

· 临床研究 ·

m1A 相关基因构建的胰腺癌预后模型的综合分析

时 坚^{1,2}, 刘雪昂^{1,2}, 朱 岩³, 胡 冉⁴, 马文静⁴, 朱 毅^{1,2*}¹南京医科大学第一附属医院胰腺中心, 江苏 南京 210029; ²南京医科大学胰腺研究所, 江苏 南京 210029; ³南京医科大学第一附属医院病理科, ⁴公共实验中心, 江苏 南京 210029

[摘要] 目的: 基于m1A相关基因在多个数据库中差异表达情况与临床参数建立新的胰腺癌风险评估模型, 为胰腺癌治疗和预后分析提供理论依据。方法: 通过分析m1A相关基因在TCGA-GTEX数据库中的差异表达, 利用Cox分析与LASSO回归构建胰腺癌预后风险模型, 搭配GEO数据库和本中心胰腺癌病例的基因表达和临床数据验证该模型的准确性与敏感性。结果: 筛选出CRLS1与C7orf50两个基因共同构成风险模型, 本研究系统验证了该模型有预测胰腺癌预后的显著效能, 该模型联合肿瘤突变负荷(tumor mutation burden, TMB)具有更强的预后预测能力。结论: 该模型是预测胰腺癌患者预后的新工具。

[关键词] 胰腺癌; m1A; 风险模型; 预后**[中图分类号]** R735.9**[文献标志码]** A**[文章编号]** 1007-4368(2023)07-954-12**doi:** 10.7655/NYDXBNS20230708

Comprehensive analysis of pancreatic cancer prognostic model constructed by m1A related genes

SHI Jian^{1,2}, LIU Xue'ang^{1,2}, ZHU Yan³, HU Ran⁴, MA Wenjing³, ZHU Yi^{1,2*}¹Pancreas Centre, the First Affiliated Hospital of Nanjing Medical University, Nanjing 210029; ²Pancreas Institute, Nanjing Medical University, Nanjing 210029; ³Department of Pathology, ⁴Public Laboratory Center, the First Affiliated Hospital of Nanjing Medical University, Nanjing 210029, China

[Abstract] **Objective:** To establish a new pancreatic cancer risk assessment model based on the differential expression of m1A-related genes in multiple databases and clinical parameters, and provide theoretical basis for pancreatic cancer treatment and prognosis analysis. **Methods:** Differential analysis of m1A-related genes was performed in the TCGA-GTEX database. Cox analysis and LASSO regression were used to construct a pancreatic cancer prognosis risk model, and the accuracy and sensitivity of the model were verified by gene expression and clinical data from the GEO database and pancreatic cancer cases in our center. **Results:** CRLS1 and C7orf50 were selected to form the risk model. The study systematically verified that the model had significant efficacy in indicating the prognosis of pancreatic cancer, and the model combined with tumor mutation burden (TMB) had a stronger prognostic indication ability. **Conclusion:** A new risk model is constructed to indicate the prognosis of pancreatic cancer patients.

[Key words] pancreatic cancer; m1A; risk model; prognosis

[J Nanjing Med Univ, 2023, 43(07):954-965]

胰腺癌(pancreatic cancer, PC)是一种恶性程度极高的消化道恶性肿瘤,发病率和病死率不断上升,已成为我国第六大致死原因。胰腺癌起病隐匿,易复发和转移,由于缺乏有效的预测和治疗策

略,预后往往较差,即使是经过根治性手术治疗的胰腺癌患者仍会在5年内出现肿瘤复发^[1]。因此,迫切需要开发新的预后评估方法来预测患者的临床预后。

研究发现,不同生物体的RNA中存在超过160种转录后RNA修饰,其中包括非编码RNA(如核糖体RNA、转运RNA和核内小RNA等)和mRNA^[2]。这

[基金项目] 江苏省“六大人才高峰”人才项目(WSW-032)

*通信作者(Corresponding author), E-mail: zhuyijssry@126.com

些修饰包括但不限于N6-甲基腺嘌呤(m6A)、N1-甲基腺嘌呤(m1A)和5-甲基胞嘧啶(m5C)等。m1A是一种动态可逆修饰过程,类似于m6A修饰,由甲基转移酶、脱甲基酶和结合蛋白调控^[3]。最新研究报道,m6A、m5C、m1A的RNA修饰有助于癌症的发生和发展,m6A/m5C/m1A调控基因的聚类亚组和风险模型与肝细胞癌的不良预后和免疫微环境相关,有望成为评估肝细胞癌患者预后的新工具^[4]。但是迄今未见m1A相关分子标志物来预测胰腺癌的预后,限制了m1A相关基因在指导胰腺癌诊断和治疗中的应用。

本研究找到了2个m1A相关基因(CRLS1与C7orf50)有独立预测胰腺癌预后的能力,并建立了预测胰腺癌预后的风险评分模型,命名为PC-m1AScore。本研究系统验证了PC-m1AScore、PC-m1AScore联合肿瘤突变负荷(tumor mutation burden, TMB)预测胰腺癌预后的效能,初步分析了PC-m1AScore和免疫细胞浸润之间的关系,旨在找到预测胰腺癌患者预后的新工具,有利于医生选择更合理的、个性化的手术范围,制定术后随访策略,实现基于肿瘤生物学行为的个体化治疗,有望改善患者的预后。

1 资料和方法

1.1 建模数据的来源

178例胰腺癌患者的癌组织、4例患者的癌旁组织的组织转录组测序数据信息、患者临床病理参数信息(包括性别、年龄、存活时间、TNM分期等)来源于公共癌症基因组图谱数据库(TCGA, <https://portal.gdc.cancer.gov/>)。165例正常胰腺组织的转录组测序数据来源于基因型组织表达数据库(GTEx, <https://www.gtexportal.org/home/index.html>)。m1A相关基因集来自Gao等^[2]的研究。

1.2 方法

1.2.1 胰腺癌差异表达的m1A相关基因的筛选

通过“limma”包,以 $P < 0.05$ 和 $|\log_2$ 差异倍数(fold change, FC)| > 1 为条件筛选胰腺癌和正常组织之间的差异基因。基因GO和KEGG通路分析在R包“clusterProfiler”上完成。接着使用Pearson检验计算差异表达基因之间的关联,并使用R包“corrplot”显示。

1.2.2 针对m1A相关基因所构建新的胰腺癌预后模型的建立与系统验证

①通过单因素Cox分析筛选与胰腺癌患者生存

相关的胰腺癌差异表达的m1A相关基因,并使用一致性指标(C-指数)来验证准确性。②将与预后相关的m1A相关基因纳入最小绝对收缩和选择算子(LASSO)分析。通过多因素Cox分析确定与预后相关的独立风险m1A相关基因。③通过线性组合Cox回归系数乘以基因表达的方法,建立新的胰腺癌预后模型:风险评分(score) $=\beta_1 \times$ 基因1表达 $+\beta_2 \times$ 基因2表达 $+\dots+\beta_n \times$ 基因n表达。本文将新建的胰腺癌预后模型命名为PC-m1AScore。④根据PC-m1AScore的中位数将胰腺癌患者分为高风险组和低风险组。为了评估两组的生存率,使用R软件包“survival”来绘制Kaplan-Meier曲线。基于受试者工作特征(receiver operating characteristic, ROC)曲线和随时间变化的曲线下面积(area under curve, AUC),进行为期5年的生存预测实验,以评估PC-m1AScore的预测效能。⑤运用单因素和多因素Cox分析,代入PC-m1AScore、性别、年龄、病理分化程度和TNM分期等多个变量指标,找到能定义公共数据库中胰腺癌病例预后风险的独立因素。结合ROC分析、临床亚组分析等验证手段,证明所建立的PC-m1AScore预测胰腺癌预后的效能。⑥在R软件包“rms”和“survival”上构建了列线图,以可视化Cox回归模型和评估预后预测的准确性。通过校准图校准了柱线图。⑦运用主成分分析(principal components analysis, PCA)方法,证明所建立的PC-m1AScore预测胰腺癌预后的效能。

1.2.3 PC-m1AScore联合TMB预测胰腺癌预后的效能

R包“survival”用于绘制Kaplan-Meier曲线,旨在评估TMB对患者预后的影响。然后将TMB数据与风险评分结合起来评估两者对预后的综合影响。使用R包“maftools”分析TCGA体细胞变异数据并可视化前10个突变基因。

1.2.4 PC-m1AScore及其组成基因的免疫相关分析

根据PC-m1AScore的中位数将胰腺癌患者分为高风险组和低风险组,在R包“Cibersort”上比较两组之间22种免疫相关细胞类型的浸润差异。通过Pearson检验分析免疫检查点(LMTK3、TIGIT、CD274、CTLA4、PDCD1)的基因表达量与组成PC-m1AScore的m1A相关基因表达量之间的关系。

1.2.5 具有完整临床信息数据的验证样本集的获取

除了公共数据库中的胰腺癌病例的测序与临床数据外,本研究还向经ISO90012015质量认证的南京医科大学第一附属医院胰腺中心胰腺专病生物样本库申请了60例胰腺癌患者手术切除的胰腺

癌组织和配对的癌旁非肿瘤组织。本研究获得医院伦理委员会的批准(伦理号2020-SRFA-364),所用组织样本都有对应来源患者的书面知情同意书。所用组织样本的患者术前均未接受化疗或放疗,均在2019—2022年接受了胰十二指肠切除术,经病理证实为胰腺癌。总生存期(overall survival, OS)定义为手术至死亡或末次随访日期之间的时间。对手术切除标本的取材是规范统一的,将取材组织一分为二,一半组织使用液氮速冻,另一半组织甲醛固定后制作石蜡块和HE切片,组织形态由两位病理医生确认。

1.3 统计学方法

所有统计分析均在 R4.3、SPSS 20.0 和 GraphPad Prism 5 中进行,所有统计检验均进行双尾检验。Student's *t* 检验、Wilcoxon rank-sum 检验用于检测肿瘤组织与正常胰腺组织基因表达差异,Fisher 精确检验用于确定不同模式之间变量的差异,Benjamini-Hochberg(BH-FDR)用于调整多个变量多重比较的 *P* 值^[5]。对于预后分析,采用 log-rank 检验和 Cox 回归分析检验特殊变量的预后价值。 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 特征性的胰腺癌差异表达 m1A 相关基因的筛选与鉴定

将TCGA数据库中178例胰腺癌患者表达情况和GTEX数据库中169个正常胰腺样本的数据整合,得到一个肿瘤与正常组织数据量基本接近的联合表达数据库。通过分析发现其中参与m1A修饰的10种修饰蛋白在肿瘤与正常组织中均存在明显表达差异,除ALKBH3与YTHDC1在胰腺癌中表达明显降低外,其余均明显升高(图1A)。71个m1A相关基因来自于Gao等^[2]的研究,使用STRING6、BioGrid7、OmniPath8、InWeb_IM9数据库进行蛋白质-蛋白质相互作用(PPI)富集分析,PPI结果提示其主要富集在DNA合成、RNA修饰、囊泡货物装载及嘧啶代谢领域(图1B)。

在GO富集分析中得到了同样的结果,生物过程中,m1A相关基因主要富集在:①“货物装载到COP II包被覆盖的囊泡中”,此过程是细胞内物质转运和分泌的重要步骤之一;②“核酸磷酸二酯键水解”,它是一种重要的生物化学反应,指的是通过水解反应将核酸分子中的磷酸二酯键断裂,从而释放出相应的核苷酸单元或碎片。这种反应在许多生

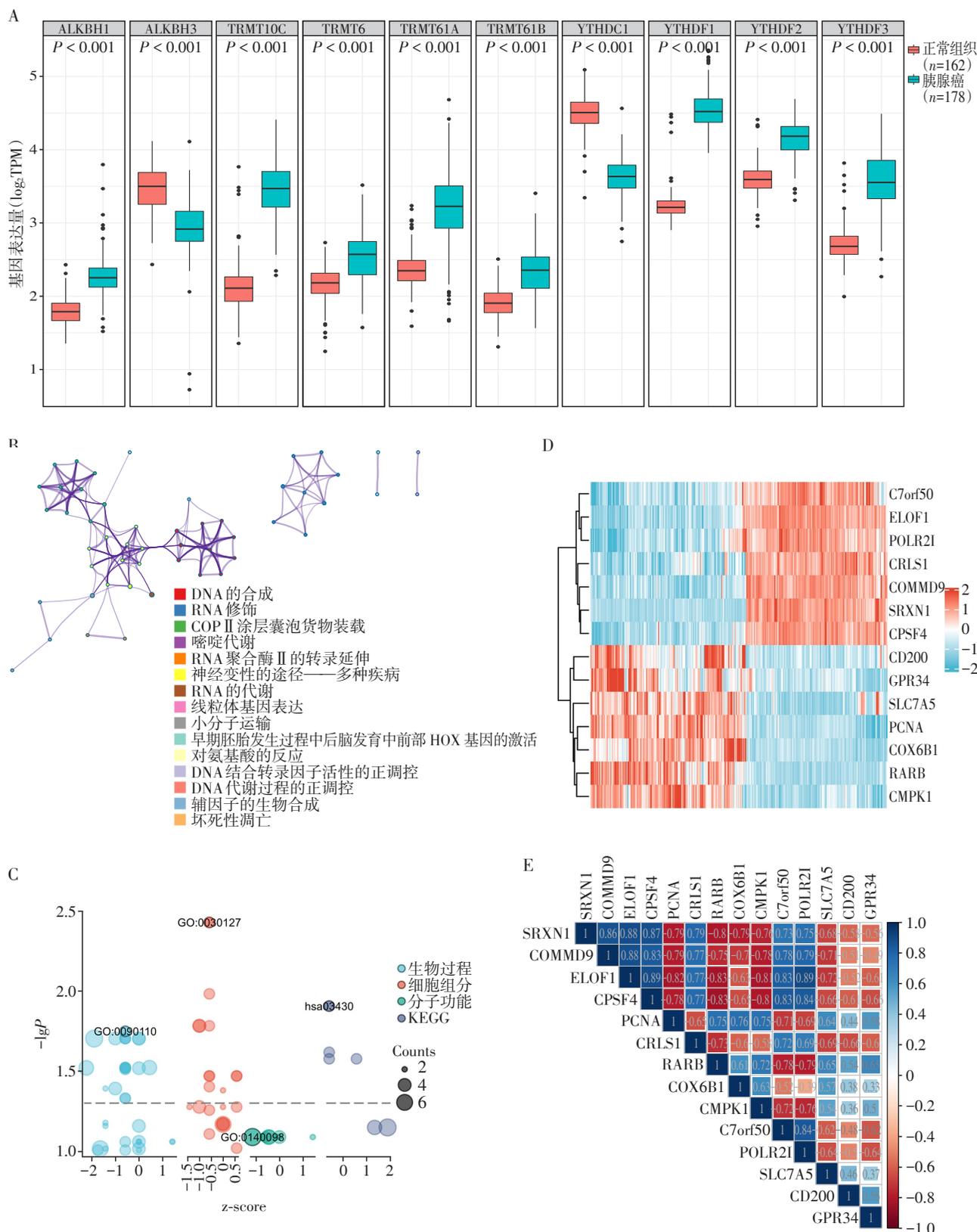
物过程中都具有重要的生物学功能,如DNA复制、RNA转录、RNA降解等。在细胞组分中,m1A相关基因主要富集在“COP II囊泡包被”,COP II囊泡包被是细胞内转运和分泌蛋白质的一种膜运输体系,它由多种蛋白质复合物组成,其中包括COP II囊泡包被蛋白,它们的聚合可以促进细胞膜上的货物进入囊泡内。COP II囊泡包被在细胞内转运和分泌过程中扮演着重要角色,其调控异常可能导致一些疾病的发生^[6]。

KEGG通路富集分析结果进一步提示,m1A相关基因主要富集在“错配修复”通路(图1C),它是细胞DNA复制过程中的一个重要机制,负责纠正在复制过程中可能出现的DNA序列不一致的错误。这种修复机制在维护基因组稳定性和避免突变方面具有重要作用,对于细胞的生存和繁殖都至关重要^[7]。

考虑到如果将71个基因的表达量都纳入预测模型,基因数量过大会限制临床转化应用,本研究尝试对基因集进行降维:通过对71个m1A相关基因在胰腺癌与正常胰腺组织中的表达分析,筛选出显著上调与下调基因各7个(图1D),并发现除CD200外,其余基因之间的相关性极高,Pearson相关系数绝对值普遍超过0.7(图1E)。

2.2 新的胰腺癌预后风险模型PC-m1AScore的构建和验证

通过单因素Cox回归分析筛选队列中胰腺癌患者的生存数据和14个差异基因表达谱,其中8个基因与总生存期相关,之前已报道多数基因与肿瘤发生发展密切相关(图2A)。随后利用LASSO回归分析和多因素Cox分析,最终找到了2个m1A相关基因(CRLS1与C7orf50)有独立预测预后的能力($P < 0.05$,图2B),用于构建预后风险模型。通过线性组合Cox回归系数乘以基因表达的方法,建立新的胰腺癌预后模型并命名为PC-m1AScore:风险评分 $= (1.798\ 5 \times CRLS1) + (-1.718\ 6 \times C7orf50)$ (图2B~D)。C-指数用于评估预后模型的预测准确性,其与患者预后的相关性优于其他临床参数(图2E)。在TCGA数据中,模型基因的表达在高风险组和低风险组中都显著不同(图2F),在计算出TCGA数据库中每个患者的PC-m1AScore风险评分后,根据中位数将他们分为高风险组和低风险组(图2G)。Kaplan-Meier分析表明,高风险组患者的预后明显比低风险组患者的预后更差($P < 0.05$,图2H、I)。在TCGA-GTEX数据集中,1年、3年和



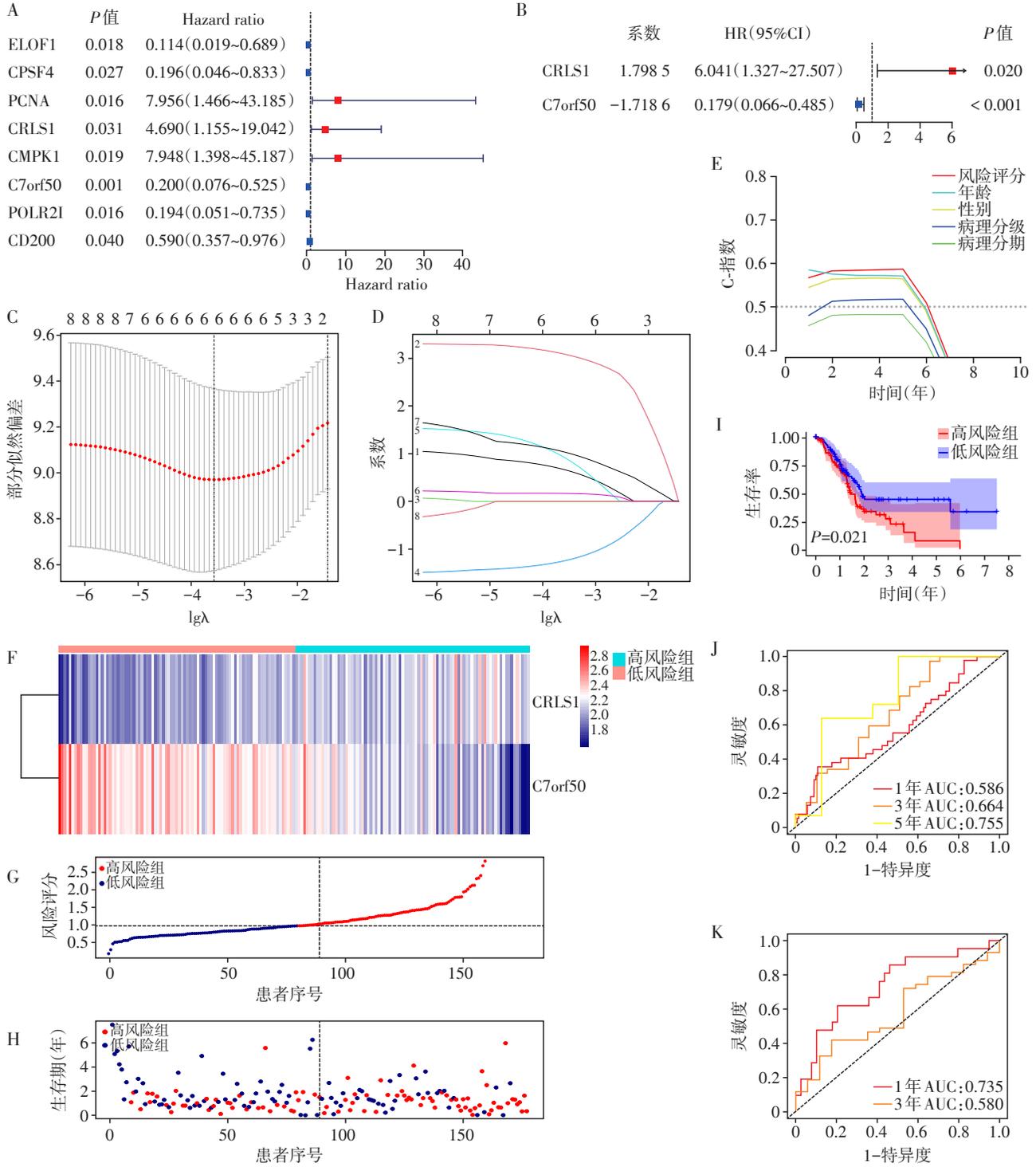
A: 10个m1A 修饰基因在TCGA-GTEx 数据库中胰腺癌与正常组织的表达差异;B: 71个m1A 相关基因蛋白质相互作用富集分析;C: 71个m1A 相关基因GO和KEGG 通路富集分析;D: 14个m1A 相关基因在TCGA-GTEx 数据库中胰腺癌与正常组织中的差异表达热图;E: 14个m1A 相关基因之间相关性分析。

图1 胰腺癌差异表达m1A 相关基因的筛选与鉴定

Figure 1 Screening and identification of differential expression of m1A-related genes in pancreatic cancer

5年的生存期预测风险评分的AUC值分别为0.586、0.664和0.755(图2J),而由于本胰腺中心60例患

者取样时间较晚,缺少5年生存期预测风险评分,其1年、3年评分分别为0.735和0.580(图2K),



A: 单因素 Cox 分析提示 14 个差异基因中有 8 个基因有预测预后的能力 ($P < 0.05$); B: 多因素 Cox 分析提示 2 个 m1A 相关基因有独立预测预后的能力 ($P < 0.05$); C、D: LASSO 回归验证多因素分析结果准确性; E: C-指数评估胰腺癌预后风险模型 PC-m1AScore 的准确性; F: 根据 PC-m1AScore 的中位数将胰腺癌患者分为高风险组和低风险组, PC-m1AScore 中 2 个权重基因表达的热图; G: 将患者以 PC-m1AScore 大小排序, 以中位数为分组标准进行分组; H: 不同分险组患者生存期; I: 不同风险组生存期 Kaplan-Meier 曲线; J: PC-m1AScore 评估 TCGA-GTEX 数据库中胰腺癌患者生存期 1、3、5 年预后的 AUC 值; K: PC-m1AScore 评估本胰腺中心胰腺癌患者生存期 1 年和 3 年预后的 AUC 值。

图2 胰腺癌预后风险模型 PC-m1AScore 的构建和验证

Figure 2 Construction and validation of a prognostic risk model (PC-m1AScore) for pancreatic cancer

综合提示PC-m1AScore模型有预测胰腺癌患者预后的效能。

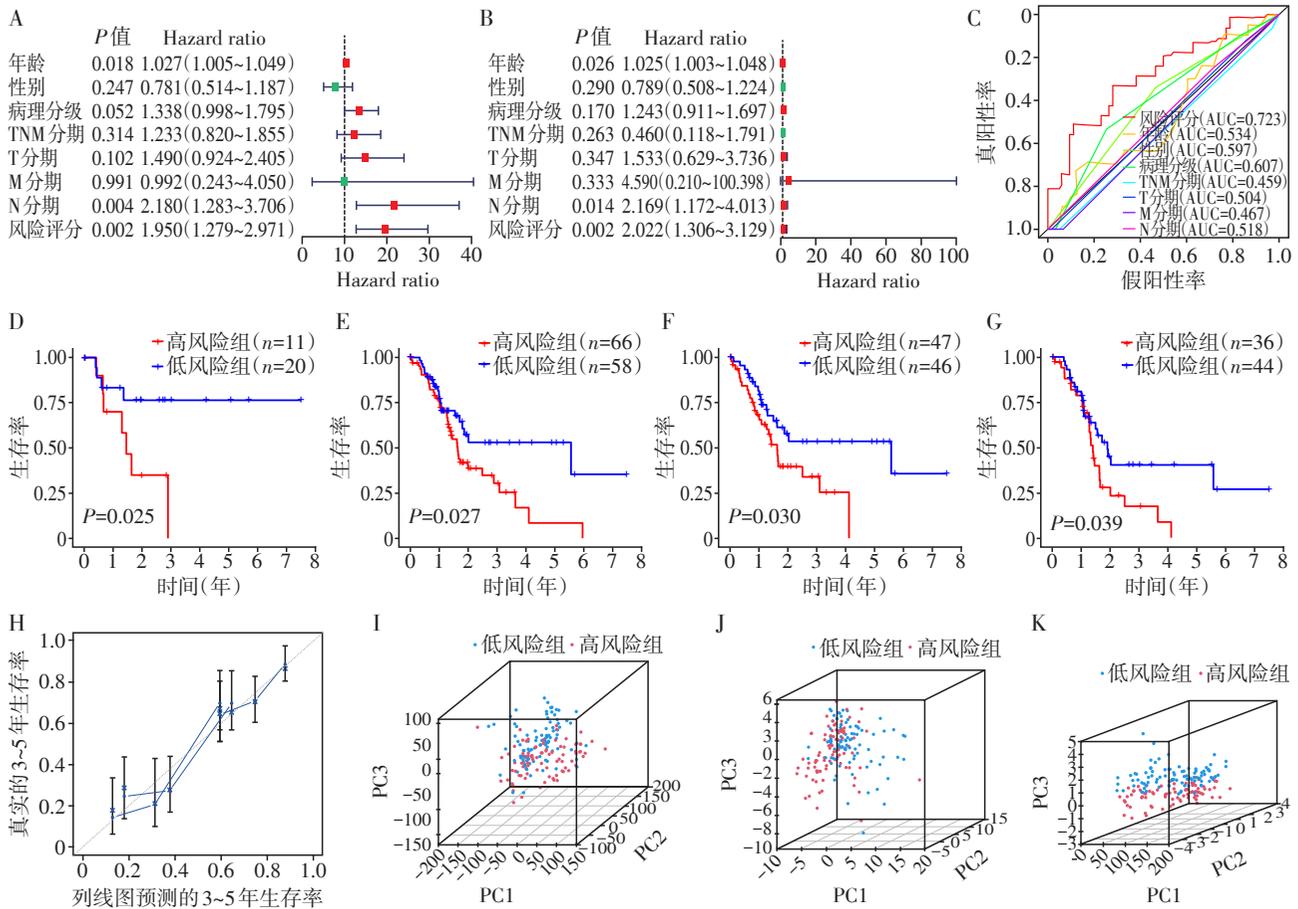
2.3 新的胰腺癌预后风险模型PC-m1AScore效能的证明

运用单因素和多因素Cox分析,代入年龄、性别、病理分级、TNM分期和PC-m1AScore风险评分多个变量指标,如图3A、B所示,多因素Cox分析结果显示年龄:HR=1.025,95%CI:1.003~1.048, $P=0.026$,性别:HR=0.789,95%CI:0.508~1.224, $P=0.290$,病理分级:HR=1.243,95%CI:0.911~1.697, $P=0.170$,TNM分期:HR=0.460,95%CI:0.118~1.791, $P=0.263$,T分期:HR=1.533,95%CI:0.629~3.736, $P=0.347$,N分期:HR=2.169,95%CI:1.172~4.013, $P=0.014$,M分期:HR=4.590,95%CI:0.210~100.398, $P=0.333$,风险评分:HR=2.022,95%CI:1.306~3.129, $P=0.002$,其中年龄、淋巴结转移(N)分期、风险评分是3个独立

的胰腺癌风险因素。如图3C的ROC分析结果显示,8个因素AUC值从小到大依次为:TNM分期(0.459)、M分期(0.467)、T分期(0.504)、N分期(0.518)、年龄(0.534)、性别(0.597)、病理分级(0.607)、风险评分(0.723),证明所建立的PC-m1AScore有预测胰腺癌预后的显著效能。此外,亚组分析结果显示,在T1~2期、G1~2期、 ≤ 65 岁患者或性别为女性的细分亚组下,PC-m1AScore风险评分能准确预测患者的生存期($P < 0.05$,图3D~G)。

为了更准确预测胰腺癌患者的生存风险,本研究构建了结合PC-m1AScore模型和临床病理学特征的预后列线图,3~5年存活率列线图的校准曲线可视化地显示预测值和实际值之间具有一致性(图3H)。

在预后分析中,多个变量通常相互关联,运用PCA主成分分析的方法在研究个体变量之间相关



A: 单因素Cox分析结果; B: 多因素Cox分析结果; C: ROC分析结果; D~G: 高、低风险组患者在不同临床亚组(D: T1~2; E: G1~2; F: 年龄 ≤ 65 岁; G: 女性)间的预后差异表现($P < 0.05$); H: 预后列线图; I~K: PCA主成分分析, 从左至右依次是以71个m1A相关基因集(I)、在胰腺癌显著差异表达的14个m1A相关基因集(J)、PC-m1AScore为权重(K)对患者进行分群。

图3 胰腺癌预后风险模型PC-m1AScore效能的验证

Figure 3 Validation of effectiveness of the prognostic risk model(PC-m1AScore) for pancreatic cancer

性的基础上,通过将数据投影到低维主成分,能降低数据的复杂性,最大限度地保留信息。结果显示:与71个m1A相关基因集和在胰腺癌显著差异表达的14个m1A相关基因集相比,应用本研究构建的胰腺癌预后风险模型PC-m1AScore,即 $(1.798\ 5 \times \text{CRLS1}) + (-1.718\ 6 \times \text{C7orf50})$,能够显著区分患者亚群(图3I~K)。

2.4 PC-m1AScore联合TMB预测胰腺癌预后的效能

TMB定义为肿瘤基因组每个编码区的突变总数,对多种肿瘤的免疫治疗具有良好的预测价值^[8]。根据178例胰腺癌患者的TMB评分对其进行分组,log-rank检验结果显示TMB水平与患者生存时间密切相关(图4A),相比于低TMB组,高TMB组具有更差的预后。Kaplan-Meier曲线也显示PC-m1AScore与TMB的联合分析具有比两者单独分析更强的预测胰腺癌预后的效能(图4B)。图4C、D显示患者的PC-m1AScore风险评分值与基因突变频率没有相关性,PC-m1AScore的高、低风险组基因突变的前10名都是KRAS>TP53>SMAD4>CDKN2A>TTN>MUC16>RNF43>HECW2>TNXB>RYR1,其中PC-m1AScore高风险组的突变频率更高(90%)。

SCNA模块提供了对给定基因具有不同体细胞拷贝数改变肿瘤之间的免疫细胞浸润水平比较。SCNA由GISTIC2.0定义,包括深度缺失(deep deletion)(-2)、臂级缺失(arm-level deletion)(-1)、二倍体/正常(0)、臂级增益(arm-level gain)(1)和高扩增(high amplification)(2)。以箱形图显示所选癌症类型中每个拷贝数状态下每个免疫子集的分布。使用双侧Wilcoxon秩和检验将每个SCNA类别的渗透水平与正常水平进行比较。结果显示CRLS1的臂级缺失与增益均会带来不同程度的B细胞、T细胞以及巨噬细胞浸润程度的改变($P < 0.05$,图4E、F)。C7orf50的体细胞拷贝数还存在高扩增改变,其臂级缺失会带来不同程度的B细胞、CD4⁺T细胞浸润的改变($P < 0.05$)。

2.5 PC-m1AScore高、低风险组之间免疫细胞浸润的差异及模型基因与免疫检查点基因表达量的相关性

根据PC-m1AScore风险评分的中位数将胰腺癌患者分为高风险组和低风险组,使用R包“Cibersort”统计分析两组之间22种免疫相关细胞类型的浸润差异,显示4种免疫细胞在两组之间存在差异($P < 0.05$,图5A),其中,中性粒细胞和CD4⁺记忆静息T细胞浸润在高风险组患者胰腺癌组织的比例更

高,而滤泡辅助T细胞与活化的NK细胞浸润在低风险组患者胰腺癌组织的比例更高。纳入PC-m1AScore预后模型的两个基因中,CRLS1的表达量与免疫检查点PDCD1的表达量呈负相关($P < 0.05$),同时与LMTK3的表达量呈正相关($P < 0.05$,图5B),而C7orf50的表达量分别与LMTK3的表达量呈正相关($P < 0.05$),与CD274和TIGIT的表达量呈负相关, ($P < 0.05$,图5C)。

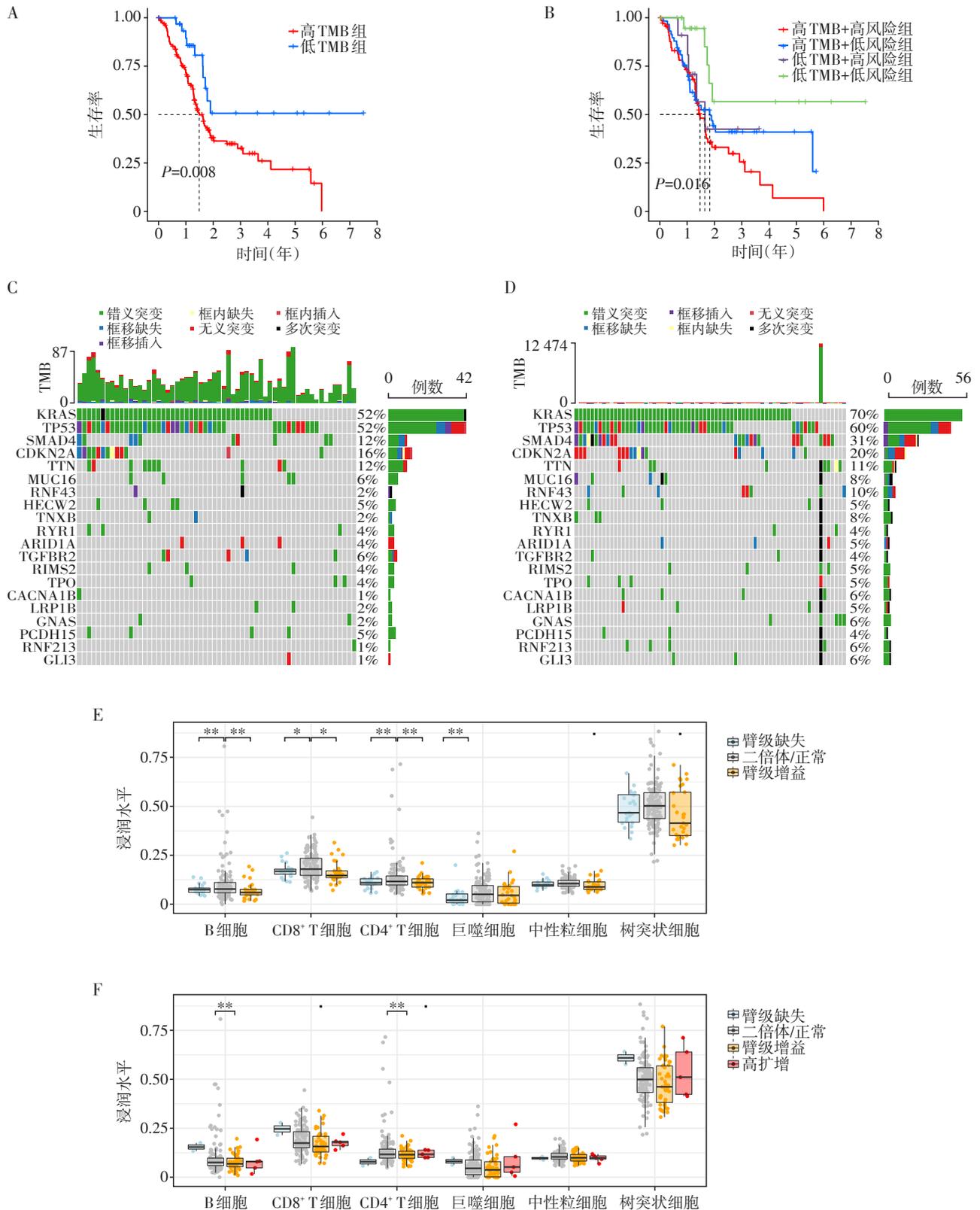
2.6 使用GEO和本中心的数据集验证PC-m1AScore预后模型的效能

除了前述来源于TCGA和GTEx的测序和临床数据以外,本研究还整理了基因表达数据库(GEO)中65例胰腺癌患者的测序和临床数据集GSE62452,使用q-PCR技术建立了本中心60例胰腺癌患者胰腺癌与癌旁非瘤组织的基因表达数据和患者生存期在内的临床数据集。使用GEO和本中心的验证数据集,根据每例患者的PC-m1AScore风险评分将患者分为高风险组和低风险组,采用log-rank检验,结果显示与低风险组相比,高风险组生存期显著缩短($P=0.020$, $P < 0.001$,图6A、B),和前述在TCGA数据库中的分析结果趋势一致,再次证明PC-m1AScore模型预测胰腺癌预后的显著效能。

采用log-rank检验,使用TCGA的数据集,组成PC-m1AScore模型的两个权重基因的表达量也都有预测胰腺癌预后的效能($P < 0.01$,图6C、D),其中高表达C7orf50基因的胰腺癌患者生存期显著延长($P < 0.001$),而高表达CRLS1基因的胰腺癌患者生存期显著缩短($P=0.007$)。60例癌与癌旁非瘤的配对组织检测结果显示CRLS1与C7orf50在胰腺癌组织中高表达(图6E、F)。根据随访的统一截止时间点,将本中心来源的患者分为生存组与死亡组,分别统计两个模型基因表达情况,结果显示,CRLS1在死亡组中的表达量显著升高,C7orf50在死亡组中的表达量显著降低($P < 0.05$,图6G)。综合图6C~G的结果,均和PC-m1AScore预测胰腺癌预后的趋势吻合,不仅再次证明PC-m1AScore模型预测胰腺癌预后的显著效能,而且提示组成PC-m1AScore模型的两个权重基因可能与胰腺癌的恶性进展有关系。

3 讨论

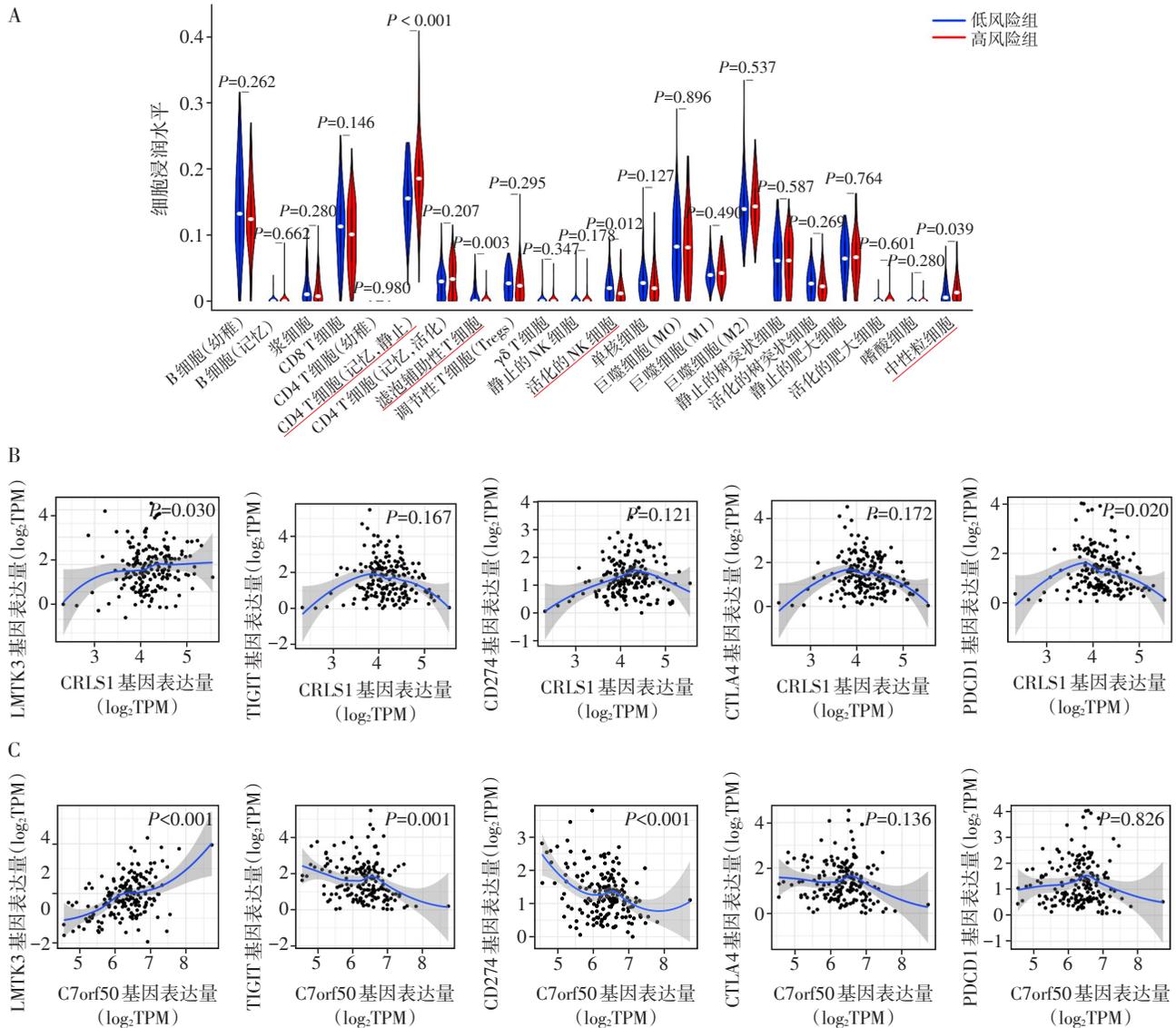
本研究通过TCGA-GTEx数据库分析胰腺癌组织中差异表达的m1A相关基因,利用Cox分析与



A: 胰腺癌患者TMB水平与生存时间的关系(log-rank 检验); B: PC-m1AScore联合TMB预测胰腺癌预后的效能; C: PC-m1AScore低风险组的基因突变情况; D: PC-m1AScore高风险组的基因突变情况; E: CRLS1基因拷贝数状态与免疫子集细胞浸润水平的关系; F: C7orf50基因拷贝数状态与免疫子集细胞浸润水平的关系; 两组比较, * $P < 0.05$, ** $P < 0.01$ 。

图4 PC-m1AScore模型及其权重基因与基因突变和免疫细胞浸润水平的关系

Figure 4 The relation of PC-m1AScore model and the levels of gene mutation and immune cell infiltration



A: 22种免疫细胞在PC-m1AScore高、低风险组胰腺癌组织中的浸润差异; 带下划线为高、低风险组间差异有统计学意义的4种免疫细胞; B、C: CRLS1(B)、C7orf50(B)表达量与免疫检查点基因表达量的相关性分析。

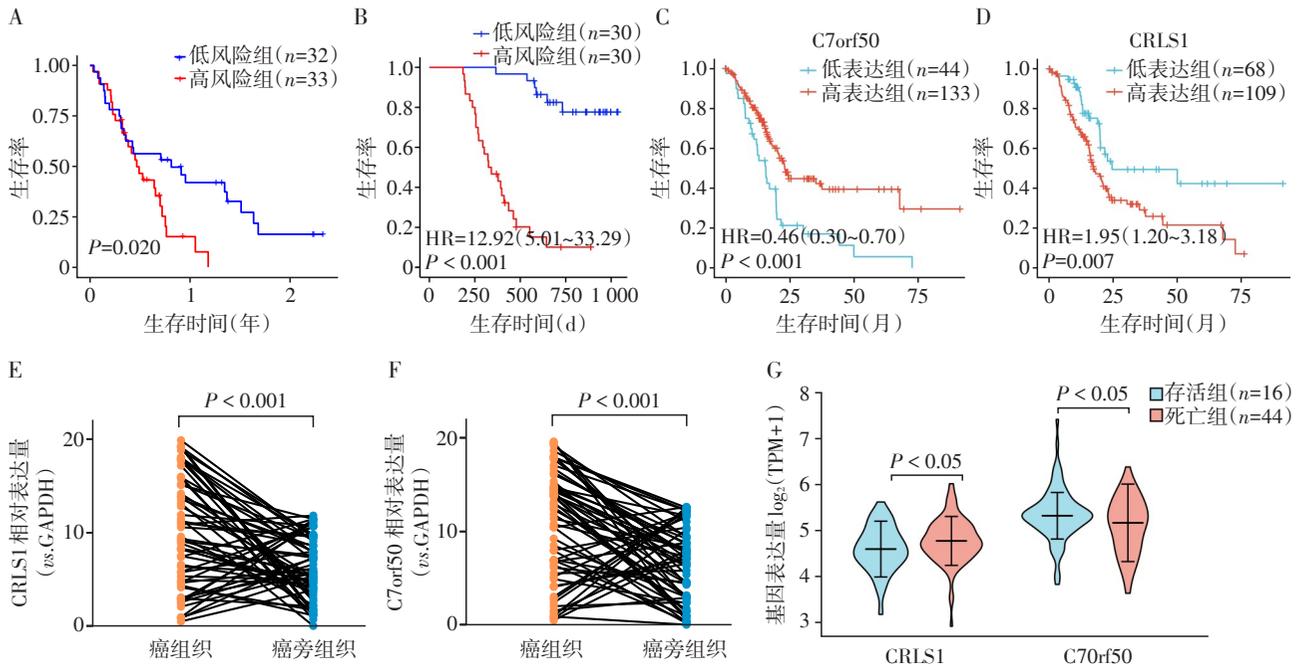
图5 PC-m1AScore高低风险组的免疫细胞浸润差异以及模型基因与免疫检查点基因表达量的相关性

Figure 5 Differences in immune infiltration between high and low risk groups and correlations between model genes and immune checkpoints

LASSO回归构建胰腺癌预后风险模型,并通过GEO数据库和本中心胰腺癌病例的基因表达和临床数据进行系统验证,找到了能预测胰腺癌患者预后的新工具,期待能转化应用于临床,助力于对胰腺癌患者的个性化监测、治疗和预后改善。

越来越多的证据表明RNA化学修饰在基本细胞过程中具有重要功能,如细胞分化、蛋白质产生、细胞信号转导和昼夜节律的维持等^[9]。已经证实,m1A失调影响多种细胞过程,包括RNA结构稳定性、折叠、与蛋白质的相互作用、细胞活力、自我更新能力受损、细胞增殖和细胞死亡^[10]。最近研究表

明,m1A修饰广泛参与许多疾病的发生和进展,例如去甲基化修饰酶ALKBH3在许多人类癌症中高度表达,与癌细胞中tRNA的m1A修饰和蛋白质合成密切相关^[11-12],甲基转移酶TRMT6重复序列中的移码突变已在结肠癌中得到证实。Gao等^[2]提出了由71个m1A相关基因组成的集群,在该集群基础上建立的m1AScore算法能预测结肠癌的预后。本研究也通过GO功能富集分析发现,这71个m1A相关基因主要富集在细胞内物质转运和分泌、DNA复制、RNA转录、RNA降解等功能上,这些通路在许多生物过程中都至关重要,稍有调控异常便可能导致



A: 利用GSE62452数据集验证PC-m1AScore的效能;B: 利用本中心数据集验证PC-m1AScore的效能;C: 利用TCGA数据集验证C7orf50预测胰腺癌预后的效能;D: 利用TCGA数据集验证CRLS1预测胰腺癌预后的效能;E、F: CRLS1和C7orf50在本中心60例患者癌组织与癌旁组织的表达差异情况;G: CRLS1与C7orf50在本中心患者以是否存活状态为分组下的差异表达情况。

图6 使用GEO和本中心的数据集验证PC-m1AScore预后模型的效能

Figure 6 The efficacy of the PC-m1AScore prognostic model was validated using GEO and our data set

疾病的发生。KEGG通路富集提示m1A相关基因主要富集在错配修复通路,是维持基因组稳定和避免突变的重要组成部分。

考虑到如果将71个基因的表达量都纳入风险预测模型,基因数量过大会限制临床转化应用,本研究尝试对基因集进行降维;通过对71个m1A相关基因在胰腺癌与正常胰腺组织中的表达分析,筛选出显著上调与下调的基因各7个,本研究结果也显示所筛选的14个基因彼此间相关性极高,提示14个基因很可能彼此协同配合发挥功能。在此基础上,通过单因素、LASSO、多因素Cox回归分析,找到了2个m1A相关基因(CRLS1与C7orf50)有独立预测胰腺癌预后的能力,并用两个基因的表达水平与其权重的乘积后求和建模,构建了新的胰腺癌预后风险模型,命名为PC-m1AScore。PCA主成分分析显示,该风险模型可以精准区分患者人群为高、低风险两个亚群,明显优于全面审视所有71个m1A相关基因。本研究通过包括log-rank检验、Cox回归分析、ROC曲线、AUC面积、列线图等在内的多种统计方法,使用TCGA、GTEX、GEO、本中心多种来源的胰腺癌病例的基因表达数据和对应的临床数据,系统证明了PC-m1AScore模型预测胰腺癌预后的效

能。值得注意的是,本研究证明了PC-m1AScore与TMB联合分析具有比两者单独分析更强的区分胰腺癌预后的效能。

此外,本研究在用PC-m1AScore进行生物学层面的免疫相关分析时发现:中性粒细胞和CD4⁺记忆静息T细胞浸润在PC-m1AScore高风险组患者胰腺癌组织中的比例更高,而滤泡辅助T细胞与活化的NK细胞浸润在低风险组患者胰腺癌组织中的比例更高。第一,众所周知,中性粒细胞的主要效应机制是构成抵御外来入侵者的第一道防线:吞噬作用、脱粒和中性粒细胞胞外陷阱形成。值得注意的是,最新的文献综述表明,中性粒细胞被募集到肿瘤微环境中并发挥炎症和肿瘤促进作用^[13-15]。作为肿瘤微环境的重要组成部分,中性粒细胞在胰腺癌中发挥着多种作用,如促进血管生成、恶性进展、转移和免疫抑制。据此,结合本研究的结果,我们推测胰腺癌细胞招募更多的中性粒细胞,可能是PC-m1AScore高风险组患者预后更差的原因之一。第二,本研究通过关键词CD4⁺记忆静息T细胞(CD4⁺ memory resting T cells)检索文献,迄今,国内外未见CD4⁺记忆静息T细胞在胰腺癌中的功能研究报告,其在肿瘤中的研究报告也很少,主要集中在

在人类免疫缺陷病毒感染方面。虽然学术界认为由CD4⁺T细胞介导的长期免疫记忆针对先前遇到的病原体或抗原提供了快速保护,但是,记忆CD4⁺T细胞是如何产生和维持的仍然存在争议^[16],本研究利用生物信息学发现PC-m1AScore高风险组患者的胰腺癌组织中有更多的CD4⁺记忆静息T细胞,也需要进一步设计生物学实验,验证和阐明观察到的现象。第三,最新研究表明,滤泡辅助T(Tfh)细胞为B细胞提供帮助,支持生发中心的形成,从而使抗体反应的亲和力成熟^[17]。Tfh尽管通常位于次级淋巴器官中,但具有Tfh细胞特征的T细胞也可以存在于人体血液中,并且它们的频率和表型在自身免疫性疾病患者中经常发生改变,该群体可能是免疫失调的核心哨兵。已有文献证实,Tfh细胞的高浸润预示着胰腺癌患者的预后较好^[18]。Tfh细胞通过分泌CXCL13和IL-21减轻胰腺癌的肿瘤生长并增强肿瘤微环境中CD8⁺T细胞、B细胞浸润和B细胞成熟。据此,结合本研究的结果,我们推测Tfh浸润更多,可能是PC-m1AScore低风险组患者预后较好的原因之一。第四,NK细胞是肿瘤免疫中重要的效应细胞,在机体内对肿瘤细胞起着免疫监视作用。已有研究证实,NK浸润不仅是预后较好的标志,NK也是对抗转移的效应群体^[19],而且NK细胞治疗可协助吉西他滨化疗抑制胰腺癌手术治疗后的局部复发,在PD-1阻断剂和吉西他滨均不能抑制远处转移的情况下,通过系统激活NK细胞可有效抑制包括肝转移在内的胰腺癌的远端转移^[20]。因此,结合本研究结果,推测NK细胞浸润更多,可能是PC-m1AScore低风险组患者预后较好的原因之一。

PC-m1AScore模型中包含两个权重基因,分别是与风险正相关的CRLS1和与风险负相关的C7orf50。CRLS1编码CDP-酒精磷脂酰转移酶I类蛋白质家族的成员。编码的酶催化心磷脂的合成,心磷脂是线粒体膜的磷脂成分,对线粒体功能至关重要^[21]。此外,CRLS1基因在癌症、心血管疾病等多种疾病中也扮演着重要角色,因此,CRLS1可能成为一种重要的治疗靶点或生物标志物^[22]。已有研究探讨了CRLS1在三阴性乳腺癌中的作用,发现CRLS1高表达促进了肿瘤进展,并赋予细胞对顺铂的耐药性^[23]。另有研究发现,CRLS1基因中的一些有害变异导致心磷脂严重缺乏,进而引起一种名为自身隐性多系统线粒体病的疾病^[24],患者表现出多个系统的症状,包括肌肉无力、眼肌麻痹、心脏问题

和神经系统障碍等^[25]。C7orf50基因产物也称为PARK16(因其与帕金森病的关联而命名),目前的研究较为有限,对其功能了解不多。C7orf50在大脑组织中高度表达,尤其是在帕金森病患者的脑组织中^[26]。C7orf50的功能尚不完全清楚,有报道它可能与细胞凋亡、线粒体功能和氧化应激有关^[27]。C7orf50蛋白可能与线粒体外膜上的磷脂酰肌醇磷酸相互作用,这可能与线粒体膜上的离子通道或其他蛋白质的功能调节有关^[28]。本研究结果提示CRLS1与C7orf50的功能可能与胰腺癌的恶性进展有关,需要开展系统的体外、体内实验与深入的机制研究来进一步证实。

[参考文献]

- [1] OWENS D K, DAVIDSON K W, KRIST A H, et al. Screening for pancreatic cancer: US preventive services task force reaffirmation recommendation statement [J]. *JAMA*, 2019, 322(5):438-444
- [2] GAO Y, WANG H, LI H, et al. Integrated analyses of m1A regulator-mediated modification patterns in tumor microenvironment-infiltrating immune cells in colon cancer[J]. *Oncoimmunology*, 2021, 10(1):1936758
- [3] CHEN C, YE L L. The m1A modification of tRNAs: a translational accelerator of T-cell activation[J]. *Cell Mol Immunol*, 2022, 19(12):1328-1329
- [4] LI D, LI K, ZHANG W, et al. The m6A/m5C/m1A regulated gene signature predicts the prognosis and correlates with the immune status of hepatocellular carcinoma [J]. *Front Immunol*, 2022, 13:918140
- [5] THISSEN D, STEINBERG L, KUANG D. Quick and easy implementation of the benjamini-hochberg procedure for controlling the false positive rate in multiple comparisons [J]. *J Educ Behav Stat*, 2002, 27(1):77-83
- [6] XU F, DU W Q, ZOU Q, et al. COPII mitigates ER stress by promoting formation of ER whorls[J]. *Cell Res*, 2021, 31(2):141-156
- [7] BATEMAN A C. DNA mismatch repair protein immunohistochemistry - an illustrated guide [J]. *Histopathology*, 2021, 79(2):128-138
- [8] SHA D, JIN Z H, BUDCZIES J, et al. Tumor mutational burden as a predictive biomarker in solid tumors [J]. *Cancer Discov*, 2020, 10(12):1808-1825
- [9] ZHU X, ZHOU C, ZHAO S, et al. Role of m6A methylation in retinal diseases [J]. *Exp Eye Res*, 2023, 231:109489
- [10] ZHAO B S, ROUNDTREE I A, HE C. Post-transcriptional gene regulation by mRNA modifications [J]. *Nat Rev*

- Mol Cell Biol,2017,18(1):31-42
- [11] CHEN Z, QI M, SHEN B, et al. Transfer RNA demethylase ALKBH3 promotes cancer progression via induction of tRNA - derived small RNAs [J]. *Nucleic Acids Res*, 2019,47(5):2533-2545
- [12] 时国东,陈 群,胡青桥,等. 基于TCGA数据库挖掘胰腺癌预后相关甲基化位点[J]. *南京医科大学学报(自然科学版)*,2020,40(8):1135-1139
- [13] MITSIS M, DROSOU P, TATSIS V, et al. Neutrophil extracellular traps and pancreatic cancer development: a vicious cycle[J]. *Cancers(Basel)*,2022,14(14):3339
- [14] JIANG W, LI X, XIANG C, et al. Neutrophils in pancreatic cancer: potential therapeutic targets [J]. *Front Oncol*, 2022,12:1025805
- [15] DEMKOW U. Neutrophil extracellular traps (NETs) in cancer invasion, evasion and metastasis [J]. *Cancers*, 2021,13(17):4495
- [16] HOJYO S, TUMES D, MURATA A, et al. Multiple developmental pathways lead to the generation of CD4 T-cell memory[J].*Int Immunol*,2020,32(9):589-595
- [17] WALKER L S K. The link between circulating follicular helper T cells and autoimmunity [J]. *Nat Rev Immunol*, 2022,22(9):567-575
- [18] LIN X, YE LY, WANG X, et al. Follicular helper T cells remodel the immune microenvironment of pancreatic cancer via secreting CXCL13 and IL-21[J]. *Cancers(Basel)*, 2021,13(15):3678
- [19] PIPER, M., VAN COURT, et al. Targeting treg-expressed STAT3 enhances NK-mediated surveillance of metastasis and improves therapeutic response in pancreatic adenocarcinoma[J]. *Clin Cancer Res*,2022,28(5):1013-1026
- [20] BROOKS J, FLEISCHMANN-MUNDT B, WOLLER N, et al. Perioperative, spatiotemporally coordinated activation of T and NK cells prevents recurrence of pancreatic cancer[J]. *Cancer Res*,2018,78(2):475-488
- [21] LEE R G, BALASUBRAMANIAM S, STENTENBACH M, et al. Deleterious variants in CRLS1 lead to cardiolipin deficiency and cause an autosomal recessive multi-system mitochondrial disease[J]. *Hum Mol Genet*,2022,31(21):3597-3612
- [22] BIDKHORI G, BENFEITAS R, ELMAS E, et al. Metabolic network-based identification and prioritization of anti-cancer targets based on expression data in hepatocellular carcinoma[J]. *Front Physiol*,2018,9:916
- [23] TU C Y, XIONG H, HU Y F, et al. Cardiolipin synthase 1 ameliorates NASH through activating transcription factor 3 transcriptional inactivation [J]. *Hepatology*, 2020,72(6):1949-1967
- [24] MAEKAWA S, TAKADA S, NAMBU H, et al. Linoleic acid improves assembly of the C II subunit and C III 2/CIV complex of the mitochondrial oxidative phosphorylation system in heart failure[J]. *Cell Commun Signal*,2019,17(1):128
- [25] GIANNI' M, GORACCI L, SCHLAEFLI A, et al. Role of cardiolipins, mitochondria, and autophagy in the differentiation process activated by all - trans retinoic acid in acute promyelocytic leukemia[J]. *Cell Death Dis*, 2022,13(1):30
- [26] AMMOUS F, ZHAO W, RATLIFF S M, et al. Epigenome-wide association study identifies DNA methylation sites associated with target organ damage in older african americans[J]. *Epigenetics*,2021,16(8):862-875
- [27] BARFIELD R, WANG H M, LIU Y M, et al. Epigenome-wide association analysis of daytime sleepiness in the multi-ethnic study of atherosclerosis reveals African-American-specific associations[J]. *Sleep*,2019,42(8): zsz101
- [28] PENG C, CARDENAS A, RIFAS-SHIMAN S L, et al. Epigenome-wide association study of total serum immunoglobulin E in children: a life course approach [J]. *Clin Epigenetics*,2018,10:55

[收稿日期] 2023-04-24

(责任编辑:蒋 莉)