

基于宫颈癌预后模型及免疫浸润分析寻找宫颈癌预后标志物的研究

王辰林¹ 姜洪苗² 伍思敏¹ 李戡文¹ 黄培康¹ 方春丽^{2*}

1. 南京医科大学 江苏南京 211166

2. 南京医科大学第二附属医院 江苏南京 210000

摘要：目的 通过构建宫颈癌预后模型，筛选预后表达差异基因，并进行肿瘤微环境相关免疫分析，寻找宫颈癌预后相关标志物。方法 利用 TCGA 及 GEO 公共数据库进行宫颈癌预后基因的筛选、建立模型并进行验证，通过免疫浸润分析研究预后表达差异基因与肿瘤免疫微环境的相互关系。结果 SPP1 作为宫颈癌预后表达基因之一，其高表达与较差的宫颈癌预后相关，在肿瘤免疫微环境中，与 M0 型巨噬细胞的浸润相关；IL3RA 的表达则提示了较好的宫颈癌预后，其在肿瘤免疫微环境中的表达与 CD8+T 细胞的浸润相关。结论 寻找宫颈癌预后相关标志物，并研究其在宫颈癌发生发展中的作用，对个体化精准治疗宫颈癌意义重大。

关键词：宫颈癌预后模型；巨噬细胞；CD8+T 细胞；SPP1；IL3RA

宫颈癌是常见的女性恶性肿瘤之一，2018 年新发宫颈癌超过 50 万例，致死超 30 万例^[1]。目前手术和放化疗仍然是主要治疗方式，对于晚期及复发转移宫颈癌的治疗效果并不理想。免疫和靶向治疗为晚期及复发转移患者带来了希望，如何筛选获益人群、制定最优的治疗策略等问题仍需不断探索^[2]。寻找宫颈癌预后分子标志物，对制定临床决策及改善预后意义重大。本研究利用 TCGA 数据库，通过生物信息学方法建立宫颈癌预后模型，并在基因层面探讨其对肿瘤微环境中免疫细胞浸润的影响，为宫颈癌个体化诊治方向提供研究思路。

1. 材料与方法

1.1 宫颈癌预后模型及差异表达基因分析

从 TCGA 数据库获取宫颈癌的基因表达谱数据，使用单因素 cox 回归分析识别宫颈癌预后相关的基因，使用 lasso 分析和多因素 cox 回归分析构建预后模型，并验证模型与宫颈癌预后相关程度。从 GEO 数据库下载宫颈癌的基因表达谱数据 GSE7803 以及 GSE63514，物种为人类，平台为 GPL96。将其分成肿瘤组织正常组织。使用 R 包 limma 进行差异分析。分析模型中差异表达基因的 ROC 曲线，以及在临床分析 III 和 IV 期分组情况下的亚组 KM 曲线，使用 Log-

rank 检验进行统计学分析。使用 TCGA 数据库中宫颈癌患者的生存数据和临床分期等数据进行危险因素评分，并绘制 KM 曲线和 nomogram 列线图评价危险评分效能。使用 HPA (Human Protein Atlas) 数据库分析模型中基因在宫颈癌患者组织中的表达情况。

1.2 宫颈癌相关预后基因的免疫浸润分析

通过 estimate 包对 TCGA- 宫颈癌预后模型数据集的表达数据计算免疫及机制评分，进而评估肿瘤的纯度，并探究模型风险高低分组下的免疫评分，基质评分，ESTIMA 评分的差异。通过 GSVA 包中单样本基因集富集分析 (single sample gene set enrichment analysis, ssGSEA) 样本中每个免疫细胞类型的浸润水平。

通过 CIBERSORT (<https://cibersort.stanford.edu/>) 算法对 TCGA- 宫颈癌数据集的免疫细胞浸润状态进行评价；我们通过 Spearman 分析预后相关基因与免疫浸润细胞的相关性，并使用 R 包 “ggplot2” 绘制热图和棒棒糖图进行可视化。

1.3 统计学分析

所有数据计算和统计均使用 R 编程 (4.0.2 版本) 进行。两组连续变量的比较，通过独立 Student t 检验统计显著性，并通过 Mann-Whitney U 检验 (即 Wilcoxon 秩和检验) 分析

非正态分布变量间的差异。使用 R 的 pROC 包绘制工作特征 (ROC) 曲线, 计算曲线下的面积 (AUC) 以评估风险评分估计预后的准确性。所有的统计 P 值均为双侧, 以 $P < 0.05$ 为有统计学意义。

2. 结果

2.1 构建宫颈癌预后模型及获取差异表达基因

对清洗好的 TCGA 数据集进行 LASSO 回归 (图 1A-C), 并使用 cox 多因素分析筛选宫颈癌相关预后标志物。将结果绘制了森林图进行展示 (图 1D), 结果显示大多数基因的预后效能较高。分别选取了 $HR > 1$ 和 $HR < 1$ 的基因绘制了生存曲线 (图 2)。检测模型基因结果以箱式图的形式展示 (图 3A-D), SPP1 作为宫颈癌患者预后的风险因素, 在肿瘤组织中高表达; 而 DES 作为宫颈癌患者预后的保护因素, 在肿瘤组织中低表达。

使用 “ConsensusClusterPlus” R 包利用上述预后模型进行一致性聚类, 预后亚型进行分组绘制热图 (图 3E), 结果显示出来基因表达差异趋势。使用 limma 包差异分析得到两组数据的差异表达基因, 筛选表达差异的基因。

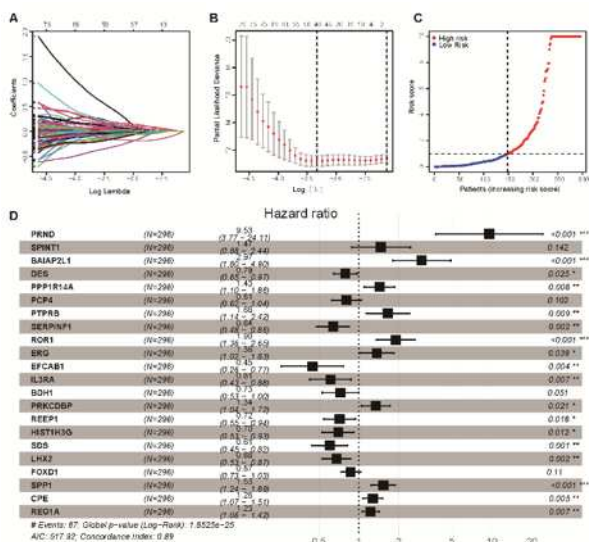


图 1 构建宫颈癌预后模型

A-C: 对 TCGA 数据集进行 lasso 回归 + cox 多因素回归分析, lasso 回归 (A), 最佳 λ 值 (B) 以及风险曲线 (C)。D: 森林图展示得到的预后模型基因。

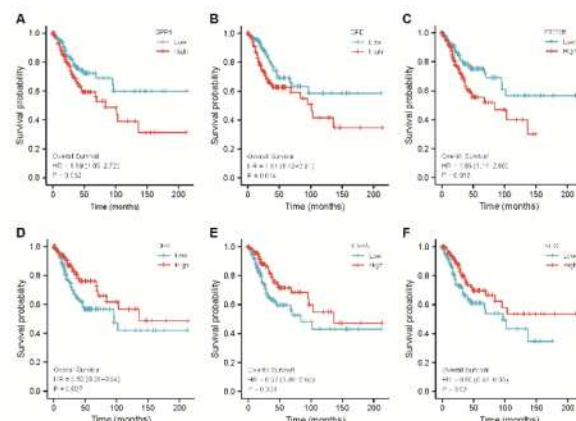


图 2 预后相关基因的生存分析

A: SPP1 的 KM 生存曲线。B: CPE 的 KM 生存曲线。C: PTPRB 的 KM 生存曲线。D: DES 的 KM 生存曲线。E: IL3RA 的 KM 生存曲线。F: SDS 的 KM 生存曲线。

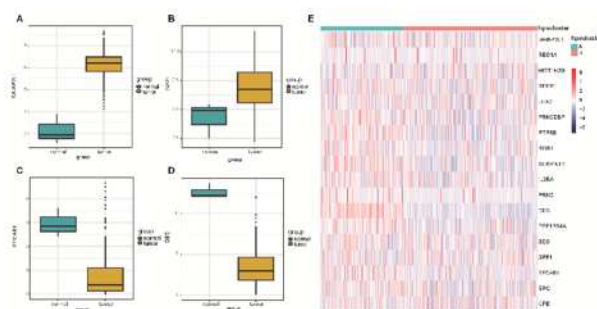


图 3 宫颈癌亚型构建

A-D: BAIAP2L1(A)、SPP1(B)、DES(C) 和 EFCAB1(D) 在肿瘤组与对照组的表达差异箱式图。E: 根据宫颈癌亚型分组绘制的热图。热图中红色表示激活, 蓝色表示抑制。

2.2 宫颈癌预后模型的临床相关性

进一步评估了预后模型对于宫颈癌的区分识别程度, 绘制了 ROC 曲线 (图 4A-C), 绘制了亚组生存曲线, 选择临床分期 III 期或者 IV 期进行亚组分析 (图 4D-F), 结果表明保护性因素的 IL3RA 在亚组分析中仍然作为宫颈癌的保护性因素, 高表达 IL3RA 患者明显比低表达该基因的患者总体生存时间更长。而危险性因素 SPP1 也显示出了高表达该基因的患者总体生存时间明显较短。

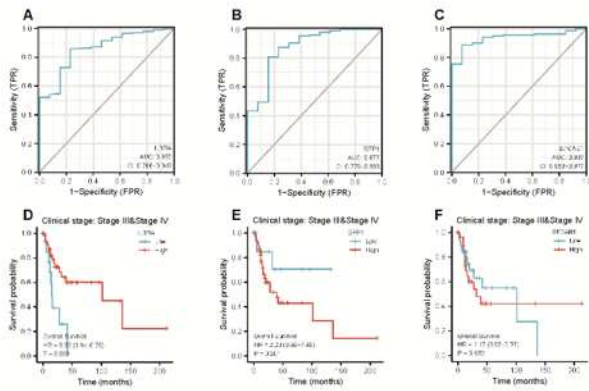


图4 预后模型的临床相关性

A-C: IL3RA(A)、SPP1(B) 和 EFCAB1(C) 的 ROC 曲线。

D-F: IL3RA(D)、SPP1(E) 和 EFCAB1(F) 的亚组 KM 总体生存曲线。

2.3 预后危险评分的预后相关分析

接下来纳入了总体生存时间、年龄、临床分期等临床因素进行危险评分分析, 使用 lasso+cox 回归分析计算危险评分, 并根据危险评分绘制 KM 生存曲线 (图 5A), 结果显示该危险评分能够很好地区分宫颈癌患者的预后情况。对该危险评分绘制了 nomogram (列线图, 图 5B), 发现风险评分能够很好地对患者预后情况进行判断, 并且效能远远超过临床分期的作用。calibration 图 (图 5C-E) 显示 1 年、3 年、5 年的预测情况都与真实值十分贴近。

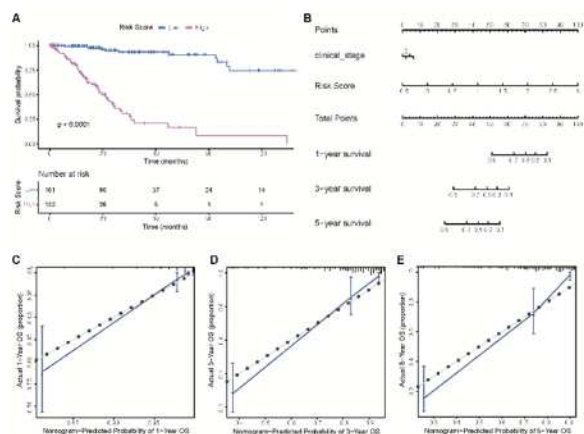


图5 预后危险评分的预后相关分析

A: 危险评分的总体生存曲线。B: 危险评分相关的列线图。C-E: 1 年 (C)、3 年 (D)、5 年 (E) 的 calibration 图。

2.4 预后相关基因的免疫组化结果

在 HPA 数据库中分别分析了 SPP1 和 IL3RA 在正常宫颈组织和宫颈癌组织中的免疫组化结果 (图 6)。SPP1 在

宫颈癌组织中的表达明显多于正常组织, 而 IL3RA 在宫颈癌和正常组织中均有较高程度的表达。

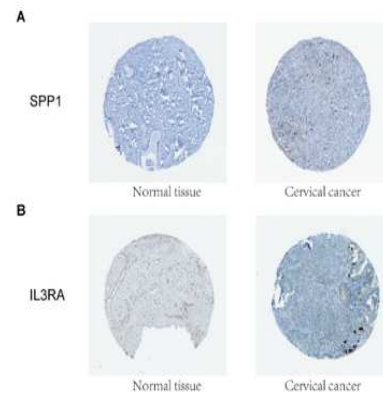


图6 预后相关基因的免疫组化结果

A: SPP1 在正常宫颈组织与宫颈癌组织中免疫组化结果。

B: IL3RA 在正常宫颈组织与宫颈癌组织中免疫组化结果。

2.4.1 TCGA- 宫颈癌预后模型数据集的免疫评分

结合 TCGA 宫颈癌预后模型数据集高低风险的分组, 我们将 Stromal Score, Immune Score, ESTIMATE Score 评分结果通过分组比较图展示出来 (图 7A-C), 结果显示, TCGA- 宫颈癌数据集的 Immune Score 评分 (图 7B) 与 TCGA- 宫颈癌数据集样本的不同分组间存在统计学显著意义 ($P < 0.05$)。

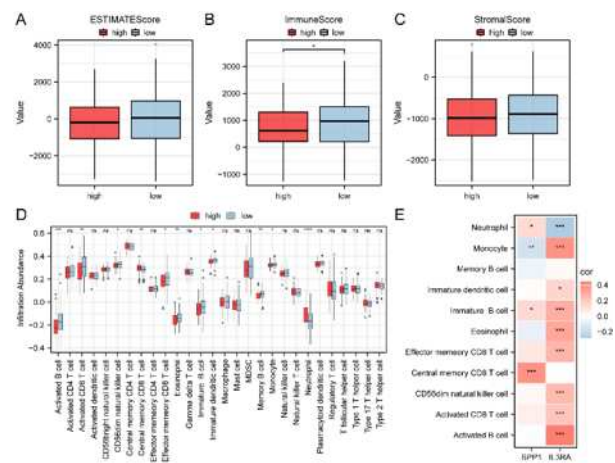


图7 免疫评分及免疫浸润分析

A. low 组、high 组的 ESTIMATE Score 评分 (A), Immune Score 评分 (B), Stromal Score 评分 (C) 分组比较图。D. 免疫细胞在 low 组, high 组的分组比较图。E. 免疫细胞与 hub genes 的相关性热图。

2.4.2 免疫浸润分析

使用 ssGSEA 算法计算 28 种免疫细胞与数据集里高低风险的分组表达谱数据之间的相关性。根据免疫细胞的浸润程度差异绘制分组比较图(图 7D)，分析免疫细胞与预后基因间的相关性。结果显示，共有 11 种免疫细胞存在显著浸润差异。

绘制免疫细胞在样本中的浸润丰度堆积柱状图(图 8A)。然后根据免疫细胞的浸润丰度分析免疫细胞在 high 和 low 组的差异(图 8B)，其中免疫细胞 T cells CD8, T cells CD4 naive, T cells CD4 memory activated, Monocytes, Macrophages M0, Dendritic cells resting, Dendritic cells activated, Mast cells resting, Mast cells activated, Neutrophils 存在显著浸润差异。我们绘制这 10 个免疫细胞的相关性热图(图 8C)以及预后基因(SPP1, IL3RA)与 10 个免疫细胞的相关性热图(图 8D)。结果显示，SPP1 与 Macrophages M0 存在最高的正相关性，IL3RA 与 T cells CD8 存在最高的正相关性。

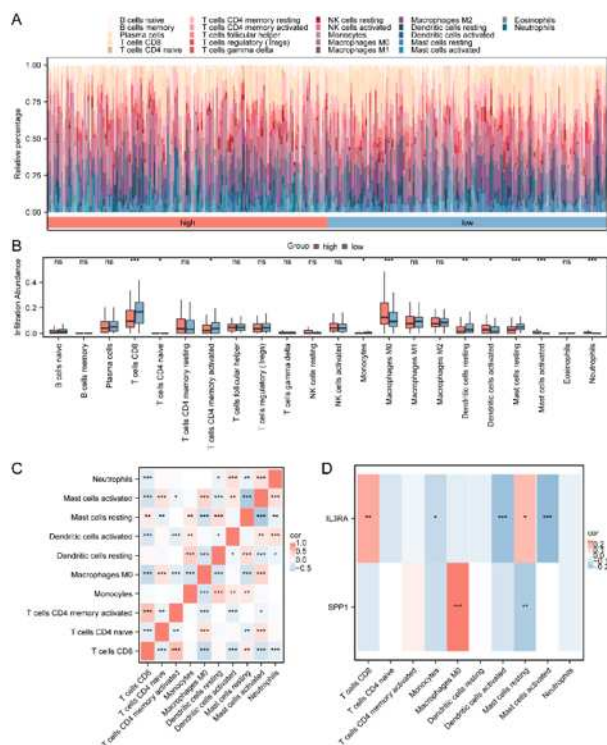


图 8 CIBERSORT 分析

A. 堆积柱状图。B. 不同分组比较图。C. 相关性热图。D. 棒棒糖图。

3. 讨论

肿瘤浸润免疫细胞(tumor infiltrates immune cells, TICs)是肿瘤微环境的重要组成部分^[3]。肿瘤细胞和 TICs 之间的相互作用最终导致可能加速肿瘤生长和侵袭的过程，如免疫逃逸、抗凋亡和异常增殖^[4]。免疫相关基因的异常表达及免疫细胞浸润在肿瘤免疫逃逸过程中的作用成为肿瘤免疫治疗的热点。

巨噬细胞是肿瘤微环境中的一种代表性细胞类型，浸润肿瘤微环境的巨噬细胞被称为肿瘤相关巨噬细胞(TAMs)。高 TAMs 浸润通常与多种肿瘤的不良预后相关，会使放疗、化疗和靶向治疗等标准治疗的疗效变差。操纵细胞形态影响巨噬细胞极化，从而诱导抗肿瘤反应，可能是一个潜在的治疗方向^[5]。磷蛋白 1(SPP1)又称骨化蛋白(Osteopontin, OPN)，由 SPP1 基因编码的一种多功能分泌型磷酸化糖蛋白，其生物功能多种多样且具有特异性。SPP1 基因表达的上调通常与炎症事件有关，在不同类型免疫细胞的免疫反应中发挥着重要作用[6, 7]。SPP1 也在癌细胞中表达，以往的研究表明，循环中 SPP1 的水平和高/或肿瘤细胞中 SPP1 表达的增加与多种癌症的不良预后以及化疗耐药之间存在相关性^[8]；与宫颈癌也关系密切，SPP1 mRNA 和蛋白在宫颈癌中的表达量明显较高，这可能会导致癌细胞对顺铂产生耐药性^[8]。我们的研究结果显示，相比正常的宫颈组织，宫颈癌组织中 SPP1 的表达明显升高；生存分析试验显示晚期宫颈癌的亚组分型中，SPP1 基因表达高的患者总生存期相对较低；宫颈癌肿瘤微环境中 SPP1 的表达与 M0 型巨噬细胞的浸润密切相关，推测该基因表达可能在肿瘤的发展过程中，分子调控功能发生了改变，并且可能影响 TAMs 的功能变化从而影响宫颈癌患者的预后。因此认为，SPP1 可能作为预测总生存率和免疫检查点抑制剂治疗宫颈癌的一种有希望的生物标记物^[9]。

在应对感染的过程中，CD8+ T 细胞的新陈代谢与细胞外代谢物和免疫信号驱动的转录、翻译和表观遗传变化相关联；在疾病情况下，细胞内在和细胞外在代谢线索都会导致 CD8+ T 细胞功能障碍。T 细胞衰竭(T cell exhaustion, Tex)，最早是在慢性病毒感染描述的，即这些细胞持续存在，但无法成功清除致病威胁^[10]。研究表明，阻断 CD8+ Tex 表达的程序性死亡 1(PD-1)等表面协同抑制受体，可重振细胞溶解性细胞介导的免疫反应，从而根除某些持久性病毒^[11]。在癌症中发现，在肿瘤微环境(TME)中的表达

高水平 PD-1 的功能障碍 CD8+Tex，对抗肿瘤免疫疗法同样反应低下；表达 PD-1 的细胞被认为可以通过简单的单向逆转从无反应的衰竭状态被免疫检查点抑制剂治疗（Immune checkpoint inhibitor therapy, ICB）拯救^[12]。IL3RA，又称 CD123，由染色体 Xp22.3 和 Yp11.3 编码^[13]，被认为是白血病的潜在治疗靶点^[14]，其在宫颈癌中的作用报道的很少。一项研究显示，宫颈癌患者外周血中 CD123 阳性树突状细胞的比例低于宫颈上皮内瘤变和健康对照组，循环树突状细胞的减少可能会导致细胞因子、抗原递呈和幼稚 T 细胞等免疫功能的改变，从而引起疾病的进展^[15]。本研究显示，相比正常的宫颈组织，宫颈癌组织中 IL3R 的表达明显降低；生存分析试验显示晚期宫颈癌的亚组分型中，IL3R 基因表达高的患者总生存期相对较长；宫颈癌肿瘤微环境中 IL3R 的表达与 CD8+ T 细胞的浸润密切相关。因此，我们推测，在宫颈癌发生发展以及 HPV 感染致病过程中，表达 IL3R 较高的患者可能会诱导更多杀伤性 CD8+ T 细胞的浸润，从而起到抗病毒以及抗肿瘤进展的作用，或可能成为预测 HPV 感染进展以及宫颈癌的总生存率预测一种有希望的生物标记物。

参考文献：

- [1] 刘文敏, 陈茹雁, 程冬, 付振明。中美两国宫颈癌发病率与死亡率长期趋势比较分析。中国预防医学杂志, 2024, Vol. 25 No. 6: 782–787。
- [2] 王伟, 郝敏。晚期及复发转移性子宫颈癌的免疫和靶向治疗。中国妇产科临床杂志, 2024, Vol. 25, No. 5: 385–387。
- [3] Hinshaw DC, Shevde LA. 2019. The Tumor Microenvironment Innately Modulates Cancer Progression. *Cancer Res.* Sep 15;79(18):4557–4566。
- [4] Xu F, Shen J, et al. 2021. Multi-Omics Data Analyses Construct a Six Immune-Related Genes Prognostic Model for Cervical Cancer in Tumor Microenvironment. *Front Genet.* May 24;12:663617。
- [5] 李雨凝, 张乐蒙。肿瘤相关巨噬细胞在抗肿瘤免疫治疗中的应用及展望。中山大学学报（医学科学版），2024, Vol. 45, No. 6: 983–993。
- [6] Lin EY, Xi W, Aggarwal N, Shinohara ML. Osteopontin (OPN)/SPP1: from its biochemistry to biological functions in the innate immune system and the central nervous system (CNS). *Int Immunol.* 2023 Apr 4;35(4):171–180。
- [7] Matsubara E, Yano H, Pan C, Komohara Y, Fujiwara Y, Zhao S, Shinchi Y, Kurotaki D, Suzuki M. The Significance of SPP1 in Lung Cancers and Its Impact as a Marker for Protumor Tumor-Associated Macrophages. *Cancers (Basel).* 2023 Apr 12;15(8):2250。
- [8] Xu X, Jiang X, Chen L, Zhao Y, Huang Z, Zhou H, et al. MiR-181a promotes apoptosis and reduces cisplatin resistance by inhibiting osteopontin in cervical cancer cells. *Cancer Biotherapy & Radiopharmaceuticals.* 2019; 34: 559 – 565。
- [9] Zhao K, Ma Z, Zhang W. Comprehensive analysis to identify SPP1 as a prognostic biomarker in cervical cancer. *Front Genet.* 2022; 12: 732822。
- [10] Blank CU, Haining WN, Held W, Hogan PG, Kallies A, Lugli E, et al. Defining ‘T Cell Exhaustion’. *Nat Rev Immunol* (2019) 14:768。
- [11] Virgin HW, Wherry EJ, Ahmed R. Redefining Chronic Viral Infection. *Cell* (2009) 138:30 – 50。
- [12] Dolina JS, Van Braeckel-Budimir N, Thomas GD, Salek-Ardakani S. CD8+ T Cell Exhaustion in Cancer. *Front Immunol.* 2021 Jul 20;12:715234。
- [13] Milatovich A, Kitamura T, Miyajima A, Francke U. Gene for the alpha subunit of the human interleukin-3 receptor (IL3RA) localized to the XY pseudoautosomal region. *American Journal of Human Genetics.* 1993;53: 1146 – 53。
- [14] Shi M, Su RJ, Parmar K, Chaudhry R, Sun K, Rao J, et al. CD123: a novel biomarker for diagnosis and treatment of leukemia. *Cardiovascular&Hematological Disorders-Drug Targets.* 2019; 19: 195–204。
- [15] Van Seters M, Beckmann I, Heijmans-Antonissen C, van Beurden M, Ewing PC, Zijlstra FJ, et al. Disturbed patterns of immunocompetent cells in usual-type vulvar intraepithelial neoplasia. *Cancer Research.* 2008; 68:6617 – 22。

作者简介：

王辰林（2003—），男，汉族，南京医科大学临床医学本科，研究方向肿瘤学。

基金项目：江苏省高等学校大学生创新创业训练计划（202410312105Y）。