

DOI: 10.3872/j.issn.1007-385x.2020.12.013

· 临床研究 ·

食管腺癌相关关键基因的筛选及其临床意义

颜朝阳^a, 徐同欣^b, 吴峥^c, 郭炜^c, 董稚明^c (河北医科大学第四医院 a. 胸外科; b. CT室; c. 河北省肿瘤研究所, 河北 石家庄 050011)

[摘要] **目的:**通过TCGA和GEO数据库筛选与食管腺癌相关的关键基因,并分析其生物学功能、相关信号通路和临床意义。**方法:**综合TCGA数据库食管腺癌数据和GEO数据库GSE92396芯片数据,使用R软件的DESeq2包和Limma包进行差异表达基因分析,获得共同差异表达基因。利用R软件的clusterProfiler包对共同差异表达基因进行GO功能富集分析及KEGG通路富集分析。运用string网站和Cytoscape3.7.2软件进行蛋白互作网络分析,筛选出调节食管腺癌蛋白表达量的关键节点基因,再结合TCGA数据库分析关键节点基因与患者生存的关系。**结果:**通过数据库中90例食管腺癌组织和18例正常食管组织标本的基因芯片数据的分析,获得共同差异表达基因521个,其中高表达基因356个,低表达基因165个,它们主要与表皮发育和表皮细胞分化的代谢过程等相关功能和细胞因子及其受体相互作用等信号通路密切相关。蛋白互作网络分析得出15个关键节点基因,其中CXCL8和CCL20低表达的食管腺癌患者生存期显著长于高表达者(中位生存期32.4 vs 19.7个月, $P < 0.05$; 32.4 vs 13.9个月, $P < 0.05$)。**结论:**数据库挖掘显示CXCL8与CCL20基因可能在食管腺癌的发生发展及预后中起着重要作用,可以作为判断患者预后的潜在指标。

[关键词] 食管腺癌;TCGA数据库;GEO数据库;差异表达基因;生存分析

[中图分类号] R735.1; R730.7 **[文献标识码]** A **[文章编号]** 1007-385X(2020)12-1393-06

Identification of key genes in esophageal adenocarcinoma and its clinical significance

YAN Zhaoyang^a, XU Tongxin^b, WU Zheng^c, GUO Wei^c, DONG Zhiming^c (a. Department of Thoracic Surgery; b. Department of CT; c. Cancer Institute of Hebei Province, the Fourth Hospital of Hebei Medical University, Shijiazhuang 050011, Hebei, China)

[Abstract] **Objective:** To screen the key genes associated with esophageal adenocarcinoma by using TCGA and GEO databases, and to analyze their biological functions, relevant signaling pathways and clinical significance. **Methods:** The esophageal adenocarcinoma data downloaded from TCGA database and GSE92396 microarray data from GEO database were integrated. The analysis of differentially expressed genes (DEGs) were performed by using DESeq2 and Limma packages of R software to obtain the co-differentially expressed genes, which were then chosen for the GO function enrichment analysis and KEGG pathway analysis with clusterProfiler package of R software. The key node genes that regulate the protein expressions in esophageal adenocarcinoma were screened out by protein-protein interaction (PPI) network analysis using the string website and Cytoscape 3.7.2 software. The correlation between key node genes and the survival of patients was further analyzed by combining with TCGA database. **Results:** By analyzing the chip data of 90 cases of adenocarcinoma tissues and 18 cases of normal esophageal tissues from databases, a total of 521 co-differentially expressed genes were obtained, including 356 upregulated genes and 165 downregulated genes. These genes were closely related to the metabolic-associated functions mainly involving epidermis development, epidermal cell differentiation and signaling pathways involving cytokine-cytokine receptor interaction, etc. The PPI network analysis revealed 15 key node genes. The survival time for patients with low CXCL8 and CCL20 expression was significantly longer compared with the patients with high expression level (median survival: 32.4 vs 19.7 months, $P < 0.05$; 32.4 vs 13.9 months, $P < 0.05$). **Conclusion:** These results show that CXCL8 and CCL20 may play an important role in the occurrence, development and prognosis of esophageal adenocarcinoma, and may be used as potential indicators to judge the prognosis of patients.

[Key words] esophageal adenocarcinoma; TCGA database; GEO database; differentially expressed gene; survival analysis

[Chin J Cancer Biother, 2020, 27(12): 1393-1398. DOI: 10.3872/j.issn.1007-385X.2020.12.013]

[基金项目] 河北省自然科学基金资助项目(No. H2019206664)。Project supported by the Natural Science Foundation of Hebei Province (No.H2019206664)

[作者简介] 颜朝阳(1990-),男,硕士,主治医师,主要从事胸部肿瘤的基础及临床研究,E-mail:490255409@qq.com

[通信作者] 董稚明(DONG Zhiming, corresponding author),教授,博士生导师,主要从事肿瘤病因学研究,E-mail:dongzhiming2000@163.com

食管癌是最常见的恶性肿瘤之一,在所有癌症病死率中居第6位^[1]。在西方国家,食管腺癌是食管癌的主要亚型,而中国食管鳞状细胞癌(简称食管鳞癌)的发病率占食管恶性肿瘤的90%以上,但随着生活方式的改变,食管鳞癌的发病率逐年下降,食管腺癌的发病率逐渐升高^[2-4]。食管腺癌发病早期症状不明显,临床发现时多为中晚期,目前治疗手段主要是手术联合放化疗。在过去的20年里,尽管治疗模式不断完善,食管腺癌患者的预后改善却十分有限,5年总生存率不足25%^[5-6]。食管腺癌的早期诊断有助于提升患者的生存率,因此筛选出食管腺癌的生物标志物对其预防和早期诊治具有重要的意义。由于肿瘤发生发展过程中的复杂性,传统的单基因分析因无法揭示肿瘤复杂的信号转导网络而存在一定的局限性,高通量测序技术的发展为肿瘤的研究开辟了新的方向^[7]。本研究结合癌症和肿瘤基因图谱计划(TCGA)和人类肿瘤相关的基因表达汇编(GEO)两大数据库的基因数据集,运用生物信息学分析技术筛选共同差异表达基因,为食管腺癌的发生与预后预测提供分子标志物。

1 资料与方法

1.1 数据来源

从TCGA数据库(<http://cancergenome.nih.gov/>)下载食管腺癌的RNA-seq数据及相应的随访数据,筛选出78例食管腺癌组织和9例正常食管组织的基因表达数据。在GEO数据(<http://www.ncbi.nlm.nih.gov/geo>)中以“esophageal cancer; esophageal neoplasms; esophageal adenocarc-inoma”为关键词筛选包含癌组织与癌旁组织的食管腺癌基因芯片数据集,最终纳入GSE92396芯片数据进行分析。GSE92396数据中包含12例食管腺癌组织和9例正常食管组织。

1.2 食管腺癌组织中的差异表达基因筛选

分别使用R软件的DESeq2包和Limma包对TCGA数据库下载的RNA-seq数据及GEO数据库中的GSE92396基因芯片数据进行差异表达分析。差异表达基因筛选条件为:Fold change (FC)>2, $P_{\text{adjust}} < 0.05$ (FDR校正)。差异分析结果绘制火山图,并采用维恩图取交集获得共同差异表达基因。

1.3 食管腺癌相关共同差异表达基因的GO和KEGG富集分析

利用R软件的clusterProfiler包^[8]对共同差异表达基因进行GO功能富集分析及KEGG通路富集分析,以 $P_{\text{adjust}} < 0.05$ (FDR校正)为阈值筛选差异表达基因的主要富集功能和通路。

1.4 食管腺癌相关共同差异表达基因的蛋白互作网络分析

将共同差异表达基因导入string(<https://string-db.org/>)在线分析网站进行蛋白质互作网络分析,其结果运用Cytoscape3.7.2软件筛选关键的蛋白表达模块与关键节点基因。

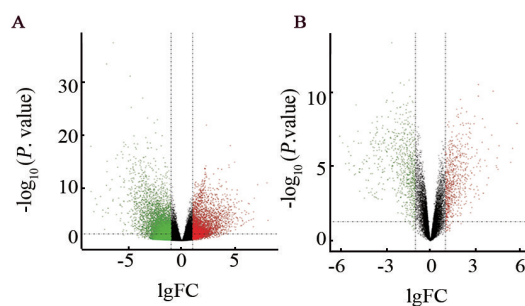
1.5 食管腺癌相关的关键节点基因与患者的生存分析

提取TCGA数据库中所有食管腺癌患者的生存资料,利用R软件survival包,将上述过程中获得的关键节点基因分为高表达组和低表达组,以总生存期(overall survival, OS)为结局终点,进行生存分析。将两组数据绘制Kaplan-Meier生存曲线,并进行Log-rank统计检验,以 $P < 0.05$ 或 $P < 0.01$ 表示差异有统计学意义。

2 结果

2.1 筛选出的食管腺癌相关差异表达基因

TCGA数据库中食管腺癌的转录组测序数据分析结果显示,表达上调的基因1909个,表达下调的基因2260个。GSE92396基因芯片筛选结果显示,表达上调的基因679个、表达下调的基因769个。使用火山图(图1)描述差异表达基因的整体分布情况。使用维恩图(图2)筛选出共同差异表达基因521个,其中表达上调的基因356个、表达下调的基因165个。



A: TCGA; B: GSE92396

图1 火山图分析差异表达基因

Fig.1 Volcano map of differentially expressed genes

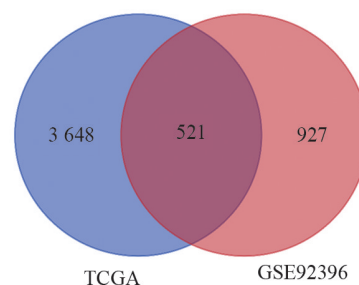


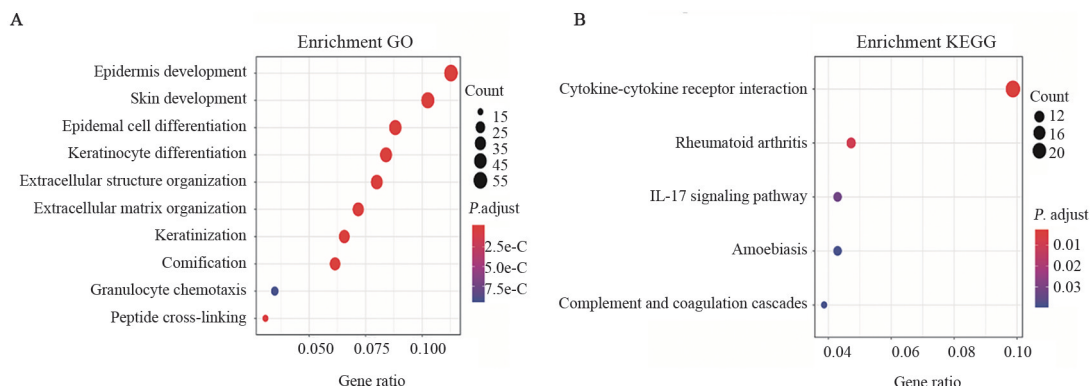
图2 维恩图显示共同差异表达基因

Fig.2 Venn diagram showing the co-differentially expressed genes

2.2 共同差异表达基因涉及的生物学作用和主要信号通路

利用 R 软件的 clusterProfiler 包对筛选出来的 521 个共同差异表达基因进行 GO 功能分析与 KEGG

通路富集分析(图 3),结果显示差异表达基因主要与表皮发育和表皮细胞分化的代谢过程等功能相关(表 1),主要富集表达于细胞因子-细胞因子受体相互作用等信号通路(表 2)。



A: GO function enrichment results; B: KEGG pathway enrichment results

图3 共同差异表达基因的功能与通路富集分析结果

Fig.3 Function and pathway enrichment analysis results of co-differentially expressed genes

2.3 蛋白互作网络分析获得 15 个关键节点基因

将共同差异表达基因导入 string 在线分析网站进行蛋白互作网络分析,其结果进一步导入 Cytoscape3.7.2 软件进行蛋白互作网络可视化处理(图 4),利用 CytoHubba 插件对蛋白互作网络中的基因进行节点得分计算,以 Degree \geq 6 作为阈值,筛选得

到前 15 个关键基因 CXCL8、VEGFA、MMP9、CXCL1、TIMP1、ICAM1、MMP1、SPP1、IVL、MMP3、PKP1、CCL20、AGT、COL1A1 及 ANXA1 基因。以默认参数 Degree \geq 6、Node Score Cutoff \geq 0.2、K-core \geq 2 及 Max. Depth=100 作为标准,使用 MCODE 插件进行重点蛋白表达模块的筛选(图 5)。

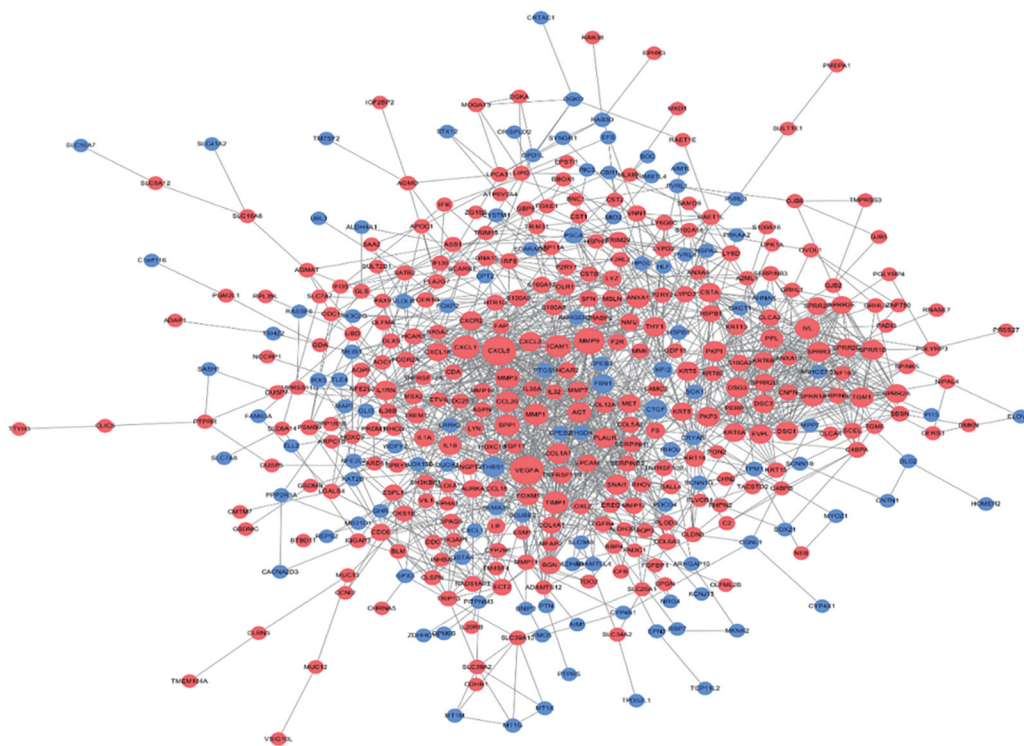


图4 共同差异表达基因的蛋白质互作网络

Fig.4 Protein-protein interaction network of co-differentially expressed genes

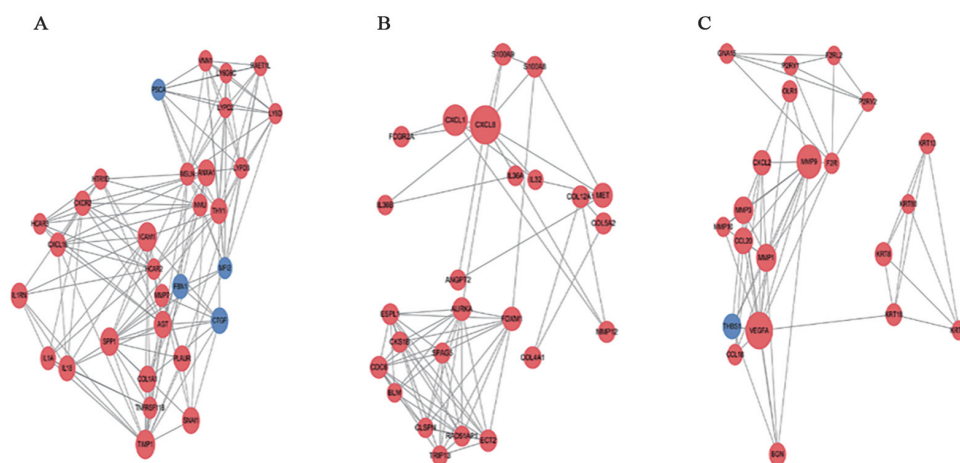


图5 食管腺癌相关关键节点基因的3个重点蛋白表达模块(A、B、C)

Fig.5 Three key protein expression modules (A, B, C) of esophageal adenocarcinoma related key node genes

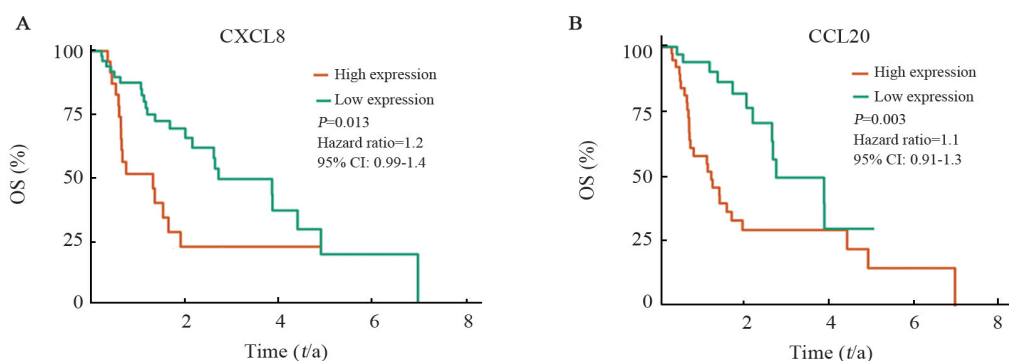


图6 食管腺癌相关关键节点基因CXCL8(A)和CCL20(B)与患者生存期的分析

Fig.6 Survival analysis of CXCL8 (A) and CCL20 (B) in esophageal adenocarcinoma

表1 GO富集分析显示的前10位的食管腺癌相关共同差异表达基因

Tab.1 The top 10 co-differentially expressed genes in esophageal adenocarcinoma in GO enrichment analysis

ID	Description	Count	P. adjust
GO:0008544	Epidermis development	55	8.00E-18
GO:0043588	Skin development	50	2.83E-16
GO:0009913	Epidermal cell differentiation	43	5.11E-14
GO:0030216	Keratinocyte differentiation	41	6.04E-15
GO:0043062	Extracellular structure organization	39	5.09E-09
GO:0030198	Extracellular matrix organization	35	2.18E-08
GO:0031424	Keratinization	32	3.42E-12
GO:0070268	Cornification	30	1.26E-18
GO:0071621	Granulocyte chemotaxis	17	9.30E-06
GO:0018149	Peptide cross-linking	15	1.41E-08

表2 食管腺癌相关共同差异表达基因涉及的信号通路

Tab.2 Signaling pathways associated with the co-differentially expressed genes in esophageal adenocarcinoma

ID	Description	Count	P. adjust
hsa04060	Cytokine-cytokine receptor interaction	23	0.003479955
hsa05323	Rheumatoid arthritis	11	0.009415404
hsa04657	IL-17 signaling pathway	10	0.032126996
hsa05146	Amoebiasis	10	0.03899629
hsa04610	Complement and coagulation cascades	9	0.03899629

2.4 食管腺癌相关的关键节点基因与患者的生存分析

利用TCGA数据库中食管腺癌患者生存数据对筛选出的关键基因进行生存分析,结果发现CXCL8与CCL20与食管腺癌患者的预后密切相关($P<0.05$)。食管腺癌患者中CXCL8与CCL20低表达者,总生存时间显著长于高表达者(CXCL8表达,中位生存期32.4 vs 19.7个月, $P<0.05$; CCL20表达,32.4 vs 13.9个月, $P<0.05$)。

3 讨论

食管腺癌是高病死率的恶性肿瘤,且全球发病率呈上升趋势^[9]。大多数食管腺癌患者起病隐匿,发现时多已发展为中晚期。随着诊断技术的不断发展,已在多种肿瘤中发现了有效的分子标志物,但食管腺癌的诊断及预后仍缺乏有效的生物标志物^[10-11]。本研究基于TCGA和GEO两大数据库的基因数据集,运用生物信息学方法,对食管腺癌的基因数据集进行分析,筛选出共同差异表达基因521个,共同差异表达基因主要汇集于表皮发育和表皮细胞分化的代谢过程等相关功能,且主要富集于细胞因子-细胞因子受体相互作用等信号通路。对共同差异表达基因进行蛋白互作网络分析,结果显示CXCL8、VEGFA、MMP9、CXCL1、TIMP1、ICAM1、MMP1、SPP1、IVL、MMP3、PKP1、CCL20、AGT、COL1A1及ANXA1这15个基因可能在食管腺癌的生物标志物中发挥重要的作用。蛋白互作网络分析联合生存分析筛选结果显示,CXCL8与CCL20可能在食管腺癌的发生发展及预后中起着重要的作用。

CXCL8是促炎因子CXC家族中的一种趋化因子,在多种肿瘤细胞、内皮细胞、浸润的中性粒细胞以及肿瘤相关巨噬细胞中都表达上调,它在肿瘤微环境中通过与细胞表面G蛋白偶联受体CXCR1和CXCR2的结合发挥促肿瘤作用^[12]。唐欣等^[13]通过对高通量数据挖掘发现,CXCL8 mRNA在宫颈鳞状上皮组织中表达上调,且其高表达与患者生存期缩短有关联。龚珍强等^[14]的研究结果表明:成纤维细胞来源的CXCL12能够增强CXCL8在结肠癌细胞和人脐静脉内皮细胞(HUVEC)中的分泌,并且CXCL8和CXCL12均能促进结肠癌细胞的增殖和侵袭能力。HOSONO等^[15]的研究结果表明,食管癌患者CXCL8高表达与淋巴结转移密切相关,且肿瘤微环境中的CXCL8上调可增强食管癌细胞的转移及侵袭能力。ŁUKASZEWICZ-ZAJĄC等^[16]的研究结果表明,与健康人群相比,食管癌患者血清中的CXCL8表达显著上调,其诊断价值优于传统肿瘤标志物。亚组分析结果表明,血清CXCL8水平与肿瘤T分期呈正相关。

趋化因子CCL20,又名巨噬细胞浸润相关蛋白30、肝激活调节趋化因子,在多种肿瘤中参与疾病的进展。趋化因子受体CCR6是目前发现的CCL20唯一受体^[17]。程先硕等^[18]的研究显示,趋化因子CCL20通过AKT/MMP3信号轴可促进结肠癌SW480细胞的侵袭和迁移。刘检等^[19]的研究结果表明,CCR6在食管癌淋巴结转移患者中呈高表达,CCR6及CCR6-CCL20趋化轴参与调控肿瘤细胞的增殖、侵袭和迁移,可能参与调节食管癌细胞上皮-间质转化的发生。LIU等^[20]的研究结果表明,CCL20可将不同种类的T淋巴细胞亚群募集到食管癌中,是患者生存的潜在预测指标。

综上所述,本研究通过对TCGA和GEO数据库的基因芯片进行分析,发现CXCL8与CCL20很可能是食管腺癌的潜在靶基因,对食管腺癌的发生发展及患者的预后评估具有重要的意义,为今后食管腺癌的诊断和预后评估提供了潜在的生物标志物。

[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. DOI:10.3322/caac.21492.
- [2] HUANG F L, YU S J. Esophageal cancer: Risk factors, genetic association, and treatment[J]. *Asian J Surg*, 2018, 41(3): 210-215. DOI:10.1016/j.asjsur.2016.10.005.
- [3] YANG W L, HAN Y, ZHAO X H, et al. Advances in prognostic biomarkers for esophageal cancer[J]. *Expert Rev Mol Diagn*, 2019, 19(2): 109-119. DOI:10.1080/14737159.2019.1563485.
- [4] SHAO D, VOGTMANN E, LIU A, et al. Microbial characterization of esophageal squamous cell carcinoma and gastric cardia adenocarcinoma from a high-risk region of China[J/OL]. *Cancer*, 2019, 125(22): 3993-4002[2020-02-18]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7285383/>. DOI: 10.1002/cncr.32403.
- [5] KATZKA D A, FITZGERALD R C. Time to challenge current strategies for detection of barrett's esophagus and esophageal adenocarcinoma[J]. *Dig Dis Sci*, 2020, 65(1): 18-21. DOI: 10.1007/s10620-019-05965-0.
- [6] NOBEL T B, LIVSCHITZ J, XING X X, et al. Surveillance implications of recurrence patterns in early node-negative esophageal adenocarcinoma[J/OL]. *Ann Thorac Surg*, 2019, 108(6): 1640-1647[2020-02-18]. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC6878144/>. DOI:10.1016/j.athoracsur.2019.05.066.
- [7] SHAO M, LI W, WANG S, et al. Identification of key genes and pathways associated with esophageal squamous cell carcinoma development based on weighted gene correlation network analysis [J]. *J Cancer*, 2020, 11(6): 1393-1402. DOI:10.7150/jca.30699.
- [8] YU G, WANG L G, HAN Y, et al. ClusterProfiler: an R package for comparing biological themes among gene clusters[J]. *Omics*, 2012, 16(5): 284-287. DOI:10.1089/omi.2011.0118.
- [9] STROES C I, SCHOKKER S, CREEMERS A, et al. Phase II feasibility

- and biomarker study of neoadjuvant trastuzumab and pertuzumab with chemoradiotherapy for resectable human epidermal growth factor receptor 2-positive esophageal adenocarcinoma: TRAP study[J/OL]. *J Clin Oncol*, 2020, 38(5): 462-471[2020-02-18]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7007286/>. DOI: 10.1200/JCO.19.01814.
- [10] GOTO M, LIU M. Chemokines and their receptors as biomarkers in esophageal cancer[J]. *Esophagus*, 2020, 17(2): 113-121. DOI:10.1007/s10388-019-00706-8.
- [11] YAMASAKI A, SHIMIZU T, KAWACHI H, et al. Endoscopic features of esophageal adenocarcinoma derived from short-segment versus long-segment Barrett's esophagus[J/OL]. *J Gastroenterol Hepatol*, 2020, 35(2): 211-217[2020-02-18]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7027738/>. DOI: 10.1111/jgh.14827.
- [12] 曹玲. 食管癌中 ID01 通过激活 KYN/NFATC2/CXCL8 通路增加中性粒细胞的招募[D]. 河南: 郑州大学, 2018.
- [13] 唐欣, 焦扬, 刘建军, 等. 高通量数据分析 CXCL8 在宫颈癌中组织表达与预后关联分析[J]. *中华肿瘤防治杂志*, 2018, 25(21): 1495-1499. DOI: 10.16073/j.cnki.cjcp.2018.21.005.
- [14] 龚珍强, 马家驰, 苏河, 等. CXCL12 和 CXCL8 调控结肠癌增殖和转移的机制研究[J]. *中华普通外科杂志*, 2018, 33(3): 244-247. DOI:10.3760/cma.j.issn.1007-631X.2018.03.018.
- [15] HOSONO M, KOMA Y I, TAKASE N, et al. CXCL8 derived from tumor-associated macrophages and esophageal squamous cell carcinomas contributes to tumor progression by promoting migration and invasion of cancer cells[J]. *Oncotarget*, 2017, 8(62): 106071-106088. DOI: 10.18632/oncotarget.22526.
- [16] ŁUKASZEWICZ-ZAJĄC M, PAĆZEK S, MUSZYŃSKI P, et al. Comparison between clinical significance of serum CXCL-8 and classical tumor markers in oesophageal cancer (OC) patients[J]. *Clin Exp Med*, 2019, 19(2): 191-199. DOI: 10.1007/s10238-019-00548-9.
- [17] SKOVDAHL H K, DAMAS J K, GRANLUND A V B, et al. C-C motif ligand 20 (CCL20) and C-C motif chemokine receptor 6 (CCR6) in human peripheral blood mononuclear cells: dysregulated in ulcerative colitis and a potential role for CCL20 in IL-1 β release [J/OL]. *Int J Mol Sci*, 2018, 19(10): 3257E3257[2020-02-18]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6214005/>. DOI: 10.3390/ijms19103257.
- [18] 程先硕, 杨芳, 董坚, 等. CCL20 通过 AKT/MMP3 轴而非 EMT 途径诱导结肠癌 SW480 细胞的侵袭和转移[J]. *中国肿瘤生物治疗杂志*, 2019, 26(6): 650-655. DOI:10.3872/j.issn.1007-385x.2019.06.006.
- [19] 刘检, 邓海峰, 徐斌, 等. 趋化因子受体 6 在食管癌组织的表达及对食管癌细胞上皮-间充质转化的影响[J]. *中华实验外科杂志*, 2018, 35(1): 141-144. DOI: 10.3760/cma.j.issn.1001-9030.2018.01.046.
- [20] LIU J Y, LI F, WANG L P, et al. CTL- vs Treg lymphocyte-attracting chemokines, CCL4 and CCL20, are strong reciprocal predictive markers for survival of patients with oesophageal squamous cell carcinoma[J/OL]. *Br J Cancer*, 2015, 113(5): 747-755[2020-02-18]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4559838/>. DOI: 10.1038/bjc.2015.290.

[收稿日期] 2020-09-06

[修回日期] 2020-10-18

[本文编辑] 韩丹