

DOI:10.12113/202108002

# 基于 SNP 共表达网络肝癌分子分型及预后分析

张思嘉,蔡挺,张顺\*

(中国科学院大学 宁波华美医院医学实验部,浙江 宁波 315100)

**摘要:**基于 SNP 突变数据与 mRNA 表达谱关联分析,构建一种肝癌分子分型方法并对不同分型预后的差异,并对不同分型肝癌的发生发展机制进一步研究。首先通过 TCGA 数据库收集 359 例肝细胞癌患者的 SNP 突变数据和 mRNA 表达数据,采用 Wilcoxon 秩和检验,筛选突变后差异表达基因,并通过生物信息学工具 String 和 Cytoscape 构建差异表达基因的蛋白互作网络,筛选连接度最高的 10 个 Hub 基因。利用 Consensus Cluster Plus 软件包,基于 Hub 基因 mRNA 表达水平构建 NMF 分子分型模型,再结合生存数据评估各分型患者的预后。最后利用加权基因共表达网络分析(WGCNA),识别与肝癌分子分型相关的模块,并针对关键模块的基因进行通路富集,从而对不同分型肝癌的基因表达谱进行比较。结果:NMF 模型将肝癌分为高危、低危 2 个分型,其中 *CDKN2A* 和 *FOXO1* 基因对分型贡献度高。生存分析显示低危组患者的生存情况显著优于高危组,高危组富集多个与肿瘤细胞侵蚀、转移、复发过程相关的信号通路,低危组则与细胞周期和胰液分泌相关。本研究在无先验性信息的前提下,基于突变后显著差异表达的 Hub 基因表达水平构建的肝癌分子分型对肝癌患者预后评估具有一定的指导意义,其中 *CDKN2A* 和 *FOXO1* 突变是肝癌患者的不良预后因素,针对二者的靶向药研发,可能为肝癌患者提供新的治疗策略。

**关键词:**肝癌;分子分型;共表达网络;SNP

中图分类号:Q42 文献标志码:A 文章编号:1672-5565(2022)04-247-10

## Molecular typing of HCC based on SNP co-expression network and prognosis analysis

ZHANG Sijia, CAI Ting, ZHANG Shun\*

(Department of Experimental Medical Science, Huamei Hospital, University of Chinese Academy of Sciences, Ningbo 315100, Zhejiang, China)

**Abstract:** A molecular classification method of hepatic carcinoma was constructed based on the correlation analysis between single nucleotide polymorphism(SNP) mutation data and mRNA expression profile, and the difference of prognosis between different subtypes was compared. The occurrence and development mechanism of different molecular classifications of hepatocellular carcinoma(HCC) were investigated. The SNP mutation data and mRNA expression profile data of 359 HCC patients were collected from the TCGA database, and differentially expressed genes after mutations were screened by the Wilcoxon rank sum test, Then, two bioinformatics tools (String, Cytoscape) were applied to construct protein-protein interaction (PPI) network and identify the top ten hub genes with high centrality degree. The non negative matrix factorization(NMF) molecular classification model was built with Consensus Cluster Plus package based on hub gene mRNA expression profile, and the prognosis of different molecular subgroups of HCC patients was assessed combined with survival data. Finally, weighted gene co-expression network analysis (WGCNA) was used to identify co-expression gene modules associated with HCC molecular subtypes. Kyoto Encyclopedia of Genes and Genomes(KEGG) enrichment analysis was performed for the comparison of gene expression profiles of HCC patients from different molecular subtypes. Results showed that according to the molecular classification based on NMF, HCC was divided into two subgroups, high-risk and low-risk, and *CDKN2A* and *FOXO1* were two major contributing genes for the subgroups. Survival analysis showed that

收稿日期:2021-08-03 修回日期:2021-11-01.

基金项目:华美研究基金项目(No.2019HMKY60);浙江省消化系统肿瘤诊治及研究重点实验室基金项目(No.2019E10020);宁波市消化系统肿瘤临床医学研究中心基金项目(No.2019A21003).

作者简介:张思嘉,女,研究方向:消化系统肿瘤、生物信息学. E-mail:zs13616hm@163.com.

\* 通信作者:张顺,男,主任技师,研究方向:生物治疗、分子生物学. E-mail:zhangshun@ucas.ac.cn.

the overall survival rate of low-risk patients was significantly better than that of high-risk group. High-risk subgroup was enriched in several pathways related to tumor invasion, metastasis, and recurrence. While low-risk subgroup was enriched in cell cycle and pancreatic secretion. Without any prior information, the molecular subtypes of HCC based on differentially expressed hub gene expression after mutation can be used for treatment decision and prognostic prediction. In this research, mutations of *CDKN2A* and *FOXO1* are considered as two adverse prognostic factors for HCC patients, and targeted drug development for these two genes may provide new therapeutic strategies for HCC patients.

**Keywords:** Hepatic carcinoma; Molecular classification; Co-expression network; SNP

肝癌是目前我国第4位常见恶性肿瘤及第2位肿瘤致死病因,全世界每年近一半的新发和死亡病例发生在中国<sup>[1]</sup>。目前原发性肝癌的治疗方法虽然很多,但肿瘤治疗的预后效果仍不理想;HCC首选方法是手术切除,但由于其起病隐蔽,确诊时患者多已中晚期,失去手术切除机会<sup>[2]</sup>。而传统的放疗等手段预后较差,不良反应较多,这使目前我国HCC的临床治疗充满挑战,急需新的治疗策略来改善肝癌患者的生存质量。

肝癌发生是多因素、多步骤和受多种机制调控的复杂过程,具有高度异质性。目前针对肝癌的大体分型和组织病理学分型是制定临床治疗方案,预测判断患者的预后与转归的重要依据。但实践表明许多具有相同类型和相同分期的肝癌的患者应用相似的临床治疗手段,其预后的差别很大,这与肝癌极其复杂的肿瘤异质性密切相关。因此迫切需要新的分期分型指标以助力肝癌精准诊断与治疗,以提高患者生存率。随着二代测序(Next-generation Sequencing, NGS)技术,基因组、转录组、单细胞测序等多组学的发展,这些技术逐渐被应用于肿瘤发病机制的研究以及分子分型的判断中,并为诊治手段及预后分析提供重要信息<sup>[3-4]</sup>。而基于肿瘤分子异质性的个体化精准诊疗已成为未来恶性肿瘤诊疗的发展方向;该诊疗方式将传统的肝癌临床病理分型和分期与肝癌的分子表型相结合,对肝癌患者进行更加细致的亚群划分,并在这一基础上将手术、放疗、化疗、免疫治疗、分子靶向治疗等手段,按照患者的个体化特征定制精准治疗方案,从而大幅度提高治疗手段的针对性和治疗效果<sup>[5-6]</sup>。

单核苷酸多态性(SNPs)是指基因组DNA序列中由于单个核苷酸替换而引起的多态性,是最普遍的遗传变异形式。绝大多数SNP并不影响蛋白序列,而是通过对基因表达的调控对生物个体产生影响<sup>[7]</sup>。越来越多的临床研究证实,通过监测肿瘤患者中生物标志物的SNP突变、mRNA及蛋白的表达水平的异常变化,来评价预后并指导临床个体化治疗,可提高疗效,减轻不良反应,促进医疗资源的

合理利用<sup>[8-10]</sup>。本研究利用肝癌患者的转录组和SNP突变数据,对样本进行分子分型并分析不同分型间的生物学差异,有利于进一步认识肝癌发生和进展的过程,并对肝癌的临床诊断和治疗选择以及预后预测具有一定的参考价值。

## 1 材料和方法

### 1.1 数据来源

本研究涉及的359例肝细胞癌患者的mRNA表达数据、单核苷酸突变数据及临床数据下载于TCGA在线公共数据库(<https://portal.gdc.cancer.gov>)。

### 1.2 分析方法

#### 1.2.1 SNP突变与mRNA表达谱联合分析

单核苷酸突变(SNP)可以通过影响基因编码和剪接影响基因表达<sup>[11]</sup>。本研究通过对肝癌患者SNP突变数据和mRNA表达谱的联合分析,筛选出因单核苷酸突变引起表达水平变化的差异基因,具体实施方式如下:首先根据SNP突变数据中基因是否发生突变,将肝癌患者划分为野生型和突变型两组,再通过wilcoxon秩和检验比较野生型和突变型肝癌患者中的表达水平,其中表达水平存在显著差异的基因作为单核苷酸突变差异表达基因(取 $P < 0.01$ 为差异具有统计学意义)。最后运用KEGG数据库通路富集和基因功能注释的信息,对单核苷酸突变表达差异基因进行功能聚类分析和代谢途径分析,并选取与癌症相关的单核苷酸突变差异表达基因进行后续分析。

#### 1.2.2 蛋白互作网络的构建与枢纽基因的筛选

将与癌症相关的单核苷酸突变差异表达基因导入STRING数据库(<https://www.string-db.org>),得到相互作用的网络节点文件,再通过Cytoscape软件对蛋白-蛋白相互作用网络PPI(Protein-protein Interaction)进行可视化处理,同时用插件CytoHubba对PPI网络进行模块分析,选取出10个连接度最高的Hub基因作为肝癌分子分型的潜在特征和重要依据。

### 1.2.3 肝癌患者分子分型与预后分析

非负矩阵分解(Nonnegative Matrix Factorization, NMF)属于双向聚类,具有良好的可解释性和数值结果,该方法已广泛用于基因表达谱数据的癌症分类<sup>[12-14]</sup>。本研究利用 R 语言中的“Consensus ClusterPlus”分析包,根据肝癌患者 10 个 hub 基因的 mRNA 表达量构建 NMF 分子分型模型,聚类数  $k$  值取 2~8,根据聚类效果选取具有较好聚类稳定性的  $k$  值用于分子分型的划分,并进一步将样本的基因表达数据和临床预后信息整合,去除生存时间小于 30 天的样本,再利用 R 软件中的“Survival”包,对不同分子分型的肝癌患者进行 Kaplan-Meier 生存分析及 log-rank 检验,判断不同分子分型患者的生存预后是否具有显著性差异。

### 1.2.4 不同分子分型肝癌患者转录组特征对比

首先采用 R 软件的 DESeq2 软件包,对不同分子分型患者的基因转录组数据进行差异表达分析,差异基因的筛选标准为:算法矫正假发现率(FDR) < 0.05,差异倍数(Fold change) > 1。再结合样本的分子分型分组,针对差异表达基因利用 R 软件的 WGCNA 软件包进行加权共表达网络分析,具体步骤如下:1) 构建基因共表达相似性矩阵,确定软阈值后再将相似性矩阵转换为邻接矩阵。2) 将邻接矩阵转换成拓扑矩阵,并采用拓扑覆盖法对差异基因进行层次聚类,按照混合动态剪切树的方法确定基因模块,同时绘制树状图并对基因模块进行可视化,其中每个模块的最小基因数目为 30。3) 将基因模块与分子分型进行关联,寻找与 2 种分子分型最显著相关的基因模块<sup>[15]</sup>。并对与分子分型相关的基因模块中的基因利用 R 软件“Cluster Profiler”包进行 KEGG 富集分析,预测与 2 种肝癌分子分型相关的功能和信号通路。

## 2 结果

### 2.1 单核苷酸突变差异表达基因的筛选与蛋白互作网络的构建

通过对癌症基因组图谱(TCGA)数据库中获得肝癌患者 SNP 突变数据的统计分析,共发现 1 932 种突变率大于 5%的突变基因。根据基因是否发生突变,并通过对突变基因在突变型和野生型两组患者 mRNA 表达谱的差异分析,确定 561 种在突变前后的表达水平发生显著性变化的基因作为单核苷酸突变差异表达基因,进一步通过 Gene ontology (GO) 富集分析进行筛选,最终共得到 64 种与癌症发生发展相关的单核苷酸突变差异表达基因。将上述基因输入

STRING 网站,导入 Cytoscape 软件得到可视化 PPI 网络图,显示有 58 个节点和 220 个边(见图 1)。运用 CytoHubba 插件,计算每个基因得连接度,得分最高的前 10 个基因作为 hub 基因(见表 1)。其中大多数核心基因的 SNP 突变与多种肿瘤有关,比如单核苷酸突变引起的 KRAS 激活可导致多种恶性肿瘤,包括肺癌、粘液腺瘤、胰腺导管癌和大肠癌;SMAD4 作为一种肿瘤抑制基因,其突变失活引起的 TGF- $\beta$  信号转导紊乱导致对肿瘤细胞生长抑制作用的逃逸<sup>[16]</sup>。TSC2 是常染色体基因,在体内作为抑癌基因广泛表达<sup>[17]</sup>。STK11 突变导致的表达缺失已被证明是肠道息肉综合征(PJS)重要致病原因并与散发性的结直肠癌直接相关<sup>[18]</sup>。TP53 基因(编码 p53 蛋白)作为一个重要的抑癌基因,通过调控一系列信号转导通路广泛参与了多种恶性肿瘤的发生发展,TP53 突变在肝癌中有着较为明显的特征<sup>[19]</sup>。CTNNB1 是 Wnt 通路的关键成员,参与细胞间黏附以及细胞间信号传递,与肿瘤的形成和浸润,转移密不可分<sup>[20]</sup>。mTOR 控制蛋白质合成、细胞生长和增殖,是 PI3K-Akt-mTOR 信号通路的核心基因<sup>[21]</sup>。对 10 种 hub 基因进行进一步生存分析,发现共有 7 种 hub 基因与肝癌患者生存预后显著相关。其中高表达的 CDKN2A、KARS 是预后的危险因素,低表达的 FOXO1、CDKN1A、MTOR、STK11、TP53 是预后的危险因素。由此可见 Hub 基因的表达异常与肝癌患者的预后关系密切,这些 Hub 基因有望成为肝癌患者早期诊断、治疗及预后判断的重要靶点。

表 1 蛋白质互作网络中连接度排名前 10 核心基因及生存分析结果<sup>4</sup>

Table 1 Ten hub genes in PPI network and survival analysis (top ten in connectivity)

基因 <sup>1</sup>	连接度	风险率(HR 值) <sup>2</sup>	P 值 <sup>3</sup>
FOXO1	31	0.36	2.30×10 <sup>-5</sup>
CDKN2A	13	1.64	5.30×10 <sup>-3</sup>
CDKN1A	18	0.62	8.90×10 <sup>-3</sup>
CTNNB1	15	1.17	3.70×10 <sup>-1</sup>
KARS	23	1.99	3.70×10 <sup>-2</sup>
MTOR	12	0.52	4.40×10 <sup>-2</sup>
SMAD4	17	0.78	1.80×10 <sup>-1</sup>
STK11	25	0.54	6.50×10 <sup>-2</sup>
TP53	18	0.65	2.90×10 <sup>-2</sup>
TSC2	14	0.79	2.10×10 <sup>-1</sup>

注:1) FOXO1: Forkhead Box O1; KRAS: KRAS Proto-Oncogene, GTPase; SMAD4: SMAD Family Member 4; CDKN1A: Cyclin Dependent Kinase Inhibitor 1A; TSC2: TSC Complex Subunit 2; STK11: Serine/Threonine Kinase 11; TP53: Tumor Protein P53; CTNNB1: beta Catenin 1; MTOR: Mechanistic Target Of Rapamycin Kinase; CDKN2A: Cyclin Dependent Kinase Inhibitor 2A; 2) HR 值即风险率,HR > 1 时表示该基因的高表达是一种危险因素,HR < 1 表示该基因的低表达是一种危险因素; 3) Log Rank 法比较两组患者的生存曲线是否有差异; 4) 10 种 Hub 基因的 Kaplan-Meier 生存曲线图见附录。



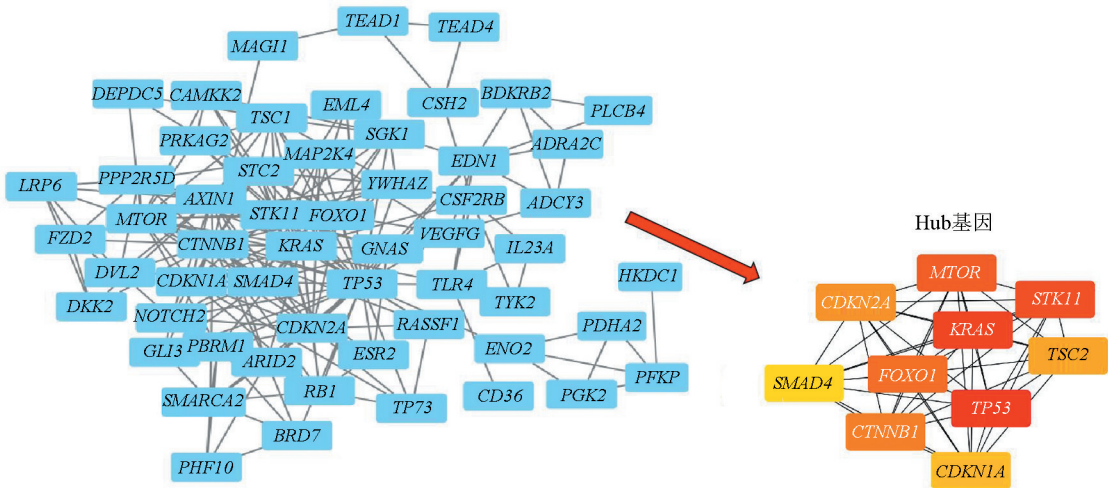


图1 突变后表达显著差异的癌症相关基因构建的 PPI 网络和 hub 基因模块

Fig.1 PPI network and hub gene module of cancer-associated genes which are significant differentially expressed after mutation

## 2.2 肝癌分子分型的建立与预后分析

基于 2.1 筛选出的 10 个 hub 基因的表达量,运用非负矩阵因子分解 (NMF) 算法对肝癌患者进行分子分型。综合判断聚类稳定性,发现当  $k=2$  时,即肝癌样本分为 2 种分子分型时,模型的稳定性较好且样本分布均匀。生存曲线 (见图 2) 以及 log-rank 检验结果显示 2 种肝癌分子分型患者的预后情况具有显著性差异,其中 cluster1 的 5 年生存率显著低于 cluster2 ( $P=0.039$ )。因此本研究按照 2 种肝癌分子分型患者生存率的高低将 cluster1 作为高危组,cluster2 作为低危组。由于 NMF 算法无法直接对因子的贡献度进行统计,为进一步寻找对肝癌分子分型的影响最为显著的 Hub 基因,本文以 Heatmap 的形式对比 10 个 Hub 基因在高危组 (Cluster1) 和低危组 (Cluster2) 的表达趋势,其中基因表达趋势与样本分子分型相关性最为明显的 Hub 基因将作为后续研究重点。突变基因 *CDKN2A* 和 *FOXO1* 表达量与样本分子分型的分布最具有明显的趋势性 (见图 3)。具体表现为 *CDKN2A* 在低危组 (Cluster2) 中的表达量明显高于高危组 (Cluster1), *FOXO1* 在高危组 (Cluster1) 中的表达量明显高于低危组 (Cluster2)。并且进一步研究发现,在不考虑分子分型对全部肝癌患者进行生存分析, *CDKN2A* 基因高表达和 *FOXO1* 基因低表达的患者生存时间显著降低 (见图 4)。这与高、低危组两种分子分型组间的生存率高低差异相一致。由此可以初步推测:由突变导致的 *CDKN2A* 与 *FOXO1* 表达量变化对肝癌样本分子分型贡献较大,其中 *CDKN2A* 为预后不良基因 (Unfavorable gene), *FOXO1* 为预后良好基因 (Favorable gene),因此以 *CDKN2A*、*FOXO1* 突变为基础的分子分型方法具有较大的研究潜力。

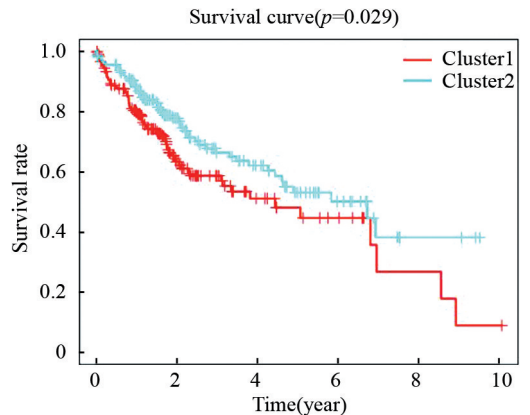


图2 2个肝细胞癌分子分型的生存曲线

Fig.2 Survival analysis of two HCC molecular typing

由以上分析表明,基于 10 种 hub 基因表达水平预测的分子分型有望成为潜在肝癌的有效预后指标。本研究通过进一步构建肝癌 1~5 年死亡概率预测列线图,比较肝癌分子分型与其它临床预后变量的关系,从而为肝癌患者的预后情况提供个体化预测。其中临床预后变量均为分类变量,包括肝癌的 TNM 分期变量、性别、年龄分层,以及治疗方法等,并按照临床预后变量的重要性由下至上进行排列,每一个预后变量均包括若干个变量取值,每一个变量取值对应分值标尺上的一个分值,总分值标尺和患者死亡概率变量均为连续性变量。患者肝癌分子分型在预测模型中的重要程度仅次于 T 分期,且优于 M、N 分期以及性别年龄分层 (见图 5)。该列线图模型的预测准确性较高,1 年生存率的 AUC 为 0.762,3 年生存率的 AUC 为 0.749,5 年生存率的 AUC 为 0.732。由此可以看出基于 10 中 hub 基因构建的肝癌分子分型方法在对肝癌患者的病情评估和预后判断具有一定的临床指导意义。

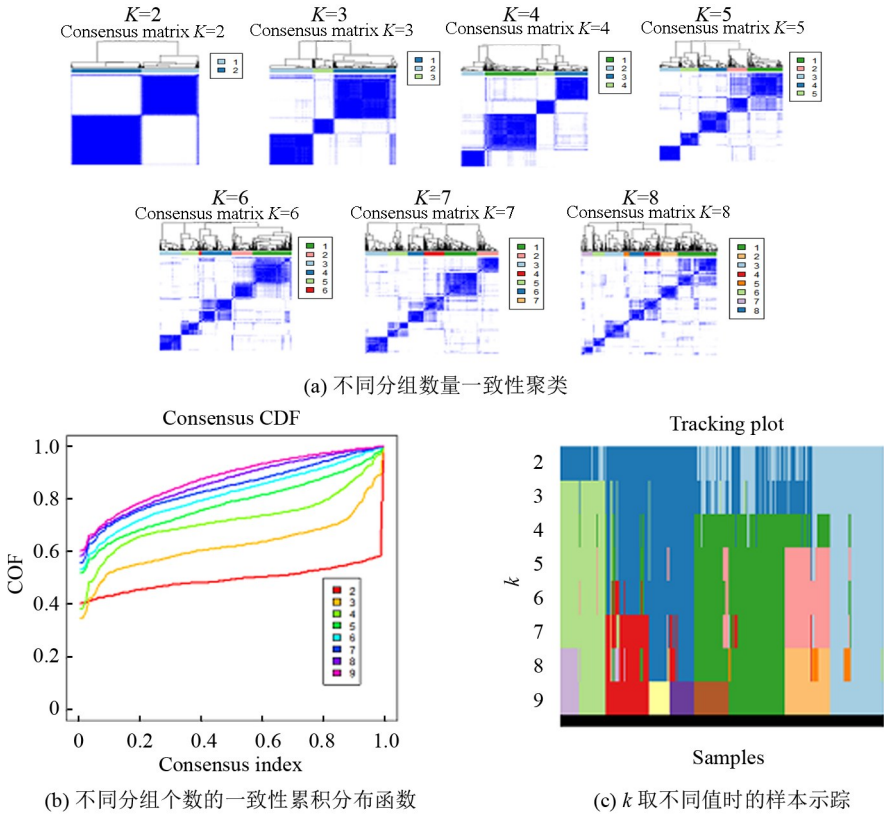


图 3 基于 NMF 模型构建肝细胞癌分子分型

Fig.3 Molecular classification of HCC based on NMF model

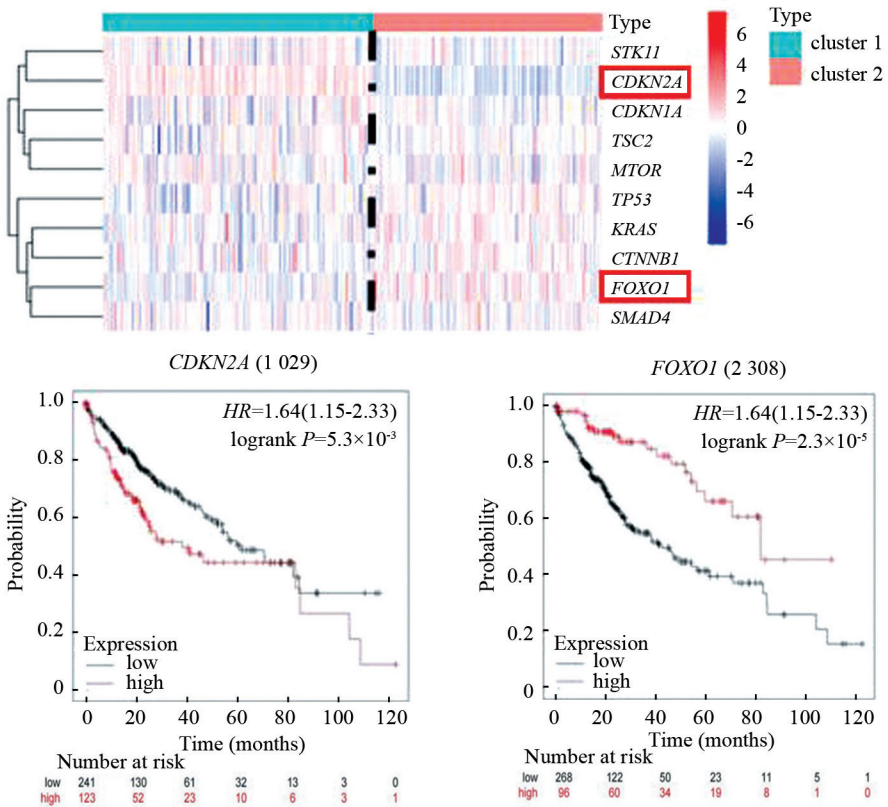


图 4 10 种 hub 基因在不同肝细胞癌分子分型表达水平

Fig.4 Expression level of ten hub genes in different molecular typing of HCC

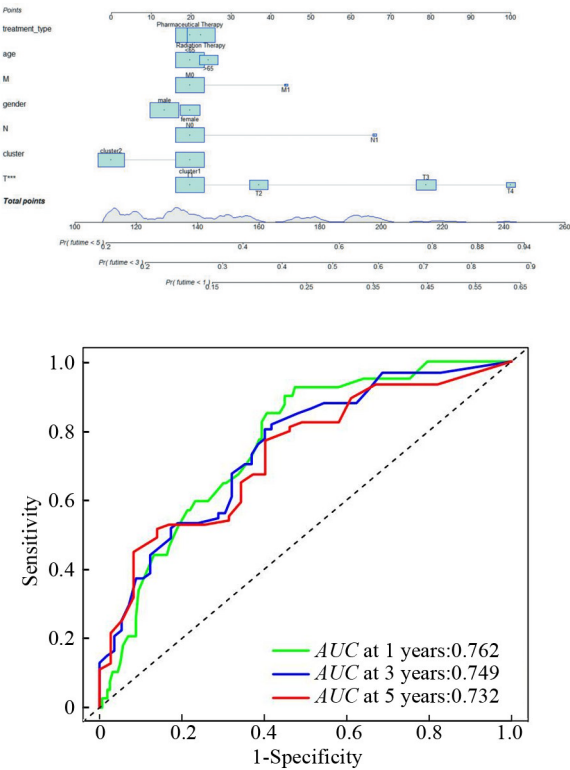


图5 包含年龄(age)、性别(gender)、肿瘤分期(TMN)、治疗方法(treatment type)和分子分型(cluster)等预测因素的肝癌患者1年、3年、5年死亡率的列线图预测模型  
Fig.5 Nomogram including age, gender, TMN stage, treatment type, and molecular cluster for 1-, 3-, and 5-year overall death rate in patients with HCC

为进一步验证分子分型方法的可靠性和稳定性,基于GEO(Gene Expression Omnibus)数据库的GSE76427数据集测序数据及临床信息,采用相同分子分型方法对该数据集中的115例肝癌患者进行预测,并对预测结果中的不同分型患者进行生存分析,同时针对分子分型方法中的关键基因*CDKN2A*和*FOXO*的表达水平进行差异分析。如图5所示验证集GSE76427与TCGA数据库的结论相似:生存分析结果均提示cluster2(低风险组)的预后显著优于cluster1(高风险组),且关键基因*CDKN2A*和*FOXO*的表达水平也分别符合在cluster2(低风险组)和cluster1(高风险组)高表达。

## 2.4 WGCNA 构建基因共表达模块

为进一步分析影响2种分子分型的肝癌患者分子分型生存率差异的原因,探究高、低风险组患者的分子遗传调控机制的不同。本研究首先在转录组水平上,对高风险组(Cluster1)和低风险组(Cluster2)肝癌患者的转录组表达谱进行差异分析,共筛选出显著差异表达的mRNA共186个(差异倍数变化值大于2或小于0.50且 $P < 0.05$ ),其中高风险组(Cluster1)相较于低风险组(Cluster2)共有120个下调基因及17个

上调基因(见图5),10种hub基因表达在高风险组,低风险组中均具有统计学差异,其中*CDKN1A*和*FOXO1*在高风险组的表达显著低于低风险组,*KARS*和*CDKN2A*在高风险组的表达显著高于低风险组,这些hub基因的在高、低风险组的表达趋势与生存分析结果相一致(见表1)。其次利用加权基因共表达网络分析算法(WGCNA),针对差异表达基因构建共表达模块,并分别对模块进行富集分析,重点挖掘和癌症发生发展相关的通路,并展开更深层次的探索。

通过WGCNA分析共鉴定出5个共表达模块,由于灰色模块(MEgrey)由没有共表达的游离基因组成,因此最终确定的有效模块数目为4个(见图6),每个模块包含基因数目大于30。为确定与分子分型显著相关的特异性基因模块,将基因模块与肝癌分子分型进行关联分析,并计算基因模块与高、低风险两种分型的相关性系数和 $P$ 值。结果显示(见图7)MEblue模块与高风险组(Cluster1)具有显著正相关性( $r = 0.34, P = 3 \times 10^{-11}$ ),而MEbrown、MEgreen模块与高风险组(Cluster1)具有显著的负相关性( $r = -0.4, P = 2 \times 10^{-15}$ )。此外不同基因模块具有特异性的代谢通路及生物学过程,其中MEblue模块基因在TECM-受体相互作用、PI3K-Akt信号通路、黏附信号等3个通路上显著富集,而MEgreen、MEyellow模块则分别与胰腺分泌、糖胺聚糖生物合成-硫酸乙酰肝素/肝素等通路密切相关,MEbrown模块中的基因则主要参与细胞周期的调控(见图8)。结合WGCNA分析和KEGG富集结果,推测高风险组富集多种与肿瘤发生发展相关的基因与通路,其中包括ECM-受体相互作用,黏附信号通路,以及与肿瘤细胞增殖相关的PI3K-Akt信号通路,而低风险组中主要是与细胞周期的改变相关。类似结果在曹颖颖等人对比高风险组和中低风险胃癌患者分析也有所提及(见图9)<sup>[22]</sup>。

## 3 讨论

通过SNP和mRNA表达谱关联分析及分子分型模型的构建,本研究基于10种核心基因的表达水平将患者分为高风险组和低风险组两类,二者的生存率存在显著差异。其中核心基因*CDKN2A*和*FOXO1*在肝癌分子分型中起到决定性作用,因此我们推测由突变导致的*CDKN2A*高表达和*FOXO1*低表达可能与肝癌患者的不良预后密切相关。*CDKN2A*即细胞周期依赖性激酶抑制基因(Cyclin-dependent kinase inhibitor),是一种直接参与细胞周期调控的抑癌基因,由Kamb等于1994年首次报道<sup>[23]</sup>,并以其在肿瘤细胞中的高突变率和抑癌作用而在肿瘤遗传学



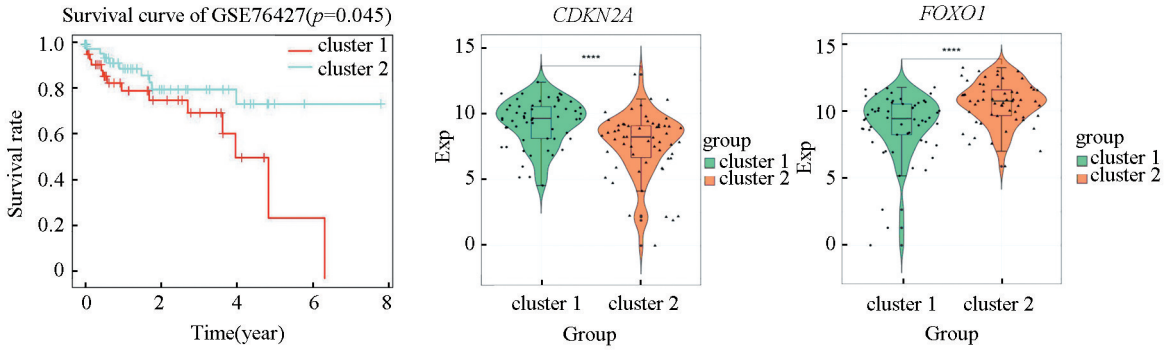


图 6 GSE76427 数据集对肝癌分子分型方法的验证

Fig.6 Validation for molecular classification method of HCC by GSE76427 dataset

表 2 主要基因模块的 KEGG 通路富集结果

Table 2 KEGG analysis of genes in major gene modules

基因模块	代谢通路	校正后 P 值	基因 ID
blue	ECM-受体相互作用	$5.451 \times 10^{-3}$	<i>IBSP/LAMC3/TNXB/THBS2/LAMA2/COL6A6</i>
	PI3K-Akt 信号通路	$5.451 \times 10^{-3}$	<i>IBSP/LAMC3/NGFR/TCL1A/TEK/FGF7/TNXB/CHRM2/THBS2/LAMA2/COL6A6</i>
	黏附信号通路	$3.841 \times 10^{-2}$	<i>IBSP/LAMC3/PAK3/TNXB/THBS2/LAMA2/COL6A6</i>
brown	细胞周期	$7.50 \times 10^{-14}$	<i>E2F2/MCM2/SMC1B/ORC1/ORC6/CDC45/CDC6/CDC7/E2F1/CCNE1/TTK/CDC20/CDKN2C/PKMYT1/CCNA2/CDKN2A/BUB1B/CCNB2/CDC25C/PTTG1/PLK1/BUB1/SFN</i>
green	胰液分泌	$3.20 \times 10^{-6}$	<i>CPA1/CELA3A/GPA2/CTR2/PRSS1/PRSS2</i>
yellow	糖胺聚糖生物合成-硫酸乙酰肝素/肝素	$3.33 \times 10^{-2}$	<i>NDST4/NDST3</i>

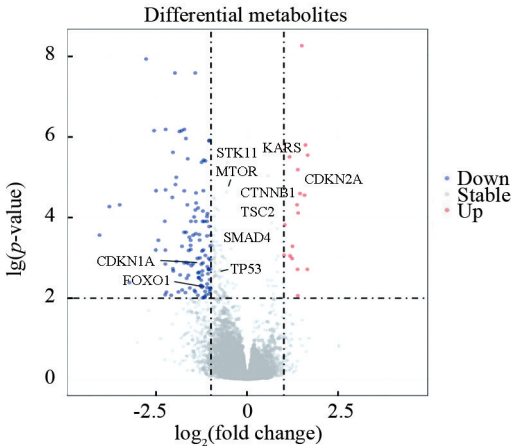


图 7 肝癌分子分型差异表达基因的火山图

Fig.7 Volcanic map of differentially expressed genes of different molecular typing of HCC

注:左侧蓝点为高危组表达水平显著高于低危组的基因,右侧红点为高危组表达水平显著低于低危组的基因;10 种 Hub 基因具有文字标注。

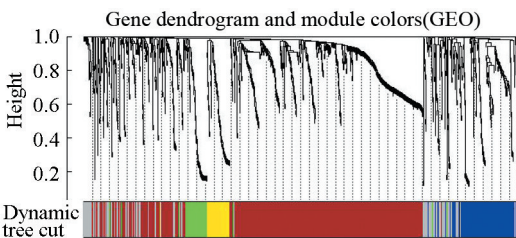


图 8 联合加权共表达网络:基因层次聚类树及基因模块 Fig.8 WGCNA: clustering dendrogram of genes with assigned modules

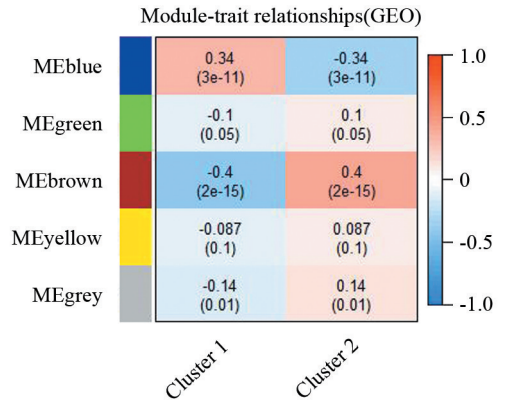


图 9 基因模块与肝细胞癌不同分子分型的相关分析

Fig.9 Associations between gene modules and different molecular typing of HCC

和肿瘤分子生物学的研究中受到了广泛关注。研究表明 CDKN2A 的缺失、突变和甲基化可导致其编码两种细胞周期抑制蛋白 p16<sup>INK4a</sup> 和 p14<sup>ARF</sup> 异常,其中 p16<sup>INK4a</sup> 不能与 CDKN2A 和 CDK6 结合,使 pRb 蛋白磷酸化和转录调节因子 E2F 释放,诱导 G1 期停滞,有助于肿瘤的发生<sup>[24]</sup>,而 p14<sup>ARF</sup> 功能的分子机制比较复杂,可能是通过 P14<sup>ARF</sup>-MDM<sup>2</sup>-p53 途径起作用<sup>[25-26]</sup>,这两条途径的异常普遍存在于各种肿瘤。但本研究基于 TCGA 数据库的分析结果显示,CDKN2A 在肝癌中属于不良预后因子并在高危组中

高表达。此外通过对 TCGA 数据库中其它癌症类型的数据分析结果,发现在子宫内膜癌、肾癌中 *CDKN2A* 同样属于不良预后因子,类似的结果在 Larque A B 等<sup>[27]</sup>关于 hpv 阴性的喉部鳞状细胞癌的研究中也有所提及。总而言之 *CDKN2A* 作为一种重要的抑癌基因,其纯合缺失、启动子甲基化或基因点突变与肿瘤的发生密切相关,但其 mRNA 表达水平作为肝癌患者预后因子的研究还不充分,*CDKN2A* 是否为肝癌的不良预后因子还需更加系统、深入的研究和验证。FOXO 又名叉头蛋白 (Forkhead box protein) 是一类转录因子,广泛参与到细胞新陈代谢、分化、凋亡、增殖等生命活动中,尤其在细胞周期进程的调控和程序化死亡中起到重要作用。许多研究表明,FOXO 的失活与肿瘤的发生发展显著相关,是哺乳动物细胞最重要的一类抑癌基因,并且在多种肿瘤中都可观察到 FOXO 转录因子的异常表达。*FOXO1* 是叉头转录因子 FOXO 家族的一个重要成员,*FOXO1* 在肝癌组织中低表达,导致肝癌细胞的细胞周期失控和凋亡异常抵抗,进而加速肿瘤的进展。提高 *FOXO1* 的活性,可影响上调细胞周期抑制蛋白 p21、p27 和下调细胞周期蛋白 cyclinD1 的表达从而抑制肝癌细胞增殖,并通过激活促凋亡蛋白 Bim,诱导肝癌细胞凋亡<sup>[28-30]</sup>。研究发现,由 *FOXO1* 突变引发的 mRNA 表达水平的变化,对肝癌患者的分子分型起到重要作用同时与预后存在较强的相关性。基于以上结果我们推断,*FOXO1* 可能是肝癌的肿瘤抑制基因,该基因突变导致的表达水平降低可导致肝癌患者生存期缩短。

此外本研究还发现,两种分子分型的差异表达基因功能,与肿瘤细胞侵蚀、转移、复发过程相关的信号通路相关,包括与高危组显著相关的信号通路包括:ECM-受体相互作用,黏附信号通路<sup>[31-32]</sup>,以及与肿瘤细胞增殖相关的 PI3K - Akt 信号通路<sup>[33-34]</sup>;而低危组中异常通路信号则主要与细胞周期、胰液分泌异常有关。研究表明,肝癌的发病早期往往就出现门静脉侵袭、肝内转移以及肝外肺脏和骨组织的转移,而肝癌的侵袭、转移和术后复发是影响患者预后的主要因素。恶性肿瘤发生发展常常伴有细胞外基质 (Extracellular Matrix, ECM) 及其细胞表面受体表达的变化<sup>[35-36]</sup>;而黏附信号通路的激活则在细胞分化,发育以及增殖,凋亡方面起重要作用,并参与肿瘤的侵袭,运动和转移过程<sup>[37]</sup>;PI3K/Akt/mTOR 信号通路则作为细胞内重要信号传导通路之一,可通过影响下游多种效应分子的活化状态,维持肿瘤细胞恶性增殖的生物学特性<sup>[32]</sup>。由此可以看出高危组肿瘤的恶性程度高,侵袭性强,极易复

发和转移,是造成患者预后较差、生存率较低的主要原因。

利用 TCGA 数据库中肝癌患者 SNP 突变数据和转录组芯片数据的联合分析,通过蛋白互作网络筛选出核心基因,并构建肝癌分子分型方法。在没有任何先验信息的情况下对肝癌患者进行分子分型,分析结果表明这种分子分型方法对肝癌患者的预后评估具有一定的作用。另外通过 WGCNA 和 KEGG 富集分析探寻了高危、低危两种分型间的基因表达和异常代谢通路的差异,对造成不同肝癌分子分型患者生存率差异的原因进行进一步研究,同时筛选出的核心基因 *CDKN2A*、*FOXO1* 也可作为肝癌早期诊断,预后监测的新型分子标志物以及分子治疗的新靶点。

## 4 结论

1) *CDKN2A* 和 *FOXO1* 在肝癌分子分型中起到决定性作用,且 *CDKN2A* 高表达和 *FOXO1* 低表达可能与肝癌患者的不良预后密切相关。

2) 高危组异常信号通路 ECM-受体相互作用,黏附信号通路,PI3K - Akt 信号通路可能与恶性程度高,侵袭性强,极易复发和转移,预后较差有关。

## 参考文献 (References)

- [1] 全国多中心前瞻性肝癌极早期预警筛查项目 (PreCar) 专家组. 中国肝癌早筛策略专家共识 [J]. 中华肝脏病杂志, 2021, 29 (6): 515 - 522. DOI: 10.3760/cma.j.cn501113-20210605-00264. Expert Group of National Multicenter Prospective Liver Cancer Very Early Warning Screening Project (PreCar). Expert consensus on early screening strategy for liver cancer in China [J]. Chinese Journal of Hepatology, 2021, 29(6): 515-522. DOI: 10.3760/cma.j.cn501113-20210605-00264.
- [2] 陈孝平, 陈汉. 原发性肝癌外科治疗方法的选择 [J]. 中华普通外科杂志, 2001, 23 (10): 762 - 763. DOI: 10.3969/j.issn.1005-6947.2005.02.001. CHEN Xiaoping, CHEN Han. Selection of surgical treatment methods for primary liver cancer [J]. Chinese Journal of General Surgery, 2001, 23(10): 762-763. DOI: 10.3969/j.issn.1005-6947.2005.02.001.
- [3] ER T K, SU Y F, WU C C, et al. Targeted next-generation sequencing for molecular diagnosis of endometriosis-associated ovarian cancer [J]. Journal of Molecular Medicine, 2016, 94 (7): 835-847. DOI: 10.1007/s00109-016-1395-2.
- [4] FRANSVEA E, PARADISO A, ANTONACI S, et al. HCC heterogeneity: Molecular pathogenesis and clinical impli cat-



- ions[J]. *Cellular Oncology*, 2014, 31(3): 227-233. DOI: 10.3233/CLO-2009-0473.
- [5] 周华邦, 胡和平. 基于基因突变的原发性肝癌精准治疗的挑战[J]. *临床肝胆病杂志*, 2017, 33(7): 1209-1210. DOI: 10.3969/j.issn.1001-5256.2017.07.001.
- ZHOU Huabang, HU Heping. Challenges of precision therapy for primary liver cancer based on gene mutation[J]. *Journal of Clinical Hepatobiliary Diseases*, 2017, 33(7): 1209-1210. DOI: 10.3969/j.issn.1001-5256.2017.07.001.
- [6] 危荣沥, 徐骁. 肝细胞癌肝移植分子分型与精准治疗[J]. *实用器官移植电子杂志*, 2019, 7(1): 78-81. DOI: CNKI:SUN:YZQG.0.2019-01-021.
- WEI Rongli, XU Xiao. Molecular typing and precise treatment of hepatocellular carcinoma in liver transplantation[J]. *Practical Journal of Organ Transplantation (Electronic Version)*, 2019, 7(1): 78-81. DOI: CNKI:SUN:YZQG.0.2019-01-021.
- [7] 晏光荣. SNP 与肿瘤易感相关性研究进展[J]. *国外医学(生理,病理科学与临床分册)*, 2004, 24(2): 175-177. DOI: 10.3969/j.issn.1673-2588.2004.02.022.
- YAN Guangrong. Research progress on the correlation between SNP and tumor susceptibility[J]. *Foreign Medicine (Physiology, Pathology and Clinical Division)*, 2004, 24(2): 175-177. DOI: 10.3969/j.issn.1673-2588.2004.02.022.
- [8] ZHU Y. An evolutionary perspective on single-nucleotide polymorphism screening in molecular cancer epidemiology[J]. *Cancer Research*, 2004, 64(6): 2251-2257. DOI: 10.1158/0008-5472.CAN-03-2800.
- [9] KLEIN R J, HALLDEN C, CRONIN A M, et al. Blood biomarker levels to aid discovery of cancer-related single-nucleotide polymorphisms: kallikreins and prostate cancer[J]. *Cancer Prevention Research*, 2010, 3(5): 611-619. DOI: 10.1158/1940-6207.CAPR-09-0206.
- [10] LAYTRAGOON-LEWIN N, CEDERBLAD L, ANDERSSON B, et al. Single-nucleotide polymorphisms and cancer risk, tumor recurrence, or survival of head and neck cancer patients[J]. *Oncology*, 2016, 92(3): 161-169. DOI: 10.1159/000452278.
- [11] FRITSCH L G, IGI W, BAILEY J, et al. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants[J]. *Nature Genetics*, 2016, 48(2): 134-143. DOI: 10.1038/ng.3448.
- [12] 任重鲁, 李金明. 非负矩阵分解在微阵列数据分类和聚类发现中的应用[J]. *计算机工程与科学*, 2014, 36(7): 1389-1397. DOI: 10.3969/j.issn.1007-130X.2014.07.030.
- REN Zhonglu, LI Jinming. Application of nonnegative matrix factorization in microarray data classification and cluster discovery[J]. *Computer Engineering and Science*, 2014, 36(7): 1389-1397. DOI: 10.3969/j.issn.1007-130X.2014.07.030.
- [13] 杜芳. 非负矩阵因子分解算法处理癌症基因表达数据的研究[D]. 成都: 电子科技大学, 2009. DOI: 10.7666/d.Y1463927.
- DU Fang. Research on non-negative matrix factorization algorithm to process cancer gene expression data[D]. Chengdu: University of Electronic Science and Technology of China, 2009. DOI: 10.7666/d.Y1463927.
- [14] BRUNET J P, TAMAYO P, GOLUB T R, et al. Metagenes and molecular pattern discovery using matrix factorization[J]. *Proceedings of the National Academy of Sciences*, 2004, 101(12): 4164-4169. DOI: 10.1073/pnas.0308531101.
- [15] 阮班来, 王桂玲. 利用 WGCNA 构建与胃癌发展进程相关的关键 lncRNA 网络[J]. *解剖科学进展*, 2019, 25(06): 70-74. DOI: 10.16695/j.cnki.1006-2947.2019.06.019.
- RUAN Banlai, WANG Guiling. Construction of key lncRNA networks related to gastric cancer development using WGCNA[J]. *Progress of Anatomical Sciences*, 2019, 25(06): 70-74. DOI: 10.16695/j.cnki.1006-2947.2019.06.019.
- [16] 李冬妹, 杨磊. 抑癌基因 Smad4 与胃肠道肿瘤[J]. *国外医学(消化系疾病分册)*, 2005, 25(4): 254-254. DOI: CNKI:SUN:GWXH.0.2005-04-022.
- LI Dongmei, YANG Lei. Tumor suppressor gene Smad4 and gastrointestinal tumors[J]. *Foreign Medicine (Digestive System Disorders)*, 2005, 25(4): 254-254. DOI: CNKI:SUN:GWXH.0.2005-04-022.
- [17] 阳琼. TSC2 在人肝癌 HepG2 细胞中的表达及机制的初步探讨[D]. 衡阳: 南华大学, 2013.
- YANG Qiong. The expression and mechanism of TSC2 in human hepatoma HepG2 cells[D]. Hengyang: University of South China, 2013.
- [18] WESTERMAN A M, ENTIUS M M, BOOR P P, et al. Novel mutations in the LKB1/STK11 gene in Dutch Peutz-Jeghers families[J]. *Human Mutation*, 2015, 13(6): 476-481. DOI: 10.1002/(SICI)1098-1004(1999)13:6<476::AID-HUMU7>3.0.CO;2-2.
- [19] NOGUEIRA J A, ONO-NITA S K, NITA M E, et al. 249 TP53 mutation has high prevalence and is correlated with larger and poorly differentiated HCC in Brazilian patients[J]. *BMC Cancer*, 2009, 9(1): 204-218. DOI: 10.1186/1471-2407-9-204.
- [20] 田媛媛. CTNNB1 基因多态性与胃癌预后的分子流行病学研究[D]. 南京: 南京医科大学, 2010.
- TIAN Yuanyuan. Molecular epidemiological study of CTNNB1 gene polymorphism and prognosis of gastric cancer[D]. Nanjing: Nanjing Medical University, 2010.
- [21] 郭琳, 王强. PI3K/Akt/mTOR 信号传导通路与恶性肿瘤浸润和转移的研究进展[J]. *现代肿瘤医学*, 2009(8):

- 1585-1589. DOI:10.3969/j.issn.1672-4992.2009.08.079.
- GUO Lin, WANG Qiang. Research progress of PI3K/Akt/mTOR signaling pathway and malignant tumor invasion and metastasis [J]. *Journal of Modern Oncology*, 2009 (8): 1585-1589. DOI:10.3969/j.issn.1672-4992.2009.08.079.
- [22] 曹颖颖, 朱小强, 陈豪燕. 基于非负矩阵分解模型构建胃癌分子分型及预后评估的案例分析 [J]. *上海交通大学学报(医学版)*, 2017, 37(9): 1187-1194. DOI: 10.3969/j.issn.1674-8115.2017.09.001.
- CAO Yingying, ZHU Xiaoqiang, CHEN Haoyan. A case study of molecular typing and prognostic assessment of gastric cancer based on non-negative matrix factorization model [J]. *Journal of Shanghai Jiaotong University (Medical Science)*, 2017, 37(9): 1187-1194. DOI: 10.3969/j.issn.1674-8115.2017.09.001.
- [23] KNUDSON A G. All in the (cancer) family [J]. *Nature Genetics*, 1993, 5(1): 103-104. DOI: 10.1038/ng1093-103.
- [24] RUAS M, PETERS G. The p16INK4a/CDKN2A tumor suppressor and its relatives [J]. *Biochimica et Biophysica ACTA-Biomembranes*, 1998, 1378(2): 115-177. DOI: 10.1016/S0304-419X(98)00017-1.
- [25] EISCHEN C M, WEBER J D, ROUSSEL M F, et al. Disruption of the ARF-Mdm2-p53 tumor suppressor pathway in Myc-induced lymphomagenesis [J]. *Genes & Development*, 1999, 13(20): 2658-2669. DOI: 10.1101/gad.13.20.2658.
- [26] STOTT F J, BATES S, JAMES M C, et al. The alternative product from the human CDKN2A locus, p14<sup>ARF</sup>, participates in a regulatory feedback loop with p53 and MDM2 [J]. *The EMBO Journal*, 1998, 17(17): 5001-5014. DOI:10.1093/emboj/17.17.5001.
- [27] LARQUE A B, CONDE L, HAKIM S, et al. p16INK4a overexpression is associated with CDKN2A mutation and worse prognosis in HPV-negative laryngeal squamous cell carcinomas [J]. *Virchows Archiv*, 2015, 466(4): 375-382. DOI:10.1007/s00428-015-1725-8.
- [28] ZHANG Y, GAN B, LIU D, et al. FoxO family members in cancer [J]. *Cancer Biology & Therapy*, 2011, 12(4): 253-259. DOI:10.4161/cbt.12.4.15954.
- [29] 周振琪, 王恬, 石放雄. FOXO 转录因子调控哺乳动物的细胞周期和凋亡 [J]. *细胞生物学杂志*, 2007, 12(2): 22-25. DOI:10.3969/j.issn.1674-7666.2007.02.004.
- ZHOU Zhenqi, WANG Tian, SHI Fangxiong. FOXO transcription factors regulate cell cycle and apoptosis in mammals [J]. *Chinese Journal of Cell Biology*, 2007, 12(2): 22-25. DOI:10.3969/j.issn.1674-7666.2007.02.004.
- [30] GONZALEZ-GALLEGO J, A GARCÍA-PALOMO, MAURIZ J L, et al. FoxO Proteins: Regulation and molecular targets in liver cancer [J]. *Current Medicinal Chemistry*, 2014, 21(10): 1231-1246. DOI: 10.2174/0929867321666131228205703.
- [31] BAO Yulong, WANG Li, SHI Lin, et al. Transcriptome profiling revealed multiple genes and ECM-receptor interaction pathways that may be associated with breast cancer [J]. *Cellular & Molecular Biology Letters*, 2019, 24(1): 24-38. DOI: 10.1186/s11658-019-0162-0.
- [32] YU W, ZHAO S, WANG Y, et al. Identification of cancer prognosis-associated functional modules using differential co-expression networks [J]. *Oncotarget*, 2017, 8(68): 112928-112941. DOI: 10.18632/oncotarget.22878.
- [33] 吴日平, 黄昌明. PI3K/Akt 信号通路与肿瘤关系的研究进展 [J]. *医学综述*, 2009, 15(10): 1501-1501. DOI: 10.3969/j.issn.1006-2084.2009.10.020.
- WU Riping, HUANG Changming. Research progress on the relationship between PI3K/Akt signaling pathway and tumor [J]. *Medical Recapitulate*, 2009, 15(10): 1501-1501. DOI: 10.3969/j.issn.1006-2084.2009.10.020.
- [34] LUO J, MANNING B D, CANTLEY L C. Targeting the PI3K-Akt pathway in human cancer: rationale and promise [J]. *Cancer Cell*, 2003, 4(4): 257-262. DOI: 10.1016/S1535-6108(03)00248-4.
- [35] 蒋新农, 周柔丽. 肿瘤细胞粘附, 迁移与转移的相关性 [J]. *生物化学与生物物理进展*, 1998, 25(5): 404-407. DOI:CNKI;SUN;SHSW.0.1998-05-004.
- JIANG Xinnong, ZHOU Rouli. Correlation of tumor cell adhesion, migration and metastasis [J]. *Progress in Biochemistry and Biophysics*, 1998, 25(5): 404-407. DOI:CNKI;SUN;SHSW.0.1998-05-004.
- [36] GEIGER B, ABU-AMARA S N, WINOGRAD-KATZ S, et al. Cell adhesion to the extra cellular matrix (ECM) in motility and metastasis [J]. *European Journal of Cancer Supplements*, 2008, 6(9): 62-63. DOI:10.1016/S1359-6349(08)71413-7.
- [37] KORNBERG L J. Focal adhesion kinase and its potential involvement in tumor invasion and metastasis [J]. *Head Neck*, 2015, 20(8): 745-752. DOI:10.1002/(sici)1097-0347(199812)20:8.