

# 基于TCGA数据库构建肝细胞癌双硫死亡相关基因(DRGs)预后风险模型及评价

王结珍, 梁培松 (中山市人民医院检验医学中心, 广东中山 528403)

**摘要:** **目的** 基于癌症基因组图谱 (the cancer genome atlas, TCGA) 数据库构建肝细胞癌 (hepatocellular carcinoma, HCC) 双硫死亡相关基因 (disulfidptosis-related genes, DRGs) 预后风险模型及评价。**方法** 通过生物信息学方法分析 TCGA 数据库中 371 例 HCC 样本及 50 例癌旁样本中 15 个 DRGs 的表达情况, 并进行基因本体 (gene ontology, GO) 功能注释和京都基因和基因组百科全书 (Kyoto encyclopedia of genes and genomes, KEGG) 富集分析、Kaplan-Meier (KM) 生存分析; 通过单因素 COX 回归分析筛选出有统计学意义的 DRGs, 通过 LASSO 回归分析及多因素 COX 回归分析筛选出关键 DRGs 构建预后风险模型, 并根据风险评分将 HCC 患者分为高风险组和低风险组, 制作 KM 生存曲线和时间依赖的受试者工作特征 (receiver operator characteristic, ROC) 曲线进行验证评价。**结果** 与癌旁样本相比, HCC 样本 15 个 DRGs 中 FLNA, MYH9, TLN1, ACTB, MYL6, CAPZB, DSTN, ACTN4, SLC7A11, INF2, CD2AP, PDLIM1 和 FLNB 均表达上调, 且差异具有统计学意义 ( $t=1\ 793 \sim 6\ 310$ , 均  $P<0.001$ ); 经 GO 功能注释和 KEGG 富集分析显示, DRGs 主要与肌动蛋白细胞骨架和细胞黏附相关的生物过程或途径密切相关。经 KM 生存分析结果显示, SLC7A11, INF2, CD2AP, MYL6, ACTB 高表达组生存率低于低表达组 [ $HR=1.46(1.03 \sim 2.07) \sim 1.93(1.36 \sim 2.75)$ , 均  $P<0.05$ ]。通过单因素 COX 回归分析、LASSO 分析及多因素 COX 回归分析构建预后风险模型  $riskscore=(0.247 \times SLC7A11)+(0.289 \times INF2)+(0.076 \times CD2AP)+(0.06 \times MYL6)$ ; 计算样本的风险评分, 风险评分越高, 预后不良的 HCC 患者人数越多; KM 生存分析显示高风险组的总生存率比低风险组低; 1, 3, 5 年的 AUC 值分别为 0.709, 0.661 和 0.648; 通过多因素 COX 回归分析表明 SLC7A11 [ $HR=1.832(1.274 \sim 2.636)$ ,  $P=0.001$ ] 是独立的预后危险因素。**结论** 四个 DRGs 构建的预后风险模型在预测 HCC 患者预后情况有一定的作用。

**关键词:** 肝细胞癌; 癌症基因组图谱; 双硫死亡相关基因; 预后风险模型

**中图分类号:** R735.7; R730.43 **文献标识码:** A **文章编号:** 1671-7414 (2024) 02-086-05

**doi:**10.3969/j.issn.1671-7414.2024.02.016

## Construction and Evaluation of the Disulfidptosis-Related Genes (DRGs) Prognostic Risk Model for Hepatocellular Carcinoma Based on TCGA Database

WANG Jiezheng, LIANG Peisong (Medical Laboratory Center, Zhongshan People's Hospital, Guangdong Zhongshan 528403, China)

**Abstract:** **Objective** To construct and evaluate a disulfidptosis-related genes (DRGs) prognostic risk model for hepatocellular carcinoma (HCC) based on the cancer genome atlas (TCGA) database. **Methods** The expression of 15 DRGs in 371 HCC samples and 50 adjacent cancer samples in the TCGA database was analyzed using bioinformatics methods, and then gene ontology (GO) functional annotation, Kyoto encyclopedia of genes and genomes (KEGG) enrichment analysis and Kaplan-Meier (KM) survival analysis were performed. Statistical significant DRGs were screened through univariate COX regression analysis, and key DRGs were selected through LASSO regression analysis and multivariate COX regression analysis to construct a prognostic risk model. HCC patients were divided into high-risk and low-risk groups based on risk scores, and the KM survival curves and time-dependent receiver operator characteristic (ROC) curves were created to validate and evaluate prognostic risk models. **Results** Compared with the adjacent cancer samples, FLNA, MYH9, TLN1, ACTB, MYL6, CAPZB, DSTN, ACTN4, SLC7A11, INF2, CD2AP, PDLIM1, and FLNB were all upregulated in the 15 DRGs of HCC samples, and the differences were significant ( $t=1\ 793 \sim 6\ 310$ , all  $P<0.001$ ). According to GO functional annotation and KEGG enrichment analysis, DRGs were closely related to biological processes or pathways related to actin cytoskeleton and cell adhesion. The results of KM survival analysis showed that the survival rate of the high expression group of SLC7A11, INF2, CD2AP, MYL6, and ACTB were lower than that of the low expression group [ $HR=1.46(1.03 \sim 2.07) \sim 1.93(1.36 \sim 2.75)$ , all  $P<0.05$ ]. Univariate COX regression analysis, LASSO analysis, and multivariate COX regression analysis were used

**作者简介:** 王结珍 (1985-), 女, 本科, 副主任技师, 研究方向: 临床检验, E-mail: wanglei8511@163.com。

**通讯作者:** 梁培松 (1976-), 男, 本科, 主任技师, 主要从事临床基础免疫工作, E-mail: mark113388@163.com。

to construct a prognostic risk model, with risk score=(0.247×SLC7A11)+(0.289×INF2)+(0.076×CD2AP)+(0.06×MYL6). The risk score of the sample in this model was calculated, and the higher the risk score, the more HCC patients with poor prognosis. KM survival analysis showed that the overall survival rate of the high-risk group was lower than that of the low-risk group. The AUC values for 1, 3, and 5 years were 0.709, 0.661, and 0.648, respectively. Multivariate COX regression analysis showed that SLC7A11 [HR=1.832(1.274~2.636),  $P=0.001$ ] was an independent prognostic risk factor. **Conclusion** The prognostic risk model was constructed by four DRGs, which has a certain role in predicting the prognosis of HCC patients.

**Keywords:** hepatocellular carcinoma; the cancer genome atlas; disulfidptosis-related genes; prognostic risk model

肝癌已成为全球第二大癌症相关死亡的原因, 而我国的新增肝癌病例及死亡率也有所上升<sup>[1]</sup>, 因此迫切需要有针对性的措施以遏制肝癌发病及死亡率的增加。肝癌包括原发性肝癌、转移性肝癌。原发性肝癌是指来自于肝脏自身的恶性肿瘤, 包括肝细胞癌、肝内胆管癌和混合性肝癌, 而其中肝细胞癌占原发性肝癌 85% ~ 90%。双硫死亡 (disulfidptosis) 是一种新型的细胞死亡方式, 独立于现存的细胞凋亡、铁死亡、铜死亡等细胞程序性死亡, 是由细胞内过量胱氨酸堆积引起的二硫化物应激导致的快速死亡<sup>[2]</sup>。双硫死亡的潜在能力使其成为癌症的治疗靶标<sup>[3]</sup>。本研究将通过生物信息学方法构建肝细胞癌 (hepatocellular carcinoma, HCC) 双硫死亡相关基因 (disulfidptosisrelated genes, DRGs) 的预后风险模型, 为 HCC 的预后预测及治疗提供新的思路。

## 1 材料与方法

**1.1 资料来源** 研究所用数据资源来源于癌症基因组图谱 (the cancer genome atlas, TCGA) 数据库下载的 HCC(TCGALIHCC) 相关数据, 包括表达谱数据及相应的临床数据, 去除重复数据及无临床信息的数据, 最终获得 371 例 HCC 样本及 50 例癌旁样本作为研究对象。DRGs 来源于文献报道<sup>[4]</sup>, 分别为 FLNA, FLNB, MYH9, TLN1, ACTB, MYL6, MYH10, CAPZB, IQGAP1, DSTN, ACTN4, SLC7A11, INF2, CD2AP, PDLIM1。

### 1.2 方法

**1.2.1 DRGs 在 HCC 中的表达情况:** 根据数据格式特征情况选择合适的统计方法, 采用 “stats” “car” 包进行统计分析 (采用 WilCOXon rank sum test 秩和检验), 用 “ggplot2” 包对数据进行可视化, 比较 TCGA 数据库中 DRGs 在 HCC 及癌旁组织中的表达差异情况,  $P < 0.05$  为差异有统计学意义。

**1.2.2 DRGs 的 GO 功能注释和 KEGG 富集分析:** 采用 “DESeq2” 包获取两组间的差异表达基因,  $\text{padj} < 0.05$  表示差异有统计学意义; 采用 “clusterProfiler” “ggplot2” 包对 DRGs 进行 GO 功能注释和 KEGG 富集分析及可视化,  $P < 0.05$  表示差异有统计学意义。GO 功能注释包含生物学过程 (biological process, BP)、细胞成分 (cellular

component, CC) 和分子功能 (molecular function, MF)。

**1.2.3 DRGs 的 KM 生存分析:** 采用 “survival” “survminer” 及 “ggplot2” 包进行统计分析及可视化, 绘制 KM 生存分析曲线图,  $P < 0.05$  表示差异有统计学意义。

**1.2.4 DRGs 的预后风险模型的构建:** 采用 “survival” “rms” 包对 DRGs 进行单因素 COX 回归分析, 筛选具有统计学意义的 HCC 预后 DRGs,  $P < 0.05$  表示差异有统计学意义。采用 “glmnet” 包对筛选出 DRGs 进行 LASSO 回归分析构建 DRGs 预后风险模型。每个患者的风险评分 (riskscore) =  $\sum$  (各个基因的表达水平 × 相应系数)。

**1.2.5 DRGs 的预后风险模型的评价:** 根据中位风险评分, 将患者分为高风险组和低风险组, 采用 “ggplot2” 包绘制风险因子图。使用 “survival” “ggplot2” 包进行 KM 总生存分析, 并绘制 KM 总生存曲线。使用 “timeROC” “ggplot2” 包绘制时间依赖 ROC, 评估模型 1, 3, 5 年生存的预测效能。对构成预后风险模型的 DRGs 进行多因素 COX 回归分析, 确定影响总体生存率的独立预后危险因素。 $P < 0.05$  表示差异有统计学意义。

**1.3 统计学分析** 采用 RStudio 4.2.1 版本软件对数据进行统计分析,  $P < 0.05$  为差异有统计学意义。

## 2 结果

**2.1 DRGs 的表达情况** 见图 1。在对 TCGA 数据库中 HCC 相关数据分析显示 FLNA, MYH9, TLN1, ACTB, MYL6, CAPZB, DSTN, ACTN4, SLC7A11, INF2, CD2AP, PDLIM1 和 FLNB 在 HCC 中均为高表达, 差异具有统计学意义 ( $t=5\ 419, 3\ 286, 4\ 088, 3\ 545, 1\ 820, 1\ 793, 3\ 480, 2\ 525, 1\ 888, 6\ 310, 3\ 676, 5\ 121, 3\ 860$ , 均  $P < 0.001$ )。

**2.2 GO 功能注释和 KEGG 富集分析** DRGs 的 GO 功能富集分析结果显示: BP 主要富集于肌动蛋白丝运动、肌动蛋白丝过程的调控、细胞形态发生的调控、肌动蛋白细胞骨架组织的调控和肌动蛋白细胞骨架重组等; CC 主要富集于肌动蛋白细胞骨架、肌动蛋白丝束、基于肌动蛋白的细胞投射簇、收缩肌动蛋白丝束和肌动蛋白等; MF 主要富集于肌动蛋白结合、肌动蛋白丝结合、微丝运动活

性、细胞骨架的结构成分和细胞骨架运动活性等。DRGs 的 KEGG 信号通路功能富集分析结果显示：主要富集于紧密连接、局灶性黏连、肌动蛋白细胞骨架的调节、黏合连接、癌症中的蛋白聚糖等的信号通路。

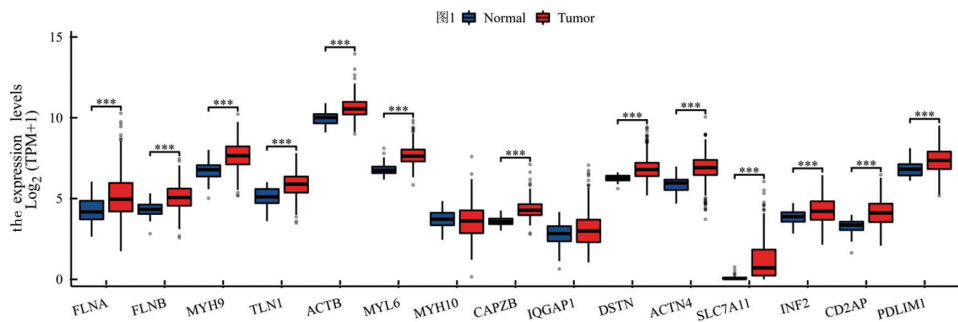


图1 15个DRGs在肝癌组织和癌旁组织中的表达情况

2.3 HCC 中 DRGs 高低两组表达量的 KM 生存分析 结果显示，SLC7A11，INF2，CD2AP，MYL6，ACTB 高表达组总生存率低于低表达组，见图2（A ~ E）。

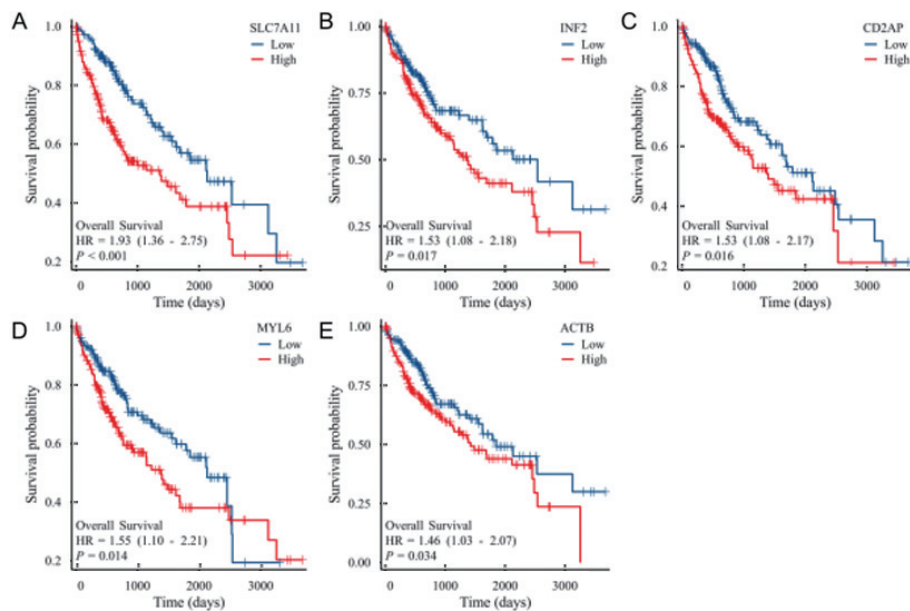


图2 SLC7A11，INF2，CD2AP，MYL6，ACTB在HCC中的KM图

2.4 预后风险模型的构建 根据HCC患者预后OS对DRGs进行单因素COX回归分析，筛选出DRGs基因SLC7A11，INF2，CD2AP，MYL6和ACTB(P < 0.05)，见图3A。进一步进行Lasso回归分析(见图3B)筛选出SLC7A11，INF2，CD2AP和MYL6用于构建预后风险模型， $\text{riskscore} = (0.247 \times \text{SLC7A11}) + (0.289 \times \text{INF2}) + (0.076 \times \text{CD2AP}) + (0.06 \times \text{MYL6})$ 。

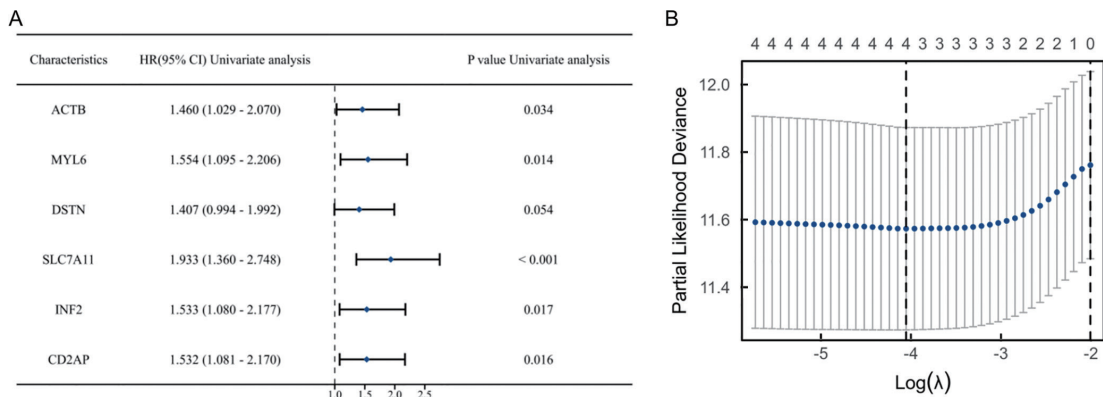
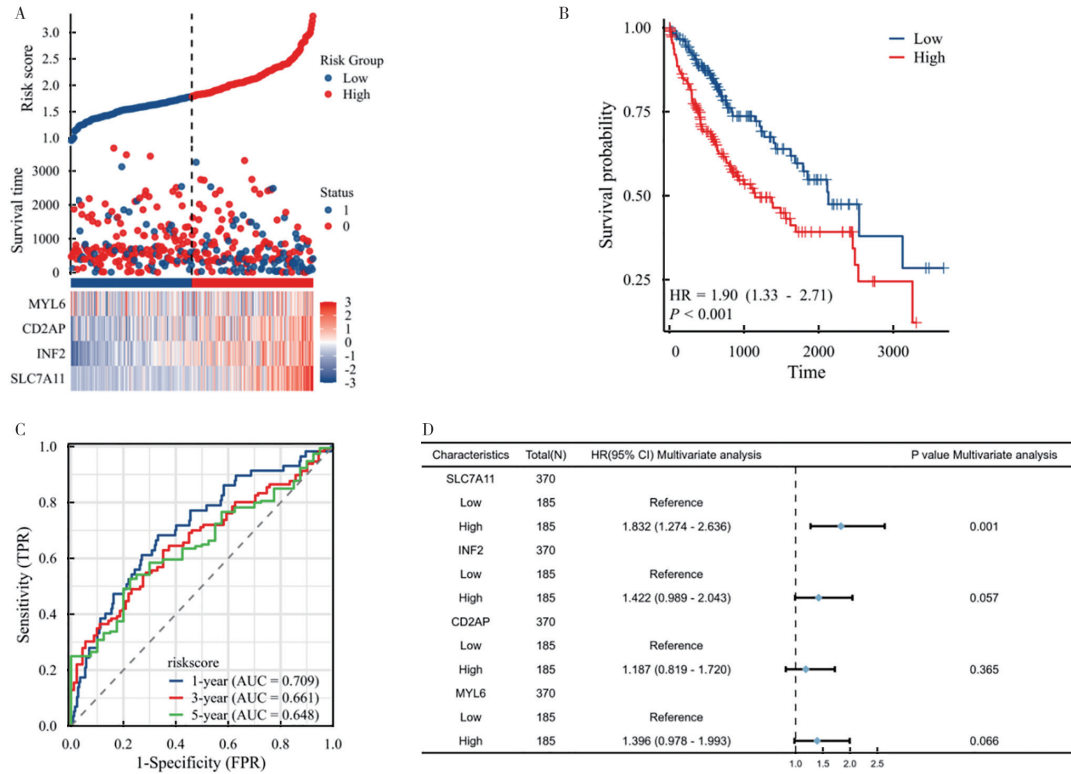


图3 单因素COX回归分析森林图及LASSO回归分析图



2.5 预后风险模型的评价 基于上述风险评分函数公式计算每个患者的风险评分,以风险评分中位数为界将患者分为高风险组和低风险组。从图 4A 中可见随着风险评分的升高,患者的生存期缩短且死亡例数增加。从 KM 生存曲线分析显示,高风险组患者的预后较低风险组差 ( $P < 0.001$ ),见图 4B。



A. 风险因子图 (Status: 1 代表死亡, 2 代表生存); B. riskscore-KM 图; C. riskscore-ROC 图; D. 对 SLC7A11, INF2, CD2AP 和 MYL6 进行多因素 COX 分析森林图。

图 4 预后风险模型

### 3 讨论

双硫死亡的主要机制是溶质载体家族 7 成员 11 (SLC7A11) 高表达的细胞由于还原型烟酰胺腺嘌呤二核苷酸磷酸 (NADPH) 供应不足导致细胞内的胱氨酸等二硫化物异常积累, 诱发二硫化物应激, 使得肌动蛋白细胞骨架中蛋白二硫键含量增高, 造成肌动蛋白丝收缩, 破坏了细胞骨架结构, 并最终导致了细胞死亡<sup>[5]</sup>。因肝细胞癌 (HCC) 的死亡率高, 且缺乏良好的早期诊断指标, 已有不少关于寻找早期特异性筛查肝癌肿瘤标志物的研究<sup>[6]</sup>。目前许多癌症疗法都是通过细胞凋亡来杀死癌细胞的, 然而, 很多癌细胞已出现逃逸, 最终导致治疗耐药和疾病复发。现今有研究发现, 许多癌症细胞中均高表达 SLC7A11, 因此双硫死亡有望成为癌症治疗的方法<sup>[7]</sup>。已有研究发现双硫死亡相关基因与肝癌的发生发展有关<sup>[8]</sup>。

本文采用生物信息学方法对 15 个双硫死亡相关基因 (DRGs) 进行数据分析显示 HCC 患者

ROC 曲线分析显示, 风险评分模型对 HCC 患者 1, 3, 5 年总生存期预测的 AUC 值分别为 0.709, 0.661 和 0.648, 见图 4C; 通过对 SLC7A11, INF2, CD2AP, MYL6 进行多因素 COX 分析可知, SLC7A11 为独立预后危险因素, 见图 4D。

的表达均不同程度升高, 其中 SLC7A11, INF2, CD2AP, MYL6 和 ACTB 过度表达预示着其生存时间缩短, 预后不良。本研究通过对 DRGs 进行 GO 和 KEGG 富集分析显示主要与肌动蛋白细胞骨架和细胞黏附相关的生物过程或途径密切相关。LIU 等<sup>[5]</sup>人的研究也发现肌动蛋白丝在双硫死亡中可能具有重要作用。目前已有研究表明细胞骨架和细胞黏附在癌症的发生、发展, 特别是癌症的侵袭和转移密切相关, 细胞骨架将有望成为新的癌症治疗靶点<sup>[9]</sup>。可见双硫死亡相关基因有望成为 HCC 治疗的靶点和预后的标志。

为进一步筛选与 HCC 预后密切相关的 DRGs 并构建预后风险模型, 对 SLC7A11, INF2, CD2AP, MYL6 和 ACTB 进行 LASS 分析及多因素 COX 回归分析, 最终得到预后风险模型:  $\text{riskscore} = (0.247 \times \text{SLC7A11}) + (0.289 \times \text{INF2}) + (0.076 \times \text{CD2AP}) + (0.06 \times \text{MYL6})$ , 且发现 SLC7A11 为独立预后危险因素。SLC7A11 基因是一种参与氨基酸代谢维持

细胞内氧化还原稳定的重要跨膜蛋白,是溶质载体家族中的一员。其既参与了铁死亡也参与了双硫死亡。SLC7A11的高表达对铁死亡有抑制作用,而对双硫死亡则有促进作用。国内已有研究表明,HCC患者血清 SLC7A11 浓度水平随着病程的发展呈增高趋势,对早期诊断及预后预测有良好的临床应用价值<sup>[10]</sup>。YAN 等<sup>[11]</sup>人的研究也发现 SLC7A11 的高表达促进了肿瘤生长,但抑制了肿瘤的转移。本研究显示 SLC7A11 在 HCC 中也呈现高表达,因此在研究如何利用双硫死亡诱导肿瘤细胞死亡时应注意其对铁死亡的抑制作用。INF2 基因是一种重要的肌动蛋白成核因子,具有调节肌动蛋白解聚和聚合的双重作用。目前越来越多的研究表明 INF2 与肾脏疾病及癌症有关,且在不同的癌症中的作用不同,目前对其机制的研究仍未明确<sup>[12-13]</sup>。CD2AP 基因编码一种调节肌动蛋白细胞骨架的“支架分子”。KURILLA 等<sup>[14]</sup>研究发现,TKS4 和 CD2AP 相结合,参与了动态肌动蛋白组装相关过程与癌症细胞的迁移,但其在肿瘤形成过程中的共同指导作用尚不清楚。MYL6 基因编码肌球蛋白碱性轻链<sup>[15]</sup>,目前其机制尚未明确。在本研究中也发现这三个基因 INF2, CD2AP 和 MYL6 在 HCC 中也呈现表达增高,且随着其增高,患者生存期缩短。

综上所述,这四个基因均与细胞骨架结构有密切的关系,其有望成为 HCC 的治疗靶点;且其在 HCC 中随着病情的发展,均有不同程度的增高,因此其有可能成为 HCC 的诊断及预测预后的新的生物学标志物。基于现今研究双硫死亡的研究较少,且本研究属于纯数据分析,缺少相关实验的验证,存在一定的局限性,因此本研究基于 TCGA 数据库所构建的 HCC 双硫死亡相关基因预后风险模型对 HCC 患者预后评估和指导作用仍需进行相关大量临床实验进行验证,但不可否认其为我们以后的 HCC 早期诊断和预后预测研究提供了方向。

#### 参考文献:

- [1] CAO Wei, CHEN Hongda, YU Yiwen, et al. Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020[J]. Chinese Medical Journal, 2021, 134(7): 783-791.
- [2] SHAO Dongpu, SHI Lei, JI Honglei. Disulfidptosis: disulfide stress mediates a novel cell death pathway via actin cytoskeletal vulnerability[J]. Molecules and Cells, 2023, 46(7): 414-416.
- [3] ZHENG Peijie, ZHOU Chuntao, DING Yuemin, et al. Disulfidptosis: a new target for metabolic cancer therapy[J]. Journal of Experimental & Clinical Cancer Research, 2023, 42(1): 103.
- [4] WANG Tianbing, GUO Kai, ZHANG Di, et al. Disulfidptosis classification of hepatocellular carcinoma reveals correlation with clinical prognosis and immune profile[J]. International Immunopharmacology, 2023, 120: 110368.
- [5] LIU Xiaoguang, NIE Litong, ZHANG Yilei, et al. Actin cytoskeleton vulnerability to disulfide stress mediates disulfidptosis[J]. Nature Cell Biology, 2023, 25(3): 404-414.
- [6] 卢美红, 杨曙梅, 王峰, 等. 血清 miRNA-574-3p, AFP, IGF-2 水平联合检测在肝癌早期诊断中的价值研究 [J]. 现代检验医学杂志, 2022, 37(2): 33-36. LU Meihong, YANG Shumei, WANG Feng, et al. Study on the value of combined detection of serum miRNA-574-3p, AFP and IGF-2 levels in the early diagnosis of liver cancer[J]. Journal of Modern Laboratory Medicine, 2022, 37(2): 33-36.
- [7] WANG Zhitang, DU Xianqiang, LIAN Weibin, et al. A novel disulfidptosis-associated expression pattern in breast cancer based on machine learning[J]. Frontiers in Genetics, 2023, 14: 1193944.
- [8] LI Xiaomin, LIU Shanpeng, LI Yu, et al. Identification of disulfidptosis-related genes with immune infiltration in hepatocellular carcinoma[J]. Heliyon, 2023, 9(8): e18436.
- [9] 方全, 韦花媚, 浦洞. 肌动蛋白丝细胞骨架重塑在肝细胞癌中的研究进展 [J]. 右江医学, 2023, 51(4): 289-293. FANG Quan, WEI Huamei, PU Jian. Research progress of actin cytoskeleton remodeling in hepatocellular carcinoma[J]. Chinese Youjiang Medical Journal, 2023, 51(4): 289-293.
- [10] 詹有芳, 张珏, 王鹏. 肝细胞癌患者血清 SLC7A11 的变化及临床意义 [J]. 肝脏, 2022, 27(12): 1288-1291. ZHAN Youfang, ZHANG Jue, WANG Peng. Changes and clinical significance of serum SLC7A11 in patients with hepatocellular carcinoma[J]. Chinese Hepatology, 2022, 27(12): 1288-1291.
- [11] YAN Yuelong, TENG Hongqi, HANG Qinglei, et al. SLC7A11 expression level dictates differential responses to oxidative stress in cancer cells[J]. Nature Communications, 2023, 14(1): 3673.
- [12] LABAT-DE-HOZ L, ALONSO M A. The formin INF2 in disease: progress from 10 years of research[J]. Cellular and Molecular Life Sciences, 2020, 77(22): 4581-4600.
- [13] ZHAO Yiting, ZHANG Hui, WANG Haibiao, et al. Role of formin INF2 in human diseases[J]. Molecular Biology Reports, 2022, 49(1): 735-746.
- [14] KURILLA A, LÁSZLÓ L, TAKÁCS T, et al. Studying the association of TKS4 and CD2AP scaffold proteins and their implications in the partial Epithelial-Mesenchymal transition (EMT) process[J]. International Journal of Molecular Sciences, 2023, 24(20): 15136.
- [15] ZHU Zhaozhong, GUO Yanjun, SHI Huwenbo, et al. Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank[J]. Journal of Allergy and Clinical Immunology, 2020, 145(2): 537-549.

收稿日期: 2023-08-01

修回日期: 2023-11-26