

生物信息分析

DOI: 10.13406/j.cnki.cyxh.002724

利用生物信息学分析食管腺癌的核心基因和通路

李畅波, 莫俊贤, 谢 坚, 杨钧铭, 谢庆茂

(广西医科大学第七附属医院心胸外科, 梧州 543000)

【摘要】目的:探讨食管腺癌(esophageal adenocarcinoma, EAC)的核心基因及通路。**方法:**结合 GEO 数据库分析 EAC 与正常食管组织间的差异表达基因。利用多种在线分析软件对中枢基因进行功能富集、通路分析、构建蛋白质互作网络及挑选核心基因。结合 TCGA、Oncomine 肿瘤数据库对核心基因进行验证以及绘制受试者工作特征(receiver operating characteristic, ROC)曲线、计算曲线下面积(area under the curve, AUC)。**结果:**4 个 GEO 数据集分析得出中枢基因共 141 个, 富集在细胞外基质中, 主要通过糖酵解、糖异生通路和酪氨酸、组氨酸、苯丙氨酸、 β -丙氨酸代谢通路。筛选出 3 个核心基因, COL1A2 的 AUC 为 0.707~0.954; TIMP1 的 AUC 为 0.795~0.984; MMP1 的 AUC 为 0.896~1.000。**结论:**利用生物信息学方法筛选出与 EAC 相关的核心基因及通路, 有助于 EAC 的早期分子诊断与基因靶向治疗。

【关键词】食管腺癌; 生物信息学; 差异表达基因; 诊断

【中图分类号】R735.1

【文献标志码】A

【收稿日期】2020-08-04

Analysis of core genes and pathways in esophageal adenocarcinoma by bioinformatics

Li Changbo, Mo Junxian, Xie Jian, Yang Junming, Xie Qingmao

(Department of Cardiothoracic Surgery, The Seventh Affiliated Hospital of Guangxi Medical University)

【Abstract】Objective: To explore the core genes and pathways of esophageal adenocarcinoma (EAC). **Methods:** The differentially expressed genes between EAC tissue and normal esophageal tissue were analyzed with GEO database. A variety of online analysis software were used for functional enrichment, pathway analysis, construction of protein-protein interaction network, and selection of core genes. Combined with TCGA and Oncomine database, the core genes were verified, the ROC curves were drawn and the area under the curve (AUC) was calculated. **Results:** A total of 141 core genes were obtained from 4 GEO datasets, which were mainly enriched in the extracellular matrix, mainly in glycolysis/gluconeogenesis pathways, and tyrosine, histidine, phenylalanine and β -alanine metabolic pathways. COL1A2, TIMP1 and MMP1 were screened out in TCGA and Oncomine databases. The AUC of COL1A2 was 0.707~0.954, that of TIMP1 was 0.795~0.984, and that of MMP1 is 0.896~1.000. **Conclusion:** Screening out core genes and pathways related to EAC by bioinformatics is helpful for early molecular diagnosis and gene targeted therapy of EAC.

【Key words】 esophageal adenocarcinoma; bioinformatics; differentially expressed gene; diagnosis

食管腺癌(esophageal adenocarcinoma, EAC)是一种高死亡率的侵袭性疾病。早期 EAC 无明显症状, 难以引起人们重视; 中晚期 EAC 常出现吞咽困难症状和伴有局部浸润或淋巴结转移。由于缺乏适当的诊断方法, EAC 患者就诊时往往已处于晚期。EAC 发病率较 40 年前增长 6 倍, 其总体的 5 年生存率不足 20%^[1-2]。早期发现及治疗能极大地改善 EAC

患者的预后, 组织学病理检查仍是 EAC 诊断的金标准。然而, 由于内镜检测和组织学检查的局限性, 尚无法在早期对 EAC 患者进行有效的诊断, 因此开发能够识别 EAC 肿瘤细胞发展的生物标记物具有重要意义^[3]。基于生物学“组学”的技术在近 20 年已被广泛应用于探索疾病进展过程中基因表达的变化, 这些随着疾病变化的基因可能成为肿瘤进展的新靶点^[4-5]。因此, 本研究结合多个肿瘤数据库资料, 利用生物信息学方法阐明 EAC 发展过程中的关键基因及潜在通路, 寻找可用于 EAC 早期诊断的生物标记物, 期望能为进一步研究提供理论基础。

作者简介: 李畅波, Email: lichangbo789@163.com,

研究方向: 胸部肿瘤的基础研究及临床应用。

优先出版: <https://kns.cnki.net/kcms/detail/50.1046.R.20210105.1400.006.html>

(2021-01-06)

1 材料与方法

1.1 数据收集

在 NCBI 的 GEO 数据库中,使用表达式“esophageal adenocarcinoma”进行数据搜索及下载 mRNA 表达谱数据,纳入标准为:癌症组样本病理诊断为 EAC;每个芯片需包含正常食管组织和 EAC 组织;每个芯片中包含至少 8 个肿瘤组织样本及 8 个正常食管组织样本;芯片需提供可用的表达谱数据;物种为人类。共获得 4 个基因芯片数据集,分别为 GSE1420、GSE13898、GSE26886 和 GSE92396。

1.2 筛选共趋势表达的中枢基因

使用 R 软件 4.0 版本对芯片数据进行提取、整理,使用 limma^[6]包分别对各个芯片进行标准化及差异表达分析,以 $\log_2|FC|>1$ 及调整后 $P<0.05$ 作为筛选标准,筛选出各自的上调及下调差异表达基因(differentially expressed genes, DEGs)。取 4 个数据集的 DEGs 进行共趋势交集,即中枢基因。

1.3 DEGs 的功能富集及通路分析

使用 DAVID (<https://david.ncifcrf.gov/>) 在线分析软件对中枢基因进行 GO 功能富集及 KEGG 通路分析。GO 分析包括 3 个方面,分别为生物学过程(biological process, BP)、细胞组成(cellular component, CC)及分子功能(molecular function, MF)。GO 及 KEGG 分析结果的数据可视化由 R 软件中 ggplot2^[7]包完成。

1.4 构建蛋白互作网络及中枢基因分析

使用 STRING (<https://string-db.org/>) 在线软件对中枢基因进行蛋白互作网络(protein protein interaction network, PPI)的构建,以 $\text{score}>0.4$ 构建网络图。使用 Cytoscape3.8 软件 CytoHubba^[8]插件对中枢基因进行拓扑分析、筛选,取排名前 20 的中枢基因作后续分析。

1.5 核心基因在多个数据库中的验证及诊断效能评价

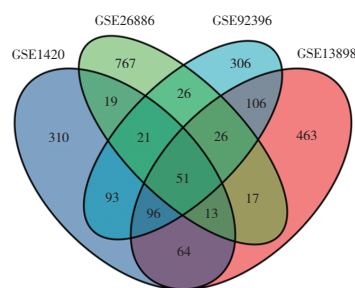
在 TCGA 数据库中下载 86 例 EAC 组织及 11 例正常食管组织的转录组数据,进行归一化处理及 $\log_2$ 转换,对前 20 的中枢基因在 EAC 组织及正常食管组织中的表达量运用 R 软件进行组间 t 检验,差异具有显著性的基因定义为核心基因,并使用 ggplot2 包作小提琴图呈现。在 Oncomine (<https://www.oncomine.org>) 在线肿瘤数据集中搜索包含核心基因在内的食管腺癌与正常组织对比的数据集,继而使用其在线分析软件对核心基因表达量标准化后进行综合对比。使用 SPSS 22.0 分析、绘制核心基因在 GEO 和 TCGA 数据库中的受试者工作特征(receiver operating characteristic, ROC)曲线,其曲线下面积(area under the curve, AUC)用以评估核心基因

的诊断效能, $0.5<AUC\leq 0.7$ 表示准确性较低, $0.7<AUC\leq 0.9$ 表示准确性中等, $0.9<AUC\leq 1.0$ 表示准确性较高。所有统计数据的检验水平 $\alpha=0.05$ 。

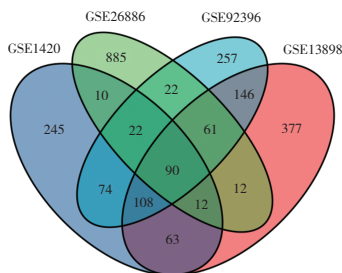
2 结果

2.1 GEO 芯片数据集及中枢基因

根据纳入及排除标准,最终入选本次研究的 GEO 芯片数据集共 4 个,分别为 GSE1420、GSE13898、GSE26886 和 GSE92396,共包含 EAC 组织样本 64 例,正常食管组织样本 105 例,其 DEGs 见表 1。在 4 个芯片数据集中呈现共趋势表达的中枢基因共 141 个,其中共趋势上调表达 51 个(图 1A),共趋势下调表达 90 个(图 1B)。



A. 韦恩图显示上调的共趋势中枢基因共 51 个



B. 韦恩图显示下调的共趋势中枢基因共 90 个

图 1 数据集中差异表达基因的韦恩图

2.2 中枢基因的功能及通路分析

中枢基因的 GO 富集分析显示,其主要参与细胞醛代谢过程、肽交联、角质细胞分化等 BP(图 2A);存在的 CC 主要包括细胞外泌体、细胞外空间及基底外侧质膜组成等(图 2B);发挥的 MF 主要包括细胞外基质结构成分、醛脱氢酶活性、结构分子活性等(图 2C)。中枢基因的 KEGG 通路分析显示其主要作用通路为糖酵解/糖异生通路;黏蛋白 O-聚糖的生物合成通路;酪氨酸、组氨酸、苯丙氨酸、 β -丙氨酸等氨基酸代谢通路(图 2D)。

表 1 纳入的 4 个 GEO 数据集详细信息

数据集(年份)	平台	标本来源	国家	数目		差异基因	
				EAC	正常	上调	下调
GSE1420(2004)	GPL96	组织	美国	8	8	681	625
GSE13898(2011)	GPL6102	组织	美国	64	28	836	869
GSE26886(2013)	GPL570	组织	德国	21	19	943	1 119
GSE92396(2016)	GPL6244	组织	美国	12	9	725	780

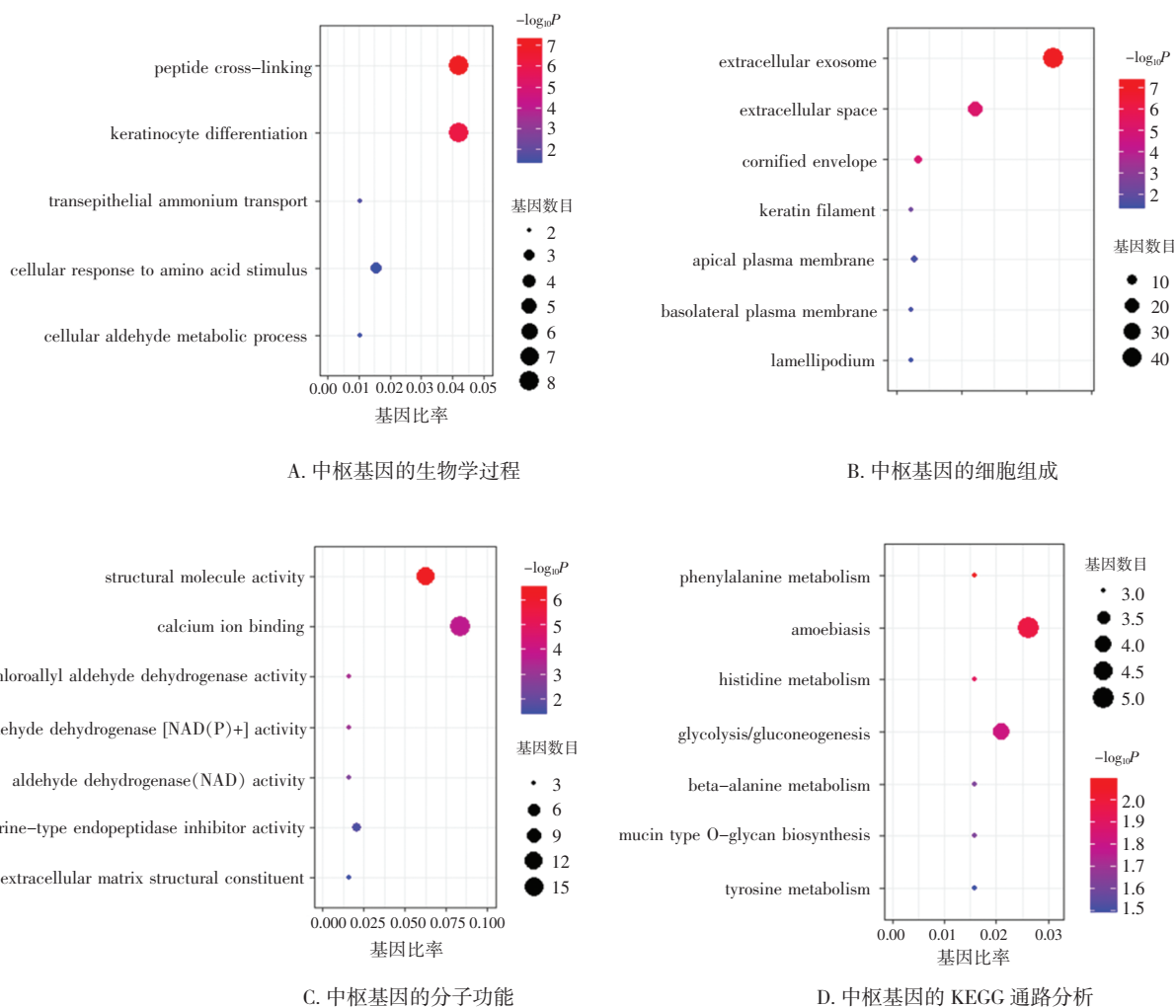


图2 中枢基因的 GO 及 KEGG 通路富集分析

2.3 中枢基因的 PPI 分析

对 141 个共趋势的中枢基因使用 STRING 在线数据库进行 PPI 构建分析,使用 Cytoscape 软件及其 CytoHubba 插件计算蛋白节点之间的相互作用,得出排名前 20 的中枢基因,包括 CLCA2、COL1A2、DSC3、DSG3、IVL、KRT13、KRT4、KRT5、KRT8、MMP1、PKP1、PPL、SCEL、SPRR1A、SPRR1B、SPRR2B、SPRR3、TGM1、TIMP1、VCAN 等(图 3)。

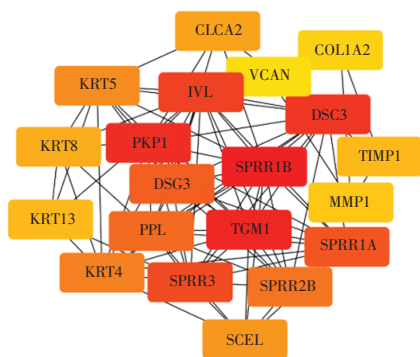


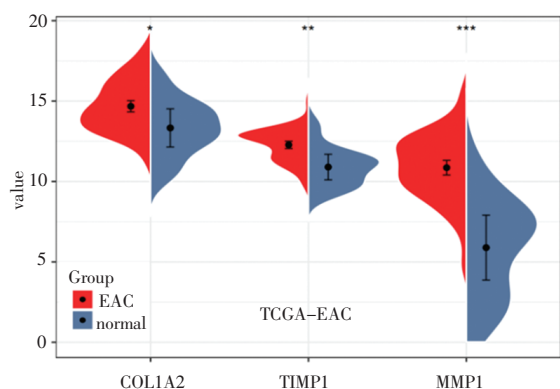
图3 蛋白质互作网络的核心模块

2.4 核心基因表达水平的验证及诊断效能评价

在 TCGA-EAC 数据库中,肿瘤组织与正常组织的表达量分析结果得出 COL1A2、TIMP1、MMP1 的差异具有统计学意义(图 4A),定义为核基因。将包含核基因的 3 个 Oncomine 肿瘤数据综合分析发现 COL1A2、TIMP1、MMP1 的表达量同样具有统计学差异(图 4B)。对 3 个核基因在 TCGA 及 GEO 数据集中进行 ROC 分析,结果显示 COL1A2 的 AUC 为 0.707~0.954(图 5A);TIMP1 的 AUC 为 0.795~0.984(图 5B);MMP1 的 AUC 为 0.896~1.000(图 5C),表明三者的表达量均能很好地区分 EAC 组织与正常食管组织。

3 讨论

EAC 是食管癌的主要类型之一,其早期诊断率不高,发病率逐年上升,总体 5 年生存率不理想^[1]。EAC 具体的发生发展机制尚不明确,探索其相关机制、寻找潜在的基因靶点有助于改善患者的生存预后。本研究通过全面检索及分析 GEO 数据库,最终选



A. 核心基因在 TCGA 数据库中的表达水平

Median Rank	P-Value	Gene
885.0	0.005	COL1A2
210.0	3.15E-4	TIMP1
178.0	2.42E-4	MMP1

1	2	3
---	---	---

Legend

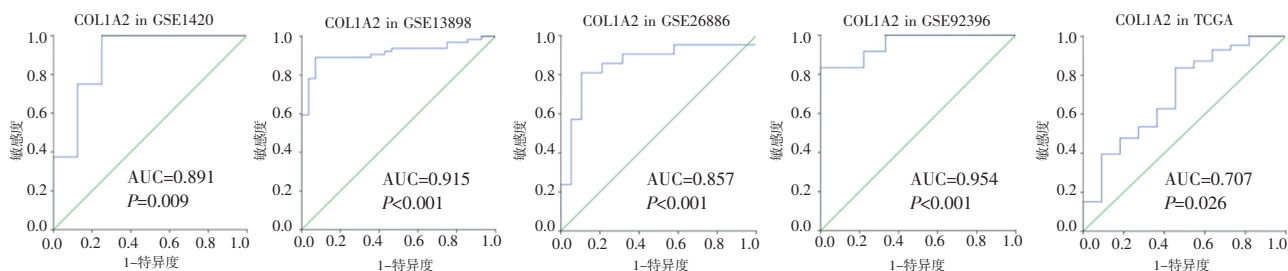
1. Esophageal Adenocarcinoma vs. Normal
Hao Esophagus, Gastroenterology, 2006
2. Esophageal Adenocarcinoma vs. Normal
Kimchi Esophagus, Cancer Res, 2005
3. Esophageal Adenocarcinoma vs. Normal
Kim Esophagus, PLoS One, 2010

1 5 10 25 25 10 5 1
%
Not measured

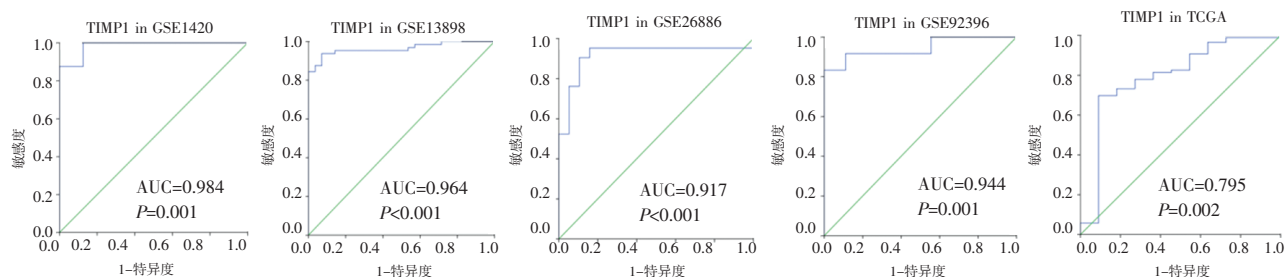
The rank for a gene is the median rank for that gene across each of the analyses.
The p-Value for a gene is its p-Value for the median-ranked analysis.

B. 核心基因在 Oncomine 肿瘤数据集表达热图

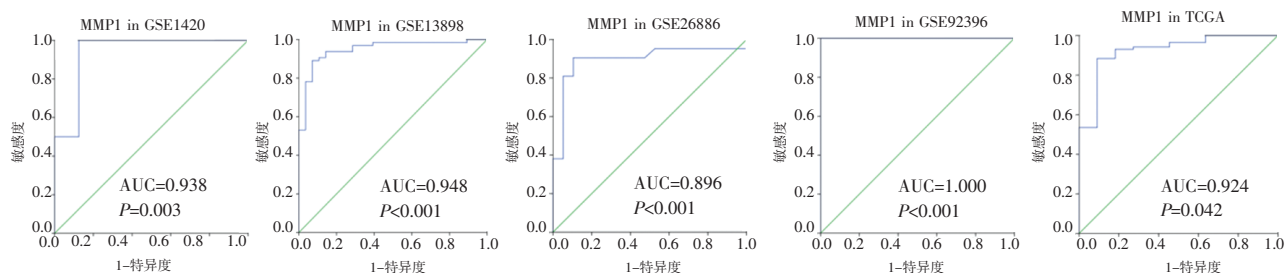
图 4 核心基因在 TCGA 和 Oncomine 数据库中的表达水平



A. COL1A2 在 GEO 和 TCGA 数据库中的 ROC 曲线



B. TIMP1 在 GEO 和 TCGA 数据库中的 ROC 曲线



C. MMP1 在 GEO 和 TCGA 数据库中的 ROC 曲线

图 5 COL1A2、TIMP1、MMP1 在 GEO 和 TCGA 数据库中的 ROC 曲线

择对 GSE1420、GSE13898、GSE26886 及 GSE92396 4 个芯片数据进行差异表达分析得出核心基因,并在 TCGA 和 Oncomine 肿瘤数据库中进行验证,结果提示 COL1A2、TIMP1、MMP1 可能为 EAC 的潜在

基因靶点。

COL1A2 是 I 型胶原 (collagen type I, COL1) 的亚型之一,位于 17 号染色体上,负责编码 I 型胶原的前 $\alpha 2$ 链^[9]。COL1 主要由成纤维细胞分泌,富含

于细胞外基质中,存在于结缔组织、骨膜和肌腱等部位,对于肿瘤生长和上皮-间质转化非常重要^[9]。COL1 可在宫颈癌肿瘤细胞中激活 Caspase-3/PI3K/Akt 通路,抑制辐射诱导的细胞凋亡^[10]。Shi Y 等^[11]敲除胃癌细胞中的 COL1 后,发现胃癌细胞增殖、侵袭能力明显降低。Fang S 等^[12]发现 COL1 可由癌细胞内源性分泌,COL1 通过调节细胞外基质以促进癌细胞黏附及肿瘤克隆形成,其表达水平与非小细胞肺癌及食管鳞状细胞癌患者预后显著相关。本研究对 COL1A2 的敏感性、特异性和 AUC 进行了评价,结果表明 COL1A2 能较好地鉴别 EAC 组织与正常组织。

MMP1 位于 11 号染色体上,是多肽酶 M10 家族成员之一,负责编码基质金属蛋白酶(matrix metalloproteinases, MMPs),在正常的生理过程中参与细胞外基质的分解,如胚胎发育、增殖和组织重建,以及炎症的发生和肿瘤的转移^[13]。此前研究显示 MMP1 参与各种癌症的进展,其高表达与癌症风险增加相关^[14]。Liu M 等^[15]通过对 611 例食管鳞癌患者研究发现,MMP1 在食管鳞癌组织中呈高表达,与癌细胞的微血管浸润、淋巴结转移和 TNM 分期呈正相关,是患者总体生存时间的独立影响因素。该研究进一步实验证实,MMP1 可通过激活 PI3K/Akt 途径促进食管鳞癌细胞增殖和迁移。此前一些报道指出,在 EAC 样本中已经观察到 MMP1 的上调,但还没有进行充分的研究^[16]。本研究结果显示 MMP1 在各个肿瘤数据库中的诊断效能较高,MMP1 可能是 EAC 发展过程中的活性蛋白,有成为 EAC 生物标志物的潜力。

TIMP1 是 TIMP 家族重要的一员。TIMPs 是上述 MMPs 的抑制剂,这些内源性分泌蛋白具有高选择性,能以 1:1 的比例与 MMPs 以非共价方式特异性结合抑制 MMP^[17]。TIMP-1 除了具有金属蛋白酶抑制结构,还具有细胞因子样信号传导活性结构,双结构域的功能导致其能与众多独特的酶和细胞表面蛋白发生功能性相互作用,引发广泛的下游效应^[18]。TIMP1 已被证实可以在乳腺癌、结直肠癌及黑色素瘤患者的体液中进行检测,血清 TIMP1 有望成为相应癌症的诊断及预后指标^[19-21]。本研究在 EAC 组织芯片数据中发现 TIMP1 的 AUC 为 0.795~0.984,能很好地区分 EAC 及非癌食管组织,TIMP1 同样有望成为 EAC 早期诊断及预后评估指标。

功能富集分析发现,包括核心基因 COL1A2、TIMP1、MMP1 在内的中枢基因的 CC 主要富集在细胞外泌体、细胞外空间及基底外侧质膜组成等,而且 MF 主要为细胞外基质结构成分。细胞外基质主要是细胞合成和分泌大分子物质的主要场所,起到传递细胞信号,调节细胞生长、发育、迁徙等功能^[22]。食管癌肿瘤细胞与其相邻基质之间的相互作用被认为是肿瘤发生与进展以及影响肿瘤细胞化疗敏感性的重要一环^[23]。

在 KEGG 通路富集中,中枢基因显著富集的通路主要有糖酵解、糖异生、O-聚糖的生物合成等通路以及酪氨酸、苯丙氨酸、组氨酸、 β -丙氨酸代谢等多种氨基酸代谢。Huang W 等^[24]对 17 项研究进行荟萃分析,结果提示糖代谢异常可增加男性患食管癌的风险。富含葡萄糖的细胞外环境促进了细胞的糖吸收及处理,高糖量负荷可引起 Na^+/H^+ 泵的逆向转运从而导致正常食管细胞恶变;体外合成的糖脂醇抑制了食管癌细胞的增殖并诱导其凋亡,但不引起正常细胞的凋亡,具有一定的抗癌作用,但糖代谢异常的患者糖脂醇合成能力下降,容易引起癌症发生^[25]。此外,O-聚糖可以调节下游的 FAK 信号传导,减弱食管癌细胞中辐射诱导的细胞凋亡,增加癌细胞的放射抵抗性^[26]。Cheng J 等^[27]发现酪氨酸及其产物苯丙氨酸可直接影响肿瘤微环境,进而导致食管癌的发生。Lee W 等^[28]研究发现人体胆汁可将组氨酸转化成组胺,从而引起肠道微生物改变,导致消化道肿瘤的产生。 β -丙氨酸虽然在食管鳞癌中没有报道,但 Pandurangan M 等^[29]通过对肾、宫颈肿瘤细胞使用流式细胞仪检测方法测定 β -丙氨酸治疗前后的细胞存活率,结果显示 β -丙氨酸有潜在的抗癌作用。

综上所述,本研究采用生物信息学方法结合多个肿瘤数据库进行分析,发现与 EAC 相关的中枢基因主要富集在细胞外基质中,通过糖酵解、糖异生通路和多种氨基酸代谢通路影响其发生与进展,而且 COL1A2、TIMP1、MMP1 对 EAC 具有较好的区别能力,极有可能成为 EAC 潜在的基因靶点,有助于 EAC 的早期分子诊断与基因靶向治疗。

参 考 文 献

- [1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020[J]. CA Can-

- cer J Clin, 2020, 70(1): 7–30.
- [2] Gupta B, Kumar N. Worldwide incidence, mortality and time trends for cancer of the oesophagus[J]. Eur J Cancer Prev, 2017, 26(2): 107–118.
- [3] Wani S, Gaddam S. Editorial; best practices in surveillance of Barrett's esophagus[J]. Am J Gastroenterol, 2017, 112(7): 1056–1060.
- [4] Kimchi ET, Posner MC, Park JO, et al. Progression of Barrett's metaplasia to adenocarcinoma is associated with the suppression of the transcriptional programs of epidermal differentiation[J]. Cancer Res, 2005, 65(8): 3146–3154.
- [5] Konda V, Souza RF. Biomarkers of Barrett's esophagus; from the laboratory to clinical practice[J]. Dig Dis Sci, 2018, 63(8): 2070–2080.
- [6] Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies[J]. Nucleic Acids Res, 2015, 43(7): e47.
- [7] Villanueva RAM, Chen ZJ. ggplot2: elegant graphics for data analysis (2nd ed.)[J]. Meas Interdiscip Res Perspect, 2019, 17(3): 160–167.
- [8] Chin CH, Chen SH, Wu HH, et al. cytoHubba: identifying hub objects and sub-networks from complex interactome[J]. BMC Syst Biol, 2014, 8(S4): S11.
- [9] Exposito JY, Valcourt U, Cluzel C, et al. The fibrillar collagen family[J]. Int J Mol Sci, 2010, 11(2): 407–426.
- [10] Liu S, Liao G, Li G. Regulatory effects of COL1A1 on apoptosis induced by radiation in cervical cancer cells[J]. Cancer Cell Int, 2017, 17: 73.
- [11] Shi Y, Duan Z, Zhang X, et al. Down-regulation of the let-7i facilitates gastric cancer invasion and metastasis by targeting COL1A1[J]. Protein Cell, 2019, 10(2): 143–148.
- [12] Fang S, Dai Y, Mei Y, et al. Clinical significance and biological role of cancer-derived type I collagen in lung and esophageal cancers[J]. Thorac Cancer, 2019, 10(2): 277–288.
- [13] Murray GI. Matrix metalloproteinases; a multifunctional group of molecules[J]. J Pathol, 2001, 195(2): 135–137.
- [14] Lu L, Sun Y, Li Y, et al. The polymorphism MMP1-1607(1G>2G) is associated with a significantly increased risk of cancers from a Meta-analysis[J]. Tumour Biol, 2015, 36(3): 1685–1693.
- [15] Liu M, Hu Y, Zhang MF, et al. MMP1 promotes tumor growth and metastasis in esophageal squamous cell carcinoma[J]. Cancer Lett, 2016, 377(1): 97–104.
- [16] Dai Y, Wang Q, Gonzalez LA, et al. Genome-wide analysis of Barrett's adenocarcinoma. A first step towards identifying patients at risk and developing therapeutic paths[J]. Transl Oncol, 2018, 11(1): 116–124.
- [17] Jackson HW, Defamie V, Waterhouse P, et al. TIMPs: versatile extracellular regulators in cancer[J]. Nat Rev Cancer, 2017, 17(1): 38–53.
- [18] Grünwald B, Schoeps B, Krüger A. Recognizing the molecular multifunctionality and interactome of TIMP-1[J]. Trends Cell Biol, 2019, 29(1): 6–19.
- [19] Lee JH, Choi JW, Kim YS. Serum TIMP-1 predicts survival outcomes of invasive breast carcinoma patients; a Meta-analysis[J]. Arch Med Res, 2011, 42(6): 463–468.
- [20] Böckelman C, Beilmann-Lehtonen I, Kaprio T, et al. Serum MMP-8 and TIMP-1 predict prognosis in colorectal cancer[J]. BMC Cancer, 2018, 18(1): 679.
- [21] Tarhini AA, Lin Y, Yeku O, et al. A four-marker signature of TNF-RII, TGF- α , TIMP-1 and CRP is prognostic of worse survival in high-risk surgically resected melanoma[J]. J Transl Med, 2014, 12: 19.
- [22] Senthilane DA, Jonker T, Rowe A, et al. The role of tumor microenvironment in chemoresistance; 3D extracellular matrices as accomplices[J]. Int J Mol Sci, 2018, 19(10): 2861.
- [23] Palumbo A, Da Costa NM, Pontes B, et al. Esophageal cancer development; crucial clues arising from the extracellular matrix[J]. Cells, 2020, 9(2): 455.
- [24] Huang W, Ren H, Ben Q, et al. Risk of esophageal cancer in diabetes mellitus; a Meta-analysis of observational studies[J]. Cancer Causes Control, 2012, 23(2): 263–272.
- [25] Hochwald JS, Zhang J. Glucose oncometabolism of esophageal cancer[J]. Anticancer Agents Med Chem, 2017, 17(3): 385–394.
- [26] Zhang C, Deng X, Qiu L, et al. Knockdown of C1GalT1 inhibits radioresistance of human esophageal cancer cells through modifying β 1-integrin glycosylation[J]. J Cancer, 2018, 9(15): 2666–2677.
- [27] Cheng J, Zheng G, Jin H, et al. Towards tyrosine metabolism in esophageal squamous cell carcinoma[J]. Comb Chem High Throughput Screen, 2017, 20(2): 133–139.
- [28] Lee W, Um J, Hwang B, et al. Assessing the progression of gastric cancer via profiling of histamine, histidine, and bile acids in gastric juice using LC-MS/MS[J]. J Steroid Biochem Mol Biol, 2020, 197: 105539.
- [29] Pandurangan M, Enkhtaivan G, Mistry B, et al. β -Alanine intercede metabolic recovery for amelioration of human cervical and renal tumors[J]. Amino Acids, 2017, 49(8): 1373–1380.

(责任编辑:唐秋姗)