

DOI: 10.3872/j.issn.1007-385x.2019.04.010

· 临床研究 ·

基于生物信息学分析的肝细胞癌预后相关基因的筛选

孙厚芳, 颜次慧, 吴磊, 李百会, 杨莉莉(天津医科大学肿瘤医院 生物技术研究室 国家肿瘤临床医学研究中心 天津市“肿瘤防治”重点实验室 天津市恶性肿瘤临床医学研究中心 天津市肿瘤免疫与生物治疗重点实验室, 天津 300060)

[摘要] **目的:** 利用生物信息学方法筛选出肝细胞癌(HCC)组织与正常肝组织之间差异表达基因(DEG), 从转录组层面分析这些候选基因参与HCC发生发展的内在机制及其与HCC患者预后相关基因的临床意义。**方法:** 分别从基因表达数据库(GEO)及人类癌症基因组图谱(TCGA)网站中下载GSE45267、GSE64041、GSE84402和TCGA中的基因表达谱, R软件和Bioconductor安装包用于筛选HCC组织与癌旁组织之间DEG, 然后对这些DEG进行基因本体(GO)富集分析、京都基因与基因组百科全书(KEGG)通路分析、蛋白质相互作用(PPI)网络分析及生存分析。**结果:** 共筛选出46个上调基因和154个下调基因, GO富集分析显示, 这些DEG主要与细胞分裂、增殖、周期调控、氧化还原过程及某些代谢途径密切相关; KEGG通路分析显示DEG主要与色氨酸、视黄醇等代谢途径及P53通路有关。在TCGA数据集中, 6个上调的中枢基因CCNA2、CDK1、DLGAP5、KIF20A、KPNA2、MELK的过表达被认为与HCC患者预后呈明显负相关(均 $P < 0.01$)。**结论:** 筛选出的一组与预后负相关的中枢上调基因对HCC诊断和治疗的临床研究可能具有潜在的指导价值。

[关键词] 肝细胞癌; 生物信息分析; 预后相关基因; 基因本体分析; 京都基因与基因组百科全书分析; 蛋白质相互作用网络
[中图分类号] R735.7; R730.7 **[文献标识码]** A **[文章编号]** 1007-385X(2019)04-0431-09

Identification of prognosis-related genes in hepatocellular carcinoma based on bioinformatical analysis

SUN Houfang, YAN Cihui, WU Lei, LI Baihui, YANG Lili (Biotechnology Lab, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy of Tianjin City, Tianjin Clinical Research Center for Malignant Cancer, Key Laboratory of Cancer Immunology and Biotherapy of Tianjin City, Cancer Hospital of Tianjin Medical University, Tianjin 300060, China)

[Abstract] Objective: To identify the differentially expressed genes (DEGs) between hepatocellular carcinoma (HCC) tissues and normal liver tissues by bioinformatic methods, and to explore the intrinsic mechanism of these candidate genes involving in the occurrence and development of HCC from transcriptome level as well as the clinical significance of their associations with the prognosis of HCC patients. **Methods:** Gene expression profiles of GSE45267, GSE64041, GSE84402 and TCGA were downloaded from GEO (Gene Expression Omnibus) and TCGA (The Cancer Genome Atlas), respectively. R software and Bioconductor packages were used to identify the DEGs between HCC tissues and para-cancer tissues, and then Gene Ontology (GO) Enrichment analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, Protein-Protein Interaction (PPI) network analysis and survival analysis were performed. **Results:** Forty-six up-regulated genes and 154 down-regulated genes were screened out, and GO enrichment analysis showed that these DEGs were mainly related to cell division, proliferation, cycle regulation, oxidation-reduction process and certain metabolic pathways. KEGG pathway analysis revealed that DEGs were mainly involved in tryptophan metabolism, retinol metabolism and other metabolic pathways as well as p53 pathway. Over-expression of a panel of up-regulated genes (CCNA2, CDK1, DLGAP5, KIF20A, KPNA2 and MELK) was shown to be significantly negatively correlated with the prognosis of HCC patients in the TCGA dataset (all $P < 0.01$). **Conclusion:** A set of up-regulated hub genes that are negatively correlated with prognosis will provide potential guiding value for the clinical research on the diagnosis and treatment of HCC.

[Key words] hepatocellular carcinoma; bioinformatical analysis; prognostic gene; gene ontology analysis; Kyoto Encyclopedia of

[基金项目] 国家自然科学基金资助项目(No.81572265)。Project supported by the National Natural Science Foundation of China(No.81572265)

[作者简介] 孙厚芳(1993-), 女, 硕士生, 主要从事肿瘤免疫基础与临床研究, E-mail: 1425267400@qq.com

[通信作者] 杨莉莉(YANG Lili, corresponding author), 博士, 研究员, 博士生导师, 主要从事肿瘤分子生物学和肿瘤免疫学研究, E-mail: yanglili@tjmuch.com

肝细胞癌(hepatocellular carcinoma, HCC), 成人肝癌最普遍(80%以上)的一种类型, 是全球第2位癌症相关死亡的原因^[1]。在中国, 由于乙型肝炎和丙型肝炎的感染使HCC发病率居高不下^[2]; 在西方国家, HCC主要源于非酒精性脂肪性肝病^[3]。尽管近年来在HCC手术治疗、化疗方面取得了一些进展, 但肿瘤转移和化疗耐药导致其预后仍然不佳^[4]。因此, 探究其潜在的分子机制并发掘新的预后相关分子标志物对临床HCC诊断和治疗而言迫在眉睫。目前, 微阵列技术已成为检测特定生物体中基因组表达水平不可或缺的工具。基因表达数据库(Gene Expression Omnibus, GEO)作为国际性的公共存储库和基因表达研究的全球资源于2000年创建, 可自行上传和下载微阵列、新一代测序及其他形式高通量功能获取的基因组数据集^[5]。人类癌症基因组图谱(The Cancer Genome Atlas, TCGA), 作为一个大规模的为研究人员提供的癌症基因组计划, 汇编了来自33种癌症的10 000多个样本的基因组学、表观基因组学和蛋白质组学数据^[6]。综合生物信息分析的优势在于可以克服样本异质性和平台差异性, 将来自不同独立微阵列研究的数据整合到一起以获得更多的临床样本, 来实现更为可靠的分析。本研究在从GEO下载的3个数据集及TCGA的数据集中分别筛选差异基因后取交集, 并进行一系列生物信息学分析和上调基因的生存分析, 为探索HCC的潜在治疗靶点提供理论依据。

1 资料与方法

1.1 数据来源

在GEO数据库(<http://www.ncbi.nlm.nih.gov/geo/>)中选取了3个同时有HCC组织和正常对照组织基因表达谱的数据集, 登记号分别为GSE45267、GSE64041、GSE84402。GSE45267和GSE84402的微阵列数据基于GPL570平台(Affymetrix Human Genome U133 Plus 2.0 Array, Affymetrix, Santa Clara, CA, USA)测得, 其中GSE45267中包含46个HCC肿瘤组织和41个正常肝组织(提交日期: 2013年3月18日), GSE84402包含14例配对的HCC肿瘤组织和正常肝组织(提交日期: 2016年7月14日)。GSE64041的微阵列数据基于GPL6244平台(Affymetrix Human Gene 1.0 ST Array, Affymetrix, Santa Clara, CA, USA)测得, 本研究从中提取了120个样本数据, 包含60例HCC肿瘤组织和相应正常肝组织(提交日期: 2014年12月10日)。所有GEO数据集芯片原始表达

数据均运用affy包中的RMA方法进行了背景校正和均一化处理^[7]。此外, 预处理后的RNA-seq数据和相应的患者临床信息从TCGA(<http://cancergenome.nih.gov/>)获得, 共372个HCC组织和52个相邻的正常组织, 后续的生存分析中包括有310例HCC患者的随访信息。

1.2 差异基因的筛选

3个GEO数据集的基因表达矩阵从GEO下载后在R 3.5.1中用limma包来筛选差异表达基因(differentially expressed gene, DEG), 而TCGA的数据则是用edgeR包来筛选。其中 $|\log_2FC(\text{fold change})| > 1$ 和 $P < 0.05$ 作为截取标准用于DEG的筛选。韦恩图由VennDiagram包绘制, 具体基因名由在线网页工具(<http://bioinformatics.psb.ugent.be/webtools/Venn/>)获取。

1.3 GO和KEGG通路富集分析

大规模基因的功能注释经常用到基因本体(gene ontology, GO)分析^[8]。DAVID(<http://david.abcc.ncifcrf.gov/>)是一个集基因注释、可视化和综合挖掘功能的数据平台, 它为研究人员提供了一套全面的基因注释工具, 用于理解基因的生物学意义^[9]。本研究将筛选得到的DEG在DAVID中进行了生物过程(BP)层面的富集分析和通路分析, 并使用京都基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG; <http://www.genome.jp/kegg/>)进行了通路分析, 以 $P < 0.05$ 作为截取标准。

1.4 PPI网络的构建

首先在STRING在线数据库(<http://string-db.org/>)中对4个数据集共有的DEG进行了蛋白质-蛋白质相互作用(protein-protein interaction, PPI)网络分析, 然后将网络图导入Cytoscape 3.5.1(<http://www.cytoscape.org/>)中使其可视化。Cytoscape作为生物信息分析的开源软件工具之一, 可将由蛋白质、基因等组成的相互作用网络可视化并进行分析, 其中每个节点(node)代表一个蛋白或基因抑或其他分子, 节点之间的连接(edge)则代表他们之间的相互作用, 而度值(degree)代表他们的连接数, 其中的MCODE插件用于筛选出重要模块, 而cytoHubba插件则用于计算度值^[10]。

1.5 统计学处理

从TCGA中下载310例HCC患者的临床信息和预后信息后, 将筛选得到的基因表达值在R中用hash包分别提取出来, 然后确定每个基因表达值的中位数, 将高于和低于中位表达值的表达值分为高表达

组和低表达组,最后用Survival包制作生存曲线。采用Log-Rank检验比较基因表达与HCC患者预后的关系。以 $P < 0.05$ 或 $P < 0.01$ 表示差异有统计学意义。

2 结果

2.1 从4个数据集中DEG筛选出46个上调、154个下调DEG

从NCBI-GEO数据库312个人类HCC组织芯片数据集中选取了GSE45267、GSE64041、GSE84402数据集,其样本均为HCC肿瘤组织与相应的癌旁正常肝组织。肿瘤样本与正常对照之间差异表达的基因由R软件生成的火山图和热图(图1)直观展现,每个火山图纵坐标为每个样本差异表达倍数(fold

change, FC)的对数值,横坐标为对相应样本 P 值取10的对数后的负值,红色和绿色的点分别代表着上调和下调的基因(HCC组织相对于正常对照),黑点则代表在肿瘤和正常组织中表达差异不甚明显的基因。此外,GSE45267、GSE64041、GSE84402及TCGA中各自筛选出的DEG分别为1317个(562个上调基因,755个下调基因)、323个(91个上调基因,232个下调基因)、1026个(308个上调基因,718个下调基因)、9058个(7501个上调基因,1557个下调基因)。将这4个独立的数据集取交集,共得到46个上调DEG和154个下调DEG(图2)。具体的上调DEG($\log FC > 1, P < 0.05$)与下调DEG($\log FC < -1, P < 0.05$)见表1。

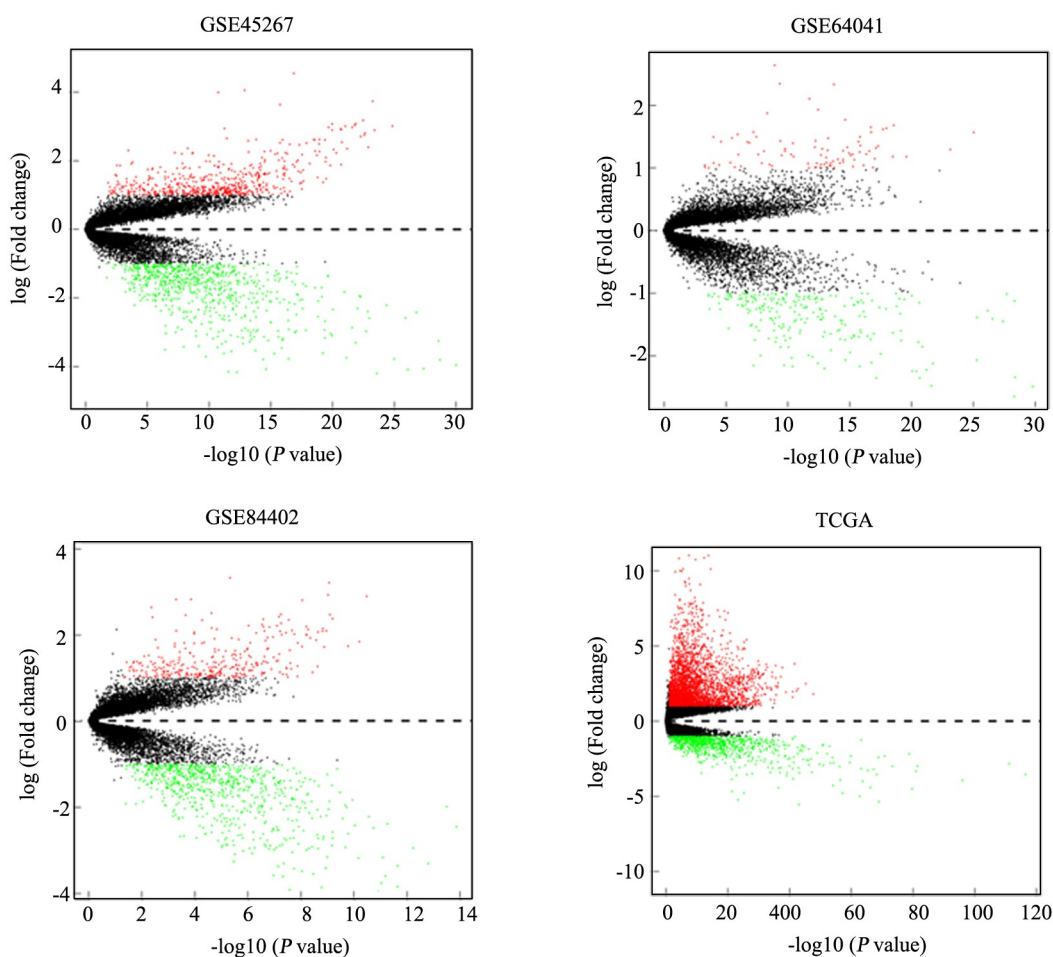


图1 反映GSE45267、GSE64041、GSE84402及TCGA中显著DEG的火山图

Fig. 1 Volcano plots reflecting significant DEGs in GSE45267, GSE64041, GSE84402 and TCGA

2.2 DEG在GO和KEGG中的富集分析

为了在分子和功能水平上对筛选到的200个DEG进行深入探究,运用在线生物学工具DAVID数据库和KEGG数据库对其进行富集分析。如表2所

示,上调DEG主要被富集在细胞分裂(有丝分裂核与细胞质分裂)、细胞周期调控(G1/M与G2/M分裂周期的转变)与细胞增殖等生物过程;而下调DEG主要参与氧化还原过程、环氧化酶P450过程、外源性药物

分解与代谢过程以及负性生长调控等过程。同时KEGG分析提示,这些DEG主要与色氨酸、视黄醇、咖啡因及某些药物的代谢途径有关,还参与了补体

和凝血级联反应。另外,发现有STEAP3、CCNE2、CDK1、CCNB2、RRM2、IGF1、IGFBP3 7个基因富集在P53转导通路,见表3。

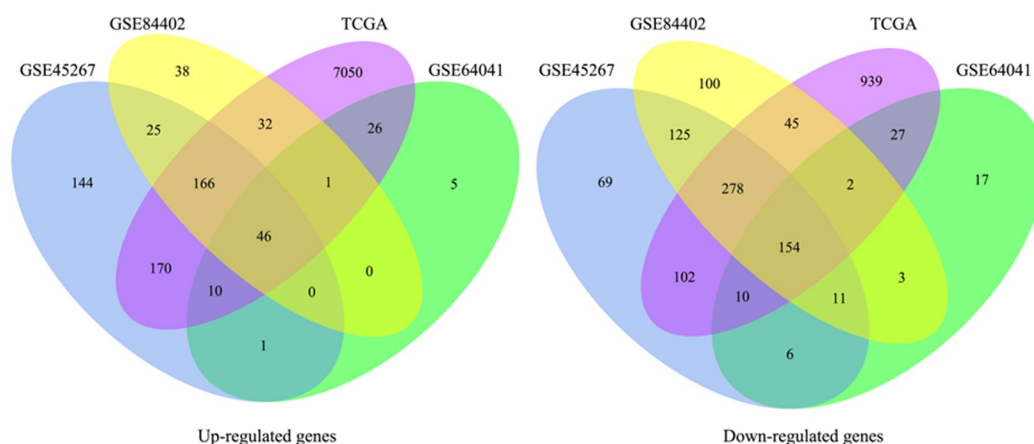


图2 GSE45267、GSE64041、GSE84402及TCGA中交集的DEG
Fig. 2 The overlapped DEGs among GSE45267, GSE64041, GSE84402 and TCGA

表1 在GSE45267、GSE64041、GSE84402和TCGA中筛选出的共有DEG
Tab.1 The common DEGs identified among GSE45267, GSE64041, GSE84402 and TCGA

DEG	Gene name
Up-regulated	GINS1, KPNA2, ANLN, FOXM1, CDK1, CDC6, AURKA, ITGA6, MELK, CCNA2, NUF2, MMP12, CKS2, ECT2, CCNB2, PRC1, CCNE2, DLGAP5, MKI67, GPC3, FLVCR1, EZH2, LEF1, CCL20, ROBO1, EDIL3, ASPM, RRM2, FANCI, DKK1, CAP2, DTL, KIF20A, PEG10, CDKN3, NCAPG, CENPF, NUSAP1
Down-regulated	BBOX1, XDH, ACOT12, CXCL14, IGF1, CYP39A1, PROZ, C8A, HRG, ZG16, MBL2, SLCO1B3, FCGR2B, BCO2, ACSM3, C3P1, TMEM45A, GHR, CLEC1B, BHMT, STEAP3, SHBG, DNASE1L3, BCHE, GPD1, CRHBP, F9, IDO2, BDH2, SLC38A4, ACADL, MT1M, SRD5A2, ADRA1A, ECM1, PBLD, KLKB1, MT1G, ABCA8, PGLYRP2, SLC22A1, STEAP4, FETUB, MFSD2A, APOA5, KCND3, CHST4, PPP1R3B, PCK1, ADH1C, CYP2C9, PIK3C2G, CLEC4M, NPY1R, RDH16, AFM, HPGD, LPAL2, THRSP, CYP4A11, AGXT2, MT1X, C7, AADAT, NNMT, MOGAT2, SERPINA4, EPHX2, HAPB2, APOF, ANGPTL1, GRAMD1C, SLC7A2, TUBE1, CNTN3, DPT, SLC25A47, MFAP4, SLC10A1, PRG4, CYP1A2, CCDC3, GYS2, CD5L, LPA, C8B, TAT, LIFR, COLEC10, VNN1, LYVE1, ALDH8A1, NAT2, AKR1D1, PAMR1, CXCL12, GNMT, WDR72, CYP3A43, AMDHD1, FBP1, ADH4, OIT3, GLYAT, CETP, SRPX, ENO3, LECT2, CYP2C8, RBMS3, SPP2, HAO2, MT1F, HHIP, LY6E, ITGA9, OLFML3, CPS1, CNDP1, FCN3, GBA3, PDGFRA, CLEC4G, CYP2B6, CCBE1, FXYP1, KMO, MASP1, ANK3, CLRN3, CFHR3, MT1H, TDO2, CLTRN, VIPR1, IGFBP3, PLAC8, HAMP, DCN, APOA1, CYP8B1, TIMD4, STAB2, HGFAC, ADGRG7, OGDHL, PZP, CYP3A4, JCHAIN, GLS2, ALPL, PTGIS, C9, CDHR2

2.3 PPI网络分析与中枢基因的筛选

PPI网络分析有利于从系统的角度发现与疾病相关的重要基因及其之间的相互联系。使用STRING在线数据库和Cytoscape软件,200个DEG中共有120个(38个上调基因和82个下调基因)被过滤到PPI网络复合体中,包含120个节点和526个边

缘(图3A)。在PPI网络中有些中心节点基因由于与多种基因之间具有相互作用,并且处于PPI网络中的关键位置,而被称为“中枢基因”,被认为是疾病发生发展的潜在驱动因素^[7],为了找出导致HCC发生的关键基因,利用CytoHubba插件计算了所有DEG的度值。如表4所示,在120个基因中有34个基因为中枢

基因(筛选标准为度值>10),其中上调基因较多(27/34),而下调基因较少(7/34)。此外,根据重要程度,利用 Cytoscape 中的 MCODE 插件筛选出了一个由 24

个节点和 272 个边缘组成的显著模块(图 3B,同时也是图 3A 中黄色区域),该模块均为上调基因且度值较高(21~30),颜色越深则度值越高。

表 2 GO 生物过程中上调和下调 DEG 显著富集的前 10 个生物过程

Tab.2 Top10 biological process terms with significantly enriched up-regulated and down-regulated DEGs in GO

GO ID	Biological process	Gene count	P	Gene symbol
Up-regulated				
GO:0051301	Cell division	10	4.98E-08	CCNE2, CDK1, CDC6, CCNB2, NCAPG, CKS2, NUF2, CENPF, AURKA, CCNA2
GO:0007067	Mitotic nuclear division	9	5.39E-08	CDK1, CDC6, CCNB2, NUF2, CENPF, AURKA, ANLN, CCNA2, ASPM
GO:0000082	G1/S transition of mitotic cell cycle	5	7.26E-05	CCNE2, CDK1, CDC6, RRM2, CDKN3
GO:0000079	Regulation of cyclin-dependent protein serine/threonine kinase activity	4	8.52E-05	CCNE2, CDC6, CDKN3, CCNA2
GO:0051726	Regulation of cell cycle	5	9.42E-06	CCNE2, CCNB2, DTL, FOXM1, CENPF
GO:0000086	G2 / M transition of mitotic cell cycle	5	2.27E-04	CDK1, CCNB2, FOXM1, AURKA, MELK
GO:0007095	Mitotic G2 DNA damage checkpoint	3	4.87E-04	CDK1, FANCI, CCNA2
GO:0000083	Regulation of transcription involved in G1 / S transition of mitotic cell cycle	3	0.001161	CDK1, CDC6, RRM2
GO:0008283	Cell proliferation	6	0.001176	CDK1, MKI67, DLGAP5, CKS2, CENPF, MELK
GO:0000281	Mitotic cytokinesis	3	0.001847	NUSAP1, ANLN, KIF20A
Down-regulated				
GO:0055114	Oxidation-reduction process	26	3.66E-11	STEAP3, CYP3A4, XDH, ALDH8A1, STEAP4, CYP2B6, KMO, BBOX1, CYP3A43, TDO2, CYP39A1, PTGIS, HHIP, SRD5A2, CYP2C9, CYP2C8, IDO2, CYP1A2, STAB2, ACADL, CYP4A11, CYP8B1, RDH16, HPGD, AKR1D1, BCO2
GO:0019373	Epoxygenase P450 pathway	6	3.39E-07	CYP4A11, CYP2B6, CYP2C9, CYP2C8, EPHX2, CYP1A2
GO:0042738	Exogenous drug catabolic process	5	2.43E-06	CYP3A4, CYP2B6, CYP2C9, CYP2C8, CYP1A2
GO:0017144	Drug metabolic process	6	3.00E-06	CYP3A4, CYP2B6, CYP2C9, CYP2C8, EPHX2, CYP1A2
GO:0006805	Xenobiotic metabolic process	8	4.68E-06	CYP3A4, GLYAT, CYP2B6, CYP2C9, CYP2C8, NAT2, EPHX2, CYP1A2
GO:0071276	Cellular response to cadmium ion	5	1.13E-05	CYP1A2, MT1H, MT1X, MT1G, MT1F
GO:0045926	Negative regulation of growth	5	1.82E-05	MT1M, MT1H, MT1X, MT1G, MT1F
GO:0071294	Cellular response to zinc ion	5	1.82E-05	MT1M, MT1H, MT1X, MT1G, MT1F
GO:0019441	Tryptophan catabolic process to kynurenine	4	2.11E-05	AADAT, TDO2, IDO2, KMO
GO:0006569	Tryptophan catabolic process	4	5.00E-05	AADAT, TDO2, IDO2, KMO

表3 KEGG中DEG显著富集的10个通路
Tab.3 Top 10 pathways with significantly enriched DEGs in KEGG

KEGG	Pathway	Gene count	P	Gene symbol
hsa01100	Metabolic pathway	38	9.05E-06	ALPL, CYP3A4, XDH, CNDP1, CYP2B6, OGDHL, ADH1C, KMO, GLS2, TDO2, PTGIS, ADH4, ENO3, BDH2, AADAT, PIK3C2G, CYP2C9, CYP2C8, NAT2, EPHX2, IDO2, FBP1, CYP1A2, CPS1, ACADL, TAT, PCK1, GBA3, ACSM3, AMDHD1, RRM2, HAO2, BHMT, AGXT2, RDH16, CYP8B1, AKR1D1, NNMT
hsa00830	Retinol metabolism	8	4.51E-05	CYP3A4, CYP2B6, CYP2C9, CYP2C8, ADH4, ADH1C, CYP1A2, RDH16
hsa04610	Complement and coagulation cascades	8	7.37E-05	C8A, MBL2, C8B, C7, C9, MASP1, KLKB1, F9
hsa05204	Chemical carcinogenesis	8	1.89E-04	CYP3A43, CYP3A4, CYP2C9, CYP2C8, ADH4, NAT2, ADH1C, CYP1A2
hsa00380	Tryptophan metabolism	6	3.10E-04	AADAT, TDO2, OGDHL, IDO2, KMO, CYP1A2
hsa04115	p53 signaling pathway	7	4.89E-04	STEAP3, CCNE2, CDK1, CCNB2, RRM2, IGF1, IGFBP3
hsa00982	Drug metabolism-cytochrome P450	7	5.29E-04	CYP3A4, CYP2B6, CYP2C9, CYP2C8, ADH4, ADH1C, CYP1A2
hsa00232	Caffeine metabolism	3	0.02198	XDH, NAT2, CYP1A2
hsa04978	Mineral absorption	5	0.004208	MT1M, MT1H, MT1X, MT1G, MT1F
hsa00980	Metabolism of xenobiotics by cytochrome P450	6	0.005059	CYP3A4, CYP2B6, CYP2C9, ADH4, ADH1C, CYP1A2

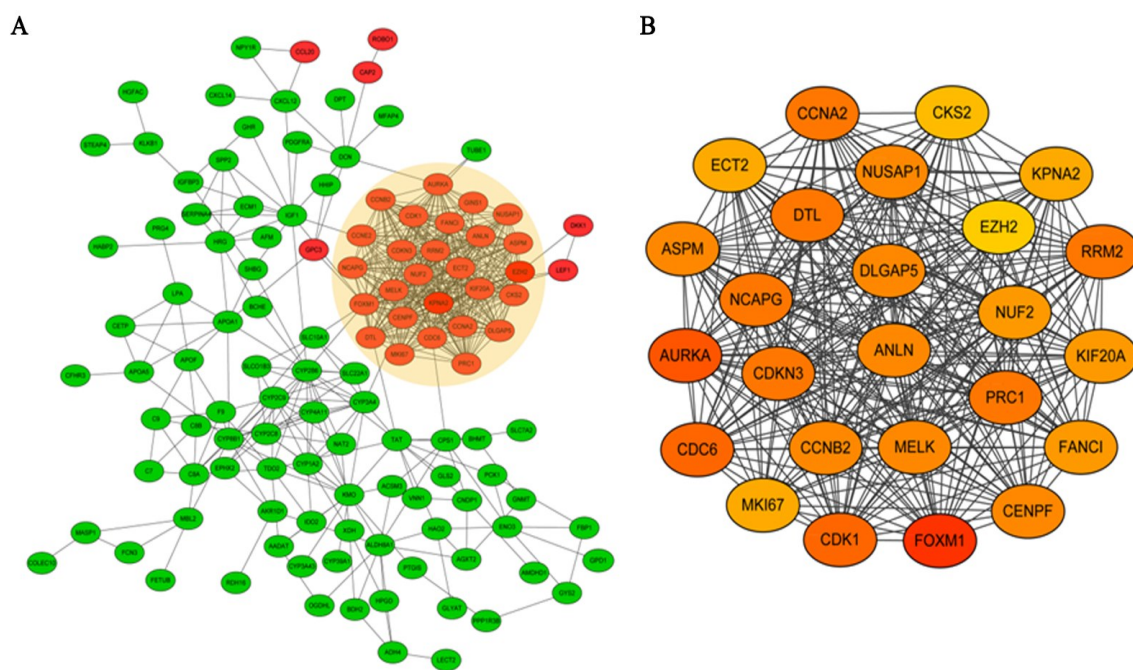


图3 DEG之间的PPI网络(A)和显著模块分析(B)
Fig.3 The constructed PPI network for DEGs(A) and analysis of significant modules(B)

2.4 中枢上调基因表达与HCC患者5年OS的关系

运用R软件中的Survival包对TCGA有临床和预后信息的310例患者进行了生存分析,其样本均为HCC肿瘤样本,并根据基因表达量分为高表达组和低表达组(图4红色和蓝色曲线),27个中枢上调基因中有6个基因的表达水平与患者5年OS率呈明显负相关,并且基因低表达组预后明显优于高表达组(CCNA2, $P < 0.01$; CDK1, $P < 0.01$; DLGAP5, $P < 0.01$; KIF20A, $P < 0.01$; KPNA2, $P < 0.01$; MELK, $P < 0.01$)。由此推测,这6个基因的表达水平可能在HCC患者预后发展中起到重要作用,且其高表达水平可能是预后不良的风险因素。

3 讨论

HCC作为世界上最常见的恶性肿瘤之一,虽然过去的几十年里众多基础和临床研究揭示了其发生发展过程中的各种诱因和潜在机制,HCC在世界范围内的发病率和病死率仍然很高^[11]。因此,挖掘未知的潜在治疗靶点与发病机制至关重要。如今,随着微阵列和高通量技术的发展,已经能够通过检测全

基因组中的基因表达量高低来预测癌症的潜在治疗靶点^[12]。

表4 PPI网络图中中枢基因连接度的统计结果

Tab.4 Statistical results of connectivity degree of the hub genes in the PPI network

Gene	Degree	Gene	Degree
FOXMI	30	FANCI	24
AURKA	28	KIF20A	24
CDC6	27	NUF2	24
CDK1	27	MKI67	23
CDKN3	26	KPNA2	23
DTL	26	ECT2	23
PRC1	26	CSK2	22
NCAPG	26	EZH2	21
RRM2	26	CYP2B6	16
CCNA2	26	KMO	16
MELK	25	CCNE2	16
ANLN	25	CYP2C9	14
ASPM	25	CYP3A4	13
NUSAP1	25	IGF1	13
CENPF	25	TAT	12
DLGAP5	25	CYP8B1	12
CCNB2	25	CYP2C8	11

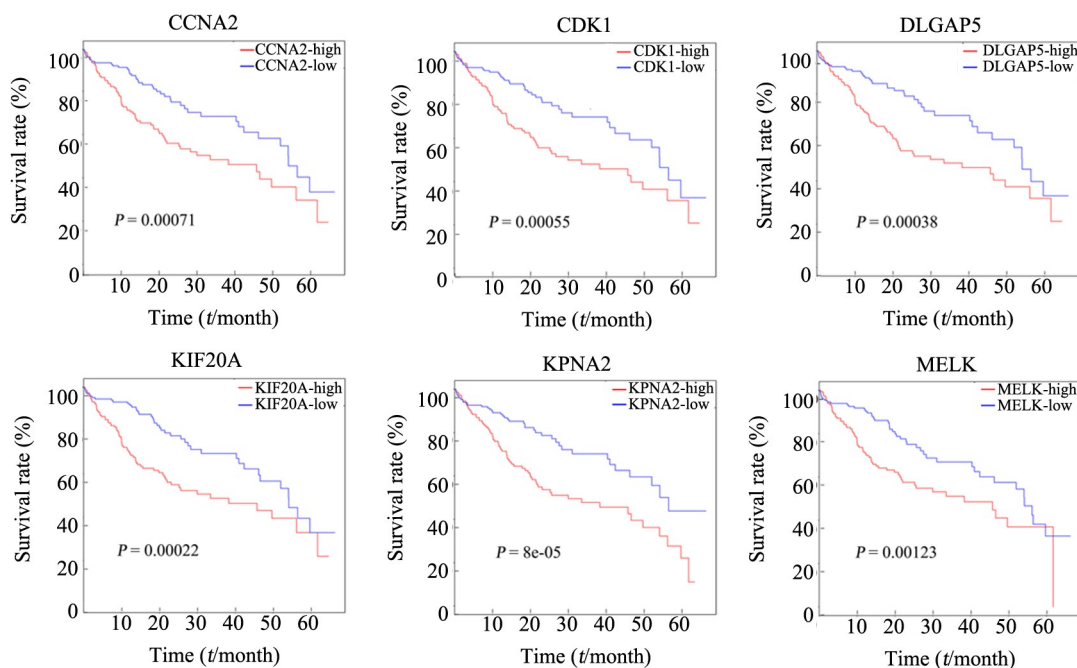


图4 6例中枢上调基因(其表达量与患者5年OS呈明显负相关)的Kaplan-Meier曲线

Fig. 4 Kaplan-Meier curves for six up-regulated hub genes, which are significantly correlated with the 5-year OS of patients

本研究中使用了标准化后的3个基因表达谱GSE45267、GSE64041、GSE84402和处理后的TCGA数据集,通过生物信息学分析,筛选出了HCC与正常组织样本之间差异表达的200个基因(46个上调基因和154个下调基因)。GO分析结果表明,DEG参与生物学过程为细胞分裂增殖、周期调控以及氧化还

原过程及某些代谢途径,这些均与肿瘤的发生发展有着密不可分的关系^[13-14]。KEGG通路富集分析表明,有7个DEG参与P53转导通路,包括3个上调基因(CCNE2、CDK1、CCNB2)和4个下调基因(STEAP3、RRM2、IGF1、IGFBP3)。此外,PPI网络与模块分析结果显示,上调基因占中枢基因中的大多

数(27/34),表明这些上调基因相比下调基因可能在HCC的发生发展中发挥更重要的作用。为了更好地探究中枢上调基因与HCC预后的关系,本研究在TCGA数据集中对其进行了生存分析,并选取了与患者预后呈显著负相关的6个基因CCNA2、CDK1、DLGAP5、KIF20A、KPNA2和MELK作为预后的潜在标志物。CCNA2编码的蛋白为细胞周期蛋白A2,主要调控细胞周期G1/S和G2/M转换,其表达被发现多种肿瘤中升高,包括乳腺癌、肝癌、前列腺癌和肺癌等。已有研究^[15]证明,一些癌前病变中CCNA2的表达与疾病的恶性进展有关(例如Barrett食管发展到食管腺癌,结肠腺瘤发展到结肠直肠癌)。最近研究^[16]表明,CCNA2基因的改变产生了一种侵袭性HCC的同质体,主要发生在非肝硬化患者,其特征在E2F和ATR途径的转录激活和RB1、PTEN的高失活频率。CDK1在二甲双胍诱导的miR-378作用下下调并且抑制HCC癌细胞增殖^[17];且CDK1抑制剂增加了索拉非尼在患者来源的异种移植(patient-derived xenotransplantation, PDX)HCC模型中的疗效^[18],以上均表现出了抗CDK1治疗在HCC靶向治疗中的应用前景。DLGAP5,其编码的蛋白又称肝癌上调蛋白(HURP)或者KIAA0008,在多种肿瘤中被检测到过表达,如HCC、非小细胞肺癌、鳞状细胞膀胱癌和移行细胞癌等,并在体外通过siRNA沉默DLGAP5后显著抑制HCC细胞的增殖和侵袭^[19];HCC中DLGAP5的上调与其启动子基因的甲基化水平有关,具体的机制有待于进一步探究^[20]。驱动蛋白家族成员20A(KIF20A)、核转运蛋白亚基 α 2(KPNA2)、母体胚胎亮氨酸拉链激酶(MELK)同样在一系列肿瘤中高表达,并已有少数文献证明其高表达与HCC的不良预后有关^[21-23]。以上均反映了这6个基因作为HCC诊断标志物和治疗靶点的潜在可能性。

综上所述,本研究通过整合多个数据集的生物信息学分析,发掘出了与HCC发生发展可能涉及的DEG和预后相关基因,而这些预后相关基因作为潜在的生物标志物,将有助于HCC的早期诊断与个体化治疗,并为后续的实验研究提供指导。此外,本研究作为现有数据库的二次挖掘,尚处于分子预测阶段,具体机制还需在以后的基础与临床研究中进一步验证。

[参考文献]

- [1] CHOU S T, CY HSIANG, HY L O, et al. Exploration of anti-cancer effects and mechanisms of Zuo-Jin-Wan and its alkaloid components in vitro and in orthotopic HepG2 xenograft immunocompetent mice[J/OL]. *BMC Complement Altern Med*, 2017, 17(1): 121 [2018-11-15]. <https://link.springer.com/article/10.1186/s12906-017-1586-6>. DOI:10.1186/s12906-017-1586-6.
- [2] CHEN W, ZHENG R, BAADE P D, et al. Cancer statistics in China, 2015[J]. *CA Cancer J Clin*, 2016, 66(2): 115-132. DOI: 10.3322/caac.21338.
- [3] BAFY G, BRUNT E M, CALDWELL S H. Hepatocellular carcinoma in non-alcoholic fatty liver disease: an emerging menace[J]. *J Hepatol*, 2012, 56(6): 1384-1391. DOI:10.1016/j.jhep.2011.10.027.
- [4] ASHHAB A A, RODIN H, POWELL J, et al. Hepatocellular carcinoma diagnosis and surveillance: socioeconomic factors don't seem to matter, unless you are an immigrant[J]. *J Hepatol*, 2017, 67(3): 648-649. DOI: 10.1016/j.jhep.2017.05.003.
- [5] CLOUGH E, BARRETT T. The gene expression omnibus database [J/OL]. *Methods Mol Biol*, 2016, 1418:93-110[2018-11-15]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4944384/>. DOI: 10.1007/978-1-4939-3578-9_5.
- [6] LEE J S. Exploring cancer genomic data from the cancer genome atlas project[J]. *BMB Rep*, 2016, 49(11): 607-611. DOI: 10.5483/BMBRep.2016.49.11.145.
- [7] XIAO Y, FENG M, RAN H, et al. Identification of key differentially expressed genes associated with non-small cell lung cancer by bioinformatics analyses[J]. *Mol Med Rep*, 2018, 17(5): 6379-6386. DOI: 10.3892/mmr.2018.8726
- [8] SONG J, PENG J, ZHU C, et al. Identification and validation of two novel prognostic lncRNAs in kidney renal clear cell carcinoma[J]. *Cell Physiol Biochem*, 2018, 48(6): 2549-2562. DOI: 10.1159/000491699.
- [9] DONG Z, WANG J, ZHAN T, et al. Identification of prognostic risk factors for esophageal adenocarcinoma using bioinformatics analysis [J/OL]. *Onco Targets Ther*, 2018, 11: 4327-4337[2018-11-15]. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC6065599/>. DOI: 10.2147/ott.s156716.
- [10] SU G, MORRIS J H, DEMCHAK B, et al. Biological network exploration with cytoscape3[J/OL]. *Curr Protoc Bioinformatics*, 2014, 47: 8.13.1-24[2018-11-15]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4174321/>. DOI: 10.1002/0471250953.bi0813s47.
- [11] TORRE L A, BRAY F, SIEGEL R L, et al. Global cancer statistics, 2012[J]. *CA Cancer J Clin*, 2015, 65(2): 87-108. DOI: 10.3322/caac.21262.
- [12] 张乐吟, 孙磊涛, 沈敏鹤. 基因变异检测技术在恶性肿瘤精准医疗中的应用[J]. *中国肿瘤生物治疗杂志*, 2019, 26(1): 22-28. DOI: 10.3872/j.issn.1007-385x.2019.01.005.
- [13] MALUMBRES M, BARBACID M. Cell cycle, CDKs and cancer: a changing paradigm[J]. *Nat Rev Cancer*, 2009, 9(3): 153-166. DOI: 10.1038/nrc2602.
- [14] DEBERARDINIS R J, CHANDEL N S. Fundamentals of cancer metabolism[J/OL]. *Sci Adv*, 2016, 2(5): e1600200[2018-10-15]. <http://www.ncbi.nlm.nih.gov/pmc/journals/2850/>. DOI: 10.1126/sciadv.1600200.
- [15] KIM D H, PARK S E, KIM M, et al. A functional single nucleotide polymorphism at the promoter region of cyclin A2 is associated with increased risk of colon, liver, and lung cancers[J]. *Cancer*, 2011, 117(17): 4080-4091. DOI: 10.1002/ncr.25930.
- [16] BAYARD Q, MEUNIER L, PENEAU C, et al. Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrange-

- ment signature of replication stress[J]. *Nat Commun*, 2018, 9(1): 5235. DOI: 10.1038/s41467-018-07552-9.
- [17] ZHOU J, HAN S, QIAN W, et al. Metformin induces miR-378 to downregulate the CDK1, leading to suppression of cell proliferation in hepatocellular carcinoma[J/OL]. *Onco Targets Ther*, 2018, 11: 4451-4459[2018-11-15]. <http://www.ncbi.nlm.nih.gov/pmc/journals/1214/>. DOI:10.2147/ott.s167614.
- [18] KUO T C, LU H P, CHAO C C, et al. The tyrosine kinase inhibitor sorafenib sensitizes hepatocellular carcinoma cells to taxol by suppressing the HURP protein[J]. *Biochem Pharmacol*, 2011, 82(2): 184-194. DOI:10.1016/j.bcp.2011.04.008.
- [19] WANG Q, CHEN Y, FENG H, et al. Prognostic and predictive value of HURP in nonsmall cell lung cancer[J]. *Oncol Rep*, 2018, 39(4): 1682-1692. DOI: 10.3892/or.2018.6280.
- [20] LIAO W, LIU W, YUAN Q, et al. Silencing of DLGAP5 by siRNA significantly inhibits the proliferation and invasion of hepatocellular carcinoma cells[J/OL]. *PLoS One*, 2013, 8(12): e80789[2018-11-15]. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3851768/>. DOI: 10.1371/journal.pone.0080789.
- [21] LU M, HUANG X, CHEN Y, et al. Aberrant KIF20A expression might independently predict poor overall survival and recurrence-free survival of hepatocellular carcinoma[J]. *IUBMB Life*, 2018, 70(4): 328-335. DOI: 10.1002/iub.1726.
- [22] JIANG P, TANG Y, HE L, et al. Aberrant expression of nuclear KPNA2 is correlated with early recurrence and poor prognosis in patients with small hepatocellular carcinoma after hepatectomy[J/OL]. *Med Oncol*, 2014, 31(8): 131[2018-11-15]. <https://link.springer.com/article/10.1007%2Fs12032-014-0131-4>. DOI:10.1007/s12032-014-0131-4.
- [23] WU S, CHEN X, HU C, et al. Up-regulated maternal embryonic leucine zipper kinase predicts poor prognosis of hepatocellular carcinoma patients in a chinese han population[J/OL]. *Med Sci Monit*, 2017, 23: 5705-5713[2018-11-15]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5721591/>. DOI:10.12659/MSM.907600.
- [收稿日期] 2018-11-25 [修回日期] 2019-03-09
[本文编辑] 党瑞山