

三种 miRNA 对肝癌预后预测的生物信息学分析模型

林思其¹, 张泽鑫², 夏睿琪¹, 刘紫凤¹, 陈林静¹, 陈栩静¹, 陈祎琦¹

(1. 广州中医药大学第二临床医学院, 广东 广州 510405;

2. 广州中医药大学第一临床医学院, 广东 广州 510405)

摘要:目的 通过 TCGA 数据库构建一种 miRNA 预后模型, 预测肝癌的预后。方法 通过 edgeR 软件包对 miRNA 和 mRNA 的表达数据进行分析并标准化, 并与生存率相结合分为两组。通过单因素和多因素 Cox 回归分析构建 3 种与总生存率明显差异相关的 miRNA, 使用 ROC 曲线评价其可靠性。通过 miRDB、TargetScan 和 miRTarBase 3 个数据库筛选出 miRNA 的靶基因, 并运用 GO 和 KEGG 探究其相关作用机制。使用 Cytoscape3.7.2 和 String 数据库筛选出前 10 个中枢基因。结果 共筛选出 47 个靶基因与生存预后显著相关; ROC 曲线结果显示, 三组的 AUC 分别为 0.789、0.730、0.763, 说明模型预测 HCC 患者预后的能力; 回归分析显示, 该模型可独立影响肿瘤的预后情况, 包括 T 分期、临床分期及远处转移的存在。结论 hsa-miR-3682-3p、hsa-miR-9-5p 和 hsa-miR-139-5p 三种 miRNA 的生物信息学分析模型可以预测 HCC 患者的预后, 且可以作为 HCC 中独立的预后因素。

关键词: 肝细胞癌; microRNA; 预后; TCGA

中图分类号: R735.7; R318.04

文献标识码: A

DOI: 10.3969/j.issn.1006-1959.2022.03.019

文章编号: 1006-1959(2022)03-0081-05

A Model of Three miRNA on the Prognosis Prediction of Hepatocellular Carcinoma by Bioinformatics Analysis

LIN Si-qi¹, ZHANG Ze-xin², XIA Rui-qi¹, LIU Zi-feng¹, CHEN Lin-jing¹, CHEN Xu-jing¹, CHEN Yi-qi¹

(1. The Second First Clinical Medical College of Guangzhou University of Chinese Medicine, Guangzhou 510405, Guangdong, China;

2. The First Clinical Medical College of Guangzhou University of Chinese Medicine, Guangzhou 510405, Guangdong, China)

Abstract: **Objective** To construct a miRNA prognostic model by TCGA database to predict the prognosis of liver cancer. **Methods** The expression data of miRNA and mRNA were analyzed and standardized by edgeR software package, and they were divided into two groups combined with survival rate. Single-factor and multi-factor Cox regression analysis was used to construct three miRNAs related to significant differences in overall survival rate, and ROC curve was used to evaluate their reliability. The target genes of miRNA were screened from miRDB, TargetScan and miRTarBase databases, and the related mechanism was explored by GO and KEGG. The top 10 central genes were screened using Cytoscape 3.7.2 and String databases. **Results** A total of 47 target genes were found to be significantly associated with survival and prognosis. ROC curve showed that the AUC of the three groups were 0.789, 0.730 and 0.763, respectively, indicating the ability of the model to predict the prognosis of HCC patients. Regression analysis showed that the model could independently affect the prognosis of tumors, including T staging, clinical staging and distant metastasis. **Conclusion** The biological analysis models of hsa-miR-3682-3p, hsa-miR-9-5p and hsa-miR-139-5p can predict the prognosis of HCC patients and can be used as independent prognostic factors in HCC.

Key words: Hepatocellular carcinoma; microRNA; Prognosis; TCGA

肝细胞癌(hepatocellular carcinoma, HCC)是世界上第 5 大常见的恶性肿瘤^[1], 全球发病率逐步上升^[2]。目前, 手术治疗是 HCC 的主要治疗方法, 然而术后 5 年复发率已达 70%^[3], 生存质量难以提高。到目前为止, 一些生物标志物已被证明与 HCC 的发生发展相关, 但其可靠性仍存在争议。miRNA 是一种非编码 RNA, 其失调会导致细胞的异常生长和生物合成, 促进肿瘤的发生发展^[4]。此外, 不同表达谱的 miRNA 可以作为肿瘤诊断和预后的分子标志物, 如血清 miR-122 可以作为 HCC 诊断标志物之一^[5]。Liu G 等^[6]已构建了 HCC 的 miRNA 特征预后模型, 但并没有进行验证和风险评估。本研究构建并验证了一种新的 miRNA 预后模型, 旨在评估 TCGA 数据

库中 HCC 患者过度生存情况, 探究三种 miRNA 信号的潜在生物学特征、肿瘤相关功能和信号通路, 为了解 HCC 模型的分子机制提供依据, 现报道如下。

1 资料与方法

1.1 数据下载和处理 miRNA 的表达数据如下:[病例(373):项目(TCGA)、项目(TCGA-LIHC)、基本位点(肝脏)、疾病类型(腺瘤和腺癌);档案(425):数据:类型(亚型表达定量分析)、类别(转录组分析)]。mRNA 表达数据如下:[病例(371):项目(TCGA)、项目(TCGA-LIHC)、基本位点(肝脏)、疾病类型(腺瘤和腺癌);档案(424):数据:类型(基因表达定量分析)、类别(转录组分析)];相关临床信息(377)(数据格式:bcr xml, 数据类别:临床)。从癌症基因组图谱(TCGA)的官方网站下载所有数据。miRNA 表达数据包含 375 个癌细胞样本和 50 个普通细胞样本, 而 mRNA 表达数据包括 374 个癌细胞样本和 50 个普通细胞样本。

基金项目: 广东大学生科技创新培育专项资金(“攀登计划”专项资金)(编号: pdjh2021b0122)

作者简介: 林思其(2000.1-), 女, 广东揭阳人, 本科

通讯作者: 张泽鑫(1995.9-), 男, 广东潮州人, 硕士研究生, 主要从事中西医结合防治恶性肿瘤方向的研究

1.2 数据分析 所有数据分析均使用 R 语言的 3.6.1 版本和相关软件包。

1.2.1 差异表达 miRNA、mRNA 的检测及与临床生存时间的联合分析 根据校正后的 P 值标准 ($|\log_2FC| > 1$ 和 $(FDR) < 0.05$), 检测出差异表达的 miRNA 和 mRNA 的数据, 使用 R 语言的边缘包进行标准化。将所有患者生存时间数据与标准化的 miRNA 和 mRNA 差异表达数据相结合。

1.2.2 样本分组、模型构建、评估和验证 使用 R 语言 v.3.6.1 的“caret”包, 将包含总生存时间和差异表达的 miRNA 数据样本, 以非特异性方式分为测试组和训练组。通过单因素 Cox 回归分析对训练组进行检验, $P < 0.05$ 。检验的结果通过 R 语言中的“Coxph”函数和“direction=both”函数进行多因素 Cox 回归分析。根据风险评分进行分组后, 该模型通过 Log-rank 检验和 Kaplan-Meier 曲线对患者生存预后进行评估, 得到中位数。使用“survivalROC”软件包评估 miRNA 模型的预测能力, 获得 3 年依赖的 ROC 曲线, 并计算曲线下面积(AUC)。

1.2.3 独立预后潜力评估 使用单因素 Cox 回归分析确定 miRNA 表达数据与 HCC 患者总生存时间, 以及其他临床信息(年龄、性别、分级、临床分期、肿瘤浸润、淋巴结和远处转移)之间的相关性。满足 $P < 0.05$ 的变量进一步使用多因素 Cox 回归进行分析, 证明其为具有独立作为预后因素的潜力。

1.2.4 三种 miRNA 的靶基因及潜在活性预测 从 miRDB、TargetScan 和 miRTarBase 三个数据库中下载 miRNA 的相关基因。使用 Perl 语言进行搜索, 按

照满足至少存在于两个数据库中的标准, 筛选出目标基因, 并绘制韦恩图, 利用 Cytoscape3.7.2 确认 miRNA 与其靶基因的相关性。对靶基因取交集, 用差异表达 mRNA 进行处理, 验证这些靶基因在 HCC 中的作用。使用 R 语言的“org.Hs.eg.db”和“clusterProfiler”软件包对所有基因进行 GO 和 KEGG 富集分析, 调整后的 $P < 0.05$, $Q < 0.05$ 。

1.2.5 中枢基因、与生存相关基因筛选 在 String 数据库中, 将中等置信度设置为 0.400, 构建 PPI 网络。使用 Cytoscape3.7.2 的 CytoHubb 插件计算基因的度值, 筛选前 10 个中枢基因。同时, 采用 Kaplan-Meier 方法, 以 $P < 0.05$ 为标准, 筛选出与生存相关的基因。

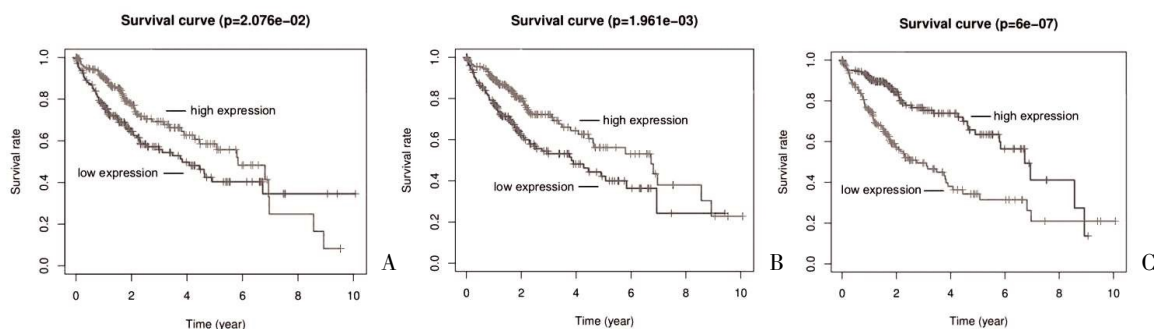
2 结果

2.1 miRNA 和 mRNA 差异表达的检测 根据以上标准, 共鉴定出 300 个差异表达 miRNA, 包括下调 40 个和上调 260 个; 6219 个差异表达 mRNA, 包括上调 4870 个和下调 1349 个。

2.2 三种 miRNA 预后模型的构建 将原始组的 miRNA ($N=371$) 结合总生存时间随机分为训练组 (184) 和测试组 (187)。训练组采用单因素 Cox 回归 ($P < 0.05$) 得到 12 个 miRNA, 见表 1; 对其进行多因素 Cox 回归分析, 并构建模型。Kaplan-Meier 方法表明 hsa-miR-139-5p、hsa-miR-9-5p、hsa-miR-3682-3p 是与患者总生存时间最显著的 3 个 miRNA ($P < 0.05$), 见图 1; 多因素 Cox 回归系数如下: miRNA 特征性风险评分 = $(-0.3325 \times \text{hsa-miR-139-5p 表达}) + (0.2717 \times \text{hsa-miR-3682-3p 表达}) + (0.0929 \times \text{hsa-miR-9-5p 表达})$ 。

表 1 单因素和多因素回归分析的差异表达 miRNA

id	单因素回归分析				多因素回归分析				
	HR	HR.95L	HR.95H	P	coef	HR	HR.95L	HR.95H	P
hsa-miR-139-3p	0.731	0.617	0.866	0.000					
hsa-miR-101-3p	0.707	0.554	0.902	0.005					
hsa-miR-1180-3p	1.326	1.074	1.637	0.009					
hsa-miR-9-5p	1.134	1.024	1.255	0.015	0.093	1.097	0.987	1.221	0.047
hsa-miR-3682-3p	1.451	1.123	1.874	0.004	0.272	1.312	1.014	1.699	0.039
hsa-miR-139-5p	0.689	0.583	0.815	0.000	-0.333	0.717	0.600	0.856	0.000
hsa-miR-9-3p	1.145	1.016	1.290	0.026					
hsa-miR-6844	1.358	1.048	1.759	0.021					
hsa-miR-3691-5p	1.313	1.006	1.715	0.045					
hsa-miR-3677-3p	1.428	1.142	1.785	0.002					
hsa-miR-301a-3p	1.331	1.054	1.681	0.017					
hsa-miR-3127-5p	1.331	1.014	1.746	0.039					



注: A: hsa-miR-9-5p; B: hsa-miR-3682-3p; C: hsa-miR-139-5p

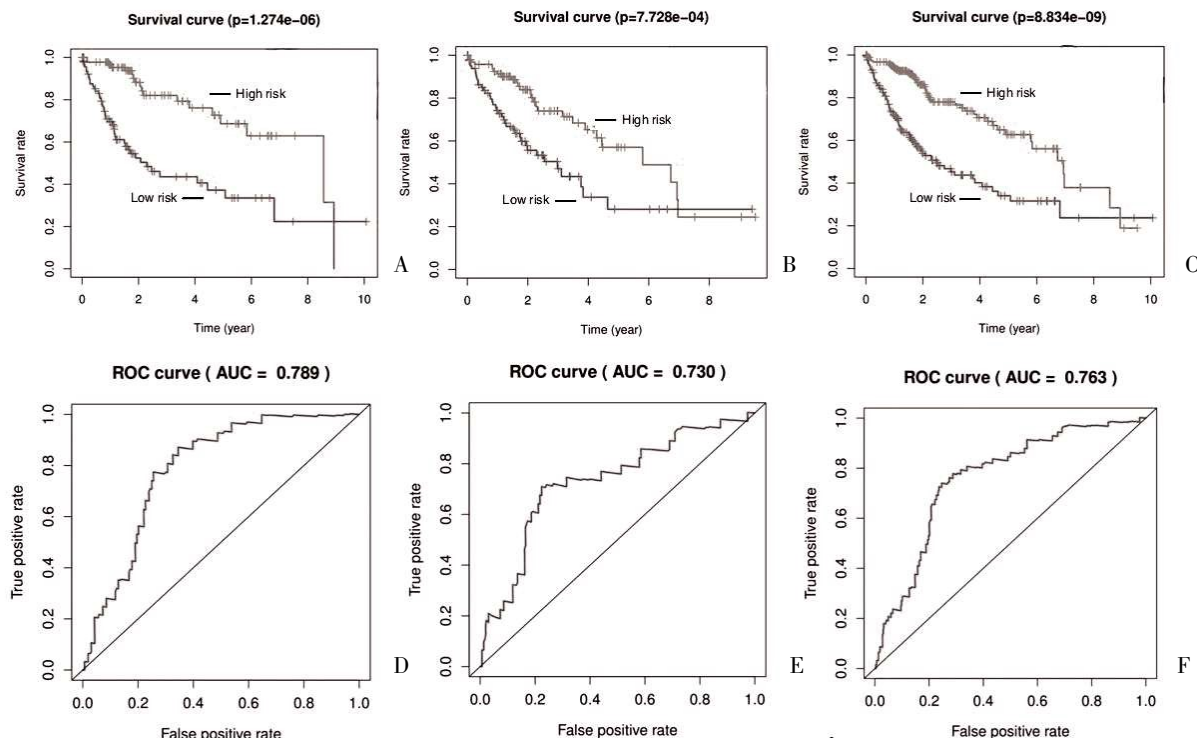
图 1 Kaplan-Meier 曲线和 Log-rank 检验筛选出的与 HCC 患者总生存期相关的 miRNA

2.3 三种 miRNA 模型在三组中的总体生存预测分组 采用中位数风险评分, Kaplan-Meier 曲线显示, 在比较低风险组和高风险组时, 三组的 P 值分别为 $P=1.274e-06$ 、 $P=7.728e-04$ 和 $P=8.834e-09$, 见图 2A~图 2C); 此外, 训练组的高风险组和低风险组的 5 年总生存率分别为 33.5% 和 68.6%; 测试组为 28.1% 和 57%; 原始组为 31.6% 和 62.7%。

2.4 对三组中的三种 miRNA 模型的评估 ROC 曲线结果显示, 三组的 AUC 分别为 0.789、0.730、0.763,

见图 2D~图 2F, 说明了该模型预测 HCC 患者生存机会的能力。在比较这三组的两个得分时, 发现高风险得分比低风险得分的死亡率高。

2.5 考虑其他临床因素时, 三种 miRNA 模型的独立性 单因素 Cox 回归分析表示, 三种 miRNA 模型与患者总生存时间明显相关; 而进一步的多因素 Cox 回归分析证明了当考虑其他临床信息时, 三种 miRNA 模型可以独立于总生存时间, 其中包括 T 分期、临床分期以及远处转移的存在, 见表 2。



注: Kaplan-Meier 曲线: A: 训练组; B: 测试组; C: 原始组; 3 年生存曲线 AUC: D: 训练组; E: 测试组; F: 原始组

图 2 三种 miRNA 预后模型的验证和评估

表2 临床特征的单因素和多因素回归分析

临床特征	单因素 Cox 回归分析				多因素 Cox 回归分析			
	HR	HR.95L	HR.95H	P value	HR	HR.95L	HR.95H	P value
Age (continuous variable)	1.006	0.988	1.024	0.538				
Gender (male vs. female)	0.776	0.483	1.247	0.295				
Grade	0.989	0.727	1.346	0.944				
Clinical stage (Ⅲ+Ⅳ vs. Ⅰ+Ⅱ)	1.864	1.452	2.392	0.000	1.014	0.393	2.620	0.977
T stage (T ₃₊₄ vs. T ₁₊₂)	1.804	1.431	2.273	0.000	1.625	0.697	3.788	0.260
M (M ₁ vs. M ₀)	3.970	1.245	12.666	0.020	1.014	0.265	3.881	0.984
N (N ₁₊₂ vs. N ₀)	2.095	0.512	8.574	0.304				
Three-miRNA signature	1.904	1.511	2.397	0.000	1.635	1.281	2.088	0.000

2.6 预测这3种miRNA的靶基因 结果显示,hsa-miR-139-5p、hsa-miR-9-5p、hsa-miR-3682-3p分别可以检测到642、1314和397个重叠基因。其中,筛选出174个基因作为三种已鉴定的miRNA的遗传靶点。为了验证这些miRNA靶基因是否参与了HCC的进展,对上调miRNA(hsa-miR-9-5p、hsa-miR-3682-3p)与下调的靶mRNA,以及下调miRNA(hsa-miR-139-5p)和上调靶mRNA取交集,并进行结果分析,最终得到了174个基因,其中包括88个上调基因和96个下调基因。

2.7 HCC相关靶基因的GO和KEGG富集分析 通过对与HCC相关的靶基因的GO注释,获得了415个结果。在这3类研究中,BP分析主要包括对神经元投射发育、轴突发生和中枢神经系统神经元分化的调控。CC分析主要包括突触膜、突触后特化和突触后膜。MF分析主要包括磷酸酯水解酶活性、DNA-结合转录激活因子活性、RNA聚合酶Ⅱ-特异性和酰胺结合。从HCC相关遗传靶点的KEGG通路中获得了6个结果,其中超过5个基因主要富集在细胞因子-细胞因子受体相互作用的信号通路中。

2.8 来自PPI网络的中枢基因和与生存相关的基因目标 从174个遗传靶点中筛选出100个,构成遗传靶点的蛋白-蛋白相互作用网络复合体,包括114条边和299个节点。共筛选10个中枢基因(ALPL、CXCL12、OIP5、TOP2A、AR、KPNA2、SLC7A2、DBT、HMGB2、MAD2L1);其中174个基因中有30个基因(CDC37L、C6、C21orf91、BEND4、CEPB3、DBT、MPDZ、HS3ST3B1、GHR、ENPEP、IL33、GPR65、NSUN6、PCDHGC5、STARD5、PDK4、OGDHL、ST8SIA6、RBMS3、RNASE4、ANXA10、ACADSB、ALPL、ANGPTL1、PDE7B、ANO1、SLC7A2、NDRG2、SELP、WNT1)与生存预后正相关,17个基因(GPSM2、EME1、COL11A1、HOXD10、LRP12、MAD2L1、KPNA2、CHST4、HMGB2、LAPTM4B、PLCB1、RAD54B、OIP5、TOP2A、TMC7、

KLHL23、MEX3A)与生存预后负相关。

3 讨论

肝细胞癌是一种高度恶性的肿瘤,极易发生肺(38.4%)、骨(32.6%)和淋巴结(24.6%)的转移^[7],5年生生存率仅18%,甚至约20%的患者在6个月后复发^[8]。因此,临床迫切需要寻找具有高敏感性和特异性的生物标志物。研究显示^[9],miRNA可能成为肿瘤生存预后的敏感生物标志物。miR-9-5p可以通过调控GOT1的表达来阻碍胰腺癌侵袭、增殖、谷氨酰胺代谢和氧化还原稳态;miR-3682-3p通过靶向肝癌中的肿瘤抑制基因GAS8,促进HCC的侵袭和迁移^[10];而miR-139-5p可以通过下调SLITRK4的表达来影响HCC细胞的侵袭和增殖能力^[11]。这三种miRNA都参与了各种肿瘤的发展调控过程,可以作为一种新的敏感生物标志物,且多种miRNA信号比单一的miRNA表现出更大的优势。本研究通过单因素和多因素Cox回归分析构建了hsa-miR-3682-3p、hsa-miR-9-5p和hsa-miR-139-5p三种miRNA特征模型,与已有的研究不同,通过miRNA分组来验证模型的可行性。

慢性炎症可以促进癌细胞免疫逃逸,而各种肿瘤细胞分泌的细胞因子与免疫细胞具有协同作用,可以促进肿瘤活性^[12]。Han KQ等^[13]发现CXCL1RNAi可抑制肿瘤细胞的凋亡和生长,提示CXCL1可能是治疗HCC的靶点。而生长激素可以上调小鼠肝脏中GAL1的表达^[14],并且影响肿瘤血管的生成^[15]。Boguszewski CL等^[16]的实验也证明了在一些动物模型和与耐药或生长激素缺乏相关的遗传缺陷患者中,生长激素对肿瘤发生具有一定的保护作用。为了深入研究HCC中三种miRNA信号的调控机制,本研究利用3个数据库对模型中三种miRNA的靶基因进行了评估。结果显示,靶基因的信号通路结果主要富集于细胞因子-细胞因子受体的相互作用途径以及生长激素的合成、分泌和作用途径中。

为了建立调控连接 HCC 的三种 miRNA 模型的关键节点,本研究共筛选出 10 个中枢基因以及 47 个与生存预后相关的基因;其中 DBT、SLC7A2、ALPL、TOP2A、MAD2L1、OIP5、KPNA2 不仅是 PPI 网络中的中枢基因,而且与患者的生存状态相关。除 DBT 基因缺乏相关研究外,SLC7A2 可能成为是乳腺癌的一个新的预后标志物,与强大的生存优势密切相关^[17];ALPL 的高表达与前列腺癌生存率较差相关^[18];KPNA2^[11]和 OIP5^[19]参与 HCC 的进展及其作用机制;TOP2A 可以影响肺腺癌细胞^[20];MAD2L1 在淋巴结转移的肺鳞癌患者表达较高,参与细胞周期的调节,可以作为非小细胞肺癌发展的预后生物标志物^[21],这 6 个基因均可以作为肿瘤的预后标志物。提示这 7 个基因可能成为 HCC 新的治疗靶点,为肝癌的治疗提供一种新的思路。

综上所述,本研究构建了能够预测 HCC 预后的 miRNA 模型,分组验证了该模型的预测能力,并且验证了该模型可以作为 HCC 中独立预后因素。最后,通过预测 miRNA 的遗传靶点进一步了解 HCC 的发生和进展。

参考文献:

- [1] Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017 [J]. CA Cancer J Clin, 2017, 67(1): 7-30.
- [2] Petrick JL, Kelly SP, Altekruse SF, et al. Future of Hepatocellular Carcinoma Incidence in the United States Forecast Through 2030 [J]. J Clin Oncol, 2016, 34(15): 1787-1794.
- [3] Zou H, Zhu CZ, Wang C, et al. Recurrence of Barcelona Clinic Liver Cancer Stage A Hepatocellular Carcinoma After Hepatectomy [J]. Am J Med Sci, 2017, 354(3): 262-267.
- [4] Ali Syeda Z, Langden SSS, Munkhzul C, et al. Regulatory Mechanism of MicroRNA Expression in Cancer [J]. Int J Mol Sci, 2020, 21(5): 1723.
- [5] Zhao XF, Li N, Lin DD, et al. Circulating MicroRNA-122 for the Diagnosis of Hepatocellular Carcinoma: A Meta-Analysis [J]. Biomed Res Int, 2020(2020): 5353695.
- [6] Liu G, Wang H, Fu JD, et al. A five-miRNA expression signature predicts survival in hepatocellular carcinoma [J]. APMIS, 2017, 125(7): 614-622.
- [7] Wang L, Ye G, Zhan C, et al. Clinical Factors Predictive of a Better Prognosis of Pulmonary Metastectomy for Hepatocellular Carcinoma [J]. Ann Thorac Surg, 2019, 108(6): 1685-1691.
- [8] Asafo-Agyei KO, Samant H. Hepatocellular Carcinoma [DB/OL]. <https://www.ncbi.nlm.nih.gov/books/NBK559177/>, 2021-08-08/2021-09-20.
- [9] Wang J, Wang B, Ren H, et al. miR-9-5p inhibits pancreatic cancer cell proliferation, invasion and glutamine metabolism by targeting GOT1 [J]. Biochem Biophys Res Commun, 2019, 509(1): 241-248.
- [10] 王亮, 姚博文, 陈天翔, 等. 微小 RNA-3682-3p 通过下调 GAS8 基因促进肝癌细胞迁移、侵袭的研究 [J]. 现代药物与临床, 2019, 34(7): 2172-2178.
- [11] Wu J, Zhang T, Chen Y, et al. MiR-139-5p influences hepatocellular carcinoma cell invasion and proliferation capacities via decreasing SLITRK4 expression [J]. Biosci Rep, 2020, 40(5): BSR20193295.
- [12] de Looft M, de Jong S, Kruijt FAE. Multiple Interactions Between Cancer Cells and the Tumor Microenvironment Modulate TRAIL Signaling: Implications for TRAIL Receptor Targeted Therapy [J]. Front Immunol, 2019(10): 1530.
- [13] Han KQ, Han H, He XQ, et al. Chemokine CXCL1 may serve as a potential molecular target for hepatocellular carcinoma [J]. Cancer Med, 2016, 5(10): 2861-2871.
- [14] Bacigalupo ML, Piazza VG, Cicconi NS, et al. Growth hormone upregulates the pro-tumorigenic galectin 1 in mouse liver [J]. Endocr Connect, 2019, 8(8): 1108-1117.
- [15] Subramani R, Nandy SB, Pedroza DA, et al. Role of Growth Hormone in Breast Cancer [J]. Endocrinology, 2017, 158(6): 1543-1555.
- [16] Boguszewski CL, Boguszewski MCDS. Growth Hormone's Links to Cancer [J]. Endocr Rev, 2019, 40(2): 558-574.
- [17] Rodriguez-Ruiz ME, Buqué A, Hensler M, et al. Apoptotic caspases inhibit abscopal responses to radiation and identify a new prognostic biomarker for breast cancer patients [J]. Oncoimmunology, 2019, 8(11): e1655964.
- [18] Rao SR, Snaith AE, Marino D, et al. Tumour-derived alkaline phosphatase regulates tumour growth, epithelial plasticity and disease-free survival in metastatic prostate cancer [J]. Br J Cancer, 2017, 116(2): 227-236.
- [19] Li Y, Xiao F, Li W, et al. Overexpression of Opa interacting protein 25 increases the progression of liver cancer via BMP2/JUN/CHEK1/RAC1 dysregulation [J]. Oncol Rep, 2019, 41(4): 2075-2088.
- [20] Kou F, Sun H, Wu L, et al. TOP2A Promotes Lung Adenocarcinoma Cells' Malignant Progression and Predicts Poor Prognosis in Lung Adenocarcinoma [J]. J Cancer, 2020, 11(9): 2496-2508.
- [21] Wei R, Wang Z, Zhang Y, et al. Bioinformatic analysis revealing mitotic spindle assembly regulated NDC80 and MAD2L1 as prognostic biomarkers in non-small cell lung cancer development [J]. BMC Med Genomics, 2020, 13(1): 112.

收稿日期: 2021-09-24; 修回日期: 2021-10-18

编辑/肖婷婷