

## • 数据挖掘与循证医学 •

## 基于生物信息学分析肝癌差异基因及潜在的中药干预

张道平<sup>1</sup>, 刘起立<sup>1</sup>, 姚昆鹏<sup>1</sup>, 李峥嵘<sup>1</sup>, 陈沛儒<sup>1</sup>, 彭建平<sup>1,2\*</sup>

1. 湖南中医药大学, 湖南 长沙 410208

2. 湖南中医药大学第一附属医院, 湖南 长沙 410007

**摘要:** **目的** 探究肝癌患者的差异基因表达量与预后的相关性, 以及差异基因在多种癌症的调控作用, 并寻找潜在的治疗靶点以及对应中药, 为临床中药的使用、中医药组方及加减用药提供依据。**方法** 从 TCGA 数据库下载肿瘤患者的转录组数据以及相应临床数据, 使用 R 语言的 edgeR 包分析得出差异表达基因, 并筛选、分析关键基因在泛癌的表达量及其表达情况; 进一步筛选得到核心基因, 研究其在泛癌中具体肿瘤的表达情况, 以及肝癌临床相关性分析, 并将核心基因提交至 TIMER、THPA 数据库, 研究免疫细胞相关性和病理学情况; 最后通过 Coremine 数据库查找核心基因对应中药, 使用 TCMSP、BATMAN-TCM 数据库查找中药有效成分, 利用 Cytoscape 软件制作药物-核心基因网络图, 并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。**结果** 共筛选出 10 个关键基因, 选取其中 5 个核心基因 *CCNB1*、*TOP2A*、*CCNA2*、*CDK1*、*CDC20*, 进行后续分析得出其与肝癌的预后生存呈负相关, 与肝癌的免疫细胞呈正相关, 并得到基因过表达病理学图片。其对应中药为丹参、枳实、枳壳、人参叶、人参芦、甘草、人参、蜂蜜。KEGG 富集分析显示中药除调控核心基因外仍通过多靶点起作用。**结论** 通过生物信息学技术分析得到肝癌核心基因及相关信号通路与治疗靶点, 并得到调控中药名称及其有效成分。为治疗肝癌以及其他癌症的中药复方研发提供思路。

**关键词:** 泛癌; 肝癌; 生物信息学; 差异基因; 丹参; 枳实; 枳壳; 人参叶; 人参芦; 甘草; 人参; 蜂蜜

**中图分类号:** R285 **文献标志码:** A **文章编号:** 0253-2670(2022)03-0788-11

**DOI:** 10.7501/j.issn.0253-2670.2022.03.019

## Analysis of differentially expressed genes of liver cancer and potential intervention of traditional Chinese medicine based on bioinformatics

ZHANG Dao-ping<sup>1</sup>, LIU Qi-li<sup>1</sup>, YAO Kun-peng<sup>1</sup>, LI Zheng-rong<sup>1</sup>, CHEN Pei-ru<sup>1</sup>, PENG Jian-ping<sup>1,2</sup>

1. Hunan University of Chinese Medicine, Changsha 410208, China

2. First Affiliated Hospital of Hunan University of Chinese Medicine, Changsha 410007, China

**Abstract: Objective** To explore the correlation between differential gene expression in patients with liver cancer and prognosis, as well as the regulatory role of differential genes in various cancers, and find potential therapeutic targets and corresponding traditional Chinese medicine (TCM), so as to provide basis for the clinical use of TCM, the formulation of TCM and its the addition or reduction use. **Methods** The transcriptome data and corresponding clinical data of cancer patients in TCGA database were downloaded, and the edgeR package of R language was used to analyze the differential genes. The expression and co-expression of key genes in pan-cancer were screened and analyzed, and the core genes were further screened. The expression of the core genes in specific tumors and the clinical correlation analysis of liver cancer were studied. The core genes were submitted to TIMER and THPA databases to study the correlation of immune cells and pathology. Finally, the core corresponding TCMS were found through the Coremine database. TCMSP and BATMAN-TCM databases were used to find the effective components of TCM. Cytoscape software was used to produce drug-core gene network diagram, and Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analysis was performed. **Results** A total of 10 key genes were screened, and five of them were selected as the core genes, namely *CCNB1*, *TOP2A*, *CCNA2*, *CDK1* and *CDC20*. Subsequent analysis showed that they were negatively correlated with the prognosis and survival of liver cancer, and positively correlated with the immune cells of liver cancer, and the pathological pictures of gene overexpression were obtained.

收稿日期: 2021-11-03

基金项目: 国家中医临床研究基地业务建设项目第二批科研专项 (JDZX2015183)

作者简介: 张道平 (1997—), 男, 在读硕士研究生, 主要从事中医药防治肝病研究。E-mail: 416367611@qq.com

\*通信作者: 彭建平 (1979—), 女, 博士, 副主任医师, 硕士生导师, 主要从事中医药防治肝病研究。E-mail: jppeng@126.com

The core genes corresponding TCM were Danshen (*Salviae Miltiorrhizae Radix et Rhizoma*), Zhishi (*Aurantii Fructus Immaturus*), Zhike (*Aurantii Fructus*), Renshenye (*Ginseng Folium*), Renshenlu (rhizome of *Panax ginseng*), Gancao (*Glycyrrhizae Radix et Rhizoma*), Renshen (*Ginseng Radix et Rhizoma*), Fengmi (*Mel*). The KEGG enrichment analysis pathway map showed that TCM still acts through multiple targets in addition to regulating core genes. **Conclusion** The core genes of liver cancer and related signaling pathways and therapeutic targets obtained by bioinformatics analysis, the names of traditional Chinese medicine and their active components were obtained. To provide ideas for the treatment of liver cancer and other cancers.

**Key words:** pan-cancer; liver cancer; bioinformatics; differentially expressed genes; *Salviae Miltiorrhizae Radix et Rhizoma*; *Aurantii Fructus Immaturus*; *Aurantii Fructus*; *Ginseng Folium*; rhizome of *Panax ginseng*; *Glycyrrhizae Radix et Rhizoma*; *Ginseng Radix et Rhizoma*; *Mel*

癌症已是全球的第2大死因, 在一些国家, 癌症是第一死因。据预测, 到2040年, 全球每年将有2750万新的癌症病例<sup>[1]</sup>, 其中原发性肝癌 (primary hepatic carcinoma, PHC) 是全球第6大常见癌症, 也是世界上死亡率最高的癌症之一<sup>[2]</sup>, 肝细胞癌 (hepatocellular carcinoma, HCC) 占有PHC的70%~90%。乙型肝炎病毒 (hepatitis B virus, HBV) 感染是导致肝癌的首要原因, 澳大利亚近年对于肝癌的流行病学调查显示<sup>[3]</sup>, 肝癌的发生率在HBV流行地区 [乙型肝炎病毒表面抗原 (hepatitis B surface antigen, HBsAg) 流行率>6%] 最高, 其他的危险因素则为丙型肝炎病毒感染、黄曲霉毒素B1、过量饮酒、肥胖和一些遗传性代谢紊乱疾病等<sup>[4]</sup>。癌症处于复杂的免疫微环境中, 目前对于癌症的单一免疫疗法的缓解率和生存率较低, 因此, 针对肝癌免疫治疗研究的重点也逐渐转向多靶点联合治疗。但免疫治疗目前仍面临着许多问题<sup>[5]</sup>。因此如何改善肿瘤免疫微环境、调节机体免疫功能以及基于基因组学指导的肿瘤免疫治疗等是未来发展的方向。随着肿瘤研

究的不断深入, 目前已经逐步进入泛癌 (pan-cancer) 研究阶段。泛癌研究是指对多种不同类型的肿瘤基因组进行分析, 寻找不同来源肿瘤中共有的异常表达基因, 进而为不同肿瘤的诊断和治疗提供广谱靶点。中医药单独治疗肝癌分为传统中药复方汤剂、中药活性成分提取物2种方式; 中西医结合治疗肝癌包括中医药联合手术、中医药联合靶向药、中医药联合免疫治疗等方式。而中西医结合运用能够更有效地防治肝癌, 并延长患者生存期。中医药治疗肝癌具有多靶点、多途径的优势, 且能有效降低药物不良反应。本研究拟探究肝癌患者的差异基因表达量与预后的相关性, 以及差异基因对多种癌症的调控作用, 并寻找潜在的治疗靶点以及对应中药, 为临床中药的使用、中药组方及加减用药提供依据。

## 1 资料与方法

### 1.1 数据平台及软件工具

本研究使用生物信息学方法分析肝癌差异基因, 并通过数据库查找中药名称及有效成分, 使用的数据平台及软件工具见表1。

表1 数据平台及软件工具

Table 1 Data platform and software tools

数据平台及软件名称	网址
癌症基因组图谱 (TCGA)	<a href="https://portal.gdc.cancer.gov/">https://portal.gdc.cancer.gov/</a>
STRING 数据平台	<a href="https://www.string-db.org/">https://www.string-db.org/</a>
中药系统药理学数据库与分析平台 (TCMSP)	<a href="https://tcmsp-e.com/">https://tcmsp-e.com/</a>
COREMINE 数据平台	<a href="https://coremine.com/">https://coremine.com/</a>
UCSC Xena 数据平台	<a href="https://xena.ucsc.edu/">https://xena.ucsc.edu/</a>
BATMAN-TCM 数据库	<a href="http://bionet.ncpsb.org.cn/batman-tcm/">http://bionet.ncpsb.org.cn/batman-tcm/</a>
TIMER 平台	<a href="https://cistrome.shinyapps.io/timer/">https://cistrome.shinyapps.io/timer/</a>
KEGG 平台	<a href="https://www.kegg.jp/">https://www.kegg.jp/</a>
GeneCards 数据库	<a href="https://www.genecards.org/">https://www.genecards.org/</a>
OMIM 数据库	<a href="https://omim.org/">https://omim.org/</a>
DrugBank 数据库	<a href="https://go.drugbank.com/">https://go.drugbank.com/</a>
TTD 数据库	<a href="http://db.idrblab.net/ttd/">http://db.idrblab.net/ttd/</a>
PharmGkb 数据库	<a href="https://www.pharmgkb.org/">https://www.pharmgkb.org/</a>
edgeR 包	<a href="http://www.bioconductor.org/packages/release/bioc/html/edgeR.html">http://www.bioconductor.org/packages/release/bioc/html/edgeR.html</a>
Cytoscape3.8.2	<a href="https://cytoscape.org/">https://cytoscape.org/</a>

## 1.2 数据来源

在癌症基因组图谱 (cancer genome atlas, TCGA) 网站中下载肝细胞癌 (liver hepatocellular carcinoma, LIHC) 的转录组数据以及相应患者的临床信息, 并使用 UCSC Xena 网站下载 TCGA 数据库其余的 32 种肿瘤转录组数据, 分别用于基因表达分析及生存分析。

## 1.3 分析及流程

**1.3.1 分析差异基因** 使用 R 语言 ver.4.0.5 的 edgeR 包分析肝癌转录组测序数据, 获得差异基因 (differentially expressed genes, DEGs), 绘制差异基因火山图。

**1.3.2 蛋白互作网络** 根据错误发现率 (false discovery rate, FDR) 升序排序, 将上调基因提交至 STRING 数据库, 得出蛋白互作网络, 下载互作文件, 导入 Cytoscap 3.8.2 中, 使用 CytoHubba 插件, 筛选出打分前 10 位的关键基因。

**1.3.3 关键基因临床相关分析** 查看关键基因在泛癌组织中表达水平、关键基因共表达情况, 并分析肝癌中关键基因表达量与临床分期相关性。进一步选择其中 5 种与肝癌最相关的基因进行具体肿瘤表达差异研究, 并探索其表达量与肝癌生存时间的关系, 以及 5 种基因表达量与 6 种免疫细胞浸润的相关性、病理学情况。

**1.3.4 获取中药有效成分** 将使用 CytoHubba 插件筛选出的前 10 位关键基因提交至 Coremine 数据库, 查找对应中药名称, 经规范后统计对应中药较多的 5 种基因, 统计出现次数在中位数 (4) 以上的药物, 在 TCMSP (traditional Chinese medicine systems pharmacology database and analysis platform)、BATMAN-TCM (a bioinformatics analysis tool for molecular mechanism of traditional Chinese medicine) 数据库中寻找中药有效成分, 并利用 Cytoscape 软件构建中药成分-核心靶点网络图。

**1.3.5 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** 将筛选后得到的高频中药的有效成分所对应的肝癌靶点进行 KEGG 通路富集分析, 明确中药治疗肝癌除影响核心基因外, 仍能通过其他分子及通路影响癌症发生发展。

## 2 结果

### 2.1 差异基因及火山图

通过  $FDR < 0.05$  以及  $\log_2FC > 2$  筛选出差异基因, 得出 2565 个上调基因, 253 个下调基因, 绘制差异基因的火山图, 见图 1。

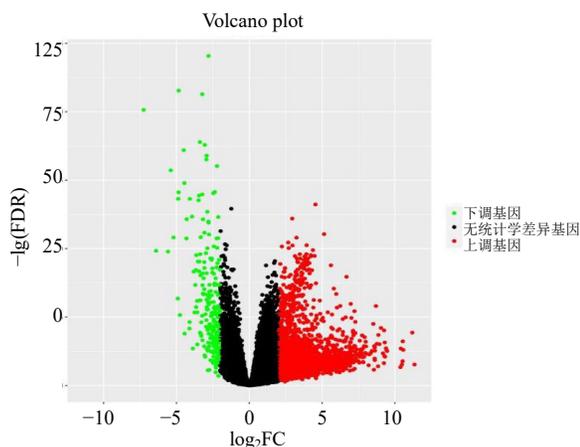
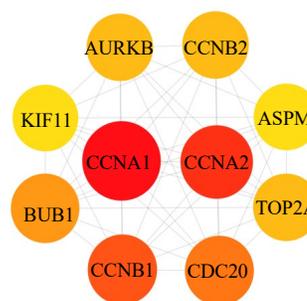


图 1 差异基因火山图

Fig. 1 Volcano map of differential genes

### 2.2 关键基因筛选

将上调基因提交至 STRING 数据库, 设置置信区间为 90%, 得到蛋白互作网络图相应的信息文件, 将文件导入 Cytoscape 软件中, 使用 CytoHubba 插件, 根据拓扑分析, 度 (degree) 值排名前 10 位的基因确定为关键基因 (hub gene), 分别是细胞周期蛋白依赖性激酶 1 (cyclin dependent kinase 1, CDK1)、细胞周期蛋白 A2 (cyclin A2, CCNA2)、细胞周期蛋白 B1 (cyclin B1, CCNB1)、细胞分裂周期蛋白 20 (cell division cycle 20, CDC20)、DNA 拓扑异构酶 II $\alpha$  (DNA topoisomerase II alpha, TOP2A)、极光激酶 B (aurora kinases B, AURKB)、细胞周期蛋白 B2 (cyclin B2, CCNB2)、人类异常纺锤体样小头畸形相关蛋白 (abnormal spindle-like microcephaly-associated protein, ASPM)、驱动蛋白超家族成员 11 (kinesin family member 11, KIF11)、有丝分裂检查点丝氨酸/苏氨酸激酶 (BUB1 mitotic checkpoint serine/threonine kinase, BUB1), 见图 2。



颜色深浅代表 degree 值大小

The shade of color represents the degree value

图 2 关键基因网络图

Fig. 2 Network map of hub genes

### 2.3 关键基因在泛癌研究中表达情况

筛选出从TCGA数据库中下载的关键基因表达量，并删除正常人的样本数据，进行基因表达的泛癌研究，绘制箱式图将数据可视化，见图3。其中横坐标为基因名称，纵坐标为矫正后基因表达数据（fragments per kilobase of exon model per million mapped fragments, FPKM）。由图3中可知，除 *BUB1*、*ASPM* 基因表达量较其余基因低外，其余8种基因在泛癌研究中均为高表达基因。

### 2.4 基因相关性分析

将关键基因进行相关性分析，如图4所示。图中横纵坐标均为关键基因的名称，上半部分是通过圆圈的大小和颜色深浅代表相关系数，蓝色越深，代表在泛癌研究中2个基因之间正相关性越强，最

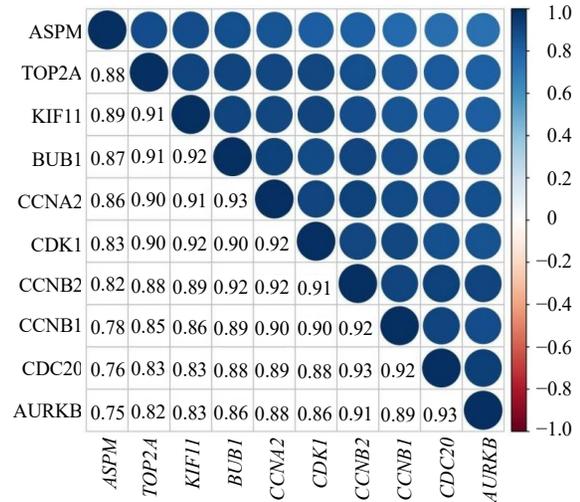


图4 关键基因相关性分析

Fig. 4 Correlation analysis of hub genes

高为1，红色越深则代表负相关越强，最低为-1；下半部分则直接通过数值代表相关系数。由图4可知，10种关键基因的相关性均较强，最高相关系数为 *CCNA2* 与 *BUB1*、*CDC20* 与 *CCNB1*、*CDC20* 与 *AURKB*，其相关系数均为0.93；最低为 *AURKB* 与 *ASPM*，其相关系数为0.75。

### 2.5 关键基因与肝癌临床分期

分析10种关键基因的肝癌临床分期与表达量的关系，其中基因名称旁的星号数反映P值大小，无星则代表该基因在对应肿瘤临床分期中无差异表达，结果见图5。从图5中可以看出，随着临床分期的进展，其基因表达量呈现上升趋势，而图中每

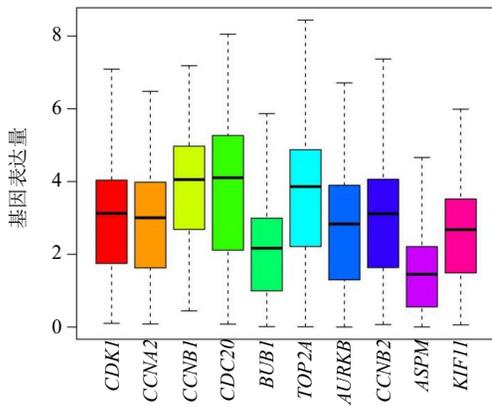


图3 关键基因在泛癌组织中的基因表达情况

Fig. 3 Expression of hub genes in pan-cancer tissues

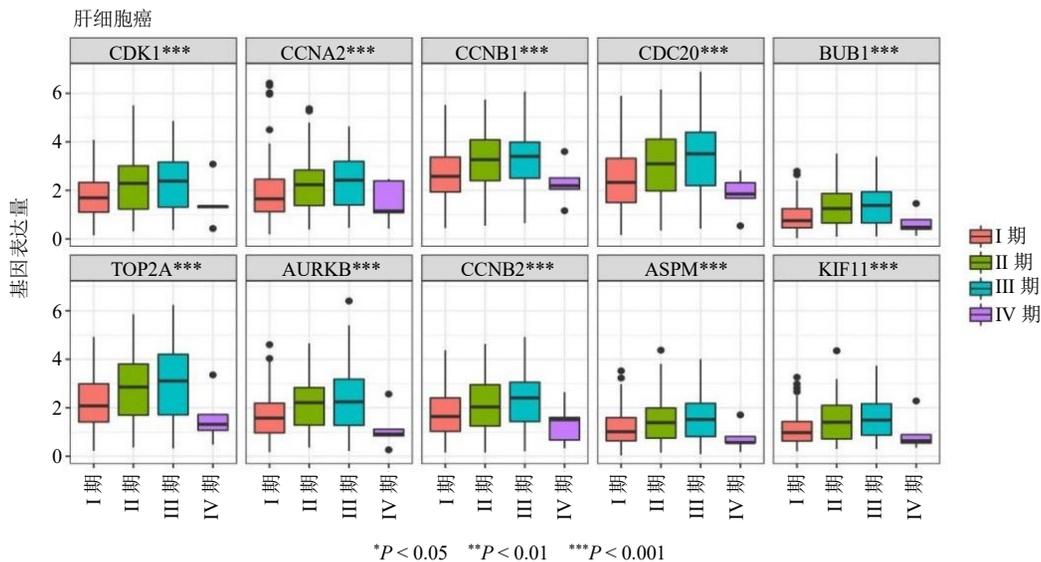


图5 关键基因表达与肝癌临床分期箱式图

Fig. 5 Box figure of hub gene expression and clinical staging of hepatocellular carcinoma

个基因 IV 分期均较前几期低，其原因可能是由于癌症进展导致患者死亡，出现样本量低导致的。

### 2.6 核心基因在肿瘤及正常组织表达量差异分析

选取 5 个在“2.3”中明显高表达的基因 *CCNB1*、*TOP2A*、*CCNA2*、*CDK1*、*CDC20* 称为核心基因，进一步进行泛癌研究。为排除样本量的影响，挑选正常样品数目  $\geq 5$  的肿瘤进行后续分析，如图 6 所示，图中横坐标为肿瘤名称简写，纵坐标为对应基因的表达量，每个肿瘤箱式图中蓝色代表正常组织，红色代表肿瘤组织，图片上方星号意义与图 5 相同，均反映 *P* 值大小。由图 6 可知，除 *CCNB1* 在肾嫌色细胞癌 (kidney chromophobe, KICH)、甲状腺癌 (thyroid carcinoma, THCA) 中，*CDC20* 在 KICH 中，*CDK1* 在肾透明细胞癌 (kidney renal clear cell carcinoma, KIRC) 中表达无差异，其余基因在肿瘤组织与正常组织中均有表达差异。

### 2.7 核心基因与肝癌临床生存分析

将 5 个核心基因进行肝癌临床 KM (Kaplan-Meier) 生存分析，如图 7 所示，红色线条代表基因

高表达的生存情况，蓝色线条代表低表达的生存情况，曲线左下角的  $P < 0.05$  说明基因的表达与生存相关。下方图中纵坐标是根据基因表达的高低，分为高、低表达组，数值代表随时间变化存活的患者例数，为风险数字表 (number at risk table)。由图 7 可知，核心基因 *P* 值均小于 0.05，并且红色线条代表的高表达组生存率均低于低表达组，说明核心基因的高表达与肝癌的生存时间缩短有明显相关性。

### 2.8 核心基因表达与肝癌组织中免疫细胞浸润相关性

将前述 5 种基因提交至 TIMER (tumor immune estimation resource) 数据库，肿瘤类型选择 LIHC，免疫浸润细胞为 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 6 种，可观察到 5 种核心基因与肿瘤纯度以及 6 种免疫细胞的散点图，如图 8 所示。图 8-A~E 分别为 *CCNB1*、*TOP2A*、*CCNA2*、*CDK1*、*CDC20* 在肝癌组织中免疫细胞浸润的相关情况。据图中 *P* 值可知，5 种核心基因与免疫细胞浸润均显著相关，其中部分细胞

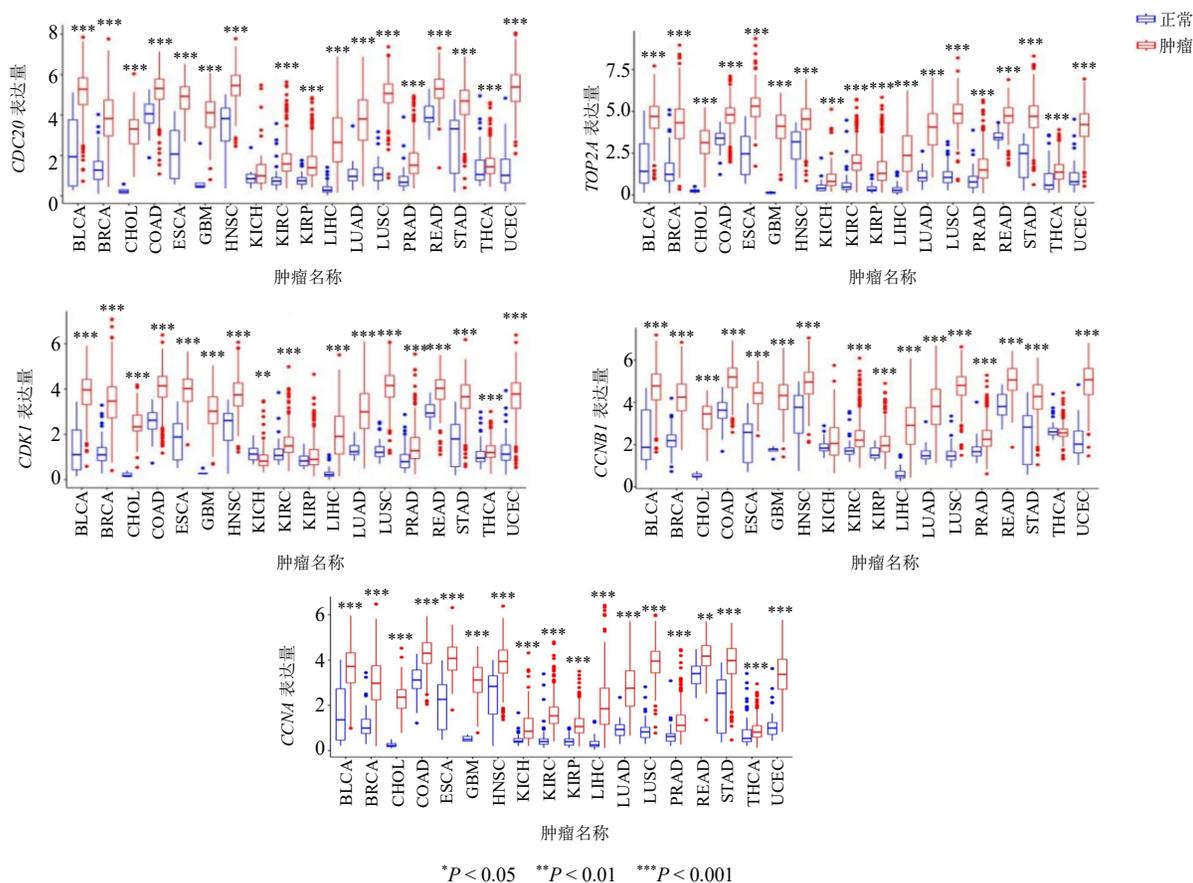


图 6 核心基因在泛癌组织以及正常组织中的表达差异

Fig. 6 Differences in expression of core genes between normal and ubiquitous tissues

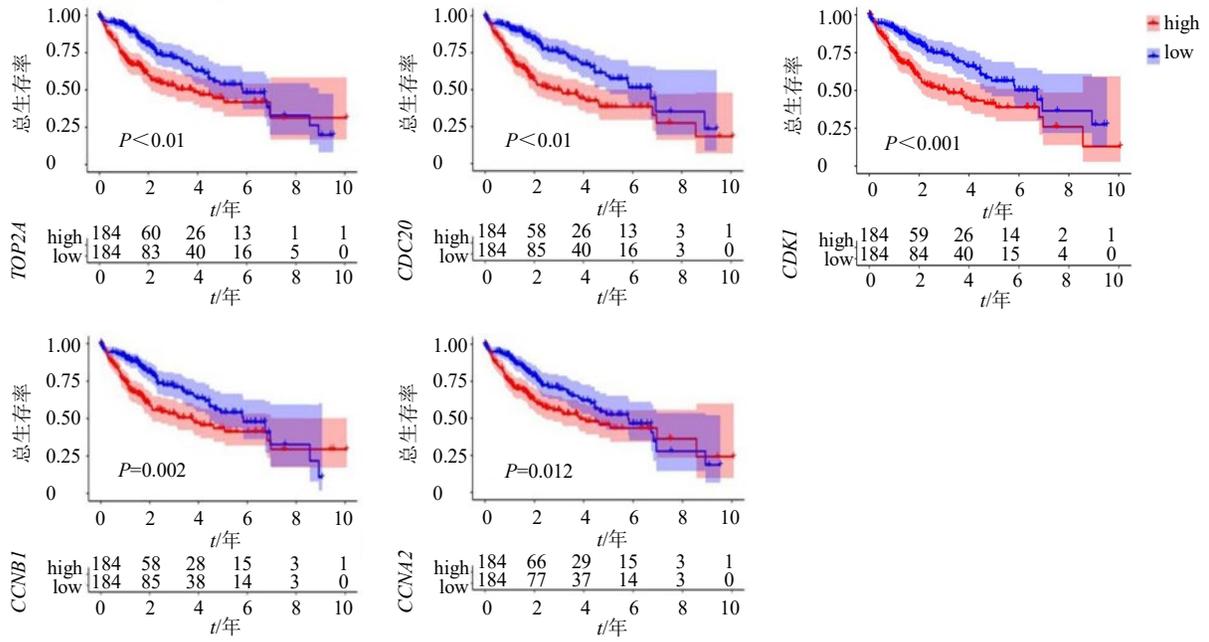


图7 核心基因与肝癌的生存分析

Fig. 7 Analysis of core genes and survival analysis of hepatocellular carcinoma

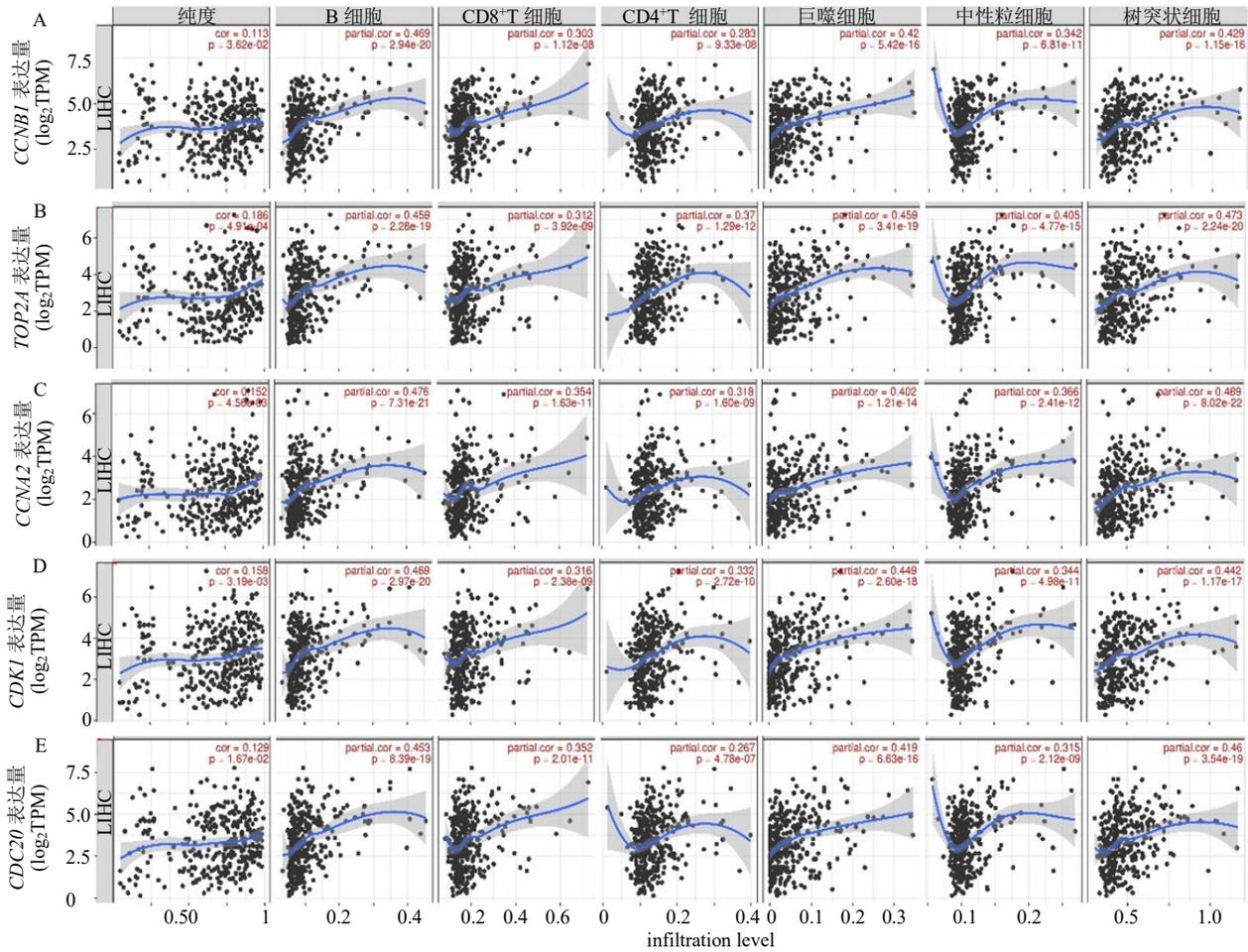


图8 核心基因与肝癌组织免疫细胞浸润相关性

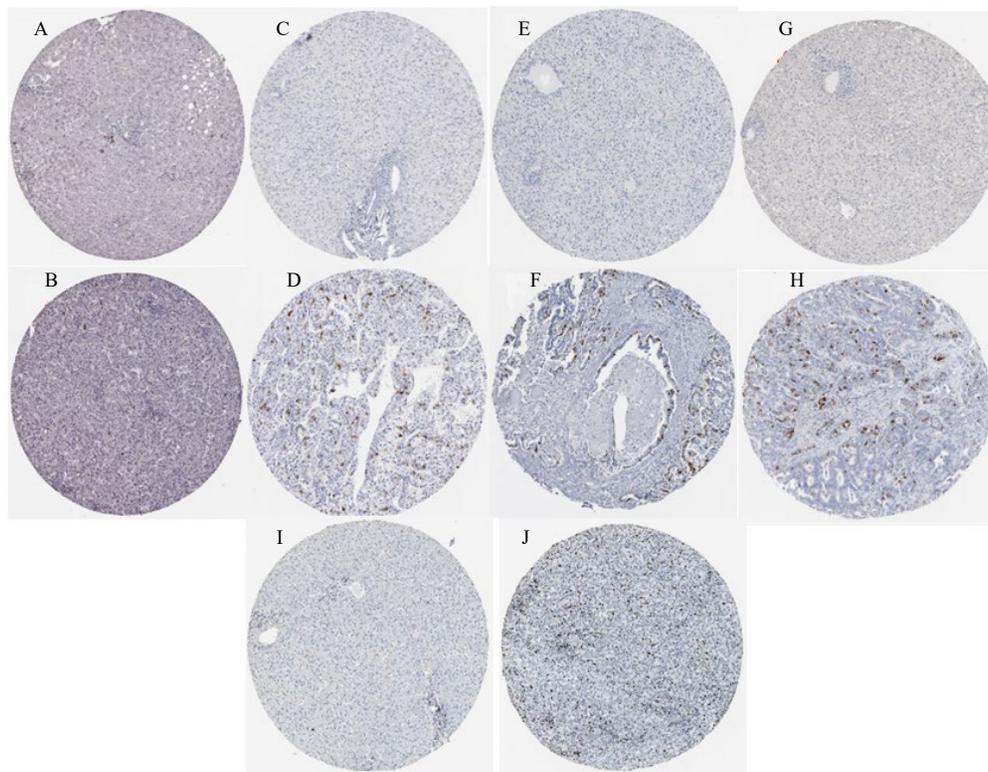
Fig. 8 Correlation between core genes and immune cell infiltration in liver cancer tissue

曲线随肿瘤纯度先上升后下降，其原因可能为肿瘤后期人体机能下降，免疫耗竭所致。

## 2.9 核心基因病理学切片研究

将 5 种核心基因提交至 THPA (human protein

atlas) 数据库，分别选择正常肝组织以及肝癌选项，得到正常肝组织及相应基因突变后肝癌组织图片，可观察到 5 种基因过表达后组织形态与正常肝组织有明显差异，如图 9 所示。



A、C、E、G、I 分别为 *CCNA2*、*CCNB1*、*CDC20*、*CDK1*、*TOP2A* 正常表达下的肝细胞组织，B、D、F、H、J 分别为 *CCNA2*、*CCNB1*、*CDC20*、*CDK1*、*TOP2A* 异常表达下的肝癌组织

A, C, E, G, and I were liver cells with normal expression of *CCNA2*, *CCNB1*, *CDC20*, *CDK1*, and *TOP2A*, respectively; B, D, F, H, and J were liver cancer tissues with abnormal expression of *CCNA2*, *CCNB1*, *CDC20*, *CDK1*, and *TOP2A*, respectively

图 9 正常组织与肝癌组织对比

Fig. 9 Comparison of normal tissue and liver cancer tissue

## 2.10 潜在中药预测

Coremine 数据库由 PubGene 公司建立，旨在为从事医学和生物学信息人士使用。该库使用网络形式展示生物医学学术语之间的关系，其中包括基因、药物、中药和化学成分等，因此可通过本数据库逆向查找基因对应的中药。将 *CDK1*、*CCNA2*、*CCNB1*、*CDC20*、*TOP2A* 5 个基因导入 Coremine 数据库，查找对应的中药，规范中药名称后进行整理，共有 137 种中药；统计中药出现次数在中位数 (4) 以上的中药 (称为高频中药)，分别有蚕砂、丹参、枳壳、枳实、土鳖虫、灵芝、人参、人参芦、人参叶、人参花、香橼、干姜、甘草、郁金、姜黄、生姜皮、生姜、天名精、茶树根、原蚕蛾、蜂蜜。在 TCMSP、BATMAN-TCM 等网站中查找以上中药的有效成分及对应靶点，

删除多个数据库中有效成分不同的中药，使用 Cytoscape 制作中药有效成分-核心基因网络图，如图 10 所示，正六边形为核心基因，其余为中药有效成分，其部分对应成分见表 2。对应的核心中药为丹参、枳实、枳壳、人参叶、人参芦、甘草、人参、蜂蜜。

## 2.11 KEGG 通路富集分析

以关键词 “liver cancer” 检索 GeneCards、PharmGkb、DrugBank、OMIM (online mendelian inheritance in man)、TTD (therapeutic target database) 5 个数据库，得到肝癌疾病靶点，将高频中药有效成分的目标基因靶点与得到的肝癌靶点基因取交集，得到交集基因，再将得到的基因进行 KEGG 通路富集分析 (hsa04110)，如图 11 所示。由图 11 可知，中药治疗肝癌除核心基因外，仍通过通路中的

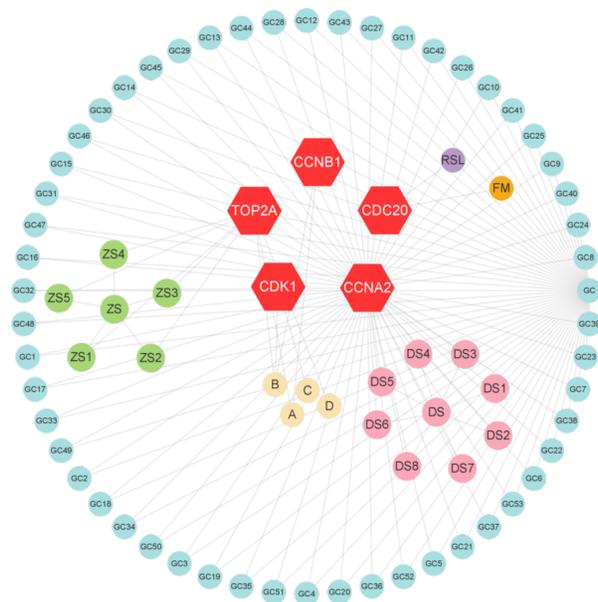


图 10 中药有效成分-核心基因网络图

Fig. 10 Network diagram of active components in traditional Chinese medicine-core gene

表 2 部分中药有效成分及靶点

Table 2 Some effective components and targets of traditional Chinese medicines

代号	药物	靶点	有效成分
DS1	丹参	CCNA2	2-isopropyl-8-methylphenanthrene-3,4-dione
DS3	丹参	CCNA2	dan-shexinkum d
DS5	丹参	CCNA2	isotanshinone II
DS6	丹参	CCNA2	miltionone I
DS7	丹参	CCNA2	neocryptotanshinone ii
FM	蜂蜜	CDC20	vitamin K
A	枳实-丹参	CCNB1	luteolin
A	枳实-丹参	TOP2A	luteolin
B	人参叶-甘草	CDK1	quercetin
B	人参叶-甘草	CCNB1	quercetin
B	人参叶-甘草	TOP2A	quercetin
C	人参叶-人参-甘草	CDK1	kaempferol
D	枳实-枳壳	TOP2A	nobiletin
RSL	人参芦	CDC20	pyrrolidine-2,5-dione
ZS1	枳实	TOP2A	sinensetin
ZS3	枳实	TOP2A	tetramethoxyluteolin
ZS4	枳实	TOP2A	isosinensetin
ZS5	枳实	TOP2A	isosakuranetin-7-rutinoside
GC1	甘草	CCNA2	isorhamnetin
GC2	甘草	CCNA2	formononetin
GC3	甘草	CCNA2	calycosin
GC4	甘草	CCNA2	licochalcone a
GC5	甘草	CCNA2	vestitol
GC6	甘草	CCNA2	glycyrol
GC9	甘草	CCNA2	7-methoxy-2-methyl isoflavone
GC15	甘草	CCNA2	kanzonols W
GC30	甘草	CCNA2	licoisoflavanone
GC41	甘草	CCNA2	1-methoxyphaseollidin
GC43	甘草	CCNA2	3'-methoxyglabridin
GC49	甘草	CCNA2	glycyrrhiza flavonol A

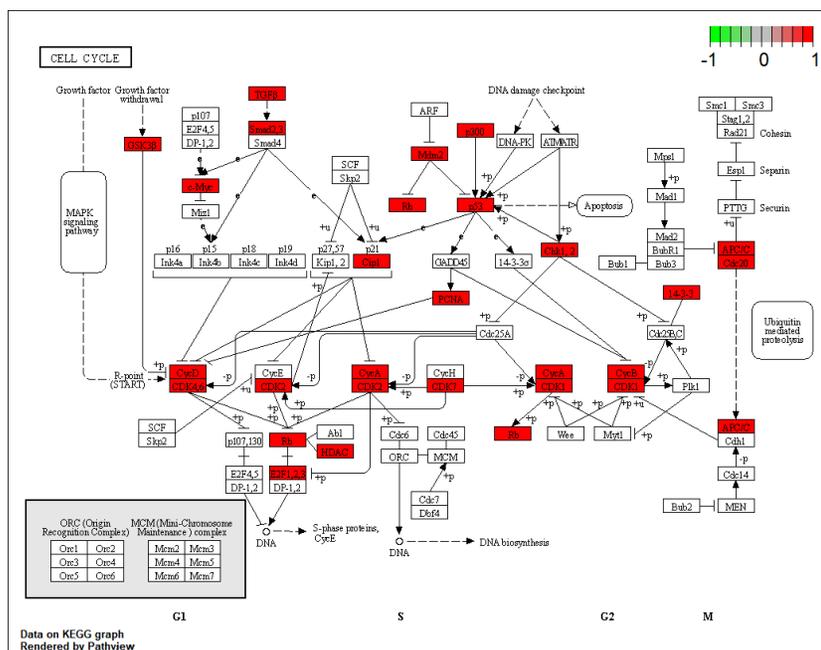


图 11 中药目标基因 KEGG 通路富集分析

Fig. 11 Enrichment analysis of KEGG pathway of target genes of traditional Chinese medicine

多种分子进行调控，如肿瘤蛋白 p53 (tumor protein p53, TP53)、β 转化生长因子 (transforming growth factor-β, TGF-β) 等。

### 3 讨论

癌症是威胁人类健康的一大因素，而在不同的国家和地区，恶性肿瘤的发病率则有很大差异，其中原发性肝癌是世界上常见的恶性肿瘤之一，也是死亡率高居前列的肿瘤，据估计，每年大约有 84 万例 HCC 新发病例，至少有 78 万人死于 HCC<sup>[6]</sup>，目前对于肝癌的治疗方法通常为以下 5 种：手术切除、肝移植、序贯射频消融 (radiofrequency ablation, RFA)、肝动脉化疗栓塞 (transcatheter arterial chemoembolization, TACE) 和靶向治疗、免疫治疗。然而由于早期肝癌的症状不明显，因此肝癌难以在早期发现，治疗时往往难以实施手术根除，而 RFA 和 TACE 则可提升中期患者的生存率，但效果仍不尽人意。索拉非尼是一种多激酶抑制剂，能够阻断多达 40 个激酶，进而延长晚期肝癌患者的生存时间，将晚期肝细胞癌患者的存活率从 7.9 个月大幅提高到 10.7 个月<sup>[4]</sup>，然而其不良反应如手足皮肤反应、腹泻、肝功能异常、高血压、皮疹等，以及其开始治疗 6 个月后即可出现耐药性，故而限制了其对肝癌的临床疗效<sup>[7]</sup>。

肝癌处于复杂的免疫微环境中，其肿瘤微环境往往可以导致免疫细胞的功能抑制，同样也是癌症

发生过程中的重要因素<sup>[8]</sup>，肿瘤通过多种方式促进免疫耐受的形成，如通过细胞成分、免疫检查点受体、抑制性细胞因子等方法损害免疫细胞的功能，而由于肿瘤的免疫微环境错综复杂，故而单抗治疗的缓解率并不理想<sup>[9]</sup>，因此多靶点治疗则是未来的可能方向之一。中医药能够预防肝癌发生，抑制细胞增殖、上皮间质转化，阻滞细胞周期，诱导肿瘤细胞自噬及衰老，抑制肿瘤细胞转移，逆转耐药性，调节机体免疫功能等<sup>[10-11]</sup>，同时其具有低毒性、安全、多靶点作用等优点。

肝癌属祖国医学“癥瘕”“积聚”等病名范畴<sup>[12]</sup>。因其治疗难度大，故而历代医家对其防治较为重视，现代中医学对肝癌的研究又有长足进步。HCC 病因包括内、外 2 个方面：内因主要是肝脾虚损，外因主要是湿热、瘀血、毒邪留滞肝胆脾胃，因此病机为正虚邪实，正虚多在肝脾，邪实多为湿、热、瘀、毒。多由脾不化湿或外湿内侵，湿聚成痰，蕴生热毒，结于肝胆；或肝气郁结，气滞血瘀，继蕴热毒，或湿热毒邪直中肝胆；而湿盛易侮肝、肝胆气滞血瘀则可导致肝不疏土或肝胆湿热；其病机和湿、痰、热、瘀、毒等病因又互为因果，从而加剧病情的发展，诸邪互结于肝而形成癌肿。而癌毒则容易累及先天之本，加之现代医学的放化疗副作用消耗身体，日久易伤阴、伤肝，肝肾同源，同属下焦，因而造成肝肾阴虚<sup>[13]</sup>，消耗人体正气进而导致病情恶化。

现代中医认为,“癌毒”病位辨证分别为肾、血、脾、气、肺、肝、心<sup>[14]</sup>,因此扶正固本、通瘀化湿是一种可行的治疗方法。然而中医药治疗的缺点则是效果不稳定、涉及的通路及分子并不明确等。

本研究使用生物信息学技术,得到肝癌上调基因 2565 个,下调基因 253 个,并进一步筛选出 10 种关键基因: *CDK1*、*CCNA2*、*CCNB1*、*CDC20*、*TOP2A*、*AURKB*、*CCNB2*、*ASPM*、*KIF11*、*BUB1*, 10 种关键基因在泛癌组织中均为高表达,其相关性也均较强,并在肝癌中关键基因表达量与临床分期阶段大部分显著相关。本研究筛选出的关键基因有望成为肝癌新的生物标志物,为肝癌以及多种癌症的早期诊断提供新思路。挑选出 5 种关键基因 *CDK1*、*CCNA2*、*CCNB1*、*CDC20*、*TOP2A* 进行泛癌分析,可知在多种肿瘤中,关键基因在肿瘤组织与正常组织中表达均有明显差异。且核心基因表达与肝癌生存分析、免疫细胞浸润、病理学结果均有明显相关。

中药有效成分能够调节一个通路中的多个靶点或同时调控多种通路。如丹参中的丹参酮II<sub>A</sub>可以在不同浓度梯度下,分别通过不同的作用机制,如影响基质金属蛋白酶 2 (matrix metalloproteinase 2, MMP-2) 和 MMP-9 的表达、核因子- $\kappa$ B (nuclear factor- $\kappa$ B, NF- $\kappa$ B) 的表达,以及对 Hippo 通路的调节作用治疗肝癌<sup>[15]</sup>。通过 KEGG 通路分析可知, *CDK1*、*CCNA2*、*CCNB1* 为中药有效成分调控肝癌的关键分子,研究表明, *CDK1*、*CCNA2*、*CCNB1* 为调控细胞周期的关键蛋白,其中 S/G<sub>2</sub> 期调控蛋白为细胞周期蛋白 A (cyclin A), G<sub>2</sub>/M 期调控蛋白为细胞周期蛋白 B (cyclin B), 而 *CDK1* 往往与细胞周期蛋白配对后驱动有丝分裂<sup>[16]</sup>。而 p53 信号转导途径、LINC00346-miR-199a-3p-*CDK1*/Cyclin B、SNHG16/let-7b-5p/*CDC25B*/*CDK1* 等信号通路可影响 *CDK1* 的表达,诱发肝细胞癌<sup>[17]</sup>。作为细胞周期的关键调控因子之一,针对 *CDK1* 的中药活性成分是当前研究方向之一,如从香茶菜中提取出的蓝萼甲素对于前列腺癌 DU145 细胞的细胞周期阻滞在 G<sub>2</sub>/M 期,其可能的机制为调控细胞周期蛋白分子 p21、cyclin B1、*CDK1*<sup>[18]</sup>。除此之外,中药的活性成分,如生物碱、多糖、皂苷等均有调控分子通路以及治疗癌症的作用<sup>[19-20]</sup>,尚可调控非编码 RNA,如 lncRNA、microRNA 等,影响转录组基因表达,进而影响肝癌的发生发展<sup>[21-22]</sup>。在我国,

由于中医药治疗肝癌的成本相对较低,故而一些患者仍偏向选择中医药疗法。研究表明,无论患者处于哪一肿瘤阶段,或肝功能情况,均可从中医药治疗中受益<sup>[23]</sup>。因此中医药治疗肿瘤具有巨大的潜力。

#### 4 问题与展望

本研究通过 TCGA 数据库筛选出了肝癌的差异基因,并挑选出关键基因进行泛癌研究,进一步选出核心基因进行中药预测及 KEGG 通路富集分析,然而存在以下不足:(1) 本研究的数据库来源为单库,后续可联合其他数据库进行深入研究。(2) 因不同数据库对于中药的有效成分并不统一,因此本研究删减了一些药物,后续的研究可进一步通过实验验证删减药物的有效成分是否有作用。(3) 本研究的高频药物从中医理论来看以理气活血、健脾扶正为主,而补肾药物则未出现,而晚期肝癌患者常出现的久病及肾则未有明显体现,后续仍可从中医理论中的“肾”入手,进行分子通路的研究。

**利益冲突** 所有作者均声明不存在利益冲突

#### 参考文献

- [1] Hulvat M C. Cancer incidence and trends [J]. *Surg Clin North Am*, 2020, 100(3): 469-481.
- [2] Zhu R X, Seto W K, Lai C L, et al. Epidemiology of hepatocellular carcinoma in the Asia-Pacific region [J]. *Gut Liver*, 2016, 10(3): 332-339.
- [3] Cocker F, Chien Yee K, Palmer A J, et al. Increasing incidence and mortality related to liver cancer in Australia: Time to turn the tide [J]. *Aust N Z J Public Heal*, 2019, 43(3): 267-273.
- [4] Forner A, Reig M, Bruix J. Hepatocellular carcinoma [J]. *Lancet*, 2018, 391(10127): 1301-1314.
- [5] 苏胜, 黄晓武. 肝癌细胞免疫治疗的研究进展 [J]. *中华肝脏病杂志*, 2020, 28(6): 461-465.
- [6] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. *CA Cancer J Clin*, 2018, 68(6): 394-424.
- [7] Anwanwan D, Singh S K, Singh S, et al. Challenges in liver cancer and possible treatment approaches [J]. *Biochim Biophys Acta Rev Cancer*, 2020, 1873(1): 188314.
- [8] Liu Y, Cheng Y, Xu Y, et al. Increased expression of programmed cell death protein 1 on NK cells inhibits NK-cell-mediated anti-tumor function and indicates poor prognosis in digestive cancers [J]. *Oncogene*, 2017, 36(44): 6143-6153.
- [9] 张锋, 殷欣. 肝癌免疫耐受机制研究进展 [J]. *复旦学*

- 报: 医学版, 2020, 47(2): 280-287.
- [10] Hu B, An H M, Wang S S, *et al.* Preventive and therapeutic effects of Chinese herbal compounds against hepatocellular carcinoma [J]. *Molecules*, 2016, 21(2): 142.
- [11] 冯晓灵, 蒋亚生, 李柏. 肝癌局部注射丝裂霉素与中药“肝复康”的临床及实验对照分析 [J]. *肿瘤*, 1994, 14(5): 291-292.
- [12] 谢晶日, 赵刚, 梁国英. 原发性肝癌的中医药现代研究进展 [J]. *中国医药导报*, 2008, 5(7): 17-18.
- [13] 旋静. 中医治疗原发性肝癌的研究进展 [J]. *中国医药指南*, 2015, 13(13): 45-46.
- [14] 刘声, 王笑民, 杨霖, 等. 浅析中医“癌毒”本质 [J]. *中医学报*, 2018, 33(10): 1827-1830.
- [15] Li Z B, Zou J, Cao D, *et al.* Pharmacological basis of tanshinone and new insights into tanshinone as a multitarget natural product for multifaceted diseases [J]. *Biomed Pharmacother*, 2020, 130: 110599.
- [16] Crncec A, Hochegger H. Triggering mitosis [J]. *FEBS Lett*, 2019, 593(20): 2868-2888.
- [17] 叶凯丽, 郑雯, 叶启发, 等. CDK1 参与肝细胞癌的发展机制及其抑制剂应用价值 [J]. *中国临床药理学与治疗学*, 2021, 26(9): 1086-1094.
- [18] 朱琳琳, 徐祉轩, 张明明. 蓝萼甲素通过 cyclinB1/CDK1 途径诱导 DU145 细胞 G2/M 期阻滞 [J]. *中国免疫学杂志*, 2021, 37(18): 2235-2239.
- [19] Liu C Y, Yang S S, Wang K L, *et al.* Alkaloids from traditional Chinese medicine against hepatocellular carcinoma [J]. *Biomed Pharmacother*, 2019, 120: 109543.
- [20] Tang K Y, Du S L, Wang Q L, *et al.* Traditional Chinese medicine targeting cancer stem cells as an alternative treatment for hepatocellular carcinoma [J]. *J Integr Med*, 2020, 18(3): 196-202.
- [21] Chen C Y, Chen C C, Shieh T M, *et al.* Corylin suppresses hepatocellular carcinoma progression via the inhibition of epithelial-mesenchymal transition, mediated by long noncoding RNA GAS5 [J]. *Int J Mol Sci*, 2018, 19(2): 380.
- [22] Wang Y P, Fu X Q, Yin C L, *et al.* A traditional Chinese medicine formula inhibits tumor growth in mice and regulates the miR-34b/c-Met/ $\beta$ -catenin pathway [J]. *J Ethnopharmacol*, 2020, 260: 113065.
- [23] LIU X, LI M, WANG X, *et al.* Effects of adjuvant traditional Chinese medicine therapy on long-term survival in patients with hepatocellular carcinoma [J]. *Phytomedicine*, 2019, 62: 152930.

[责任编辑 潘明佳]