

胃癌前病变相关关键基因的生物信息学分析

蒙秀东¹, 李昕¹, 袁旭¹, 苗苗¹, 张月林¹, 袁红霞^{2*}

1. 天津中医药大学 研究生院, 天津 301617

2. 天津中医药大学 管理学院, 天津 301617

【摘要】 **目的** 利用生物信息学方法探讨胃癌前病变的差异基因及潜在治疗靶点。**方法** 使用 GEO2R 挖掘 GEO 数据库中的基因芯片数据(GSE55696 和 GSE130823)筛选差异表达基因;然后使用 DAVID 数据库对 DEGs 进行 GO 和 KEGG 途径富集分析;使用 STRING 数据库构建 DEGs 的 PPI 网络,并将其导入 Cytoscape 进行网络分析和模块分析;使用 Kaplan-Meier、GEPIA 和 HPA 数据库分析核心 DEG。**结果** 共筛选出 273 个差异表达基因,其中 74 个上调基因和 199 个下调基因。这些基因主要参与胃酸分泌、白细胞跨内皮迁移、甘油酯代谢、细胞黏附分子、PI3K-Akt 信号通路和糖酵解/糖异生通路;生存分析发现 4 个危险基因(*EFNA3*、*PPP2R3A*、*THBS2*、*EPOR*)和 5 个保护基因(*CCND2*、*MYB*、*OSM*、*ITGB8* 和 *PIK3R3*),并且发现 *EFNA3*、*ITGB8*、*PIK3R3*、*PCK1* 和 *THBS2* 在细胞质/膜组织中过度表达。**结论** *EFNA3*、*PPP2R3A*、*THBS2* 和 *EPOR* 可能是 PLGC 发病的危险基因,*CCND2*、*MYB*、*OSM*、*ITGB8* 和 *PIK3R3* 可能是 PLGC 发病的保护基因。

【关键词】 胃癌前病变; 蛋白质组学; 生物信息学分析; 生物标志物

Identification of key genes associated with precancerous lesions of gastric cancer by bioinformatics analysis

Meng Xiudong¹, Li Xin¹, Yuan Xu¹, Miao Miao¹, Zhang Yuelin¹, Yuan Hongxia^{2*}

Graduate College Tianjin University of Traditional Chinese Medicine, Tianjin 301617, China

【Abstract】 **Objective** To explore the differential genes of precancerous lesions of gastric cancer (PLGC) patients and predict the potential therapeutic targets of PLGC by bioinformatics analysis. **Methods** Two microarray data (GSE55696 and GSE130823) GEO database were selected to screen the differentially expressed genes (DEGs) by using GEO2R online analysis tool, the cut-off criteria of DEGs were $|\log_2 FC| > 1$ and adjusted value < 0.05 ; Then the GO and KEGG pathway enrichment analysis of DEGs were performed using the DAVID database. The PPI network of DEGs was constructed using the STRING database, which was imported into Cytoscape for network analysis and module analysis; and core DEGs were analyzed using Kaplan Meier Plotter, GEPIA and HPA databases. **Results** A total of 273 differentially expressed genes were screened including 74 up-regulated genes and 199 down-regulated genes. These genes were mainly involved in gastric acid secretion, leukocyte trans endothelial migration, glyceride metabolism, cell adhesion molecules, PI3K-Akt signaling pathway and glycolysis / gluconeogenesis pathway, it was found that 4 risky genes (*EFNA3*, *PPP2R3A*, *THBS2*), *EPOR* and 5 protected genes (*CCND2*, *MYB*, *OSM*, *ITGB8* and *PIK3R3*). The overexpression of *EFNA3*, *ITGB8*, *PIK3R3*, *PCK1* and *THBS2* was found in cytoplasmic/membranous. **Conclusion** *EFNA3*, *PPP2R3A*, *THBS2* and *EPOR* in PI3K-Akt pathway may be risk genes for PLGC, while *CCND2*, *MYB*, *OSM*, *ITGB8* and *PIK3R3* may be protective genes for PLGC.

【Key words】 Precancerous lesions of gastric cancer; Proteomics; Bioinformatics analysis; Biomarkers

胃癌(gastric cancer, GC)在全球癌症相关死亡原因中排名第三,是最常见的癌症之一。GC患者常出现消化不良和反流等症状。该疾病的危险因

素包括幽门螺杆菌感染、高盐摄入、年龄以及低水果和蔬菜摄入^[1-3]。其常见的治疗方式包括手术、全身化疗、免疫治疗、放疗和靶向治疗。GC发现时

基金项目:天津市科技计划项目(17XMFYSY00110);天津中医药大学研究生科研创新项目(YJSC-20201012);天津市研究生科研创新项目(2020YJSB199)

第一作者:蒙秀东,博士研究生,E-mail:xiudong19@qq.com

*通信作者:袁红霞,E-mail:yhx1877@163.com

大多处于晚期,5年生存率仅为7%~34%。肠型胃癌一般经历:慢性胃炎、萎缩、肠化生、异型增生、癌变和转移等过程^[4]。其中慢性萎缩性胃炎伴有肠化生或异型增生为胃癌前病变(precancerous lesions of gastric cancer, PLGC)。据报道萎缩性胃炎、肠化生和异型增生发展为GC的概率分别是正常胃黏膜的2.8、3.4和6.5倍^[5]。因此,探索PLGC诊断及治疗靶点对于降低GC的发病率和死亡率具有重要意义。近年来,许多研究使用微阵列技术探索GC差异表达基因(differentially expressed genes, DEG),并用生物信息学方法分析其生物学过程、分子功能和细胞成分等途径,进一步确定GC相关的基因,这些基因可作为GC诊断或治疗的潜在生物标志物^[7-9]。然而PLGC相关生物信息学研究不够充分。

本研究从GEO数据库的两个微阵列数据集中确定PLGC的DEG。然后通过DAVID对DEGs进行GO富集分析和KEGG富集分析;利用GEPIA对Hub基因进行生存分析,并使用HPA在线数据库验证Hub基因的表达水平。

1 资料和方法

1.1 基因收集与筛选 从GEO数据库下载2个独立的PLGC基因表达谱微阵列数据集GSE55696(包括19个胃炎组织和39个PLGC组织样本)和GSE130823(包括31个胃炎和31个PLGC组织样本)(<https://www.ncbi.nlm.nih.gov/geo/>)。使用GEO2R筛选PLGC和胃炎样本之间的DEG(<http://www.ncbi.nlm.nih.gov/geo/geo2r>),FDR是通过使用Benjamini-Hochberg方法调整原始P值获得的,调整后P值(adjusted P value) <0.05 和 $|\log_2 \text{FC}|>1$ 的筛选标准的基因被视为DEG。GSE55696和GSE130823中的上调/下调基因重叠部分视为PLGC常见的DEG。

1.2 富集分析和PPI网络 使用DAVID(<https://david.ncifcrf.gov/>)数据库进行GO富集分析和KEGG富集分析。通过STRING数据库(<https://www.string-db.org/>)选择要求得分大于0.4的基因构建PPI网络,并选择全字符串网络为网络特征,然后使用Cytoscape 3.8.2软件对PPI网络进行分析和可视化。

1.3 Hub基因的生存分析和RNA测序表达 使用Kaplan-Meier绘图工具(www.kmplot.com)^[10]测

量Hub基因的预后价值,其中包括来自TCGA 6个数据集的875例GC患者的生存信息和基因表达谱,包括GSE14210、GSE15459、GSE22377、GSE29272、GSE51105和GSE62254。根据中位基因表达(高和低),GC患者被分为两组。使用Bioconductor软件包计算并绘制Kaplan-Meier生存图 and 95%CI和对数秩P值的危险比。Kaplan-Meier生存曲线是根据每个Hub蛋白的值来显示的,风险数显示在曲线下方。通过基因表达谱交互分析(GEPIA)网站验证DEGs的表达(<http://gepia2.cancer-pku.cn/#index>)。GEPIA提供了可定制的功能,如肿瘤/正常差异表达分析、根据癌症类型或病理分期进行分析、患者生存分析、相似基因检测、相关性分析和降维分析。

1.4 Hub基因的表达 使用人类蛋白质图谱(HPA)(<https://www.proteinatlas.org/>)研究Hub基因在胃癌组织中的表达和分布,并与正常组织进行比较,HPA显示了蛋白质水平对癌症患者生存率的影响。

1.5 统计学方法 PLGC基因表达谱微阵列数据集的DEG使用Benjamini-Hochberg方法调整原始P值,以调整后P值 <0.05 和 $|\log_2 \text{FC}|>1$ 视为DEG,生存分析使用Bioconductor包的“survplot”函数进行分析并制图。

2 结果

2.1 PLGC的差异表达基因 通过使用GEO2R在线工具分析微阵列数据集,本研究从GSE55696获得了1875个差异表达基因,包括802个上调基因和1073个下调基因,从GSE130823获得了1404个差异表达基因,包括663个上调基因和741个下调基因(图1)。绘制维恩图最终获得273个常见差异表达基因,包括PLGC组织中74个上调基因和199个下调基因(图2)。

2.2 GO和KEGG通路分析 273个DEGs的GO富集分析分为3个部分:细胞成分、分子功能和生物过程。按P值最小值排名靠前10个GO术语如图3所示。DEGs的BP主要富集于胚胎前肢形态发生、碳酸氢盐转运、前后模式规范、双细胞紧密连接组装和RNA聚合酶II启动子的转录。

KEGG分析表明DEG主要富集于6种信号途径,包括胃酸分泌、白细胞跨内皮迁移、甘油脂、代谢、细胞黏附分子(CAMs)、PI3K Akt信号途径、糖

酵解/糖异生(图4)。这表明10个基因(*EFNA3*、*CCND2*、*MYB*、*OSM*、*ITGB8*、*PIK3R3*、*PPP2R3A*、*PCK1*、*THBS2*、*EPOR*)在PI3K Akt信号通路中显著富集(图5)。

2.3 PPI网络 STRING数据库构建了由260个节点和1827条边组成的PPI网络。然后将PPI网

络导入Cytoscape软件进行可视化分析。该PPI网络的平均介数为477.44,平均度值为15.63(图6)。其中*CDX3*、*CXCL12*、*CFTR*等基因度值较高。

2.4 Hub基因的生存分析和表达 Kaplan-Meier平台用于研究10个潜在与PI3K-Akt通路相关Hub基因的预后价值。发现4个基因为危险基因,

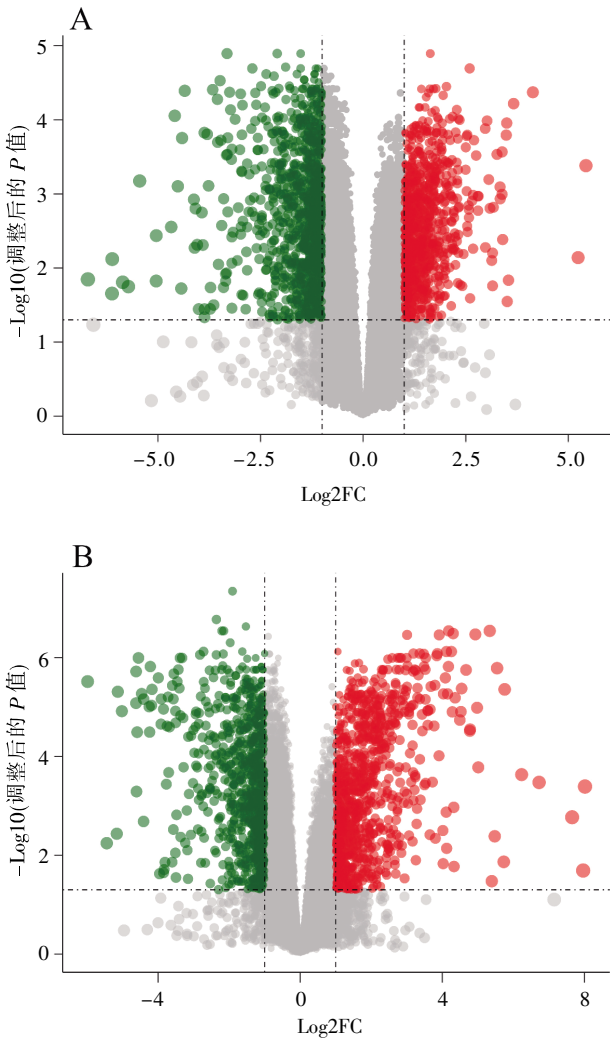


图1 两个数据集的DEGs

注:A为GSE55696,B为GSE130823;红点代表上调基因;绿点代表下调的基因。

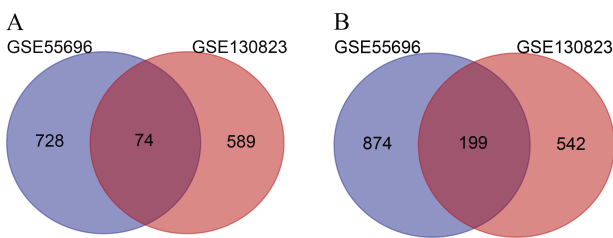


图2 差异基因 Venn图

注:A为74个共上调基因,B为199个共下调基因。

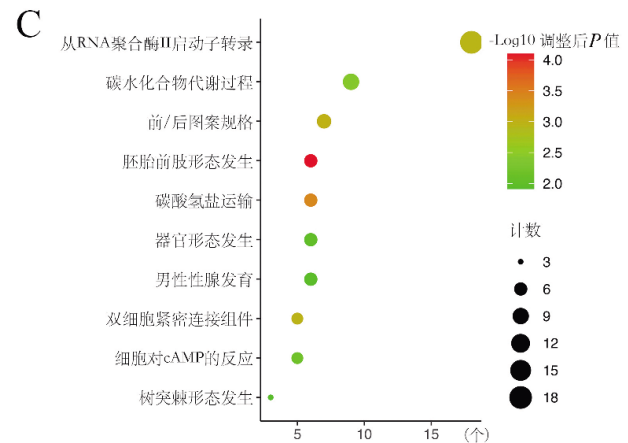
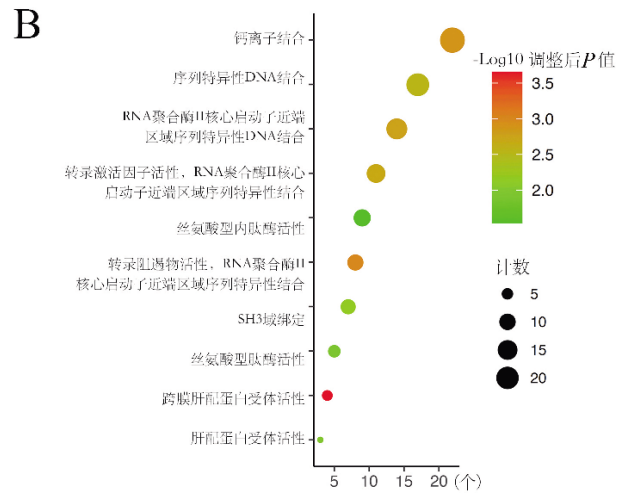
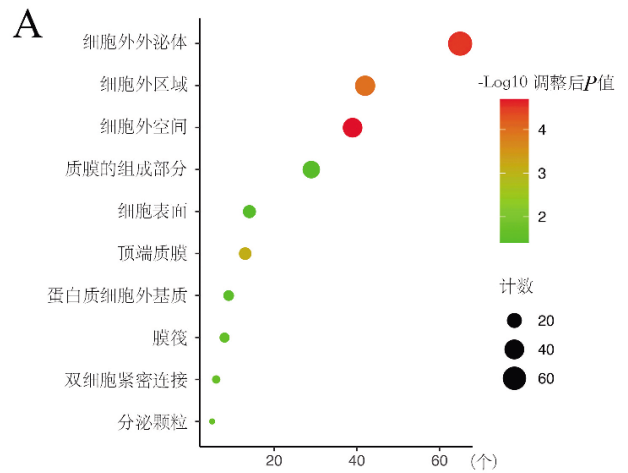


图3 DEG的GO富集分析

注:A为细胞成分,B为分子功能,C为生物过程。

包括 *EFNA3* ($HR=1.33, P=0.0016$)、*PPP2R3A* ($HR=1.39, P=0.00051$)、*THBS2* ($HR=1.55, P=1.2e-06$)、*EPOR* ($HR=1.52, P=3.8e-06$); 5个基因为保护基因, 包括 *CCND2* ($HR=0.74, P=0.0018$)、*MYB* ($HR=0.63, P=2.6e-07$)、*OSM* ($HR=0.65, P=0.00046$)、

ITGB8 ($HR=0.8, P=0.04$)、*PIK3R3* ($HR=0.78, P=0.0054$) (图7)。然后利用 GEPIA 平台检测 GC 组织与正常组织之间 Hub 基因的表达水平。发现 *EFNA3*、*CCND2*、*MYB*、*ITGB8*、*PCK1* 和 *THBS2* 等6个基因在胃癌组织中高表达; 和 *EPOR* 在 GC 组织中低表达 (图8)。

2.5 HPA 验证 Hub 基因 HPA 数据库用于检测 GC 组织样本和正常胃黏膜组织样本中的蛋白质表达水平。由于在 HPA 数据库中未发现 *OSM* 和 *EPOR* 的免疫组织化学信息, 本研究显示剩余8个蛋白的染色的结果。*EFNA3*、*ITGB8*、*PIK3R3*、*PCK1* 和 *THBS2* 在细胞质/膜组织中过度表达; 在细胞核中发现 *CCND2* 和 *MYB* 的过度表达。Hub 基因的蛋白质表达水平与 mRNA 表达结果一致 (图9)。

3 讨论

GC 的发病率和疾病负担逐年增加, 然而从

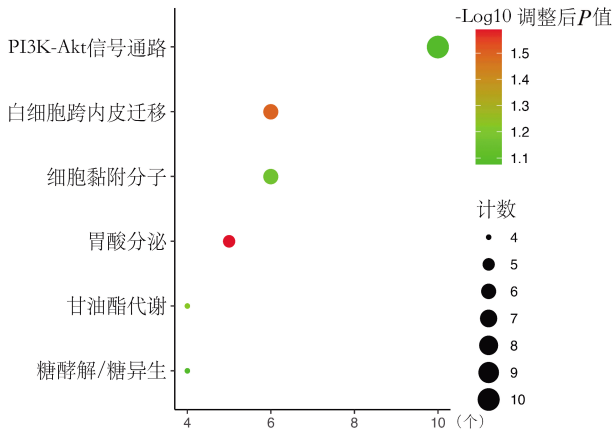


图4 DEG的KEGG途径富集分析

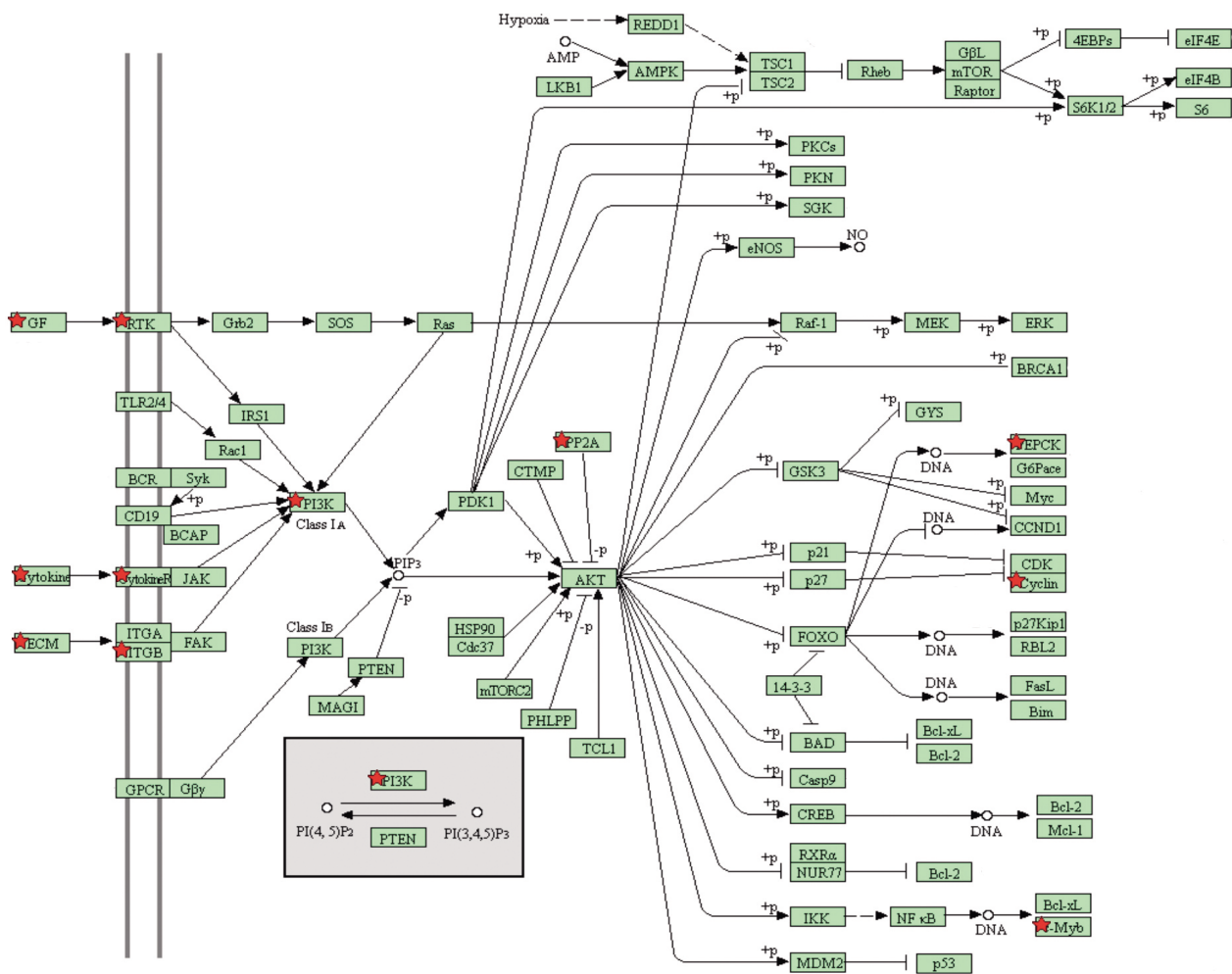


图5 PI3K Akt 信号通路中富集的 10 个基因

注:带红色五角星为 PLGC 中出现的 DEG。

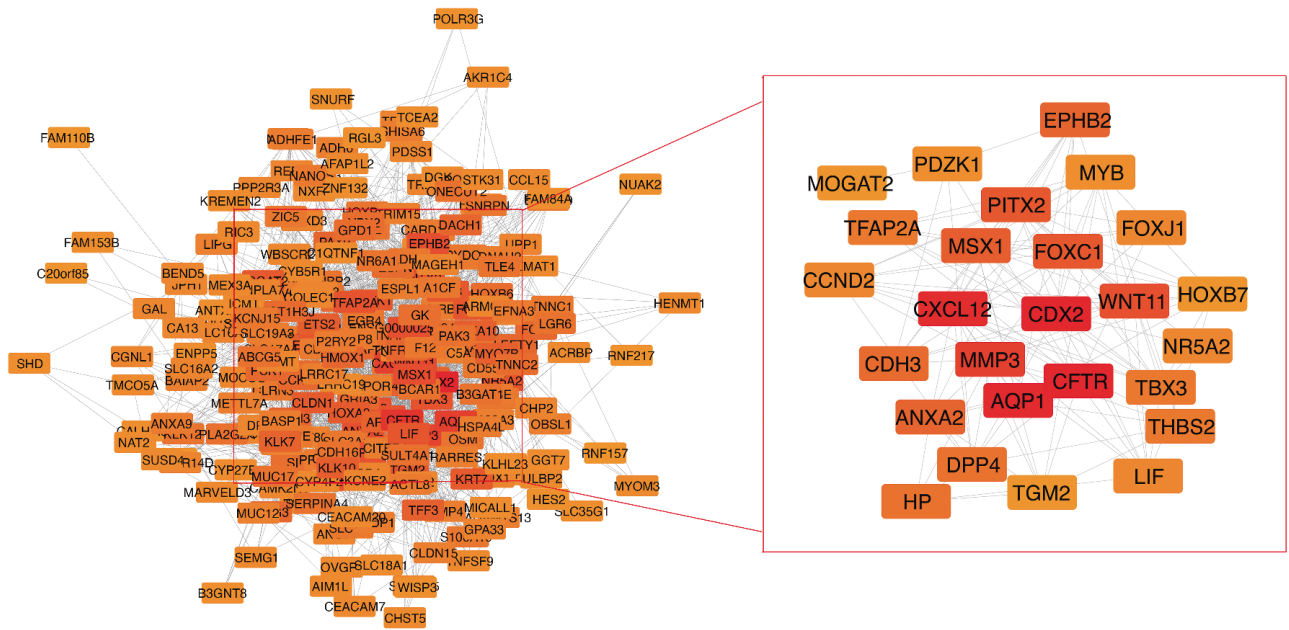


图6 DEGs的PPI网络

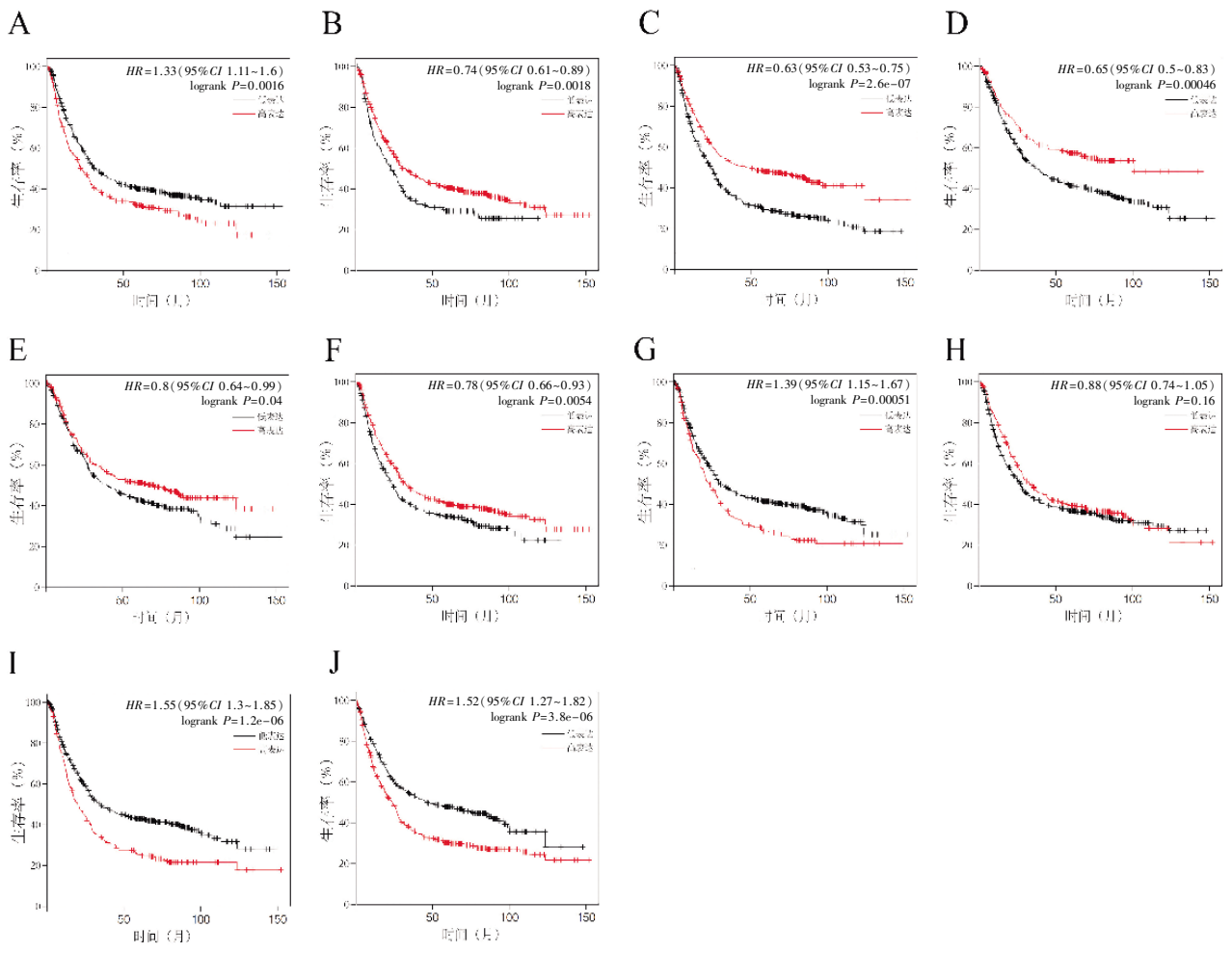


图7 胃癌核心基因表达和预后的存活曲线

注:A为EFNA3, B为CCND2, C为MYB, D为OSM, E为ITGB8, F为PIK3R3, G为PPP2R3A, H为PCK1, I为THBS2, J为EPOR。红色曲线表示高表达组,黑色曲线表示低表达组。

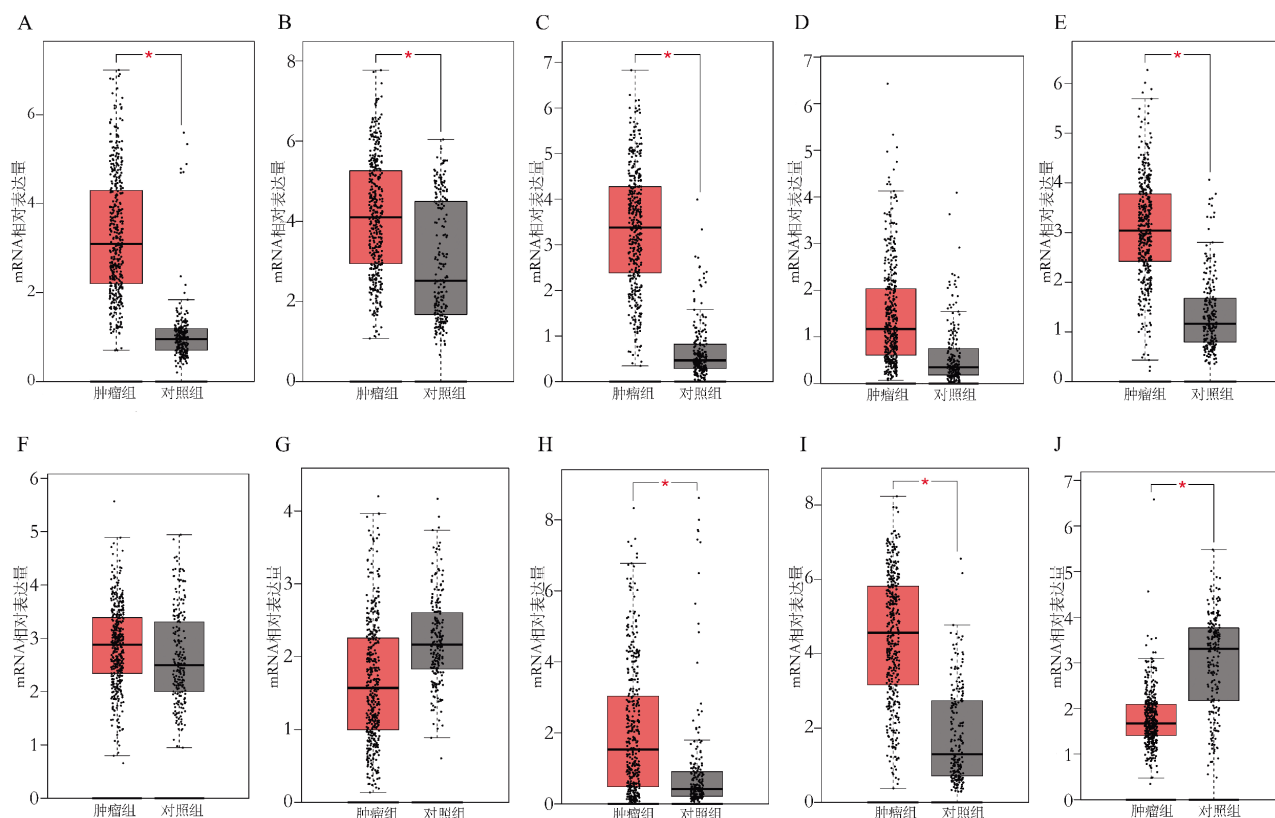


图8 基因 mRNA 表达水平的验证

注:A为EFNA3, B为CCND2, C为MYB, D为OSM, E为ITGB8, F为PIK3R3, G为PPP2R3A, H为PCK1, I为THBS2, J为EPOR。红色框表示GC样本,灰色框表示正常样本。

PLGC 到 GC 进展过程的相关生物学机制尚未完全阐明。生物信息学分析在癌症研究中起着重要作用,然而目前大多数研究仍集中在 GC 的诊断和治疗上,缺乏对 PLGC 的生物信息学研究。DNA 微阵列基因表达谱被广泛用于探索 DEGs 在肿瘤生长、诊断和治疗中的作用^[11],本研究通过两个基因表达谱探索 PLGC 的 DEGs,最终发现 273 个常见 DEG,包括 74 个上调基因和 199 个下调基因。细胞外基质是在癌症的所有特征中起着积极作用重要成分,GO 富集分析表明,DEGs 主要通过细胞外基质参与 PLGC 的过程^[12]。KEGG 通路分析表明,DEGs 主要参与胃酸分泌、白细胞跨内皮迁移、甘油酯代谢、细胞黏附分子(CAMs)、PI3K-Akt 信号通路和糖酵解/糖异生。胃黏膜萎缩→肠化生→异型增生过程中,胃酸分泌逐渐减少,其中以胃酸分泌减少为基础的胃泌素-17 测定已广泛应用于临床诊断^[13]。炎症过程涉及白细胞的渗出和聚集-激活后的吞噬和免疫-介导组织损伤,大量研究表明各种炎症因子参与 PLGC 发生发展过程,而白

细胞渗出是炎症反应的主要特征且与白细胞跨内皮迁移密切相关^[14,15]。已知 PI3K-Akt 信号通路具有广泛的生物学效应,参与调控细胞分裂、增殖、凋亡、代谢等活动过程,并且与 GC 细胞增殖、凋亡以及侵袭和转移有关^[16]。本研究结果表明这些信号通路也与 PLGC 密切相关。

本研究通过 PPI 网络鉴定出部分 Hub 基因。其中 HOX 可以通过激活癌基因促进 GC 细胞的增殖并促进 GC 细胞的侵袭和转移^[17]。CDX2 作为 H⁺/K⁺-ATP 酶的启动子不仅会导致胃黏膜肠上皮化生并最终发展为腺癌,而且 CDX2 作为 FOX3 的启动子,也可导致肠化生^[18]。B3GAT1 与人类蛋白质糖基化密切相关并参与糖基转移酶的编码^[19],糖基化过程中 HMGB1 和 RAGE 的高表达可能与 GC 的侵袭和转移正相关^[20]。AQP 是一种专门运输水分子的蛋白质家族,它可以帮助水分子通过特殊的漏斗状三级蛋白质结构快速通过胃肠道细胞膜。有研究发现 CAG 胃中 AQP3 mRNA 的表达显著降低,提示 AQP 与 PLGC 的进程有一定关联^[21]。

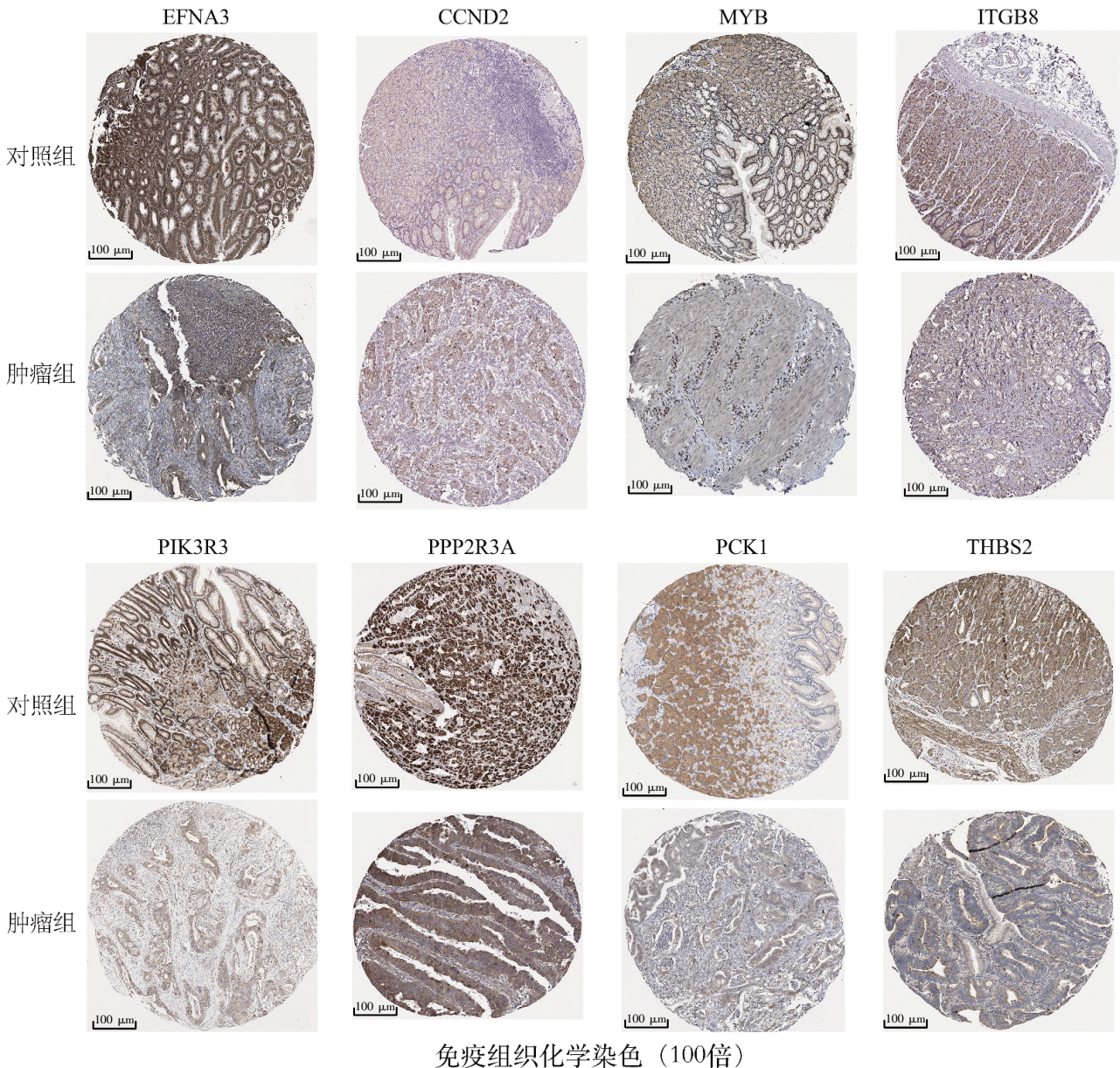


图9 HPA数据库中GC组织和非癌组织中Hub蛋白免疫组织化学染色情况

利用 Kaplan-Meier 和 GEPIA 数据库对 PLGC 中 DEGs 与 PI3K-Akt 信号通路交叉基因进行鉴定,发现 CCND2、MYB、OSM、ITGB8 和 PIK3R3 是 GC 的保护因子,提示这部分 DEGs 的低表达可能是 GC 进展的关键因素;而 EPOR 和 PPP2R3A 是 GC 的危险因素,表明这些基因的高表达可能是加重 GC 进展的因素。CCND2 的主要功能是促进细胞增殖,促进细胞周期从 G1 期进入 S 期,且由 CCND2 和 CDK2 形成的细胞周期蛋白 D1-CDK2 复合物可能在促进细胞增殖和肿瘤形成中发挥重要作用^[22]。抑癌基因 P53 可下调糖异生途径中 PCK1 和 G6PC 基因的表达,可通过抑制糖异生在

肿瘤抑制中发挥作用^[23]。THBS 家族基因通过多种靶蛋白的相互作用,参与血管生成、细胞运动与凋亡、细胞骨架形成和细胞外基质反应平台构建等多种生物学过程。沉默 THBS2 表达可促进胃癌细胞凋亡,抑制胃癌细胞的增殖、迁移和侵袭^[24]。作为 EPO 的特异性膜受体,EPOR 与肿瘤血管生成和胃癌进展呈正相关^[25]。而肿瘤细胞系中 EPOR 突变的存在可能导致 EPOR 调控的丧失,进而导致肿瘤细胞的无限增殖^[26]。

本研究首先筛选了 273 个可能与 PLGC 进展有关的 DEG。通过构建 PPI 网络和生存分析,鉴定出 PLGC 与 PI3K-Akt 信号通路的交叉基因,这有

助于研究 PLGC 的生物学机制和寻找新的诊断治疗靶点。由于目前的研究仍局限于生物信息学数据挖掘研究, 研究结果尚需通过实验进一步验证。

参考文献

- [1] SMYTH EC, NILSSON M, GRABSCH HI, et al. Gastric cancer[J]. *The Lancet*, 2020(8):635-648.
- [2] JOSHI SS, BADGWELL BD. Current treatment and recent progress in gastric cancer [J]. *CA Cancer J Clin*, 2021,71(3): 264-279.
- [3] SUNG H, FERLAY J, SIEGEL RL, 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.
- [4] CORREA P, SHIAO YH. Phenotypic and genotypic events in gastric carcinogenesis[J]. *Cancer Res*, 1994,54(7):1941s-1943s.
- [5] SONG H, EKHEDEEN IG, ZHENG Z, et al. Incidence of gastric cancer among patients with gastric precancerous lesions: observational cohort study in a low risk Western population[J]. *BMJ*, 2015,27;351: h3867.
- [6] KULASINGAM V, DIAMANDIS EP. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies[J]. *Nat Clin Pract Oncol*, 2008,5(10): 588-599.
- [7] XIE N, BAI Y, QIAO L, et al. ARL4C might serve as a prognostic factor and a novel therapeutic target for gastric cancer: bioinformatics analyses and biological experiments[J]. *J Cell Mol Med*, 2021,25(8): 4014-4027.
- [8] LI Y, WANG JS, ZHANG T, et al. Identification of New Therapeutic Targets for Gastric Cancer With Bioinformatics [J]. *Front Genet*, 2020,18(11):865.
- [9] NIE K, SHI L, WEN Y, et al. Identification of Hub genes correlated with the pathogenesis and prognosis of gastric cancer via bioinformatics methods [J]. *Minerva Med*, 2020,111(3): 213-225.
- [10] NAGY Á, MUNKÁCSY G, GYÓRFFY B. Pancancer survival analysis of cancer hallmark genes [J]. *Sci Rep*, 2021,11(1): 6047.
- [11] FAN X, SHI L, FANG H, et al. DNA microarrays are predictive of cancer prognosis: a re-evaluation [J]. *Clin Cancer Res*, 2010,16(2):629-636.
- [12] ANDREUZZI E, CAPUANO A, POLETTI E, et al., Role of extracellular matrix in gastrointestinal cancer-associated angiogenesis[J]. *Int J Mol Sci*, 2020,21(10):3686.
- [13] LI Y, ZHAO Y, LI Y, et al. Gastrin-17 induces gastric cancer cell epithelial-mesenchymal transition via the Wnt/ β -catenin signaling pathway[J]. *J Physiol Biochem*, 2021,77(1):93-104.
- [14] JU X, ZHANG H, ZHOU Z, et al. Tumor-associated macrophages induce PD-L1 expression in gastric cancer cells through IL-6 and TNF- α signaling [J]. *Exp Cell Res*, 2020, 396(2):112315.
- [15] LU G, TIAN S, SUN Y, et al. NEK9, a novel effector of IL-6/STAT3, regulates metastasis of gastric cancer by targeting ARHGEF2 phosphorylation [J]. *Theranostics*, 2021,11(5): 2460-2474.
- [16] BAGHERY SAGHCHY KHORASANI A, POURBAGHERI -SIGAROODI A, PIRSALEHI A, et al. The PI3K/Akt/mTOR signaling pathway in gastric cancer; from oncogenic variations to the possibilities for pharmacologic interventions [J]. *Eur J Pharmacol*, 2021,898:173983.
- [17] LI Y, ZHANG Z, YANG Y, et al. Long Noncoding RNA HOX Transcript Antisense RNA Gene rs17720428 Single Nucleotide Polymorphism Is Associated with Gastric Cancer Risk and Prognosis[J]. *Genet Test Mol Biomarkers*, 2020,24(1):38-46.
- [18] WANG Y, JIANG RK, WANG Q, et al. Silencing LINC01021 inhibits gastric cancer through upregulation of KISS1 expression by blocking CDK2-dependent phosphorylation of CDX2 [J]. *Mol Ther Nucleic Acids*, 2021,26(24):832-844.
- [19] SHARAPOV SZ, SHADRINA AS, TSEPILOV YA, et al. Replication of 15 loci involved in human plasma protein N-glycosylation in 4802 samples from four cohorts[J]. *Glycobiology*, 2021,31(2):82-88.
- [20] HUFFMAN JE, KNEZEVIC A, VITART V, et al. Polymorphisms in B3GAT1, SLC9A9 and MGAT5 are associated with variation within the human plasma N-glycome of 3533 European adults[J]. *Hum Mol Genet*, 2011,20(24):5000-5011.
- [21] ZHUO S, SUN M, BAI R, et al. Long intergenic non-coding RNA 00473 promotes proliferation and migration of gastric cancer via the miR-16-5 p/CCND2 axis and by regulating AQP3[J]. *Cell Death Dis*, 2021,12(5):496.
- [22] LI WC, WU YQ, GAO B, et al. MiRNA-574-3 p inhibits cell progression by directly targeting CCND2 in colorectal cancer[J]. *Biosci Rep*, 2019, 39(12): BSR20190976.
- [23] ZHANG P, TU B, WANG H, et al. Tumor suppressor p53 cooperates with SIRT6 to regulate gluconeogenesis by promoting Fox O1 nuclear exclusion [J]. *Proc Natl Acad Sci USA*, 2014,111(29): 10684-10689.
- [24] AO R, GUAN L, WANG Y, et al. Silencing of COL1A2, COL6A3, and THBS2 inhibits gastric cancer cell proliferation, migration, and invasion while promoting apoptosis through the PI3 k-Akt signaling pathway [J]. *J Cell Biochem*. 2018 Jun, 119(6):4420-4434.
- [25] NISWANDER LM, LOFTUS JP, LAINEY é, et al. Therapeutic potential of ruxolitinib and ponatinib in patients with EPOR-rearranged Philadelphia chromosome-like acute lymphoblastic leukemia [J]. *Haematologica*, 2021,106(10):2763-2767.
- [26] LI Y, LI M, WEI R, et al. Identification and Functional Analysis of EPOR + Tumor-Associated Macrophages in Human Osteosarcoma Lung Metastasis [J]. *J Immunol Res*, 2020 (18): 9374240.