

DOI:10.19296/j.cnki.1008-2409.2022-06-001

· 论 著 ·

基于 GEO 数据挖掘分析肝硬化-肝癌的演变机制^①

陈 燕^{1②}, 刘 军², 莫之婧^{1③}

(1. 桂林医学院广西高校生物化学与分子生物学重点实验室, 广西 桂林 541199; 2. 桂林优利特医疗电子有限公司, 广西 桂林 541004)

摘要 目的:应用生物信息学方法挖掘肝硬化-肝癌演变的关键基因,并探究其机制。方法:使用 GEO2R,从数据集 GSE6764 中筛选出在肝硬化组织和极早期肝癌组织中差异表达的基因,随后进行 GO 和 KEGG 分析;构建蛋白质-蛋白质相互作用网络后通过 Cytoscape 软件中的 MCODE 插件检测网络中最显著模块所含关键基因;对关键基因进行聚类分析和预后分析。结果:共筛选出 169 个差异表达基因,其中表达上调和下调的基因数目分别为 19 个和 150 个。MCODE 筛选后获得 16 个关键基因,主要与细胞因子和免疫等因素相关,其中关键基因 MPEG1 与肝癌患者无病生存期相关。结论:本研究发现的关键基因,有助于认识肝硬化-肝癌的演变机制,并可作为早期肝癌防治的潜在标志物。

关键词: 肝硬化;肝癌;生物信息学

中图分类号:R735.7

文献标志码:A

文章编号:1008-2409(2022)06-0001-05

Analysis of the evolution mechanism of liver cirrhosis and liver cancer based on GEO data mining^①

CHEN Yan^{1②}, LIU Jun², MO Zhijing^{1③}. (1. Key Laboratory of Biochemistry & Molecular Biology of Guangxi Institutions of Higher Learning, Guilin Medical University, Guilin 541199; 2. Urit Medical Electronic Co., Ltd., Guilin 541004, China)

Abstract Objective: To screen the hub genes in the evolution of liver cirrhosis and liver cancer by bioinformatics method so as to explore its mechanism. Methods: Differentially expressed genes in liver cirrhosis tissues and early hepatocellular carcinoma tissues were screened from dataset GSE6764 by using GEO2R, followed by GO and KEGG analysis. After constructing the protein-protein interaction network, the hub genes contained in the most significant modules in the network were detected by the MCODE plug-in with Cytoscape software. Then, cluster analysis and prognostic analysis were performed on hub genes. Results: A total of 169 differentially expressed genes were screened, and the numbers of up-regulated and down-regulated genes were 19 and 150, respectively. After MCODE screening, 16 hub genes were

① 基金项目:国家自然科学基金项目(32060159);广西自然科学基金项目(2019JJA140462)。

② 作者简介:陈燕(1998—),女,广西鹿寨人,2022年桂林医学院生物医学工程专业毕业。研究方向:系统生物学。

③ 通信作者:莫之婧,E-mail:mozhiqing@glmc.edu.com。

obtained, which were mainly related to cytokines and immune factors, among which, the hub gene MPEG1 was related to the disease-free survival of hepatocellular carcinoma (HCC) patients. Conclusion: The hub genes identified in this study will help us to understand the evolution mechanism of liver cirrhosis and cancer, which can be used as potential markers for the prevention and treatment of patients with early liver cancer.

Keywords: liver cirrhosis; liver cancer; bioinformatics

肝硬化是在临床上比较常见的一种慢性肝病,其病变部位范围极其广泛。这种病变主要是因为人体长期受到疾病影响而引发的,这些疾病通常是一种或多种类型。原发性肝细胞癌在世界各地普遍发生,且恶性程度极高。截止到2020年,在所有癌症中,肝癌发病人数世界排名第6位,死亡人数排第3位,死亡人数与新发病人非常接近^[1]。在我国,肝癌更是导致死亡的第二大癌症^[2]。越来越多的研究发现,肝硬化与肝癌之间有非常密切的关系。在临床病史上,由于身患肝硬化而一步步发展成为肝癌的患者占80%~90%。与此相反的是,没有患上肝硬化的患者发生肝癌的概率很小^[3-7]。可见研究肝硬化-肝癌的演变过程对肝癌的防治意义重大。

目前,对肝癌分子机制的研究一直都在进行中,但常规方法仅限于单个基因的深入分析,导致研究结果不具备代表性,存在一定的局限。而基因芯片作为一种新型前沿技术,开启了高通量表现基因变化的新模式,为全局理解基因功能和基因之间的相互关系及致病机制提供了新线索。通过该技术使全面综合分析肝硬化-肝癌的演变机制成为可能。本研究通过对基因表达汇编(gene expression omnibus, GEO)数据库中的基因芯片进行生物信息学分析,筛选出肝硬化与早期肝癌的差异表达基因,并挖掘出差异表达基因的生物信息学功能和信号通路,探讨肝硬化-肝癌演变的潜在分子机制,为肝癌的诊断、治疗和机制研究提供理论依据。

1 材料与方法

1.1 微阵列数据集

从GEO数据库中下载基因芯片GSE6764,内含10例癌旁肝硬化组织、8例极早期肝癌组织样本。

1.2 筛选差异表达基因

使用GEO2R,对GSE6764的10例癌旁肝硬化组织样本和8例极早期肝癌组织样本进行基因差异表达分析,设置筛选的标准: $|\log_2FC| > 1, P_{adj} < 0.01$ 。

1.3 对差异表达基因的GO和KEGG分析

使用DAVID数据库,对筛选出的差异表达基因进行GO和KEGG分析,以 $P < 0.05$ 为差异具有统计学意义。

1.4 蛋白质-蛋白质相互作用网络图的构建及关键基因的筛选

使用STRING数据库,将之前获取到的所有上调和下调的差异表达基因输入,获得蛋白质相互作用网络图并进行下载,结果保存并导入到Cytoscape软件,使用插件MCODE筛选出最显著模块,其包含的基因为关键基因。

1.5 关键基因的预后分析

使用GEPIA网站,对每个关键基因进行总生存期和无病生存期分析,以 $P < 0.05$ 为差异具有统计学意义。

2 结果

2.1 筛选肝硬化组织和极早期肝癌组织中的差异表达基因

GEO2R分析数据集GSE6764中肝硬化组织和极早期肝癌组织中的差异表达基因,结果显示,相对于肝硬化组织,在极早期肝癌组织中表达水平上调的基因有19个,下调的基因有150个,一共169个差异表达基因,见图1。

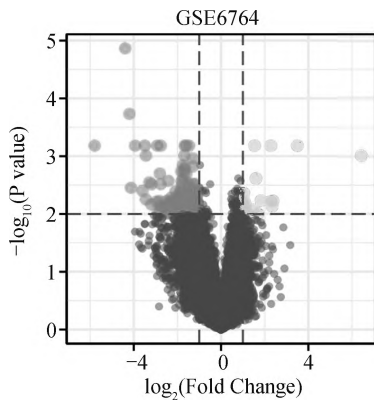


图 1 差异表达基因的火山图

2.2 差异表达基因的 GO 和 KEGG 分析

将 169 个差异表达基因,利用 DAVID 数据库进行 GO 分析和 KEGG 信号通路富集分析。GO 分析结果发现,169 个差异表达基因主要与中性粒细胞活化、中性粒细胞介导的免疫、免疫球蛋白结合、MHC 蛋白复合物结合等因素相关。KEGG 分析结果表明,差异表达基因主要参与吞噬体、金黄色葡萄球菌感染、利什曼病等,见图 2。

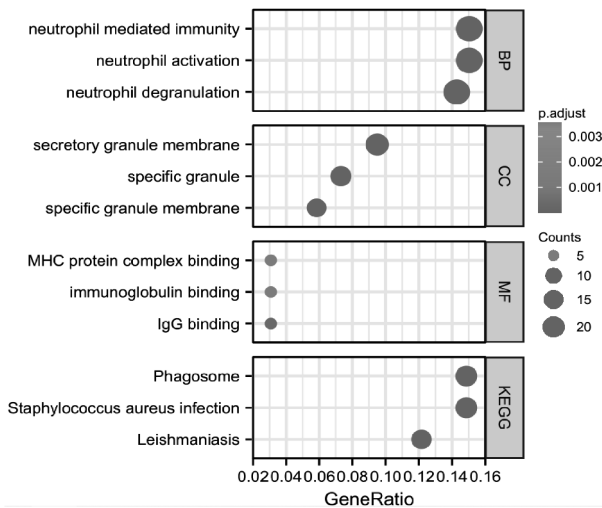


图 2 差异表达基因的 GO 和 KEGG 分析

2.3 蛋白质-蛋白质相互作用网络进行可视化及筛选最显著模块所含关键基因

STRING 数据库下载的蛋白质-蛋白质互相作用关系文件导入 Cytoscape 软件后,一共得到了由 97 个

节点和 385 个边组成的蛋白质-蛋白质相互作用网络图。使用 MCODE 插件从上述图中筛选出 16 个节点和 112 个边组成的最显著模块,该模块包含了 16 个在极早期肝癌组织中表达下调的基因:CYBB、FCGR3B、AIF1、ITGAM、LCP2、TLR4、TYROBP、CSF1R、MPEG1、C3AR1、FCER1G、HCK、C1QB、CD74、FCGR2B 和 FCGR2A,见图 3。

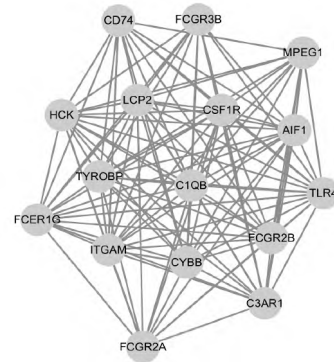


图 3 筛选的最显著模块

2.4 关键基因的聚类分析

对上述 16 个下调的关键基因进行聚类分析,结果显示,关键基因主要与细胞因子产生的正向调节、巨噬细胞活化、吞噬作用、免疫球蛋白结合等因素相关,见图 4。

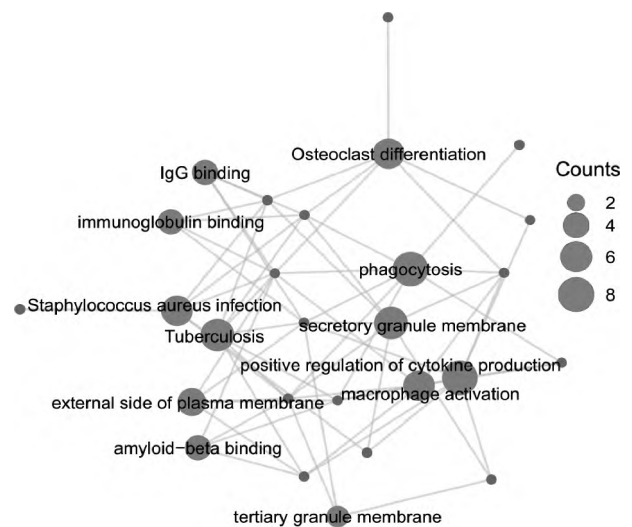


图 4 16 个关键基因的聚类分析网络图

2.5 关键基因 MPEG1 的预后分析

16个关键基因的生存曲线结果显示,高表达 MPEG1 的肝癌患者则生存率显著升高($P<0.05$),见图5。

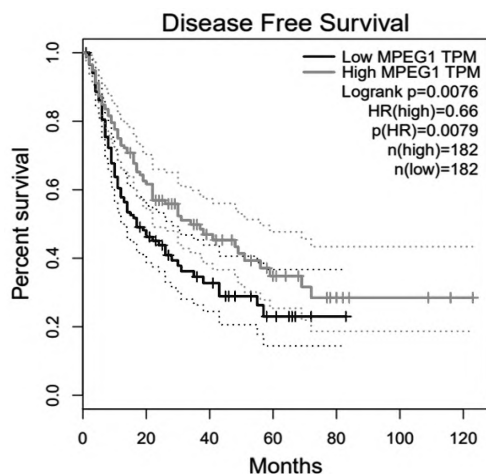


图5 肝癌患者 MPEG1 的表达与预后的生存曲线

3 讨论

肝硬化组织与极早期肝癌组织中差异表达基因有169个,这些基因主要与中性粒细胞活化、中性粒细胞介导的免疫、免疫球蛋白结合、MHC蛋白复合物结合等因素相关。中性粒细胞是先天免疫系统的关键细胞,通过包括吞噬、脱粒作用在内的一系列效应功能介导宿主防御。在癌症环境中,中性粒细胞的过度释放可能会调节组织微环境,最终导致肿瘤发生、生长和转移^[8]。KEGG分析结果表明,差异表达基因主要参与金黄色葡萄球菌感染、吞噬体等信号通路。金黄色葡萄球菌感染在癌症患者是严重的临床事件,死亡率极高^[9]。肿瘤微环境的细胞吞噬蛋白可以通过影响相关吞噬作用参与T淋巴细胞的免疫抑制,从而诱导肿瘤相关巨噬细胞在吞噬垂死的肿瘤细胞时控制肿瘤生长^[10]。这些研究结果提示,肝硬化-肝癌的演化过程可能与中性粒细胞活化及其介导的免疫相关。

从蛋白质-蛋白质相互作用网络中筛选出的16个关键基因均在极早期肝癌组织中表达下调。这些关键基因的聚类分析结果显示,主要与细胞因

子产生的正向调节、巨噬细胞活化、吞噬作用、免疫球蛋白结合等因素相关。细胞因子是由免疫细胞(单核、巨噬细胞等)和某些非免疫细胞经刺激而合成、分泌的一类具有广泛生物学活性的小分子蛋白质。细胞因子可调节细胞生长、分化等效应,调控免疫应答。如白细胞介素是肿瘤微环境中主要的一类细胞因子,已有多篇报道^[11]指出白细胞介素在癌症中表达失调的问题。炎症和癌症的关系十分紧密,主要体现在肿瘤微环境中,细胞因子的表达失调会影响各种信号途径以及代谢过程,从而引发肿瘤的发生和发展,这些重要的信号途径包括细胞的凋亡、存活、增殖和血管的生成、侵袭、转移等。

为了进一步探究关键基因在肝癌进展中的作用,将16个关键基因进行生存预后分析,发现在肝癌患者样本中高表达 MPEG1 患者的无病生存率显著升高。巨噬细胞基因1蛋白 MPEG1 是一种成孔蛋白,在防御细菌病原体中起关键作用,对于吞噬作用后的细菌杀伤至关重要^[12-13]。复旦大学中山医院对274例肝细胞癌患者的研究发现 MPEG1 在预后不良的患者中下调^[14],这与本研究结果趋势一致,提示 MPEG1 除了影响肝癌患者预后,可能还与肝硬化-肝癌的演变相关。

本研究通过生物信息学方法分析了肝硬化-肝癌基因芯片数据,确定了有差异表达的16个关键基因,揭示了肝硬化-肝癌的演化机制与多个基因、多种代谢途径、多条重要信号通路转导的异常密切相关,为肝癌的早期防治提供了理论基础。

参考文献:

- [1] SUNG H, FERLAY J, SIEGEL R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2021, 71(3): 209-249.
- [2] CAO W, CHEN H D, YU Y W, et al. Changing profiles of cancer burden worldwide and in china: a secondary analysis of the global cancer statistics 2020[J]. Chin Med J (Engl), 2021, 134(7): 783-791.
- [3] KUSNIK A, HUNTER N, RASBACH E, et al. Co-medica-

- tion and nutrition in hepatocellular carcinoma: potentially preventative strategies in hepatocellular carcinoma[J]. *Dig Dis*,2021,39(5):526-533.
- [4] CELSA C, CABIBBO G, PAGANO D, et al. Sicily network for Liver cancer: a multidisciplinary network model for the management of primary liver tumors [J]. *J Laparoendosc Adv Surg Tech A*,2020,30(10):1048-1053.
- [5] OH S, JO Y, JUNG S, et al. From genome sequencing to the discovery of potential biomarkers in liver disease [J]. *BMB Rep*,2020,53(6):299-310.
- [6] HUANG Z Y, ZHANG E L, CHEN X P. [Staging the severity of liver cirrhosis and decision of surgical treatment for hepatocellular carcinoma: Tongji experience][J]. *Zhonghua Wai ke Za zhi*,2019,57(6):408-411.
- [7] LIU Y, LI H, YE N, et al. Non-cirrhotic liver is associated with poor prognosis of hepatocellular carcinoma: a literature review[J]. *Med Sci Monit*,2019,25:6615-6623.
- [8] RAWAT K, SYEDA S, SHRIVASTAVA A. Neutrophil-derived granule cargoes: paving the way for tumor growth and progression[J]. *Cancer Metastasis Rev*,2021,40(1):221-244.
- [9] KANG C I, SONG J H, KO K S, et al. Clinical features and outcomes of *Staphylococcus aureus* infections in non-neutropenic cancer patients[J]. *Support Care Cancer*,2012,20(3):483-488.
- [10] CUNHA L D, YANG M, CARTER R, et al. LC3-associated phagocytosis in myeloid cells promotes tumor immune tolerance[J]. *Cell*,2018,175(2):429-441.
- [11] KUMARI N, DWARAKANATH B S, DAS A, et al. Role of interleukin-6 in cancer progression and therapeutic resistance[J]. *Tumour Biol*,2016,37(9):11553-11572.
- [12] HOWER S, MCCORMACK R, BARTRA S S, et al. LPS modifications and AvrA activity of salmonella enterica serovar typhimurium are required to prevent perforin-2 expression by infected fibroblasts and intestinal epithelial cells[J]. *Microb Pathog*,2021,154:104852.
- [13] BAYLY-JONES C, PANG S S, SPICER B A, et al. Ancient but not forgotten: new insights into MPEG1, a macrophage perforin-like immune effector[J]. *Front Immunol*,2020,11:581906.
- [14] ZHU G Q, YANG Y, CHEN E B, et al. Development and validation of a new tumor-based gene signature predicting prognosis of HBV/HCV-included resected hepatocellular carcinoma patients[J]. *J Transl Med*,2019,17(1):203.

[收稿日期:2022-07-12]

[责任编辑:李中原 英文编辑:阳雨君]