

·基础研究·

基于生物信息学途径食管鳞癌 lncRNA 预后风险模型的构建

蔡河源, 谢春莹, 邹健勇, 罗红鹤

(中山大学附属第一医院胸外科, 广东 广州 510080)

摘要:【目的】基于癌症基因组图谱(TCGA)数据库食管鳞癌 lncRNA 差异表达,构建食管鳞癌预后风险模型。【方法】从 TCGA 肿瘤数据库下载食管鳞癌基因表达数据及临床信息,使用 R 4.0.3 软件提取差异表达 lncRNA,进一步使用 COX 回归分析筛选模型 lncRNA,并构建预后风险模型 Riskscore;按照风险评分中位值将食管鳞癌患者分为低风险组和高风险组,进一步比较高低风险组生存预后;分析比较风险模型及其他临床特征对食管鳞癌生存预后的预测性能;分析风险评分与其他临床特征相关性;采用主成分分析(PCA)及基因富集分析(GSEA)探索高低风险组基因分布差异。【结果】得到差异表达 lncRNA 174 个,食管鳞癌组织表达上调 126 个,表达下调 48 个。COX 回归分析显示 *AL033384.1*, *AC108449.2* 可作为食管鳞癌的独立预后因素,风险模型公式为 $\text{Riskscore} = 1.303 \times \text{AL033384.1} - 1.525 \times \text{AC108449.2}$,且低风险组总体生存率高于高风险组($P < 0.001$),主成分分析显示模型 lncRNA 能较好区分高低风险组,是高低风险组分布的主要影响因素。高低风险组存在细胞通路差异,高风险组主要富集在表皮细胞的角质化和细胞内的高代谢等过程。该风险模型 1、2 和 3 年预测性能分别为 0.750、0.768 和 0.796,优于其他临床特征如 TNM 分期的预测性能。【结论】基于 TCGA 数据库筛选得到的 *AL033384.1*, *AC108449.2* 建立的预后风险模型对预测食管鳞癌患者的预后具有一定的临床价值,低风险组患者总体生存率更长,预后更好。

关键词:食管鳞癌;长链非编码 RNA;风险评分;预后

中图分类号:R735.1

文献标志码:A

文章编号:1672-3554(2021)05-0729-09

DOI: 10.13471/j.cnki.j.sun.yat-sen.univ(med.sci).2021.0511

Establishment of a Prognostic Risk Model of lncRNA in Esophageal Squamous Cell Carcinoma by Bioinformatics

CAI He-yuan, XIE Chun-ying, ZOU Jian-yong, LUO Hong-he

(Department of Thoracic Surgery, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou 510080, China)

Correspondence to: LUO Hong-he; E-mail: luohhzm@163.com

Abstract:【Objective】To establish a prognostic risk-model of lncRNA in esophageal squamous cell carcinoma based on cancer genome atlas (TCGA) database.【Methods】The gene expression data and clinical information of esophageal squamous cell carcinoma were downloaded from the TCGA database. The R 4.0.3 software was used to identify the differential expression of lncRNA, Cox regression analysis was used to screen the lncRNA of model, and a prognostic risk model - Riskscore was established. According to the median of Riskscore, the patients were divided into high and low risk group, and the survival prognosis between the two groups were further compared. The diagnostic performance of Riskscore and other clinical features of survival prognosis were further analyzed. In addition, the correlation between Riskscore and other clinical features was analyzed. Finally, principal component analysis (PCA) and gene enrichment analysis (GSEA) were used to explore the differences in gene distribution between the two groups.【Results】Using R 4.0.3 software, 174 lncRNA were differentially expressed, in which 126 were up-regulated and 48 were down-regulated. Univariate and multivariate COX regression analysis showed *AL033384.1* and *AC108449.2* could be independent prognostic factors of esophageal squa-

收稿日期:2021-04-28

基金项目:卫生部医药卫生科技发展中心项目(W2014RQ22)

作者简介:蔡河源,硕士生,研究方向:胸外科肿瘤机制研究;罗红鹤,通信作者,E-mail:luohhzm@163.com

mous cell carcinoma. The risk-model formula is $Riskscore = 1.303 \times AL033384.1 - 1.525 \times AC108449.2$, and the overall survival in the low risk group was longer than that of the high risk group ($P < 0.001$). Principal component analysis showed the model-lncRNA could better distinguish the high and low risk groups, and it was the main influencing factor for the distribution in the two groups. Further GSEA enrichment analysis revealed gene functional annotation differences between the two groups, and the high risk group was mainly enriched in the process of keratinization and intracellular hypermetabolism. Besides, the 1-year, 2-year and 3-year prediction performance of this risk-model was 0.750, 0.768 and 0.796, respectively, which was better than those of other clinical characteristics such as TNM staging. 【Conclusion】 Based on TCGA database, the prognostic risk-model, established by *AL033384.1* and *AC108449.2*, has important clinical value in predicting the prognosis of patients with esophageal squamous cell carcinoma, and the model proves the low risk group has longer overall survival and better prognosis.

Key words: esophageal squamous cell carcinomas; lncRNA; riskscore; prognosis

[J SUN Yat-sen Univ (Med Sci), 2021, 42(5): 729-737]

食管癌是全球最常见肿瘤之一,最新全球癌症统计,其发病率占第11位,而死亡率高居第6位^[1]。食管癌按照病理类型分为食管鳞癌(esophageal squamous cell carcinoma, ESCC)和食管腺癌(esophageal adenocarcinoma, EAC)。在我国,ESCC占食管癌的90%左右,且90%病理分期为中晚期^[2],预后非常差,癌症死亡率居第4位^[3],5年整体生存率不到30%^[4-6],亟需寻找新的早期诊断及治疗方法。长链非编码RNA(long non-coding RNA, lncRNA)是指长度大于200 bp,且转录后不编码翻译产生蛋白质的一类RNA,参与并调节人体70%的基因的复制,转录,翻译等细胞生物学过程,从而调节机体各项生命活动^[7-10]。*Linc02471*, *UCA1*、*CASC9*、*DLEU1*等,lncRNA通过调控转录翻译促进ESCC的发生发展,并可以作为ESCC的诊断治疗潜在靶点^[11-14];但lncRNA与ESCC预后关系尚不明确。本研究基于癌症基因组图谱(the cancer genome atlas, TCGA)数据库食管癌转录组基因表达谱,分析ESCC组织与正常食管上皮组织lncRNA表达差异,并构建ESCC预后风险模型,旨在预测ESCC患者预后,寻找ESCC潜在生物标记物,指导ESCC临床诊治。

1 材料与方法

1.1 数据下载整理

从TCGA数据库(<https://cancergenome.nih.gov/>)下载食管癌基因表达数据及临床资料。剔除

病理类型为腺癌的病例,剔除生存时间小于30 d的病例,数据集包含基因转录表达数据样本90例,80例为ESCC组织样本,10例为癌旁正常组织样本,进一步提取数据得到食管鳞癌lncRNA表达矩阵。数据采集时间为2021年1月16日。

1.2 差异基因筛选

采用R 4.0.3软件limma包对lncRNA表达矩阵进行差异基因筛选,并设定错误发现率(false discovery rate, FDR) < 0.05和log₂ |fold change| > 1作为临界值,得到ESCC与正常食管组织表达差异的lncRNA。并使用pheatmap包和ggplot2包绘制差异基因热图及火山图。

1.3 预测模型的构建

合并ESCC样本临床生存资料及差异lncRNA表达数据,使用R 4.0.3软件survival包进行单因素COX及多因素COX回归分析,α = 0.05。选取单因素COX及多因素COX分析均有统计学差异的lncRNA构建lncRNA风险预测模型。

模型构建公式为:

$$Riskscore = \sum_{i=1}^N (expi \times coefi),$$

其中*N*, *expi*, *coefi*分别代表模型lncRNA, lncRNA表达量和表达系数。

计算80例ESCC样本的Riskscore,并按中位数排列由低到高将80例ESCC样本分为低风险组及高风险组,进一步使用Kaplan-Meier生存曲线分析高低风险组生存预后差异。进一步使用pheatmap包绘制高低风险组生存状态分布图及模型lncRNA表达热图,使用survival ROC包绘制ESCC样本1

年、2年和3年总体生存率ROC曲线图。

1.4 临床特征与 Riskscore 模型预测性能比较

使用 R 4.0.3 软件 survival 包对 ESCC 临床特征及 Riskscore 分别进行单因素 COX 及多因素 COX 回归分析,探索 ESCC 独立预后影响因素,并绘制 ROC 曲线比较不同临床特征和风险模型的预测性能。

1.5 临床特征与 Riskscore 模型评分相关性分析

分别提取 ESCC 样本临床特征如性别、年龄、肿瘤 TMN 分期与对应模型 Riskscore 评分合并,使用 ggpubr 包 wilcoxon 秩和检验进行相关性分析, $\alpha=0.05$,并绘制相关性箱式图。

1.6 高低风险组基因表达矩阵主成分分析

采用主成分分析(Principal component analysis, PCA)分别对 ESCC 样本 lncRNA 表达矩阵及模型 lncRNA 矩阵进行降维分析,并使用 scatterplot3d 包绘制高低风险组 3D 散点图。

1.7 GO,KEGG 功能富集分析

通过基因富集分析(Geneset enrichment analysis, GSEA)对高低风险组进一步行基因本体(gene ontology, GO),京都基因与基因组百科全书(kyoto encyclopedia of genes and genomes, KEGG)功能富集分析,探索高低风险组在不同细胞通路基因富集的差异性。从(<http://www.gsea-msigdb.org/>)下载 GSEA 软件,GO 基因集,KEGG 基因集,分析并绘制基因富集图。

2 结果

2.1 食管鳞癌样本临床特征

90 例食管上皮测序样本纳入分析,80 例为 ESCC 组织样本,10 例为癌旁正常组织样本。80 例 ESCC 组织样本中,男性 68 人,女性 12 人,平均年龄 58.43(30~90)岁,肿瘤分期以 II, III 期多见,T 分期中 T2-3 占 85%,N 分期中 N0 期最为高发,N1 次之,且 87.5% ESCC 患者未见远处转移,其他临床特征见表 1。

2.2 差异表达的 lncRNA 筛选结果

80 例 ESCC 组织样本,10 例癌旁正常组织样本经筛选得到差异 lncRNA 174 个,其中 ESCC 样本表达上调 lncRNA 126 个,表达下调 lncRNA 48 个(图 1)。

2.3 差异 lncRNA 单因素及多因素 COX 回归分析

结果显示单因素 COX 回归有 3 个 lncRNA 有统计学意义($P < 0.05$),多因素 COX 回归分析有 2 个 lncRNA 有统计学意义($P < 0.05$),*AC108449.2* 和 *AL033384.1* 可作为 ESCC 的独立预后因素,*AC108449.2* 的风险比为 0.218,是 ESCC 预后的保护因素,*AL033384.1* 的风险比为 3.681,是 ESCC 预后的危险因素(表 2、3)。

2.4 lncRNA 风险模型构建及高低风险组生存分析

经单因素及多因素 COX 回归分析得到模型算式

$$\text{Riskscore} = 1.303 \times \text{AL033384.1} - 1.525 \times \text{AC108449.2},$$

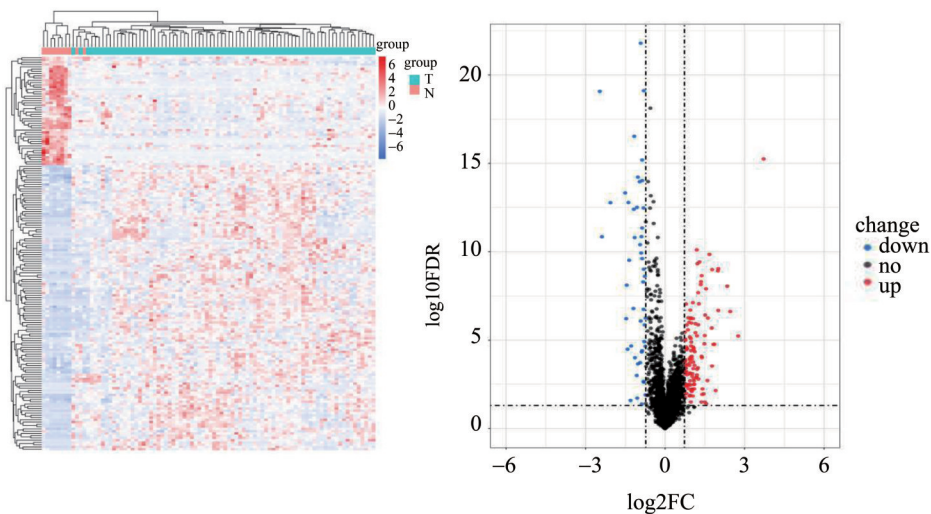


图 1 差异表达 lncRNA 热图及火山图

Fig. 1 Heat map and volcano plot of differentially expressed lncRNA

表1 80例食管鳞癌组织样本临床特征

Table 1 Clinical characteristics of 80 cases of esophageal squamous cell carcinoma [n(%)]

Characteristics	Sample size	Characteristics	Sample size
Gender		Age	
Male	68(85.00)	< 60 years	47(58.75)
Female	12(15.00)	≥ 60 years	33(41.25)
Tumor staging		N staging	
I	7(8.75)	N0	46(57.50)
II	47(58.75)	N1	25(31.25)
III	21(26.25)	N2	5(6.25)
IV	3(3.75)	N3	1(1.25)
Not sure	2(2.50)	Nx	3(3.75)
T staging		M staging	
T1	8(10.00)	M0	70(87.50)
T2	28(35.00)	M1	4(5.00)
T3	40(50.00)	Mx	6(7.50)
T4	4(5.00)		

Riskscore 中位值为 1.025, 根据中位值由低到高排列将 ESCC 样本分为低风险组和高风险组。生存状态图显示相较于高风险组, 低风险组患者存活

数更多, 死亡患者存活时间更长, 且低风险组 *AL033384.1* 低表达, *AC108449.2* 高表达, 而高风险组相反, 结果见图 2。进一步 Kaplan-Meier 生存分析显示低风险组 1 年, 3 年, 5 年生存率均高于高风险组 ($P < 0.001$), 预后较好(图 3)。Riskscore 模型时间依赖性 ROC 曲线图显示该模型预测 ESCC 患者 1 年、2 年和 3 年总体生存率效能分别为 0.750、0.768 和 0.796, 具有良好的预测性能(图 4A)。

2.5 临床特征单因素及多因素 COX 回归分析及 ROC 曲线

对年龄、stage 肿瘤分期、T 分期、M 分期、N 分期和 Riskscore 进行单因素及多因素 COX 回归分析。单因素 COX 回归分析显示 stage 分期、N 分期和 Riskscore 有统计学意义 ($P < 0.05$), 进一步对 stage 分期、N 分期和 Riskscore 做多因素 COX 回归分析, 显示只有 Riskscore 有统计学意义 ($P < 0.05$), 提示风险模型 Riskscore 可作为 ESCC 生存独立预后因素(表 4、5)。进一步比较风险模型 Riskscore 和各临床特征的预测效能, 预测 1 年生存率的 ROC 曲线下面积显示 Riskscore 模型为 0.750, 高于临床特征 N 分期 0.695, stage 肿瘤分期 0.667, T 分期 0.550, M 分期 0.549, 年龄 0.520, 提示该模型相比于传统 TNM 分期能更准确判断 ESCC 患者预后(图 4B)。

表2 差异表达 lncRNA 单因素 COX 回归分析

Table 2 Univariate Cox regression analysis of differentially expressed lncRNA

Variables	<i>b</i>	<i>S_b</i>	Wald χ^2	<i>P</i>	HR	HR95%CI
<i>LINC00365</i>	1.268	0.600	4.472	0.034	3.555	(1.097, 11.518)
<i>AC108449.2</i>	-1.604	0.723	4.919	0.027	0.201	(0.049, 0.830)
<i>AL033384.1</i>	1.304	0.524	6.203	0.013	3.684	(1.320, 10.282)

HR: hazard ratio; HR95%CI: 95% confidence interval of hazard ratio

表3 差异表达 lncRNA 多因素 COX 回归分析

Table 3 Multivariate Cox regression analysis of differentially expressed lncRNA

Variables	<i>b</i>	<i>S_b</i>	Wald χ^2	<i>P</i>	HR	HR95%CI
<i>AC108449.2</i>	-1.525	0.719	4.501	0.034	0.218	(0.053, 0.890)
<i>AL033384.1</i>	1.303	0.537	5.893	0.015	3.681	(1.285, 10.543)

HR: hazard ratio; HR95%CI: 95% confidence interval of hazard ratio

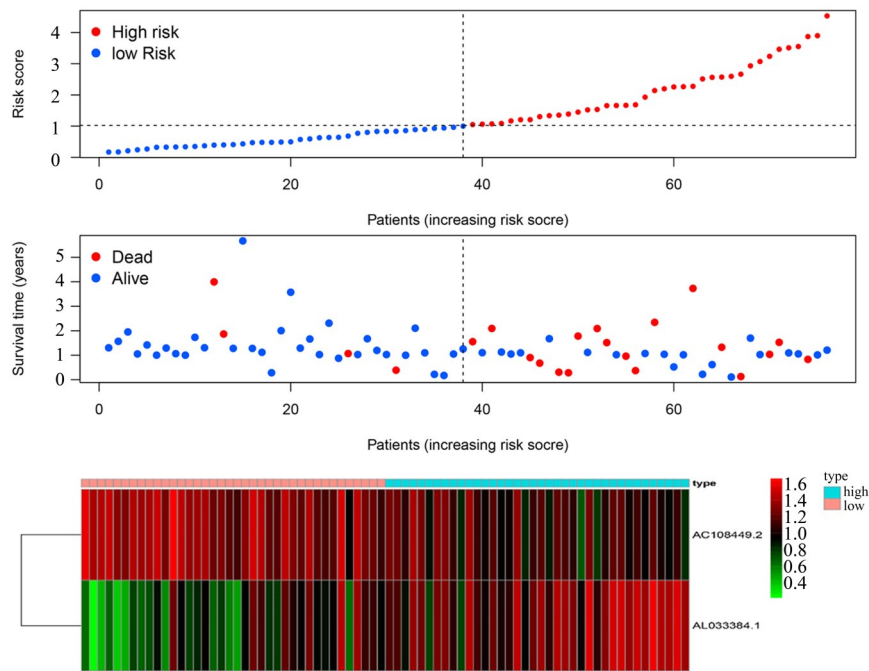


图2 高低风险组生存状态分布及 lncRNA 表达

Fig. 2 Survival-distribution plot and lncRNA-expression in high and low risk groups

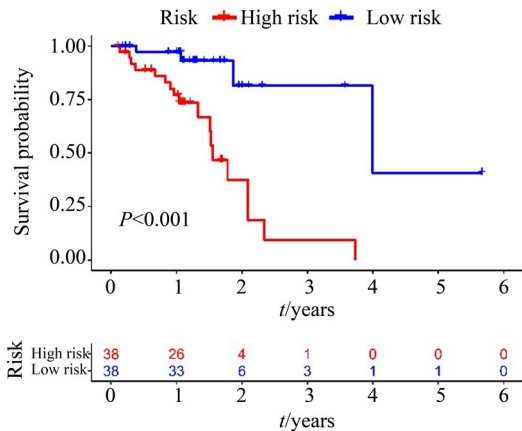


图3 高风险组和低风险组生存率的 Kaplan-Meier 生存分析曲线

Fig. 3 Kaplan-Meier survival analysis curve in high and low risk groups

2.6 临床特征与 Riskscore 相关性分析

分析各 ESCC 样本 Riskscore 评分与性别、年龄、生存状态、肿瘤分期、T 分期、N 分期和 M 分期等临床特征的相关性,结果显示男性及死亡患者的 Riskscore 评分更高($P < 0.05$),女性及生存患者的 Riskscore 评分相对较低($P < 0.05$),在年龄和 stage 分期中 Riskscore 评分差异没有统计学意义($P > 0.05$),在 T 分期, N 分期中差异性不显著, M 分期 (M0/M1: 70/4)未行相关性检验(图 5)。

2.7 高低风险组主成分分析

对 ESCC 样本 lncRNA 表达矩阵和模型 lncRNA (AC108449.2, AL033384.1) 表达矩阵按照高低风险组进行主成分分析, 3D 散点图显示原有 ESCC lncRNA 表达矩阵经降维后仍不能有效区分高低风险组患者, 而模型 lncRNA 表达矩阵经降维后高低风险患者区分较为明显, 提示模型 lncRNA (AC108449.2, AL033384.1) 是 ESCC 高低风险组分布的主要影响因素(图 6)。

2.8 高低风险组 GO 及 KEGG 基因功能富集分析

使用 GSEA 软件对高低风险组行 GO、KEGG 分析, 富集标准为 $INESI > 1$, $NOM\ p\text{-value} < 0.05$, $FDR\ q\text{-value} < 0.25$, 结果显示高风险组主要富集在表皮细胞的角质化和细胞内的高代谢等过程, 低风险组主要富集在细胞内氨基酸的甲基化过程, 提示高风险组患者食管上皮磷化过程及细胞高代谢促进了 ESCC 的发生发展, 是导致高风险组不良预后的内在影响因素(图 7)。

3 讨论

食管癌是全世界最常见的肿瘤之一, 早期难以发现, 一经诊断 90% 的患者即为中晚期, 5 年生存率不到 30%。然而, 对于食管鳞癌发病机制研究

表4 临床特征单因素COX回归分析

Table 4 Univariate Cox regression analysis of clinical characteristics

Variables	<i>b</i>	<i>S_b</i>	Wald χ^2	<i>P</i>	HR	HR95%CI
Age	0.009	0.028	0.108	0.743	1.009	(0.956, 1.066)
Stage	0.757	0.306	6.128	0.013	2.132	(1.171, 3.881)
T	0.067	0.323	0.043	0.835	1.070	(0.568, 2.015)
M	0.970	0.659	2.167	0.141	2.637	(0.725, 9.585)
N	0.819	0.277	8.708	0.003	2.268	(1.316, 3.906)
Riskscore	0.686	0.221	9.660	0.002	1.986	(1.288, 3.061)

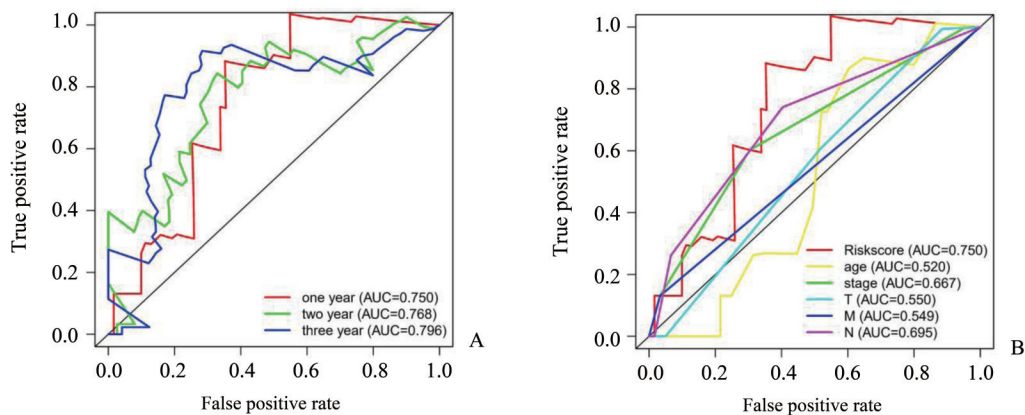
HR: hazard ratio; HR95%CI: 95% confidence interval of hazard ratio

表5 临床特征多因素COX回归分析

Table 5 Multivariate Cox regression analysis of clinical characteristics

Variable	<i>b</i>	<i>S_b</i>	Wald χ^2	<i>P</i>	HR	HR95%CI
Age	0.025	0.037	0.472	0.492	1.025	(0.955, 1.102)
Stage	0.191	1.059	0.033	0.857	1.211	(0.152, 9.653)
T	0.494	0.621	0.634	0.426	1.639	(0.486, 5.531)
M	0.154	2.610	0.003	0.953	1.166	(0.007, 194.182)
N	0.784	0.478	2.688	0.101	2.190	(0.858, 5.590)
Riskscore	0.760	0.270	7.940	0.005	2.139	(1.260, 3.629)

HR: hazard ratio; HR95%CI: 95% confidence interval of hazard ratio



A: Time-dependent ROC curves of Riskscore model for predicting 1-, 2-, and 3-year OS; B: ROC curves of Riskscore model, age, tumor stage, T stage, M stage and N stage for predicting 1-year OS.

图4 Riskscore模型及临床特征的ROC曲线图

Fig. 4 ROC curves of Riskscore model and clinical characteristics

尚有争议,近年来,随着分子研究的不断发展, lncRNA 作为明星分子被发现在基因复制、转录、表达等各个环节影响着各种癌症的发生发展。大量研究证明 lncRNA 与食管鳞癌的发生发展密不可分。Li 等^[15]人研究发现 lncRNA *NLIPMT* 可通过调节

miR-320/survivin 轴抑制食管鳞癌细胞增殖和迁移并促进细胞周期停滞、凋亡; Xu 等^[16]人指出 lncRNA *ZEB2-AS1* 可通过调节 *miR-574-3p/HMGA2* 轴促进食管鳞癌的增殖, 迁移和侵袭; Li 等^[17]研究证实 lncRNA *MIR205HG* 通过调节 *miR-214/SOX4*

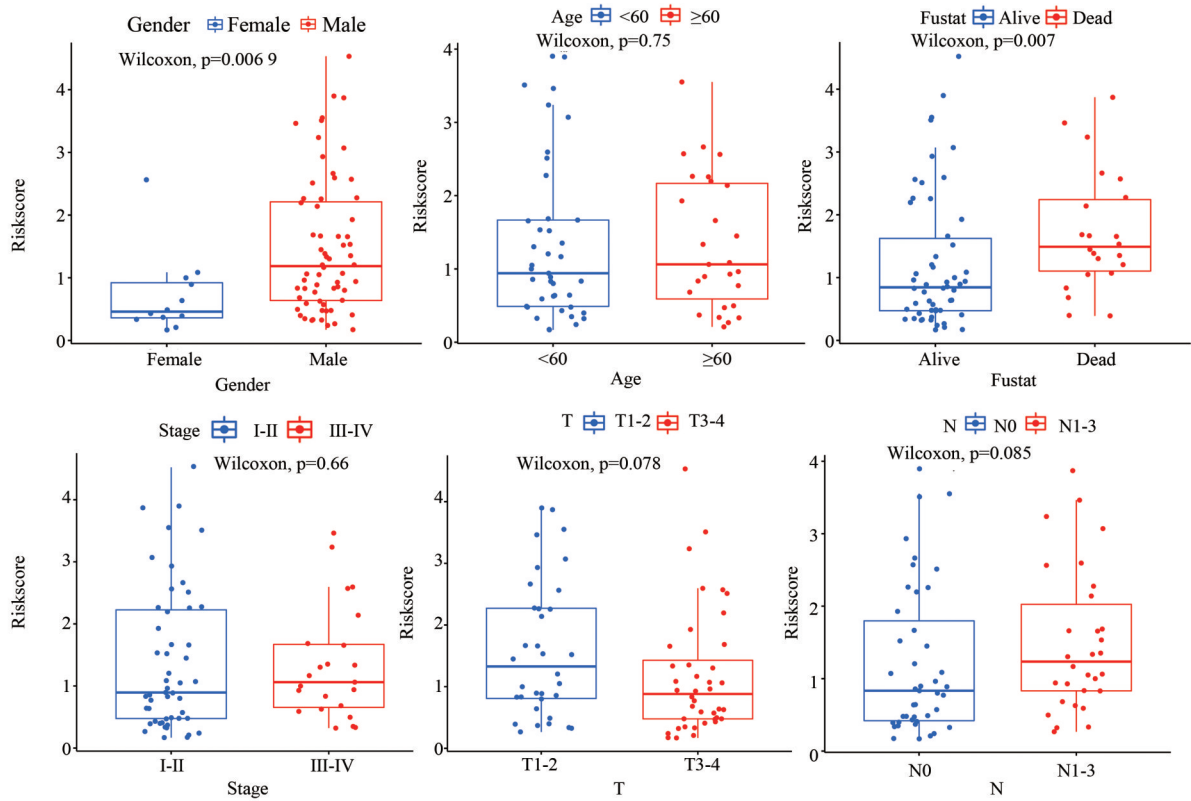
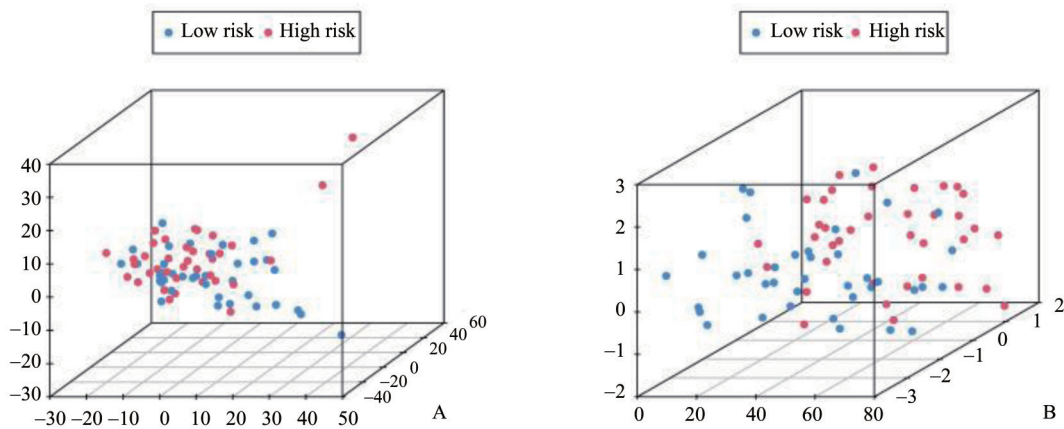


图5 临床特征与 Riskscore 相关性分析

Fig. 5 Correlation analysis between clinical characteristics and Riskscore



A: LncRNA matrix of esophageal squamous cell carcinoma; B: LncRNA matrix of Riskscore model.

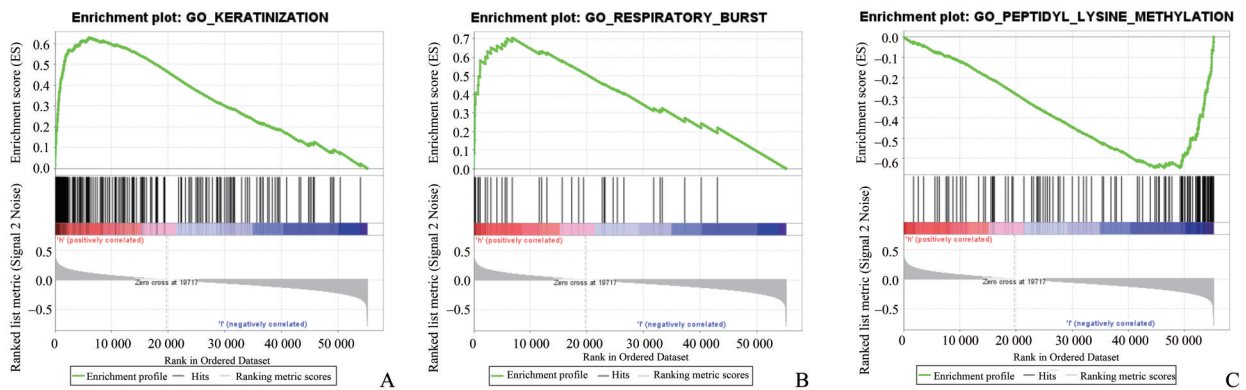
图6 高低风险组主成分分析图

Fig. 6 Principal component analysis plot in high and low risk group

轴加速食管鳞癌的进展。可见食管鳞癌发生发展与lncRNA的调控息息相关。

然而,关于lncRNA与食管鳞癌预后的关系国内尚无相关报导,已有研究提示基于4个lncRNA构建的风险模型可以有效预测胶质母细胞瘤的预后^[18],胃癌患者中表达的AC097478.1、AC097478.3、AL354719.2也可有效预测胃癌患者的预后^[19]。本

研究旨在通过分析TCGA食管鳞癌表达样本基因分子特征差异,寻找食管鳞癌预后的生物标志物,以指导临床诊疗。我们发现食管鳞癌上皮与正常食管上皮基因表达存在差异,单因素及多因素COX回归分析提示AC108449.2、AL033384.1这两个lncRNA与食管鳞癌的预后相关,且高风险组AC108449.2表达下调,AL033384.1表达上调,总体



A: Epithelial cell keratinization; B: Hypermetabolism in cell; C: Amino acid methylation.

图7 高低风险组GSEA富集分析图

Fig. 7 GSEA enrichment analysis plot in high and low risk group

生存率较差,而低风险组相反,已有研究显示 *AC108449.2* 是肾脏透明细胞癌预后因素,在高风险组中低表达,在低风险组中高表达^[20],另有研究指出 *AL033384.1* 与食管鳞癌预后相关,*AL033384.1* 高表达的食管鳞癌患者预后较差^[21],这与我们的研究发现类似。我们进一步根据预后相关基因构建了食管鳞癌预后风险模型,并将食管鳞癌样本按照模型中位值划分为高低风险组,预后风险模型的预测效能均高于TNM分期等临床特征,具有良好的预测效能,可作为食管鳞癌患者临床预后分析的补充。进一步主成分分析(PCA)提示模型 lncRNA (*AC108449.2*, *AL033384.1*) 能较好区分高低风险组的患者,GSEA富集分析提示高风险组的患者基因富集在表皮细胞的角质化及细胞的高代谢活动,提示高风险组患者通过促进表皮鳞化和加速肿瘤增殖导致了高风险组食管鳞癌的发生发展,这为进一步深入研究其内在机制奠定了基础。

然而,本研究仍存在一定的局限性。首先,纳

入模型构建的食管鳞癌样本均来自TCGA数据库,可能与我国食管鳞癌的情况有所差异。其次,我们的研究发现 Stage 肿瘤分期与模型 Riskscore 评分不相关,T分期和N分期与模型 Riskscore 具有较弱的相关性,我们认为样本量不足与其他临床因素如术后有无辅助治疗等导致了这种现象,后续扩大样本量以及进一步亚组分析术后治疗对预后的影响可能有助于探讨TNM分期与风险 Riskscore 评分的关系。最后,作为一项初步的探索性研究,只对 lncRNA 与食管鳞癌患者预后风险鉴别有定性作用,尚未能精确定量 lncRNA 与食管鳞癌患者预后的关系,应用价值有限,仍需要多中心的大样本的研究进一步验证。

综上所述,通过差异表达 lncRNA 与食管鳞癌患者预后关系揭示食管鳞癌 2 个 lncRNA 潜在生物标志物,本研究成功构建了预后风险模型,是鉴别和预测食管鳞癌患者预后的重要尝试,是现有基于 TNM 分期预测食管鳞癌患者预后的有益补充手段。

参考文献

- [1] Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2021, 71(3): 209-249.
- [2] Yang Z, Zeng H, Xia R, et al. Annual cost of illness of stomach and esophageal cancer patients in urban and rural areas in China: A multi-center study[J]. Chin J Cancer Res, 2018, 30(4): 439-448.
- [3] 兰蓝, 赵飞, 蔡玥, 等. 中国居民2015年恶性肿瘤死亡率流行病学特征分析[J]. 中华流行病学杂志, 2018, 39(1): 32-34.
Lan L, Zhao F, Cai Y, et al. Epidemiological analysis on mortality of cancer in China, 2015[J]. Chin J Epidemiol, 2018, 39(1): 32-34.
- [4] Zeng H, Chen W, Zheng R, et al. Changing cancer survival in China during 2003-15: a pooled analysis of

- 17 population-based cancer registries [J]. *Lancet Glob Health*, 2018, 6(5): e555-e567.
- [5] Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015 [J]. *CA Cancer J Clin*, 2016, 66(2): 115-132.
- [6] Fitzmaurice C, Dicker D, Pain A, et al. The global burden of cancer 2013 [J]. *JAMA Oncol*, 2015, 1(4): 505-527.
- [7] Batista PJ, Chang HY. Long noncoding RNAs: cellular address codes in development and disease [J]. *Cell*, 2013, 152(6): 1298-1307.
- [8] Kornienko AE, Guenzl PM, Barlow DP, et al. Gene regulation by the act of long non-coding RNA transcription [J]. *BMC Biol*, 2013, 11: 59.
- [9] Shi X, Sun M, Liu H, et al. Long non-coding RNAs: a new frontier in the study of human diseases [J]. *Cancer Lett*, 2013, 339(2): 159-166.
- [10] Mercer TR, Mattick JS. Structure and function of long noncoding RNAs in epigenetic regulation [J]. *Nat Struct Mol Biol*, 2013, 20(3): 300-307.
- [11] 林姜鑫, 王伟伟, 王莉, 等. 长链非编码 RNA (linc02471) 对食管鳞状细胞癌细胞增殖和侵袭的影响 [J]. *中国卫生检验杂志*, 2018, 28(5): 516-519.
- Ling J, Wang W, Wang L, et al. Study on effect of long noncoding RNA linc02471 on cell proliferation and invasion of esophageal squamous cell carcinoma [J]. *Chin J Health Lab Technol*, 2018, 28(5): 516-519.
- [12] 王献, 王海兵. 食管鳞癌组织中 lncRNA UCA1 表达及其对癌细胞黏附和迁移影响 [J]. *青岛大学学报 (医学版)*, 2020, 56(3): 355-358.
- Wang X, Wang HB. Expression of the long noncoding RNA UCA1 in esophageal squamous cell carcinoma and its effect on the adhesion and migration of cancer cells [J]. *J Qingdao Univ (Med Sci)*, 2020, 56(3): 355-358.
- [13] Liang Y, Chen X, Wu Y, et al. LncRNA CASC9 promotes esophageal squamous cell carcinoma metastasis through upregulating LAMC2 expression by interacting with the CREB-binding protein [J]. *Cell Death Differ*, 2018, 25(11): 1980-1995.
- [14] 段玉青, 王郁, 王雪晓, 等. 食管鳞癌中 lncRNA DLEU1 的表达及对食管鳞癌细胞增殖和迁移的影响 [J]. *肿瘤防治研究*, 2021, 48(3): 255-260.
- Duan YQ, Wang Y, Wang XX, et al. Expression of lncRNA DLEU1 in esophageal squamous cell carcinoma and its effect on proliferation and migration of esophageal squamous cell carcinoma cells [J]. *Cancer Res Prev Treat*, 2021, 48(3): 255-260.
- [15] Li D, Li D, Meng L, et al. LncRNA NLIPMT inhibits tumorigenesis in esophageal squamous-cell carcinomas by regulating miR-320/Survivin axis [J]. *Cancer Manag Res*, 2020, 12: 12603-12612.
- [16] Xu JH, Chen RZ, Liu LY, et al. LncRNA ZEB2-AS1 promotes the proliferation, migration and invasion of esophageal squamous cell carcinoma cell through miR-574-3p/HMGA2 axis [J]. *Eur Rev Med Pharmacol Sci*, 2020, 24(10): 5391-5403.
- [17] Li H, Jia J, Yang L, et al. LncRNA MIR205HG drives esophageal squamous cell carcinoma progression by regulating miR-214/SOX4 axis [J]. *Onco Targets Ther*, 2020, 13: 13097-13109.
- [18] 彭慧, 秦凯, 戴宇翊, 等. 基于 TCGA 数据库的胶质母细胞瘤 lncRNA 风险预测模型的建立 [J]. *肿瘤防治研究*, 2019, 46(5): 417-420.
- Peng H, Qin K, Dai YH, et al. Establishment of lncRNA risk prediction model for glioblastoma based on TCGA database [J]. *Cancer Res Prev Treat*, 2019, 46(5): 417-420.
- [19] 史爽, 李娟, 王雪, 等. 胃癌 lncRNA 预后风险评估模型的构建 [J]. *中国现代普通外科进展*, 2020, 23(4): 253-257; 262.
- Shi S, Li J, Wang X, et al. Establishment of a lncRNA signature to predict prognosis for gastric cancer [J]. *Chin J Curr Adv Gen Surg*, 2020, 23(4): 253-257; 262.
- [20] Sun Z, Jing C, Xiao C, et al. Long non-coding RNA profile study identifies an immune-related lncRNA prognostic signature for kidney renal clear cell carcinoma [J]. *Front Oncol*, 2020, 10: 1430.
- [21] Shi X, Li Y, Sun Y, et al. Genome-wide analysis of lncRNAs, miRNAs, and mRNAs forming a prognostic scoring system in esophageal squamous cell carcinoma [J]. *Peer J*, 2020, 8: e8368.