

# 基于生物信息学筛选宫颈癌血管生成基因及构建预后模型

夏娜娜<sup>1,2</sup>, 杨京蕊<sup>1</sup>, 康敏<sup>1</sup>, 余敏敏<sup>1\*</sup>

(1. 南京市第二医院 妇科, 中国江苏 南京 210003; 2. 中国人民解放军总医院第七医学中心 药剂科, 中国北京 100700)

**摘要:** 利用生物信息学方法筛选与宫颈癌发生、发展和预后相关的血管生成相关基因(angiogenesis related gene, ARG), 并进行相关预后风险模型的构建与验证。首先, 从TCGA数据库中检索宫颈癌患者的表达谱和临床特征, 并提取差异表达的ARG; 其次, 采用Lasso Cox回归筛选预后ARG, 构建相关预后模型; 再次, 使用GSE52903和GSE44001数据集进行外部验证; 最后, 利用基因集富集分析(gene set enrichment analysis, GSEA)探讨宫颈癌预后机制。筛选结果显示, 共获得15个预后ARG, 分别为 *EFNA1*、*ITGA5*、*EPHB4*、*NRP1*、*CDH5*、*PLAU*、*BMP6*、*DLL4*、*JUN*、*CA9*、*MMP1*、*BAIAP2L1*、*SERPINF1*、*F2RL1* 和 *FGFR2*。GSE52903和GSE44001数据集的Kaplan-Meier生存曲线显示, 高风险组的总生存期(overall survival, OS) ( $P=0.005$ )和无病生存期(disease-free survival, DFS) ( $P<0.001$ )显著低于低风险组。受试者操作特征(receiver operating characteristic, ROC)曲线分析结果显示, GSE52903验证集在1年、3年和5年的曲线下面积(area under the curve, AUC)值分别为0.84、0.77和0.73, C-指数为0.72; GSE44001验证集在1年、3年和5年的AUC值分别为0.71、0.72和0.70, C-指数为0.70, 说明该模型对患者预后具有很强的预测效能。GSEA分析富集的通路主要涉及DNA复制、细胞外基质(extracellular matrix, ECM)受体相互作用、补体和凝血级联等, 这些过程与宫颈癌发生、发展紧密相关。以上结果表明, 这15个关键ARG可能是宫颈癌预后潜在的生物标志物。

**关键词:** 宫颈癌(CC); 生物信息学; 血管生成相关基因(ARG); 预后模型

中图分类号: Q811.4, R737.3

文献标志码: A

文章编号: 1007-7847(2024)02-0179-10

## Screening of Angiogenesis Related Genes and Construction of Prognostic Risk Model of Cervical Cancer Based on Bioinformatics

XIA Nana<sup>1,2</sup>, YANG Jingrui<sup>1</sup>, KANG Min<sup>1</sup>, YU Minmin<sup>1\*</sup>

(1. Department of Gynaecology, the Second Hospital of Nanjing, Nanjing 210003, Jiangsu, China; 2. Department of Pharmacy, the Seventh Medical Center of Chinese PLA General Hospital, Beijing 100700, China)

**Abstract:** Angiogenesis related genes (ARGs) associated with the occurrence, development and prognosis of cervical cancer (CC) were screened using bioinformatics methods, and then the related prognostic risk model was constructed and verified. Firstly, the expression profile and clinical characteristics of CC patients were searched and differentially expressed ARGs were extracted from TCGA database. Through Lasso Cox regression analysis, the ARGs for predicting prognosis were then selected to construct a relevant model. Furthermore, external validation was performed with the GSE52903 and GSE44001 datasets. Finally, the mechanism of CC prognosis was discussed by gene set enrichment analysis (GSEA). Fifteen prognosis-related ARGs were selected out, including *EFNA1*, *ITGA5*, *EPHB4*, *NRP1*, *CDH5*, *PLAU*, *BMP6*, *DLL4*, *JUN*, *CA9*, *MMP1*, *BAIAP2L1*, *SERPINF1*, *F2RL1* and *FGFR2*. Kaplan-Meier survival curves based on GSE52903 and GSE44001 datasets showed that the overall survival (OS) ( $P=0.005$ ) and disease-free survival (DFS) ( $P<0.001$ ) of high-risk group were significantly lower than those of low-risk group. Receiver operating characteristic (ROC) curve analysis showed that, in the GSE52903 validation set, the area under the curve (AUC) values at 1 year,

收稿日期: 2022-10-31; 修回日期: 2023-02-26; 网络首发日期: 2023-04-17

基金项目: 国家自然科学基金资助项目(81472431); 南京中医药大学自然科学基金项目(XZR2020070)

作者简介: 夏娜娜(1994—), 女, 河南商丘人, 硕士研究生; \*通信作者: 余敏敏(1967—), 女, 江苏南京人, 博士, 教授, 主任医师, 主要从事妇科肿瘤研究, E-mail: njyy022@njuem.edu.cn

3 years and 5 years were 0.84, 0.77 and 0.73, respectively, and the C-index was 0.72, and in the GSE44001 validation set, the AUC values were 0.71, 0.72 and 0.70, respectively, and the C-index was 0.70, indicating that the model has a strong predictive effect on the prognosis of CC patients. The pathways enriched by GSEA are mainly involved in DNA replication, extracellular matrix (ECM) receptor interactions, complement and coagulation cascades, the processes of which are closely related to cervical carcinogenesis and progression. These results suggest that the 15 key ARGs may be potential biomarkers for the prognosis of CC.

**Key words:** cervical cancer (CC); bioinformatics; angiogenesis related gene (ARG); prognostic model

(*Life Science Research*, 2024, 28(2): 179–188)

宫颈癌(cervical cancer, CC)的发病率和死亡率在女性恶性肿瘤中排名第二<sup>[1-2]</sup>, 仅次于乳腺癌。目前, 宫颈癌的治疗方法多为早期手术和中晚期放化疗, 对于晚期、反复发作或转移患者, 虽有放化疗但预后尚差。虽然肿瘤靶向治疗、免疫治疗在近几年开展得如火如荼, 但是现有靶向药物对宫颈癌的治疗需要在特定群体中才能显效; 针对宫颈癌的治疗, 新的、更加可靠的靶点还有待寻找以制定新的治疗方案。血管生成是指在原有血管系统基础之上形成一个以毛细血管为主要成分的新血管系统<sup>[3]</sup>。多种血管生成信号通路的激活可使肿瘤细胞在抗血管生成治疗后存活<sup>[4-5]</sup>, 这使得宫颈癌患者的临床治疗效果不佳。近年来, 人们已经研发出许多通过靶向破坏新生血管抑制肿瘤生长和改变肿瘤微环境的治疗药物<sup>[6]</sup>, 但血管生成相关基因(angiogenesis related gene, ARG)在宫颈癌中的临床相关性尚未完全阐明。因此, 进一步识别肿瘤血管生成的生物学机制, 寻找新的生物标志物进行靶向血管生成抑制治疗以改善宫颈癌预后具有重要意义。

本研究首先通过TCGA数据库下载了宫颈鳞癌和腺癌(cervical squamous cell carcinoma and endocervical adenocarcinoma, CESC)患者的表达谱与临床信息, 提取了差异表达的血管生成相关基因(differentially expressed ARG, DEARG); 然后, 采用Cox回归和Lasso回归筛选了预后血管生成相关基因(prognostic ARG, PARG), 构建了相关预后模型; 最后, 使用GSE52903和GSE44001数据集进行了外部验证, 并对模型基因进行了富集分析。相关结果为宫颈癌的发展和预后评估提供了有用的参考资料。

## 1 材料与方法

### 1.1 数据来源

从TCGA数据库<sup>[7]</sup>(<https://portal.gdc.cancer.gov/>)

下载已处理的CESC原始mRNA表达数据, 包括正常组和肿瘤组。利用R软件(3.6.3版)的limma包<sup>[8]</sup>对下载的mRNA level 3的FPKM (fragments per kilobase of exon model per million mapped reads)数据进行整合和标准化, 分析差异表达基因(differentially expressed gene, DEG)及其表达水平, 差异基因筛选条件为 $|\log_2(\text{FC})| > 1$  (FC: fold change)和 $P < 0.05$ <sup>[9]</sup>。从GEO公共数据库<sup>[10]</sup>(<https://www.ncbi.nlm.nih.gov/geo/>)下载GSE52903和GSE44001的Series Matrix File数据文件, 注释平台为GPL6244。通过GeneCards数据库<sup>[11]</sup>(<https://www.genecards.org>)获取ARG。数据获取时间为2021年12月。

### 1.2 GO和KEGG功能分析

使用clusterProfiler包<sup>[12]</sup>对预后基因进行功能注释, 以全面探讨这些预后基因的功能相关性。基因本体论(Gene Ontology, GO)及京都基因和基因组数据库(Kyoto Encyclopedia of Genes and Genomes, KEGG)被用来评估相关的功能类别。 $P < 0.05$ 和 $q < 0.05$ 被认为有统计学意义。

### 1.3 预后模型的构建与验证

选择DEARG, 采用Lasso回归来建立其预后相关模型。结合每个特定基因的表达评估, 我们不仅收集了每个患者的风险评分标准, 而且通过Lasso回归分析对其所有预测的回归系数进行了测量。根据风险评分公式, 以中位数风险评分值为分界点, 将患者分为低风险组和高风险组。采用Kaplan-Meier分析评估两组间的生存差异, 并采用log-rank检验法进行组间比较。然后, 采用Lasso回归模型分析和层次分析来评估相对风险评分在预测患者预后中的作用, 并采用受试者操作特征(receiver operating characteristic, ROC)曲线来衡量其建模的性能。

### 1.4 基于预后模型的独立预后分析及列线图的构建

使用单因素和多因素Cox回归分析来评估预

后模型是否可以独立于宫颈癌患者的其他临床病理变量,包括 TNM (tumor-node-metastasis)分期、分级、年龄以及风险评分。选择临床特征作为自变量,总生存期(overall survival, OS)作为因变量,计算危险比(hazard ratio, HR)、95%置信区间(confidence interval, CI)以及  $P$  值。选择年龄、分期、TNM 分期、风险评分作为因素,构建能够评估宫颈癌患者 5 年和 7 年 OS 概率的列线图。

### 1.5 GSEA 富集分析

对宫颈癌患者的表达谱进行基因集富集分析(gene set enrichment analysis, GSEA) (<http://www.broadinstitute.org/gsea>),以确定在高风险组和低风险组患者之间差异表达的基因。基因集不满足参数(15~500 个基因)的条目被抛弃。进行 100 次排列后,基于  $P < 0.05$  和错误发现率(false discovery rate, FDR)值小于 0.25 获得丰富的基因集。

## 2 结果

### 2.1 ARG 在 CESC 队列中的表达

采用 limma 包对 CESC 的 mRNA 表达数据进行组间差异分析,共获得 3 602 个差异基因(图 1A)。同时,通过 GeneCards 数据库获得 281 个相关性分数(relevance score)大于 3 的 ARG。将 281 个 ARG 与前面获得的宫颈癌差异基因取交集,发现有 115 个基因发生了重叠(图 1B)。

### 2.2 预后基因的筛选以及功能富集分析

为了找出 ARG 中的关键基因,我们收集了 CESC 患者的临床信息,并使用 Cox 单因素回归进行筛选。结果显示,经 Cox 单因素回归共筛选出 20 个预后相关基因( $P < 0.01$ ),依次为: *EFNA1*、

*ITGA5*、*TGFA*、*EPHB4*、*NRP1*、*CDH5*、*PLAU*、*BMP6*、*DLL4*、*JUN*、*CA9*、*MMP1*、*CD248*、*COL4A2*、*BCL2*、*BAIAP2L1*、*SERPINF1*、*F2RL1*、*MMP2* 和 *FGFR2*。GO 分析表明,这些预后基因可能与肽基酪氨酸磷酸化、肽基酪氨酸修饰、上皮细胞增殖调控、血管生成调控和血管发育调控的生物学特性密切相关(图 2A)。KEGG 分析表明,这些预后基因直接或间接地参与内分泌抵抗、EGFR 酪氨酸激酶抑制剂抵抗、糖尿病并发症中的 AGE-RAGE 信号通路、PI3K-Akt 信号通路和雌激素信号通路等(图 2B)。此外,通过 STRING 数据库 (<http://string-db.org>)分析了这些预后基因数据集中基因的蛋白质相互作用网络(图 2C)。进一步的 GeneMANIA 数据库(<https://genemania.org/>)分析表明,这 20 个基因及其相关分子(如 *EFNB2*、*COL4A1*、*SOX8*、*KDR*、*FDF10*、*BBS7*、*WARS1* 等)的功能主要与上皮管芽端分裂、血管生成、上皮管和上皮结构的形态发生、出芽式血管生成等有关(图 2D)。

### 2.3 ARG 的预后模型构建

根据前文提取的 20 个预后相关基因与 TCGA 数据集中患者生存之间的关系,选择一个 Lasso Cox 回归模型来识别有价值的预后基因。基于 TCGA 数据库中 20 个血管生成相关核心基因的 Lasso 系数剖面图(图 3A)和部分似然偏差图(图 3B),我们选择了 15 个关键 ARG 来构建预后模型: *EFNA1*、*ITGA5*、*EPHB4*、*NRP1*、*CDH5*、*PLAU*、*BMP6*、*DLL4*、*JUN*、*CA9*、*MMP1*、*BAIAP2L1*、*SERPINF1*、*F2RL1*、*FGFR2*,这 15 个模型基因对应的参数见表 1 和图 3C。通过 Lasso 回归分析得到每个样本对应的最佳风险评分值(risk score,  $S_r$ ),预

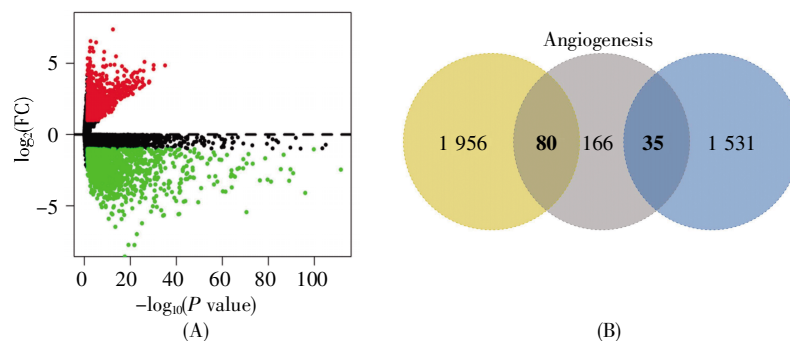


图 1 宫颈癌血管生成 DEGs 的分析

(A) 正常组与肿瘤组间 DEGs 的火山图。红点和绿点分别代表上调和下调的 DEGs; (B) 交叉基因的维恩图。黄色表示下调基因,蓝色表示上调基因,灰色相交部分表示重叠基因。

### Fig.1 Analysis of angiogenesis DEGs in CC

(A) Volcano map of DEGs between the normal and tumor groups. Red and green dots represent the upregulated and downregulated DEGs, respectively; (B) Venn diagram of intersection genes. Yellow indicates the downregulated genes, blue indicates the upregulated genes and gray intersecting parts indicate the genes that overlap.

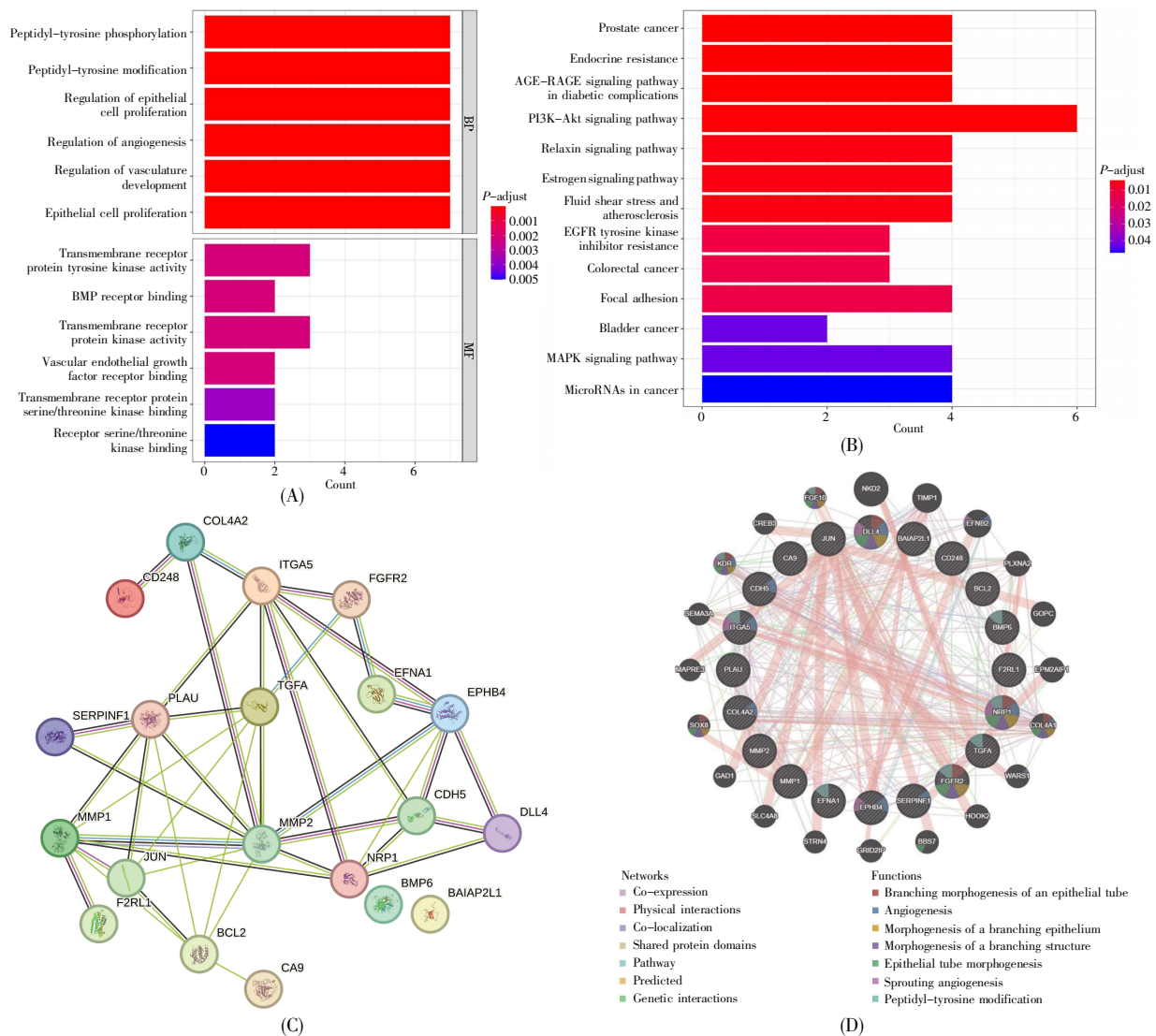


图2 宫颈癌预后基因集的富集分析

(A) 预后基因的GO富集分析; (B) 预后基因的KEGG通路富集分析; (C) 预后基因的蛋白质互作网络; (D) 预后基因与驱动基因的功能富集。

Fig.2 Enrichment analysis of the CC prognostic gene set

(A) GO enrichment analysis of prognostic genes; (B) KEGG pathway enrichment analysis of prognostic genes; (C) Protein-protein interaction network of prognostic genes; (D) Functional enrichment of prognostic genes and driving genes.

后模型为  $S_R = -0.188281925 \times S_{R-FGFR2} - 0.038250892 \times S_{R-SERPINF1} + 0.172882991 \times S_{R-EFNA1} + 0.033082229 \times S_{R-ITGA5} + 0.034745903 \times S_{R-EPHB4} + 0.017054977 \times S_{R-NRP1} + 0.126317951 \times S_{R-CDH5} + 0.092617820 \times S_{R-PLAU} + 0.129831454 \times S_{R-BMP6} + 0.065320131 \times S_{R-DLL4} + 0.075208882 \times S_{R-JUN} + 0.120662338 \times S_{R-CA9} + 0.094936232 \times S_{R-MMP1} + 0.157804210 \times S_{R-BAIAP2L1} + 0.024313692 \times S_{R-F2RL1}$ 。根据风险评分的中位数, 将宫颈癌患者分为高风险组和低风险组。高、低风险组患者的体细胞突变图谱(由 maftools 包绘制)显示, 高风险组患者 *SYNE1*、*HUWE1*、*RYSR2*、*LRP2* 和 *LRP1B* 等基因的突变频率显著高于低风险组

患者(15% vs. 12%、16% vs. 8%、14% vs. 11%、11% vs. 8%和 14% vs. 11%), 低风险组患者 *TTN*、*PIK-3CA*、*FBXW7* 和 *CSMD1* 等基因的突变频率显著高于高风险组患者(35% vs. 30%、33% vs. 26%、16% vs. 12%和 11% vs. 8%) (图 3D), 可见宫颈癌患者存在多种体细胞突变基因, 且高、低风险患者的体细胞突变基因之间存在显著性差异, 提示后续可以对该模型患者的体细胞突变情况进行深入研究。

#### 2.4 外部数据集验证预后模型的鲁棒性

下载 GEO 数据库中已处理的带有生存资料的 CESC 患者的数据(GSE52903、GSE44001), 根据模型预测 GEO 数据库中 CESC 患者的临床分型,

表 1 15 个血管生成预后模型基因的相关参数  
Table 1 Related parameters of 15 angiogenesis prognosis model genes

Gene	Coefficient	HR	Lower limit of CI	Upper limit of CI	P value
<i>FGFR2</i>	-0.188 281 925	0.692 253 912	0.480 296 205	0.997 749 876	0.048 604 807
<i>F2RL1</i>	0.024 313 692	1.239 101 227	1.003 917 551	1.529 380 425	0.045 893 328
<i>SERPINF1</i>	-0.038 250 892	0.674 516 630	0.461 816 758	0.985 180 109	0.041 628 772
<i>BAIAP2L1</i>	0.157 804 210	1.232 002 475	1.010 933 107	1.501 414 967	0.038 665 595
<i>MMP1</i>	0.094 936 232	1.215 371 197	1.040 846 987	1.419 158 787	0.013 657 302
<i>CA9</i>	0.120 662 338	1.288 881 287	1.075 720 436	1.544 281 316	0.005 936 119
<i>JUN</i>	0.075 208 882	1.326 614 545	1.085 573 722	1.621 176 081	0.005 735 695
<i>DLL4</i>	0.065 320 131	1.258 157 323	1.071 080 787	1.477 908 920	0.005 173 724
<i>BMP6</i>	0.129 831 454	1.228 642 394	1.063 783 157	1.419 050 604	0.005 093 065
<i>PLAU</i>	0.092 617 820	1.327 653 133	1.088 973 121	1.618 646 786	0.005 064 001
<i>CDH5</i>	0.126 317 951	1.249 885 681	1.076 251 610	1.451 532 524	0.003 467 708
<i>NRP1</i>	0.017 054 977	1.309 270 294	1.100 082 242	1.558 236 863	0.002 414 379
<i>EPHB4</i>	0.034 745 903	1.371 852 589	1.120 223 010	1.680 004 348	0.002 227 807
<i>ITGA5</i>	0.033 082 229	1.425 499 532	1.215 843 050	1.671 308 576	0.000 012 500
<i>EFNA1</i>	0.172 882 991	1.495 046 363	1.248 381 869	1.790 448 647	0.000 012 300

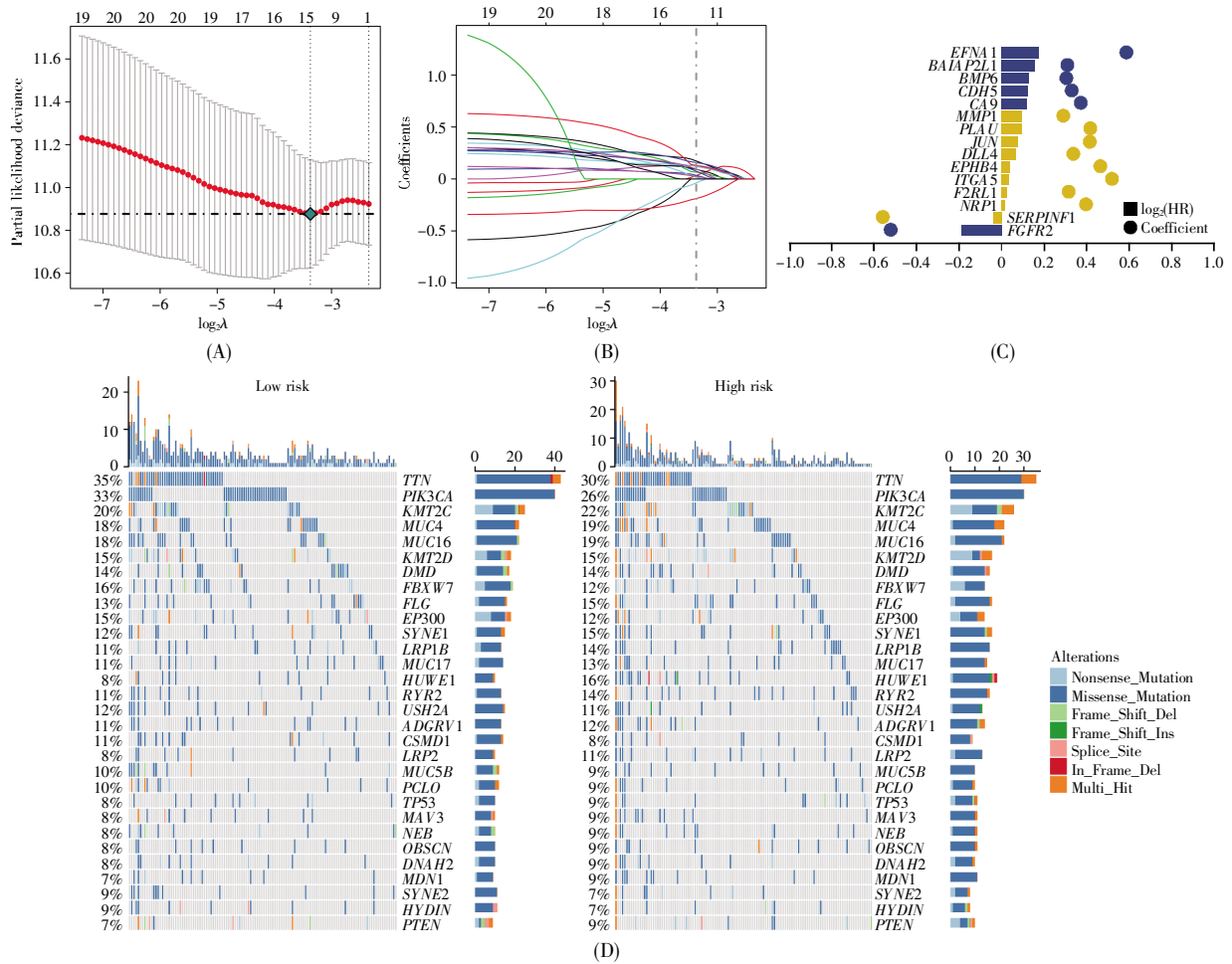


图 3 宫颈癌 ARG 和预后标志物的评估

(A) 筛选与最佳部分似然偏差值对应的最佳  $\lambda$  值, 左边的虚线表示 15 个最有价值的基因的阈值; (B) 通过最佳  $\lambda$  值得到各基因的最佳系数值; (C) 模型基因系数图; (D) 高风险组和低风险组患者的体细胞突变图谱。

Fig.3 Evaluation of ARGs and prognostic markers of CC

(A) Filtration of the best  $\lambda$  value corresponding to the best partial likelihood deviation value, the dashed line on the left represents the threshold for the 15 most valuable genes; (B) The best coefficient value corresponding to each gene obtained through the best  $\lambda$  value; (C) Model gene coefficient diagram; (D) Somatic mutation map of patients in the high-risk and low-risk groups.

通过 Kaplan-Meier 分析评估两组间的生存差异, 探讨预测模型的稳定性。结果显示, 在两个 GEO 外部验证数据集中, 高风险组的 OS ( $P=0.005$ ) 和无病生存期 (disease-free survival, DFS) ( $P<0.001$ ) 均显著低于低风险组 (图 4A~4B)。为了验证模型的准确性, 我们使用外部验证数据集对模型做 ROC 曲线分析。结果显示, GSE52903 验证集在 1 年、3 年和 5 年的曲线下面积 (area under the curve, AUC) 值分别为 0.84、0.77 和 0.73, C-指数为 0.72 (图 4C); GSE44001 验证集在 1 年、3 年和 5 年的 AUC 值分别为 0.71、0.72 和 0.70, C-指数为 0.70 (图 4D), 说明该模型对宫颈癌患者的预后具有很强的预测效能。

## 2.5 发病风险及独立预后分析

根据风险评分的中位数将宫颈癌患者分为高风险和低风险两组, 回归分析的结果以列线图的

形式呈现, 结果显示宫颈癌患者的多个临床指标和风险评估在打分过程中均有分布 (图 5A)。进一步的校准曲线分析显示, 该模型对宫颈癌患者 5 年和 7 年 OS 的预测情况接近于真实情况 (图 5B)。此外, 单因素和多因素 Cox 回归分析结果显示, 通过模型分析得出的风险评估是宫颈癌患者的独立预后因素 ( $P<0.001$ ) (表 2, 图 5C)。

## 2.6 预后模型基因富集的相关信号通路

为了探究风险评估影响肿瘤进展的潜在分子机制, 我们对预后模型基因进行了 GSEA 分析。结果显示, 预后模型基因在许多通路都有显著富集: GO 分析中富集到的通路有细胞外基质 (extracellular matrix, ECM) 降解、胶原纤维组织、核苷酸切除修复和血液凝固的负反馈等 (图 6A); KEGG 分析中富集到的通路有 DNA 复制、ECM 受体相互作用、补体和凝血级联等 (图 6B), 提示这些信号通路的扰

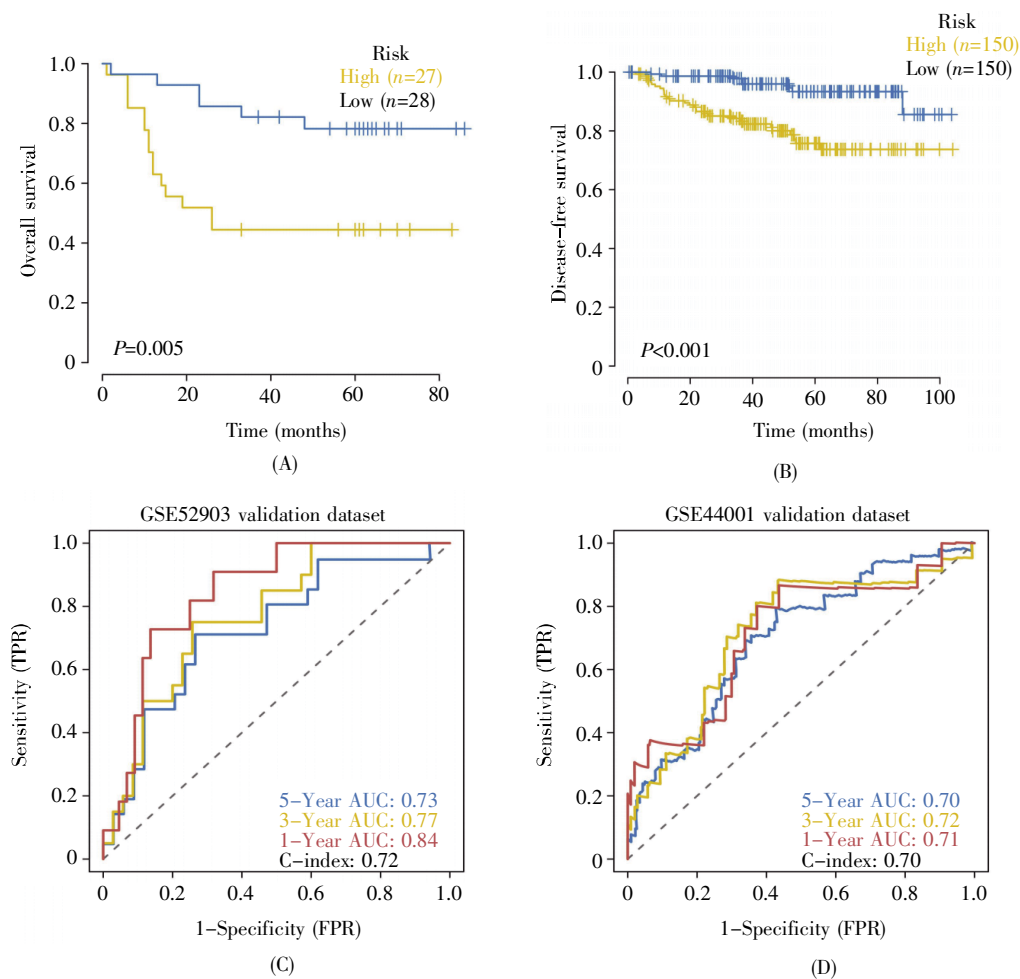


图 4 模型有效性的验证

(A~B) 两个 GEO 外部数据集的 Kaplan-Meier 生存曲线; (C~D) 两个 GEO 外部数据集的 1 年、3 年和 5 年生存率的 ROC 曲线。

### Fig.4 Validation of the model

(A~B) Kaplan-Meier survival curves in the two GEO external datasets; (C~D) ROC curves for 1-, 3-, and 5-year survival rates in the two GEO external datasets.

表 2 基于单因素和多因素 Cox 回归分析的宫颈癌患者临床特征与风险评分  
**Table 2 Clinical characteristics and risk scores of CC patients based on univariate and multivariate Cox regression analysis**

Type	HR	HR.95L	HR.95H	P value
<b>Univariate Cox regression analysis</b>				
Age	1.011 356 236	0.966 925 975	1.057 828 068	0.622 263 141
Grade	1.281 313 013	0.621 238 269	2.642 726 824	0.502 139 531
Tumor	1.508 045 183	0.758 441 729	2.998 516 810	0.241 394 934
Node	2.043 749 412	0.774 336 857	5.394 179 056	0.148 882 975
Risk score	5.627 643 472	2.330 672 267	13.588 513 280	0.000 122 399
<b>Multivariate Cox regression analysis</b>				
Age	0.998 823 279	0.947 327 719	1.053 118 075	0.965 225 978
Grade	0.665 657 284	0.291 810 427	1.518 450 266	0.333 416 547
Tumor	1.515 388 634	0.796 010 652	2.884 889 425	0.205 717 421
Node	3.086 682 825	1.031 512 299	9.236 546 059	0.043 858 171
Risk score	8.496 779 744	2.851 767 593	25.315 971 120	0.000 122 395

注: HR.95L 表示风险比的 95% 置信区间下限; HR.95H 表示风险比的 95% 置信区间上限。

Notes: HR.95L, lower limit of 95% confidence interval for risk ratio; HR.95H, higher limit of 95% confidence interval for risk ratio.

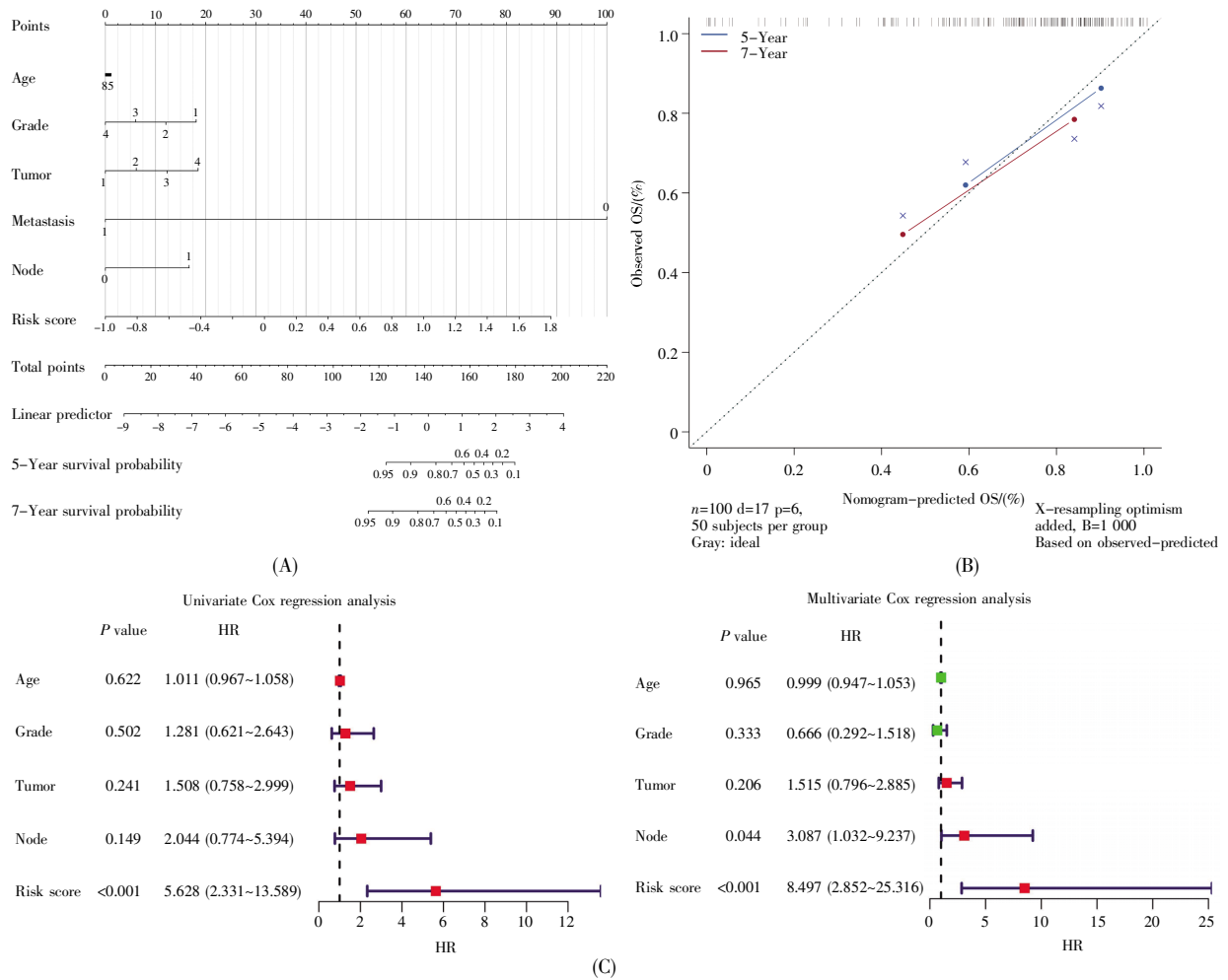


图 5 宫颈癌患者预后的列线图 and 临床预后分析

(A) 预后列线图; (B) 预后模型的校准曲线(5 年和 7 年); (C) 基于单因素和多因素 Cox 回归分析的宫颈癌预后特征森林图。

**Fig.5 A nomogram and clinical prognostic analysis of CC patients**

(A) Prognostic nomogram; (B) The correlation curve of the ARG model (5-year and 7-year); (C) Forest map of CC prognostic characteristics analyzed by both univariate and multivariate Cox regression.

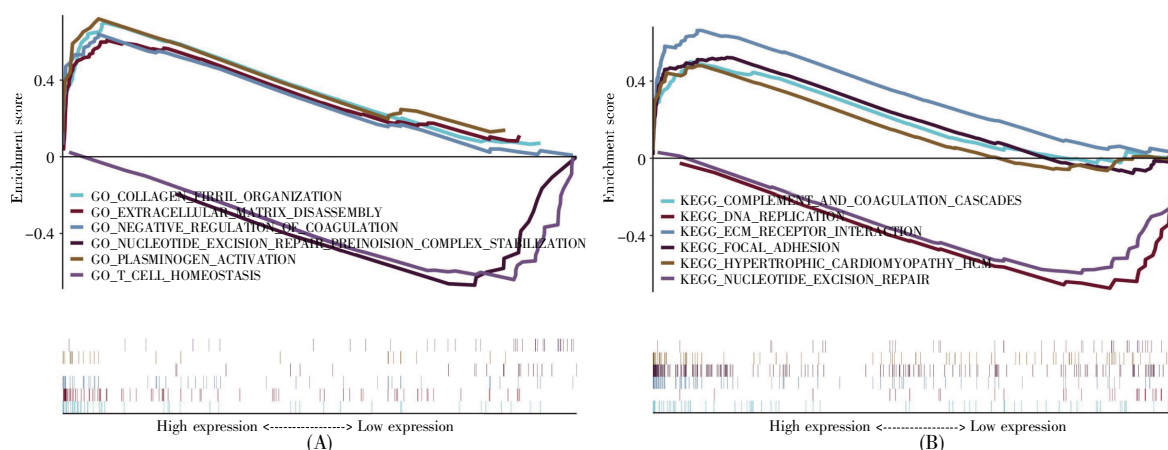


图6 预后模型基因的GSEA富集分析

(A) 通过GSEA进行的GO分析; (B) 通过GSEA进行的KEGG分析。

Fig.6 GSEA of the prognosis-related ARG

(A) GO analysis via GSEA; (B) KEGG analysis via GSEA.

动会影响宫颈癌患者的预后。

### 3 讨论

在过去的几十年里,发展中国家宫颈癌的发病率仍然较高,患者预后较差<sup>[1]</sup>。研究表明,血管生成在宫颈癌的发生和进展中起着关键作用<sup>[13]</sup>。因此,识别新的可靠的血管生成相关生物标志物并用其来预测宫颈癌患者的预后非常有必要。本研究构建了一个宫颈癌ARG的预后模型,以供正在进行的研究使用。首先,我们通过TCGA和GeneCards数据库提取了与血管生成相关的DEG。然后,通过Lasso Cox回归分析( $P < 0.01$ )筛选了20个预后相关基因。最后,基于这20个基因的Lasso系数剖面图和部分似然偏差图结果,我们选择了15个关键ARG构建预后模型。此外,本文通过STRING数据库分析找到连接度较前且属于预后模型基因集的5个关键基因: *CDH5*、*EPHB4*、*PLAU*、*ITGA5*、*SERPINF1*, 研究显示这些基因与宫颈癌发生发展密切相关。

钙黏着蛋白5 (cadherin 5, *CDH5*)是经典II型钙黏着蛋白中的一员,在细胞间黏附、接触抑制生长、抑制内皮细胞迁移和凋亡中起重要作用<sup>[14-15]</sup>。张星海等<sup>[16]</sup>的研究显示, *CDH5*可促进微血管生成,为肿瘤细胞提供适合的支持微环境,并能促进肿瘤细胞增殖和转移。酪氨酸激酶 *EPHB4* 是Eph受体家族的一员,与肿瘤血管生成、生长和转移有关<sup>[17-18]</sup>。研究显示, *EPHB4*的表达程度及mRNA水平与宫颈癌患者临床分级、肿瘤大小、淋巴结转移显著正相关,高表达 *EPHB4* 的宫颈癌患者生

存率低下<sup>[19]</sup>。尿激酶型纤溶酶原激活物(U-plasminogen activator, *PLAU*)属于S1丝氨酸蛋白酶,它可以水解ECM重构相关蛋白质和激活生长因子<sup>[20]</sup>,其表达水平与肿瘤细胞淋巴结和远处器官转移显著相关<sup>[9, 21]</sup>。Liu等<sup>[22]</sup>研究了与宫颈癌缺氧和血管生成相关的基因: *MOCSI*、*PPP1R14A*、*ESM1*、*DES*、*ITGA5*、*SERPINF1*; 通过与本研究进行比对,我们发现整合素亚基5 (*integrin alpha 5*, *ITGA5*)和丝氨酸蛋白酶抑制剂 *F1* (*serpin family F member 1*, *SERPINF1*)在这两项研究中都有涉及,推测 *ITGA5* 和 *SERPINF1* 在宫颈癌血管生成中有着重要作用。*ITGA5* 主要参与细胞表面黏附和信号转导过程<sup>[23-24]</sup>,被证明在膀胱癌<sup>[25]</sup>和胶质瘤<sup>[26]</sup>中具有预后作用。董晶等<sup>[27]</sup>发现,靶向 *ITGA5* 的干扰小RNA (small interfering RNA, siRNA) 敲低内源性 *ITGA5* 的表达后,宫颈癌细胞系的增殖能力显著降低,凋亡水平增加。另有文献报道, *ITGA5* 在肿瘤耐药、上皮间质转化、血管生成拟态和淋巴管生成等方面具有关键作用<sup>[23, 28]</sup>。*SERPINF1* 又称色素上皮衍生因子 (pigment epithelium-derived factor, PEDF), 是一种蛋白酶抑制剂,可分泌相关蛋白质并强烈抑制血管生成<sup>[29]</sup>。彭钧等<sup>[30]</sup>的研究显示,宫颈癌组织 *SERPINF1* 蛋白阳性表达率低于正常宫颈组织,且 *SERPINF1* 高表达的宫颈癌患者的5年生存率高于 *SERPINF1* 低表达的宫颈癌患者。另有研究显示, *SERPINF1* 的高表达与肿瘤相关的成纤维细胞和内皮细胞的浸润能力呈负相关,其在卵巢癌、乳腺癌、前列腺癌及胰腺癌等肿瘤中发挥主要抑癌作用<sup>[31-32]</sup>,因此,研发能够上调 *SERPINF1* 表

达的磷酸铂化合物可为癌症化疗提供更多的选择。从以上分析可知, *CDH5*、*EPHB4*、*PLAU*、*ITGA5* 和 *SERPINF1* 与肿瘤预后及血管生成紧密相关, 它们在宫颈癌发生发展中的作用及机制值得深入探究。

根据 GSEA 分析正常化富集评分(normalized enrichment score, NES)的正负性结果, 核苷酸切除修复、ECM 受体相互作用、DNA 复制和 T 细胞稳态等途径与宫颈癌患者预后良好有关, 胶原纤维组织、ECM 降解、凝血负调节、纤溶酶原激活途径、补体和黏着斑等途径与患者预后不良有关。核苷酸切除修复是一种主要的 DNA 修复方法, 它可以消除多种环境诱变剂产生的 DNA 损伤, 包括紫外线照射和有毒化合物等造成的损伤<sup>[33]</sup>。浸润和转移是引起宫颈癌难治的重要原因<sup>[34]</sup>, 而 ECM 的降解在浸润、转移中具有重要作用<sup>[35]</sup>。人体的补体及凝血级联系统由肝脏产生凝血物质与血管内皮维持, 其过程包括凝血、抗凝、纤溶、补体激活<sup>[36-37]</sup>, 这些过程与肿瘤发生、发展紧密相关。

综上所述, 本研究通过模型的建立和验证, 筛选了 15 个与宫颈癌血管生成相关的预后基因: *EFNA1*、*ITGA5*、*EPHB4*、*NRP1*、*CDH5*、*PLAU*、*BM-P6*、*DLL4*、*JUN*、*CA9*、*MMP1*、*BAIAP2L1*、*SERPINF1*、*F2RL1*、*FGFR2*, 并评估了与这些基因相关的功能和通路; 该预测模型可有效地预测宫颈癌患者的预后, 这将为进一步探索宫颈癌发生发展机制及寻找新型分子标志物提供依据。

## 参考文献(References):

- [1] SUNG H, FERLAY J, SIEGEL R L, *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. *CA: A Cancer Journal for Clinicians*, 2021, 71(3): 209–249.
- [2] LIU X J, SONG J W, ZHANG Y N, *et al.* ASF1B promotes cervical cancer progression through stabilization of CDK9[J]. *Cell Death & Disease*, 2020, 11(8): 705.
- [3] LI M, ZHANG F R, SU Y J, *et al.* Nanoparticles designed to regulate tumor microenvironment for cancer therapy[J]. *Life Sciences*, 2018, 201: 37–44.
- [4] VIALARD C, LARRIVÉE B. Tumor angiogenesis and vascular normalization: alternative therapeutic targets[J]. *Angiogenesis*, 2017, 20 (4): 409–426.
- [5] FLEGG J A, MENON S N, BYRNE H M, *et al.* A current perspective on wound healing and tumour-induced angiogenesis[J]. *Bulletin of Mathematical Biology*, 2020, 82(2): 23.
- [6] HU H Y, CHEN Y, TAN S T, *et al.* The research progress of antiangiogenic therapy, immune therapy and tumor microenvironment[J]. *Frontiers in Immunology*, 2022, 13: 802846.
- [7] GAO S, ZHAO Z Y, ZHANG Z Y, *et al.* Prognostic value of microRNAs in esophageal carcinoma: a meta-analysis[J]. *Clinical and Translational Gastroenterology*, 2018, 9(11): 203.
- [8] RITCHIE M E, PHIPSON B, WU D, *et al.* *limma* powers differential expression analyses for RNA-seq and microarray studies[J]. *Nucleic Acids Research*, 2015, 43(7): e47.
- [9] DING H, FAN G L, YI Y X, *et al.* Prognostic implications of immune-related genes' (IRGs) signature models in cervical cancer and endometrial cancer[J]. *Frontiers in Genetics*, 2020, 11: 725.
- [10] HE Z K, WANG X J, YANG Z M, *et al.* Expression and prognosis of CDC45 in cervical cancer based on the GEO database[J]. *PeerJ*, 2021, 9: e12114.
- [11] REBHAN M, CHALIFA-CASPI V, PRILUSKY J, *et al.* GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support[J]. *Bioinformatics*, 1998, 14(8): 656–664.
- [12] YU G C, WANG L G, HAN Y Y, *et al.* clusterProfiler: an R package for comparing biological themes among gene clusters[J]. *OMICS: A Journal of Integrative Biology*, 2012, 16(5): 284–287.
- [13] 中国抗癌协会宫颈癌专业委员会. 子宫颈癌抗血管生成药物临床应用指南(2023 年版)[J]. *中国实用妇科与产科杂志* (Cervical Cancer Committee of China Anti-Cancer Association. Clinical application guidelines for anti angiogenic drugs in cervical cancer (2023 edition)[J]. *Chinese Journal of Practical Gynecology and Obstetrics*), 2023, 39(12): 1201–1209.
- [14] MICHELINI S, RICCI M, AMATO B, *et al.* *CDH5*, a possible new candidate gene for genetic testing of lymphedema[J]. *Lymphatic Research and Biology*, 2022, 20(5): 496–506.
- [15] YÜCEL E I, SAHIN M. Fenretinide reduces angiogenesis by downregulating *CDH5*, *FOXM1* and *eNOS* genes and suppressing microRNA-10b[J]. *Molecular Biology Reports*, 2020, 47(3): 1649–1658.
- [16] 张星海, 方天逸, 边秀兰, 等. *CDH5* 在恶性肿瘤中的研究进展[J]. *国际免疫学杂志* (ZHANG Xinghai, FANG Tianyi, BIAN Xiulan, *et al.* Research progress of *CDH5* in malignant tumors[J]. *International Journal of Immunology*), 2023, 46(1): 66–70.
- [17] CHEN D, HUGHES E D, SAUNDERS T L, *et al.* Angiogenesis depends upon *EPHB4*-mediated export of collagen IV from vascular endothelial cells[J]. *JCI Insight*, 2022, 7(4): e156928.
- [18] PAN X Y, LIANG L Y, SI R, *et al.* Discovery of novel anti-angiogenesis agents. Part 10: multi-target inhibitors of VEGFR-2, Tie-2 and EphB4 incorporated with 1,2,3-triazol[J]. *European Journal of Medicinal Chemistry*, 2019, 163: 1–9.
- [19] ALAM S M, FUJIMOTO J, JAHAN I, *et al.* Coexpression of EphB4 and ephrinB2 in tumor advancement of uterine cervical cancers[J]. *Gynecologic Oncology*, 2009, 114(1): 84–88.
- [20] MINAEI E, MUELLER S A, ASHFORD B, *et al.* Cancer progression gene expression profiling identifies the urokinase plasminogen activator receptor as a biomarker of metastasis in cutaneous squamous cell carcinoma[J]. *Frontiers in Oncology*, 2022, 12: 835929.
- [21] WU M, WEI B, DUAN S L, *et al.* Methylation-driven gene *PLAU* as a potential prognostic marker for differential thyroid carcinoma[J]. *Frontiers in Cell and Developmental Biology*, 2022, 10: 819484.
- [22] LIU L L, ZHU H C, WANG P, *et al.* Construction of a six-gene prognostic risk model related to hypoxia and angiogenesis for cervical cancer[J]. *Frontiers in Genetics*, 2022, 13: 923263.
- [23] LI X Q, ZHANG R, LU H, *et al.* Extracellular vesicle-packaged *CDH11* and *ITGA5* induce the premetastatic niche for bone colonization of breast cancer cells[J]. *Cancer Research*, 2022, 82(8): 1560–1574.
- [24] LIU G L, XU X T, GENG H, *et al.* FGA inhibits metastases and induces autophagic cell death in gastric cancer via inhibiting *ITGA5* to regulate the FAK/ERK pathway[J]. *Tissue and Cell*, 2022, 76: 101767.
- [25] XIONG Y Y, SONG X H, KUDUSI, *et al.* Oncogenic *GBX2* promotes the malignant behaviors of bladder cancer cells by binding to the *ITGA5* promoter and activating its transcription[J]. *Functional & Integrative Genomics*, 2022, 22(5): 937–950.

- [26] LI S Y, ZHANG N, LIU S Y, *et al.* ITGA5 is a novel oncogenic biomarker and correlates with tumor immune microenvironment in gliomas[J]. *Frontiers in Oncology*, 2022, 12: 844144.
- [27] 董晶, 商双, 谢锋. RNA 干扰 ITGA5 的表达对宫颈癌细胞生长和侵袭的影响[J]. *复旦学报(医学版)* (DONG Jing, SHANG Shuang, XIE Feng. Effect of RNA interference with ITGA5 expression on cervical cancer cell growth and invasion[J]. *Fudan University Journal of Medical Sciences*), 2023, 50(3): 330-336.
- [28] SHI Y, WU M W, LIU Y Y, *et al.* ITGA5 predicts dual-drug resistance to temozolomide and bevacizumab in glioma[J]. *Frontiers in Oncology*, 2021, 11: 769592.
- [29] XU B, LI J, LIU X X, *et al.* TXNDC5 is a cervical tumor susceptibility gene that stimulates cell migration, vasculogenic mimicry and angiogenesis by down-regulating SERPINF1 and TRAF1 expression[J]. *Oncotarget*, 2017, 8(53): 91009-91024.
- [30] 彭钧, 鲍方, 李小月, 等. MDR1、PEDF、LRP 在宫颈癌中的表达及意义[J]. *中国现代医学杂志* (PENG Jun, BAO Fang, LI Xiaoyue, *et al.* Expression of MDR1, PEDF, and LRP in cervical cancer and its clinical significance[J]. *China Journal of Modern Medicine*), 2021, 31(13): 11-16.
- [31] BELKACEMI L, ZHANG S X. Anti-tumor effects of pigment epithelium-derived factor (PEDF): implication for cancer therapy. A mini-review[J]. *Journal of Experimental & Clinical Cancer Research*, 2016, 35: 4.
- [32] ANSARI D, ALTHINI C, OHLSSON H, *et al.* The role of PEDF in pancreatic cancer[J]. *Anticancer Research*, 2019, 39(7): 3311-3315.
- [33] ZHANG X P, YIN M D, HU J C. Nucleotide excision repair: a versatile and smart toolkit[J]. *Acta Biochimica et Biophysica Sinica*, 2022, 54(6): 807-819.
- [34] VINK J, YU V, DAHAL S, *et al.* Extracellular matrix rigidity modulates human cervical smooth muscle contractility—new insights into premature cervical failure and spontaneous preterm birth[J]. *Reproductive Sciences*, 2021, 28(1): 237-251.
- [35] ZHUANG Y, HUANG Y H, HE Z J, *et al.* Effect of substrate stiffness on the mechanical properties of cervical cancer cells[J]. *Archives of Biochemistry and Biophysics*, 2022, 725: 109281.
- [36] ZHANG J, CHEN M S, ZHAO Y Y, *et al.* Complement and coagulation cascades pathway correlates with chemosensitivity and overall survival in patients with soft tissue sarcoma[J]. *European Journal of Pharmacology*, 2020, 879: 173121.
- [37] BAUER A T, GORZELANNY C, GEBHARDT C, *et al.* Interplay between coagulation and inflammation in cancer: limitations and therapeutic opportunities[J]. *Cancer Treatment Reviews*, 2022, 102: 102322.

## 《生命科学研究》生态学专栏征稿启事

### 一、刊登范围

1. 群落结构及生物多样性的资源调查与维持或保护机制研究;
2. 基于生物有机体开展的生态系统空间分布格局与动态变化机制研究;
3. 基于动物、植物、微生物开展的生态系统对全球变化的响应与适应机制研究;
4. 生物地球化学循环与生物环境控制机制研究;
5. 生物防治和生态毒理方面的基础与应用研究;
6. 生物资源的开发与利用;
7. 生物资源在生态环境建设中的应用;
8. 聚焦生物生态学的各种技术手段的开发与应用;等等。

研究内容可从以上几方面考虑(但不局限于此)。

### 二、来稿要求及权益

1. 来稿均须原创,严禁抄袭,字数控制在 6 000 字左右,要求层次清晰,逻辑严谨,行文流畅,数据可靠,图片无版权纠纷,观点鲜明,创新性强,无政治性差错。
2. 来稿须使用规范汉字,严格按本刊格式要求规范撰写。
3. 来稿免收审稿费,通过本刊外审和终审的稿件,电子版于 15 个工作日内在“中国知网”网络首发平台全文出版,纸质版论文在 4 个月内优先发表。
4. 正式见刊的稿件均由本刊负责付费在同方知网平台上进行大力推广,增加曝光率和引用率。

### 三、联系方式

投稿网址: <http://smkx.hunnu.edu.cn>

E-mail: [smkxyj@hunnu.edu.cn](mailto:smkxyj@hunnu.edu.cn); [life@hunnu.edu.cn](mailto:life@hunnu.edu.cn)

地 址:长沙市湖南师范大学生命科学学院 1 号楼 105 房《生命科学研究》编辑部,邮编:410081

电 话:0731-88872616