

HCV 相关肝硬化关键基因的交互分析

张立洁¹, 彭 泉², 芦东徽¹, 张明金²

(解放军第901医院肿瘤中心¹, 普外科², 安徽 合肥 230031)

摘要:目的 使用生物信息学方法挖掘丙型肝炎病毒(HCV)相关肝硬化(HCV-Cirr)的关键基因并探索其作用机制。方法 通过分析数据集 GSE14323 和 GSE40744, 筛选在 HCV-Cirr 形成过程中存在表达差异的基因(DEGs)和 miRNA(DEMs), 使用 miRnet 数据库预测 DEMs 的靶基因并与 DEGs 取交集得到候选基因, 使用 NetworkAnalyst 数据库进行基因的富集分析, 使用 STRING 数据库和 Cytoscape 软件构建蛋白交互网络(PPI)和模块分析。结果 共得到 66 个候选基因; GO 和 KEGG 通路富集分析显示, 候选基因主要富集在对各种因素刺激的反应、细胞外区域、脂质结合, 并主要参与了 MAPK 信号通路、粘着斑通路; 基于 PPI 网络得到 10 个关键基因 (MMP2、CCL2、JUN、COL1A1、CAV1、TIMP3、SPARC、COL4A2、FBN1、FOS), 均为在 HCV-Cirr 中表达上调 DEGs, 其中 CAV1、SPARC、FBN1、FOS 为 HCV-Cirr 中表达下调的 hsa-miR-192-5p 的靶基因。结论 最终筛选出 4 个核心基因及其上游的调控 miRNA, 这些基因可能用于 HCV-Cirr 的早期诊断和治疗。

关键词: 丙型肝炎; 肝硬化; HCV 相关肝硬化; 微小 RNA

中图分类号: R512.63

文献标识码: A

DOI: 10.3969/j.issn.1006-1959.2023.11.002

文章编号: 1006-1959(2023)11-0010-06

Interaction Analysis of Key Genes in HCV-related Liver Cirrhosis

ZHANG Li-jie¹, PENG Quan², LU Dong-hui¹, ZHANG Ming-jin²

(Tumor Diagnosis and Treatment Center¹, General Surgery Department², No.901 Hospital of the PLA Joint Logistics Support Force, Hefei 230031, Anhui, China)

Abstract: **Objective** To explore the key genes of HCV-related cirrhosis (HCV-Cirr) and its mechanism by bioinformatics. **Methods** By analyzing the data sets GSE14323 and GSE40744, differentially expressed genes (DEGs) and miRNAs (DEMs) were screened during the formation of HCV-Cirr. The miRnet database was used to predict the target genes of DEMs and intersect with DEGs to obtain candidate genes. The NetworkAnalyst database was used for gene enrichment analysis. The STRING database and Cytoscape software were used to construct protein interaction network (PPI) and module analysis. **Results** A total of 66 candidate genes were obtained; GO and KEGG pathway enrichment analysis showed that candidate genes were mainly enriched in response to various factors, extracellular regions, lipid binding, and mainly involved in MAPK signaling pathway and focal adhesion pathway. Based on the PPI network, 10 key genes (MMP2, CCL2, JUN, COL1A1, CAV1, TIMP3, SPARC, COL4A2, FBN1, FOS) were obtained, all of which were up-regulated DEGs in HCV-Cirr, while CAV1, SPARC, FBN1 and FOS were the target genes of hsa-miR-192-5p down-regulated in HCV-Cirr. **Conclusion** Four core genes and their upstream regulatory miRNAs are finally screened out, which may be used for early diagnosis and treatment of HCV-Cirr.

Key words: Hepatitis C; Liver cirrhosis; HCV-related cirrhosis; microRNA

据世界卫生组织统计,全球范围内丙型肝炎病毒(hepatitis C virus, HCV)的感染率约为3%,其中有30%的患者可能在30年内发展为肝硬化(HCV-Cirr),而HCV-Cirr出现后,HCV相关肝癌(HCV-related hepatocellular carcinoma, HCV-HCC)的发生率会以每年4%~7%的速度上升,因此对HCV-Cirr患者的早期诊断和治疗是非常重要的^[1,2]。HCV感染肝组织后,可导致宿主基因组的表达变化和基因的突

变,从而影响细胞的信号转导、氧化应激、增殖凋亡等生物学过程,使得肝脏组织出现代谢失调、纤维化及异常增殖等病理现象。在HCV-Cirr的发生、进展过程中会有多基因参与,其中一些基因的表达改变,会在疾病的进程中发挥决定性的作用。由于技术及研究成本的限制,既往难以进行大规模的基因表达分析,而随着测序技术、基因芯片技术和大数据分析技术的发展^[3],对发现HCV-Cirr关键基因和这些基因的调控机制提供了保障。本研究旨在通过生物信息学的方法,筛选出在HCV-Cirr与正常肝脏组织之间存在差异表达的关键miRNA和基因,为鉴定HCV-Cirr发生、发展过程中的关键生物标志物和潜在的调控机制研究提供依据。

1 材料与方法

1.1 数据来源 从GEO数据库查询包含正常肝脏组

基金项目: 中国人民解放军联勤保障部队第901医院院管课题(编号:2021YGYB06)

作者简介: 张立洁(1983.5-),女,辽宁辽阳人,硕士,副主任医师,主要从事肿瘤的早期诊断及综合治疗

通讯作者: 彭泉(1983.12-),男,安徽合肥人,硕士,副主任医师,主要从事表观遗传学及生物信息分析应用研究

组织和HCV相关肝硬化组织的基因表达谱和miRNA表达谱原始数据集。为保证数据集的可信度,芯片表达数据需同时包含HCV-Cirr样本和正常肝脏组织样本,总样本量 ≥ 20 且各分组需 ≥ 5 。最终选择GSE14323和GSE40744数据集。GSE14323基于GPL571平台,包括41个HCV相关肝硬化组织样品,19个正常肝脏组织样品。GSE40744基于GPL14613平台,包括18个HCV样品和7个正常肝脏组织样本。

1.2 差异miRNA、差异基因分析 将数据集中的样本分为两组(HCV-Cirr组 vs.正常肝脏组织)进行差异表达分析,使用GEO数据库自带GEO2R分析软件,鉴定出在两组间存在差异表达的基因(DEGs)和miRNA(DEMs)。本研究中将探针与基因关系为一对多或多对一的情况予以删除,将 $FDR < 0.01$, $|\log_2 FC|$ 绝对值 > 1.5 作为筛选DEMs或DEGs标准,并保留TXT格式的结果用于后续分析。

1.3 候选基因的筛选 miRnet数据库(<https://www.mirnet.ca/miRNet/home.xhtml>)可以提供miRNA和其靶基因的查询,其数据来源于已被实验验证的miRNA和下游靶基因关系对^[4]。在本研究中,将GSE40744数据集分析得到的DEMs在miRnet数据库中进行靶基因预测,预测采用以下标准:Organism:H.sapiens;Tissue:liver;Target:genes,将得到下游基因与GSE14323数据集分析得到的DEGs取交集,得到的基因作为候选基因进行进一步研究。

1.4 候选基因的富集分析 GO分析用于注释基因,基因产物和潜在生物现象序列的共同方法;KEGG

通路分析用于基因组序列和其他高通量数据的生物学解释的整合数据库资源。在本研究中,使用NetworkAnalyst在线软件工具(<https://www.networkanalyst.ca/>)对候选基因进行GO和KEGG通路分析^[5],其中GO分析包括以下3个方面:生物过程(biological process,BP)、细胞组分(cellular component,CC)和分子功能(molecular function,MF),以 $P < 0.01$,基因count > 2 为选择标准。

1.5 PPI网络构建和模块分析 STRING数据库(<https://string-db.org/cgi/input.pl>)作为目前应用最为广泛的蛋白质相互作用检索工具,包含了1100多个物种的5200多万蛋白质,可使用多种计算工具构建PPI互作关系^[6]。本研究中为进一步在候选基因中获得关键基因,使用STRING在线数据库构建PPI网络,并使用Cytoscape软件(版本号3.6.1)将其进行可视化。而在PPI网络中,具有更多分度的节点被认为是中枢基因,并被作为是具有重要生理调节作用的核心蛋白或关键蛋白,使用Cytoscape软件中的CytoHubba插件在PPI网络中进一步筛选关键基因。

2 结果

2.1 DEM和DEG鉴定 按上述差异基因筛选标准,在GSE14323中得到288个DEGs,其中248个表达上调,40个表达下调;在GSE40744中得到30个DEMs,其中24个表达上调,6个表达下调,通过查询miRBase数据库(<http://www.mirbase.org/>)获得DEMs成熟体的官方名称。基于表达情况,构建差异分析的火山图,见图1。

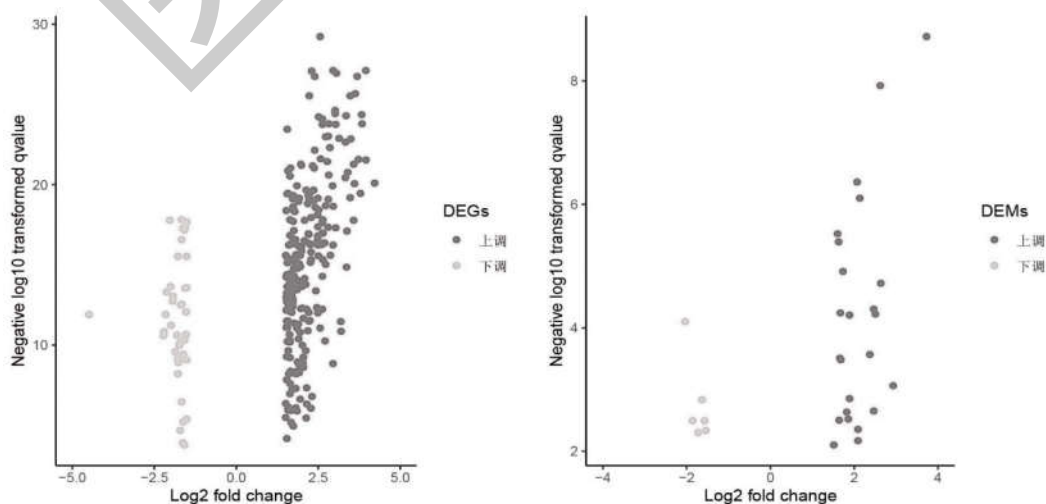


图1 数据集中存在差异表达的基因

2.2 候选基因筛选 按照标准, 共得到 9 个 DEMs (hsa-miR-192-5p, hsa-miR-193a-3p, hsa-miR-34a-5p, hsa-miR-212-3p, hsa-miR-21-5p, hsa-miR-223-3p, hsa-miR-29b-3p, hsa-miR-155-5p, hsa-miR-31-5p) 下游的 3323 个靶基因, 其中只有 hsa-miR-192-5p 为表达下调的 DEM, 其余均为表达上调的 DEMs。将预测的 DEMs 的靶基因与差异分析得到的 DEGs 做 Venn 图, 得到共有的 66 个基因作为候选基因进行下一步分析, 见图 2。

2.3 富集分析结果 使用 NetworkAnalyst 在线软件工具对上述得到的 66 个候选基因进行富集分析, 在 GO-BP 中, 候选基因主要富集在对各种因素刺激的反应、血管发育和形态结构的形成; 在 GO-CC 中, 基因主要富集在细胞外区域、细胞表面、内质网腔和基底膜; 在 GO-MF 中, 基因富集在脂质结合、G_蛋白偶联受体结合并参与结构分子活性和细胞因子活

性。KEGG 通路分析显示候选基因主要参与了 MAPK、TNF、IL-17、雌激素信号通路传导、粘着斑、动脉粥样硬化及糖类代谢等, 见图 3。

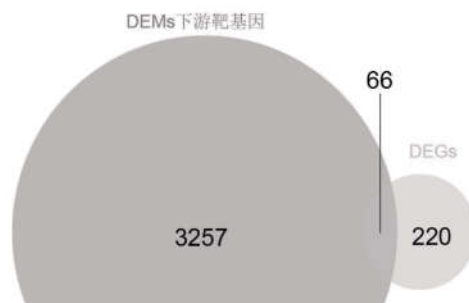
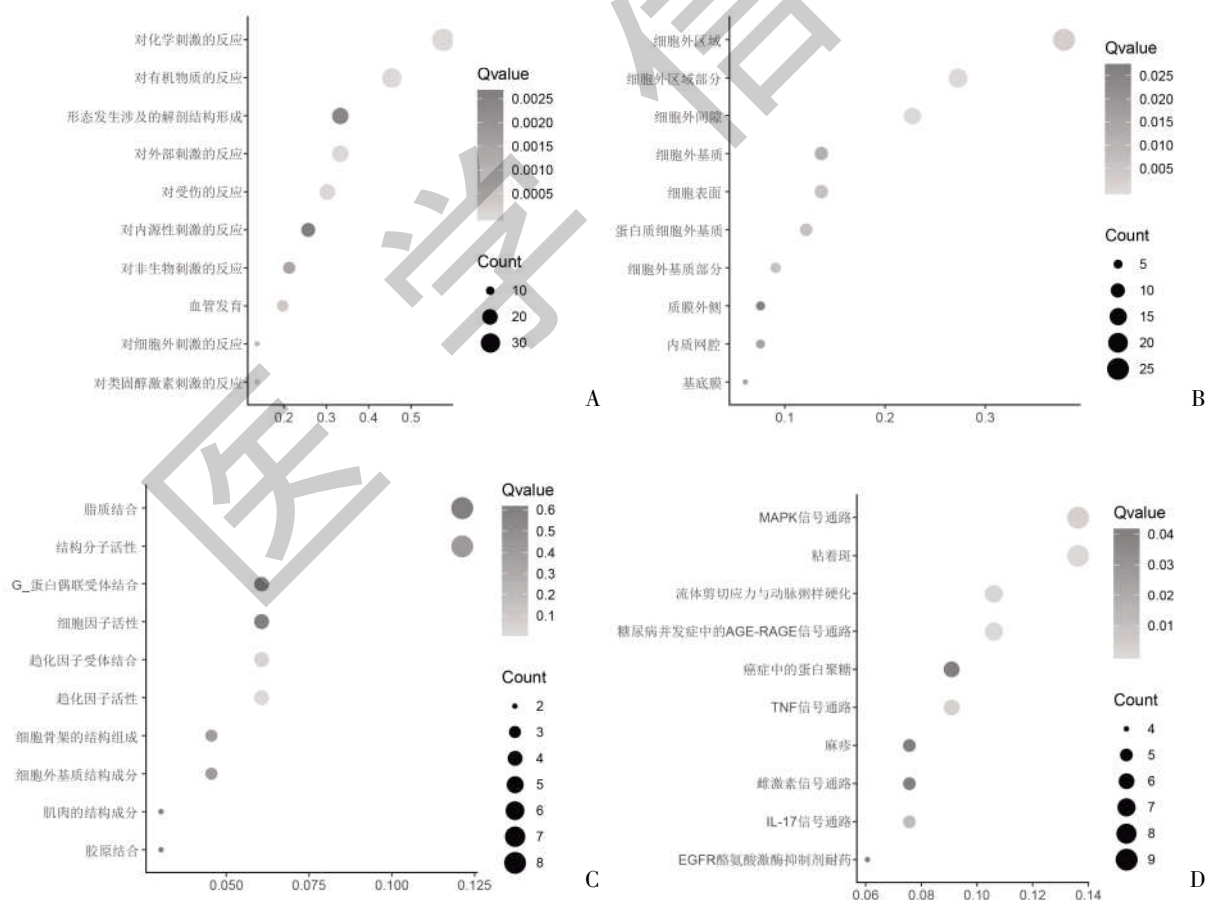


图 2 DEMs 的靶基因和 DEGs 的韦恩图

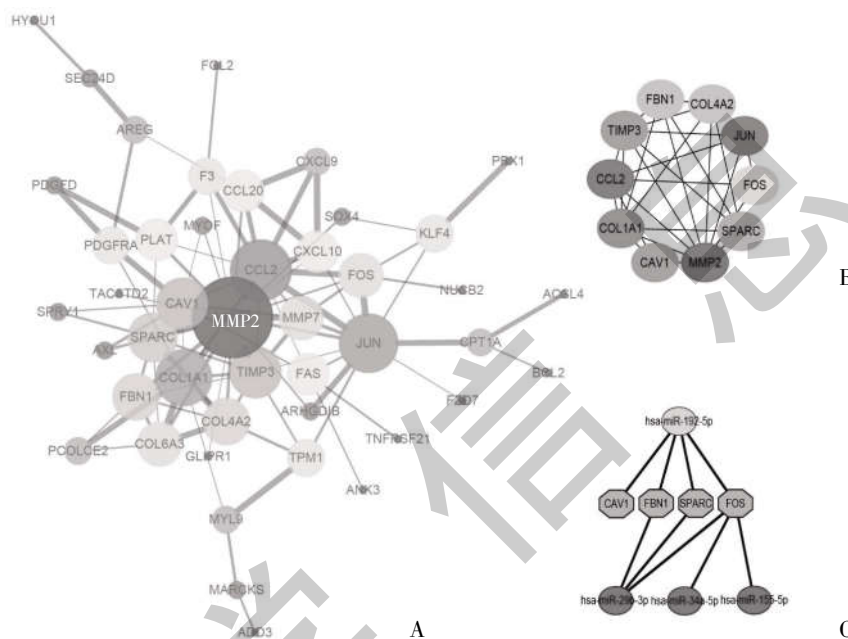


注: A: GO-BP; B: GO-CC; C: GO-MF; D: KEGG 通路分析

图 3 候选基因富集分析

2.4 PPI 网络构建和枢纽基因鉴定 使用 STRING 进行 PPI 网络构建,选择交互分数 ≥ 0.4 ,得到包含 66 个节点,99 个分度的蛋白互作网络,去除单个基因和不在主网络的基因后采用 Cytoscape 软件进行可视化处理,使用 CytoHubba 插件基于分度进行核心基因选择,最终得到 10 个的枢纽基因: MMP2、CCL2、JUN、COL1A1、CAV1、TIMP3、SPARC、COL4A2、FBN1、FOS,见图 4。进一步分析发现,这

10 个枢纽基因在 HCV-Cirr 中均是表达上调基因,其中仅有 CAV1、SPARC、FBN1、FOS 为 HCV-Cirr 中表达下调的 hsa-miR-192-5p 的靶基因。通过对枢纽基因的 GO 富集分析发现,这些基因主要参与细胞对有机物和内分泌性刺激的反应、分子功能的调节、胞外结构等,KEGG 通路分析显示枢纽基因参与糖尿病并发症中的 AGE-RAGE 信号通路、流体剪切应力与动脉粥样硬化以及松弛素信号通路。



注:A:候选基因 PPI 网络,节点越大代表分度越大,两点间连线越粗代表交互分数越高;B: CytoHubba 插件所选择出的枢纽基因;C: 核心基因的调控网络

图 4 PPI 网络和核心基因

3 讨论

慢性 HCV 感染是 HCC 发生的主要因素之一,在 HCV 感染到 HCC 发生的过程中,比较重要的一个阶段是 HCV-Cirr 的形成。因此,研究 HCV-Cirr 的发生、发展的分子机制具有重要意义^[7,8]。在本研究中,采用生物信息学交互分析的方法鉴定出 HCV-Cirr 的候选生物标志物和潜在的治疗靶点,将分析得出的 DEMs 在经实验验证的数据库中查询肝脏组织特异性的靶基因,并与分析得出的 DEGs 取交集,得出 66 个可能在 HCV-Cirr 中发挥重要生物学作用的基因靶点。

GO 富集分析表明候选基因参与了对各类刺激的反应,细胞外结构的组成和相关受体的结合,这些基因可能在 HCV 感染后,肝脏细胞在长期炎症刺激下发生细胞变性,结构改变到出现肝脏纤维化的过程中发挥重要的生理作用。KEGG 通路分析提示候

选基因主要参与 MAPK 信号通路和粘着斑通路,其中 MAPK 信号在多种疾病中发挥重要作用,而在肝硬化的疾病进展中,该通路也发挥了重要作用。MAPK 通路不仅是肝纤维化形成的信号调节通路,而且也可能参与了细胞的自噬,这表明 MAPK 通路还可作为抗纤维化治疗的潜在靶点^[9]。在大鼠模型的研究表明,大鼠肝纤维化模型中,MAPK 通路激活,经阿魏酸或和厚朴酚治疗后,MAPK 通路相关基因表达下调,信号转导受到抑制,起到了抗肝脏纤维化的作用^[9,10]。此外,MAPK 通路激活在肝癌的发生,发展中也有重要的生物学作用,其靶向药物有望作为肝癌治疗的新方向^[11]。

通过构建 PPI 网络,本研究确定了 10 个枢纽基因,均为在 HCV-Cirr 中表达上调的基因,分别为 4 个上调 DEMs (hsa-miR-34a-5p, hsa-miR-21-5p, hsa-miR-29b-3p, hsa-miR-155-5p) 和 1 个下调

DEMs (hsa-miR-192-5p) 的靶基因。其中 CAV1、SPARC、FBN1、FOS 为 HCV-Cirr 中表达下调的 hsa-miR-192-5p 的下游基因, 同时 SPARC、FBN1 也是 hsa-miR-29b-3p 的靶基因, FOS 也是 hsa-miR-155-5p, hsa-miR-29b-和 hsa-miR-34a-5p 的靶基因。CAV1 是质膜中小窝的结构蛋白, 目前已被认为是肝脏功能的调节剂, CAV1 可以调节肝脏内的脂类和糖类代谢, 参与了肝细胞的增殖过程与肝脏星状细胞的活化过程, 与胆汁淤积, 肝炎, 肝硬化和肝癌的发生有关^[12,13], 而且有研究表明在表达 CAV1 的肝硬化组织更容易发生肝癌, 因此可以作为包括肝硬化和肝癌在内的多种肝脏疾病的生物标志物和治疗靶点^[14]。SPARC 编码的蛋白参与了细胞外基质的合成并促进细胞形态的改变, 与肿瘤细胞的侵袭转移相关。目前, SPARC 在肝硬化中研究较少, 少量的研究表明 SPARC 及其家族蛋白参与了肝脏细胞的脂肪变性和纤维化^[15], 动物实验还表明 SPARC 可通过细胞外基质的沉积和由此产生的疤痕化在肝硬化的发病机理和进展中发挥作用^[16]。基于本研究分析所得, SPARC 在 HCV-Cirr 中的作用机制值得进一步研究。FBN1 可编码细胞外基质成分原纤维蛋白, 这些蛋白可在人体的弹性及非弹性结缔组织中提供结构支撑作用, 在人体和大鼠模型的研究中都表明了 FBN1 在肝脏纤维化的过程中起到了促进作用^[17,18], 通过阻止 FBN1 编码蛋白的活化, 可以抑制肝脏纤维化, 作为治疗肝硬化的潜在疗法^[19]。FOS 基因家族由 4 个成员组成: FOS、FOSB、FOSL1 和 FOSL2。FOS 编码蛋白已被认为是细胞增殖, 分化和转化的调节剂, 某些情况下 FOS 基因的表达也与细胞凋亡有关。但在肝硬化疾病中, FOS 基因的研究较少, 有报道称 FOS 可通过参与竞争性内源性 RNA 相互作用网络 (ceRNA) 在肝硬化的免疫反应中发挥调控作用^[20]。此外, FOS 家族成员还参与了 HCC 的发生^[21,22]。总之, 本研究分析得出的核心基因在肝硬化中的作用与文献报道相符, 但目前为止在 HCV 相关肝硬化中的研究仍十分罕见, 其与上游 miRNA 之间的调控关系也鲜有报道^[23], 因此值得进一步深入研究。

肝脏组织受到 HCV 感染, 进而出现 HCV-Cirr 到最终发生 HCV-HCC, 其过程涉及基因组的表达失调, 信号通路的激活或抑制以及多种生物学过程, 整合的生物信息学分析有助于研究和理解其分子机制。通过交互分析, 本研究确定了一些可能有助于预测 HCV-Cirr 发生和发展的生物标志物, 此外还

确立了一些有望用于治疗 HCV-Cirr 的信号通路和机制。枢纽基因与上游 miRNA 之间的调节关联可能为 HCV-Cirr 的诊断和治疗提供新的研究方向。但值得注意的是, 基因在 miRNA 调控下的表达受到多种因素影响, 本研究也发现在 HCV-Cirr 中表达上调的多个枢纽基因同时受到上调和下调 DEMs 的调控, 因此需要开展进一步的实验, 将分析得出的 miRNA 和基因在组织样本中进行表达检测, 并随后确认它们在 HCV-Cirr 诊断和分子治疗中的实用性。

参考文献:

- [1]王文鑫,刘红虹,毕京峰,等.慢性 HCV 感染 DAA 应答特点和长期预后观察研究[J].传染病信息,2021,34(3):208-212.
- [2]施新焕,唐睿珠,刘子杰,等.云南地区丙型肝炎病毒基因亚型及 RNA 载量在丙型肝炎进展中的变化[J].昆明医科大学学报,2018,39(6):57-61.
- [3]黄成,易尚辉,查文婷,等.基于生物信息学分析筛选舌鳞状细胞癌核心基因及其预后价值[J].医学信息,2020,33(3):6-12.
- [4]Fan Y,Xia J.miRNet-Functional Analysis and Visual Exploration of miRNA-Target Interactions in a Network Context[J].Methods in Molecular Biology,2018,1819:215-233.
- [5]Zhou G,Soufan O,Ewald J,et al.NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis [J].Nucleic Acids Research,2019,47 (W1): W234-W241.
- [6]Szkarczyk D,Gable AL,Lyon D,et al.STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets [J].Nucleic Acids Research,2019,47(D1):D607-D613.
- [7]任珊,徐鹏,张莹,等.血清 AFP 水平和组织 GPC3、VEGF 蛋白表达在丙肝相关肝癌中的意义 [J].广东医学,2018,39(19): 2890-2894.
- [8]Xu B,Chen B,Ma ZH,et al.Application of anti-HCV and HCV RNA detection in intravenous drug users[J].Chinese Journal of Hepatology,2021,29(5):415-420.
- [9]王媛媛,朱浩,史玉欢,等.阿魏酸通过抑制 MAPK 和 NF- κ B/I κ B α 信号通路减轻大鼠肝纤维化的机制研究[J].中草药,2021,19(12):2489-2495.
- [10]符健,高艺,林锋,等.和厚朴酚对大鼠肝纤维化及 P38 MAPK/Nrf2 信号通路的影响 [J].中西医结合肝病杂志,2021,31(3):246-249.
- [11]龙雪,邱婉婷,曹世杰,等.补骨脂酚通过 MAPK 途径诱导人肝癌 HepG2 细胞凋亡的研究[J].现代肿瘤医学,2020,28(5): 729-735.
- [12]刘振华,李小宁,陈平圣.Cav-1 在肝癌组织的表达及其意义[J].东南大学学报(医学版),2018,37(1):35-40.
- [13]Ilha M,Moraes KDS,Rohden F.Exogenous expression of

caveolin-1 is sufficient for hepatic stellate cell activation[J]. J Cell Biochem, 2019, 120(11): 19031-19043.

[14] Yim SY, Hae NJ, Shin JH, et al. Identification of prognostic biomarker in predicting hepatocarcinogenesis from cirrhotic liver using protein and gene signatures [J]. Exp Mol Pathol, 2019, 111: 104319.

[15] Athwal VS, Pritchett J. SOX9 regulated matrix proteins are increased in patients serum and correlate with severity of liver fibrosis[J]. Sci Rep, 2018, 8(1): 17905.

[16] Yuting Y, Lifeng F, Qiwei H. Secreted modular calcium-binding protein 2 promotes high fat diet (HFD)-induced hepatic steatosis through enhancing lipid deposition, fibrosis and inflammation via targeting TGF- β 1 [J]. Biochem Biophys Res Commun, 2019, 509(1): 48-55.

[17] Lou Y, Tian GY, Song Y, et al. Characterization of transcriptional modules related to fibrosing-NAFLD progression [J]. Sci Rep, 2017, 7(1): 4748.

[18] Tiedemann K, Muthu ML, Reinhardt DP, et al. Male Marfan mice are predisposed to high-fat diet-induced obesity, diabetes,

and fatty liver [J]. Am J Physiol Cell Physiol, 2022, 323(2): C354-C366.

[19] Eser Karlidag G, Arslan Solmaz O. Are adropin, apelin, elabela, asprosin and betatrophin biomarkers for chronic hepatitis and staging of fibrosis? [J]. Biotech Histochem, 2020, 95(2): 152-159.

[20] Gao B, Zhang X, Huang Y, et al. Coding and non-coding gene regulatory networks underlie the immune response in liver cirrhosis [J]. PLoS One, 2017, 12(3): e0174142.

[21] 李文宏, 曾宪成, 王捷. miRNA-135a 在肝癌组织中的表达及临床意义 [J]. 岭南现代临床外科, 2017, 17(3): 261-266.

[22] 王挺帅, 王娜, 张荣臻, 等. 基于微小 RNA-mRNA-信号通路调控网络探讨肝细胞癌的潜在生物学标志物 [J]. 广西医学, 2021, 43(14): 1723-1729.

[23] Pascut D, Hoang M, Nguyen N, et al. HCV Proteins Modulate the Host Cell miRNA Expression Contributing to Hepatitis C Pathogenesis and Hepatocellular Carcinoma Development [J]. Cancers, 2021, 13(10): 2485.

收稿日期: 2022-08-12; 修回日期: 2022-09-06

编辑/成森