

基于关键致病基因筛选的肝癌临床预后及中药靶向治疗的生物信息学研究

吴佳^{1,2}, 吴嫚婷^{1,3}, 龙荣^{1,2},
仇婧玥^{1,3}, 喻昶^{1,3}, 熊萌^{1,3}, 曾梅艳^{3*}, 宋厚盼^{1,3*}

¹湖南中医药大学中医诊断学湖南省重点实验室; ²湖南中医药大学医学院; ³湖南中医药大学中医学学院, 长沙 410208

摘要:通过生物信息学方法筛选肝癌关键致病基因,探究其表达水平、致病机制、免疫浸润、突变及预后情况,并进一步挖掘潜在的可用于肝癌靶向治疗的中药。通过数据集 GSE19665 分析正常和肝癌组织的差异表达基因(differentially expressed genes, DEGs);对 DEGs 进行 GO 和 KEGG 分析,构建 DEGs 蛋白互作网络并筛选肝癌发病关键基因;对关键基因进行突变、免疫浸润、表达及预后分析;挖掘潜在的治疗肝癌的中药和天然活性成分。筛选得到显著性 DEGs 159 个,主要涉及免疫、炎症反应,参与 p53、mTOR 等信号通路;筛选出 10 个关键基因, CDC20 等四个基因在肝癌组织的表达显著升高,高表达组患者的总生存期明显降低;关键基因突变患者的总生存期和无病生存期均明显降低;关键基因表达与免疫细胞浸润密切相关;以关键基因为靶点筛选得到 36 种中药和 105 种中药活性成分。本研究鉴定出了肝癌关键致病基因,其异常表达与肝癌的不良预后、免疫细胞浸润密切相关,挖掘出潜在的治疗肝癌的中药及其活性成分,研究结果可为肝癌的诊断、治疗及预后提供科学依据。

关键词:肝癌;关键基因;生物信息学;临床预后;免疫细胞浸润;中药

中图分类号:R285

文献标识码:A

文章编号:1001-6880(2022)8-1415-15

DOI:10.16333/j.1001-6880.2022.8.016

Bioinformatics study on clinical prognosis of liver cancer and targeted therapy of traditional Chinese medicine based on screening of core pathogenic genes

WU Jia^{1,2}, WU Man-ting^{1,3}, LONG Rong^{1,2},
QIU Jing-yue^{1,3}, YU Chang^{1,3}, XIONG Meng^{1,3}, ZENG Mei-yan^{3*}, SONG Hou-pan^{1,3*}

¹Hunan Provincial Key Laboratory of Diagnostic Research in Chinese Medicine, Hunan University of Chinese Medicine;

²Faculty of Medicine, Hunan University of Chinese Medicine; ³Faculty of Traditional Chinese Medicine, Hunan University of Chinese Medicine, Changsha 410208, China

Abstract: The key pathogenic genes of liver cancer were screened by bioinformatics methods to explore their expression level, pathogenesis, immune infiltration, mutation and prognosis, and to further explore the potential traditional Chinese medicine for targeted treatment of liver cancer. The differentially expressed genes (DEGs) in normal and hepatoma tissues were analyzed by data set GSE19665; Go and KEGG analysis were performed on DEGs to construct the interaction network of DEGs protein and screen the key genes of liver cancer; Mutation, immune infiltration, expression and prognosis of key genes were analyzed; Excavate potential traditional Chinese medicine and natural active ingredients for the treatment of liver cancer. 159 significant DEGs were screened, which were mainly involved in immune and inflammatory responses and involved in p53, mTOR and other signal pathways; Ten key genes were screened out. The expression of four genes such as CDC20 in HCC was significantly increased, and the overall survival time of patients in the high expression group was significantly reduced; The overall survival and disease-free survival of patients with key gene mutations were significantly reduced; The expression of key genes was closely related to the infiltration of immune cells; 36 kinds of traditional Chinese medicine and 105 kinds of active ingredients

收稿日期:2022-04-24 接受日期:2022-07-01

基金项目:国家自然科学基金(81703920, 82004427); 湖湘青年科技创新人才项目(2021RC3101); 湖南省教育厅科学研究重点项目(21A0240)

*通信作者 E-mail: zengmy2018@126.com, hpsong2015@126.com

were screened by the key gene target. In this study, the key pathogenic genes of liver cancer were identified, and their abnormal expression was closely related to the poor prognosis and immune cell infiltration of liver cancer. The potential traditional Chinese medicine and its active ingredients for the treatment of liver cancer were excavated. The research results can provide a scientific basis for the diagnosis, treatment and prognosis of liver cancer.

Key words: liver cancer; core gene; bioinformatics; clinical prognosis; immune cell infiltration; traditional Chinese medicine

原发性肝癌是全球癌症致死的第三大原因^[1], 肝细胞癌 (hepatocellular carcinoma, HCC) 是原发性肝癌最主要的类型, 其危险因素主要有乙型肝炎病毒 (HBV)/丙型肝炎病毒 (HCV) 感染、长期过度饮酒、食用过多黄曲霉菌污染的食品、多种原因导致的肝硬化及肝细胞癌家族史^[2]。肝细胞癌的早期诊断常借助于生物标志物, 在临床上被广泛运用的生物标志物主要来源于血清、血浆、组织的 microRNA、突变的基因、蛋白等^[3]。很多确诊为肝癌的患者通常处于晚期不可切除阶段, 采用姑息治疗后, 中位生存期仅为 6~12 个月, 5 年存活率仅为 10%^[4]。因此, 筛选肝癌关键致病基因对肝癌患者进行早期诊断及分析肝癌预后情况具有重要的临床意义。

高通量测序技术的迅速发展改变了生物医学的研究模式^[5]。基因测序为肿瘤发病机制等肿瘤生物学问题提供了全新的认识, 并且对肿瘤的诊断、预后和治疗的选择具有重要的参考价值^[6]。生物信息学算法对于处理高通量组学数据至关重要^[7]。本研究以高通量基因芯片数据挖掘为切入点, 引入生物信息学算法知识, 从 GEO 数据库获取肝癌相关数据集, 分析肝癌组织与正常肝组织的差异表达基因。进一步从差异表达基因中筛选出肝癌关键致病基因, 探讨关键基因的生物学功能及其涉及的信号通路、突变情况、免疫浸润和对临床预后的影响, 并进一步筛选潜在的治疗肝癌的中药, 旨在为肝癌的临床诊断、治疗及预后判断提供理论参考和科学依据。

1 材料与方法

1.1 数据获取与处理

使用美国国家生物技术信息中心 (National Center for Biotechnology Information, NCBI) 基因表达数据库 (gene expression omnibus database, GEO) (<https://www.ncbi.nlm.nih.gov/gds/>), 检索 HCC 相关数据集, 下载得到 GSE19665 mRNA 表达芯片数据^[8]。该数据集共有样本 20 例, 其中正常样本 10 例, 分别为 GSM490987、GSM490989、GSM490991、GSM490993、GSM490995、GSM490997、GSM490999、GSM491001、GSM491003、GSM491005; 肝癌样本 10

例, 分别为 GSM490988、GSM490990、GSM490992、GSM490994、GSM490996、GSM490998、GSM491000、GSM491002、GSM491004、GSM491006。20 例样本均由 GPL570 平台检测提交, 运用 R 软件中的 affy 程序包对下载的数据进行背景矫正、均一化处理。

1.2 肝癌与正常组织差异表达基因筛选

使用在线分析工具 GEO 2R 设置分组, 即正常样本组和肝癌样本组, 对基因表达谱芯片 GSE19665 进行分析, 保存分析结果。导出数据至 Excel 中进行两次筛选, 初次筛选条件为 $-\log_{10}P > 1.4$, $|\log_2FC| \geq 1$, 绘制火山图; 再次筛选条件为 $P < 0.01$, $|\log_2FC| \geq 2$, 绘制层次聚类热图, 得到肝癌和正常组织显著性 DEGs。

1.3 肝癌与正常组织 DEGs GO 富集分析和 GSEA 分析

使用富集分析工具 Metascape (<https://metascape.org/>), 选择物种为“homo sapiens”, 对筛选得到的显著性差异表达的 mRNA 进行基因本体论 (gene ontology, GO) 功能富集分析; 使用 OmicShare 平台 GSEA 动态工具对差异表达基因进行 KEGG 通路富集分析, 设置筛选条件为 $|NES| \geq 1$, $P < 0.05$, 依次对富集分析后的数据进行可视化处理。

1.4 肝癌与正常组织显著性 DEGs 蛋白互作网络构建与关键基因筛选

在蛋白数据库 STRING (the searcher tool for the retrieval of interacting genes) (<http://string-db.org/>) 检索框中输入筛选得到的显著性差异表达基因, 将最低要求相互作用分设置为中等置信度 (0.04), 构建蛋白质相互作用网络 (protein-protein interaction, PPI), 导出 tsv 格式数据文件。通过 Cytoscape 对导出的数据进行可视化处理, 借助 Cytohubba 插件中的 MCC、MNC、DMNC 算法, 分析出网络中度值 (degree) 排名前 10 位的差异表达基因, 即为肝癌致病的关键基因。

1.5 肝癌致病关键基因生存分析

利用生存分析数据库 Kaplan-Meier plotter (<http://kmplot.com/analysis/>) 中肝癌相关的 mRNA 芯

片数据,将筛选后的关键基因依次导入数据库。根据肝癌组织中关键基因表达的中位数进行分组,选择“自动选择最佳截止”(auto select best cutoff)分析关键基因表达的高低对肝癌患者总生存率(overall survival, OS)的影响,并以 $P < 0.05$ 为筛选条件,绘制 Kaplan-Meier 生存曲线。

1.6 肝癌关键致病基因表达验证

使用癌症组学数据库 UALCAN (<http://ualcan.path.uab.edu/>) 中 TCGA (the cancer genome atlas) 分析模块,输入关键致病基因,TCGA dataset 设置为 liver hepatocellular carcinoma,选择 Expression 表达分析模块,进一步分析关键基因在不同的 (sample types) 组织样本和 (individual cancer stages) 肿瘤分期中的表达情况。运用人类蛋白图谱数据库 (Human Protein Atlas, HPA) (<http://www.proteinatlas.org>),输入关键基因,分别选择 Tissue (组织) 和 Pathology (病理),在 Tissue 模块中选择 Liver,在 Pathology 模块中选择 Liver cancer,根据 Staining (染色强度)、Intensity (染色密度)、Quantity (定量)、Location (定位) 分析关键基因的蛋白质表达情况。

1.7 肝癌关键致病基因的突变及其相关性分析

使用多维癌症基因组数据平台 cBioPortal (<http://www.cbioportal.org>),选择癌症类型为 liver hepatocellular carcinoma,数据类型为 TCGA, Firehose Legacy,输入关键基因,设置样本类型为 Samples with mRNA data,采用 Z-评分法,分析肝癌关键致病基因的突变与预后情况。运用 R 软件读取标准化后的关键基因数据,通过 corplot 程序包对关键基因进行相关性分析,并将分析结果进一步可视化。

1.8 肝癌关键致病基因免疫细胞浸润分析

利用癌症免疫浸润数据库 TIMER ([trome.dfci.harvard.edu/TIMER\) 中的 Gene \(基因\) 板块,在 Gene Symbol 中输入关键基因, Cancer Types 设置为 LIHC \(liver hepatocellular carcinoma\)。分析肝癌组织中关键致病基因的表达与 6 种免疫细胞浸润程度的关系,具体包括 B 细胞、 \$CD4^+\$ T 细胞、 \$CD8^+\$ T 细胞、中性粒细胞、巨噬细胞、树突状细胞。](http://cis-</p>
</div>
<div data-bbox=)

1.9 治疗肝癌药物的筛选方法

Coremine Medical (<https://coremine.com/medical/>) 数据库是一个开放的生物医学数据分析平台,记录了大量生物学术语间的关系。在 Coremine Medical 数据库中导入肝癌关键致病基因,下载 traditional Chinese medicine 模块中的数据,设置筛选条件 $P < 0.05$,筛选可用于肝癌治疗的中药。比较毒物基因学数据库 (comparative toxicogenomics database, CTD) (<http://ctdbase.org/>) 综合整理了来自各个物种的毒理学数据,具有分析化学-基因/蛋白相互作用、化学-疾病、基因-疾病的功能。在 CTD 中导入核心基因,筛选具有潜在治疗 HCC 作用的中药活性成分,通过 cytoscape 软件绘制潜在治疗中药-核心基因-中药活性成分相互作用的网络图。

2 结果

2.1 芯片数据处理

通过 R 语言 affy 包对 GSE19665 数据集进行均一化处理,并绘制小提琴图,结果如图 1 所示。小提琴图中的横坐标表示芯片数据中的样本编号,包括 10 例正常样本和 10 例肝癌样本,纵坐标表示芯片数据中样本的表达值。小提琴图中央白色条形的范围即下四分位点到上四分位点,中心的实点表示中位数值,可见各样本中位数基本位于同一水平 (1.42 左右),说明该芯片数据的样本结果可靠,可用于下一步分析。

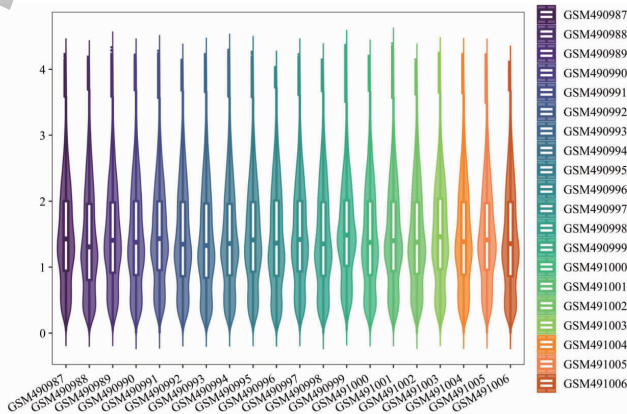


图 1 各样本标准化处理后小提琴图

Fig. 1 Violin diagram of each sample after standardized treatment

2.2 肝癌致病基因筛选

利用 GEO 2R 在线分析工具对 GSE19665 数据集进行 DEGs 分析,共纳入 54 638 个基因,初次筛选后得到差异表达基因 4 000 个,绘制火山图如图 2 所示,其中与肝癌发病相关的上调基因 874 个,下调

基因 3 126 个。通过比较 $|\log_2FC|$ 值进行再次筛选,分别选取上调基因中排名前 100 的基因、下调基因中排名前 100 的基因,剔除重复值后,绘制聚类热图。图 3 结果显示,其中与肝癌发病相关的显著性上调基因 81 个,显著性下调基因 80 个。

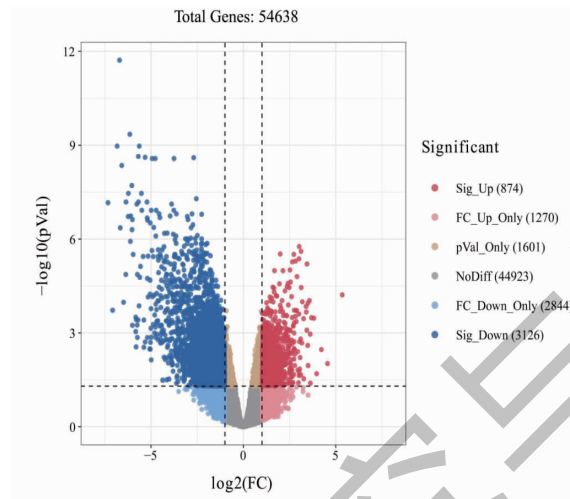


图 2 肝癌与正常肝组织差异表达基因分布火山图

Fig. 2 Volcanic map of differentially expressed genes in hepatocellular carcinoma and normal liver tissue

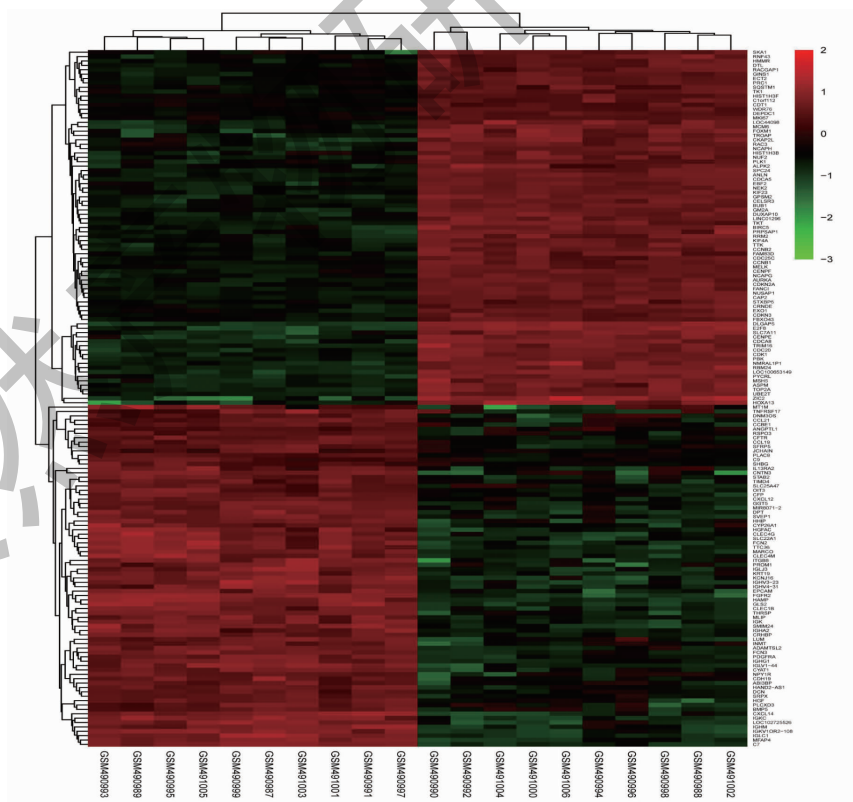


图 3 肝癌与正常肝组织显著性差异表达基因聚类热图

Fig. 3 Clustering heatmap of significantly differentially expressed genes between hepatocellular carcinoma and normal liver tissue

2.3 肝癌与正常肝组织显著性 DEGs GO 功能富集分析

表1结果显示,肝癌与正常肝组织显著性 DEGs 生物学过程(biological process, BP)主要涉及免疫反应、炎症反应、细胞粘附、细胞趋化作用、适应性免疫

应答等;细胞组分(cell component, CC)主要含细胞外区域、细胞外间隙、蛋白质性细胞外基质、质膜组成部分、细胞外基质等;分子功能(molecular function, MF)主要涉及肝素结合、受体活性、丝氨酸型内肽酶活性、糖结合、血红素结合等。

表1 肝癌与正常肝组织显著性 DEGs GO 功能富集分析

Table 1 GO function enrichment analysis of significantly differentially expressed genes between hepatocellular carcinoma and normal liver tissue

项目 Category	ID	条目 Term	基因数 Number of genes	P 值 P value
BP	GO:0006955	免疫反应 Immune response	69	5.56E-18
BP	GO:0006954	炎症反应 Inflammatory response	55	3.73E-12
BP	GO:0007155	细胞粘附 Cell adhesion	58	2.42E-10
BP	GO:0060326	细胞趋化作用 Cell chemotaxis	19	1.87E-09
BP	GO:0002250	适应性免疫应答 Adaptive immune response	25	3.52E-07
CC	GO:0005576	细胞外区域 Extracellular region	164	6.27E-33
CC	GO:0005615	细胞外空间 Extracellular space	148	8.14E-21
CC	GO:0005578	蛋白质性细胞外基质 Proteinaceous extracellular matrix	50	1.20E-15
CC	GO:0005887	质膜组成部分 Integral component of plasma membrane	138	7.64E-15
CC	GO:0031012	细胞外基质 Extracellular matrix	46	1.52E-11
MF	GO:0008201	肝素结合 Heparin binding	36	3.33E-14
MF	GO:0004872	受体活性 Receptor activity	31	2.05E-07
MF	GO:0004252	丝氨酸型内肽酶活性 Serine-type endopeptidase activity	33	7.61E-07
MF	GO:0030246	糖结合 Carbohydrate binding	28	8.89E-07
MF	GO:0020037	血红素结合 Heme binding	22	2.58E-06

2.4 基于 GSEA 的肝癌与正常肝组织显著性 DEGs KEGG 富集分析

图4结果显示,GSEA分析筛选出肝癌与正常肝组织显著性 DEGs 相关的 KEGG 通路92条,根据 P 值大小排序,最具统计学意义的信号通路主要涉及细胞因子-细胞因子受体相互作用通路、p53 信号通路、细胞色素 P450 相关通路、mTOR 信号通路、ErbB 信号通路、肿瘤相关通路、JAK-STAT 信号通路、MAPK 信号通路、Wnt 信号通路、Toll 样受体信号通路等。提示这些信号通路异常表达与肝癌的发病密切相关。

2.5 肝癌与正常肝组织显著性 DEGs PPI 网络构建及关键基因筛选

将肝癌与正常肝组织显著性 DEGs 导入 STRING 数据库,得到由 111 个蛋白节点和 1 284 条蛋白相互作用边所构成的网络,如图 5A 肝癌与正

常肝组织显著性 DEGs PPI 网络所示。借助 cytohubba 插件对该网络进行筛选,选择前度值排名前 10 的基因,结果如图 5B 肝癌致病关键基因所示。10 个关键基因分别为细胞分裂周期蛋白 20 (cell division cycle 20, CDC20)、胞周期蛋白 B2 (cyclin B2, CCNB2)、细胞周期蛋白 B1 (cyclin B1, CCNB1)、有丝分裂关卡基因 (budding uninhibited by benzimidazoles 1, BUB1)、杆状病毒凋亡抑制蛋白 5 (baculoviral inhibitor of apoptosis repeat-containing protein 5, BIRC5)、极光激酶 A (Aurora-A kinase, AURKA)、拓扑异构酶 II α (topoisomerase II α , TOP2A)、染色体非结构维持凝聚素 I 复合亚单位 H (Non-structural maintenance of chromosomes condensin I complex subunit H, NCAPH)、Discs 大同源物关联蛋白 5 (Discs large-associated protein 5, DLGAP5)、细胞周期蛋白激酶 1 (cyclin-dependent kinase 1, CDK1)。

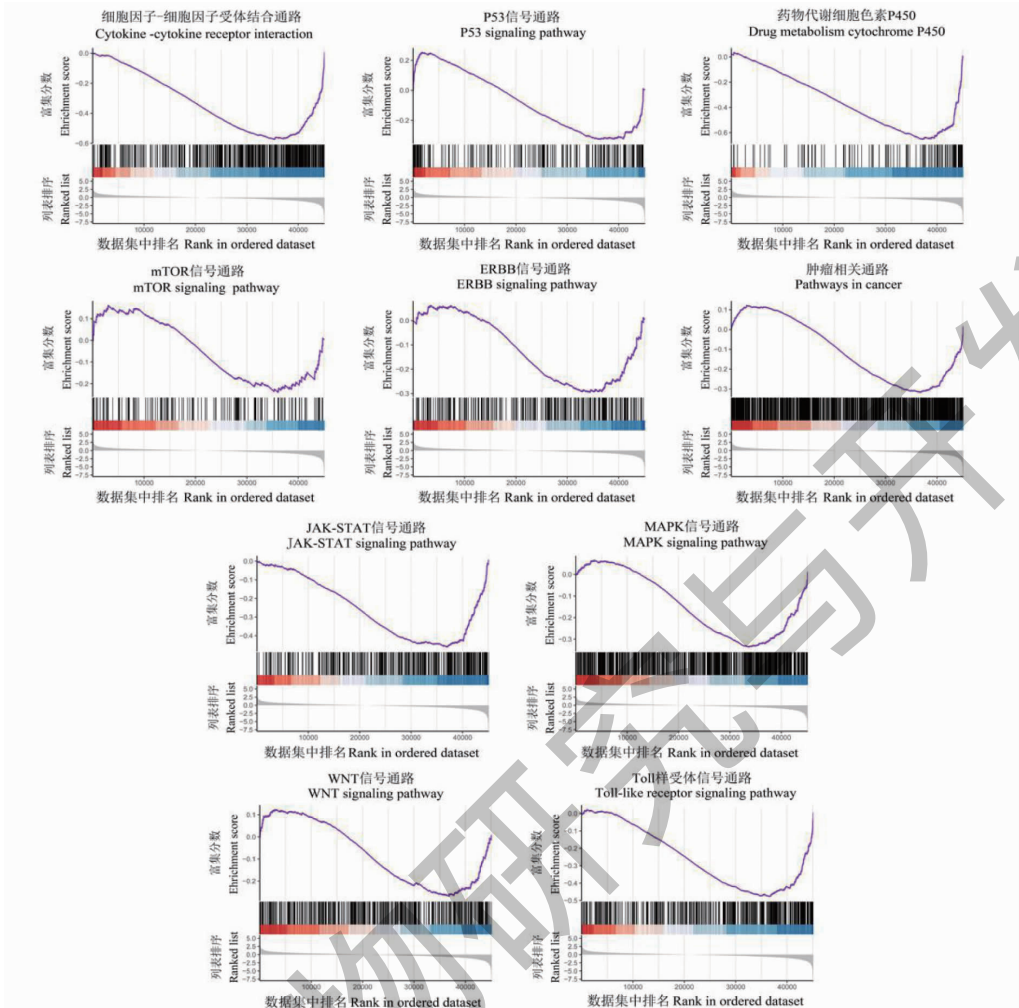


图4 基于GSEA的肝癌与正常肝组织显著性DEGs KEGG富集分析

Fig. 4 Signal pathway enrichment analysis of significantly differentially expressed genes in hepatocellular carcinoma and normal liver tissues based on GSEA

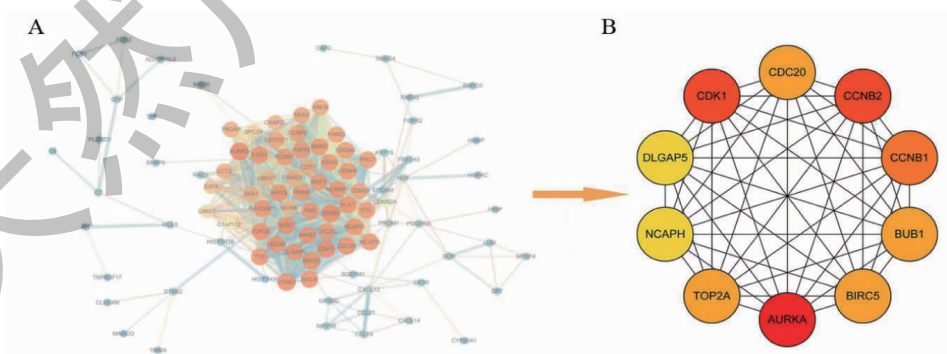


图5 肝癌致病关键基因筛选示意图。

Fig. 5 Schematic diagram of screening of pathogenic core genes of liver cancer.

2.6 肝癌致病关键基因对肝癌预后的影响

将关键基因分别导入在线工具 Kaplan Meier-Plotter。图6结果显示,与肝癌发病密切相关的10

个关键基因高表达组的肝癌患者的总生存期(overall survival, OS)均低于低表达组。CDC20、CCNB2、CCNB1、BUB1、BIRC5、AURKA、TOP2A、NCAPH、DL-

GAP5、CDK1 对应 P 值分别为 0.000 000 51、0.001 3、0.000 034、0.000 058、0.000 000 74、0.001 1、0.000 12、0.000 28、0.000 02、0.000 011, 差异均具有统计学意义。

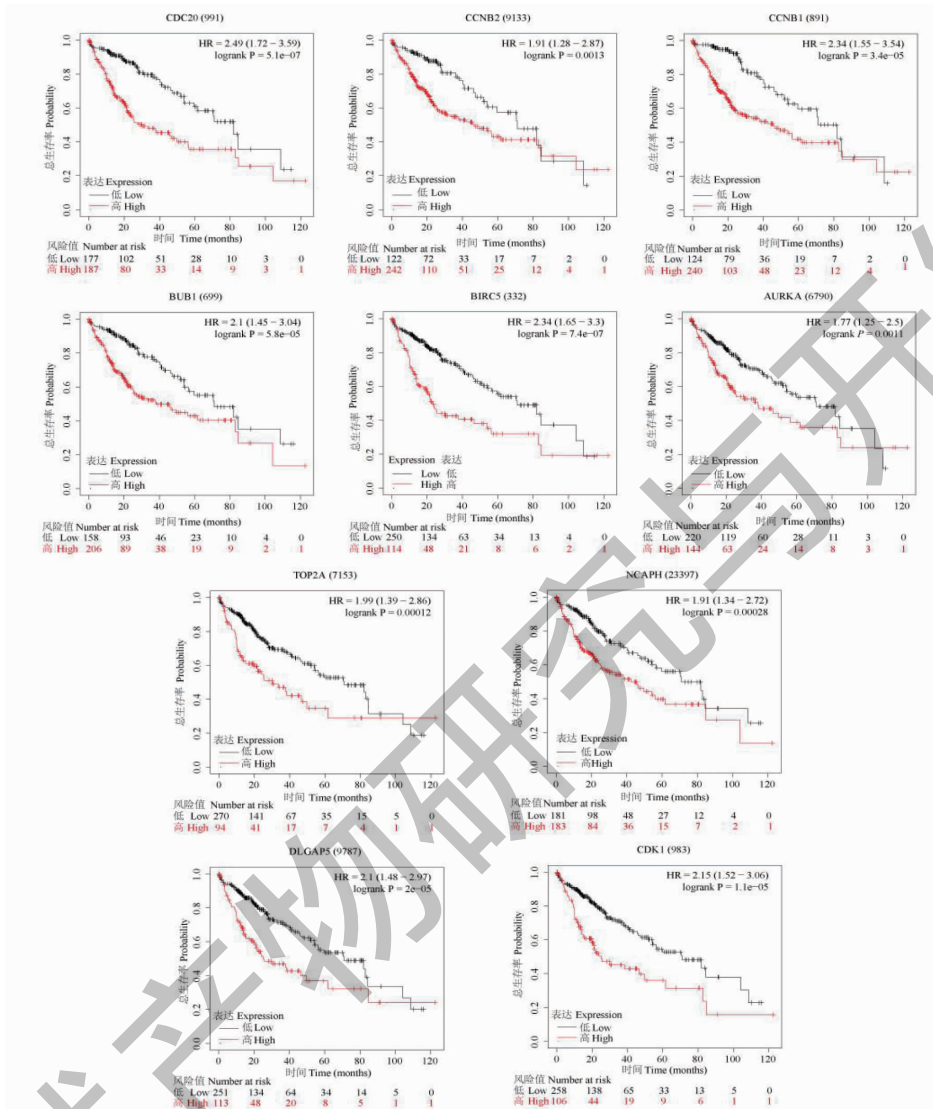


图 6 肝癌致病关键基因表达水平对肝癌预后影响的生存曲线图

Fig. 6 Survival curve of the effect of the expression levels of the core genes involved in liver cancer on the prognosis of liver cancer

2.7 肝癌致病关键基因表达验证

使用在线分析工具 UALCAN 比较 10 个关键基因的表达水平, 结果表明, 10 个关键基因在肝癌组织的表达水平显著高于正常组织, 差异均具有统计学意义 ($P < 0.05$), 结果如图 7 所示。10 个关键基因在 1、2、3、4 期肝癌组织中的表达的水平均高于正常组织, 且在第 3 期表达水平最高, 结果如图 8 所示。运用蛋白表达数据库 HPA 对 10 个关键基因进行分析, 结果发现, 肝癌组织中 CDC20、CCNB1、TOP2A、NCAPH 的蛋白表达量均高于正常组织, 结果如图 9 所示, 4 个关键基因的蛋白染色情况见表 2。

2.8 肝癌致病关键基因突变及预后分析

利用基因组学在线分析工具 cBioPortal 数据库对来源于 TCGA 的 373 例肝癌样本中 CDC20、CCNB2、CCNB1、BUB1、BIRC5、AURKA、TOP2A、NCAPH、DLGAP5、CDK1 等 10 个关键致病基因进行分析。结果发现共有 34.17% 的样本 (126 例) 中上述基因发生改变, 其中有 2.22% 的样本 (8 例) 发生突变, 4.17% 的样本 (15 例) 发生多种改变, 6.67% 的样本 (24 例) 发生扩增, 1.39% 的样本 (5 例) 发生重度缺失, 19.72% 的样本 (71 例) mRNA 表达上调, 结果如图 10A、10B 所示。关键基因的生存分析结

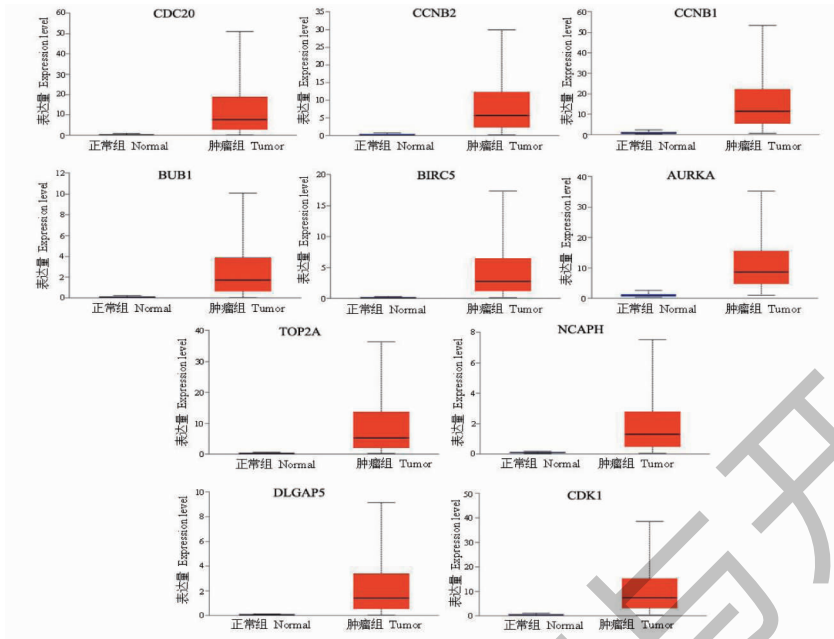


图 7 正常组织和肝癌组织中关键基因的表达情况

Fig. 7 Expression of core genes in normal tissues and liver cancer tissues

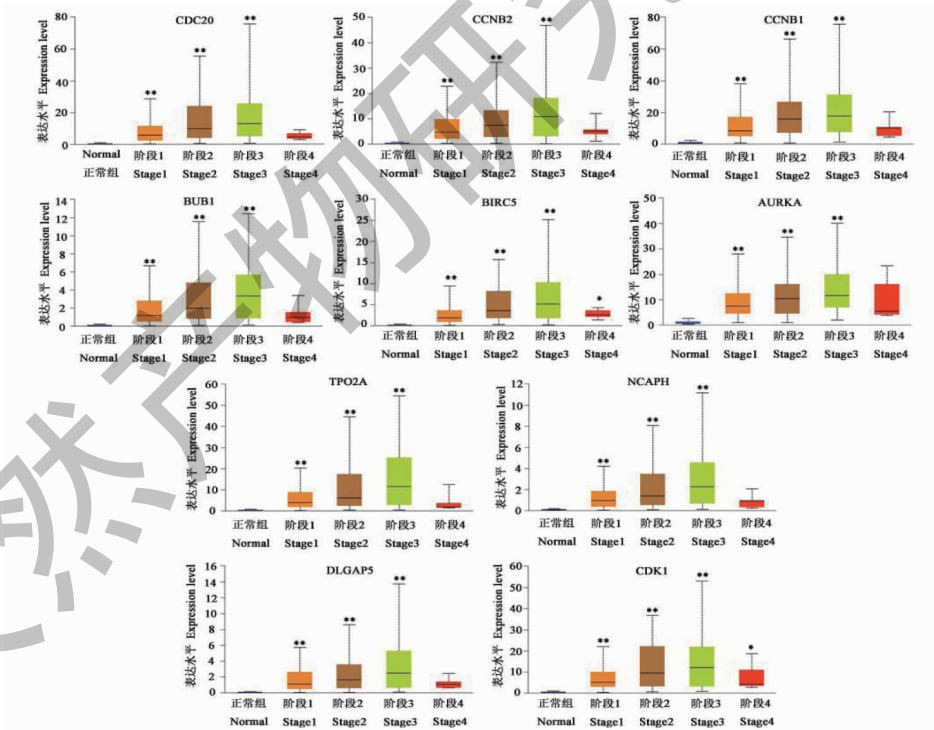


图 8 关键基因在肝癌病理分期中的表达情况

Fig. 8 Expression of core genes in pathological staging of hepatocellular carcinoma

注:与正常组比较, * $P < 0.05$, ** $P < 0.01$ 。Note: Compared with the normal group, * $P < 0.05$, ** $P < 0.01$.

果显示,34.17% 的样本(126 例)关键基因发生改变,为基因改变组,且关键基因改变组的总生存期(overall survival)和无病生存期(disease free survi-

al)均显著低于未改变组($P < 0.05$),结果如图 10C ~ 10D 所示。

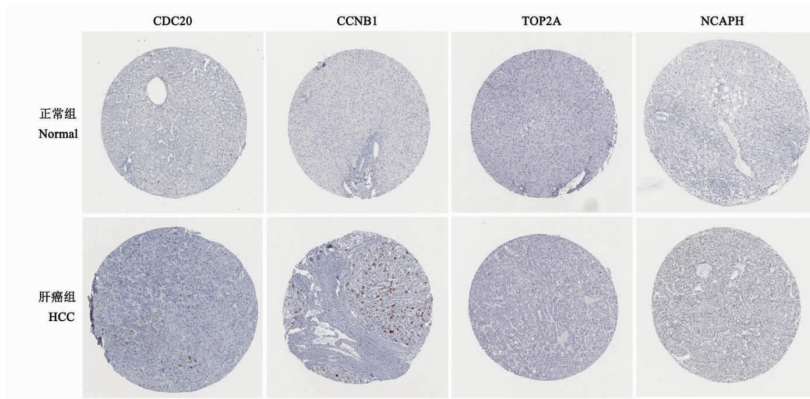


图 9 肝癌致病关键基因在正常组织和肝癌组织中的免疫组化图

Fig. 9 Immunohistochemical maps of the core genes of hepatocellular carcinoma in normal and hepatocellular carcinoma tissues

表 2 核心基因在正常组织和肝癌中的蛋白染色情况

Table 2 Protein staining of core genes in normal tissues and hepatocellular carcinoma

基因名 Gene	染色 Staining		密度 Intensity		定量 Quantity		定位 Location	
	正常组 Normal	肝癌组 HCC	正常组 Normal	肝癌组 HCC	正常组 Normal	肝癌组 HCC	正常组 Normal	肝癌组 HCC
CDC20	未检测到	低	无	中	无	<25%	无	细胞质
CCNB1	未检测到	中	无	强	无	<25%	无	细胞质/细胞膜
TOP2A	未检测到	中	无	强	无	<25%	无	细胞核
NCAPH	未检测到	中	无	强	无	>75%	无	细胞质/细胞膜

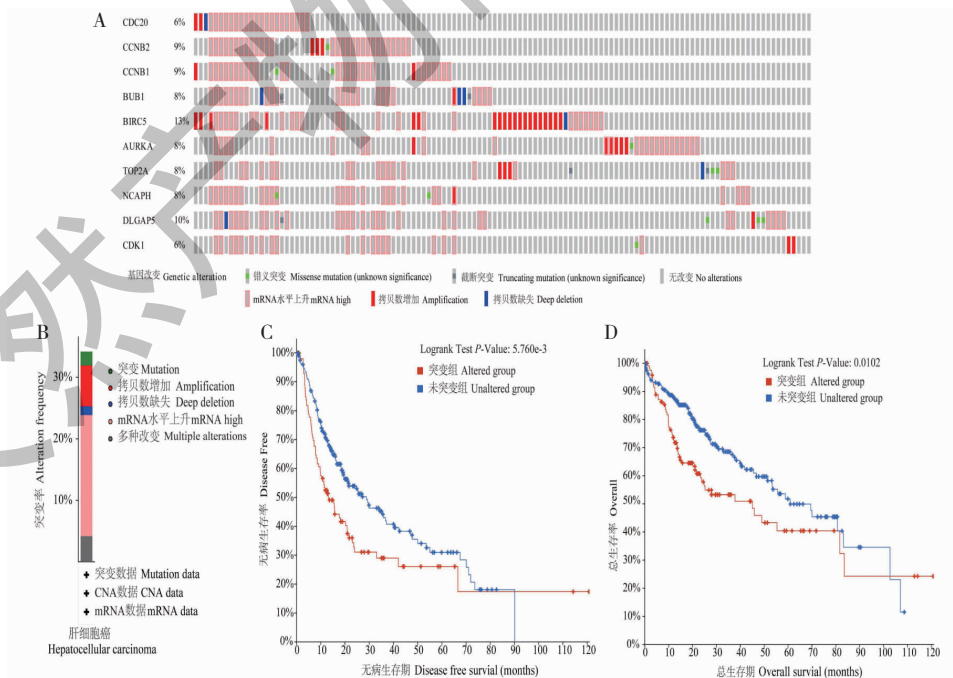


图 10 关键致病基因在肝癌组织中的突变及其生存预后情况

Fig. 10 Mutations of core pathogenic genes in hepatocellular carcinoma tissues and their survival prognosis

注: A. 各关键基因的突变; B. 关键基因总体突变; C. 无病生存期; D. 总生存期。Note: A. Overall mutation of core genes; B. Mutations of each core gene; C. Disease free survival; D. Overall survival.

2.9 肝癌致病关键基因间的相关性分析

运用 R 软件的 `corrplot` 包对与肝癌发病及预后密切相关的 10 个关键基因进行相关性分析,结果表明,关键基因之间的表达呈正相关关系,如图 11 所

示,且表达的相关性具有统计学意义($P < 0.01$),统计学检验结果见表 3。以上结果提示,关键致病基因相互影响,共同诱导和促进肝癌的发生、发展。

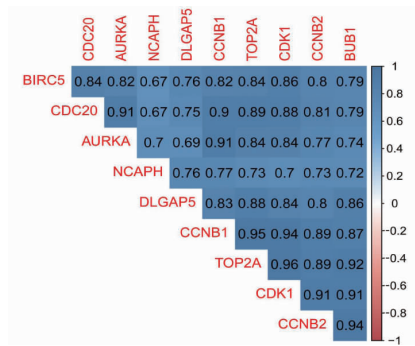


图 11 肝癌致病关键基因表达相关性示意图

Fig. 11 Schematic diagram of correlation between expression of pathogenic core genes of liver cancer

表 3 肝癌致病关键基因间相关性的统计学分析结果

Table 3 Statistical analysis of correlation between pathogenic core genes of liver cancer

基因 Gene	关键基因 Cor-gene	<i>P</i>	基因 Gene	关键基因 Cor-gene	<i>P</i>
CDC20	CCNB2	1.50E-05	CCNB1	CDK1	5.50E-10
	CCNB1	9.12E-08	BUB1	BIRC5	3.63E-05
	BUB1	4.09E-05	AURKA	TOP2A	0.000169
	BIRC5	3.35E-06	TOP2A	NCAPH	1.17E-08
	AURKA	3.78E-08	NCAPH	DLGAP5	0.00034
	TOP2A	1.14E-07	DLGAP5	CDK1	1.55E-06
	NCAPH	0.001138	CDK1	BIRC5	3.87E-08
	DLGAP5	0.000142	BIRC5	AURKA	1.12E-05
	CDK1	2.29E-07	TOP2A	TOP2A	4.36E-06
	CCNB2	CCNB1	1.52E-07	NCAPH	DLGAP5
	BUB1	5.83E-10	DLGAP5	CDK1	8.58E-05
	BIRC5	2.57E-05	CDK1	AURKA	1.08E-06
	AURKA	6.69E-05	AURKA	TOP2A	3.17E-06
	TOP2A	1.23E-07	NCAPH	NCAPH	0.000634
	NCAPH	0.000243	DLGAP5	DLGAP5	0.000768
	DLGAP5	2.44E-05	CDK1	CDK1	3.19E-06
	CDK1	2.17E-08	TOP2A	NCAPH	0.000288
CCNB1	BUB1	6.61E-07	DLGAP5	DLGAP5	2.58E-07
	BIRC5	9.08E-06	CDK1	CDK1	3.04E-11
	AURKA	2.79E-08	NCAPH	DLGAP5	9.79E-05
	TOP2A	8.30E-11	DLGAP5	CDK1	0.000556
	NCAPH	7.94E-05	DLGAP5	CDK1	4.54E-06
	DLGAP5	7.08E-06			

2.10 肝癌致病关键基因免疫浸润结果

运用免疫浸润数据库 TIMER 对 10 个关键基因进行分析, 结果发现, CDC20、CCNB2、CCNB1、BUB1、BIRC5、AURKA、TOP2A、NCAPH、CDK1 的表达与肝癌细胞纯度呈正相关关系, 差异均具有统计

学意义 ($P < 0.05$)。CDC20、CCNB2、CCNB1、BUB1、BIRC5、AURKA、TOP2A、NCAPH、DLGAP5、CDK1 的表达与 B 细胞、CD8⁺ T 细胞、CD4⁺ T 细胞、巨噬细胞、中性粒细胞、树突状细胞浸润程度呈正相关关系, 差异具有统计学意义 ($P < 0.05$), 结果见表 4。

表 4 肝癌关键致病基因与免疫细胞浸润相关性的统计学分析

Table 4 Statistical analysis of correlation between key pathogenic genes and immune cell infiltration in hepatocellular carcinoma

基因 Gene	纯度 Purity	B 细胞 B cell	CD8 ⁺ T 细胞 CD8 ⁺ T cell	CD4 ⁺ T 细胞 CD4 ⁺ T cell	巨噬细胞 Macrophage	中性粒细胞 Neutrophil	树突状细胞 Dendritic cell
CDC20	△	△	△	△	△	△	△
CCNB2	△	△	△	△	△	△	△
CCNB1	△	△	△	△	△	△	△
BUB1	△	△	△	△	△	△	△
BIRC5	△	△	△	△	△	△	△
AURKA	△	△	△	△	△	△	△
TOP2A	△	△	△	△	△	△	△
NCAPH	△	△	△	△	△	△	△
DLGAP5		△	△	△	△	△	△
CDK1	△	△	△	△	△	△	△

注: △代表 $P < 0.05$, 表示基因与免疫细胞浸润密切相关

Note: △ represents $P < 0.05$, indicating that gene is closely related to immune cell infiltration.

2.11 潜在的可用于肝癌治疗的药物筛选结果

借助 Coremine Medical 数据库、药物遗传学 (CTD) 数据库对 10 个关键基因进行分析, 筛选潜在的可用于肝癌治疗的中药、天然活性成分, 共得到中药 36 种、天然治疗成分 105 种。利用 cytoscape 软件绘制潜在治疗中药-核心基因-天然活性成分的相互作用网络, 如图 12 所示, 靶向作用于关键基因两

次及以上的中药有 8 种, 分别为青蒿、高良姜、冬凌草、雷丸、野马追、鹅不食草、蟾酥、九香虫, 以 $P < 0.01$ 为筛选条件, 共获得潜在治疗肝癌的中药 36 种, 结果见表 5。靶向作用于 6 个以上关键基因的天然活性成分有白藜芦醇、金雀异黄素、槲皮素、医用棕榈树、鱼藤酮、金复康、紫杉醇、姜黄素、长春新碱、表没食子儿茶素没食子酸酯, 共 10 种。

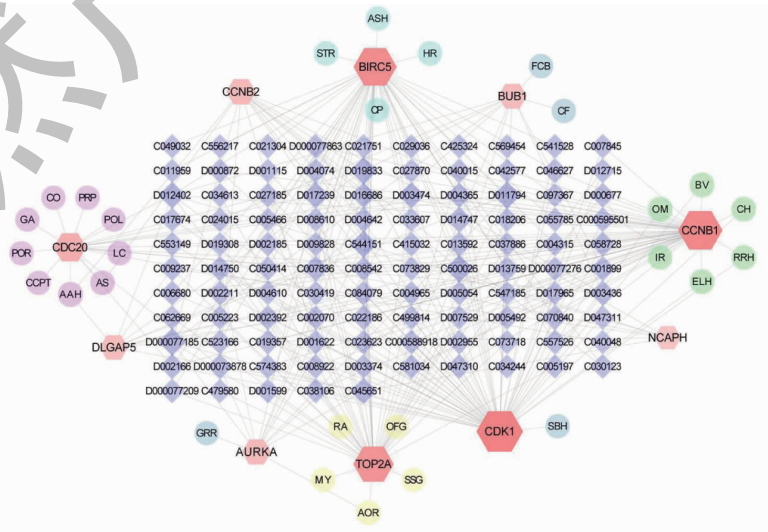


图 12 潜在治疗中药-核心基因-天然治疗成分的作用关系网络

Fig. 12 Relationship network between traditional Chinese medicine and pathogenic core genes of liver cancer

表5 靶向作用于肝癌致病关键基因的中药信息

Table 5 Information of traditional Chinese medicine targeting the pathogenic core gene of liver cancer

基因名称 Gene	中药名称 Chinese medicine	P 值 P value
CDC20	鹿筋 Ligamentum Cervi (LC)	0.000 112
	猪苓 Polyporus (POL)	0.005 254
	茯神 Poria cum Radix Pini (PRP)	0.015 422
	云芝 Coriolus (CO)	0.017 576
	灵芝 Ganoderma (GA)	0.018 327
	九香虫 Aspongopus (AS)	0.023 816
	茯苓 Poria (POR)	0.025 460
	鹿茸 Cervi Cornu Panto Trichum (CCPT)	0.028 196
	青蒿 Artemisiae Annuae Herba (AAH)	0.049 074
	CCNB2	九香虫 Aspongopus (AS)
青蒿 Artemisiae Annuae Herba (AAH)		0.017 778
CCNB1	野马追 Eupatorii Lindleyani Herba (ELH)	0.000 747
	冬凌草 Rabdosiae Rubescentis Herba (RRH)	0.001 064
	鹅不食草 Centipeda Herba (CH)	0.001 877
	土木香 Inulae Radix (IR)	0.002 223
	蟾酥 Bufonis Venenum (BV)	0.002 379
	雷丸 Omphalia (OM)	0.002 622
BUB1	川贝母 Fritillariae Cirrhosae Bulbus (FCB)	0.001 406
	香橼 Citri Fructus (CF)	0.008 158
BIRC5	山慈菇 Cremastrae Pseudobulbus (CP)	0.001 871
	红芪 Hedysari Radix (HR)	0.001 885
	茵陈 Artemisiae Scopariae Herba (ASH)	0.002 288
AURKA	防己 Stephaniae Tetrandrae Radix (STR)	0.003 121
	高良姜 Alpiniae Officinarum Rhizoma (AOR)	0.006 094
	甘草 Glycyrrhizae Radix (GRR)	0.006 094
TOP2A	大豆黄卷 Sojae Semen Germinatum (SSG)	0.004 747
	高良姜 Alpiniae Officinarum Rhizoma (AOR)	0.021 956
	稻芽 Oryzae Fructus Germinatus (OFG)	0.024 139
	紫草 Radix Arnebiae (RA)	0.041 436
DLGAP5	没药 Myrrha (MY)	0.042 536
	青蒿 Artemisiae Annuae Herba (AAH)	0.013 322
CDK1	冬凌草 Rabdosiae Rubescentis Herba (RRH)	0.001 141
	雷丸 Omphalia (OM)	0.001 37
	野马追 Eupatorii Lindleyani Herba (ELH)	0.001 567
	鹅不食草 Centipeda Herba (CH)	0.001 748
	半枝莲 Scutellariae Barbatae Herba (SBH)	0.003 604
	蟾酥 Bufonis Venenum (BV)	0.005 049

3 讨论与结论

肝癌是全球范围内最常见的癌症之一,每年罹患肝癌的人数约为 84 万,而每年因肝癌死亡的人数至少为 78 万^[9]。中国是一个肝炎大国,肝炎发病率的上升增加了肝癌患者的数量,世界上约有 50% 的肝癌患者来自中国^[10]。目前,早期肝癌的治疗主要采取手术切除、射频消融和肝脏移植,中期通常采用手术、射频消融、靶向治疗、免疫治疗以及化疗等多种方式结合的手段;而晚期大多采取支持治疗^[11]。肝癌是一种复发率极高的恶性肿瘤,采取手术切除等根治性治疗手段,其 5 年内的复发率仍高达 77%^[12],故临床可考虑采用中西医结合治疗、中医药治疗的方法。中医对肝癌并无具体的称谓,依据其症候表征,多将其归属为“癥瘕”“肝积”“黄疸”“臌胀”等疾病范畴。中医药治疗肝癌包括中药复方治疗、单位中药/中成药治疗以及针灸、穴位敷贴等外治法;中西医结合治疗主要是中医联合手术治疗、中医联合射频消融、中医联合靶向治疗、中医联合化疗等,以达到减少并发症、减轻副作用、延缓肿瘤复发转移的目的^[13,14]。本文通过采用生物信息学相关知识,从分子机制层面对肝癌的芯片数据集进行分析,挖掘其致病关键基因,并分析关键基因在肝癌预后中的意义,进一步通过关键基因筛选潜在治疗肝癌的中药。本研究旨在为肝癌的鉴别诊断、预后和治疗提供科学依据和参考。

GO 功能和 KEGG 通路富集分析发现,差异表达基因主要参与免疫反应、炎症反应和细胞粘附等生物学过程,涉及的信号通路包括 p53、mTOR、ErbB 等。研究表明,肿瘤睾丸抗原和 Sal 样蛋白 4 与特异性 T 细胞的反应在控制早期肝癌中可能发挥着重要作用^[15];肝脏中核苷酸去结合寡聚化结构域 2 的缺失可通过炎症反应、DNA 损伤和基因组不稳定性来诱导肝癌的发生^[16];NK 细胞来源的干扰素- γ 通过 HBV 转基因小鼠上皮细胞黏附分子-上皮间充质转换过程,可促进肝癌的发生^[17];此外,有研究显示 miR-621 可通过激活 p53 信号通路提高肝癌细胞的放射敏感性^[18];高表达水平的蛋白酶体活性亚单位 4 通过 mTOR 信号通路可促进肝癌细胞的增殖^[19];上皮 V 样抗原 1 能够上调 ErbB3-PI3K 信号通路进而促进肝癌的进展和转移^[20]。

蛋白相互作用分析得到 10 个在肝癌组织中高

表达的关键基因,分别为 CDC20、CCNB2、CCNB1、BUB1、BIRC5、AURKA、TOP2A、NCAPH、CDK1。对此 10 个关键基因进行预后分析发现,关键基因高表达组肝癌患者总生存期明显降低,这些关键基因与肝癌的病理分析结果存在一定的正相关关系。研究表明,CDC20 在肝癌组织中的表达高于正常组织($P < 0.05$),且高表达 CDC20 的肝癌患者总体的生存率较低^[21];研究显示,高 CCNB2 水平的肝癌患者 5 年总生存期和无病生存期均短于低 CCNB2 水平者^[22];另有研究表明,在肝癌组织中可检测到较高水平的 BUB1B,BUB1B 过表达与不良的临床病理表现呈正相关^[23]。

进一步对 10 个关键基因进行表达验证和突变分析发现,关键基因在肝癌组织中高表达,其中 CDC20、CCNB1、TOP2A、NCAPH 在肝癌组织中的蛋白表达量有不同程度的增加;此外,关键基因在肝癌组织中存在一定程度的突变,携带此突变关键基因的肝癌患者有较差的预后。关键基因的免疫细胞浸润分析发现,关键基因的表达与 B 细胞、CD8⁺ T 细胞、CD4⁺ T 细胞、巨噬细胞、中性粒细胞、树突状细胞的浸润程度呈正相关关系。Yang Gang 等采用免疫组织化学法检测肝癌组织中 CDC20 的蛋白表达水平,结果显示在 59.2% 的肝细胞癌样本中观察到 CDC20 高表达^[24]。Rong Min-Hua 等研究发现 CCNB1 的 mRNA 和蛋白在肝细胞癌组织中过度表达,且 CCNB1 可能通过调节 DNA 复制参与 HCC 的细胞周期^[25]。这些结果提示本文预测结果具有较好的可信性与准确性。

潜在治疗药物的筛选结果发现,CCNB1、BIRC5、CDK1、CDC20 筛选出的中药数目多,CDK1、CCNB1、BIRC5、TOP2A 所结合的潜在天然治疗成分数量多,CDK1、CCNB1、BIRC5 可能为药物治疗的通用靶点。青蒿、高良姜、冬凌草等可作用于 2~3 个关键基因,为潜在治疗肝癌的中药。研究表明,青蒿的有效成分为青蒿素,其可通过 EGFR 等多种途径调节肝癌细胞的增殖、凋亡、血管生成^[26];高良姜的有效成分为高良姜素,其可通过内质网应激和线粒体依赖性凋亡来预防肝细胞癌^[27];冬凌草可通过靶向 AKT 通路来增强肝癌对索拉非尼抗癌作用的敏感性^[28]。中医理论认为,肝细胞癌的主要病机为“瘀”“毒”“虚”,中药在分子水平发挥扶正祛邪、攻

毒抗癌的功效,可达到机体阴阳平衡的状态^[29]。相较于传统的西医治疗,中医药在治疗肝细胞癌中具有潜在的优势,具体体现在中药副作用小,不良反应发生率低,肿瘤生长减缓,复发转移减少,患者生存质量提升,生存周期延长^[30]。

综上所述,本文研究发现肝癌的发生发展具有多基因、多通路、多功能的特点,共筛选获得 10 个肝癌致病关键基因;这些基因及其蛋白产物在肝癌患者中的表达水平均有升高,且与肝癌的不良预后、免疫细胞浸润程度密切相关;进一步挖掘得到青蒿、高良姜、冬凌草等 36 味潜在的肝癌靶向治疗中药和 105 个天然活性成分。本文研究结果可为肝癌的临床诊断、预后判断及相关治疗提供科学依据。

参考文献

- 1 Wang WL, Liu XY, Liu FJ, et al. Review on molecular mechanism of active components from natural products against hepatocellular carcinoma[J]. *Nat Prod Res Dev*(天然产物研究与开发), 2020, 32:1606-1613.
- 2 Zhou J, Sun H, Wang Z, et al. Guidelines for the diagnosis and treatment of hepatocellular carcinoma (2019 Edition) [J]. *Liver cancer*, 2020, 9(6):682-720.
- 3 Mansouri V, Razzaghi M, Nikzamir A, et al. Assessment of liver cancer biomarkers [J]. *Gastroenterol Hepatol Bed Bench*, 2020, 13(Suppl1):S29-S39.
- 4 Alqahtani A, Khan Z, Alloghbi A, et al. Hepatocellular carcinoma: molecular mechanisms and targeted therapies[J]. *Medicina*(Kaunas), 2019, 55(9):526.
- 5 Wratten L, Wilm A, Cöke J. Reproducible, scalable, and shareable analysis pipelines with bioinformatics workflow managers[J]. *Nature Methods*, 2021, 18:1161-1168.
- 6 Meggendorfer M, Jobanputra V, Wrzeszczynski KO, et al. Analytical demands to use whole-genome sequencing in precision oncology [J]. *Semin Cancer Biol*, 2021, 3(21):179-186.
- 7 Mangul S, Martin LS, Langmead B, et al. How bioinformatics and open data can boost basic science in countries and universities with limited resources [J]. *Nat Biotechnol*, 2019, 37:324-326.
- 8 Deng YB, Nagae G, Midorikawa Y, et al. Identification of genes preferentially methylated in hepatitis C virus-related hepatocellular carcinoma [J]. *Cancer Sci*, 2010, 101:1501-1510.
- 9 Wang H, Lu Z, Zhao X. Tumorigenesis, diagnosis, and therapeutic potential of exosomes in liver cancer [J]. *J Hematol Oncol*, 2019, 12(1):133.
- 10 Zuo M, Huang J. The history of interventional therapy for liver cancer in China [J]. *J Interv Med*, 2018, 1(2):70-76.
- 11 2019 Expert Committee on Clinical Coordination of Traditional Chinese and Western Medicine for Liver Cancer. Interventional expert consensus on integrated traditional Chinese and Western medicine for the diagnosis and treatment of hepatocellular carcinoma(trial edition I) [J]. *J Interv Radiol*(介入放射学杂志), 2021, 30:1079-1090.
- 12 Liu L, Cheng QL, Bai CC, et al. Advances of traditional Chinese medicine in prevention and treatment of recurrence and metastasis of liver cancer [J]. *J Clin Hepatol*(临床肝胆病杂志), 2021, 37:2009-2015.
- 13 Zhan MY, Zhang GL. Breakthrough points and difficulties in traditional Chinese medicine treatment of primary liver cancer [J]. *J Clin Hepatol*(临床肝胆病杂志), 2021, 37:2016-2024.
- 14 Gao WQ, Gan ZH, Yin CJ. A methodological study of traditional Chinese medicine treatment of primary liver cancer [J]. *J Clin Hepatol*(临床肝胆病杂志), 2021, 37:2009-2015.
- 15 Zang C, Zhao Y, Qin L, et al. Distinct tumour antigen-specific T-cell immune response profiles at different hepatocellular carcinoma stages [J]. *BMC cancer*, 2021, 21(1):1007.
- 16 Zhou Y, Hu L, Tang W, et al. Hepatic NOD2 promotes hepatocarcinogenesis via a RIP2-mediated proinflammatory response and a novel nuclear autophagy-mediated DNA damage mechanism [J]. *J Hematol Oncol*, 2021, 14(1):9.
- 17 Chen Y, Hao X, Sun R, et al. Natural killer cell-derived interferon-gamma promotes hepatocellular carcinoma through the epithelial cell adhesion molecule-epithelial-to-mesenchymal transition axis in hepatitis B virus transgenic mice [J]. *Hepatology*, 2019, 69:1735-1750.
- 18 Shao Y, Song X, Jiang W, et al. MicroRNA-621 acts as a tumor radiosensitizer by directly targeting SETDB1 in hepatocellular carcinoma [J]. *Mol Ther*, 2019, 27:355-364.
- 19 Ge S, Huang H, Huang W, et al. PSME4 activates mTOR signaling and promotes the malignant progression of hepatocellular carcinoma [J]. *Int J Gen Med*, 2022, 15:885-895.
- 20 Ni Q, Chen Z, Zheng Q, et al. Epithelial V-like antigen 1 promotes hepatocellular carcinoma growth and metastasis via the ERBB-PI3K-AKT pathway [J]. *Cancer Sci*, 2020, 111:1500-1513.

- 21 Zhang X, Zhang X, Li X, et al. Connection between CDC20 expression and hepatocellular carcinoma prognosis [J]. *Med Sci Monit*, 2021, 27; e926760.
- 22 Li R, Jiang X, Zhang Y, et al. Cyclin B2 overexpression in human hepatocellular carcinoma is associated with poor prognosis [J]. *Arch Med Res*, 2019, 50(1): 10-17.
- 23 Qiu J, Zhang S, Wang P, et al. BUB1B promotes hepatocellular carcinoma progression via activation of the mTORC1 signaling pathway [J]. *Cancer Med*, 2020, 9: 8159-8172.
- 24 Yang G, Wang G, Xiong Y, et al. CDC20 promotes the progression of hepatocellular carcinoma by regulating epithelial-mesenchymal transition [J]. *Mol Med Rep*, 2021, 24(1): 483.
- 25 Rong MH, Li JD, Zhong LY, et al. CCNB1 promotes the development of hepatocellular carcinoma by mediating DNA replication in the cell cycle [J]. *Exp Biol Med (Maywood)*, 2022, 247: 395-408.
- 26 Zhang S, Mo Z, Zhang S, et al. A Network pharmacology approach to reveal the underlying mechanisms of *Artemisia annua* on the treatment of hepatocellular carcinoma [J]. *Evidence-based Compl Alt*, 2021, 2021: 8947304.
- 27 Fang D, Xiong Z, Xu J, et al. Chemopreventive mechanisms of galangin against hepatocellular carcinoma: a review [J]. *Biomed Pharmacother*, 2019, 109: 2054-2061.
- 28 Li X, Chen W, Liu K, et al. Oridonin sensitizes hepatocellular carcinoma to the anticancer effect of sorafenib by targeting the Akt pathway [J]. *Cancer Manag Res*, 2020, 12: 8081-8091.
- 29 Ning L, Sun JG. Research progress of traditional Chinese medicine and western medicine in precancerous lesions of primary hepatic carcinoma [J]. *Mod Tradit Chin Med Mater Med World Sci Technol (世界科学技术-中医药现代化)*, 2021, 23: 3590-3598.
- 30 Li Z, Wang K, Wang KQ, et al. Research progress on mitochondrial prevention and treatment of liver cancer mediated by traditional Chinese medicine [J]. *Chin Tradit Herb Drug (中草药)*, 2022, 53: 575-589.