

• 临床研究 •

结直肠癌差异表达及预后相关基因分析*

王思毓, 刘珊, 刘艳艳, 郭梦玲, 张思宇, 安倩, 蒋永新[△]

650118 昆明, 昆明医科大学第三附属医院 分子诊断中心/云南省肺癌研究重点实验室/教育部高原区域性高发肿瘤国际合作联合实验室

[摘要] 目的: 结直肠癌是发病率和病死率较高的消化系统恶性肿瘤, 但结直肠癌发病的分子机制尚不明确。本研究基于生物信息学分析鉴定结直肠癌的关键基因、信号通路及预后相关基因, 以更好地了解结直肠癌的分子机制, 为结直肠癌的诊断、预后评估和治疗提供有价值的分子标志物。方法: 从 GEO 数据库下载结直肠癌基因芯片数据集, 分析筛选出差异表达基因(differentially expressed gene, DEGs), 通过 GO 和 KEGG 富集分析对 DEGs 进行功能注释和通路分析, 构建蛋白质相互作用网络并筛选出结直肠癌的关键枢纽基因。对结直肠癌的关键枢纽基因进行生存预后分析, 找到与预后相关基因, 并验证预后相关基因的表达, 同时检测预后相关基因的启动子甲基化水平。结果: 共筛选出 256 个 DEGs, 其中 73 个表达上调, 183 个表达下调。DEGs 主要富集在细胞外基质降解、分子黏附等生物学过程, 胶原蛋白类细胞外基质、基底细胞膜等细胞组分, 受体配体相互作用、趋化因子活动等分子功能, PPAR 信号通路、氮代谢等信号通路。从 PPI 网络中筛选出 1 个连接最紧密的核心模块, 包含 13 个关键枢纽基因, 进一步生存预后分析结果显示 CXCL2 和 GNG4 的表达与结直肠癌患者总生存时间显著相关。验证 CXCL2 和 GNG4 在结直肠癌中的表达与 GEO 数据集中的表达情况一致。相比于正常对照组织, 结直肠癌组织中的 CXCL2 和 GNG4 启动子甲基化水平显著降低。结论: CXCL2 和 GNG4 是结直肠癌中的核心基因, 与结直肠癌预后显著相关, 其表达可能受启动子甲基化调控, 有望作为潜在的分子标志物用于结直肠癌的诊断、预后评估和治疗, 为后续研究提供新的方向。

[关键词] 结直肠癌; 生物信息学; 差异表达基因; 分子标志物

[中图分类号] R319; R735.3⁺5; R735.3⁺7 **[文献标志码]** A doi:10.3969/j.issn.1674-0904.2021.06.005

引文格式: Wang SY, Liu S, Liu YY, et al. Differentially expressed and prognostic genes in colorectal cancer[J]. J Cancer Control Treat, 2021, 34(6): 509-519. [王思毓, 刘珊, 刘艳艳, 等. 结直肠癌差异表达及预后相关基因分析[J]. 肿瘤预防与治疗, 2021, 34(6): 509-519.]

Differentially Expressed and Prognostic Genes in Colorectal Cancer

Wang Siyu, Liu Shan, Liu Yanyan, Guo Mingling, Zhang Siyu, An Qian, Jiang Yongxin

Molecular Diagnostic Center, Yunnan Provincial Key Laboratory on Lung Cancer, Joint Laboratory for International Cooperation on High-Incidence Tumors in Plateaus, Yunnan Cancer Hospital & The Third Affiliated Hospital of Kunming Medical University & Yunnan Cancer Center, Kunming 650118, Yunnan, China

Corresponding author: Jiang Yongxin, E-mail: 598104374@qq.com

This study was supported by National Natural Science Foundation of China (No. 17947530).

[Abstract] **Objective:** Colorectal cancer is a digestive system malignancy with high morbidity and mortality, but the molecular pathogenesis of colorectal cancer is still unclear. This study aims to better understand the molecular pathogenesis of colorectal cancer, and provide valuable molecular markers for the diagnosis, prognosis assessment and treatment of colorectal cancer, based on key genes, signaling pathways and prognosis-related genes of colorectal cancer identified by bioinformatics analysis. **Methods:** The colorectal cancer gene chip dataset was downloaded from the GEO database. Differentially expressed genes (DEGs) were analyzed. Functional annotation and pathway analysis of DEGs were analyzed by GO and KEGG analyses. Protein-protein interaction (PPI) network was established to identify hub genes of colorectal cancer. Survival and prognosis analyses were performed on those genes to locate prognostic genes and verify their expressions. Meanwhile, promoter methylation level of prognostic genes was detected. **Results:** Among 256 DEGs selected, 73 cases showed up-reg-

[收稿日期] 2020-11-29 **[修回日期]** 2021-03-23

[基金项目] * 国家自然科学基金(编号:17947530)

[通讯作者] [△] 蒋永新, E-mail: 598104374@qq.com

prognosis analyses were performed on those genes to locate prognostic genes and verify their expressions. Meanwhile, promoter methylation level of prognostic genes was detected. **Results:** Among 256 DEGs selected, 73 cases showed up-reg-

ulated expression, and 183 showed down-regulated. DEGs were mainly concentrated in biological processes such as extracellular matrix degradation and molecular adhesion, cell components such as collagen extracellular matrix and basal cell membrane, and molecular functions such as receptor ligand interaction and chemokine activity, as well as signaling pathways such as PPAR and nitrogen metabolism. One core module with the closest connection, which contained 13 key hub genes, was selected from PPI network. Further survival and prognosis analyses showed that *CXCL2* and *GNG4* expressions were significantly correlated to the overall survival time of colorectal cancer patients. It was verified that *CXCL2* and *GNG4* expressions in colorectal cancer was consistent with those in the GEO dataset. The levels of *CXCL2* and *GNG4* promoter methylation in colorectal cancer tissues were significantly lower than those in normal tissues. **Conclusion:** *CXCL2* and *GNG4* are core genes in colorectal cancer, and are significantly correlated to the prognosis of colorectal cancer. Their expressions may be regulated by promoter methylation, and are expected to be used as potential molecular markers for early diagnosis, prognosis assessment and treatment of colorectal cancer, providing a new direction for future studies.

[Key words] Colorectal cancer; Bioinformatics; Differentially expressed genes; Molecular marker

据最新统计数据,结直肠癌是全球第三大常见肿瘤和第二大肿瘤相关死因,每年有近 180 万新发病例和近 90 万与之相关的死亡,5 年总体存活率不足 50%^[1-3]。结直肠癌的预后在很大程度上取决于能否早期发现肿瘤。Liu 等^[4]发现,早期诊断的结直肠癌患者 5 年生存率 >90%,但由于患者早期无特殊症状,同时缺乏足够有效的诊断方法,结直肠癌往往在晚期才被发现。近年来随着手术及各种治疗方法的进步,结直肠癌患者的治疗疗效有了很大的提高,然而晚期结直肠癌患者的 5 年生存率仍很低,不足 12%^[5]。因此,迫切需要了解结直肠癌发生进展的分子机制,寻找有效的生物标志物用于结直肠癌的早期诊断、预后评估和治疗,以提高患者的存活率。

近年来基于基因芯片结合生物信息学的工具广泛用于研究复杂疾病的发病机制,是识别与肿瘤发生、进展、诊断和预后相关的生物标志物的有效途径^[6-7]。在本研究中,我们整合了从 GEO 数据库中筛选出的两个数据集,以获得结直肠癌特有的差异表达基因(differentially expressed genes, DEGs)。对 DEGs 进行功能和通路富集分析来揭示 DEGs 的功能和作用机制,构建蛋白质相互作用(protein-protein interaction, PPI)网络以确定与结直肠癌相关的关键枢纽 Hub 基因。利用在线工具 Kaplan-Meier plotter 和 GEPIA 数据库对筛选出的关键枢纽 Hub 基因进行生存分析,探讨 Hub 基因与结直肠癌预后的关系,筛选出与预后相关的核心基因,并对与预后相关的核心基因的表达水平进行验证,分析预后相关的核心基因的启动子甲基化水平,以期对结直肠癌的早期诊断、预后评估和治疗提供新的分子标志物。

1 资料和方法

1.1 数据收集

本研究的工作流程如图 1 所示。通过在美国国

家生物技术信息中心(National Center for Biotechnology Information, NCBI)平台下的 GEO 数据库(<https://www.ncbi.nlm.nih.gov/geo/>)中输入关键词“colorectal cancer”“Homo sapiens”和“expression profiling by array”来进行数据集搜索。数据集需要满足以下标准:1)数据来自临床研究,而不是侧重于细胞系或特定免疫细胞的研究;2)数据集的样本大小必须大于 100;3)比较对象必须是结直肠肿瘤组织和邻近的正常结直肠组织。最终从搜索结果中获得两个数据集 GSE87211^[8]和 GSE44076^[9],其中 GSE87211 数据集基于 GPL13497 平台(Agilent-026652 Whole Human Genome Microarray 4x44K v2),包含 203 例结直肠癌组织和 160 例正常对照组织。GSE44076 数据集基于 GPL13667 平台([HG-U219] Affymetrix Human Genome U219 Array)包括 98 例结直肠癌组织和 98 例正常对照组织。

1.2 数据预处理与 DEGs 筛选

使用 R 软件(3.6.3 版, <https://www.r-project.org/>)中的 GEOquery 软件包对 GEO 中的两个数据集进行标准化预处理。去除没有对应基因的探针集,根据注释信息进一步转换成相应的基因,同一基因有多个探针时计算其均值。R 软件中的 limma 软件包分析两数据集中结直肠癌组织和正常组织的 DEGs, DEGs 的筛选标准为: $|\log FC(\text{fold change})| > 2$ 和 $\text{adjusted } P < 0.01$ 。分别用 R 软件中 pheatmap 软件包和 ggplot2 软件包进行聚类热图和火山图的绘制分析。最后为降低结果中的假阳性率,使用 R 软件中的 VennDiagram 软件包对两个数据集取交集绘制 Venn 图,确定共有的 DEGs。

1.3 GO 功能注释和 KEGG 通路富集分析

为了揭示 DEGs 的功能,利用 Enrichr 数据库(<https://maayanlab.cloud/Enrichr/>)对 DEGs 进行

了基因本体 (Gene Ontology, GO) 注释和京都基因和基因组百科全书 (Kyoto Encyclopedia of Genes and Genomes, KEGG) 通路富集分析, 将富集结果用 R 软件可视化, 以同时满足 $P < 0.05$ 和 $FDR < 0.05$ 作为显著富集的标准。

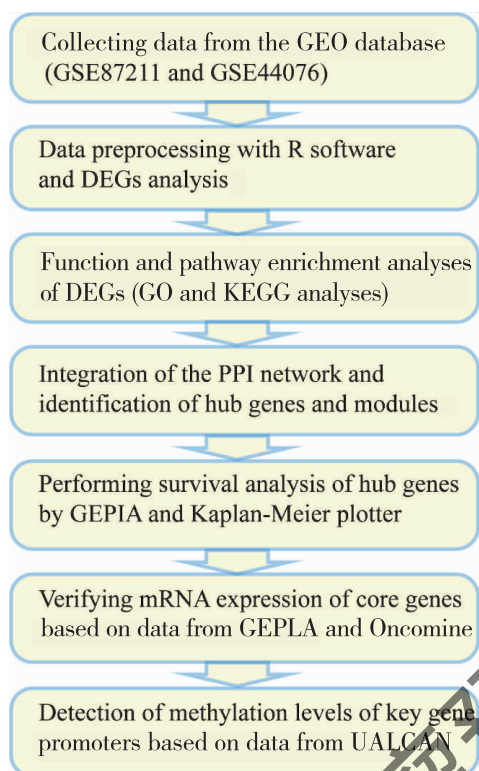


图 1 研究工作流程图

Figure 1. Flow of the Research

GEPIA, Oncomine and UALCAN are web resources.

GEO: Gene Expression Omnibus; DEGs: Differentially expressed genes; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; PPI: Protein-protein interaction; GEPIA: Gene Expression Profiling Interactive Analysis.

1.4 PPI 网络构建和 Hub 基因的筛选

通过向 STRING 数据库网站 (<http://string-db.org>) 上传 DEGs 来构建及可视化编码蛋白之间的 PPI。为了避免不准确的 PPI 网络, 我们设置所需的最低交互得分为 0.9 (cutoff ≥ 0.9 , 高置信度) 来获得显著的 PPI, 并去除与其他蛋白无相互作用的蛋白节点。采用 Cytoscape 软件中的 MCODE 插件进一步分析及可视化关键模块和 Hub 基因, Hub 基因筛选的标准为: MCODE 评分 > 10 分。

1.5 Hub 基因在结直肠癌中的生存分析

Kaplan-Meier plotter (<http://kmplot.com/analysis/>) 网站基于来自 GEO、EGA 和 TCGA 数据库的肿

瘤样本, 广泛应用于探索基因对 21 种肿瘤生存的影响。GEPIA (<http://gepia.cancer-pku.cn/index.html>) 提供了基于 TCGA 和 GTEx 数据库的多功能分析。应用 Kaplan-Meier plotter (含 165 名结直肠癌患者) 和 GEPIA (含 270 名结直肠癌患者) 来评估 Hub 基因表达与结直肠癌患者总生存期 (overall survival, OS) 的关系, 筛选与预后相关核心基因。根据 Hub 基因表达的中位数将结直肠癌患者分为高表达组和低表达组, 以基因表达量的下限 50% 和上限 50% 作为分析标准, Log-rank 检验 $P < 0.05$ 为差异有统计学意义。

1.6 与预后相关核心基因的 mRNA 表达验证

应用 Oncomine 数据库 (<https://www.oncomine.org>) 和 GEPIA 数据库对与预后相关的核心基因在结直肠癌样本和正常样本中的 mRNA 表达量进行验证及可视化。同时满足 $|\log FC (\text{fold change})| > 1$ 和 $P < 0.01$ 为差异有统计学意义。

1.7 核心基因的启动子甲基化水平检测

Ualcan 数据库 (<http://ualcan.path.uab.edu/index.html>) 是基于 TCGA 中的相关癌症数据进行基因甲基化、相关性分析等多功能分析数据库^[10]。应用 Ualcan 数据库, 对结直肠癌组织和正常对照组织中核心基因的启动子甲基化水平进行分析及可视化, 设置 $P < 0.05$ 为差异有统计学意义。

2 结 果

2.1 筛选出结直肠癌的 DEGs

从 GEO 数据库获取两个符合纳入标准的数据集, 其中 GSE87211 共纳入 363 例样本 (203 例结直肠癌组织和 160 例正常组织), GSE44076 共纳入 196 例样本 (98 例结直肠癌组织和 98 例正常组织)。根据筛选标准 $|\log FC (\text{fold change})| > 2$ 和 adjusted $P < 0.01$, 从 GSE87211 数据集中得到 853 个 DEGs (357 个上调基因和 496 个下调基因), GSE44076 数据集中得到 467 个 DEGs (119 个上调基因和 348 个下调基因)。数据集的前 50 个 DEGs 的聚类热图见图 2A、C, DEGs 的分布火山图见图 2B、D。Venn 图 (图 2E) 显示了两个数据集取交集的基因数, 共有 256 个共同的 DEGs, 其中包括 73 个表达上调基因和 183 个表达下调基因。

2.2 DEGs 的 GO 功能注释和 KEGG 通路富集分析

GO 富集分析结果显示, DEGs 在生物过程方面 (biological process, BP) 主要富集在细胞外基质降解、分子黏附等; 在细胞组分方面 (cellular compo-

ment, CC) 主要富集在胶原蛋白类细胞外基质、基底细胞膜等;在分子功能方面(molecular function, MF) 方面主要富集在受体配体相互作用、趋化因子活动等(图 3A ~ C)。KEGG 富集分析结果显示, DEGs

主要富集在 PPAR 信号通路、氮代谢等(图 3D)。同时分别对上调和下调的 DEGs 进行了富集分析, 结果见表 1 和表 2。

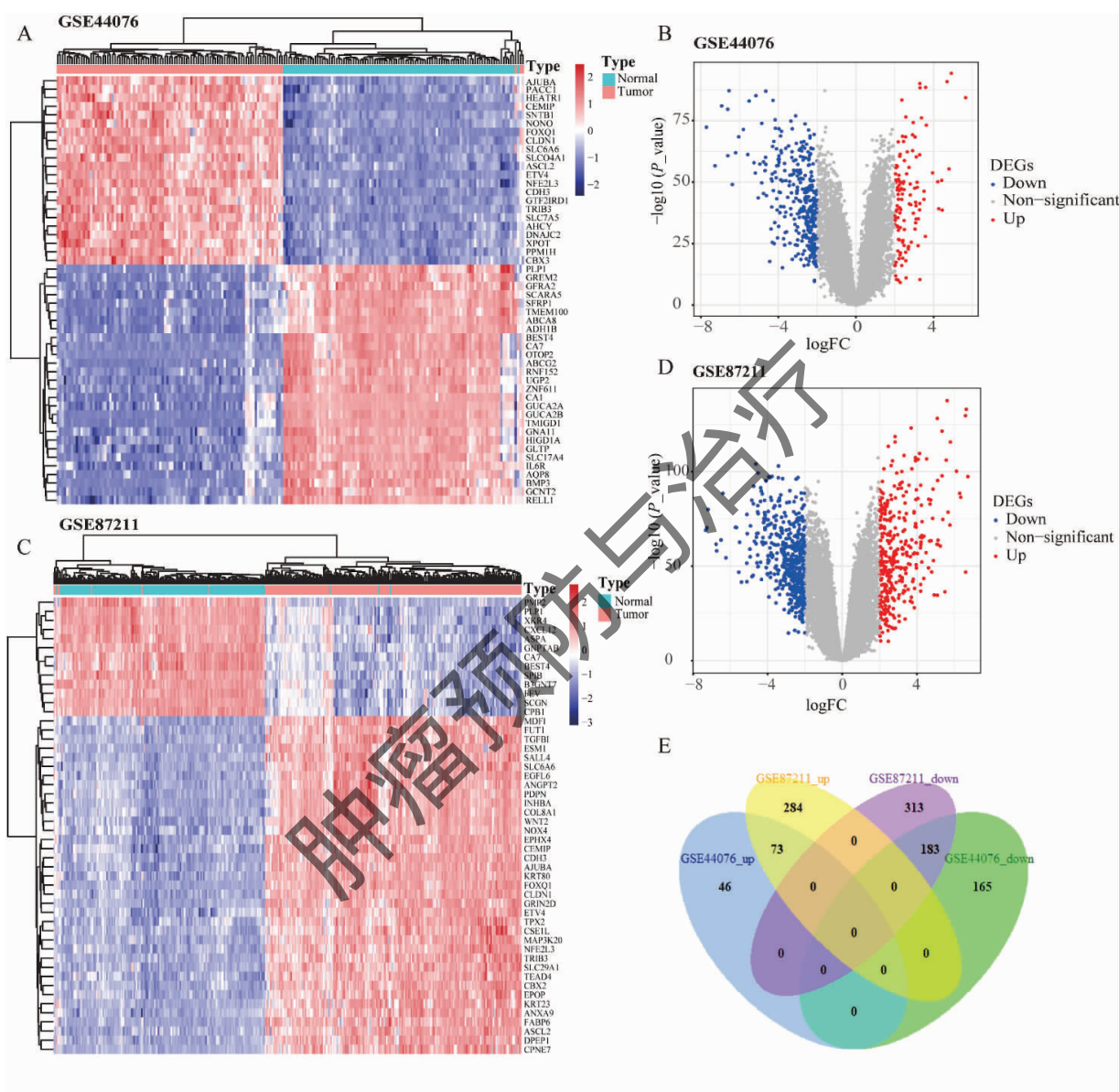


图 2 结直肠癌 DEGs

Figure 2. DEGs in Colorectal Cancer

A. Clustering heatmap of the top 50 most significant DEGs in GSE44076 dataset; B. Volcano plot of GSE44076 dataset; C. Clustering heatmap of the top 50 most significant DEGs in GSE87211 dataset; D. Volcano plot of GSE87211 dataset; E. The Venn diagram of overlapping DEGs from an intersection of the two datasets.

Abbreviations as indicated in figure 1.

2.3 构建 DEGs 的 PPI 网络和筛选 Hub 基因

将 DEGs 上传至 STRING, 设置 cut off ≥ 0.9 (高置信度), 构建 PPI 网络。去除与其他蛋白明显无相互作用的蛋白节点, 如图 4A 所示, PPI 网络中共含

64 个节点, 286 条相互作用的连线, PPI 网络富集 $P < 0.001$ 运用 Cytoscape 软件中的 MCODE 插件, 设置 MCODE 评分 > 10 分, 筛选得到连接最紧密的 1 个核心模块, 共包含 13 个 Hub 基因(图 4B)。

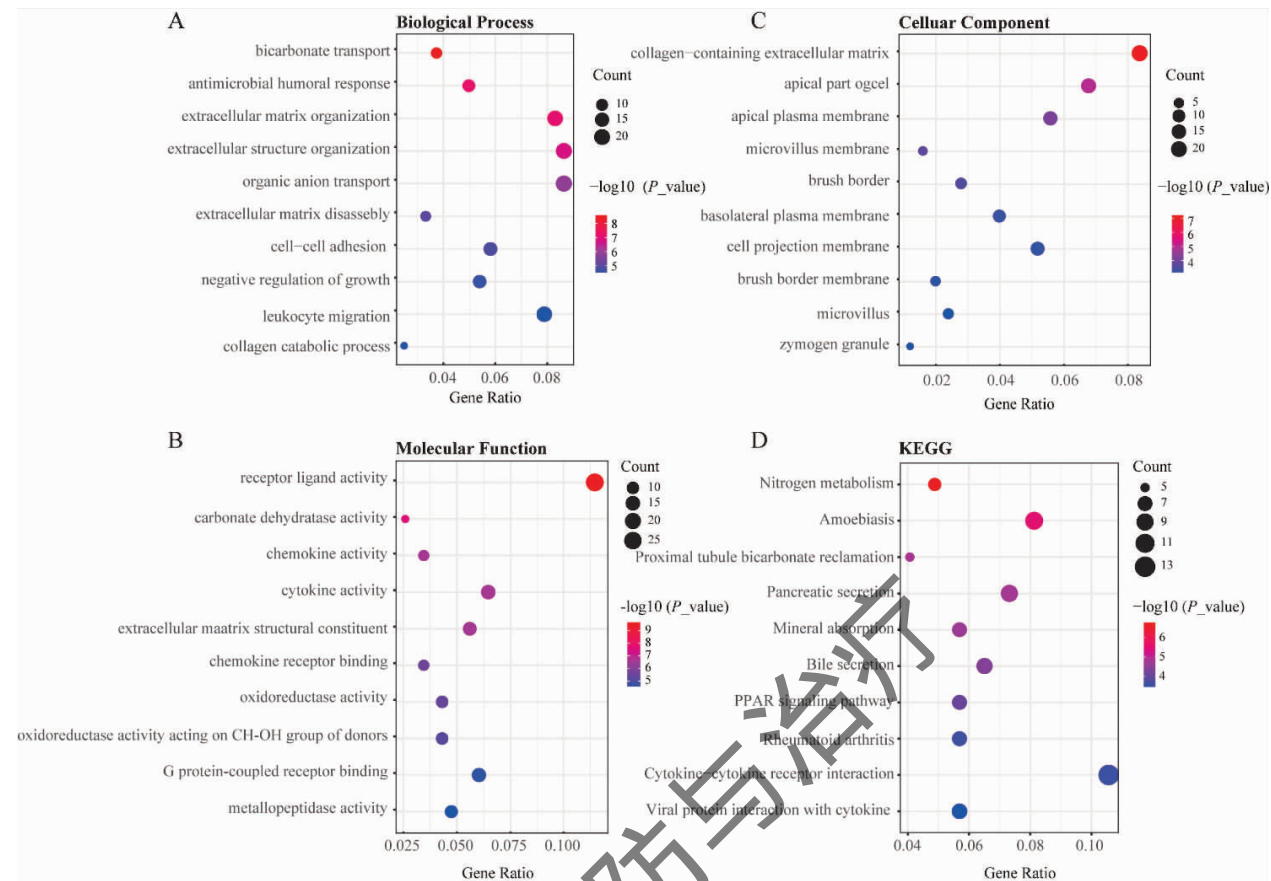


图3 DEGs 的 GO 和 KEGG 富集

Figure 3. GO and KEGG Enrichment of DEGs

A. Biological process; B. Molecular function; C. Cellular component; D. KEGG signaling pathway. Abbreviations as indicated in figure 1.

表1 上调 DEGs 的 GO 和 KEGG 富集

Table 1. GO and KEGG Enrichment of Up-Regulated DEGs

Category	ID	Description	Count	P
BP	GO:0030198	Extracellular matrix organization	14	<0.001
	GO:0043062	Extracellular structure organization	14	<0.001
	GO:0032963	Collagen metabolic process	8	<0.001
	GO:0030574	Collagen catabolic process	6	<0.001
	GO:0022617	Extracellular matrix disassembly	7	<0.001
	GO:0061844	Antimicrobial humoral immune response	6	<0.001
	GO:0043588	Skin development	10	<0.001
	GO:0042908	Xenobiotic transport	4	<0.001
	GO:0019730	Antimicrobial humoral response	6	<0.001
	GO:1990868	Response to chemokine	5	<0.001
	GO:0031528	Microvillus membrane	4	<0.001
	GO:0062023	Collagen-containing extracellular matrix	10	<0.001
CC	GO:0016323	Basolateral plasma membrane	7	<0.001
	GO:0045178	Basal part of cell	4	<0.001

(Table 1 continues on next page)

(Continued from previous page)

Category	ID	Description	Count	P
MF	GO:0031253	Cell projection membrane	7	<0.001
	GO:0005581	Collagen trimer	4	<0.001
	GO:0005902	Microvillus	4	<0.001
	GO:0005583	Fibrillar collagen trimer	2	<0.001
	GO:0098643	Banded collagen fibril	2	<0.001
	GO:0005788	Endoplasmic reticulum lumen	6	<0.001
	GO:0045236	CXCR chemokine receptor binding	4	<0.001
	GO:0005125	Cytokine activity	8	<0.001
	GO:0005201	Extracellular matrix structural constituent	7	<0.001
	GO:0048018	Receptor ligand activity	10	<0.001
	GO:0030546	Signaling receptor activator activity	10	<0.001
	GO:0008009	Chemokine activity	4	<0.001
	GO:0004222	Metalloendopeptidase activity	5	<0.001
	GO:0008237	Metallopeptidase activity	6	<0.001
	GO:0042910	Xenobiotic transmembrane transporter activity	3	<0.001
KEGG	GO:0042379	Chemokine receptor binding	4	<0.001
	hsa05323	Rheumatoid arthritis	6	<0.001
	hsa04657	IL-17 signaling pathway	6	<0.001
	hsa05146	Amoebiasis	6	<0.001
	hsa05134	Legionellosis	4	<0.001
	hsa05120	Epithelial cell signaling in Helicobacter pylori infection	4	<0.001
	hsa04061	Viral protein interaction with cytokine	4	<0.001
	hsa04933	AGE-RAGE signaling pathway in diabetic complications	4	<0.001
	hsa04064	NF-kappa B signaling pathway	4	<0.001
	hsa04062	Chemokine signaling pathway	5	0.001
	hsa05167	Kaposi sarcoma-associated herpesvirus infection	5	0.001

BP: Biological process; CC: Cellular component; MF: Molecular function; KEGG: Kyoto Encyclopedia of Genes and Genomes.

表2 下调 DEGs 的 GO 和 KEGG 富集

Table 2. GO and KEGG Enrichment of Down-Regulated DEGs

Category	ID	Description	Count	P
BP	GO:0015701	Bicarbonate transport	8	<0.001
	GO:0006730	One-carbon metabolic process	6	<0.001
	GO:0006821	Chloride transport	8	<0.001
	GO:0010273	Detoxification of copper ion	4	<0.001
	GO:1990169	Stress response to copper ion	4	<0.001
	GO:0061687	Detoxification of inorganic compound	4	<0.001
	GO:0097501	Stress response to metal ion	4	<0.001
	GO:0015698	Inorganic anion transport	4	<0.001
	GO:0034754	Cellular hormone metabolic process	9	<0.001
	GO:0042445	Hormone metabolic process	8	<0.001

(Table 2 continues on next page)

(Continued from previous page)

Category	ID	Description	Count	P
CC	GO:0005903	Brush border	7	<0.001
	GO:0098992	Neuronal dense core vesicle	3	<0.001
	GO:0031526	Brush border membrane	5	<0.001
	GO:0045177	Apical part of cell	13	<0.001
	GO:0016324	Apical plasma membrane	11	<0.001
	GO:0062023	Collagen-containing extracellular matrix	12	<0.001
	GO:0098862	Cluster of actin-based cell projections	7	<0.001
	GO:0031045	Dense core granule	3	0.003
MF	GO:0016616	Oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	10	<0.001
	GO:0004089	Carbonate dehydratase activity	5	<0.001
	GO:0016614	Oxidoreductase activity, acting on CH-OH group of donors	10	<0.001
	GO:0048018	Receptor ligand activity	17	<0.001
	GO:0030546	Signaling receptor activator activity	17	<0.001
	GO:0033764	Steroid dehydrogenase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	5	<0.001
	GO:0016229	Steroid dehydrogenase activity	5	<0.001
	GO:0018455	Alcohol dehydrogenase [NAD(P) ⁺] activity	3	<0.001
	GO:0004303	Estradiol 17-beta-dehydrogenase activity	3	<0.001
	GO:0016836	Hydro-lyase activity	5	<0.001
KEGG	hsa00910	Nitrogen metabolism	5	<0.001
	hsa04972	Pancreatic secretion	9	<0.001
	hsa04978	Mineral absorption	7	<0.001
	hsa04964	Proximal tubule bicarbonate reclamation	5	<0.001
	hsa04976	Bile secretion	7	<0.001
	hsa00140	Steroid hormone biosynthesis	5	<0.001
	hsa04960	Aldosterone-regulated sodium reabsorption	4	<0.001
	hsa00830	Retinol metabolism	5	<0.001
	hsa00982	Drug metabolism-cytochrome P450	5	0.001
	hsa00620	Pyruvate metabolism	4	0.002

Abbreviations as indicated in Table 2.

2.4 Hub 基因的生存预后分析

运用 GEPIA 和 Kaplan-Meier plotter 分析与绘制生存预后曲线可视化 Hub 基因与结直肠癌患者生存预后的关系,设置 Log-rank $P < 0.05$ 为差异有统计学意义。其中 *CXCL2* 和 *GNG4* 基因在两种网站的分析结果中均表现出与患者 OS 的显著相关性(图 5),认为是与患者预后相关的核心基因。高表达 *CXCL2* 的患者相较于低表达患者预后更好,高表达 *GNG4* 的患者相较于低表达患者预后更差。其余基因与患者 OS 无明显相关性,或在两种网站分析

中结果未达成一致(表 3)。

2.5 核心基因 *CXCL2* 和 *GNG4* 的 mRNA 表达验证

采用 GEPIA 与 Oncomine 两种数据库分别对核心基因 *CXCL2* 和 *GNG4* 在结直肠癌样本与正常对照样本中的 mRNA 表达进行验证,设置 $|\log_2FC| > 1$ 和 $P < 0.01$ 为差异有统计学意义,*CXCL2* 和 *GNG4* 在结直肠癌组织中均显著高表达(图 6A、B),与在 GEO 数据库筛选出的 GSE872211 和 GSE44076 数据集中的表达结果一致。

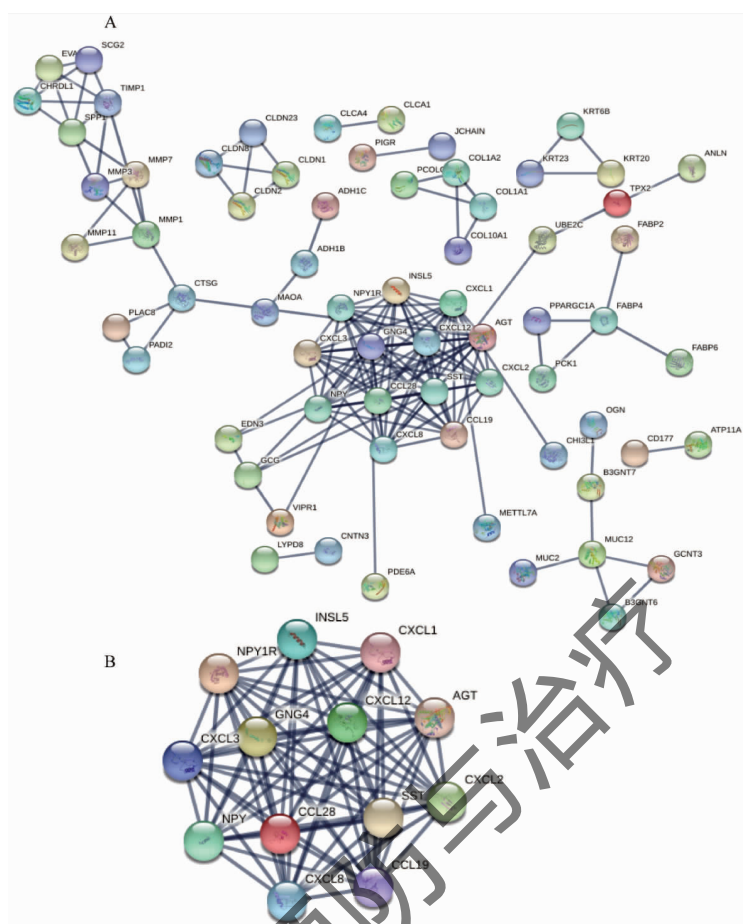


图 4 DEGs 的 PPI 网络 (A) 和核心模块的 Hub 基因 (B)

Figure 4. PPI Network of DEGs (Panel A) and the Hub Genes of Its Core Module (Panel B)

Abbreviations as indicated in figure 1.

表 3 基于 GEPIA 和 Kaplan-Meier plotter 的 13 个 Hub 基因的生存预后分析

Table 3. Overall Survival Analysis of 13 Hub Genes Based on GEPIA and Kaplan-Meier Plotter

Gene symbol	Overall survival by GEPIA (<i>P</i>)	Overall survival by Kaplan-Meier plotter (<i>P</i>)
<i>SST</i>	0.750	0.100
<i>CXCL3</i>	0.047 *	0.110
<i>INSL5</i>	NA	0.140
<i>CXCL8</i>	0.050	0.330
<i>CCL19</i>	0.610	0.070
<i>CCL28</i>	0.370	0.073
<i>NPYR1R</i>	0.710	0.310
<i>CXCL2</i>	0.042 *	0.037 *
<i>CXCL1</i>	0.130	0.110
<i>GNG4</i>	0.027 *	0.008 *
<i>CXCL12</i>	0.680	0.220
<i>NPY</i>	0.310	0.110
<i>AGT</i>	0.073	0.024 *

* *P* < 0.05

Abbreviations as indicated in figure 1.

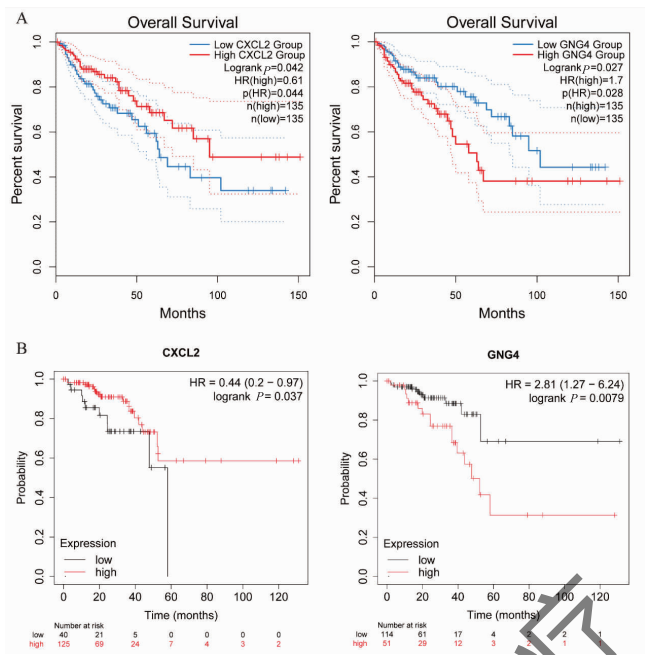


图5 基于 GEPIA (A) 和 Kaplan-Meier plotter (B) 绘制的预后相关核心基因的生存曲线
Figure 5. Survival Curves of Core Genes Related to Prognosis Based on GEPIA (Panel A) and Kaplan-Meier Plotter (Panel B)
Abbreviations as indicated in figure 1.

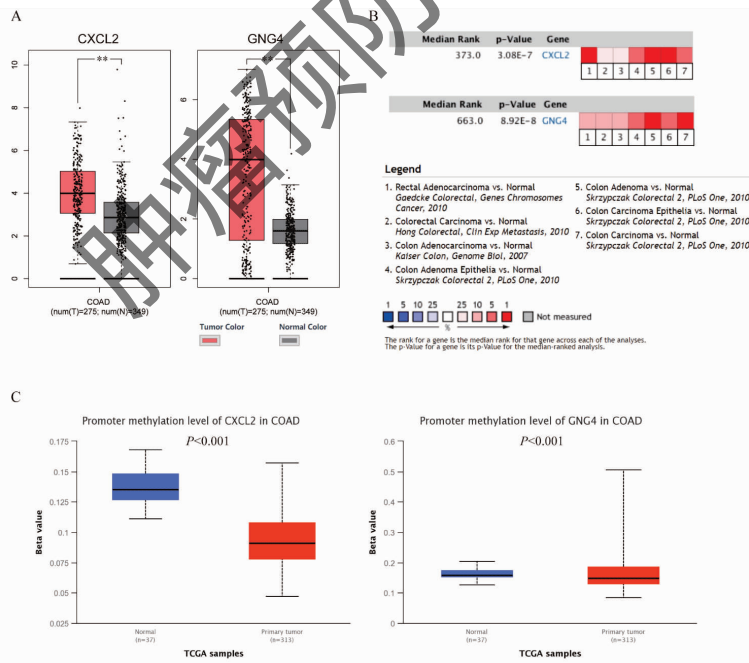


图6 基于 GEPIA 中 TCGA 和 GTEx 的数据、Oncomine 中的 7 个研究和 UALCAN 中的 TCGA 数据分析验证 CXCL2 和 GNG4 在结直肠癌和正常组织中的 mRNA 表达水平 (A、B) 和甲基化水平 (C)
Figure 6. mRNA Expression Levels (Based on 7 Studies from Oncomine, and TCGA and GTEx Data from GEPIA; Panel A and B) and Methylation Levels (Based on TCGA Data from UALCAN; Panel C) of CXCL2 and GNG4 in Colorectal Cancer and Normal Tissues
** $P<0.01$.

TCGA and GTEx are databases; GEPIA, Oncomine and UALCAN are web resources.
Abbreviations as indicated in figure 1.

2.6 核心基因 *CXCL2* 和 *GNG4* 的启动子甲基化水平

通过 Ualcan 分析 *CXCL2* 和 *GNG4* 的启动子甲基化水平, $P < 0.05$ 为差异有统计学意义。如图 6 C 所示, y 轴的 Beta 值表示甲基化水平从 0 (未甲基化) 到 1 (完全甲基化), *CXCL2* 的中位 Beta 值在正常组织中为 0.14, 结直肠癌组织中为 0.09, *GNG4* 的中位 Beta 值在正常组织中为 0.16, 结直肠癌组织中为 0.15。与正常组织相比, 结直肠癌组织的 *CXCL2* 和 *GNG4* 甲基化水平均显著降低。

3 讨论

晚期结直肠癌患者预后极差, 其 5 年生存率不足 12%^[5,11]。深入研究结直肠癌发生进展的生物学过程、分子机制、涉及的信号通路, 以发现有效的生物标志物用于结直肠癌的早期诊断、预后评估和治疗具有重要意义。

本研究从 GEO 数据库中筛选出含大量结直肠癌和正常对照组织的基因芯片的信息数据集 GSE87211 和 GSE44076, 共包含 559 例样本 (301 例结直肠癌样本和 258 例正常对照样本)。先对每个数据集筛选各自 DEGs, 再取交集得到共同的 256 个 DEGs (73 个表达上调基因和 183 个表达下调基因)。对 DEGs 进行 GO 和 KEGG 富集分析, 发现 DEGs 主要富集参与细胞外基质降解、分子黏附等生物过程; 存在于胶原蛋白类细胞外基质、基底细胞膜等细胞组分; 发挥受体配体相互作用、趋化因子活动等分子功能; KEGG 通路主要富集在 PPAR 信号通路、氮代谢等。由此, 我们对 DEGs 的功能、作用的信号通路有了更深入的了解。进一步对 DEGs 的 PPI 分析得到 13 个可能在结直肠癌的发生进展中起关键作用的 Hub 基因, 其中 *CXCL2* 和 *GNG4* 基因与结直肠癌患者的预后密切相关。

CXCL2 是趋化因子家族的成员, 属于内源性配体, 是一种小分子分泌蛋白^[12]。目前多项研究表明, *CXCL2* 作为癌基因或抑癌基因与多种肿瘤密切相关。Ding 等^[13]发现, 肝癌组织中 *CXCL2* 的表达水平显著低于配对正常肝组织, 与肝癌的不良预后相关。其过表达对肝癌细胞的增殖和生长有明显的抑制作用, 并能通过细胞核和线粒体凋亡途径诱导肝癌细胞凋亡, 通过 ERK1/2 信号通路对肝癌细胞的细胞周期进行负调控。此外 Subat 等^[14]对肝癌细胞 DNA 脱甲基化处理后 *CXCL2* 显著上调, 提示 *CXCL2* 的作用机制可能受甲基化调控。另有研究^[15]发现, *CXCL2* 的表达和 CD33 + 髓源性抑制细胞

(myeloid-derived suppressor cells, MDSCs) 的聚集在膀胱癌中明显增加, 且与疾病分期和预后不良显著相关。*CXCL2* 可通过介导膀胱癌中 MDSC 的趋化作用募集 MDSC, 进而提高肿瘤的免疫抑制能力, 促进肿瘤发生发展。在本研究中, 相较于正常组织, *CXCL2* 在结直肠癌中明显过表达。但有趣的是, 生存分析显示, *CXCL2* 高表达的结直肠癌患者表现出更好的生存预后。Acosta 等^[16]的报道指出, *CXCR2* 在癌前病变和早期癌症中表达升高, 而晚期癌症中表达缺失, 抑制 *CXCR2* 能抑制衰老和 DNA 损伤反应, 与恶性肿瘤发展中逃避衰老的观点一致。而 *CXCL2* 信号可通过 *CXCR2* 受体在肿瘤发生的早期加强衰老, 起到阻止肿瘤进一步恶性转化的作用, 从而抑制肿瘤生长, 或许可以解释 *CXCL2* 在结直肠癌发生和预后中的双重作用。

GNG4 是 G 蛋白复合体家族中的一员^[17]。Pal 等^[18]的研究表明, *GNG4* 是胶质母细胞瘤 (glioblastoma, GBM) 中甲基化程度最高、表达水平最低的基因之一, 通过甲基化抑制剂的处理, 可以逆转 GBM 细胞中 *GNG4* 转录水平的降低。过表达 *GNG4* 可抑制 GBM 细胞的增殖、集落形成、迁移和侵袭, 提示 *GNG4* 在 GBM 中具有潜在的抑癌作用。Kishibuchi 等^[19]发现, *GNG4* 在胸腺癌中也表现出高度甲基化, DNA 甲基化程度较高的胸腺癌患者的无复发生存率明显低于 DNA 甲基化水平较低的患者。另外, *GNG4* 的高表达与左半结肠癌患者的病理分期密切相关^[20], 还是影响直肠癌患者的新辅助放化疗疗效的重要因素^[21]。我们的研究发现 *CXCL2* 和 *GNG4* 在结直肠癌组织中过表达, 处于结直肠癌 PPI 网络中的核心地位, 对患者生存预后存在显著影响, 但其在结直肠癌中具体的作用及作用机制有待进一步的实验研究。

利用 GEPIA 和 Oncomine 中数据对 *CXCL2* 和 *GNG4* 在结直肠癌和正常组织中的 mRNA 表达进行验证, 结果与 GEO 数据库两数据集结果一致。另外, *CXCL2* 和 *GNG4* 在结直肠癌组织高表达, 同时表现出低甲基化水平, 提示 *CXCL2* 和 *GNG4* 的启动子甲基化可能是调控两基因表达的重要因素。

综上所述, 本研究通过生物信息学技术筛选获得了可能影响结直肠癌发生、进展、预后的核心基因, 为结直肠癌的发病和进展提供了新的见解, 为探索结直肠癌的发生发展的机制提供了新的思路, 并初步证实启动子甲基化是调控 *CXCL2* 和 *GNG4* 基因表达的重要因素。但他们能否在结直肠癌中发挥

具体作用以及要了解其具体作用机制需要进一步的实验探索。

作者声明:本文全部作者对于研究和撰写的论文出现的不端行为承担相应责任;并承诺论文中涉及的原始图片、数据资料等已按照有关规定保存,可接受核查。

学术不端:本文在初审、返修及出版前均通过中国知网(CNKI)科技期刊学术不端文献检测系统的学术不端检测。

同行评议:经同行专家双盲外审,达到刊发要求。

利益冲突:所有作者均声明不存在利益冲突。

文章版权:本文出版前已与全体作者签署了论文授权书等协议。

[参考文献]

- [1] Bray F, Ferlay J, Soerjomataram I, *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. *CA Cancer J Clin*, 2018, 68(6):394-424.
- [2] Arnold M, Sierra MS, Laversanne M, *et al.* Global patterns and trends in colorectal cancer incidence and mortality [J]. *Gut*, 2017, 66(4):683-691.
- [3] Jung G, Benítez-Ribas D, Sánchez A, *et al.* Current treatments of metastatic colorectal cancer with immune checkpoint inhibitors-2020 update [J]. *J Clin Med*, 2020, 9(11):3520.
- [4] Liu J, Liu Z, Li J, *et al.* Personalizing prognostic prediction in early-onset colorectal cancer [J]. *J Cancer*, 2020, 11(22):6727-6736.
- [5] Kow AWC. Hepatic metastasis from colorectal cancer [J]. *J Gastrointest Oncol*, 2019, 10(6):1274-1298.
- [6] Philibert A. The future of intelligence or the intelligence of the future? [J]. *Orthod Fr*, 2020, 91(1-2):139-144.
- [7] Gauthier J, Vincent AT, Charette SJ, *et al.* A brief history of bioinformatics [J]. *Brief Bioinform*, 2019, 20(6):1981-1996.
- [8] Hu Y, Gaedcke J, Emons G, *et al.* Colorectal cancer susceptibility loci as predictive markers of rectal cancer prognosis after surgery [J]. *Genes Chromosomes Cancer*, 2018, 57(3):140-149.
- [9] Moreno V, Alonso MH, Closa A, *et al.* Colon-specific eQTL analysis to inform on functional SNPs [J]. *Br J Cancer*, 2018, 119(8):971-977.
- [10] Chandrashekar DS, Bashel B, Balasubramanya SAH, *et al.* UALCAN: A portal for facilitating tumor subgroup gene expression and survival analyses [J]. *Neoplasia*, 2017, 19(8):649-658.
- [11] Arnold M, Abnet CC, Neale RE, *et al.* Global burden of 5 major types of gastrointestinal cancer [J]. *Gastroenterology*, 2020, 159(1):335-349.
- [12] Rajarathnam K, Schnoor M, Richardson RM, *et al.* How do chemokines navigate neutrophils to the target site: Dissecting the structural mechanisms and signaling pathways [J]. *Cell Signal*, 2019, 54:69-80.
- [13] Ding J, Xu K, Zhang J, *et al.* Overexpression of CXCL2 inhibits cell proliferation and promotes apoptosis in hepatocellular carcinoma [J]. *BMB Rep*, 2018, 51(12):630-635.
- [14] Subat S, Mogushi K, Yassen M, *et al.* Identification of genes and pathways, including the CXCL2 axis, altered by DNA methylation in hepatocellular carcinoma [J]. *J Cancer Res Clin Oncol*, 2019, 145(3):675-684.
- [15] Zhang H, Ye YL, Li MX, *et al.* CXCL2/MIF-CXCR2 signaling promotes the recruitment of myeloid-derived suppressor cells and is correlated with prognosis in bladder cancer [J]. *Oncogene*, 2017, 36(15):2095-2104.
- [16] Acosta JC, O'Loughlin A, Banito A, *et al.* Chemokine signaling via the CXCR2 receptor reinforces senescence [J]. *Cell*, 2008, 133(6):1006-1018.
- [17] Bonham LW, Evans DS, Liu Y, *et al.* Neurotransmitter pathway genes in cognitive decline during aging: Evidence for GNG4 and KCNQ2 genes [J]. *J Alzheimers Dis Other Dement*, 2018, 33(3):153-165.
- [18] Pal J, Patil V, Mondal B, *et al.* Epigenetically silenced GNG4 inhibits SDF1 α /CXCR4 signaling in mesenchymal glioblastoma [J]. *Genes Cancer*, 2016, 7(3-4):136-147.
- [19] Kishibuchi R, Kondo K, Soejima S, *et al.* DNA methylation of GHSR, GNG4, HOXD9 and SALL3 is a common epigenetic alteration in thymic carcinoma [J]. *Int J Oncol*, 2020, 56(1):315-326.
- [20] Song J, Yang J, Lin R, *et al.* Molecular heterogeneity of guanine nucleotide binding-protein γ subunit 4 in left- and right-sided colon cancer [J]. *Oncol Lett*, 2020, 20(6):334.
- [21] Palma P, Cano C, Conde-Muiño R, *et al.* Expression profiling of rectal tumors defines response to neoadjuvant treatment related genes [J]. *PLoS One*, 2014, 9(11):e112189.