

· 论 著 ·

基于 TCGA 数据库肺腺癌自噬相关基因预后模型的建立与验证*

吴双丽¹, 吴铁成², 喻光², 徐敬宣², 李保健², 邢龙^{2△}

(青岛滨海学院附属医院: 1. 神经电生理科; 2. 肿瘤内科, 山东 青岛 266500)

[摘要] 目的 探究肺腺癌自噬相关基因(ARGs), 并构建肺腺癌 ARGs 的预后模型。方法 肺腺癌的 RNA 高通量转录组数据下载于癌症基因组图谱(TCGA)数据库和人类自噬基因(HADb)数据库, 获取 ARGs, 并基于肺腺癌差异表达 ARGs 构建肺腺癌预后模型并验证, 进一步构建列线图及校准曲线探究模型在临床中的应用价值。对差异表达 ARGs 进行基因本体和京都基因与基因组百科全书富集分析。对获得的具有预后意义的差异表达 ARGs 进行 Lasso 回归分析并构建肺腺癌预后模型。绘制 Kaplan-Meier 生存曲线。结果 共获得 31 个差异表达 ARGs, 筛选得到 10 个具有预后意义的差异表达 ARGs。高风险组患者与较差的总生存期明显相关, 差异有统计学意义($P < 0.05$)。T 分期、N 分期、风险评分为肺腺癌患者的独立预后因素。校准曲线和列线图全局一致性为 0.710, 说明模型预测结果与实际情况具有较高的符合度。结论 基于差异表达 ARGs 构建的风险模型可作为肺腺癌患者的一种预后特征或可为肺腺癌患者的个体化治疗提供参考依据。

[关键词] 肺腺癌; 自噬相关基因; 预后模型; 癌症基因组图谱数据库; 计算生物学

DOI: 10.3969/j.issn.1009-5519.2024.10.001

中图法分类号: R734.2

文章编号: 1009-5519(2024)10-1621-06

文献标识码: A

Establishment and validation of a prognostic model for autophagy related genes in lung adenocarcinoma based on TCGA database*

WU Shuangli¹, WU Tiecheng², YU Guang², XU Jingxuan², LI Baojian², XING Long^{2△}

(1. Department of Neuroelectrophysiology; 2. Department of Oncology, Affiliated Hospital of Qingdao Binhai University, Qingdao, Shandong 266500, China)

[Abstract] **Objective** To explore autophagy-related genes (ARGs) in lung adenocarcinoma and construct a prognostic model for lung adenocarcinoma based on ARGs. **Methods** RNA high-throughput transcriptome data of lung adenocarcinoma were obtained from The Cancer Genome Atlas (TCGA) database and HADb database to acquire ARGs. A prognostic model for lung adenocarcinoma was constructed and validated based on differentially expressed ARGs, followed by the construction of column line graphs and calibration curves to explore the clinical application value of the model. Gene ontology and Kyoto Encyclopedia of Genes and Genomes enrichment analyses were performed on differentially expressed ARGs. Lasso regression analysis was conducted on differentially expressed ARGs with prognostic significance to construct the prognostic model for lung adenocarcinoma. Kaplan-Meier survival curves were plotted. **Results** A total of 31 differentially expressed ARGs were obtained, and 10 differentially expressed ARGs with prognostic significance were selected. Patients in the high-risk group were significantly associated with poorer overall survival, with statistical significance ($P < 0.05$). T stage, N stage and risk score was an independent prognostic factor for patients with lung adenocarcinoma. The global consistency of the calibration curve column line graph was 0.710, indicating a high level of agreement between the model's predicted results and actual outcomes. **Conclusion** The risk model constructed based on differentially expressed ARGs can serve as a prognostic feature for patients with lung adenocarcinoma or provide a reference for individualized treatment for patients with lung adenocarcinoma.

[Key words] Lung adenocarcinoma; Autophagy-related genes; Prognosis model; The Cancer Genome Atlas database; Computational biology

我国国家癌症中心的最新数据显示, 我国新发肺癌患者约为 78.7 万例/年, 其中男 52.0 万例, 女 26.7 万例, 因肺癌死亡的患者约为 63.0 万例/年, 其中男 43.3 万例, 女 19.7 万例^[1]。随着肺腺癌发病机制研

* 基金项目: 山东省医药卫生科技发展计划项目(202203100054); 青岛西海岸新区 2020 年度科技惠民项目(2020-44)。

作者简介: 吴双丽(1988—), 本科, 主治医师, 主要从事老年性疾病的诊断与治疗工作。△ 通信作者, E-mail: xinglong0307@163.com。

究的逐渐深入,新治疗方法取得了较大的突破,但中晚期肺腺癌患者 5 年生存率依然很低^[2]。因此,针对其发生、发展过程中分子机制的研究任重道远。

自噬是一种只发生在真核细胞中的溶酶体降解途径,对生存、分化、发育、稳态至关重要。在某些情况下其被认为是一种非凋亡细胞死亡途径。有充分的证据表明,自噬涉及多种疾病,包括癌症、神经变性、感染等^[3]。对自噬的研究对进一步理解恶性肿瘤的病理生理过程至关重要,基于此,本研究利用从癌症基因组图谱(TCGA)数据库和人类自噬基因(HADb)数据库获得的信息,建立了基于自噬相关基因(ARGs)的肺腺癌患者的预后模型,旨在为肺腺癌的诊治提供参考依据。

1 资料与方法

1.1 研究对象 将基于 TCGA 数据库和 HADb 数据库获取的肺腺癌 ARGs 作为研究对象。

1.2 方法

1.2.1 转录组数据与临床数据的获取 从 TCGA 数据库(<https://portal.gdc.cancer.gov/>)下载肺腺癌患者肿瘤组织和癌旁组织的基因转录组高通量测序数据及临床资料。

1.2.2 肺腺癌中差异表达 ARGs 筛选 使用 R 软件的 DESeq2 包筛选肺腺癌和正常肺组织样本的 ARGs,筛选条件为错误发现率小于 0.05, $|\log_{2}(\text{Foldchange})| \geq 1$ 。使用 Venndiagram 包进一步绘制差异表达基因和 ARGs 的韦恩图,得到在肺腺癌中差异表达的 ARGs,然后整合肺腺癌和正常肺组织中差异表达的 ARGs 数据及其相应的临床信息,通过 ComplexHeatmap 包绘制差异表达的 ARGs 的热图并进行可视化,进一步通过 stats 包和 car 包分析差异表达的 ARGs 在肺腺癌及正常肺组织中的表达,并通过 ggplot2 包进行可视化。

1.2.3 差异表达的 ARGs 功能富集分析 使用 R 软件 ggplot2 包对差异表达的 ARGs 进行富集分析,并对结果进行可视化,对满足 $P_{\text{adj}} < 0.05$ 的结果以气泡图的形式进行展示。

1.2.4 预后模型中 ARGs 的筛选 使用 survival 包进行单变量 Cox 回归分析,同时,对预后相关的 ARGs 使用 ggplot2 包绘制森林图。进一步通过 glmnet 包和 survival 包进行 Lasso 系数筛选及绘制 Lasso 变量轨迹图,得到构建预后模型的基因及对应的风险系数(Coef 值),然后根据风险评分公式(风险评分=基因 1 表达量 \times 相应的回归系数+...+基因 n 表达量 \times 相应的回归系数)计算每例患者的风险评分,并以评分的中位值将患者分为高、低风险组。

1.2.5 自噬基因预后模型的性能评估 分别使用 R 语言中 survival 包和 survminer 包对患者的生存资料进行统计分析可视化,对高、低风险组进行生存分析

并绘制 Kaplan-Meier 生存曲线。通过 ggplot2 包绘制风险因子图及可视化。使用 survival 包对年龄、性别、T 分期、N 分期、M 分期、风险评分进行单因素及多因素独立预后分析,将单因素中样本满足 $P < 0.1$ 的因素纳入多因素 Cox 中构建模型,以验证风险评分是否能独立于其他临床因素成为独立的预测因子。

1.2.6 列线图 and 校准曲线的构建 通过 R 语言中 rms 包和 survival 包绘制预后列线图及校准曲线,以风险评分、年龄、性别、T 分期、N 分期、M 分期、临床分期构建列线图,以预测肺腺癌患者 1、3、5 年总生存期(OS),同时,采用校准曲线评估预测模型在临床中的实际应用价值。

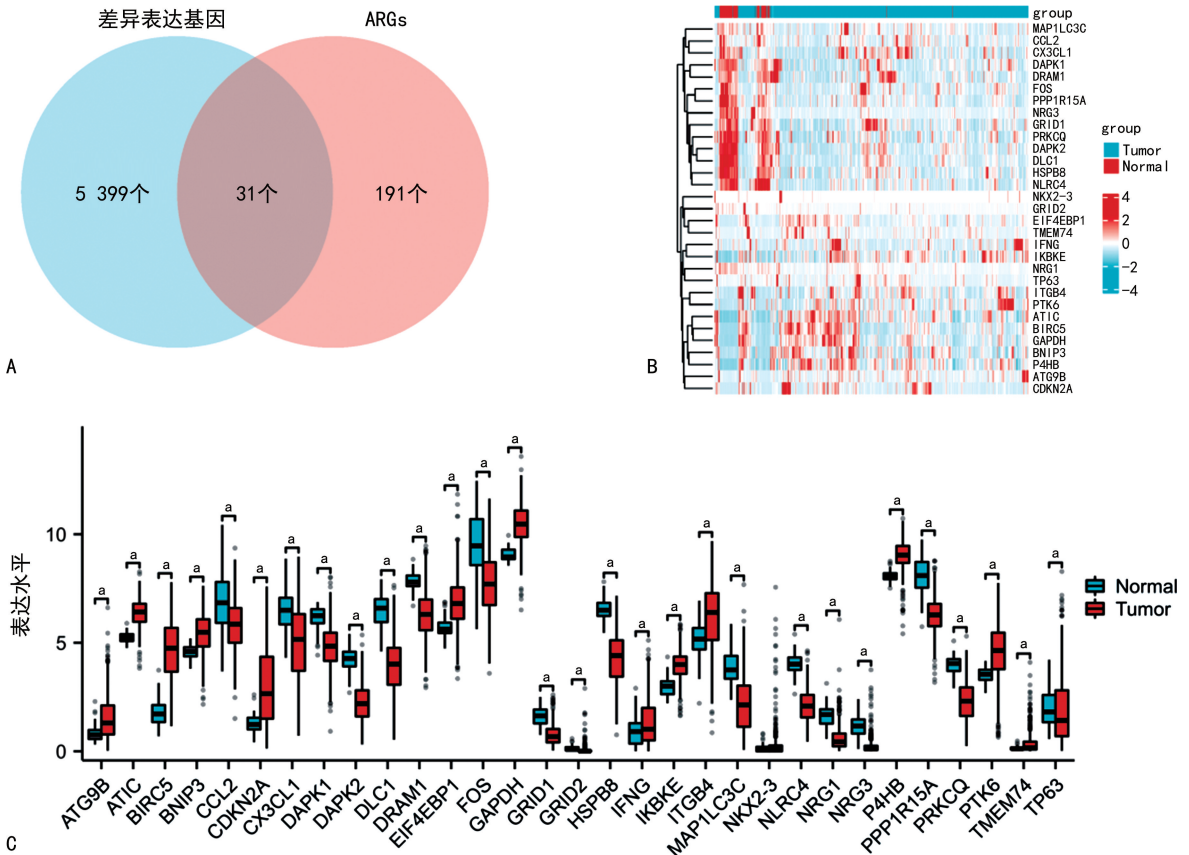
1.3 统计学处理 应用 R4.2.1 软件进行数据分析,绘制 Kaplan-Meier 生存曲线描述高低风险人群生存率,使用 Lasso 回归降维处理,以 OS 为因变量,年龄、性别、T 分期、N 分期、M 分期、临床分期、风险评分作为协变量进行单变量 Cox 分析和多变量 Cox 分析,确定独立预后因子。以 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 肺腺癌测序数据和临床资料 从 TCGA 数据库中下载、整理及分析 594 个样本数据,包括 535 例肺腺癌肿瘤组织标本和 59 例正常肺组织数据,共筛选出 5 430 个差异表达基因。ARGs 下载自 AgingAtlas (<https://ngdc.cncb.ac.cn/aging/index>),共获得 222 个 ARGs。韦恩图取交集后获得 31 个差异表达的 ARGs。见图 1A。其中 ATG9B、ATIC、BIRC5、BNIP3、CDKN2A、EIF4EBP1、GAPDH、IFNG、IKBKE、ITGB4、NKX2-3、P4HB、PTK6、TMEM74 在肿瘤组织中高表达,CCL2、CX3CL1、DAPK1、DAPK2、DLC1、DRAM1、FOS、GRID1、GRID2、HSPB8、MAP1LC3C、NLRC4、NRG1、NRG3、PPP1R15A、PRKCQ、TP63 在肿瘤组织中显著低表达。见图 1B、C。

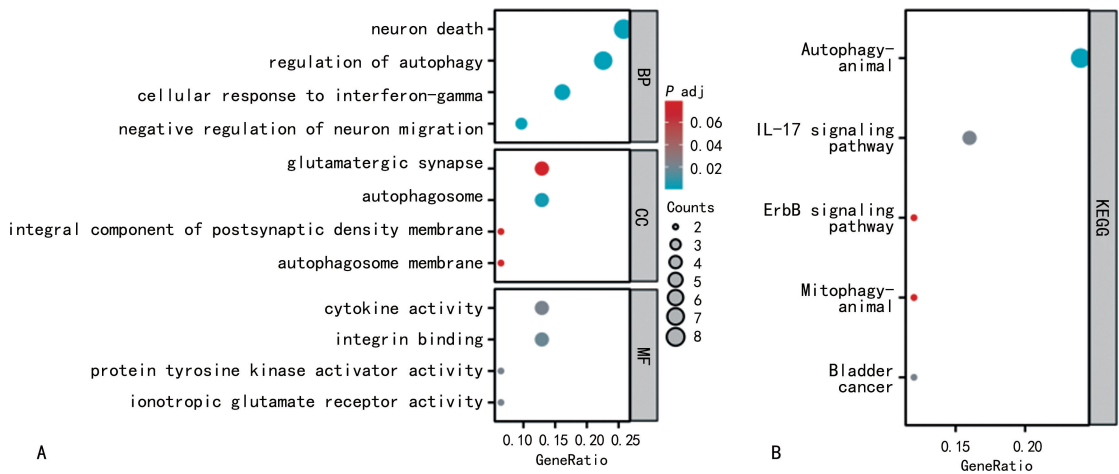
2.2 差异表达 ARGs 的功能富集分析 基因本体(GO)功能富集提示生物过程主要在神经元死亡、神经元迁移的负调控、自噬调节、细胞对干扰素 γ 的反应方面显著富集,细胞组成主要在自噬体、谷氨酸能突触、自噬体膜、突触后密度膜的组成部分方面显著富集,分子功能主要在整合素结合、离子型谷氨酸受体活性、蛋白酪氨酸激酶激活剂活性、细胞因子活性方面显著富集。京都基因与基因组百科全书(KEGG)通路主要在动物自噬、IL-17 信号通路、膀胱癌、动物线粒体吞噬、ErbB 信号通路上富集。见图 2。

2.3 预后相关 ARGs 在 31 个差异表达的 ARGs 中选出 10 个 $P < 0.05$ 的 ARGs,包括 5 个潜在风险基因(ATIC、EIF4EBP1、GAPDH、ITGB4 和 P4HB)和 5 个潜在保护基因(DAPK2、DLC1、DRAM1、MAP1LC3C 和 NLRC4)。见图 3。



注:A. 差异表达基因和 ARGs 的韦恩图;B. 肺腺癌与正常标本差异表达 ARGs 的热图,红色表示 ARGs 在相应组织中的高表达,绿色表示低表达;C. 肺腺癌中不同 ARGs 的表达情况,红色代表肺腺癌组织,蓝色代表正常肺组织,* $P < 0.05$ 。

图 1 肺腺癌中差异表达的 ARGs



注:A. GO;B. KEGG。

图 2 差异表达 ARGs 的 GO 和 KEGG 富集分析

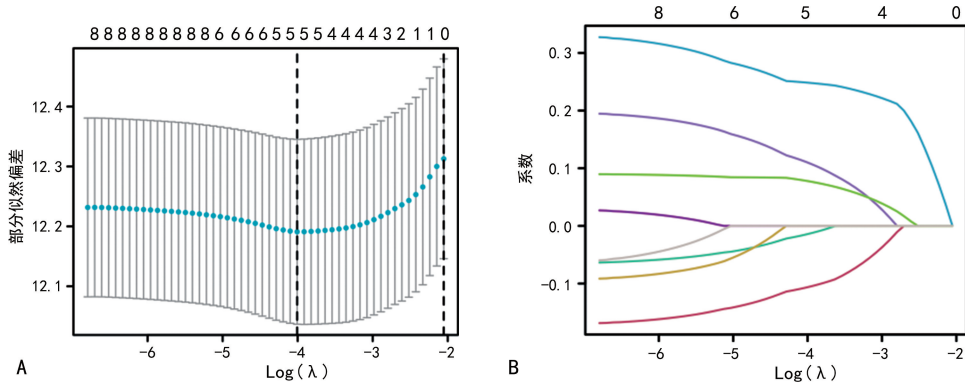
| 特征 | HR(95%CI) | P |
|----------|---------------------|---------|
| ATIC | 1.591 (1.259~2.010) | < 0.001 |
| DAPK2 | 0.717 (0.601~0.855) | < 0.001 |
| DLC1 | 0.886 (0.795~0.987) | 0.027 |
| DRAM1 | 0.864 (0.765~0.975) | 0.018 |
| EIF4EBP1 | 1.154 (1.022~1.304) | 0.021 |
| GAPDH | 1.590 (1.347~1.877) | < 0.001 |
| ITGB4 | 1.203 (1.093~1.323) | < 0.001 |
| MAP1LC3C | 0.894 (0.800~0.999) | 0.049 |
| NLRC4 | 0.801 (0.661~0.971) | 0.024 |
| P4HB | 1.284 (1.043~1.581) | 0.018 |

图 3 差异表达的 ARGs 单因素 Cox 分析森林图

2.4 Lasso 回归模型建立及变量筛选 对筛选得到的 10 个与肺腺癌患者预后相关的差异表达 ARGs 进行 Lasso 回归降维处理,通过 Lasso 参数筛选及绘制 Lasso 变量轨迹图得到构建预后模型的基因及对应的 Coef 值。见图 4。最终 ATIC、DAPK2、DRAM1、GAPDH、ITGB4 被鉴定为最有价值的预测基因,相应的 Coef 值分别为 0.110、-0.107、-0.014、0.248、0.078。风险评分 = ATIC 表达量 \times 0.110 + DAPK2 表达量 \times (-0.107) + DRAM1 表达量 \times (-0.014) + GAPDH 表达量 \times 0.248 + ITGB4 表达量 \times 0.078。

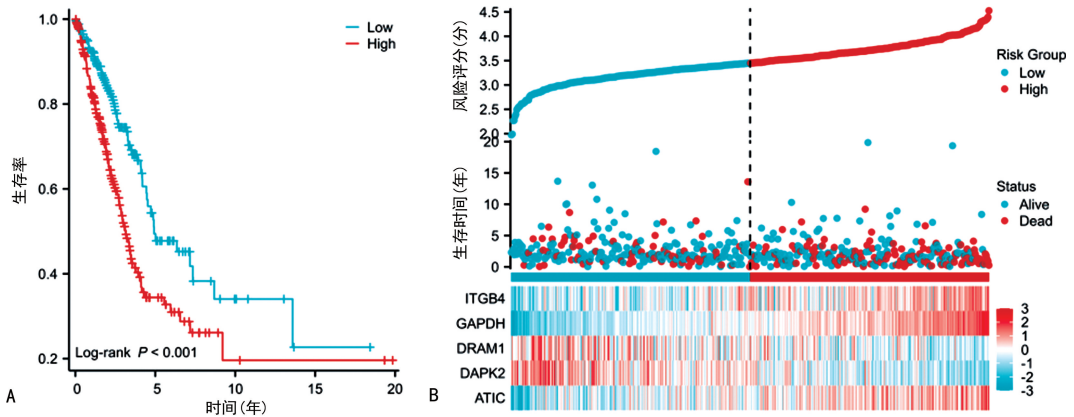
2.5 临床预后模型及验证 高风险组患者与较差的 OS 明显相关, 差异有统计学意义 ($P < 0.05$)。见图 5A。随着风险评分的增加, 高风险组患者死亡人数明显高于低风险组, DRAM1、DAPK2 为风险评分的保护因素, ITGB4、GAPDH、ATIC 为风险评分的危险因素。

因素。见图 5B。T 分期、N 分期、风险评分是肺腺癌患者的独立预后因素, 差异均有统计学意义 (危害比 = 1.931、1.674、2.650, 95% 可信区间: 1.278 ~ 2.920、1.197 ~ 2.376、1.823 ~ 3.853, $P = 0.002$ 、0.004、 < 0.001)。见图 5C、D。



注: A. Lasso 参数筛选; B. Lasso 变量轨迹图。

图 4 交叉验证调整 Lasso 回归模型中的参数筛选和 Lasso 回归模型中的系数分布



| 特征 | HR (95%CI) | P |
|------|-----------------------|---------|
| 年龄 | 0.999 (0.982 ~ 1.016) | 0.935 |
| 性别 | 1.137 (0.820 ~ 1.578) | 0.442 |
| T分期 | 2.445 (1.629 ~ 3.669) | < 0.001 |
| N分期 | 2.408 (1.734 ~ 3.345) | < 0.001 |
| M分期 | 1.900 (1.072 ~ 3.368) | 0.028 |
| 风险评分 | 3.254 (2.297 ~ 4.610) | < 0.001 |

| 特征 | HR (95%CI) | P |
|------|-----------------------|---------|
| T分期 | 1.931 (1.278 ~ 2.920) | 0.002 |
| N分期 | 1.674 (1.179 ~ 2.376) | 0.004 |
| M分期 | 1.526 (0.859 ~ 2.712) | 0.150 |
| 风险评分 | 2.650 (1.823 ~ 3.853) | < 0.001 |

注: A. 基于 TCGA 数据库肺腺癌患者高、低风险组绘制 Kaplan-Meier 生存曲线; B. 风险因子图; C. 单因素 Cox 回归分析的森林图; D. 多因素 Cox 回归分析的森林图。

图 5 临床预后模型及验证

2.6 列线图模型 基于 TCGA 数据库预测肺腺癌患者 1、3、5 年 OS 的列线图和校准曲线, 见图 6。模型

全局一致性为 0.710, 说明模型预测结果与实际情况具有较高的符合度, 可为肺腺癌患者预后提供相对稳

定、可靠的参考依据。

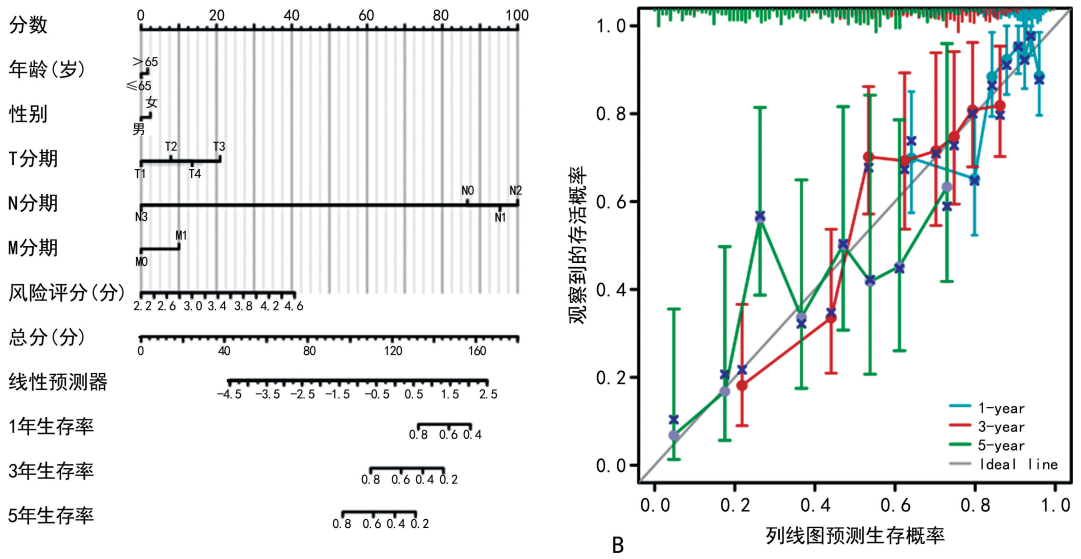


图 6 基于 TCGA 数据库预测肺腺癌患者 1、3、5 年 OS 的列线图 and 校准曲线

3 讨 论

在我国,每年肺癌发病人数超过 70 万,死亡人数超过 60 万。肺腺癌已成为发病率最高的肺癌病理类型,其发病机制复杂,对肺腺癌的诊治相关标志物的探索,如 EGFR、ALK、ROS1、RET、MET、NTRK、KRAS、BRAF、HER2、TP53、KEAP1、STK11 等,极大地改善了肺腺癌患者的预后^[1-6]。基于现有诊治水平的进步,肺癌患者生存期已有较大限度的提高,但原位期、局部期、晚期肺癌患者 5 年生存率仍分别只有 60%、33% 和 6%,在将恶性肿瘤变成慢性疾病愿景方面仍任重道远。

自噬是一种多步骤溶酶体降解途径,已被认为是调节癌症的过程。在肿瘤发生的早期,自噬作为一种生存途径和质量控制机制,阻止肿瘤的发生和抑制癌症进展。一旦肿瘤进展至晚期,自噬作为一种动态的降解和循环系统,有助于形成的肿瘤的生存和生长,并通过促进转移而促进癌症的侵袭性。已有众多证据表明,调节自噬可作为有效的癌症治疗策略^[7-8]。且已有越来越多的证据表明,自噬参与了免疫逃逸过程^[9],诱导或抑制自噬有助于提高免疫治疗的疗效^[10]。但挑战和机遇仍然存在,对肿瘤微环境中自噬更深入的了解可使得新的诊治策略成为可能。

本研究的目的是深入了解差异表达 ARGs 在肺腺癌中的作用,并基于差异表达 ARGs 构建肺腺癌患者的预后模型,为判断肺腺癌患者预后提供参考依据。共筛选出 5 个可作为肺腺癌患者的预后相关 ARGs,并进一步构建了风险模型,依据中位风险评分值将患者分为高、低风险组;同时,绘制 Kaplan-Meier 生存曲线,结果显示,高、低风险组患者 OS 差异显著。风险因子图显示,随风险评分的增加,高风险组患者死亡人数明显多于低风险组。单因素、多因素 Cox 分

析结果显示,风险评分为肺腺癌的独立预后因素。进一步基于以风险评分、年龄、性别、T 分期、N 分期、M 分期构建列线图及校准曲线,校准曲线提示,构建的模型全局一致性为 0.710,说明模型预测结果与实际情况具有较高的符合度,可为肺腺癌患者预后提供相对稳定、可靠的参考依据。表明自噬可能在肺腺癌发生、发展中具有非常重要的作用,对肺腺癌中自噬基因的研究,可能有助于进一步理解肺腺癌的发生、发展机制,为肺腺癌的诊疗提供新的思路。

本研究发现,用于构建预后模型的 5 个差异表达 ARGs 参与了诸多恶性肿瘤的疾病进程,在肿瘤的发生、发展中发挥着重要作用。有研究发现,ATIC 可通过蛋白激酶 B/FOXO3 信号通路抑制自噬并促进肝癌进展^[11]。在许多癌症中 DAPK2 可通过影响自噬过程进而影响许多细胞功能,包括增殖、迁移、侵袭和存活^[12]。GENG 等^[13]发现,DRAM1 在非小细胞肺癌中表达下降,与本研究结果相符,并且体外实验证明 DRAM1 可抑制非小细胞肺癌细胞的生长、迁移和侵袭。此外,DRAM1 可通过影响自噬-上皮间质转化过程调控肝癌细胞的迁移和侵袭^[14]。同时,GAPDH 可能在癌症中发挥促生存因子作用诱导自噬的发生,以支持快速增殖肿瘤细胞的能量消耗^[15]。BUTERA 等^[16]也发现,GAPDH 在恶性肿瘤中的自噬调节作用可作为癌症治疗的靶点。ITGB4 是一种跨膜受体,在肿瘤发生、发展中具有关键作用^[17]。ITGB4 mRNA 在结肠癌组织中表达显著上调,且与微卫星高度不稳定性有关,并且与癌症患者不良总生存率显著相关^[18]。与非肿瘤组织比较,口腔鳞癌组织 ITGB4 mRNA 及蛋白均显著过表达,提示 ITGB4 可能在口腔鳞癌的发生、发展中发挥关键作用^[19]。

综上所述,通过肺腺癌差异表达基因和 ARGs 构

建肺腺癌预后模型,能为肺腺癌患者的个体化诊治提供新的视角。但本研究只是基于 TCGA 数据库进行的信息挖掘,未经临床进一步验证,存在一定局限性,期望后续通过进一步临床研究验证预后模型的可靠性。此外,5 个 ARGs 的具体分子机制还有待于进一步研究。

参考文献

- [1] TERASHIMA M, ICHIKAWA W, OCHIAI A, et al. TOP2A, GGH, and PECAM1 are associated with hematogenous, lymph node, and peritoneal recurrence in stage II/III gastric cancer patients enrolled in the ACTS-GC study [J]. *Oncotarget*, 2017, 8(34):57574-57582.
- [2] LI Y, GUO X B, WEI Y H, et al. Serum CX-CL13 and PECAM-1 can be used as diagnostic and prognostic markers in elderly patients with gastric cancer [J]. *Clin Transl Oncol*, 2021, 23(1):130-138.
- [3] MIZUSHIMA N, KOMATSU M. Autophagy: Renovation of cells and tissues [J]. *Cell*, 2011, 147(4):728-741.
- [4] DAVIDSON M R, GAZDAR A F, CLARKE B E. The pivotal role of pathology in the management of lung cancer [J]. *J Thorac Dis*, 2013, 5(Suppl 5):S463-S478.
- [5] LANGER C J, BESSE B, GUALBERTO A, et al. The evolving role of histology in the management of advanced non-small-cell lung cancer [J]. *J Clin Oncol*, 2010, 28(36):5311-5320.
- [6] SEGUIN L, DURANDY M, FERLAL C C. Lung adenocarcinoma tumor origin: A guide for personalized medicine [J]. *Cancers (Basel)*, 2022, 14(7):1759.
- [7] AMARAVADI R K, KIMMELMAN A C, DEBNATH J. Targeting autophagy in cancer: Recent advances and future directions [J]. *Cancer Discov*, 2019, 9(9):1167-1181.
- [8] PATEL N H, BLOUKH S, ALWOHOSH E, et al. Autophagy and senescence in cancer therapy [J]. *Adv Cancer Res*, 2021, 150:1-74.
- [9] DUAN Y L, TIAN X Q, LIU Q, et al. Role of autophagy on cancer immune escape [J]. *Cell Commun Signal*, 2021, 19(1):91.
- [10] LEI Y H, ZHANG E X, BAI L L, et al. Autophagy in cancer immunotherapy [J]. *Cells*, 2022, 11(19):2996.
- [11] ZHANG H, XIA P, LIU J, et al. ATIC inhibits autophagy in hepatocellular cancer through the AKT/FOXO3 pathway and serves as a prognostic signature for modeling patient survival [J]. *Int J Biol Sci*, 2021, 17(15):4442-4458.
- [12] MOVAHHED P, SABERIYAN M, SAFI A, et al. The impact of DAPK1 and mTORC1 signaling association on autophagy in cancer [J]. *Mol Biol Rep*, 2022, 49(6):4959-4964.
- [13] GENG J, ZHANG R, YUAN X, et al. DRAM1 plays a tumor suppressor role in NSCLC cells by promoting lysosomal degradation of EGFR [J]. *Cell Death Dis*, 2020, 11(9):768.
- [14] CHEN C, LIANG Q Y, CHEN H K, et al. DRAM1 regulates the migration and invasion of hepatoblastoma cells via autophagy-EMT pathway [J]. *Oncol Lett*, 2018, 16(2):2427-2433.
- [15] COLELL A, RICCI J E, TAIT S, et al. GAPDH and autophagy preserve survival after apoptotic cytochrome c release in the absence of caspase activation [J]. *Cell*, 2007, 129(5):983-997.
- [16] BUTERA G, MULLAPPILLY N, MASETTO F, et al. Regulation of autophagy by nuclear GAPDH and its aggregates in cancer and neurodegenerative disorders [J]. *Int J Mol Sci*, 2019, 20(9):2062.
- [17] JIANG X, WANG J, WANG M Y, et al. ITGB4 as a novel serum diagnosis biomarker and potential therapeutic target for colorectal cancer [J]. *Cancer Med*, 2021, 10(19):6823-6834.
- [18] LI M, JIANG X, WANG G Q, et al. ITGB4 is a novel prognostic factor in colon cancer [J]. *J Cancer*, 2019, 10(21):5223-5233.
- [19] WANG X M, CHEN G, DANG Y W, et al. A comprehensive investigation regarding the clinical significance of ITGB4 in oral squamous cell carcinoma combining immunohistochemistry, RNA-seq, and microarray data [J]. *Comput Biol Chem*, 2023, 104:107846.

(收稿日期:2023-11-09 修回日期:2024-02-20)