

• 临床研究 •

基于多基因构建胰腺癌预后模型及临床价值

高过, 华永飞, 江寅, 王高卿

【摘要】目的 利用数据集分析对胰腺癌具有预后价值的重要基因, 并构建预后模型。**方法** 通过单因素 Cox 回归分析获得与胰腺癌预后明显相关的基因, 利用 Lasso Cox 回归构建基因预后模型, 并结合相关临床因素绘制列线图。**结果** 共鉴定出着丝粒蛋白 E(CENPE)、肌动蛋白结合蛋白(ANLN)和细胞骨架相关蛋白 2 类似物(CKAP2L) 3 个基因, 均在胰腺癌组织中高表达, 且与胰腺癌患者的不良预后相关。基于三基因构建的预后模型将患者分为高危组和低位组, 采用单因素和多因素 Cox 回归分析, 结果显示预后模型和化疗均是独立预测因素(均 $P < 0.05$)。列线图预测 1 年总体生存时间的曲线面积为 0.78。**结论** CENPE、ANLN 和 CKAP2L 的高表达预示着胰腺癌患者的不良预后, 基于三基因表达的列线图模型对于指导临床治疗具有重要价值。

【关键词】 胰腺癌; 预后模型; 数据库

doi:10.3969/j.issn.1671-0800.2024.03.011

【中图分类号】 R735.9 **【文献标志码】** A **【文章编号】** 1671-0800(2024)03-0319-05

胰腺癌(pancreatic cancer, PC)是发病率和死亡率逐年上升的恶性肿瘤之一^[1]。其是一个高度异质性的恶性肿瘤^[2], 传统的分期如 TNM 分期, 在预测患者生存差异方面并未表现出很好效果^[3]。精准医学为 PC 患者的治疗提供了许多有价值的参考, 也让个体化治疗成为了现实^[4-5]。二代测序技术的使用, 使利用相关特征对患者进行风险分层成为了可能^[6]。本研究通过不同数据集获得了和 PC 预后相关的基因, 并建立了着丝粒蛋白 E (centromere protein E, CENPE)、肌动蛋白结合蛋白 (anillin, actin binding protein, ANLN) 和细胞骨架相关蛋白 2 类似物 (cytoskeleton associated protein 2 like, CKAP2L) 在内的预后模型, 适用于指导患者预后评估和治疗决策, 现报道如下。

1 资料与方法

1.1 数据来源 本研究严格遵守根据赫尔辛基宣言, 所使用的数据来源于 TCGA (<https://portal.gdc.com>)、ICGC (<https://dcc.icgc.org/releases/current/Projects>) 和 HPA (<https://www.proteinatlas.org/>) 数据库。本研究从 ICGC 和 TCGA 数据库获取 PC 的 RNAseq 数据 (level3) 和相应的临床信息, 从 GTEx

V8 版数据库 (<https://gtexportal.org/home/datasets>) 获取正常胰腺组织 RNAseq 数据, 从 HPA 数据库获取相关蛋白在 PC 中表达的免疫组化结果。

1.2 筛选和 PC 预后相关的基因 基于 ICGC 数据库所获取的两个数据集中 PC 的 RNAseq 数据 (level3) 和相应的临床信息, 通过单因素 Cox 回归分析与 PC 预后相关的基因, 两个数据集取交集并结合文献资料最终选择所要研究的目的基因。

1.3 目的基因在 PC 和正常胰腺组织中的表达水平 通过 TCGA 数据库进一步验证所筛选出的目的基因在 PC 和正常胰腺组织中的表达水平。从 TCGA 数据库获取在 PC 及癌旁组织中 RNAseq 数据 (level3), 从 GTEx 数据库获取正常胰腺组织 RNAseq 数据。

1.4 目的基因表达水平对 PC 患者预后的影响 基于 TCGA 数据库获取的数据, 进一步验证目的基因对 PC 患者预后的影响。

1.5 构建基因相关的预后模型 基于 1.4 中所获取的 TCGA 数据集, 使用 R 包“glmnet”执行 Lasso Cox 回归以及 10 倍交叉验证; 采用 R 软件 survival 包进行多因素 Cox 回归分析构建预后模型。

1.6 基因预后模型独立于其他临床因素 为了证明 PC 患者存在其他临床变量影响预后, 本研究进行了单因素和多因素 Cox 回归分析, 将临床特征和基因预后模型设置为自变量, 将总体生存时间 (overall survival, OS) 设置为因变量。

基金项目: 浙江省医药卫生科技计划项目 (2022KY1077)

作者单位: 315040 宁波, 宁波市医疗中心李惠利医院

通信作者: 华永飞, Email: lhlhuayongfei@nbu.edu.cn

1.7 构建和验证列线图 为了提高模型的适用性,将具有预后能力的临床变量与基因特征模型相结合,并应用列线图来预测 PC 患者的预后。基于单因素和多因素 Cox 回归分析选择所有独立预后因素来构建列线图来评估 PC 患者 1 年 OS 的概率。使用 R 包“timeROC”获得 1 年 OS 的 ROC 曲线和 AUC。

1.8 统计方法 采用 R V4.0.3 软件进行统计学分析,使用的 R 包主要包括 survival、survminer、timeROC、glmnet 和 rms 等。两组间基因表达水平的比较采用 Wilcoxon 检验。 $P < 0.05$ 表示差异有统计学意义。

2 结果

2.1 PC 差异表达基因 通过 ICGC 数据库中的两个数据集筛选与 PC 预后相关的差异基因,其中澳大利亚数据集和加拿大数据集分别获取了 2 355 和 2 828 个差异基因,将两个数据集所获差异基因最显著的前 30 个基因取交集,最终获得 3 个候选目的基因,分别 CENPE、ANLN 和 CKAP2L,见图 1。

2.2 差异基因在 PC 和正常组织中的表达水平 基于 TCGA 和 GTEx 数据集比较 2.1 中所选 3 个差异基因的表达水平,结果显示相比正常胰腺组织,CENPE、ANLN 和 CKAP2L 在胰腺癌组织中的表达明显升高(均 $P < 0.05$),见图 2。HPA 数据库的免疫组织化学染色结果也提示 CENPE、ANLN 和 CKAP2L 蛋白表达水平呈强阳性,见图 3。

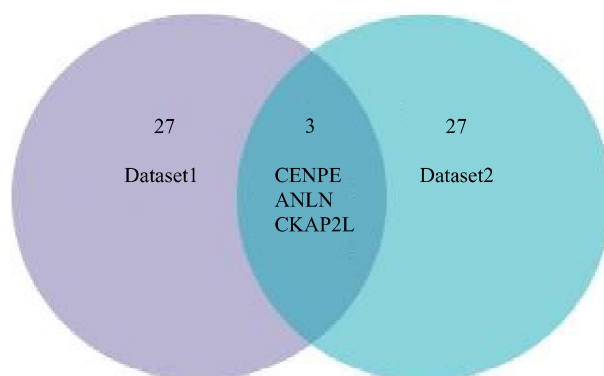
2.3 差异基因表达水平与 PC 不良预后的相关性 基于 TCGA 数据集进一步验证和分析 CENPE、ANLN 和 CKAP2L 的表达对 PC 患者 OS 的影响,结果显示高表达 CENPE、ANLN、CKAP2L 的 PC 患者的中位 OS 短于低表达患者(均 $P < 0.05$),见图 4。

2.4 构建预测模型 使用 TCGA 数据集生成了基于上述三个基因表达的预测模型,随后计算了三个基因表达风险评分,根据风险评分截断值,89 例患者被分为高风险组,其余 90 例患者被分为低风险组。基于三个基因的两组 K-M 曲线提示高风险组的预后明显短于低风险组($P < 0.05$)。预后模型的 1 年生存时间的 AUC 为 0.67,见图 5。

2.5 构建和验证列线图 采用单因素和多因素 Cox 回归分析来检测基于三基因的预后模型在 TCGA 队列中的独立预测能力。单因素 Cox 回归分析结果提

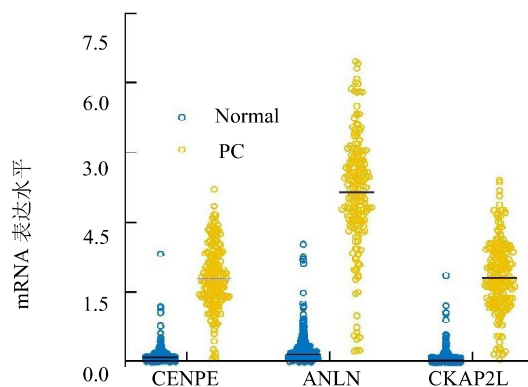
示:预后模型、年龄、肿瘤分级(分化程度)和放疗具有预后价值,而其他变量与 OS 没有显著相关;将上述具有统计学差异的变量纳入到多因素 Cox 回归分析,显示预后模型和放疗是独立预后因素(均 $P < 0.05$),见表 1。

为了使预后模型适用于临床,应用列线图来预测 TCGA 队列中患者 1 年 OS 的概率,列线图中的预测因素包括预后模型和放疗,1 年 OS 概率校准图提示列线图性能较好,见图 6。进一步计算了相应模型预测 1 年 OS 的 AUC 为 0.78,见图 7。



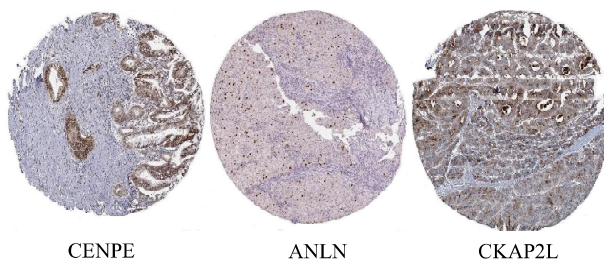
注: Dataset 1 为澳大利亚数据集; Dataset 2 为加拿大数据集

图 1 差异表达基因



注: Normal 为正常胰腺组织; PC 为胰腺癌

图 2 CENPE、ANLN 和 CKAP2L 基因在胰腺癌和正常组织中的表达水平



注: 免疫组织化学染色结果

图 3 CENPE、ANLN 和 CKAP2L 蛋白在胰腺癌中的表达 (IHC 染

3 讨论

PC 是最常见的癌症死亡原因之一,长期存活率仍然很低,随着治疗水平的不断提高,虽然在一定程度上改善了 PC 患者的预后,但通过危险分层识别高危患者将更有助于提高 PC 患者的存活时间。基因的异常表达是肿瘤始发的内在因素,筛选出 PC 中异

常表达的基因对于开发新标记物和构建肿瘤预后模型至关重要。本研究利用不同数据集筛选出与 PC 预后密切相关的差异表达基因,进而构建了多基因预后模型,并且纳入了相关临床因素形成了具有较好预测能力的列线图。

本研究结果显示 CENPE、ANLN 和 CKAP2L 与 PC 密切相关。CENPE 是着丝粒的核心成分,在有

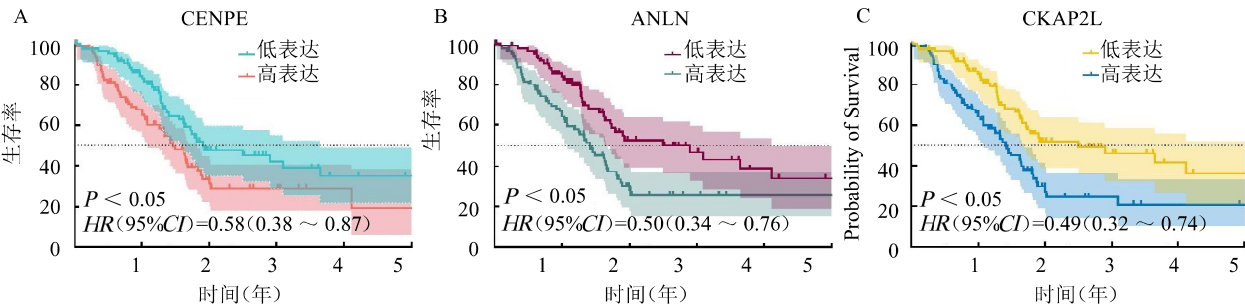


图 4 基于基因表达和风险模型的生存分析

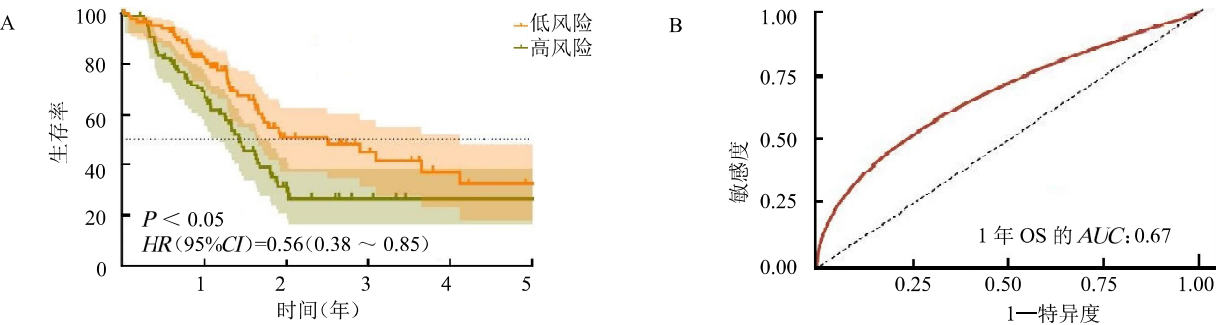


图 5 以最佳截断点为界,高风险组的总生存率低于低风险组

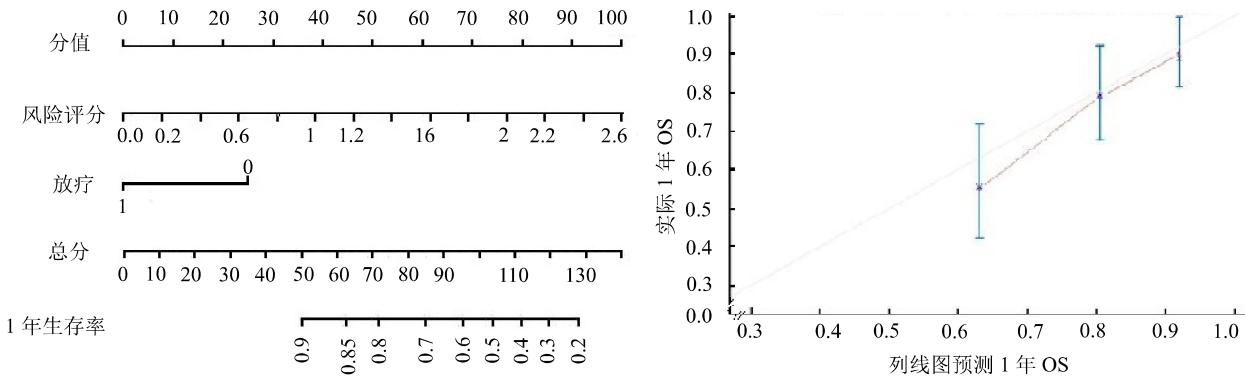


图 6 列线图的构建和验证

表 1 单因素和多因素 Cox 回归分析

指标	单因素回归分析		多因素回归分析	
	HR 值(95%CI)	P 值	HR 值(95%CI)	P 值
预后模型	3.24(2.02 ~ 5.19)	< 0.05	3.36(1.85 ~ 6.09)	< 0.05
年龄	1.03(1.01 ~ 1.05)	< 0.05	1.02(1.00 ~ 1.05)	> 0.05
性别	0.81(0.54 ~ 1.22)	> 0.05		
种族	0.90(0.59 ~ 1.37)	> 0.05		
分期	1.22(0.85 ~ 1.76)	> 0.05		
肿瘤分级	1.45(1.09 ~ 1.93)	< 0.05	1.40(0.99 ~ 1.99)	> 0.05
吸烟状态	1.07(0.68 ~ 1.69)	> 0.05		
放疗	0.39(0.20 ~ 0.77)	< 0.05	0.40(0.20 ~ 0.78)	< 0.05

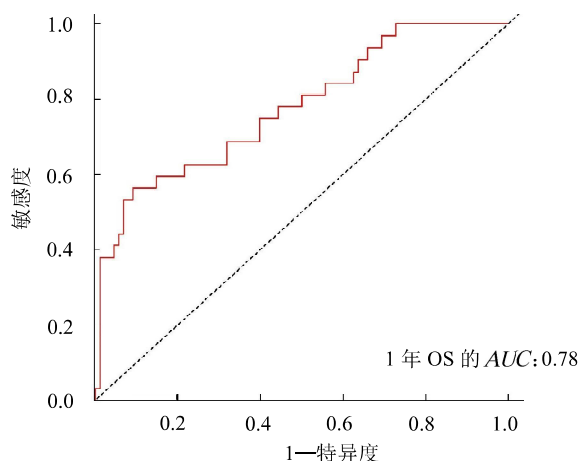


图7 列线图模型预测PC患者1年OS的时间依赖性ROC曲线

丝分裂过程中将染色体连接到纺锤体微管上,在纺锤体微管捕获和染色体会聚中发挥作用^[7-8]。相关研究已发现CENPE与多种肿瘤的恶性表型有关,如肺癌^[9]、食管癌^[10]、乳腺癌^[11]和视网膜母细胞瘤^[12]等。紫杉醇和秋水仙碱通过抑制有丝分裂纺锤体可塑性来抑制肿瘤细胞分裂,而CENPE蛋白在有丝分裂过程中起着重要作用,因此,CENPE很可能成为实体肿瘤治疗的靶点^[13]。本研究也发现CENPE在PC组织中表达明显升高,且与PC患者的不良预后相关,与Glyven^[14]研究结果相符。

既往相关研究已发现^[15-16],ANLN过表达于PC组织和细胞系中,且与肿瘤大小、肿瘤分化程度、TNM分期、淋巴结转移、远处转移和不良预后相关,ANLN通过调节EZH2/miR-218-5p/LASP1信号轴促进PC进展,这提示ANLN可能是潜在治疗靶点。Olakowski等^[17]也发现ANLN和HIST1H1C在PC中高表达,并基于这两个基因构建了预后模型且具有较好性能。本研究通过筛选也发现ANLN明显与PC患者的不良预后有关,进一步提示ANLN在PC中的关键作用。

CKAP2L是一种细胞周期相关蛋白,与许多肿瘤的进展密切相关。一项泛癌研究表明,在大多数癌症中,CKAP2L表达和活性显著升高,CKAP2L表达升高导致患者预后不良,并且是大多数肿瘤的独立危险因素;高表达CKAP2L的患者对免疫治疗更加敏感,可作为预测肿瘤免疫治疗的生物标志物^[18]。本研究结果也显示CKAP2L和PC患者的不良预后密切相关。杨潞颀^[19]对CKAP2L在PC中的作用机

制进行详细探讨,发现CKAP2L通过调节MYC通路促进PC细胞的恶性表型,且促进了PC细胞的体内成瘤和肝转移。

基于临床相关因素构建的PC预后模型已被诸多研究报道,尽管所报道的模型效能较好^[20-21],但考虑基因异常在肿瘤中的关键作用,似乎有必要将相关异常基因纳入到模型中从而更能提高其效能。本研究结合上述三基因构建的危险分层和预后模型,可以较准确对患者预后作出预测。Olakowski^[17]等基于ANLN和HIST1H1C双基因构建的PC预后模型,列线图模型的曲线下面积在0.5年时为0.83,在1年时为0.73,并且列线图模型的C指数比单独的其他变量高0.66。说明了结合将异常表达的基因和临床相关因素结合构建肿瘤预后模型的重要性。

然而,本研究存在一些局限性。首先,考虑到PC的预后不良,本文构建了1年OS的预测模型;其次,由于缺乏特定的临床数据,无法对列线图进行外部验证,这意味着列线图应该使用多中心临床试验和前瞻性研究进一步验证。

综上所述,本研究使用相对严格的回归模型方法确定三个基因在PC中具有预后意义,构建了包含预后模型的列线图,以帮助临床医生以个性化方式治疗PC患者。

致谢 感谢TCGA、ICGC、GTEx和HPA数据库提供的平台和资源,同时感谢中南大学湘雅二医院的武强博士在数据处理上提供的指导和帮助。

利益冲突 所有作者声明无利益冲突

作者贡献声明 高过:实验操作、论文撰写;高过、江寅:数据整理、统计学分析;江寅、王高卿、华永飞:研究指导、论文修改、经费支持

参 考 文 献

- [1] KLEIN A P. Pancreatic cancer epidemiology: understanding the role of lifestyle and inherited risk factors[J]. Nature reviews Gastroenterology & Hepatology, 2021, 18(7): 493-502.
- [2] KAMISAWA T, WOOD L D, ITOI T, et al. Pancreatic cancer[J]. The Lancet, 2016, 388(10039): 73-85.
- [3] VAN ROESSEL S, KASUMOVA G G, VERHEIJ J, et al. International validation of the eighth edition of the American Joint Committee on Cancer (AJCC) TNM staging system in patients with resected pancreatic cancer[J]. JAMA Surgery, 2018, 153(12): e183617.
- [4] LOMBERK G, DUSETTIN I, IOVANNA J, et al. Emerging epigenomic landscapes of pancreatic cancer in the era of precision medicine[J]. Nature Communications, 2019, 10(1): 3875.
- [5] TORRES C, GRIPPO P J. Pancreatic cancer subtypes: A roadmap for precision medicine[J]. Annals of Medicine, 2018, 50(4):

- 277-287.
- [6] YU J, SADAKARI Y, SHINDO K, et al. Digital next-generation sequencing identifies low-abundance mutations in pancreatic juice samples collected from the duodenum of patients with pancreatic cancer and intraductal papillary mucinous neoplasms[J]. *Gut*, 2017, 66(9): 1677-1687.
- [7] YEN T J, LI G, SCHAAR B T, et al. CENP-E is a putative kinetochore motor that accumulates just before mitosis[J]. *Nature*, 1992, 359(6395): 536-539.
- [8] WOOD K W, SAKOWICZ R, GOLDSTEIN L S, et al. CENP-E is a plus end-directed kinetochore motor required for metaphase chromosome alignment[J]. *Cell*, 1997, 91(3): 357-366.
- [9] SHAN L, ZHAO M, LU Y, et al. CENPE promotes lung adenocarcinoma proliferation and is directly regulated by FOXM1[J]. *Int J Oncol*, 2019, 55(1): 257-266.
- [10] ZHU X, LUO X, FENG G, et al. CENPE expression is associated with its DNA methylation status in esophageal adenocarcinoma and independently predicts unfavorable overall survival[J]. *PLoS One*, 2019, 14(2): e0207341.
- [11] KUNG P P, MARTINEZ R, ZHU Z, et al. Chemogenetic evaluation of the mitotic kinesin CENP-E reveals a critical role in triple-negative breast cancer[J]. *Mol Cancer Ther*, 2014, 13(8): 2104-2115.
- [12] IEGIANI G, GAI M, DI CUNTO F, et al. CENPE inhibition leads to mitotic catastrophe and DNA damage in medulloblastoma cells[J]. *Cancers*, 2021, 13(5): 1028.
- [13] EL-ARABEY A A, SALAMA S A, ABD-ALLAH A R. CENP-E as a target for cancer therapy: Where are we now[J]. *Life Sci*, 2018, 208: 192-200.
- [14] GUVEN E. Gene expression characteristics of tumor and adjacent non-tumor tissues of pancreatic ductal adenocarcinoma (PDAC) in-silico[J]. *Iran J Biotechnol*, 2022, 20(1): e3092.
- [15] WANG A, DAI H, GONG Y, et al. ANLN-induced EZH2 upregulation promotes pancreatic cancer progression by mediating miR-218-5p/LASP1 signaling axis[J]. *J Exp Clin Cancer Res*, 2019, 38(1): 347.
- [16] ZHOU S, YAN Y, CHEN X, et al. A two-gene-based prognostic signature for pancreatic cancer[J]. *Aging (Albany NY)*, 2020, 12(18): 18322-18342.
- [17] OLAKOWSKI M, TYSZKIEWICZ T, JARZAB M, et al. NBL1 and anillin (ANLN) genes over-expression in pancreatic carcinoma[J]. *Folia Histochem Cytobiol*, 2009, 47(2): 249-255.
- [18] YI B, FU Q, ZHENG Z, et al. Pan-cancer analysis reveals the prognostic and immunotherapeutic value of cytoskeleton-associated protein 2-like[J]. *Sci Rep*, 2023, 13(1): 8368.
- [19] 杨璐嶝. CKAP2L 在胰腺癌发生发展中的作用及机制研究[D]. 重庆:西南大学, 2022.
- [20] 董仕, 施华清, 李昕, 等. 胰腺癌患者预后风险模型的构建及验证[J]. *兰州大学学报(医学版)*, 2022, 48(12): 22-28.
- [21] STOCKEN D, HASSAN A, ALTMAN D, et al. Modelling prognostic factors in advanced pancreatic cancer[J]. *Br J Cancer*, 2008, 99(6): 883-893.

收稿日期:2023-12-29

(本文编辑:吴迪汉)