

文章编号 1006-8147(2020)06-0514-04

论 著

## 基因表达谱分析肝细胞癌的特征基因

张萃萃<sup>1,2</sup>, 邓为民<sup>1</sup>

(1.天津医科大学基础医学院免疫学系,天津 300070;2.天津市血液中心输血科,天津 300110)

**摘要** 目的:生物信息学方法建立全面的肝细胞癌(HCC)差异表达基因谱以及初步筛选基于该恶性肿瘤样本的特征基因。方法:利用R语言中的edgeR包分析不同组之间差异表达的基因(DEGs)。利用STRING数据库分析预测蛋白质的功能联系和蛋白质相互作用。R语言的clusterProfiler包做GO(gene ontology)(包括biological process、molecular function和cellular component)及KEGG(kyoto encyclopedia of genes and genomes)通路富集分析。筛选差异基因显著富集的GO和KEGG分子途径。利用SLC7A11(solute carrier family 7 member 11)和CCDC14(coiled-coil domain-containing 14)基因作为肝细胞癌的标志物,利用ELISA法进行检测(确诊的HCC阳性患者为HCC组,共195例;HCC阴性正常人群为对照组,共107名),计算阳性率。结果:与癌旁对照样本相比,一共鉴定出454个差异表达基因,其中高表达基因234个,低表达基因220个。PPI方法进一步筛选出排名前10的重要潜在基因,其中SLC7A11和CCDC14两个基因排名靠前。利用富集分析,差异基因主要集中在neuroactive ligand-receptor interaction通路中。与对照组相比,HCC组SLC7A11和CCDC14存在显著的高表达[SLC7A11:(70.12±14.3) ng/mL比(23.12±7.11) ng/mL,  $t=6.17$ ,  $P<0.001$ ; CCDC14:(6.17±2.31) pg/mL比(2.13±0.84) pg/mL,  $t=5.35$ ,  $P<0.001$ ]。SLC7A11针对HCC检测阳性率达到80.2%,而CCDC14达到68.8%。结论:SLC7A11和CCDC14可以作为HCC的两个潜在的有效临床生物标志物。

**关键词** 肝细胞癌;PPI方法;ELISA;SLC7A11;CCDC14

中图分类号 R730.3

文献标志码 A

## Gene expression profiling reveals important characteristic genes in hepatocellular carcinoma

ZHANG Cui-cui<sup>1,2</sup>, DENG Wei-min<sup>1</sup>

(1. Department of Immunology, School of Basic Medical Sciences, Tianjin Medical University, Tianjin 300070, China; 2. Department of Blood Distribution, Tianjin Blood Center, Tianjin 300110, China)

**Abstract Objective:** To establish a comprehensive differentially expressed gene profile of hepatocellular carcinoma(HCC) by bioinformatics method and to screen the potential characteristic genes based on the malignant tumor samples. **Methods:** The edge package of R language was performed to analyze the differentially expressed genes(DEGs) among different groups. The STRING database was generated for better predicting functional relationship of protein and protein-protein interaction. Meanwhile, the cluster profiler functional package of R language was processed to generate GO(gene ontology) (including biological process, molecular function and cellular component) and KEGG(kyoto encyclopedia of genes and genomes) enrichment assay. GO and KEGG molecular pathways with significant enrichment of differentially expressed genes were screened. In addition, SLC7A11(solute carrier family 7 member 11) and CCDC14(coiled-coil domain-containing 14) genes were developed as specific biomarkers for HCC. ELISA method was utilized to analyze selected gene expression in external clinical experiments (HCC positive patients with 195 cases were HCC group, as well as HCC negative normal people with 107 cases were control group), and positive rate was calculated. **Results:** Compared with the control samples, 454 differentially expressed genes were identified, including 234 upregulated genes and 220 downregulated genes. PPI method further screened out the top 10 primary potential genes, of which SLC7A11 and CCDC14 ranked the top. Using the enrichment analysis, the differentially expressed genes were mainly concentrated in neuroactive ligand-receptor interaction pathway. The concentrations of SLC7A11 and CCDC14 were significantly higher in HCC positive patients group[SLC7A11:(70.12±14.3) ng/mL vs. (23.12±7.11) ng/mL,  $t=6.17$ ,  $P<0.001$ ; CCDC14:(6.17±2.31) pg/mL vs. (2.13±0.84) pg/mL,  $t=5.35$ ,  $P<0.001$ ]. The positive rate of SLC7A11 was 80.2% and of CCDC14 was 68.8% for HCC respectively. **Conclusion:** SLC7A11 and CCDC14 can be used as two potential effective clinical biomarkers for HCC detection.

**Key words** hepatocellular carcinoma; PPI method; ELISA; SLC7A11; CCDC14

肝细胞癌(hepatocellular carcinoma, HCC)是肝癌中最常见的原发性恶性肿瘤,占肝癌病例的75%~85%<sup>[1]</sup>。在世界范围内每年约有841 000例HCC新

增病例和782 000例死亡病例<sup>[2]</sup>。乙型肝炎病毒(hepatitis B virus, HBV)和丙型肝炎病毒(hepatitis C virus, HCV)是导致HCC发展的重要原因<sup>[3]</sup>。病毒诱导的发病机制涉及多种机制,例如HBV-DNA整合进入宿主遗传系统、DNA甲基化以及氧化应激<sup>[4-5]</sup>。肝脏慢性疾病是由于病毒通过各种受体介导的机制持续进

基金项目 天津市自然科学基金重点项目(18YFZCSY00040)

作者简介 张萃萃(1986-),女,主管技师,硕士在读,研究方向:肿瘤免疫;通信作者:邓为民, E-mail: dengweimin@tmu.edu.cn。

入宿主细胞而导致的,这些机制包括感染免疫防御控制中心、病毒抑制抗原呈递、选择性免疫抑制、病毒基因表达下调和病毒突变能够通过识别 HBV 抗原使病毒特异性 T 细胞功能失效等<sup>[6]</sup>。尽管 HBV 和 HCV 引起的病毒性肝炎与 HCC 高度相关,但仍有一些非病毒因素可诱发 HCC 的发生和发展,如糖尿病、酒精滥用、心血管疾病、肝脏炎症、肥胖、血脂异常和非酒精性脂肪性肝病(nonalcoholic fatty liver disease, NAFLD)等<sup>[7-9]</sup>。

如果在早期发现,HCC 患者可以得到良好的治疗,主要的治疗手段是手术切除或者肝移植<sup>[10]</sup>。但是,大多数 HCC 患者都在晚期才得到确诊,此时已经无法进行有效的手术治疗,转而采取化疗的方式破坏癌细胞并抑制癌细胞的增殖,治疗效果不尽如人意<sup>[11]</sup>。综上所述,HCC 治疗的关键是及早发现、及时诊断和手术治疗。然而目前,HCC 仍然缺少最为有效的肿瘤标志物。临床应用的肿瘤标志物阳性率尚无法令人满意,误诊情况屡有出现。为此,本研究结合生物信息学和临床检验方法,筛选出 HCC 中两个最为重要的靶点基因,然后构建了诊断模型,并且在临床样本中进行一对一精确验证。该研究丰富了 HCC 的早期诊断,为该疾病的研究提供了理论基础。

## 1 资料和方法

**1.1 差异基因筛选** 数据源 HCC 患者样本数据获取自 TCGA(the cancer genome atlas)数据库。本项目下载了 TCGA 数据库中 426 例 HCC 的 mRNA 数据,426 例样本中包含 340 例癌症样本和 86 例配对癌旁样本。TCGA 数据库链接为:<https://tcga-data.nci.nih.gov/tcga/>。本实验使用 R 语言中的 edgeR 包分析不同组之间差异表达的基因(differentially expressed genes, DEGs),以对数转换后的差异表达倍数(Log2FC)的绝对值>1 和  $P<0.05$  为标准进行筛选。

### 1.2 富集通路分析

**1.2.1 蛋白质-蛋白质相互作用(PPI)网络 and 核心基因的识别** 本项目利用 STRING(<https://string-db.org/>, version 11.0)分析蛋白质的功能联系及蛋白质的相互作用,保留综合得分大于等于 0.4 的相互作用对。用 Cytoscape(<https://cytoscape.org/>, version 3.7.2)可视化 PPI 网络。Cytoscape 软件中的 MCODE 插件识别显著的聚类模块,以 MCODE score > 2 为阈值进行筛选。

**1.2.2 功能富集分析** 用 R 语言的 clusterProfiler 包做 GO (gene ontology)(包括 biological process, molecular function 和 cellular component)及 KEGG(kyoto encyclopedia of genes and genomes)通路富集分析,以  $P<$

0.05 为标准筛选显著富集的 GO 和 KEGG 通路。

### 1.3 SLC7A11 和 CCDC14 浓度检测

**1.3.1 研究对象** 回顾性分析天津医科大学肿瘤医院 2018 年 1 月-2019 年 12 月经病理确诊的 HCC 患者。纳入标准:所有病例均经病理证实;有完整的影像学检查资料;未经放化疗。排除标准:存在其他肿瘤病史;存在其他肝脏慢性疾病。

本研究经本院伦理委员会批准(KY2018C278),并与患者或家属签署知情同意书。根据以上标准,将试验对象分为两组:HCC 阳性患者为 HCC 组,共 195 例,其中男性 114 例,女性 81 例;HCC 阴性正常人群为对照组,共 107 名,其中男性 60 名,女性 47 名。

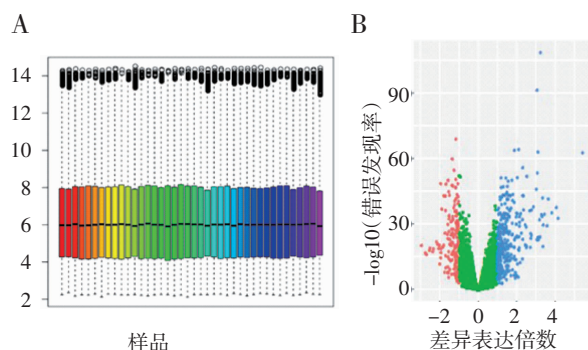
**1.3.2 ELISA 检测** SLC7A11(solute carrier family 7 member 11)和 CCDC(Coiled-coil domain-containing)14 浓度测定采用 ELISA 双抗体夹心法,具体操作严格按照试剂盒说明书进行(试剂盒选自 Abcam 公司),本项目的待测样品提取自外周血。为了减少待测人群冻存后细胞数量和检测因子受损严重情况,影响指标检测,本项目样本采用新鲜提取的外周血,分批次对待测人群进行检验,最终的数据进行统一整理。标准品的稀释与加样:在酶标包被板上设标准品孔 10 孔,在第一、第二孔中分别加标准品 100  $\mu\text{L}$ ,然后在第一、第二孔中加标准品稀释液 50  $\mu\text{L}$ ,混匀;以此类推,逐步稀释,最终稀释后各孔加样量都为 50  $\mu\text{L}$ ,浓度分别为 24  $\mu\text{g/L}$ 、16  $\mu\text{g/L}$ 、8  $\mu\text{g/L}$ 、4  $\mu\text{g/L}$  和 2  $\mu\text{g/L}$ ;加样:分别设空白孔(空白对照孔不加样品及酶标试剂,其余各步操作相同)、待测样品孔。在酶标包被板上待测样品孔中先加样品稀释液 40  $\mu\text{L}$ ,然后再加待测样品 10  $\mu\text{L}$ (样品最终稀释度为 5 倍)。将样品加于酶标板孔底部,尽量不触及孔壁,轻轻晃动混匀。温育:用封板膜封板后置 37℃温育 30 min。配液:将 30 倍(48T 的 20 倍)浓缩洗涤液用蒸馏水 30 倍(48T 的 20 倍)稀释后备用。洗涤:小心揭掉封板膜,弃去液体,甩干,每孔加满洗涤液,静置 30 s 后弃去,如此重复 5 次,拍干。加酶:每孔加入酶标试剂 50  $\mu\text{L}$ ,空白孔除外。进一步温育和洗涤。显色:每孔先加入显色剂 A 50  $\mu\text{L}$ ,再加入显色剂 B 50  $\mu\text{L}$ ,轻轻震荡混匀,37℃避光显色 15 min。终止:每孔加终止液 50  $\mu\text{L}$ ,终止反应(此时蓝色立转黄色)。测定:以空白调零,450 nm 波长依序测量各孔的吸光度(OD 值)。测定应在加终止液后 15 min 以内进行。

**1.4 统计学处理** 采用 Excel 2013 软件建立数据库,SAS 9.4 和 SPSS 19.0 统计软件进行数据分析。对连续性变量进行正态分布检验,符合正态分布的

变量以  $\bar{x} \pm s$  表示,采用 SAS 9.4 中的 *Paired-t test* 方法用于数据比较。 $P < 0.05$  为差异有统计学意义。

## 2 结果

**2.1 差异表达基因的分析结果** 为了消除批次误差效应,本项目首先对芯片原始数据进行标准化(图 1A)。340 例 HCC 样本和 86 例癌旁样本进行基因差异化分析结果显示:454 个基因表现出显著性差异,可列为 DEGs,其中上调基因 234 个,下调基因 220 个(图 1B)。

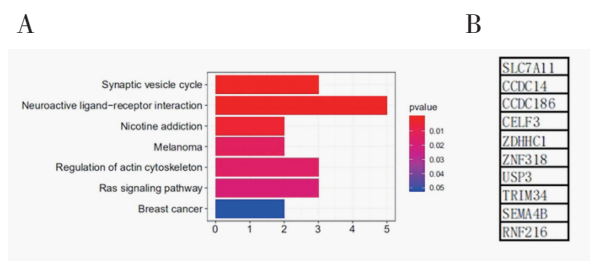


注: A:数据进行标准化的结果;B:差异表达基因的火山图,横轴表示差异表达的倍数,纵轴表示 $-\log_{10}$ (错误发现率),蓝点表示基因上调,红点为下调

图1 HCC 和癌旁组织差异基因分析

Fig 1 Analysis of differentially expressed genes in HCC cancer samples and adjacent tissues

**2.2 富集分析结果** 通过基因富集分析,差异表达基因主要集中在 neuroactive ligand-receptor interaction 通路中(图 2A)。PPI 网络筛选出针对 HCC 发展最重要的 10 个基因,其中 SLC7A11 和 CCDC14 排名最靠前,意味着这两个基因在 HCC 形成中的重要性最高(图 2B)。



注:A:显著富集的7条KEGG通路;B:通过PPI网络构建筛选出的重要性排名前10的基因

图2 KEGG 富集分析结果和 PPI 网络筛选的重要基因

Fig 2 Results of KEGG enrichment analysis and important genes screened by PPI network

**2.3 SLC7A11 和 CCDC14 检测结果** 与对照组相比,HCC 组 SLC7A11 浓度显著升高;CCDC14 浓度亦有明显提高。以发现的正常最高值作为检测标准,SLC7A11 和 CCDC14 针对 HCC 检测阳性率达到标准(表 1)。

表1 SLC7A11 和 CCDC14 浓度检测( $\bar{x} \pm s$ )

Tab 1 Analysis of concentration for SLC7A11 and CCDC14( $\bar{x} \pm s$ )

组别	SLC7A11(ng/mL)	CCDC14(pg/mL)
HCC 组	70.12±14.30*	6.17±2.31*
对照组	23.12±7.11	2.13±0.84
t	6.17	5.35
P	<0.001	<0.001
阳性率(%)	80.20	68.80

注:HCC:肝细胞癌;与对照组相比,\* $P < 0.05$

## 3 讨论

HCC 治疗效果主要取决于早期精确诊断、及时的手术治疗以及有效的预后预测模型。然而,目前临床上该恶性肿瘤尚缺乏令人满意的临床标志物。甲胎蛋白(alpha fetoprotein,AFP)是目前全世界应用最广泛的 HCC 肿瘤标志物,诊断 HCC 的敏感度为 39%~65%,特异度为 76%~94%。其敏感度和特异度均不令人十分满意,尤其在具有小肿块的 HCC 早期阶段,80%的患者血清 AFP 并未见明显升高。近几年 HCC 的基础和临床研究也提出了若干新型肿瘤标志物<sup>[12]</sup>。例如高尔基体蛋白 73(golgi protein 73,GP73)和甲胎蛋白异质体 3(alpha fetoprotein-L3,AFP-L3)等。GP73 在正常人的肝细胞中表达量极低或者不表达,而在 HCC 患者血清中明显升高,然而有诸多报道证实肝炎和脂肪肝患者中也存在着 GP73 水平显著升高。AFP-L3 是 AFP 异质体的一种。AFP-L3 为肝癌细胞特有,当临界值设为 10%~15% 时,约 1/3 的小肝癌(<3 cm)患者血清 AFP-L3 阳性;而>15%时,AFP-L3 的敏感度为 75%~96.9%,特异度为 90%~92%。然而有文献对于 AFP-L3 在 HCC 诊断方面有效性的报道却不一致,一项纳入 12 篇文章的 Meta 分析指出 AFP-L3 比 AFP 有更高的特异度,但敏感度较低<sup>[13]</sup>。基于这些事实,有必要寻找更有效、更可靠的 HCC 临床标志物。

本研究对 TCGA 中 HCC 样品和癌旁对照样品进行了差异表达基因分析,以揭示潜在的关键性差异基因。实验从 340 个癌症样本和 86 个相邻对照样本中筛选了 454 个差异基因,其中高表达的 234 个,低表达的 220 个。这些差异基因主要集中在神经活性配体-受体通路中。Liu 等<sup>[14]</sup>实验证明神经活性配体-受体通路与 HCC 相关,因为诸多在人肝中表达的基因参与神经活性配体-受体相互作用途径。另外,Zhao 等<sup>[15]</sup>发现,肝癌的早期、中期和晚期都存在神经活性的配体-受体相互作用。因此,该途径在 HCC 进展中似乎很重要。

本项目确定了若干个有显著表达差异的基因,



其中最为突出的两个基因是 SLC7A11 和 CCDC14, 这为将来的研究提供了非常有价值的起点。SLC7A11 是广泛存在于肝细胞膜上面的特异性离子进出通道, 之前有报道称其在 HCC 存在差异性表达。SLC7A11 在 HCC 中的差异性表达可能与 HCC 变异细胞中特定改变的分子信号有密切联系<sup>[6]</sup>。CCDC 卷曲螺旋结构域结构基因, 其表观遗传的改变已经证实与包括鼻咽癌、胰腺癌、前列腺癌在内的多种恶性肿瘤息息相关。其在 HCC 中的作用尚未明了<sup>[17]</sup>。

总体而言, 本项目全面揭示 HCC 的差异性基因图谱, 并且提出了两个可行的创新临床肿瘤标志物, 为 HCC 临床诊断和治疗提供了新的思路。本研究为单中心研究, 样本量有限, 尚存在一定的局限性。为了使 SLC7A11 和 CCDC14 作为 HCC 特异性的标志物广泛地应用于临床检验, 未来还需要进行多中心、大样本量研究。

#### 参考文献:

- [1] El-Serag H B. Hepatocellular carcinoma[J]. N Engl J Med, 2011, 365(1):1118
- [2] Lurje I, Czigany Z, Bednarsch J, et al. Treatment strategies for hepatocellular carcinoma a multidisciplinary approach[J]. Int J Mol Sci, 2019, 20(1):1
- [3] Colquhoun S D. Hepatocellular carcinoma[J]. Crit Rev Oncog, 2016, 21(1):93
- [4] Forner A, Reig M, Bruix J. Hepatocellular carcinoma[J]. Lancet, 2018, 391(1):1301
- [5] Li S, Yang F, Ren X. Immunotherapy for hepatocellular carcinoma[J]. Drug Discov Ther, 2015, 9(1):363
- [6] Villanueva A. Hepatocellular Carcinoma[J]. N Engl J Med, 2019, 38(1):1450
- [7] Golabi P, Rhea L, Henry L, et al. Hepatocellular carcinoma and non-alcoholic fatty liver disease[J]. Hepatol Int, 2019, 13(688):94
- [8] Kudo M. Targeted and immune therapies for hepatocellular carcinoma: predictions for 2019 and beyond [J]. World J Gastroenterol, 2019, 25(789):807
- [9] Schlachterman A, Craft W W, Hilgenfeldt E, et al. Current and future treatments for hepatocellular carcinoma[J]. World J Gastroenterol, 2015, 21(84):78
- [10] Greten T F, Lai C W, Li G, et al. Targeted and immune-based therapies for hepatocellular carcinoma[J]. Gastroenterol, 2019, 156(1):510
- [11] Grandhi M S, Kim A K, Ronnekleiv-Kelly S M, et al. Hepatocellular carcinoma: from diagnosis to treatment[J]. Surg Oncol, 2016, 25(1):74
- [12] 杨贵敏, 赵运胜, 王春华, 等. 肝细胞癌肿瘤标志物的研究进展[J]. 临床肝胆病杂志, 2018, 34(1):199
- [13] Wang L, Huang J, Jiang M, et al. AFP computational secreted network construction and analysis between human hepatocellular carcinoma (HCC) and no-tumor hepatitis/cirrhotic liver tissues[J]. Tumor Biol, 2010, 31(1):417
- [14] Liu Z, Gartenhaus R B, Tan M, et al. Gene and pathway identification with Lp penalized Bayesian logistic regression[J]. BMC Bioinformatics, 2008, 9(1):412
- [15] Zhao Y, Xue F, Sun J, et al. Genome-wide methylation profiling of the different stages of hepatitis B virus-related hepatocellular carcinoma development in plasma cell-free DNA reveals potential biomarkers for early detection and high-risk monitoring of hepatocellular carcinoma[J]. Clin Epigenetics, 2014, 6(1):30
- [16] Yue C, Ren Y, Ge H, et al. Comprehensive analysis of potential prognostic genes for the construction of a competing endogenous RNA regulatory network in hepatocellular carcinoma[J]. OncoTargets Ther, 2019, 12(1):561
- [17] 许建辉, 殷德涛. CCDC 与恶性肿瘤的研究进展[J]. 国际外科学杂志, 2014, 41(9):1

(2020-06-04 收稿)

·读者·作者·编者·

## 《天津医科大学学报》对缩略语的使用说明

文题原则上不能使用缩略语, 文中应尽量减少缩略语。公认的缩略语在文中可以直接使用。未公布的名词术语, 请按照如下规则进行缩写: 原词过长且在文中出现 3 次以上者, 可在第一次出现时写出全称, 并在括号内写出缩略语。不超过 5 个汉字的名称不宜使用缩略语, 以免影响文章的可读性。

缩略语	中文名称	缩略语	中文名称
ADA	美国糖尿病协会	MRI	磁共振成像
CT	电子计算机体层扫描	MtDNA	线粒体 DNA
ELISA	酶联免疫吸附试验	OR	优势比
HE	苏木素-伊红	PCR	聚合酶链反应
HIV	人类免疫缺陷病毒	PET	正电子发射断层摄影术
HbA1c	糖化血红蛋白	Real-time PCR	实时定量聚合酶链反应
HR	风险比	RT-PCR	反转录聚合酶链反应
ICU	重症监护治疗病房	WHO	世界卫生组织

本刊编辑部