



doi:10.7659/j.issn.1005-6947.2022.08.008
http://dx.doi.org/10.7659/j.issn.1005-6947.2022.08.008
Chinese Journal of General Surgery, 2022, 31(8):1048-1063.

· 基础研究 ·

基于生物信息学的肝内胆管癌差异表达基因谱中关键基因的筛选及分析

陈伟毅, 陈立军

(湖南医药学院医学院, 湖南 怀化 418000)

摘要

背景与目的: 肝内胆管癌 (ICC) 是指来源于肝内胆管上皮的一种恶性肿瘤, 其发病隐匿, 恶性程度高。ICC 早期无明显临床表现, 大多数患者确诊时往往已失去手术机会, 因此预后极差。寻找 ICC 的早期诊断和治疗靶标具有重要意义, 因此, 本研究通过生物信息学方法对影响 ICC 发生发展的关键基因进行筛选。

方法: 从 GEO 数据库下载 2 个 ICC 转录组数据集 (GSE107943、GSE119336)。用 R 语言的 edgeR 包筛选出差异表达基因, 然后对这些差异表达基因进行 GO 和 KEGG 通路富集分析。通过 STRING 数据库建立差异表达基因的蛋白质-蛋白质相互作用 (PPI) 网络, 使用 Cytoscape 中的 MCODE 插件发掘出关键调控基因。分析关键调控基因在肿瘤组织中的表达, 采用 UALCAN、GEPIA 数据库进行验证。通过 UCSC XENA 数据库分析关键调控基因在泛癌中的表达。利用 TCGA 数据库分析关键调控基因共表达基因。采用 UALCAN、GEPIA 数据库分析关键调控基因与患者预后、肿瘤分级、分期、淋巴转移的关系。使用 R 语言 GSVA 包计算关键调控基因表达与免疫浸润相关性。绘制 ROC 曲线评价关键调控基因对 ICC 的预测能力。采用细胞实验验证关键调控基因的表达。

结果: 共筛选出 1 094 个共同差异表达基因, 其中共同上调的基因为 567 个, 共同下调的基因为 527 个, 主要参与小分子分解代谢、有机酸生物合成、碳代谢等过程。通过 PPI 网络挖掘出 3 个关键调控基因 Polo 样激酶 1 (PLK1)、羟基酸氧化酶 2 (HAO2)、纤维胶凝蛋白 (FCN2), 其中 PLK1 基因在肿瘤组织中明显上调, HAO2 和 FCN2 基因在肿瘤组织明显下调, 通过 UALCAN、GEPIA 数据库验证发现 3 个基因表达与分析结果一致。通过 UCSC XENA 数据库分析发现 PLK1 在 28 种肿瘤中表达显著增高, HAO2 在 24 种肿瘤表达显著降低, FCN2 在 27 种肿瘤表达显著降低。TCGA 数据库分析发现 PLK1 与 CCNA2、GTSE1 等基因共表达, HAO2 与 MTTP、CPS1 等基因共表达, FCN2 与 FAM99A、GDF2 等基因共表达。UALCAN 数据库分析发现 3 个基因表达与肿瘤的分期和分级、淋巴转移有关, 其中 PLK1 高表达、HAO2、FCN2 低表达提示肿瘤分期更高、分化更差、更易出现淋巴转移。相关性分析发现 PLK1 表达与 Th2 cells 浸润呈显著正相关, FCN2 表达与 aDC 浸润呈显著负相关。绘制 ROC 曲线显示这 3 个基因都可以很好地诊断 ICC, 其中 HAO2 的诊断能力最好。细胞实验结果发现 PLK1 在 RBE 中表达明显升高, HAO2、FCN2 在 RBE 中表达明显降低 (均 $P < 0.01$)。

结论: PLK1、HAO2、FCN2 可能是影响 ICC 发生发展的关键调控基因, 这 3 个基因可能成为 ICC 诊治的新靶点。

关键词

胆管肿瘤; 胆管, 肝内; 基因表达谱; 计算生物学

中图分类号: R735.8

基金项目: 湖南省自然科学基金资助项目 (2022JJ50290); 湖南省卫生健康委科研课题基金资助项目 (202104081483); 2021、2022 年度湖南省大学生创新创业训练计划资助项目 (4314、5021); 湖南医药学院 2020 年科研孵化库建设项目第二批基金资助项目。

收稿日期: 2022-02-23; **修订日期:** 2022-07-22。

作者简介: 陈伟毅, 湖南医药学院讲师, 主要从事消化系统肿瘤防治方面的研究。

通信作者: 陈伟毅, Email: cwyeason@163.com

Screening and analysis of key genes in differential expression gene profile of intrahepatic cholangiocarcinoma based on bioinformatics

CHEN Weiyi, CHEN Lijun

(Medical College, Hunan University of Medicine, Huaihua Hunan 418000, China)

Abstract

Background and Aims: Intrahepatic cholangiocarcinoma (ICC) is a malignant tumor arising from the intrahepatic bile duct epithelium, which has an insidious onset and high degree of malignancy. Because there are no obvious clinical symptoms in the early stage of ICC, and most patients have lost the opportunity for surgery at the time of diagnosis, so its prognosis is very poor. Exploration of targets for early diagnosis and treatment of ICC is of great significance. Therefore, this study was conducted to screen the key genes involved in the occurrence and development of ICC.

Methods: Two ICC transcriptome datasets (GSE107943, GSE119336) were downloaded from the GEO database. The differentially expressed genes were screened by edgeR package of R language, and then these genes underwent GO and KEGG pathway enrichment analysis. The protein-protein interaction (PPI) networks of these genes was constructed by using STRING database, and the key protein regulatory genes were mined by using the MCODE plug-in of Cytoscape. The expressions of key protein regulatory genes in tumor tissues were analyzed and verified by UALCAN and GEPIA databases. UCSC XENA database was used to analyze the expressions of key regulatory genes in generalized carcinoma. TCGA database was used to analyze the co-expression genes of the key regulatory genes. UALCAN and GEPIA databases were used to analyze the relationship between key regulatory genes and patient prognosis, tumor grade, stage and lymphatic metastasis. The correlation between the expressions of the regulatory genes and immune infiltration were calculated using R language GSEA package. The predictive abilities of the key protein regulatory genes to ICC were evaluated by drawing ROC curve. Cell experiments were performed to verify the expressions of the key regulatory genes.

Results: A total of 1 094 common differentially expressed genes were screened, including 567 up-regulated genes and 527 down-regulated genes, which were mainly involved in the process of small molecule catabolism, organic acid biosynthesis, carbon metabolism and so on. Three key genes Polo-like kinase 1 (PLK1), hydroxyacid oxidase 2 (HAO2) and ficolin-2 (FCN2) were mined through PPI networks. PLK1 gene was significantly up-regulated in tumor tissues, and HAO2 and FCN2 genes were significantly down-regulated in tumor tissues, which were verified by UALCAN and GEPIA databases. The analysis of UCSC XENA database showed that the expression of PLK1 was significantly increased in 28 types of tumors, the expression of HAO2 was significantly decreased in 24 types of tumors, and the expression of FCN2 was significantly decreased in 27 types of tumors. The analysis of TCGA database showed that PLK1 was co-expressed with CCNA2 and GTSE1, HAO2 was co-expressed with MTTP and CPS1, and FCN2 was co-expressed with FAM99A and GDF2. The analysis of the UALCAN database found that the expression of three genes was related to the stage and grade of the tumor and lymph node metastasis. Among them, high expression of PLK1, low expression of HAO2 and FCN2 were associated with higher tumor stage, worse differentiation and more prone to lymph node metastasis. The correlation analysis found that the expression of PLK1 was significantly positively correlated with the infiltration of Th2 cells, and the expression of FCN2 was significantly negatively correlated with the infiltration of aDC cells. The ROC curve showed that all the three genes could diagnose ICC well,

among which HAO2 had the best diagnostic ability. The results of cell experiments showed that the expression of PLK1 was significantly increased, while the expression of HAO2 and FCN2 were significantly decreased in RBE (all $P < 0.01$).

Conclusion: LK1, HAO2 and FCN2 may be the key protein regulatory genes involved in the occurrence and progression of ICC. These three genes may probably become new targets for the diagnosis and treatment of ICC.

Key words Bile Duct Neoplasms; Bile Ducts, Intrahepatic; Gene Expression Profiling; Computational Biology

CLC number: R735.8

肝内胆管癌 (intrahepatic cholangiocarcinoma, ICC) 是指起源于肝内胆管上皮的一种恶性肿瘤。ICC 约占肝脏原发恶性肿瘤的 10%~15%，是发病率仅次于肝细胞肝癌的肝脏原发恶性肿瘤^[1]。ICC 早期往往无典型临床症状，大多数患者确诊时往往已失去了手术的机会，而未行手术治疗者，预后极差^[2]。因此对 ICC 的早期诊断对于改善患者的预后具有重要意义。

目前临床上诊断 ICC 所使用的肿瘤标记物特异性及敏感度欠缺，诊断价值很低^[3]，寻找新的具有临床诊断价值的潜在生物标志物对提高 ICC 患者的早期诊断率具有重要作用。

随着生物学技术的发展，二代测序技术已广

泛应用于探索肿瘤新的生物标志物和治疗靶点^[4]。现今研究^[5]表明，基因组的不稳定性以及变异性可能会对 ICC 的发生和发展产生影响。本研究通过生物信息学方法筛选 ICC 的关键调控基因，评估关键调控基因对 ICC 的预测能力，为诊断和治疗 ICC 患者提供新的思路，为深入认识 ICC 疾病提供新的方法。

1 材料与方法

1.1 原始数据下载

从公共数据库 GEO 中下载 2 个 ICC 转录组数据集，编号分别为 GSE107943 和 GSE119336。数据集信息、检测平台以及样本分类等信息具体见表 1。

表 1 ICC 转录组数据信息和样本分类

Table 1 ICC transcriptome data information and sample classification

数据集编号	样本总数	肿瘤样本数	正常样本数	检测平台
GSE107943	57	27	30	GPL18573
GSE119336	30	15	15	GPL11154

1.2 方法

1.2.1 转录组数据预处理过程 GSE107943 数据集预处理过程：(1) 使用 MAP-R Seq 管道进行分析 RNA 测序样本。(2) 采用 bowtie1 软件进行选项比对，双端读段由 TopHatv2 对比人类参考基因组 hg19 后构建。(3) 使用 FeatureCounts 软件量化基因表达，每个样本的映射读段与 ENSEMBL 的 GRCh37.75 RNA 特征参考基因定义文件对齐，然后量化。(4) 将与每个 RNA 特征对齐的原始读取数标准化为 RPKM 值。GSE119336 数据集预处理过程：(1) 测序读段被修剪为适配器序列，屏蔽低复杂性或低质量序列。(2) 使用 TopHat (v2.0.10) 将序列与人类参考基因

组 (hg19 版本) 和 RefSeq 注释基因进行比对，发现转录本剪接位点。(3) 根据 TopHat 生成的基因图谱，使用 Cufflinks (v2.1.1) 评估基因表达水平。(4) 将每个基因的表达水平标准化为 FPKM 值，以便于样本之间的转录水平比较。

1.2.2 差异表达基因筛选 转录组数据预处理后，采用 R 语言的 edgeR 包对 ICC 组织和正常组织进行差异表达基因筛选，设置校正 $P < 0.01$ 且 $|\log_2 FCI| \geq 2$ 为阈值条件。

1.2.3 差异表达基因的 GO 和 KEGG 富集分析 使用 R 语言的 clusterProfiler、org.Hs.eg.db 程序包对差异表达基因进行 GO 功能 (包括生物学过程 BP、细胞

组分CC、分子功能MF)和KEGG通路富集分析,并将TOP5的显著富集结果以气泡图展示。

1.2.4 差异表达基因的蛋白质-蛋白质相互作用网络(Protein-Protein Interaction, PPI) Networks构建与关键模块挖掘 采用STRING数据库(<https://www.string-db.org>)对差异表达基因进行PPI网络分析,设置combined score>0.9为阈值条件,然后使用Cytoscape软件对其进行可视化。采用Cytoscape软件中的MCODE插件对PPI网络中的关键子网络进行辨别,并从中挖掘出关键调控基因,设置参数为:Degree Cutoff=2, Node Score Cutoff=0.2, Max. Depth=100, K-Core=2。

1.2.5 关键调控基因的表达分析与验证 通过SPSS26.0分析关键调控基因在ICC组织和正常组织中的表达,分析方法采用独立样本 t 检验,采用UALCAN、GEPIA数据库对基因表达进行验证。

1.2.6 关键调控基因的泛癌表达分析 在UCSC XENA数据库(<https://xenabrowser.net/datapages>)下载经统一处理的TCGA和GTEx的RNAseq数据,使用R语言的ggplot2包对关键调控基因进行泛癌表达分析。

1.2.7 关键调控基因在ICC中的共表达基因分析、PPI网络分析、富集分析 在TCGA数据库(<https://portal.gdc.cancer.gov>)下载胆管癌RNAseq数据,使用R语言的stat包分析编码基因与关键调控基因的相关系数,选取相关系数TOP10的编码基因进行可视化。采用STRING数据库对相关系数TOP10的编码基因和关键调控基因进行PPI网络分析,利用Metascape网站(<http://metascape.org>)进行富集分析。

1.2.8 关键调控基因的表达与肿瘤分期、分级、淋巴转移、预后关系的分析 利用UALCAN数据库分析关键调控基因与肿瘤分期、分级、淋巴转移的相关性。利用GEPIA数据库分析关键调控基因与患者预后的相关性。

1.2.9 关键调控基因表达对ICC免疫细胞浸润的影响分析 在TCGA数据库下载胆管癌RNAseq数据,使用R语言GSVA包计算免疫浸润相关性及其显著性。

1.2.10 绘制ROC曲线评价关键调控基因对ICC的诊断能力 使用Graphpad 8.0软件绘制受试者工作特征曲线(ROC曲线)。ROC曲线下面积(AUC)越接近于1.0,其诊断效果越好。

1.2.11 细胞及培养 人肝内胆管上皮细胞(HIBEC)购于北京北纳创联生物技术研究院,人肝内胆管癌细胞(RBE)购于中科院上海细胞库,两种细胞株均采用含10%的胎牛血清的RPMI 1640,培养于37℃、5%CO₂孵箱中,每48h进行消化传代。

1.2.12 RT-qPCR检测mRNA表达 采用TRIzol法(美国Invitrogen公司)提取细胞总RNA,用分光光度计检测RNA浓度,用逆转录试剂盒(日本TaKaRa公司)将RNA逆转录成cDNA,然后用SYBR Green法(美国Thermo Fisher公司)进行实时荧光定量PCR,以GAPDH为内参,采用2^{- $\Delta\Delta C_t$} 法计算目的基因的相对表达量。Polo样激酶1(Polo-like Kinase, PLK1)上游引物:5'-GCT GGG CAA CCT TTT CCT G-3',下游引物:5'-CCA GTG GGA TCT GTC TGA AGC-3';羟基酸氧化酶(hydroxyacid oxidase 2, HAO2)上游引物:5'-TGA CAG ACT TTC AGG CCC AT-3',下游引物:5'-CAC TGA TCT CCT CCC CTT GG-3';纤维胶凝蛋白(ficolin-2, FCN2)上游引物:5'-CTG CCA TGT GT CAA ACC TGA A-3',下游引物:5'-TTC CCC GAC TTC CAG TTG ATG-3';GAPDH上游引物:5'-GAG TCA ACG GAT TTG GTC GT-3',下游引物:5'-GAC AAG CTT CCC GTT CTC AG-3'。

1.2.13 Western blot检测蛋白表达 提取细胞蛋白,采用BCA法测量蛋白浓度,每组取60g蛋白样品进行电泳,之后将蛋白转移到PDVF膜上,5%脱脂奶粉封闭后加入一抗4℃孵育过夜,再加入二抗孵育2h,曝光显影,以GAPDH为内参计算蛋白相对表达量。PLK1、HAO2、FCN2、GAPDH抗体均购自美国Abcam公司。

2 结果

2.1 ICC差异表达基因分析结果

采用R语言的edgeR包对两个转录组数据进行分析,设置校正 $P<0.01$ 且 $|\log_2FCI|\geq 2$ 为阈值条件。GSE107943数据集中共筛选出2242个(3.88%)差异表达基因,其中上调基因1320个,下调基因922个(图1A)。GSE119336数据集中共筛选出1663个(6.50%)差异表达基因,其中上调基因904个,下调基因759个(图1B)。两个数据集共同的差异表达基因为1123个(图1C),其中有

567个基因为共同上调基因(图1D), 527个基因为共同下调基因(图1E)。表2罗列了2个数据集中差异表达最大的5个共同上调和共同下调基因。

共同上调和共同下调的基因共1094个。在ICC的发生机制中这1094个差异表达基因可能起着非常重要的作用, 故后续分析都根据此。

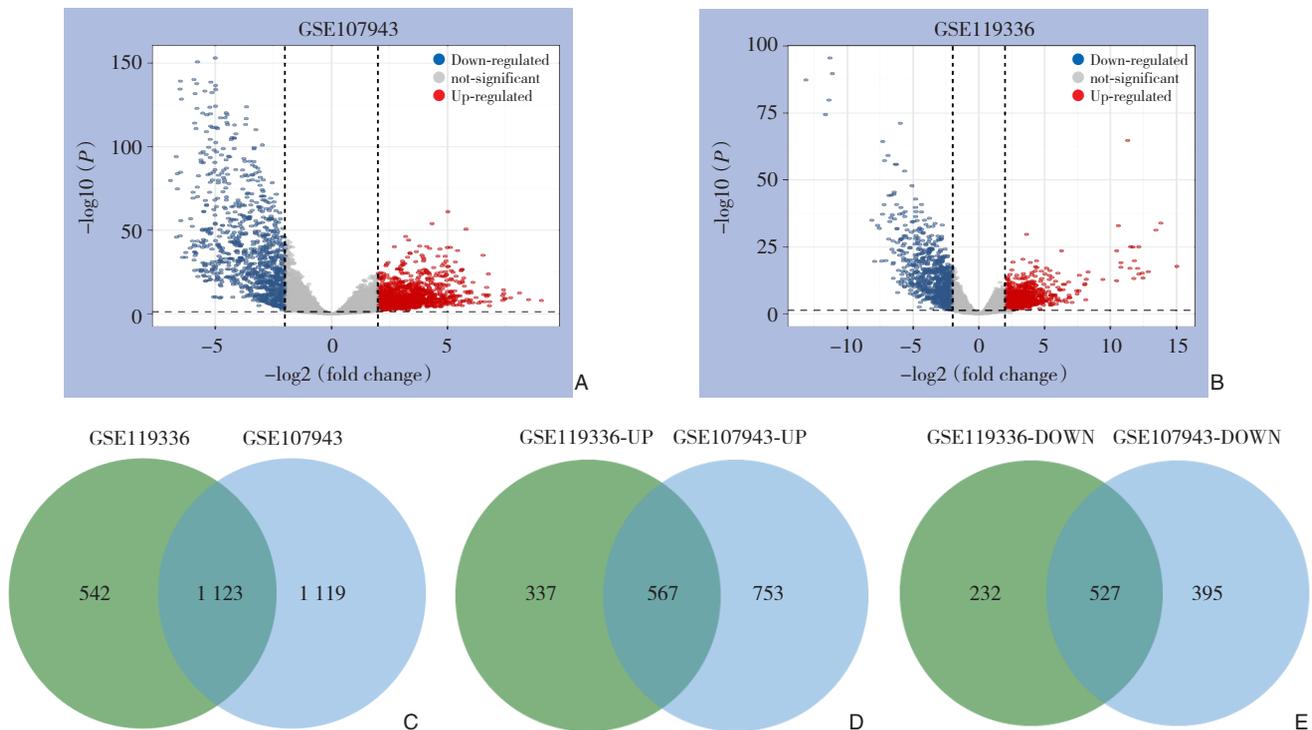


图1 ICC差异表达基因的分析结果 A: GSE107943差异表达基因火山图; B: GSE119336差异表达基因火山图; C: 两个数据集共同差异表达基因 Venn图; D: 两个数据集共同上调基因 Venn图; E: 两个数据集共同下调基因 Venn图

Figure 1 Results of analysis of differentially expressed genes in ICC A: Volcano map of GSE107943; B: Volcano map of GSE119336; C: Venn diagram of differentially expressed genes both in GSE107943 and GSE119336; D: Venn diagram of up-regulated differentially expressed genes; E: Venn diagram of down-regulated differentially expressed genes

表2 两个数据集中差异表达最大的5个共同上调和下调基因
Table 2 The five co-up-regulated and co-down-regulated genes

GSE119336	校正P	\log_2FC	GSE107943	校正P	\log_2FC
上调基因			上调基因		
LY6D	5.29E-06	8.1210	FDCSP	3.03E-09	8.5009
CST1	5.58E-12	8.0653	OLFM4	2.36E-10	7.7281
RPAEP	2.30E-09	7.6106	EG4	6.91E-10	7.4770
FDCSP	3.63E-08	7.4690	MSLN	5.68E-13	7.4348
CEACAM7	2.26E-09	7.4238	KRT17	4.61E-15	7.4193
下调基因			下调基因		
CYP2A6	8.38E-34	-7.7034	SLC22A1	2.10E-80	-6.9442
HAMP	1.81E-20	-7.3721	CYP3A4	1.24E-46	-6.6824
CYP4A22	6.06E-38	-7.3140	GNMT	1.82E-75	-6.6456
APOF	4.04E-65	-7.2976	UROC1	4.05E-135	-6.5283
UROC1	1.58E-41	-7.2034	TAT	2.31E-47	-6.5206

2.2 ICC差异表达基因功能和通路富集分析

对1094个差异表达基因进行GO富集分析结果显示: 1094个共同差异表达基因主要参与BP的小分子分解代谢、有机酸生物合成、羧酸生物合成、羧酸分解代谢、有机酸分解代谢等过程(图2A); CC主要与含胶原的细胞外基质、细胞顶端、内质网内腔、顶端质膜、血液微粒等有关(图2B); MF主要与酶结合、硫化物结合、羧酸结合、维生素结合、单加氧酶活性等有关(图2C)。KEGG通路分析发现, 1094个共同差异表达基因主要参与碳代谢、补体与凝血级联、氨基酸生物合成、缬氨酸、亮氨酸和异亮氨酸降解、甘氨酸、丝氨酸和苏氨酸代谢等(图2D)。

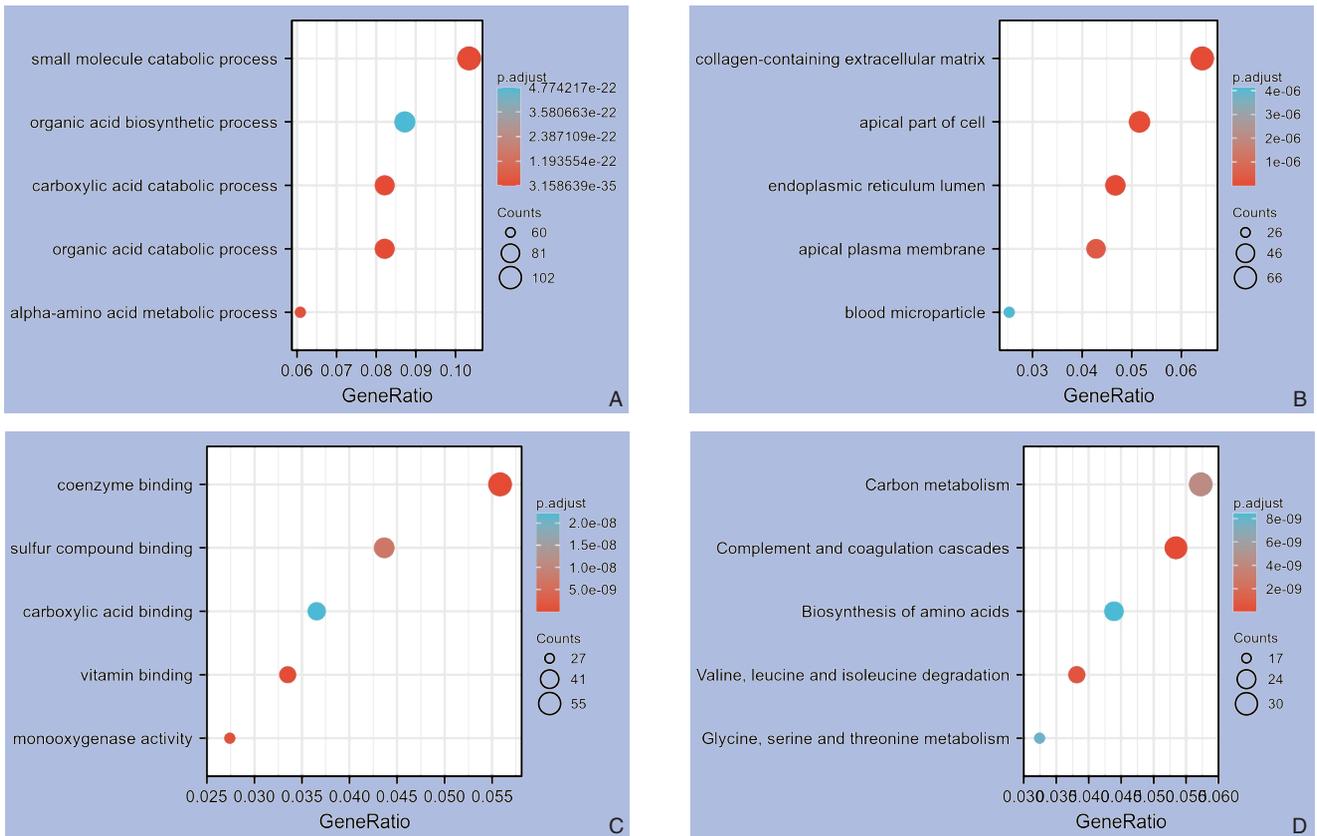


图2 ICC共同差异表达基因的功能和通路富集 A: 共同差异表达基因的GO_BP富集; B: 共同差异表达基因的GO_CC富集; C: 共同差异表达基因的GO_MF富集; D: 共同差异表达基因的KEGG通路富集

Figure 2 GO and KEGG enrichment of common differentially expressed genes in ICC A: GO_BP enrichment; B: GO_CC enrichment; C: GO_MF enrichment; D: GO_BP enrichment

2.3 ICC差异表达基因的PPI网络分析及关键子网络分析

采用STRING数据库对1 094个共同差异表达基因进行PPI网络的构建,使用Cytoscape进行可视化,该PPI网络共有451个节点和1288条边(图3)。

使用MCODE插件一共筛选出17个关键子网络,取得分最高的前3个子网络进行分析(表3)(图4),从中挖掘出3个关键调控基因PLK1、HAO2、FCN2(表4)。

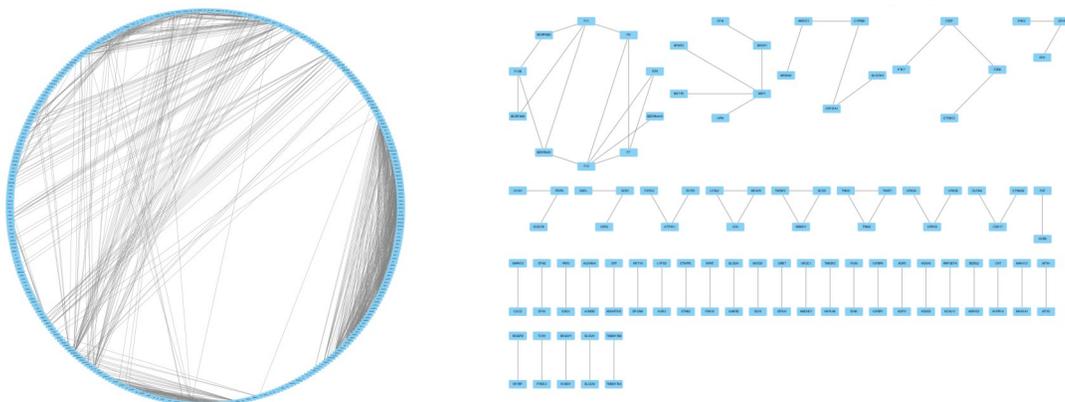


图3 ICC差异表达基因的PPI网络

Figure 3 PPI network of the differentially expressed genes in ICC

表 3 得分前三的子网络模块信息

Table 3 Information of the top 3 sub-network modules

模块	得分	节点	边	节点基因名称
1	25.481	28	344	CENPF、TPX2、KIF11、KIF23、PLK1、BIRC5、PTTG1、CENPA、KIF15、CCNB2、UBE2C、CDCA8、CDC20、NDC80、NUSAP1、KIF2C、MKI67、TTK、BUB1B、CCNA2、FOXM1、NUF2、CDCA3、AURKA、KIF20A、CEP55、CDK1、TOP2A
2	10.762	22	113	SLC27A2、CYP4A11、HMGCL、CYP2A6、ACOX1、EHHADH、MLYCD、CYP26A1、CYP3A5、HAO2、CYP1A1、HAO1、AGXT、CYP3A7、EPHX2、CYP4A22、ACOX2、AOX1、NUDT7、CYP3A4、PECR、CYP2B6
3	6.667	7	20	FCN3、C4A、MBL2、C2、MASP2、MASP1、FCN2

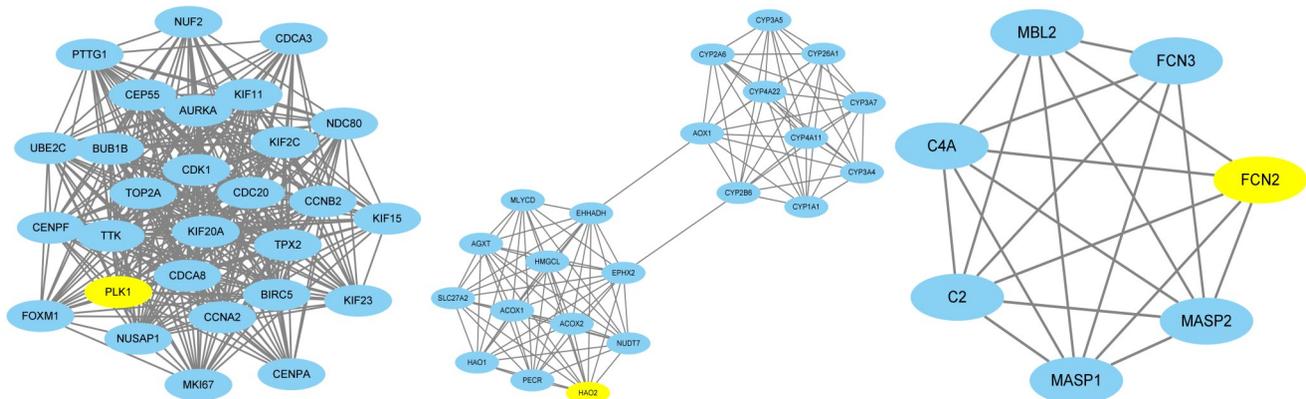


图 4 ICC 差异表达基因 PPI 网络中 3 个关键子网络

Figure 4 Three key subnetworks in ICC differential expression gene PPI network

表 4 关键蛋白调控功能基因节点信息

Table 4 Node information of the key genes

节点类型	基因名称	度值	得分
种子节点	PLK1	25	21.0
种子节点	HAO2	11	11.0
种子节点	FCN2	5	5.0

2.4 关键调控基因表达分析及验证

采用独立样本 *t* 检验比较 ICC 组织和正常肝脏组织中 PLK1、HAO2、FCN2 表达，ICC 组织中 PLK1 表达明显高于正常组织，HAO2、FCN2 表达明显低于正常组织，差异均有显著性（均 $P < 0.001$ ）（图 5A）。结果表明 PLK1 在 ICC 中呈高表达、HAO2、FCN2 在 ICC 中呈低表达。利用 UALCAN 中的 TCGA 数据库对 3 个基因在 ICC 中的表达进行验证，发现 PLK1 在 ICC 中高表达，HAO2、FCN2 在 ICC 中低表达，与分析结果一致（图 5B）。利用 GEPIA 数据库对 3 个基因的表达进行验证，发现 PLK1 在 ICC 中高表达，HAO2、FCN2 在 ICC 中低表达，与分析结果也是一致的（图 5C）。

2.5 关键调控基因在肿瘤中的表达分析

采用 UCSC XENA 数据库分析 PLK1、HAO2、FCN2 在肿瘤中的表达情况。发现在 33 种肿瘤中，

PLK1 在 28 种肿瘤表达明显增高（图 6A），HAO2 在 24 种肿瘤表达明显降低（图 6B），FCN2 在 27 种肿瘤表达明显降低（图 6C）。

2.6 关键调控基因在 ICC 中的共表达基因分析

采用 TCGA 数据库分析关键调控基因在 ICC 中的共表达基因，发现 PLK1 与 CCNA2、GTSE1 等基因共表达相关性最强（图 7A），HAO2 与 MTTP、CPS1 等基因共表达相关性最强（图 7B），FCN2 与 FAM99A、GDF2 等基因共表达相关性最强（图 7C）。

利用 STRINGS 数据库构建共表达基因和关键基因的 PPI 网络，该 PPI 网络共有 30 个节点和 110 条边（图 7D）。采用 Metascape 网站对共表达基因进行富集分析，发现共表达基因主要参与细胞分裂、补体激活、凝集素途径、甘油三酯代谢过程等（图 7E）。

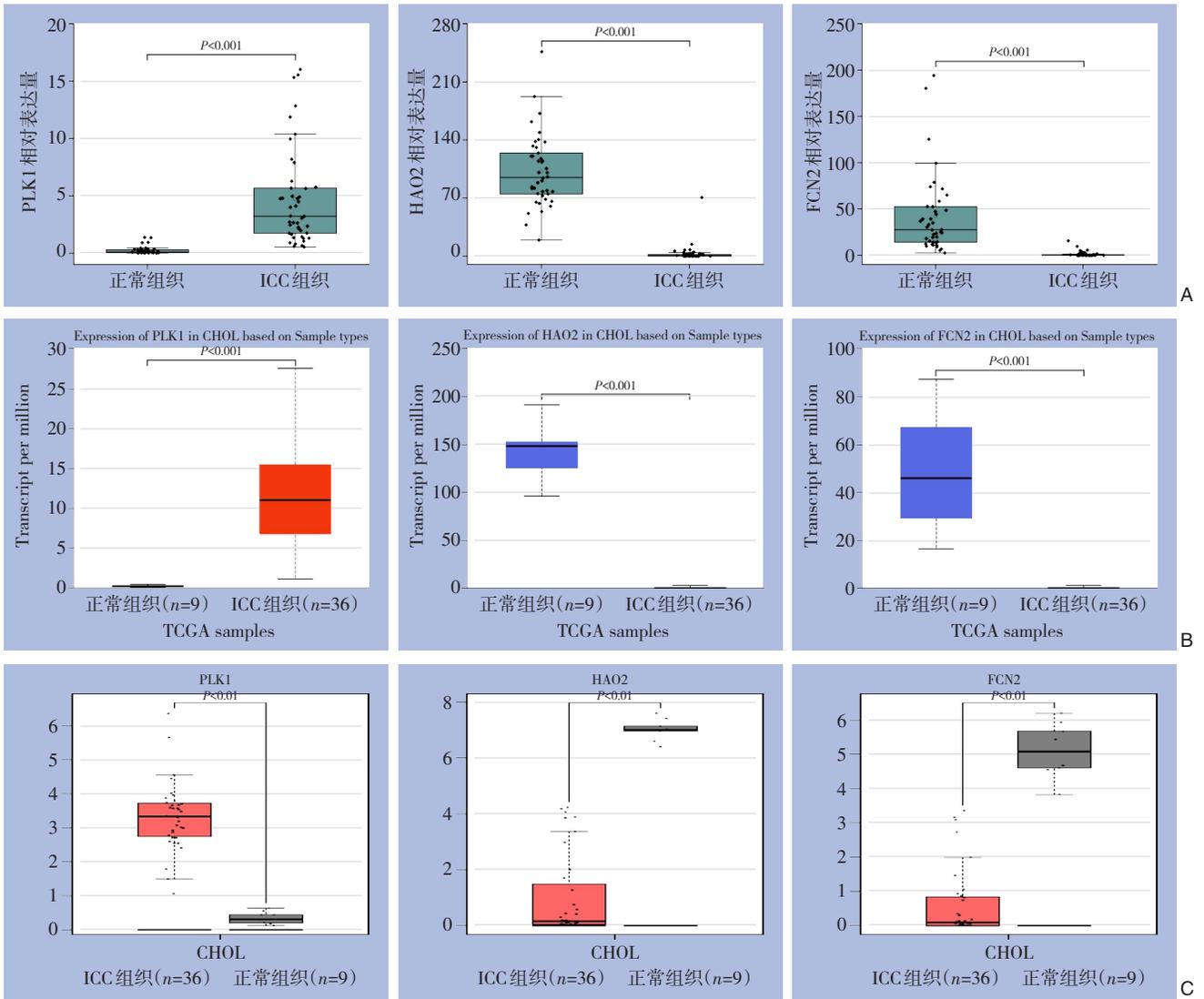


图5 关键调控基因的表达分析及验证 A: PLK1、HAO2、FCN2在ICC组织和正常组织中的表达差异; B: PLK1、HAO2、FCN2在UALCAN数据库中的表达差异; C: PLK1、HAO2、FCN2在GEPIA数据库中的表达差异

Figure 5 Expression analysis and validation of the key regulatory genes A: The expression differences of PLK1, HAO2 and FCN2 between ICC and normal tissues; B: The expression differences of PLK1, HAO2 and FCN2 in UALCAN database; C: The expression differences of PLK1, HAO2 and FCN2 in GEPIA database

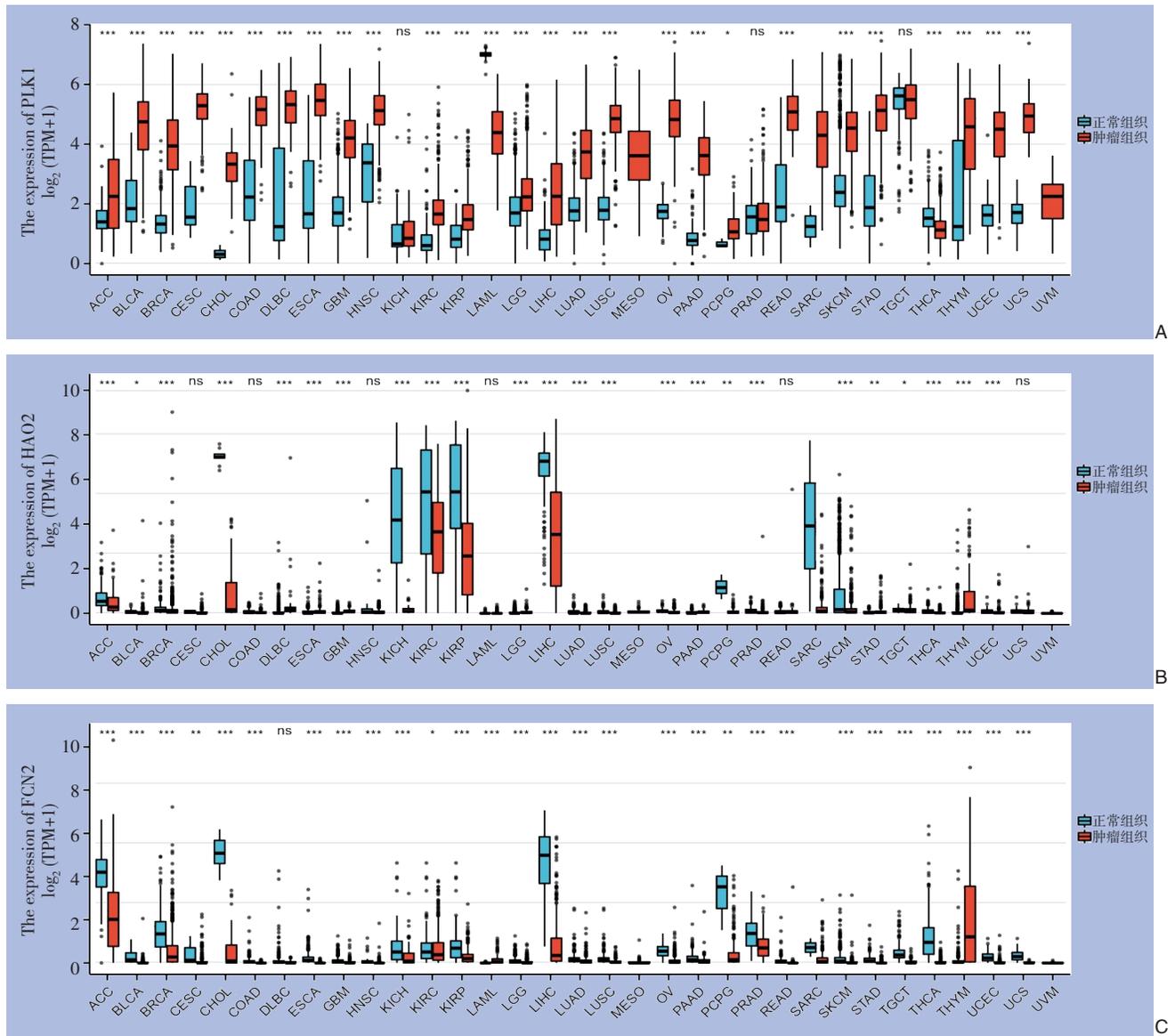


图6 PLK1、HAO2、FCN2在肿瘤中的表达情况分析 A: PLK1在28/33种肿瘤中表达上调; B: HAO2在24/33种肿瘤中表达下调; C: FCN2在27/33种肿瘤中表达下调

Figure 6 Analysis of the expressions of PLK1, HAO2 and FCN2 in tumors A: Up-regulation of PLK1 in 28/33 tumors; B: Down-regulation of HAO2 in 24/33 tumors; C: Down-regulation of FCN2 in 27/33 tumors

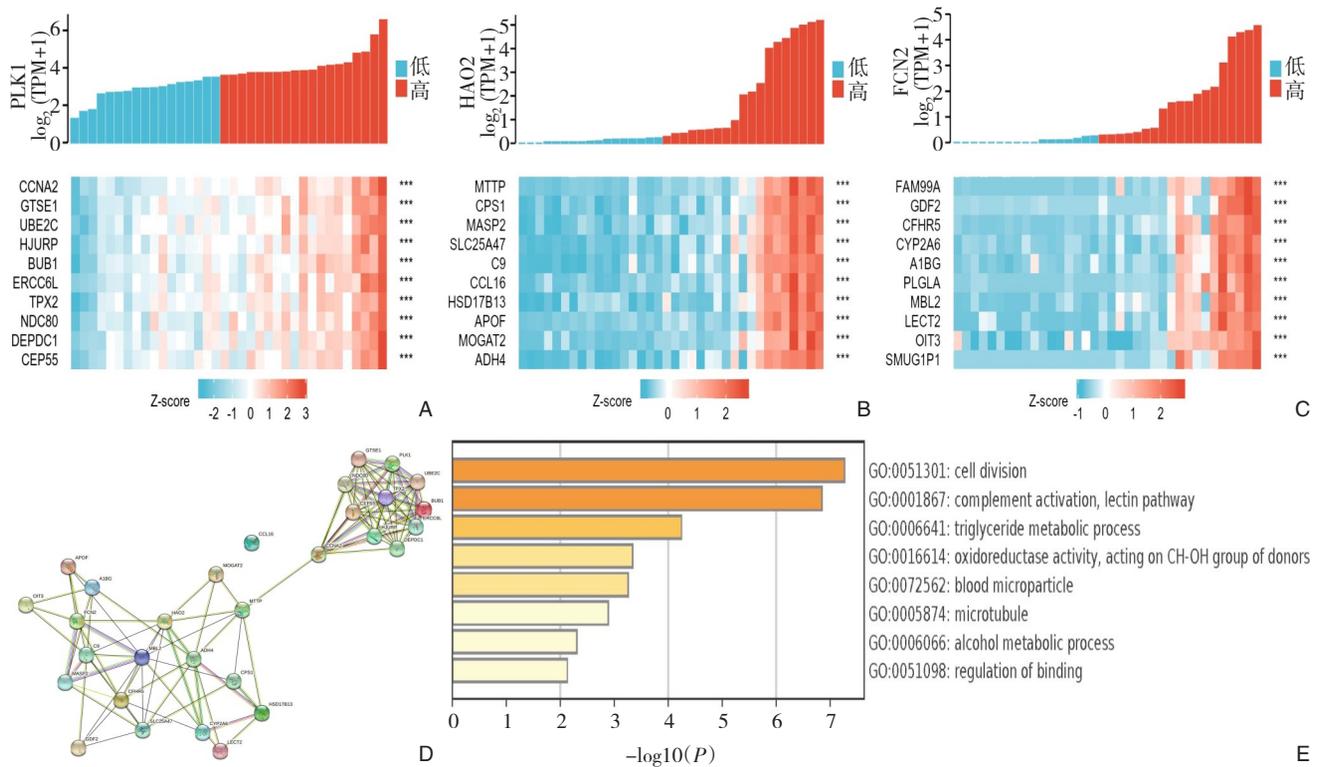


图 7 关键调控基因在 ICC 中的共表达基因分析 A: PLK1 共表达基因热图; B: HAO2 共表达基因热图; C: FCN2 共表达基因热图; D: 共表达基因和关键基因的 PPI 网络; E: 共表达基因 GO 和 KEGG 富集分析

Figure 7 Co-expression gene analysis of the key regulatory genes in ICC A: Heat map of PLK1 co-expressed genes; B: Heat map of HAO2 co-expressed genes; C: Heat map of FCN2 co-expressed genes; D: PPI network of co-expressed genes and key regulatory genes; E: GO and KEGG enrichment of co-expressed genes

2.7 关键调控基因对 ICC 的发展、预后的影响分析

利用 UALCAN 中的 TCGA 数据库分析关键调控基因对 ICC 发展的影响。结果显示, PLK1、HAO2、FCN2 表达与肿瘤分期、分级、淋巴转移有关, PLK1 表达越高、HAO2、FCN2 表达越低, 肿瘤分期越高、分化越差、越易出现淋巴转移 (图 8A-C)。

利用 GEPIA 数据库分析关键调控基因对 ICC 患者预后的影响。结果显示 PLK1、HAO2、FCN2 与 ICC 患者总体生存率无相关性 (图 8D)。

2.8 关键调控基因表达对 ICC 免疫细胞浸润的影响分析

分析关键调控基因表达与免疫细胞浸润水平的相关性。结果发现 PLK1 表达与 Th2 细胞浸润正相关性最大, 与 NK 细胞浸润负相关性最大;

HAO2 表达与中性粒细胞浸润正相关性最大, 与 CD8 T 细胞浸润负相关性最大; FCN2 表达与 Tcm 浸润正相关性最大, 与 aDC 浸润负相关性最大 (图 9A)。

进一步分析关键调控基因表达与正、负相关性最大细胞浸润的显著性。结果发现 PLK1 表达与 Th2 细胞浸润呈显著正相关 ($r=0.517, P=0.001$), PLK1 表达与 NK 细胞浸润相关性不显著 ($P=0.090$) (图 9B); HAO2 表达与 CD8 T 细胞、中性粒细胞浸润相关性均不显著 ($P=0.103, P=0.070$) (图 9C); FCN2 表达与 Tcm 浸润相关性不显著 ($P=0.126$), FCN2 表达与 aDC 浸润呈负相关 ($r=-0.435, P=0.008$) (图 9D)。

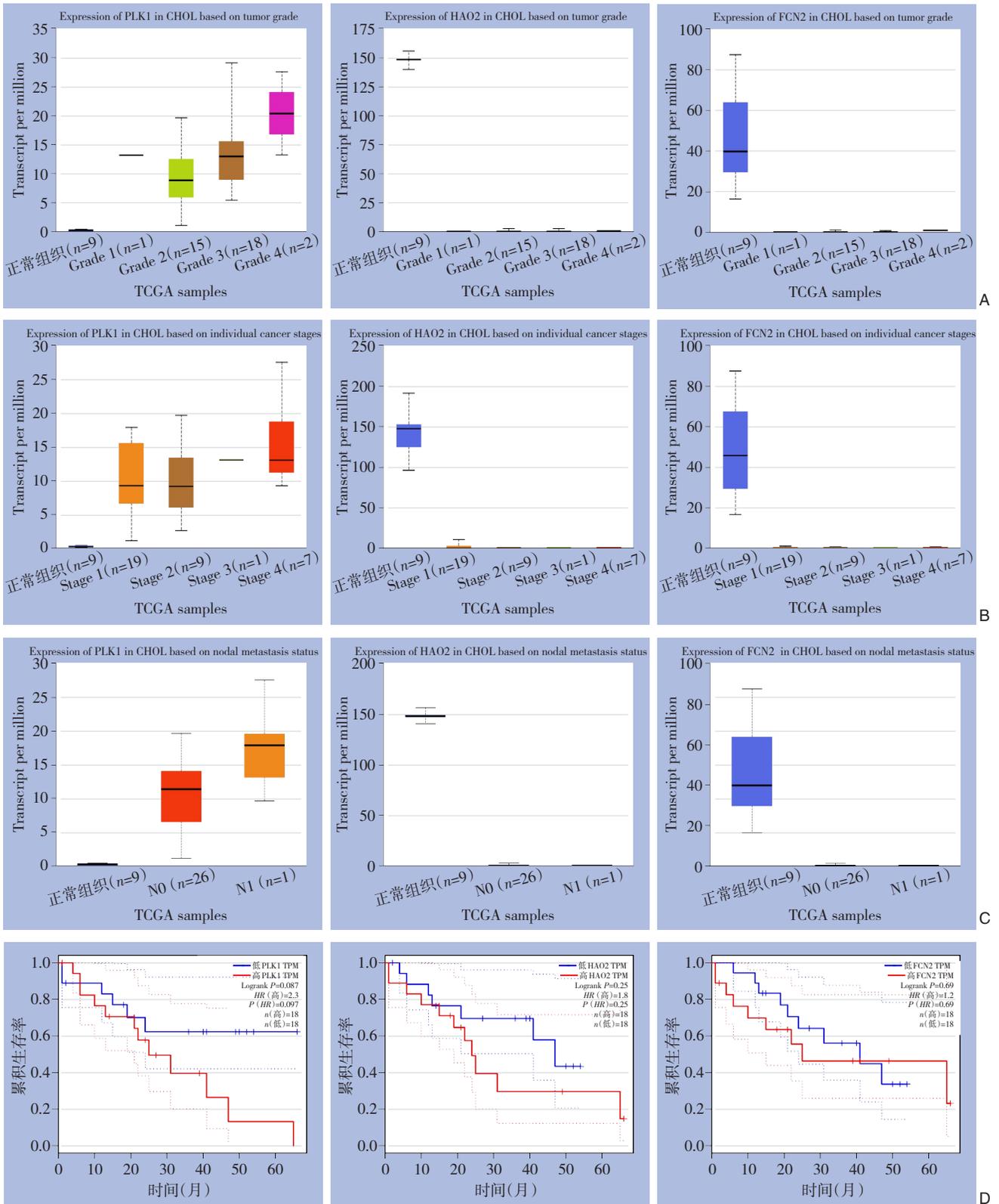


图 8 关键调控基因对 ICC 发展、预后的影响 A: PLK1、HAO2、FCN2 对肿瘤分级的影响; B: PLK1、HAO2、FCN2 对肿瘤分期的影响; C: PLK1、HAO2、FCN2 对肿瘤淋巴结转移的影响; D: PLK1、HAO2、FCN2 对患者预后的影响

Figure 8 Effects of the key regulatory genes on the development and prognosis of ICC A: Effect of PLK1, HAO2, FCN2 on tumor grade; B: Effect of PLK1, HAO2, FCN2 on tumor stages; C: Effect of PLK1, HAO2, FCN2 on tumor nodal metastasis; E: Effect of PLK1, HAO2, FCN2 on prognosis of patients

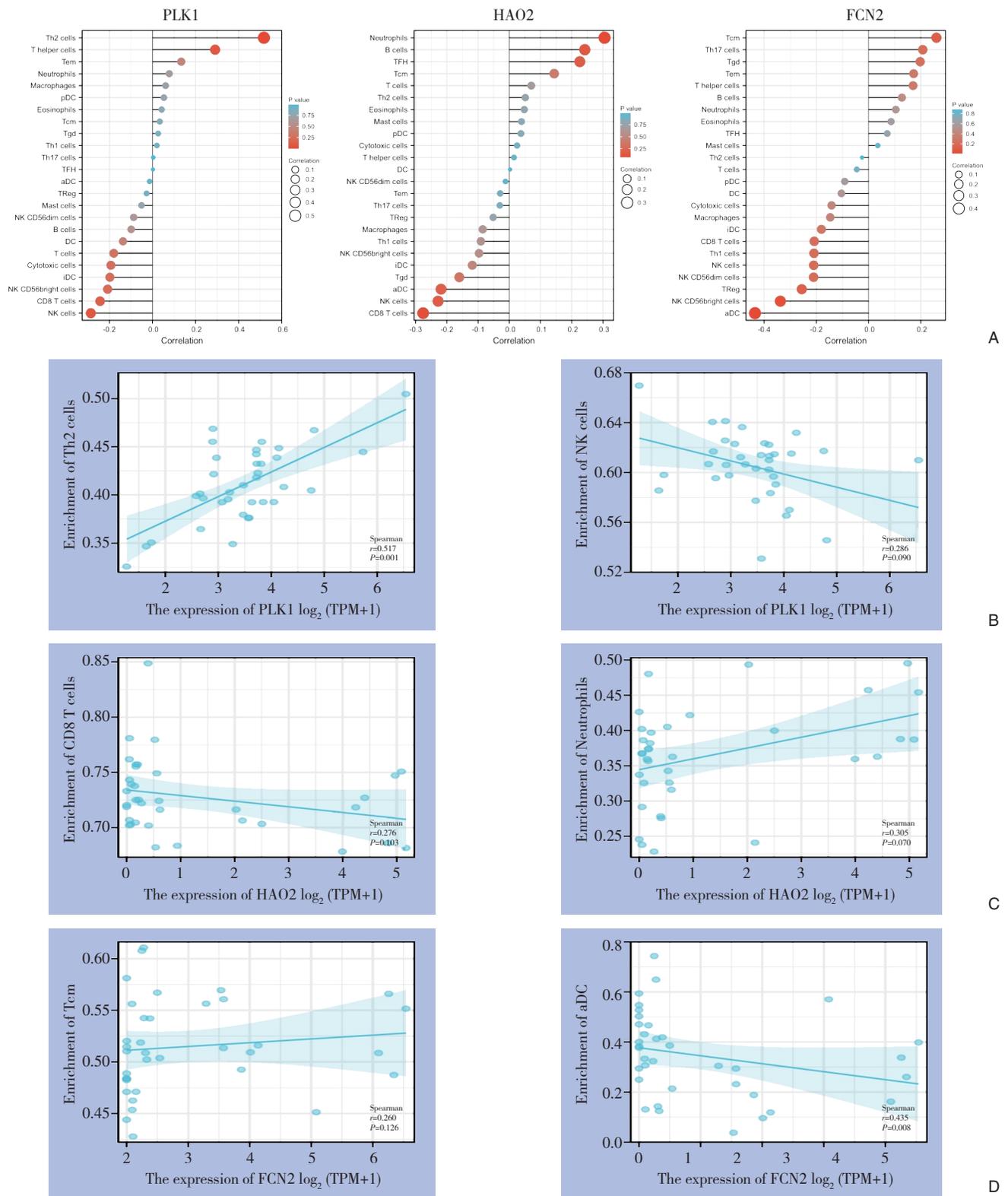


图9 关键调控基因对 ICC 免疫浸润水平的影响 A: PLK1、HAO2、FCN2 表达与免疫细胞浸润水平的相关性; B: PLK1 表达与 Th2 细胞、NK 细胞浸润相关性; C: HAO2 表达与中性粒细胞、CD8 T 细胞浸润相关性; D: FCN2 表达与 Tcm、aDC 浸润相关性

Figure 9 Effect of key regulatory genes on ICC immune infiltration level A: Correlation between PLK1, HAO2, FCN2 expression and immune cell infiltration level; B: Correlation between PLK1 expression and Th2 cells as well as NK cell infiltration; C: Correlation between HAO2 expression and neutrophils as well as CD8 T cells infiltration; D: Correlation between FCN2 expression and Tcm as well as aDC

2.9 评估关键调控基因对 ICC 的诊断能力

绘制 ROC 曲线评估关键调控基因诊断 ICC 的效果。结果发现 PLK1、HAO2、FCN2 3 个基因对 ICC 的 AUC 值均 >0.98，表明 3 个基因对 ICC 均有较好的诊断能力，其中 HAO2 的 AUC 值最大，表明 HAO2 诊断能力最好。PLK1 联合 HAO2 或 FCN2 后

AUC 值均为 1.000，表明 PLK1 联合 HAO2 或 FCN2 对 ICC 具有非常好的诊断能力。HAO2 和 FCN2 联合应用后 AUC 值为 0.999 5，诊断能力略微下降。PLK1、HAO2 和 FCN2 3 个基因联合应用后 AUC 值为 1.000，表明 3 个基因联合应用同样具有非常好的诊断能力（图 10）。

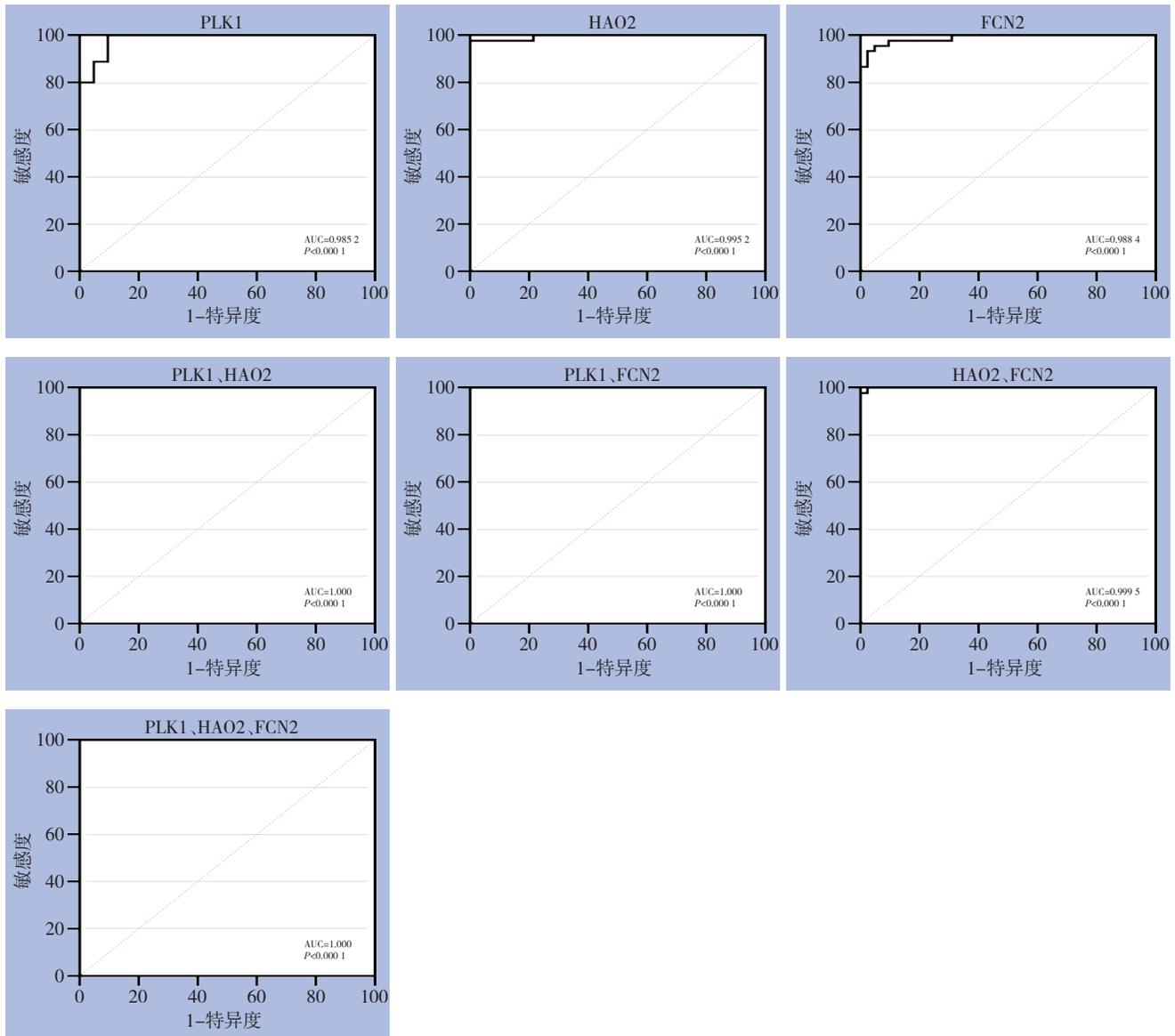


图 10 关键调控基因对 ICC 的诊断能力分析

Figure 10 Diagnostic abilities of the key regulatory genes for ICC

2.10 关键调控基因在肝胆管癌细胞中的表达验证

检测关键调控基因在 HIBEC 和 RBE mRNA 和蛋白表达差异，结果发现 PLK1 mRNA 和蛋白在 RBE 中表达显著高于 HIBEC ($P<0.01$, $P<0.001$),

HAO2、FCN2 mRNA 和蛋白在 RBE 系中表达显著高于 HIBEC ($P<0.01$, $P<0.001$) (图 11)。实验结果表明关键调控基因在细胞中表达差异与生物信息学分析结果一致。

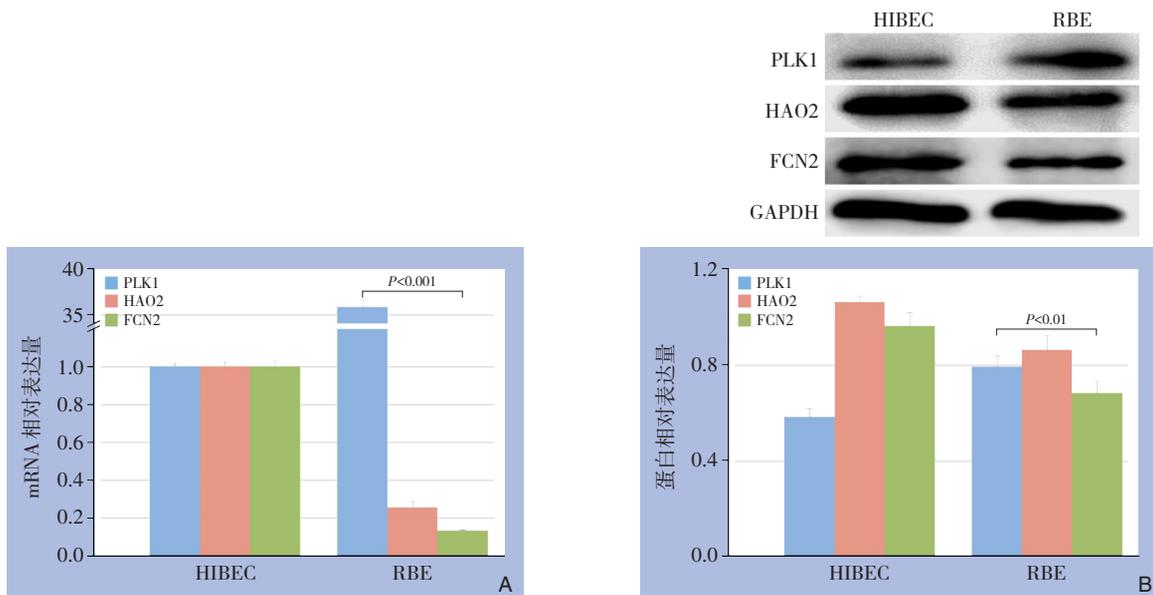


图11 关键调控基因在HIBEC和RBE表达差异 A: 两种细胞株PLK1、HAO2、FCN2 mRNA表达差异; B: 两种细胞株PLK1、HAO2、FCN2蛋白表达差异

Figure 11 Differential expression of key regulatory genes in HIBEC and RBE A: Differential expression of PLK1, HAO2, FCN2 mRNA in two cell lines; B: Differential expression of PLK1, HAO2, FCN2 protein in two cell lines

3 讨论

ICC约占肝脏原发性恶性肿瘤的10%~15%，近年来发病率呈上升趋势^[6]。由于缺乏特异典型的临床表现和敏感可靠的诊断标志物，多数ICC患者被确诊时已进入晚期，因此急切面临对ICC分子机制的切实解析与探究^[7-8]。本研究首先在GEO数据库中筛选出ICC转录组层面的2个高通量数据集GSE107943和GSE119336，从中筛选出3个关键调控基因PLK1、HAO2、FCN2。对关键调控基因表达进行分析，发现PLK1在ICC组织中呈高表达，HAO2、FCN2在ICC组织中呈低表达。采用TCGA和GTex数据库联合分析关键调控基因在泛癌中的表达，发现PLK1在28种肿瘤显著高表达，HAO2在24种肿瘤显著低表达，FCN2在27种肿瘤显著呈低表达。通过UALCAN数据库分析关键调控基因与ICC肿瘤分级和分期、淋巴转移的相关，发现PLK1高表达、HAO2、FCN2低表达提示分化差、分期高、易淋巴转移。分析关键调控基因表达与免疫细胞浸润水平的相关性，发现PLK1表达与Th2细胞浸润呈显著正相关，FCN2表达与aDC浸润呈显著负相关。绘制ROC曲线评估PLK1、HAO2、FCN2对ICC的诊断效果，发现3个基因对ICC均有较好的诊断能力。采用细胞实验验证关键调控基

因在ICC中的表达，发现在RBE中PLK1表达显著上调，HAO2、FCN2表达显著下调。

PLK1是一种高度保守的丝/苏氨酸激酶，其在有丝分裂进程中的重要作用已经被阐明^[9]。最近的研究发现，PLK1有诱导DNA合成、DNA完整性的检修以及防止细胞凋亡方面的作用^[10]。PLK1还能通过磷酸化p53而抑制其转活性，进而抑制p53发挥检验点蛋白和诱导细胞凋亡的功能^[11]。另外，PLK1与肿瘤的发生发展密切相关，研究发现PLK1在多种人类肿瘤中呈高表达^[12]。本研究发现PLK1在ICC中的表达明显高于正常组织，与上述研究结果一致。同时本研究发现PLK1高表达与ICC患者肿瘤分期、分级和淋巴转移、免疫细胞浸润有关，且对ICC有较好的诊断能力，表明PLK1有望成为ICC诊断和治疗的新靶点。

过氧化物酶体普遍存在于真核细胞内，参与多种代谢途径^[13]。羟羧酸氧化酶(hydroxyacid oxidase, HAO)是过氧化物酶体中所包含的一类氧化酶，一共有3个亚型，其中HAO2主要在肝脏和肾脏中表达^[14]。研究^[15-16]发现HAO2在肝癌、肾透明细胞癌中表达下降，与患者的不良预后相关。本研究发现HAO2在ICC中表达同样下降，且HAO2低表达提示患者生存较差。HAO2是催化含羟基的脂肪酸的酶类，直接参与脂肪酸的氧化分

解过程,而肿瘤代谢的过程脂肪酸分解减少、合成增加^[17]。本研究发现HAO2在ICC中低表达,提示ICC可能通过降低HAO2表达适应其脂代谢重编程的需要,HAO2可能是纠正ICC代谢异常的一个潜在靶点。

FCN2能够与凝集素相关丝氨酸蛋白酶形成复合物,是少数能激活凝集素补体途径的分子^[18-19]。目前FCN2研究多集中在感染性疾病,研究报道其单核苷酸多态性与肺结核、越南登革热、意大利风湿热、风湿性心脏病等疾病有关^[20-23]。关于肿瘤方面只有少量报道,Yang等^[24]发现FCN2在肝癌组织中呈表达下调,FCN2低表达与肝癌的侵袭转移相关。本研究发现FCN2在ICC组织中同样表达下降,与ICC患者肿瘤分期、分级和淋巴转移、免疫细胞浸润等相关,提示FCN2在ICC中发挥重要作用,可能是ICC的一个潜在的作用靶点。

免疫细胞在调节肿瘤细胞行为中发挥重要作用,越来越多的证据支持它们在预测许多癌症类型的结果和治疗效果方面的具有重要价值^[25]。本研究分析了PLK1、HAO2、FCN2基因表达与ICC中各种肿瘤免疫细胞浸润的相关性,发现PLK1表达与Th2 cells浸润呈显著正相关,FCN2表达与aDC浸润呈显著负相关,因此PLK1和FCN2可能通过调节肿瘤免疫影响ICC进展。

本研究筛选出3个对ICC具有诊断能力的基因,这3个基因在ICC中均未见报道但在其他肿瘤中均有报道,且报道结果与本研究结果基本一致,表明本研究筛选结果具有较高的可信度。但是本研究也存在着以下不足。第一,该研究是通过生物信息学方法进行的,只进行了细胞实验验证,缺乏临床试验和动物实验验证,因此后续还需临床试验和动物实验来进一步验证。第二,因为ICC自身的发生机制具有异质性,而研究样本数量是有限的,可能要依赖单细胞水平的下一代测序技术迅速发展来解决这一难题^[26-27]。第三,由于ICC疾病在发生发展过程中受到多种因素影响^[28-30],只通过转录组层面来看的话,具有一定的片面性,本研究尚未把代谢组、基因组、蛋白质组等多组学之间的数据信息进行结合分析,多组学的数据信息是否有相关性还需进一步探讨。

综上所述,本文通过生物信息学方法筛选出ICC中3个关键调控基因PLK1、HAO2、FCN2,这3个基因对ICC具有很好的诊断和预测能力,有可

能成为ICC潜在的治疗靶点。本研究筛选出的3个关键调控基因在肿瘤中均有报道,研究结果具有较高的可信度,可以为深入理解ICC的分子机制提供新的观点和思路。

利益冲突:所有作者均声明不存在利益冲突。

参考文献

- [1] 陈小彬,吴嘉艺,黄理铭,等.肝内胆管细胞癌中hsa_circ_0089153的表达及其生物学功能[J].中国普通外科杂志,2022,31(2):199-207. doi: 10.7659/j.issn.1005-6947.2022.02.008.
Chen XB, Wu JY, Huang LM, et al. Expression of hsa_circ_0089153 in intrahepatic cholangiocarcinoma and its biological function[J]. Chinese Journal of General Surgery, 2022, 31(2): 199-207. doi: 10.7659/j.issn.1005-6947.2022.02.008.
- [2] 刘继升,刘河,刘浩,等. TIP30在肝内胆管癌中的表达及意义[J]. 中国普通外科杂志, 2021, 30(2): 187-194. doi: 10.7659/j.issn.1005-6947.2021.02.008.
Liu JS, Liu H, Liu H, et al. Expression of TIP30 in intrahepatic cholangiocarcinoma and its significance[J]. Chinese Journal of General Surgery, 2021, 30(2): 187-194. doi: 10.7659/j.issn.1005-6947.2021.02.008.
- [3] 李雪鹏,刘苏来,蒋波.肝内胆管癌分子水平研究进展[J].中国普通外科杂志,2019,28(2):219-226. doi: 10.7659/j.issn.1005-6947.2019.02.013.
Li XP, Liu SL, Jiang B. Progress of molecular profiling of intrahepatic cholangiocarcinoma[J]. Chinese Journal of General Surgery, 2019, 28(2): 219-226. doi: 10.7659/j.issn.1005-6947.2019.02.013.
- [4] Rego SM, Snyder MP. High throughput sequencing and assessing disease risk[J]. Cold Spring Harb Perspect Med, 2019, 9(1): a026849. doi: 10.1101/cshperspect.a026849.
- [5] Moeini A, Sia D, Bardeesy N, et al. Molecular pathogenesis and targeted therapies for intrahepatic cholangiocarcinoma[J]. Clin Cancer Res, 2016, 22(2):291-300. doi: 10.1158/1078-0432.CCR-14-3296.
- [6] Massarweh NN, El-Serag HB. Epidemiology of hepatocellular carcinoma and intrahepatic cholangiocarcinoma[J]. Cancer Control, 2017, 24(3):1073274817729245. doi: 10.1177/1073274817729245.
- [7] 柏杨,刘洪,方征,等.高迁移率族蛋白1在肝内胆管癌中的作用及其与肿瘤微血管形成的关系[J].中国普通外科杂志,2020,29(2):171-178. doi: 10.7659/j.issn.1005-6947.2020.02.008.
Bai Y, Liu H, Fang Z, et al. Function of high mobility group protein 1 in intrahepatic cholangiocarcinoma and its relationship with tumor microangiogenesis[J]. Chinese Journal of General Surgery, 2020, 29(2):171-178. doi: 10.7659/j.issn.1005-6947.2020.02.008.
- [8] Rodrigues PM, Olaizola P, Paiva NA, et al. Pathogenesis of

- cholangiocarcinoma[J]. *Annu Rev Pathol Mech Dis*, 2021, 16:433–463. doi: [10.1146/annurev-pathol-030220-020455](https://doi.org/10.1146/annurev-pathol-030220-020455).
- [9] Singh P, Pesenti ME, Maffini S, et al. BUB1 and CENP-U, primed by CDK1, are the main PLK1 kinetochore receptors in mitosis[J]. *Mol Cell*, 2021, 81(1):67–87. doi: [10.1016/j.molcel.2020.10.040](https://doi.org/10.1016/j.molcel.2020.10.040).
- [10] Peng B, Shi RF, Bian J, et al. PARP1 and CHK1 coordinate PLK1 enzymatic activity during the DNA damage response to promote homologous recombination-mediated repair[J]. *Nucleic Acids Res*, 2021, 49(13):7554–7570. doi: [10.1093/nar/gkab584](https://doi.org/10.1093/nar/gkab584).
- [11] Jung Y, Kraikivski P, Shafiekhani S, et al. Crosstalk between Plk1, p53, cell cycle, and G2/M DNA damage checkpoint regulation in cancer: computational modeling and analysis[J]. *Npj Syst Biol Appl*, 2021, 7(1): 1–13. doi: [10.1038/s41540-021-00203-8](https://doi.org/10.1038/s41540-021-00203-8).
- [12] Iliaki S, Beyaert R, Afonina IS. Polo-like kinase 1 (PLK1) signaling in cancer and beyond[J]. *Biochem Pharmacol*, 2021, 193: 114747. doi: [10.1016/j.bcp.2021.114747](https://doi.org/10.1016/j.bcp.2021.114747).
- [13] He AY, Dean JM, Lodhi IJ. Peroxisomes as cellular adaptors to metabolic and environmental stress[J]. *Trends Cell Biol*, 2021, 31(8):656–670. doi: [10.1016/j.tcb.2021.02.005](https://doi.org/10.1016/j.tcb.2021.02.005).
- [14] Zhuo H, Xia JG, Zhang J, et al. CircASPH promotes hepatocellular carcinoma progression through methylation and expression of HAO2[J]. *Front Oncol*, 2022, 12: 911715. doi: [10.3389/fonc.2022.911715](https://doi.org/10.3389/fonc.2022.911715).
- [15] Mattu S, Fornari F, Quagliata L, et al. The metabolic gene HAO2 is downregulated in hepatocellular carcinoma and predicts metastasis and poor survival[J]. *J Hepatol*, 2016, 64(4):891–898. doi: [10.1016/j.jhep.2015.11.029](https://doi.org/10.1016/j.jhep.2015.11.029).
- [16] Xiao W, Wang XG, Wang T, et al. HAO2 inhibits malignancy of clear cell renal cell carcinoma by promoting lipid catabolic process[J]. *J Cell Physiol*, 2019, 234(12): 23005–23016. doi: [10.1002/jcp.28861](https://doi.org/10.1002/jcp.28861).
- [17] Li ZY, Zhang HF. Reprogramming of glucose, fatty acid and amino acid metabolism for cancer progression[J]. *Cell Mol Life Sci*, 2016, 73(2):377–392. doi: [10.1007/s00018-015-2070-4](https://doi.org/10.1007/s00018-015-2070-4).
- [18] Bjarnadottir H, Arnardottir M, Ludviksson BR. Frequency and distribution of FCN2 and FCN3 functional variants among MBL2 genotypes[J]. *Immunogenetics*, 2016, 68(5):315–325. doi: [10.1007/s00251-016-0903-4](https://doi.org/10.1007/s00251-016-0903-4).
- [19] Garred P, Honoré C, Ma YJ, et al. MBL2, FCN1, FCN2 and FCN3—the genes behind the initiation of the lectin pathway of complement[J]. *Mol Immunol*, 2009, 46(14): 2737–2744. doi: [10.1016/j.molimm.2009.05.005](https://doi.org/10.1016/j.molimm.2009.05.005).
- [20] Xu DD, Wang C, Jiang F, et al. Association of the FCN2 gene single nucleotide polymorphisms with susceptibility to pulmonary tuberculosis[J]. *PLoS One*, 2015, 10(9): e0138356. doi: [10.1371/journal.pone.0138356](https://doi.org/10.1371/journal.pone.0138356).
- [21] Giang NT, Tong HV, Nghia TH, et al. Association of FCN2 polymorphisms and Ficolin-2 levels with dengue fever in Vietnamese patients[J]. *Int J Infect Dis*, 2020, 95: 253–261. doi: [10.1016/j.ijid.2020.02.029](https://doi.org/10.1016/j.ijid.2020.02.029).
- [22] Marzetti V, di Battista C, Ferrante R, et al. MBL2 and FCN2 gene polymorphisms in a cohort of Italian children with rheumatic fever: a case-control study[J]. *Semin Arthritis Rheum*, 2017, 47(2):264–268. doi: [10.1016/j.semarthrit.2017.04.006](https://doi.org/10.1016/j.semarthrit.2017.04.006).
- [23] Elshamaa MF, Hamza H, El Rahman NA, et al. Association of ficolin-2 (FCN2) functional polymorphisms and protein levels with rheumatic fever and rheumatic heart disease: relationship with cardiac function[J]. *Arch Med Sci Atheroscler Dis*, 2018, 3: e142–e155. doi: [10.5114/amsad.2018.80999](https://doi.org/10.5114/amsad.2018.80999).
- [24] Yang GC, Liang YJ, Zheng TS, et al. FCN2 inhibits epithelial-mesenchymal transition-induced metastasis of hepatocellular carcinoma via TGF- β /smad signaling[J]. *Cancer Lett*, 2016, 378(2): 80–86. doi: [10.1016/j.canlet.2016.05.007](https://doi.org/10.1016/j.canlet.2016.05.007).
- [25] Sabry M, Lowdell MW. Killers at the crossroads: the use of innate immune cells in adoptive cellular therapy of cancer[J]. *Stem Cells Transl Med*, 2020, 9(9):974–984. doi: [10.1002/sctm.19-0423](https://doi.org/10.1002/sctm.19-0423).
- [26] El-Diwany R, Pawlik TM, Ejaz A. Intrahepatic cholangiocarcinoma[J]. *Surg Oncol Clin N Am*, 2019, 28(4): 587–599. doi: [10.1016/j.soc.2019.06.002](https://doi.org/10.1016/j.soc.2019.06.002).
- [27] Saleh M, Virarkar M, Bura V, et al. Intrahepatic cholangiocarcinoma: pathogenesis, current staging, and radiological findings[J]. *Abdom Radiol*, 2020, 45(11): 3662–3680. doi: [10.1007/s00261-020-02559-7](https://doi.org/10.1007/s00261-020-02559-7).
- [28] 谢之豪, 李俊, 夏勇, 等. 肝内胆管癌靶向治疗研究现状[J]. *中华外科杂志*, 2020, 58(4): 289–294. doi: [10.3760/cma.j.cn112139-20200128-00048](https://doi.org/10.3760/cma.j.cn112139-20200128-00048).
- Xie ZH, Li J, Xia Y, et al. Recent progress in molecular targeted therapies for intrahepatic cholangiocarcinoma[J]. *Chinese Journal of Surgery*, 2020, 58(4): 289–294. doi: [10.3760/cma.j.cn112139-20200128-00048](https://doi.org/10.3760/cma.j.cn112139-20200128-00048).
- [29] Zhang M, Yang H, Wan LF, et al. Single-cell transcriptomic architecture and intercellular crosstalk of human intrahepatic cholangiocarcinoma[J]. *J Hepatol*, 2020, 73(5): 1118–1130. doi: [10.1016/j.jhep.2020.05.039](https://doi.org/10.1016/j.jhep.2020.05.039).
- [30] Fabris L, Sato K, Alpini G, et al. The tumor microenvironment in cholangiocarcinoma progression[J]. *Hepatology*, 2021, 73(Suppl 1): 75–85. doi: [10.1002/hep.31410](https://doi.org/10.1002/hep.31410).

(本文编辑 宋涛)

本文引用格式: 陈伟毅, 陈立军. 基于生物信息学的肝内胆管癌差异表达基因谱中关键基因的筛选及分析[J]. *中国普通外科杂志*, 2022, 31(8):1048–1063. doi: [10.7659/j.issn.1005-6947.2022.08.008](https://doi.org/10.7659/j.issn.1005-6947.2022.08.008)

Cite this article as: Chen WY, Chen LJ. Screening and analysis of key genes in differential expression gene profile of intrahepatic cholangiocarcinoma based on bioinformatics[J]. *Chin J Gen Surg*, 2022, 31(8): 1048–1063. doi: [10.7659/j.issn.1005-6947.2022.08.008](https://doi.org/10.7659/j.issn.1005-6947.2022.08.008)