

研究报告

Research Report

基于生物信息学分析的肝癌核心基因的筛选及验证

李怡敏^{1,2} 汪洋^{1,2} 袁琳³ Nur Fazleen Binti Idris^{1,2} 黄国旺^{1,2} 涂增^{1,2*}

1 重庆医科大学基础医学院病原生物学教研室, 重庆, 400016; 2 重庆医科大学分子医学与肿瘤研究中心, 重庆, 400016; 3 西安市第一医院, 西安, 710000

* 通信作者, tuz1980@126.com

摘要 本研究基于 GEO 数据库, 选取由慢性乙型肝炎诱导的肝细胞癌芯片数据 GSE121248 为研究对象, 利用 GEO2R 软件分析数据, 筛选出表达具有差异的基因, 利用 DAVID 数据库进行 GO 分析和 KEGG pathway 富集分析。利用 STRING 数据库构建 PPI 网络, 分析筛选核心基因。利用 GEPIA 对核心基因的表达进行验证, Kaplan Meier Plotter 在线分析工具对核心基因与患者生存情况的相关性进行验证。通过上述方法筛选出 309 个 DEGs, 其中上调基因 94 个, 下调基因 215 个。差异基因功能分析显示上调的 DEGs 主要参与细胞周期和卵母细胞减数分裂通路等途径, 下调的 DEGs 则在补体和凝血级联、代谢途径以及咖啡因代谢途径富集。筛选出 15 个具有高度关联性的核心基因(BUB1, BUB1B, BIRC5, CCNB1, CCNB2, CDC20, CDK1, KIF20A, MAD2L1, NCAPG, ZWINT, PBK, BTL, TTK 和 NUSAP1), 它们与肝癌患者的总体生存率明显具有相关性, 并为其构建了 miRNA 调控网络。我们的研究通过生物信息学方法有效分析了肝细胞癌发生、发展相关的差异表达基因, 筛选出 15 个核心基因, 分析其生物学相关功能, 以期探索肝细胞癌发病机制, 并为临床诊断标志物的改进以及筛选提供一定的理论基础。

关键词 肝癌, 差异基因表达, 生物信息学

Identification of Hub Genes in Hepatocellular Carcinoma Based on Bioinformatics Analysis

Li Yimin^{1,2} Wang Yang^{1,2} Yuan Lin³ Nur Fazleen Binti Idris^{1,2} Huang Guowang^{1,2} Tu Zeng^{1,2*}

1 Department of Pathogen Biology, College of Basic Medical Sciences, Chongqing Medical University, Chongqing, 400016; 2 Molecular Medicine and Tumor Research Center, College of Basic Medical Sciences, Chongqing Medical University, Chongqing, 400016; 3 Xi'an NO.1 Hospital, Xi'an, 710000

* Corresponding author, tuz1980@126.com

DOI: 10.5376/gmb.cn.2020.10.0007

Abstract Our study is based on the GEO database, the hepatocellular carcinoma chip profile GSE121248 induced by chronic hepatitis B was selected as the study subject. The differential expressed genes were screened by using GEO2R tools. GO analysis and KEGG pathway analysis were performed for differential expressed genes by using DAVID. Then, PPI network was constructed to picked out hub genes. Finally, the expression of hub genes and the patient survival were verified by the GEPIA and Kaplan Meier Plotter analysis tools. Through the above method, a total of 309 DEGs were screened, consisting of 94 up-regulated genes and 215 down-regulated genes. Functional analysis of DEGs showed that up-regulated DEGs are mainly involved in the cell cycle and oocyte

本文首次发表在 International Journal of Molecular Medical Science 上, 现依据版权所有人授权的许可协议, 采用 Creative Commons Attribution License, 协议对其进行授权, 再次发表与传播

收稿日期: 2020 年 2 月 25 日; 接受日期: 2020 年 5 月 6 日; 发表日期: 2020 年 5 月 18 日

引用格式: 李怡敏, 汪洋, 袁琳, Nur Fazleen Binti Idris, 黄国旺, 涂增, 2020, 基于生物信息学分析的肝癌核心基因的筛选及验证, 基因组学与医学生物学, 10(1): 1-10 (doi: 10.5376/gmb.cn.2020.10.0007) (Li Y.M., Wang Y., Yuan L., Nur Fazleen B.I., Huang G.W., and Tu Z., 2020, Identification of Hub genes in hepatocellular carcinoma based on bioinformatics analysis, Jiyinzhexue Yu Yixue Shengwuxue (Genomics and Medical Biology), 10(1): 1-10 (doi: 10.5376/gmb.cn.2020.10.0007))

meiosis pathways, down-regulated DEGs are enriched in the complement and coagulation cascades, metabolic pathways, and caffeine metabolic pathways. Moreover, fifteen hub genes with high correlation were screened, including BUB1, BUB1B, BIRC5, CCNB1, CCNB2, CDC20, CDK1, KIF20A, MAD2L1, NCAPG, ZWINT, PBK, DTL, TTK and NUSAP1. They were found to be associated with the overall survival in patients with HCC and constructed a miRNA regulatory network. Bioinformatics can effectively analyze the DEGs associated with the occurrence and development of HCC. The identification of the 15 Hub genes can provide theoretical guidance for further research on the molecular mechanism and molecular marker screening of HCC.

Keywords Hepatocellular carcinoma, Differential expressed genes, Bioinformatics

肝细胞癌(hepatocellular carcinoma, HCC)死亡率居世界第四,严重损害人类的生命健康(Segal et al., 2018)。中国作为肝病大国,肝癌发病率位居第四,死亡率居第二(孙可欣等, 2019)。甲胎蛋白(alpha-feto-protein, AFP)作为全世界目前应用最为广泛的肝癌检测标志物,但其作为肝癌标志物仍存在较大争议(Mohammed and Roberts, 2017)。其他肝癌标志物,例如甲胎蛋白异质体(fucosylated fraction of AFP, AFP-L3)和异常凝血酶(des-gamma-carboxy prothrombin, DCP)等也被报道在HCC诊断中具有一定的局限性(Han, 2012)。因此,了解肝癌的发病机制,寻找更为高效敏感的诊断标志物及治疗新靶点尤为重要。基于生物信息技术飞速发展,愈来愈多的科研工作者通过生物信息分析的方法来挖掘在肿瘤发生发展中发挥重要作用的分子及通路,从而提高对癌症发病机制的认识(Tinker et al., 2006)。本研究通过GEO (Gene Expression Omnibus)数据库基于肝癌组织和癌旁组织数据发掘其表达存在差异的基因,利用多种生物信息学分析方法对调控肝癌的核心基因进行筛选及分析,以期更进一步的理解肝细胞癌发生发展的机制并为其提供一定的理论基础。

1 结果与分析

1.1 差异表达基因的筛选

利用GEO2R分析肝细胞癌及癌旁样本后,共筛选出309个DEGs,其中在肝癌组织中表达上调的有94个,下调的有215个(图1)。

1.2 DEGs GO和KEGG结果分析

我们使用DAVID分析了差异基因的GO和KEGG pathway富集。GO富集分析表明:上调的DEGs显著富集于细胞分裂、有丝分裂、姐妹染色单体粘连、细胞周期调控等生物过程,参与细胞核、细胞质、中间体、浓缩的染色体着丝粒等细胞成分。上调的DEGs并未在分子功能中发现显著富集(图2)。

下调的DEGs则在氧化还原过程、免疫过程、受体介导的内吞作用、补体的激活、异生素代谢过程及环氧化酶P450途径等生物过程,参与胞外体、胞外区、细胞外间隙等细胞成分,分子功能主要集中在铁离子结合、血红素结合、丝氨酸型内肽酶活性、氧化还原

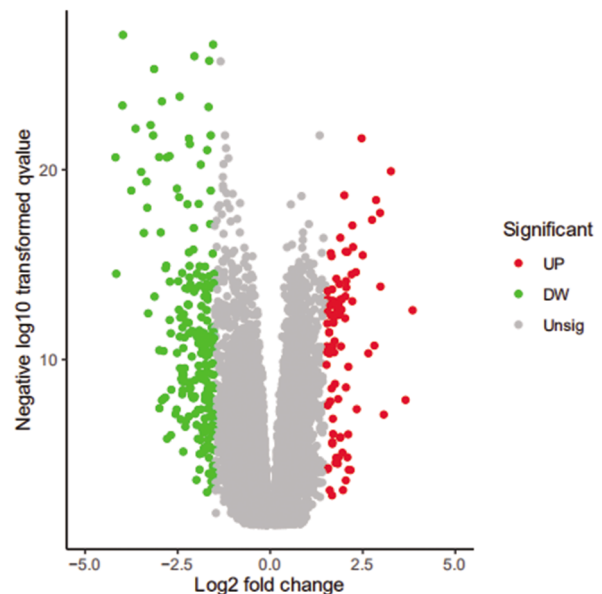


图1 肝癌差异表达基因分析

Figure 1 Differential expressed genes analysis of HCC

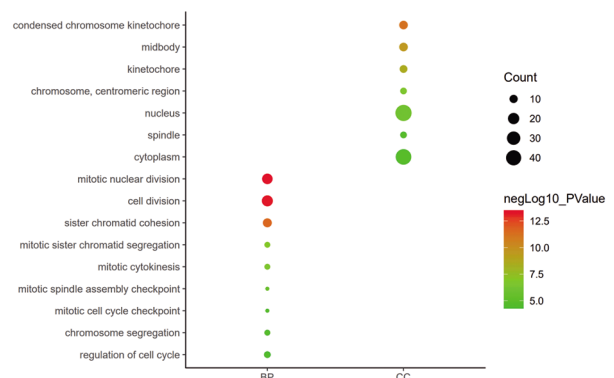


图2 与肝癌相关的上调差异表达基因的基因本体论分析

Figure 2 Gene ontology analysis of up-regulated differentially expressed genes associated with HCC

酶活性及氧结合等(图 3)。

KEGG 结果分析表明：上调的 DEGs 显著富集于细胞周期和卵母细胞减数分裂通路，下调的 DEGs 则在补体和凝血级联、代谢途径以及咖啡因代谢途径富集(图 4)。

1.3 PPI 网络构建和功能模块分析

基于 STRING 数据库，我们将 DEGs 的 PPI 网络利用 Cytoscape 进行可视化。DEGs 的 PPI 网络由 268 个节点蛋白和 1671 条相互作用关系构成。然后，我们通过 MCODE 插件对 DEGs 的 PPI 网络进行聚类分析，获得了 9 个不同的 PPI 功能模块，其中有 6 个主要模块(图 5)。我们利用 DAVID 数据库对这 6 个模块中的基因进行了 KEGG pathway 富集分析(表 1)。分析结果表明模块 1 中的基因主要富集于细胞周期、卵母细胞减数分裂过程和黄体酮介导的卵母细胞成熟等途径，模块 2 中所包含的基因参与补体和凝血级联途径，模块 3 的基因则主要富集于药物代谢 - 细胞色素 P450、化学致癌作用、视黄醇代谢和

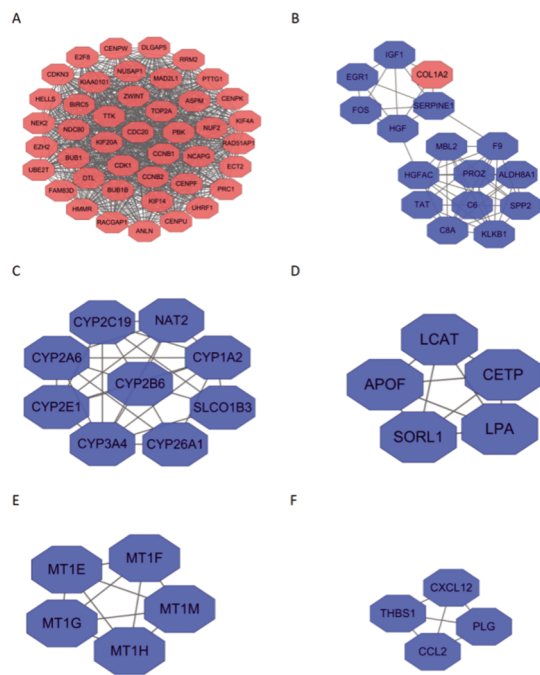


图 5 蛋白质 - 蛋白质相互作用网络的前 6 个模块

注：A: 模块 1; B: 模块 2; C: 模块 3; D: 模块 4; E: 模块 5; F: 模块 6

Figure 5 Top 6 modules from the protein-protein interaction network

Note: A: Modules 1; B: Modules 2; C: Modules 3; D: Modules 4; E: Modules 5; F: Modules 6

细胞色素 P450 对异生素的代谢等途径，模块 5 的基因富集于矿物质吸收途径，模块 4 与模块 6 未能发现富集于在 KEGG pathway 分析中具有统计学意义的信号通路。

1.4 核心基因分析

基于 DEGs 的 PPI 网络，我们利用 cytoHubba 插件的 MCC 算法，获得了 15 个关联度最高的基因作为核心基因，为它们构建蛋白质互作网络(图 6)。并用 GEPIA 数据库进一步验证了这 15 个核心基因在肝癌组织(369 例)与正常组织(160 例)中的表达情况。非 SMC 缩合蛋白 I 复合物亚基 G(non-structural maintenance of chromosomes condensin I complex subunit G, NCAPG)、驱动蛋白家庭成员 20A (kinesin family member 20A, KIF20A)、细胞周期蛋白依赖性激酶 1(cyclin-dependent kinase 1, CDK1)、杆状病毒 I-AP 重复序列 (baculoviral IAP repeat containing 5, BIRC5)、细胞周期蛋白 B1(cyclin B1, CCNB1)、细胞分裂周期蛋白 20 (cell division cycle 20, CDC20)、细胞周期蛋白 B2(cyclin B2, CCNB2)、有丝分裂阻滞缺陷 2 样蛋白 1 (MAD2 mitotic arrest deficient-like 1,

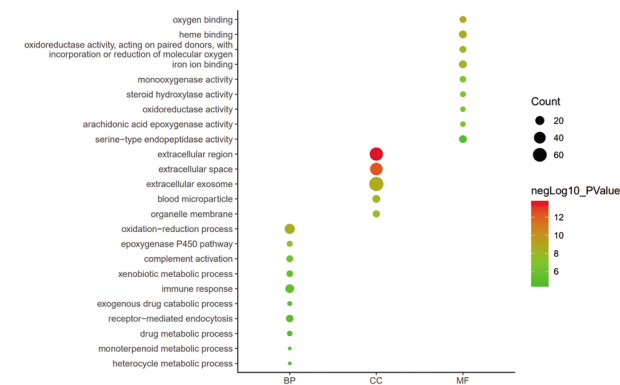


图 3 与肝癌相关的下调差异表达基因的基因本体论分析

Figure 3 Gene ontology analysis of down-regulated differentially expressed genes associated with HCC

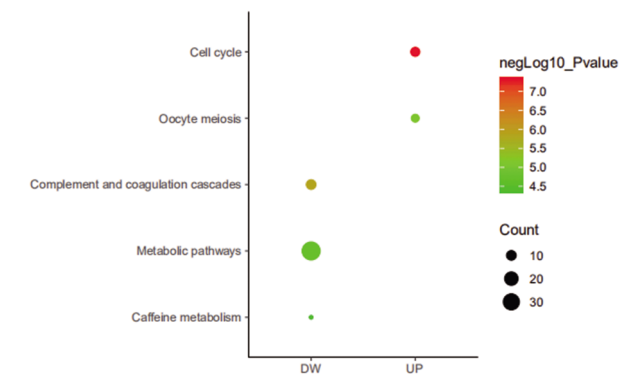


图 4 与肝癌相关的差异表达基因的 KEGG 通路分析

Figure 4 KEGG pathway of differentially expressed genes associated with HCC

表 1 前 6 个模块的 KEGG 通路分析

Table1 KEGG pathway analysis of Top 6 modules

组别	通路名	聚集数	P 值	伪发现率
Sample Group	Term	Count	P value	FDR
模块 1	hsa04110: 细胞周期	9	1.06E-11	8.19E-09
Module1	hsa04110: Cell cycle			
	hsa04114: 卵母细胞减数分裂过程	7	2.41E-08	1.86E-05
	hsa04114: Oocyte meiosis			
	hsa04914: 黄体酮介导的卵母细胞成熟	5	1.56E-05	0.012080
	hsa04914: Progesterone-mediated oocyte maturation			
模块 2	hsa04610: 补体和凝血级联	6	3.87E-08	3.98E-05
Module2	hsa04610: Complement and coagulation casca3es			
模块 3	hsa00982: 药物代谢 - 细胞色素 P450	6	4.45E-09	3.02E-06
Module3	hsa00982: Drug metabolism-cytochrome P450			
	hsa05204: 化学致癌作用	6	1.02E-08	6.94E-06
	hsa05204: Chemical carcinogenesis			
	hsa00830: 视黄醇代谢	5	4.64E-07	3.15E-04
	hsa00830: Retinol metabolism			
	hsa00980: 细胞色素 P450 对异生素的代谢	5	8.36E-07	5.68E-04
	hsa00980: Metabolism of xenobiotics by cytochrome P450			
	hsa00591: 亚油酸代谢	4	3.72E-06	0.002527
	hsa00591: Linoleic acid metabolism			
	hsa00232: 咖啡因代谢	3	1.18E-05	0.008026
	hsa00232: Caffeine metabolism			
	hsa01100: 代谢途径	8	3.66E-05	0.024859
	hsa01100: Metabolic pathways			
模块 5	hsa04978: 矿物质吸收途径	5	1.46E-09	1.46E-07
Module5	hsa04978: Mineral absorption			

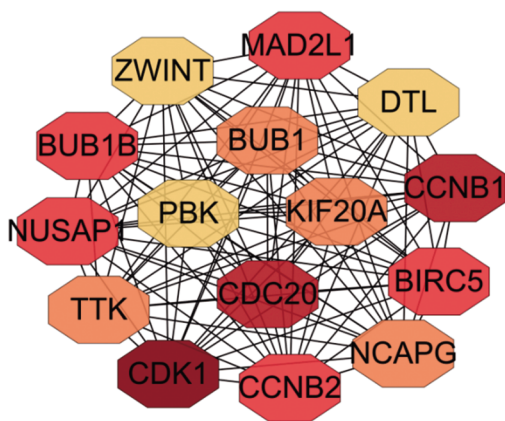


图 6 15 个核心基因编码蛋白质的互作网络

Figure 6 Protein-protein interaction network of top 15 hub genes

MAD2L1)、核仁纺锤体相关蛋白 1 (nucleolar and spindle associated protein1, NUSAP1)、Zw10 结合因子 (zeste white 10 interactor, ZWINT)、BUB1 有丝分裂检查点丝氨酸 / 苏氨酸激酶 (BUB1 mitotic checkpoint serine/threonine kinase, BUB1)、TTK 蛋白激酶

(TTK protein kinase, TTK)、BUB1 有丝分裂检查点丝氨酸 / 苏氨酸激酶 B (BUB1 mitotic checkpoint serine/threonine kinase B, BUB1B)、PDZ 结合激酶 (PDZ-binding kinase, PBK)、无齿状 E3 泛素蛋白连接酶 (denticleless E3 ubiquitin protein ligase, DTL), 在肝癌组织, 这 15 个核心基因均表达增高 ($p < 0.01$) (图 7)。随后, 我们利用 Kaplan Meier Plotter 软件为这 15 个基因绘制生存曲线。生存分析结果显示, 15 个核心基因过表达与患者的不良预后具有相关性。在核心基因高低表达的两组比较发现, 患者总体生存率存在差异, 显示为低表达组优于高表达组, 且具有统计学意义 (图 8)。最后我们利用 MiRDB 数据库预测了可能调控这 15 个核心基因的 miRNA, 利用 Cytoscape 将其可视化 (图 9)。该 miRNA 调控网络共涉及 26 个 miRNA, 并且 hsa-miR-802 同时调控 PBK 和 CDK1 两个核心基因。我们利用 GEPIA 数据库分析了 PBK 与 CDK1 在肝癌组织中的相关性, 结果显示两者高度相关 (图 10)。

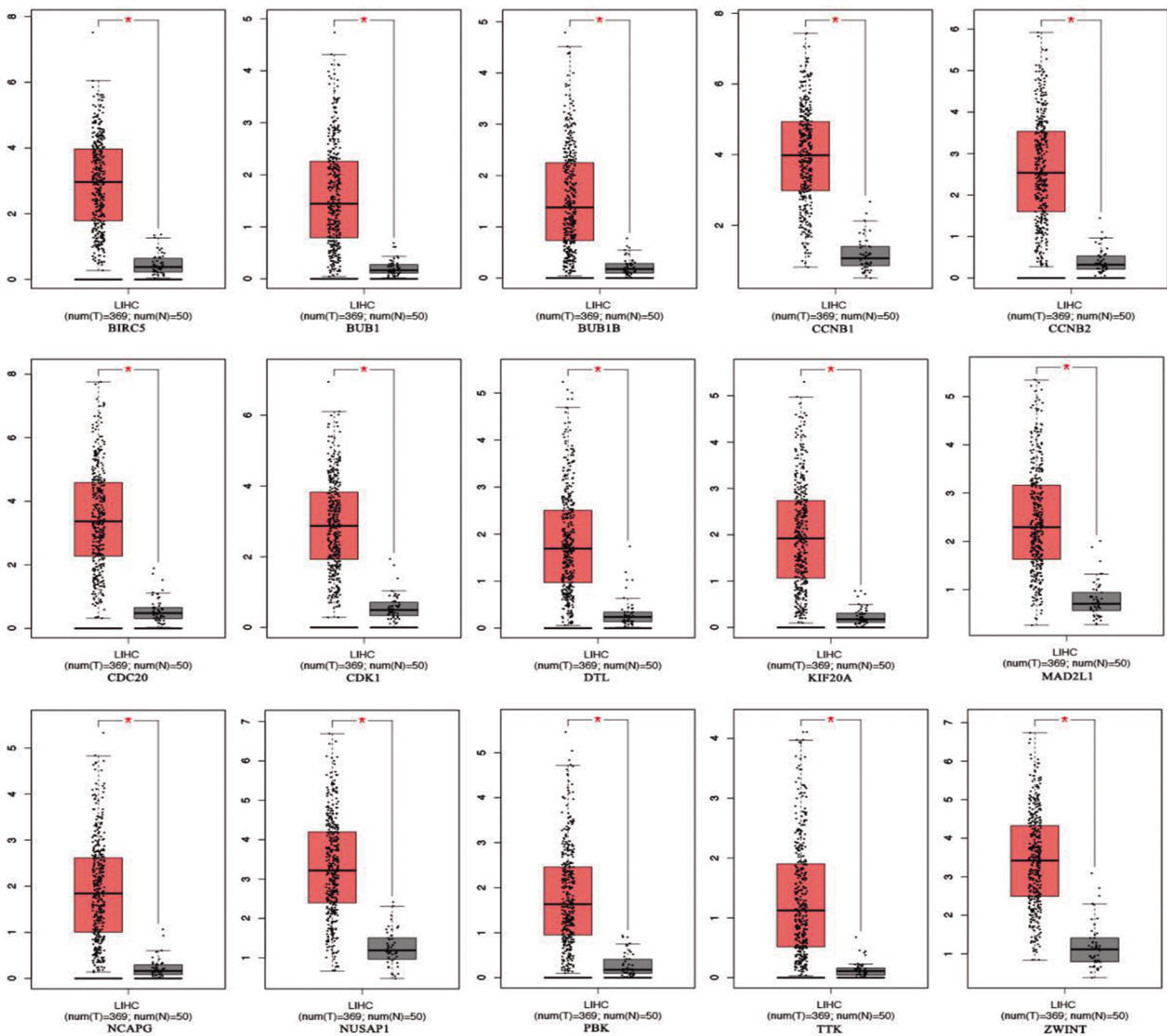


图 7 肝癌组织与癌旁组织 15 个核心基因的表达验证(* $p < 0.01$)

Figure 7 Expression levels of 10 hub genes in Tumor (T) and normal tissues (N) of hepatocellular carcinoma (* $p < 0.01$)

2 讨论

肝癌作为世界常见恶性肿瘤，其发病过程是一个动态生物学过程(Segal et al., 2018)。目前人们对于肝癌的发病机制仍不明确。近年来，现代生物技术飞速发展，高通量测序等技术帮助人们从分子水平探索疾病发病的机制，从而更进一步探索疾病肿瘤的发生、发展过程。本文基于 GEO 数据库通过对肝癌组织和癌旁组织的数据进行分析，筛选出 309 个 DEGs，其中有 94 个 DEGs 表达上调的，215 个 DEGs 下调。我们对这些差异基因进行了 GO 功能富集、KEGG pathway 富集分析以及 PPI 网络分析，筛选出 15 个核心基因并验证其在肝癌组织与正常组织中的表达情况，发现均在肝癌组织中高表达。随后，为它们绘制了 Kaplan-Meier 生存分析曲线。15 个核心基

因基因低表达组优于高表达组，且差异具有统计学意义，可能为促癌基因。

异常的细胞周期调控以及肿瘤细胞过度分裂是肿瘤发生的重要原因之一，而这与细胞有丝分裂的异常增加有关。其中，已有报道显示，BIRC5 具有抑制细胞凋亡、调控细胞周期尤其是有丝分裂过程的功能(Cao et al., 2013)，相较于普通肝组织，BIRC5 在肝癌组织中过表达，且定位于肝癌细胞的细胞质和细胞核中(Yamamoto et al., 2008)。CCNB1 调节有丝分裂中的 G2 到 M 转变，其表达在整个细胞周期中呈周期性变化。研究显示 CCNB 1 基因敲除抑制肝癌细胞的增殖、迁移和侵袭(Gu et al., 2019)。CDK1 可以与 CCNB1 结合形成 CDK1-CyclinB 激酶复合体，从而调节细胞周期中 G1/S 和 G2/M 期的转变(Dash

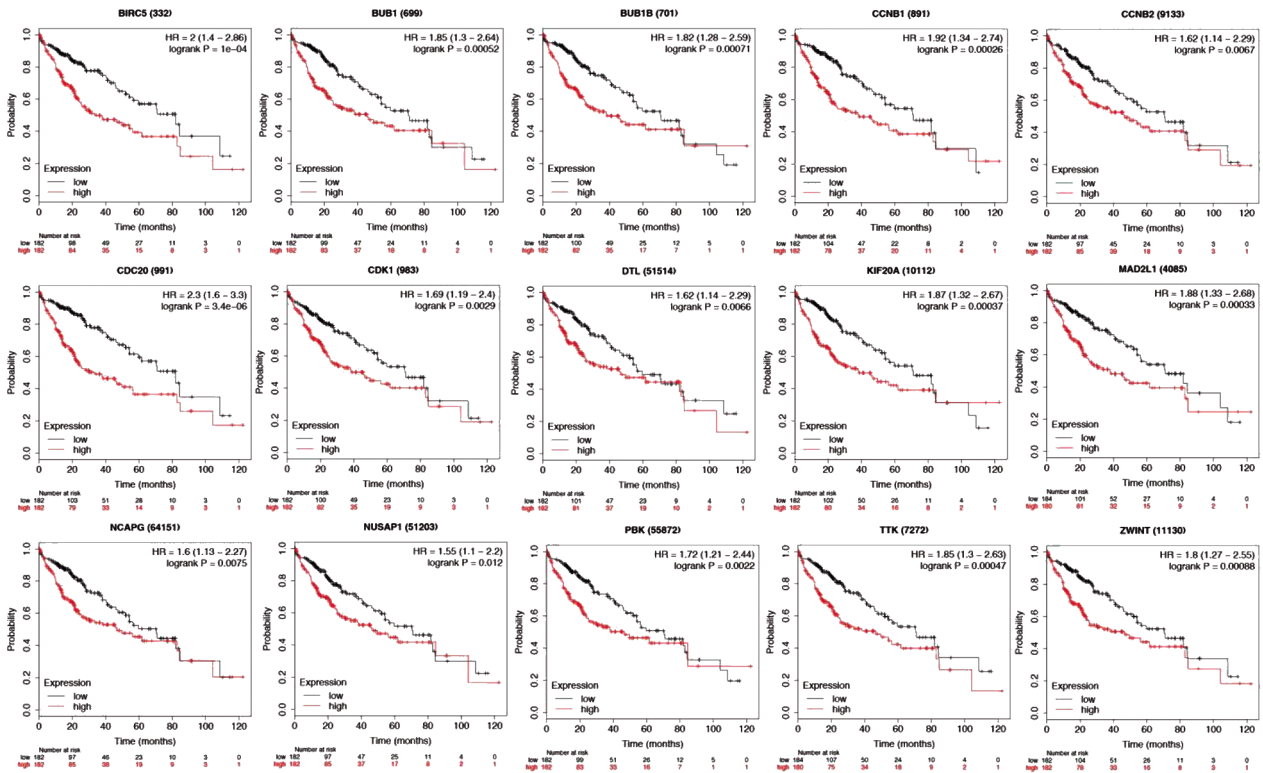


图 8 肝细胞癌核心基因的 Kaplan-Meier 生存曲线分析
Figure 8 Kaplan-Meier survival curve analysis of hub genes of Hepatocellular carcinoma

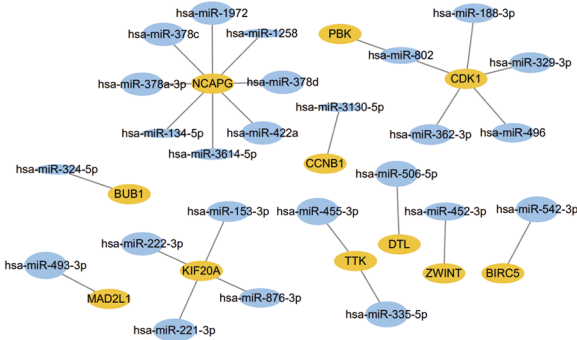


图 9 核心基因的 miRNA 调控网络
Figure 9 The miRNA regulating network of hub genes

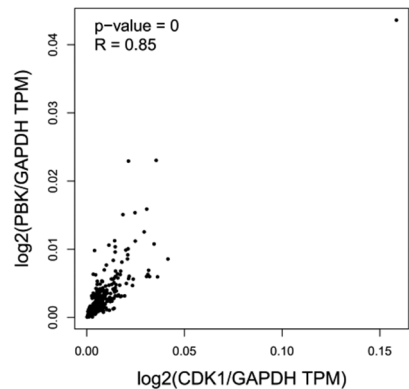


图 10 肝细胞癌中 PBK 和 CDK1 的相关性研究
Figure 10 The correlation between PBK and CDK1 in HCC

and El-Deiry, 2005), 而乙型肝炎病毒 X 蛋白(Hepatitis B virus X protein, HBx) 可以通过持续激活 CDK1-CyclinB 激酶复合体诱导 G2/M 期阻滞和凋亡,抑制 HCC 细胞的增殖(Dash and El-Deiry, 2005)。CCNB2 在肝癌细胞中高表达, 下调 CCNB2 表达可抑制 BEL-7404 细胞增殖和迁移,促进细胞凋亡。并且与肿瘤数目及大小、癌栓和丙氨酸氨基转移酶水平显著相关(Li et al., 2019)。CDC20 是参与细胞周期的调控蛋白(Wang et al., 2013), 其在肝癌组织中高表达, 这与肿瘤的分化和 TNM 分期相关, 沉默 CDC20 的表达可以显著抑制 HCC 细胞的增殖 (Li et al., 2014)。KIF20A 作为驱动蛋白家族的一员 Gli2 通过激活 FOXM1 增强 KIF20A 表达有助于 HCC 细胞的增殖和细胞周期进程。这说明 KIF20A 在 HCC 的发生发展中起到了关键的作用, 可作为预测临床 HCC 预后的生物标志物 (Shi et al., 2016; Lu et al., 2018)。Li 等人的实验发现, MAD2L1 在肝癌组织中表达上调, 它促进 HCC 细胞的活力, miR-200c-5p 的过表达下调了 MAD2L1 的表达并诱导细胞凋亡和细胞周期停滞, 从而抑制 HCC 细胞的增殖(Li et al., 2017)。NCAPG 是缩合素复合物的一个亚基, 仅在入肝癌组织和细胞中表达, 在正常人肝组织中不表达, 下调

NCAPG 的表达抑制 HCC 细胞生长和增殖。值得注意的是,NCAPG 的表达与 CCNB1 的表达正相关,这说明 NCAPG 与 CCNB1 之间可能存在相互作用(Zhang et al., 2018)。ZWINT 是染色体运动和有丝分裂检查站的重要调控蛋白,可以调节如 CDK1,CCNB1,PCNA 等与细胞周期调控有关的蛋白促进肝癌细胞增殖(Ying et al., 2018; 李慧等, 2018)。TTK 为有丝分裂检查点蛋白,其在有丝分裂时期起到准确分离染色体的作用,在 77.63%的肝癌组织中过表达,体内外实验均证明 TTK 可以促进肿瘤的生长及迁移(Liu et al., 2015; Miao et al., 2016)。BUB1 是一种丝氨酸/苏氨酸激酶,其在配对染色体和建立有丝分裂纺锤体检查点方面发挥重要作用,MiR-490-5p 可以通过抑制 BUB1 来调节 TGF β / Smad 信号通路,从而抑制 HCC 细胞的增殖,侵袭和迁移(Xu et al., 2017)。PBK 属于丝裂原活化蛋白激酶(mitogen-active protein kinase, MAPK)家族成员。研究表明其在肝癌标本及细胞系中高表达,通过调节 ETV4/uPAR 信号通路促进肝癌细胞的侵袭(Yang et al., 2019)。NUSAP1 作为重要的有丝分裂调节因子,HBx 通过抑制 miR-18b 上调 NUSAP1 的表达从而促进肝癌细胞的增殖,且其高表达与癌症的恶性程度密切相关(Yang et al., 2019)。DTL 在细胞周期的 S 期调节中起重要作用。研究表明,DTL 的耗竭降低了 G2/M 期肝细胞比例,诱使细胞变大变平,并且 DTL 可以通过下调 TPX2 来抑制肝癌细胞的生长(Chen et al., 2018)。BUB1B 在前列腺癌细胞中的过表达促进了细胞的增殖和迁移,其高表达与前列腺癌不良的临床病理特征相关(Fu et al., 2016)。在多发性骨髓瘤中,BUB1B 表达上调促进了多发性骨髓瘤细胞增殖,并且 BUB1B 的表达与 CDC20、CCNB1/2 表达高度相关(Yang et al., 2015)。BUB1B 的敲除显著地抑制了肺腺癌细胞系的锚定非依赖性生长,在小鼠模型研究中,抑制 BUB1B 抑制原发肿瘤生长,减少肺和淋巴结转移(Chen et al., 2015)。但其在肝癌的研究鲜有报道。

综上所述,核心基因均在肝癌组织或细胞中表达增高,这与我们的生物信息学分析结果一致。这些基因在肝癌的发病过程中发挥重要的作用,可能是潜在的治疗靶点及预测肿瘤发生发展的生物标志物。但其之间的调节机制仍未阐明,我们预测了可能调控这些基因的 miRNA,为核心基因在肝癌发挥作用的机制提供新的思路。并且我们发现 hsa-miR-802 同时调控 PBK 和 CDK1 两个核心基因。我们利用

GEPIA 数据库分析显示 PBK 与 CDK1 在肝癌组织中高度相关,已有文献报道 CDK1 介导的 PBK 有丝分裂磷酸化抑制了其在乳腺癌细胞中的致癌活性(Stauffer et al., 2017),这提示 PBK/CDK1 轴可能成为肝癌发生发展过程中的重要环节。并且通过文献的查阅,我们发现核心基因之间也存在相互作用,但其相互调控的机制以及这种肝癌相关基因间的相互调控在肝癌的发生发展中起到的作用仍需要进一步的试验探索。

3 材料和方法

3.1 数据获取

本研究从 GEO (Gene Expression Omnibus, <http://www.ncbi.nlm.gov/geo/>) 数据库中筛选样本量大于 100 且由慢性乙型肝炎诱导的肝细胞癌的癌组织与癌旁组织相对照的数据。本文下载了 GSE121248 数据集(Wang et al., 2007)。该数据集包含 107 个样本,其中癌组织样本 70 个,癌旁组织样本 37 个。

3.2 DEGs 数据处理

GEO2R (<http://www.ncbi.nlm.gov/geo/geo2r/>) 是一个交互式的可访问的在线工具。它使用 Bioconductor 项目中的 GEOquery 和 limma R 软件包对研究者提供的处理数据表进行比较,允许研究者比较两个或多个数据集,从而获得差异表达基因(Differentially expressed genes, DEGs)。本研究利用 GEO2R 分析肝细胞癌及癌旁样本,以 $\text{adjust } P < 0.05, |\log \text{FC}| \geq 1.5$ 为标准,得到两组间的差异表达基因。

3.3 DEGs 数据的 GO 分析和 KEGG 分析

DAVID (the Database for Annotation, Visualization and Integrated Discovery, <http://david.ncifcrf.gov/>) 为提供大规模基因和蛋白系统综合的生物学功能信息的在线免费生物信息数据库(Menyhart et al., 2018)。在本研究中,对于 DEGs,通过 DAVID 数据库首先进行了基因本体论(Gene Ontology, GO)分析,获得了在生物过程 (Biological Process, BP)、细胞组成 (Cellular Component, CC)、分子功能(Molecular Function, MF)三方面 DEGs 富集的结果,其次进行了京都基因和基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG)信号通路富集分析并得到通路富集结果。差异以 $p < 0.05$ 且 $\text{FDR} < 0.05$ 为具有统计学意义。

3.4 PPI 网络构建和模块分析

STRING (Search Tool for the Retrieval of Interacting Genes, <http://string-db.org/>) 是一个搜索蛋白质相互作用的在线数据库 (Szkarczyk et al., 2019)。将 DEGs 导入 STRING 后得到蛋白相互作用的网络结构, 并通过 Cytosecape 将其可视化。并利用其 MCODE 插件, 将 degree cutoff 为 2, node score cutoff 为 0.2, k-core 为 2, max.depth 为 100 作为截断标准, 筛选出 PPI 模块, 并以综合评分为 4 作为截断标准, 利用 DAVID 对筛选出的 PPI 模块进行 KEGG 分析。

3.5 核心基因的筛选和验证

通过利用 Cytosecape (Shannon et al., 2003) 软件中的 cytoHubba 插件, 使用 MCC 算法获得了综合评分最高, 即在 PPI 网络中连接度最高的前 15 个基因作为核心基因。GEPIA (Gene Expression Profiling Interactive Analysis, <http://gepia.cancerpku.cn/>) 是基于 TCGA 转录组数据库的高度可视化分析工具 (Tang et al., 2017)。本研究中, 我们利用 GEPIA 在线分析软件分析核心基因在肝癌组织和正常组织中的表达情况以及在肝癌组织中的相关性研究。

3.6 核心基因的生存分析

Kaplan Meier Plotter (<http://kmplot.com/analysis/>) 是一款在线生存分析工具 (Nagy et al., 2018)。它包含了 54675 个基因的 10 461 个癌症样本数据, 可以依据不同的病理特征、患病种群及诱发因素等对乳腺癌、胃癌、肺癌、卵巢癌和肝癌相关的 mRNA 和 miRNA 进行生存分析。本研究以核心基因的表达情况将肝癌患者分为两组, 利用 Kaplan Meier Plotter 绘制核心基因的 Kaplan-Meier 生存分析曲线来对比这两组患者, 从而分析核心基因表达对于肝癌患者的预后的影响。差异以 $p < 0.05$ 为有统计学意义。

3.7 构建核心基因的 miRNA 调控网络

miRDB (<http://mirdb.org/>) 是一个用于 miRNA 靶标预测的在线数据库 (Liu and Wang, 2019)。其所有靶标均通过生物信息学工具 MirTarget 预测, 可以预测包括人、大鼠、小鼠、狗和鸡五个物种的 miRNA 靶标。本研究中以作用评分高于 80 为截取标准, 预测了可能调控核心基因的 miRNA 靶标。

作者贡献

李怡敏完成论文的数据分析和文章初稿的撰写

工作; 涂增完成论文的构思以及论文的修改和定稿; 袁琳和汪洋参与完成了样本数据收集和数据库工作; Nur Fazleen Binti Idris 和黄国旺参与了论文的撰写和修改。全体作者都阅读并同意最终的文本。

致谢

本研究由重庆市基础科学与前沿技术研究专项 (CSTC2015jcyjA10006)、国家自然科学基金青年科学基金 (81501751) 和重庆市博士后科学基金 (Xm2014006) 共同资助。

参考文献

- Cao L., Li C., Shen S., Yan Y., Ji W., Wang J., Qian H., Jiang X., Li Z., Wu M., Zhang Y., and Su C., 2013, OCT4 increases BIRC5 and CCND1 expression and promotes cancer progression in hepatocellular carcinoma, *BMC cancer*, 13 (1): 82
- Chen H., Lee J., Kljavin N.M., Haley B., Daemen A., Johnson L., and Liang Y., 2015, Requirement for BUB1B/BUBR1 in tumor progression of lung adenocarcinoma, *Genes & cancer*, 6 (3-4): 106-118
- Chen Y., Chen L., Huang G., Kang C., Wang K., Tsao M., and Pan H., 2018., Targeting DTL induces cell cycle arrest and senescence and suppresses cell growth and colony formation through TPX2 inhibition in human hepatocellular carcinoma cells, *OncoTargets and Therapy*, 11: 1601-1616
- Cheng P., Li Y., Yang L., Wen Y., Shi W., Mao Y., Chen P., Lv H., Tang Q., and Wei Y., 2009, Hepatitis B virus X protein (HBx) induces G2/M arrest and apoptosis through sustained activation of cyclin B1-CDK1 kinase, *Oncology Reports*, 22 (5): 1101
- Dash B.C., and El-Deiry W.S., 2005, Phosphorylation of p21 in G2/M promotes cyclin B-Cdc2 kinase activity, *Molecular and Cellular Biology*, 25(8): 3364-3387
- Fu X., Chen G., Cai Z., Wang C., Liu Z., Lin Z., Wu Y., Liang Y., Han Z., and Liu J., 2016, Overexpression of BUB1B contributes to progression of prostate cancer and predicts poor outcome in patients with prostate cancer, *OncoTargets and Therapy*, 9: 2211
- Gu J., Liu X., Li J., and He Y., 2019, MicroRNA-144 inhibits cell proliferation, migration and invasion in human hepatocellular carcinoma by targeting CCNB1, *Cancer Cell International*, 19(1): 10-15
- Li H., Wang G., Zheng J., Cai J., Zhang J., Cheng D., Zhou Q., and Yang Y., 2018, Effect of high expression of Zwint on the proliferation of hepatoma cells and the prognosis of liver transplantation for hepatocellular carcinoma, *Organ. Trans-*

- plantation, 9(2): 122-129 (李慧, 汪根树, 郑俊, 蔡建业, 张俊斌, 程道柔, 周奇, 杨扬, 2018, *Zwint* 高表达对肝癌细胞增殖和肝癌肝移植预后的影响, *器官移植*, 9(2): 122-129)
- Han K., 2012, Epidemiology and surveillance of hepatocellular carcinoma, *Liver Cancer*, 1(1): 2-14
- Sun K., Zheng R., Zhang S., Ceng H., Zou X., Chen R., Gu X., Wei W., and He J., 2019, Report of Cancer Incidence and Mortality in Different Areas of China, 2015, *China Cancer*, 28(1): 1-11 (孙可欣, 郑荣寿, 张思维, 曾红梅, 邹小农, 陈茹, 顾秀琪, 魏文强, 赫捷, 2019, 2015 年中国分地区恶性肿瘤发病和死亡分析, *中国肿瘤*, 28(1): 1-11)
- Li J., Gao J., Du J., Huang Z., and Wei L., 2014, Increased CDC20 expression is associated with development and progression of hepatocellular carcinoma, *International Journal of Oncology*, 45(4): 1547
- Li R., Jiang X., Zhang Y., Wang S., Chen X., Yu X., Ma J., and Huang X., 2019, Cyclin B2 overexpression in human hepatocellular carcinoma is associated with poor prognosis, *Archives of Medical Research*, 50(1): 10-17
- Li Y., Bai W., and Zhang J., 2017, MiR-200c-5p suppresses proliferation and metastasis of human hepatocellular carcinoma (HCC) via suppressing MAD2L1, *Biomedicine & Pharmacotherapy*, 92: 1038-1044
- Liu W., and Wang X., 2019, Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data, *Genome Biology*, 20(1): 10-18
- Liu X., Liao W., Yuan Q., Ou Y., and Huang J., 2015, TTK activates Akt and promotes proliferation and migration of hepatocellular carcinoma cells, *Oncotarget*, 6(33): 34309
- Lu M., Huang X., Chen Y., Fu Y., Xu C., Xiang W., Li C., Zhang S., and Yu C., 2018, Aberrant KIF20A expression might independently predict poor overall survival and recurrence - free survival of hepatocellular carcinoma, *IUBMB Life*, 70(4): 328-335
- Menyhárt O., Nagy Á., and Györfy B., 2018, Determining consistent prognostic biomarkers of overall survival and vascular invasion in hepatocellular carcinoma, *Royal Society open science*, 5(12): 181006
- Miao R., Wu Y., Zhang H., Zhou H., Sun X., Csizmadia E., He L., Zhao Y., Jiang C., and Miksad R A., 2016, Utility of the dual-specificity protein kinase TTK as a therapeutic target for intrahepatic spread of liver cancer, *Scientific Reports*, 6: 33121
- Mohammed H.F.A., and Roberts L.R., 2017, Should AFP (or any biomarkers) be used for HCC surveillance? *Current hepatology reports*, 16(2): 137-145
- Nagy Á., Lánckzy A., Menyhárt O., and Györfy B., 2018, Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets, *Scientific reports*, 8(1): 1-9
- Segal R.L., Miller K.D., and Jemal A., 2018, Cancer statistics, 2018, *Ca. Cancer J. Clin.*, 68(1): 7-30
- Shannon P., Markiel A., Ozier O., Baliga N.S., Wang J.T., Ramage D., Amin N., Schwikowski B., and Ideker T., 2003, Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Research*, 13(11): 2498-2504
- Shi C., Huang D., Lu N., Chen D., Zhang M., Yan Y., Deng L., Lu Q., Lu H., and Luo S., 2016, Aberrantly activated Gli2-KIF20A axis is crucial for growth of hepatocellular carcinoma and predicts poor prognosis, *Oncotarget*, 7(18): 26206
- Stauffer S., Zeng Y., Zhou J., Chen X., Chen Y., and Dong J., 2017, CDK1-mediated mitotic phosphorylation of PBK is involved in cytokinesis and inhibits its oncogenic activity, *Cellular Signalling*, 39: 74-83
- Szklarczyk D., Gable A.L., Lyon D., Junge A., Wyder S., Huerta-Cepas J., Simonovic M., Doncheva N.T., Morris J.H., Bork P., Jensen L.J., and Mering C.V., 2019, STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, *Nucleic Acids Research*, 47 (D1): D607-D613
- Tang Z., Li C., Kang B., Gao G., Li C., and Zhang Z., 2017, GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses, *Nucleic Acids Research*, 45(W1): W98-W102
- Tinker A.V., Boussioutas A., and Bowtell D.D.L., 2006, The challenges of gene expression microarrays for the study of human cancer, *Cancer Cell*, 9(5): 333-339
- Wang S.M., London L.P.J.O., and Hui K.M., 2007, Identification and validation of a novel gene signature associated with the recurrence of human hepatocellular carcinoma, *Clinical Cancer Research*, 13(21): 6275-6283
- Wang Z., Wan L., Zhong J., Inuzuka H., Liu P., Sarkar F.H., and Wei W., 2013, Cdc20: a potential novel therapeutic target for cancer treatment, *Current Pharmaceutical Design*, 19(18): 3210
- Xu B., Xu T., Liu H., Min Q., Wang S., and Song Q., 2017, MiR-490-5p suppresses cell proliferation and invasion by targeting BUB1 in hepatocellular carcinoma cells, *Pharmacology*, 100(5-6): 269-282
- Yamamoto H., Ngan C.Y., and Monden M., 2008, Cancer cells survive with survivin, *Cancer Science*, 99(9): 1709-1714
- Yang Q., Zhong S., He L., Jia X., Tang H., Cheng S., Ren J., Yu H., Zhou L., and Zhou H., 2019, PBK overexpression promotes metastasis of hepatocellular carcinoma via activating ETV4-uPAR signaling pathway, *Cancer Letters*, 452:

90-102

Yang Y., Gu C., Luo C., Li F., and Wang M., 2015, BUB1B promotes multiple myeloma cell proliferation through CDC20/CCNB axis, *Medical Oncology*, 32(3): 1-5

Yang Z., Li J., Feng G., Wang Y., Yang G., Liu Y., Zhang S., Feng J., and Zhang X., 2019, Hepatitis B virus X protein enhances hepatocarcinogenesis by depressing the targeting of NUSAP1 mRNA by miR-18b, *Cancer Biology & medicine*, 16(2): 276-287

Ying H., Xu Z., Chen M., Zhou S., Liang X., and Cai X., 2018, Overexpression of Zwint predicts poor prognosis and promotes the proliferation of hepatocellular carcinoma by regulating cell-cycle-related proteins, *Onco. Targets and Therapy*, 11: 689-702

Zhang Q., Su R., Shan C., Gao C., and Wu P., 2018, Non-SMC Condensin I Complex, Subunit G (NCAPG) is a novel mitotic gene required for hepatocellular cancer cell proliferation and migration, *Oncology Research*, 26(2): 269