

基于 TCGA 数据库构建肺腺癌自噬相关基因预后风险模型

王雪芹^{1,2}, 刘亚锋^{1,2}, 吴静^{2,3}, 周家伟^{1,2}, 邢应如^{1,4}, 张鑫^{1,2}, 李丹婷^{1,2}, 谢军⁴, 丁选胜², 胡东^{2,3}

摘要 目的 基于 TCGA 数据库结合人类自噬数据库构建预测肺腺癌患者预后的风险模型,探讨自噬相关基因预后风险模型对肺腺癌患者预后的预测性能及与免疫微环境的相关性。方法 从癌症基因组图谱数据库下载肺腺癌患者临床信息和转录组数据,结合人类自噬数据库筛选出 232 个自噬相关基因。通过 Cox 回归分析筛选出 4 个独立与预后相关的自噬基因,并采用风险评分构建肺腺癌预后预测模型,用 ROC 曲线评价预测模型性能。使用 ESTIMATE 和 CIBERSORT 在线网站(<https://cibersort.stanford.edu/>)探索风险评分与肿瘤免疫微环境之间的关系。结果 肺腺癌中

有 30 个差异表达的自噬相关基因,其中 4 个自噬基因(*BIRC5*、*ERO1A*、*ITGB4*、*NLR4*)具有预测患者预后的功能。依据风险评分进行分组,Kaplan-Meier 分析表明高风险组生存率低于低风险组($P < 0.0001$)。ROC 曲线表明风险评分模型判断肺腺癌预后的准确性(AUC = 0.757)。ESTIMATE 和 CIBERSORT 分析提示风险评分模型与肿瘤微环境中的多个免疫细胞亚群浸润相关。结论 自噬相关基因预后风险模型较临床数据能更好地预测肺腺癌患者预后。在高风险组中,CD4⁺记忆静止细胞可改善肺腺癌患者预后。

关键词 肺腺癌;自噬;免疫细胞;免疫浸润;生存预后
中图分类号 R 734.2

文献标志码 A 文章编号 1000-1492(2022)04-0528-06
doi: 10.19405/j.cnki.issn1000-1492.2022.04.005

2021-07-19 接收

基金项目:国家自然科学基金(编号:81971483);安徽省高校拔尖人才项目(编号:gxhjZD12);安徽理工大学研究生创新基金(编号:2020CX2083)

作者单位:¹安徽理工大学医学院免疫教研室,淮南 232000

²安徽省职业健康安全工程实验室,淮南 232000

³安徽理工大学工业粉尘防控与职业安全健康教育重点实验室,淮南 232000

⁴安徽理工大学附属肿瘤医院,淮南 232000

作者简介:王雪芹,女,硕士研究生;

邢应如,男,副教授,硕士生导师,责任作者,E-mail: austxingyr@126.com;

胡东,男,教授,博士生导师,责任作者,E-mail: austhudong@126.com

自噬是维持哺乳动物体内稳态及细胞应激适应性反应的关键组成部分。在饥饿环境中,自噬捕获并降解溶酶体中的细胞器和蛋白质从而保护生物体免受疾病侵害,主要包括心脏病、感染和神经变性^[1]。但有研究^[2]表明,黑素瘤和乳腺癌在内的癌症会诱导自噬并依赖自噬生存。自噬介导的循环维持能量稳态和线粒体功能可以促进肿瘤细胞的生长和增殖。因此,抑制自噬功能可能对癌症治疗产生有益效果。

mg/kg), and young rats were used as Control group (Control). The levels of heart rate (HR), left ventricular ejection fraction (LVEF) and fractional shortening (FS) were detected by echocardiography. The levels of creatine kinase MB (CK-MB), myohemoglobin (Mb), cardiac troponin I (cTnI), superoxide dismutase (SOD), malondialdehyde (MDA) and lactate dehydrogenase (LDH) were detected by ELISA. The pathological morphology of heart tissue was observed by HE staining. The apoptosis of myocardial cells was observed by TUNEL staining. Western blot detected the protein expression of cysteine aspartic protease 9 (Caspase 9), cleaved cysteine aspartic protease 9 (cleaved Caspase 9), cysteine aspartic protease 3 (Caspase 3), cleaved cysteine aspartic protease 3 (cleaved Caspase 3), nuclear factor E2 related factor 2 (Nrf2), phosphorylation nuclear factor E2 related factor 2 (p-Nrf2), quinone oxidoreductase 1 (NQO1), heme oxygenase 1 (HO-1). **Results** Compared with the elderly model group, HR, LVEF, FS increased, CK-MB, Mb, cTnI levels decreased, MDA, LDH content decreased, SOD activity increased, cleaved Caspase 9/Caspase 9, cleaved Caspase 3/Caspase 3 expression decreased, The expressions of Nrf2, NQO1 and HO-1 were increased, and the pathological morphology of cardiac tissue and myocardial apoptosis were improved in Cor groups. **Conclusion** Cor can improve cardiac function of elderly rats, relieve oxidative stress response, inhibit myocardial cell apoptosis and have a certain protective effect on cardiac tissue injury in rats, and the mechanism may be related to the promotion of Nrf2/HO-1 pathway related protein expression.

Key words cordycepin; heart function; Nrf2/HO-1 pathway; NQO1

肺癌占世界新癌症发病率和病死率的第一位^[3]。近年来肿瘤切除术、免疫治疗和放疗化疗综合应用于治疗肺腺癌,但是肺腺癌患者平均5年生存率仍低于20%^[4]。因此,鉴定疾病特异性生物标志物,提高对肺腺癌分子机制的了解,开发新的治疗决策,以改善患者的预后是非常有必要的。

1 材料与方法

1.1 数据采集 于2020年8月从癌症基因组图谱(the cancer genome atlas,TCGA)数据库(<https://tcga-data.nci.nih.gov/tcga/>)下载肺腺癌RNA-seq数据以及临床数据。从人类自噬数据库(human autophagy database,HADDb)(<http://www.autophagy.lu/>)获得了232个自噬相关基因(autophagy related genes,ARGs)。TCGA数据库中肺腺癌患者的基本临床特征:女性213例(45.7%),男性254例(54.3%);65岁以上有227例(48%),65岁及以下有240例(52%);临床I、II、III期分别为251例(54%)、108例(23%)、75例(16%);T1、T2、T3、T4型分别为162例(34%)、244例(52%)、39例(8%)、19例(6%);M0、M1型分别为314例(92%)、24例(8%),N分型中N0、N1、N2~3型分别为300例(66%)、87例(19%)、68例(15%)。

1.2 鉴定差异表达自噬基因 自TCGA数据库下载肺腺癌RNAs表达谱,并从中提取人自噬相关基因的表达数据,使用R软件“limma”软件包对正常样本和癌症样本鉴定差异表达自噬基因(DEARGs),通过“pheatmap”软件包绘制热图。

1.3 自噬相关基因预后模型的构建 用R语言“survival”程序包进行单因素Cox分析,并将初步筛选的基因与肺腺癌预后相关的自噬基因($P < 0.05$)纳入多因素Cox分析,以进一步验证基因对肺腺癌患者预后的价值。经过上述分析,获得风险评分模型。风险评分计算如下:

$$\text{风险评分} = \sum_{i=1}^n \text{Coef}_i \times x_i$$

其中Coef是系数, x 是每个选定自噬基因的表达水平。

1.4 高、低风险组之间免疫细胞浸润程度的比较 为探讨风险评分与肺腺癌预后之间的关系,对风险评分与肿瘤微环境之间的关系进行分析。使用R软件ESTIMATE算法确定每个样品的免疫评分,通过Wilcoxon检验比较高危组和低危组之间免疫细胞浸润程度差异。此外使用CIBERSORT软件包评估22种免疫细胞亚型比例。

2 结果

2.1 筛选差异表达自噬基因 使用R软件整理TCGA数据库中497个肿瘤样本及54个肿瘤患者的癌旁正常样本。转录组数据并从中提取232个自噬基因(HADDb数据库获得)的表达数据,进行差异表达分析,筛选出30个差异自噬基因($FDR < 0.05$ 且 $|\log_2FC| > 1$)。其中有18个上调的ARGs[含有杆状病毒IAP重复序列5(baculoviral IAP repeat containing 5,*BIRC5*)、*CDKN2A*、*TMEM74*、内质网氧化还原酶1- α (endoplasmic reticulum oxidoreductase 1 alpha,*ERO1A*)、*ATG9B*、*GAPDH*、*PTK6*、*EIF4EBP1*、整合素4(integrin subunit beta 4,*ITGB4*)、*VMP1*、*IFNG*、*ATIC*、*BNIP3*、*IKBKE*、*P4HB*、*PARP1*、*ERBB2*和*HSPA5*]和12个下调的ARGs[*MAP1LC3C*、*GRID1*、*CCL2*、*PRKCQ*、*NRG1*、*FOS*、*PPP1R15A*、*DAPK2*、*HSPB8*、NLR家族含有CARD结构域4(NLR family CARD domain containing 4,*NLRC4*)、*DLC1*和*NRG3*]。见图1。

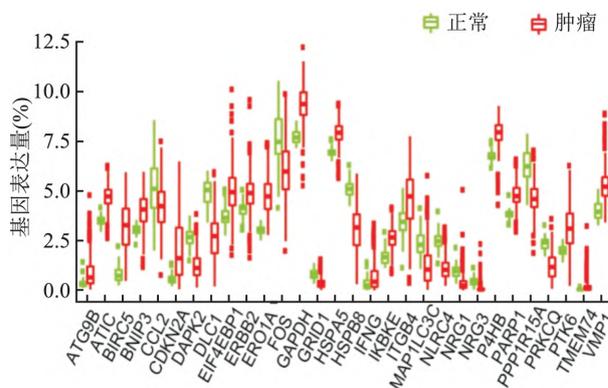


图1 肺腺癌和正常组织间差异表达的自噬基因

2.2 肺腺癌患者预后相关自噬基因鉴定及风险评分模型构建 为初步筛选与肺腺癌预后相关的自噬基因,对差异表达的30个自噬mRNAs进行单因素Cox回归分析,共发现有9个自噬基因与患者预后相关($P < 0.05$) (图2A)。经过多变量Cox回归分析确定4个自噬基因用于构建预后风险模型(图2B)。风险评分计算方法如下: 风险评分 = $(0.123746 \times \text{BIRC5 表达量}) + (0.324086 \times \text{ERO1A 表达量}) + (0.121162 \times \text{ITGB4 表达量}) + (-0.547430 \times \text{NLRC4 表达量})$ 。

通过风险预后模型来计算497例患者的风险值并以中位值将患者分为高、低风险两组。 HR (风险比) > 1 的基因(*BIRC5*、*ERO1A*、*ITGB4*)被认为是危

险基因,而 $HR < 1$ 的基因(*NLRC4*)被认为是保护性基因。随风险值升高,病例病死率呈升高趋势(图 2C)。高风险组患者表达危险基因增多,表达保护基因减少(图 2D)。

2.3 风险评分模型中基因的蛋白表达水平 为了进一步确定这 4 个自噬基因在肺腺癌中的表达,利用人类蛋白质图谱数据库(HPA)的免疫组织化学数据加以验证,结果显示,与正常肺组织相比,肺腺癌中的 *BIRC5*、*ERO1A*、*ITGB4* 表达增加,而 *NLRC4* 在肿瘤细胞中则未发现(图 3)。肺腺癌中 *BIRC5*、*ERO1A*、*ITGB4*、*NLRC4* 蛋白水平的变化与其基因水平的变化一致。

2.4 风险评分模型的性能检测 Log-rank 检验对低、高风险组进行 Kaplan-Meier 分析,结果表明高风险组的总生存期较差 ($P < 0.0001$),见图 4A。ROC 曲线下面积(AUC)用来评估风险评分模型预测效能,该风险评分模型预测患者 1 年生存率的 AUC 值为 0.757(95% CI: 0.727 ~ 0.787),3 年生存率的 AUC 值为 0.723(95% CI: 0.667 ~ 0.779),说明该模型具有良好的短期预测能力(图 4B)。然而 5 年生存率的 AUC 值仅为 0.581(95% CI: 0.516 ~ 0.646),提示该模型不能很好地预测患者长期(5 年以上)生存状态。采用 ROC 曲线比较风险评分模型

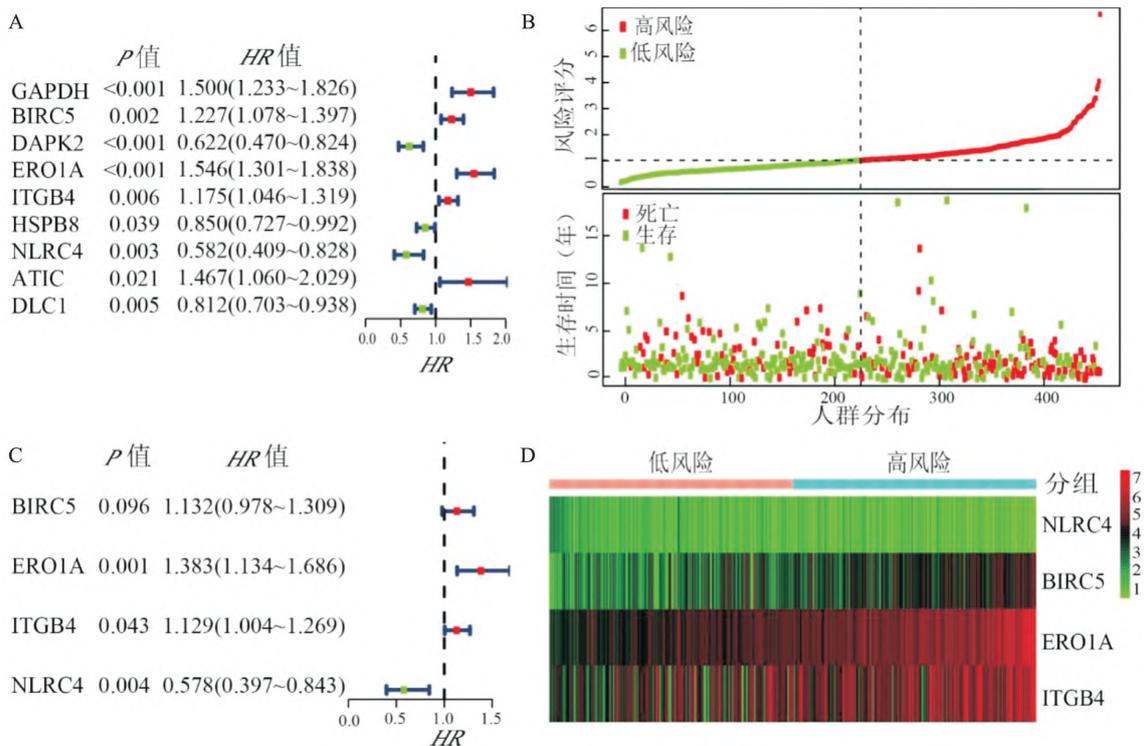


图 2 预后相关自噬基因筛选及表达特征

A: 自噬相关基因的单变量 Cox 回归分析; B: 自噬相关基因的多变量 Cox 回归分析; C: 风险评分和患者生存时间以及生存/死亡状态的分布; 黑色虚线: 将患者分为低风险和高风险组的最佳临界值; D: 低风险和高风险组自噬相关基因表达谱

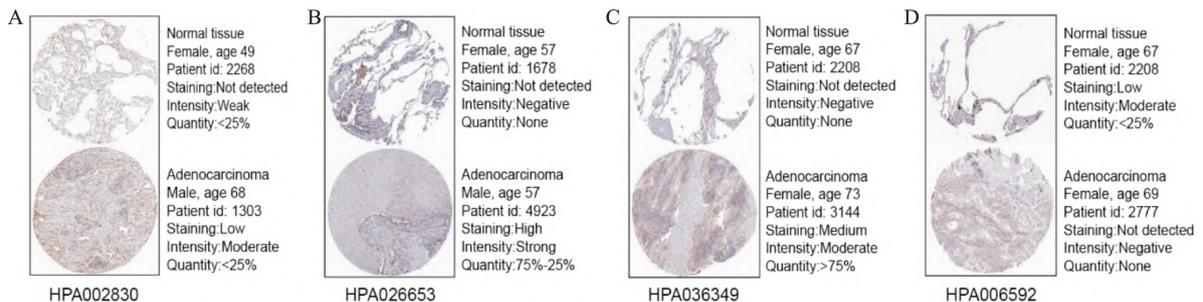


图 3 风险评分模型中 4 个基因的蛋白表达水平

A: BIRC5; B: ERO1A; C: ITGB4; D: NLRC4

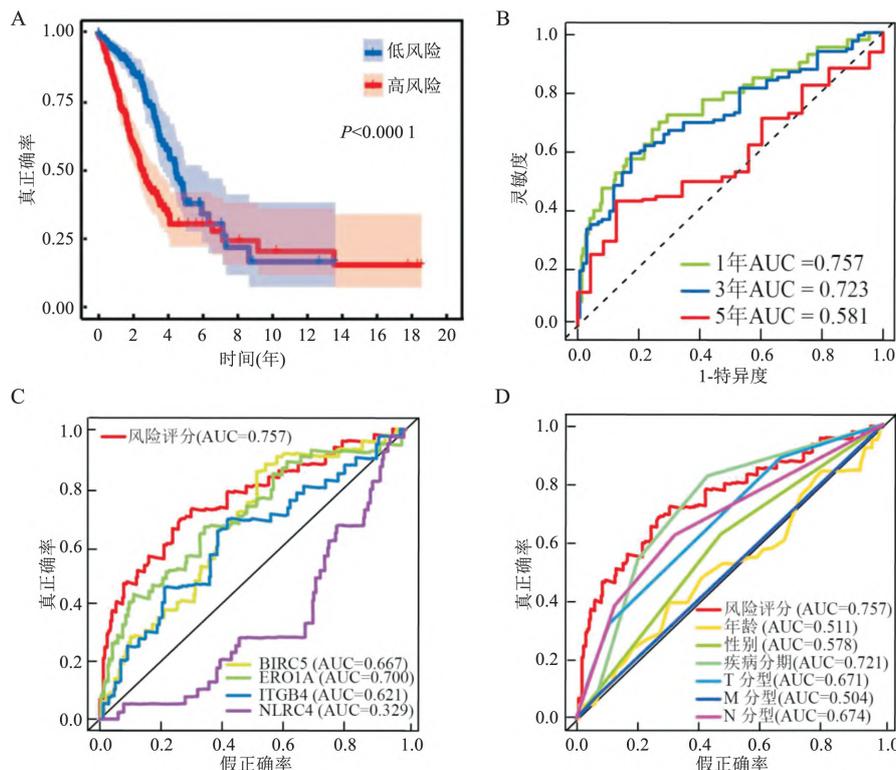


图4 风险评分模型高、低风险组 Kaplan-Meier 曲线、ROC 曲线及临床参数的相关性

A: 基于4个自噬基因的 KM 曲线; B: 基于4个自噬基因 ROC 曲线; C: 基于4个自噬基因与单独自噬基因的 ROC 曲线; D: 基于4个自噬基因与临床参数的 ROC 曲线

与四个单基因及临床参数的预测能力, 风险评分模型的 AUC 为 0.757 (95% CI: 0.727 ~ 0.787); *BIRC5* 的 AUC 为 0.667 (95% CI: 0.609 ~ 0.725); *ERO1A* 的 AUC 为 0.700 (95% CI: 0.649 ~ 0.751); *ITGB4* 的 AUC 为 0.621 (95% CI: 0.555 ~ 0.687); *NLRC4* 的 AUC 为 0.329 (95% CI: 0.258 ~ 0.400); 年龄的 AUC 为 0.511 (95% CI: 0.442 ~ 0.580); 性别的 AUC 为 0.578 (95% CI: 0.512 ~ 0.644); 疾病分期的 AUC 为 0.721 (95% CI: 0.670 ~ 0.772); T 分型的 AUC 为 0.671 (95% CI: 0.617 ~ 0.725); M 分型的 AUC 为 0.504 (95% CI: 0.434 ~ 0.574); N 分型的 AUC 为 0.674 (95% CI: 0.622 ~ 0.726) (图 4C、4D)。DeLong 检验表明, 风险评分模型的 AUC 高于年龄 ($P = 0.033$)、性别 ($P = 0.039$)、T 分型 ($P = 0.047$)、M 分型 ($P = 0.031$) 及 N 分型 ($P = 0.042$) 的 AUC, 但是风险评分模型与疾病分期的 AUC 差异无统计学意义 ($P = 0.083$) 其原因可能是样本量不足。由此可见, 无论与单基因或者其他临床参数相比, 该模型具有较好的预测能力和准确性。

2.5 高、低风险组之间的免疫浸润评分 免疫评分

显示高、低风险评分组之间存在差异(图 5A)。Kaplan-Meier 分析显示, 肺癌中不同的免疫评分具有不同的生存期(图 5B)。肺癌的 22 种免疫细胞比例, 其中 M0 巨噬细胞、M2 巨噬细胞和 T 细胞 CD4、CD8 为肺癌浸润主要免疫细胞亚群(图 5C)。在高、低风险组中, 有 10 群免疫细胞水平存在差异, 其中 CD4 记忆静止细胞、CD4 记忆激活细胞、单核细胞、M0 巨噬细胞、静止树突状细胞、肥大静止细胞及肥大激活细胞差异最为明显(图 5D)。

3 讨论

自噬是肿瘤起始和癌症进展的抑制因子, 而在肿瘤晚期自噬作为促进生存、生长和侵略性的保护因子。越来越多的证据表明自噬与癌症密切相关。如 *NLRC5* 通过抑制自噬能够促进子宫内膜癌细胞的增殖^[5]; 孕酮和脂联素受体 3 (*PAQR3*) 通过调节表皮生长因子受体 (*EGFR*) 介导的自噬抑制非小细胞肺癌细胞生长^[6]; 下调自噬通过抑制 *GLI* 系列锌指 2 (*GLI2*) 的产生, 从而促进 *Shh* 途径抑制的抗肺癌功效^[7]。

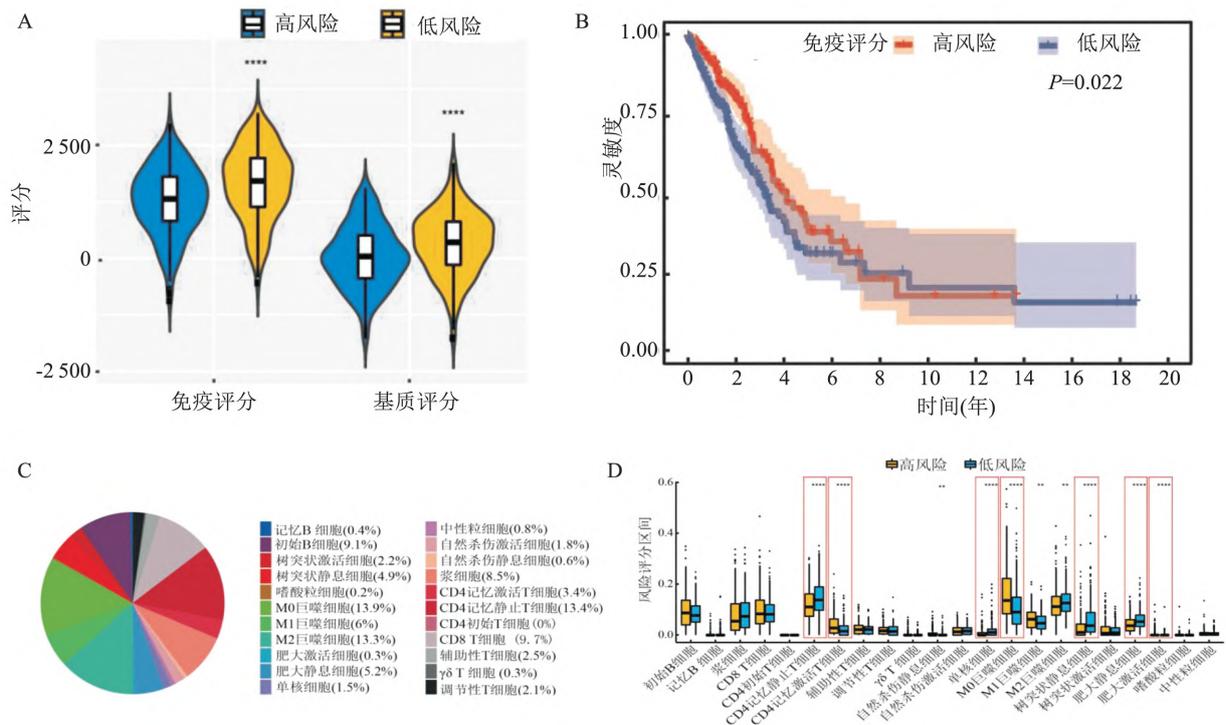


图5 风险评分模型与肿瘤免疫微环境的相关性

A: 免疫评分、基质评分和风险评分的关联; B: 免疫评分对肺腺癌总体生存影响的 KM 分析; C: 肺腺癌中 22 种免疫细胞的比例; D: 高风险组和低风险组之间免疫细胞差异性表达; 与高风险组比较: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$

本文基于 TCGA 数据库中的 RNA-seq 转录组数据及临床资料 结合 HADBB 进行生物信息学分析, 表明 *BIRC5*、*ERO1A* 和 *ITGB4* 是肺腺癌危险因素, *NLRC4* 是肺腺癌保护因素。 *BIRC5* 能够通过启动抗凋亡机制引起肺癌细胞增殖、肿瘤形成并可增强肿瘤细胞对化疗和放射治疗的耐药性^[8-9]; *ERO1A* 能够通过上调血管内皮生长因子, 激活肿瘤血管生成信号从而促进肿瘤的生长^[10]; *ITGB4* 能够调控肿瘤、呼吸系统疾病和神经系统病变等多种疾病的进展^[11]。 *NLRC4* 通过抑制细胞增殖和促进细胞死亡来预防结肠癌^[12]。本研究利用自噬相关基因结合风险评分构建肺腺癌预后预测模型, 总结出高风险组的总生存期较差, 且无论与单基因或其他临床参数相比, 该模型能较好地预测肺腺癌患者预后。

目前, 免疫疗法是肺腺癌的一种治疗方案, 已有研究证明免疫细胞在肿瘤细胞上的作用。此外, 肺腺癌组织中的免疫成分可用于评估患者的治疗效果和肺腺癌预后^[13]。该研究结果表明 4 个自噬基因组合的风险评分与免疫浸润评分具有相关性。使用 CIBERSORT 算法计算免疫细胞亚型, 结果显示, 高、低风险评分组表达不同的免疫细胞亚型, 表明免疫

细胞成分比例失衡与癌症患者的不良预后和低存活率相关。肺腺癌中的免疫细胞主要包括 M0 巨噬细胞、M2 巨噬细胞、CD4⁺ 记忆静止 T 细胞和 CD8⁺ T 细胞。其中, 高风险组 CD4⁺ 记忆静止 T 细胞和 M2 巨噬细胞水平较低, 而 M0 巨噬细胞水平较高。这暗示着高危组中 CD4⁺ 记忆静止 T 细胞与 M0 巨噬细胞比例的失衡可能降低患者的生存率。

参考文献

- [1] Picca A, Lezza A M S, Leeuwenburgh C, et al. Fueling inflammation through mitochondrial dysfunction: mechanisms and molecular targets [J]. Int J Mol Sci 2017, 18(5): 933.
- [2] Amaravadi R, Kimmelman A C, White E. Recent insights into the function of autophagy in cancer [J]. Genes Dev 2016, 30(17): 1913-30.
- [3] Siegel R L, Miller K D, Jemal A. Cancer statistics, 2020 [J]. CA Cancer J Clin 2020, 70(1): 7-30.
- [4] Behera M, Owonikoko T K, Gal A A, et al. Lung adenocarcinoma staging using the 2011 IASLC/ATS/ERS classification: a pooled analysis of adenocarcinoma in situ and minimally invasive adenocarcinoma [J]. Clin Lung Cancer 2016, 17(5): e57-64.
- [5] 姚顺, 詹磊, 孙士莹等. NLRC5 抑制自噬促进子宫内膜癌细胞增殖的研究 [J]. 安徽医科大学学报, 2019, 54(7): 1022-7.

- [6] Cao Q, You X, Xu L, et al. PAQR3 suppresses the growth of non-small cell lung cancer cells *via* modulation of EGFR-mediated autophagy [J]. *Autophagy*, 2020, 16(7): 1236–47.
- [7] Fan J, Zhang X, Wang S, et al. Regulating autophagy facilitated therapeutic efficacy of the sonic Hedgehog pathway inhibition on lung adenocarcinoma through GLI2 suppression and ROS production [J]. *Cell Death Dis*, 2019, 10(9): 626.
- [8] 李雪菲, 张丁丁, 董丹丹, 等. BIRC5 高表达在胃癌中增强细胞活力、抑制凋亡并与预后不良相关 [J]. *中国病理生理杂志*, 2020, 36(11): 2013–9.
- [9] 赵欢, 马小莉, 张萱, 等. BIRC5 抗体在非小细胞肺癌血浆中的表达及其与预后的关系 [J]. *中国免疫学杂志*, 2021, 37(1): 93–7.
- [10] Varone E, Decio A, Chernorudskiy A, et al. The ER stress response mediator ERO1 triggers cancer metastasis by favoring the angiogenic switch in hypoxic conditions [J]. *Oncogene*, 2021, 40(9): 1721–36.
- [11] Soung Y H, Clifford J L, Chung J. Crosstalk between integrin and receptor tyrosine kinase signaling in breast carcinoma progression [J]. *Bmb Reports* 2010, 43(5): 311–8.
- [12] Yu X, Liu W, Chen S, et al. Immunologically programming the tumor microenvironment induces the pattern recognition receptor NLR4-dependent antitumor immunity [J]. *J Immunother Cancer*, 2021, 9(1): e001595.
- [13] 成天琼, 陈学农. 基于 TCGA 挖掘肺腺癌微环境预后相关基因 [J]. *遵义医科大学学报*, 2020, 43(2): 221–8.

Construction of prognostic risk model of autophagy related genes in lung adenocarcinoma based on TCGA database

Wang Xueqin^{1,2}, Liu Yafeng^{1,2}, Wu Jing^{2,3}, Zhou Jiawei^{1,2}, Xing Yingru^{1,4},
Zhang Xin^{1,2}, Li Danting^{1,2}, Xie Jun⁴, Ding Xuansheng², Hu Dong^{2,3}

¹Dept of Immunology, School of Medicine, Anhui University of Science and Technology, Huainan 232000;

²Anhui Occupational Health and Safety Engineering Laboratory, Huainan 232000;

³Key Laboratory of Industrial Dust Control and Occupational Safety and Health, Ministry of Education, Anhui University of Science and Technology, Huainan 232000;

⁴Affiliated Cancer Hospital of Anhui University of Science and Technology, Huainan 232000)

Abstract Objective A prognostic risk model for lung adenocarcinoma patients was established based on the cancer genome atlas (TCGA) database to explore the prognostic performance of autophagy related gene risk model for lung adenocarcinoma patients and its correlation with immune microenvironment. **Methods** Clinical information and transcriptome data of lung adenocarcinoma patients were downloaded and extracted from TCGA database, and 232 autophagy-related genes were screened from the human autophagy database. cox regression analysis was used to screen out four autophagy genes independently associated with prognosis. The prognostic prediction model of lung adenocarcinoma was constructed by risk score, and the performance of prediction model was evaluated by ROC curve. The relationship between risk scores and tumor immune microenvironment was explored using ESTIMATE (estimation of stromal and immune cells in malignant tumour tissues using expression data) and CIBERSORT algorithms. **Results** Thirty differentially expressed autophagy-related genes were identified in lung adenocarcinoma, of which four autophagy genes (*BIRC5*, *ERO1A*, *ITGB4*, *NLR4*) could predict the prognosis of the patients. Grouped by risk score, the Kaplan–Meier analysis demonstrated that the survival rate of high-risk group was significantly lower than that of low-risk group ($P < 0.0001$). The ROC curve proved the accuracy of the model in predicting the prognosis of lung adenocarcinoma ($AUC = 0.757$). The ESTIMATE and CIBERSORT analyses revealed that the risk scoring model was associated with multiple immune cells and immune infiltrates in the tumor microenvironment. **Conclusion** Compared with clinical data, the autophagy gene prognostic risk model can better predict the prognosis of patients with lung adenocarcinoma. In the high-risk group, CD4⁺ memory quiescent cells can improve prognosis in lung adenocarcinoma patients.

Key words lung adenocarcinoma; autophagy; immune cells; immune infiltration; survival and prognosis