

肺腺癌中胆固醇代谢亚型的探索与生存预测模型构建[△]

李敬儒*, 曹邦伟#(首都医科大学附属北京友谊医院肿瘤科, 北京 100050)

中图分类号 R979.1 文献标志码 A 文章编号 1672-2124(2024)12-1409-04

DOI 10.14009/j.issn.1672-2124.2024.12.001



摘要 目的:通过生物信息学方法分析胆固醇代谢相关基因在肺腺癌中的表达变化,并基于这些基因构建一个预后风险模型,以期对肺腺癌的诊断和药物治疗提供新的思路与靶点。方法:从癌症基因组图谱计划(TCGA)中获取600例肺腺癌样本的临床生存数据及RNA测序数据,通过差异表达分析鉴定在肿瘤和正常组织中表达不同的基因,并从分子特征数据库筛选胆固醇代谢相关基因。利用Boruta算法和多因素Cox比例风险模型构建胆固醇代谢评分模型,并在GEO数据库的GSE30219数据集中进行验证。结果:在TCGA数据库中,共鉴定出28个胆固醇代谢相关的差异表达基因。通过Boruta算法筛选出对肺腺癌患者总生存期有显著影响的18个基因,构建了胆固醇代谢评分模型。该模型在训练集和验证集中均表现出良好的预测能力,Kaplan-Meier生存曲线分析显示低代谢评分组的生存显著优于高代谢评分组。结论:本研究构建的胆固醇代谢评分模型在肺腺癌预后评估中具有显著的预测能力,揭示了胆固醇代谢在肺腺癌中的重要作用,为预后评估和靶向药物治疗提供了新的研究方向。

关键词 肺腺癌;胆固醇代谢;Boruta算法;Cox比例风险模型

Exploration of Cholesterol Metabolism Subtypes and Construction of Survival Prediction Models in Lung Adenocarcinoma[△]

LI Jingru, CAO Bangwei (Dept. of Oncology, Beijing Friendship Hospital, Capital Medical University, Beijing 100050, China)

ABSTRACT **OBJECTIVE:** To analyze the expression changes of cholesterol metabolism-related genes in lung adenocarcinoma through bioinformatics methods, and to construct a prognostic risk model based on these genes, in order to provide new thoughts and targets for the diagnosis and drug treatment of lung adenocarcinoma. **METHODS:** Clinical survival data and RNA sequencing data of 600 lung adenocarcinoma samples were obtained from the Cancer Genome Atlas (TCGA) project, genes differentially expressed in tumors and normal tissues were identified by differential expression analysis, and cholesterol metabolism-related genes were screened from the molecular signature database. The cholesterol metabolism scoring model was constructed by using the Boruta algorithm and the multifactorial Cox proportional hazard model, which were validated in the GSE30219 dataset of the GEO database. **RESULTS:** A total of 28 cholesterol metabolism-related differentially expressed genes were identified in the TCGA database. The Boruta algorithm screened out 18 genes that significantly affect the overall survival of lung adenocarcinoma patients, and constructed a cholesterol metabolism scoring model. This model showed good predictive ability in both the training set and validation set, and Kaplan-Meier survival curve analysis showed that the survival of the low metabolism score group was significantly better than that of the high metabolism score group. **CONCLUSIONS:** The cholesterol metabolism scoring model constructed in this study has significant predictive ability in the prognostic evaluation of lung adenocarcinoma, which reveals the important role of cholesterol metabolism in lung adenocarcinoma, provides a new research direction for prognostic evaluation and targeted drug therapy.

KEYWORDS Lung adenocarcinoma; Cholesterol metabolism; Boruta algorithm; Cox proportional hazard model

随着肿瘤代谢重编程概念的提出^[1],越来越多的研究发现肿瘤组织中存在显著的代谢改变,例如有氧糖酵解(Warburg效应)^[2]、谷氨酰胺代谢^[3]和乳酸代谢^[4]等。在这些代谢重编程中,胆固醇代谢的改变也有着不容忽视的作用,尤其对于发病率最高的恶性肿瘤之一——非小细胞肺癌,有

研究发现肺癌细胞中存在活跃的胆固醇代谢^[5],他汀类药物对于非小细胞肺癌的治疗也存在一定的辅助功能^[6]。但目前仍未明确这背后的生物学原理。因此,本研究采用生物信息学方法,基于胆固醇代谢相关基因构建肺腺癌预后风险模型,并选用外部数据库进行验证,以期对肺腺癌的诊断和治疗提供新的思路与靶点。

1 资料与方法

1.1 数据收集

从癌症基因组图谱计划(TCGA)中以肺腺癌为关键词检

△ 基金项目:国家自然科学基金资助项目(No. 82173056)

* 硕士研究生。研究方向:恶性肿瘤的代谢重编程。E-mail: jingruli8000@gmail.com

通信作者:主任医师。研究方向:恶性肿瘤的综合诊治。E-mail: oncology@cemu.edu.cn

索,获取 600 例肺腺癌样本的临床生存数据及 RNA 测序 (RNA-seq) 数据,其中包括 541 例肿瘤组织信息,59 例健康组织信息。从分子特征数据库 (MSigDB) 中的 C5 集以胆固醇代谢 (cholterol metabolism) 为关键词检索到 434 个相关基因。并从基因表达综合数据库 (GEO) 中的 GSE30219 数据集筛选出 85 例肺腺癌样本作为验证集。

1.2 鉴定肺腺癌组织与正常组织差异表达基因

分别使用 RStudio 中 DESeq2、edgeR 和 limma 包对肿瘤组织与正常组织中的 RNA-seq 进行差异分析,获取表达有所差异的基因。并与从 MSigDB 数据库中检索到的胆固醇代谢相关基因取交集,获得在肺腺癌中胆固醇代谢基因。

1.3 构建预测模型与使用验证集评价

使用 bourate 算法以生存时间为指标对 RNA-seq 及临床数据进行特性选择,筛除低重要性变量。根据重要性对基因进行赋值,将标准化后的测序值乘重要性并依次相加,总值计为胆固醇代谢评分。将代谢评分与临床数据联合构建多因素 COX 比例风险模型。计算验证集的胆固醇代谢评分,根据评分中位数分组,绘制 Kaplan-Meier (K-M) 生存曲线。基于 timeROC 包分析验证集代谢评分预测总生存期 (OS)。

1.4 统计学方法

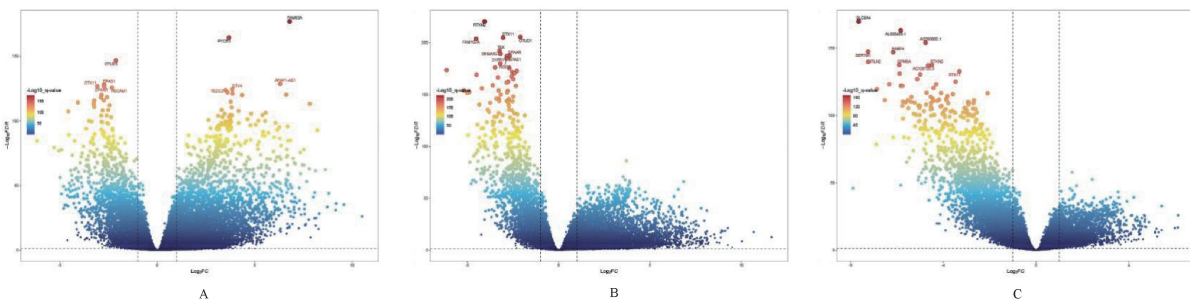
使用 R version 4.3.3 对数据进行统计分析。使用 DESeq2、edgeR 和 limma 对肿瘤组织与正常组织中的 RNA-seq 进行差异

分析,阈值为 $\log_{2}FC = 2.5$, P . Value = 0.01。使用 Boruta 算法对筛选后的变量进行特征选择,将重要性大于 shadowmax 值的变量纳入模型,根据重要性及基因表达计算胆固醇代谢评分。将 TCGA 数据集作为训练集构建多因素 cox 模型,GSE30219 数据集作为验证集进行模型评价,并根据代谢评分划分高低组别绘制 K-M 曲线比较两组间的生存。检验水准 $\alpha = 0.05$ 。

2 结果

2.1 肺腺癌中胆固醇代谢相关基因

于 TCGA 数据库中获取 541 例肺腺癌肿瘤组织基因表达水平及 59 例正常组织基因表达水平。(1) 分别使用 DESeq2、edgeR 和 limma 包对上述信息进行差异分析 [图 1(A)—图 1(C)],筛选阈值设定为 $\log_{2}FC = 2.5$, P . Value = 0.01。通过上述分析方法,确定了肿瘤组织相较于正常组织表达水平有差异的基因。再取 3 种分析方法结果的交集,得到 739 个表达有所差异的基因。(2) 从 MSigDB 数据库中的 C5 集以“cholterol metabolism”为关键词检索到 434 个相关基因。(3) 将 739 例表达差异基因,与 434 例代谢相关基因取交集,见图 2。结果得到以下 28 个基因,① 上调基因: *TOP2A*、*KIF4A*、*KIF20A*、*HJUBP*、*ASPM*、*DLGAP5*、*CENPF*、*CDC6*、*CDC20*、*COL11A1*、*CEP55*、*RRM2*、*CCNB1*、*CCNB2*、*NDC80*、*HMMR*、*NCAPG*、*FGL1*、*PBK*、*CHRNA5*、*COLA1*、*MUC5B*; ② 下调基因: *HBB*、*CD36*、*SYNPO2L*、*DES*、*CNTN6*、*MTIM*。



A. DESeq2; B. edgeR; C. limma。

图 1 基于 RStudio 的肺腺癌组织与正常组织 RNA-seq 水平差异比较

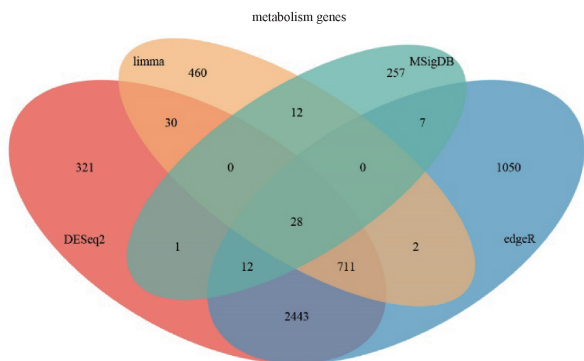


图 2 差异基因与胆固醇代谢相关基因取交集

2.2 特征选择

从 TCGA 数据集中筛选出合适的 517 例肿瘤样本。将样本的基因表达量、年龄、性别及肿瘤 TNM 分期纳入特征选择,使用 survial 包建立以 OS 与生存状态为指标的生存函数,再使用 bourate 算法进行特性选择。结果显示, *ASPM*、*CCNB1*、

CCNB2、*CDC20*、*CDC6*、*CENPF*、*CEP55*、*COL11A1*、*DLGAP5*、*HJURP*、*HMMR*、*KIF20A*、*KIF4A*、*NCAPG*、*NDC80*、*PBK*、*RRM2* 和 *TOP2A* 的表达水平对肺腺癌患者的 OS 存在显著影响,见图 3。根据重要性建立代谢评分,公式为: $ASPM \times 3.58 + CCNB1 \times 5.91 + CCNB2 \times 4.77 + CDC20 \times 6.32 + CDC6 \times 5.26 + CENPF \times 3.58 + CEP55 \times 5.2 + COL11A1 \times 5.93 + DLGAP5 \times 4.88 + HJURP \times 4.22 + HMMR \times 4.19 + KIF20A \times 3.51 + KIF4A \times 6.35 + NCAPG \times 5.72 + NDC80 \times 5.45 + PBK \times 4.72 + RRM2 \times 4.48 + TOP2A \times 5.8$ 。

2.3 模型构建与验证

将 517 例 TCGA 样本作为训练集,以年龄、性别、TNM 分期和代谢评分进行多因素 cox 分析,可以明确胆固醇代谢评分是肺腺癌患者生存的独立危险因素 ($P = 0.0086$),见图 4。

使用从 GSE30219 数据集中筛选出 85 例肺腺癌样本作为验证集,并根据胆固醇代谢评分进行分组及预测。使用 timeROC 包,以胆固醇代谢评分为指标,对验证集数据绘制随时间改变的受试者工作特征 (ROC) 曲线,对 OS 进行预测评估,结果显示,第 12、24 及 36 个月预测效果均较为良好,计算

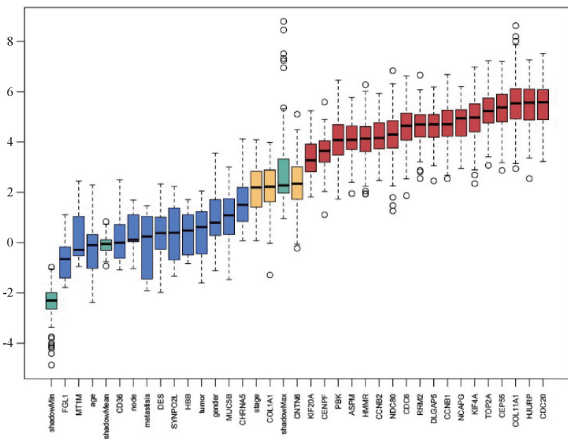


图3 基于 Boruta 算法的样本基因表达水平及临床信息 (按照生存时间及状态进行重要性排序)

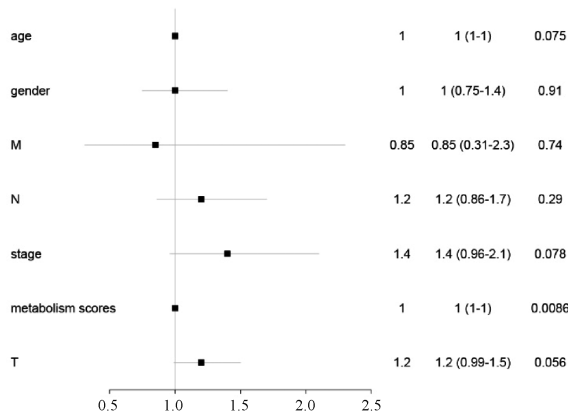


图4 多因素 COX 分析下胆固醇代谢评分及余信息的风险率 (HR)、置信区间 (CI) 及显著性概率

曲线下面积 (AUC), 分别达 0.68、0.72 和 0.74, 见图 5。按照胆固醇代谢评分的中位数将验证集分为高代谢组和低代谢组, 绘制 K-M 生存曲线, 低代谢评分组生存显著优于高代谢评分组 ($P=0.016$), 见图 6。

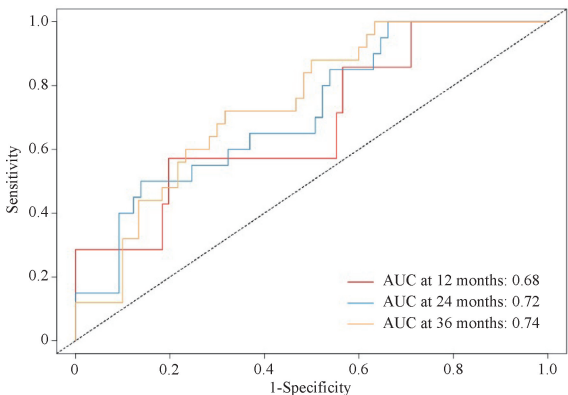


图5 验证集 ROC 曲线 (以胆固醇代谢评分为指标)

3 讨论

随着肿瘤代谢重编程相关研究的不断深入, 发现代谢途径的改变对肿瘤细胞生长、生存有着广泛且巨大的影响。而代谢重编程与逃避免疫破坏在维持增殖信号传导、逃避生长抑制

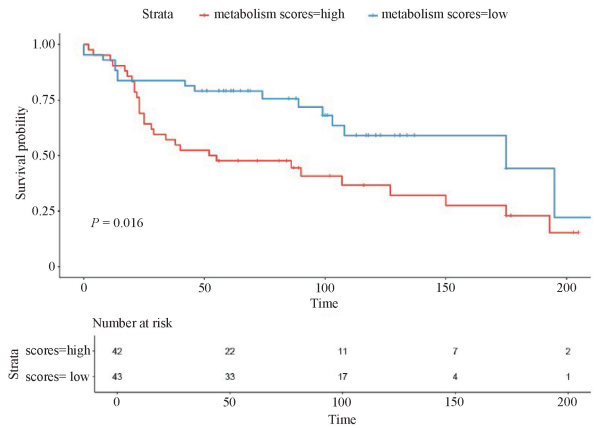


图6 验证集 (高代谢组和低代谢组) K-M 生存曲线

因子、抵抗细胞死亡、实现复制永生、诱导/获取血管供应和激活侵袭转移等六大肿瘤细胞共性后, 被确立为第七、八肿瘤细胞特性^[1]。

胆固醇代谢重编程是肿瘤细胞特性中不容忽视的一部分。有研究表明, 增加细胞内游离胆固醇可以增强 CD8⁺ T 细胞的效应功能和增殖^[7]; 但相反, 肿瘤微环境中胆固醇浓度升高会导致 CD8⁺ T 细胞耗竭^[8]。巨噬细胞中的胆固醇积累可以阻止肿瘤浸润巨噬细胞的凋亡, 防止其从 M2 型转变为 M1 型。若敲除小鼠中负责这一转运途径的 *ABCG1* 基因, 小鼠则表现出卓越的抗肿瘤特性^[9]。此外, 主要由胆固醇构成的脂筏, 是恶性肿瘤发生与发展过程中多种活跃信号通路所必需的结构基础^[10]。

近年来, 以胆固醇为靶点的他汀类药物因其抗肿瘤能力而受到认可, 在体外和临床研究中表现出优异的抗肿瘤性能^[6,11]。他汀类药物不仅通过香叶基香叶基焦磷酸介导的机制抑制肿瘤细胞的异常胆固醇摄取和合成, 还能促进肿瘤细胞凋亡^[12]。此外, 他汀类药物还可以通过破坏小鸟苷三磷酸酶的香叶基化和法尼基化来抑制肿瘤转移^[13]。

同时, 与胆固醇关联的脂质与脂蛋白也表现出一定的肿瘤相关性, 例如高密度脂蛋白 (HDL) 中关键成分载脂蛋白 A1 可以抑制肿瘤相关的血管生成, 并防止抑制性细胞从骨髓向肿瘤部位的募集^[14]。此外, 肿瘤细胞可以通过上调清道夫受体-B1 (SR-B1) 的表达, 加速从循环 HDL 中摄取胆固醇, 同时减少 ATP 结合盒转运蛋白 A1 的表达, 从而协同降低恶性肿瘤细胞的胆固醇输出^[15-16]。还有低密度脂蛋白 (LDL) 通过多种恶性肿瘤中的信号通路调节肿瘤生长和迁移, 许多恶性肿瘤中 LDL 受体 (LDLR) 的表达上调, 主要是为肿瘤细胞的脂质代谢重编程导致对胆固醇需求显著增加^[17-19]。研究表明, 三酰甘油 (TG) 可以导致由自然杀伤细胞组成的免疫监视系统的功能障碍^[20]。此外, 在肺癌患者中, 外周血树突状细胞 (DC) 数量显著减少, 细胞内脂质主要为 TG 的积累导致 DC 抗原呈递功能的降低^[21]。

本研究重点关注了胆固醇代谢途径的变化, 结合 TCGA 数据库和外部验证数据, 发现胆固醇代谢相关基因在肺癌中的表达变化显著, 并且通过多因素 Cox 回归分析, 明确了这些基因对患者预后的独立影响。本研究识别出的关键基因均在多个癌种中被报道与肿瘤进展相关。 *ASPM*、*CCNB1*、*CCNB2*、

CDC20 和 *CDC6* 等基因的胆固醇代谢相关功能的注释均来自一项卵巢癌研究,在该研究中,通过将卵巢癌细胞暴露于高孕酮环境中,从而揭示了一系列在肿瘤细胞中存在潜在调控胆固醇功能的基因^[22]。本研究的发现表明,这些基因在肺腺癌中同样可能通过胆固醇代谢途径发挥调控作用,为肺腺癌的机制研究提供了重要参考,为后续研究开辟了方向。

目前,临床诊治中使用的程序性细胞死亡配体表达水平^[23-25]和肿瘤突变负荷^[26]等用于肺腺癌预后预测的生物标志物,因各种原因,预测效果仍存在一定的局限性。本研究构建的胆固醇代谢评分模型,通过结合多种代谢相关基因,提供了一种新的预后评估工具,显著提升了对肺腺癌患者生存期的预测能力。模型在外部验证集中的良好表现,对第 12、24 和 36 个月的 OS 预测 AUC 达到 0.68、0.72 和 0.74,进一步证实了其稳健性和实用性。

本研究不仅为肺腺癌的预后评估提供了新的思路,同时也揭示了胆固醇代谢在肿瘤生物学中的潜在作用,为未来针对代谢途径的靶向治疗提供了理论依据和研究基础。然而,本研究也存在一些局限性,例如数据来源的多样性和基因表达调控机制的复杂性,未来研究需要进一步探讨这些代谢基因的具体功能及其在肿瘤进展中的作用机制。

参考文献

[1] HANAHAN D. Hallmarks of cancer: new dimensions[J]. *Cancer Discov*, 2022, 12(1): 31-46.

[2] LEBELO M T, JOUBERT A M, VISAGIE M H. Warburg effect and its role in tumorigenesis [J]. *Arch Pharm Res*, 2019, 42(10): 833-847.

[3] LI T, COPELAND C, LE A. Glutamine metabolism in cancer[J]. *Adv Exp Med Biol*, 2021, 1311: 17-38.

[4] FAUBERT B, LI K Y, CAI L, et al. Lactate metabolism in human lung tumors[J]. *Cell*, 2017, 171(2): 358-371. e9.

[5] MASLYANKO M, HARRIS R D, MU D. Connecting cholesterol efflux factors to lung cancer biology and therapeutics[J]. *Int J Mol Sci*, 2021, 22(13): 7209.

[6] CANTINI L, PECCI F, HURKMANS D P, et al. High-intensity statins are associated with improved clinical activity of PD-1 inhibitors in malignant pleural mesothelioma and advanced non-small cell lung cancer patients[J]. *Eur J Cancer*, 2021, 144: 41-48.

[7] YANG W, BAI Y B, XIONG Y, et al. Potentiating the antitumor response of CD8⁺ T cells by modulating cholesterol metabolism[J]. *Nature*, 2016, 531(7596): 651-655.

[8] MA X Z, BI E G, LU Y, et al. Cholesterol induces CD8⁺ T cell exhaustion in the tumor microenvironment [J]. *Cell Metab*, 2019, 30(1): 143-156. e5.

[9] SAG D, CEKIC C, WU R P, et al. The cholesterol transporter ABCG1 links cholesterol homeostasis and tumour immunity[J]. *Nat Commun*, 2015, 6: 6354.

[10] MOLLINEDO F, GAJATE C. Lipid rafts as major platforms for signaling regulation in cancer[J]. *Adv Biol Regul*, 2015, 57: 130-146.

[11] JIANG W, HU J W, HE X R, et al. Statins: a repurposed drug to fight cancer[J]. *J Exp Clin Cancer Res*, 2021, 40(1): 241.

[12] GÖBEL A, ZINNA V M, DELL'ENDICE S, et al. Anti-tumor effects of mevalonate pathway inhibition in ovarian cancer[J]. *BMC*

Cancer, 2020, 20(1): 703.

[13] ZHONG W B, WANG C Y, CHANG T C, et al. Lovastatin induces apoptosis of anaplastic thyroid cancer cells via inhibition of protein geranylgeranylation and *de novo* protein synthesis [J]. *Endocrinology*, 2003, 144(9): 3852-3859.

[14] ZAMANIAN-DARYOUSH M, LINDNER D, TALLANT T C, et al. The cardioprotective protein apolipoprotein A1 promotes potent anti-tumorigenic effects[J]. *J Biol Chem*, 2013, 288(29): 21237-21252.

[15] MOOBERRY L K, SABNIS N A, PANCHOO M, et al. Targeting the SR-B1 receptor as a gateway for cancer therapy and imaging [J]. *Front Pharmacol*, 2016, 7: 466.

[16] KONTUSH A, LINDAHL M, LHOMME M, et al. Structure of HDL: particle subclasses and molecular components [M]//VON ECKARDSTEIN A, KARDASSIS D. *High Density Lipoproteins: From Biological Understanding to Clinical Exploitation*. Cham: Springer International Publishing, 2015: 3-51.

[17] DOS SANTOS C R, DOMINGUES G, MATIAS I, et al. LDL-cholesterol signaling induces breast cancer proliferation and invasion [J]. *Lipids Health Dis*, 2014, 13: 16.

[18] NAITO S, MAKHOV P, ASTSATUROV I, et al. LDL cholesterol counteracts the antitumour effect of tyrosine kinase inhibitors against renal cell carcinoma [J]. *Br J Cancer*, 2017, 116(9): 1203-1207.

[19] JUNG Y Y, KO J H, UM J Y, et al. LDL cholesterol promotes the proliferation of prostate and pancreatic cancer cells by activating the STAT3 pathway[J]. *J Cell Physiol*, 2021, 236(7): 5253-5264.

[20] SUN J C, BEILKE J N, LANIER L L. Adaptive immune features of natural killer cells[J]. *Nature*, 2009, 457(7229): 557-561.

[21] ARAI R, SODA S, OKUTOMI T, et al. Lipid accumulation in peripheral blood dendritic cells and anticancer immunity in patients with lung cancer[J]. *J Immunol Res*, 2018, 2018: 5708239.

[22] WILCOX C B, FEDDES G O, WILLETT-BROZICK J E, et al. Coordinate up-regulation of TMEM97 and cholesterol biosynthesis genes in normal ovarian surface epithelial cells treated with progesterone: implications for pathogenesis of ovarian cancer[J]. *BMC Cancer*, 2007, 7: 223.

[23] BORGHAEI H, PAZ-ARES L, HORN L, et al. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer[J]. *N Engl J Med*, 2015, 373(17): 1627-1639.

[24] HORN L, SPIGEL D R, VOKES E E, et al. Nivolumab versus docetaxel in previously treated patients with advanced non-small-cell lung cancer: two-year outcomes from two randomized, open-label, phase III trials (CheckMate 017 and CheckMate 057) [J]. *J Clin Oncol*, 2017, 35(35): 3924-3933.

[25] RITTMAYER A, BARLESI F, WATERKAMP D, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial [J]. *Lancet*, 2017, 389(10066): 255-265.

[26] RIZVI N A, HELLMANN M D, SNYDER A, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer [J]. *Science*, 2015, 348(6230): 124-128.

(收稿日期:2024-05-20 修回日期:2024-08-12)