

DOI:10.16781/j.CN31-2187/R.20230337

· 论 著 ·

基于索拉非尼敏感性基因的肝细胞癌预后模型的构建及临床意义

纪琳琳¹, 胡定涛², 高 鹏², 丁 劲², 褚晓源^{1*}

1. 南京大学医学院附属金陵医院肿瘤科, 南京 210000

2. 海军军医大学(第二军医大学)转化医学中心临床肿瘤研究所, 上海 200433

[摘要] **目的** 基于生物信息学方法构建并验证肝细胞癌(HCC)对索拉非尼敏感性相关基因的预后模型, 探究该模型对HCC患者预后和索拉非尼敏感性的预测能力。**方法** 对基因表达汇编数据库GSE109211数据集、癌症基因组图谱(TCGA)数据库LIHC队列、国际肿瘤基因组协作组(ICGC)数据库LIRI队列进行差异基因分析, 通过交集筛选出HCC索拉非尼敏感性相关基因, 进行京都基因与基因组百科全书(KEGG)信号通路富集分析。利用单因素Cox回归与最小绝对收缩和选择算子(LASSO)回归构建预后模型, 依据风险评分中位值将患者分为高、低风险组。通过Kaplan-Meier法和多因素Cox回归分析进行生存分析。通过癌症药物敏感性基因组学(GDSC)数据库分析索拉非尼的IC₅₀并探索其与风险评分的关系。**结果** 筛选出365个HCC索拉非尼敏感性相关基因, KEGG富集分析显示存在与药物代谢相关的信号通路。通过单因素Cox回归分析获得221个与HCC预后相关的基因, 通过LASSO回归构建了一个包括含TCPI1伴侣蛋白亚单位3(CCT3)、红细胞生成素(erythropoietin, EPO)、甲酰转移酶环脱氨酶(FTCD)、葡萄糖-6-磷酸脱氢酶(G6PD)、驱动蛋白家族成员20A(KIF20A)、磷脂酰肌醇聚糖锚定生物合成U类基因(PIGU)、分泌型磷蛋白1(SPPI)7个关键基因的预后风险评分模型: 风险评分 = CCT3×0.032+EPO×0.055+FTCD×(-0.026)+G6PD×0.083+KIF20A×0.039+PIGU×0.144+SPPI×0.009。验证结果显示高风险组生存时间短于低风险组(P<0.001), 多因素Cox回归分析显示风险评分是独立预后因素。高风险组的索拉非尼IC₅₀比低风险组低, 提示高风险组对索拉非尼的治疗可能更敏感。**结论** 基于索拉非尼敏感性相关基因构建的HCC预后模型具有良好的预测价值。

[关键词] 肝细胞癌; 索拉非尼; 药物敏感性; 生物信息学; 预后模型

[引用本文] 纪琳琳, 胡定涛, 高鹏, 等. 基于索拉非尼敏感性基因的肝细胞癌预后模型的构建及临床意义[J]. 海军军医大学学报, 2023, 44(7): 830-838. DOI: 10.16781/j.CN31-2187/R.20230337.

Construction of a prognostic prediction model for hepatocellular carcinoma based on sorafenib sensitivity-related genes and its clinical significance

JI Linlin¹, HU Dingtao², GAO Peng², DING Jin², CHU Xiaoyuan^{1*}

1. Department of Oncology, Jinling Hospital, Medical School of Nanjing University, Nanjing 210000, Jiangsu, China

2. Clinical Cancer Institute, Center of Translational Medicine, Naval Medical University (Second Military Medical University), Shanghai 200433, China

[Abstract] **Objective** To construct and validate a prognostic model for sorafenib sensitivity-related genes in hepatocellular carcinoma (HCC) based on bioinformatics methods, and to explore the predictive ability of the model for prognosis and the efficacy for sorafenib treatment in HCC. **Methods** Differential gene analysis was performed on GSE109211 data set of Gene Expression Omnibus, LIHC cohort of The Cancer Genome Atlas (TCGA) and LIRI cohort of International Cancer Genome Consortium (ICGC). Sorafenib sensitivity-related genes in HCC were screened by intersection, and Kyoto Encyclopedia of Genes and Genomes (KEGG) signal pathway enrichment analysis was performed. A prognostic model was constructed using univariate Cox regression and least absolute shrinkage and selection operator (LASSO) regression, and the patients were divided into high-risk and low-risk groups based on the median risk score. Survival analysis was conducted using Kaplan-Meier method and multivariate Cox regression analysis. The half inhibitory

[收稿日期] 2023-06-15 [接受日期] 2023-07-07

[基金项目] 国家自然科学基金面上项目(82072725), 江苏省卫生健康委员会重点项目(ZD2021039). Supported by General Program of National Natural Science Foundation of China (82072725) and Key Program of Health Commission of Jiangsu (ZD2021039).

[作者简介] 纪琳琳, 硕士生. E-mail: jll199801@163.com

*通信作者(Corresponding author). Tel: 025-80864770, E-mail: chuxiaoyuan000@163.com

concentration (IC₅₀) of sorafenib was analyzed through the Genomics of Drug Sensitivity in Cancer (GDSC) and its relationship with risk score was explored. **Results** A total of 365 sorafenib sensitivity-related genes in HCC were screened, and the KEGG enrichment analysis revealed the presence of pathways associated with drug metabolism. Univariate Cox analysis identified 221 genes associated with prognosis, and a prognostic model containing 7 key genes (chaperonin containing TCP1 subunit 3 [CCT3], erythropoietin [EPO], formimidoyltransferase cyclodeaminase [FTCD], glucose-6-phosphate dehydrogenase [G6PD], kinesin family member 20A [KIF20A], phosphatidylinositol glycan anchor biosynthesis class U [PIGU], and secreted phosphoprotein 1 [SPP1]) was constructed by LASSO regression. Risk score = CCT3 × 0.032 + EPO × 0.055 + FTCD × (-0.026) + G6PD × 0.083 + KIF20A × 0.039 + PIGU × 0.144 + SPP1 × 0.009. The results showed that the high-risk group had a shorter survival time compared to the low-risk group ($P < 0.001$). Multivariate Cox analysis demonstrated that risk score was an independent prognostic factor. The IC₅₀ of sorafenib in the high-risk group was lower than that in the low-risk group, suggesting that the high-risk group may be more sensitive to sorafenib. **Conclusion** The constructed prognostic model for HCC based on sorafenib sensitivity-related genes has good prognostic value for HCC.

[**Key words**] hepatocellular carcinoma; sorafenib; drug susceptibility; bioinformatics; prognostic model

[**Citation**] JI L, HU D, GAO P, et al. Construction of a prognostic prediction model for hepatocellular carcinoma based on sorafenib sensitivity-related genes and its clinical significance [J]. Acad J Naval Med Univ, 2023, 44(7): 830-838. DOI: 10.16781/j.CN31-2187/R.20230337.

肝癌是威胁人类健康的主要恶性肿瘤之一。对 GLOBOCAN 2020^[1]数据的二次分析显示, 2020年中国肝癌新发病例约为 41 万例, 位居中国恶性肿瘤发病第 5 位, 而死亡病例约为 39.1 万例, 位居中国癌症相关死亡原因第 2 位^[2]。肝细胞癌 (hepatocellular carcinoma, HCC) 是肝癌的主要病理类型, 其治疗手段包括手术切除、肿瘤消融、肝移植等局部疗法, 靶向治疗、免疫治疗等全身疗法, 手术切除、消融等手段是潜在的治愈疗法, 但多限于早期 HCC 患者, 且几乎 70% 的患者会复发^[3]。大部分 HCC 患者发现时已处于中晚期, 不具备手术指征, 需通过全身疗法延缓病情进展^[4]。索拉非尼是第 1 个被批准用于晚期 HCC 一线全身治疗的靶向药物, 是晚期 HCC 患者的标准疗法, 能够延长晚期 HCC 患者的中位生存期, 然而许多 HCC 患者在接受数月的索拉非尼治疗后出现耐药, 导致临床获益受限^[5-6]。目前, 索拉非尼耐药的具体机制尚不明确, 涉及 MAPK14 信号、肿瘤启动细胞的富集和胰岛素样生长因子/成纤维细胞生长因子信号的重激活等^[7], 尚缺乏有效的疗效预测指标。本研究应用生物信息学技术筛选 HCC 索拉非尼耐药相关基因并构建预后模型, 为预测 HCC 的预后及索拉非尼敏感性提供依据。

1 资料和方法

1.1 数据收集 从基因表达汇编 (Gene Expression Omnibus, GEO) 数据库 GSE109211 数据集^[8]下载索拉非尼治疗队列的转录组测序数据和相关临

床信息 (包含 67 例索拉非尼治疗患者)。从癌症基因组图谱 (The Cancer Genome Atlas, TCGA) 数据库 (<https://portal.gdc.cancer.gov/>) LIHC (TCGA-LIHC) 队列下载 377 例 HCC 患者转录组测序数据和相对应的临床信息数据。从基因型组织表达 (Genotype-Tissue Expression Consortium, GTEx) 数据库下载正常人肝组织样本的转录组测序数据。从国际肿瘤基因组协作组 (International Cancer Genome Consortium, ICGC) 数据库 (<https://dcc.icgc.org/>) LIRI (ICGC-LIRI) 队列下载 243 例 HCC 患者 (240 例为原发 HCC, 3 例为转移 HCC) 肿瘤样本转录组测序数据和临床信息数据。利用 Sangerbox 3.0 (<http://sangerbox.com/home.html>) 对下载的转录组测序数据进行清洗, 去除重复样本, 并对原始数据进行 $\log_2(X+1)$ 标准化处理。差异基因分析样本纳入标准: (1) 诊断为 HCC; (2) 转录组表达数据无缺失。最终获得来自 TCGA-LIHC 队列的 371 例肿瘤样本和来自 GTEx 数据库的 276 例正常肝组织样本, 以及来自于 ICGC-LIRI 队列的 240 例原发 HCC 肿瘤样本和 202 例正常肝组织样本。预后模型构建样本纳入标准: (1) 原发疾病诊断为 HCC; (2) 生存资料完整。在剔除无相关信息患者后, TCGA-LIHC 队列中有 368 例 HCC 患者被纳入训练集, ICGC-LIRI 队列中的 243 例 HCC 患者被纳入验证集。

1.2 HCC 索拉非尼敏感性相关基因的获取与分析 使用 R4.3.0 软件 limma 包^[9]分析 GSE109211 队列中索拉非尼敏感组和耐受组之间的差异表达

基因, P 值经 Benjamini-Hochberg (BH) 法校正, 设置校正后 $P < 0.05$ 筛选出与索拉非尼敏感性相关的基因。使用 R4.3.0 软件 limma 包^[9] 识别 TCGA-LIHC 和 ICGC-LIRI 队列中肿瘤组织与正常肝组织之间的差异表达基因, P 值用 BH 法校正, 设置校正后 $P < 0.05$ 和 $|\log_2FC| > 1$ (FC 为差异倍数, fold change) 筛选出差异表达基因。将 3 个队列的差异表达基因利用 R4.3.0 软件 VennDiagram 包^[10] 绘制维恩图并获取交集基因。采用 R4.3.0 软件 clusterProfiler 包^[11] 对交集基因进行京都基因与基因组百科全书 (Kyoto Encyclopedia of Genes and Genomes, KEGG) 通路富集分析, 用 BH 法校正 P 值。

1.3 基于索拉非尼敏感性基因的 HCC 预后模型的建立与验证 通过 R4.3.0 软件 survival 包对上述交集基因在 TCGA-LIHC 队列中进行单因素 Cox 回归分析, 设置 $P < 0.05$ 筛选出与样本生存显著相关的基因。以 TCGA-LIHC 队列为训练集, 基于上述候选基因使用 R4.3.0 软件 glmnet 包进行最小绝对收缩和选择算子 (least absolute shrinkage and selection operator, LASSO) 回归分析^[12-13], 以减少过拟合的可能性。在惩罚参数 λ 为 0.061 时, 基于 7 个与索拉非尼敏感性和预后相关的基因构建风险模型。风险评分为 7 个基因表达量与回归系数乘积之和, 将样本按照风险评分的中位值分为高、低风险组。以 ICGC-LIRI 队列为验证集, 采用同样的方法计算验证集中每个样本的风险得分并进行分组。利用 R4.3.0 软件 prcomp 函数对训练集和验证集进行主成分分析 (principal component analysis, PCA) 并通过 R4.3.0 软件 ggplot2 包^[14] 进行可视化以探索不同组别的分布情况。在训练集和验证集中通过 R4.3.0 软件 survminer 包和 timeROC 包分别绘制 Kaplan-Meier 生存曲线和 ROC 曲线, 评估该模型的预测能力。绘制热图分析风险评分和临床病理特征 (年龄、性别、临床分期) 与生存结局的相关性, 绘制小提琴图评价风险评分与生存结局、组织病理学分级、T 分期、临床分期等的相关性。在 TCGA-LIHC、ICGC-LIRI 队列中, 纳入临床病理特征和风险评分进行单因素和多因素 Cox 回归分析, 筛选预后的独立危险因素。

1.4 索拉非尼敏感性分析 用于拟合的数据来自癌症药物敏感性基因组学 (Genomics of Drug Sensitivity in Cancer, GDSC) 数据库 (<https://www.cancerrxgene.org/>) GDSC1 和 GDSC2 数据集。通

过 R4.3.0 软件 OncoPredict 包^[15] 根据基因表达水平预测 TCGA-LIHC 队列中每例患者的药物敏感性, 获得索拉非尼的 IC_{50} , 比较高风险组与低风险组 IC_{50} 的差异。

1.5 统计学处理 所有数据的可视化和统计分析均通过 R4.3.0 软件完成。两组之间计量资料的比较采用独立样本 t 检验, 多组之间计量资料的比较采用单因素方差分析; 计数资料的比较采用 χ^2 检验。两个变量之间的相关性采用 Spearman 相关分析。单因素生存分析采用 Kaplan-Meier 法和 log-rank 检验, 多变量生存分析采用 Cox 回归分析。检验水准 (α) 为 0.05。

2 结果

2.1 基于索拉非尼敏感性相关基因的 HCC 预后模型的构建 从 GEO 数据库 GSE109211 数据集筛选出与索拉非尼敏感性相关的差异基因, 在 TCGA-LIHC 和 ICGC-LIRI 队列中鉴定出正常肝组织和肿瘤组织的差异表达基因, 将两者进行交集, 获得 365 个基因 (图 1A)。KEGG 富集分析显示这些基因与药物代谢和癌症发生等相关 (图 1B), 说明筛选出的基因与 HCC 的发生和索拉非尼的敏感性相关。通过单因素 Cox 回归分析上述 365 个基因在 TCGA-LIHC 队列中与总生存期之间的相关性, 获得 221 个与 HCC 预后相关候选基因。以 TCGA-LIHC 队列为训练集, 基于这些候选基因进行 LASSO 回归分析构建模型, 根据最小准则确定的惩罚参数 λ 得到含 TCP1 伴侣蛋白亚单位 3 (chaperonin containing TCP1 subunit 3, CCT3)、红细胞生成素 (erythropoietin, EPO)、甲酰转移酶环脱氨酶 (formimidoyltransferase cyclodeaminase, FTCD)、葡萄糖-6-磷酸脱氢酶 (glucose-6-phosphate dehydrogenase, G6PD)、驱动蛋白家族成员 20A (kinesin family member 20A, KIF20A)、磷脂酰肌醇聚糖锚定生物合成 U 类基因 (phosphatidylinositol glycan anchor biosynthesis class U, PIGU)、分泌型磷蛋白 1 (secreted phosphoprotein 1, SPPI) 7 个关键基因 (图 1C、1D), 其中 FTCD 为保护因素, 而 CCT3、EPO、G6PD、KIF20A、PIGU、SPPI 为危险因素 (图 1E)。

2.2 基于索拉非尼敏感性基因的 HCC 预后模型的评价和验证 计算训练集、验证集的风险评分, 依据中位值将样本分为高、低风险组 (图 2A、

2B), PCA分析显示训练集的高、低风险组呈不同方向分布(图2C)。在训练集中,与低风险组相比,高风险组患者总生存期更短($P < 0.002$,图2D);同时,采用ROC曲线评价风险评分对总生存期的预测性能,预测1、3、4年总生存期的AUC分别为0.797、0.721、0.730(图2E)。与训

练集结果相似,验证集的高、低风险组在PCA上也呈现为不同的方向(图2F),高风险组的生存期短于低风险组($P < 0.001$,图2G),风险评分预测1、3、4年总生存期的AUC分别为0.751、0.741、0.780(图2H)。

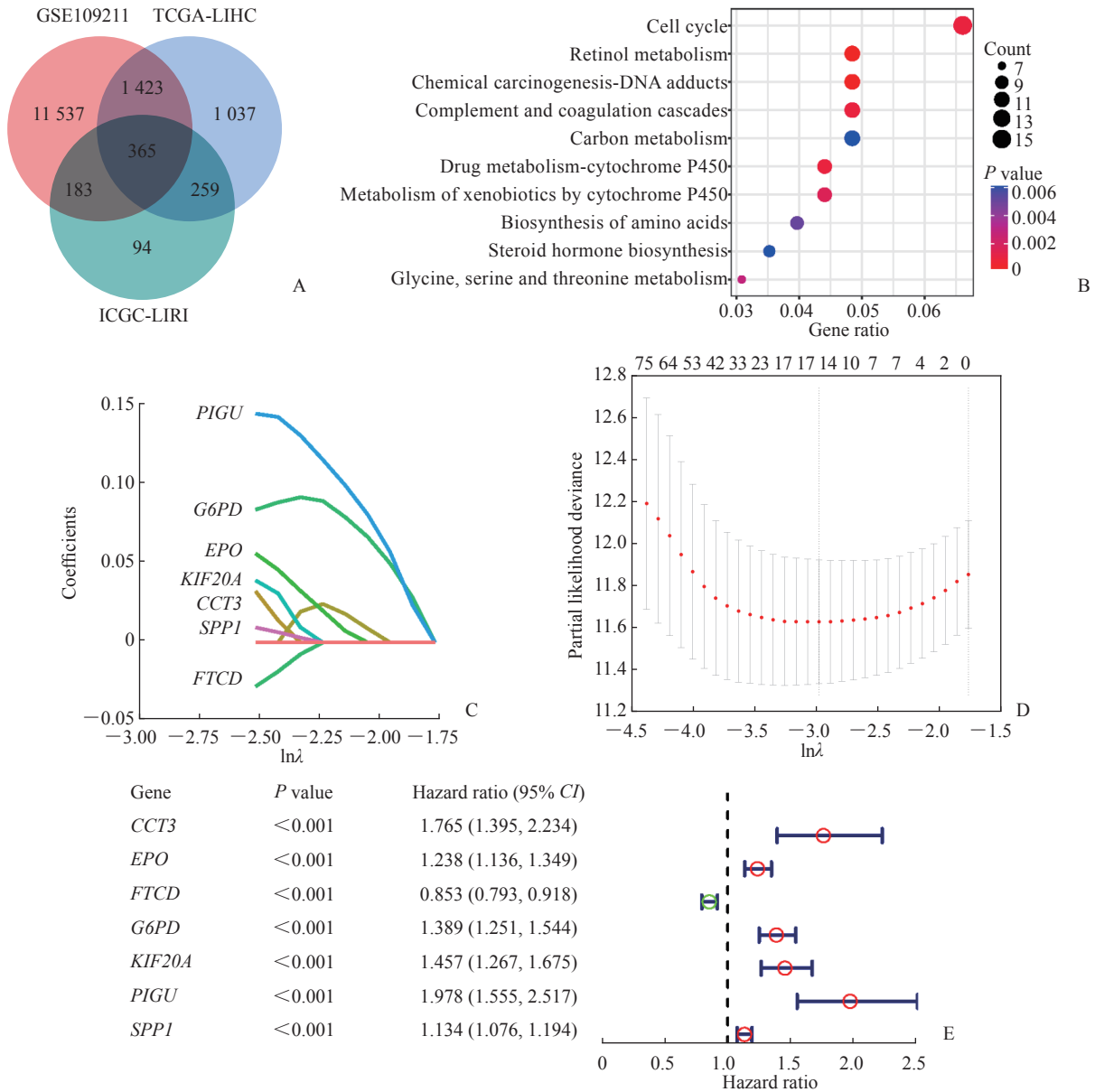


图1 HCC 索拉非尼敏感性相关基因的筛选及预后模型构建

Fig 1 Screening of sorafenib sensitivity-related genes and construction of prognostic model in HCC

A: Venn diagram to identify differentially expressed genes correlated with sorafenib sensitivity between tumor and adjacent normal tissue. B: Bubble diagram of the top 10 terms in Kyoto Encyclopedia of Genes and Genomes enrichment analysis of the differentially expressed genes. C: LASSO coefficient profiles of the expression of the candidate genes. D: Selection of the penalty parameter (λ) in the LASSO model via 10-fold cross-validation. The optimal λ value is 0.061. E: Univariate Cox regression analysis of 7 key genes in the TCGA-LIHC cohort. HCC: Hepatocellular carcinoma; LASSO: Least absolute shrinkage and selection operator; TCGA: The Cancer Genome Atlas; ICGC: International Cancer Genome Consortium; PIGU: Phosphatidylinositol glycan anchor biosynthesis class U; G6PD: Glucose-6-phosphate dehydrogenase; EPO: Erythropoietin; KIF20A: Kinesin family member 20A; CCT3: Chaperonin containing TCP1 subunit 3; SPP1: Secreted phosphoprotein 1; FTCD: Formimidoyltransferase cyclodeaminase; CI: Confidence interval.

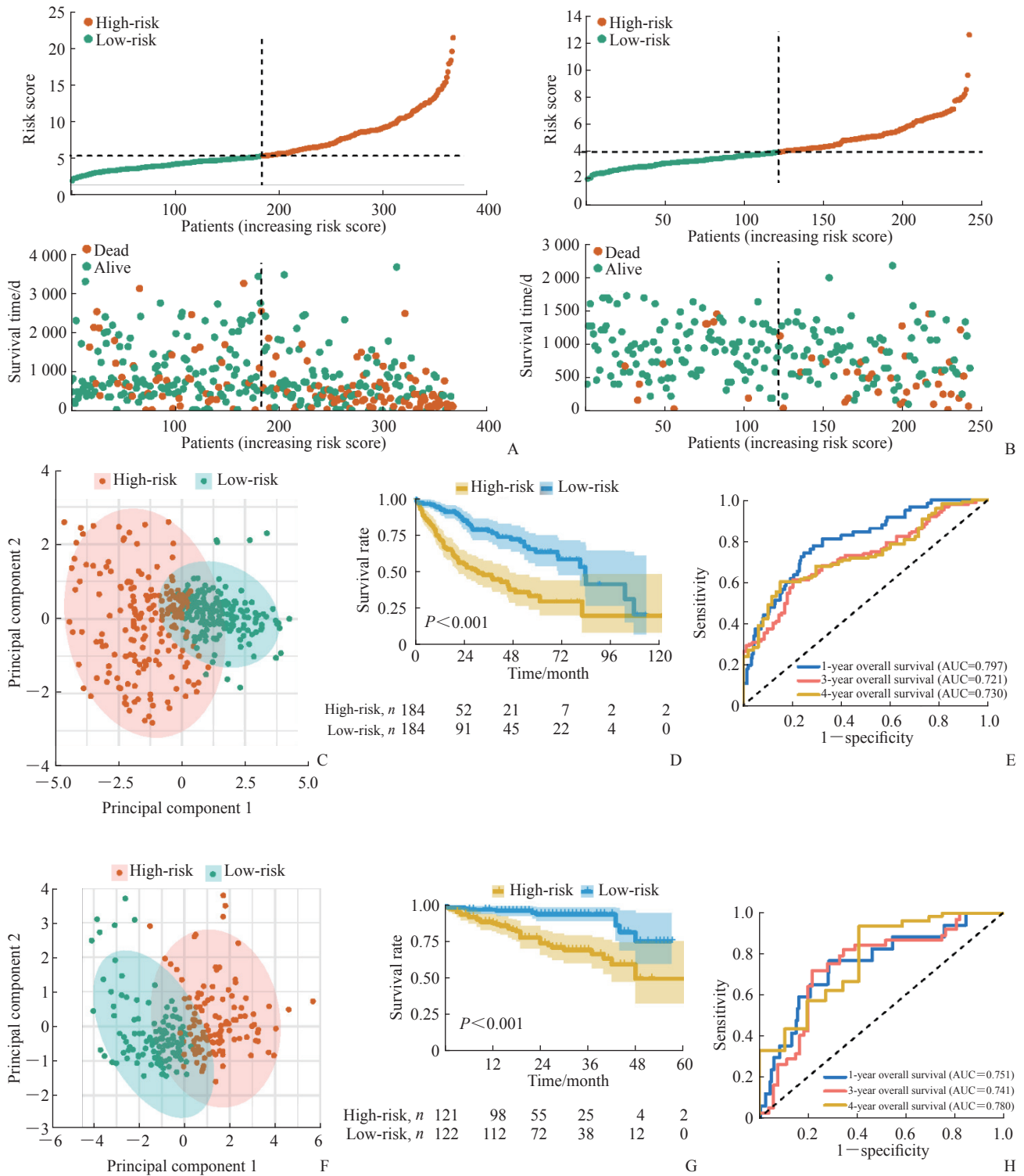


图2 基于索拉非尼敏感性相关基因的HCC预后模型的评价

Fig 2 Evaluation of prognostic model based on sorafenib sensitivity-related genes in HCC

A: The risk score distribution and patient survival status in the TCGA-LIHC cohort were depicted in ranked dot and scatter plots; B: The risk score distribution and patient survival status in the ICGC-LIRI cohort were depicted in ranked dot and scatter plots; C: PCA plot of the TCGA-LIHC cohort; D: Kaplan-Meier curves for the overall survival of patients between the high-risk group and low-risk group in the TCGA-LIHC cohort; E: AUC of time-dependent ROC curves verified the prognostic performance of the risk score in the TCGA-LIHC cohort; F: PCA plot of the ICGC-LIRI cohort; G: Kaplan-Meier curves for the overall survival of patients between the high-risk group and low-risk group in the ICGC-LIRI cohort; H: AUC of time-dependent ROC curves verified the prognostic performance of the risk score in the ICGC-LIRI cohort. HCC: Hepatocellular carcinoma; TCGA: The Cancer Genome Atlas; ICGC: International Cancer Genome Consortium; PCA: Principal component analysis; AUC: Area under curve; ROC: Receiver operating characteristic.

生存分析结果显示,死亡的HCC患者多属于高风险组,并且 *CCT3*、*EPO*、*G6PD*、*KIF20A*、*PIGU*、*SPP1* 基因的表达水平较高,而 *FTCD* 基因的表达水平较低(图 3A)。风险评分与年龄和性别无关,而发生死亡事件、组织病理学分级较高及临床分期为晚期的患者风险评分较高(图 3B~3E),并且死亡事件的发生和临床分期与风险评分的关系在验证集中获得一致的结论(图 3F、3G)。对可用变量进行单因素和多因素 Cox

回归分析,单因素 Cox 回归分析显示,在 TCGA-LIHC 和 ICGC-LIRI 队列中低风险评分为总生存期的保护因素 ($HR=0.34$, $95\% CI 0.24\sim0.50$, $P<0.001$; $HR=0.24$, $95\% CI 0.12\sim0.48$, $P<0.001$); 多因素 Cox 回归分析显示,在 TCGA-LIHC 队列和 ICGC-LIRI 中低风险评分为总生存期的独立保护因素 ($HR=0.40$, $95\% CI 0.27\sim0.59$, $P<0.001$; $HR=0.29$, $95\% CI 0.14\sim0.60$, $P<0.001$)。因此,风险评分对 HCC 患者预后具有独立的预测价值。

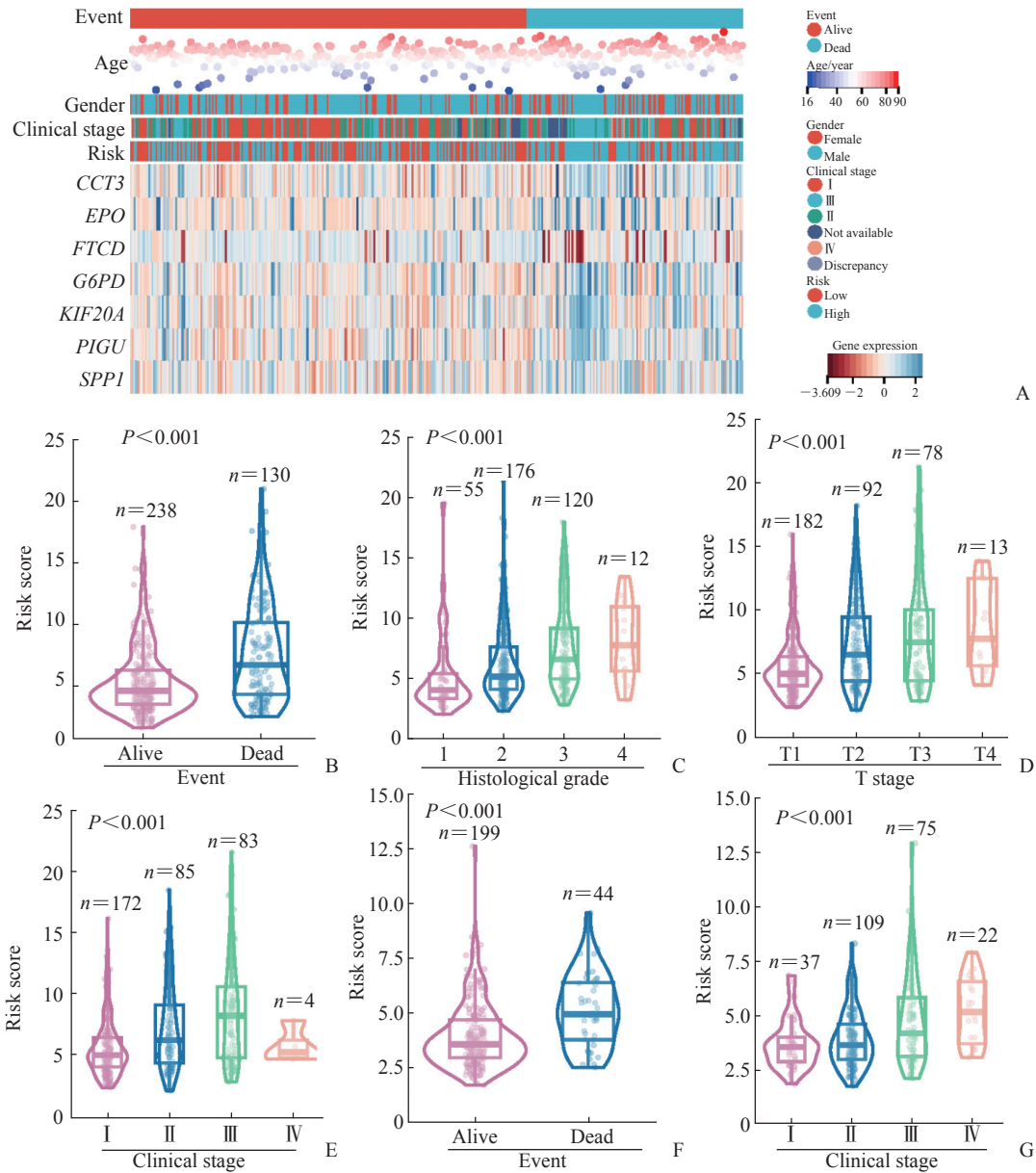


图 3 基于索拉非尼敏感性基因的 HCC 预后模型风险评分与临床特征的关系

Fig 3 Relationship between clinicopathological features and risk scores of HCC prognostic model based on sorafenib sensitivity-related genes in HCC

A: Heatmap of prognostic and clinicopathological features of 7 sorafenib sensitivity-related genes in 2 risk subgroups; B-E: Associations between the risk scores and clinicopathological features, including event (B), histological grade (C), clinical stage (D) and T stage (E) in the TCGA-LIHC cohort; F, G: Associations between the risk scores and clinicopathological features, including event (F) and clinical stage (G) in the ICGC-LIRI cohort. HCC: Hepatocellular carcinoma; TCGA: The Cancer Genome Atlas; ICGC: International Cancer Genome Consortium.

2.3 基于索拉非尼敏感性的HCC预后 结果显示不论在GDSC1还是GDSC2数据集中,高风险组的索拉非尼IC₅₀均低于低风险组,风险评分与索拉非尼的IC₅₀呈负相关(图4A、4B),说明高风险评分的患者对索拉非尼更敏感。将患者依据7个关键基因表达中位值分为高、低表达组,其中6个基因的高表达组与低表达组之间索拉非尼的IC₅₀存

在明显差异(图4C~4I)。相关性分析结果显示,7个关键基因的表达水平均与索拉非尼的IC₅₀相关,其中CCT3、EPO、G6PD、KIF20A、PIGU、SPP1表达水平与索拉非尼的IC₅₀呈负相关,而FTCD表达水平与索拉非尼的IC₅₀呈正相关(图4J)。以上数据表明,该模型可以预测HCC患者对索拉非尼的敏感性。

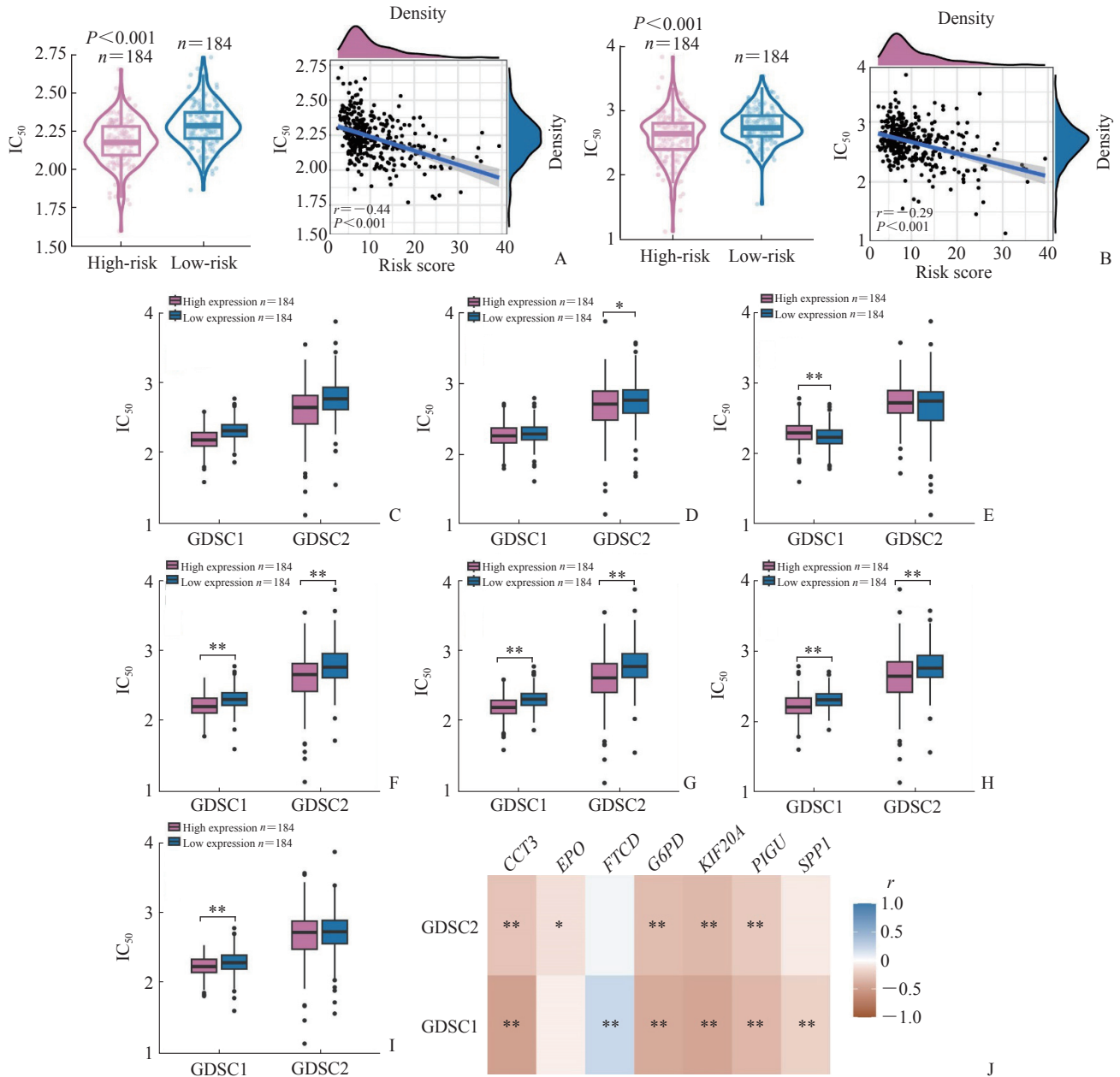


图4 基于索拉非尼敏感性相关基因的HCC预后模型与HCC中索拉非尼IC₅₀的关系

Fig 4 Relationship between sorafenib IC₅₀ and prognostic model based on sorafenib sensitivity-related genes in HCC

A: Analysis of correlations between risk scores and sorafenib IC₅₀ of GDSC1 database in the TCGA-LIHC cohort; B: Analysis of correlations between risk scores and sorafenib IC₅₀ of GDSC2 database in the TCGA-LIHC cohort; C-I: The influence on the sorafenib IC₅₀ of CCT3 (C), EPO (D), FTCD (E), G6PD (F), KIF20A (G), PIGU (H) and SPP1 (I) expression; J: The correlations between the sorafenib IC₅₀ and the expression of CCT3, EPO, FTCD, G6PD, KIF20A, PIGU and SPP1. *P<0.05, **P<0.01. HCC: Hepatocellular carcinoma; IC₅₀: Half inhibitory concentration; GDSC: Genomics of Drug Sensitivity in Cancer; CCT3: Chaperonin containing TCP1 subunit 3; EPO: Erythropoietin; FTCD: Formimidoyltransferase cyclodeaminase; G6PD: Glucose-6-phosphate dehydrogenase; KIF20A: Kinesin family member 20A; PIGU: Phosphatidylinositol glycan anchor biosynthesis class U; SPP1: Secreted phosphoprotein 1.

3 讨论

本研究通过 GEO、TCGA 和 ICGC 数据库筛选出 365 个在 HCC 中差异表达且与索拉非尼敏感性相关的基因,通过单因素 Cox 回归和 LASSO 回归分析,构建了包含 7 个关键基因 (*CCT3*、*EPO*、*FTCD*、*G6PD*、*KIF20A*、*PIGU*、*SPP1*) 的预后模型,有望为临床个体化治疗 HCC 提供理论依据。

CCT3 是分子伴侣家族的重要成员,参与多种恶性肿瘤的发展与侵袭过程,在 HCC 中能够启动肿瘤细胞的转录和增殖^[16]。*EPO* 的主要功能是刺激红细胞生成,研究发现 *EPO* 与其受体在乳腺癌、膀胱癌中能够促进肿瘤细胞的增殖和恶性表型^[17-18]。*FTCD* 是一种双功能酶,有研究表明, *FTCD* 的过表达能够抑制 HCC 细胞的增殖并促进其凋亡^[19]。*G6PD* 是磷酸戊糖途径的第 1 个限速酶,在多种代谢过程中发挥重要作用。*G6PD* 在多种肿瘤中表达上调,并且与癌细胞的转化、转移及药物耐药等过程相关^[20]。另有研究发现在索拉非尼耐药的细胞中 *G6PD* 的表达下调,与耐药细胞对葡萄糖需求增高从而抑制葡萄糖进入磷酸戊糖途径相关^[21]。*KIF20A* 是驱动蛋白超家族的成员,能够调节细胞有丝分裂。*KIF20A* 几乎在所有肿瘤中高表达,可能是一种预后标志物。在乳腺癌、前列腺癌和结直肠癌中, *KIF20A* 能够促进肿瘤细胞耐药^[22]。在 HCC 中,有研究者发现 *KIF20A* 敲低能够抑制 HCC 细胞增殖并增强其对顺铂和索拉非尼的敏感性^[23]。*PIGU* 是糖基磷脂酰肌醇转氨酶 (glycosylphosphatidylinositol transamidase, GPI-T) 复合物的重要亚基,参与糖基磷脂酰肌醇中长链脂肪酸的识别,并通过增加 GPI-T 的活性促进癌症进展。在 HCC 中, *PIGU* 通过 NF- κ B 信号通路促进肝细胞活力、细胞周期进程、迁移和侵袭,并抑制细胞凋亡^[24]。*SPP1* 又称骨桥蛋白,在与癌症发展相关的慢性炎症期间调节宿主免疫、促进肿瘤增殖、抑制肿瘤细胞凋亡,与多种肿瘤预后不良相关^[25]。有证据表明, *SPP1* 的表达在某些癌种中与化疗耐药性相关^[26]。本研究结果与上述研究基本一致,除了 *FTCD* 为 HCC 预后的保护因素外,其余 6 个关键基因均为 HCC 预后的危险因素。

为进一步探索预后模型在 HCC 中的应用,本研究分析了训练集 TCGA-LIHC 队列和验证集

ICGC-LIRI 队列中风险评分对 HCC 患者预后的影响,结果表明,与低风险组的患者相比,高风险组患者的总生存期短,并且风险评分与临床分期和组织病理学分级等临床特征相关,是 HCC 的独立预后因素,提示该模型对 HCC 患者的预后具有一定的预测能力。由于这些基因是基于对索拉非尼治疗的反应性筛选出来的,本研究还利用 GDSC 数据库中的药物 IC₅₀ 信息预测了 TCGA-LIHC 队列 HCC 患者对索拉非尼的敏感性,探索了不同风险分组与索拉非尼 IC₅₀ 的关系,结果显示高风险组的 HCC 患者有较低的 IC₅₀,表明高风险组的患者可能对索拉非尼更加敏感。

综上所述,本研究基于 *CCT3*、*EPO*、*FTCD*、*G6PD*、*KIF20A*、*PIGU*、*SPP1* 基因构建的预后模型对 HCC 患者的预后及索拉非尼敏感性具有一定的预测能力。虽然生物信息学结果显示这些基因与索拉非尼敏感性相关,但大部分基因与索拉非尼耐药的关系还未见报道,值得进一步探索。本研究的局限在于所有分析是建立在生物信息学基础上的,缺少相关分子生物学实验的验证,后续本课题组将开展相关实验工作,进一步探索关键基因与索拉非尼耐药的关系及机制,并评估该模型在 HCC 的风险分层、预后评价和疗效评估等方面的作用。

[参考文献]

- [1] SUNG H, FERLAY J, SIEGEL R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2021, 71(3): 209-249. DOI: 10.3322/caac.21660.
- [2] CAO W, CHEN H D, YU Y W, et al. Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020[J]. Chin Med J, 2021, 134(7): 783-791. DOI: 10.1097/CM9.0000000000001474.
- [3] YANG J D, HAINAUT P, GORES G J, et al. A global view of hepatocellular carcinoma: trends, risk, prevention and management[J]. Nat Rev Gastroenterol Hepatol, 2019, 16(10): 589-604. DOI: 10.1038/s41575-019-0186-y.
- [4] CASAK S J, DONOGHUE M, FASHOYIN-AJE L, et al. FDA approval summary: atezolizumab plus bevacizumab for the treatment of patients with advanced unresectable or metastatic hepatocellular carcinoma[J]. Clin Cancer Res, 2021, 27(7): 1836-1841. DOI: 10.1158/1078-0432.CCR-20-3407.

- [5] HUANG A, YANG X R, CHUNG W Y, et al. Targeted therapy for hepatocellular carcinoma[J]. *Signal Transduct Target Ther*, 2020, 5(1): 146. DOI: 10.1038/s41392-020-00264-x.
- [6] KULIK L, EL-SERAG H B. Epidemiology and management of hepatocellular carcinoma[J]. *Gastroenterology*, 2019, 156(2): 477-491.e1. DOI: 10.1053/j.gastro.2018.08.065.
- [7] LLOVET J M, ZUCMAN-ROSSI J, PIKARSKY E, et al. Hepatocellular carcinoma[J]. *Nat Rev Dis Primers*, 2016, 2: 16018. DOI: 10.1038/nrdp.2016.18.
- [8] PINYOL R, MONTAL R, BASSAGANYAS L, et al. Molecular predictors of prevention of recurrence in HCC with sorafenib as adjuvant treatment and prognostic factors in the phase 3 STORM trial[J]. *Gut*, 2019, 68(6): 1065-1075. DOI: 10.1136/gutjnl-2018-316408.
- [9] RITCHIE M E, PHIPSON B, WU D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies[J]. *Nucleic Acids Res*, 2015, 43(7): e47. DOI: 10.1093/nar/gkv007.
- [10] CHEN H, BOUTROS P C. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R[J]. *BMC Bioinformatics*, 2011, 12: 35. DOI: 10.1186/1471-2105-12-35.
- [11] YU G, WANG L G, HAN Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters[J]. *OMICS*, 2012, 16(5): 284-287. DOI: 10.1089/omi.2011.0118.
- [12] GAO J, KWAN P W, SHI D. Sparse kernel learning with LASSO and Bayesian inference algorithm[J]. *Neural Netw*, 2010, 23(2): 257-264. DOI: 10.1016/j.neunet.2009.07.001.
- [13] FRIEDMAN J, HASTIE T, TIBSHIRANI R. Regularization paths for generalized linear models via coordinate descent[J]. *J Stat Softw*, 2010, 33(1): 1-22. DOI: 10.18637/jss.v033.i01.
- [14] ITO K, MURPHY D. Application of ggplot2 to pharmacometric graphics[J]. *CPT Pharmacometrics Syst Pharmacol*, 2013, 2(10): e79. DOI: 10.1038/psp.2013.56.
- [15] MAESER D, GRUENER R F, HUANG R S. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data[J]. *Brief Bioinform*, 2021, 22(6): bbab260. DOI: 10.1093/bib/bbab260.
- [16] LIU W, LU Y, YAN X, et al. Current understanding on the role of CCT3 in cancer research[J]. *Front Oncol*, 2022, 12: 961733. DOI: 10.3389/fonc.2022.961733.
- [17] REINBOTHE S, LARSSON A M, VAAPIL M, et al. EPO-independent functional EPO receptor in breast cancer enhances estrogen receptor activity and promotes cell proliferation[J]. *Biochem Biophys Res Commun*, 2014, 445(1): 163-169. DOI: 10.1016/j.bbrc.2014.01.165.
- [18] PARK S L, WON S Y, SONG J H, et al. EPO gene expression induces the proliferation, migration and invasion of bladder cancer cells through the p21WAF1-mediated ERK1/2/NF- κ B/MMP-9 pathway[J]. *Oncol Rep*, 2014, 32(5): 2207-2214. DOI: 10.3892/or.2014.3428.
- [19] CHEN J, YANG Y, LIN B, et al. Hollow mesoporous organosilica nanotheranostics incorporating formimidoyltransferase cyclodeaminase (FTCD) plasmids for magnetic resonance imaging and tetrahydrofolate metabolism fission on hepatocellular carcinoma[J]. *Int J Pharm*, 2022, 612: 121281. DOI: 10.1016/j.ijpharm.2021.121281.
- [20] YANG H C, STERN A, CHIU D T Y. G6PD: a hub for metabolic reprogramming and redox signaling in cancer [J]. *Biomed J*, 2021, 44(3): 285-292. DOI: 10.1016/j.bj.2020.08.001.
- [21] YOU X, JIANG W, LU W, et al. Metabolic reprogramming and redox adaptation in sorafenib-resistant leukemia cells: detected by untargeted metabolomics and stable isotope tracing analysis[J]. *Cancer Commun (Lond)*, 2019, 39(1): 17. DOI: 10.1186/s40880-019-0362-z.
- [22] JIN Z, PENG F, ZHANG C, et al. Expression, regulating mechanism and therapeutic target of KIF20A in multiple cancer[J]. *Heliyon*, 2023, 9(2): e13195. DOI: 10.1016/j.heliyon.2023.e13195.
- [23] WU C, QI X, QIU Z, et al. Low expression of KIF20A suppresses cell proliferation, promotes chemosensitivity and is associated with better prognosis in HCC[J]. *Aging*, 2021, 13(18): 22148-22163. DOI: 10.18632/aging.203494.
- [24] WEI X, YANG W, ZHANG F, et al. PIGU promotes hepatocellular carcinoma progression through activating NF- κ B pathway and increasing immune escape[J]. *Life Sci*, 2020, 260: 118476. DOI: 10.1016/j.lfs.2020.118476.
- [25] LAMORT A S, GIOPANO I, PSALLIDAS I, et al. Osteopontin as a link between inflammation and cancer: the Thorax in the spotlight[J]. *Cells*, 2019, 8(8): 815. DOI: 10.3390/cells8080815.
- [26] PANG X, GONG K, ZHANG X, et al. Osteopontin as a multifaceted driver of bone metastasis and drug resistance[J]. *Pharmacol Res*, 2019, 144: 235-244. DOI: 10.1016/j.phrs.2019.04.030.