

胃癌组蛋白 H3 第 4 位赖氨酸的三甲基化修饰相关长链非编码 RNA 预后模型的构建与验证

胡震¹, 祁玉忠¹, 赵绍基², 王光熙¹, 孙开宇², 吴文辉^{1*}

1. 中山大学附属第七医院 消化医学中心, 广东 深圳 518107

2. 中山大学附属第一医院 胃肠外科, 广东 广州 510080

【摘要】 目的 探索胃癌组蛋白 H3 第 4 位赖氨酸的三甲基化 (trimethylation of lysine 4 on histone H3, H3K4me3) 修饰相关长链非编码 RNA (long non-coding RNA, LncRNA) 特征, 构建相关预后模型并预测胃癌免疫治疗疗效。方法 从癌症基因组图谱数据库下载胃癌相关转录组测序数据和对应的患者临床资料, 通过构建 H3K4me3 相关调节因子基因与 LncRNA 的共表达网络识别 H3K4me3 修饰相关 LncRNA, 并将癌症基因组图谱数据库中 370 例符合筛选标准的胃癌患者样本 (整体组) 按 1:1 随机抽样划分为训练组 ($n=185$) 和验证组 ($n=185$)。随后基于单因素 Cox 回归、Lasso 回归分析构建 H3K4me3 相关 LncRNA 预后风险评分模型并进行内部验证。Kaplan-Meier 生存分析和受试者操作特征曲线 (receiver operating characteristic curve, ROC 曲线) 被用于验证模型的预测性能。通过单因素和多因素 Cox 回归分析评估风险评分等临床指标的预后预测价值。结合风险评分、年龄和肿瘤 TNM 分期构建预测胃癌患者总生存率的列线图模型, ROC 曲线与校准图被用于评估列线图的预测准确性。借助共识聚类识别异质性聚类亚群并进行免疫治疗疗效预测。结果 基于共表达网络关系识别了 14 个具有预后价值的 H3K4me3 相关 LncRNA 并构建了相关风险模型及评价体系。根据训练组预后风险评分模型获得的中位风险评分将训练组、验证组和整体组胃癌患者划分为高、低风险, Kaplan-Meier 生存曲线显示低风险患者的总体生存情况要优于高风险患者 ($P<0.05$)。此外, 该模型在训练组中预测胃癌患者 1、3、5 年总生存率的曲线下面积 (area under curve, AUC) 分别为 0.708、0.730、0.770, 在验证组中分别为 0.690、0.648、0.713, 而在整体组中分别为 0.697、0.670、0.724。多因素 Cox 回归分析显示基于 H3K4me3 相关 LncRNA 构建的风险评分模型是预测胃癌患者预后的独立因素 ($P<0.001$)。构建的列线图预测胃癌患者 1、2、3 年总生存率的 AUC 分别为 0.727、0.780、0.717, 且其校准曲线与理想曲线相接近。基于共识聚类算法进一步识别了 2 种具有异质性免疫特征的 H3K4me3-LncRNA 亚群, 其中亚组 II 具有更高的免疫细胞浸润水平和更强的免疫应答潜力, 并对 5-氟尿嘧啶、奥沙利铂等显示出更高的药物敏感性, 而亚组 I 则可能对磷脂酰肌醇 3-激酶特异性抑制剂具有更高的敏感性。结论 本研究构建了一个 H3K4me3-LncRNA 风险评分模型以预测胃癌患者的预后, 并揭示了其异质性微观特征及在预测免疫治疗疗效方面的潜在价值。

【关键词】 胃癌; 组蛋白 H3 第 4 位赖氨酸的三甲基化修饰; 组蛋白修饰; 长链非编码 RNA; 免疫治疗

Construction and verification of trimethylation modification of lysine 4 on histone H3 related long non-coding RNA prognostic model for gastric cancer

Hu Zhen¹, Qi Yuzhong¹, Zhao Shaoji², Wang Guangxi¹, Sun Kaiyu², Wu Wenhui^{1*}

1. Gastroenterology Center, the Seventh Affiliated Hospital of Sun Yat-sen University, Shenzhen 518107, Guangdong, China

2. Department of Gastrointestinal Surgery, the First Affiliated Hospital of Sun Yat-sen University, Guangzhou

基金项目: 国家自然科学基金 (82203642); 深圳市医疗卫生“三名工程” (SZSM201911010); 广州市科技计划项目 (2023A04J2212)

* 通信作者: 吴文辉, E-mail: doctorwusysu@126.com

510080, Guangdong, China

*Corresponding author: Wu Wenhui, E-mail: doctorwusysu@126.com

【Abstract】 Objective To explore the characteristics of long non-coding RNA (LncRNA) associated with trimethylation modification of lysine 4 on histone H3 protein (H3K4me3) in gastric cancer (GC), construct a related prognostic model and predict the efficacy of immunotherapy for GC. **Method** The RNA transcriptome sequencing data and clinical information of GC patients were downloaded from the cancer genome atlas database, and the H3K4me3 modification related LncRNA was identified by constructing a co-expression network of H3K4me3-related regulatory factor genes and LncRNA. In addition, 370 GC patients' samples (integral set) from the cancer genome atlas database meeting the screening criteria were randomly divided into training set ($n=185$) and verification set ($n=185$) according to 1:1 random sampling. Subsequently, the H3K4me3-related LncRNA prognostic risk score model was constructed based on univariate Cox regression and Lasso regression analysis, and internal verification was conducted. Kaplan-Meier survival analysis and receiver operating characteristic curve (ROC curve) were used to verify the predictive performance of the model. Univariate and multivariate Cox regression analysis were used to evaluate the predictive prognosis value of clinical indicators such as risk score. Combined with risk score, age and tumor TNM stage, a nomogram model was constructed to predict the overall survival rates of GC patients. The predictive accuracy of the nomogram was assessed by applying ROC curve and calibration curves. The heterogeneous cluster subsets were identified by consensus clustering and the therapeutic effect of immunotherapy was predicted. **Result** Based on the co-expression network relationship, 14 H3K4me3-related LncRNA with prognostic value were identified and the related risk model and evaluation system were constructed. According to the median risk score obtained by the training set's prognostic risk score model, GC patients in the training set, verification set and integral set were divided into high risk and low risk. The overall survival curves of GC patients with low-risk drawn by Kaplan-Meier method were better than those of the GC patients with high-risk in the training set, validation set, and integral set ($P<0.05$). The area under curve (AUC) of the prognostic model for predicting the 1-year, 3-year and 5-year overall survival rates of GC patients in the training set were 0.708, 0.730 and 0.770 respectively; in the verification set, they were 0.690, 0.648 and 0.713, respectively; and in the integral set, they were 0.697, 0.670 and 0.724, respectively. Multivariate Cox regression analysis confirmed that the risk score model based on H3K4me3 modification related LncRNA was an independent factor for predicting the prognosis of GC patients ($P<0.001$). The AUC of the constructed nomogram for predicting the 1-year, 2-year and 3-year overall survival rates of GC patients were 0.727, 0.780 and 0.717, respectively, and the calibration curves were close to the ideal curves. Based on consensus clustering algorithm, 2 H3K4me3-LncRNA subsets with heterogeneous immune characteristics were further identified, of which subgroup II had higher levels of immune cell infiltration and stronger immune response potential, and showed higher drug sensitivity to 5-fluorouracil and oxaliplatin, whereas subgroup I was likely to be more sensitive to phosphatidylinositol 3-kinase specific inhibitors. **Conclusion** In this study, we constructed an H3K4me3-LncRNA risk score model to predict the prognosis of patients with GC, and revealed its heterogeneous microscopic characteristics and potential value in predicting the efficacy of immunotherapy.

【Key words】 Gastric cancer; Trimethylation of lysine 4 on histone H3; Histone modification; Long non-coding RNA; Immunotherapy

胃癌是起源于胃黏膜上皮细胞且极具侵袭性的恶性肿瘤,位居肿瘤发病及相关死亡谱的前列,具有侵袭性强、恶性程度高、易复发转移等特性^[1-2]。生物性状是基因和环境共同作用的结果,通过特定的转录程序实现基因组DNA序列的精准传递并受到包括转录因子和表观遗传修饰在内的多层次因素的调节^[3-4]。组蛋白修饰作为表观遗传

修饰的重要组成部分,其可通过调节组蛋白上氨基酸残基的甲基化、乙酰化、泛素化等修饰来参与转录因子调控,进而影响靶蛋白的募集和功能的呈现^[5]。组蛋白H3第4位赖氨酸的三甲基化(trimethylation of lysine 4 on histone H3, H3K4me3)是在活性基因上发现的一种特殊的甲基化标志,通常由甲基转移酶、去甲基化酶和结合蛋白直接

调节,其在增强子或启动子区域的富集可促进靶基因的转录进程^[6]。既往研究多认为H3K4me3修饰在调节基因转录起始阶段发挥作用,但近来也有研究表明H3K4me3可通过动态调节启动子近端蛋白质暂停-释放进程来调节基因表达^[7]。长链非编码RNA(long non-coding RNA, lncRNA)是重要的表观遗传调节因子,在肿瘤发生、转移、耐药等病理过程中发挥着至关重要的作用^[8-9]。lncRNA SATB2-AS1可通过诱导启动子区域靶蛋白的募集来促进H3K4me3修饰以激活SATB2转录^[10]。也有研究报道lnc_3712可抑制KDM5B介导的H3K4me3核编程^[11]。这些报道提示lncRNA可通过直接或间接方式调节H3K4me3修饰进而干扰生物体内的细胞信号转导。然而目前尚未有研究系统阐述H3K4me3相关lncRNA在调控胃癌病理进程中的潜在机制和预后价值。因此,本文旨在利用生物信息学工具筛选H3K4me3相关lncRNA并进行预后模型的构建,以期对胃癌免疫治疗和预后预测提供有效帮助。

1 材料与方法

1.1 数据获取

从癌症基因组图谱(the cancer genome atlas, TCGA)(<https://portal.gdc.cancer.gov/>)数据库检索获得了412例胃癌和36例正常组织的转录组测序数据及相关临床资料,并进行初步筛选,过滤低丰度基因(TPM<0.2)及随访时间少于30d的样本,最终370例胃癌患者样本数据被纳入后续分析。通过癌症免疫组图谱(the cancer immunome atlas, TCIA)(<https://tcia.at/home>)数据库获取胃癌免疫表型评分(immunophenoscore, IPS)以预测其对免疫检查点阻断治疗[包括细胞毒性T淋巴细胞相关蛋白4(cytotoxic T lymphocyte-associated antigen-4, CTLA-4)和程序性死亡受体1(programmed death-1, PD-1)抑制剂]的应答疗效,IPS与肿瘤微环境中免疫细胞的浸润程度和免疫应答效应呈正相关。

1.2 H3K4me3相关lncRNA的筛选

通过检索文献资料获得了29个H3K4me3相关调节因子基因,其中包括7个甲基转移酶基因(KMT2A、KMT2B、KMT2C、KMT2D、SETD2、SETD7、EZH2)、9个去甲基化酶基因(KDM1A、KDM2A、KDM2B、KDM4A、KDM5A、KDM5B、KDM5C、KDM5D、

KDM7A)和13个结合蛋白基因(SPIN1、CXXC1、C17orf49、PHF23、ZCWPW1、BRD4、SPP1、PHF13、ZMYND8、ING1、ING2、ING4、PHF7)^[12-14]。随后运用“limma”R包基于Pearson相关性分析筛选与H3K4me3显著相关的lncRNA,筛选标准为|Pearson R|>0.5且P<0.001。

1.3 H3K4me3相关lncRNA预后模型构建

使用“caret”R包将370例胃癌患者样本(整体组)按照1:1随机划分为训练组(n=185)和验证组(n=185)。其中训练组用于构建预后模型,而验证组和整体组则用于模型验证。随后通过单因素Cox回归分析筛选出与胃癌预后相关的H3K4me3-lncRNA,并在训练组中进行Lasso回归分析,重复1000次10倍交叉验证,并以P<0.05作为筛选标准识别关键lncRNA并构建预后模型。计算每例胃癌患者的风险评分:风险评分= $\sum_{i=1}^n \text{Exp}_{\text{lncRNA}_i} \times \text{Coef}_{\text{lncRNA}_i}$,其中n为lncRNA数量,Exp_{lncRNA_i}为对应lncRNA的表达量,Coef_{lncRNA_i}为相关系数。目标lncRNA分别为:RRN3P2、CHROMR、BNC2-AS1、AC018752.1、LINC01932、AC131391.1、LINC01140、AC004596.1、AC068790.7、AC009318.2、TMEM75、BX890604.1、AC004522.5、Z73420.1。

1.4 基因互作网络的构建

将H3K4me3相关调节因子基因与筛选获得的lncRNA及相互作用关系(|Pearson R|>0.5且P<0.001)读入Cytoscape软件(版本3.9.1)构建相关基因互作网络图。

1.5 预后模型生存分析

基于训练组的中位风险评分值,将胃癌患者划分为高、低风险。通过运用热图、风险曲线、散点图和Kaplan-Meier生存曲线实现对风险亚群总体生存状态及风险评分的刻画。“survival”“survminer”以及“pheatmap”R包用于相关图形的可视化。

1.6 预后模型有效性验证

使用受试者操作特征曲线(receiver operating characteristic curve, ROC曲线)和曲线下面积(area under curve, AUC)进一步预测训练组、验证组以及整体组的1、3、5年总生存率并验证模型的预后预测价值。时间依赖的ROC曲线的AUC由“timeROC”R包计算获得。AUC值通常介于0.5~1之间,其值越大表明模型分类性能越佳,区分能力越好。

1.7 独立预后分析

通过单因素和多因素Cox回归分析识别包括

风险评分(赋值:低风险=0,高风险=1)、年龄(连续变量,赋值从小到大)、性别(赋值:女=0,男=1)、肿瘤分级(赋值: $G_1=0, G_2=1, G_3=2, G_x=4$)、肿瘤TNM分期(赋值:I期=0,II期=1,III期=2,IV期=3,X期=4)、T分期(赋值: $T_1=0, T_2=1, T_3=2, T_4=3, T_x=4$)、N分期(赋值: $N_0=0, N_1=1, N_2=2, N_3=3, N_x=4$)、M分期(赋值: $M_0=0, M_1=1, M_x=2$)在内的临床指标的独立预后预测价值,单因素和多因素Cox回归分析中 P 值均小于0.05的因素被判定为具有独立预后预测价值的指标。

1.8 列线图的构建

为扩展预后模型的实用性和可及性,结合包括风险评分在内的具有独立预后预测价值的指标,借助“rms”和“survival”R包构建具有预后意义的列线图模型,并预测胃癌患者的1、2、3年总生存率。ROC曲线与校准图被用于进一步评估列线图的精准性并验证预测结果与实际结果的符合程度。

1.9 基于H3K4me3-LncRNA的共识聚类

共识聚类是一种基于重复抽样获取最佳集群数目的无监督聚类方法。本研究基于筛选获得的H3K4me3-LncRNA表达量,使用R语言工具运行“ConsensusClusterPlus”命令,重复1000次进行共识聚类(最大集群数设置为9且聚类一致性得分>0.8)以探索胃癌潜在的分子亚型,并借助主成分分析进行降维、聚类以验证不同分子亚型的区分度。最佳聚类数 k 则通过评估不同 k 值下的累积分布函数获得,当 $k=2$ 时聚类效果最佳,此时胃癌患者样本被重新划分为2个异质性亚群,即亚组I($n=167$)和亚组II($n=203$)。

1.10 药物敏感性分析

使用“oncoPredict”R包拟合TCGA胃癌样本的基因表达谱和来自癌症药物敏感性基因组学(genomics of drug sensitivity in cancer, GDSC)(<https://www.cancerrxgene.org/>)数据库的潜在联合免疫治疗药物的半抑制浓度(half maximal inhibitory concentration, IC50),并通过非配对 t 检验对不同聚类亚群的药物敏感性进行比较,以 $P<0.001$ 为显著性的阈值。

1.11 统计学方法

本研究涉及的所有统计学分析及可视化均基于R软件(版本4.3.1)和SPSS 26进行,且使用的软件包均在相应方法学部分进行了描述。非配对 t

检验用于不同聚类亚群间药物敏感性和IPS的比较;Wilcox.test函数则被用于进行Mann Whitney U 检验以比较不同聚类亚群间的免疫细胞浸润特征差异。使用Kaplan-Meier曲线对不同组别进行生存分析。所有检验均为双侧, $P<0.05$ 为差异有统计学意义。

2 结果

2.1 H3K4me3相关LncRNA的筛选

提取胃癌转录组中29个H3K4me3调节因子基因和16302个LncRNA的表达量进行共表达分析,并筛选 $|Pearson R|>0.5$ 且 $P<0.001$ 的LncRNA,最终识别了291个H3K4me3功能相关的LncRNA用于后续分析(包括PHF23、ZCWPW1和ING2在内的3个H3K4me3调节因子基因与LncRNA的共表达分析结果不符合上述筛选条件,因而未纳入作后续分析),其共表达关系通过桑基图进行了可视化(图1A)。随后基于单因素Cox回归分析确定了29个具有预后价值的LncRNA(图1B)。然后将370例胃癌患者样本按1:1随机抽样划分为训练组($n=185$)和验证组($n=185$)。在训练组中,基于Lasso回归生成的惩罚函数对纳入筛选的LncRNA进行变量压缩,逐步回归消除多重共线性变量,避免过度拟合,最后确定了14个具有预后价值的H3K4me3相关LncRNA(图1C、1D和表1)。通过构建基因互作网络,明确了14个目标LncRNA与H3K4me3调节因子间的潜在关联(图1E)。

2.2 H3K4me3相关LncRNA预后模型的构建与验证

基于训练组确定的目标LncRNA及其相关回归系数,使用风险评分公式分别计算了训练组、验证组和整体组中每例胃癌患者的风险评分并构建了预后模型。以训练组中位风险评分0.079 122 212作为临界值将胃癌患者划分为高风险组和低风险组,其中训练组分别为高风险92例和低风险93例,验证组分别为高风险104例和低风险81例,而整体组分别为高风险196例和低风险174例。不同亚群其H3K4me3-LncRNA表达模式呈现高度的相似性(图2A~2C)。在训练组、验证组和整体组中,其14个LncRNA表达、生存状态和生存时间在高风险组和低风险组间显示出清晰的区分度,且高风险组的生存结局较差,预示不良预后(图2D~2I)。ROC曲线分析结果显示,训练组患者

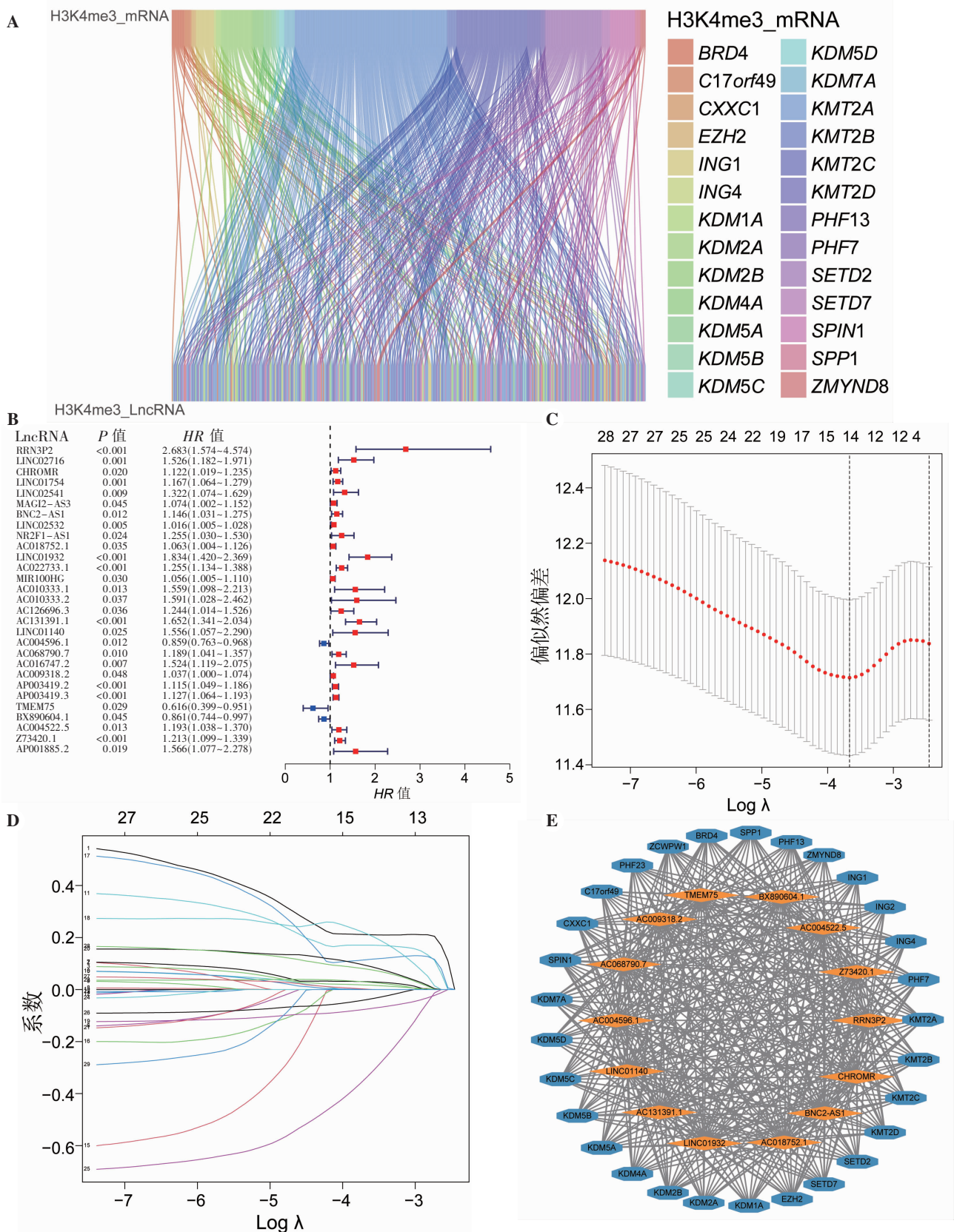


图1 鉴定具有预后价值的H3K4me3相关LncRNA

注:A, H3K4me3 调节因子相关基因与识别的H3K4me3-LncRNA 的桑基图;B, 单因素 Cox 回归分析确定了 29 个具有预后价值的 H3K4me3-LncRNA;C, H3K4me3-LncRNA 的回归系数路径图;D, Lasso 回归的交叉验证曲线;E, H3K4me3 调节因子和相关 LncRNA 的基因共表达网络关系图。

表1 单因素 Cox 及 Lasso 回归分析得出的 14 个与胃癌 H3K4me3 相关的 LncRNA 及其风险比

H3K4me3-LncRNA	Lasso 回归系数	HR (95%CI)	P 值
AC004522.5	0.001 521	1.193 (1.038~1.370)	0.013
AC004596.1	-0.074 990	0.859 (0.763~0.968)	0.012
AC009318.2	0.028 799	1.037 (1.000~1.074)	0.048
AC018752.1	0.000 601	1.063 (1.004~1.126)	0.035
AC068790.7	0.091 923	1.189 (1.041~1.357)	0.010
AC131391.1	0.114 854	1.652 (1.341~2.034)	<0.001
BNC2-AS1	0.030 263	1.146 (1.031~1.275)	0.012
BX890604.1	-0.038 530	0.861 (0.744~0.997)	0.045
CHROMR	0.017 369	1.122 (1.019~1.235)	0.020
LINC01140	0.214 191	1.556 (1.057~2.290)	0.025
LINC01932	0.170 822	1.834 (1.420~2.369)	<0.001
RRN3P2	0.211 127	2.683 (1.574~4.574)	<0.001
TMEM75	-0.336 630	0.616 (0.399~0.951)	0.029
Z73420.1	0.078 575	1.213 (1.099~1.339)	<0.001

1、3、5 年总生存率的 AUC 分别为 0.708、0.730、0.770, 验证组 AUC 分别为 0.690、0.648、0.713, 而整体组的 AUC 分别为 0.697、0.670、0.724 (图 2J~2L)。上述结果表明基于 H3K4me3 相关 LncRNA 构建的风险评分模型对胃癌预后具有良好的预测价值。

2.3 独立预后分析

运用单因素和多因素 Cox 回归分析以识别基于 H3K4me3 相关 LncRNA 构建的风险评分模型是否具有独立于其他临床指标的预后预测价值。单因素 Cox 回归分析结果显示, 年龄、肿瘤分期、T 分期、N 分期、M 分期及风险评分是患者预后的预测因素 ($P < 0.05$), 其中高龄、肿瘤浸润加深、发生淋巴结转移或远处转移、风险评分为高风险提示患者预后较差。随后将上述变量纳入多因素 Cox 回归分析, 结果表明基于 H3K4me3 相关 LncRNA 模型获得的风险评分具有独立于其他临床指标的预后预测价值, 风险评分为高风险可作为预测胃癌患者预后的独立危险因子 ($HR = 2.312, 95\% CI 1.583 \sim 3.379, P < 0.001$) (表 2)。

2.4 列线图的建立与验证

结合基于 Cox 回归分析识别的具有独立预后预测价值的临床指标 (年龄、风险评分) 和肿瘤 TNM 分期构建了预测胃癌患者 1、2、3 年总生存率的多因素列线图, 以进一步提高预后模型的可及性和准确性 (图 3A)。ROC 曲线分析结果显示, 列线图 1、2、3 年总生存率的 AUC 均大于 0.7 (图

3B), 而列线图 2 年总生存率的 AUC 为 0.780, 具有优于风险评分 ($AUC = 0.696$)、肿瘤 TNM 分期 ($AUC = 0.670$)、T 分期 ($AUC = 0.600$)、N 分期 ($AUC = 0.639$) 以及 M 分期 ($AUC = 0.533$) 等单一临床指标的预测价值 (图 3C)。同时校准曲线验证了列线图具有较高的预测准确性, 并显示出与实际生存期相符的预期生存状况和较高的净收益 (图 3D~3F)。上述结果表明结合风险评分构建的多因素列线图具有较强的预后预测潜能, 对于指导胃癌的临床实践具有潜在的应用价值。

2.5 H3K4me3-LncRNA 的聚类分析与免疫预测

将筛选获得的 14 个 H3K4me3-LncRNA 表达水平, 基于共识聚类算法进行二次聚类, 当 $k = 2$ 时聚类效果最佳, 此时样本被重新划分为两个异质性亚群, 即亚组 I 和亚组 II (图 4A)。主成分分析显示两个亚群之间具有可识别的区分度 (图 4B)。匹配生存时间和生存状况后, 使用 Kaplan-Meier 曲线对这两个亚群进行了生存分析, 结果表明亚组 II 的总体生存状况要优于亚组 I, 且差异具有统计学意义 ($P = 0.025$) (图 4C)。随后进一步探索了不同亚群其肿瘤微环境中免疫细胞的浸润异质性, 发现包括 $CD8^+$ T 细胞、激活自然杀伤细胞 (natural killer cell, NK 细胞)、M1 样巨噬细胞在内的多种免疫细胞在亚组 II 中上调 ($P < 0.05$), 预示其可能具有更活跃的免疫状态 (图 4D)。此外, H3K4me3-LncRNA 聚类特征下的免疫疗效预测分析显示, 亚组 II 具有更高的 IPS, 这表明亚组 II 具

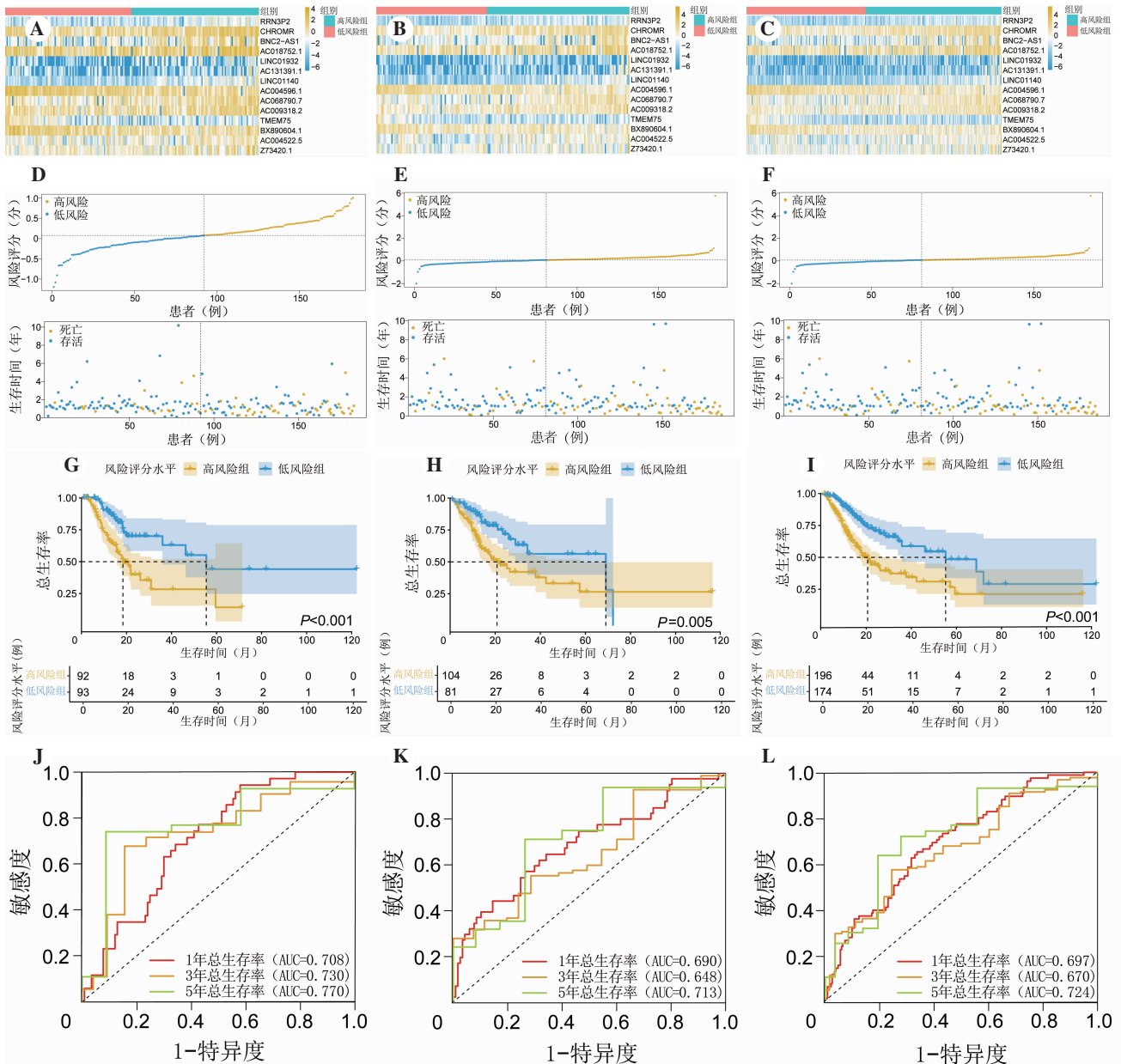


图2 H3K4me3-LncRNA 风险评分模型在训练组、验证组、整体样本组中的预后价值

注:A~C, 训练组(A)、验证组(B)、整体组(C)中14个H3K4me3-LncRNA的表达热图;D~F, 训练组(D)、验证组(E)、整体组(F)的风险评分曲线(上)和胃癌患者生存状态散点图(下);G~I, 训练组(G)、验证组(H)、整体组(I)中高、低风险组的Kaplan-Meier生存曲线;J~L, 训练组(J)、验证组(K)、整体组(L)中1、3、5年总生存率的ROC曲线。

备较强的免疫应答潜能,特别是对单独PD-1抑制剂治疗和单独CTLA4抑制剂治疗尤为敏感(图4E~4H)。基于GSDC数据库中肿瘤细胞对抗癌药物敏感性和反应性的数据,对胃癌亚群间潜在联合免疫治疗药物的IC50进行对比,发现包括5-氟尿嘧啶、奥沙利铂在内的多种药物在亚组II呈现出更低的IC50,这与亚组II更活跃的“热肿瘤”特性相呼应。同时分析结果也显示亚组I患者可能对塔塞利布、阿培利司等磷脂酰肌醇3-激酶

(phosphatidylinositol 3-kinases, PI3K)特异性抑制剂具有更高的敏感性(图4I~4P)。上述发现表明基于H3K4me3-LncRNA特征构建的分类模型对于指导胃癌临床的精准用药具有现实的应用价值。

3 讨论

随着生物学技术的发展,越来越多的研究表明LncRNA的异常表达和功能失调是包括胃癌在内多种肿瘤发生和发展的关键^[15-16]。一些在肿瘤

表2 预测胃癌患者预后的相关临床指标的单因素和多因素 Cox 回归分析

临床指标	单因素 Cox 回归分析			多因素 Cox 回归分析		
	HR	95%CI	P 值	HR	95%CI	P 值
年龄	1.020	1.003~1.037	0.024	1.034	1.014~1.054	0.001
性别						
女比男	0.764	0.526~1.110	0.157	0.822	0.555~1.218	0.329
肿瘤分级			0.548			0.373
G ₂ 比 G ₁	1.132	0.351~3.651	0.835	2.105	0.551~8.041	0.276
G ₃ 比 G ₁	1.453	0.458~4.608	0.526	2.624	0.699~9.858	0.153
G _x 比 G ₁	1.620	0.326~8.041	0.555	3.244	0.567~18.543	0.186
肿瘤 TNM 分期			<0.001			0.123
II 期比 I 期	1.474	0.732~2.965	0.227	0.956	0.348~2.624	0.930
III 期比 I 期	2.239	1.175~4.269	0.014	1.439	0.377~5.495	0.594
IV 期比 I 期	4.309	2.807~8.897	<0.001	2.920	0.689~12.365	0.146
X 期比 I 期	5.325	2.250~12.602	<0.001	2.486	0.618~10.004	0.200
T 分期			0.003			0.526
T ₂ 比 T ₁	6.085	0.825~44.879	0.077	3.368	0.434~26.153	0.246
T ₃ 比 T ₁	8.643	1.197~62.386	0.032	4.648	0.541~39.941	0.162
T ₄ 比 T ₁	9.237	1.267~67.352	0.028	4.119	0.471~36.048	0.201
T _x 比 T ₁	37.152	4.130~334.201	0.001	6.571	0.566~76.283	0.132
N 分期			0.003			0.195
N ₁ 比 N ₀	1.498	0.911~2.463	0.111	0.869	0.409~1.844	0.714
N ₂ 比 N ₀	1.552	0.909~2.648	0.107	0.854	0.345~2.115	0.733
N ₃ 比 N ₀	2.504	1.522~4.119	<0.001	1.289	0.518~3.205	0.585
N _x 比 N ₀	3.308	1.270~8.616	0.014	2.470	0.711~8.573	0.155
M 分期			0.015			0.725
M ₁ 比 M ₀	2.339	1.314~4.163	0.004	1.405	0.584~3.381	0.448
M _x 比 M ₀	1.186	0.484~2.907	0.709	0.877	0.340~2.264	0.786
风险评分						
高风险比低风险	2.420	1.685~3.474	<0.001	2.312	1.583~3.379	<0.001

注:G_x、肿瘤 TNM 分期的 X 期、T_x、N_x、M_x 为未知分型。

中表达过度或缺失的 LncRNA 具有成为肿瘤预后标志物和治疗靶点的潜力。既往研究发现 LncRNA MALAT1 在肿瘤中的表观遗传调控与肿瘤侵袭和转移密切相关,靶向 MALAT1 具有潜在的临床应用价值^[17]。LncRNA HOTAIR 也被证实其在多种肿瘤中的过度表达与肿瘤的不良预后相关^[18]。此外,全基因组染色质状态分析结果显示,包括胃癌、结直肠癌、前列腺癌、乳腺癌等多种癌症的发生过程均存在组蛋白 H4 第 16 位赖氨酸的乙酰化、H3K4me3 和组蛋白 H4 第 20 位赖氨酸的三甲基化的普遍缺失以及组蛋白 H3 第 9 位赖氨酸的甲基化、组蛋白 H3 第 27 位赖氨酸的三甲基化修饰增加^[19-20]。H3K4me3 是组蛋白修饰的一种形式,其可通过与转录因子 II D 复合物相互作用来调控基因激活,

被公认为是一种与转录起始有关的组蛋白标签。一项来自肺腺癌的研究表明,基于 H3K4me3 相关 LncRNA 特征构建的评分模型能够有效识别不同亚群的免疫背景,在指导免疫治疗方面显示出独特的应用价值^[21]。虽然 H3K4me3 相关的表观遗传修饰在胃癌病理进程中扮演着至关重要的角色,但目前对于 H3K4me3 相关 LncRNA 在当中的作用缺乏系统阐述。

在本研究中,笔者构建了 1 个 H3K4me3 相关 LncRNA 风险评分模型,该模型在训练组中预测胃癌患者 1、3、5 年总生存率均具有一定的准确性,其 AUC 分别为 0.708、0.730 和 0.770,且多因素 Cox 回归分析结果显示该风险评分模型是预测胃癌患者预后的独立因素。此外,该模型在验证组和

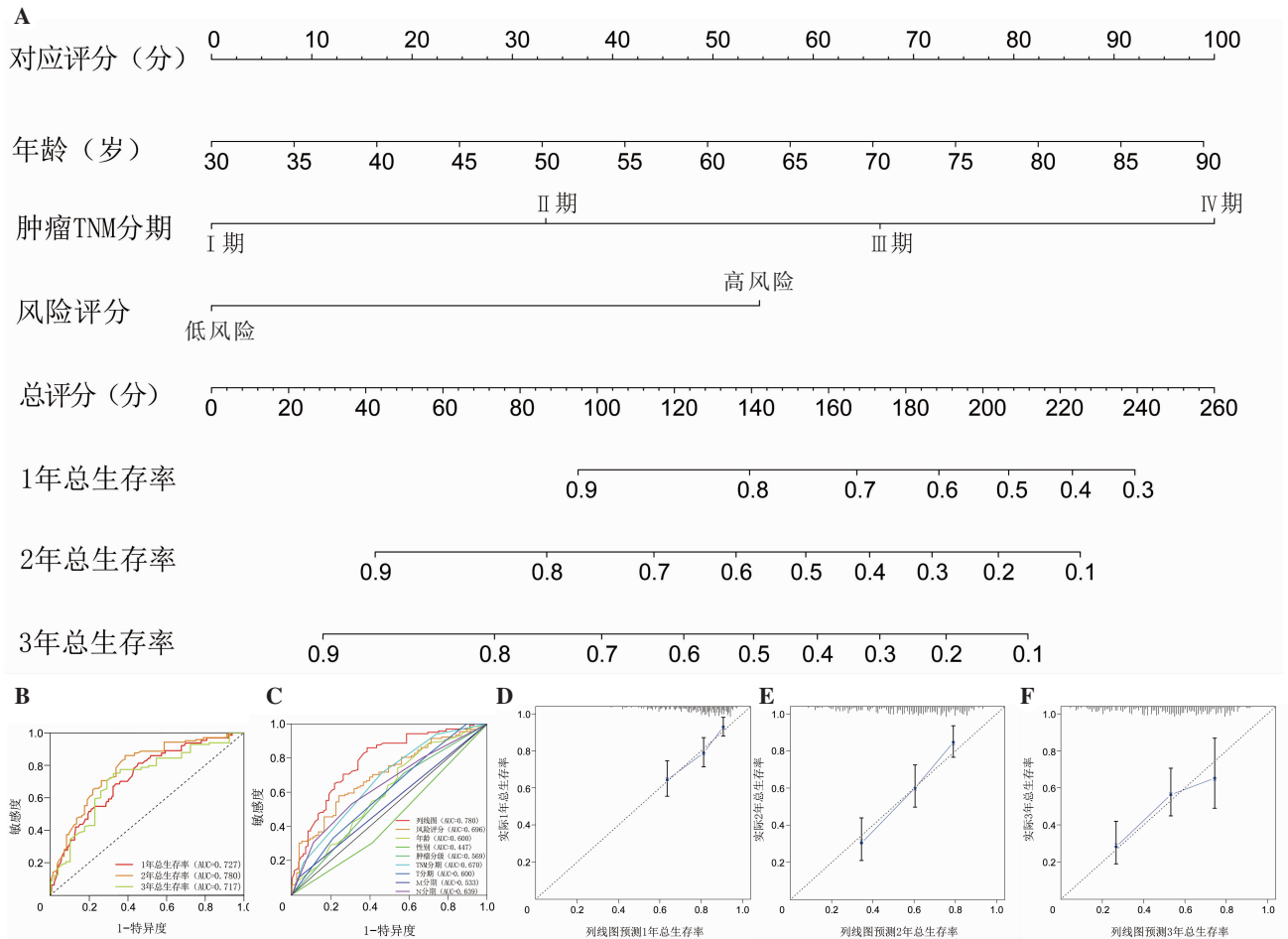


图3 预测胃癌患者1、2、3年总生存率的多因素列线图的构建与验证

注:A,基于年龄、肿瘤TNM分期和风险评分构建的列线图;B,列线图评估胃癌患者1、2、3年总生存率的ROC曲线;C,列线图、风险评分与其他临床指标评估胃癌患者2年总生存率的ROC曲线;D-F,列线图预测胃癌患者1、2、3年总生存率的校准图。

整体组中也具有良好的预测性能,进一步证实了模型的稳健性。为深化模型的临床可及性和实用性,笔者结合具有独立预后预测价值的指标和传统肿瘤TNM分期构建了1个预测胃癌患者1、2、3年总生存率的多因素列线图,校准曲线显示列线图对胃癌患者1、2、3年总生存率的预测与实际生存情况具有较高的吻合度,且其1、2、3年总生存率的AUC均大于0.7,优于传统的肿瘤TNM分期及其他预后指标。鉴于胃癌微观环境的高度异质性,笔者试图进一步分析其微观异质性在免疫治疗方面的应用价值。通过共识聚类,当k=2时聚类亚群呈现出较高的清洁性和区分度,此时肿瘤样本被划分为最佳类别。随后进一步证实了不同聚类亚群免疫细胞的浸润水平差异,发现包括CD8⁺T细胞在内的免疫细胞在不同亚群中的浸润

水平差异有统计学意义。CD8⁺T细胞被认为是抗肿瘤免疫的主要驱动因素,并可作为中枢效应细胞调节免疫检查点阻断治疗的疗效。进一步评估不同亚群的IPS差异及其对潜在联合免疫治疗药物的敏感性,结果表明亚组II具有较强的免疫应答潜力,并显示出对5-氟尿嘧啶、奥沙利铂更高的敏感性,而亚组I则可能从靶向PI3K治疗中获益。

总之,基于H3K4me3相关LncRNA构建的预后模型具有较好的预后预测能力,能够有效筛选联合免疫治疗相关药物,对指导临床个体化用药具有一定的意义。本研究是基于公共数据库建立的,虽然在内部验证中得到证实,但受限于目标LncRNA在外部数据集的缺乏,无法得到TCGA以外数据集的有效验证,因此其临床实用性仍需进一步检验及前瞻性实验的证实。

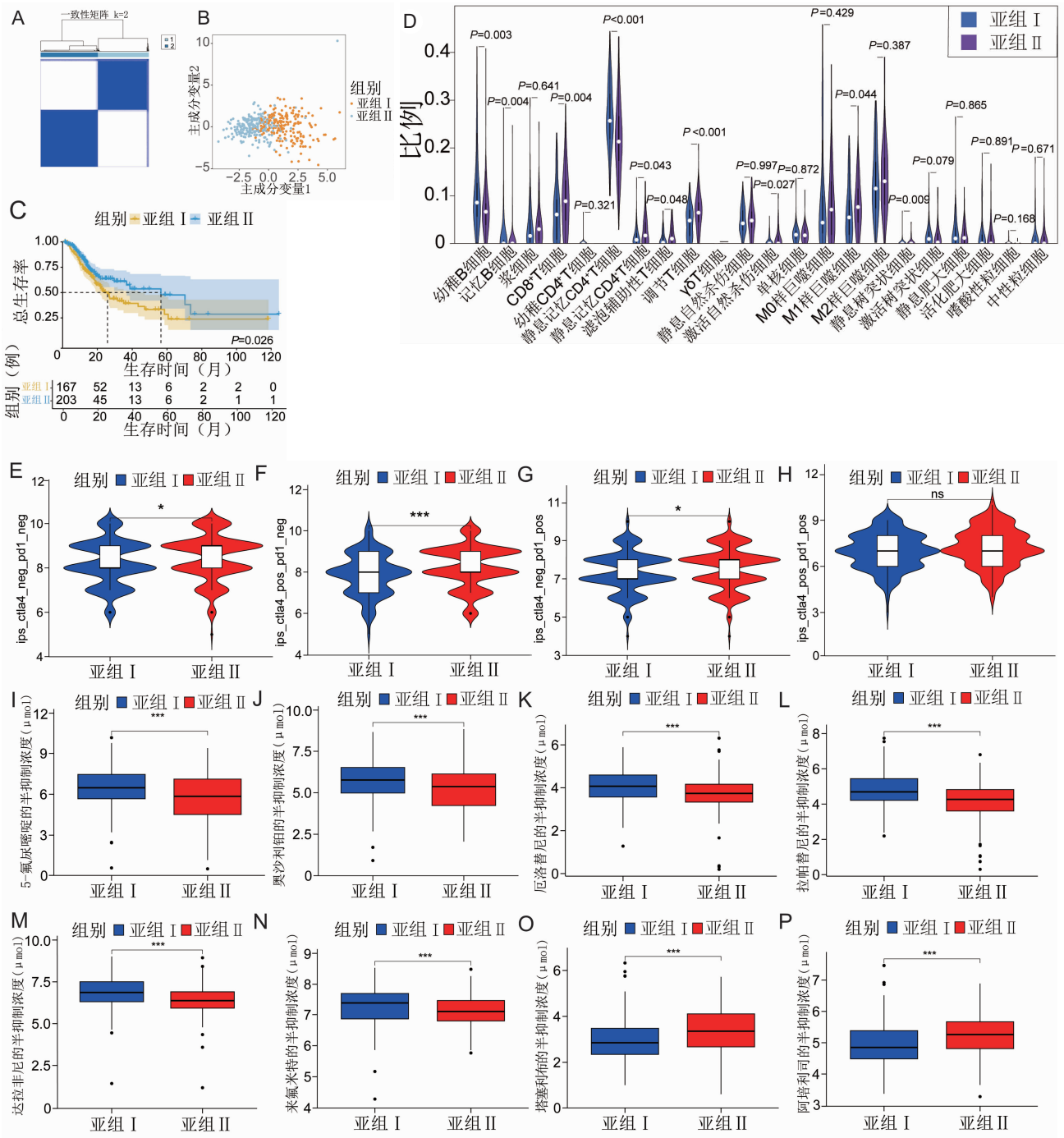


图 4 H3K4me3-LncRNA 的聚类分析与免疫治疗疗效预测

注:A,基于 H3K4me3-LncRNA 的共识聚类;B,亚组 I 和亚组 II 的主成分聚类图;C,亚组 I 和亚组 II 的 Kaplan-Meier 生存曲线;D,22 种免疫细胞在亚组 I 和亚组 II 的分布情况,NK 细胞为自然杀伤细胞;E~H,免疫表型评分在亚组 I 和亚组 II 的分布情况,ips_ctla4_neg_pd1_neg(E)表示对细胞毒性 T 淋巴细胞相关蛋白 4 抑制剂和程序性死亡受体 1 抑制剂均无应答,ips_ctla4_pos_pd1_neg (F) 表示对细胞毒性 T 淋巴细胞相关蛋白 4 抑制剂应答而对程序性死亡受体 1 抑制剂无应答,ips_ctla4_neg_pd1_pos (G) 表示对细胞毒性 T 淋巴细胞相关蛋白 4 抑制剂无应答而对程序性死亡受体 1 抑制剂应答,ips_ctla4_pos_pd1_pos (H) 表示对细胞毒性 T 淋巴细胞相关蛋白 4 抑制剂和程序性死亡受体 1 抑制剂均应答;I~P,来自 GDSC 数据库的 8 种联合免疫治疗药物的半抑制浓度在亚组 I 和亚组 II 的分布。ns, $P>0.05$;* $P<0.05$;** $P<0.01$;*** $P<0.001$ 。

参考文献

- [1] SMYTH EC, NILSSON M, GRABSCH HI, et al. Gastric cancer [J]. *Lancet*, 2020, 396(10251): 635–648.
- [2] 中国抗癌协会胃癌专业委员会. 胃癌腹膜转移诊治中国专家共识 (2023 版) [J]. *中华胃肠外科杂志*, 2023, 26(8): 717–728.
- [3] SUN L, ZHANG H, GAO P. Metabolic reprogramming and epigenetic modifications on the path to cancer [J]. *Protein Cell*, 2022, 13(12): 877–919.
- [4] FLAVAHAN WA, GASKELL E, BERNSTEIN BE. Epigenetic plasticity and the hallmarks of cancer [J]. *Science*, 2017, 357(6348): eaal2380.
- [5] BANNISTER AJ, KOUZARIDES T. Regulation of chromatin by histone modifications [J]. *Cell Res*, 2011, 21(3): 381–395.
- [6] CHEN K, CHEN Z, WU D, et al. Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes [J]. *Nat Genet*, 2015, 47(10): 1149–1157.
- [7] WANG H, FAN Z, SHLIAHA PV, et al. H3K4me3 regulates RNA polymerase II promoter-proximal pause-release [J]. *Nature*, 2023, 615(7951): 339–348.
- [8] SINGH D, ASSARAF YG, GACCHE RN. Long non-coding RNA mediated drug resistance in breast cancer [J]. *Drug Resist Updat*, 2022, 63: 100851.
- [9] HASHEMI M, MOOSAVI MS, ABED HM, et al. Long non-coding RNA (lncRNA) H19 in human cancer: From proliferation and metastasis to therapy [J]. *Pharmacol Res*, 2022, 184: 106418.
- [10] XU M, XU X, PAN B, et al. LncRNA SATB2-AS1 inhibits tumor metastasis and affects the tumor immune cell microenvironment in colorectal cancer by regulating SATB2 [J]. *Mol Cancer*, 2019, 18(1): 135.
- [11] DENG M, WAN Y, CHEN B, et al. Long non-coding RNA lnc_3712 impedes nuclear reprogramming via repressing Kdm5b [J]. *Mol Ther Nucleic Acids*, 2021, 24: 54–66.
- [12] XU Q, XIANG Y, WANG Q, et al. SETD2 regulates the maternal epigenome, genomic imprinting and embryonic development [J]. *Nat Genet*, 2019, 51(5): 844–856.
- [13] ZHOU Q, ZHANG Y, WANG B, et al. KDM2B promotes IL-6 production and inflammatory responses through Brg1-mediated chromatin remodeling [J]. *Cell Mol Immunol*, 2020, 17(8): 834–842.
- [14] SHA Q, ZHU Y, XIANG Y, et al. Role of CxxC-finger protein 1 in establishing mouse oocyte epigenetic landscapes [J]. *Nucleic Acids Res*, 2021, 49(5): 2569–2582.
- [15] OLIVERO CE, MARTÍNEZ-TERROBA E, ZIMMER J, et al. p53 Activates the Long Noncoding RNA Pvt1b to Inhibit Myc and Suppress Tumorigenesis [J]. *Mol Cell*, 2020, 77(4): 761–774.e768.
- [16] 王晓龙, 陈亮, 曹洪涛, 等. 长链非编码 RNA NPPA-AS1 通过靶向 miR-372-3p 调控结直肠癌 SW480 细胞的增殖和凋亡 [J/CD]. *消化肿瘤杂志(电子版)*, 2022, 14(2): 141–147.
- [17] GOYAL B, YADAV SRM, AWASTHEE N, et al. Diagnostic, prognostic, and therapeutic significance of long non-coding RNA MALAT1 in cancer [J]. *Biochim Biophys Acta Rev Cancer*, 2021, 1875(2): 188502.
- [18] RAJU GS, PAVITRA E, BANDARU SS, et al. HOTAIR: a potential metastatic, drug-resistant and prognostic regulator of breast cancer [J]. *Mol Cancer*, 2023, 22(1): 65.
- [19] PAL D, PATEL M, BOULET F, et al. H4K16ac activates the transcription of transposable elements and contributes to their cis-regulatory function [J]. *Nat Struct Mol Biol*, 2023, 30(7): 935–947.
- [20] MARSOLIER J, PROMPSY P, DURAND A, et al. H3K27me3 conditions chemotolerance in triple-negative breast cancer [J]. *Nat Genet*, 2022, 54(4): 459–468.
- [21] FAN T, ZHU M, MUHAMMAD S, et al. H3K4me3-related lncRNAs signature and comprehensive analysis of H3K4me3 regulating tumor immunity in lung adenocarcinoma [J]. *Respir Res*, 2023, 24(1): 122.

收稿日期: 2023-12-28