



# 结肠腺癌诊断与预后相关关键基因的鉴定

周延杰<sup>1</sup> (通信作者\*), 蒋露<sup>2</sup>, 王珂<sup>1\*</sup>

(1. 江阴市人民医院 检验科, 江苏 江阴 214401; 2. 无锡市疾病预防控制中心 检验科, 江苏 无锡 214400)

**摘要:** **目的** 运用生物信息学技术鉴定新的用于结肠腺癌诊断和预后评估的生物学标志物。**方法** 从基因表达综合数据库(Gene Expression Omnibus, GEO)中下载3个结肠癌基因表达谱数据, 利用GEO2R工具对差异表达基因(Differential expression genes, DEGs)进行分析。采用WebGestalt在线工具进行差异表达基因功能及通路富集研究。采用SRING数据库和Cytoscape软件进行蛋白相互作用(PPI)网络构建并使用MCC算法从蛋白相互作用网络中筛选出前20个关键基因。采用GEPIA2在线工具从20个基因中筛选影响预后的基因。最后, 使用TCGA-COAD临床数据集验证核心基因的表达水平和诊断意义。**结果** 本研究共筛选了413个差异表达基因, 通过构建蛋白相互作用网络, 我进一步筛选出20个关键基因, 并通过生存分析选择了其中3个有预后价值的基因。最后, 用TCGA-COAD数据集的临床分析, 验证了这些核心基因的诊断价值。**结论** SPP1、GRP和GNGT1在结肠腺癌中高表达, 可作为诊断与预后评估的标志物。

**关键词:** 结肠腺癌; 生物标志物; 诊断; 预后; SPP1、GRP、GNGT1

**中图分类号:** R446.1, R735.3

**文献标识码:** A

**DOI:** 10.3969/j.issn.1671-3141.2022.99.012

**本文引用格式:** 周延杰, 蒋露, 王珂. 结肠腺癌诊断与预后相关关键基因的鉴定[J]. 世界最新医学信息文摘, 2022, 22(099): 60-66.

## Identification of Key Biomarkers Associated with Diagnosis and Survival Prognosis in Colon Adenocarcinoma

ZHOU Yan-ji<sup>1\*</sup>, JIANG Lu<sup>2</sup>, WANG Ke<sup>1\*</sup>

(1. Department of Clinical Laboratory, Jiangyin People's Hospital, Jiangyin Jiangsu 214400; 2. Department of Laboratory, Wuxi Center for Disease Control and Prevention, Wuxi Jiangsu 214023)

**ABSTRACT: Objective** To identify novel molecular biomarkers for the early diagnosis and survival prognosis of colorectal adenocarcinoma (COAD) by bioinformatics technology. **Methods** Three gene expression profile data were downloaded from the Gene Expression Omnibus(GEO), and the differential expression genes(DEGs) were identified by GEO2R. Gene Ontology (GO) and KEGG pathway enrichment analysis were conducted by WebGestalt online tool. String database and Cytoscape software were used for protein-protein interaction (PPI) network construction and module analysis. The top 20 Hub Genes were screened from the PPI network using MCC algorithm on CytoHubba app of Cytoscape software, and were verified by ONCOMINE database then. The core genes affecting CRC prognosis were screened by GEPIA2 survival analysis web tool. Finally, the expression level and clinical indicators including core genes was analyzed by TCGA-COAD dataset. **Results** In total, 413 differentially expressed genes (DEGs) were identified, and the GO and KEGG enrichment analyses of DEGs were processed. After, the protein-protein interaction (PPI) network was constructed and 20 hub genes were identified. Furthermore, three core genes were selected via survival analysis. Finally, the diagnostic and prognostic value of these core genes was verified by clinical analysis of TCGA-COAD dataset. **Conclusion** SPP1, GRP and GNGT1 were all over-expressed in COAD, and may be regarded as novel diagnostic and prognostic biomarkers for COAD.

**KEY WORDS:** colon adenocarcinoma; biomarker; diagnosis; prognosis; SPP1, GRP, GNGT1

**基金项目:** 无锡市科技发展医疗卫生指导性计划项目 (NZ2019042)。

**作者简介:** 共同通信作者\*: 周延杰, 王珂。

## 0 引言

结直肠癌(CRC)是世界上最常见的癌症类型之一, 约占所有确诊癌症病例的11%, 同时结直肠癌还是全球第二大最致命的癌症<sup>[1]</sup>。在我国, 结直肠癌的发病率呈逐年上升趋势, 且多数结直肠癌患者在确诊时已属于中晚期<sup>[2]</sup>。结肠腺癌(COAD)是结直肠癌最常见的类型, COAD的发病率和死亡率分别为10.2%和9.2%<sup>[3]</sup>。结直肠癌发生的确切机制尚不清楚, 因此, 更好地了解肿瘤发生的确切机制, 识别新的分子生物标志物, 用于早期诊断、预后评估、疾病监测至关重要。

高通量基因芯片技术已广泛应用于医学和生物学的多个领域, 通过该技术我们能够共享和探索各种分子机制<sup>[4]</sup>。本研究通过GEO2R工具鉴定了413个结直肠癌差异表达基因(DEGs), 对这些基因进行功能富集和蛋白互作关系研究, 并鉴定了20个关键基因。最后通过生存分析选择了3个具有预后价值的关键基因, 并用数据库中的临床数据进行了验证。

## 1 资料与方法

### 1.1 微阵列数据

基因表达综合数据库(GEO)(<http://www.ncbi.nlm.nih.gov/geo>)是一个公共功能基因组数据库、芯片和微阵列。我们从GEO下载了3个样本量较大的结直肠癌和正常上皮的基因表达谱, 分别为GSE21815、GSE35279和GSE71187。这三个数据集的微阵列数据均为GPL6480平台, 其中GSE21815数据集包含132例结直肠癌样本和9例正常上皮细胞样本, GSE35279数据集包含74例结直肠癌样本和5例正常上皮细胞样本, GSE71187数据集包含47例结直肠癌组织标本和12个正常结直肠组织标本。

### 1.2 差异表达基因的识别

基因表达综合数据库分析工具GEO2R(<http://www.ncbi.nlm.nih.gov/geo/geo2r>)被认为是一种交互式在线工具, 旨在比较GEO系列中的两个或多个数据集, 以便在不同实验条件下识别差异表达基因。我们采用GEO2R检测结直肠癌组织与正常上皮组织之间的差异, 选择 $P < 0.05$ 和 $|\log \text{fold change (FC)}| > 2$ 作为阈值, 将筛选出的数据使用“ggplot2” R语言包绘制差异表达基因的火山图及韦恩图。

### 1.3 功能富集分析

基因间交互关系检索工具WebGestalt(<http://www.webgestalt.org/>)是一个在线生物信息软件工具, 它集成了来自中央和公共管理的数据库以及计算分析的功能类别。我们用WebGestalt分析了差异表达基因的功能(GO)和信号转导通路(KEGG)途径富集分析。功能途径的富集分析可分为生物过程(BP)、细胞组分(CC)和分子功能(MF)。

### 1.4 PPI网络建设及关键基因筛选

基因间交互关系检索工具STRING(<http://string-db.org>)在线数据库可用来预测蛋白质相互作用网络、解释疾病的发生机制和发展。我们筛选了STRING数据库中差异表达基因蛋白质相互作用的综合得分 $> 0.7$ 的数据, 运用Cytoscape 3.7.1软件绘制了差异表达基因的蛋白互作网络, 并根据最大团系中心性(MCC)筛选出前20位的关键基因。随后我们使用GEPIA2(<http://gepia2.cancer-pku.cn/#index>)在线工具中的“生存分析”功能筛选了影响结直肠癌预后的关键基因。

### 1.5 关键基因的验证

为了验证这3个临床研究数据集筛选出的关键基因在结肠腺癌中普遍的诊断与预后价值, 我们使用TCGA-COAD数据库中的临床数据, 分析了所筛选的关键基因在结直肠腺癌和正常肠上皮组织中的表达差异及其诊断价值, 运用R4.0.4软件绘制了箱线图和受试者工作曲线。

## 2 结果

### 2.1 差异表达基因的鉴定

通过GEO2R在线工具,在基因表达谱数据处理和标准化后,以 $P < 0.05$ 和 $|\log FC| > 2$ 为截止标准,提取三个数据集中的差异表达基因(GSE35279中1357个、GSE21815中1493个和GSE71187中1884个)。火山图和维恩图是用R4.0.4中的“ggplot2”软件包绘制的。这三个数据集中重叠的差异表达基因共有413个(图1)。

### 2.2 差异表达基因的富集分析

为了阐明重叠差异表达基因的生物学功能,我们通过Webgalt在线工具进行了功能注释和通路富集分析。结果表明,生物过程功能富集分析中重叠的差异表达基因与组织发育、管腔形态发生、组织形态发生、管腔发育、骨骼系统发育、参与形态发生、动物器官形态发育、循环系统发育、细胞增殖和上皮发育显著相关(图2A)。在分子功能富集分析中,差异表达基因与受体调节活性、受体配体活性、RNA聚合酶II调控区序列特异性DNA结合、RNA聚合酶II调节区DNA结合、细胞外基质结构成分、转录调控区序列特异性DNA结合、序列特异性双链DNA结合、赋予拉伸强度的细胞外基质结构成分、激素活性、转录调控区DNA结合显著相关(图2B)。在细胞成分分析中,差异表达基因与含胶原蛋白、细胞外基质、基底膜、细胞外基质成分、胶原蛋白三聚体复合物、基底外侧质膜、质膜区、内质网腔超分子聚合物、超分子复合物相关(图2C)。此外,KEGG信号通路富集分析表明,这些差异表达基因在Wnt信号通路、糖尿病并发症中的AGE-RAGE信号通路、肿瘤的中心碳代谢、氨基酸生物合成、乳腺癌、疟疾、胃癌、ECM受体相互作用、TGF-beta信号通路、MiRNA调节通路中发挥着关键作用(图2D)。

### 2.3 PPI网络构建与关键基因筛选

为了评估差异表达基因间的相互关系,我们首先在STRING网站获取了这些差异基因的相互作用网络数据,当置信值为0.7作为临界标准,蛋白相互作用网络包含388个节点和334条边。我们用Cytoscape 3.7.1软件绘制了差异表达基因的蛋白互作网络(图3A),并最大团系中心性(MCC)筛选出前20位的关键基因:SST、PPBP、SPP1、GCG、GCG、INSL5、CXCL8、EDN3、CXCL11、PYY、P2RY1、HTR1D、GPR4、SAA1、IL6、PLCB1、APLN、GRP、ADRA2C、GNGT1、GAL(图3B)。

接下来我们使用GEPIA2在线工具分析了20个关键基因的预后作用,发现3个关键基因SPP1、GRP和GNGT1的表达水平与结肠腺癌(COAD)患者的生存显著相关(图4)。

### 2.4 关键基因的验证

最后,我们通过TCGA-COAD临床数据集验证了上述3个关键基因在结肠腺癌和正常肠上皮组织中的表达水平,并探讨了它们的诊断价值。结果表明,3个基因在临床病例中表达上调,且受试者工作曲线(ROC)显示这种差异表达具有诊断意义(图5)。

## 3 讨论

根据2018年的全球疾病负担研究,结直肠癌对人民的健康构成了重大威胁,它一直是全球第三大常见癌症。近十年来,人们对结直肠癌的发病机制的研究越来越多,其治疗策略不断发展。然而,由于原发肿瘤的术后复发和转移,结直肠癌的死亡率和发病率仍然很高。因此,识别新的分子靶点对结直肠癌的早期诊断、治疗和预后具有重要意义。

本研究通过GEO2R从三个基因芯片微阵列数据集中获得了结直肠癌和非癌变组织之间的差异表达基因,三个数据集共有413个共同

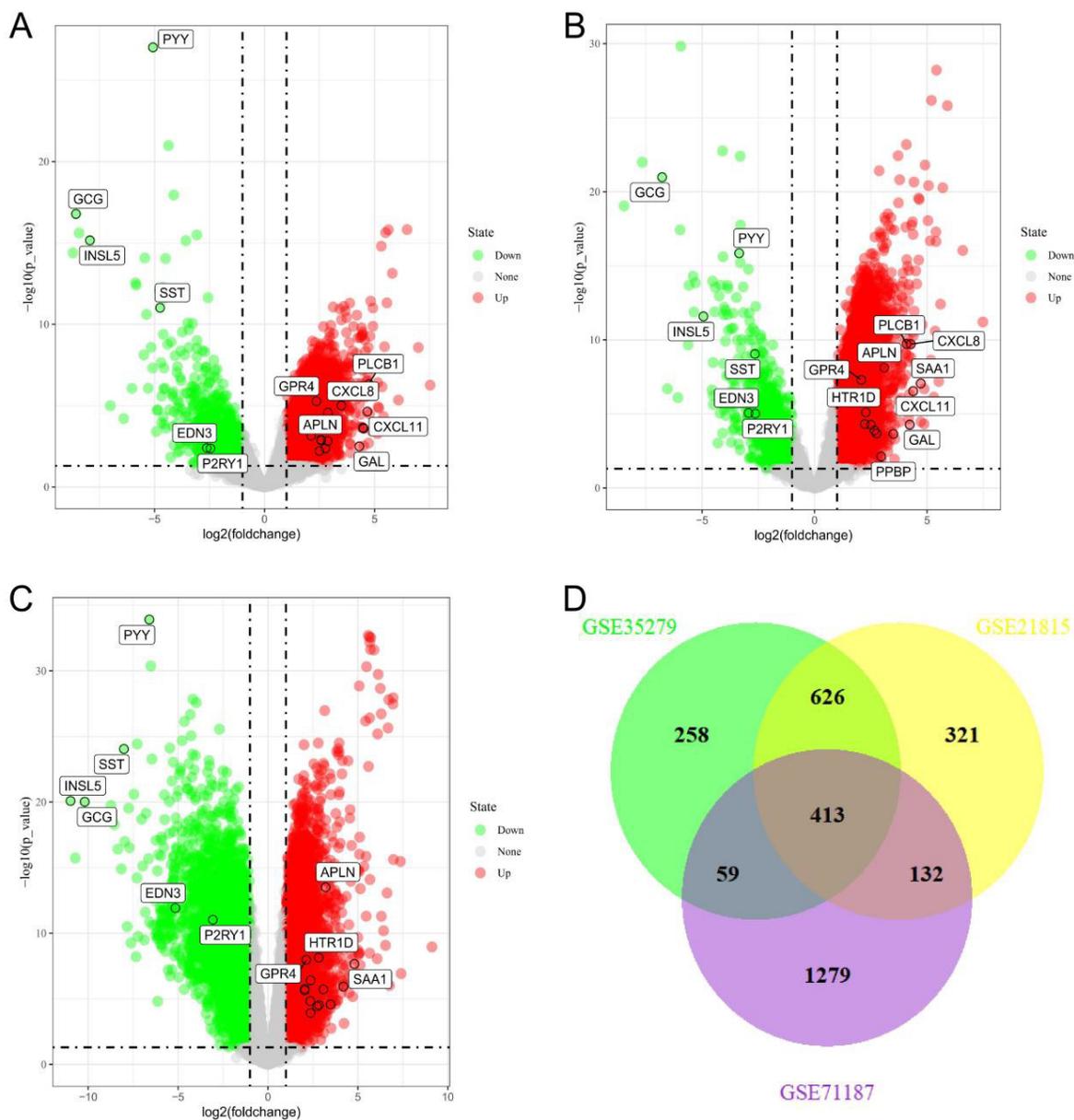


图 1 三个结肠癌研究基因表达数据集中的差异表达基因

的差异表达基因。通过基因功能和信号通路的富集分析，我们发现差异表达基因在与肿瘤相关的功能和通路中富集，如细胞增殖、细胞外基质结构成分、Wnt信号通路<sup>[5]</sup>、癌症中心碳代谢、ECM-受体相互作用<sup>[6]</sup>和TGF- $\beta$ 信号通路<sup>[7]</sup>，因此我们有理由相信，这些差异表达基因在结直肠癌的发生发展过程中发挥了重要的作用。

我们通过STRING在线工具构建了差异表达基因的蛋白相互作用网络，并通过CytoHubba筛选了最重要的20个基因，有通过

GEPIA的预后分析功能筛选出3个具有预后作用的基因：SPP1，GRP，GNGT1。最后通过TCGA-COAD数据库中的临床数据验证了这3个基因在直肠腺癌中呈高表达，且ROC曲线这种差异表达具有诊断价值。

分泌型磷蛋白1(SPP1，又称骨桥蛋白)是一种细胞外基质趋化因子样磷糖蛋白，促进细胞-基质相互作用，在各种恶性肿瘤中过表达，并在肿瘤发生和转移中发挥作用。Wei等人发现SPP1过表达可增强KRAS基因突变细胞的非锚定生长、细胞迁移和侵袭，过表达也可诱导

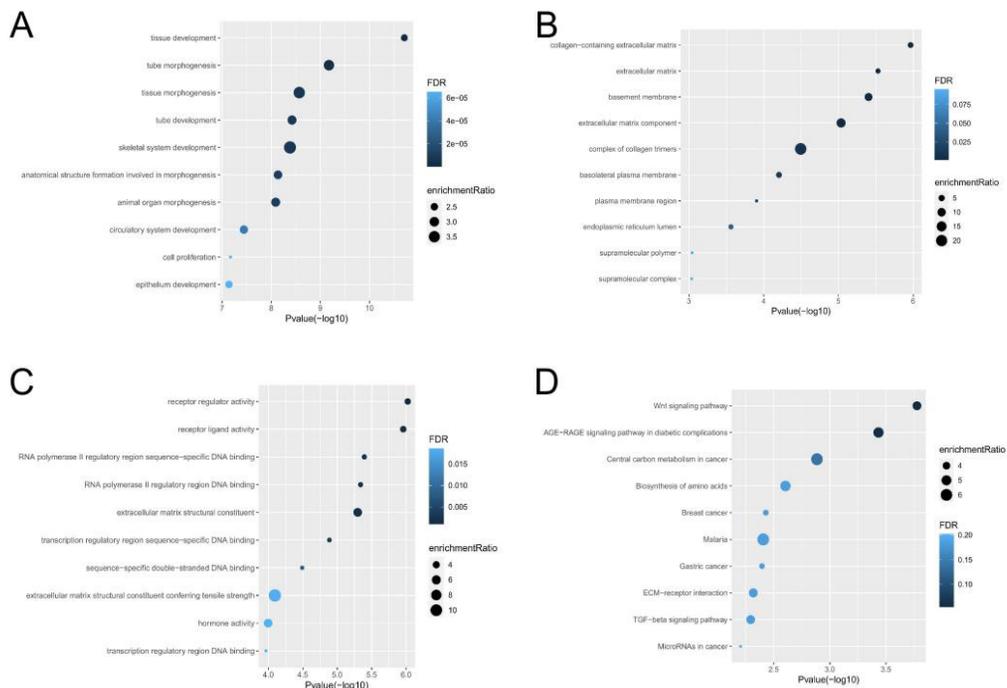


图 2 差异表达基因的功能及信号通路富集分析

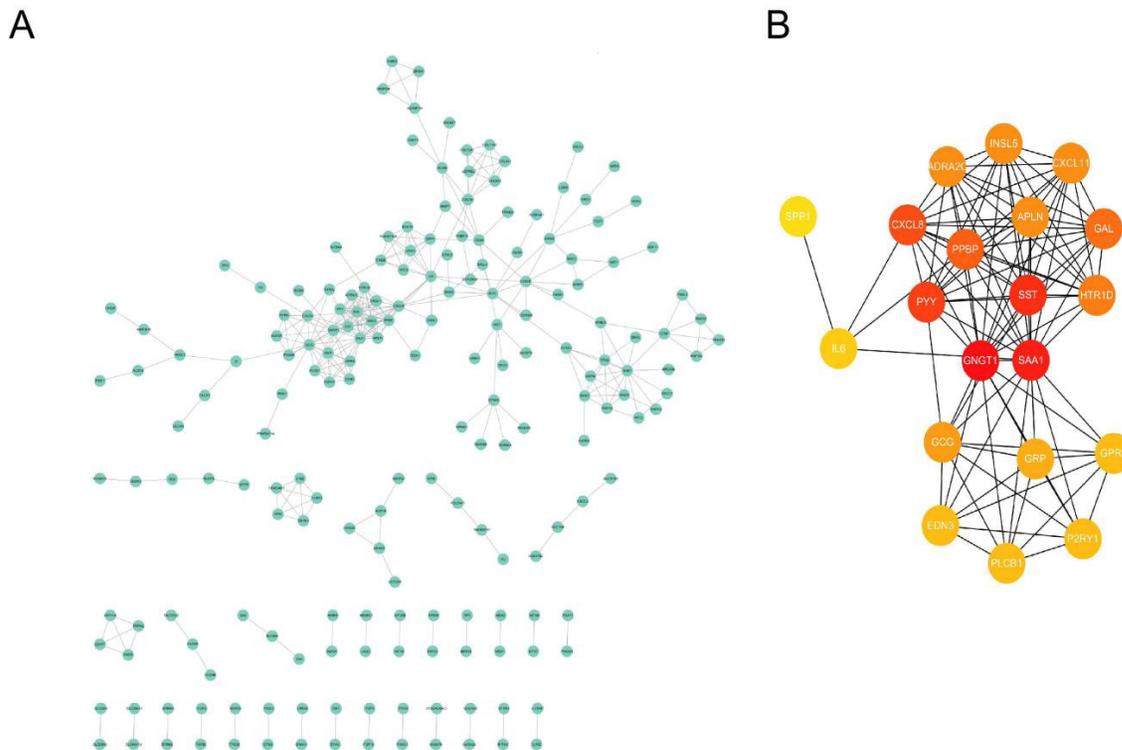


图 3 差异表达基因的蛋白 - 蛋白相互作用网路与其中的 20 个关键基因

PI3K信号转导和基质金属肽酶9(MMP9)的表达，并抑制KRAS突变细胞中E-钙连蛋白的表达<sup>[8]</sup>。Zeng等人发现SPP1的表达可激活整合素  $\beta 1$ /FAK/AKT通路，促进卵巢癌的进展<sup>[9]</sup>。Zhang等人发现SPP1可上调PD-L1，介

导肺癌中巨噬细胞极化，促进免疫逃逸<sup>[10]</sup>。

胃泌素释放肽(GRP)是一种与炸弹素具有同源性的调节肽。GRP及其受体在多种生理功能中发挥重要作用，包括感觉传递、中枢自主通路调节、体温调节、垂体激素分泌、胃和胰腺

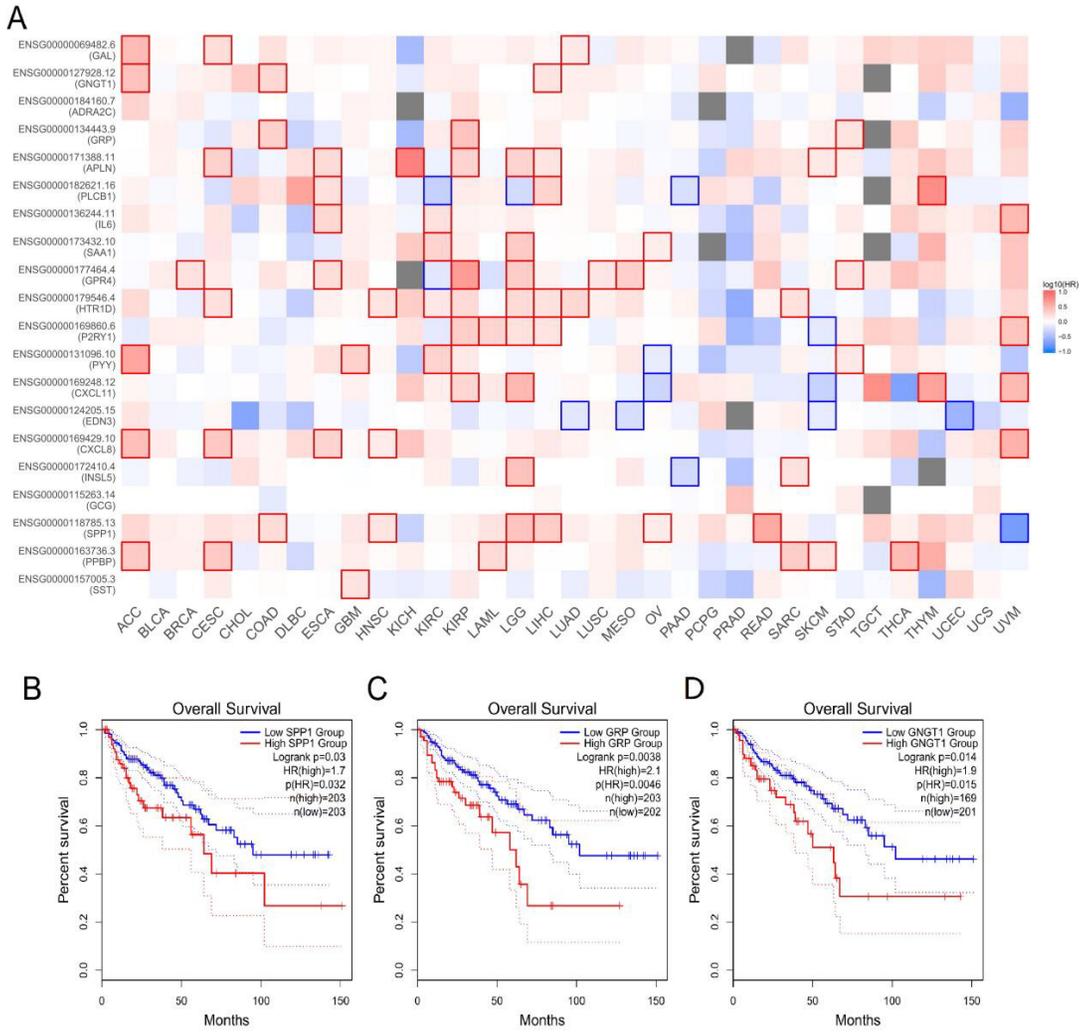


图 4 SPP1,GRP 与 GNGT1 与结肠腺癌的预后相关

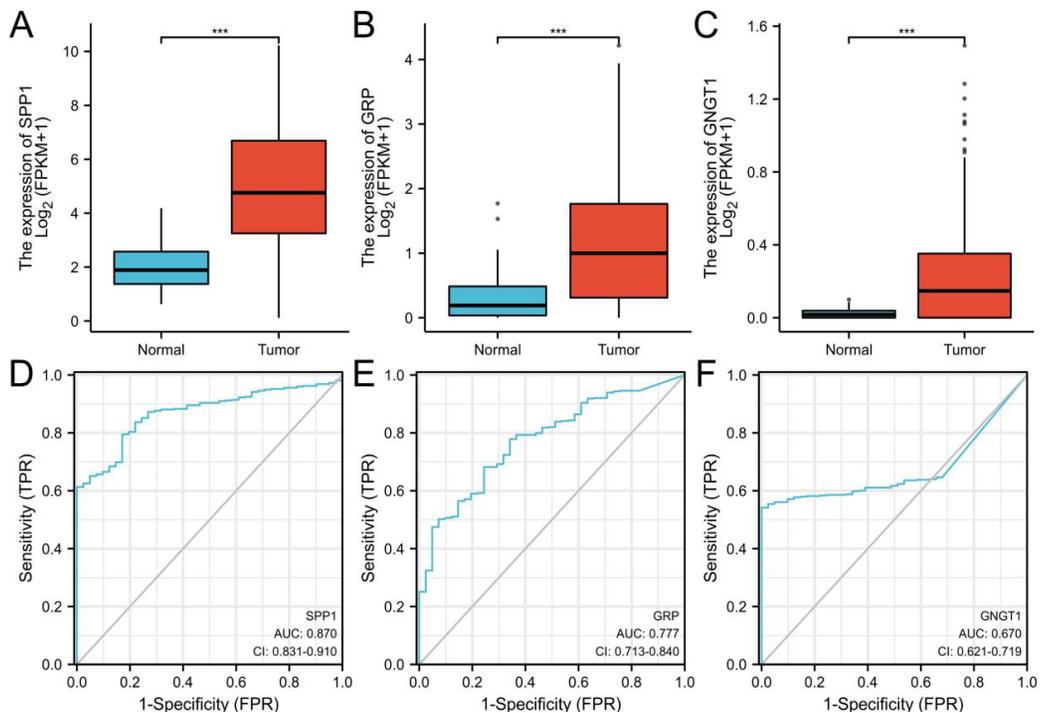


图 5 SPP1,GRP 与 GNGT1 在直肠腺癌组织中呈高表达且具有诊断价值

分泌、食物摄入和饱腹感。越来越多的证据表明, GRP在某些癌症中作为有丝分裂原、形态原和促血管生成因子<sup>[11]</sup>。Patel等人发现前GRP C端非酰胺化肽在结直肠癌细胞系和肿瘤中大量表达, 刺激结直肠癌细胞和正常结肠黏膜的增殖, 这些肽是新型结直肠癌治疗的有吸引力的靶点<sup>[12]</sup>。Ni等人发现GRP在有淋巴结转移的乳腺癌患者中高表达。此外, 在有淋巴结转移的患者中, 胃泌素释放肽表达量较高的患者生存时间较短。过表达胃泌素释放肽显著增强了细胞的侵袭性。相反, 通过短发夹RNA方法敲除胃泌素释放肽, 显著减少了MCF-7细胞的侵袭<sup>[13]</sup>。

鸟嘌呤核苷酸结合蛋白(G蛋白)  $\gamma$  转导活性肽1(GNGT1)编码异三聚体G蛋白  $\gamma$  亚基家族的一个成员( $\gamma$ 1)。Alsaleem等发现e-钙粘蛋白表达减少/丢失与差异表达基因相关, 包括调控GPIK3-AKT信号通路的GNGT1<sup>[14]</sup>。此外, Mucaki等发现GNGT1是一种对卡铂化疗有有效反应的基因<sup>[15]</sup>。

综上所述, 我们通过对GEO数据库中基因表达谱的综合生物信息学分析, 挖掘出了三个与结肠腺癌预后相关的核心分子, 包括SPP1、GRP和GNGT1。这些核心基因在结肠中呈高表达, 可被认为是新的COAD诊断和预后生物标志物。

#### 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] 王露尧,张鹭鹭.中国结直肠癌发病和死亡情况及防控策略[J].解放军医院管理杂志,2021,28(12):1195–1197.
- [3] Masoomi H, Ziogas A, Lin BS, et al. Population-based evaluation of adenosquamous carcinoma of the colon and rectum[J]. Dis Colon Rectum,2012, 55:509–14.
- [4] Dai GP, Wang LP, Wen YQ, et al. Identification of key genes for predicting colorectal cancer prognosis by integrated bioinformatics analysis[J]. Oncol Lett,2020,19:388–98.
- [5] Cheng X, Xu X, Chen D, et al. Therapeutic potential of targeting the Wnt/ $\beta$ -catenin signaling pathway in colorectal cancer[J]. Biomed Pharmacother,2019,110:473–81.
- [6] Walker C, Mojares E, Del R í o Hernández A. Role of Extracellular Matrix in Development and Cancer Progression[J]. Int J Mol Sci,2018,19(10):3028.
- [7] Itatani Y, Kawada K, Sakai Y. Transforming Growth Factor- $\beta$  Signaling Pathway in Colorectal Cancer and Its Tumor Microenvironment[J]. Int J Mol Sci,2019,20(23):5822.
- [8] Wei R, Wong JPC, Lyu P, et al. In vitro and clinical data analysis of Osteopontin as a prognostic indicator in colorectal cancer[J]. J Cell Mol Med,2018,22:4097–105.
- [9] Zeng B, Zhou M, Wu H, et al. SPP1 promotes ovarian cancer progression via Integrin  $\beta$ 1/FAK/AKT signaling pathway[J]. Onco Targets Ther,2018,11:1333–43.
- [10] Zhang Y, Du W, Chen Z, et al. Upregulation of PD-L1 by SPP1 mediates macrophage polarization and facilitates immune escape in lung adenocarcinoma[J]. Exp Cell Res,2017,359:449–57.
- [11] Patel O, Shulkes A, Baldwin GS. Gastrin-releasing peptide and cancer[J]. Biochim Biophys Acta,2006,1766:23–41.
- [12] Patel O, Clyde D, Chang M, et al. Pro-GRP-derived peptides are expressed in colorectal cancer cells and tumors and are biologically active in vivo[J]. Endocrinology,2012,153:1082–92.
- [13] Ni C, Zhao X, Sun T, et al. Role of gastrin-releasing peptides in breast cancer metastasis[J]. Hum Pathol,2012,43:2342–7.
- [14] Alsaleem M, Toss MS, Joseph C, et al. The molecular mechanisms underlying reduced E-cadherin expression in invasive ductal carcinoma of the breast: high throughput analysis of large cohorts[J]. Mod Pathol,2019,32:967–76.
- [15] Mucaki EJ, Zhao JZL, Lizotte DJ, et al. Predicting responses to platinum chemotherapy agents with biochemically-inspired machine learning[J]. Signal Transduct Target Ther,2019,4:1.