

肝细胞癌预后、诊断和免疫细胞浸润相关关键基因及其潜在治疗药物的生物信息学分析

黎金连^{1,2}, 黄岚珍³, 黄希仕³, 李康智¹, 蒋佳丽⁴, 张苗苗¹, 吴群英¹

(1. 桂林医学院智能医学与生物技术学院细胞与遗传学教研室, 广西 桂林 541004; 2. 南宁中心血站输血传播疾病联合实验室, 广西 南宁 530007; 3. 桂林医学院科学实验中心, 广西 桂林 541004; 4. 桂林医学院全科医学院, 广西 桂林 541004)

[摘要] **目的:** 利用生物信息学分析方法筛选与肝细胞癌(HCC)患者预后、诊断和免疫细胞浸润相关的关键基因, 并分析其潜在治疗药物。**方法:** 从基因表达汇编(GEO)数据库和癌症基因组图谱(TCGA)数据库下载HCC基因表达谱数据及相应的HCC患者临床信息。采用R软件limma包筛选HCC差异表达基因(DEGs), 对DEGs进行基因本体论(GO)功能和京都基因与基因组百科全书(KEGG)信号通路富集分析, 利用STRING数据库构建蛋白-蛋白互作(PPI)网络, 使用Cytoscape软件对PPI网络进行可视化并筛选关键基因。通过Kaplan-Meier生存曲线和LASSO回归算法鉴定HCC预后相关关键基因, 并利用外部数据集对其表达进行验证和诊断效能分析。采用CIBERSORT算法评估预后相关的关键基因表达与HCC免疫细胞浸润的关系。利用MiRNet和Network Analyst数据库分别构建微小RNA(miRNA)-关键基因mRNA和转录因子(TFs)-关键基因mRNA分子调控网络。利用CMap数据库筛选潜在治疗HCC的小分子药物。**结果:** 共筛选出146个DEGs, 其中表达上调基因30个, 表达下调基因116个。GO功能和KEGG信号通路富集分析, DEGs明显富集在类固醇、烯化合物和激素代谢等生物过程(BP)及视黄醇代谢、药物代谢-细胞色素P450(CYP450)、补体和凝血级联反应等信号通路。PPI网络分析筛选得到14个关键基因, 其中乙酰氨基转移酶环化脱氨酶(FTCD)、分泌型磷蛋白2(SPP2)、凝血酶-抗凝血酶复合物(TAT)、补体C6(C6)和CYP450家族成员2C9(CYP2C9)与HCC患者预后、临床病理分期和组织学分级有明显相关性, 在HCC诊断中具有较高的诊断效能, 与HCC的免疫细胞浸润有密切关联。Hsa-mir-182-5p、同源框CUT样蛋白1(CUX1)、早期生长反应1(EGR1)、SMAD家族成员4(SMAD4)和肿瘤蛋白P53(TP53)是靶向上述预后相关关键基因的重要调控因子。DL-硫沙芬(DL-thiorphan)、异丙嗪(promethazine)和芹菜素(apigenin)可能对HCC有治疗作用。**结论:** FTCD、SPP2、TAT、C6和CYP2C9可能是HCC诊断、预后判断和治疗的潜在靶点, 预测得到的3种小分子药物DL-thiorphan, promethazine和apigenin可为HCC治疗药物的研发提供参考。

[关键词] 肝细胞癌; 生物标志物; 小分子药物; 生物信息学

[中图分类号] Q811.4; R735.7 **[文献标志码]** A

[收稿日期] 2023-07-18

[基金项目] 广西壮族自治区教育厅中青年提升项目(2018KY0407); 广西壮族自治区教育厅大学生创新创业训练计划项目(S202310601092); 桂林医学院博士科研启动基金项目(20501021028)

[作者简介] 黎金连(1994—), 女, 广西壮族自治区梧州市人, 技师, 医学硕士, 主要从事疾病系统生物信息学分析和分子机制方面的研究。

[通信作者] 吴群英, 副教授(E-mail: wuqunying@glmc.edu.cn)

Bioinformatics analysis on key genes related to prognosis, diagnosis, and immune cell infiltration of hepatocellular carcinoma and their potential therapeutic drugs

LI Jinlian^{1,2}, HUANG Lanzhen³, HUANG Xishi³, LI Kangzhi¹, JIANG Jiali⁴,
ZHANG Miaomiao¹, WU Qunying¹

(1. Department of Cell and Genetics, School of Intelligent Medicine and Biotechnology, Guilin Medical University, Guilin 541004, China; 2. Joint Laboratory of Transfusion-Transmitted Diseases, Nanning Blood Center, Nanning 530007, China; 3. Research Center for Science, Guilin Medical University, Guilin 541004, China; 4. School of General Medicine, Guilin Medical University, Guilin 541004, China)

ABSTRACT Objective: To screen the key genes related to the prognosis, diagnosis, and immune infiltration of the hepatocellular carcinoma (HCC) patients by bioinformatics analysis methods, and to analyze their potential therapeutic drugs. **Methods:** The HCC gene expression profile data and corresponding clinical informations of the HCC patients were downloaded from the Gene Expression Omnibus (GEO) database and The Cancer Genome Atlas (TCGA) database. The R software package limma was used to screen the differentially expressed genes (DEGs) in HCC. Gene Ontology (GO) functional enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed on the DEGs. The STRING database was used to construct the protein-protein interaction (PPI) network; the Cytoscape software was used to visualize the PPI network and screen the key genes; Kaplan-Meier survival curve and LASSO regression algorithm were used to identify the key genes related to the HCC prognosis; external data sets were used to validate their expressions and analyze the diagnostic efficacy; CIBERSORT algorithm was used to detect the relationship between the expression of prognosis-related key genes and HCC immune cell infiltration. The MiRNet and Network Analyst databases were used to construct the microRNA (miRNA)-key gene mRNA and transcription factors (TFs)-key gene mRNA molecular regulatory networks; CMap database was used to screen the potential small molecule drugs for HCC treatment. **Results:** A total of 146 DEGs were screened, including 30 upregulated genes and 116 downregulated genes. The GO functional enrichment analysis and KEGG pathway enrichment analysis results showed that the DEGs were significantly enriched in biological processes (BP) such as steroid, alkene compound, and hormone metabolism, as well as signaling pathways such as retinol metabolism, drug metabolism-cytochrome P450 (CYP450), complement and coagulation cascades. The PPI network analysis identified 14 key genes, among which formimidoyltransferase cyclodeaminase (FTCD), secreted phosphoprotein 2 (SPP2), thrombin-antithrombin complex (TAT), complement C6 (C6), and cytochrome CYP450 family member 2C9 (CYP2C9) were significantly associated with the prognosis, clinical pathological stage, and histological grade of the HCC patients and also had high diagnostic efficacy for HCC and were closely related to immune cell infiltration in HCC. Hsa-mir-182-5p, CUT-like homeobox 1 (CUX1), early growth response 1 (EGR1), SMAD family member 4 (SMAD4), and tumor protein P53 (TP53) were identified as the important regulators targeting the above-mentioned prognosis-related key genes. DL-thiorphan, promethazine, and apigenin may have the therapeutic effects on HCC. **Conclusion:** FTCD, SPP2, TAT, C6, and CYP2C9 may be the potential targets for the diagnosis, prognosis, and treatment of HCC. Three predicted small molecule drugs, DL-thiorphan, promethazine, and apigenin, may provide the references for the development of therapeutic drugs for HCC.

KEYWORDS Hepatocellular carcinoma; Biomarker; Small molecule drug; Bioinformatics

肝癌是一种常见的恶性肿瘤,也是癌症相关死亡的重要原因^[1]。据估计,2025年将会有超过100万新发肝癌病例^[2]。在肝癌的组织学形式中,超过90%的原发性肝癌为肝细胞癌(hepatocellular carcinoma, HCC)^[3]。目前,HCC的治疗手段仍然有限,尽管近年来提出了许多治疗策略,但由于其复发率高和治疗耐药,HCC患者5年生存率仍低于20%^[4]。因此寻找HCC患者早期诊断和预后评估的分子标志物及潜在的治疗HCC的小分子药物研究至关重要。

近年来,测序技术的迅速发展和公共数据库的广泛普及,为挖掘HCC的生物标志物提供了机会,通过生物信息学分析已经发现了一些与HCC预后密切相关的分子生物标志物^[5]。研究^[6]显示:血液和神经表达1(hematological and neurological expressed 1, HN1)、Ras相关核蛋白(Ras-related nuclear protein, RAN)、受体活性修饰蛋白3(receptor activity-modifying protein 3, RAMP3)、角蛋白19(Keratin 19, KRT19)和TATA-box结合蛋白相关因子9(TATA-box binding protein associated factor 9, TAF9)5个基因的表达水平可以预测HCC患者的预后,其异常表达与HCC的发生发展有密切关联^[6]。另外,保罗样激酶1(Polo-like kinase 1, PLK1)、富脯氨酸的有丝分裂检查点控制因子(proline rich mitotic check point control factor, PRCC)、pre-mRNA加工因子4(pre-mRNA processing factor 4, PRPF4)和蛋白酶体20S亚基alpha 7(proteasome 20S subunit alpha 7, PSMA7)可能是HCC潜在的生物标志物或治疗靶点^[7]。但上述HCC分子标志物的表达缺乏大样本数据进行独立验证,单一的Kaplan-Meier生存分析方法可能存在假阳性。本研究利用多个独立的HCC数据集进行联合分析和验证,并采用Kaplan-Meier生存曲线联合LASSO回归算法进行分析,以期筛选出有效的HCC生物标志物和小分子药物,为HCC患者的早期诊断、预后评估和临床治疗提供参考。

1 资料与方法

1.1 数据集来源 从基因表达汇编(Gene Expression Omnibus, GEO)数据库(<https://www.ncbi.nlm.nih.gov/geo/>)获取6个HCC基因表达谱芯片数据集,包括GSE76427、GSE57957、GSE39791、GSE36376、GSE29721和GSE45267,其中GSE76427、GSE57957、GSE39791和

GSE36376作为训练集,GSE29721和GSE45267作为验证集。此外,从癌症基因组图谱(The Cancer Genome Atlas, TCGA)数据库(<https://portal.gdc.cancer.gov/>)中获取396例HCC组织样本和160例癌旁组织样本的转录组测序数据,并下载了相应的HCC患者临床信息。

1.2 差异表达分析 获得差异倍数(fold change, FC),以 $\log_2|FC|>1$ 和校正后 $P<0.05$ 为筛选条件,利用limma R软件包筛选HCC组织和癌旁组织之间的差异表达基因(differentially expressed genes, DEGs)。利用ggplot2和pheatmap包绘制DEGs的火山图和热图,可视化差异分析结果^[8]。采用Venny 2.1.0软件(<http://bioinfogp.cnb.csic.es/tools/Venny/index.html>)^[9]绘制4个训练集中DEGs的韦恩图,以获得同时在4个数据集中存在的DEGs,用于后续研究。

1.3 基因本体论(Gene Ontology,GO)功能和京都基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes,KEGG)信号通路富集分析

GO功能富集分析包括生物过程(biological process, BP)、细胞成分(cellular component, CC)和分子功能(molecular function, MF)。利用R软件clusterProfiler包进行GO功能和KEGG信号通路富集分析。以 $P<0.05$ 作为筛选条件,选取GO功能和KEGG信号通路富集分析结果中排名前10位的条目,以气泡图表示。

1.4 蛋白-蛋白互作(protein-protein interaction, PPI)网络构建和关键基因筛选 为了进一步分析DEGs之间的相互作用,将共同的DEGs输入STRING数据库(<http://STRING-db.org>),以互作评分 >0.4 为标准,构建PPI网络。采用Cytoscape软件可视化PPI网络,其中节点(node)代表节点基因,边(edge)代表基因之间的相互作用情况。通过CytoHubba插件中MCC、MNC、Degree和EPC 4种不同的拓扑算法筛选出PPI网络排名前19位基因,取交集获得HCC相关的关键基因。

1.5 关键基因表达水平验证及其预后价值分析

基于TCGA数据库,应用R软件limma包^[10]对配对和非配对样本中的关键基因表达水平进行分析。应用survival包绘制Kaplan-Meier生存曲线,分析关键基因表达水平与HCC患者总生存期(overall survival, OS)之间的关系,用survminer及ggplot2包将结果进行可视化。应用glmnet包进

行LASSO回归分析, 通过调节正则化参数 λ , 可以筛选出HCC患者预后相关基因。通过上述2种方法得出的交集基因为HCC患者预后相关关键基因, 采用corrplot软件包对上述基因的相关性进行分析。

1.6 预后相关关键基因表达水平验证 利用验证集GSE29721和GSE45267的外部数据验证HCC预后相关关键基因表达情况, 采用Mann-Whitney U 检验进行统计学分析, 以 $P < 0.05$ 为差异有统计学意义。为了评估预后相关关键基因对HCC的诊断价值, 采用pROC包绘制受试者工作特性(receiver operating characteristic, ROC)曲线, 并计算曲线下面积(area under curve, AUC)^[11]。

1.7 预后相关关键基因与HCC患者临床病理特征的关系 使用R语言stats和car包, 采用Kruskal-Wallis检验方法, 分析预后相关关键基因表达水平与HCC患者临床病理特征之间的关系。采用plot2包对分析结果进行可视化, 以 $P < 0.05$ 为差异有统计学意义。

1.8 预后相关关键基因与HCC患者免疫细胞浸润的关系 利用CIBERSORT算法分析TCGA数据库中的HCC数据集, 研究预后相关关键基因表达与22种不同类型免疫细胞浸润之间的关系^[12]。采用ggplot2软件包计算关键基因表达与免疫细胞浸润得分, 绘制棒棒糖图对结果进行可视化。

1.9 构建微小RNA(micro-RNA, miRNA)-关键基因和转录因子(transcription factors, TFs)-关键基因调控网络 利用MiRNet数据库(<https://www.mirnet.ca/>)分析靶向预后相关关键基因的miRNAs, 并采用Network Analyst数据库(<https://www.networkanalyst.ca/>), 以cutoff degree >1 为标准^[13], 筛选靶向调控HCC预后相关关键基因的TFs。构建miRNA-关键基因mRNA和TFs-关键基因mRNA调控网络, 采用Cytoscape软件对分析结果进行可视化。

1.10 预测潜在治疗HCC的小分子药物 联通图谱(Connectivity Map, CMap)数据库(<https://clue.io/>)是基于干预基因表达的表达谱数据库, 可以揭示疾病、基因和小分子化合物之间的关系^[14]。本研究将“1.2”中分析得到的DEGs分为表达上调和表达下调2组基因, 并将2组基因分别输入CMap数据库, 计算小分子化合物的富集分数($-1 \sim 1$), 并筛选出富集分数 < -0.700 且 $P <$

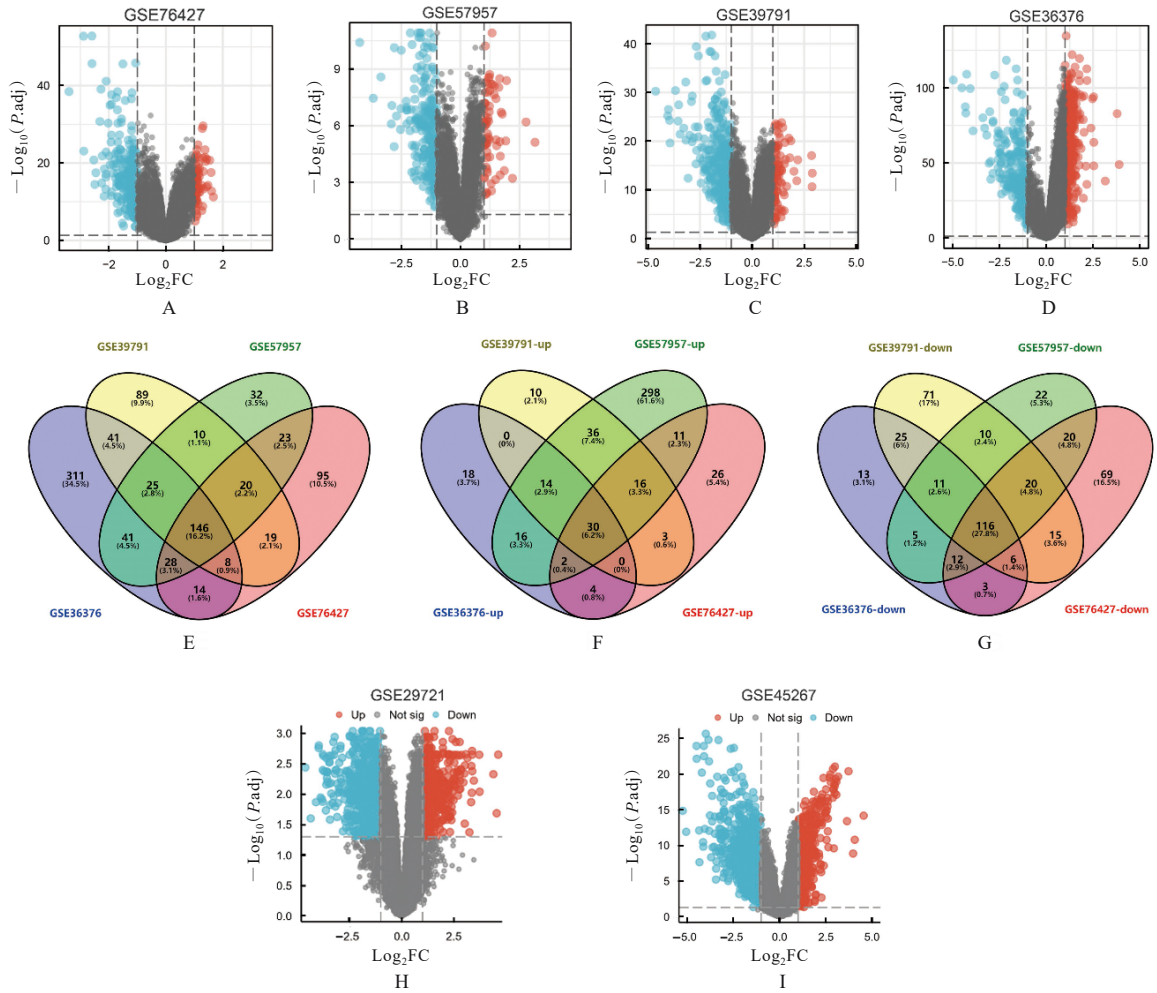
0.05 的前20个小分子药物。

2 结果

2.1 DEGs的筛选 对4个基因表达谱芯片数据集进行差异表达分析, 结果显示: GSE76427、GSE57957、GSE39791和GSE36376的DEGs分别为353、325、358和614个(图1A~1D), 对上述DEGs取交集, 最终获得了146个共同的DEGs(图1E), 其中30个基因表达上调(图1F), 116个基因表达下调(图1G)。验证集GSE29721和GSE45267的基因差异表达结果见图1H和1I, 排名前50位的DEGs见图2。

2.2 GO功能和KEGG信号通路富集分析 GO功能富集分析结果显示: 146个DEGs在类固醇、烯化合物和激素代谢等BP中明显富集(图3A); DEGs主要分布于含胶原蛋白的细胞外基质、血液微粒和内质网等CC中(图3B); DEGs的主要MF包括单加氧酶活性、铁离子结合和亚铁血红素结合等(图3C)。KEGG信号通路富集分析显示: DEGs主要富集在视黄醇代谢、药物代谢-细胞色素P450(cytochrome P450, CYP450)、补体和凝血级联反应等信号通路中(图3D)。

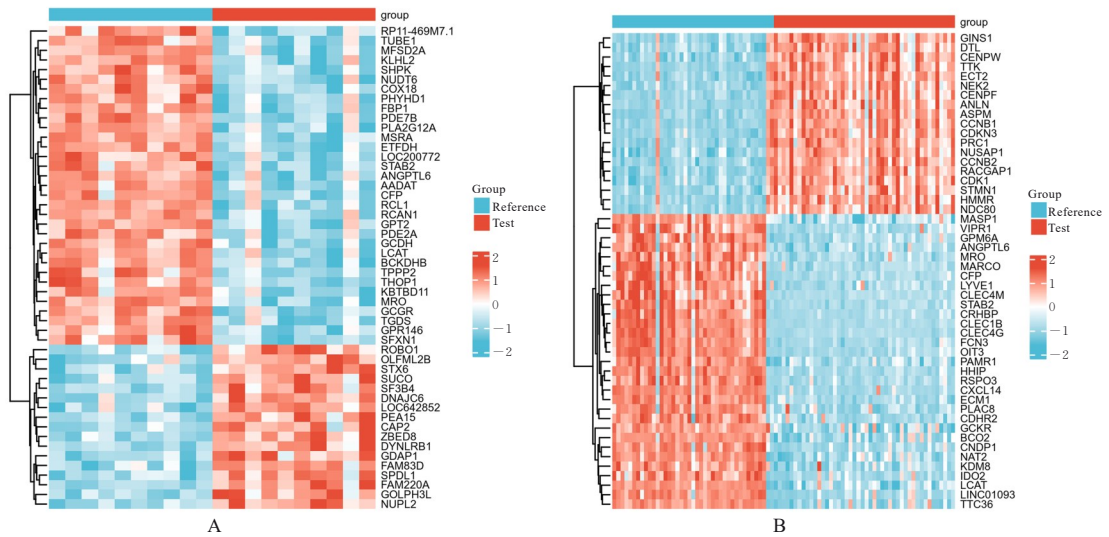
2.3 PPI网络构建和关键基因鉴定 利用STRING数据库构建PPI网络, 结果显示: 146个DEGs之间形成144个节点和367条边(图4A)。采用CytoHubba插件的MCC、MNC、Degree和EPC 4种不同算法, 筛选得到14个关键基因, 包括补体C8 α 链(complement C8 alpha chain, C8A)、CYP450家族成员3A4(CYP450 family member 3A4, CYP3A4)、凝血因子Ⅸ(coagulation factor Ⅸ, F9)、UDP葡萄糖醛酸转移酶家族2成员B10(UDP glucuronosyltransferase 2 family, polypeptide B10, UTG2B10)、甘露糖结合凝集素2(mannose binding lectin 2, MBL2)、甲酰氨基转移酶环化脱氨酶(ormimidoyltransferase cyclodeaminase, FTCD)、CYP450家族成员2E1(CYP450 family member 2E1, CYP2E1)、分泌型磷酸蛋白2(secreted phosphoprotein 2, SPP2)、激肽释放酶B1(kallikrein B1, KLKB1)、凝血酶-抗凝血酶复合物(thrombin-antithrombin complex, TAT)、补体C8 β 链(complement C8 beta chain, C8B)、补体C6(complement component 6, C6)、CYP450家族成员2C9(CYP450 family member 2C9, CYP2C9)和CYP450家族成员1A2



A-D: Volcano plots of GSE76427, GSE57957, GSE39791, and GSE36376; E: Venn diagram of DEGs; F: Venn diagram of up-DEGs; G: Venn diagram of down-DEGs; H: Volcano plot of GSE29721; I: Volcano plot of GSE45267.

图1 DEGs的筛选

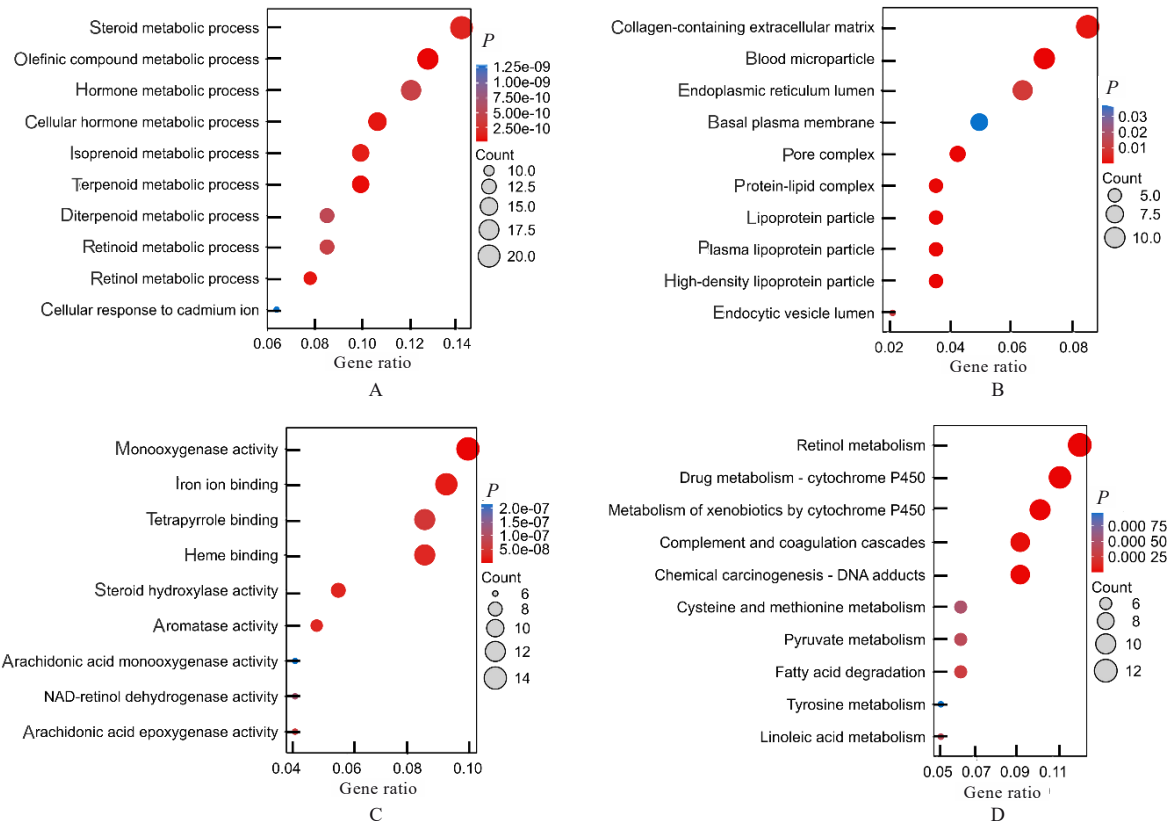
Fig. 1 Screening for DEGs



A: GSE29721; B: GSE45267.

图2 排名前50位DEGs的热图

Fig. 2 Heat map of top 50 DEGs



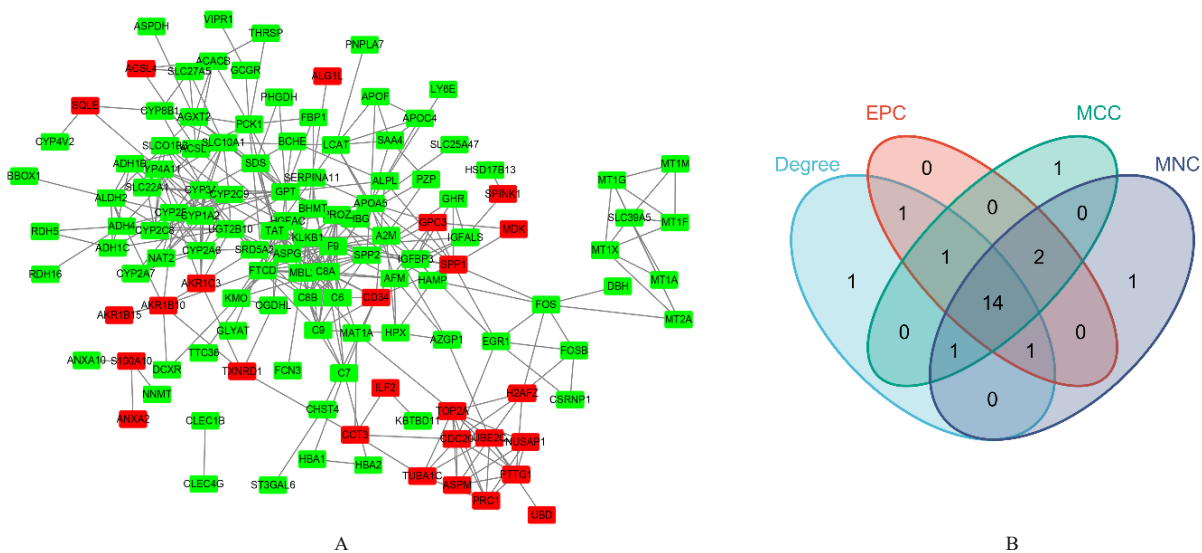
A: BP; B: CC; C: MF; D: KEGG signaling pathway enrichment analysis.

图3 DEGs的GO功能和KEGG信号通路富集分析

Fig. 3 GO functional and KEGG signaling pathway enrichment analysis on DEGs

(CYP450 family member 1A2, CYP1A2) (图4B)。上述14个关键基因在HCC配对样本和非配对样本之间的表达差异见图5, 与癌旁组织比较,

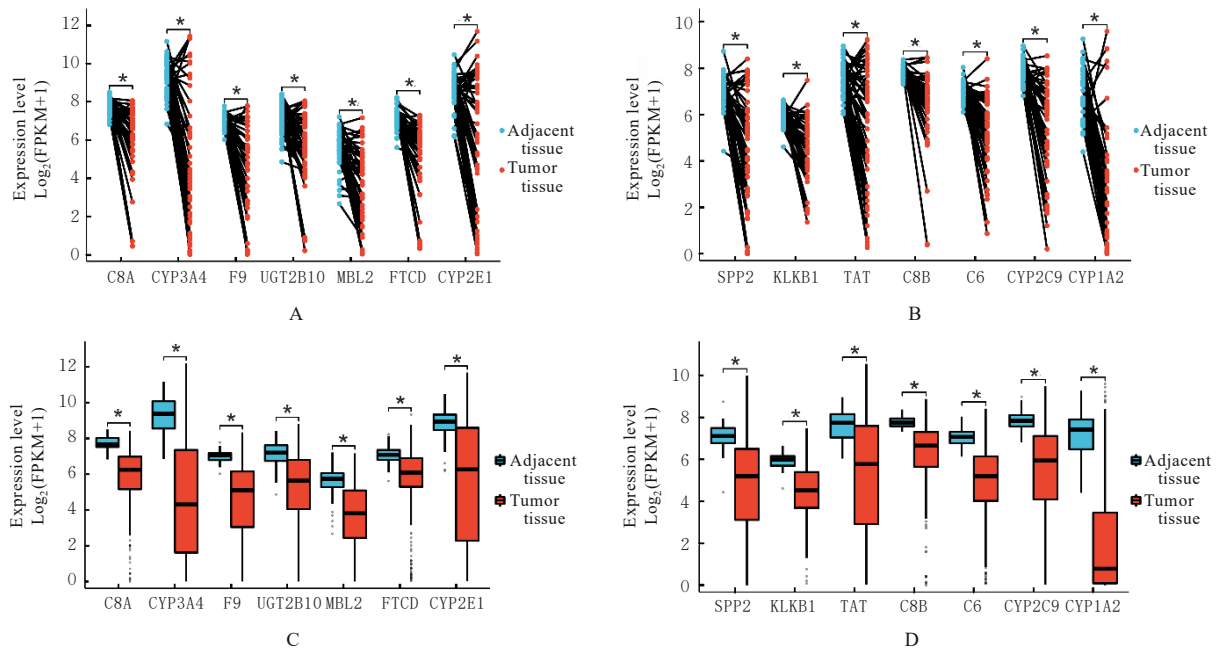
肿瘤组织中14个关键基因表达水平均明显降低 ($P < 0.01$)。



A: PPI network of DEGs; B: Identification of hub genes.

图4 PPI网络构建和关键基因鉴定

Fig. 4 Construction of PPI network and identification of key genes



A, B: Expression levels of genes in liver cancer tissue and matched adjacent tissue; C, D: Expression levels of genes in liver cancer tissue and no-matched adjacent tissue. * $P < 0.01$ vs adjacent tissue.

图5 14个关键基因表达水平分析

Fig. 5 Analysis on expression levels of 14 key genes

2.4 关键基因与HCC患者预后的相关性

Kaplan-Meier生存曲线分析结果显示: 14个关键基因中, UGT2B10、MBL2、KLKB1和CYP1A2表达差异与HCC患者的预后无明显相关性 ($P > 0.05$), 而C8A、CYP3A4、F9、FTCD、CYP2E1、SPP2、TAT、C8B、C6和CYP2C9表达差异与HCC患者预后有明显的相关性 ($P < 0.05$), 为HCC患者预后相关的关键基因(图6)。通过LASSO回归模型对上述基因进行优化, 结果显示: 识别出由5个基因(FTCD、SPP2、TAT、C6和CYP2C9)组成的HCC患者预后预测模型(图7A和7B)。对2种分析方法获得的预后相关基因取交集, 结果表明: FTCD、SPP2、TAT、C6和CYP2C9可作为HCC患者预后相关关键基因(图7C)。相关性分析结果表明: HCC患者5个预后相关关键基因表达水平呈明显正相关关系 ($0.31 \leq r \leq 0.66$), 其中TAT与C6表达的相关性最高 ($r = 0.66$)(图7D)。

2.5 HCC预后相关关键基因的表达验证及其与HCC患者临床病理特征的关系

以GSE29721和GSE45267作为外部数据集, 分析训练集获得的HCC预后相关关键基因的表达, 结果表明: 5个HCC预后相关关键基因的表达模式与训练集相似,

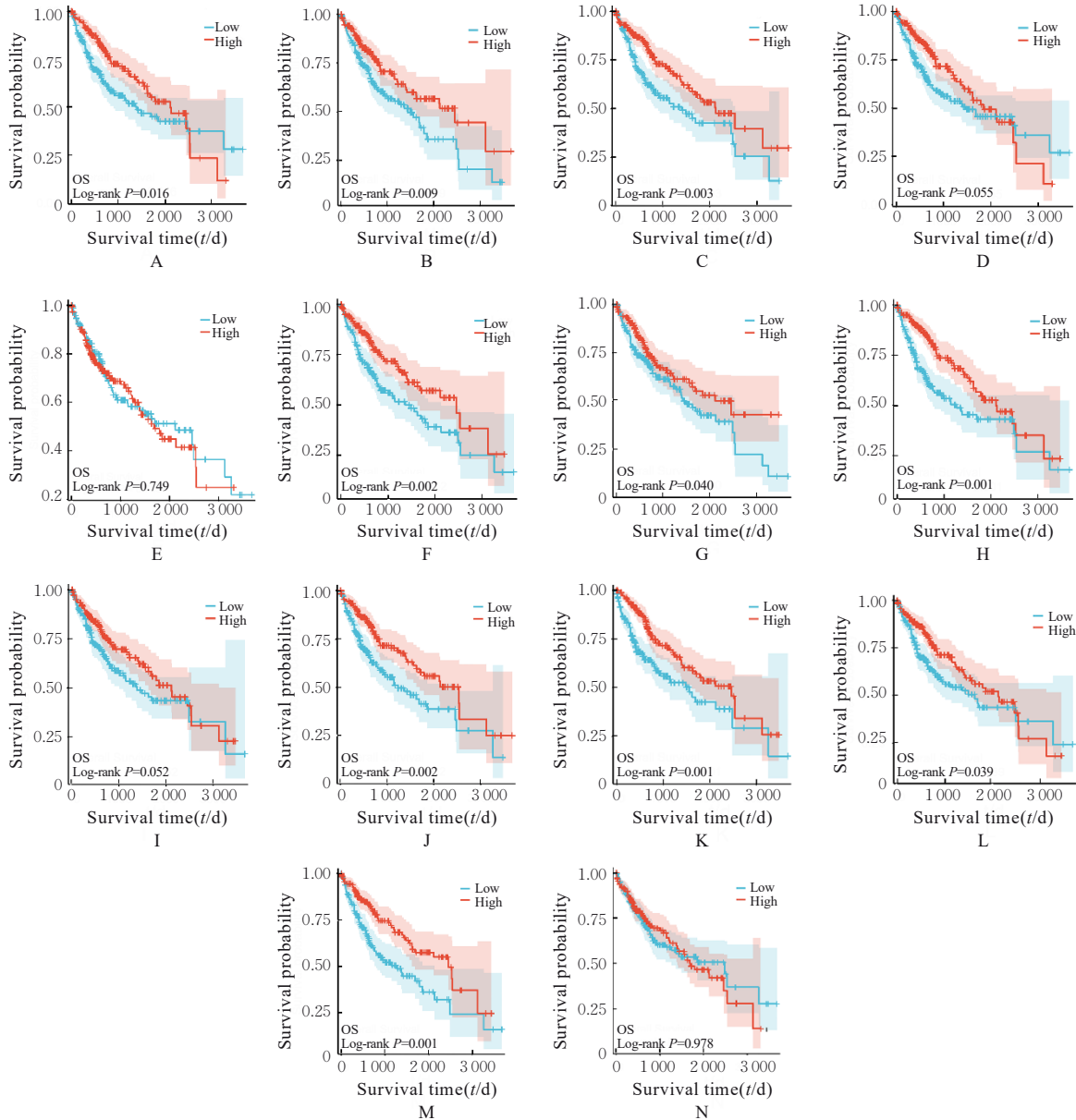
在GSE29721数据集中, 与癌旁正常组织比较, 肿瘤组织中FTCD、SPP2、TAT、C6和CYP2C9表达水平明显下调 ($P < 0.05$ 或 $P < 0.01$)(图8A~8E)。GSE45267数据集中得到相似的结果(图8F~8J)。进一步分析显示: 5个HCC预后相关关键基因与HCC患者病理分期和组织学分级有明显相关关系 ($P < 0.05$ 或 $P < 0.01$)(图9A和9B)。ROC曲线分析结果表明: 5个HCC预后相关关键基因在预测HCC患者病理分期和肿瘤分级方面具有良好的准确性(图9C和9D)。

2.6 HCC预后相关关键基因的诊断效能

利用ROC曲线评估HCC预后相关关键基因FTCD、SPP2、TAT、C6和CYP2C9在HCC诊断中的价值, 其AUC值 > 0.60 , 表明上述5个HCC预后相关关键基因具有较高的诊断价值和准确性, 可以作为HCC的潜在诊断标志物。见图10。

2.7 HCC预后相关关键基因与HCC患者免疫细胞浸润的相关性

应用CIBERSORT算法分析HCC预后相关关键基因与HCC患者免疫细胞浸润的关系, 结果显示: HCC预后相关关键基因均在HCC的免疫调节中发挥作用, 5个预后相关关键基因均与辅助性T细胞17(T helper cell 17, Th17)浸润呈明显正相关关系 ($0.294 \leq r \leq 0.440$), 并与辅助



A: C8A; B: CYP3A4; C: F9; D: UGT2B10; E: MBL2; F: FTCD; G: CYP2E1; H: SPP2; I: KLKB1; J: TAT; K: C8B; L: C6; M: CYP2C9; N: CYP1A2.

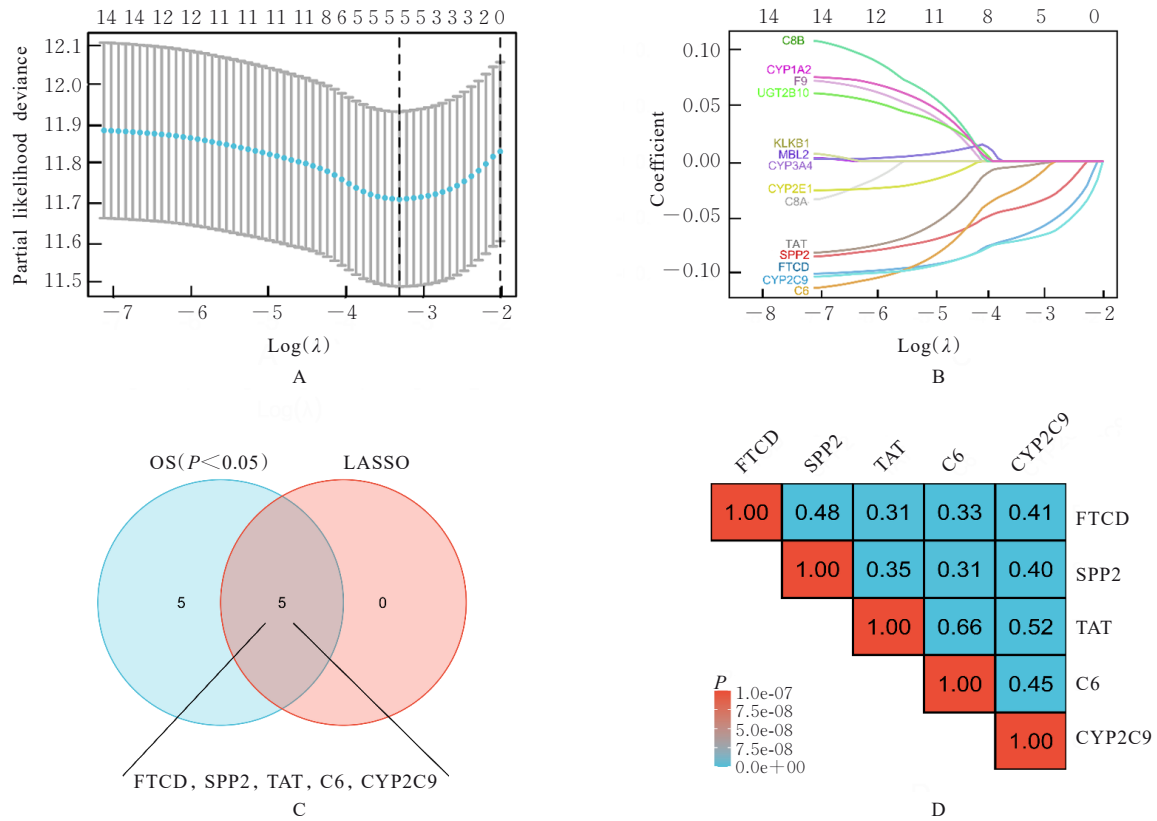
图6 关键基因在HCC患者中的OS曲线
Fig. 6 OS curves of key genes in HCC patients

性T细胞2 (T helper cell 2, Th2) 浸润呈明显负相关关系 ($-0.386 \leq r \leq -0.197$)。见图11。

2.8 miRNA-关键基因和TFs-关键基因调控网络 miRNA-关键基因mRNA网络分析显示: has-miR-182-5p是FTCD、SPP2和C6基因的重要调控miRNA (图12A)。TFs-关键基因mRNA网络分析显示: 有37个TFs可能参与了HCC预后相关关键基因的表达调控 (图12B), 其中调控SPP2、C6和CYP2C9的TFs为转录因子4 (transcription factor 4, TCF4) 和SMAD家族成员4 (SMAD

family member 4, SMAD4); 调控SPP2、TAT和C6的TFs为同源框CUT样蛋白1 (CUT-like homeobox 1, CUX1); 调控FTCD、TAT和C6的TFs为早期生长反应1 (early growth response 1, EGR1); 调控FTCD、C6和CYP2C6的TFs为肿瘤蛋白53 (tumor protein p53, TP53)。表明TCF4、CUX1、EGR1和TP53是关键TFs, 均靶向3个基因。

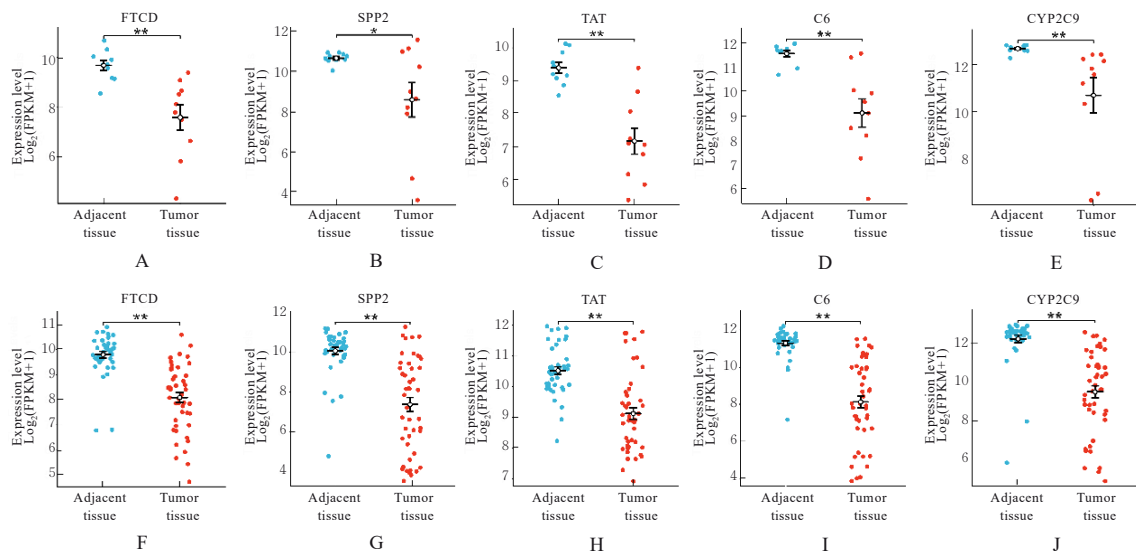
2.9 小分子药物的筛选 通过CMap数据库筛选出呈明显负相关关系的小分子药物20个 (表1),



A, B: Key genes identified by LASSO regression; C: Venn diagram of prognostic gene obtained from Kaplan-Meier analysis and LASSO; D: Correlation analysis on key gene expression.

图7 HCC患者预后相关关键基因的鉴定

Fig. 7 Identification of prognosis-related key genes of HCC patients

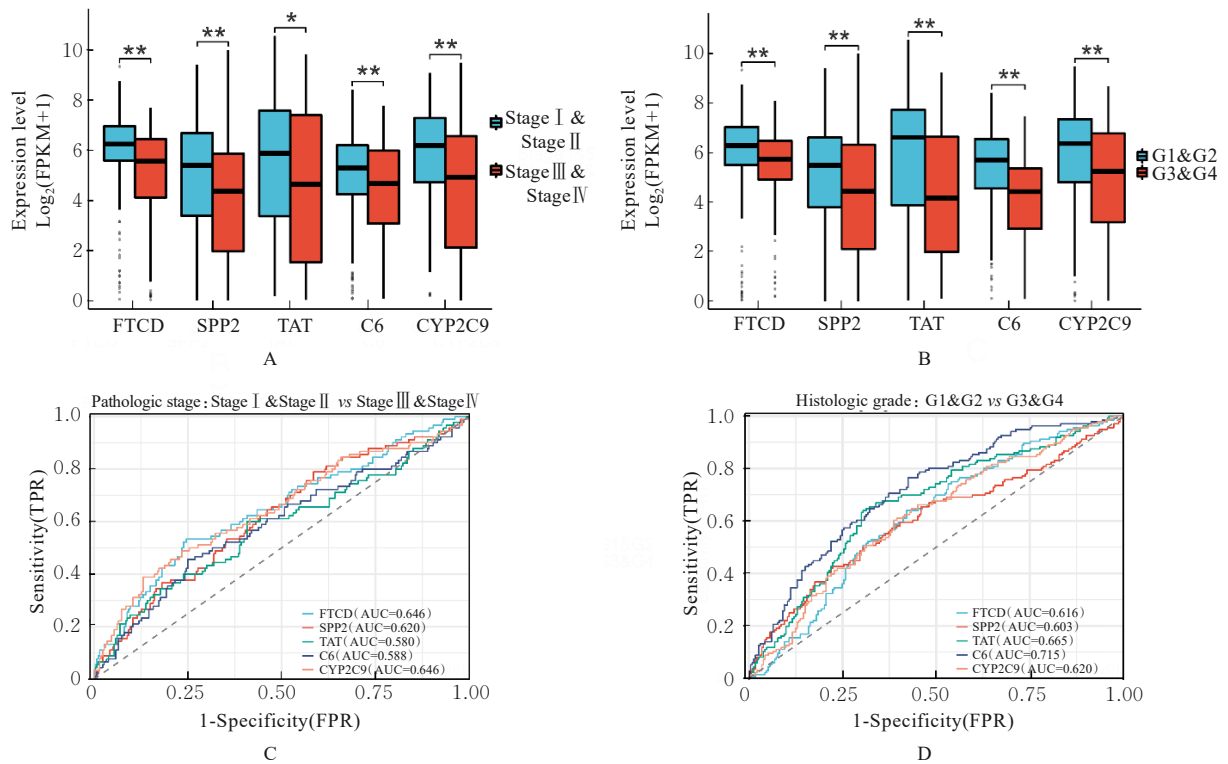


A-E: Expression levels of FTCD, SPP2, TAT, C6, and CYP2C9 in GSE29721 dataset; F-J: Expression levels of FTCD, SPP2, TAT, C6, and CYP2C9 in GSE45267 dataset. * $P < 0.05$, ** $P < 0.01$.

图8 HCC预后相关关键基因表达水平验证

Fig. 8 Validation of prognosis-related key gene expression levels of HCC

DL-硫沙芬 (DL-thiorphan)、异丙嗪 (promethazine) 和芹菜素 (apigenin) 具有较高的富集分数, 分别为 -0.929 , -0.882 和 -0.879 , 其中 DL-thiorphan 的富集分数绝对值最高, 富集分



A: Pathological stage; B: Histologic grade; C: AUC values of pathological stages; D: AUC values of histologic grades. * $P < 0.05$, ** $P < 0.01$.

图9 FTCD、SPP2、TAT、C6和CYP2C9表达水平与HCC患者临床病理特征的相关性

Fig. 9 Correlations between expression levels of FTCD, SPP2, TAT, C6, and CYP2C9 and clinicopathological characteristics of HCC patients

数接近-1。

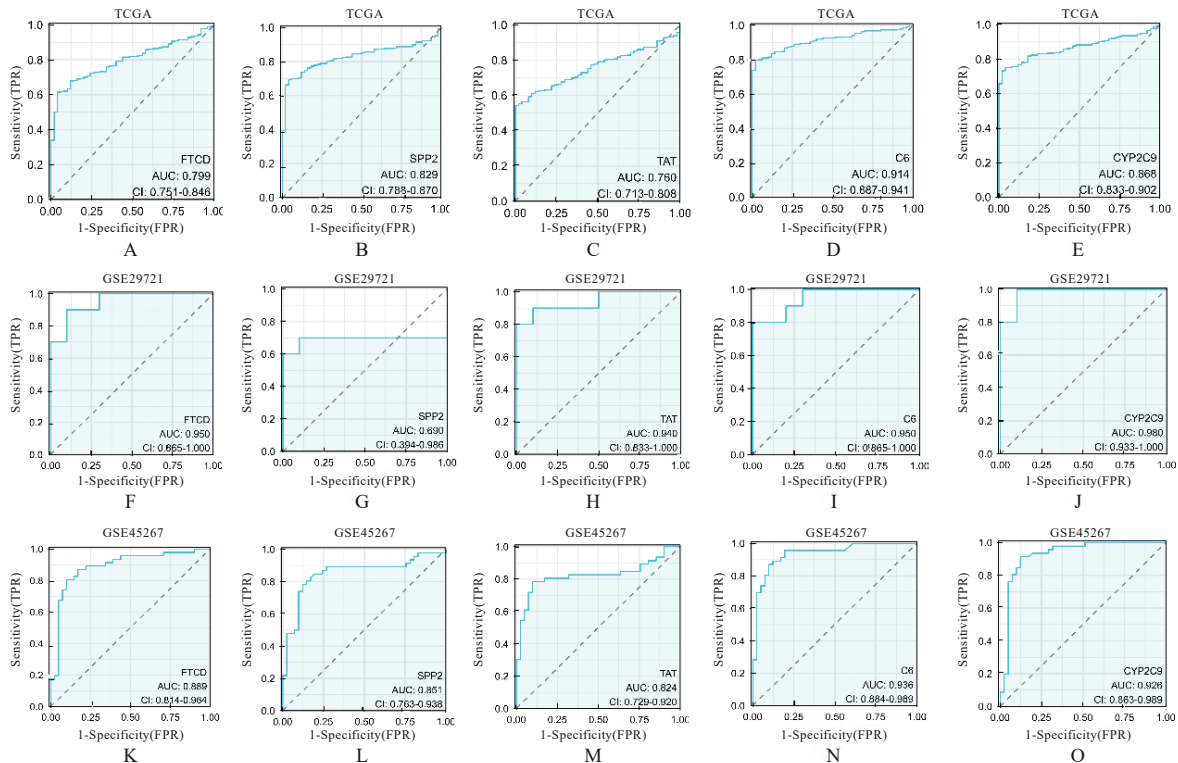
3 讨论

越来越多的证据^[15-16]表明：代谢异常与肿瘤的发生发展和转移有密切关联。本研究利用生物信息学分析方法，鉴定了146个HCC中的DEGs，GO功能富集分析显示其主要富集在类固醇代谢、烯化合物代谢和激素代谢等BP，提示上述代谢过程异常可能促进HCC的发生；KEGG信号通路富集分析结果显示：DEGs主要参与药物代谢-CYP450、视黄醇代谢、补体和凝血级联反应等信号通路。CYP450主要存在于人体肝脏和肠道中，可以催化多种内源性和外源性物质的代谢，在哺乳动物类固醇、药物、致癌物和许多其他化合物的代谢中起重要作用^[17]。研究^[18]显示：在HCC组织中，CYP450表达模式发生改变，酶活性严重受损，监测其表达和活性对于HCC的个体化治疗非常必要。本研究结果提示：DEGs可能通过调节代谢影响HCC的发生，针对上述生物学过程和信号通路开展HCC的靶向治疗可能具有可行性。

基于DEGs的PPI网络，联合Kaplan-Meier生

存分析和LASSO回归算法，本研究确定了在HCC中明显低表达的FTCD、SPP2、TAT、C6和CYP2C9 5个基因是HCC患者预后相关关键基因，ROC曲线结果显示其具有较高的诊断效能。研究^[19-22]显示：FTCD、SPP2、TAT和CYP2C9基因表达下调与HCC的不良预后有关，与本研究结果一致。补体系统形成了抵御入侵病原体的第一道防线，已被证明可以抵御机体的肿瘤细胞^[23]。C6是补体系统的重要成员，主要以糖蛋白的形式在肝脏中合成^[24]。研究^[25-26]显示：食管癌中C6表达水平明显降低，C6中rs9200的杂合GA变异体与癌症复发有关。本研究结果显示：C6与肝脏患者的分期和分级有关，其表达可能受到hsa-miR-182-5p及CUX1、ERG1、SMAD4和TP53的调节，C6可能是HCC新的预后标志物和治疗靶点。

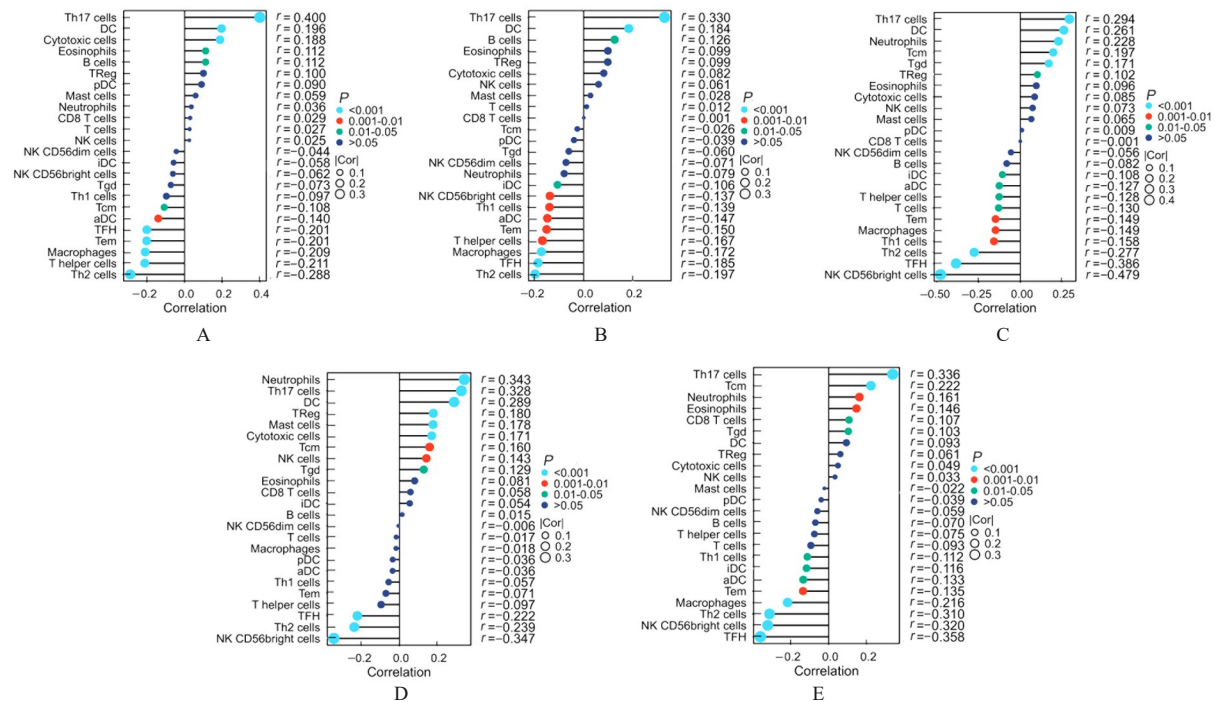
本研究结果显示：FTCD、SPP2、TAT、C6和CYP2C9与HCC中多种浸润性免疫细胞有密切关联，均与Th17浸润呈正相关关系，与Th2浸润呈负相关关系，其中TAT与多种免疫细胞如



A—E: ROC curve analysis on FTCD, SPP2, TAT, C6, and CYP2C9 in TCGA cohort; F—J: ROC curve analysis on FTCD, SPP2, TAT, C6, and CYP2C9 in GSE29721 dataset; K—O: ROC curve analysis on FTCD, SPP2, TAT, C6, and CYP2C9 in GSE45267 dataset.

图 10 HCC 预后相关关键基因的诊断效能

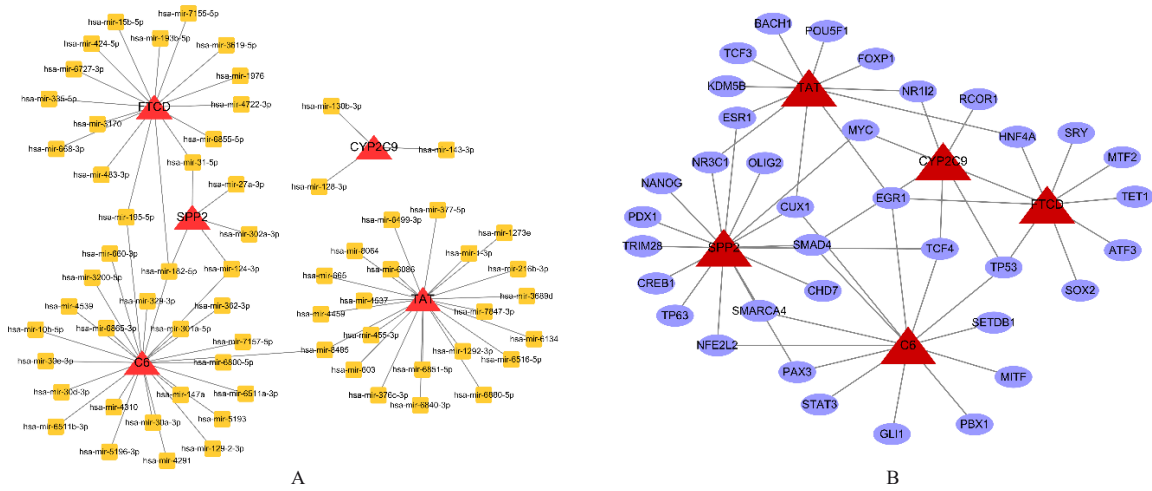
Fig. 10 Diagnostic efficiencies of prognosis-related key genes of HCC



A: FTCD; B: SPP2; C: TAT; D: C6; E: CYP2C9.

图 11 HCC 预后相关关键基因与 HCC 患者免疫细胞浸润的相关性

Fig. 11 Correlations between prognosis-related key genes of HCC and immune cell infiltration of HCC patients



A: MiRNA network; B: TFs-genes network. Prognostic genes were denoted by red triangle nodes; TFs were by purple oval, and miRNAs denoted by yellow square.

图 12 miRNAs 和 TFs 与预后相关关键基因的调控网络

Fig. 12 Regulatory networks of miRNAs and TFs and prognosis-related key genes

表 1 潜在治疗 HCC 的 20 个小分子药物

Tab. 1 Twenty small molecule drugs for potential treatment of HCC

Cmap name	Mean	n	Enrichment	P
DL-thiorphan	-0.787	2	-0.929	0.010
Promethazine	-0.852	4	-0.882	0.000
Apigenin	-0.814	4	-0.879	0.001
Sanguinarine	-0.742	2	-0.867	0.035
Skimmianine	-0.777	4	-0.864	0.001
Verteporfin	-0.724	3	-0.858	0.006
Repaglinide	-0.792	4	-0.831	0.001
Tracazolate	-0.733	4	-0.814	0.002
Milrinone	-0.728	3	-0.809	0.014
Talampicillin	-0.734	4	-0.805	0.003
Dequalinium chloride	-0.730	4	-0.800	0.003
Estriol	-0.736	4	-0.791	0.004
Chrysin	-0.716	3	-0.790	0.019
Monobenzone	-0.721	4	-0.787	0.004
Piperidolate	-0.740	3	-0.786	0.020
Puromycin	-0.712	4	-0.781	0.005
Phenoxybenzamine	-0.698	4	-0.778	0.005
Ethaverine	-0.690	4	-0.776	0.005
Ursodeoxycholicacid	-0.650	3	-0.775	0.023
Trioxysalen	-0.707	4	-0.756	0.007

Th17、树突状细胞 (dendritic cells, DC) 和自然杀伤细胞 (natural killer cell, NK) 呈正相关关系; C6 与多种免疫细胞如中性粒细胞、DCs-调节性 T 细胞 (regulatory T cells, Treg)、肥大细胞、细

胞毒性细胞、中央记忆型 T 细胞 (central memory T cells, TCM) 和 NK 细胞浸润呈正相关关系, 而与辅助性 T 细胞 (T helper cell, Th) 浸润呈负相关关系。研究^[27-29]证实: 肿瘤免疫细胞浸润影响 HCC 患者的预后。Th17 在 HCC 组织中浸润可促进血管生成, 进一步促进疾病进展^[29]。DCs 浸润在 HCC 中表现出强大的抗肿瘤活性, 可能是 HCC 免疫治疗的潜在靶点^[30]。本文作者推测: HCC 预后相关关键基因 FTCD、SPP2、TAT、C6 和 CYP2C9 可能通过介导 Th17 和 Th2 的浸润对 HCC 的发生发展产生影响, TAT 可能是 HCC 免疫治疗干预的靶点。但以上研究结论仍需要进一步的验证。

本研究结果显示: hsa-miR-182-5P 是调节 FTCD、SPP2 和 C6 基因的关键 miRNA。miRNA 在 HCC 发生和耐药性进展中起重要作用^[31]。研究^[32-33]表明: hsa-miR-182-5P 与肝纤维化发生进展有密切关联, 在 HCC 患者中其表达水平明显上调, 增强了 HCC 的侵袭能力, 同时也是接受治疗性手术 HCC 患者早期复发的有价值的预测因子。TFs 包含与特定基因启动子或增强子区域特异性结合的结构域, 其表达可导致靶基因表达的上调或下调^[34]。本研究结果显示: CUX1、ERG1、SMAD4 和 TP53 可能是调节 HCC 预后相关关键基因的转录调节因子, 进一步研究并阐明其调控的分子机制可为 HCC 治疗提供潜在靶点。

HCC 对传统化疗具有耐药性, 所以开发有效的治疗 HCC 的药物具有重要意义。研究^[35]显示:

基于CMap数据库的数据, DL-thiorphan能够逆转非小细胞肺癌(non-small cell lung cancer, NSCLC)中的基因表达变化。本研究结果表明: DL-thiorphan(富集分数为-0.929)是逆转HCC基因异常表达最有前途的小分子药物, 但DL-thiorphan在HCC发展和治疗中的确切作用机制需要进一步研究证实。

综上所述, 本研究通过生物信息学分析, 鉴定了HCC预后相关关键基因特征, 预测了潜在治疗HCC的候选小分子药物。本研究结果为HCC的诊断、预后判断和治疗提供了有潜力的生物标志物、治疗靶点和潜在治疗药物。

利益冲突声明:

所有作者声明不存在利益冲突。

作者贡献声明:

黎金连参与研究设计、数据分析和论文撰写, 黄岚珍和黄希仕参与网络构建和关键基因验证分析, 李康智和蒋佳丽参与数据收集和统计学分析, 张苗苗参与图片整合和论文校对, 吴群英参与研究设计、论文撰写指导和论文审校。

[参考文献]

- [1] 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.
- [2] LLOVET J M, KELLEY R K, VILLANUEVA A, et al. Hepatocellular carcinoma[J]. *Nat Rev Dis Primers*, 2021, 7(1): 6.
- [3] MALUCCIO M, COVEY A. Recent progress in understanding, diagnosing, and treating hepatocellular carcinoma[J]. *CA Cancer J Clin*, 2012, 62(6): 394-399.
- [4] BURKHART R A, RONNEKLEIV-KELLY S M, PAWLIK T M. Personalized therapy in hepatocellular carcinoma: molecular markers of prognosis and therapeutic response[J]. *Surg Oncol*, 2017, 26(2): 138-145.
- [5] LI B H, XU T C, LIU C H, et al. Liver-enriched genes are associated with the prognosis of patients with hepatocellular carcinoma[J]. *Sci Rep*, 2018, 8(1): 11197.
- [6] NAULT J C, REYNIÈS A D, VILLANUEVA A, et al. A hepatocellular carcinoma 5-gene score associated with survival of patients after liver resection [J]. *Gastroenterology*, 2013, 145(1): 176-187.
- [7] TU J X, CHEN J J, HE M M, et al. Bioinformatics analysis of molecular genetic targets and key pathways for hepatocellular carcinoma [J]. *Onco Targets Ther*, 2019, 12: 5153-5162.
- [8] 刘 迁, 祁国萍, 于华裔, 等. 结肠癌核心基因和独立预后因子筛选的生物信息学分析[J]. *吉林大学学报(医学版)*, 2022, 48(3): 755-765.
- [9] 李 楠, 陈 蕾, 许天敏, 等. 子宫内膜异位症患者细胞外基质相关基因筛选的生物信息学分析[J]. *吉林大学学报(医学版)*, 2022, 48(1): 188-194.
- [10] RITCHIE M E, PHIPSON B, WU D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies [J]. *Nucleic Acids Res*, 2015, 43(7): e47.
- [11] MANDREKAR J N. Receiver operating characteristic curve in diagnostic test assessment[J]. *J Thorac Oncol*, 2010, 5(9): 1315-1316.
- [12] LI M H, XIN S Y, GU R Y, et al. Novel diagnostic biomarkers related to oxidative stress and macrophage ferroptosis in atherosclerosis [J]. *Oxid Med Cell Longev*, 2022, 2022: 8917947.
- [13] SHEN L, ZHOU K G, LIU H, et al. Prediction of mechanosensitive genes in vascular endothelial cells under high wall shear stress[J]. *Front Genet*, 2021, 12: 796812.
- [14] WANG Y Y, KANG H E, XU T Y, et al. CeDR Atlas: a knowledgebase of cellular drug response [J]. *Nucleic Acids Res*, 2022, 50(D1): D1164-D1171.
- [15] SNAEBJORNSSON M T, JANAKI-RAMAN S, SCHULZE A. Greasing the wheels of the cancer machine: the role of lipid metabolism in cancer[J]. *Cell Metab*, 2020, 31(1): 62-76.
- [16] CHAN A W, GILL R S, SCHILLER D, et al. Potential role of metabolomics in diagnosis and surveillance of gastric cancer[J]. *World J Gastroenterol*, 2014, 20(36): 12874-12882.
- [17] CHENG X G, GU J, KLAASSEN C D. Adaptive hepatic and intestinal alterations in mice after deletion of NADPH-cytochrome P450 Oxidoreductase (Cpr) in hepatocytes [J]. *Drug Metab Dispos*, 2014, 42(11): 1826-1833.
- [18] YAN T M, LU L L, XIE C, et al. Severely impaired and dysregulated cytochrome P450 expression and activities in hepatocellular carcinoma: implications for personalized treatment in patients[J]. *Mol Cancer Ther*, 2015, 14(12): 2874-2886.
- [19] YU Z H, GE Y Y, XIE L, et al. Using a yeast two-

- hybrid system to identify FTCD as a new regulator for HIF-1 α in HepG2 cells[J]. *Cell Signal*, 2014, 26(7): 1560-1566.
- [20] LU C Y, FANG S J, WENG Q Y, et al. Integrated analysis reveals critical glycolytic regulators in hepatocellular carcinoma [J]. *Cell Commun Signal*, 2020, 18(1): 97.
- [21] FU L, DONG S S, XIE Y W, et al. Down-regulation of tyrosine aminotransferase at a frequently deleted region 16q22 contributes to the pathogenesis of hepatocellular carcinoma[J]. *Hepatology*, 2010, 51(5): 1624-1634.
- [22] WANG Q, TANG Q, ZHAO L J, et al. Time serial transcriptome reveals Cyp2c29 as a key gene in hepatocellular carcinoma development [J]. *Cancer Biol Med*, 2020, 17(2): 401-417.
- [23] ROUMENINA L T, DAUGAN M V, PETITPREZ F, et al. Context-dependent roles of complement in cancer[J]. *Nat Rev Cancer*, 2019, 19(12): 698-715.
- [24] HOBART M J, FERNIE B A, DISCIPIO R G, et al. A physical map of the C6 and C7 complement component gene region on chromosome 5p13[J]. *Hum Mol Genet*, 1993, 2(7): 1035-1036.
- [25] OKA R, SASAGAWA T, NINOMIYA I, et al. Reduction in the local expression of complement component 6 (C6) and 7 (C7) mRNAs in oesophageal carcinoma[J]. *Eur J Cancer*, 2001, 37(9): 1158-1165.
- [26] WANG Z, LIAO J, WU S, et al. Recipient C6 rs9200 genotype is associated with hepatocellular carcinoma recurrence after orthotopic liver transplantation in a Han Chinese population[J]. *Cancer Gene Ther*, 2016, 23(6): 157-161.
- [27] GARNELO M, TAN A, HER Z, et al. Interaction between tumour-infiltrating B cells and T cells controls the progression of hepatocellular carcinoma [J]. *Gut*, 2017, 66(2): 342-351.
- [28] CHEN Q F, LI W, WU P H, et al. Significance of tumor-infiltrating immunocytes for predicting prognosis of hepatitis B virus-related hepatocellular carcinoma [J]. *World J Gastroenterol*, 2019, 25(35): 5266-5282.
- [29] KUANG D M, PENG C, ZHAO Q Y, et al. Activated monocytes in peritumoral stroma of hepatocellular carcinoma promote expansion of memory T helper 17 cells[J]. *Hepatology*, 2010, 51(1): 154-164.
- [30] SUN K, WANG L, ZHANG Y Y. Dendritic cell as therapeutic vaccines against tumors and its role in therapy for hepatocellular carcinoma [J]. *Cell Mol Immunol*, 2006, 3(3): 197-203.
- [31] WEI L, WANG X W, LV L Y, et al. The emerging role of microRNAs and long noncoding RNAs in drug resistance of hepatocellular carcinoma [J]. *Mol Cancer*, 2019, 18(1): 147.
- [32] VAN KEUREN-JENSEN K R, MALENICA I, COURTRIGHT A L, et al. MicroRNA changes in liver tissue associated with fibrosis progression in patients with hepatitis C[J]. *Liver Int*, 2016, 36(3): 334-343.
- [33] CAO M Q, YOU A B, ZHU X D, et al. MiR-182-5p promotes hepatocellular carcinoma progression by repressing FOXO3a[J]. *J Hematol Oncol*, 2018, 11(1): 12.
- [34] LAMBERT S A, JOLMA A, CAMPITELLI L F, et al. The human transcription factors[J]. *Cell*, 2018, 175(2): 598-599.
- [35] WU Q, ZHANG B, SUN Y D, et al. Identification of novel biomarkers and candidate small molecule drugs in non-small-cell lung cancer by integrated microarray analysis[J]. *Onco Targets Ther*, 2019, 12: 3545-3563.