

基于失巢凋亡相关基因构建膀胱癌预后模型的研究

Construction of a prognostic model for bladder cancer based on loss – of – nest apoptosis – related genes

汪露^{1,2a}, 陈琳^{2a,2b},
谷延伦^{2a,2b}, 董秉琪^{2a,2b},
陈洁^{2a,2b}, 崔一民^{1,2a,2b}□

(1. 温州医科大学 药学院, 浙江 温州 325035;
2. 北京大学 第一医院, a. 临床药理研究所;
b. 药学部, 北京 100034)

WANG Lu^{1,2a}, CHEN Lin^{2a,2b},
GU Yan – lun^{2a,2b},
DONG Bing – qi^{2a,2b},
CHEN Jie^{2a,2b},
CUI Yi – min^{1,2a,2b}

(1. School of Pharmacy, Wenzhou Medical University, Wenzhou 325035, Zhejiang Province, China; 2. a. Institute of Clinical Pharmacology; b. Department of Pharmacy, Peking University First Hospital, Beijing 100034, China)

作者简介: 汪露(1999 –), 女, 硕士研究生, 主要从事肿瘤临床与基础方面的相关研究

通信作者: 崔一民, 教授, 博士生导师

Tel: (010) 82802440

E – mail: cui. pharm@pkuhf. com

摘要:目的 在膀胱癌中建立失巢凋亡相关基因(ANRs)预后风险模型, 计算风险评分, 并分析高、低风险评分的膀胱癌患者与肿瘤微环境的关系。方法 通过公共数据库信息和Cox回归分析筛选预后相关的ANRs和临床独立风险因素。用最小绝对值收敛和选择算子算法(LASSO)分析和列线图建立预后风险模型。通过Kaplan – Meier生存分析和曲线下面积(AUC)验证预后风险模型准确度。通过CIBERSORT(<https://cibersortx.stanford.edu/>)和单样本基因集富集分析(ssGSEA)探究风险评分与肿瘤微环境的关系。结果 预后相关的ANRs分别是B淋巴细胞瘤 – 2相关启动子(BAD)、细胞周期蛋白依赖性激酶抑制因子3(CDKN3)和增殖细胞核抗原(PCNA), 临床独立风险因素分别是性别、年龄、临床分期(T, N)和风险评分。预后风险模型表达为风险评分 = (0.155 2 × BAD表达量) + (0.228 6 × CDKN3表达量) + (0.011 4 × PCNA表达量)和列线图。风险评分越低表示膀胱癌患者预后越好, 生存曲线1、3、5年的AUC分别为0.732、0.620、0.541, 列线图1、3、5年校准曲线几乎为对角对应, 体现该模型准确。预后风险模型高、低风险组在膀胱癌肿瘤微环境中免疫细胞浸润方面呈现巨大差异。结论 建立的膀胱癌失巢凋亡相关基因预后风险模型准确度高, 可以较好地评估膀胱癌患者预后, 高、低风险评分的膀胱癌患者与肿瘤微环境关系密切。

关键词: 膀胱癌; 失巢凋亡; 治疗靶点; 肿瘤微环境; 预后价值

DOI: 10.13699/j.cnki.1001 – 6821.2025.02.018

中图分类号: R979.1 文献标志码: B

文章编号: 1001 – 6821(2025)02 – 0240 – 05

Abstract: Objective To develop a prognostic risk model for anoikis – related genes (ANRs) in bladder cancer, calculate risk scores, and analyze the relationship between bladder cancer patients with high and low risk scores and the tumor microenvironment. **Methods** Prognosis – related ANRs and clinically independent risk factors were screened by public database information and Cox regression analysis. Prognostic risk modeling was performed by least absolute shrinkage and selection operator (LASSO) analysis and column – line diagrams. Prognostic risk model accuracy was validated by Kaplan – Meier survival analysis and area under receiver operating characteristic curve (ROC) curve (AUC). The relationship between risk score and tumor microenvironment was explored by CIBERSORT (<https://cibersortx.stanford.edu/>) and single sample gene set enrichment analysis (ssGSEA). **Results** The prognostically relevant ANRs were B – lymphoblastoma – 2 – associated promoter (BAD), cell cycle protein – dependent kinase inhibitor 3 (CDKN3), and proliferating cell nuclear antigen (PCNA), and the clinically independent risk factors were gender, age, clinical stage (T, N), and risk score. The

prognostic risk model was expressed as risk score = $(0.1552 \times BAD \text{ expression}) + (0.2286 \times CDKN3 \text{ expression}) + (0.0114 \times PCNA \text{ expression})$ and column line graph. The lower the risk score the better the prognosis of bladder cancer patients, the AUC of the survival curves for 1, 3 and 5 years were 0.732, 0.620 and 0.541, respectively, and the column line graphs of the 1-, 3- and 5-year calibration curves almost corresponded diagonally, reflecting the accuracy of the model. The high and low risk groups of the prognostic risk model showed great differences in immune cell infiltration in the tumor microenvironment of bladder cancer. **Conclusion** The established prognostic risk model for bladder cancer loss of apoptosis-related genes is highly accurate and can better assess the prognosis of bladder cancer patients, and bladder cancer patients with high and low risk scores are closely related to the tumor microenvironment.

Key words: bladder cancer; anoikis; therapeutic target; tumour microenvironment; prognostic value

膀胱癌 (bladder cancer, BLCA) 作为常见的泌尿系统恶性肿瘤, 世界范围内发病率居恶性肿瘤的第9位, 且超过60%的患者会肿瘤复发^[1]。因此, 寻找新的治疗靶点和预后生物标志物仍然是目前研究的重点。细胞外基质和邻近细胞导致的细胞程序性死亡, 称为失巢凋亡^[2]。尽管失巢凋亡已被证明与多种实体肿瘤的进展和转移有关, 但仅有少数研究系统分析失巢凋亡相关基因在膀胱癌中的作用^[3]。探索新的预后标志物并给予不同治疗策略具有重要临床意义^[4], 本研究旨在探索膀胱癌失巢凋亡相关基因的差异表达及潜在亚型, 构建基于失巢凋亡相关基因差异表达的风险模型来预测患者预后, 为制定个性化和精准的膀胱癌治疗策略提供参考。

资料与方法

1 数据收集及构建预后风险模型

1.1 数据收集

从癌症基因组图谱 (the cancer genome atlas, TCGA)^[5] 和基因表达综合 (gene expression omnibus, GEO)^[6] 数据库下载 BLCA 的转录组表达谱和相关临床数据 (包括年龄、性别、总体生存时间、生存状态和临床分期)。获得 TCGA-BLCA 队列和 GEO-BLCA 队列 (GSE13507 队列) 进行后续分析。对原始文件进行调整和分位数归一化后将2个数据集合并 (排除没有所需信息的样品)。

失巢凋亡相关基因下载自 GeneCard 数据库 (<https://www.genecards.org/>) 和 Harmonizome 网站 (<https://maayanlab.cloud/Harmonizome/>), 筛选条件为相关性评分 > 0.4 ^[7-8]。以错误发现率 (false discovery rate, FDR) < 0.05 和 $|\log_2 FC| = 0.585$ 作为阈值, 用 “limma” R 包和 Wilcoxon 符号秩和检验获得膀胱癌患者中差异表达的失巢凋亡相关基因。

1.2 构建预后风险模型

用单因素 Cox 回归分析确定与预后显著相关的失巢凋亡相关基因, $P < 0.05$ 被认为是预后相关基因。基因风险比 (hazard ratio, HR) < 1 为低风险基因, 表示基因表达越高, 患者风险越低; HR > 1 为高风险基因, 表示基因表达越高, 患者风险越高。

用 “caret” R 包将最终纳入的 BLCA 患者按 1:1 随机分为训练集和验证集。用最小绝对值收敛和选择算子算法 (least absolute shrinkage and selection operator, LASSO) 并根据最佳 λ 值筛选出最显著的失巢凋亡相关基因, 以创建预后风险模型。选取二元离散变量模型和 R 包 “glmnet” 的默认参数构建模型, 将表达式数据标准化后计算预测模型的风险评分, 用以下公式计算每例膀胱癌患者的风险评分。

$$\text{RiskScore} = \sum_{j=1}^n \text{Coe}(\text{gene}_j) \times \text{Exp}(\text{gene}_j)$$

其中, Coe (gene_j) 是基因系数的缩写, Exp (gene_j) 是基因表达的缩写。

用 “glmnet” R 包构建预后风险模型^[9]。

体细胞拷贝数变异 (copy number variation, CNV) 信息下载自 XENA-UCSC (<http://xena.ucsc.edu/>) 数据库。用 “Rcircos” R 包绘制圈图展示拷贝数变异基因于染色体上的位置。用 GISTIC 2.0 在线分析用于处理 CNV 数据, 根据变异类型赋予 GISTIC 评分, 计算拷贝数变异频率并绘制箱线图。

2 验证预后风险模型

以中位风险评分为分界点, 根据预后风险模型计算的风险评分将膀胱癌患者分为高风险组和低风险组。

用 “survminer” R 包对高风险组和低风险组患者进行 kaplan-meier 生存分析, 从而进一步验证模型对膀胱癌患者预后预测的准确性。用 “survival” R 包和 “timeROC” R 包绘制受试者 1 年、3 年和 5 年工作特征 (receiver operating characteristic curve, ROC) 曲

线,根据曲线下面积(area under ROC curve, AUC)数值大小,越接近1则准确度越高,验证预后风险模型的准确度。

3 构建列线图

整理TCGA-BLCA和GSE13507数据集中所有肿瘤样本的临床病理信息,包括患者的生存时间、生存状态、年龄、性别、肿瘤分期和风险评分变量等。用“survival”R包对所有变量进行单因素Cox回归分析。随后将初步筛选到的临床病理因素作为变量与风险评分共同纳入多因素Cox回归分析,构建多因素回归模型,获得独立风险因素以构建列线图。用R软件中“rms”包的Lrm函数绘制列线图,将上述得到的复杂的回归模型转变为可视化的图形。在列线图中,可以根据刻度找到计算每个变量在不同情况下对应的单项得分,将所有变量在不同情况下得到的单项得分相加即为总得分,总得分对应预测率即为BLCA患者的生存预测概率。用“rms”包绘制BLCA患者1、3、5年的校准曲线来预测列线图准确度,校准曲线越符合对角对应列线图准确度越高。

4 肿瘤微环境

为探索此预后风险模型高、低风险组与BLCA肿瘤微环境之间的关系,将TCGA-BLCA队列与GSE13507队列合并,与包含22种免疫细胞的参考数据集共同作为输入文件。通过CIBERSORT工具可得到高、低风险组每个BLCA样本的免疫细胞浸润比例,每个样本中所有免疫细胞比例之和等于1。通过“ggpubr”“vioplot”“corrplot”和“pheatmap”R包对结果进行可视化,绘制相应的小提琴图、热图和柱状图。

基于单样本基因集富集分析(single sample gene set enrichment analysis, ssGSEA)对BLCA的样本进行免疫评分(immune score)和基质评分(stromal score),联合免疫评分和基质评分得到“ESTIMATE score”。通过“estimate”R包分析高、低风险组各个样本免疫细胞和基质细胞的丰度,绘制小提琴图,进一步探究BLCA的肿瘤微环境。

通过“tidyverse”“ggExtra”和“ggpubr”R包分析免疫细胞与风险评分之间的关系,其相关性通过spearman相关性分析来检验,并以散点图和热图的形式呈现。

5 统计学处理

本实验所有统计分析均用R 4.2.3进行。Kaplan-meier生存分析用于比较不同组别间总生存期(overall

survival, OS)差异。单因素、多因素Cox回归及Lasso回归分析用于预后变量筛选和构建模型。基因表达差异比较用 t 检验。用spearman相关性分析计算失巢凋亡相关基因(anoikis-related genes, ANRs)间的相关性。用Wilcoxon检验和Kruskal-Wallis检验进行非参数数据的两独立样本间比较和多样本间比较。从Bio Conductor包或R包下载相关R包,包括“ggplot2”“ggpubr”“survival”“survminer”等。

结 果

1 预后相关基因的获取及预后风险模型的构建

从GeneCard和Harmonizome数据库中下载652个ANRs,在肿瘤组织和正常组织间鉴定出206个差异表达的ANRs。单因素Cox回归分析在206例ANRs中筛选出B淋巴细胞瘤-2相关启动子(B-cell lymphoma-2 associated death promoter, BAD)、E2F转录因子1(E2F transcription factor 1, E2F1)、细胞周期蛋白依赖性激酶抑制因子3(cyclin-dependent kinase inhibitor 3, CDKN3)、果蝇zeste基因增强子人类同源物2(enhancer of zeste homolog 2, EZH2)、3型 β 微管蛋白(Beta tubulin type III, TUBB3)、细胞分裂周期蛋白25同源蛋白C(cell division cycle 25 homolog C, CDC25C)和增殖细胞核抗原(proliferating cell nuclear antigen, PCNA)这7个与BLCA预后相关的ANRs,且都是高风险基因(见图1A)。

用LASSO回归算法,获得最佳 λ 值为3(见图1B),即从7个BLCA预后相关的ANRs中选出3个(BAD、CDKN3和PCNA)纳入预后风险模型,根据系数分布图(见图1C)得到具体计算公式:风险评分 = $(0.1552 \times \text{BAD 表达量}) + (0.2286 \times \text{CDKN3 表达量}) + (0.0114 \times \text{PCNA 表达量})$, BAD、CDKN3和PCNA均是高风险基因。

拷贝数变异分析揭示拷贝数变异普遍存在于BLCA患者的ANRs中,其中EZH2、CDKN3和PCNA显示明显的CNV丢失, BAD、E2F1、TUBB3和CDC25C表现出显著的CNV扩增(见图2)。

2 预后风险模型的验证

在测试集上对该模型进行验证, kaplan-meier曲线显示,低风险组的BLCA患者比高风险组的OS更长($P < 0.05$),见图3,即风险评分越低BLCA患者的预后越好,证明模型预测结果准确。其1、3、5年的AUC分别为0.732、0.620和0.541,这表明对于BLCA患者预后的预测,该模型具有高准确度。

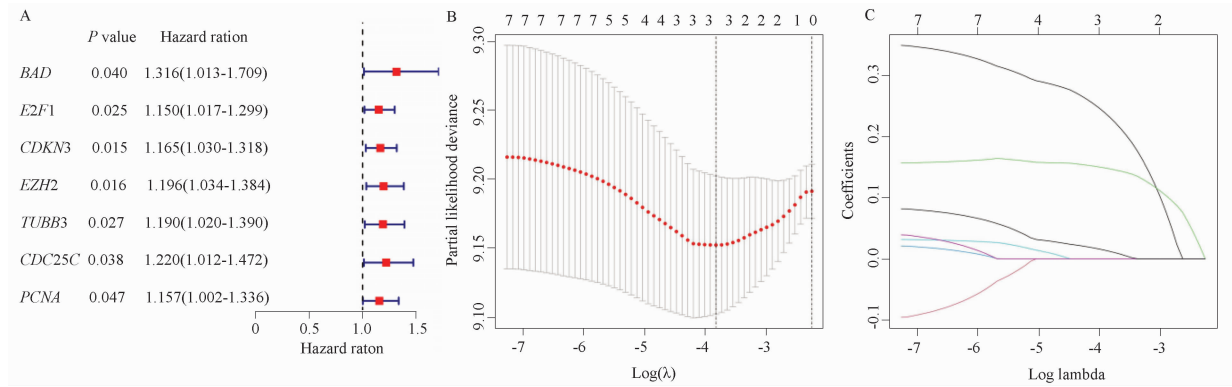


图1 预后风险模型的构建

Figure 1 Construction of prognostic risk models

A: One - way Cox regression analysis; B: Lasso regression cross - validation; C: Distribution of coefficients from Lasso analysis; *BAD*: Bcl - 2 associated death promoter; *E2F1*: E2F transcription factor 1; *CDKN3*: Cyclin - dependent kinase inhibitor 3; *EZH2*: Enhancer of zeste homolog 2; *TUBB3*: Beta tubulin type III; *CDC25C*: Cell division cycle 25 homolog C; *PCNA*: Proliferating cell nuclear antigen.

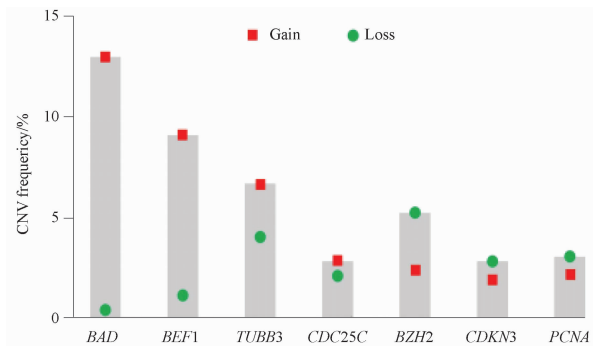


图2 失巢凋亡相关基因 (ANRs) 的拷贝数变异 (CNV) 分析

Figure 2 Frequency of copy number variation (CNV) in anoikis - related gene (ANRs)

3 构建列线图

Cox 回归分析筛选出性别、年龄、临床分期 (T、N) 和风险评估共 5 项独立风险因素建立列线图, 如图 4A, 列线图展示了 1 例 65 岁 (8 分)、N3 期 (4 分)、T3

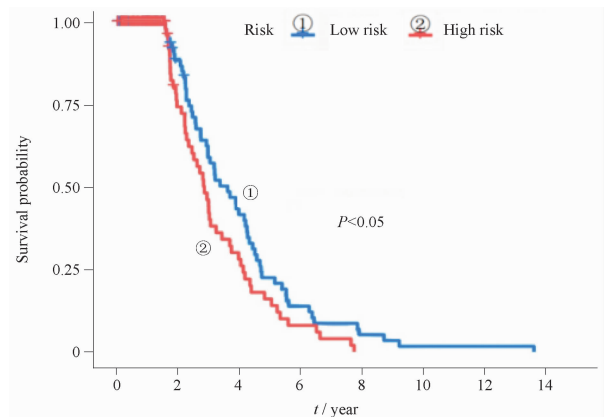


图3 测试集生存曲线

Figure 3 Test set survival curve

Low - risk group: Group of patients with risk scores calculated by the prognostic risk model that are lower than the median risk score; High - risk group: Group of patients with prognostic risk model - calculated risk scores higher than the median risk score; BLCA patients in the low - risk group had longer overall survival than those in the high - risk group, $P < 0.05$.

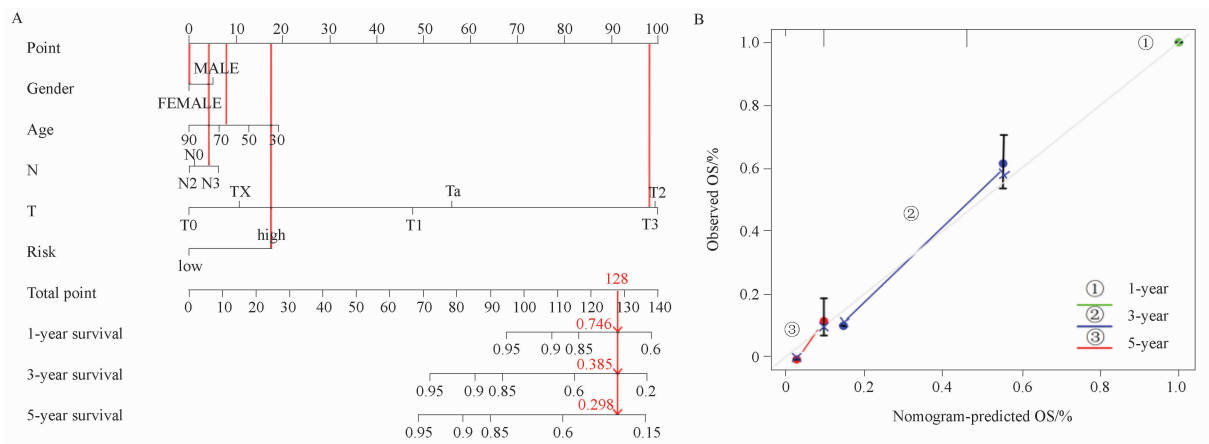


图4 列线图的构建及评价

Figure 4 Construction and evaluation of column line charts

A: Column line charts; B: Calibration curve charts; T: Tumor invasion; N: Lymph node metastasis; OS: Overall survival.

期(98分)、高风险评分(18分)的女性(0分)膀胱癌患者(总分128分)在1、3、5年的生存概率分别为0.746、0.385和0.298。1、3、5年校准曲线几乎为对角对应,展现列线图预测生存率和实际生存率较为一致,预测结果接近理想结果(见图4B)。

4 肿瘤微环境

小提琴图、热图和柱状图显示:幼稚B细胞、调节性T细胞、单核细胞3类免疫细胞在低风险组患者浸润更明显;而M1巨噬细胞、CD4幼稚型T细胞、CD4记忆激活型T细胞和滤泡辅助型T细胞在高风险组浸润明显。

散点图和热图显示: γ - δ T细胞和CD4记忆激活型T细胞与风险评分呈正相关,调节性T细胞和单核细胞与风险评分呈负相关。

小提琴图提示:高风险组中免疫细胞和基质细胞丰度均显著高于低风险组(均 $P<0.01$)。

讨 论

近年来,膀胱癌的发病率和死亡率均呈上升趋势,然而能够准确预测BLCA患者预后的生物标志物和模型很少。失巢凋亡已被证明在肿瘤(如乳腺癌^[10])的转移和生长中起关键作用,但是关于失巢凋亡相关基因对膀胱癌侵袭性的影响及其在预测BLCA预后中的作用尚不完全清楚。

本研究首先建立失巢凋亡相关基因的预后模型,得到风险评分公式,单、多因素Cox回归分析显示,ANRs预后模型所计算的风险评分可以独立于其他临床因素,作为BLCA患者的独立预后因子,从而构建列线图进一步探究与临床因素之间的关系。用校准曲线检测列线图的预后价值,提示该预后模型有望帮助一线医师在个体化的膀胱癌治疗上提供参考。膀胱癌肿瘤微环境不仅是膀胱癌发展和进展的关键因素,也是潜在的治疗靶点^[11]。本研究分析了22种免疫细胞在高、低风险评分组中的浸润差异。在高风险组中,M1巨噬细胞、CD4幼稚型T细胞、CD4记忆激活型T细胞和滤泡辅助型T细胞浸润水平显著上调,表明它们在膀胱癌的恶性进展中可能发挥作用。既往大量研究都提示乳腺癌属于高免疫原性肿瘤,本文的分析进一步印证了此理论,同时表明其对免疫治疗,如免疫检查点抑制药,可能有较好的响应。另外,

3个模型ANRs都与 γ - δ T细胞和CD4记忆激活型T细胞具有相关性,其中CDKN3与这2种免疫细胞相关性最强。因此,提示CDKN3/ γ - δ T细胞和CD4记忆激活型T细胞可能是膀胱癌的肿瘤微环境中值得进一步探索的通路。

本研究存在以下局限:首先,本研究基于对公共数据库的挖掘,样本量小,重要临床变量不完整,可能导致构建的模型存在一定的偏倚;其次,本研究还需要从实验上验证相关基因的表达及其调控膀胱癌进展的分子机制。

参考文献:

- [1] SIEGEL R L, MILLER K D, FUCHS H E, et al. Cancer statistics, 2023 [J]. *CA Cancer J Clin*, 2023, 73(1): 17—48.
- [2] FRISCH S M, FRANCIS H. Disruption of epithelial cell - matrix interactions induces apoptosis [J]. *J Cell Biol*, 1994, 124(4): 619—626.
- [3] RAEISI M, ZEHTABI M, VELAEI K, et al. Anoikis in cancer: The role of lipid signaling [J]. *Cell Biol Int*, 2022, 46(11): 1717—1728.
- [4] OSTROM Q T, GITTLEMAN H, LIAO P, et al. CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the united states in 2010 - 2014 [J]. *Neuro Oncol*, 2017, 19(suppl_5): v1—v88.
- [5] HUTTER C, ZENKLUSEN J C. The cancer genome atlas: Creating lasting value beyond its data [J]. *Cell*, 2018, 173(2): 283—285.
- [6] BARRETT T, TROUP D B, WILHITE S E, et al. NCBI GEO: Mining tens of millions of expression profiles—database and tools update [J]. *Nucleic Acids Res*, 2007, 35(3): D760—D765.
- [7] REBHAN M, CHALIFA - CASPI V, PRILUSKY J, et al. GeneCards: Integrating information about genes, proteins and diseases [J]. *Trends Genet*, 1997, 13(4): 163.
- [8] ROUILLARD A D, GUNDERSEN G W, FERNANDEZ N F, et al. The harmonizome: A collection of processed datasets gathered to serve and mine knowledge about genes and proteins [J/OL]. *Database (Oxford)*, 2016, 2016: baw100. 2016 - 07 - 03 [2024 - 05 - 10]. <https://pubmed.ncbi.nlm.nih.gov/27374120/>.
- [9] FRIEDMAN J, HASTIE T, TIBSHIRANI R. Regularization paths for generalized linear models via coordinate descent [J]. *J Stat Softw*, 2010, 33(1): 1—22.
- [10] YU Y, SONG Y, CHENG L, et al. CircCEMIP promotes anoikis - resistance by enhancing protective autophagy in prostate cancer cells [J]. *J Exp Clin Cancer Res*, 2022, 41(1): 188.
- [11] HU J, YU A, OTHMANE B, et al. Siglec15 shapes a non - inflamed tumor microenvironment and predicts the molecular subtype in bladder cancer [J]. *Theranostics*, 2021, 11(7): 3089—3108.

(收稿日期 2024 - 07 - 26; 本文编辑 孟海峰)