



doi:10.7659/j.issn.1005-6947.2023.01.005
http://dx.doi.org/10.7659/j.issn.1005-6947.2023.01.005
China Journal of General Surgery, 2023, 32(1):64-73.

· 专题研究 ·

基于E2F靶点基因集和免疫亚型差异的肝细胞癌预后 风险评分模型的建立

何锶¹, 赵杨², 朱永乾², 吴卓翼³, 吴英⁴, 谢君蓉⁴, 郑登焯⁴, 简红梅⁴

(中国人民解放军陆军军医大学基础医学院 1. 学员三大队十一队 2. 学员四大队十二队 3. 学员三大队九队, 重庆 400038;
4. 中国人民解放军陆军军医大学第一附属医院 肝胆外科, 重庆 400038)

摘要

背景与目的: 肝细胞癌(HCC)是肝癌中最常见的种类, HCC患者的预后生存情况较差, 其有效的预后预测也面临巨大挑战。许多研究已证实E2F基因家族和免疫微环境相关的基因标志物是癌症的重要预后因素, 因此, 本研究利用TCGA数据库筛选E2F基因家族和免疫微环境相关的HCC基因标志物, 建立新的HCC风险评分模型, 并预测HCC潜在治疗靶点。

方法: TCGA数据库中下载大型HCC(LIHC)队列(424例样本)。进行了基因集富集分析、基因集单样本富集分析和基因集单样本富集分析分数聚类后的基因表达差异分析, 通过Lasso回归筛选标志基因并建模, 根据模型计算患者得分并将患者分为高风险组和低风险组。使用受试者工作特征曲线(ROC)、Kaplan-Meier生存曲线、Cox回归分析等多种统计学方法以验证模型的可靠性。所有统计分析均使用R语言软件。最后在Cbioportal数据库查询风险模型的标志基因在TCGA-HCC样本中的基因变异情况, 从String数据库中下载蛋白互作信息并用Cytoscape软件进行可视化分析。

结果: 确认了与HCC密切相关的E2F靶点基因组和免疫相关差异基因后, 从中筛选到了与HCC患者总生存率明显相关的7个基因(CYR61, FBLN5, LPA, SAA1, SDC3, SERPINE1, SSRP1), 建立7-mRNA预后模型: 风险评分 = $-0.55 \times \text{CYR61 表达} - 0.18 \times \text{FBLN5 表达} - 0.17 \times \text{LPA 表达} - 0.06 \times \text{SAA1 表达} + 0.31 \times \text{SDC3 表达} + 0.38 \times \text{SERPINE1 表达} + 1.08 \times \text{SSRP1 表达}$ 。该模型ROC的AUC值为0.846。Kaplan-Meier生存曲线显示, 高风险评分患者预后不良($P < 0.001$), 高、低风险评分对预后的区分度与肿瘤大小和UICC分期相似, 而比淋巴转移、远处转移和BMI值更好。多因素Cox回归分析显示, 7-mRNA预后模型的预测能力独立于临床因素。此外, 联合蛋白组学找到7个基因中的关键基因SERPINE1和LPA, 预测抑制纤溶酶原激活可能是治疗HCC的新的靶途径。

结论: 本研究揭示了7个基因与E2F靶点和免疫的相关关系, 为HCC患者的不良预后提供了新的生物标志物, 并建立了有较高预测准确性预后风险评分模型。然而, 多基因预后模型的预测能力仍需大量多中心的循证医学证据证实, 被纳入的多基因模型的基因功能和参与的机制仍需进行更深入的研究。

关键词

癌, 肝细胞; E2F转录因子类; 免疫; 预后; 危险因素

中图分类号: R735.7

基金项目: 重庆市科技局技术创新与应用发展专项基金资助项目(CSTC2021jscx-gksb-N0009)。

收稿日期: 2022-04-06; **修订日期:** 2022-06-17。

作者简介: 何锶, 中国人民解放军陆军军医大学基础医学院本科生, 主要从事生物信息学分析方面的研究。

通信作者: 简红梅, Email: jhm702087727@163.com

Construction of prognostic assessment model for hepatocellular carcinoma based on E2F targets and immune subtype differences

HE Si¹, ZHAO Yang², ZHU Yongqian², WU Zhuoyi³, WU Ying⁴, XIE Junrong⁴, ZHENG Dengye⁴, JIAN Hongmei⁴

(1. The Eleventh Squadron of the Third Student Brigade 2. the Twelfth Squadron of the Fourth Student Brigade 3. the Ninth Squadron of the Third Student Brigade, Basic Medicine College of Army Medical University, Chongqing 400038, China; 4. Department of Hepatobiliary Surgery, the First Affiliated Hospital of Army Medical University, Chongqing 400038, China)

Abstract

Background and Aims: Hepatocellular carcinoma (HCC) is the most common type of liver cancer. The prognosis of HCC patients is poor, and its effective prognosis prediction is also facing significant challenges. Several studies have shown that the genetic markers associated with the E2F gene family and immune microenvironment are important prognostic factors for cancers. Therefore, this study was conducted to screen the HCC gene signatures related to the E2F gene family and immune microenvironment using the TCGA database, establish a new risk assessment model for HCC and predict the potential therapeutic targets for HCC.

Methods: A large HCC (LIHC) dataset ($n=424$) from the TCGA database was downloaded. Gene set enrichment analysis, single sample gene set enrichment analysis, and differential gene expression analysis was performed, marker genes were screened and modeled by Lasso regression, patient scores were calculated according to the model, and patients were divided into high-risk and low-risk groups. Multiple statistical methods, such as the receiver operating characteristic (ROC) curve, Kaplan-Meier survival curve, and Cox regression analysis, were used to verify the model's reliability. R language software was used for all statistical analyses. Finally, genetic alterations of the marker genes from the risk model were queried in the TCGA-HCC samples in the Cbioportal database. The protein interaction information was downloaded from the String database and visualized in Cytoscape software.

Results: After identification of the E2F target genome and immune-related differential genes which were closely related to HCC, seven genes (CYR61, fbln5, LPA, SAA1, SDC3, serpine1, ssrp1) significantly associated with the overall survival rate of HCC patients were screened, and a prognostic 7-mRNA signature model was established: risk score= $-0.55 \times \text{CYR61 expression} - 0.18 \times \text{FBLN5 expression} - 0.17 \times \text{LPA expression} - 0.06 \times \text{SAA1 expression} + 0.31 \times \text{SDC3 expression} + 0.38 \times \text{SERPINE1 expression} + 1.08 \times \text{SSRP1 expression}$. The ROC AUC value of the model was 0.846. Kaplan-Meier survival curve showed that patients with high-risk scores had a poor prognosis ($P < 0.001$). The degree of discrimination for prognosis of high and low-risk scores was similar to those of tumor size and UICC stage and higher than those of lymph node metastasis, distant metastasis, and BMI. Multivariate Cox regression analysis showed that the predictive ability of the 7-mRNA signature model was independent of clinical factors. In addition, the key genes SERPINE1 and LPA in the 7 genes were found by combining proteomics, which predicted that inhibiting plasminogen activation was probably a new target approach for treating HCC.

Conclusion: This study reveals the correlation between seven genes and E2F targets and immunity, provides new biomarkers for poor prognosis of HCC patients and establishes a prognostic risk score model with high predictive accuracy. However, the predictive ability of the polygenic prognosis model still needs to be confirmed by many evidence-based medical practices from multiple centers, and the gene function and participation mechanism of the included polygenic models still need to be further studied.

Key words Carcinoma, Hepatocellular; E2F Transcription Factors; Immunity; Prognosis; Risk Factors

CLC number: R735.7

原发性肝癌是目前我国第4位常见恶性肿瘤及第2位肿瘤致死病因,严重威胁我国人民的生命和健康^[1]。其中原发性肝细胞癌(hepatocellular carcinoma, HCC)是肝癌最常见的类型,发生率占原发性肝癌的75%~85%^[2]。尽管近年来原发性HCC的临床治疗模式不断发展,肝癌患者的预后生存情况仍然较差,患者5年生存率仅18%左右^[3],预后预测仍然面临巨大挑战。预后风险评估模型可以预测个体在将来是否发生某种结局。近年来,预测模型研究呈明显上升趋势,在肿瘤领域,如乳腺癌、前列腺腺癌、肺腺癌等中已有诸多应用^[4-6]。目前,在HCC领域,已有基于基因的风险评分模型被成功建立以预测患者预后^[7],但大多只是大范围地筛选基因,没有聚焦在某一已知与癌症发生相关的基因集上,这会为进一步深入的机制研究带来困难。因此,有目的地筛选HCC治疗的关键基因及相关通路,建立新的基因预后模型,预测HCC的潜在治疗靶点具有重大意义和可行性。

转录因子E2F相关基因在调节肿瘤进展,参与血管生成促进肿瘤转移等各方面起着重要作用^[8]。另一方面免疫疗法最近受到广泛关注,越来越多的证据^[9]表明免疫相关基因可预测肝癌预后。近年来,许多研究已证实E2F基因家族和免疫微环境相关的基因标志物是癌症的重要预后因素,E2F-1基因缺失可导致PTEN丢失诱导的锯齿状肿瘤的发生率和恶化程度增加^[10],17个免疫相关基因建立的分类系统能有效预测早期肺鳞状细胞癌患者的预后^[11],又如E2F2、3、6、8的高表达与胰腺癌患者的肿瘤分期有关且E2F基因的表达与免疫浸润显著相关^[12]。因此本研究旨在利用TCGA数据库开发一种新的E2F基因家族和免疫微环境相关的基因标志物,建立风险评估模型以预测HCC患者的预后并找到治疗靶点。

1 资料与方法

1.1 数据来源与数据预处理

本研究从TCGA数据库(<https://portal.gdc.cancer.gov>)中下载了424例Liver Hepatocellular Carcinoma

(LIHC)患者的临床特征以及60 484个mRNA的表达数据集。从分子特征数据库(MSIGDB)中下载共50组hallmark基因集(癌症特征基因集合),该基因集合表示了明确的生物状态或过程,下载免疫标志基因集,该基因集可表示免疫系统内的细胞状态和调动。基于上述详细数据,将424例LIHC患者根据TCGA数据库的样本编号原则分为有癌组(374例)和无癌组(50例)。此外,还记录了以下临床信息:性别、年龄、肿瘤大小、淋巴结转移数、远处器官转移情况、UICC分期和BMI。最后,对261例有明确生存时间且生存时间不为零的患者进行了分类。一般临床特征详见表1。

1.2 基因集富集分析与基因集单样本富集分析

使用“DESeq2”包进行基因表达量的差异分析得到9 149个差异基因($P<0.05$),进一步使用“clusterProfiler”包对这9 149个差异基因进行GSEA基因富集分析,以识别在有癌组和无癌组中存在显著统计学差异且富集程度最高的基因组用于进一步研究。基因集单样本富集分析(ssGSEA)是通过扩展GSEA实现的,使用签名中的基因和其余基因的经验累积分布函数生成富集分数^[13]。用“GSVA”包对374例有癌组样本进行基因集单样本富集分析,可得到各样本在免疫系统内不同细胞状态和调动下的富集得分,再使用“hclust”包对富集得分聚类得到两种免疫相关亚型,最后使用R包“DESeq2”对两组免疫亚型样本的mRNA表达量进行差异分析得到存在显著统计学差异的免疫亚型差异基因($P<0.05$)。

1.3 构建预后特征模型

使用glmnet包,采用Cox法进行10倍似然交叉验证,纳入LIHC中的261例有明确生存时间且生存时间不等于零的样本,分别对富集分析中富集得分最高的基因集的基因和免疫亚型差异基因通过Lasso回归进行变量筛选。同样使用glmnet包,采用Cox法进行十倍似然交叉验证,将两者筛选后的基因一起基于表达水平与Lasso回归分析得出的系数加权的线性组合,建立了预后风险评估模型。风险评估=基因1的表达 $\times \beta_1$ +基因2的表达 $\times \beta_2$ +...+基因 $n \times \beta_n$ 的表达,并绘制可用于临床评分的列线图。

表1 本研究中HCC患者的临床特征[n (%)]

Table 1 Clinical characteristics of HCC patients in this study [n (%)]

特征	数值
年龄(岁)	
≥59	145(55.6)
<59	115(44.1)
x	1(0.4)
性别	
男	183(70.1)
女	78(29.9)
T分期	
T1	141(61.7)
T2	69(26.4)
T3	46(17.6)
T4	3(1.1)
Tx	2(0.8)
N分期	
N0	180(69.0)
N1	3(1.1)
Nx	78(29.9)
M分期	
M0	199(76.2)
Mx	62(23.8)
UICC分期	
I	133(51.0)
II	65(24.9)
III	50(19.2)
IV	1(0.4)
x	12(4.6)

注:x为未知或无法评估

Note: x means unknown or cannot be assessed

1.4 预后模型的评估与验证

使用“ROCR”包对风险预测模型对截止时间内生存死亡结局的预测能力通过受试者工作特征曲线(ROC)及曲线下面积(AUC)值进行检验,以验证模型预测生存状态和生存时间的敏感度和特异度。以中位风险评分为分界点,通过中位值将261例患者分为高风险和低风险亚组。使用“survival”包和“survminer”包,绘制Kaplan-Meier生存曲线并进行Log-rank检验。相同地,对BMI值、淋巴转移、远处转移、肿瘤大小和UICC分期的常见临床指标绘制Kaplan-Meier生存曲线并进行Log-rank检验,与风险预测模型相比较。将通过Log-rank检验的各项指标进行单因素和多因素Cox回归分析,以检查风险评分是否是可用数据范围

内的一个独立预后因素。在考虑不同UICC分期亚型时,本研究还绘制了I、II分期和III、IV分期两亚型的高、低风险的Kaplan-Meier生存曲线,确立风险评估模型的适用范围。随后,本研究检测了有癌组和无癌组之间构成模型的基因的差异表达和基因改变情况,通过蛋白互作网络寻找模型中的关键基因以求发现新的可能药物靶点。372个TCGA HCC样本的构成风险模型的基因的改变量从Cbioportal数据库(<http://www.cbioportal.org>)查询,蛋白互作信息从String数据库(<https://cn.string-db.org>)中下载并在Cytoscape软件中分析可视化。

1.5 统计学处理

所有统计分析均使用R软件(版本4.1.0, <https://www.r-project.org>)进行, $P < 0.05$ 为差异有统计学意义。整体流程图如图1。

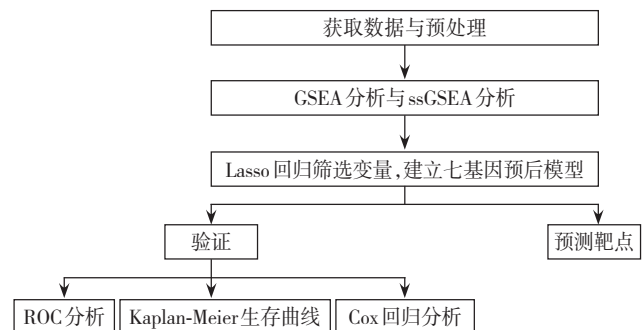


图1 整体流程图

Figure 1 Overall flow chart

2 结果

2.1 富集得分最高的E2F靶点及其相关基因

在差异基因中, E2F靶点基因组的200个基因富集程度高, 归一化后的富集得分最高, 标准化 P 值为 $1.00E-10$ 和多重假设检验校正后的 P 值(q 值)为 $6.25E-10$ (图2A-B)。Lasso回归分析筛选得到两个基因: SSRP1, KIF2C(图2C-D)。

2.2 免疫亚型及免疫相关差异表达基因

根据样本在免疫系统内不同细胞状态和调动下的富集得分进行聚类, 将样本分成了富集得分差异明显的两组(图3A)。两组免疫亚型样本差异存在统计学意义的基因($P < 0.05$)有242个, 其中23个下调, 219个上调(图3B)。Lasso回归分析筛选得到6个基因, 即CYR61, FBLN5, LPA, SAA1, SDC3, SERPINE1(图3C-D)。

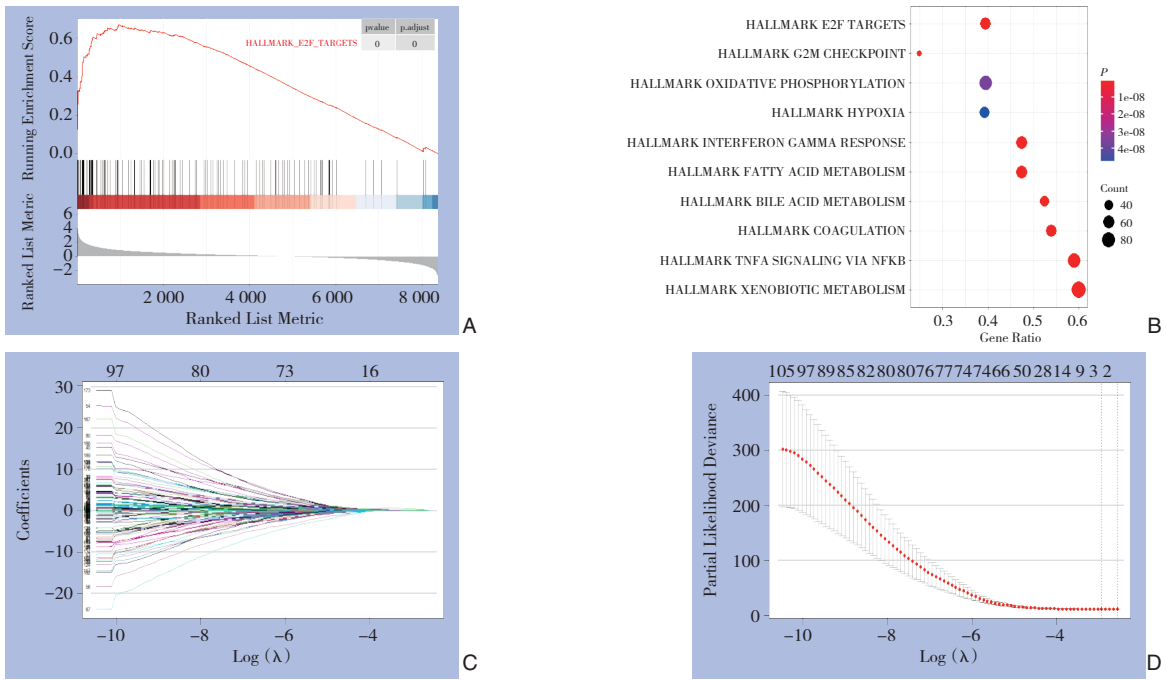


图2 富集得分最高的E2F靶点及其相关基因的鉴定 A: 区分有、无癌组的E2F靶点基因集富集图; B: 按归一化后的富集得分排序的10个标志基因集的P值及基因集包含基因数量; C: 所选特征的系数由λ参数表示; D: 偏似然偏差和对数(λ)用Lasso-Cox回归模型绘制

Figure 2 E2F targets with the highest enrichment score and identification of related genes A: Enrichment plots of E2F target gene sets differentiated between in cancer and non-cancer group; B: The P-value and the number of genes of top 10 marker gene sets with the highest normalized enrichment score; C: The coefficients of the selected features represented by the λ parameter; D: Partial likelihood deviation and logarithm (λ) constructed with the Lasso-Cox regression model

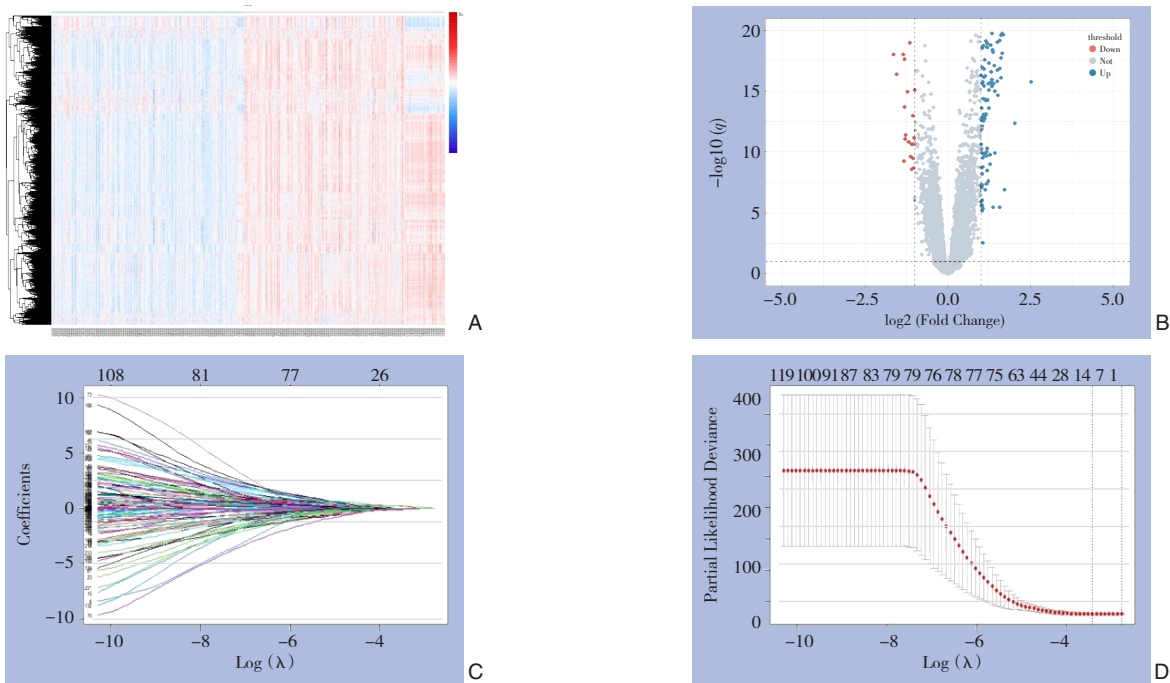


图3 免疫亚型及免疫相关差异表达基因的鉴定 A: 有癌组样本按富集得分的聚类分组; B: 两免疫亚组间差异表达mRNA的火山图; C: 所选特征的系数由λ参数表示; D: 偏似然偏差和对数(λ)用Lasso-Cox回归模型绘制

Figure 3 Immune subtypes and identification of immune-related differentially expressed genes A: The samples in the cancer group distinguished according to the enrichment score; B: Volcano map of differentially expressed mRNAs between two immune subgroups; C: The coefficients of the selected features represented by the λ parameter; D: Partial likelihood deviation and logarithm (λ) constructed with the Lasso-Cox regression model

2.3 构建预后模型

将E2F靶点基因集筛选得到的两个基因和免疫亚型差异筛选得到的6个基因共同进行Lasso回归,完成变量筛选和预后模型建立,得到7-mRNA预后模型: 风险评分 = $-0.55 \times \text{CYR61 表达} - 0.18 \times \text{FBLN5 表达} - 0.17 \times \text{LPA 表达} - 0.06 \times \text{SAA1 表达} + 0.31 \times \text{SDC3 表达} + 0.38 \times \text{SERPINE1 表达} + 1.08 \times$

SSRP1 表达 (图4A)。7个基因通过等比例风险检验 ($P > 0.1$),对生存风险的作用在不同时间点变化的比例基本一致。还构建了多因素Cox回归模型,做出列线图(图4B)。对该7-mRNA预后模型对截止时间内生存死亡结局的预测能力进行ROC分析,AUC值为0.846(图4C-D)。

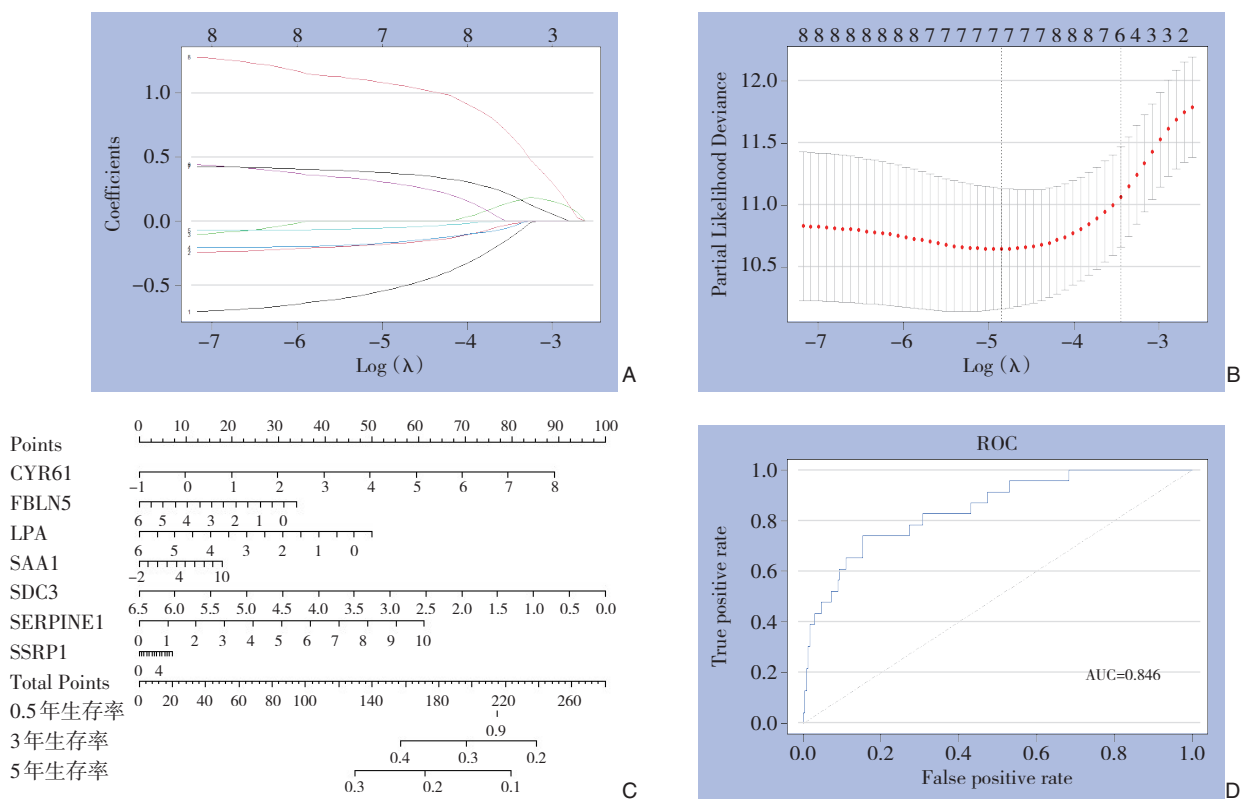


图4 构建预后模型 A: 所选特征的系数由 λ 参数表示; B: 偏似然偏差和对数(λ)用Lasso-Cox回归模型绘制; C: 多因素Cox回归模型的构建及其列线图; D: ROC检测7-mRNA预后模型的预后能力

Figure 4 Construction of prognosis model A: The coefficients of the selected features represented by the λ parameter; B: Partial likelihood deviation and logarithm (λ) constructed with the Lasso-Cox regression model; C: Construction of multivariate Cox regression model and its nomogram; D: Prognostic ability of the 7-mRNA signature model detected by ROC

2.4 7-mRNA模型预后能力验证分析

将风险评分高于中位数3.394 911的标记为高风险,低于中位数的为低风险,高风险样本生存时间普遍更长,截止期内死亡发生更少(图5A-B)。对比高、低风险组绘制Kaplan-Meier生存曲线,曲线显示高风险评分患者预后不良($P < 0.001$),高低风险评分对预后的区分度明显,与临床指标中的肿瘤大小和UICC分期具有相似的良好区分度,而比淋巴转移、远处转移和BMI值区分度要更好(图5C-H)。对风险评分、肿瘤大小和UICC分

期3个指标进行Cox回归分析,风险评分无论在单因素Cox回归分析中或是在多因素Cox回归分析中都被视为独立的预后指标($P < 0.001$)(表2)。在考虑不同UICC分期亚型时,本研究绘制了I、II分期和III、IV分期两亚型的高、低风险Kaplan-Meier生存曲线,在I、II期患者中评分对预后的区分度不理想($P > 0.05$),而在III、IV期患者中有明显的区分能力($P < 0.01$)(图5I-J),提示本研究建立的7-mRNA预后模型更适用于较危重患者的生存结局预测。

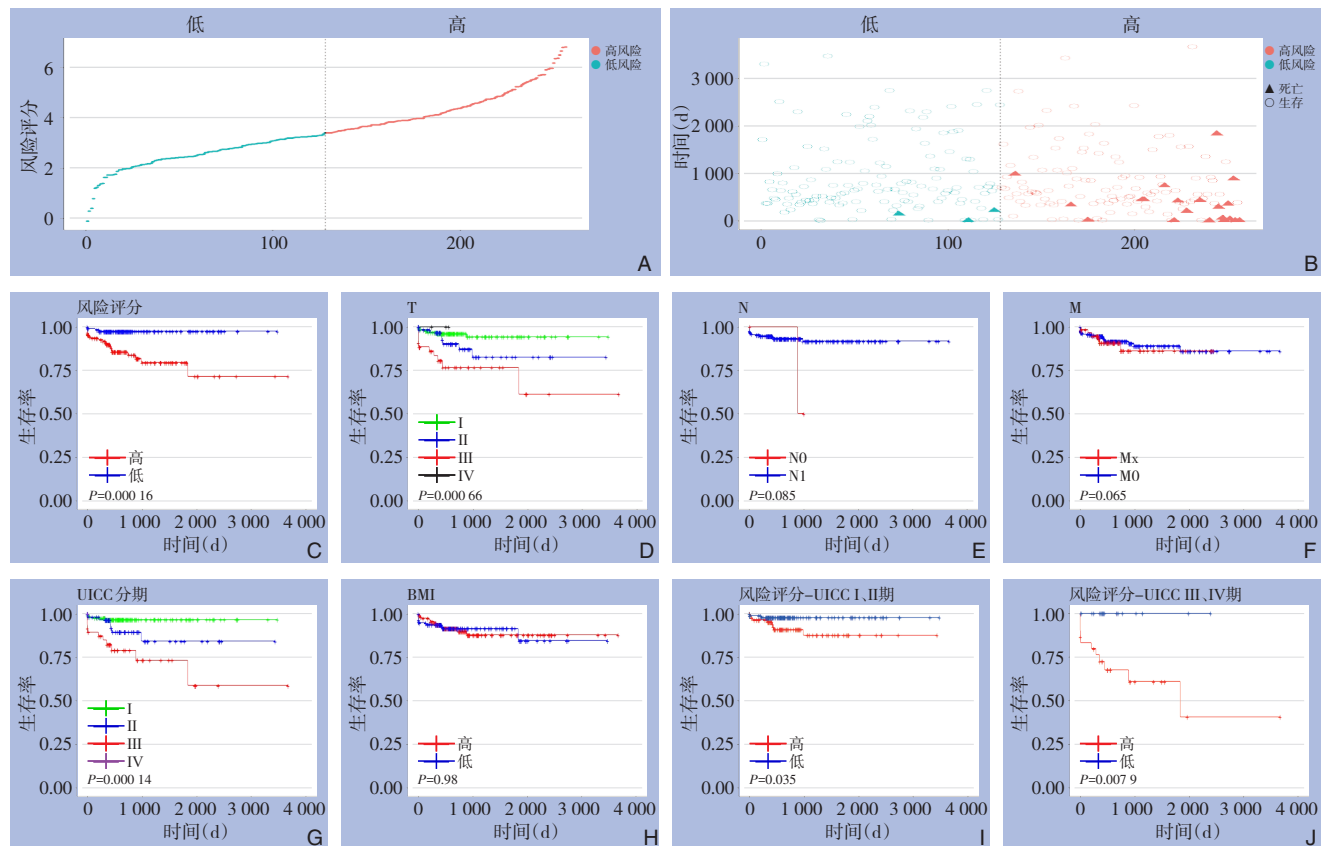


图5 7-mRNA 预后模型预后能力分析验证 A-B: 将风险评分高于中位数的标记为高风险, 低于中位数的为低风险; C-H: 按高低风险评分、肿瘤大小 (T)、淋巴转移 (N)、远处转移 (M)、UICC 分期和 BMI 值分别绘制的 Kaplan-Meier 生存曲线; I-J: 两 UICC 亚型 (I、II 期和 III、IV 期) 分别绘制的高低风险 Kaplan-Meier 生存曲线

Figure 5 Analysis and verification of the prognostic ability of the 7-mRNA signature model A-B: Marking the risk score higher than the median as high risk, and the risk score lower than the median as low risk; C-H: Constructing Kaplan-Meier survival curves for the high and low-risk score, tumor size (T), lymphatic metastasis (N), distant metastasis (M), UICC stage and BMI value; I-J: Constructing high and low-risk Kaplan-Meier survival curves between the two UICC subtypes (stages I, II, III and IV)

表2 与临床指标进行 Cox 回归分析

Table 2 Cox regression analysis with clinical indicators

项目	单因素 Cox 回归分析		多因素 Cox 回归分析	
	P	HR(95% CI)	P	HR(95% CI)
风险评分	4.9e-10	12(5.6~27)	6.5e-07	10(4~25)
肿瘤大小	0.000 38	2.3(1.5~3.7)	0.48	1.4(0.56~3.4)
UICC 分期	8.6e-05	2.9(1.7~4.9)	0.45	1.3(0.64~2.8)

2.5 分析模型的7个基因及寻找药物靶点

在 424 例 LIHC 样本中风险模型的 7 个基因 RNA 表达量在有无癌样本中差异均较明显, 其中有癌样本的 CYR61, FBLN5, LPA, SAA1, SDC3, SERPINE1 6 个基因表达量均下调, 仅 SSRP1 基因表达量上调 (图 6A)。7 个基因在 Cbioportal 数据库样本中都存在一定程度的不同形式突变 (表 3)。分析风险模型的 7 个基因的蛋白互作信息, 找到度数

较高的核心基因 SERPINE1 和 LPA (图 6B)。LPA 具有丝氨酸蛋白酶活性, 能够自体蛋白溶解, 属于纤溶酶原亚科; SERPINE1 是纤溶酶原激活剂抑制剂 1、丝氨酸蛋白酶抑制剂, 与 PLAT 的快速相互作用可以作为调节纤维蛋白溶解的主要控制点。因此预测抑制纤溶酶原激活可能是 HCC 的新的药物靶点。

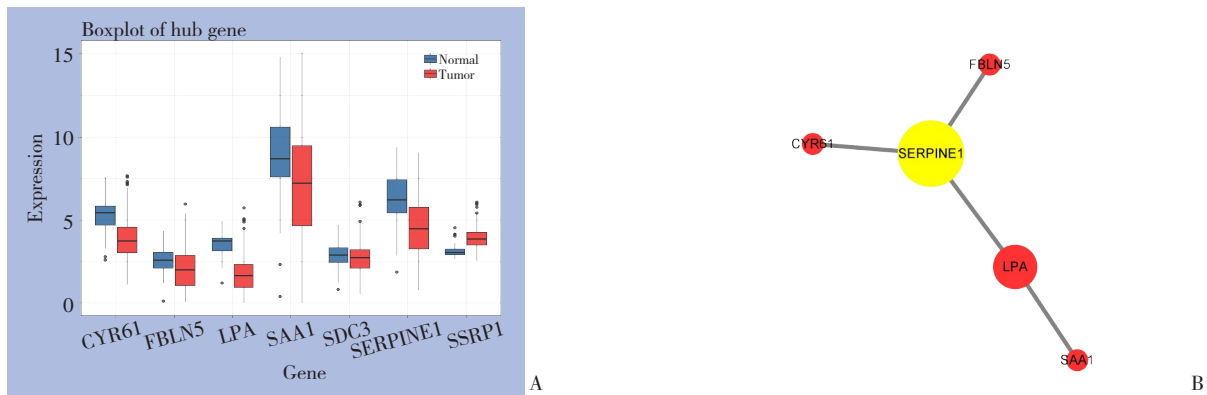


图 6 模型的 7 个基因分析及药物靶点寻找 A: 有、无癌样本 7 个基因表达量对比箱线图; B: 蛋白互作并找出核心基因

Figure 6 Analysis of seven genes and identification of drug targets A: Box-plot of comparison of 7 genes expression between cancer and non-cancer samples; B: Protein interaction and identification of the core genes

表 3 模型的 7 个基因在 372 个 TCGA-HCC 样本中的基因变异情况 (%)

Table 3 Genetic variation of 7 genes in 372 TCGA-HCC samples (%)

基因	扩增	深度缺失	突变
CYR61	0.27	0.00	0.00
FBLN5	0.27	0.54	0.00
LPA	1.63	0.27	1.08
SAA1	0.54	0.00	0.54
SDC3	0.00	0.27	0.27
SERPINE1	0.54	0.00	0.27
SSRP1	0.27	0.27	0.27

3 讨论

在我国, HCC 是最常见的肝癌类型, 最近的研究表明肝硬化患者中约 85% 有原发性 HCC 发生, 其 5 年生存率仅为 18%, 仅次于胰腺癌。其重要危险因素有乙型肝炎病毒 (HBV) 感染、过度饮酒等^[3]。因此急切面临对 HCC 分子机制的切实解析与探究。

本研究整理了 HCC 表达谱数据, 通过富集分析识别出了疾病相关的 E2F 靶点基因组和免疫相关差异基因, 进一步分析得到与 HCC 患者预后生存相关的风险基因并构建 HCC 预后风险模型, 并证实其具有较高的预测准确性。此外, 本研究深挖了构成模型的基因, 找到核心基因 SERPINE1 和 LPA, 预测了可能的治疗新靶点: 纤溶酶原激活。

CDK-RB-E2F 轴形成驱动细胞周期进展的核心转录机制, 决定了基因组复制的时间和保真度, 其最终效应子是 E2F 基因家族^[14]。E2F 靶基因包括

细胞 DNA 合成及细胞周期进程的限速调节剂、原癌基因及肿瘤抑制基因等, 其对许多细胞过程如细胞周期调节, 血管生成, DNA 损伤反应和凋亡都发挥重要作用^[15], 考虑到 HCC 是典型的多血供肿瘤, 其血管生成的作用极有可能影响肿瘤自身生长与肿瘤微环境的改变, 肿瘤微环境与肿瘤的发生发展、侵袭转移有密切关系。同样, 作为肿瘤微环境的重要组成部分, 微环境中的免疫相关细胞也对 HCC 的发展和进展具有重大影响, 如肝动脉化疗栓塞术后调节性 T 细胞 (T_H1) 水平高的患者较水平低的患者的无进展生存期明显缩短^[16], 而主要由原发性 HCC 环境中预先存在的活化 CD8⁺ 细胞毒性 T 细胞触发的 PD-L1 过表达则是原发性 HCC 的良好预后因子^[17]。

目前, 虽然已有较多的免疫相关生物标志物被发掘与 HCC 有关, 但由于肿瘤微环境内基因调控与免疫浸润的复杂性, 基于 mRNA 表达谱的生物标志物很少用于估计其在 HCC 中的进展。在本研究中, 基因集富集分析用于识别整体和癌症发生相关性最大的基因集, 而基因集单样本富集分析评分用于量化癌症样本中免疫特征的活性, 并以此分数聚类以得到免疫状态不同的两个亚组, 再在其中找到差异表达的基因。之后, 基于上述 7 个基因构建了 E2F 靶点和免疫相关风险评估模型。值得注意的是, 它能够区分高风险人群与低风险人群, 并且预后估计具有较高的敏感度和特异性。在临床实用性方面, 预后模型显示与患者的肿瘤大小和肿瘤 UICC 分期有显著相关性, 且在高分期患者中的预后估计能力更为准确。因此, 本研究构建的 E2F 靶点和免疫相关特征极有可能参与 HCC

的发生和进展,使其具有作为高效的临床生物标志物的潜力。尽管多基因风险评估预测能力在复杂的肿瘤微环境中有一定局限性,但具有较高预测准确性和可操作性的评分系统仍然可在筛查指南的修改、高危人群的预防中转化为临床效用^[18]。

最终纳入模型的7个基因中,CYR61(富含半胱氨酸的血管生成诱导剂61)也即CCN1,属于CCN基质细胞蛋白家族,在增殖、分化、细胞凋亡、血管生成和纤维化的调节中起着关键作用^[19],许多研究表明CYR61是肿瘤血管形成的关键因素,可能参与肿瘤细胞的迁移和侵袭,在多种癌症如肺癌、乳腺癌、女性生殖系统肿瘤的发生中起重要作用^[20-22],但同时与本文类似的可能在非小细胞肺癌中作为肿瘤抑制因子^[23];FBLN5作为母细胞糖蛋白的成员在富含弹性蛋白的组织中表达,对血管生成有抑制作用,可能通过控制细胞增殖、运动和血管生成来抑制肿瘤形成^[24-25];SAA1编码血清淀粉样蛋白A,并且与C反应蛋白一起参与急性期反应,已被发现在脂质代谢中起重要作用,并有助于细菌清除,炎症调节和肿瘤发病机制^[26-27];SDC3的表达在神经元、炎症性疾病,血管生成等疾病相关过程中具有重要作用^[28];SSRP1沉默可影响癌症凋亡和细胞增殖^[29],许多癌症相关病例中,SSRP1表达升高被证明与转移性肿瘤有关,使SSRP1成为肿瘤抑制的潜在预后标志物和抗癌靶标,与我们的结果相一致^[30]。以及在蛋白互作分析中处于核心地位的SERPINE和LPA。已发表的文献表明,在胃癌中SERPINE1表达上调,增加胃癌发生风险进而影响其预后^[31];在头颈部癌中SERPINE1的过表达增强肿瘤细胞迁移和侵袭,在转移发展中起关键作用^[32];LPA通过促进前列腺癌中的钙网蛋白表达介导肿瘤淋巴血管生成^[33],SERPINE1也曾被分别纳入自噬相关和P53途径相关的HCC预后特征与治疗靶点中^[34]。因此猜测SERPINE和LPA可能通过调节血管生成和促进细胞移动影响HCC的进展。

然而,本研究仍存在一定的问題亟待解决:多基因预后模型的预测能力仍需大量多中心的循证医学证据证实;样本中截止时间内死亡样本较小可能对模型建立的准确性产生一定影响;被纳入的多基因模型的基因功能和参与的机制尚不明确,与HCC的发生、发展的关系仍需要大量的临床研究进一步印证;同时对纤溶酶原激活是如何

影响HCC也需要进行更深入的研究。

利益冲突:所有作者均声明不存在利益冲突。

参考文献

- [1] Zhou MG, Wang HD, Zeng XY, et al. Mortality, morbidity, and risk factors in China and its provinces, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017[J]. *Lancet*, 2019, 394(10204): 1145-1158. doi: 10.1016/S0140-6736(19)30427-1.
- [2] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. *CA Cancer J Clin*, 2018, 68(6):394-424. doi: 10.3322/caac.21492.
- [3] Lewis DR, Chen HS, Cockburn MG, et al. Early estimates of SEER cancer incidence, 2014[J]. *Cancer*, 2017, 123(13):2524-2534. doi: 10.1002/cncr.30630.
- [4] Yang YM, Li ZY, Zhong QY, et al. Identification and validation of a novel prognostic signature based on transcription factors in breast cancer by bioinformatics analysis[J]. *Gland Surg*, 2022, 11(5):892-912. doi: 10.21037/gs-22-267.
- [5] Fei HJ, Chen XM. Development of a novel five-gene immune-related risk model for the prognosis evaluation of prostate adenocarcinoma patients[J]. *Am J Cancer Res*, 2022, 12(5):2337-2349.
- [6] Zhao RC, Ding D, Ding Y, et al. Predicting differences in treatment response and survival time of lung adenocarcinoma patients based on a prognostic risk model of glycolysis-related genes[J]. *Front Genet*, 2022, 13:828543. doi: 10.3389/fgene.2022.828543.
- [7] Wang WH, Wang LC, Xie XS, et al. A gene-based risk score model for predicting recurrence-free survival in patients with hepatocellular carcinoma[J]. *BMC Cancer*, 2021, 21(1): 6. doi: 10.1186/s12885-020-07692-6.
- [8] Xie D, Pei Q, Li JY, et al. Emerging role of E2F family in cancer stem cells[J]. *Front Oncol*, 2021, 11: 723137. doi: 10.3389/fonc.2021.723137.
- [9] Budhu A, Forgues M, Ye QH, et al. Prediction of venous metastases, recurrence, and prognosis in hepatocellular carcinoma based on a unique immune response signature of the liver microenvironment[J]. *Cancer Cell*, 2006, 10(2): 99-111. doi: 10.1016/j.ccr.2006.06.016.
- [10] Dosil MA, Navaridas R, Mirantes C, et al. Tumor suppressive function of E2F-1 on PTEN-induced serrated colorectal carcinogenesis[J]. *J Pathol*, 2019, 247(1): 72-85. doi: 10.1002/path.5168.

- [11] Fan T, Lu ZL, Liu Y, et al. A novel immune-related seventeen-gene signature for predicting early stage lung squamous cell carcinoma prognosis[J]. *Front Immunol*, 2021, 12: 665407. doi: [10.3389/fimmu.2021.665407](https://doi.org/10.3389/fimmu.2021.665407).
- [12] Liu XS, Gao Y, Liu C, et al. Comprehensive analysis of prognostic and immune infiltrates for E2F transcription factors in human pancreatic adenocarcinoma[J]. *Front Oncol*, 2020, 10: 606735. doi: [10.3389/fonc.2020.606735](https://doi.org/10.3389/fonc.2020.606735).
- [13] Barbie DA, Tamayo P, Boehm JS, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1[J]. *Nature*, 2009, 462(7269): 108–112. doi: [10.1038/nature08460](https://doi.org/10.1038/nature08460).
- [14] Kent LN, Leone G. The broken cycle: E2F dysfunction in cancer[J]. *Nat Rev Cancer*, 2019, 19(6): 326–338. doi: [10.1038/s41568-019-0143-7](https://doi.org/10.1038/s41568-019-0143-7).
- [15] Hinshaw DC, Shevde LA. The tumor microenvironment innately modulates cancer progression[J]. *Cancer Res*, 2019, 79(18): 4557–4566. doi: [10.1158/0008-5472.CAN-18-3962](https://doi.org/10.1158/0008-5472.CAN-18-3962).
- [16] Park H, Jung JH, Jung MK, et al. Effects of transarterial chemoembolization on regulatory T cell and its subpopulations in patients with hepatocellular carcinoma[J]. *Hepatol Int*, 2020, 14(2): 249–258. doi: [10.1007/s12072-020-10014-4](https://doi.org/10.1007/s12072-020-10014-4).
- [17] Xie QK, Zhao YJ, Pan T, et al. Programmed death ligand 1 as an indicator of pre-existing adaptive immune responses in human hepatocellular carcinoma[J]. *Oncoimmunology*, 2016, 5(7): e1181252. doi: [10.1080/2162402X.2016.1181252](https://doi.org/10.1080/2162402X.2016.1181252).
- [18] Klein RJ, Gümüş ZH. Are polygenic risk scores ready for the cancer clinic? -a perspective[J]. *Transl Lung Cancer Res*, 2022, 11(5): 910–919. doi: [10.21037/tlcr-21-698](https://doi.org/10.21037/tlcr-21-698).
- [19] Kim KH, Won JH, Cheng NY, et al. The matricellular protein CCN1 in tissue injury repair[J]. *J Cell Commun Signal*, 2018, 12(1): 273–279. doi: [10.1007/s12079-018-0450-x](https://doi.org/10.1007/s12079-018-0450-x).
- [20] Zhu Y, Almutashiri S, Han Y, et al. The roles of CCN1/CYR61 in pulmonary diseases[J]. *Int J Mol Sci*, 2020, 21(21): E7810. doi: [10.3390/ijms21217810](https://doi.org/10.3390/ijms21217810).
- [21] Ahmed KA, Hasib TA, Paul SK, et al. Potential role of CCN proteins in breast cancer: therapeutic advances and perspectives[J]. *Curr Oncol*, 2021, 28(6): 4972–4985. doi: [10.3390/currenol28060417](https://doi.org/10.3390/currenol28060417).
- [22] Yang R, Chen Y, Chen D. Biological functions and role of CCN1/Cyr61 in embryogenesis and tumorigenesis in the female reproductive system (Review)[J]. *Mol Med Rep*, 2018, 17(1): 3–10. doi: [10.3892/mmr.2017.7880](https://doi.org/10.3892/mmr.2017.7880).
- [23] Tong X, Xie D, O'Kelly J, et al. Cyr61, a member of CCN family, is a tumor suppressor in non-small cell lung cancer[J]. *J Biol Chem*, 2001, 276(50): 47709–47714. doi: [10.1074/jbc.M107878200](https://doi.org/10.1074/jbc.M107878200).
- [24] Albig AR, Schiemann WP. Fibulin-5 function during tumorigenesis[J]. *Future Oncol*, 2005, 1(1): 23–35. doi: [10.1517/14796694.1.1.23](https://doi.org/10.1517/14796694.1.1.23).
- [25] Tang JC, Xie AY, Cai XJ. Diverse functions of fibulin-5 in tumors[J]. *Mol Biol (Mosk)*, 2014, 48(6): 875–880. doi: [10.7868/s0026898414060172](https://doi.org/10.7868/s0026898414060172).
- [26] Sun L, Ye RD. Serum amyloid A1: structure, function and gene polymorphism[J]. *Gene*, 2016, 583(1): 48–57. doi: [10.1016/j.gene.2016.02.044](https://doi.org/10.1016/j.gene.2016.02.044).
- [27] Sack GH Jr. Serum amyloid A (SAA) proteins[J]. *Subcell Biochem*, 2020, 94: 421–436. doi: [10.1007/978-3-030-41769-7_17](https://doi.org/10.1007/978-3-030-41769-7_17).
- [28] Arokiasamy S, Balderstone MJM, De Rossi G, et al. Syndecan-3 in inflammation and angiogenesis[J]. *Front Immunol*, 2019, 10: 3031. doi: [10.3389/fimmu.2019.03031](https://doi.org/10.3389/fimmu.2019.03031).
- [29] Garcia H, Miecznikowski JC, Safina A, et al. Facilitates chromatin transcription complex is an "accelerator" of tumor transformation and potential marker and target of aggressive cancers[J]. *Cell Rep*, 2013, 4(1): 159–173. doi: [10.1016/j.celrep.2013.06.013](https://doi.org/10.1016/j.celrep.2013.06.013).
- [30] Jia SN, Guo BF, Wang LH, et al. The Current status of SSRP1 in cancer: tribulation and road ahead[J]. *J Healthc Eng*, 2022, 2022: 3528786. doi: [10.1155/2022/3528786](https://doi.org/10.1155/2022/3528786).
- [31] Resende C, Thiel A, Machado JC, et al. Gastric cancer: basic aspects[J]. *Helicobacter*, 2011, 16(Suppl 1): 38–44. doi: [10.1111/j.1523-5378.2011.00879.x](https://doi.org/10.1111/j.1523-5378.2011.00879.x).
- [32] Pavón MA, Arroyo-Solera I, Céspedes MV, et al. uPA/uPAR and SERPINE1 in head and neck cancer: role in tumor resistance, metastasis, prognosis and therapy[J]. *Oncotarget*, 2016, 7(35): 57351–57366. doi: [10.18632/oncotarget.10344](https://doi.org/10.18632/oncotarget.10344).
- [33] Lin YC, Chen CC, Chen WM, et al. LPA1/3 signaling mediates tumor lymphangiogenesis through promoting CRT expression in prostate cancer[J]. *Biochim Biophys Acta Mol Cell Biol Lipids*, 2018, 1863(10): 1305–1315. doi: [10.1016/j.bbalip.2018.07.005](https://doi.org/10.1016/j.bbalip.2018.07.005).
- [34] Chen W, Hu MJ, Zhong XL, et al. Screening of a novel autophagy-related prognostic signature and therapeutic targets in hepatocellular carcinoma[J]. *J Gastrointest Oncol*, 2021, 12(6): 2985–2998. doi: [10.21037/jgo-21-664](https://doi.org/10.21037/jgo-21-664).

(本文编辑 姜晖)

本文引用格式:何锬,赵杨,朱永乾,等.基于E2F靶点基因集和免疫亚型差异的肝细胞癌预后风险评估模型的建立[J].中国普通外科杂志,2023,32(1):64–73. doi:10.7659/j.issn.1005-6947.2023.01.005

Cite this article as: He S, Zhao Y, Zhu YQ, et al. Construction of prognostic assessment model for hepatocellular carcinoma based on E2F targets and immune subtype differences[J]. *Chin J Gen Surg*, 2023, 32(1):64–73. doi:10.7659/j.issn.1005-6947.2023.01.005