

·生物信息学·

食管癌预后相关 lncRNA 的特征构建

黄向东,张兴伟

(内蒙古赤峰市医院胸外科,内蒙古 赤峰 024000)

摘要:目的 基于癌症基因组图谱(TCGA)数据库,探索食管癌预后相关 lncRNA,构建患者预后预测特征,为食管癌患者的潜在治疗靶点提供新的证据支持。**方法** 获取 TCGA 数据库中食管癌患者表达谱及临床信息。使用 R 语言鉴定癌组织与正常组织之间差异表达的 mRNAs(DEGs)和差异表达的 lncRNAs(DELs)。基于 Cox 回归和 LASSO 回归分析构建预后预测特征,通过 K-M 生存分析评估预后模型中高风险和低风险人群体的生存差异。另外,通过 Pearson 相关分析来识别与预后模型中 lncRNAs 具有共表达关系的 mRNAs,对 lncRNA 的功能进行分析。**结果** 基于 TCGA 数据库,共获得 169 个样本,鉴定出 592 个 DELs 和 1219 个 DEGs。基于 LASSO 回归分析,构建了包含 13 个 lncRNA 的稳健预测特征。根据风险评分将患者分为高低风险组,K-M 生存曲线结果显示,高危组的总生存时间(OS)短于低危组($P=3e-10$),该风险特征对 1、3 和 5 年生存率预测的 AUC 分别为 0.762、0.773 和 0.925,显示出极好的可预测性。AC002331.1、LINC01068 和 RP11-2N1.2 的高表达与患者的不良预后密切相关。**结论** 本研究中构建的 lncRNA 特征可较好的预测食管癌患者的总生存时间,可能是食管癌患者的潜在治疗靶点。

关键词:TCGA 数据库;食管癌;lncRNA;预测特征;生物标志物

中图分类号:R735.1

文献标识码:A

DOI:10.3969/j.issn.1006-1959.2023.05.001

文章编号:1006-1959(2023)05-0001-07

Feature Construction of lncRNA Related to Prognosis of Esophageal Cancer

HUANG Xiang-dong,ZHANG Xing-wei

(Department of Thoracic Surgery,Chifeng Municipal Hospital,Chifeng 024000,Inner Mongolia,China)

Abstract: Objective To explore the prognostic lncRNA of esophageal cancer and construct the prognostic prediction features of patients based on the Cancer Genome Atlas (TCGA) database, and to provide new evidence support for the potential therapeutic targets of esophageal cancer patients.**Methods** The expression profile and clinical information of esophageal cancer patients in TCGA database were obtained. R language was used to identify differentially expressed mRNAs (DEGs) and differentially expressed lncRNAs (DELs) between cancer tissues and normal tissues. Based on Cox regression and LASSO regression analysis, prognostic predictive features were constructed, and survival differences between high-risk and low-risk groups in the prognostic model were evaluated by K-M survival analysis. In addition, Pearson correlation analysis was used to identify mRNAs that were co-expressed with lncRNAs in the prognostic model, and the function of lncRNAs was analyzed.**Results** Based on the TCGA database, a total of 169 samples were obtained, and 592 DELs and 1219 DEGs were identified. Based on LASSO regression analysis, robust prediction features containing 13 lncRNAs were constructed. According to the risk score, the patients were divided into high-risk group and low-risk groups. The K-M survival curve showed that the overall survival time (OS) of the high-risk group was shorter than that of the low-risk group ($P=3e-10$). The AUC of the risk feature for predicting 1, 3 and 5-year survival rates was 0.762, 0.773 and 0.925, respectively, showing excellent predictability. The high expression of AC002331.1, LINC01068 and RP11-2N1.2 was closely related to the poor prognosis of patients.**Conclusion** The lncRNA features constructed in this study can better predict the overall survival time of patients with esophageal cancer and may be a potential therapeutic target for esophageal cancer patients.

Key words:TCGA database;Esophageal cancer;lncRNA;Prediction features;Biomarkers

食管癌(esophageal cancer, EC)是世界上第 7 大常见癌症,也是全球第 6 大致命的癌症,主要包括食管鳞状细胞癌(ESCC)和食管腺癌(EAC)两种亚型^[1]。食管腺癌在欧美国家占主导地位,而食管鳞状细胞

癌在我国占主导地位^[2]。目前,手术是治疗早期食管癌的首选方法,但因其恶性程度高、进展快、预后差,很多患者一经诊断就失去了手术机会。虽然现有的治疗方法在一定程度上提高了食管癌的生存率,但其 5 年生存率仅为 20%^[3]。现阶段,尚无针对食管癌作用的靶向药物。因此,寻找早期诊断食管癌分子标志物,探索发病机制和治疗靶点具有重要意义。长链非编码 RNA(long non-coding RNA, lncRNA)是长度超过 200 个核苷酸的非编码 RNA,广泛存在于基因组中,由于其在许多生物过程中起关键作用而受

作者简介:黄向东(1985.11-),男,河南永城人,硕士,主治医师,主要从事胸外科疾病的诊疗工作及胸部肿瘤的研究

通讯作者:张兴伟(1967.3-),男,辽宁朝阳县人,本科,主任医师,副教授,主要从事胸外科疾病的诊疗工作

到越来越多研究的关注^[4]。目前有研究表明^[5], lncRNAs 在肿瘤增殖、转移和微环境中发挥重要作用,并有望成为诊断癌症、预后、转移和耐药性的生物标志物。此外,lncRNA 在不同肿瘤组织中存在着表达差异,并与肿瘤细胞的增殖侵袭以及不良预后相关^[6]。本研究基于癌症基因组图谱(The Cancer Genome Atlas,TCGA)中的食管癌表达数据,探索与食管癌预后相关的 lncRNA,构建食管癌患者预后预测特征,期待为食管癌患者的潜在治疗靶点提供新

的证据支持。

1 资料与方法

1.1 数据来源 基于癌症基因组图谱(TCGA)数据库(<https://cancergenome.nih.gov/>)中下载的食管癌患者的 RNA-seq 数据和相应的临床信息。排除没有完整临床数据和生存信息的样本,共得到了 158 个食管癌组织样本和 11 个正常邻近组织进行后续分析,见表 1。基于 GENCODE 项目基因注释文件(版本 33, GRCh38)^[7]对基因符号和生物型进行注释。

表 1 样本中患者的临床数据统计

Clinical characteristics	Patients(n=158)	Percentage(%)	Clinical characteristics	Patients(n=158)	Percentage(%)
T			Age(y)		
T ₁	25	15.82	≤65	98	62.03
T ₂	41	25.95	>65	60	37.97
T ₃	85	53.80	Gender		
T ₄	5	3.16	Female	23	14.56
T _x	2	1.27	Male	135	85.44
N			Stage		
N ₀	66	41.77	Stage I	17	10.76
N ₁	72	45.57	Stage II	71	44.94
N ₂	11	6.96	Stage III	55	34.81
N ₃	5	3.17	Stage IV	15	9.49
N _x	4	2.53			
M					
M ₀	143	90.51			
M ₁	15	9.49			

1.2 差异表达 lncRNAs 和 mRNAs 筛选 通过使用“edgeR”包,采用 $\log_2FC \geq 2$ 且 $P \leq 0.05$ 的筛选标准,鉴定食管癌和正常样品之间差异表达的 mRNAs (DEGs)和差异表达的 lncRNAs (DELs)。并通过“heatmap”和“ggplot2”包绘制 DEG 和 DEL 的热图和火山图。

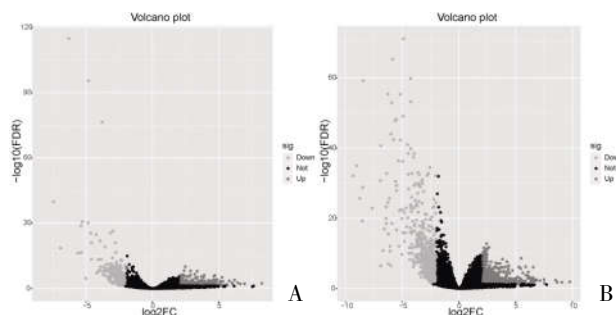
1.3 预后模型构建 基于单变量 Cox 回归分析评估总生存期(Overall survival,OS)和 lncRNA 表达水平之间的相关性。对于 $P < 0.05$ 的 lncRNA,进行 LASSO 回归分析和多变量 Cox 回归分析,计算各选择 lncRNA 的预后总和。基于 Cox 系数和基因表达值的线性组合计算风险评分,风险评分计算如下:风险评分=∑回归系数(lncRNA)×表达值(lncRNA)。基于“survival”和“glmnet”包,根据风险评分将样本分为高风险组或低风险组,构建预测特征。

1.4 预后模型临床价值评估 使用“timeROC”和“survival”包绘制 ROC 曲线并计算该模型的 C 指数,评估预测签名的稳定性和准确性。并通过 K-M 生存分析评估预后模型中高风险和低风险人群体的生存差异。分析构建模型中 lncRNAs 与临床病理特征(包括年龄、性别、病理分期和 TNM 状态)之间的相关性。

1.5 功能分析 通过 Pearson 相关分析来识别与预后模型中 lncRNAs 具有共表达关系的 mRNAs,并且 $|R| > 0.5$ 被认为是有意义的。利用“clusterProfiler”包对共表达的 mRNAs 进行基因本体论(GO)和京都基因与基因组百科全书(KEGG)功能富集分析,间接探讨 lncRNAs 在高风险组和低风险组人群之间的生物学功能差异。

2 结果

2.1 DELs 和 DEGs 的识别 基于 TCGA 数据库,共获得 169 个样本(158 个 EC 样本和 11 个正常样本)的数据,根据 $|\log_2FC| > 2$ 且 $P \leq 0.05$ 的纳入标准,最终鉴定出 592 个 DELs 和 1219 个 DEGs,见图 1A、图 1B。



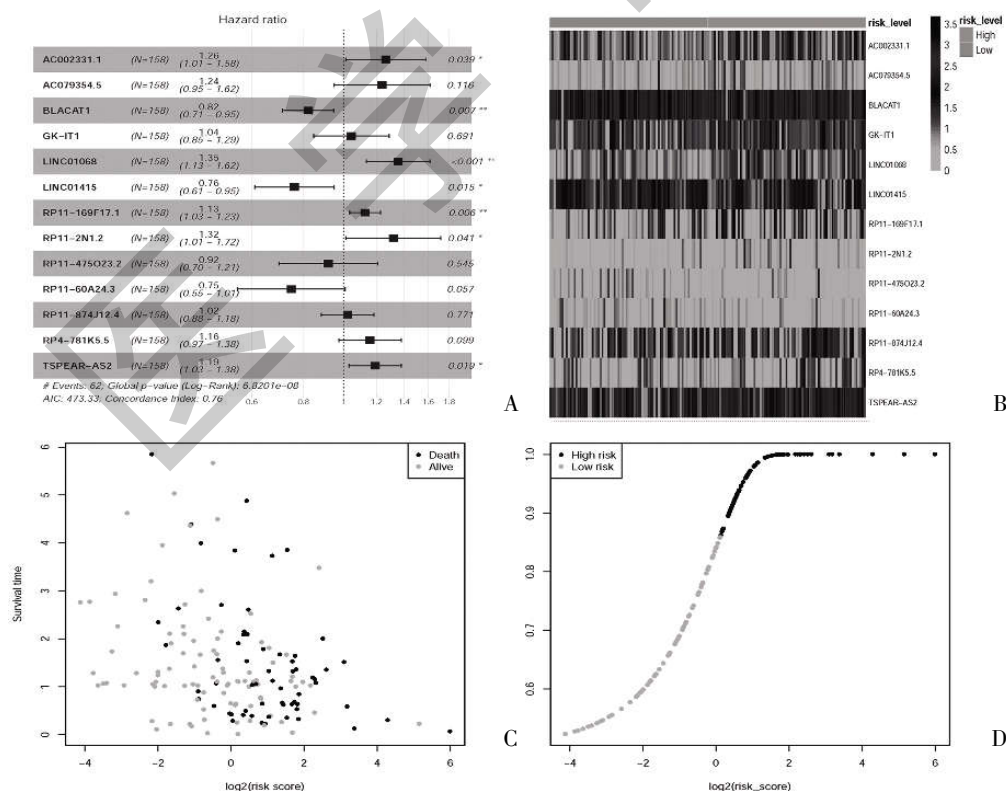
注:A;mRNAs 表达火山图;B:lncRNAs 表达火山图

图 1 基因表达火山图

2.2 预后特征构建及验证 基于单变量 Cox 回归分析和 LASSO 回归分析,总共获得了 13 个构建预后特征的 DELs 见图 2A。随后,基于多变量 Cox 回归分析获得 13 个 DELs 的回归系数(Coef)以计算风险

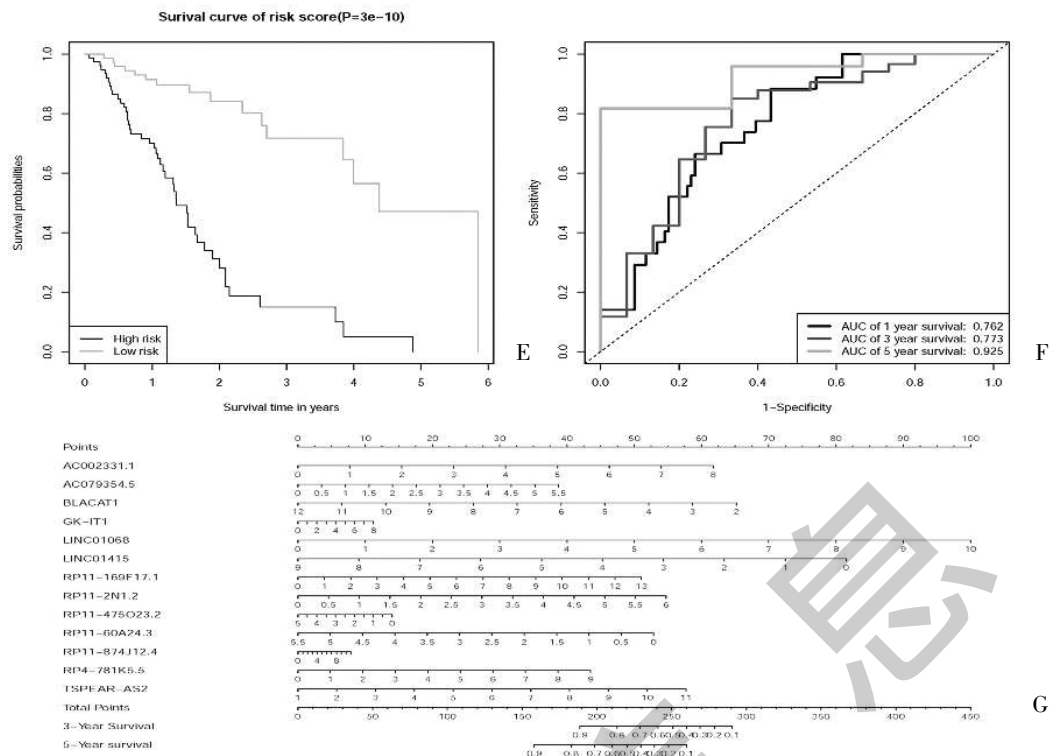
评分并构建预测特征。根据风险评分的中位数值,将 158 患者分为高危组和低危组。分析高风险和低风险组中患者的分布及 DELs 的表达水平,见图 2B~图 2D。另外,高危组患者的 OS 低于低危组,见图 2E。ROC 曲线结果表明 1、3 和 5 年的 ROC(AUC)曲线下面积分别为 0.762、0.773 和 0.925,见图 2F。同时构建了对应的列线图评估患者的 1~5 年生存率,见图 2G。

2.3 特征的临床相关性 根据多元 Cox 回归和 K-M 生存曲线,发现 AC002331.1、LINC01068 和 RP11-2N1.2 的高表达与患者的不良预后密切相关,见图 3A~图 3C。此外,13 个 DELs 与食管癌临床特征分析表明,BLACAT1 和 LINC01068 在 M0 的中位表达值高于 M1。RP4-781K5.5 和 LINC01415 在 65 岁以上人群中高于 65 岁以下人群,而 GK-IT1 在 65 岁以下人群中高于 65 岁以上人群。TSPEAR-AS2 和 GK-IT1 的高表达水平与患者淋巴结转移相关。LINC01415 在无淋巴结转移患者中的表达中值高于有淋巴结转移患者。此外,RP4-781K5.5 在男性中的中值表达值高于女性,见图 3D~图 3L。



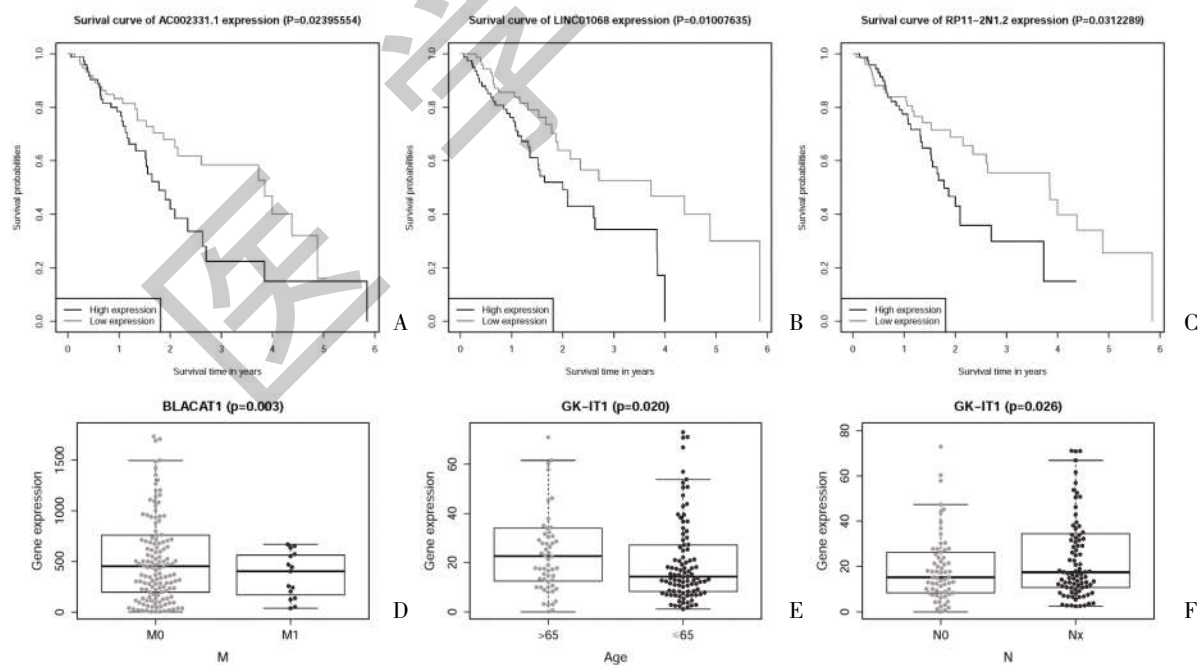
注:A:森林图展示了关键 lncRNAs 的 HR 和 95%置信区间。* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; B:lncRNAs 表达谱热图;C:低危组和高危组患者的生存状况;D 风险特征的排序和风险组的分布;E:高、低危组患者的总生存分析;F:1、3、5 年患者 ROC 曲线;G:用于预测 EC 患者 1、3 和 5 年 OS 概率的综合列线图

图 2 lncRNA 预后模型



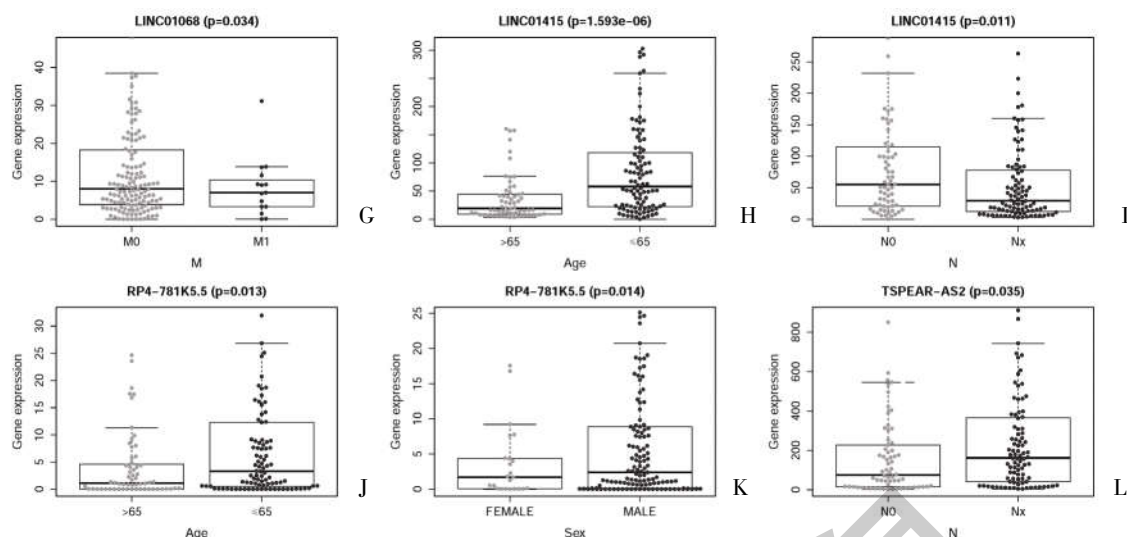
注:A:森林图展示了关键 lncRNAs 的 HR 和 95%置信区间。 $^*P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$;B:lncRNAs 表达谱热图;C:低危组和高危组患者的生存状况;D 风险特征的排序和风险组的分布;E:高、低危组患者的总生存分析;F:1、3、5 年患者 ROC 曲线;G:用于预测 EC 患者 1、3 和 5 年 OS 概率的综合列线图

图 2 lncRNA 预后模型(续)



注:A~C:AC002331.1、LINC01068、RP11-2N1.2;D:BLACAT1 在 M 分期中的表达;E:GK-IT1 表达和年龄;F:GK-IT1 表达和淋巴结转移;G:LINC01068 表达和 M 分期;H:LINC01415 表达和年龄;I:LINC01415 表达淋巴结转移;J:RP4-781K5.5 表达和年龄;K:RP4-781K5.5 表达和性别;L:TSPEAR-AS2 表达和淋巴结转移

图 3 Kaplan Meier 生存分析及 EC 患者 lncRNAs 表达与临床病理因素的关系

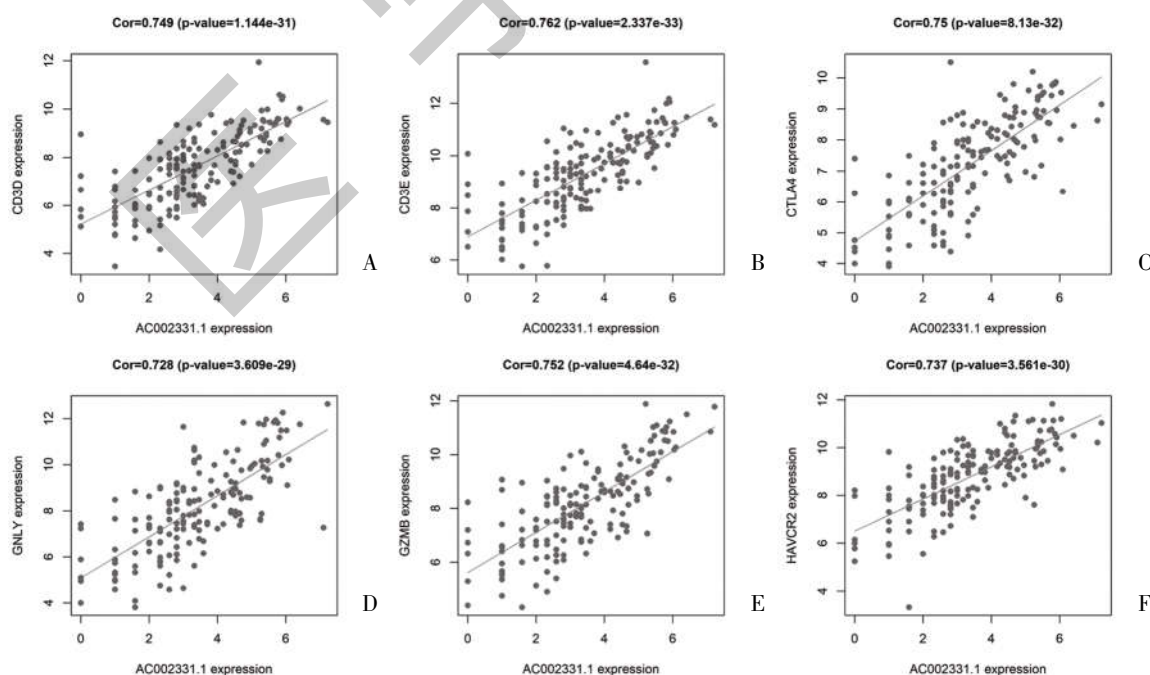


注:A~C:AC002331.1、LINC01068、RP11-2N1.2;D:BLACAT1 在 M 分期中的表达;E:GK-IT1 表达和年龄;F:GK-IT1 表达和淋巴结转移;G:LINC01068 表达和 M 分期;H:LINC01415 表达和年龄;I:LINC01415 表达淋巴结转移;J:RP4-781K5.5 表达和年龄;K:RP4-781K5.5 表达和性别;L:TSPEAR-AS2 表达和淋巴结转移

图 3 Kaplan Meier 生存分析及 EC 患者 lncRNAs 表达与临床病理因素的关系(续)

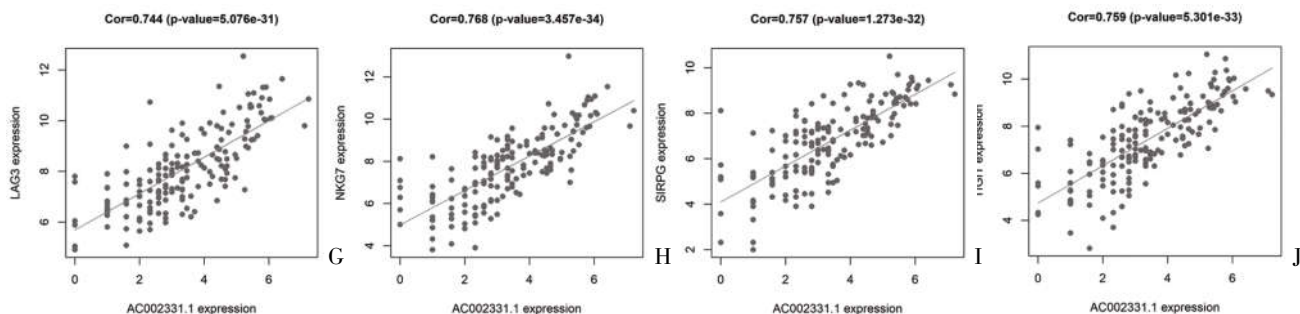
2.4 功能分析 Pearson 相关性分析分析了与 lncRNA 共表达关系的 DEGs,探索 3 种与预后相关 lncRNAs 的功能和潜在分子机制。以 $R>0.5$ 作为筛选标准。对于 LINC01068, 没有 DEGs 符合标准。RP11-2N1.2 有 9 个符合标准的 DEGs。对于 AC002331.1,共收集了 375 个 DEGs,并列出了前 10 个共表达 DEGs,见

图 4。根据 GO 分析结果显示,ACC02331.1 与 MHC 蛋白结合、MHC 蛋白复合物结合、细胞因子受体活性等功能有关,KEGG 分析结果表明与 ACC02331.1 具有共表达的 DEGs 涉及到细胞粘附分子和抗原加工和呈递等相关通路,见图 5。



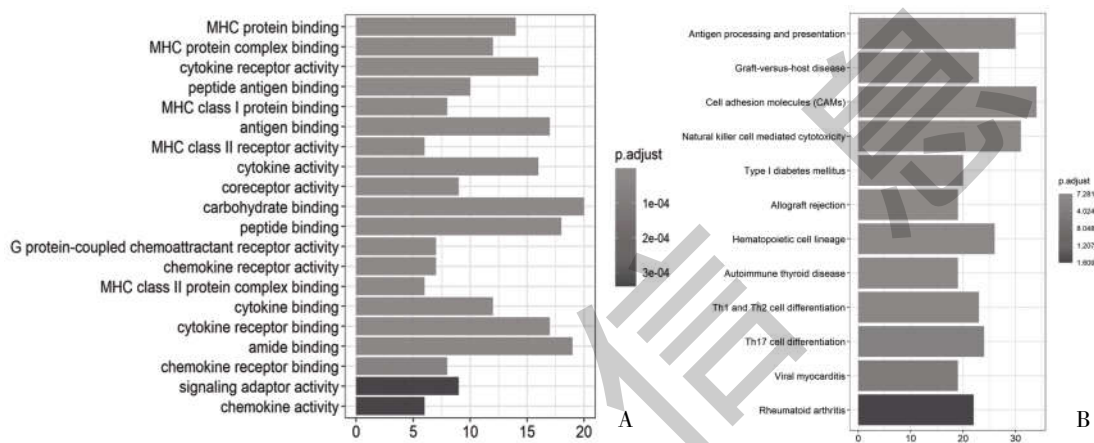
注:A~J:CD3D、CD3E、CTLA4、GNLY、GZMB、HAVCR2、LAG3、NKG7、SIRPG、TIGIT

图 4 与 AC002331.1 相关的前 10 个共表达基因的 Pearson 相关分析



注:A~J:CD3D、CD3E、CTLA4、GNLY、GZMB、HAVCR2、LAG3、NKG7、SIRPG、TIGIT

图 4 与 AC002331.1 相关的前 10 个共表达基因的 Pearson 相关分析(续)



注:A:GO 分析;B:KEGG 分析

图 5 AC002331.1 的共表达蛋白编码 mRNAs 的功能富集分析

3 讨论

食管癌是消化系统恶性肿瘤,预后不良。虽然,近年来对于食管癌的治疗取得了一定进展,但由于食管癌的恶性程度高、复发转移率高,因此预后较差^[8]。基于来自 TCGA 公共数据库的数据集,多项研究评估了 lncRNA 在各种癌症类型中的预后价值^[9-11]。这些研究结果表明该方法是可行的。例如,PVT1 的高表达与食管癌预后不良有关^[12],而 LINC01296 可促进细胞增殖、迁移和侵袭,与食管癌预后不良有关^[6]。鉴于 lncRNA 与多种肿瘤有着密切的关系,本研究建立了 lncRNA 相关食管癌患者预后预测特征,特征中 lncRNAs 对未来食管癌的诊断及治疗提供了新的方向。

本研究基于 TCGA 数据库中食管癌患者的表达数据,构建了包含 13 个 lncRNA 的食管癌的预后预测特征,包括 AC002331.1、AC079354.5、BLACAT1、GK-IT1、LINC01068、LINC01415、RP11-169F17.1、RP11-2N1.2、RP11-475O23.2、RP11-60A24.3、

RP1174-17RP115.5、TSPEAR-AS2。根据风险评分将食管癌患者分为高、低风险两组,通过绘制 K-M 生存曲线发现高风险组患者的生存时间显著低于低风险组($P=3e-10$)。特征的 1、3 和 5 年的 AUC 值分别为 0.762、0.773 和 0.925,该特征的 C 指数为 0.764,均表现出较好的预测能力。

在预测模型和临床病理数据相关性分析中发现,BLACAT1 和 LINC01068 在 M0 患者中的中位表达值高于 M1。既往研究表明^[13],BLACAT1 的高表达与实体瘤患者较短的总生存期、淋巴结转移、较差 TNM 分期以及肿瘤分级相关。而 LINC01068 在肿瘤中的作用未得到进一步探究。此外,本研究发现 RP4-781K5.5 和 LINC01415 在 65 岁以下患者的中位表达值低于 65 岁以上患者,说明对于 65 岁以下患者,这两种 lncRNA 的异常表达可能会增加食管癌风险。相反,GK-IT1 不同于 RP4-781K5.5 和 LINC01415,它在 65 岁以上的患者中具有更高的表达水平。本研究还发现存在远处淋巴结转移的患者

中 GK-IT1 的表达水平高于无淋巴结转移的患者。LINC01415 在没有淋巴结转移的患者中显示出更高的表达水平。此外,TSPEAR-AS2 的高表达可能与远处淋巴结转移有关,AC002331.1、LINC01068 和 RP11-2N1.2 在肿瘤组织中的高表达与患者的不良预后密切相关,并进一步分析它们的潜在功能。

基于 Pearson 相关系数分析获得了与 lncRNAs (AC002331.1、LINC01068、RP11-2N1.2)具有共表达关系的 mRNAs。其中,对与 AC002331.1 具有相关 mRNAs 进行功能分析,发现具有多种生物学功能,主要与 MHC 蛋白结合、MHC 蛋白复合物结合、细胞因子结合受体相关。KEGG 分析表明,AC002331.1 可能与细胞粘附分子、抗原加工和呈递、JAK-STAT 信号通路和 NF- κ B 信号通路密切相关。而这些功能与癌症的发生、发展存在着密切的关联性。并且 JAK-STAT 信号通路在细胞免疫、分裂和死亡以及肿瘤形成中起着至关重要的作用^[14]。JAK-STAT 信号通路的突变和信号调控与白血病^[15]和骨肉瘤^[16]密切相关。此前的研究也表明,JAK-STAT 通路可能参与了食管癌的发生^[17]。Li B 等^[18]和 Huang H 等^[19]的研究也表明,NF- κ B 信号通路食管癌的发展过程密切相关。

综上所述,通过分析 TCGA 数据库中食管癌的数据,探索其中与食管癌预后相关的 lncRNA,建立食管癌患者预后预测特征,该特征中的 lncRNAs 可能为食管癌患者的诊断及治疗提供一定思路。

参考文献:

- [1]Siewert JR,Ott K.Are squamous and adenocarcinomas of the esophagus the same disease? [J].Semin Radiat Oncol,2007,17(1): 38-44.
- [2]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.
- [3]Siegel RL,Miller KD,Jemal A.Cancer statistics, 2020 [J].CA Cancer J Clin,2020,70(1):7-30.
- [4]Esteller M.Non-coding RNAs in human disease [J].Nat Rev Genet,2011,12(12):861-874.
- [5]Schmitt AM,Chang HY.Long Noncoding RNAs in Cancer Pathways[J].Cancer Cell,2016,29(4):452-463.
- [6]Wang B,Liang T,Li J.Long noncoding RNA LINC01296 is associated with poor prognosis in ESCC and promotes ESCC

cell proliferation, migration and invasion [J].Eur Rev Med Pharmacol Sci,2018,22(14):4524-4531.

- [7]Frankish A,Diekhans M,Ferreira AM,et al.GENCODE reference annotation for the human and mouse genomes [J].Nucleic Acids Res,2019,47(D1):D766-D773.
- [8]Liu J,Xie X,Zhou C,et al.Which factors are associated with actual 5-year survival of oesophageal squamous cell carcinoma? [J].Eur J Cardiothorac Surg,2012,41(3):e7-e11.
- [9]Yang S,Zhou Y,Zhang X,et al.The prognostic value of an autophagy-related lncRNA signature in hepatocellular carcinoma [J].BMC Bioinformatics,2021,22(1):217.
- [10]Liu Y,Liu B,Jin G,et al.An Integrated Three-Long Non-coding RNA Signature Predicts Prognosis in Colorectal Cancer Patients [J].Front Oncol,2019,9:1269.
- [11]Wang JJ,Huang YQ,Song W,et al.Comprehensive analysis of the lncRNA-associated competing endogenous RNA network in breast cancer [J].Oncol Rep,2019,42(6):2572-2582.
- [12]Xu Y,Li Y,Jin J,et al.LncRNA PVT1 up-regulation is a poor prognosticator and serves as a therapeutic target in esophageal adenocarcinoma [J].Mol Cancer,2019,18(1):141.
- [13]Liu N,Hu G,Wang H,et al.LncRNA BLACAT1 regulates VASP expression via binding to miR-605-3p and promotes glioma development [J].J Cell Physiol,2019,234 (12):22144 - 22152.
- [14]Aaronson DS,Horvath CM.A road map for those who don't know JAK-STAT [J].Science,2002,296(5573):1653-1655.
- [15]Venugopal S,Bar-Natan M,Mascarenhas JO.JAKs to STATs: A tantalizing therapeutic target in acute myeloid leukemia [J].Blood Rev,2020,40:100634.
- [16]Subramaniam D,Angulo P,Ponnuram S,et al.Suppressing STAT5 signaling affects osteosarcoma growth and stemness [J].Cell Death Dis,2020,11(2):149.
- [17]You Z,Xu D,Ji J,et al.JAK/STAT signal pathway activation promotes progression and survival of human oesophageal squamous cell carcinoma [J].Clin Transl Oncol,2012,14(2):143-149.
- [18]Li B,Li YM,He WT,et al.Knockdown of DDX46 inhibits proliferation and induces apoptosis in esophageal squamous cell carcinoma cells [J].Oncol Rep,2016,36(1):223-230.
- [19]Huang H,Wei L,Qin T,et al.Circular RNA ciRS-7 triggers the migration and invasion of esophageal squamous cell carcinoma via miR-7/KLF4 and NF- κ B signals [J].Cancer Biol Ther, 2019,20(1):73-80.

收稿日期:2022-05-05;修回日期:2022-06-09

编辑/肖婷婷