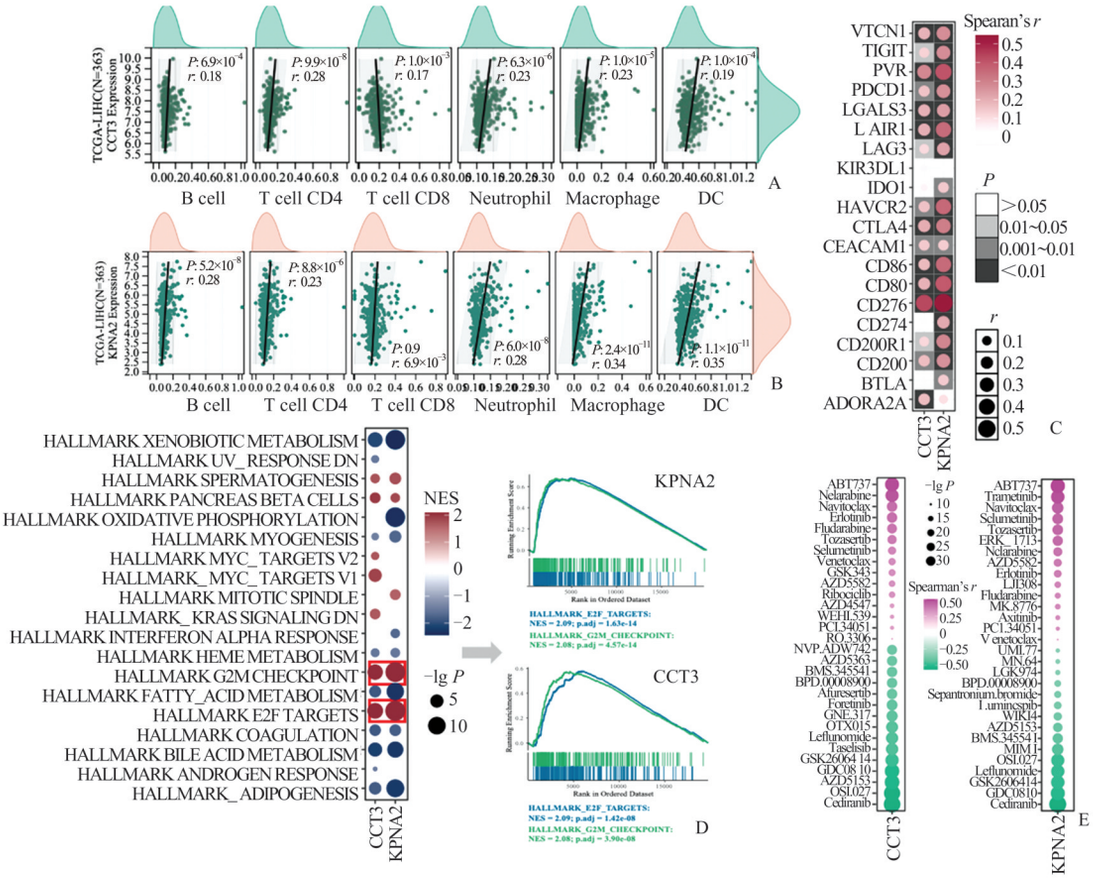


图 9 核心基因的肿瘤免疫相关分析

A、B. *CCT3*、*KPNA2* 与 6 种免疫细胞浸润水平的相关性；C. *CCT3* 和 *KPNA2* 的免疫检查点分析；D. *CCT3* 和 *KPNA2* 的通路富集分析；E. *CCT3* 和 *KPNA2* 与药物敏感性分析×10^4

进行在 mRNA 水平差异表达分析、生存分析和 ROC 曲线分析，综合结果筛选出了 7 个核心基因，它们分别是 *CCT3*、*KPNA2*、*HSP90AB1*、*ANXA2*、*ESR1*、*PCK1*。这些基因可能与肝细胞癌(HCC)的诊断和治疗相关，并有潜力成为生物标志物。具体而言，*CCT3* 是 TCP-1 环状复合物的一个亚单位，参与蛋白质的折叠和组装过程[12]。在 HCC 组织中 mRNA 和蛋白质表达高于癌旁肝组织，可能通过调控 JAK/STATs 信号通路来影响肝癌的增殖和侵袭转移[13]，对于甲胎蛋白(alpha-fetoprotein，AFP)阴性的 HCC 患者，CCT3 蛋白水平与 HCC 的肿瘤大小、TNM 阶段与 CCT3 表达的程度呈正相关[14]。KPNA2 是一种核蛋白，在 HCC 组织中高表达，与肿瘤分化、血管浸润，早期复发和转移有关[15]。ANXA2 蛋白在肝癌组织中高表达，其表达水平与患者是否感染乙肝、肿瘤大小及门脉癌栓密切相关[16]。ECT2 在许多其他癌症中也发挥着促进肿瘤的作用，如非小细胞肺癌和乳腺癌[17-18]，其可通过 ECT2/PLK1/PTEN 途径促进肝癌中肿瘤相关巨噬细胞的极化，因此可作为诊断和预后的生物标志物[19]。此外，PCK1 和 ESR1 蛋白在肝癌组织中的表达水平较低，可能作为抑癌基因参与肝癌的发展[20]。

通过对这 7 个关键基因的进一步分析，本研究筛选出了两个高危风险分子，分别是 *CCT3* 和 *KPNA2*，并构建了双分子联合 Cox 模型与 Nomogram 图。结果表明，双分子联合 Cox 模型的评分与肿瘤分期是独立的临床风险评估指标。该双分子模型与诺莫图显示在预测患者生存概率上有较好的预测价值。此外，基于 *CCT3* 和 *KPNA2* 的诊断模型在早期肝癌的诊断中展现出极高的诊断效能。研究[21]发现，肿瘤细胞与肿瘤微环境的影响在肿瘤发生过程中有重要作用，对免疫治疗的效果也有重大影响。LI 等[22]研究显示在肝细胞癌中，*CCT3* 的高表达与免疫细胞浸润有显著相关性，并且认为在肝细胞癌中巨噬细胞主要以 M2 型存在，M2 巨噬细胞在肿瘤微环境中有促进肿瘤增殖和逃避免疫监视的作用，因此巨噬细胞浸润较多 HCC 患者的预后不佳。刘晓红等[23]研究表明，在肝癌细胞中 *KP-*

*NA2*影响多种免疫细胞的浸润水平，*KPNA2*通过下调巨噬细胞M1水平促进HCC细胞免疫逃逸，促进肿瘤细胞增殖。本研究结果中*CCT3*和*KPNA2*与巨噬细胞浸润显著关联（$P<0.000\ 1$）提示*CCT3*和*KPNA2*可能通过调控巨噬细胞浸润，*CCT3*和*KPNA2*主要富集E2F和G2M通路；E2F是细胞周期关键调控因子，G2M是检查点相关的基因，它们在调控细胞周期、DNA修复、细胞凋亡、和代谢等过程中发挥着关键作用，促进肝癌增殖和转移。*CCT3*和*KPNA2*都与免疫检查点CD276高度正相关，CD276为肿瘤高表达的抗原，能够抑制免疫细胞的活性，促进肿瘤进展、介导肿瘤免疫逃逸[24]。药物敏感性分析显示，高表达*CCT3*和*KPNA2*的患者与西地尼布呈高度负相关，表明其对西地尼布的药物敏感。西地尼布是一种VEGFR酪氨酸激酶的高效抑制剂，可通过阻断肿瘤血管生成抑制肿瘤生长，该药物在肝癌治疗中的应用已进入临床研究阶段，但其副作用较大，一定程度上限制了临床应用，还需进一步研发。

综上所述，本研究成功筛选出HCC特异性生物标志物*CCT3*和*KPNA2*。基于双分子构建的逻辑回归诊断模型、Cox风险模型和Nomogram图表明其可有效用于HCC诊断及预后预测。双分子的表达水平与肝癌的预后不良呈正相关，主要通过E2F和G2M通路与免疫检查点CD276相关通路促进肿瘤发展和免疫逃逸，其高表达患者对西地尼布的药物敏感。*CCT3*和*KPNA2*可作为HCC诊断和预后的新生物标志物和潜在的免疫治疗靶点，同时对临床药物治疗有指导作用。但尚存在不足，需要更多的队列及相应基础实验对本研究的预测结果进行验证并探讨其作用机制。

参考文献

[1] ZHOU J, SUN H C, WANG Z, et al. Guidelines for the diagnosis and treatment of primary liver cancer (2022 edition)[J]. Liver Cancer, 2023, 12(5): 405-444.

[2] AIZARANI N, SAVIANO A, SAGAR, et al. A human liver cell atlas reveals heterogeneity and epithelial progenitors[J]. Nature, 2019, 572(7768): 199-204.

[3] SU T, ZHANG N S, WANG T, et al. Super enhancer-regulated LncRNA LINC01089 induces alternative splicing of $DIAPH_3$ to drive hepatocellular carcinoma metastasis[J]. Cancer Res, 2023, 83(24): 4080-4094.

[4] SELÇUK H. Prognostic factors and staging systems in hepatocellular carcinoma[J]. Exp Clin Transplant, 2017, 15(Suppl 2): 45-49.

[5] GROHMANN M, WIEDE F, DODD G T, et al. Obesity drives STAT-1-dependent NASH and STAT-3-dependent HCC[J]. Cell, 2018, 175(5): 1289-1306. e20.

[6] PROKOP J W, MAY T, STRONG K, et al. Genome sequencing in the clinic: The past, present, and future of genomic medicine[J]. Physiol Genomics, 2018, 50(8): 563-579.

[7] CHEN C T, CHEN C F, LIN T Y, et al. Traditional Chinese medicine Kuan-Sin-Yin Decoction inhibits cell mobility via downregulation of CCL2, CEACAM1 and PIK3R3 in hepatocellular carcinoma cells[J]. J Ethnopharmacol, 2023, 317: 116834.

[8] SHAHINI E, PASCULLI G, SOLIMANDO A G, et al. Updating the clinical application of blood biomarkers and their algorithms in the diagnosis and surveillance of hepatocellular carcinoma: A critical review[J]. Int J Mol Sci, 2023, 24(5): 4286.

[9] ZHOU J M, XU W Q, WU Y B, et al. GPR37 promotes colorectal cancer liver metastases by enhancing the glycolysis and histone lactylation via Hippo pathway [J]. Oncogene, 2023, 42(45): 3319-3330.

[10] LI B H, CAO Y J, MENG G, et al. Targeting glutaminase 1 attenuates stemness properties in hepatocellular carcinoma by increasing reactive oxygen species and suppressing Wnt/beta-catenin pathway[J]. EBioMedicine, 2019, 39: 239-254.

[11] LIU Z Y, LI Y N, WANG Y F, et al. Exosomes in HBV infection[J]. Clin Chim Acta, 2023, 538: 65-69.

[12] CUI X, HU Z P, LI Z, et al. Overexpression of chaperonin containing TCP1, subunit 3 predicts poor prognosis in hepatocellular carcinoma[J]. World J Gastroenterol, 2015, 21(28): 8588-8604.

[13] LIU Y, ZHANG X, LIN J F, et al. CCT3 acts upstream of YAP and TFCP2 as a potential target and tumour biomarker in liver cancer[J]. Cell Death Dis, 2019, 10(9): 644.

[14] QIAN E N, HAN S Y, DING S Z, et al. Expression and diagnostic value of CCT3 and IQGAP3 in hepatocellular carcinoma[J]. Cancer Cell Int, 2016, 16: 55.

[15] GUO X G, WANG Z H, ZHANG J N, et al. Upregulated KPNA2 promotes hepatocellular carcinoma progression and indicates prognostic significance across human cancer types[J]. Acta Biochim Biophys Sin, 2019, 51(3): 285-292.

[16] HERRERA-LÓPEZ E E, GUERRERO-ESCALERA D, AGUIRRE-MALDONADO I, et al. Annexins A2 and A5 are potential early biomarkers of hepatocarcinogenesis[J]. Sci Rep, 2023, 13(1): 6948.

[17] WANG H K, LIANG J F, ZHENG H X, et al. Ex-

pression and prognostic significance of ECT2 in invasive breast cancer[J]. J Clin Pathol, 2018, 71(5): 442-445.
[18] BAI X G, YI M, XIA X C, et al. Progression and prognostic value of ECT2 in non-small-cell lung cancer and its correlation with PCNA[J]. Cancer Manag Res, 2018, 10: 4039-4050.
[19] XU D F, WANG Y, WU J C, et al. ECT2 overexpression promotes the polarization of tumor-associated macrophages in hepatocellular carcinoma via the ECT2/PLK1/PTEN pathway[J]. Cell Death Dis, 2021, 12(2): 162.
[20] MENG J, WANG L H, HOU J, et al. CCL23 suppresses liver cancer progression through the CCR1/AKT/ESR1 feedback loop[J]. Cancer Sci, 2021, 112(8): 3099-3110.
[21] CHANDRA R, KARALIS J D, LIU C, et al. The colorectal cancer tumor microenvironment and its impact on liver and lung metastasis[J]. Cancers, 2021, 13(24): 6206.
[22] LI W L, LIU J, ZHAO H T. Prognostic power of a chaperonin containing TCP-1 subunit genes panel for hepatocellular carcinoma[J]. Front Genet, 2021, 12: 668871.
[23] 刘晓红,李紫荆,李琨,等. 核转运蛋白α-2在肝细胞癌中的表达及临床意义[J]. 中华实验外科杂志,2022, 39(10):1848-1852.
LIU X H, LI Z J, LI K, et al. Expression and clinical significance of nuclear transporter alpha-2 in hepatocellular carcinoma [J]. Chin J LAB Surgery, 2022, 39(10): 1848-1852.
[24] FENG Y, LEE J, YANG L P, et al. Engineering CD276/B7-H3-targeted antibody-drug conjugates with enhanced cancer-eradicating capability[J]. Cell Rep, 2023, 42(12): 113503.

(2024-01-18收稿)
编辑 沈建国

(上接第643页)

[9] WANG J C, SHI G F, CHEN C, et al. TLR2 derangements likely play a significant role in the inflammatory response and thrombosis in patients with Ph (−) classical myeloproliferative neoplasm[J]. Mediators Inflamm, 2024, 2024: 1827127.
[10] VINCELETTE N D, YU X Q, KUYKENDALL A T, et al. Trisomy 8 defines a distinct subtype of myeloproliferative neoplasms driven by the MYC-alarmin axis[J]. Blood Cancer Discov, 2024, 5(4): 276-297.
[11] KAWANO T, SHIMAMURA M, NAKAGAMI H, et al. Therapeutic vaccine against S100A9 (S100 calcium-binding protein A9) inhibits thrombosis without increasing the risk of bleeding in ischemic stroke in mice [J]. Hypertension, 2018, 72(6): 1355-1364.
[12] WANG X, GUAN M, ZHANG X H, et al. The association between S100A8/A9 and the development of very late stent thrombosis in patients with acute myocardial infarction[J]. Clin Appl Thromb Hemost, 2020, 26: 1076029620943295.
[13] KOVAČIĆ M, MITROVIĆ-AJTIĆ O, BELESLIN-ČOKIĆ B, et al. TLR4 and RAGE conversely mediate pro-inflammatory S100A8/9-mediated inhibition of proliferation-linked signaling in myeloproliferative neoplasms[J]. Cell Oncol (Dordr), 2018, 41(5): 541-553.
[14] KRAAKMAN M J, LEE M K, AL-SHAREA A, et al. Neutrophil-derived S100 calcium-binding proteins A8/A9 promote reticulated thrombocytosis and atherogenesis in diabetes[J]. J Clin Invest, 2017, 127(6): 2133-2147.
[15] CHEN X N, TAO T, WANG H Y, et al. Arterial thrombosis is accompanied by elevated mitogen-activated protein kinase (MAPK) and cyclooxygenase-2 (COX-2) expression *via* toll-like receptor 4 (TLR-4) activation by S100A8/A9[J]. Med Sci Monit, 2018, 24: 7673-7681.
[16] CARNEVALE R, SCIARRETTA S, VALENTI V, et al. Low-grade endotoxaemia enhances artery thrombus growth *via* Toll-like receptor 4: Implication for myocardial infarction[J]. Eur Heart J, 2020, 41(33): 3156-3165.
[17] SPELETAS M, LIADAKI K, KALALA F, et al. TLR4 single nucleotide polymorphisms and thrombosis risk in patients with myeloproliferative disorders[J]. Thromb Res, 2008, 122(1): 27-32.

(2024-10-18收稿)
编辑 张 峻