基于生物信息学途径探究肝细胞癌的关键基因

曾　洁，邓　伟，黄天壬
(广西医科大学附属肿瘤医院肿瘤实验研究部，广西 南宁 530000)

摘　要：【目的】采用生物信息学方法探究肝细胞癌(HCC)发生的关键基因，望有助于了解 HCC 发生发展的分子机制。【方法】从 GEO 数据库中选择基因表达谱 GSE76427，该数据集包含 115 例肝癌组织样本和 52 例肝癌临近非肿瘤组织样本。通过 GEO2R 工具在线分析，筛选出肝癌组织中差异表达基因(DEG)，利用 GO 数据库获取 DEG 的功能注释，利用 KEGG 数据库进行通路富集分析。基于 STRING 数据库，利用 MCC 算法，筛选具有高度连接性的关键基因，基于 TCGA 数据库在线软件分析关键基因的预后效应。【结果】共筛选出 190 个 DEG，其中，上调基因有 16 个，下调基因有 174 个。功能富集分析显示，下调的差异基因显著富集于细胞对铬离子、锌离子的反应，也可能参与环氧化酶 P450 途径和氧化还原反应等功能。下调的差异基因可能参与视黄醇的新陈代谢和矿物质的吸收等通路功能。筛选出 15 个具有高度连接性的关键基因，发现 *CDC20*、*KIAA0101*、*PRC1*、*PTTG1* 和 *UBE2C* 等 5 个基因与乙肝相关性肝癌的患者总体生存率(OS)相关。*PTTG1* 的诊断效能最佳。【结论】*CDC20*、*KIAA0101*、*PRC1*、*PTTG1* 和 *UBE2C* 可能在乙型肝炎病毒相关性肝癌的发生发展中发挥重要作用。

关键词：肝细胞癌；TCGA；GEO；关键基因
中图分类号：R73　　**文献标志码**：A　　**文章编号**：1672-3554(2018)05-0780-11

Key Core Genes of Hepatocellular Carcinoma Based on Bioinformatics Approach

ZENG Jie, DENG Wei, HUANG Tian-Ren
(Department of Tumor Research, Affiliated Tumor Hospital of Guangxi Medical University, Nanning 530000, China)
Corresponding to: HUANG Tian-Ren; E-mail: tianrenhuang@sina.com

Abstract:【Objectives】Use bioinformatics methods to explore the key genes in HCC is helpful to understand the molecular mechanisms of The occurrence of hepatocellular carcinoma (HCC) development.【Methods】The present study selected gene expression profile of GSE76427 from the GEO database. 115 cases of hepatocellular carcinoma tissue samples and 52 cases of hepatocellular carcinoma adjacent non-tumor tissue samples were included in this analysis. Differentially expressed genes (DEG) were picked out using GEO2R tools, then we used gene ontology (GO) analysis and Kyoto Encyclopedia of Gene and Genome (KEGG) pathway analysis. Moreover, based on the STRING database, the MCC algorithm was used to collect key genes with high connectivity. The TCGA database online software was used to analyze the prognostic effect of key genes.【Results】Total 190 DEG, including 16 up-regulated genes and 174 down-regulated genes, were screened in this study. Functional enrichment analysis showed that the down-regulated differential genes were significantly enriched in the response of cells to chromium ions and zinc ions, and may also be involved in the functions of the epoxygenase P450 pathway and redox reactions. Down-regulated differential genes may participate in pathway functions such as retinol metabolism and mineral absorption. In this study, 15 key genes with high connectivity were screened and 5 genes including *CDC20*, *KIAA0101*, *PRC1*, *PTTG1*, and *UBE2C* were found to be associated with overall survival (OS) in patients with HBV-related HCC. *PTTG1* has the best diagnostic performance.【Conclusion】*CDC20*, *KIAA0101*, *PRC1*, *PTTG1* and *UBE2C* may play an important role in the development of hepatitis B associated liver cancer.

Key words: hepatocellular carcinoma; TCGA; GEO; key core genes
[J SUN Yat-sen Univ(Med Sci), 2018, 39(5): 780-790]

收稿日期：2018-04-13
基金项目：广西壮族人民自治区科学基金项目(81660561)
作者简介：曾洁，硕士，E-mail：zengjiehcc@163.com；黄天壬，通信作者，教授，研究方向：肿瘤流行病学，E-mail：tianrenhuang@sina.com

肝细胞癌（hepatocellular carcinoma，HCC）是全球最常见的恶性肿瘤之一，其发病率居世界第五，死亡率居世界第三[1]。肝癌的发生与多种因素有关，例如，人的遗传易感性、乙型肝炎病毒（hepatitis B virus，HBV）和丙型肝炎病毒等病毒感染、饮酒、黄曲霉毒素污染、代谢综合征等。中国是肝癌高发地区，患者常以肝区疼痛为首发症状就诊，常伴有乙型肝炎病毒感染和肝硬化的基础疾病。甲胎蛋白（a-Fetoprotein，AFP）、甲胎蛋白异质体（Lens culinaris agglutinin-reactive fraction of a-fetoprotein，AFP-L3）、异常凝血酶原（Des-y-carboxyprothrombin，DCP）是常用于检测肝癌发生发展的标志物，但是，AFP是否应该作为肝癌的检测指标仍然存在较大争议[2]。近年来，肝癌的肿瘤标志物层出不穷，从mRNA、蛋白到ctDNA[3]、miRNA[4]、LncRNA[5]等。虽然，新发现的肝癌标志物灵敏度和特异度均优于AFP，由于依赖高新技术，不如AFP经济实用。例如，ctDNA藏匿于正常DNA片段中，需要依赖数字PCR（Digital PCR）、BEAMing技术、标记扩增深度测序（TAM-Seq）、癌症个体化深度侧粗（CAPP-Seq）等手段。癌症基因组图谱（The Cancer Genome Atlas，TCGA）是由美国国立癌症研究所（National Cancer Institute，NCI）和国立人类基因组研究所（National Human Genome Research Institute，NHGRI）于2006年联合启动的项目，也是全球最大的癌症基因数据库，收录有大量的临床信息。Gene Expression Ominbus（GEO）是隶属于美国国立卫生研究院NCBI的公共开放的基因表达谱数据库。通过对公共数据库的挖掘，医学研究人员可以提高对癌症发病分子机制的认识[6]，有利于提高癌症的早发现、早诊断、早治疗的能力。本研究拟通过对GEO数据库挖掘，发现肝细胞癌的差异表达关键基因，采用相关分析方法，探究关键基因在肝细胞癌发生发展过程的可能分子机制。

1　材料与方法

1.1　微阵列数据获取

从GEO（Gene Expression Omnibus，https://www.ncbi.nlm.nih.gov/geo/）数据库筛选样本量大于60、乙肝相关性肝细胞癌的癌组织和癌旁组织对照的芯片数据。本文下载了GSE76427中115个肝细胞癌组织样本和52个非肿瘤的肝细胞癌癌旁组织样本的基因表达数据。其中，感染HBV的HCC患者占46%。

1.2　DEG数据处理

GEO2R是一个交互式的在线工具，允许用户比较GEO系列中的两个或更多样品组，并利用默认的Benjamini－Hochberg方法测定错误发现率。本研究应用GEO2R（https://www.ncbi.nlm.nih.gov/geo/geo2r/）检测肝癌样本和正常样本两组之间的差异表达基因，以$P<0.01$且$|\log FC| \geqslant 1.5$为截断标准。随后，本研究使用Cytoscape软件中cytoHubba插件的MCC法，获取连接度最高的前15个基因作为关键基因。

1.3　GO注释富集和KEGG信号通路富集

GO（Gene Ontology）是一个提供基因及产物功能注释信息的数据库。KEGG（Kyoto Encyclopedia of Gene and Genome）是一个包含基因、通路、疾病、药物的综合数据库。DAVID（The Database for Annotation，Visualization and Integrated Discovery，https://david.ncifcrf.gov/）整合了许多在线生物信息学资源，可注释基因和蛋白功能。本研究设定$P<0.01$且FDR＜0.05为截断标准，用DAVID可将相应DEG的核心生物学过程、分子功能、细胞组分和通路可视化。

1.4　PPI网络和模块分析

相互作用基因库检索工具（Search Tool for the Retrieval of Interacting Genes，）可用于评估蛋白质相互作用（protein-protein interaction，PPI）。本研究设定最低可信度得分为0.4。将PPI数据导入Cytoscape，利用其中的分子复合体检测（MCODE）插件，以degree cutoff = 2，node score cutoff = 0.2，k-core = 2，and max. depth = 100截断标准，筛选出PPI模块，用DAVID分析每个模块中的基因的信号途径。

1.5　TCGA数据下载与分析

从TCGA数据库中下载公开的肝癌转录组数据与肿瘤样本临床数据，筛选整理罹患HBV相关性HCC的亚洲人，其肝癌组织中关键基因的表达量及临床资料。以基因表达量的75%为界，划分高表达组与低表达组。使用SPSS 22.0软件进行统计分析，计量资料若服从正态分布采用均数±标准差（$\bar{x} \pm s$），计量资料若服从偏态分布采用M（Q_L-Q_U）或表示；计数资料用n表示样本数。计数

资料采用χ^2检验，采用Kaplan-Meier生存分析绘制生存曲线，采用Log-rank检验比较组间差异，$P<0.05$表示差异有统计学意义。

2 结 果

2.1 DEG和关键基因的鉴定

基于指定的分析标准，通过GEO2R在线分析工具，检测出DEG 190个，其中，上调基因有16个，下调基因有174个。所有DEG编码的蛋白质的PPI网络，红色节点表示上调基因，蓝色节点表示下调基因，根据节点大小表示|log FC|大小，线条粗细表示相关程度(图1)。利用MCC算法，筛选出前15个关键基因(hub genes，表1)，用颜色深浅表示MCC算法预测的评分高低(图2)。

2.2 GO富集和KEGG通路富集

GO分析结果显示，下调的DEG显著富集于细胞对镉离子的反应、细胞对锌离子的反应、监管的负增长、氧化还原过程、环氧化酶P450途径等生物过程(BP)。下调的DEG在分子功能(MF)中富集，包括作用于成对的供体，并入或减少分子氧的氧化还原酶活性，以及结合铁离子、结合血红

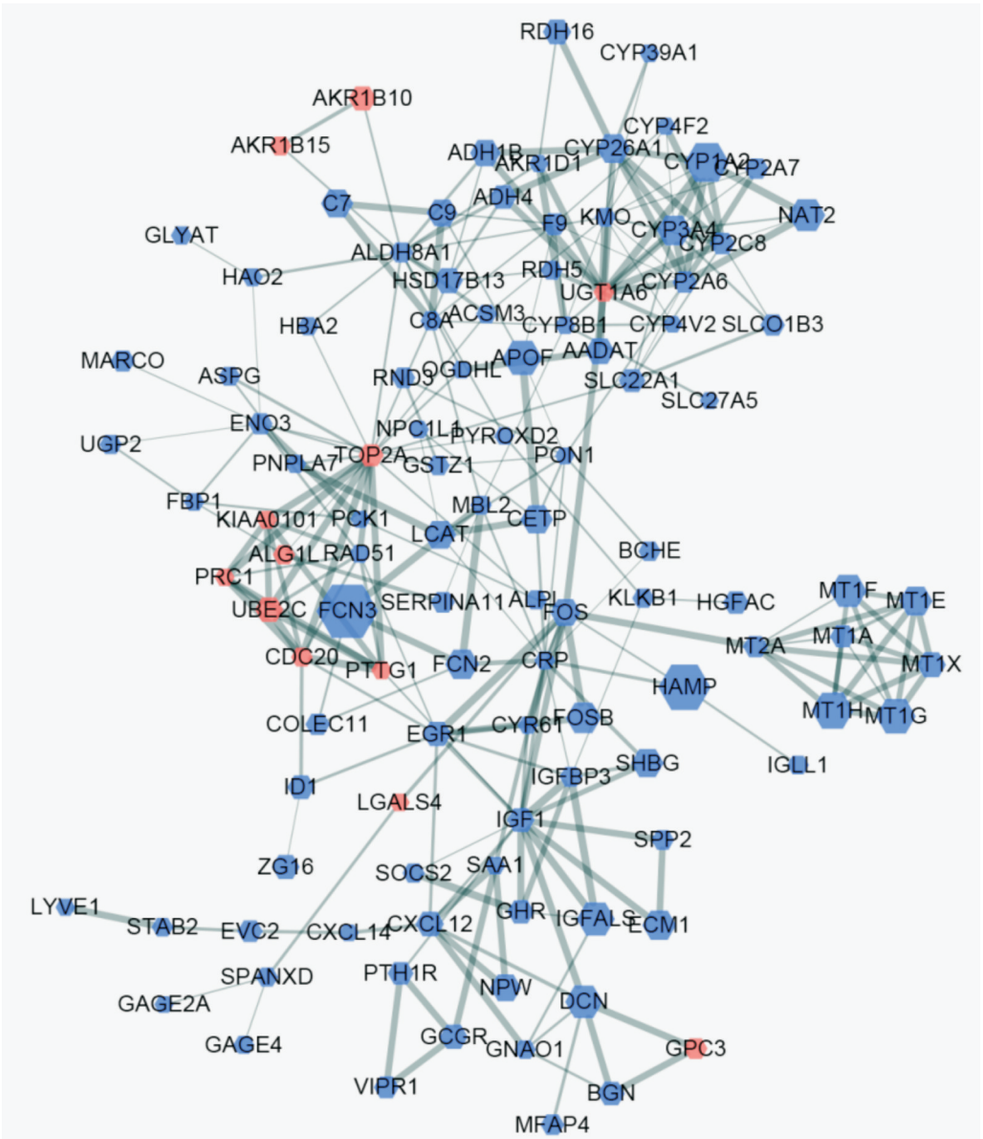


图1 差异基因编码蛋白质相互作用网络

Fig.1 Protein-protein interaction network coded by differentially expressed genes

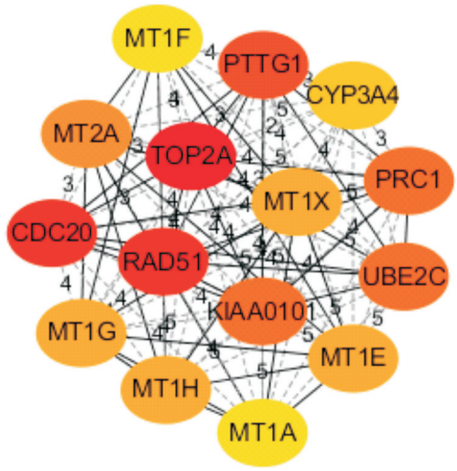


图2　15个关键基因编码蛋白质的相互作用网络

Fig.2　Protein-protein interaction network of top 15 hub genes

素、单加氧酶活性和氧结合。GO细胞成分（CC）分析显示，下调的DEG参与细胞外区域、细胞外空间、细胞外外来体等细胞成分（表2）。下调DEG显著富集的KEGG通路包括：视黄醇的新陈代谢和矿物质的吸收（表3）。由于上调基因数量较少，在GO和KEGG中均未有显著的富集结果。

2.3　关键基因及PPI网络中功能模块

基于公共数据库STRING中蛋白质信息的查询，我们构建了15个连接程度较高的关键基因相应的蛋白质-蛋白质相互作用网络（protein-protein interaction network，图2）。DEG中有4个主要的功能模块（图3），KEGG通路富集分析发现，模块2中的基因参与矿物质的吸收，模块3的基因主要参与视黄醇代谢、药物代谢-细胞色素P450、化学致癌作用、异生素的细胞色素P450代谢以及代谢途径（表4-5）。模块1和模块4未能在KEGG中富集出有统计学意义的信号通路。

2.4　HBV相关性HCC患者的临床基线资料及关键基因的基线资料

基于TCGA数据库下载分析，罹患HBV相关性HCC的亚洲人134例，仅HBV感染者92例，HBV合并HCV感染者42例。患者的临床基线资料详见表6。患者肝癌组织15个关键基因表达量的基线数据详见表7。15个关键基因与AFP诊断效能的比较详见表8。

2.5　关键基因的Kaplan　Meier生存曲线绘制及与病理学参数的相关性分析

基于TCGA数据库中HBV相关性HCC的数据，本研究采用Kaplan-Meier生存曲线的方法绘制关键基因的生存曲线。其中，*CDC20*［HR：3.903（1.893-33.00），$P=0.0049$］、*KIAA0101*［HR：

表1　基于MCC算法的前15个关键基因

Table 1　TOP15 hub genes based on MCC algorithm

Gene symbol	Gene title	P[1)]	Score
TOP2A	topoisomerase (DNA) II alpha	1.8E-16	739
CDC20	cell division cycle 20	2.4E-11	724
RAD51	RAD51 associated protein 1	2.2E-09	724
PTTG1	pituitary tumor-transforming 1 interacting protein	3.5E-04	721
UBE2C	ubiquitin conjugating enzyme E2 C	8.8E-22	720
PRC1	protein regulator of cytokinesis 1	2.0E-20	720
KIAA0101	PCNA clamp associated factor	3.0E-19	720
MT2A	metallothionein 2A	5.0E-19	242
MT1X	metallothionein 1X	1.0E-16	240
MT1H	metallothionein 1H	7.3E-16	240
MT1E	metallothionein 1E	1.1E-14	240
MT1G	metallothionein 1G	1.2E-18	240
CYP3A4	cytochrome P450 family 3 subfamily A member 4	2.9E-12	130
MT1F	metallothionein 1F	1.9E-15	120
MT1A	metallothionein 1A	9.2E-18	120

1）E-n：$\times 10^{-n}$

表2 与肝癌相关的下调差异表达基因的基因本体论分析

Table 2 Gene ontology analysis of down-regulated differentially expressed genes associated with Hepatocellular carcinoma

Category	Term	Count	%	P[1)]	Genes	FDR[1)]
Biological Process	GO: 0071276~cellular response to cadmium ion	7	4.40	3.8E-9	*MT1A*, *MT1E*, *CYP1A2*, *MT1H*, *MT1X*, *MT1G*, *MT1F*	5.9E-6
	GO: 0071294~cellular response to zinc ion	7	4.40	8.2E-9	*MT1A*, *MT2A*, *MT1E*, *MT1H*, *MT1X*, *MT1G*, *MT1F*	1.3E-5
	GO: 0045926~negative regulation of growth	7	4.40	8.2E-9	*MT1A*, *MT2A*, *MT1E*, *MT1H*, *MT1X*, *MT1G*, *MT1F*	1.3E-5
	GO: 0055114~oxidation-reduction process	22	13.8	2.6E-8	*STEAP3*, *CYP3A4*, *ALDH8A1*, *ASPDH*, *PYROXD2*, *CYP2C8*, *HSD17B13*, *CYP26A1*, *KMO*, *STAB2*, *CYP1A2*, *DBH*, *BBOX1*, *RDH5*, *CYP39A1*, *KDM8*, *CYP2A6*, *CYP2A7*, *CYP4F2*, *RDH16*, *CYP8B1*, *AKR1D1*	4.1E-5
	GO: 0019373~epoxygenase P450 pathway	5	3.14	1.4E-5	*CYP2C8*, *CYP2A6*, *CYP2A7*, *CYP4F2*, *CYP1A2*	0.021
Molecular Function	GO: 0016705~oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	10	6.29	1.1E-9	*CYP3A4*, *CYP39A1*, *CYP2C8*, *KLKB1*, CYP26A1, *CYP2A6*, *CYP2A7*, *CYP4V2*, *CYP4F2*, *CYP8B1*	1.5E-6
	GO: 0005506~iron ion binding	13	8.18	6.8E-9	*CYP3A4*, *CYP39A1*, *CYP2C8*, *KLKB1*, CYP26A1, *CYP2A6*, *HBA2*, *CYP2A7*, *CYP4V2*, *CYP4F2*, *CYP1A2*, *CYP8B1*, *BBOX1*	9.3E-6
	GO: 0020037~heme binding	12	7.55	2.3E-8	*CYP3A4*, *CYP39A1*, *CYP2C8*, *KLKB1*, CYP26A1, *CYP2A6*, *HBA2*, *CYP2A7*, *CYP4V2*, *CYP4F2*, *CYP1A2*, *CYP8B1*	3.2E-5
	GO: 0004497~monooxygenase activity	9	5.66	2.9E-8	*CYP3A4*, *CYP39A1*, *CYP2C8*, *KLKB1*, *CYP2A7*, *CYP4V2*, *CYP4F2*, *CYP1A2*, *CYP8B1*	4.0E-5
	GO: 0019825~oxygen binding	7	4.40	2.7E-6	*CYP3A4*, *CYP2C8*, *CYP26A1*, *HBA2*, *CYP2A7*, *CYP1A2*, *CYP8B1*	0.0036
Cellular Component	GO: 0005576~extracellular region	44	27.67	5.1E-13	*C7*, *MBL2*, *C9*, *HSD17B13*, *PAMR1*, *CRP*, *DCN*, *DNASE1L3*, *CXCL12*, *PRSS8*, *AZGP1*, *FCN3*, *BCHE*, *SAA1*, *HAMP*, *FCN2*, *LCAT*, *APOF*, *KLKB1*, *C1QTNF1*, *PROZ*, *CETP*, *HGFAC*, *SPP2*, *GHR*, *CYR61*, *SHBG*, *F9*, *SAA4*, *IGF1*, *IGFALS*, *HBA2*, *COLEC11*, *DBH*, *ECM1*, *C8A*, *AFM*, *CCL14*, *BGN*, *CXCL14*, *NPW*, *PON1*, *MFAP4*, *IGFBP3*	6.1E-10
	GO: 0005615~extracellular space	33	20.75	1.7E-8	*ALPL*, *MBL2*, *SERPINA11*, *CRHBP*, *CRP*, *DCN*, *CXCL12*, *PRSS8*, *AZGP1*, *HAMP*, *SAA1*, *LCAT*, *KLKB1*, *C1QTNF1*, *PROZ*, *ENO3*, *CETP*, *HGFAC*, *GHR*, *F9*, *IGFALS*, *IGF1*, *SAA4*, *DBH*, *ECM1*, *C8A*, *AFM*, *CCL14*, *CXCL14*, *TACSTD2*, *PON1*, *IGFBP3*	2.0E-5
	GO: 0031090~organelle membrane	10	6.29	3.0E-8	*CYP3A4*, *CYP39A1*, *CYP2C8*, *CYP26A1*, *CYP2A6*, *CYP2A7*, *CYP4F2*, *CYP1A2*, *RDH16*, *CYP8B1*	3.6E-5
	GO: 0070062~extracellular exosome	48	30.19	2.4E-7	*ALPL*, *ALDH8A1*, *C7*, *C9*, *PTH1R*, *CRP*, *KMO*, *CXCL12*, *BBOX1*, *PRSS8*, *AZGP1*, *ADIRF*, *SAA1*, *TKFC*, *FCN2*, *LCAT*, *KLKB1*, *PGLYRP2*, *PROZ*, *IGLL1*, *ENO3*, *CETP*, *NDRG2*, *SPP2*, *SHBG*, *CDHR2*, *FBP1*, *F9*, *SAA4*, *IGFALS*, *HBA2*, *ECM1*, *PCK1*, *C8A*, *RND3*, *LYVE1*, *AFM*, *GLYAT*, *BGN*, *TACSTD2*, *HAO2*, *PON1*, *CYFIP2*, *MFAP4*, *AKR1D1*, *IGFBP3*, *UGP2*	2.9E-4
	GO: 0034364~high-density lipoprotein particle	6	3.77	7.0E-7	*SAA1*, *APOF*, *LCAT*, *PON1*, *SAA4*, *CETP*	8.4E-4
	GO: 0072562~blood microparticle	9	5.66	3.0E-5	*C8A*, *AFM*, *C9*, *BCHE*, *FCN3*, *FCN2*, *PON1*, *IGLL1*, *HBA2*	0.0354

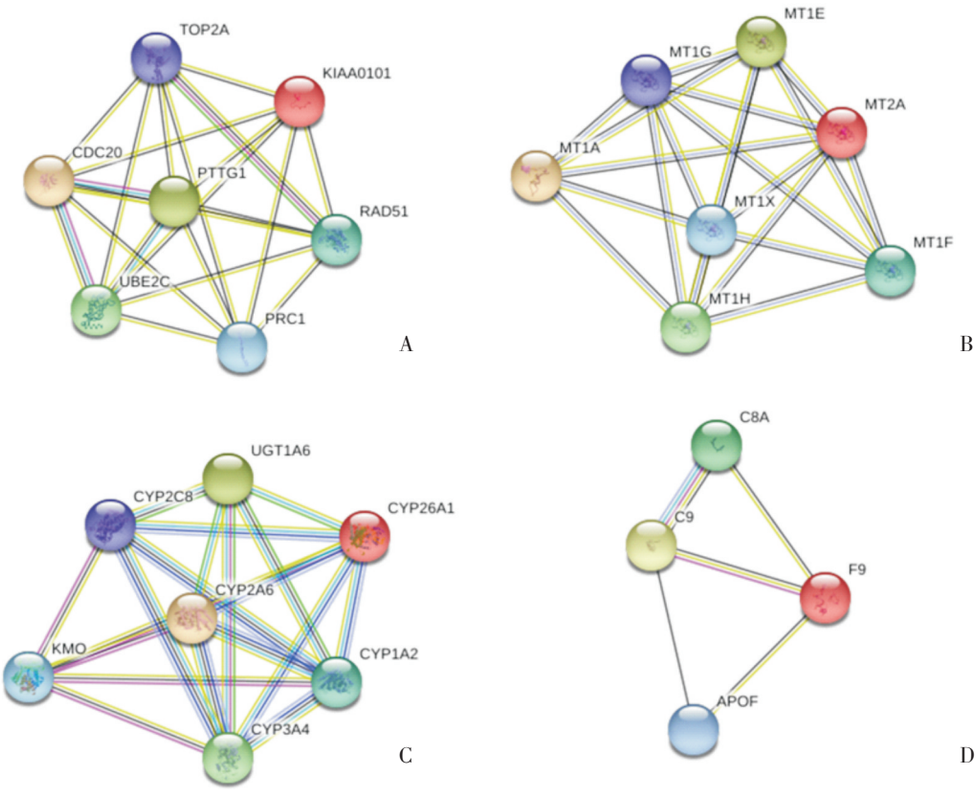
1) E-n: $\times 10^{-n}$

表3　与肝细胞癌相关的下调差异表达基因的KEGG通路分析

Table 3　KEGG pathway analysis of down-regulated differentially expressed genes associated with Hepatocellular carcinoma

Term	Count	%	$P^{1)}$	Genes	FDR[1)]
hsa00830:Retinol metabolism	9	5.66	7.4E-7	*CYP3A4*, *CYP2C8*, *ADH4*, *ADH1B*, *CYP26A1*, *CYP2A6*, *CYP1A2*, *RDH16*, *RDH5*	8.8E-4
hsa04978:Mineral absorption	7	4.40	1.4E-5	*MT1A*, *MT2A*, *MT1E*, *MT1H*, *MT1X*, *MT1G*, *MT1F*	0.0165

KEGG: Kyoto Encyclopedia of Genes and Genomes; FDR: False Discovery Rate. 1) E-*n*: $\times 10^{-n}$



A: module 1; B: module 2; C: module 3; D: module 4

图3　蛋白质-蛋白质相互作用网络的前4个模块

Fig.3　Top 4 modules from the protein-protein interaction network

表4　模块2的KEGG通路富集分析

Table 4　KEGG pathway enrichment analysis of module 2

Term	*P*	FDR	Genes
hsa04978: Mineral absorption	6.2E-14[1)]	6.21E-12[1)]	*MT1A*, *MT2A*, *MT1E*, *MT1H*, *MT1X*, *MT1G*, *MT1F*

1) E-*n*: $\times 10^{-n}$

4.102（2.083- 38.68），*P* = 0.0034］、*PCR1*［HR：3.386（1.439- 22.07），*P* = 0.0135］、*PTTG1*［HR：8.151（7.812-127.6），*P* < 0.0001］、*UEB2C*［HR：2.979（1.129-20.21），*P* = 0.035］的总体生存率在其高表达组和低表达组组间比较差异具有统计学意义，且低表达组均优于高表达组（图4）。这5个

表5 模块3的KEGG通路富集分析

Table 5 KEGG pathway enrichment analysis of module 3

Term	P[1)]	FDR[1)]	Genes
hsa00830: Retinol metabolism	3.8E−10	2.6E−7	*CYP3A4*, *UGT1A6*, *CYP2C8*, *CYP26A1*, *CYP2A6*, *CYP1A2*
hsa00982: Drug metabolism−cytochrome P450	1.3E−7	8.8E−5	*CYP3A4*, *UGT1A6*, *CYP2C8*, *CYP2A6*, *CYP1A2*
hsa05204: Chemical carcinogenesis	2.5E−7	1.7E−4	*CYP3A4*, *UGT1A6*, *CYP2C8*, *CYP2A6*, *CYP1A2*
hsa00980: Metabolism of xenobiotics by cytochrome P450	2.3E−5	0.0160	*CYP3A4*, *UGT1A6*, *CYP2A6*, *CYP1A2*
hsa01100: Metabolic pathways	3.1E−5	0.0217	*CYP3A4*, *UGT1A6*, *CYP2C8*, *CYP26A1*, *CYP2A6*, *KMO*, *CYP1A2*

1) E−n: $\times 10^{-n}$

表6 乙型肝炎病毒相关性肝癌患者临床基线资料

Table 6 Clinical baseline data of HBV−related HCC patients (n/n or $\bar{x} \pm s$)

Characteristics	Number of patients (n=134)	Characteristics	Number of patients (n=134)
Sex(Male/Female)	105/29	PT	1.0(0.8−10.9)
Age	55.5 ± 10.4	INR	1.0(1.0−12.0)
Height	164.6 ± 7.4	AFP(≥400/ < 400)	70/30
Weight	64.4 ± 10.4	Fibrosis ishak score[1)] (0/1 & 2/3 & 4/5/6)	6/12/14/2/32
BMI(< 18.5/18.5−22.9/23−24.9/25−29.9/≥30/≥40)	10/58/35/23/5/0	Child−Pugh score(A/B)	84/4
Relative family cancer history (Yes/No)	14/110	T stage(T1/T2/T3/T3a/T4)	65/29/22/12/6
Viral hepatitis(HBV/HBV+HCV)	92/42	N stage(N0/N1/NX)	123/1/10
TBil	1.2(1.0−1.4)	M stage(M0/M1/MX)	127/1/6
ALB	4.1(0.2−6.9)	New neoplasm event occurrence anatomic site(Intrahepatic/Extrahepatic/Both)	107/21/6
PLT	180.8 ± 55.6	PVTT(Yes/No)	71/35
Cr	0.9(0.5−54.0)	Neoplasm histologic grade (G1/G2/G3/G4)	13/52/59/10

BMI: Body mass index (kg/m^2); TBil: total bilirubin; ALB: albumin; PLT: blood platelet; Cr: creatinine; PT: prothrombin time; INR: International Normalized Ratio; AFP: alpha fetoprotein; PVTT: portal vein tumor thrombus. 1)0−No Fibrosis; 1 & 2−Portal Fibrosis; 3 & 4−Fibrous Speta; 5−Nodular Formation and Incomplete Cirrhosis; 6−Established Cirrhosis

关键基因与病理学参数的相关性分析详见表9。

3 讨 论

本研究共筛选出190个DEG中,上调基因有16个,下调基因有174个。功能富集分析显示,下调的差异基因显著富集于细胞对铬离子、锌离子的反应,也可能参与环氧化酶P450途径和氧化还原反应等功能。下调的差异基因可能参与视黄醇的新陈代谢和矿物质的吸收等通路功能。基于TCGA数据库,15个具有高度连接性的关键基因中,*CDC20*、*KIAA0101*、*PCR1*、*PTTG1*、*UBE2C*,其低表达组的生存率优于高表达组,且与肿瘤大小相关(表7),可能为促癌基因。*PTTG1*的诊断效能最佳,AUC曲线下面积为0.728,可能作为乙肝相关性肝癌潜在的诊断指标。

表7　肝癌组织中关键基因表达量的基线资料

Table 7　Baseline data of hub genes expression in hepatocellular carcinoma tissues　mean(range)

Genes	Expression level
TOP2A	1458(52-16896)
CDC20	531(28-7310)
RAD51	126(13-571)
PTTG1	611(56-5196)
UBE2C	454(31-4848)
PRC1	631(68-4621)
KIAA0101	337(29-2192)
MT2A	6464(602-382055)
MT1X	1950(22-115223)
MT1H	23(0-82269)
MT1E	670(14-87900)
MT1G	418(2-636622)
CYP3A4	2754(3-336881)
MT1F	122(4-59472)
MT1A	128(20-32921)

表8　AFP及15个关键基因单项诊断HBV相关性HCC的ROC曲线下面积比较

Table 8　AFP and 15 hub genes were compared under the ROC curve of HBV-related HCC

	AUC	95%CI	*P*
AFP	0.568	0.415-0.722	0.415
PTTG1	0.728	0.592-0.863	0.004
CDC20	0.637	0.485-0.789	0.084
KIAA0101	0.584	0.426-0.742	0.291
MT2A	0.583	0.444-0.722	0.295
UBE2C	0.577	0.443-0.710	0.334
MT1A	0.576	0.443-0.710	0.335
MT1X	0.573	0.436-0.709	0.361
CYP3A4	0.565	0.412-0.718	0.411
RAD51	0.559	0.414-0.704	0.459
MT1E	0.537	0380-0.695	0.639
MT1G	0.535	0.383-0.686	0.662
TOP2A	0.525	0.354-0.696	0.754
MT1H	0.520	0.373-0.668	0.797
PRC1	0.517	0.339-0.695	0.830
MT1F	0.502	0.357-0.647	0.980

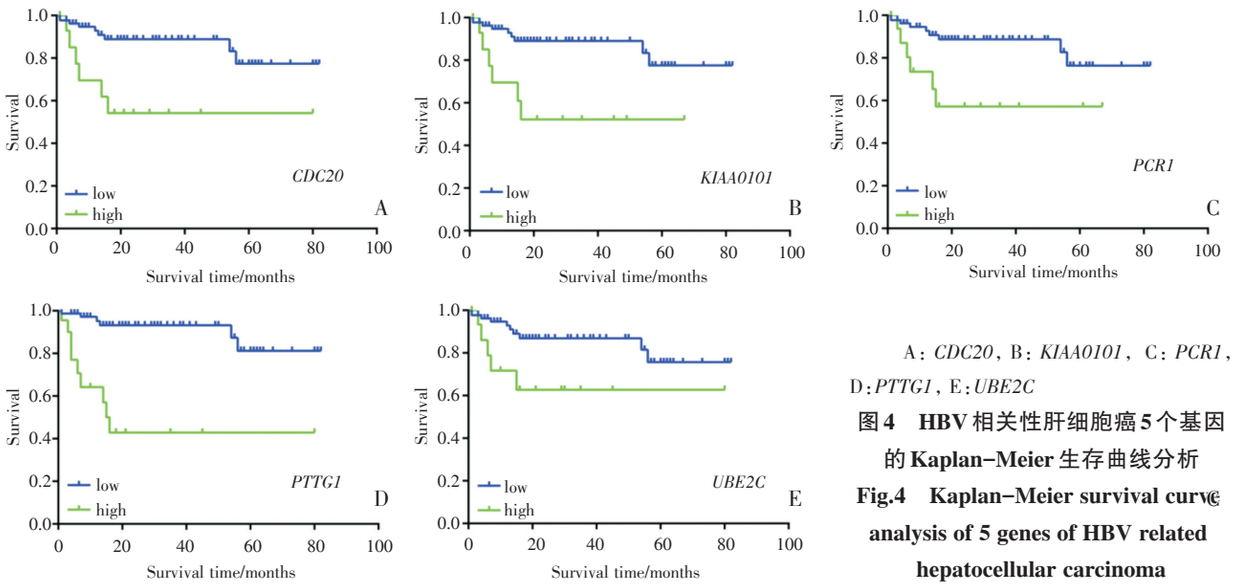
DEG的GO富集显示，HBV相关性HCC的发生与氧化还原反应、氧结合相关，提示肿瘤的发生与细胞能量代谢有关。Hanahan等[7]提出，细胞长期、频繁地增殖失控是肿瘤的本质，增殖失控将引起相应的能量新陈代谢变化。细胞色素P450酶(cytochrome P450，CYP)是肝脏中药的代谢酶，参与多种内源性物质(如类花生四烯酸、脂肪酸、甾醇等)和外源性物质(致癌物、药物等)的I相生物转化[8]。环境毒物(如肝炎病毒、亚硝胺类、黄曲霉毒素等)与CYP酶密切相关[9]。

CDC20 (Cell Division Cycle 20)是参与细胞周期的调控蛋白，在染色体分离之前，与核运动所需的许多其他蛋白相互作用[10]。Li等[11]研究表明，沉默CDC20表达显著抑制HCC细胞增殖和周期进展。CDC20可能成为癌症治疗的潜在靶点[12]。

KIAA0101是一种增殖细胞核抗原相关因子，参与细胞增殖，主要定位于线粒体，部分位于细胞核[13]。216例HCC组织mRNA、蛋白水平研究显示，KIAA0101高表达是与肿瘤分级高、血管侵犯、早期肿瘤复发的独立危险因素，导致不良预后[14]。KIAA0101存在KIAA0101转录变异体1和KIAA0101转录变异体2两种结构。其中，KIAA0101转录变异体1通过抑制p53激活来预防阿霉素诱导的细胞凋亡，使肝癌细胞存活[15]。而KIAA0101转录变异体2可竞争性结合P53，在肝癌中发挥抗肿瘤作用[16]。

胞质分裂素1(protein regulator of cytokinesis 1，PRC1)是一种微管相关蛋白，它调节反平行的微管交联 促进微管的形成[17]。在有丝分裂的S和G2/M期，PRC1蛋白质含量高，G1期时，其蛋白水平下降。PCR1可能是反应肝癌细胞有丝分裂是否活跃的一个指标。此外，PRC1削弱5-FU诱导的细胞凋亡，消除了5-FU诱导的G2 / M期阻滞，增加了化疗耐药，因此，若接受化疗的HCC患者PRC1高表达，其预后不良[18]。PRC1还可以增强Wnt信号通路，以促进HCC早期复发[19]。

垂体肿瘤转化基因1(pituitary tumor-transforming 1，*PTTG1*)抑制姊妹染色单体分离，参与肿瘤的发生发展。在肝癌组织中，*PTTG1*的表达量和AFP水平，门静脉癌栓、肿瘤分期和成纤维细胞生长因子bFGF蛋白水平相关[20-21]。因此，*PTTG1*可能作为HCC诊断的候选指标。*PTTG1*与HBV可能存在密切关系。一方面，HBV抑制miR-



A: *CDC20*, B: *KIAA0101*, C: *PCR1*, D: *PTTG1*, E: *UBE2C*

图4 HBV相关性肝细胞癌5个基因的Kaplan-Meier生存曲线分析

Fig.4 Kaplan-Meier survival curve analysis of 5 genes of HBV related hepatocellular carcinoma

表9 5个关键基因与病理学参数的相关性

Table 9 Correlation between 5 hub genes and pathological parameters

Characteristics	*n*	*CDC20* expression			*KIAA0101* expression			*PRC1* expression			*PTTG1* expression			*UBE2C* expression		
		Low	High	*P*	Low	High	*P*	Low	High	*P*	Low	High	*P*	Low	High	*P*
T stage				0.0026			0.0012			0.0093			0.0113			0.0006
T1	65	57	8		57	8		53	12		56	9		57	8	
T2	29	21	8		22	7		24	5		20	9		21	8	
T3	22	10	12		10	12		10	12		11	11		9	12	
T3a	12	9	3		9	3		10	2		10	2		9	3	
T4	6	4	2		3	3		4	2		4	2		5	1	
N stage				0.1337			0.2040			0.1021			0.4332			0.2040
N0	123	95	28		93	30		95	28		94	29		93	30	
N1	1	1	0		0	1		0	1		1	0		0	1	
NX	10	5	5		8	2		6	4		6	4		8	2	
M stage				0.2898			0.7581			0.7501			0.7501			0.7581
M0	127	97	30		95	32		96	31		96	31		95	32	
M1	1	1	0		1	0		1	0		1	0		1	0	
MX	6	3	3		5	1		4	2		4	2		5	1	

122，上调*PTTG1*，促进肝细胞癌的生长和侵袭[22]。另一方面，HBV X蛋白可促进PTTG1的异常积累，可能导致HBV相关HCC发病[23]。

泛素结合酶E2C（ubiquitinconjugating enzyme E2C，UBE2C）是泛素-蛋白酶体系统的组成部分，其基因位于染色体20q13上。尽管UBE2C的蛋白大小仅19.6 ku，但其结构复杂，可能与泛素激活酶E1、泛素连接酶E3相互作用[24]。UBE2C的表达与卵巢癌、乳腺癌、肺癌、膀胱癌、脑癌等多种肿瘤分化不良有关。在HCC患者中，UBE2C在肝癌组织中的表达高于癌旁组织，且UBE2C高表达与肿瘤浸润、门静脉侵犯、肿瘤分化相关，无瘤生存率较UBE2C低表达组更低[25]。

综上所述，*CDC20*、*KIAA0101*、*PCR1*、*PTTG1*、*UBE2C*高表达均可引起HCC的不良预后，与本文生物信息学分析结果一致。尽管*CDC20*、*KI-*

AA0101、*PCR1*、*PTTG1*均显示与细胞周期有关，但其相互调控机制仍需进一步的试验探索。

现阶段，肿瘤化疗在临床上的应用主要是经验性的，未充分考虑个体肿瘤的生物学特点，而肿瘤的异质性常导致肿瘤耐药，引起化疗效果不佳[26]。寻找预测肿瘤发生发展的分子标志物，研发新的抗肿瘤靶向药物，将有助于肿瘤的个体化治疗。

（感谢麦荣云、陈文杰在论文修改、数据处理、理论提出中给予的重大帮助）

参考文献

[1] Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017[J]. CA Cancer J Clin, 2017, 67(1): 7-30.

[2] Ahmed MHF, Roberts LR. Should AFP (or any biomarkers) be used for HCC surveillance?[J]. Curr Hepatol Rep, 2017, 16(2): 137-145.

[3] Xu RH, Wei W, Krawczyk M, et al. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma[J]. Nat Mater, 2017, 16(11): 1155-1161.

[4] Yang N, Ekanem NR, Sakyi CA, et al. Hepatocellular carcinoma and microRNA: new perspectives on therapeutics and diagnostics[J]. Adv Drug Deliv Rev, 2015, 81(9): 62-74.

[5] Yang Y, Chen L, Gu J, et al. Recurrently deregulated lncRNAs in hepatocellular carcinoma[J]. Nat Commun, 2017, 8(1): 144-151.

[6] Noor AM, Holmberg L, Gillett C, et al. Big Data: the challenge for small research groups in the era of cancer genomics[J]. Br J Cancer, 2015, 113(10): 1405-1412.

[7] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation[J]. Cell, 2011, 144(5): 646-674.

[8] Guengerich FP. Reactions and significance of cytochrome P-450 enzymes[J]. J Biol Chem, 1991, 266(16): 10019-10022.

[9] Fontham ET, Thun MJ, Ward E, et al. American Cancer Society perspectives on environmental factors and cancer[J]. CA Cancer J Clin, 2009, 59(6): 343-351.

[10] Yamaguchi M, VanderLinden R, Weissmann F, et al. Cryo-EM of mitotic checkpoint complex-bound APC/C reveals reciprocal and conformational regulation of ubiquitin ligation[J]. Mol Cell, 2016, 63(4): 593-607.

[11] Li J, Gao JZ, Du JL, et al. Increased CDC20 expression is associated with development and progression of hepatocellular carcinoma[J]. Int J Oncol, 2014, 45(4): 1547-1555.

[12] Wang Z, Wan L, Zhong J, et al. Cdc20: a potential novel therapeutic target for cancer treatment[J]. Curr Pharm Des, 2013, 19(18): 3210-3214.

[13] Guo M, Li J, Wan D, et al. KIAA0101 (OEACT-1), an expressionally down-regulated and growth-inhibitory gene in human hepatocellular carcinoma[J]. BMC Cancer, 2006, 6(1): 109-117.

[14] Yuan RH, Jeng YM, Pan HW, et al. Overexpression of KIAA0101 predicts high stage, early tumor recurrence, and poor prognosis of hepatocellular carcinoma[J]. Clin Cancer Res, 2007, 13(18 Pt 1): 5368-5376.

[15] Liu L, Chen X, Xie S, et al. Variant 1 of KIAA0101, overexpressed in hepatocellular carcinoma, prevents doxorubicin-induced apoptosis by inhibiting p53 activation[J]. Hepatology, 2012, 56(5): 1760-1769.

[16] Liu L, Liu Y, Chen X, et al. Variant 2 of KIAA0101, antagonizing its oncogenic variant 1, might be a potential therapeutic strategy in hepatocellular carcinoma[J]. Oncotarget, 2017, 8(27): 43990-44003.

[17] Subramanian R, Wilson-Kubalek EM, Arthur CP, et al. Insights into antiparallel microtubule crosslinking by PRC1, a conserved nonmotor microtubule binding protein[J]. Cell, 2010, 142(3): 433-443.

[18] Wang Y, Shi F, Xing GH, et al. Protein regulator of cytokinesis PRC1 confers chemoresistance and predicts an unfavorable postoperative survival of hepatocellular carcinoma patients[J]. J Cancer, 2017, 8(5): 801-808.

[19] Chen J, Rajasekaran M, Xia H, et al. The microtubule-associated protein PRC1 promotes early recurrence of hepatocellular carcinoma in association with the Wnt/β-catenin signalling pathway[J]. Gut, 2016, 65(9): 1522-1534.

[20] Liang M, Chen X, Liu W, et al. Role of the pitu-

itary tumor transforming gene 1 in the progression of hepatocellular carcinoma [J]. Cancer Biol Ther, 2011, 11(3): 337-345.
[21] Fujii T, Nomoto S, Koshikawa K, et al. Overexpression of pituitary tumor transforming gene 1 in HCC is associated with angiogenesis and poor prognosis [J]. Hepatology, 2006, 43(6): 1267-1275.
[22] Li C, Wang Y, Wang S, et al. Hepatitis B virus mRNA-mediated miR-122 inhibition upregulates PTTG1-binding protein, which promotes hepatocellular carcinoma tumor growth and cell invasion [J]. J Virol, 2013, 87(4): 2193-2205.
[23] Molina-Jiménez F, Benedicto I, Murata M, et al. Expression of pituitary tumor-transforming gene 1 (PTTG1)/securin in hepatitis B virus (HBV)-associated liver diseases: evidence for an HBV X protein-mediated inhibition of PTTG1 ubiquitination and degradation [J]. Hepatology, 2010, 51(3): 777-787.
[24] Xie C, Powell C, Yao M, et al. Ubiquitin-conjugating enzyme E2C: a potential cancer biomarker [J]. Int J Biochem Cell Biol, 2014, 47(1): 113-117.
[25] Ieta K, Ojima E, Tanaka F, et al. Identification of overexpressed genes in hepatocellular carcinoma, with special reference to ubiquitin-conjugating enzyme E2C gene expression [J]. Int J Cancer, 2007, 121(1): 33-38.
[26] Mroz EA, Rocco JW. The challenges of tumor genetic diversity [J]. Cancer, 2017, 123(6): 917-927.

(编辑 刘清海)

(上接第779页 from page 779)

65-87.
[23] Liu F, Chen N, Gong Y, et al. The long non-coding RNA NEAT1 enhances epithelial-to-mesenchymal transition and chemoresistance via the miR-34a/c-Met axis in renal cell carcinoma [J]. Oncotarget, 2017, 8(38): 62927-62938.
[24] Ning L, Li Z, Wei D, et al. LncRNA, NEAT1 is a prognosis biomarker and regulates cancer progression via epithelial-mesenchymal transition in clear cell renal cell carcinoma [J]. Cancer Biomark, 2017, 19(1): 75-83.
[25] Chacon-cortes D, Smith RA, Lea RA, et al. Association of microRNA 17-92 cluster host gene (MIR17HG) polymorphisms with breast cancer [J]. Tumour Biol, 2015, 36(7): 5369-5376.
[26] Guo C, Song W Q, Sun P, et al. LncRNA-GAS5 induces PTEN expression through inhibiting miR-103 in endometrial cancer cells [J]. J Biomed Sci, 2015, 22(1): 100-123.
[27] Sunamura N, Ohira T, Kataoka M, et al. Regulation of functional KCNQ1OT1 lncRNA by beta-catenin [J]. Sci Rep, 2016, 6(20): 690-700.
[28] Tantai J, Hu D, Yang Y, et al. Combined identification of long non-coding RNA XIST and HIF1A-AS1 in serum as an effective screening for non-small cell lung cancer [J]. Int J Clin Exp Pathol, 2015, 8(7): 7887-7895.

(编辑 刘清海)