

• 临床研究 •

基于生物信息学分析胃癌与癌旁组织中差异基因的表达及验证*

段宝军, 王晓霏, 高艺, 白俊, 普彦淞, 靳荣辉, 张蓉, 路凯[△]

710068 西安, 陕西省人民医院 肿瘤内科(段宝军、白俊), 普外科(普彦淞), 放疗科(靳荣辉), 消化内科(张蓉), 麻醉科(路凯); 710061 西安, 西安交通大学 医学部(王晓霏、高艺)

[摘要] 目的: 分析胃癌(gastric cancer, GC)与癌旁组织间的差异表达基因, 寻找与 GC 的发生、发展及预后相关的关键基因。方法: 基于基因表达数据库(gene expression omnibus, GEO)筛选 GC 与癌旁组织间显著差异表达的基因, 利用京都基因与基因组百科全书(Kyoto encyclopedia of genes and genomes, KEGG)行通路富集分析。在癌症基因组图谱数据库(the cancer genome atlas, TCGA)中分析差异基因在 GC 中的表达及其与 GC 分期和预后的关系。最后在 GC 临床样本中验证。结果: 在 GEO 中共筛选出 1 353 个差异表达基因。对 GC 中显著高表达的 31 个基因进行 KEGG 分析发现, 其主要参与细胞外基质受体相互作用和蛋白消化与吸收。TCGA 中分析显示此通路中的 3 个基因(COL1A1、COMP 及 THBS2)在 GC 中显著高表达, 其中 COL1A1 高表达与 GC 的分期、预后不良显著相关。免疫组化结果证实 COL1A1 在 GC 中显著高表达, 其表达与 GC 患者的预后相关。结论: COL1A1 在 GC 中显著高表达, 其与 GC 的发生发展及预后有重要关系。

[关键词] 胃癌; 生物信息学; COL1A1; COMP; THBS2

[中图分类号] R735.2; R319 **[文献标志码]** A doi:10.3969/j.issn.1674-0904.2020.10.005

引文格式: Duan BJ, Wang XF, Gao Y, et al. Validation of differentially expressed genes in gastric cancer and paracancerous tissues based on bioinformatics[J]. J Cancer Control Treat, 2020, 33(10): 843-849. [段宝军, 王晓霏, 高艺, 等. 基于生物信息学分析胃癌与癌旁组织中差异基因的表达及验证[J]. 肿瘤预防与治疗, 2020, 33(10): 843-849.]

Validation of Differentially Expressed Genes in Gastric Cancer and Paracancerous Tissues Based on Bioinformatics

Duan Baojun, Wang Xiaofei, Gao Yi, Bai Jun, Pu Yansong, Jin Ronghui, Zhang Rong, Lu Kai
Department of Medical Oncology, Shaanxi Provincial People's Hospital, Xi'an 710068, Shaanxi, China (Duan Baojun, Bai Jun); Department of Surgery, Shaanxi Provincial People's Hospital, Xi'an 710068, Shaanxi, China (Pu Yansong); Department of Radiotherapy, Shaanxi Provincial People's Hospital, Xi'an 710068, Shaanxi, China (Jin Ronghui); Department of Digestive System Disease, Shaanxi Provincial People's Hospital, Xi'an 710068, Shaanxi, China (Zhang Rong); Department of Anesthesiology, Shaanxi Provincial People's Hospital, Xi'an 710068, Shaanxi, China (Lu Kai); Health Science Center, Xi'an Jiaotong University, Xi'an 710061, Shaanxi, China (Wang Xiaofei, Gao Yi)

Corresponding author: Lu Kai, E-mail: 281969461@qq.com

This study was supported by General Program of Natural Science Foundation of Shaanxi Province (NO. 2018JM7113); and by grants from Ministry of Finance of the People's Republic of China and Ministry of

Education of the People's Republic of China (NO. sxzy012019014), Xi'an Science & Technology Department (NO. 2017114SF/YX008) and Shaanxi Provincial People's Hospital (NO. 2017YX-11).

[收稿日期] 2020-04-21 **[修回日期]** 2020-07-05

[基金项目] * 陕西省自然科学基金面上项目(编号:2018JM7113); 中央高校基本科研业务费专项资金资助(编号: sxzy012019014); 西安市科技计划项目(编号: 201714SF/YX008); 陕西省人民医院科研基金项目(编号: 2017YX-11)

[通讯作者] [△]路凯, E-mail: 281969461@qq.com

[Abstract] **Objective:** To analyze differentially expressed genes (DEGs) in gastric cancer (GC) and paracancerous tis-

sues, and identify these vital genes which participate in the occurrence and progression of GC. **Methods:** Significant DEGs in GC and paracancerous tissues were screened in gene expression omnibus (GEO). Vital DEGs in GC were analyzed, and related pathways in Kyoto Encyclopedia of Genes and Genomes (KEGG) were investigated. The expression patterns of these vital DEGs in GC and normal/paracancerous tissues and their correlation with the stage and prognosis of GC were analyzed based on the Cancer Genome Atlas (TCGA). In addition, vital DEGs were validated in specimens of patents with GC. **Results:** 1,353 DEGs were identified in GEO, while 31 of these DEGs were up-regulated significantly in GC. KEGG analysis implied that 31 DEGs were associated with extracellular matrix receptor interaction, digestion and absorption of proteins. Based on TCGA analysis, 3 DEGs (*COL1A1*, *COMP* and *THBS2*) involved in these pathways were up-regulated significantly in GC compared to those in normal/paracancerous tissues; and *COL1A1* expression was significantly associated with pathological stage and prognosis of GC. Immunohistochemistry also validated that the expression of *COL1A1* was associated with the prognosis of patents with GC. **Conclusion:** *COL1A1* is significantly up-regulated in GC, which may play a key role in the occurrence, development and prognosis of GC.

[Key words] Gastric cancer; Bioinformatics; *COL1A1*; *COMP*; *THBS2*

在世界范围内,胃癌(gastric cancer, GC)的发病率在所有恶性肿瘤中居第 5 位,死亡率居第 3 位。我国每年 GC 的新诊断患者约 46 万人,发病率居所有恶性肿瘤的第 3 位,因 GC 死亡者约 39 万人,居所有恶性肿瘤的第 2 位^[1]。由于早期 GC 缺乏特异的症状和体征,导致超过 70% 的患者诊断时已处于进展期^[2]。虽然近年来在 GC 的诊断及治疗方面已有较多的进展^[3-4],但其发生发展的分子机制仍不十分清楚,研究与 GC 相关的基因异常改变对此疾病的诊断、治疗及预后评估具有重要的意义。

本研究通过检索基因表达数据库(gene expression omnibus, GEO)中的基因表达芯片数据,分析 GC 与癌旁组织样本中的差异表达基因(differentially expressed genes, DEGs),利用京都基因与基因组百科全书(Kyoto encyclopedia of genes and genomes, KEGG)对 GC 中显著高表达的差异基因进行通路富集分析,并在癌症基因组图谱数据库(the cancer genome atlas, TCGA)中研究差异基因在 GC 中的表达及其与 GC 分期及预后的关系,最后在病理组织样本中验证 *COL1A1* 在 GC 组织中的表达及其与 GC 患者预后的关系。

1 资料与方法

1.1 基因表达芯片的检索及差异基因筛选

在 GEO 数据库(<https://www.ncbi.nlm.nih.gov/gds>)中以“Gastric cancer”为关键词检索并筛选 GC 的基因表达芯片。通过计算各基因在 GC 与癌旁组织样本间表达的倍数改变值和 *P* 值得到显著差异表达的基因。

1.2 差异基因的 KEGG 通路富集与关键基因的确定

利用在线工具 Venny(<https://bioinfogp.cnb.csic.es/tools/venny/index.html>)将 GC 与癌旁组织

样本间差异表达的基因取交集,获得 GC 组织中显著高表达的基因。使用 Metascape 对 GC 中显著高表达的基因进行 KEGG 通路富集分析,在最显著富集的通路中确定 GC 中高表达的关键基因。

1.3 关键基因在 GC 中的表达及其与 GC 分期及预后的关系

通过在线工具 GEPIA(<http://gepia.cancer-pku.cn/>)分析 TCGA 数据库中 GC 样本中关键基因的表达差异,并进一步分析关键基因的表达水平与 GC 分期及预后的关系。

1.4 临床样本验证

GC 及癌旁组织标本($n = 113$),平均年龄:(52.31 ± 10.82)岁,男/女:64/49)来自陕西省人民医院胃肠道肿瘤生物样本库。所有患者有完整的治疗和随访资料,随访截止日期 2020 年 1 月 1 日,中位随访时间 46.4 月(11 ~ 120 月)。所有纳入病例的病理诊断均经两名不同的病理医生确定,病理分期依据美国癌症联合会第七版分期。*COL1A1* 抗体(No. 67288-1-Ig)购自 Proteintech 公司,二抗试剂购自北京中杉金桥公司。病理组织切片厚度为 4μm,经脱水、抗原修复、去除内源性过氧化物酶及山羊血清封闭后,滴加鼠抗人 *COL1A1* 一抗(1:100),4°C 过夜孵育,加生物素化二抗,DAB 显色,苏木素复染,透明后封片,镜下观察并拍照,Image J 软件通过测量免疫组化切片的光密度值计算阳性细胞面积百分比^[5]。

2 结果

2.1 GC 与正常/癌旁胃粘膜组织的基因表达

经检索、筛选后下载 2 个 GC 与癌旁组织样本的基因表达芯片,分别为 GSE54129(GC 组织 111 例,癌旁组织 21 例)和 GSE79973(GC 组织 10 例,癌

旁组织 10 例)。经过分析 GSE54129 的差异基因共 954 个,其中 GC 中高表达的 548 个,低表达的 406 个(图 1A)。GSE79973 的差异基因共 569 个,其中 GC 中高表达的 117 个,低表达的 452 个(图 1B)。

2.2 差异基因的 KEGG 通路富集及关键基因确定

在 GSE54129 和 GSE79973 中,GC 组织均高表

达的基因共 31 个,均低表达的基因共 54 个(图 1C、D)。KEGG 通路富集分析发现这 2 个芯片的 GC 组织中均高表达的 31 个基因主要富集于细胞外基质受体相互作用和蛋白的消化与吸收信号通路,此通路包括 3 个基因,分别为 *COL1A1*、*COMP* 及 *THBS2*,确定其为与 GC 有关的关键基因(图 2)。

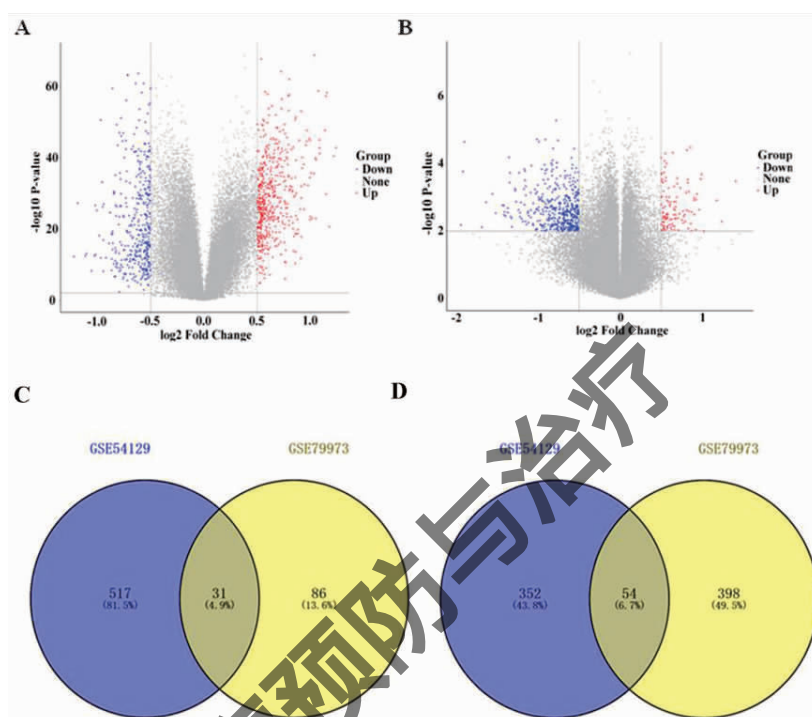


图 1 GC 与正常/癌旁组织的差异表达基因及韦恩图。

Figure 1. DEGs in Gastric Cancer and Normal/Paracancerous Tissues and DEGs in GSE54129 and GSE79973

A. Volcano plot for DEGs in GSE54129; B. Volcano plot for DEGs in GSE79973; C. Up-regulated DEGs; D. Down-regulated DEGs.

DEGs: differentially expressed genes; GC: gastric cancer.

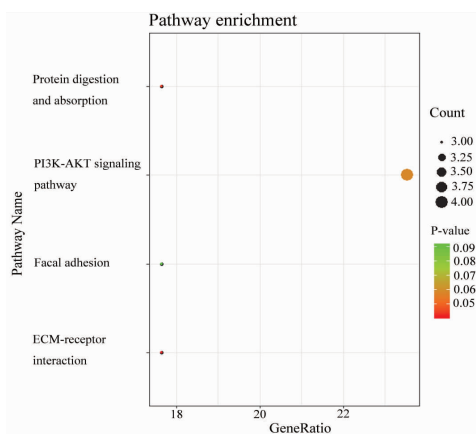


图 2 GSE54129 和 GSE79973 中高表达基因的 KEGG 通路富集分析

Figure 2. KEGG Pathway Enrichment of Up-Regulated Genes in 2 Datasets (GSE54129 and GSE79973)

2.3 关键基因在 GC 中的表达及其与 GC 分期和预后的关系

与癌旁组织相比,*COL1A1*、*COMP* 及 *THBS2* 在 GC 组织中的表达显著升高($P < 0.001$, $P < 0.001$, $P < 0.001$)(图 3A ~ C)。*COL1A1* 及 *THBS2* 在 GC 组织中的表达与 GC 的临床分期显著相关,*COL1A1* 及 *THBS2* 表达水平越高,分期越晚 ($P = 0.029$, $P = 0.025$)(图 3D ~ F)。进一步分析发现,在 GC 患者中,*COL1A1* 及 *COMP* 的高表达与 GC 的预后不良显著相关($P = 0.024$, $P = 0.045$)(图 3G ~ I)。

2.4 COL1A1 在 GC 临床组织样本中的表达及其与预后的关系

COL1A1 呈棕黄色,主要在细胞间质中表达。

与癌旁组织相比, *COL1A1* 在 GC 组织中显著高表达 ($P < 0.001$, 图 4A ~ C)。 *COL1A1* 的表达与 GC 的分期明显相关 ($P = 0.022$), *COL1A1* 表达水平越高,

分期越晚(图 4D), 且 *COL1A1* 与 GC 的预后显著相关 ($P = 0.001$) (图 4E)。

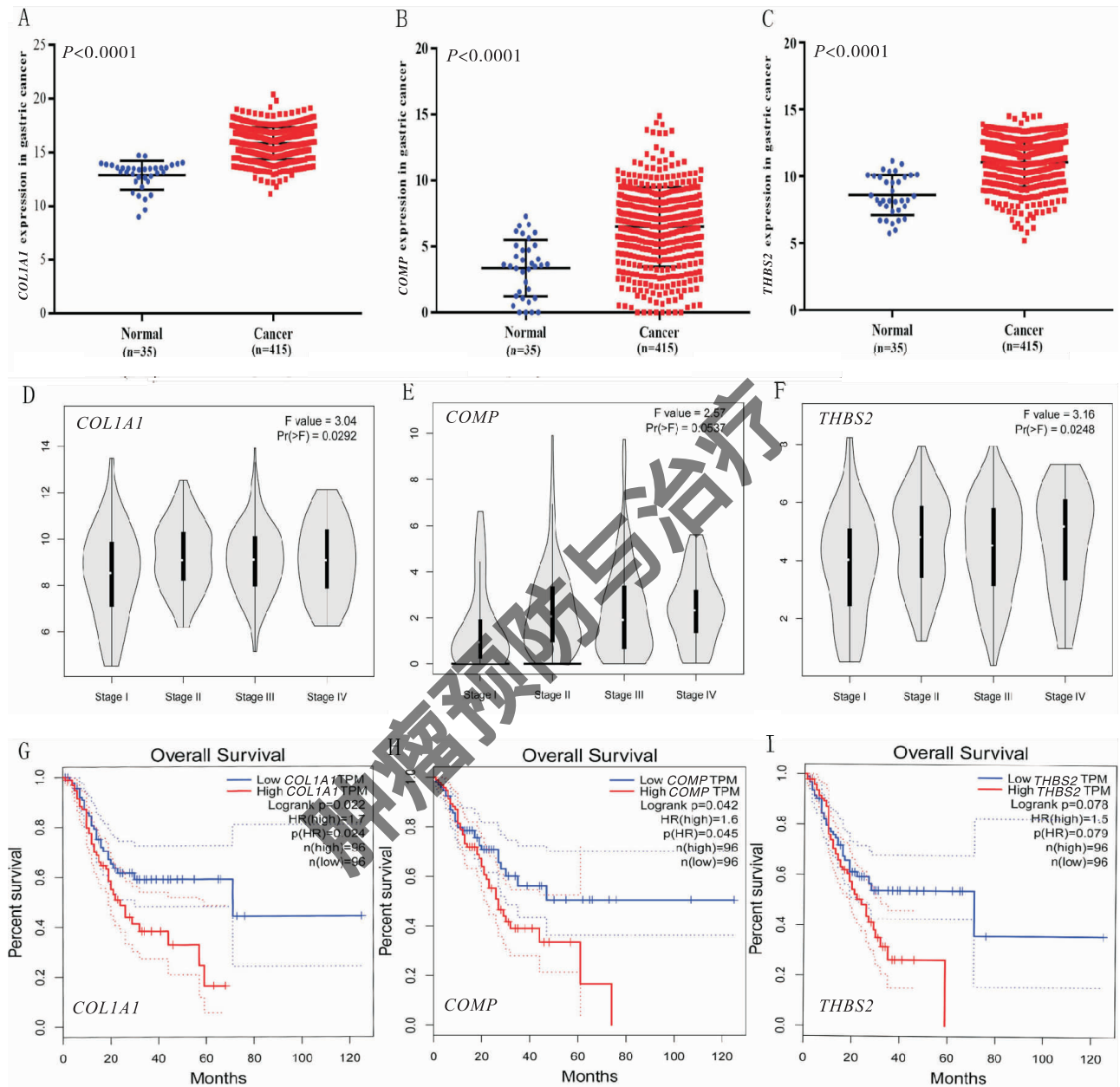


图 3 *COL1A1*、*COMP* 及 *THBS2* 在 GC 和癌旁组织中的表达及其与 GC 分期及预后的关系

Figure 3. Expression of *COL1A1*, *COMP* and *THBS2* in GC and Paracancerous Tissues and Their Relationship with Stage and Prognosis of GC

A. Expression of *COL1A1* in GC and paracancerous tissues; B. Expression of *COMP* in GC and paracancerous tissues; C. Expression of *THBS2* in GC and paracancerous tissues; D. Relationship between *COL1A1* expression and stage of GC; E. Relationship between *COMP* expression and stage of GC; F. Relationship between *THBS2* expression and stage of GC; G. Relationship between *COL1A1* expression and prognosis of GC; H. Relationship between *COMP* expression and prognosis of GC; I. Relationship between *THBS2* expression and prognosis of GC.

Abbreviations as indicated in Figure 1.

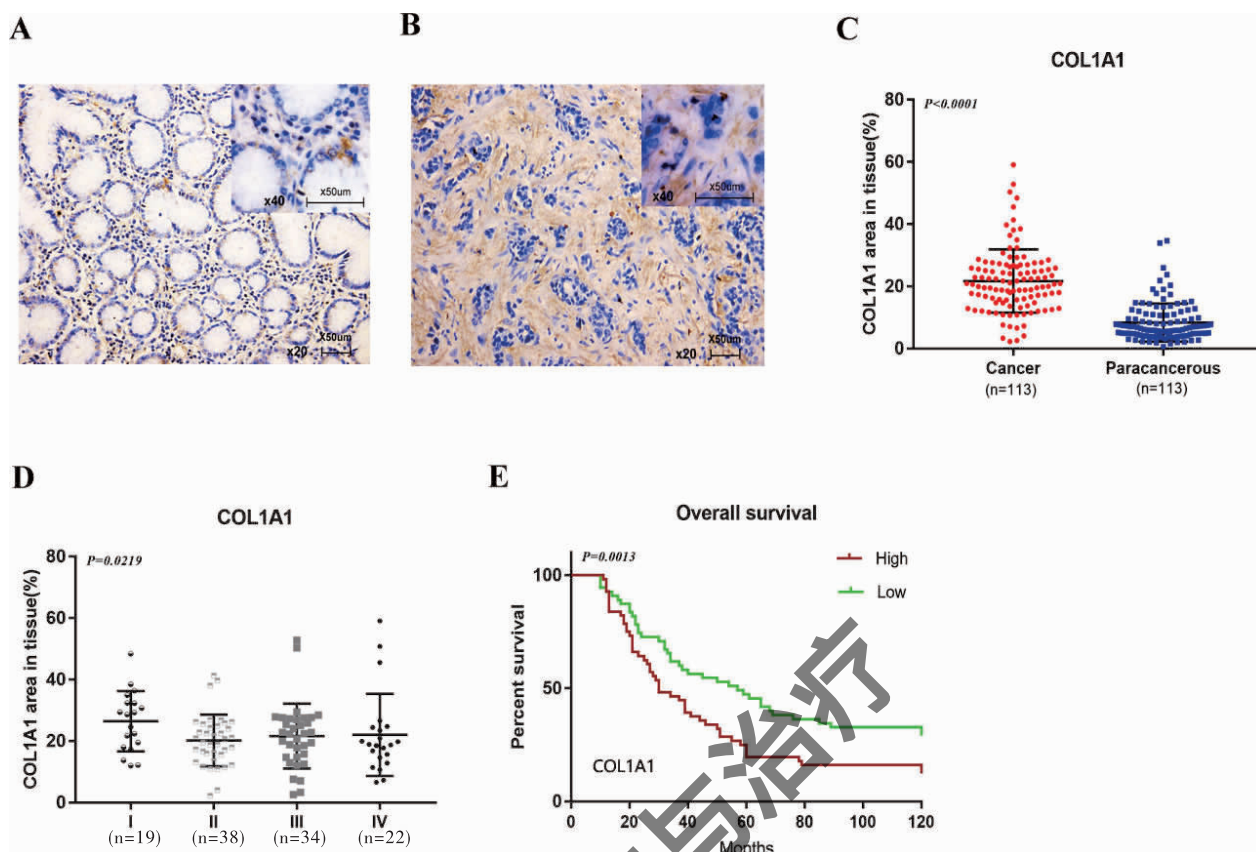


图 4 *COL1A1* 在 GC 中的表达及其与预后的关系

Figure 4. *COL1A1* Expression and Its Correlation with Prognosis of GC

A. *COL1A1* expression in GC tissues; B. *COL1A1* expression in paracancerous tissues; C. Analysis of positively expressed *COL1A1* in GC and paracancerous tissues; D. Relationship between *COL1A1* expression and stage of GC; E. Relationship between *COL1A1* expression and prognosis of GC.

Abbreviations as indicated in Figure 1.

3 讨论

生物信息学分析包括芯片表达谱、基因功能注释及基因互作网络等,是研究恶性肿瘤发病机制及筛选潜在诊断标志物和治疗靶点的有力研究工具^[6-8]。Zhang 等^[9]通过生物信息学分析发现,与癌旁/正常胃组织标本相比,99 个长链非编码 RNA 及 2 857 个 mRNA 在 GC 中的表达存在显著差异,其中 13 个长链非编码 RNA 及 823 个 mRNA 与总生存显著相关,且差异表达的 mRNA 与黏附斑及 PI3K-Akt 等信号通路有关,提示长链非编码 RNA 与 mRNA 的表达网络可作为 GC 的转录后分子特征。有学者通过差异共表达分析(differential coexpression analysis, DCEA)、分期特异基因调节网络(gene regulatory network, GRN)与差异调节网络(differential regulation networking, DRN)的方法分析 GEO 及 TCGA 数据库中的数据,发现胃内因子(gastric intrinsic fac-

tor, GIF)及 X 盒结合蛋白 1(X-box binding protein 1, XBP1)与 GC 有关,研究结果提示从差异调节网络的层面进行生物信息学分析为理解 GC 的发病机制、挖掘新的药物靶点及标志物提供了新的视角^[10]。Liao 等^[11]的研究发现 Serpin 家族 E 成员 1(serpin family E member 1, SERPINE1)及分泌蛋白酸性和富含半胱氨酸(secreted protein acidic and cysteine rich, SPARC)在 GC 中显著高表达,且其与 GC 的不良预后显著相关。

我们通过在 GEO 数据库中检索后筛选出 2 个 GC 与癌旁组织基因表达的芯片,分析发现 GC 与癌旁组织中的基因表达存在显著差异,使用 Venny 筛选出 31 个在 2 个芯片的 GC 组织中均显著高表达的基因,KEGG 分析提示 GC 中高表达的基因主要富集于细胞外基质受体相互作用和蛋白的消化、吸收信号通路,通路中包含的 3 个基因分别为 *COL1A1*、*COMP* 及 *THBS2*。吴茜等^[12]通过生信分

析发现 *INHBA*、*UGT2B15*、*ITGA2*、*ITGB1*、*SH3GL2* 及其相关的生物过程可能是 GC 的潜在生物标志物和治疗靶标。本研究的结果与其存在一定的差异,考虑主要与所分析的芯片数据及使用的生信分析方法不同有关。

COL1A1 为 I 型胶原 $\alpha 1$, 为 I 型胶原的主要成份, 在多种组织中广泛存在, *COL1A1* 的异常表达与多种肿瘤有关, 包括肾癌、黑色素瘤及肝癌[13-14]。Li 等[15] 运用 RT-PCR 的方法在正常/癌旁胃粘膜组织、GC 前病变及 GC 组织中研究发现, *COL1A1* 在 GC 中显著高表达, 其高表达提示 GC 预后不良。THBS2 为血栓素 2, 属钙离子结合糖蛋白, 其通过调节细胞间及细胞与基质间的相互作用, 调控血管生成、细胞凋亡、细胞增殖及细胞间黏附[16-17]。Zhuo 等[17] 的小组在 GC 的切除标本中研究发现, THBS2 在 GC 中表达显著上调, 生存分析结果显示与 THBS2 低表达患者相比, THBS2 高表达者预后更差。COMP 为软骨寡聚基质蛋白, 在多种细胞及组织中表达, 通过与细胞表面的整合素结合发挥生物活性, 目前研究表明 COMP 的异常表达在骨骼疾病、血管生成、乳腺癌及前列腺癌中发挥重要作用[18]。结合我们生物信息学分析的结果与既往的研究报道, 提示 *COL1A1*、*COMP* 及 *THBS2* 可能与 GC 的发生发展具有重要关系。

我们进一步分析 TCGA 数据库中 GC 与癌旁组织样本发现, 与癌旁组织相比, *COL1A1* 与 *THBS2* 在 GC 组织中的表达水平与 GC 的临床分期显著正相关。床分期相关。生存分析结果提示 *COL1A1* 及 *COMP* 高表达与 GC 的预后不良显著相关。综合 GEO 及 TCGA 数据库中的数据分析结果, 我们发现 *COL1A1* 在 GC 中显著高表达, 且其表达水平与 GC 的临床分期及生存显著相关。

虽然生物信息学分析是研究肿瘤分子生物学异常变化的有效切入点, 但由于生信分析本身属于二次数据挖掘, 其研究的对象为基因表达层面的改变, 与执行生物学功能的蛋白质分子之间没有必然的因果联系, 因此我们运用免疫组化的方法在组织学层面对 *COL1A1* 的表达进行了验证, 结果提示 *COL1A1* 主要在细胞间质中表达, 其在 GC 中显著高表达, 且其高表达与 GC 的临床分期相关, 提示预后不良。

综上所述, 既往资料及我们的研究结果提示 *COL1A1* 可能在 GC 的发生、发展中发挥重要作用, 但具体的作用机制尚不清楚。另一方面, 生物信息

学分析为研究 GC 发病相关的分子机制研究提供了新的切入点。本研究未在 GC 组织标本中研究 *COL1A1* 的表达水平, 且未深入研究 *COL1A1* 参与 GC 发生发展的分子机制, 这有待进一步的探索。

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] Feng RM, Zong YN, Cao SM, et al. Current cancer situation in China: Good or bad news from the 2018 Global Cancer Statistics? [J]. *Cancer Commun (Lond)*, 2019, 39(1): 22.
- [2] Song Z, Wu Y, Yang J, et al. Progress in the treatment of advanced gastric cancer [J]. *Tumour Biol*, 2017, 39(7): 1010428317714626.
- [3] Yang C, Chen F, Wang S, et al. Circulating tumor cells in gastrointestinal cancers: Current status and future perspectives [J]. *Front Oncol*, 2019, 9: 1427.
- [4] Matsuzaki J, Tsugawa H, Suzuki H. Precision medicine approaches to prevent gastric cancer [J]. *Gut Liver*, 2020. doi:10.5009/gnl19257.
- [5] Duan BJ, Bai J, Qiu J, et al. Histone-lysine N-methyltransferase SETD7 is a potential serum biomarker for colorectal cancer patients [J]. *EBioMedicine*, 2018, 37: 134-143.
- [6] 夏得淳, 雷子贤, 赵娟, 等. 人 TERT 基因启动子区生物信息学分析 [J]. *肿瘤预防与治疗*, 2020, 33(3): 215-223.
- [7] Liu W, Ouyang SY, Zhou ZG, et al. Identification of genes associated with cancer progression and prognosis in lung adenocarcinoma: Analyses based on microarray from Oncomine and The Cancer Genome Atlas databases [J]. *Mol Genet Genomic Med*, 2019, 7(2): e00528.
- [8] Oh SC, Sohn BH, Cheong JH, et al. Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype [J]. *Nat Commun*, 2018, 9(1): 1777.
- [9] Zhang J, Yuan YJ, Wei ZW, et al. Crosstalk between prognostic long noncoding RNAs and messenger RNAs as transcriptional hallmarks in gastric cancer [J]. *Epigenomics*, 2018, 10(4): 433-443.
- [10] Dai WT, Li QX, Liu BY, et al. Differential networking meta-a-

- analysis of gastric cancer across Asian and American racial groups [J]. BMC Syst Biol, 2018, 12(S4) : 51.
- [11] Liao P, Li W, Liu RZ, *et al.* Genome-scale analysis identifies SERPINE1 and SPARC as diagnostic and prognostic biomarkers in gastric cancer[J]. Onco Targets Ther, 2018, 11 : 6969-6980.
- [12] 吴茜, 宋兴勃, 钟慧钰, 等. 胃癌关键基因和通路的生物信息学和功能分析[J]. 肿瘤预防与治疗, 2020, 33(2) : 131-139.
- [13] Ma HP, Chang HL, Bamodu OA, *et al.* Collagen 1A1 (COL1A1) is a reliable biomarker and putative therapeutic target for hepatocellular carcinogenesis and metastasis[J]. Cancers (Basel), 2019, 11(6) : 786.
- [14] Leask A. A centralized communication network: Recent insights into the role of the cancer associated fibroblast in the development of drug resistance in tumors[J]. Semin Cell Dev Biol, 2020, 101 : 111-114.
- [15] Li J, Ding Y, Li A. Identification of COL1A1 and COL1A2 as candidate prognostic factors in gastric cancer[J]. World J Surg Oncol, 2016, 14(1) : 297.
- [16] Tian QG, Liu Y, Zhang YG, *et al.* THBS2 is a biomarker for AJCC stages and a strong prognostic indicator in colorectal cancer [J]. J BUON, 2018, 23(5) : 1331-1336.
- [17] Zhuo CJ, Li XB, Zhuang HQ, *et al.* Elevated THBS2, COL1A2, and SPPI expression levels as predictors of gastric cancer prognosis [J]. Cell Physiol Biochem, 2016, 40(6) : 1316-1324.
- [18] Papadakos KS, Bartoschek M, Rodriguez C, *et al.* Cartilage oligomeric matrix protein initiates cancer stem cells through activation of jagged1-notch3 signaling[J]. Matrix Biol, 2019, 81 : 107-121.

· 读者 · 作者 · 编者 ·

关于假冒《肿瘤预防与治疗》编辑的声明

近期发现有人或机构通过电子邮件或微信假冒《肿瘤预防与治疗》的责任编辑, 散布虚假信息。对此, 本刊特此声明:

- 1) 本刊仅通过①在线采编系统; ②官方邮箱 zlyfyzl@163.com 接收稿件并通知版面费缴纳等事宜。
- 2) 本刊不会向作者收取除版面费之外的任何费用。
- 3) 一切疑问均以《肿瘤预防与治疗》编辑部官方电话 028 - 85420233 解释为准。

请广大作者注意甄别, 谨防假冒。

《肿瘤预防与治疗》

网站: www.zlyfyzl.cn

邮箱: zlyfyzl@163.com

电话: 028 - 85420233

本刊编辑部