

基于 GEO 芯片数据的肝癌特征基因生物信息学及预后关联性分析

郁乐^{1,2}, 张轶雯³, 辛文秀³, 潘宗富³, 钟里科³, 黄萍^{1,3*} (1.浙江中医药大学药学院, 杭州 310053; 2.杭州市西溪医院, 杭州 310023; 3.浙江省肿瘤医院药剂科, 杭州 310022)

摘要: 目的 通过生物信息学方法利用 GEO 芯片数据分析肝癌组织中的特征基因及预后的相关性。方法 在 GEO 数据库中获取肝癌芯片数据, 并利用 GEO2R 工具分析肝癌组织与正常肝组织之间的差异表达基因; 利用 GO 数据库和 KEGG 数据库对差异表达基因进行富集和功能注释; 基于 STRING 数据库和 Cytoscape 软件构建蛋白质互相作用网络 (protein-protein interaction, PPI), 分析其关键基因; 用在线工具 Kaplan Meier-Plotter 对这些关键基因进行生存分析; 用 The Human Protein Atlas 数据库对这些关键基因在肝癌组织中的蛋白质表达进行免疫组化分析。结果 共鉴定出 338 个差异表达基因, 其中 97 个为上调基因, 241 个为下调基因。其中上调的差异表达基因显著富集在细胞核有丝分裂、细胞分裂、有丝分裂成对染色单体分离、成对染色单体凝聚等生物过程; 下调的差异表达基因显著富集在环氧合酶 P450 途径、氧化还原过程、外源性药物分解代谢过程等生物过程。根据 PPI 网络, 对关键模块的 24 个关键基因进行鉴定, 发现这些关键基因的高表达与肝癌患者的生存率低有相关性, 并截取关键差异表达基因免疫染色代表性图像。结论 本研究发现的关键差异基因有助于更全面地了解肝癌发生的分子机制, 可作为肝癌预后的生物标志物以及潜在的肝癌治疗分子靶点。**关键词:** 肝癌; 差异表达基因; 生物信息学分析; 关键基因

中图分类号: R966 文献标志码: A 文章编号: 1007-7693(2019)17-2177-06

DOI: 10.13748/j.cnki.issn1007-7693.2019.17.012

引用本文: 郁乐, 张轶雯, 辛文秀, 等. 基于 GEO 芯片数据的肝癌特征基因生物信息学及预后关联性分析[J]. 中国现代应用药学, 2019, 36(17): 2177-2182.

Bioinformatics and Prognosis Correlation Analysis of Liver Cancer Genes Based on GEO Chip Data

YU Le^{1,2}, ZHANG Yiwen³, XIN Wenxiu³, PAN Zongfu³, ZHONG Like³, HUANG Ping^{1,3*} (1.School of Pharmacy, Zhejiang Chinese Medical University, Hangzhou 310053, China; 2.Xixi Hospital of Hangzhou, Hangzhou 310023, China; 3.Department of Pharmacy, Zhejiang Cancer Hospital, Hangzhou 310022, China)

ABSTRACT: OBJECTIVE To use the bioinformatics method based on data of GEO chip to analyze the development of the molecular markers and prognosis in liver cancer. **METHODS** Liver cancer chip data was obtained in GEO database. The differentially expressed genes in liver cancer tissues and normal liver tissues were identified by GEO2R; the pathway enrichment of differentially expressed genes were performed by using KEGG and GO. The protein-protein interaction(PPI) network of differentially expressed genes was constructed by STRING database and visualized by Cytoscape. The perform survival analysis of these key genes were analyzed by online tool Kaplan Meier-Plotter. The immunohistochemistry analysis of these key genes were analyzed by The Human Protein Atlas Database. **RESULTS** A total of 338 differentially expressed genes were identified, including 97 up-regulated genes and 241 down-regulated genes. These up-regulated differentially expressed genes were significantly enriched in mitotic nuclear division, cell division, mitotic sister chromatid segregation, sister chromatid cohesion. These down-regulated differentially expressed genes were significantly enriched in epoxygenase P450 pathway, oxidation-reduction process, exogenous drug catabolic process. The PPI network was constructed a key module by 24 key genes. These key genes were found to be associated with poor survival in patients with liver cancer. Representative images of key differentially expressed genes immunostained were taken. **CONCLUSION** These key genes are found by this study which contributed to understanding the molecular mechanism of liver cancer, and can be used as a biomarker for the prognosis of liver cancer and a molecular target for the treatment of liver cancer.

KEYWORDS: liver cancer; differentially expressed gene; bioinformatics analysis; key genes

肝癌是世界第 5 位常见癌症, 在全世界消化系统肿瘤中, 死亡率占第 3 位^[1]。在我国, 发病人 数约占全球 55%, 其病死率在消化系统恶性肿瘤中排第 2 位^[2-3]。由于肝癌的进展快、复发转移率

基金项目: 国家自然科学基金项目(81503165); 浙江省自然科学基金项目(LQ18H160018, LQ17H310002); 浙江省卫生高层次创新人才(黄萍); 浙江省 151 人才工程第二层次(黄萍)

作者简介: 郁乐, 男, 药师 Tel: (0571)86481604 E-mail: 993832953@qq.com *通信作者: 黄萍, 女, 博士, 主任药师 Tel: (0571)88122438 E-mail: huangping1841@zjcc.org.cn

高,除较早期小肝癌切除手术外,迄今仍无有效治疗方法,5年生存率只有30%左右^[3]。近年来,以索拉菲尼为代表的靶向药物,为肝癌的靶向治疗开辟了新的道路,然而靶向药物仍然有其局限性,长期使用会产生耐药性^[4-5]。因此,对肝癌发病机制的研究,不断发现新的靶点仍是医学界的研究热点^[6]。

高通量基因芯片和测序技术的发展,为快速研究肝癌的基因表达谱、发现肝癌组织和细胞在特定状态下的基因表达情况和关键基因变化规律提供了可能^[7]。生物信息学以基因芯片研究为基础,将基因芯片技术产生的海量的、复杂的生物信息数据,运用序列比对、统计分析、可视化作图、生物聚类、生物分子网络及通路分析等方法,进行生物信息数据挖掘,从而能更加全面地对疾病进行系统研究^[8-9]。

本研究利用生物信息学的初步挖掘,筛选出有关的功能基因,希望能获得更多与肝癌发生转化相关的分子机制的生物学信息。

1 材料与方法

1.1 材料

利用美国国立生物技术信息中心(NCBI)平台下的基因表达综合数据库 GEO (<http://www.ncbi.nlm.gov/geo/>),检索含有人源肝癌样本的数据芯片,选取含有肝癌病因信息的数据芯片 GSE62232^[9]。芯片信息: Affymetrix Human Genome U133 Plus 2.0 Array,平台是 GPL570,该芯片包含 81 例肝癌组织和 10 例正常肝组织的基因表达阵列数据,且 81 例肝癌组织已标明患者发病原因、性别、年龄及癌症分期等信息。

1.2 差异表达基因筛选

利用在线工具 GEO2R (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>)对所选的芯片进行数据下载及差异表达基因筛选。筛选标准为 $\text{adj. } P\text{-Val} < 0.01$, $\log \text{FC} > 2$ 或 $\log \text{FC} < -2$,并将探针名称转化为标准基因名称。

1.3 基因功能富集和注释

通过 DAVID 6.8 数据库(<https://david.ncifcrf.gov/>)可对基因进行注释、可视化和综合探索。将已分析好的差异表达基因导入该数据库,依据 GO(Gene Ontology)数据库,以人源基因为背景,对差异表达基因进行生物学功能注释。并利用 KEGG(Kyoto Encyclopedia of Genes and Genomes)通路数据库进行信号通路的富集,寻找差异表达

基因所富集的关键通路。

1.4 差异表达基因的相互作用分析

通过 STRING 10.5 数据库(<http://string-db.org/>),构建肝癌差异表达基因蛋白质相互作用(protein protein interaction, PPI)网络,并应用 Cytoscape 3.6.1 软件进行可视化分析。使用 CytoHubba 插件对蛋白质相互作用网络进行关联度分析,并根据节点度大小进行排序。

1.5 差异表达基因的预后价值分析

利用 Kaplan Meier-Plotter(<http://kmplot.com/analysis/>)在线软件,分析差异表达基因对应肝癌患者的总生存期(overall survival, OS)。根据肝癌患者差异表达基因的表达情况及中位值分为高表达和低表达 2 组,并计算危险比(HR)及其 95%置信区间和对应的 P 值,绘制生存曲线。

1.6 免疫组织化学分析

利用 The Human Protein Atlas 数据库 (<http://www.proteinatlas.org/>)分析关键差异表达基因在肝癌组织中蛋白质表达情况。根据蛋白质在组织中的染色强度和染色细胞的百分比,比较正常组织和肿瘤组织中关键基因蛋白质表达的差异,并截取具有代表性的免疫染色图像。

2 结果

2.1 数据预处理与差异表达分析

将 GSE62232 芯片中的 10 例正常肝组织为空白组,81 例肝癌组织为对照组,共筛选出 338 个基因,其中 97 个为上调基因,241 个为下调基因,见图 1。

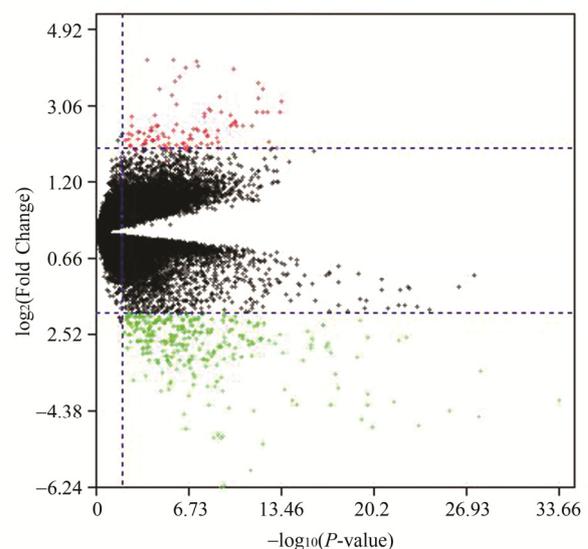


图 1 基因的表达水平及分布火山图

Fig. 1 Volcano map of the gene's expression level and distribution

2.2 GO 富集分析

通过 GO 富集分析差异基因相关的生物进程发现, 上调基因显著富集在细胞核有丝分裂、细胞分裂、有丝分裂成对染色单体分离、成对染色单体凝聚等生物过程, 见图 2A; 下调基因显著富集在环氧化酶 P450 途径、氧化还原过程、外源性药物分解代谢过程等生物过程, 见图 2B。

2.3 KEGG 富集分析

进一步对差异表达基因进行 KEGG 通路富集分析发现, 上调的基因主要富集于细胞周期、P53 信号通路、孕酮介导的卵母细胞成熟通路、ECM-受体相互作用通路、PI3K-Akt 信号通路等, 见图 3A, 而下调的基因主要富集于视黄醇代谢通路、化学致癌作用通路、代谢途径、矿物质吸收通路

等, 见图 3B。

2.4 蛋白质相互作用网络的构建和关键节点分析

用 STRING 10.5 数据库和 Cytoscape 构建差异表达基因之间的蛋白质相互作用网络。去除游离的蛋白影响后, 得到了由 182 个点, 931 条边构成的网络, 见图 4A。用 Cyto Hubba 插件对网络进行分析, 取其中相互作用程度最高的前 40 个基因。发现其中关键模块有 24 个基因并列排名第 1, 见图 4B, 且这 24 个关键基因均被上调。这 24 个基因分别是: DTL、KIAA0101、ANLN、ASPM、NUSAP1、CDKN3、PBK、PRC1、HMMR、TOP2A、NCAPG、DLGAP5、RRM2、KIF20A、NDC80、NEK2、RACGAP1、ZWINT、TTK、MAD2L1、CCNA2、BUB1B、CCNB1、CDK1。

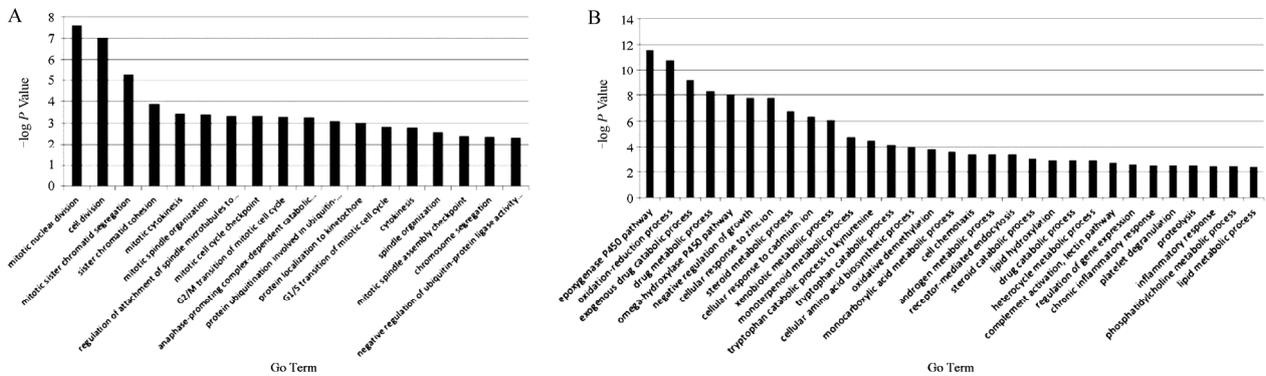


图 2 差异表达基因注释结果
Fig. 2 Differentially expressed gene annotation results

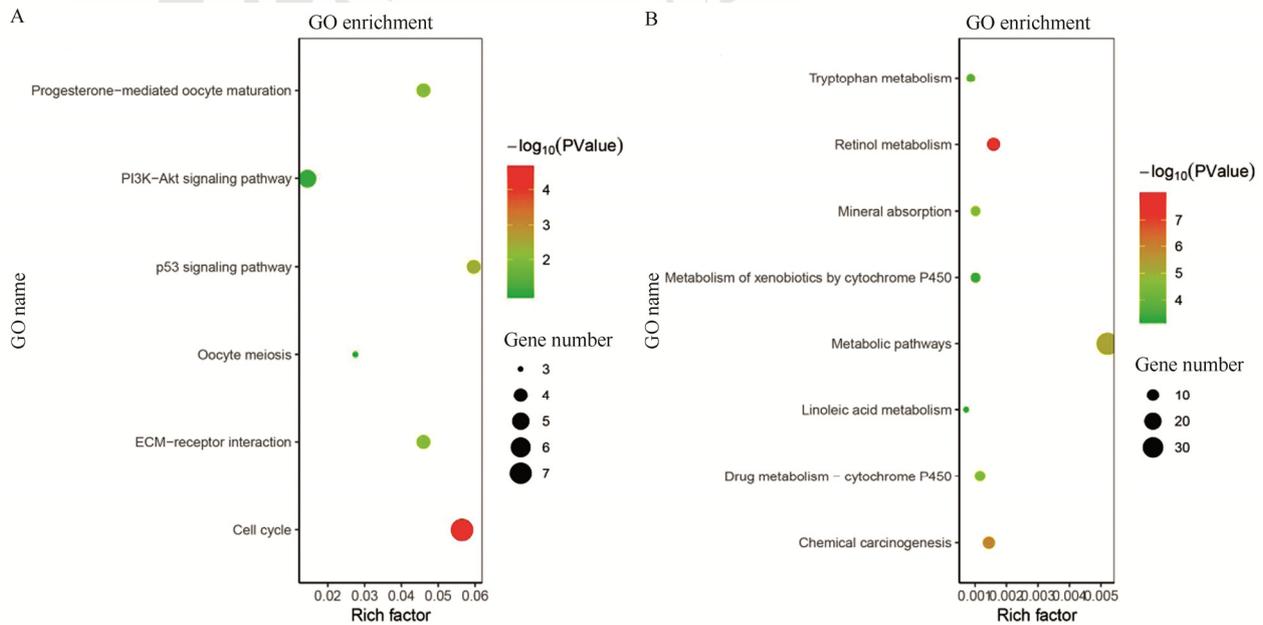


图 3 差异表达基因通路注释结果
Fig. 3 Differentially expressed gene pathway annotation results

2.5 预后价值分析

利用在线工具 Kaplan Meier-Plotter 对关键模块中的 24 个差异表达基因的预后价值进行评价。

结果显示, 差异表达最显著的 4 个基因 RRM2、TOP2A、DTL、CDK1 的高表达组 OS 显著低于低表达组, 见图 5。

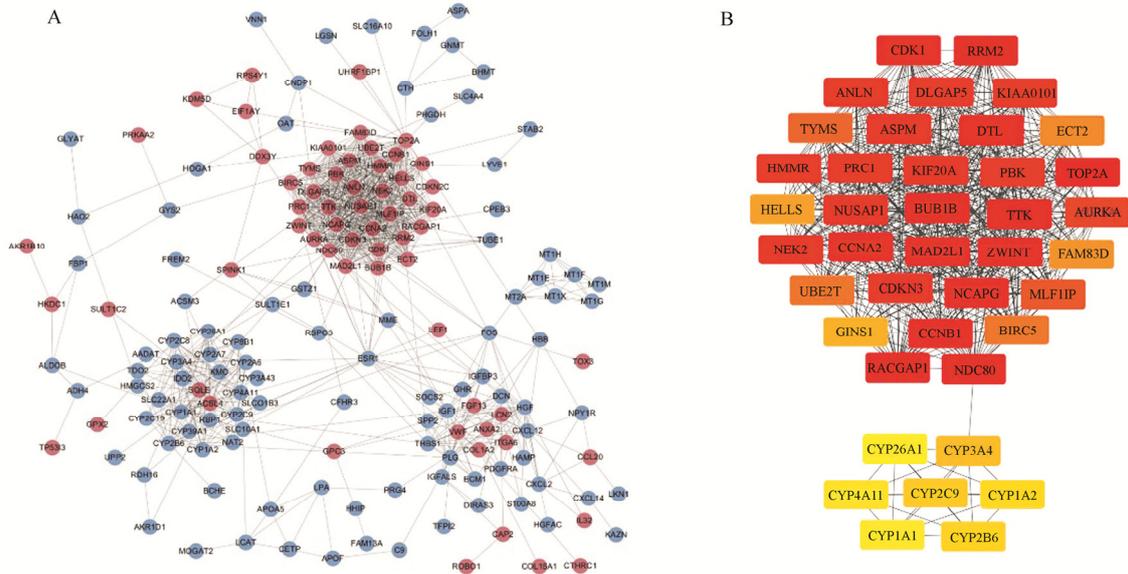


图 4 差异表达基因蛋白质相互作用网络可视化结果

Fig. 4 Differentially expressed gene protein interaction network visualization results

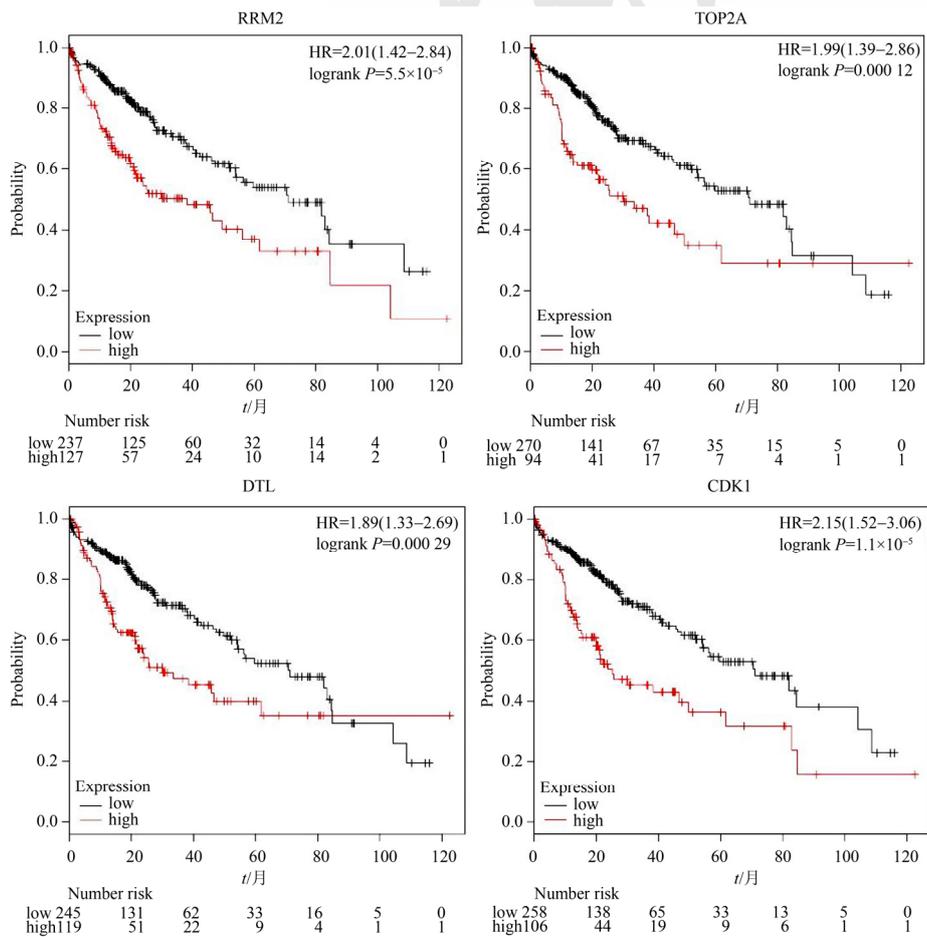


图 5 RRM2、TOP2A、DTL、CDK1 在肝癌中的预后价值

Fig. 5 Prognosis of RRM2, TOP2A, DTL and CDK1 in liver cancer

2.6 免疫组织化学分析

通过 The Human Protein Atlas 数据库对差异表达最显著的 4 个基因进行搜索,发现这些基因在肝癌组织中均呈阳性反应。利用抗体 HPA056994 在正常肝组织中无法检测到 RRM2,而在肝癌组织中呈现中等强度染色,染色细胞比例为 25%~75%。利用抗体 HPA00 6458、HPA026773 以及 CAB002448 在正常肝组织中无法检测到 TOP2A,而在肝癌组织中呈现中等到高强度染色,染色细胞比例 25%~75%。利用抗体 HPA028016 在正常肝组织中可以低强度检测到 DTL,而在肝癌组织中呈现中等强度染色,染色细胞比例为 25%~75%。利用抗体 HPA003387 和 CAB00 3799 在正常肝组织中无法检测到 CDK1,而在肝癌组织中呈强染色,染色细胞比例为<25%。4 个基因代表性免疫染色图像,见图 6。

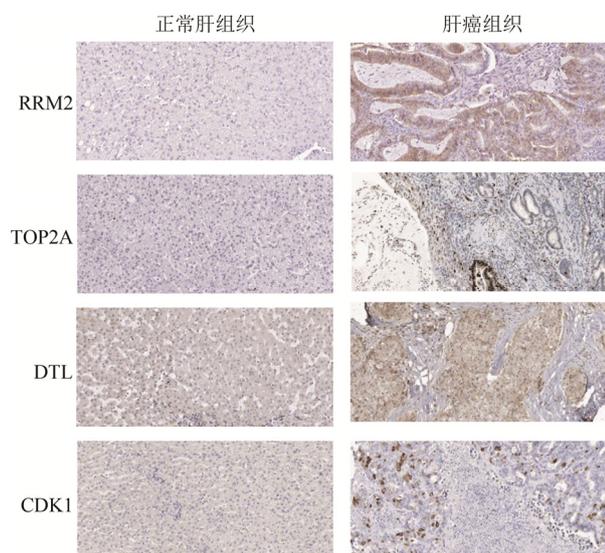


图 6 RRM2、TOP2A、DTL、CDK1 在肝癌中免疫组化代表性图像

Fig. 6 Representative images of RRM2, TOP2A, DTL and CDK1 in liver cancer

3 讨论

肝癌的发生、发展与转移被认为是一个非常复杂的过程,为深入了解其中的分子机制,筛选可作为肝癌诊断的关键基因和药物治疗靶点,本研究利用生物信息学分析方法使用 GEO 数据库中的 GSE62232 芯片来筛选肝癌与正常组织的差异表达基因,共获得 338 个基因,其中 97 个为上调基因,241 个为下调基因。

通过 GO 和 KEGG 富集分析,发现上调的差异表达基因富集在细胞核有丝分裂、细胞分裂、

细胞周期通路、P53 信号通路等生物过程,反映了肝癌组织中细胞增殖旺盛,细胞凋亡调控发生异常。近年来的研究已经证实^[10],细胞凋亡调控与肝癌的发生、发展、消退等存在着紧密联系,而靶向细胞凋亡信号通路的抗肿瘤新药开发也愈来愈受到关注。另外,在 P53 信号通路中,P53 作为抑癌基因参与了细胞生长、分化及凋亡的调控。P53 信号通路异常,能够促进肝癌细胞恶性增殖和凋亡抑制^[11]。而下调的差异表达基因主要富集在环氧化酶 P450 途径、氧化还原过程、外源性药物分解代谢、视黄醇代谢通路、化学致癌作用通路等生物过程,这些过程主要涉及 3 大营养物质代谢途径、外源物质的代谢等,与肝癌患者出现的肝功能紊乱所表现的症状一致^[12]。肝脏是人体重要的代谢器官,肝脏功能受损影响细胞内氧化还原状态的失衡,可直接导致细胞内生物分子损伤。

STRING 10.5 数据库和 Cytoscape 软件分析结果显示,关键模块中的基因与其他差异表达基因之间存在大量相互作用。对这些基因进行文献挖掘,证实上述基因与肝癌的发生、浸润、转移及预后关系密切。RRM2 富集在细胞周期通路上,调控细胞的分裂。RRM2 的高表达增强了 AKT 信号通路,增加了肿瘤细胞的耐药性^[13]。RRM2 可作为一个新的治疗靶点,增加肿瘤细胞对化疗药物的敏感性。TOP2A、DTL、CDK1 与 DNA 的复制和细胞的有丝分裂有关。TOP2A 的高表达,会导致癌细胞的不断分裂增殖^[14]。可以针对 TOP2A 设计药物,以控制肝癌的发展。有研究^[15]阻断 DTL 的表达后,能阻止细胞周期进程,减少 DNA 损伤,并导致癌细胞的生长抑制。这可能作为一种新的癌症治疗思路。由于 CDK1 对细胞生存具有多方面的作用^[16],CDK1 抑制剂有潜力成为有效的抗癌药物,联合化疗药物使用,将增强多种其他化疗药物的细胞毒性。

综上所述,本研究采用生物信息学的方法对肝癌基因芯片数据进行挖掘。这些信息有助于人们对于肝癌发生、发展和转移的进一步认识,并且可以为临床检测和治疗肝癌提供潜在的生物标志物及靶点。然而,缺乏进一步的分子生物学实验来确认这些基因在肝细胞中的功能,是这些分析所局限的。尽管如此,这项研究仍有助于全面了解肝癌的潜在分子机制,并为后续的实验研究提供指导。

REFERENCES

- [1] SAID A, WILLIAMS J, HOLDEN J, et al. Model for end stage liver disease score predicts mortality across a broad spectrum of liver disease [J]. *Hepatology*, 2004(40): 897-903.
- [2] IOANNOU G N, PERKINS J D, CARITHERS R L. Liver transplantation for hepatocellular carcinoma: Impact of the MELD allocation system and predictors of survival [J]. *Gastroenterology*, 2008(134): 1342-1351.
- [3] POMFRET EA, WASHBURN K, WALD C, et al. Report of a national conference on liver allocation in patients with hepatocellular carcinoma in the United States [J]. *Liver Transpl*, 2010(16): 262-278.
- [4] 董志珍, 朱小东, 李照, 等. 2016 年肝癌基础与临床研究进展[J]. *中华肝脏病杂志*, 2017, 25(2): 85-93.
- [5] XU X, LING Q, WANG J, et al. Donor miR-196a-2 polymorphism is associated with hepatocellular carcinoma recurrence after liver transplantation in a Han Chinese population [J]. *Int J Cancer*, 2016, 138(3): 620-629.
- [6] SCHULZE K, IMBEAUD S, LETOUZE E, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets [J]. *Nat Genet*, 2015, 47(5): 505-511.
- [7] ZHANG Y W, ZHONG L K, LOU Q W, et al. Study on the HPV-positive oropharyngeal cancer features gene based on GEO database by bioinformatics [J]. *Chin J Mod Appl Pharm*(中国现代应用药理学), 2018, 5(35): 638-641.
- [8] PINKEL D, ALBERTSON D G. Array comparative genomic hybridization and its applications in cancer [J]. *Nature Genetics Supplement*, 2005(37): S11-S17.
- [9] FUJIMOTO A, TOTOKI Y, ABE T, et al. Whole-genome sequencing of liver cancers identifies etiological influences mutation patterns and recurrent mutations in chromatin regulators [J]. *Nature Genetics Nat Genet*, 2012, 44(7): 760-764.
- [10] SIA D, VILLANUEVA A, FRIEDMAN S L, et al. Liver cancer cell of origin, molecular class, and effects on patient prognosis [J]. *Gastroenterology*, 2017, 152(4): 745-761.
- [11] DHAR D, ANTONUCCI L, NAKAGAWA H, et al. Liver cancer initiation requires p53 inhibition by CD44-enhanced growth factor signaling [J]. *Cancer Cell*, 2018, 11(33): 1061-1077.
- [12] SHOSTAK A. Circadian clock, cell division, and cancer: from molecules to organism [J]. *Int J Mol Sci*, 2017(18): 873-888.
- [13] WANG Y, ZHI Q, YE Q, et al. SCYL1-BP1 affects cell cycle arrest in human hepatocellular carcinoma cells via Cyclin F and RRM2 [J]. *Anticancer Agents Med Chem*, 2016, 16(4): 440-446.
- [14] LIU L M, XIONG D O, LIN P, et al. DNA topoisomerase 1 and 2A function as oncogenes in liver cancer and may be direct targets of nitidine chloride [J]. *Int J Oncol*, 2018 53(5): 1897-1912.
- [15] CHEN Y C, CHEN I S, HUANG G J, et al. Targeting DTL induces cell cycle arrest and senescence and suppresses cell growth and colony formation through TPX2 inhibition in human hepatocellular carcinoma cells [J]. *Oncol Targets Ther*, 2018, 21(11): 1601-1616.
- [16] WANG J, ZHANG Y S, THAKUR K, et al. Licochalcone A from licorice root, an inhibitor of human hepatoma cell growth via induction of cell apoptosis and cell cycle arrest [J]. *Food Chem Toxicol*, 2018, 10(120): 407-417.

收稿日期: 2018-11-01

(本文责编: 曹粤锋)