

膀胱癌溶酶体相关基因的预后模型构建与分析

安义均¹⁾, 余立丹²⁾, 赵美素³⁾, 马冬梅⁴⁾, 杨春花⁴⁾, 孔 瑶⁴⁾

(1)昭通市第一人民医院血管肿瘤科; 2)神经内科; 3)南院妇科; 4)放疗科, 云南 昭通 657000)

[摘要] **目的** 探索基于溶酶体相关基因的预后模型在膀胱癌患者中应用的可行性。**方法** 通过下载癌症基因组图谱(the cancer genome atlas program, TCGA)数据库中膀胱癌数据和基因表达综合数据库(gene expression omnibus, GEO)中GSE13507数据集。利用R语言通过差异分析、单因素比例风险模型(COX)回归分析筛选出TCGA数据库中膀胱癌生存相关的差异表达的溶酶体相关基因,采用最小绝对值收敛和选择算子算法(Lasso)回归模型构建出预后模型。根据构建模型风险评分的中位值划分出高、低风险组。使用生存分析比较高、低风险2组患者的生存差异并在GEO数据集中进行验证。采用单因素及多因素Cox回归分析验证风险评分是否为影响膀胱癌患者预后的独立危险因素。受试者工作特征曲线用于评估预后模型预测的准确性。GO及KEGG富集分析用于探索高、低风险组差异基因的生物学功能及信号通路。免疫分析用于探索高、低风险组免疫功能差异。**结果** 共筛选出44个差异表达的溶酶体相关基因,其中9个与预后相关基因用于预后模型构建,生存分析显示低风险组预后明显优于高风险组($P < 0.05$),并在GEO数据库中得到验证。构建模型预测膀胱癌患者1 a、3 a、5 a生存的ROC曲线下面积(area under the curve, AUC)分别为0.696、0.717、0.738。独立预后分析提示构建模型为膀胱癌患者独立预后影响因素;GO富集分析提示高、低风险组差异基因主要参与细胞结构功能相关;KEGG富集分析提示差异基因主要富集于PI3K-Akt信号通路。免疫分析提示2组患者免疫细胞浸润情况及免疫功能具有明显差异($P < 0.05$)。**结论** 膀胱癌溶酶体相关基因风险模型能准确有效地预测膀胱癌患者预后。

[关键词] 膀胱癌; 溶酶体相关基因; 预后模型; 生物信息学

[中图分类号] R737.14 **[文献标志码]** A **[文章编号]** 2095-610X(2024)05-0066-07

The Application of Prognostic Model of Lysosomal Related Genes in Bladder Cancer

AN Yijun¹⁾, YU Lidan²⁾, ZHAO Meisu³⁾, MA Dongmei⁴⁾, YANG Chunhua⁴⁾, KONG Yao⁴⁾

(1) Dept. of Vascular Oncology; 2) Dept. of Neurology; 3) Dept. of Gynaecology; 4) Dept. of Radiotherapy, The 1st People's Hospital of Zhaotong, Zhaotong Yunnan 657000, China)

[Abstract] **Objective** To explore the feasibility of applying the prognosis model based on lysosomal related genes(LRGs) in patients with bladder cancer (BC). **Methods** Bladder cancer data were downloaded from The Cancer Genome Atlas (TCGA) program and dataset of GSE13507 were downloaded from Gene Expression Omnibus (GEO). Differentially expressed LRGs related to the survival of BC in the TCGA database were screened by differential analysis and single factor proportional hazards model (COX) regression analysis via R software. LASSO regression was used to construct a prognostic risk model. BC patients were divided into the high and low risk groups according to the median risk score. Survival analysis were used to compare the survival differences between the high-risk and low-risk groups of BC patients, and validate in GEO database. Univariate and multivariate cox regression analysis were used to verify whether the risk scores were an independent risk factor affecting the prognosis of BC patients. The receiver operating characteristic (ROC) curve was used to evaluate the accuracy of prognostic model

[收稿日期] 2024-01-15

[基金项目] 云南省教育厅科学研究基金资助项目(2023Y0708)

[作者简介] 安义均(1991~), 男, 云南昭通人, 医学硕士, 主治医师, 主要从事恶性肿瘤放化疗研究工作。

[通信作者] 余立丹, E-mail: 1145930630@qq.com

predictions. GO and KEGG enrichment analysis were used to explore the biological functions and signaling pathways of differentially expressed genes between the high-risk and low-risk groups. Immunoassay was used to explore the differences in immune function between the high-risk and low-risk groups. **Results** A total of 44 differentially expressed lysosomal related genes were screened, of which 9 genes related to the prognosis were used to construct the prognosis model in this study. Survival analysis showed that the prognosis of the low-risk group was significantly better than that of the high-risk group ($P < 0.05$), which was verified in GEO database. The area under the ROC curve (AUC) of the BC prognosis risk scoring model to predict the 1-, 3- and 5-year survival of patients were 0.696, 0.717 and 0.738, respectively. Independent prognostic analysis showed that this prognostic risk model was an independent prognostic factor for BC patients. GO enrichment analysis indicated that the differential genes between the high-risk and low-risk groups were mainly involved in cell structure and function correlation. KEGG enrichment analysis suggested that the differential genes were mainly enriched in PI3K-Akt signaling pathway. Immunological analysis showed that there were significant differences in immune cell infiltration and immune function between the two groups. **Conclusion** Risk model of lysosomal related genes in BC can accurately and effectively predict the prognosis of BC patients.

[**Key words**] Bladder cancer; Lysosomal related genes; Prognostic model; Bioinformatics

膀胱癌是泌尿系统常见的恶性肿瘤, 根据相关的调查数据显示 2018 年全球每年新发的膀胱癌患者数达到 54 万例, 而每年因膀胱癌死亡的患者也接近 20 万例。在中国, 膀胱癌是发病率及死亡率最高的泌尿系肿瘤^[1-2]。

溶酶体是存在于真核细胞内的单层囊泡状细胞器, 其广泛分布于细胞质中, 是细胞内参与降解的主要细胞器^[3], 同时相关研究表明溶酶体广泛参与多种肿瘤增殖、侵袭、转移以及治疗耐药的调控过程。溶酶体相关基因的表达异常是调控溶酶体功能的重要因素, 进一步探索膀胱癌中差异表达的溶酶体相关基因与预后的关系有利于寻找潜在的干预靶点, 改善膀胱癌患者的预后。

本研究拟通过生物信息学方式构建膀胱癌溶酶体相关基因的预后模型用于预测膀胱癌患者预后, 为后续进一步探索膀胱癌发病机制, 寻找潜在的治疗靶点提供参考。

1 资料与方法

1.1 数据收集与整理

从 TCGA 数据库中下载膀胱癌转录组表达数据及相对应患者的临床信息数据, 数据下载截止日期为 2022 年 10 月 19 日, 共下载获得膀胱癌肿瘤样本 412 例, 癌旁正常组织样本 19 例。将 TCGA 数据库所下载的膀胱癌数据作为本研究的训练集。

从 GEO 数据库中下载 GSE13507 数据集的表达数据与临床数据并整理。将 GSE13507 作为本

研究的验证集。

从 KEGG 数据库中检索并下载溶酶体相关的基因(lysosomal related genes, LRGs), 共获得溶酶体相关基因 132 个。

1.2 LRGs 的筛选

使用 R 语言中的 Limma 包对下载整理的 TCGA 膀胱癌及癌旁正常组织样本表达数据与 KEGG 数据库中下载的溶酶体相关基因取交集, 从而获得溶酶体相关基因的表达数据。继续使用 Limma 包对溶酶体相关基因表达数据进行差异分析, 按照预先设定阈值 $|\log_2 \text{Fold change}| \geq 1$ 且 $P < 0.05$ 筛选获得膀胱癌差异表达的 LRGs, 使用 R 语言 pheatmap 包和 plot 包绘制膀胱癌差异表达的溶酶体相关基因的热图和火山图。

1.3 提取与预后相关的 LRGs 并构建风险模型

将 TCGA 差异表达的膀胱癌 LRGs 表达数据与对应的生存数据进行合并, 使用“Survival”和“Survminer”包, 设置过滤条件为 0.05, 采用单因素分析寻找与预后相关的 LRGs 并绘制森林图。

使用 R 语言“Glmnet”包和“Survival”包, 以 TCGA 样本为训练集、GEO 样本为验证集进行模型构建。首先对训练集数据进行 lasso 回归模型构建并绘制图形, 同时对数据进行交叉验证并得到交叉验证误差最小点所对应的基因及模型构建公式, 并绘制交叉验证图形。根据 TCGA 构建模型基因的表达量对其进行风险评分, 根据风险评分的中位值划分出高、低风险组。根据 TCGA 构建的模型公式提取 GEO 数据集中模型基因的表达

量并对其风险评分，同样按风险评分的中位值划分为高、低风险组。

1.4 生存分析

使用 R 语言“Survival”和“Survminer”包分别对构建的 TCGA 和 GEO 风险模型进行高、低风险组生存差异分析并绘制生存曲线图。

1.5 独立预后分析

分别采用单因素及多因素 Cox 回归分析对膀胱癌患者进行独立预后分析以寻找影响膀胱癌患者的预后因素。

1.6 ROC 曲线绘制

使用 R 软件“Survival”、“Survminer”、“TimeROC”包对构建的模型与其他临床形状进行生存的联合 ROC 曲线绘制来评价该模型的敏感性与特异性。同时对构建的模型分别绘制出膀胱癌患者 1 a、3 a、5 a 的生存 ROC 曲线。

1.7 功能富集分析

为了解所构建模型高、低风险 2 组差异表达基因的生物学功能及可能涉及的通路，使用 R 语

言“ClusterProfiler”包对高、低风险组中差异表达基因进行 GO 及 KEGG 富集分析。

1.8 免疫分析

使用“Limma”包对训练集中 412 例膀胱癌样本的免疫细胞浸润情况进行分析，然后使用“Limma”、“Reshape2”、“Ggpubr”包，采用 CIBERSORT 算法，设置过滤条件 $P=0.05$ ，对所构建模型的高、低风险 2 组样本中免疫细胞表达情况进行比较分析，并绘制结果箱线图。使用“Limma”、“GSVA”、“GSEABase”包，采用 ss GSEA 算法对高、低风险 2 组的免疫相关功能进行差异分析并绘制箱线图。

2 结果

2.1 膀胱癌中溶酶体相关基因的差异表达情况

通过差异分析获得差异表达的 LRGs 共 44 个，其中差异表达的上调基因共 35 个，下调基因 9 个，见图 1。

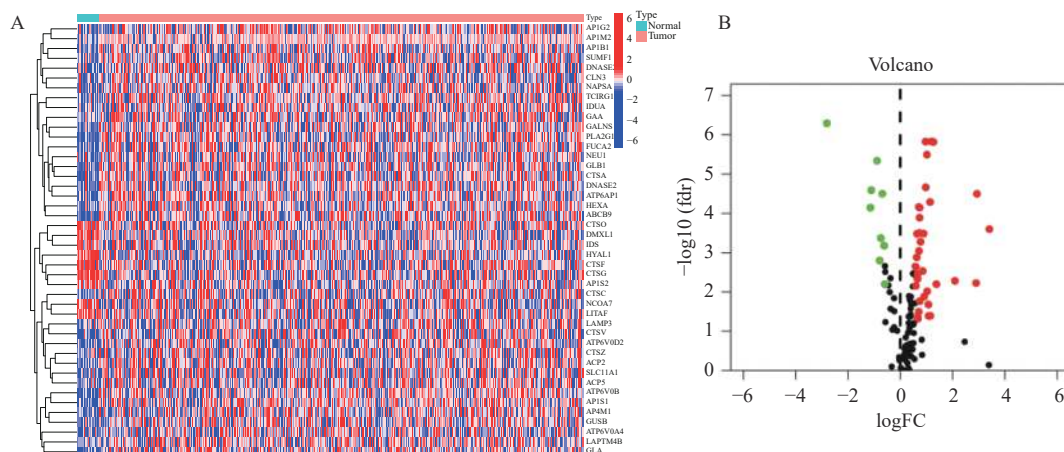


图 1 差异表达基因热图(左)及火山图(右)

Fig. 1 Heat map (left) and volcano map (right) of differentially expressed genes

A: 差异表达基因热图; B: 火山图。

2.2 溶酶体相关基因的预后模型构建

通过单因素分析得到与预后相关的 LRGs 共有 10 个 (TCIRG1、ATP6V0B、CTSA、NAPSA、IDUA、HYAL1、DNASE2B、ABCB9、CLN3、AP1G2)，见图 2。

通过 Lasso 回归分析及交叉验证最终确定 9 个与预后相关的差异表达的溶酶体相关基因 (TCIRG1、ATP6V0B、CTSA、NAPSA、IDUA、HYAL1、DNASE2B、ABCB9、CLN3) 用于预后模型构建，见图 3。

2.3 生存分析

生存分析发现通过模型构建后的高、低风险组在 TCGA 和 GEO 数据集中 2 组生存差异均有统计学意义 ($P < 0.05$)，见图 4。

2.4 独立预后分析

单因素及多因素分析均发现所构建的预后模型为膀胱癌患者的独立预后因素 ($P < 0.001$)，可有效预测患者预后，见图 5。

2.5 ROC 曲线

通过联合 ROC 曲线发现所构建的预后模型预

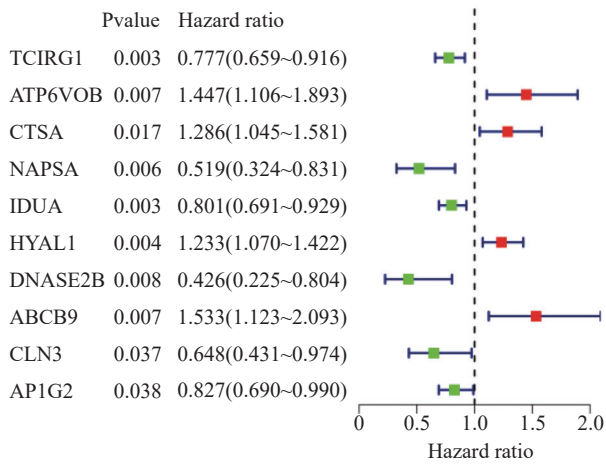


图 2 单因素 COX 分析

Fig. 2 Univariate COX analysis

测膀胱癌患者生存预后相比其他临床性状预测曲线下面积最大(AUC=0.738)。所构建模型预测 1 a、3 a、5 a 生存 ROC 曲线 AUC 分别为 0.696、0.717、

0.738, 见图 6。

2.6 基因富集分析结果

GO 富集分析显示高、低风险组差异表达基因主要富集在表皮发育、皮肤发育角质化、中间灯丝组织、角质形成细胞分化、外部封装结构组织、细胞外基质组织、细胞外结构组织等生物过程中。KEGG 富集分析提示高低风险组差异表达基因在 PI3K-Akt 信号通路中有统计学意义($P < 0.05$), 见图 7。

2.7 免疫分析结果

免疫分析结果显示 CD8⁺ T 细胞、滤泡辅助性 T 细胞(T cells follicular helper, Tfh)、调节性 T 细胞(T cells regulatory, Tregs)等 3 种免疫细胞在高、低风险组中表达具有统计学意义($P < 0.05$)。免疫相关功能差异分析显示: 抗原呈递细胞共抑制剂(APC co inhibition)、抗原呈递细胞共刺激(APC co stimulation)、趋化因子受体(CCR)、免疫检查点

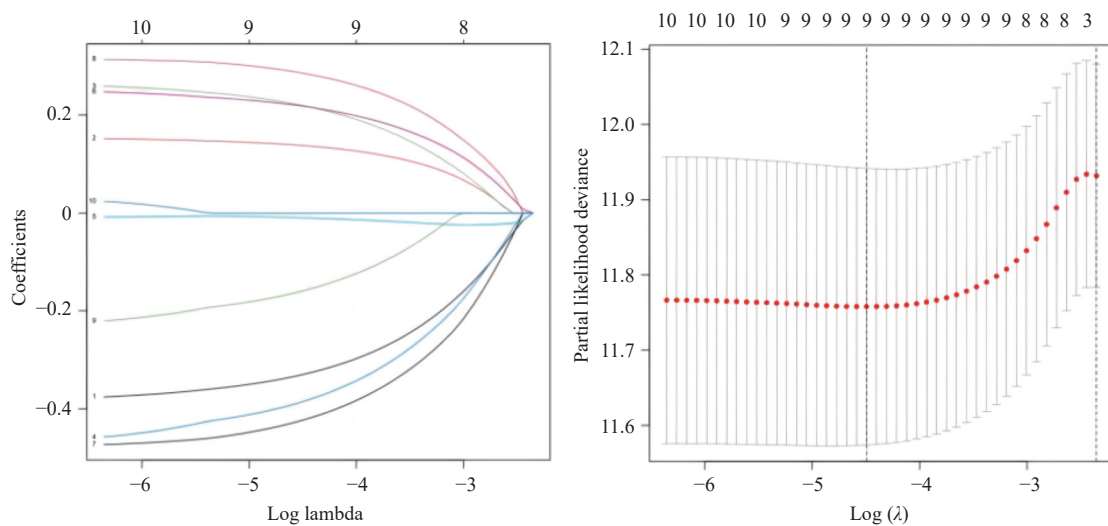


图 3 lasso 回归(左)和交叉验证(右)图形

Fig. 3 Lasso regression (left) and cross-validation (right) graph

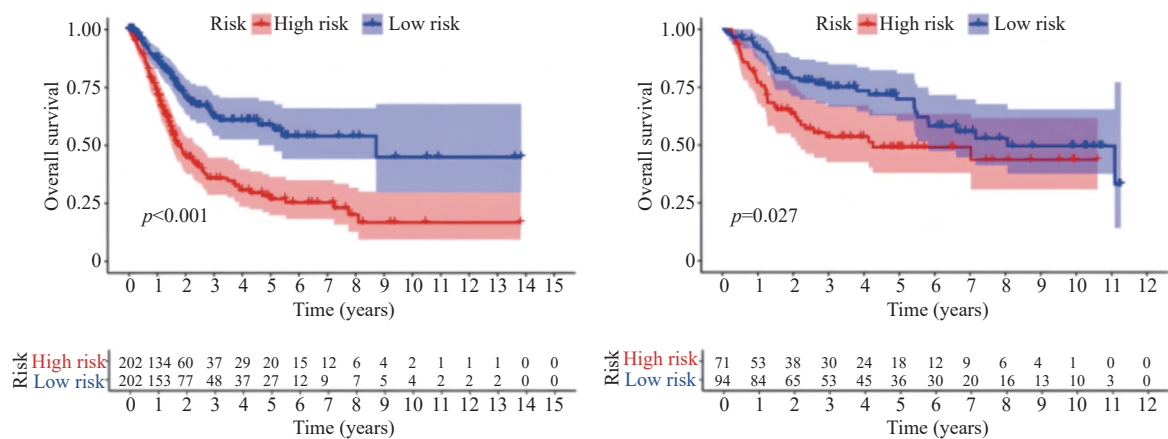


图 4 训练集和验证集的 Kaplan-Meier 分析曲线

Fig. 4 Kaplan-Meier analysis curve of training and verification sets

(Check-point)、炎症促进(Inflammation-promoting)、MHC 主要组织相容性复合体 I 类分子(MHC class I)、副炎症(Parainflammation)、T 细胞共抑制(T cell co-inhibition)、T 细胞共刺激(T cell co-stimulation)、

I 型干扰素(T cell co-stimulation)等 10 种免疫相关功能在高、低风险 2 组中存在统计学差异($P < 0.05$), 见图 8。

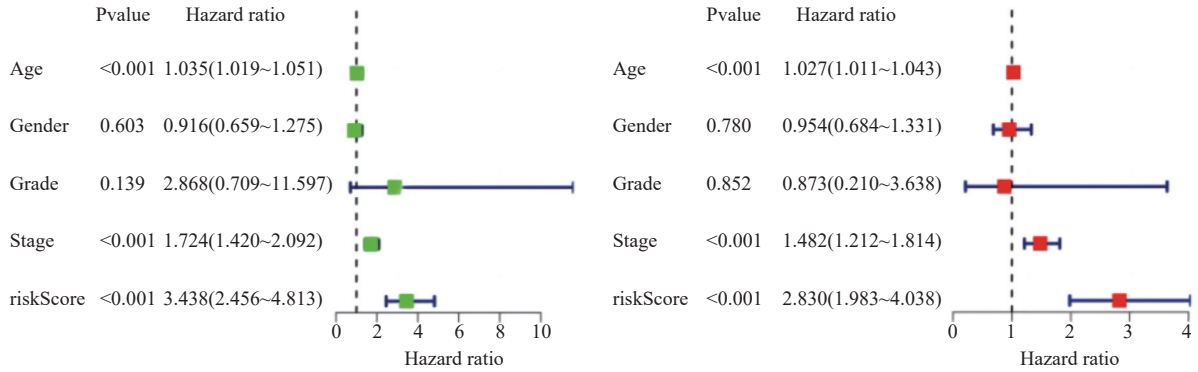


图 5 单因素(左)及多因素(右)分析

Fig. 5 Univariate (left) and multivariate (right) analysis

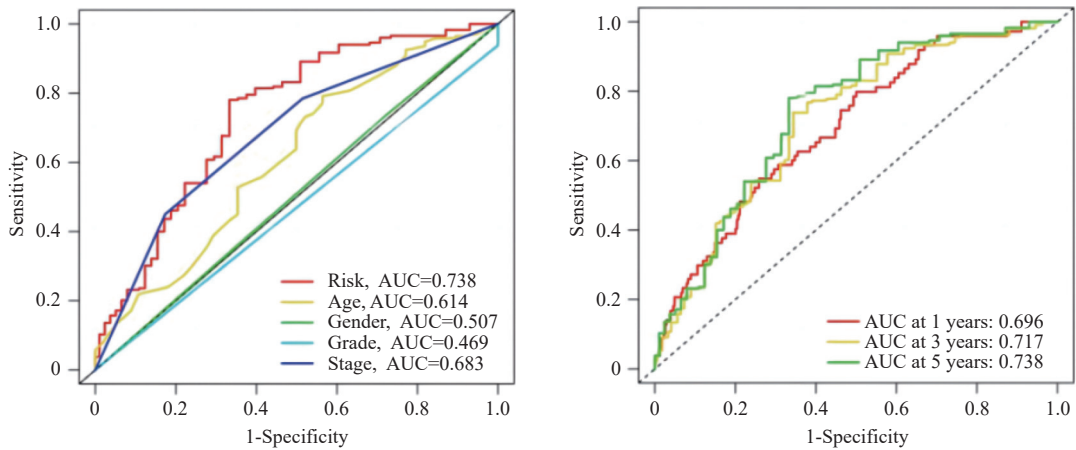


图 6 联合 ROC 曲线(左)和训练集 ROC 曲线(右)

Fig. 6 Joint (left) and training set (right) ROC curves

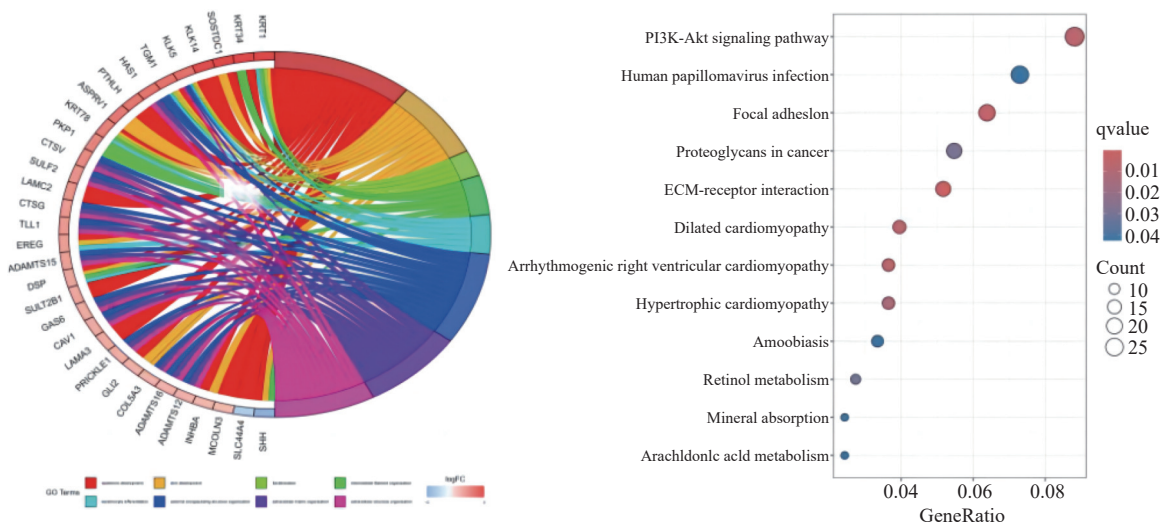


图 7 GO(左)和KEGG(右)富集分析

Fig. 7 GO (left) and KEGG (right) enrichment analysis

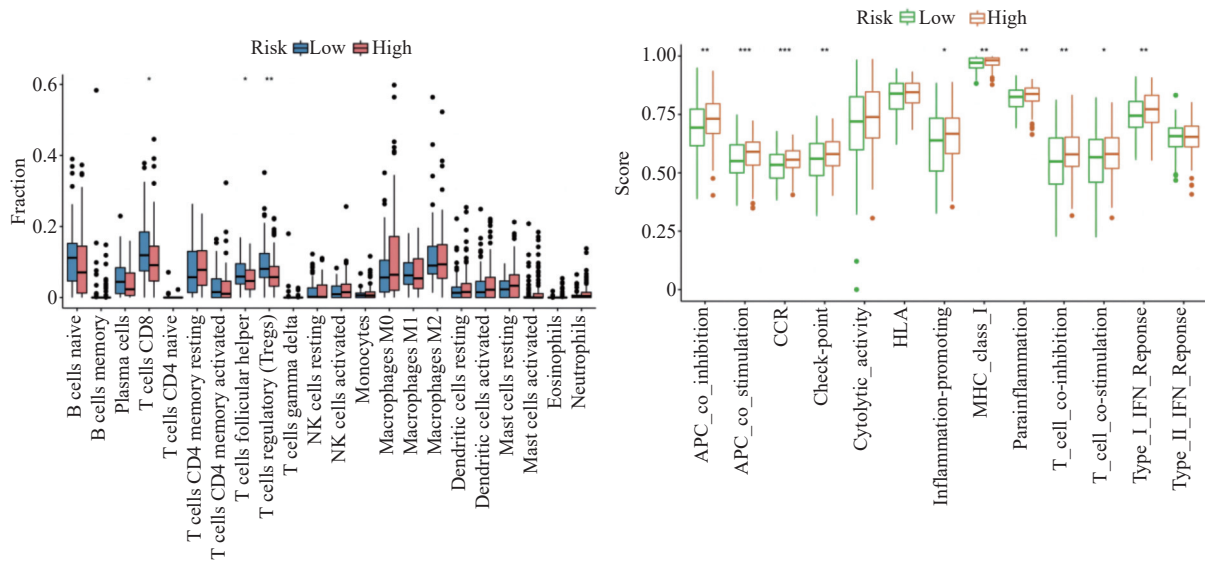


图 8 免疫浸润细胞分析(左)及免疫相关功能分析结果(右)

Fig. 8 The analysis results of immune infiltrating cells (left) and immune-related functions (right)

3 讨论

膀胱癌是目前发病率最高的泌尿系肿瘤,老年男性高发。近年来随着各种分子靶向治疗、免疫治疗等发展,膀胱癌患者的预后得到了一定的改善^[4],但由于膀胱癌早期症状隐匿、缺乏有效的标记物,整体上预后仍较差,进一步寻找膀胱癌预后相关的标志物对于膀胱癌患者预后判断极为重要。本研究通过 Lasso 回归筛选出 9 个与预后相关的差异表达的溶酶体相关基因用于预后模型构建,该模型可有效预测膀胱癌患者的生存预后并在 GEO 数据库中 GSE13507 数据集中得到了进一步验证,而且相比于其他临床性状,笔者所构建的模型其 ROC 曲线下面积最大,进一步表明了该模型预测的准确性及有效性。

本研究通过差异分析发现有 44 个溶酶体相关基因在 TCGA 膀胱癌与癌旁组织中差异表达,单因素分析发现这其中的 10 个差异表达的溶酶体相关基因与膀胱癌患者预后相关,而 Lasso 回归分析后最终确定了 9 个基因与预后相关的差异表达基因用于预后模型构建,并在 GEO 数据集中得到了验证。这 9 个基因通过影响溶酶体功能继而发挥作用,如参与溶酶体存储与转运、溶酶体糖蛋白的降解等。在癌症的转化过程中溶酶体的数量、形态、pH 值、酶的含量以及位置等都将发生改变^[5]。溶酶体在肿瘤微环境,特别是酸性环境的诱导下发生离心运动,从细胞质中央向周围运动,继而发生胞吐作用,溶酶体内的酸性水解酶释放

到细胞间质,导致细胞基质降解,从而加速肿瘤细胞的迁徙与转移^[6-8]。本研究中预后模型基因 HYAL1 基因所编码的透明质酸酶通过降解透明质酸进一步增强肿瘤细胞的侵袭能力,加速肿瘤转移^[9]。另外,溶酶体是分解代谢的主要场所,而分解代谢亢进是肿瘤的重要特征,溶酶体通过对自噬作用产生的底物进行降解,从而为肿瘤的生长与增殖提供物质基础,当溶酶体功能异常时这种作用将进一步增强,促进了肿瘤的增殖^[10],有研究表明本研究中的模型构建基因 ATP6V0B 基因可以通过 E2F1 介导溶酶体运输、mTORC1 系统信号激活、自噬,从而参与肿瘤转移过程^[11]。总之溶酶体及其功能异常是肿瘤发生、发展重要影响因素,所以对于溶酶体的靶向调控可能是肿瘤治疗的潜在靶点。

为进一步了解高、低风险 2 组的生物学功能差异,笔者进行了差异基因的 GO 和 KEGG 富集分析,通过富集分析发现 2 组主要集中在细胞结构组成、细胞外基质与结构等生物学方面存在差异。而 KEGG 富集分析发现 2 组差异基因主要富集于 PI3K-Akt 信号通路中,而相关的研究表明 PI3K-Akt 信号通路是膀胱癌增殖、转移的重要通路^[12]。同时对 2 组进行免疫分析发现 2 组的免疫细胞浸润情况及免疫功能存在较大差异,表明 2 组患者生存结局的不同可能与肿瘤免疫微环境等有关。

综上所述,本研究基于 TCGA 膀胱癌数据集成功鉴定了 9 个溶酶体相关基因作为膀胱癌患者预后预测的潜在标志物,同时在 GEO 数据库中得到了进一步验证,通过该模型可以筛选出高危患

者, 为治疗提供参考依据。

[参考文献]

- [1] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. *CA: A Cancer Journal for Clinicians*, 2018, 68(6): 394–424.
- [2] Zheng R, Zhang S, Zeng H, et al. Cancer incidence and mortality in China, 2016[J]. *Journal of the National Cancer Center*, 2022, 2(1): 1–9.
- [3] De Duve C. The lysosome turns fifty[J]. *Nature Cell Biology*, 2005, 7(9): 847–849.
- [4] Patel V G, Oh W K, Galsky M D. Treatment of muscle -invasive and advanced bladder cancer in 2020[J]. *CA: A Cancer Journal for Clinicians*, 2020, 70(5): 404–423.
- [5] Kallunki T, Olsen O D, Jäättelä M. Cancer-associated lysosomal changes: Friends or foes?[J]. *Oncogene*, 2013, 32(16): 1995–2004.
- [6] Reddy A. Plasma membrane repair is mediated by calcium (2+) -regulated exocytosis of lysosomes[J]. *Yale University*, 2001, 106(2): 157–169.
- [7] Mohamed M M, Sloane B F. Multifunctional enzymes in cancer[J]. *Nature Reviews Cancer*, 2006, 6(10): 764–775.
- [8] Dykes S S, Gray A L, Coleman D T, et al. The Arf-like GTPase Arl8b is essential for three-dimensional invasive growth of prostate cancer in vitro and xenograft formation and growth in vivo[J]. *Oncotarget*, 2016, 7(21): 31037–31052.
- [9] Tan J X, Wang X Y, Su X L, et al. Upregulation of HYAL1 expression in breast cancer promoted tumor cell proliferation, migration, invasion and angiogenesis[J]. *PLoS One*, 2011, 6(7): e22836.
- [10] Zhu S, Yao R, Li Y, et al. Lysosomal quality control of cell fate: A novel therapeutic target for human diseases[J]. *Cell Death & Disease*, 2020, 11(9): 1–13.
- [11] Meo-Evoli N, Almacellas E, Massucci F A, et al. V-ATPase: A master effector of E2F1-mediated lysosomal trafficking, mTORC1 activation and autophagy[J]. *Oncotarget*, 2015, 6(29): 28057–28070.
- [12] Rezaei S, Nikpanjeh N, Rezaee A, et al. PI3K/Akt signaling in urological cancers: Tumorigenesis function, therapeutic potential, and therapy response regulation[J]. *European Journal of Pharmacology*, 2023, 955(15): 175909.