

基于生物信息学探究CXCL3在结肠腺癌中的临床价值

陶俊安, 宋政烨, 张云*, 胡慧娴*

皖南医学院医学工程学教研室, 安徽 芜湖

收稿日期: 2024年6月11日; 录用日期: 2024年7月4日; 发布日期: 2024年7月11日

摘要

目的: 通过生物信息学方法分析趋化因子CXCL3在结肠腺癌(Colon adenocarcinoma, COAD)中的临床价值。方法: 从TCGA数据库中提取出趋化因子CXCL3在COAD患者组织中的基因表达谱, 通过R语言进行数据整理, 并在COAD患者肿瘤组织与癌旁正常组织中进行差异表达分析, 再对CXCL3进行相关性分析。为了进一步探讨CXCL3在COAD中可能参与的分子机制, 通过DAVID平台展开CXCL3的功能富集分析, 并呈现可视化结果, 再结合临床病理因素分析, 最后通过GEO外部数据库进行验证。结果: 经分析发现CXCL3及其共表达基因集在肿瘤组织中高表达($p < 0.05$), 且对患者生存有不良的潜在影响。结论: CXCL3基因结肠腺癌患者中高表达, 对患者的预后有一定影响。

关键词

CXCL3, 结肠腺癌, 临床价值

Exploration of the Clinical Value of CXCL3 in Colon Adenocarcinoma Based on Bioinformatics

Jun'an Tao, Zhengye Song, Yun Zhang*, Huixian Hu*

Department of Medical Engineering, Wannan Medical College, Wuhu Anhui

Received: Jun. 11th, 2024; accepted: Jul. 4th, 2024; published: Jul. 11th, 2024

Abstract

Objective: To analyze the clinical value of chemokine CXCL3 in colon adenocarcinoma (COAD) by

*通讯作者。

文章引用: 陶俊安, 宋政烨, 张云, 胡慧娴. 基于生物信息学探究 CXCL3 在结肠腺癌中的临床价值[J]. 临床医学进展, 2024, 14(7): 305-312. DOI: 10.12677/acm.2024.1472015

bioinformatics methods. Methods: The gene expression profile of chemokine CXCL3 in the tissues of COAD patients was extracted from the TCGA database, and the data were collated by R language, and the differential expression analysis was carried out in the tumor tissues of COAD patients and adjacent normal tissues, and then the correlation analysis of CXCL3 was carried out. To further explore the possible molecular mechanisms of CXCL3 involvement in COAD, the functional enrichment analysis of CXCL3 was expanded by the DAVID platform and the visual results were presented. After that, it was combined with clinic pathological factor analysis. Finally, it is verified through the GEO external database. **Results:** The analysis showed that CXCL3 and its co-expressed gene sets were highly expressed in tumor tissues ($p < 0.05$), and had a potential adverse effect on the survival of patients. **Conclusion:** The high expression of CXCL3 gene in patients with colon adenocarcinoma has a certain impact on the prognosis of patients.

Keywords

CXCL3, Colon Adenocarcinoma, Clinical Value

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

结直肠癌(colorectal cancer, CRC)是世界上第三大常见恶性肿瘤和第二大致死性癌症, 2020 年全球约有 190 万结直肠癌发病病例和 90 万结直肠癌死亡病例[1]。结直肠癌的致病原因主要与遗传因素、饮食习惯、生活方式、环境因素以及生活方式有关[2]。

CRC 早期患者缺乏明显且特异的临床特征[3]。1968 年, Wilson 和 Ungner 就针对 CRC 的筛查提出了重要的 10 项原则[4], 同时 Dobrow 等[5]针对肿瘤筛查工作进行了总结和回顾, 进一步衍生出了肿瘤筛查 12 项原则。一些研究表明, 与未接受筛查的患者相比, 接受筛查的患者死亡率更低[6]。目前, 临床上主要的筛查方式有愈创木脂粪便潜血检测(fecaloccult blood tests, FOBT)和软式乙状结肠镜检查(FS), 已有的研究表明筛查方法具有一定的效果, gFOBT 这种化学检测在筛选无症状的可能患癌人群中的敏感性很低[7]。但 FS 筛查有潜在的危害, 例如穿孔、大出血和轻微出血, 以及手术或后续结肠镜检查导致的死亡[7]。近年来, 有较多文献报道, 生物标志物对不同恶性肿瘤的辅助诊断与鉴别方面有重要应用价值。在众多学者过去的工作中, 已鉴定出许多结直肠癌的生物标记物, 如 CA199、PS、TAG-72 和 CEA 等[8]。但是结直肠癌的高死亡率表示仍需要更多的高可靠性的生物标记物来给予 CRC 患者更精确的诊断与更好的治疗。

趋化因子是一种很小的可溶性蛋白分子, 介导各种生物过程, 可诱导多种类型的细胞进行趋化[9]。人类所表达的趋化因子多达 50 多种, 根据其两个保守的 N-末端半胱氨酸残基位置可将其分为 CC、CXC、CX3C 和 C 四个家族。趋化因子 3 (Chemokines 3, CXCL3)属于 CXC 趋化因子家族的单链蛋白[10]。田雨、Cui C 等多位学者通过生物化学分析发现趋化因子 CXCL3 可能是早期 CRC 的诊断与预后的关键基因[11]。同时据报道 CXCL3 及其相关基因可能是一种新的诊断和预后的生物标志物[12]。虽然通过研究, CXCL3 与结直肠癌之间的关系得到了一定肯定[13], 但 CXCL3 的具体关系并未全部研究清楚。

本研究探究趋化因子 CXCL3 在结直肠癌中的临床价值。对 TCGA (The cancer genome atlas)结肠腺癌数据库(Colon adenocarcinoma, COAD)数据进行提取, 通过使用 R 语言进行数据处理并综合分析, 确定趋

化因子 CXCL3 表达水平在 COAD 癌旁组织(Normal)和肿瘤组织(Tumor)中的表达量差异, 分析趋化因子 CXCL3 对于 COAD 的诊断的研究价值, 绘制 Kaplan-Meier (KM)生存曲线分析 CXCL3 的表达水平与 CRC 患者预后的相关性, GO 功能富集分析和 KEGG 通路富集分析探索 CXCL3 在 CRC 中的潜在的功能及信号通路。同时使用 GEO 数据库(Gene Expression Omnibus data base)验证 CXCL3 在 CRC 中表达量。本研究通过分析 CXCL3 在结肠腺癌中的临床价值, 为结肠癌的诊断、治疗、预后提供可靠依据。

2. 资料与方法

2.1. 数据来源

研究所用的数据集来自美国国家癌症研究所(National Cancer Institute, NCI)的 TCGA 数据库结肠腺癌项目(COAD) (<https://portal.gdc.cancer.gov/>), 包含 453 个结肠腺癌组织样本(tumor)和 41 个正常组织样本(normal)。验证所用数据集来自美国国立生物技术信息中心(National Center for Biotechnology Information, NCBI)的 GEO 数据库中的 GSE39582 (<https://www.ncbi.nlm.nih.gov/geo/>), 包含 566 个结肠腺癌组织样本和 19 个正常组织样本。表达谱矩阵为 RNA-seq 数据。使用 R (version 4.3.2)及 R 软件包分析与处理。

2.2. 表达量分析

在 R 软件包中, 针对 CXCL3 基因在结直肠组织样本和正常组织样本中的不同表达量, 利用 R 包 ggplot2 绘制箱线图, 利用 R 包 ROC 绘制受试者工作特征(Receiver Operating Characteristic, ROC)曲线, 并计算 ROC 曲线下面积(Area Under Curve, AUC)评判其诊断效能。

2.3. 相关性分析

在 TCGA 数据库中, 使用 spearman 算法将 CXCL3 与其他编码基因批量做相关性分析, 设置阈值 $p < 0.05$ 进行筛选后, 分别选取相关性系数最高(正相关)和最低(负相关)的十个基因进行绘图。利用 R 包 pheatmap 绘制热图。

2.4. 功能富集分析

对相关性分析得到的基因进一步筛选, 取相关性系数绝对值前 200 个基因($p < 0.05$)。将这批基因在 DAVID 平台(<https://david.ncifcrf.gov/>)进行基因本体论(Gene Ontology, GO)和京都基因及基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG)富集分析。其中 GO 包括生物学过程(Biological Processes, BP)、细胞成分(Cellular Components, CC)和分子功能(Molecular Functions, MF), 可视化采用 R 包 ggplot2。

2.5. 基因表达水平验证

在 GEO 数据库(GSE39582)中, 利用 R 包 ggplot2 绘制 CXCL3 基因在结肠腺癌组织样本和正常组织样本中表达量的箱线图, 并绘制其对应的 ROC 曲线、计算 AUC 值。计算其他基因与 CXCL3 的相关性, 并进行初步筛选, 并绘制热图。进一步筛选 CXCL3 共表达基因, 在 DAVID 平台进行富集分析, 将得到的结果与 TCGA 数据库的结果进行比较, 作为实验结果的验证。

2.6. 患者资料分析

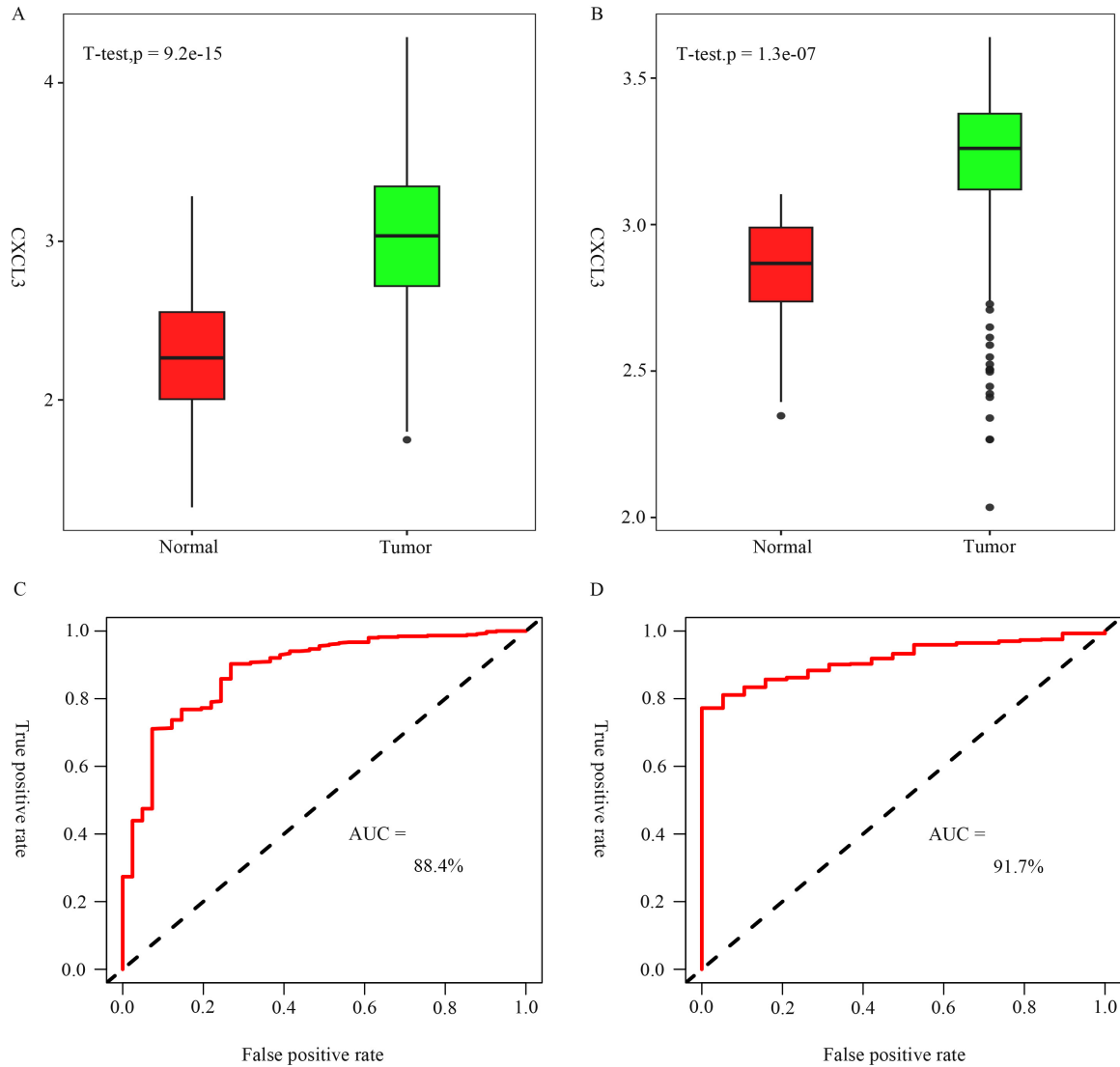
在 TCGA 数据库中, 根据 CXCL3 在结肠腺癌组织中相对表达量的中位数为界, 分成高表达组(H-CXCL3, $n = 246$)和低表达组(L-CXCL3, $n = 246$)。依据患者生存情况进一步分为存活组(Alive)和死亡组(Dead)。与年龄、性别、肿瘤分期、肿瘤浸润深度、淋巴结转移情况和远处转移情况等临床病理因素进

行分析比较，并绘制成表格。

3. 结果

3.1. 表达量分析

在 TCGA 数据中将 CXCL3 基因两组中的表达差异绘制成箱线图(图 1(A))。结果显示 CXCL3 基因在结肠腺癌组织中显著高表达($p < 0.05$)。采用 ROC 曲线评判其诊断效能，曲线下面积 AUC 为 88.4% (图 1(C))。选择 GEO 数据库中与结肠腺癌相关数据集 GSE39582 进行外部验证。结果显示 CXCL3 基因在结肠腺癌组织中显著高表达($p < 0.05$) (图 1(B))，ROC 曲线下面积 AUC 为 91.7% (图 1(D))。这表明 CXCL3 基因有成为结肠腺癌潜在标志物的可能。

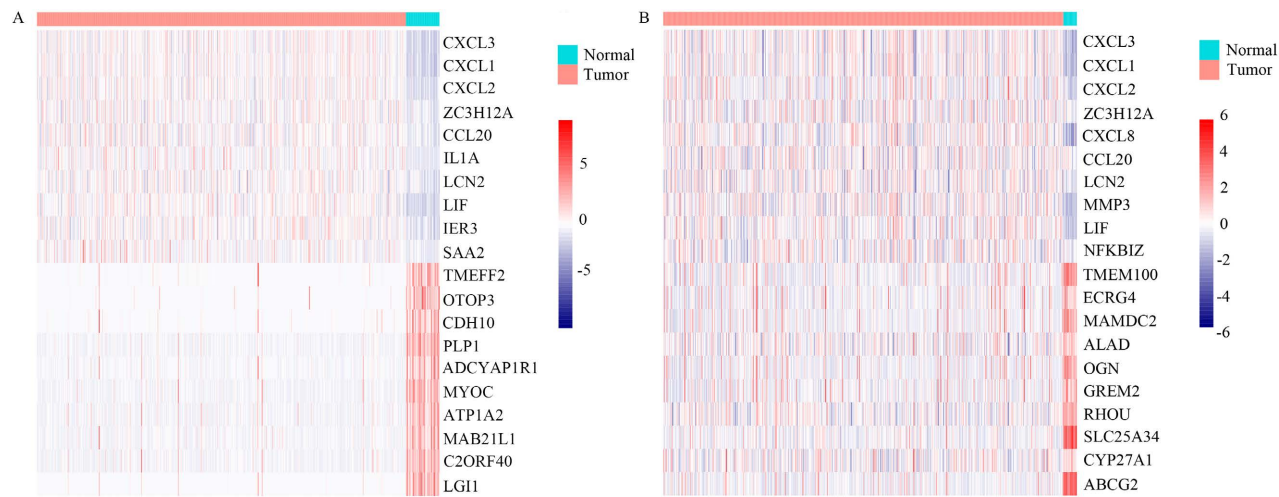


A 为 TCGA 数据库中 CXCL3 在正常组织和结肠腺癌组织的表达差异箱线图; B 为 GEO 数据库验证 CXCL3 在正常组织和结肠腺癌组织的表达差异箱线图; C 为 TCGA 数据库 CXCL3 表达差异的 ROC 曲线; D 为 GEO 数据库验证 CXCL3 表达差异的 ROC 曲线。

Figure 1. CXCL3 expression difference and ROC curve
图 1. CXCL3 表达差异与 ROC 曲线

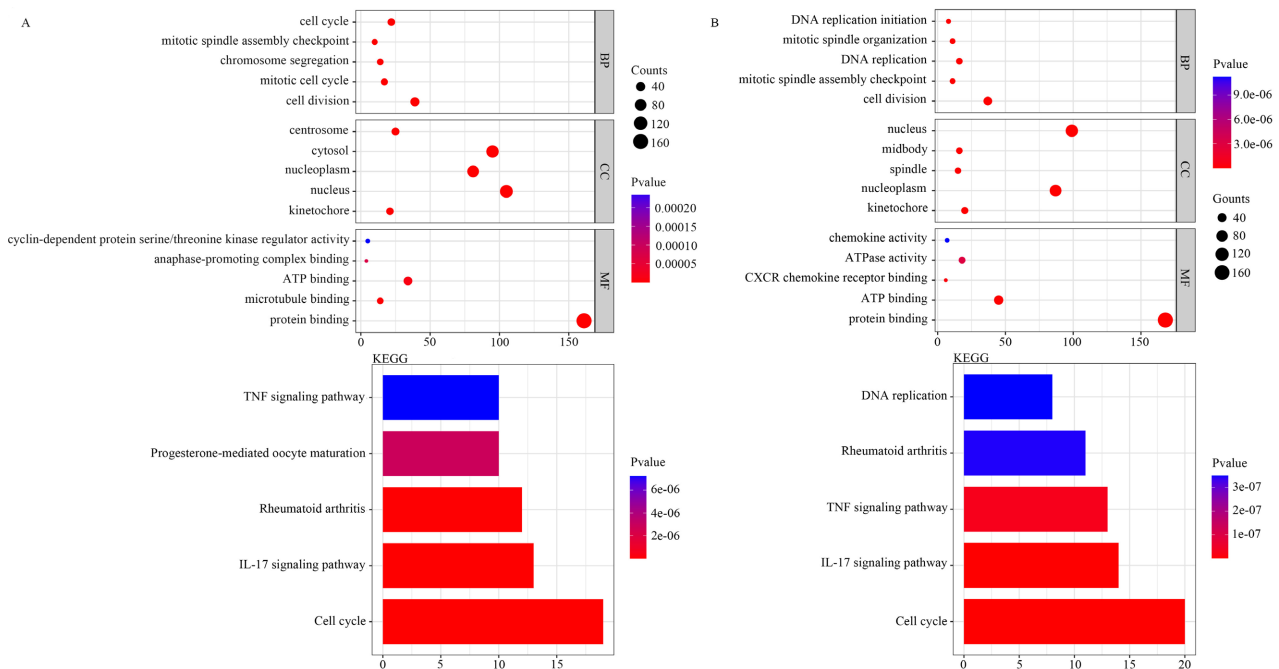
3.2. 相关性分析

在 TCGA 数据库中, 经过筛选后的 CXCL3 相关基因绘制出热图($p < 0.05$), 结果显示 CXCL3 及其正相关基因在结肠腺癌组织样本中表达量明显高于正常组织样本, 负相关基因的结果与之相反(图 2(A))。在 GSE39582 数据库中进行验证, CXCL3 及其正相关基因表达关系同数据组(图 2(B))。此结果证明 CXCL3 基因可以作为有效的结肠腺癌诊断标志物。



A 为 TCGA 数据库中 CXCL3 共表达基因集表达差异的热图; B 为 GEO 数据库验证 CXCL3 共表达基因集表达差异的热图。

Figure 2. Heat map of CXCL3 co-expressed gene set expression differences
图 2. CXCL3 共表达基因集表达差异的热图



A 为 TCGA 数据库下 BP、CC、MF 前 5 个富集条目气泡图和 KEGG 通路前 5 个富集条目柱状图; B 为 GEO 数据库验证 BP、CC、MF 前 5 个富集条目气泡图和 KEGG 通路前 5 个富集条目柱状图。

Figure 3. GO and KEGG enrichment analysis
图 3. GO 和 KEGG 富集分析

3.3. 功能富集结果

在 TCGA 数据库中将相关性分析得到的 200 个基因进行 GO 功能富集分析，结果显示其主要参与的生物学过程有细胞分裂、有丝分裂细胞周期、染色体分离、有丝分裂纺锤体组装检查点和细胞周期等，KEGG 富集结果显示其参与的主要通路为细胞周期、IL-17 信号通路、类风湿性关节炎、黄体酮介导的卵母细胞成熟和 TNF 信号通路等(图 3(A))。

在 GSE39582 数据库中进行结果的验证，结果显示其主要参与的生物学过程同样有细胞分裂和有丝分裂纺锤体组装检查点，KEGG 富集结果中同样有细胞周期、IL-17 信号通路、TNF 信号通路和类风湿性关节炎等(图 3(B))。

3.4. 患者资料分析

通过将 CXCL3 的表达量与患者临床病理因素结合分析，可见 CXCL3 高表达组相较于低表达组更容易进入肿瘤 II 期、肿瘤浸润深度 T3 期、淋巴转移情况 N0 期，肿瘤远处转移 M0 期且死亡率更高(表 1)。结果可知高表达的 CXCL3 对患者的危害性更强，致死率更高。

Table 1. Comparison of clinicopathological factors between high and low groups of CXCL3 expression in colon adenocarcinoma

表 1. 结肠腺癌中 CXCL3 表达量高低分组的临床病理因素比较

Clinical Parameters	H-CXCL3		L-CXCL3	
	Alive (N = 202)	Dead (N = 44)	Alive (N = 177)	Dead (N = 69)
Age				
Mean (SD)	67.4 (12.2)	69.8 (12.9)	65.3 (13.8)	70.7 (13.0)
Median [Min, Max]	68.0 [34.0, 90.0]	73.5 [40.0, 90.0]	67.0 [31.0, 90.0]	74.0 [34.0, 90.0]
Gender				
female	92 (45.5%)	18(40.9%)	90 (50.8%)	33(47.8%)
male	110 (54.5%)	26 (59.1%)	87 (49.2%)	36 (52.2%)
Tumor stage				
I	46 (22.8%)	2 (4.5%)	27 (15.3%)	4 (5.8%)
II	90 (44.6%)	14 (31.8%)	75 (42.4%)	19 (27.5%)
III	51 (25.2%)	11 (25.0%)	48 (27.1%)	23 (33.3%)
IV	11 (5.4%)	15 (34.1%)	25 (14.1%)	19 (27.5%)
NA	4 (2.0%)	2 (4.5%)	2 (1.1%)	4 (5.8%)
Pathologic T				
T1	7 (3.5%)	1 (2.3%)	2 (1.1%)	1 (1.4%)
T2	43 (21.3%)	2 (4.5%)	32 (18.1%)	5 (7.2%)
T3	142 (70.3%)	33 (75.0%)	117 (66.1%)	45 (65.2%)
T4	9 (4.5%)	8 (18.2%)	26 (14.7%)	18 (26.1%)
NA	1 (0.5%)	0 (0%)	0 (0%)	0 (0%)
Pathologic N				
N0	143 (70.8%)	18 (40.9%)	106 (59.9%)	27 (39.1%)

续表

N1	41 (20.3%)	10 (22.7%)	40 (22.6%)	19 (27.5%)
N2	18 (8.9%)	16 (36.4%)	31 (17.5%)	23 (33.3%)
Pathologic M				
M0	165 (81.7%)	24 (54.5%)	133 (75.1%)	36 (52.2%)
M1	11 (5.4%)	15(34.1%)	25 (14.1%)	19 (27.5%)
NA	26 (12.9%)	5 (11.4%)	19 (10.7%)	14 (20.3%)

4. 讨论

CXCL3 是趋化因子 CXC 家族中的一员，是一种非常强大的中性粒细胞的趋化剂[13]。CXCL3 通过 G 蛋白偶联受体 CXC 受体 2 (CXCR2)发出信号并与之结合促进其向肿瘤微环境的迁移，从而参与肿瘤的生长、增殖、转移和血管生成[14]-[16]。Xiong、Ruan 等人的研究也表明，CXCL3 高表达与 CRC 患者低 OS (Overall survival, OS)关[17]，且高表达的 CXCL3 与结肠癌患者的 OS 时间缩短有关[18]。由此可见趋化因子 CXCL3 高表达与 CRC 患者的总生存时间缩短，由此可见 CXCL3 是一种极具研究价值的生物标记物。

本研究通过对 CXCL3 在 tumor 组和 normal 组的差异分析，发现 CXCL3 基因在结肠腺癌组织中显著高表达(p < 0.05)，这表明 CXCL3 基因有成为结肠腺癌潜在标志物的可能。对 CXCL3 进行相关性分析，提取相关性系数绝对值前 200 个基因(p < 0.05)并进行 GO 功能富集分析。结果显示其主要参与的生物学过程有细胞分裂、有丝分裂细胞周期、染色体分离、有丝分裂纺锤体组装检查点和细胞周期等，与肿瘤细胞的发展密切相关[18]。KEGG 富集结果显示其参与的主要通路为 IL-17 信号通路，有研究表明，其与 CRC 发展密切相关[19]。TCGA 数据库的热图结果显示 CXCL3 及其正相关基因(CXCL1、CXCL2、ZC3H12A、CCL20、IL1A、LCN2、LIF、IER3、SAA2)在结肠腺癌组织样本中表达量明显高于正常组织样本，负相关基因(TMEFF2, OTOP3, CDH10, PLP1, ADCYAP1R1, MYOC, ATP1A2, MAB21L1, C2ORF40, LGI1)的结果与之相反。与田雨、Cui C、Ershov P 等学者研究结果相似[11] [20] [21]。此结果证明 CXCL3 基因可以作为有效的结肠腺癌诊断标志物。然而，本研究中也有一定的局限性。本研究基于统计分析的方法，对临床数据进行分析与研究，但尚未进行临床试验的证实，因此仍需设计包含临床验证的实验探究来进一步增强结果的可靠性。

综上所述，本研究通过生物信息学的方法研究 CXCL3 及其相关基因进行综合分析，为 CRC 的早期诊断及其愈后提供一定参考。

基金项目

皖南医学院省级大学生创新创业训练计划项目(S202310368033)。

参考文献

[1] Xi, Y. and Xu, P. (2021) Global Colorectal Cancer Burden in 2020 and Projections to 2040. *Translational Oncology*, **14**, Article 101174. <https://doi.org/10.1016/j.tranon.2021.101174>

[2] 郭士祥. 结直肠癌的发病因素及高危人群的预防对策[J]. 求医问药(下半月), 2012, 10(6): 239-240.

[3] 邢洁, 李鹏, 张澍田. 中国结直肠癌防治现状[J]. 中华结直肠疾病电子杂志, 2013, 2(6): 288-291.

[4] Harris, R., Sawaya, G.F., Moyer, V.A., et al. (2011) Reconsidering the Criteria for Evaluating Proposed Screening Programs: Reflections from 4 Current and Former Members of the U.S. Preventive Services Task Force. *Epidemiologic Reviews*, **33**, 20-35. <https://doi.org/10.1093/epirev/mxr005>

- [5] Dobrow, M.J., Hagens, V., Chafe, R., *et al.* (2018) Consolidated Principles for Screening Based on a Systematic Review and Consensus Process. *Canadian Medical Association Journal*, **190**, E422-E429. <https://doi.org/10.1503/cmaj.171154>
- [6] Dekker, E. and Rex, D.K. (2018) Advances in CRC Prevention: Screening and Surveillance. *Gastroenterology*, **154**, 1970-1984. <https://doi.org/10.1053/j.gastro.2018.01.069>
- [7] Fitzpatrick-Lewis, D., Ali, M.U., Warren, R., Kenny, M., Sherifali, D. and Raina, P. (2016) Screening for Colorectal Cancer: A Systematic Review and Meta-Analysis. *Clinical Colorectal Cancer*, **15**, 298-313. <https://doi.org/10.1016/j.clcc.2016.03.003>
- [8] 杜晓帆, 沙如拉, 李文新. 趋化因子在结肠癌中的研究进展[J]. 内蒙古医学杂志, 2023, 55(10): 1194-1196.
- [9] Miller, M.C. and Mayo, K.H. (2017) Chemokines from a Structural Perspective. *International Journal of Molecular Sciences*, **18**, Article 2088. <https://doi.org/10.3390/ijms18102088>
- [10] Zhang, L., Zhang, L.X., Li, H., *et al.* (2016) CXCL3 Contributes to CD133⁺ CSCs Maintenance and Forms a Positive Feedback Regulation Loop with CD133 in HCC via Erk1/2 Phosphorylation. *Scientific Reports*, **6**, Article No. 27426. <https://doi.org/10.1038/srep27426>
- [11] 田雨. 趋化因子 CXCL3 在结直肠癌诊断与预后判断中的相关研究[D]: [硕士学位论文]. 无锡: 江南大学, 2021.
- [12] 龚艺贞. 结肠腺癌的发病机制探索及趋化因子 CXC 亚家族基因在结肠腺癌中的临床应用价值及机制研究[D]: [博士学位论文]. 南宁: 广西医科大学, 2020.
- [13] 刘娇. 基于生物信息学途径研究结直肠癌中 CXCL 的表达及其对预后的影响[D]: [硕士学位论文]. 大连: 大连医科大学, 2023.
- [14] 程瑶, 牟文博, 辛华. 趋化因子 CXCL3 与肿瘤的相关性研究进展[J]. 化工时刊, 2020, 34(4): 26-28.
- [15] Liao, W.T., Overman, M.J., *et al.* (2019) KRAS-IRF2 Axis Drives Immune Suppression and Immune Therapy Resistance in Colorectal Cancer. *Cancer Cell*, **35**, 559-572.E7. <https://doi.org/10.1016/j.ccell.2019.02.008>
- [16] Wu, T., Yang, W., Sun, A., Wei, Z. and Lin, Q. (2022) The Role of CXC Chemokines in Cancer Progression. *Cancers*, **15**, Article 167. <https://doi.org/10.3390/cancers15010167>
- [17] Xiong, Y., You, W., Wang, R., Peng, L. and Fu, Z. (2017) Prediction and Validation of Hub Genes Associated with Colorectal Cancer by Integrating PPI Network and Gene Expression Data. *BioMed Research International*, **2017**, Article ID: 2421459. <https://doi.org/10.1155/2017/2421459>
- [18] Ruan, G.T., Gong, Y.Z., Liao, X.W., Wang, S., Huang, W., Wang, X.K., Zhu, G.Z., Liao, C. and Gao, F. (2019) Diagnostic and Prognostic Values of C-X-C Motif Chemokine Ligand 3 in Patients with Colon Cancer. *Oncology Reports*, **42**, 1996-2008. <https://doi.org/10.3892/or.2019.7326>
- [19] 李文斌. IL-17A/G-CSF axis 对于免疫细胞的调控及其在结肠炎相关结直肠癌发生发展中的作用研究[D]: [博士学位论文]. 北京: 北京协和医学院, 2013.
- [20] Cui, C., Zhang, R., Gu, F., Pei, Y., Sun, L., Huang, Y., Niu, G. and Li, J. (2022) Plasma CXCL3 Levels Are Associated with Tumor Progression and an Unfavorable Colorectal Cancer Prognosis. *Journal of Immunology Research*, **2022**, Article ID: 1336509. <https://doi.org/10.1155/2022/1336509>
- [21] Ershov, P., Poyarkov, S., Konstantinova, Y., Veselovsky, E. and Makarova, A. (2023) Transcriptomic Signatures in Colorectal Cancer Progression. *Current Molecular Medicine*, **23**, 239-249. <https://doi.org/10.2174/1566524022666220427102048>