



doi:10.7659/j.issn.1005-6947.2021.01.005
http://dx.doi.org/10.7659/j.issn.1005-6947.2021.01.005
Chinese Journal of General Surgery, 2021, 30(1):32-43.

· 专题研究 ·

肝细胞癌发生发展关键基因及其功能的生物信息学分析

李文波^{1,2}, 孙成杰^{1,2}, 周国俊^{1,2}, 应伟^{1,2}, 冯彦超^{1,2}, 黄婷^{1,2}, 侍琳^{1,2}, 黄理政^{1,2}, 李健水^{1,2}, 冷政伟^{1,2}

(川北医学院附属医院 1. 肝胆外科二 2. 肿瘤干细胞研究中心, 四川 南充 637000)

摘要

背景与目的: 肝细胞癌(HCC)是常见的原发性肝癌,其预后较差。基因的激活与失活可促进HCC的发生发展。本研究基于生物信息学HCC发生发展的关键基因及功能并进行临床样本表达验证。

方法: 从公共基因GEO数据库中筛选HCC及癌旁组织基因芯片,通过GEO2R在线工具及Venn图筛选出差异表达基因(DEGs),对筛选出来的DEGs用DAVID网站进行GO功能分析和KEGG通路富集分析,再用STRING网站及Cytoscape软件对DEGs进行蛋白-蛋白相互作用(PPI)网络分析并筛选出核心DEGs,将核心DEGs在Kaplan-Meier Plotter网站进行生存分析,筛选出与预后相关的DEGs,将与预后相关的DEGs用GEPIA网站进行表达量分析,得到在HCC和癌旁组织中的差异表达情况;将与预后相关且在HCC中高表达的DEGs经Metascape网站进行功能及富集通路分析,得到与HCC发生发展有关的关键基因,最后选取关键基因在HCC和癌旁组织标本中进行表达验证。

结果: 从GEO数据库下载的符合要求的3个基因芯片(GSE14520, GSE41804, GSE45267)中共筛选出78个DEGs。用DAVID网站进行GO功能分析、KEGG通路富集分析、STRING网站及Cytoscape软件分析后,筛选出17个核心DEGs(CDK1、ASPM、CENPF、RRM2、CCNB1、TOP2A、PTTG1、ECT2、CDKN3、CYP2B6、SLCO1B3、CYP1A2、CYP4A11、CYP26A1、CYP2E1、NAT2、CYP3A4),将17个核心DEGs在Kaplan-Meier Plotter网站进行生存分析后显示,有9个基因(CDK1、ASPM、CENPF、RRM2、CCNB1、TOP2A、PTTG1、ECT2、CDKN3)与HCC的预后相关(均 $P<0.05$)。GEPIA网站进行表达量分析显示,9个基因在HCC组织中均高表达(均 $P<0.05$)。Metascape网站分析显示,9个高表达基因要富集在细胞有丝分裂的负调控、细胞周期、核染色体隔离和雌配子的产生方面。选取CDK1在HCC组织和癌旁组织中进行验证,结果显示,CDK1在HCC组织中的表达量明显高于癌旁组织($P<0.05$)。

结论: 本研究得到的9个基因可能是HCC发生、发展的关键基因,可为HCC的发生机制的研究以及诊断、治疗提供参考。

关键词

癌,肝细胞;基因,肿瘤;计算生物学

中图分类号: R735.7

基金项目: 国家自然科学基金资助项目(81402444);四川省科技厅应用基础基金资助项目(2017JY0170);四川省南充市重大技术攻关资助项目(18SXHZ0460)。

收稿日期: 2020-03-19; **修订日期:** 2020-12-15。

作者简介: 李文波,川北医学院附属医院硕士研究生,主要从事普通外科临床与基础方面的研究。

通信作者: 李健水, Email: ljs2005doctor@126.com; 冷政伟, Email: lengzhengwei@163.com

Bioinformatics analysis of key genes and their functions in occurrence and development hepatocellular carcinoma

LI Wenbo^{1,2}, SUN Chengjie^{1,2}, ZHOU Guojun^{1,2}, YING Wei^{1,2}, FENG Yanchao^{1,2}, HUANG Ting^{1,2}, SHI Lin^{1,2}, HUANG Lizheng^{1,2}, LI Jianshui^{1,2}, LENG Zhengwei^{1,2}

(1. The Second Department of Hepatobiliary Surgery 2. Cancer Stem Cell Research Center, Affiliated Hospital of North Sichuan Medical College, Nanchong, Sichuan 637000, China)

Abstract

Background and Aims: Hepatocellular carcinoma (HCC) is a common primary liver cancer with a poor prognosis. The activation and inactivation of genes can promote the occurrence and development of HCC. This study was conducted to investigate the key genes and their functions in the occurrence and development of HCC based on bioinformatics and verify their expressions in clinical samples.

Methods: The HCC and paracancerous tissue gene chips were downloaded from the public gene GEO database, the differentially expressed genes (DEGs) were screened through GEO2R online tools and Venn diagrams, and the GO function analysis and KEGG pathway enrichment analysis were performed on the selected DEGs using the DAVID website, and then the core DEGs were picked up from the DEGs by protein-protein interaction (PPI) network analysis using STRING website and Cytoscape software. The prognosis-related DEGs were determined by survival analysis using Kaplan-Meier Plotter website, and the expression levels of the prognosis-related DEGs were analyzed using the GEPIA website to obtain the differential expressions between HCC and paracancerous tissue, and then the key genes associated with the occurrence and development of HCC were screened from the highly expressed prognosis-related DEGs in HCC by function and enrichment pathway analysis using Metascape website. Finally, the expression verification of select key genes was performed in the specimens of HCC and paracancerous tissue.

Results: A total of 78 DEGs were screened from the three eligible gene chips (GSE14520, GSE41804, GSE45267) downloaded from GEO. Then, 17 core DEGs (CDK1, ASPM, CENPF, RRM2, CCNB1, TOP2A, PTTG1, ECT2, CDKN3, CYP2B6, SLCO1B3, CYP1A2, CYP4A11, CYP26A1, CYP2E1, NAT2, CYP3A4) were screened out after using DAVID website GO function analysis and KEGG pathway enrichment analysis as well as STRING website and Cytoscape software analysis. After the survival analysis of the 17 core DEGs on Kaplan-Meier Plotter website, 9 genes (CDK1, ASPM, CENPF, RRM2, CCNB1, TOP2A, PTTG1, ECT2, CDKN3) were detected to be associated with the prognosis of HCC (all $P < 0.05$). The expression level analysis by GEPIA website showed that all the 9 genes were highly expressed in HCC tissue (all $P < 0.05$). Metascape website analysis showed that the 9 highly expressed genes were mainly enriched in the processes of negative regulation of mitotic cell cycle, nuclear chromosome segregation and female gamete generation. The CDK1 was selected to verify in HCC and paracancerous tissues, and the result showed that the CDK1 expression level was significantly higher than that in paracancerous tissue ($P < 0.05$).

Conclusion: The 9 genes obtained in this study may be the key genes in the occurrence and development of HCC, which provide a reference for the study of the pathogenetic mechanism as well as the diagnosis and treatment of HCC.

Key words

Carcinoma, Hepatocellular; Genes, Neoplasm; Computational Biology

CLC number: R735.7

肝细胞癌 (hepatocellular carcinoma, HCC) 是消化系统最常见的恶性肿瘤之一, 据美国癌症协会统计, HCC在男性癌症中的发病率排第六

位, 但在女性中仍以每年超过2%的速度增长, 其病死率在男性排第五位, 在女性中排第六位^[1-2], 与其他癌症相似, HCC与潜在的危险因素相关,

如超重、吸烟、酗酒、乙型肝炎病毒（HBV）、丙型肝炎病毒（HCV）、非酒精性脂肪性肝病和某些基因突变等^[3-4]。目前HCC的常用治疗方法包括手术切除、射频消融、肝脏移植、放射栓塞、免疫治疗等^[5-7]，不同分期的肿瘤患者选择不同的治疗方法，但5年复发率仍然很高^[8-9]。近年来越来越多的研究表明，基因失活和基因突变是导致HCC发生发展的重要因素，故准确找到导致HCC的相关基因，从基因水平上研究HCC已成为研究热点之一。

近年来利用生物信息学和基因芯片技术研究因基因失活或基因突变所致肿瘤的发生、发展已成为发展趋势^[10]。基因芯片具有数据全面，样本量大等优点，在生物学及医学领域占据重要位置。但这些数据信息混杂、缺乏足够实验基础，其准确性需要进一步验证。因此如何挖掘出准确可靠的基因信息逐渐成为生物信息学研究热点。本研究通过生物信息学的方法对HCC和癌旁组织的基因芯片中的差异基因进行分析并进行临床样本表达验证，筛选出与HCC发生发展的关键基因并进行临床样本验证，以期为其早期诊断、靶向治疗等提供参考依据。

1 资料与方法

1.1 芯片数据信息

HCC芯片从NCBI-GEO (<https://www.ncbi.nlm.nih.gov/geo>) 数据库下载，GSE14520，GSE41804，GSE45267分别包含HCC组织225、20、48例，癌旁组织220、20、39例。

1.2 差异表达基因 (differentially expressed genes, DEGs) 筛选

通过GEO2R在线工具，定义HCC和癌旁组织，比较出DEGs，再筛选出满足绝对值LogFC>2且校正 $P<0.05$ 的DEGs，将筛选出LogFC>2的基因，定义为上调DEGs，LogFC<-2的基因，定义为下调DEGs，通过Venn (<http://bioinformatics.psb.ugent.be/webtools/Venn>) 在线制图工具筛选出3个芯片的上调基因的交集与下调调基因的交集。

1.3 GO 功能分析和 KEGG 通路富集分析

利用生物学信息注释数据库 (Database

for Annotation, Visualization and Integrated Discovery, DAVID) 为DEGs进行系统综合的生物学功能注释分析。通过上传DEGs到DAVID网站 (<https://david.ncifcrf.gov>)，进行GO (gene ontology, GO) 功能分析和KEGG (Kyoto Encyclopedia of Genes and Genomes) 通路富集分析，以 $P<0.05$ 且FDR<0.05为差异有统计学意义，选择分析项目为生物过程 (biological process, BP)、分子功能 (molecular function, MF)、细胞成分 (cellular component, CC) 和信号通路。

1.4 蛋白相互作用网络 PPI 分析和核心 DEGs 筛选

将所获得的共同表达DEGs用STRING网站 (<https://string-db.org>) 进行PPI网络图的构建，构建后的网络图利用Cytoscape3.7.2软件及MCODE插件分析，筛选出核心DEGs。

1.5 核心 DEGs 生存分析

在Kaplan-Meier Plotter网站 (<http://www.kmplot.com>) 中，将核心DEGs逐个进行生存分析，选择总生存率 (OS) 为指标，作出每个基因的生存曲线图，筛选出生存分析中 $P<0.05$ 的基因，其可能为与预后相关基因。

1.6 预后相关 DEGs 在 HCC 与癌旁组织中的表达情况

把Kaplan-Meier Plotter网站中筛选出的与预后相关的基因用GEPIA (<http://gepia.cancer-pku.cn>) 在线分析，选择表达图形选择Box Plots，数据来源选择LIHC，得到其在HCC组织和癌旁组织中的表达情况，以 $P<0.05$ 为具有统计学意义。

1.7 与预后相关且在 HCC 中高表达基因的功能及通路富集分析

将获得的与预后相关且在HCC中高表达基因上传至Metascape网站 (<http://metascape.org/gp/index.html#/main/step1>)，种属选择homo sapiens，得到功能和通路的富集分析结果，以 $P<0.05$ 为具有统计学意义。

1.8 免疫组化染色分析

在本研究中心选取HCC组织和癌旁组织各70例用于将石蜡包埋的组织切成5 μm 的切片。使用标准的免疫过氧化物酶染色程序进行免疫组化分析。分析染色强度 (阴性: 0分, 弱阳性: 1分, 中等阳性: 2分, 强阳性: 3分)

和阳性细胞百分比 (<5%: 0分, 5%~25%: 1分, 26%~50%: 2分, 51%~75%: 3分, 76%~100%: 4分), 最终计算公式: 强度×百分比, 其最终范围为0~12分。用这种方法对每一张玻片进行打分, 以6分为界限, 分为低表达样本和高表达样本。然后使用SPSS分别对HCC组织和癌旁组织进行数据统计, 再把统计数据导入GraphPad prism 8中绘制统计图。

2 结果

2.1 DEGs 筛选结果

芯片GSE14520, GSE41804, GSE45267分别筛选出差DEGs 252、257、497个, 其中上调的DEGs分别为54、63、116个, 下调的DEGs分别为198、194、381个, 3个芯片的共差异的上调基因为16个(图1A), 共差异的下调基因为62个(图1B), 其共同的DEGs具体详见(表1)。

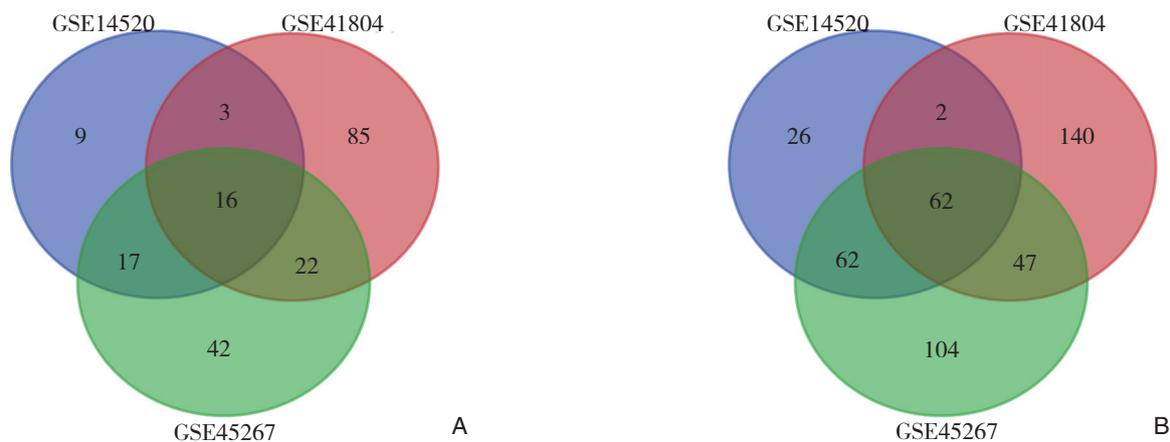


图1 三个芯片 DEGs 的 Venn 图 (上调基因 LogFc>2, 下调基因 LogFc<-2) A: 16 个上调 DEGs; B: 62 个下调 DEGs
 Figure 1 Venn diagram of DEGs in the three gene microarrays (up-regulated LogFc>2, down-regulated LogFc<-2) A: 16 up-regulated genes; B: 62 down-regulated genes

表1 78 个共同 DEGs
 Table 1 The 78 common DEGs

差异基因	基因名称
上调	CDK1 SPINK1 SFN CAP2 S100P RRM2 CCNB1 TOP2A ASPM CDKN3 AKR1B10 GPC3 SULT1C2 PTTG1 CENPF ECT2
下调	MT1G LECT2 CYP26A1 BBOX1 CYP2A6 CYP2C8 CXCL14 SLC22A1 CYP39A1 SPP2 HAO2 HPD MT1F ADH1B SLC10A1 SLC01B3 LY6E CYP1A2 INS-IGF2//IGF2 MT1E FCN3 GBA3 CYP2B6 CD5L LPA GHR CLEC1B FOSB CYP2E1 MT1H LCAT CLEC4M ESR1 VIPR1 HAMP LDOB DNASE1L3 DCN NAT2 NR1I2 RDH16 AKR1D1 CXCL12 CRHBP F9 CYP4A11 HGFAC MT1X MT2A C7 MT1M CYP3A4 GLYAT GLS2 SRD5A2 ECM1 APOF MT1HL1 C9 SRPX CDHR2 FCN2

2.2 GO 与 KEGG 分析结果

经过DAVID网站将差异表达的78个基因, 进行GO功能分析和KEGG通路富集分析。

GO功能分析中(表2), BP共涉及48个方面, 主要集中在: 细胞负增长的调控、细胞对锌离子的反应、外源性药物代谢过程、异型生物质的代谢过程、氧化还原过程、细胞对镉离子反应、P450表氧化酶通路、类固醇代谢过程; MF共有21个相关方面, 其中主要集中于氧化还原酶活性、铁离子结合、氧结合、血红素结合、单加氧酶活性、花生四烯酸环氧合酶活性、氧化还原酶

活性、类固醇羟化酶活性、咖啡因氧化酶活性、芳香酶活性; CC共涉及13个方面, 其中有统计学意义的有以下方面: 细胞器膜、细胞外泌体、细胞外区、内质网膜、细胞外间隙、胞质核周区、中间体、血液微粒、次膜攻击复合物。KEGG通路富集分析(表3)中, 共涉及15个通路, 其主要涉及于视黄醇的新陈代谢、矿物质吸收、药物代谢-细胞色素P450、化学致癌性、细胞色素P450对外源生物的代谢作用、咖啡因代谢、甾体类激素生物合成、亚油酸代谢、代谢途径、P53信号通路、药物代谢-其他酶。

表2 78个DEGs GO功能分析
Table 2 GO function analysis of the 78 DEGs

类别	功能	总数	%	P
GOTERM_BP_DIRECT	GO: 0045926~negative regulation of growth	8	0.097 752	1.28E-12
GOTERM_BP_DIRECT	GO: 0071294~cellular response to zinc ion	7	0.085 533	1.68E-10
GOTERM_BP_DIRECT	GO: 0042738~exogenous drug catabolic process	6	0.073 314	1.20E-09
GOTERM_BP_DIRECT	GO: 0006805~xenobiotic metabolic process	9	0.109 971	1.97E-09
GOTERM_BP_DIRECT	GO: 0071276~cellular response to cadmium ion	6	0.073 314	9.21E-09
GOTERM_BP_DIRECT	GO: 0019373~epoxygenase P450 pathway	6	0.073 314	1.27E-08
GOTERM_BP_DIRECT	GO: 0008202~steroid metabolic process	7	0.085 533	3.46E-08
GOTERM_BP_DIRECT	GO: 0055114~oxidation-reduction process	16	0.195 503	4.36E-08
GOTERM_BP_DIRECT	GO: 0017144~drug metabolic process	6	0.073 314	1.16E-07
GOTERM_BP_DIRECT	GO: 0008209~androgen metabolic process	4	0.048 876	4.60E-05
GOTERM_MF_DIRECT	GO: 0016705~oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	8	0.097 752	4.25E-09
GOTERM_MF_DIRECT	GO: 0005506~iron ion binding	10	0.122 19	1.70E-08
GOTERM_MF_DIRECT	GO: 0019825~oxygen binding	7	0.085 533	4.55E-08
GOTERM_MF_DIRECT	GO: 0020037~heme binding	9	0.109 971	1.18E-07
GOTERM_MF_DIRECT	GO: 0004497~monooxygenase activity	7	0.085 533	1.65E-07
GOTERM_MF_DIRECT	GO: 0008392~arachidonic acid epoxygenase activity	5	0.061 095	4.01E-07
GOTERM_MF_DIRECT	GO: 0016712~oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, reduced flavin or flavoprotein as one donor, and incorporation of one atom of oxygen	5	0.061 095	4.01E-07
GOTERM_MF_DIRECT	GO: 0008395~steroid hydroxylase activity	5	0.061 095	4.24E-06
GOTERM_MF_DIRECT	GO: 0034875~caffeine oxidase activity	3	0.036 657	1.07E-04
GOTERM_MF_DIRECT	GO: 0070330~aromatase activity	4	0.048 876	2.02E-04
GOTERM_CC_DIRECT	GO: 0031090~organelle membrane	10	0.122 19	8.19E-11
GOTERM_CC_DIRECT	GO: 0005576~extracellular region	19	0.232 16	7.06E-05
GOTERM_CC_DIRECT	GO: 0070062~extracellular exosome	24	0.293 255	6.58E-04
GOTERM_CC_DIRECT	GO: 0005789~endoplasmic reticulum membrane	12	0.146 628	7.63E-04
GOTERM_CC_DIRECT	GO: 0005615~extracellular space	15	0.183 284	0.001 068
GOTERM_CC_DIRECT	GO: 0048471~perinuclear region of cytoplasm	9	0.109 971	0.003 939
GOTERM_CC_DIRECT	GO: 0030496~midbody	4	0.048 876	0.016 16
GOTERM_CC_DIRECT	GO: 0072562~blood microparticle	4	0.048 876	0.024 818
GOTERM_CC_DIRECT	GO: 0005579~membrane attack complex	2	0.024 438	0.028 46

表3 78个DEGs的KEGG通路富集分析
Table 3 KEGG pathway enrichment analysis of the 78 DEGs

类别	功能	总数	%	P
KEGG_PATHWAY	hsa00830:Retinol metabolism	8	0.097 752	2.35E-07
KEGG_PATHWAY	hsa04978:Mineral absorption	7	0.085 533	4.83E-07
KEGG_PATHWAY	hsa00982:Drug metabolism-cytochrome P450	7	0.085 533	6.60E-06
KEGG_PATHWAY	hsa05204:Chemical carcinogenesis	7	0.085 533	1.70E-05
KEGG_PATHWAY	hsa00980:Metabolism of xenobiotics by cytochrome P450	6	0.073 314	1.50E-04
KEGG_PATHWAY	hsa00232:Caffeine metabolism	3	0.036 657	4.70E-04
KEGG_PATHWAY	hsa00140:Steroid hormone biosynthesis	5	0.061 095	6.72E-04
KEGG_PATHWAY	hsa00591:Linoleic acid metabolism	4	0.048 876	0.001 026
KEGG_PATHWAY	hsa01100:Metabolic pathways	18	0.219 941	0.002 522
KEGG_PATHWAY	hsa04115:p53 signaling pathway	4	0.048 876	0.011 176
KEGG_PATHWAY	hsa00983:Drug metabolism-other enzymes	3	0.036 657	0.040 62

2.3 PPI 分析结果

将78个DEGs基因上传至STRING网站分析后,有15个基因未出现在PPI网络分析中,剩余共有63个基因,其中包括上调基因17个,下调基因46个,共有蛋白之间相互作用关系线条156条(图2A)。

通过Cytoscape 3.7.2软件及MCODE插件分析后,得到两簇相交点最多的基因簇,分别有9、8个基因,分别有36、21条线,其分值分别为9、6,将其定义为核心DEGs。共获得22个核心DEGs(图2B)。

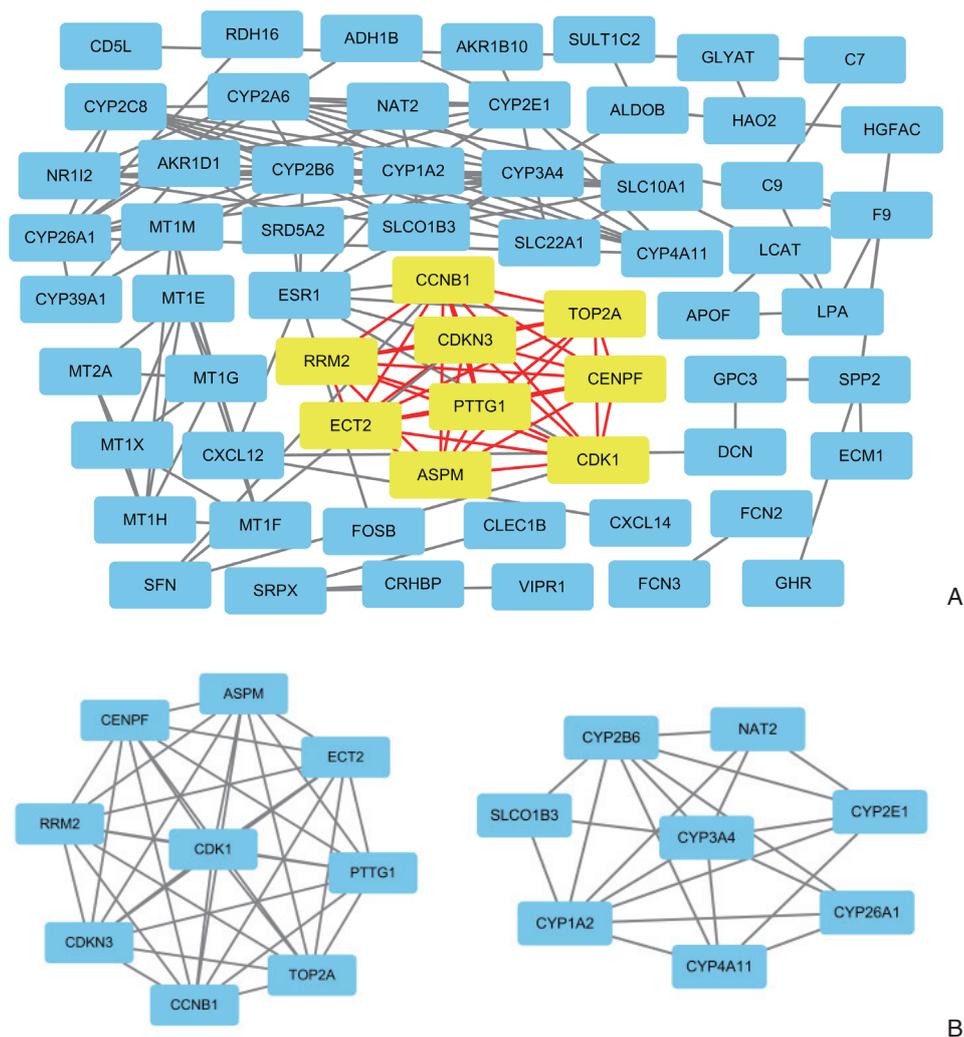


图 2 共同 DEGs 的 PPI 分析 (黄色表示上调基因, 蓝色表示下调基因) A: 63 个 DEGs 的 PPI 图; B: 22 个核心共同 DEGs

Figure 2 PPI analysis of the common DEGs (yellow color representing up-regulated genes, blue color representing down-regulated genes) A: PPI network of the 63 DEGs; B: The 22 common core DEGs

2.4 核心 DEGs 生存分析结果

17 个核心 DEGs 经过 Kaplan-Meier Plotter 网站生存分析后, 共有 9 个基因 (CDK1、ASPM、CENPF、RRM2、CCNB1、TOP2A、PTTG1、ECT2、CDKN3) 的生存分析差异有统计学意义 (均 $P < 0.05$), 另 8 个基因的生存分析差异无统计学意义 (均 $P > 0.05$) (图 3)。

2.5 预后相关 DEGs 表达量分析结果

将上述 9 个与预后相关的 DEGs 在 GEPIA 网站进行表达量分析后, 9 个基因均在 HCC 组织中较癌旁组织高表达 (均 $P < 0.05$) (图 4)。

2.6 高表达核心 DEGs 功能及通路富集分析结果

将上述 9 个在 HCC 组织中高表达的基因上传至 Metascape 网站分析, 得到功能和通路的富集分析主要集中在细胞有丝分裂的负调控、细胞周期、核染色体隔离和雌配子的产生等方面 (图 5) (表 4)。

2.7 关键 DEGs 验证结果

在筛选出来的 9 个基因中选取 CDK1 在 HCC 组织和癌旁组织中的染色, 结果显示, CDK1 在 HCC 组织中的评分为 (7.871 8 ± 1.524 9) 分, 在癌旁组织中的评分为 (3.410 3 ± 1.163 4) 分, 差异有统计学意义 ($t = 14.429, P < 0.000 1$) (图 6)。

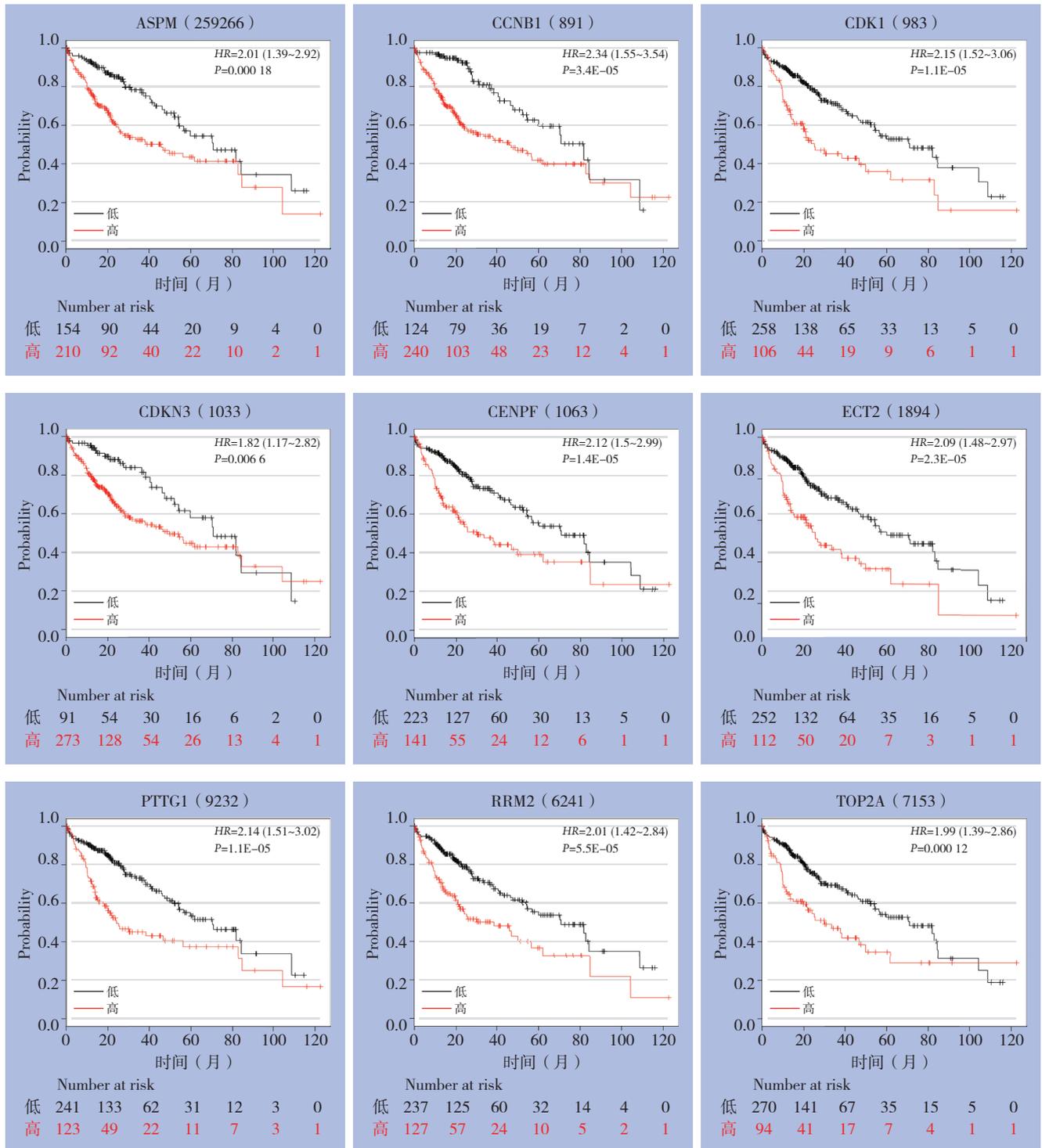


图 3 核心 DEGs 生存分析图

Figure 3 Survival curves of the core DEGs

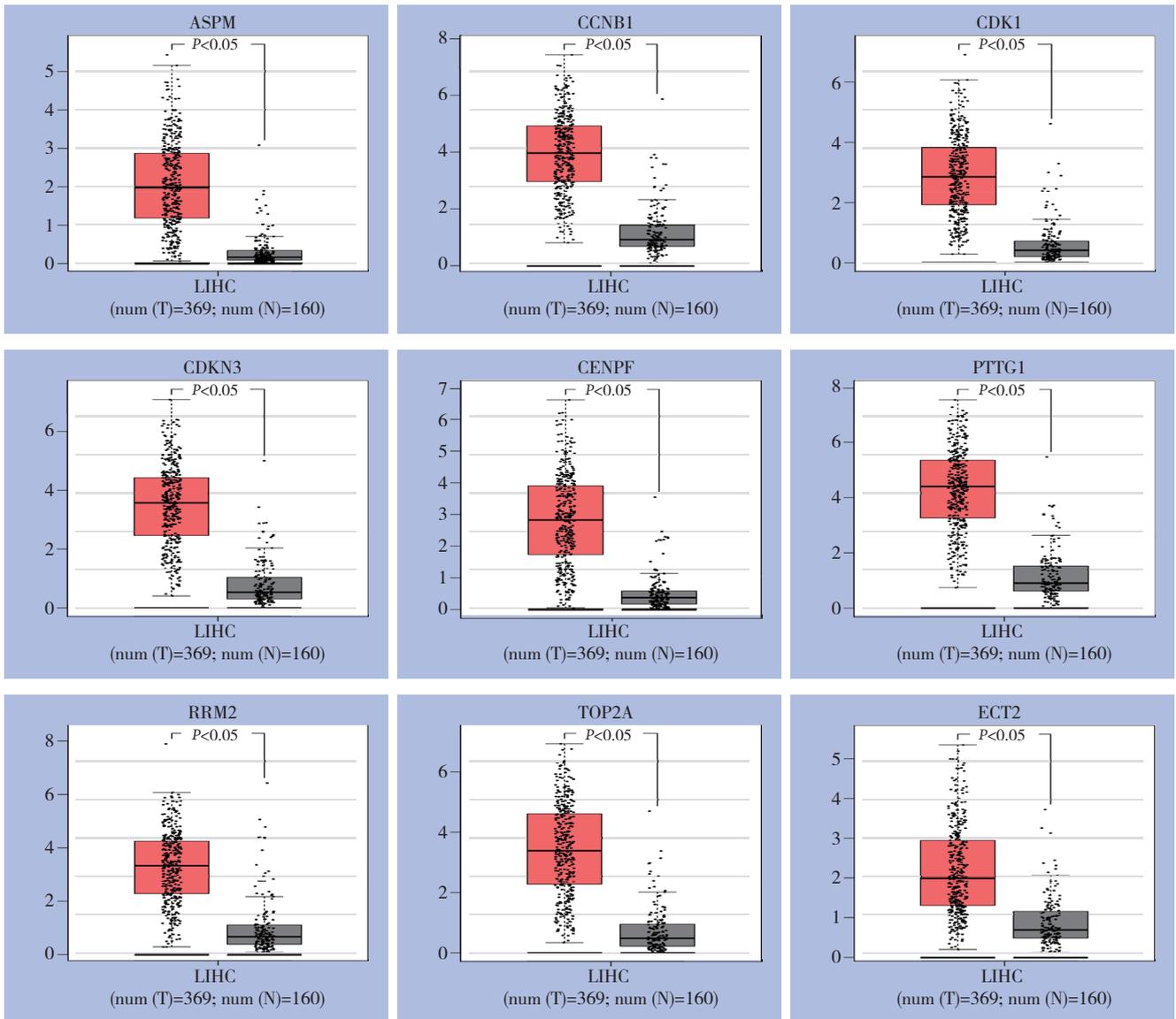


图 4 9 个 DEGs 在组织中表达盒形图
Figure 4 Box plots of expressions of the 9 DEGs

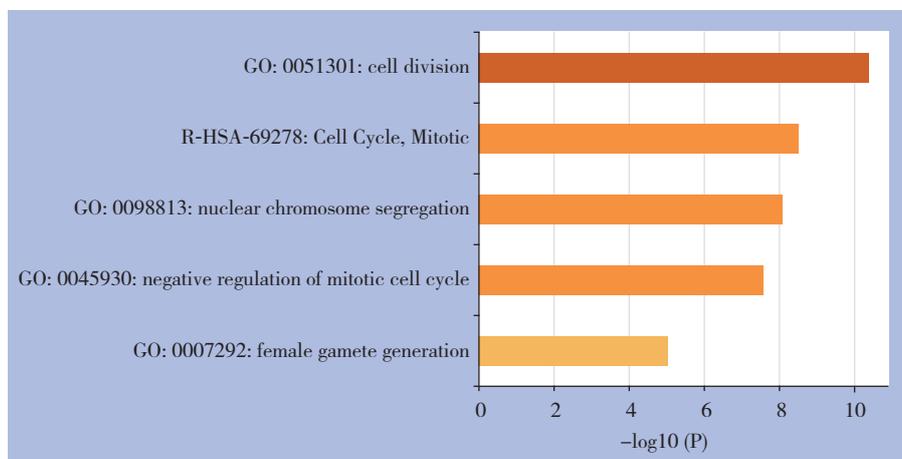


图 5 9 个核心 DEGs 功能和通路富集图
Figure 5 Function and pathway enrichment plots of the 9 core DEGs

表4 9个核心 DEGs 功能和通路富集分析数据

Table 4 Function and pathway enrichment data of the 9 core DEGs

项目	功能	总数	%	Log10 (P)
GO Biological Processes	cell division	7	87.50	-10.37
Reactome Gene Sets	cell cycle, mitotic	6	75.00	-8.49
GO Biological Processes	nuclear chromosome segregation	5	62.50	-8.08
GO Biological Processes	negative regulation of mitotic cell cycle	5	62.50	-7.55
GO Biological Processes	female gamete generation	3	37.50	-5.02



图6 免疫组化检测 CDK1 在 HCC 与癌旁组织的表达

Figure 6 Immunohistochemical staining for CDK1 expressions in HCC and adjacent tissue

3 讨论

HCC的发生发展通常是涉及基因、环境、饮食等多种因素的共同作用过程，利用生物信息学技术准确筛选出导致HCC发生发展的关键基因，这对于HCC的早期诊断、精准靶向治疗提供了重要依据。近年来伴随着生物信息学的蓬勃发展，大量基因芯片应用于研究疾病的发生发展和靶向基因的筛选等方面。例如Cao等^[11]利用3个GEO数据库进行生物信息学分析，确定参与IL-10信号转导的CXCL8、CXCL1和IL-1 β 是溃疡性结肠炎的前3个核心基因；Mo等^[12]利用生物信息学分析确定DNAJB4作为潜在的乳腺癌标记物；Xue等^[13]综合生物信息学分析确定了4个（CDC45、GINS2、MCM2和PCNA）可能与宫颈癌患者预后相关的关键基因，可作为宫颈癌潜在的预后生物标志物。本研究基于生物信息学的分析方法在GEO数据库中筛选出近几年的、样本量较大的3份HCC及癌旁组织基因芯片，并在多个生物信息分析网站中进行了系统全面的分析，最终得出CDK1（细胞周期蛋白依赖性激酶1）、ASPM（纺锤体微管组装因子）、RRM2（核糖核苷酸还原酶调节亚基M2）、TOP2A（DNA拓扑异构酶2A）、CENPF（着丝粒蛋白F）、CCNB1（细胞周期蛋白B1）、PTTG1（垂体肿瘤转化基因1）、ECT2（上皮细

胞转化序列2）、CDKN3（细胞周期蛋白依赖性激酶抑制因子3）与HCC发生、发展有重要关系，并最后在本研究中心选取临床样本进行表达验证。筛选出的9个基因主要作用于细胞有丝分裂的负调控、细胞周期、核染色体隔离和雌配子的产生等方面，从而引起细胞周期的紊乱、基因的突变，最终导致癌症的发生、发展。

CDK1属于丝氨酸/苏氨酸蛋白激酶家族，是调节细胞周期进程、DNA复制和分离、细胞成熟和增殖所必需的，CDK1的异常激活通过促进细胞增殖在肿瘤发生中起重要作用。CDK1在多种癌症中均有表达，在结直肠癌中，CDK1作为miR-769的直接靶点，在癌组织中高表达，miR-769通过直接作用CDK1来抑制肿瘤进展^[14]；CDK1的表达可被嗜酸乳杆菌CICC 6074 S层蛋白下调，阻止G₁细胞周期，从而发挥其对结肠癌细胞的细胞毒活性^[15]；CDK1可作为PKN蛋白磷酸化的对应激酶，促进肿瘤细胞贴壁依耐性生长和迁移，充当原癌基因的作用^[16]；在骨肉瘤中，Huang等^[17]研究证实，CDK1在骨肉瘤组织的细胞质中阳性表达，被确定是miR-199a-3p的潜在靶基因。在甲状腺癌中，研究者使用组织芯片证实CDK1蛋白在甲状腺癌（THCA）组织中的表达明显高于在非肿瘤组织中的表达，CDK1基因在THCA组织中共表达的KEGG分析表明：细胞周期、甲状腺激素合成、引

起自身免疫性甲状腺疾病等是CDK1在甲状腺癌中表达最丰富的途径^[18]。在乳腺癌中,由于选择性阻断CDK1单独或与其他治疗药物联合使用与有效的抗癌效果有关,因此CDK1可能被认为是乳腺癌治疗的靶点之一^[19]。在腺样囊性癌和非小细胞肺癌中的研究表明,CDK1的高表达与癌症患者的总体生存率较低相关,因此CDK1可作为诊断和预后的肿瘤标志物或药物治疗靶点之一^[20-21];在宫颈癌中,Luo等^[22]的研究揭示了CDK1在宫颈癌发展过程中对基因相互作用网络的综合作用,从而表明CDK1作为治疗靶点的潜在作用。此外,有研究^[23]表明,CDK1活跃于多种肿瘤调节细胞粘附的细胞周期,可作为多种癌症的临床预后生物标志物。

在HCC中,CDK1的异常表达可以调节凋亡素诱导的凋亡,在肿瘤进展中起着关键作用^[24]。CDK1的过度表达也被发现与HCC的门脉侵犯、甲胎蛋白水平高和预后不良直接相关^[25]。最近的一项研究发现二甲双胍可以通过诱导G₂/M期阻滞来显著抑制HCC细胞的增殖,并能有效地降低CDK1的表达^[26],提示CDK1可能参与了HCC细胞周期中的细胞增殖过程。另一项研究表明miR-582-5p通过直接抑制CDK1和Akt3的表达,间接抑制cyclin D1的表达来调控HCC的进展^[27],Wang等^[28]综合生物信息学分析发现,CDK1、CCNB1、CCNB2、MAD2L1和TOP2A等5个HUB基因可作为预测肝癌预后的生物标志物;Sun等^[29]应用生物信息学分析筛选发现CCNB1、CDK1、RRM2和BUB1B在肝癌组织中的过度表达与肝癌患者的不良生存相关,这些基因可能成为肝癌治疗的潜在靶点;He等^[30]通过生物信息学发现CDK1可能通过细胞周期和p53信号通路在肝硬化转化为HCC过程中发挥重要作用。Zou等^[31]发现CDK1、CCNB1和CCNB2是HCC潜在的预后生物标志物,并与HCC免疫细胞浸润有关。

在HCC中,通过3组芯片共369例HCC组织和160例癌旁组织的基因表达量分析发现,CDK1在HCC组织中的表达明显高于癌旁组织,后期的生存分析曲线显示CDK1高表达患者的生存时间较CDK1低表达患者明显减少,但上述实验数据均来源于生物信息学,由于其固有缺陷如平台与样本选择的不同可产生一定的差异,且数据信息混杂、缺乏足够实验基础,所以最后在本研究中

心选取70例HCC组织和癌旁组织进行表达验证,结果为CDK1在HCC组织中的评分为(7.871 8 ± 1.524 87)分,在癌旁组织中的评分为(3.410 3 ± 1.163 43)分,两组进行比较,差异具有统计学意义($t=14.429$, $P<0.000 1$),与预期结果相符。

综上所述,本研究基于生物信息学分析发现CDK1、ASPM、RRM2、TOP2A、CENPF、CCNB1、PTTG1、ECT2、CDKN3基因可能是HCC发生、发展的重要基因,且涉及细胞有丝分裂的负调控、细胞周期、核染色体隔离和雌配子的产生等方面,最后,选用CDK1在HCC组织和癌旁组织中进行验证,发现在HCC组织中CDK1的表达高于癌旁组织,与本研究预期结果相符。生物信息学在发现新靶点方面具有强大功能,但是需要分子生物学、细胞生物学、临床实验及疗效方面的研究来验证。这9个HCC相关基因均在HCC的发生、发展过程中具有巨大的作用,有望成为HCC筛查及治疗的新靶点,同时也将为研究HCC的发生、发展提供一定的理论基础。

参考文献

- [1] Siegel RL, Miller KD, Fuchs HE, et al. Cancer Statistics, 2021[J]. CA Cancer J Clin, 2021, 71(1):7-33. doi: 10.3322/caac.21654.
- [2] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020[J]. CA Cancer J Clin, 2020, 70(1):7-30. doi: 10.3322/caac.21590.
- [3] Islami F, Goding Sauer A, Miller KD, et al. Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States[J]. CA Cancer J Clin, 2018, 68(1):31-54. doi: 10.3322/caac.21440.
- [4] Degasperi E, Colombo M. Distinctive features of hepatocellular carcinoma in non-alcoholic fatty liver disease[J]. Lancet Gastroenterol Hepatol, 2016, 1(2):156-164. doi: 10.1016/S2468-1253(16)30018-8.
- [5] 张汉洋, 庄志彬, 林春冬, 等. 经皮射频消融术与腹腔镜肝切除术治疗原发性小肝癌的疗效及预后比较[J]. 中国普通外科杂志, 2019, 28(1):24-30. doi:10.7659/j.issn.1005-6947.2019.01.004. Zhang HY, Zhuang ZB, Lin CD, et al. Efficacy and prognosis comparison of percutaneous radiofrequency ablation and laparoscopic hepatectomy in treatment of small hepatocellular carcinoma[J]. Chinese Journal of General Surgery, 2019, 28(1):24-30. doi:10.7659/j.issn.1005-6947.2019.01.004.
- [6] 代柏树, 雷世雄, 杨振宇, 等. 肝癌合并门静脉癌栓患者术后肝动脉化疗栓塞术联合门静脉化疗的有效性及其安全性[J]. 中国

- 普通外科杂志, 2019, 28(2):188-194. doi:10.7659/j.issn.1005-6947.2019.02.009.
- Dai BS, Lei SX, Yang ZY, et al. Efficacy and safety of postoperative adjuvant transcatheter arterial chemoembolization plus portal vein chemotherapy for hepatocellular carcinoma patients associated with portal vein tumor thrombus[J]. Chinese Journal of General Surgery, 2019, 28(2):188-194. doi:10.7659/j.issn.1005-6947.2019.02.009.
- [7] 宁克, 殷香保. 肝细胞癌免疫治疗的研究进展[J]. 中国普通外科杂志, 2019, 28(2):234-240. doi:10.7659/j.issn.1005-6947.2019.02.015.
- Ning K, Yin XB. Research progress in immunotherapy of hepatocellular carcinoma[J]. Chinese Journal of General Surgery, 2019, 28(2):234-240. doi:10.7659/j.issn.1005-6947.2019.02.015.
- [8] Kulik L, El-Serag HB. Epidemiology and management of hepatocellular carcinoma[J]. Gastroenterology, 2019, 156(2):477-491. doi:10.1053/j.gastro.2018.08.065.
- [9] 乐琪, 朱同恩, 莫志远, 等. 原发性肝癌患者手术切除术后早期复发影响因素分析[J]. 中国普通外科杂志, 2019, 28(1):18-23. doi:10.7659/j.issn.1005-6947.2019.01.003.
- Le Q, Zhu TE, Mo ZY, et al. Analysis of factors for early recurrence of patients with hepatocellular carcinoma after resection[J]. Chinese Journal of General Surgery, 2019, 28(1):18-23. doi:10.7659/j.issn.1005-6947.2019.01.003.
- [10] Kristensen VN, Lingjærde OC, Russnes HG, et al. Principles and methods of integrative genomic analyses in cancer[J]. Nat Rev Cancer, 2014, 14(5):299-313. doi: 10.1038/nrc3721.
- [11] Cao F, Cheng YS, Yu L, et al. Bioinformatics Analysis of Differentially Expressed Genes and Protein-Protein Interaction Networks Associated with Functional Pathways in Ulcerative Colitis[J]. Med Sci Monit, 2021, 27:e927917. doi: 10.12659/MSM.927917.
- [12] Mo L, Liu J, Yang Z, et al. DNAJB4 identified as a potential breast cancer marker: evidence from bioinformatics analysis and basic experiments[J]. Gland Surg, 2020, 9(6):1955-1972. doi: 10.21037/gs-20-431.
- [13] Xue H, Sun Z, Wu W, et al. Identification of Hub Genes as Potential Prognostic Biomarkers in Cervical Cancer Using Comprehensive Bioinformatics Analysis and Validation Studies[J]. Cancer Manag Res, 2021, 13:117-131. doi: 10.2147/CMAR.S282989.
- [14] Wang L, Xu M, Lu P, et al. microRNA-769 is downregulated in colorectal cancer and inhibits cancer progression by directly targeting cyclin-dependent kinase 1[J]. Onco Targets Ther, 2018, 11:9013-9025. doi: 10.2147/OTT.S183847.
- [15] Zhang T, Pan D, Yang Y, et al. Lactobacillus acidophilus Effect of CICC 6074 S-Layer Protein on Colon Cancer HT-29 Cell Proliferation and Apoptosis[J]. J Agric Food Chem, 2020, 68(9):2639-2647. doi: 10.1021/acs.jafc.9b06909.
- [16] Zeng R, Wang Z, Li X, et al. Cyclin-dependent kinase 1-mediated phosphorylation of protein kinase N1 promotes anchorage-independent growth and migration[J]. Cell Signal, 2020, 69:109546. doi: 10.1016/j.cellsig.2020.109546.
- [17] Huang WT, Liu AG, Cai KT, et al. Exploration and validation of downregulated microRNA-199a-3p, downstream messenger RNA targets and transcriptional regulation in osteosarcoma[J]. Am J Transl Res, 2019, 11(12):7538-7554.
- [18] Zheng HP, Huang ZG, He RQ, et al. Integrated assessment of CDK1 upregulation in thyroid cancer[J]. Am J Transl Res, 2019, 11(12):7233-7254.
- [19] Izadi S, Nikkhoo A, Hojjat-Farsangi M, et al. CDK1 in Breast Cancer: Implications for Theranostic Potential[J]. Anticancer Agents Med Chem, 2020, 20(7):758-767. doi: 10.2174/1871520620666200203125712.
- [20] Liu HB, Huang GJ, Luo MS. Transcriptome analyses identify hub genes and potential mechanisms in adenoid cystic carcinoma[J]. Medicine (Baltimore), 2020, 99(2):e18676. doi: 10.1097/MD.00000000000018676.
- [21] Chen W, Zhu S, Zhang Y, et al. Identification of key candidate tumor biomarkers in non-small-cell lung cancer by in silico analysis[J]. Oncol Lett, 2020, 19(1):1008-1016. doi: 10.3892/ol.2019.11169.
- [22] Luo Y, Wu Y, Peng Y, et al. Systematic analysis to identify a key role of CDK1 in mediating gene interaction networks in cervical cancer development[J]. Ir J Med Sci, 2016, 185(1):231-239. doi: 10.1007/s11845-015-1283-8.
- [23] Jones MC, Askari JA, Humphries JD, et al. Cell adhesion is regulated by CDK1 during the cell cycle[J]. J Cell Biol, 2018, 217(9):3203-3218. doi: 10.1083/jcb.201802088.
- [24] Zhao J, Han SX, Ma JL, et al. The role of CDK1 in apoptin-induced apoptosis in hepatocellular carcinoma cells[J]. Oncol Rep, 2013, 30(1):253-259. doi: 10.3892/or.2013.2426.
- [25] Ito Y, Takeda T, Sakon M, et al. Expression and prognostic role of cyclin-dependent kinase 1 (cdc2) in hepatocellular carcinoma[J]. Oncology, 2000, 59(1):68-74. doi: 10.1159/000012140.
- [26] Zhou J, Han S, Qian W, et al. Metformin induces miR-378 to downregulate the CDK1, leading to suppression of cell proliferation in hepatocellular carcinoma[J]. Onco Targets Ther, 2018, 11:4451-4459. doi: 10.2147/OTT.S167614.
- [27] Zhang Y, Huang W, Ran Y, et al. miR-582-5p inhibits proliferation of hepatocellular carcinoma by targeting CDK1 and AKT3[J]. Tumour Biol, 2015, 36(11):8309-8316. doi: 10.1007/s13277-015-3582-0.
- [28] Wang M, Wang L, Wu S, et al. Identification of Key Genes and Prognostic Value Analysis in Hepatocellular Carcinoma by

- Integrated Bioinformatics Analysis[J]. *Int J Genomics*, 2019, 2019:3518378. doi: 10.1155/2019/3518378.
- [29] Sun Q, Liu P, Long B, et al. Screening of significant biomarkers with poor prognosis in hepatocellular carcinoma via bioinformatics analysis[J]. *Medicine (Baltimore)*, 2020, 99(32):e21702. doi: 10.1097/MD.00000000000021702.
- [30] He B, Yin J, Gong S, et al. Bioinformatics analysis of key genes and pathways for hepatocellular carcinoma transformed from cirrhosis[J]. *Medicine (Baltimore)*, 2017, 96(25):e6938. doi: 10.1097/MD.0000000000006938.
- [31] Zou Y, Ruan S, Jin L, et al. CDK1, CCNB1, and CCNB2 are Prognostic Biomarkers and Correlated with Immune Infiltration in Hepatocellular Carcinoma[J]. *Med Sci Monit*, 2020, 26:e925289. doi: 10.12659/MSM.925289.
- (本文编辑 姜晖)

本文引用格式: 李文波, 孙成杰, 周国俊, 等. 肝细胞癌发生发展关键基因及其功能的生物信息学分析[J]. 中国普通外科杂志, 2021, 30(1):32-43. doi:10.7659/j.issn.1005-6947.2021.01.005

Cite this article as: Li WB, Sun CJ, Zhou GJ, et al. Bioinformatics analysis of key genes and their functions in occurrence and development hepatocellular carcinoma[J]. *Chin J Gen Surg*, 2021, 30(1):32-43. doi:10.7659/j.issn.1005-6947.2021.01.005

本刊常用词汇英文缩写表

C- 反应蛋白	CRP	甲型肝炎病毒	HAV	心电图	ECG
Toll 样受体	TLRs	碱性成纤维细胞转化生长因子	bFGF	心脏监护病房	CCU
氨基末端激酶	JNK	聚合酶链反应	PCR	血管紧张素 II	AngII
白细胞	WBC	抗生物素蛋白 - 生物素酶复合物法	ABC 法	血管内皮生长因子	VEGF
白细胞介素	IL	辣根过氧化物酶	HRP	血管性血友病因子	vWF
半数抑制浓度	IC ₅₀	链霉抗生物素蛋白 - 生物素酶复合物法	SABC 法	血红蛋白	Hb
变异系数	CV	磷酸盐缓冲液	PBS	血肌酐	SCr
标记的链霉抗生物素蛋白 - 生物素法	SP 法	绿色荧光蛋白	GFP	血尿素氮	BUN
表皮生长因子	EGF	酶联免疫吸附测定	ELISA	血小板	PLT
丙氨酸氨基转移酶	ALT	美国食品药品监督管理局	FDA	血压	BP
丙二醛	MDA	脑电图	EEG	血氧饱和度	SO ₂
丙型肝炎病毒	HCV	内毒素 / 脂多糖	LPS	烟酰胺腺嘌呤二核苷酸	NADPH
超氧化物歧化酶	SOD	内皮型一氧化氮合酶	eNOS	严重急性呼吸综合征	SARS
磁共振成像	MRI	内生肌酐清除率	CCr	一氧化氮	NO
极低密度脂蛋白胆固醇	VLDL-C	尿素氮	BUN	一氧化氮合酶	NOS
低密度脂蛋白胆固醇	LDL-C	凝血酶时间	TT	乙二胺四乙酸	EDTA
动脉血二氧化碳分压	PaCO ₂	凝血酶原时间	PT	乙酰胆碱	ACh
动脉血氧分压	PaO ₂	牛血清白蛋白	BSA	乙型肝炎病毒	HBV
二甲亚砜	DMSO	热休克蛋白	HSP	乙型肝炎病毒 e 抗体	HBeAb
反转录 - 聚合酶链反应	RT-PCR	人类免疫缺陷病毒	HIV	乙型肝炎病毒 e 抗原	HBeAg
辅助性 T 细胞	Th	人绒毛膜促性腺激素	HCG	乙型肝炎病毒表面抗体	HBsAb
肝细胞生长因子	HGF	三磷酸腺苷	ATP	乙型肝炎病毒表面抗原	HBsAg
干扰素	IFN	三酰甘油	TG	乙型肝炎病毒核心抗体	HBeAb
高密度脂蛋白胆固醇	HDL-C	生理氯化钠溶液	NS	乙型肝炎病毒核心抗原	HBeAg
谷胱甘肽	GSH	世界卫生组织	WHO	异硫氰酸荧光素	FLTC
固相 pH 梯度	IPG	双蒸水	ddH ₂ O	诱导型一氧化氮合酶	iNOS
核糖核酸	RNA	丝裂原活化蛋白激酶	MAPK	原位末端标记法	TUNEL
核因子 - κ B	NF-κ B	四甲基偶氮唑盐微量酶反应	MTT	杂合性缺失	LOH
红细胞	RBC	苏木精 - 伊红染色	HE	增强化学发光法	ECL
红细胞沉降率	ESR	胎牛血清	FBS	肿瘤坏死因子	TNF
环氧酶 -2	COX-2	体质量指数	BMI	重症监护病房	ICU
活化部分凝血活酶时间	APTT	天门冬氨酸氨基转移酶	AST	转化生长因子	TGF
活性氧	ROS	脱氧核糖核酸	DNA	自然杀伤细胞	NK 细胞
获得性免疫缺陷综合征	AIDS	细胞间黏附分子	ICAM	直接胆红素	DBIL
肌酐	Cr	细胞外基质	ECM	总胆固醇	TC
基质金属蛋白酶	MMP	细胞外调节蛋白激酶	ERK	总胆红素	TBIL
计算机 X 线断层照相技术	CT	纤连蛋白	FN		