DOI:10.16016/j.2097-0927.202502080

基于机器学习的中心体扩增相关基因预测食管鳞状细胞癌的预后

李超群,郑泓量,黄 平,杨 欢,柏健鹰 陆军军医大学(第三军医大学)第二附属医院消化内科,重庆

[摘要] **目的** 通过机器学习构建中心体扩增相关基因(centrosome amplification-related genes, CARGs)的预后模型,研究CARGs对食管鳞状细胞癌(esophageal squamous cell carcinoma, ESCC)预后的影响。**方法** 从基因本体论(gene ontology, GO)和京都基因与基因组百科全书(Kyoto encyclopedia of genes and genomes, KEGG)数据库中下载CARGs。下载癌症基因组图谱(the Cancer Genome Atlas, TCGA)和基因表达综合数据库(Gene Expression Omnibus, GEO)中的ESCC转录组测序(RNA-sequencing, RNA-seq)数据,作为训练集和验证集。基于单样本基因集富集分析(single-sample gene set enrichment analysis, ssGSEA)和加权基因共表达网络(weighted gene co-expression network analysis, WGCNA)筛选CARGs。使用集成了12种机器学习算法构建CARGs的预后模型,采用单因素和多因素Cox回归分析判断模型是否可以作为独立预后因素。从陆军军医大学第二附属医院消化内科收集15对ESCC和癌旁组织样本,采用RT-qPCR和免疫组化验证基因在组织样本中的表达。**结果** 构建了包含9个CARGs的预后模型,其中基因*DENR*、*TRIP13*、*BRCA2*、*TTF2*、*TCFL5*和*NUP188*经实验验证其mRNA表达水平高于正常组织($P<0.05$),DENR、TRIP13、TTF2和TCFL5的蛋白表达水平高于正常组织。**结论** *DENR*、*TRIP13*、*TTF2*和TCFL5在ESCC中表达水平升高,与临床不良预后密切相关,提示这些基因在ESCC发生发展中可能发挥作用。

[关键词] 机器学习;食管鳞状细胞癌;中心体扩增相关基因;预后
[中图法分类号] R394.3;R730.7;R735.1 [文献标志码] A

Construction of machine learning-based prognostic model of centrosome amplification-related genes for esophageal squamous cell carcinoma

LI Chaoqun, ZHENG Hongliang, HUANG Ping, YANG Huan, BAI Jianying Department of Gastroenterology, Second Affiliated Hospital, Army Medical University (Third Military Medical University), Chongqing, China

[**Abstract**] **Objective** To construct a prognostic model of centrosome amplification-related genes (CARGs) by machine learning and evaluate its prediction performance for the prognosis of esophageal squamous cell carcinoma (ESCC). **Methods** CARGs were obtained from Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) datasets. The RNA-sequencing (RNA-seq) transcriptome datasets of ESCC were retrieved from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO), and assigned into training and validation sets, respectively. Subsequently, single-sample gene set enrichment analysis (ssGSEA) and weighted gene co-expression network analysis (WGCNA) were employed to screen CARGs. A prognostic model of CARGs was constructed by incorporating 12 machine learning algorithms, and univariate and multivariate Cox regression analyses were applied to evaluate whether the 12 models as an independent prognostic factor or not. Eventually, 15 paired ESCC and adjacent non-tumor tissue

[基金项目] 重庆英才创新领军人才项目(2022CQ010)
[通信作者] 柏健鹰,E-mail:baijianying_xq@tmmu.edu.cn

2097-0927/Copyright © 2025, 陆军军医大学(第三军医大学)。这篇开放获取文章遵循CC BY许可协议(https://creativecommons.org/licenses/by/4.0/)
引用格式:李超群,郑泓量,黄平,等. 基于机器学习的中心体扩增相关基因预测食管鳞状细胞癌的预后[J]. 陆军军医大学学报,2025,47(13):1536-1547. DOI:10.16016/j.2097-0927.202502080.
Citation:LI C Q, ZHENG H L, HUANG P, *et al*. Construction of machine learning-based prognostic model of centrosome amplification-related genes for esophageal squamous cell carcinoma[J]. J Army Med Univ, 2025, 47(13):1536-1547. DOI:10.16016/j.2097-0927.202502080.

samples were collected from the Department of Gastroenterology of the Second Affiliated Hospital of Army Medical University, and real-time quantitative PCR (RT-qPCR) and immunohistochemistry staining were performed to detect the expression of these genes ESCC samples. **Results** Our 9-CARGs prediction model for ESCC prognosis was constructed. RT-qPCR confirmed that the mRNA expression levels of *DENR*, *TRIP13*, *BRCA2*, *TTF2*, *TCFL5* and *NUP188* were significantly higher in ESCC tissues than normal tissues (P<0.05), and the protein levels of DENR, TRIP13, TTF2 and TCFL5 were also elevated when compared to normal tissues. **Conclusion** *DENR*, *TRIP13*, *TTF2* and *TCFL5* are highly expressed and closely associated with poor prognosis of ESCC, suggesting their potential roles in the pathogenesis and progression of this malignancy.

[**Key words**] machine learning; esophageal squamous cell carcinoma; centrosome amplification-related genes; prognosis

Supported by the Innovative Leading Talents of Chongqing Talent Program (2022CQ010). Corresponding author: BAI Jianying, E-mail: baijianying_xq@tmmu.edu.cn

食管鳞状细胞癌(esophageal squamous cell carcinoma, ESCC)是我国最常见的食管癌类型,占全部病例的90%[1]。由于症状不明显,大多数ESCC患者确诊时已属晚期,5年总生存率低于20%[2]。目前,尽管新辅助化疗和手术干预取得了实质性进展[3],但预后仍不理想[4]。因此,探索ESCC的关键致癌基因对于准确评估ESCC患者的预后尤为重要。

中心体是大多数动物细胞中充当主要微管组织中心功能的细胞器。除了控制细胞形状和细胞内运输等微管相关过程外,对有丝分裂纺锤体的组装和细胞分裂也至关重要[5]。中心体数量的增加即中心体扩增,可以干扰有丝分裂过程中染色体分离的保真性,并导致遗传不稳定性、非整倍体和肿瘤进展[6-8],与癌症患者的不良临床预后密切相关[9-11]。既往研究已证实中心体扩增在乳腺癌、头颈部鳞状细胞癌等多种癌症中广泛发生[12],但中心体扩增在ESCC中的作用研究较少。

机器学习是人工智能的一个分支,在肿瘤学的早期检测、疾病分类和个性化治疗方法中显示出巨大的前景[13]。如今,越来越多的算法组合用于构建癌症预后预测模型[14-15]。

鉴于中心体扩增与肿瘤发生的密切病理关联,本研究分析了中心体扩增相关基因(centrosome amplification-related genes, CARGs)是否与ESCC的发展和不良预后有关。利用公共数据库数据分析正常食管与ESCC之间差异表达基因,整合多种机器学习算法构建筛选预后模型。运用单因素、多因素Cox分析判断模型是否可以作为独立预后因素。通过实验验证基因在ESCC临床样本中的表达,以期为ESCC诊断预测和预后提供新的见解。

1 材料与方法

1.1 实验样本

本研究实验样本取自陆军军医大学第二附属医院消化内科,共纳入15对经病理确诊的ESCC组织及其配对的癌旁正常组织。经陆军军医大学第二附属医院伦理委员会审核批准(2024-研第306-01),在进行研究之前向患者提供知情同意书并取得了患者的同意。

1.2 实验试剂

Trizol、逆转录试剂购自日本Takara公司,Taq Pro Universal SYBR qPCR Master Mix购自南京诺维赞科技有限公司,引物由上海生工公司合成。免疫组化试剂盒购自北京中杉金桥公司。

1.3 数据的收集和处理

从癌症基因组图谱(the Cancer Genome Atlas, TCGA)和基因表达综合数据库(Gene Expression Omnibus, GEO)下载转录组测序(RNA-sequencing, RNA-seq)数据和患者临床数据,包括TCGA-ESCC(癌症样本81例,正常样本11例),GSE53625(癌症样本179例,正常样本179例)和GSE213565(癌症样本10例,正常样本10例)3个数据集。为了比较肿瘤与非肿瘤组织之间的基因表达差异,对TCGA-ESCC和GSE213565数据集以校正后P<0.05和|Log_2FC|>0.5为标准,使用DESeq2包进行差异表达基因(differentially expressed genes, DEGs)分析,并通过ComplexHeatmap包可视化。将TCGA-ESCC作为训练集,GSE53625作为验证集用于构建和验证预后模型。

1.4 ESCC中心体扩增相关基因的鉴定

从基因本体论(gene ontology, GO)和京都基因与基因组百科全书(Kyoto encyclopedia of genes and genomes, KEGG)数据库中下载412个CARGs[16-17]。使用单样本基因集富集分析(single-sample gene set enrichment analysis, ssGSEA)计算TCGA数据集中每个患者的CARGs评分[18]。再用加权基因共表达网络(weighted gene co-expression network analysis, WGCNA)来确定与CARGs分数最显著相关的基因模块,利用无标度拓扑标准判断计算软阈值,最佳软阈值权重设置为β=6(R^2=0.85),将最小模块基因计数设置

为200,mergeCutHeight为0.25。用Venneuler包将DEGs和WGCNA取交集,使用clusterProfiler包进行GO和KEGG富集分析[19],富集标准定义为校正后$P<0.05$。

1.5 CARGs预后模型的构建

应用单因素Cox回归分析在交集基因中筛选具有潜在预后作用的CARGs($P<0.05$)。再选择包括随机森林(random forest, RF)、最小绝对收缩和选择算子(least absolute shrinkage and selection operator, Lasso)、岭回归算法(ridge regression algorithm, Ridge)、偏最小二乘回归(partial least squares regression, plsr)、弹性网络(elastic network, Enet)、广义线性模型中的变量选择(variable selection in generalized linear models, Stepglm)、支持向量机(support vector machine, SVM)、基于似然的提升的广义线性模型(generalized linear model by likelihood based boosting, glmBoost)、线性判别分析(linear discriminant analysis, LDA)、梯度提升机(gradient boosting machine, GBM)、极限梯度提升(eXtreme gradient boosting, XGboost)和朴素贝叶斯(naiveBayes)12种组合机器学习算法对训练集和验证集进行变量选择和模型构建。采用十折交叉验证框架对预后模型进行拟合,计算训练集和验证集的预测性能的AUC指数,以确定最优的算法组合,筛选最优的预后模型。

1.6 生存分析及预测列线图的构建

根据Cox分析风险评分中位数,将TCGA-ESCC和GSE53625分为高、低风险组。随后,使用survminer包进行Kaplan-Meier曲线分析,以探讨高、低风险组在总生存期(Overall survival, OS)方面是否存在显著差异。此外,使用timeROC包进行受试者工作特征(receiver operating characteristic, ROC)曲线分析,以评估CARGs预测ESCC患者OS的敏感性和特异性并探讨了CARGs与其他临床特征的相关性。对TCGA-ESCC数据集进行了单因素和多因素Cox回归分析,以确定CARGs是否可以作为预测ESCC患者的独立预后因素。为了提高模型的预后准确性和预测能力,绘制了列线图来量化ESCC患者的预期生存时间。通过决策曲线分析(decision curve analysis, DCA)评估列线图的临床净获益。

1.7 RT-qPCR检测

使用Trizol法从11对食管组织中提取RNA,Prime Script RT reagent kit将RNA反转录成cDNA,Taq Pro Universal SYBR qPCR Master Mix进行RT-qPCR,GAPDH作为内参,引物见表1。RT-qPCR反应体系包括:SYBR qPCR Master Mix 5 μL,正、反向引物各0.2 μL,cDNA模板1 μL,ddH_2O补至10 μL。反应程序按照stage1:预变性95 ℃ 30 s;stage2:40个循环,95 ℃ 10 s;60 ℃ 10 s。采用$2^{-\Delta\Delta Ct}$法对基因表达进行定量。

表1 目的基因引物名称、序列

基因	引物(5′→3′)
TCFL5	正向:GCGCTATGCAAACAAGCACT 反向:CCAAGGGCTCTTCGCTCTAC
TTF2	正向:AACCTCAATGTGGAGGGCAC 反向:ATTCGAGCCCAGGCTATTCG
TRIP13	正向:GAGTCGCCAACGGTCCAC 反向:TGAGTAGCTTTCTAACACTCAGGT
DENR	正向:GGGAGACGAGTTGCATGTGT 反向:CCATTTCACAAACCCAGGCG
BRCA2	正向:TGCGCCTCGGGTGTCTTT 反向:GCTTGGTAAATAAGTCCGCTCC
NUP188	正向:ACTGAATAAACATTGGCGGCG 反向:TCAGCTGAACTTGGACTGGG
GAPDH	正向:GTCTCCTCTGACTTCAACAGCG 反向:ACCACCCTGTTGCTGTAGCCAA
POP1	正向:GTCACCTCAAAGCGAGCA 反向:ATTAGCACCATGAGTACCAA
GTF2H3	正向:TATGCAGCATGTGATGCAAC 反向:TTGTGGCAGGTTTGTTGGGAT
DNTTIP2	正向:TCACGGAATCCCGGCGGC 反向:TGCAGTCTCATTCCCAGGCTGTTTC

1.8 免疫组化

4对食管组织样本在4%多聚甲醛中固定24 h后进行脱水包埋。使用的一抗包括:TTF2(1∶200,ab181116,abcam)、TRIP13(1∶200,HPA005727,sigma)、TCFL5(1∶100,29404-1-AP,Proteintech)、DENR(1∶50,10656-1-AP,Proteintech)、NUP188(1∶100,bs-4404R,Bioss)和BRCA2(1∶50,YB1-07C20C,BOSTER)。石蜡切片脱蜡、梯度浓度乙醇脱水。使用柠檬酸钠(索莱宝,C1032,pH=6.0)微波加热进行抗原修复。3%过氧化氢室温孵育30 min后,室温封闭30 min,一抗孵育4 ℃过夜,二抗(1∶200,ab6721,abcam)室温孵育1 h,DAB显影,苏木素复染。

1.9 统计学分析

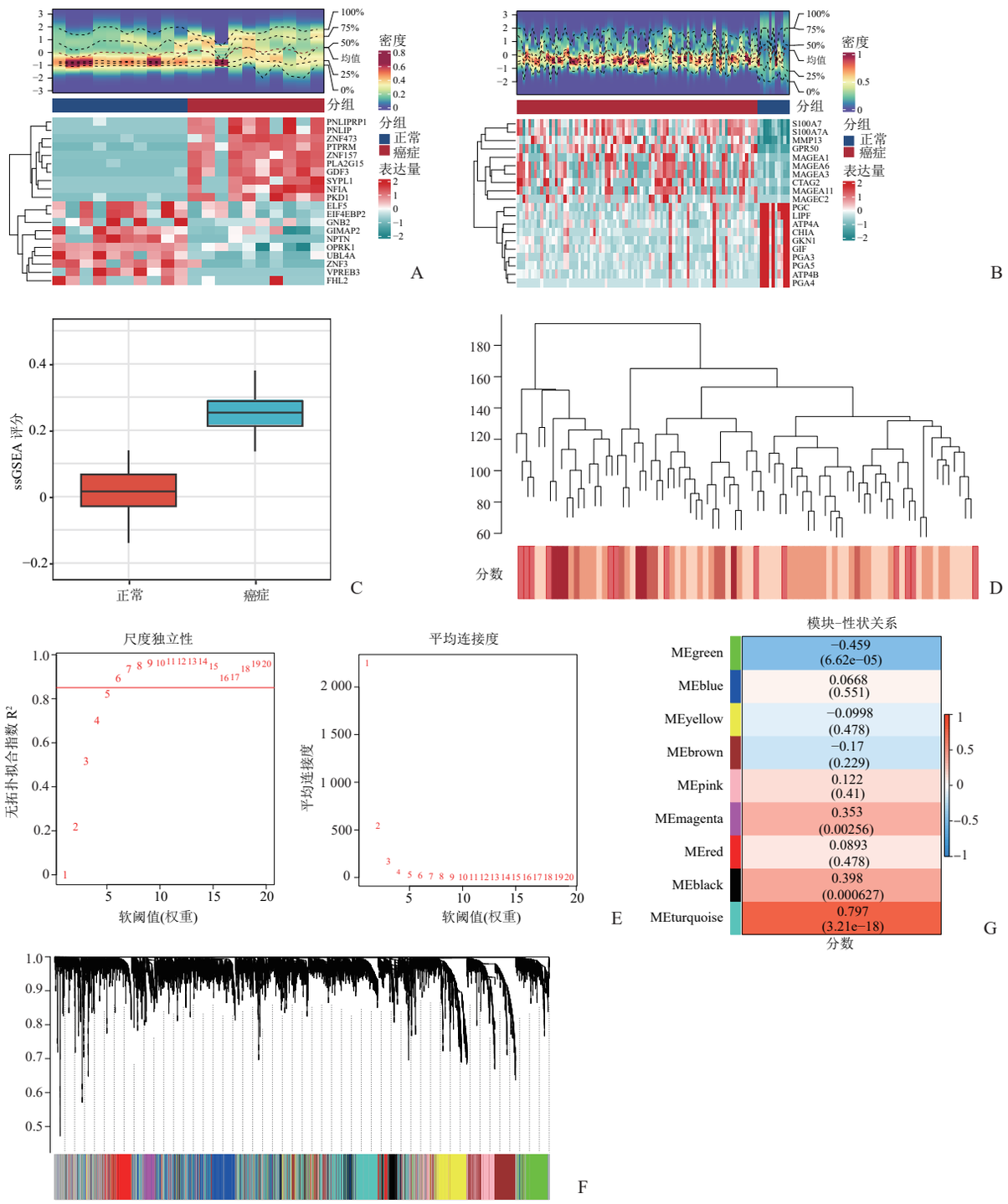
所有统计分析、绘图采用R 4.3.3和GraphPad Prism 9进行。在DEGs和富集分析中,使用校正后P值来评估DEGs的显著差异。采用Wilcoxon检验比较高危组和低危组之间非正态分布变量的差异。采用卡方检验或Fisher精确检验评估CARGs评分与临床特征之间的相关性。比较不同亚组之间的总生存期使用Kaplan-Meier生存分析和log-rank检验。RT-qPCR结果在2组间的比较采用Student's t检验,以$P<0.05$为差异有统计学意义。

2 结果

2.1 RNA-seq中CARGs基因的鉴定

对GSE213565中10个癌症组织和10个正常组织

进行差异分析,得到DEGs共2 617个,其中1 003个上调基因,1 614个下调基因(图1A)。对TCGA中81个ESCC组织和11个配对癌旁非肿瘤组织进行差异分析,得到DEGs2共7 775个差异基因,上调基因3 869个和下调基因3 906个(图1B)。ssGSEA算法通常用于评估单个样品的生物过程和途径活性的变化。利用ssGSEA算法计算每个TCGA-ESCC样本的CARGs活性评分,作为后续WGCNA分析的表型数据。发现在肿瘤组和对照组样本之间的CARGs评分存在显著差异,表明了CARGs在ESCC的发生发展中可能发挥



A、B:GSE213565和TCGA-ESCC数据集差异表达基因热图;C:TCGA-ESCC数据集的ssGSEA评分;D:TCGA-ESCC样本的层次聚类树状图;E、F:WGCNA的最佳软阈值和聚类树;G:模块-性状热图显示模块与CARGs性状的关系

图1　中心体扩增相关基因的鉴定

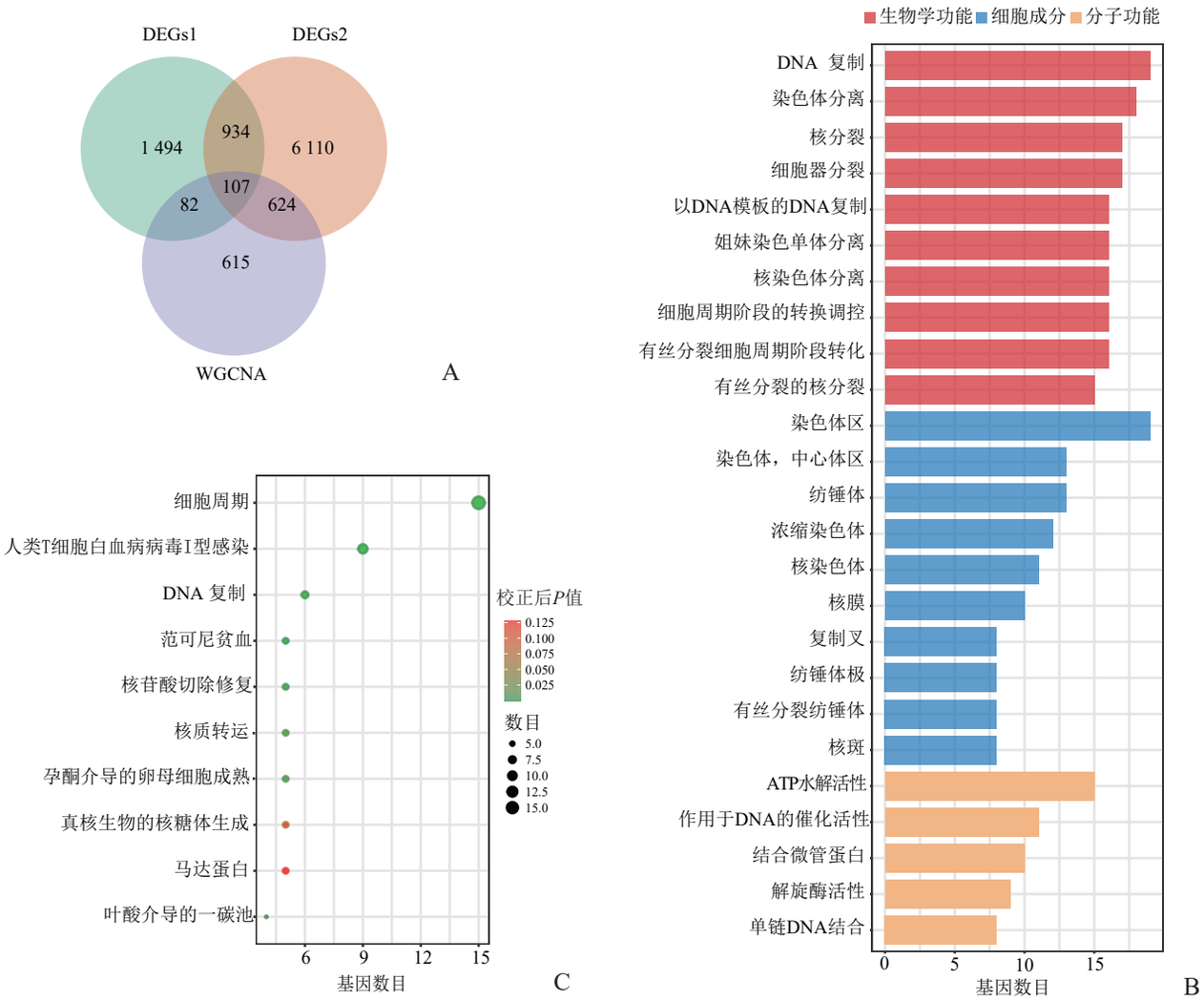
作用(图1C)。

为了识别与CARGs得分显著相关的模块，利用WGCNA分析TCGA-ESCC数据集。通过动态树切分包共鉴定出9个共表达模块(图1D~F)，结果表明青绿色模块与CARGs评分密切相关(|Cor|=0.797，P=3.21e-18，图1G)，提示青绿色模块内的基因可能具有中心体扩增相关的功能，选择其包含的1 428个基因模块与DEGs1、DEGs2取交集，得到107个基因用于接下来的分析(图2A)。

利用GO和KEGG富集分析研究这些基因所具有的分子功能以及参与特定生物现象的关键通路或过程，以了解与ESCC发展相关的潜在分子机制。结果显示，这些基因大多位于染色体、纺锤体上，并映射到与细胞周期相关的通路，包括DNA复制、染色体分离、核分裂、姐妹染色单体分离和有丝分裂细胞周期阶段转化等生物学过程(图2B、C)。随后对107个基因进行单因素Cox回归分析，得到23个显著基因(P<0.05，图3A)。

2.2　基于机器学习的预后模型的构建

为了构建预后模型，使用机器学习算法对单因素Cox回归分析中的23个预后基因进行了分析。组合了12种机器学习算法，将TCGA-ESCC和GSE53625分别作为训练集和验证集，构建了57个预后模型。使用十折交叉验证框架拟合了所有预后模型，并计算了训练集和验证集全部的AUC指数。经过全面筛选，最终确定了包含9个基因的随机森林+广义回归偏最小二乘法(RF+plsRglm)在训练集和验证集中表现出良好的预测能力，具有较高精确度的模型AUC值为0.984 5(图3B)。用模型中的线性回归系数对9个基因计算每个患者的风险评分，其风险分数绝对值越大，在



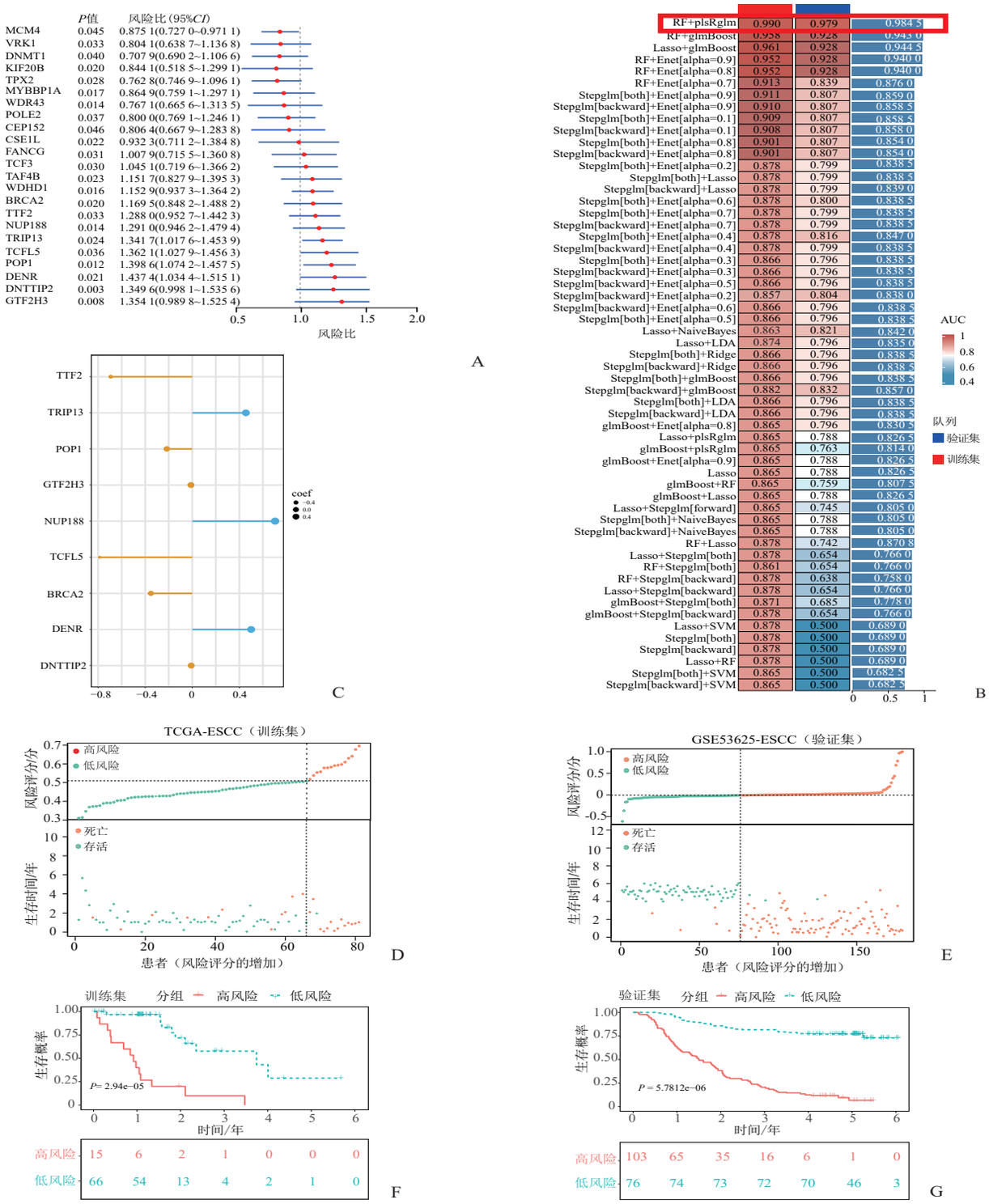
A：韦恩图显示交集基因；B、C：对107个CARGs基因进行GO、KEGG富集分析

图2　中心体扩增相关基因的功能

ESCC的发展和不良预后中的作用也就越大(图3C)。根据风险评分中位数将训练集和验证集中所有患者分配到高风险组或低风险组。随着风险评分的增加,死亡的患者人数也逐渐增加(图3D、E)。此外,高风险组患者的总生存期明显低于低风险组患者(图3F、G)。

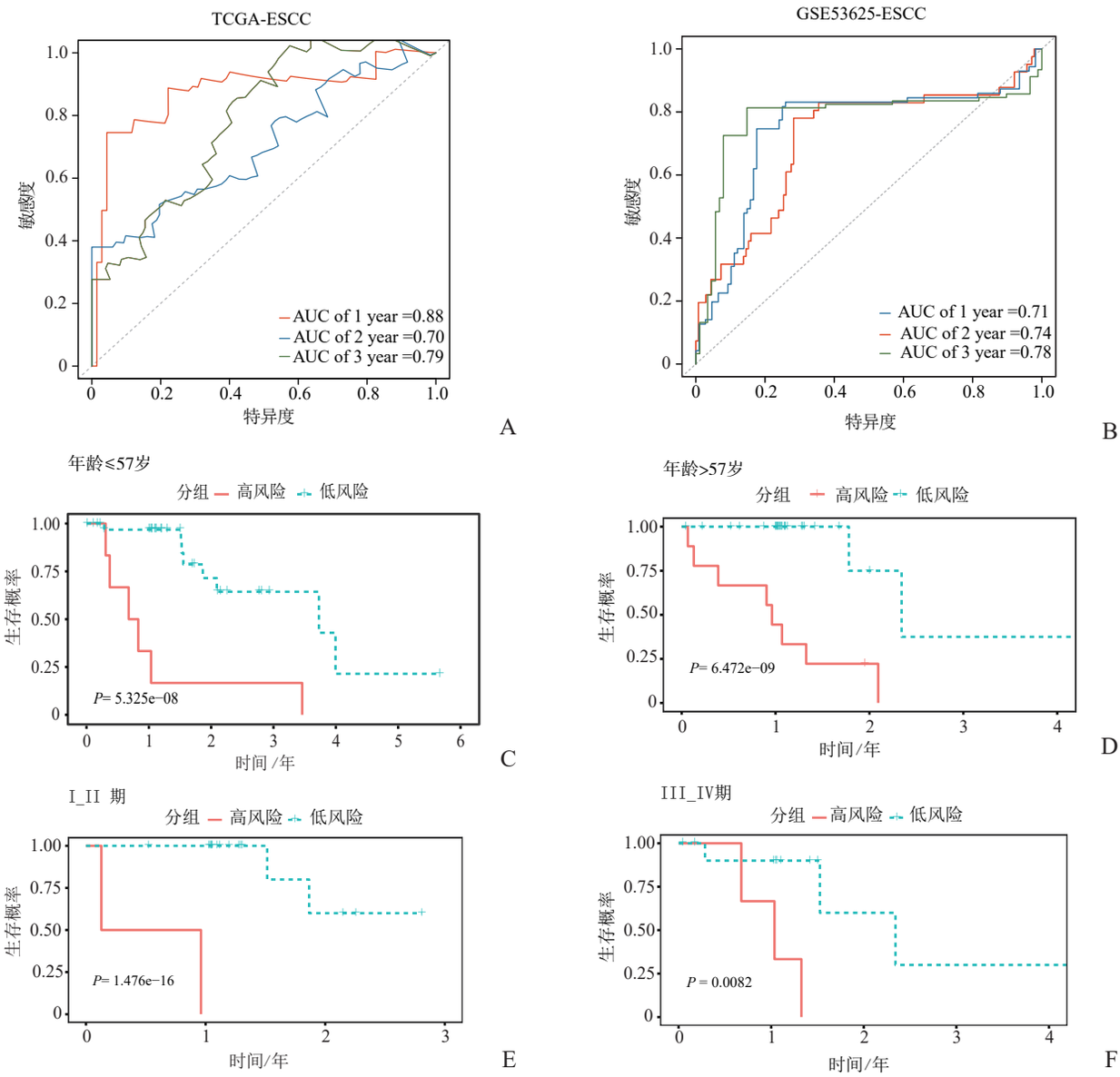
2.3　CARGs预后模型的评估

ROC曲线分析表明,在训练集中CARGs的AUC在1年、2年和3年间时分别达到0.88、0.70和0.79;



A:单因素Cox回归分析得出具有预后意义的基因;B:集成12种机器学习十折交叉验证;C:RF+plsRglm模型中的候选基因;D、E:训练集、验证集采用log-rank检验评估组间生存差异;F、G:训练集、验证集中患者的风险评分和总生存状态的分布

图3　通过机器学习构建CARGs预后模型



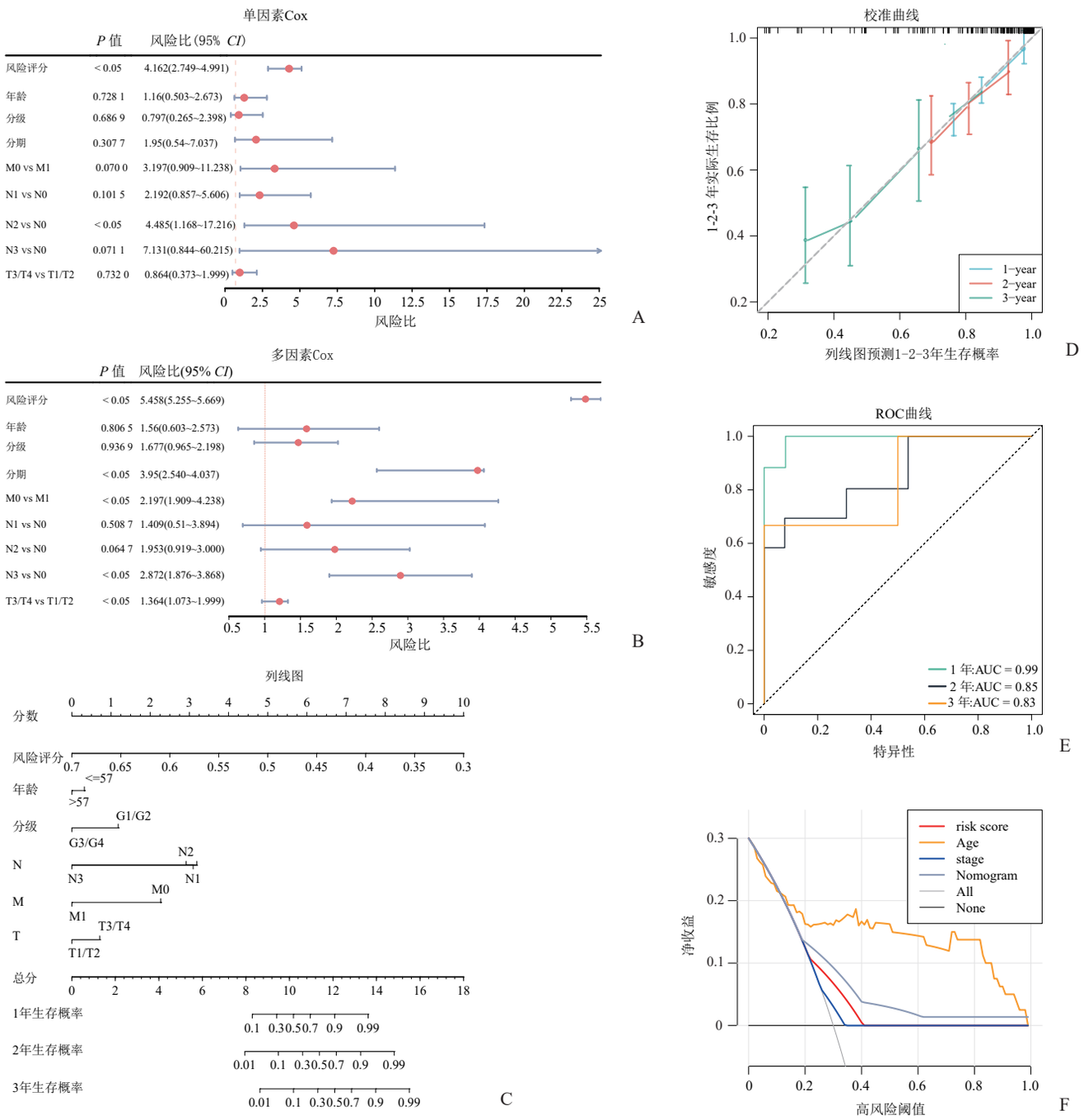
A、B:在训练集和验证集中CARGs预测1、2、3年OS的特异度和敏感度的ROC曲线;C~F: Kaplan-Meier曲线显示CARGs在ESCC患者亚组中表现结果

图4 CARGs模型的评价

验证集中为0.71、0.74和0.78(图4A、B)。这些结果表明CARGs具有评价预后结果能力。通过Kaplan-Meier曲线分析,发现CARGs在不同临床特征(包括年龄和分期)的亚组中显示出对ESCC良好的预测能力(图4C~F)。鉴于CARGs在预测ESCC患者预后方面的优越性能,进一步确定其在预测ESCC发病和进展方面的适用性。

2.4 结合临床特征构建、验证列线图

为了评估CARGs是否为ESCC的独立预后因素,在TCGA-ESCC数据集中对OS进行了单因素和多因素Cox回归分析,结果表明,在单因素Cox分析中CARGs是OS的显著危险因素(*HR*=4.162,*P*<0.05,图5A)。在多因素分析中,CARGs仍然是OS(*HR*=5.458,*P*<0.05,图5B)的独立预后因素,表明其在ESCC患者中具有预测预后的能力。为了使CARGs的预后模型更具有临床适用性,根据CARGs和临床特征构建了列线图(图5C)。由于临床上通常采用临床特征来评估ESCC患者的预后,因此对CARGs与各种临床特征之间的相关性进行了评估。校准曲线显示列线图预测结果与实际观测结果一致(图5D)。列线图的AUC在1年、2年和3年间时分别达到0.99、0.85和0.83,表明其具有较高的预测准确性(图5E)。DCA显示校准曲



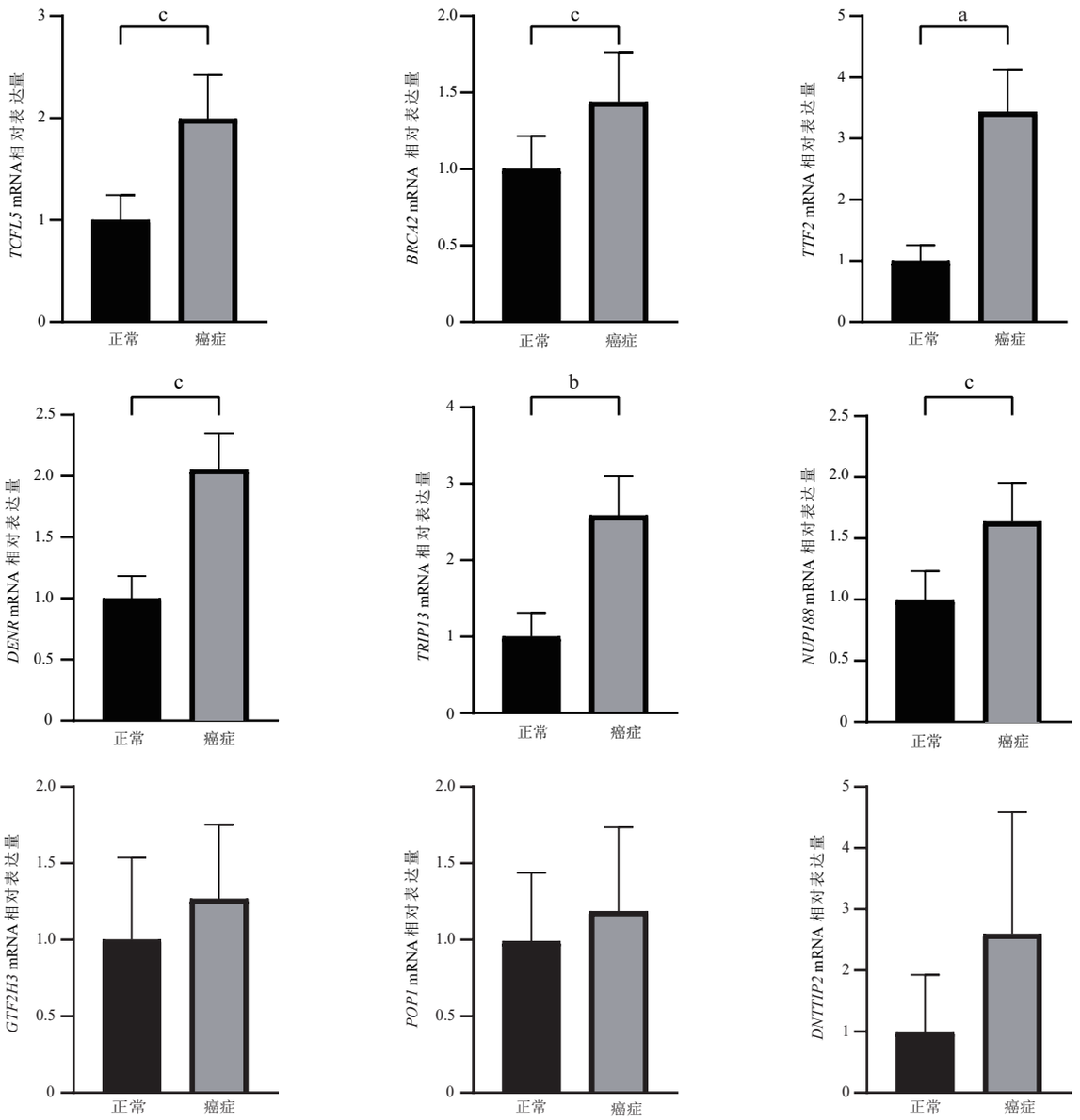
A、B:OS的临床特征和CARGs的单因素和多因素Cox分析;C:根据CARGs和临床特征构建列线图;D:1年、2年、3年OS的校准曲线;E:ROC曲线表示列线图对1年、2年、3年OS的预测性能;F:DCA通过列线图和其他临床特征来显示净收益

图5 生存列线图的构建与验证

线具有很好的临床净收益(图5F)。这些发现表明,基于CARGs的列线图为ESCC的预后预测提供了可靠且准确的工具。

2.5 模型候选基因的表达验证

评估了预后模型的效能后,对9个基因(*DENR*、*TRIP13*、*POP1*、*BRCA2*、*TTF2*、*GTF2H3*、*TCFL5*、*DNTTIP2*和*NUP188*)在ESCC临床样本和正常样本中的表达进行验证。RT-qPCR结果显示,*DENR*、*TRIP13*、*BRCA2*、*TTF2*、*TCFL5*和*NUP188*的mRNA水平在肿瘤样本中表达高于正常样本($P<0.05$),*GTF2H3*、*DNTTIP2*和*POP1*在正常样本和肿瘤样本中的表达无明显差异(图6)。进一步对6个mRNA高表达的基因进行免疫组化实验,验证这些基因在ESCC组织中的蛋白水平的表达。对正常和癌症样本中的免疫组化



a:P<0.001,b:P<0.01,c:P<0.05

图6 候选基因在正常组织和ESCC组织中的mRNA表达量

染色的结果表明，DENR、TRIP13、TTF2和TCFL5的蛋白水平在癌症样本中显著上调；BRCA2和NUP188的蛋白水平在癌症和正常样本中无显著差异(图7)。

3 讨论

ESCC是一种常见的消化道恶性肿瘤，与饮酒、热损伤、氧化应激等多种危险因素有关[20]。中心体扩增是多种恶性肿瘤中公认的肿瘤发生和进展的致癌驱动因素。越来越多的证据表明，中心体扩增与各种类型癌症的预后和治疗反应显著相关。临床研究表明在宫颈鳞状细胞癌和口腔鳞状细胞癌CARGs与肿瘤复发概率呈正相关[21-23]，但CARGs对ESCC的预后影响尚不明确。缺乏可靠的预后生物标志物削弱了ESCC的临床治疗效果，强调进一步研究阐明ESCC的分子机制的重要性。

3.1 CARGs预测ESCC预后的潜力

本研究通过公共数据库中的ESCC数据集筛选DEGs，利用ssGSEA和WGCNA算法取交集后获得107个CARGs。这些基因主要参与DNA复制和调控细胞周期的过程。基于单因素Cox回归分析筛选的23个基因，使预后模型更有针对性。在模型的构建上与之前的研究相比有着明显区别，先前已有研究开发了各

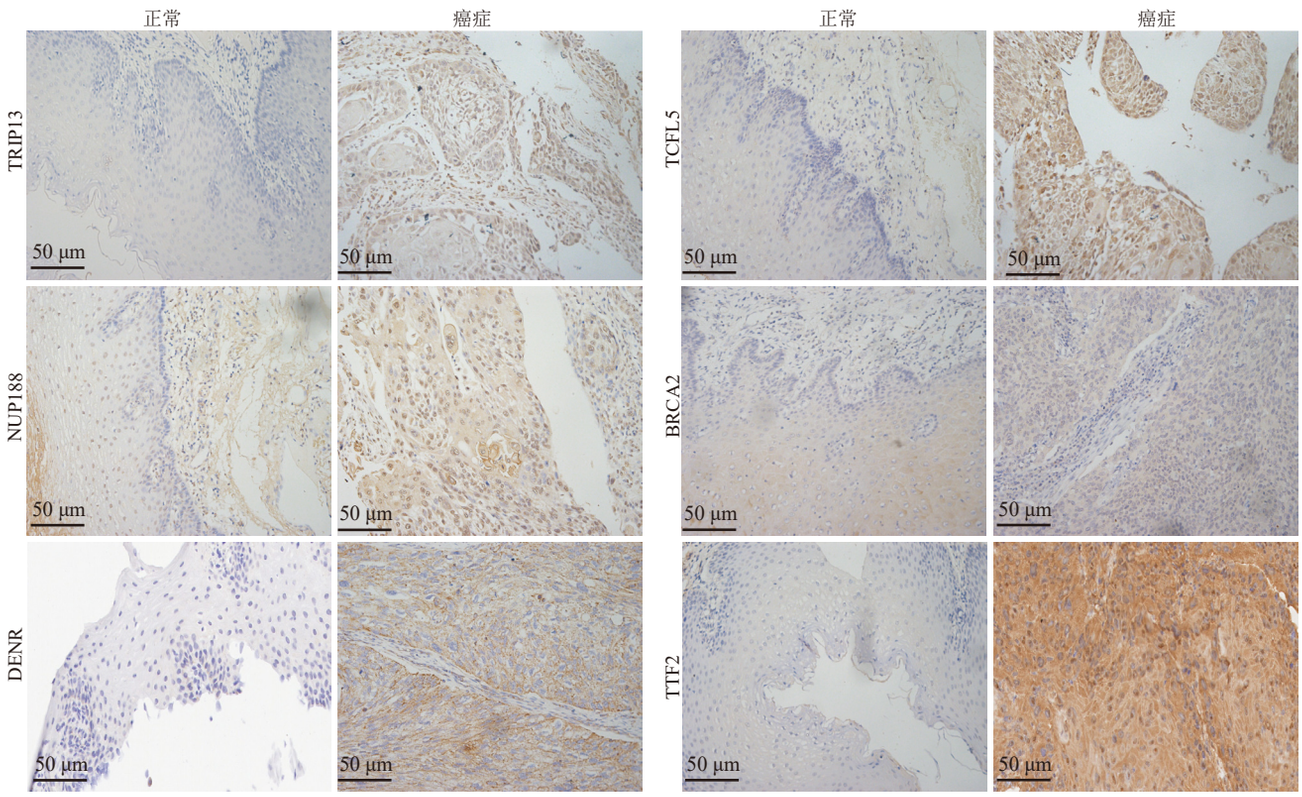


图7 验证候选基因在ESCC组织中的蛋白表达

种模型来预测ESCC患者的生存期[24],本研究全面评估了12种机器学习算法的性能以减轻算法选择对构建ESCC预后模型分析的影响,为ESCC提供了一个高度可靠的预测模型。构建了一个包含9个在ESCC中过表达的CARGs不良预后模型后,利用Kaplan-Meier曲线分析和Cox分析确定了CARGs是ESCC患者生存期的独立预后因素。校准曲线显示预测和观察生存率之间的一致性也证实了该模型的准确性。

通过ESCC临床样本的实验验证了*DENR*、*TRIP13*、*TTF2*和*TCFL5*的mRNA和蛋白的表达水平的上调。已有的文献表明,肿瘤中DENR的增加导致JAK2翻译增加、JAK-STAT信号通路激活,从而增强了肿瘤细胞中PD-L1的表达;并通过降低肿瘤微环境中$CD8^+$T细胞的抗肿瘤杀伤活性来促进肿瘤生长[25]。DI等[26]在体外和体内实验证实了TRIP13的促ESCC的作用,并且这种作用可能与β-catenin和上皮-间质转化有关。TCFL5介导的NRSN2-AS1表达上调可能增强ESCC细胞增殖、存活或迁移,从而促进肿瘤的生长和转移[27]。TTF2转录因子在调控Warburg效应中起着重要的作用,与结肠癌的不良预后有关[28]。表明了该模型筛选出的CARGs有着可作为新的生物标志物或特征用于ESCC患者的预后评估和个性化治疗的可能性,为ESCC的预后预测提供了新的视角。

3.2 机器学习在ESCC预后评估中的应用潜力与挑战

在近年来对于ESCC的研究中,传统的预后评估方法主要依赖于临床病理特征和影像学检查,然而这些方法往往具有主观性和局限性[29]。与之相比,机器学习算法能够处理大量非结构化数据,如基因组数据、影像数据等,并识别复杂的模式,从而提高预后评估的准确性和可靠性[30]。多项研究表明,应用机器学习算法在ESCC的预后模型中显示出了优势。例如,RF、LDA、GBM和SVM用于建立患者生存率的预后预测模型,结果显示这些模型在预测效果上明显优于传统的Cox回归模型。此外,如卷积神经网络在处理医学影像方面的应用,能够有效提取图像特征,进而为ESCC患者的预后评估提供更为精细的依据[31]。然而,机器学习模型的临床应用仍面临一些问题,例如模型的可解释性、数据的异质性以及模型的泛化能力等问题。因此,尽管机器学习技术在ESCC的预后评估中展现出良好的前景,但仍需进一步的研究以验证其在临床实践中的有效性与可靠性。

3.3 研究的局限性

本研究也存在一些局限性:①需要大量、全面的

数据进行分析,以确保模型构建的稳定性和准确性。②应在更大的、多中心甚至泛癌症队列中更广泛地证实CARGs的临床价值,以扩大其应用价值。③还应利用细胞功能实验、体内外实验验证*DENR*、*TRIP13*、*TCFL5*和*TTF2*基因对ESCC细胞的增殖和侵袭转移能力的影响。

综上所述,通过机器学习构建了CARGs的预后模型,利用列线图和校准曲线证明其准确性,证明了其作为预测ESCC患者生存是一个有潜力的临床应用。

利益冲突 本文所有作者声明不存在利益冲突

参考文献:

[1] MORGAN E, SOERJOMATARAM I, RUMGAY H, *et al*. The global landscape of esophageal squamous cell carcinoma and esophageal adenocarcinoma incidence and mortality in 2020 and projections to 2040: new estimates from GLOBOCAN 2020[J]. Gastroenterology, 2022, 163(3): 649-658.e2. DOI:10.1053/j.gastro.2022.05.054.

[2] LIU C Q, MA Y L, QIN Q, *et al*. Epidemiology of esophageal cancer in 2020 and projections to 2030 and 2040[J]. Thorac Cancer, 2023, 14(1): 3-11. DOI: 10.1111/1759-7714.14745.

[3] MILLER K D, NOGUEIRA L, DEVASIA T, *et al*. Cancer treatment and survivorship statistics, 2022[J]. CA Cancer J Clin, 2022, 72(5): 409-436. DOI:10.3322/caac.21731.

[4] COLLABORATORS G 2 O C. The global, regional, and national burden of oesophageal cancer and its attributable risk factors in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017[J]. Lancet Gastroenterol Hepatol, 2020, 5(6): 582-597. DOI:10.1016/S2468-1253(20)30007-8.

[5] ANDERHUB S J, KRÄMER A, MAIER B. Centrosome amplification in tumorigenesis[J]. Cancer Lett, 2012, 322(1): 8-17. DOI:10.1016/j.canlet.2012.02.006.

[6] KIERMAIER E, STÖTZEL I, SCHAPFL M A, *et al*. Amplified centrosomes-more than just a threat[J]. EMBO Rep, 2024, 25(10): 4153-4167. DOI:10.1038/s44319-024-00260-0.

[7] BAKHOUM S F, NGO B, LAUGHNEY A M, *et al*. Chromosomal instability drives metastasis through a cytosolic DNA response[J]. Nature, 2018, 553(7689): 467-472. DOI:10.1038/nature25432.

[8] BAKHOUM S F, LANDAU D A. Chromosomal instability as a driver of tumor heterogeneity and evolution[J]. Cold Spring Harb Perspect Med, 2017, 7(6): a029611. DOI: 10.1101/cshperspect.a029611.

[9] MONTEIRO P, YEON B, WALLIS S S, *et al*. Centrosome amplification fine tunes tubulin acetylation to differentially control intracellular organization[J]. EMBO J, 2023, 42(16): e112812. DOI:10.15252/embj.2022112812.

[10] ZHAO J Z, YE Q, WANG L, *et al*. Centrosome amplification in cancer and cancer-associated human diseases[J]. Biochim Biophys Acta Rev Cancer, 2021, 1876(1): 188566. DOI:10.1016/j.bbcan.2021.188566.

[11] LIU Y L, HE M, KE X R, *et al*. Centrosome amplification-related signature correlated with immune microenvironment and treatment response predicts prognosis and improves diagnosis of hepatocellular carcinoma by integrating machine learning and single-cell analyses[J]. Hepatol Int, 2024, 18(1): 108-130. DOI:10.1007/s12072-023-10538-5.

[12] MATTHEWS H K, BERTOLI C, DE BRUIN R A M. Cell cycle control in cancer[J]. Nat Rev Mol Cell Biol, 2022, 23(1): 74-88. DOI:10.1038/s41580-021-00404-3.

[13] SHAH P, KENDALL F, KHOZIN S, *et al*. Artificial intelligence and machine learning in clinical development: a translational perspective[J]. NPJ Digit Med, 2019, 2: 69. DOI:10.1038/s41746-019-0148-3.

[14] 孟祥勇, 秦嘉怡, 陈文生. 基于机器学习的早期胃癌淋巴结转移预测模型构建与验证[J]. 陆军军医大学学报, 2024, 46(21): 2432-2442. DOI: 10.16016/j.2097-0927.202403126.
MENG X Y, QIN J Y, CHEN W S. Construction and validation of a prediction model for lymph node metastasis in early gastric cancer based on machine learning[J]. J Army Med Univ, 2024, 46(21): 2432-2442. DOI: 10.16016/j.2097-0927.202403126.

[15] 陈金城, 张小勤, 刘杰, 等. 基于机器学习算法构建胸段食管鳞癌患者化疗联合免疫治疗反应预测模型的研究[J]. 陆军军医大学学报, 2025, 47(6): 591-601. DOI: 10.16016/j.2097-0927.202412113.
CHEN J C, ZHANG X Q, LIU J, *et al*. Development of a prediction model for chemotherapy and immunotherapy response in esophageal squamous cell carcinoma patients using machine learning algorithms[J]. J Army Med Univ, 2025, 47(6): 591-601. DOI: 10.16016/j.2097-0927.202412113.

[16] CONSORTIUM G O. Gene ontology consortium: going forward[J]. Nucleic Acids Res, 2015, 43(Database issue): D1049-D1056. DOI:10.1093/nar/gku1179.

[17] LIN A T, SONG L R, WANG Y, *et al*. Future prospects of deep learning in esophageal cancer diagnosis and clinical decision support (Review)[J]. Oncol Lett, 2025, 29(6): 293. DOI:10.3892/ol.2025.15039.

[18] HÄNZELMANN S, CASTELO R, GUINNEY J. GSVA gene set variation analysis for microarray and RNA-seq data[J].

BMC Bioinformatics, 2013, 14: 7. DOI: 10.1186/1471-2105-14-7.

[19] YU G C, WANG L G, HAN Y Y, *et al*. clusterProfiler: an R package for comparing biological themes among gene clusters[J]. OMICS, 2012, 16(5): 284-287. DOI:10.1089/omi.2011.0118.

[20] YANG H, WANG F, HALLEMEIER C L, *et al*. Oesophageal cancer[J]. Lancet, 2024, 404(10466): 1991-2005. DOI:10.1016/s0140-6736(24)02226-8.

[21] DIONNE L K, SHIM K, HOSHI M, *et al*. Centrosome amplification disrupts renal development and causes cystogenesis[J]. J Cell Biol, 2018, 217(7): 2485-2501. DOI:10.1083/jcb.201710019.

[22] BASTO R, BRUNK K, VINADOGROVA T, *et al*. Centrosome amplification can initiate tumorigenesis in flies[J]. Cell, 2008, 133(6): 1032-1042. DOI:10.1016/j.cell.2008.05.039.

[23] GISSELSSON D, JONSON T, YU C, *et al*. Centrosomal abnormalities, multipolar mitoses, and chromosomal instability in head and neck tumours with dysfunctional telomeres[J]. Br J Cancer, 2002, 87(2): 202-207. DOI:10.1038/sj.bjc.6600438.

[24] YU C T, BIAN Y, GAO Y, *et al*. Machine learning-based lactate-related genes signature predicts clinical outcomes and unveils novel therapeutic targets in esophageal squamous cell carcinoma[J]. Cancer Lett, 2025, 613: 217458. DOI:10.1016/j.canlet.2025.217458.

[25] CHEN B W, HU J J, HU X T, *et al*. DENR controls JAK2 translation to induce PD-L1 expression for tumor immune evasion[J]. Nat Commun, 2022, 13(1): 2059. DOI:10.1038/s41467-022-29754-y.

[26] DI S Y, LI M Y, MA Z Q, *et al*. TRIP13 upregulation is correlated with poor prognosis and tumor progression in esophageal squamous cell carcinoma[J]. Pathol Res Pract, 2019, 215(6): 152415. DOI:10.1016/j.prp.2019.04.007.

[27] YAO W J, LIU J, HOU Z Y, *et al*. Recombinant human protein TCFL5-activated NRSN2-AS1 promotes esophageal cancer progression *via* the microRNA-874-5p/RELT regulatory axis[J]. Int J Biol Macromol, 2024, 277(Pt 1): 133814. DOI:10.1016/j.ijbiomac.2024.133814.

[28] DAI W X, MENG X K, MO S B, *et al*. FOXE1 represses cell proliferation and Warburg effect by inhibiting HK2 in colorectal cancer[J]. Cell Commun Signal, 2020, 18(1): 7. DOI:10.1186/s12964-019-0502-8.

[29] EL-SHERBINI A H, CORONEOS S, ZIDAN A L, *et al*. Machine learning as a diagnostic and prognostic tool for predicting thrombosis in cancer patients: a systematic review[J]. Semin Thromb Hemost, 2024, 50(6): 809-816. DOI:10.1055/s-0044-1785482.

[30] KOLASSERI A E, B V. Comparative study of machine learning and statistical survival models for enhancing cervical cancer prognosis and risk factor assessment using SEER data[J]. Sci Rep, 2024, 14(1): 22203. DOI:10.1038/s41598-024-72790-5.

[31] YANG Z, GONG J, LI J, *et al*. The gap before real clinical application of imaging-based machine-learning and radiomic models for chemoradiation outcome prediction in esophageal cancer: a systematic review and meta-analysis[J]. Int J Surg, 2023, 109(8): 2451-2466. DOI:10.1097/JS9.0000000000000441.

(收稿:2025-02-26;修回:2025-05-19)

(编辑 邓强庭)