

基于加权基因共表达网络分析和机器学习的肛瘻 潜在生物标志物筛选及实验研究

庞雪

(山东第一医科大学第一附属医院肛肠科, 济南 250012)

摘要 目的 利用加权基因共表达网络分析(WGCNA)、机器学习算法、免疫浸润分析和动物实验筛选肛瘻(AF)潜在的生物标志物。方法 下载基因表达数据库中包含AF和瘻管旁组织(PF)的转录组数据进行差异分析,对差异表达基因(DEGs)进行基因本体(GO)和京都基因与基因组百科全书(KEGG)信号通路富集分析。整合WGCNA结果和DEGs,筛选AF相关基因。利用最小绝对收缩和选择算子(LASSO)、支持向量机递归特征消除(SVM-RFE)和随机森林(RF)等机器学习方法筛选AF的潜在生物标志物,并进行免疫浸润分析,复制AF大鼠模型进行验证。结果 共获得377个DEGs,主要富集在B细胞受体信号通路、趋化因子信号通路等。机器学习算法筛选出AF的潜在生物标志物基质金属蛋白酶13(MMP13)。AF样本中记忆B细胞、浆细胞、M0巨噬细胞、M1巨噬细胞比例高于PF样本,静息CD4记忆T细胞、静息树突细胞比例低于PF样本。MMP13与M0巨噬细胞、活肥大细胞和幼稚B细胞呈正相关;与静息肥大细胞呈负相关。实验结果显示,大鼠AF样本中MMP13表达水平高于对照组。结论 AF发病涉及多种免疫细胞和信号通路,MMP13在AF组织中表达显著增高,与多种免疫细胞具有相关性,可能成为AF潜在的生物标志物。

关键词 肛瘻;加权基因共表达网络分析;机器学习;生物标志物;免疫浸润

中图分类号 R 657.1+6

文献标志码 A **文章编号** 1000-1492(2024)11-1958-09

doi:10.19405/j.cnki.issn1000-1492.2024.11.010

肛瘻(anal fistula, AF)是指肛管与肛门周围皮肤之间的窦道,以肛门瘙痒、潮湿、疼痛和反复破溃流脓为主要临床表现^[1]。在中国,AF的发病率仅次于痔疮,约占肛门直肠疾病的1.7%~3.6%,国外AF的发病率约为8%~25%^[2]。以清除瘻管和内口为目标的手术是目前肛瘻治疗的主要方法,但术中易出现肛门括约肌损伤,术后创面易愈合不良,影响治疗效果和患者的生活质量^[3]。AF术后复发率高,据统计,AF术后1年内复发率约8.94%~12.92%^[4],瘻管反复发作,可增加癌变的风险^[5]。AF的基础研究主要聚焦于术后创面愈合,涉及炎症、血管生成和细胞凋亡等表型,AF发病的机制尚未明确^[6]。目前,挖掘AF生物标志物的研究开展较少,新的AF生物标志物亟待探索。该研究通过分析AF样本的转录组测序数据,利用加权基因共表达网络分析(weighted gene co-expression network

analysis, WGCNA)、最小绝对收缩和选择算子(least absolute shrinkage and selection operator, LASSO)、支持向量机递归特征消除(support vector machine recursive feature elimination, SVM-RFE)和随机森林(random forest, RF)等机器学习方法筛选AF的潜在生物标志物,探讨标志物和浸润性免疫细胞之间的关系,以期AF发病机制研究提供参考。

1 材料与方法

1.1 转录组数据分析

1.1.1 数据来源 登录基因表达数据库(<http://www.ncbi.nlm.nih.gov/geo>),下载基因芯片数据集“GSE157020”,该数据集包含6例非克罗恩AF和6例瘻管旁组织(perifistule, PF)。

1.1.2 差异表达基因(differential expressed genes, DEGs)筛选和富集分析 使用R语言“limma”数据包读取样本测序数据矩阵,标准化分析后,以基因差异表达倍数绝对值(|logFC|) > 1和矫正后P值 < 0.05为标准,筛选DEGs,绘制火山图和热图可视化DEGs。使用R语言“clusterProfiler”数据包对DEGs进行基因本体论(gene ontology, GO)和京都基因和基因组百科全书(kyoto encyclopedia of genes and genomes,

2024-07-29 接收

基金项目:国家自然科学基金项目(编号:82104876);山东省医药卫生科技发展计划项目(编号:2017WS802);山东省中医药科技发展计划项目(编号:2019-0379)

作者简介:庞雪,女,博士,主治医师,硕士生导师,通信作者, E-mail: pangxue_sdutcm@163.com

KEGG)信号通路富集分析。

1.1.3 WGCNA分析和AF相关DEGs筛选 利用R语言“WGCNA”数据包基于GSE157020数据集中基因表达数据,计算基因间的加权相关性,构建共表达网络。筛选最佳软阈值,根据最佳软阈值构建加权共表达网络。使用“hclust”函数将基因按照表达模式进行层次聚类,将模式相似的基因归为同一个模块,计算每个基因模块与AF病变的相关性。将DEGs和模块基因取交集,获得AF相关DEGs。

1.1.4 机器学习筛选AF生物标志物 使用R语言的“randomForest”数据包行RF分析,使用“glmnet”数据包进行LASSO回归,使用“e1071”数据包进行SVM-RFE分析,筛选与AF的生物标志物。将3种机器学习方法筛选获得的生物标志物取交集,筛选AF关键生物标志物。

1.1.5 免疫浸润分析 使用反卷积算法(CIBERSORT)分析GSE157020数据集中AF组和PF组样本免疫细胞浸润水平,计算每种免疫细胞在两组样本中的比例以及不同免疫细胞之间的相关性。

1.1.6 免疫细胞与AF生物标志物相互作用分析 使用Spearman秩相关检验分析免疫细胞与AF生物标志物之间的关系,使用“ggplot2”R包进行可视化。

1.2 动物实验

1.2.1 实验动物 雄性SPF级SD大鼠20只(6周龄,体质量范围200~220g),购自济南朋悦实验动物繁育有限公司[合格证号:370726221100545624;生产许可证号:SCXK(鲁)20190003]。

1.2.2 试剂与仪器 兔多克隆抗MMP13(AH-10317539,北京博奥森生物技术有限公司),SABC免疫组化试剂盒(SA1052,武汉博士德生物工程有限公司),蛋白浓度测定试剂盒(RPN2232,美国AMERSHAM公司),SDS-PAGE凝胶制备试剂盒(G2037,武汉赛维尔生物科技有限公司)。BX41显微镜(日本OLYMPUS公司),Image J医学图像处理系统(美国国立卫生研究院),SDS-PAGE电泳仪、Mini-PROTEAN Tetra电泳槽、Mini Trans-Blot转膜槽(美国伯乐公司)、凝胶成像系统(上海天能科技有限公司)。

1.2.3 实验动物分组及造模方法 将大鼠随机分为模型组和对照组,每组10只。随机平均分笼饲养。考虑到实验过程中可能出现的大鼠死亡情况,每组选用6只取材研究。对照组不进行任何操作。AF模型复制方法:参照Arakaki et al^[7] AF模型复制

方法,模型组大鼠禁食禁水12h后,按照3 ml/kg的剂量,以10%水合氯醛腹腔注射麻醉,碘伏消毒大鼠肛周手术区,以大鼠鼠尾根部为6点位,在大鼠膝胸位3点位,从肛内距肛缘1 cm处,用缝合针连带螺纹钢,穿透肛门括约肌肌肉,贯穿后固定钢丝。造模后每天检查并活动钢丝,避免脱落及粘连。30d后取材组织病理学检查和实验。该研究已获山东第一医科大学第一附属医院伦理委员会批准(批准文号:2024030401)。

1.2.4 取材及苏木精-伊红(hematoxylin and eosin, HE)染色 造模第30d,用10%的水合氯醛按3 ml/kg体质量进行腹腔麻醉,剔除瘰管组织,均分为两份,分别置于4%多聚甲醛组织固定液和液氮中保存。取材后大鼠予颈椎脱臼处死。取AF组织,石蜡包埋、切片、HE染色后镜下观察。

1.2.5 免疫组化法(immunohistochemistry, IHC)检测样本MMP13蛋白水平 取样本石蜡切片,脱蜡、脱水、透明后,操作步骤如下:①滴加1:200稀释的MMP13一抗孵育过夜;②滴加生物酶标记二抗孵育;③二氨基联苯着色并在显微镜下观察。阳性染色为棕黄色。使用Image J软件分析平均积分吸光度(optical density, OD)值。

1.2.6 Western blot法测定样本MMP13蛋白水平 将组织裂解,匀浆离心后,吸取上清液,进行蛋白定量。加入蛋白上样缓冲液混匀,变性并还原蛋白样本后,依次进行凝胶分离、转膜、封闭。滴加1:1 000稀释的MMP13和1:5 000稀释的GAPDH一抗孵育过夜。TBST清洗,加入1:5 000稀释的二抗孵育2h,曝光并采集图像。使用Image J软件进行灰度值分析。

1.3 统计学处理 采用2人3次测量后取平均值的方法统计IHC阳性产物OD值和Western blot条带的灰度值。使用GraphPad Prism 8.0统计学软件对数据进行分析。符合正态分布的计量资料用 $\bar{x} \pm s$ 表示,组间数据比较采用 t 检验,以 $P < 0.05$ 表示差异有统计学意义。

2 结果

2.1 转录组测序数据分析结果

2.1.1 GSE157020数据集DEGs筛选 GSE157020测序数据标准化处理后,消除了不同样本间的系统性差异,使数据更加适合进行下游分析(图1A、1B)。共获得表达量大于0的基因40 941个,DEGs 377个,其中差异表达上调基因314个,差异表达下调基因62

个。使用火山图和热图可视化 DEGs(图 1C、1D)。

2.1.2 DEGs 富集分析 富集分析结果显示,DEGs 显著富集于 B 细胞受体信号通路、趋化因子信号通路等信号通路(图 2A),以及 B 细胞活化的调节等生物学过程、免疫球蛋白复合物等细胞组分和抗原结合等细胞功能上(图 2B)。

2.1.3 WGCNA 和 AF 相关基因筛选 筛选最佳软阈值为 6,构建无尺度网络(图 3A)。聚类获得基因模块 21 个(图 3B),钢蓝色和三文鱼红色模块与 AF 相关(图 3C),差异有统计学意义。两个模块共包含 1 951 个基因。DEGs 和模块基因取交集,获得 AF 相关基因 296 个(图 3D)。

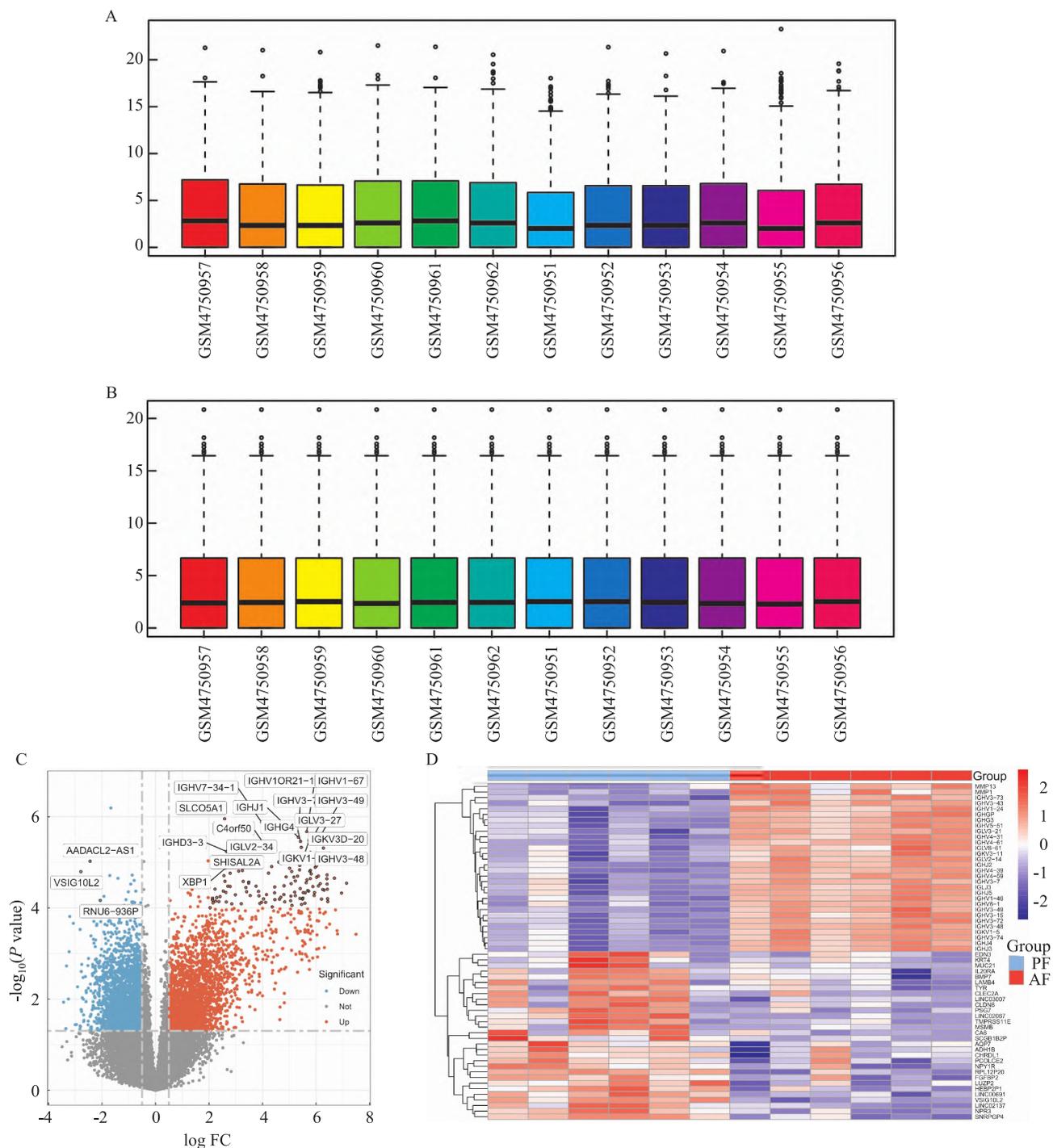


图 1 GSE157020 数据集 DEGs 筛选

Fig. 1 DEGs screening of the GSE157020 dataset

A: Gene expression bar plot before normalization; B: Gene expression bar plot after normalization; C: Volcano plot of DEGs; D: Heatmap of DEGs.

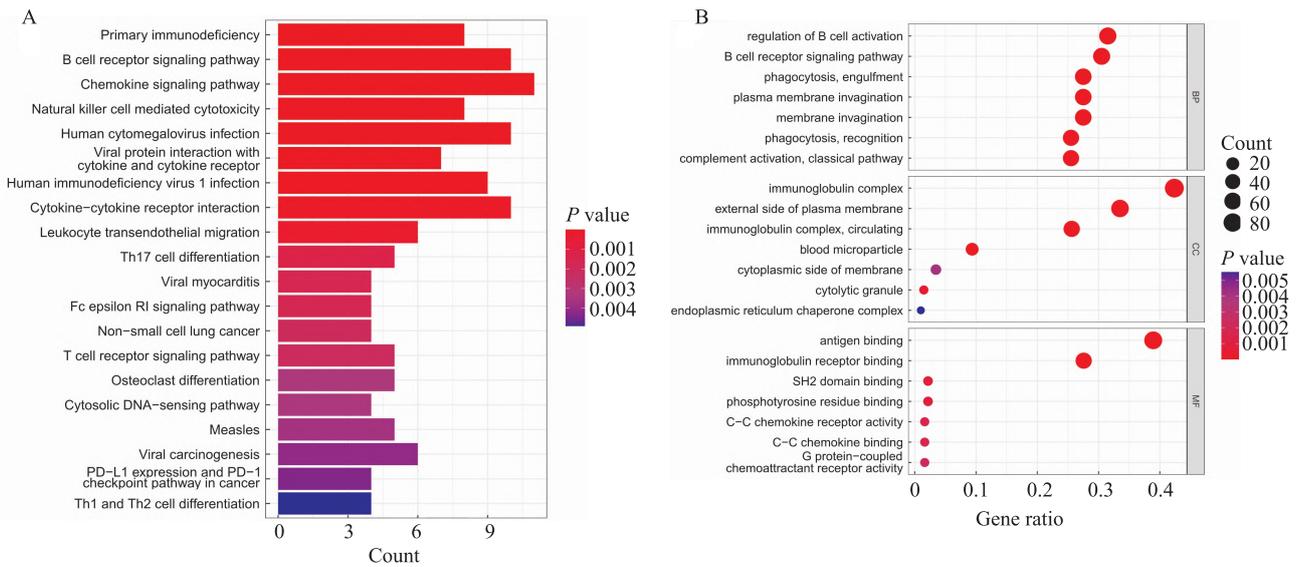


图2 DEGs的KEGG信号通路富集结果和GO富集结果
 Fig. 2 KEGG pathway and GO enrichment results for DEGs
 A: KEGG pathway enrichment results; B: GO enrichment results.

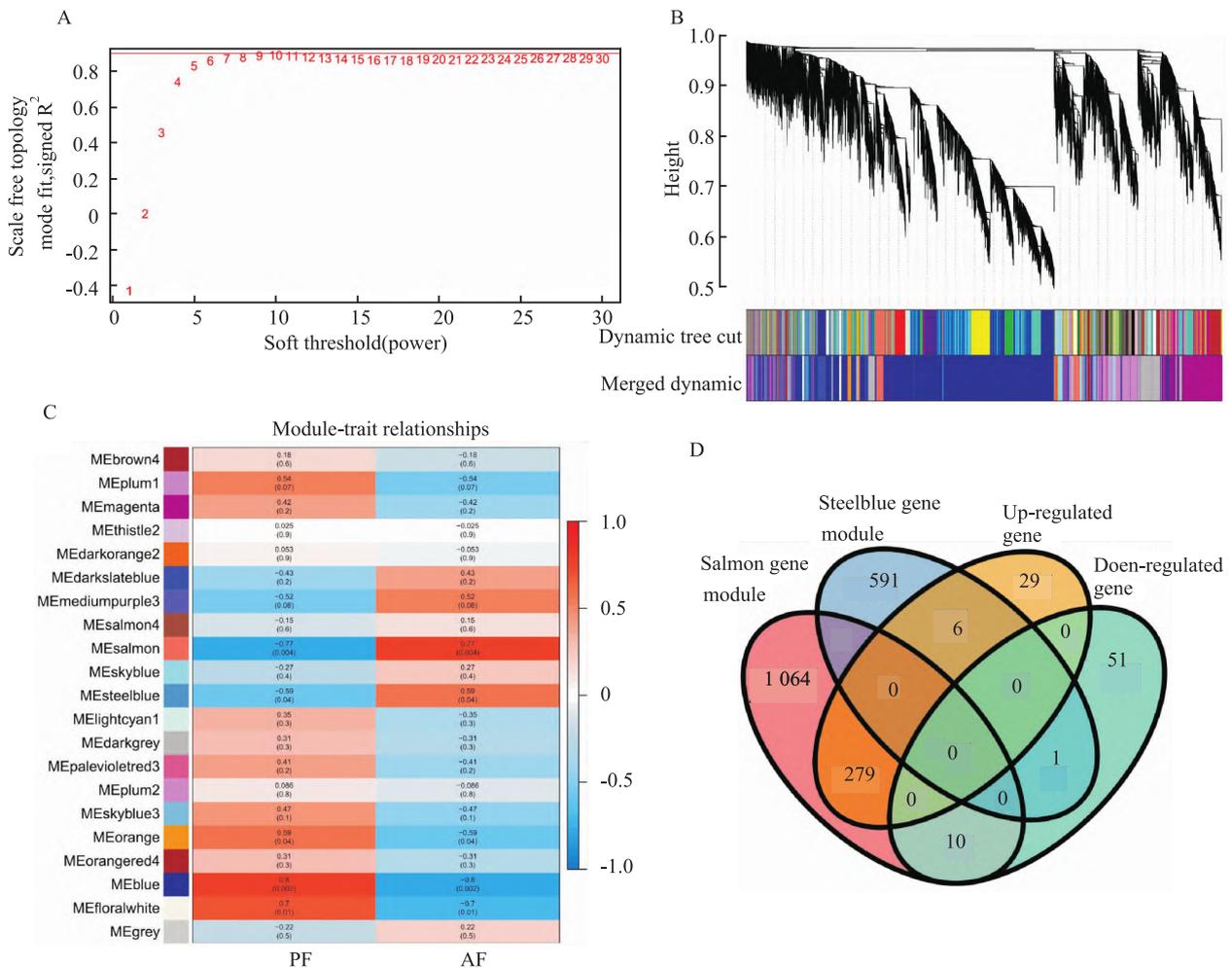


图3 WGCNA和AF相关基因筛选
 Fig. 3 WGCNA and AF-related gene screening

A: Optimal soft-threshold selection in WGCNA; B: Dendrogram of gene module selection in WGCNA; C: Heatmap of module-trait association analysis; D: Venn diagram of AF-related gene screening.

2.1.4 AF 生物标志物识别 SVM-RFE 法识别出生物标志物 12 个(图 4A),LASSO 回归识别出生物标志物 9 个(图 4B),RF 算法识别出生物标志物 16 个(图 4C,4D),将 3 种机器学习筛选出的生物标志物取交集,获得 1 个关键生物标志物基质金属蛋白酶 13(matrix metalloproteinase, MMP13)(图 4E)。

2.1.5 免疫浸润分析 使用 CIBERSORT 算法对 GSE157020 数据集 AF 和 PF 样本进行免疫浸润分析(图 5A)。22 个免疫细胞相关性热图结果显示,滤泡辅助 T 细胞与活化树突细胞、中性粒细胞,活化树突细胞与中性粒细胞呈显著正相关;M2 巨噬细胞滤泡辅助 T 细胞、活化树突细胞、中性粒细胞, M0 巨噬细胞与幼稚 CD4 T 细胞、嗜酸性粒细胞、静息肥大细胞,静息肥大细胞与活化肥大细胞、幼稚 B 细胞呈负相关(图 5B)。AF 样本中记忆 B 细胞、浆细胞、M0 巨噬细胞、M1 巨噬细胞比例高于 PF 样本;AF 样本中静息 CD4 记忆 T 细胞、静息树突细胞比例低于 PF 样本(图 5C)。

2.1.6 MMP13 与免疫细胞的相关性分析 根据 AF 与 PF 样本的相关性分析结果,MMP13 与 M0 巨噬细胞($R = 0.94, P = 0.017$)、活化肥大细胞($R =$

$0.85, P = 0.034$)和幼稚 B 细胞($R = 0.88, P = 0.021$)呈正相关;与静息肥大细胞($R = -0.93, P = 0.0077$)呈负相关(图 6)。

2.2 动物实验

2.2.1 一般情况观察 两组大鼠均无死亡。模型组大鼠可见钢丝活动度可,AF 外口可见分泌物,未见明显红肿(图 7A)。与对照组样本比较,AF 组织 HE 染色可见炎性肉芽组织、纤维结缔组织,伴大量炎性细胞浸润(图 7B,7C)。

2.2.2 大鼠 AF 组织 MMP13 表达量测定 免疫组化结果,模型组大鼠 AF 组织 MMP13 表达高于对照组,差异有统计学意义($t = 5.759, P = 0.0002$)(图 8)。Western blot 法检测结果,模型组 AF 样本 MMP13 蛋白表达量高于对照组,差异有统计学意义($t = 19.255, P < 0.001$)(图 9)。

3 讨论

据统计,约 90% 的 AF 为腺源性感染导致的,一方面,排泄物附着于肛隐窝,细菌经肛隐窝侵入肛腺,诱发感染;另一方面,克罗恩病等炎性肠病、手术创伤、放射治疗等也可损伤肛管皮肤或直肠黏膜,导

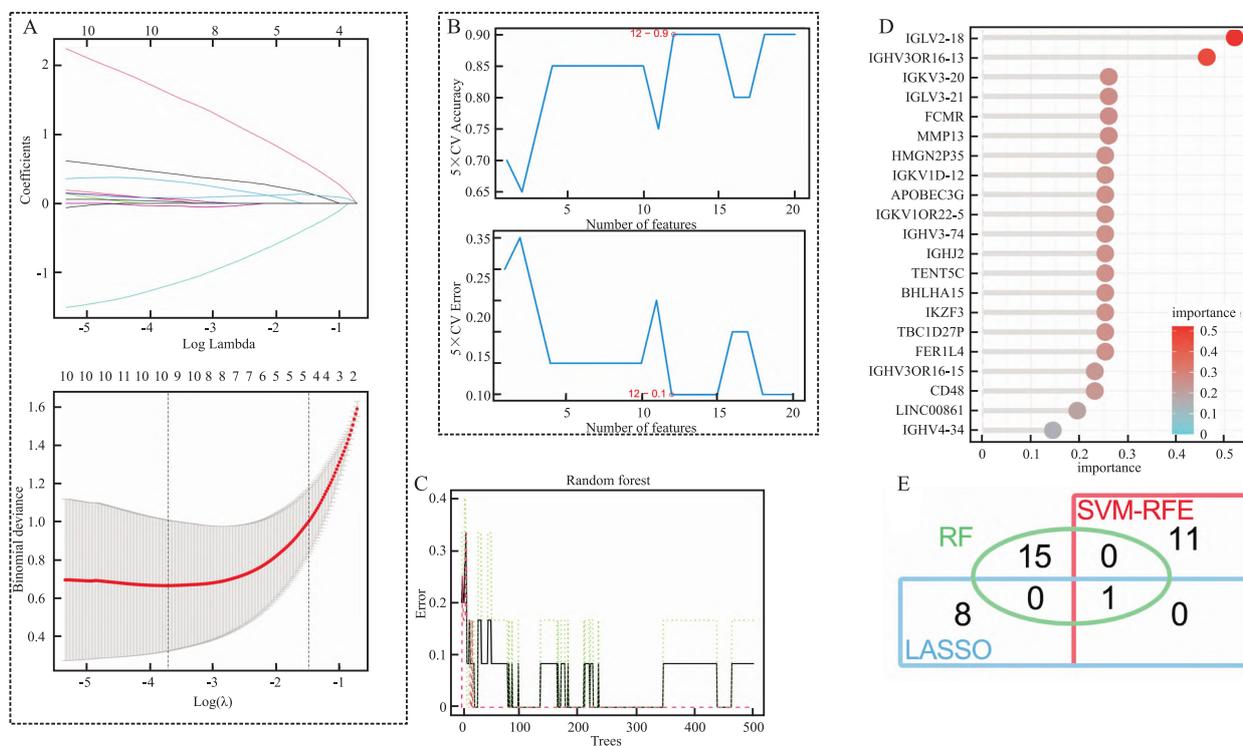


图 4 机器学习识别 AF 生物标志物

Fig. 4 Machine learning identification of AF biomarkers

A: LASSO regression identification of AF biomarkers; B: SVM-RFE identification of AF biomarkers; C: RF identification of AF biomarkers; D: Lollipop plot of the importance of AF biomarkers identified by RF; E: Venn diagram of AF biomarkers identified by machine learning.

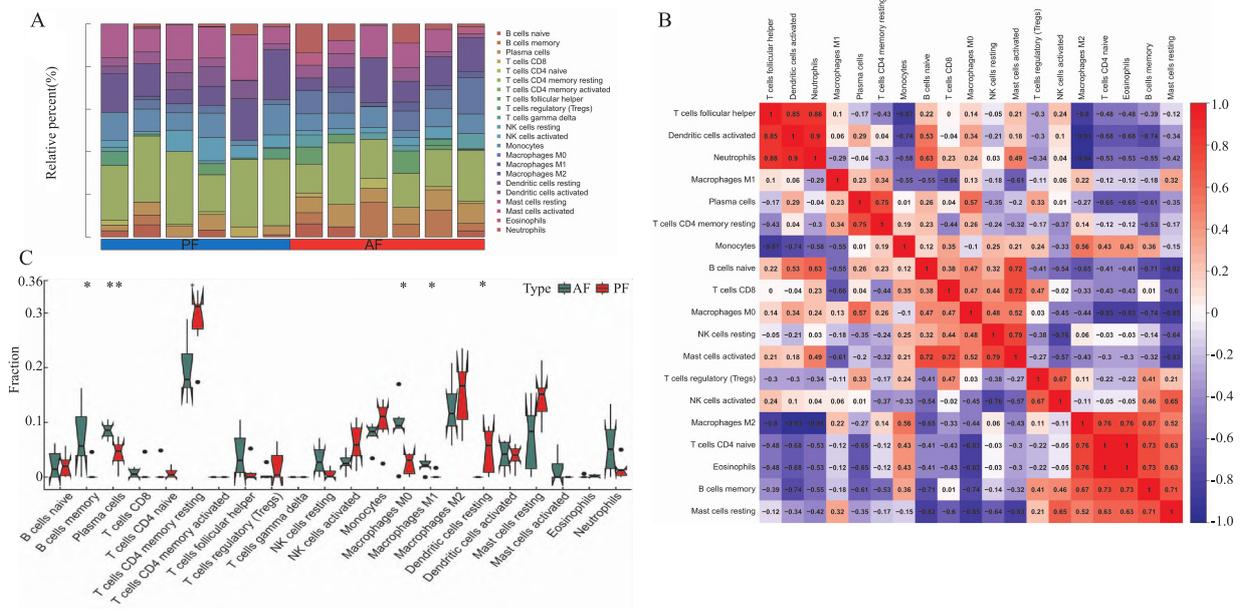


图5 免疫浸润分析
Fig. 5 Immune infiltration analysis

A: Heatmap comparing 22 immune cell subtypes between AF and PF samples; B: Heatmap of correlations among 22 different immune cells in AF and PF samples, with red indicating positive correlations, blue indicating negative correlations, and deeper colors representing stronger correlations; C: Bar plot of the proportion differences of 22 immune cell types between AF and PF samples.

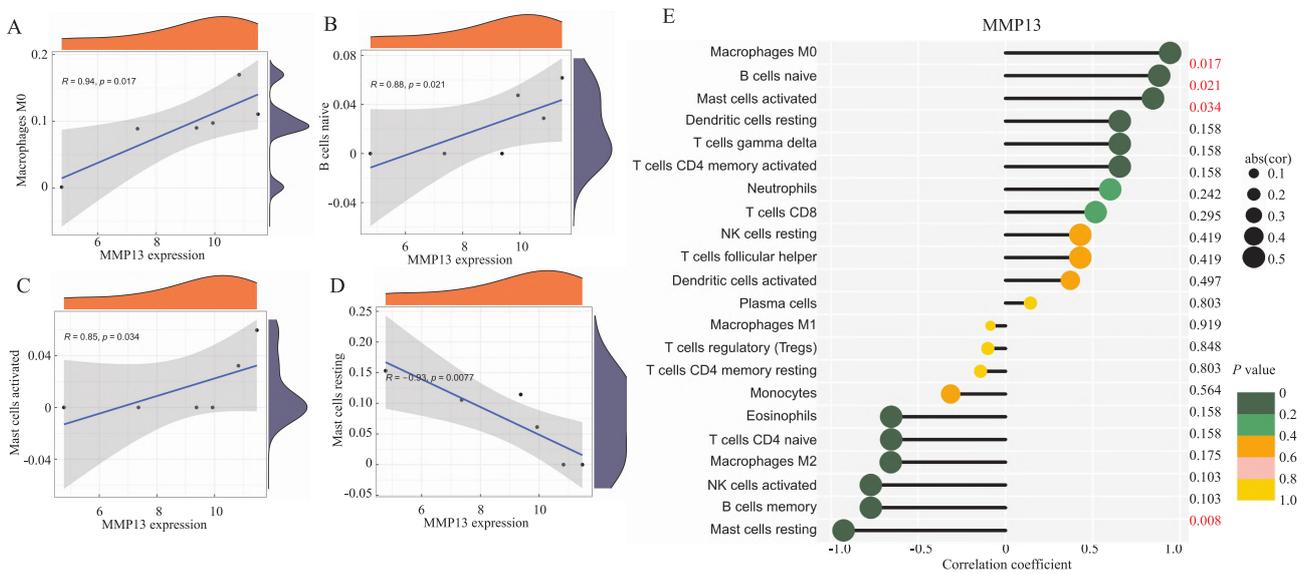


图6 MMP13与免疫细胞的相关性分析
Fig. 6 Correlation analysis of MMP13 with immune cells

A: Correlation analysis of MMP13 with M0 macrophages; B: Correlation analysis of MMP13 with naive B cells; C: Correlation analysis of MMP13 with activated mast cells; D: Correlation analysis of MMP13 with resting mast cells; E: Lollipop plot of the correlation between MMP13 and various immune cells.

致肛隐窝感染,继发 AF^[8]。

该研究通过对 AF 组织转录组测序数据分析显示,AF 相关 DEGs 显著富集于 B 细胞受体信号通路和趋化因子信号通路。B 细胞是免疫系统的重要组

成之一,主要通过 B 细胞表面的受体介导适应性免疫应答。B 细胞受体信号通路广泛参与感染过程中的免疫反应,感染后的组织抗原暴露后,机体产生免疫应答,诱导 B 细胞产生特异性抗原,阻止感染进

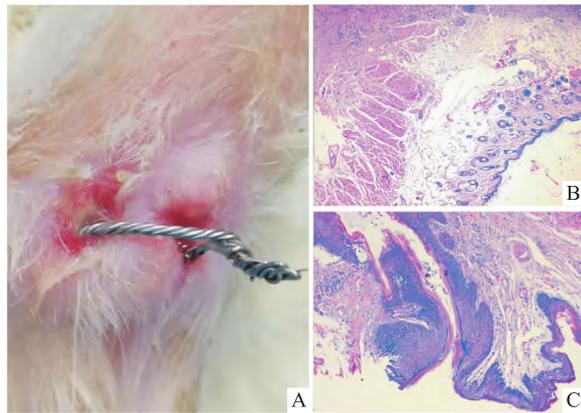


图 7 AF 大鼠模型复制和病理切片

Fig. 7 Replication of AF rat model and histopathological sections

A: AF rat model showing secretion at the AF external opening; B: HE staining of control group rat sample $\times 100$; C: HE staining of model group rat sample $\times 100$.

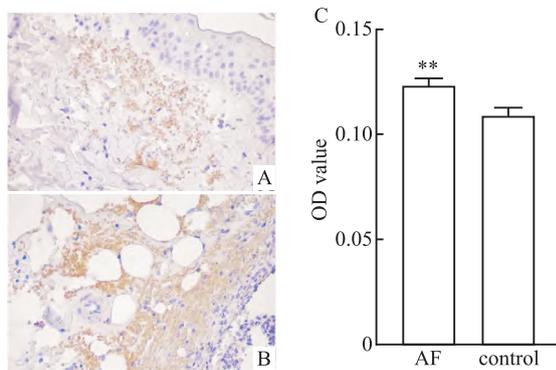


图 8 两组大鼠样本 MMP13 蛋白表达

Fig. 8 MMP13 protein expression in rat samples from both groups

A: Immunohistochemistry of control group rat sample $\times 400$ (SABC); B: Immunohistochemistry of model group rat sample $\times 400$ (SABC); C: Bar plot of OD values; ** $P < 0.01$ vs control group.

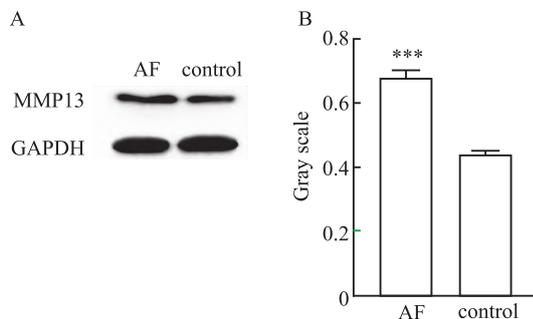


图 9 两组大鼠样本 MMP13 蛋白表达

Fig. 9 MMP13 protein expression in rat samples from both groups

A: Western blot bands; B: Bar plot of grayscale values; *** $P < 0.001$ vs control group.

展^[9]。趋化因子可诱导免疫细胞定向趋化,与免疫和炎症密切相关^[10]。研究^[11]表明,AF 患者血清趋化因子 5 水平呈上调趋势,术后趋化因子 5 水平显著下降。

机器学习是生物信息学研究中的重要手段,可从大量数据中识别并筛选出关键信息,并借助数据分析、模型训练等步骤,挖掘并构建其中最准确的预测模型^[12]。机器学习已在疾病诊断标志物筛选、治疗靶点的识别和预测模型的构建等方面广泛应用,准确度较高^[13-14]。

该研究通过 SVM-RFE、RF 和 LASSO 等机器学习方法,筛选出 AF 的生物标志物 MMP13。目前暂无 MMP13 与 AF 直接的相关性报道,MMP13 可能通过介导炎症浸润、脂肪分化等途径,参与 AF 病理过程。该研究复制 AF 大鼠模型,Western blot 和 IHC 实验结果提示,AF 模型组织中 MMP13 表达呈上调趋势,进一步验证了转录组分析结果,提示 MMP13 可能是新的 AF 发病的关键靶点。

该研究进行了免疫浸润分析,进一步筛选与 MMP13 相关的免疫细胞,结果显示,MMP13 与 MO 巨噬细胞、活肥大细胞和幼稚 B 细胞呈正相关,与静息肥大细胞呈负相关。一般状态下,巨噬细胞分化为 M1 和 M2 表型,这两种表型均参与肛肠术后创面愈合过程^[15]。此外,还存在一种非极化状态的 MO 巨噬细胞,MO 巨噬细胞和肥大细胞与 AF 的相关性尚未明确,与 AF 发病密切相关的细胞类型有待单细胞转录组测序等技术进一步明确。

综上所述,该研究通过对 AF 样本转录组测序数据分析显示,AF 发病涉及 B 细胞受体信号通路、趋化因子信号通路等途径,MMP13 可能是 AF 发病的关键基因,MO 巨噬细胞、肥大细胞和 B 细胞可能是与 AF 发生发展密切相关的免疫细胞亚群。

参考文献

- [1] 陆金根,曹永清,姚一博. 肛肠拖线疗法临床实践指南(2019)[J]. 结直肠肛门外科, 2020, 26(1): 1-4. doi:10.19668/j.cnki.issn1674-0491.2020.01.001.
- [1] Lu J G, Cao Y Q, Yao Y B. Clinical practice guidelines of suture-dragging therapy for anal fistula (2019) [J]. J Colorectal Anal Surg, 2020, 26(1): 1-4. doi:10.19668/j.cnki.issn1674-0491.2020.01.001.
- [2] 陈希琳,冯六泉,姜国丹,等. 肛瘘的诊治专家共识(2020版)[J]. 实用临床医药杂志, 2020, 24(17): 1-7. doi:10.7619/jcmp.202017001.
- [2] Chen X L, Feng L Q, Jiang G D, et al. Consensus of experts on diagnosis and treatment of anal fistula (2020 edition)[J]. J Clin Med Pract, 2020, 24(17): 1-7. doi:10.7619/jcmp.

- 202017001.
- [3] Xu H, Xiao G Z, Zheng Y H, et al. A magnetic resonance imaging-based decision-making tool for predicting complex anal fistulas healing in the early postoperative period[J]. *BMC Gastroenterol*, 2023, 23(1): 372. doi:10.1186/s12876-023-02963-5.
- [4] 徐玉宝, 郑洁, 胡丰良. 视频辅助肛瘘治疗术对比切开挂线术治疗复杂性肛瘘有效性与安全性的 meta 分析[J]. *中国普外基础与临床杂志*, 2023, 30(3): 338-44. doi:10.7507/1007-9424.202208058.
- [4] Xu Y B, Zheng J, Hu F L. Efficacy and safety of video-assisted anal fistula treatment compared with incision and thread drawing in the treatment of complex anal fistula: A meta-analysis[J]. *Chin J Bases Clin Gen Surg*, 2023, 30(3): 338-44. doi:10.7507/1007-9424.202208058.
- [5] Perez S, Eisenstein S. Cancer in anal fistulas[J]. *Clin Colon Rectal Surg*, 2024, 37(1): 41-5. doi:10.1055/s-0043-1762928.
- [6] 刘颖, 余炼, 尹园缘, 等. 基于 Fas/Fas L 通路探讨象皮生肌膏对肛瘘术后模型大鼠创面细胞凋亡的影响[J]. *世界科学技术-中医药现代化*, 2023, 25(8): 2800-9. doi:10.11842/wst.20230102003.
- [6] Liu Y, Yu L, Yin Y Y, et al. To explore the effect of Xiangpi Shengji ointment on the apoptosis of wound cells in model rats after anal fistula operation based on fas/fas L pathway[J]. *Mod Tradit Chin Med Mater Med World Sci Technol*, 2023, 25(8): 2800-9. doi:10.11842/wst.20230102003.
- [7] Arakaki M S, Santos C H M D, Falcão G R, et al. Experimental model of anal fistula in rats[J]. *J Coloproctology*, 2013, 33(3): 135-8. doi:10.1016/j.jcol.2013.04.005.
- [8] Tan Z, Zhu S, Liu C, et al. Causal link between inflammatory bowel disease and fistula: evidence from Mendelian randomization study[J]. *J Clin Med*, 2023, 12(7): 2482. doi:10.3390/jcm12072482.
- [9] Lindenbaum O, Nouri N, Kluger Y, et al. Alignment free identification of clones in B cell receptor repertoires[J]. *Nucleic Acids Res*, 2021, 49(4): e21. doi:10.1093/nar/gkaa1160.
- [10] 张倩倩, 柴向斌. 趋化因子配体 CCL21 与慢性炎症性疾病的关系[J]. *中华临床免疫和变态反应杂志*, 2021, 15(2): 207-12. doi:10.3969/j.issn.1673-8705.2021.02.014.
- [10] Zhang Q Q, Chai X B. Relationship between chemokine ligand CCL21 and chronic inflammatory diseases[J]. *Chin J Allergy Clin Immunol*, 2021, 15(2): 207-12. doi:10.3969/j.issn.1673-8705.2021.02.014.
- [11] 梁艳. 肛瘘病人手术前后血清中 IgA、血清淀粉蛋白 A、趋化因子 5 和炎症因子水平变化与肛门功能及疗效的关系[J]. *临床外科杂志*, 2020, 28(5): 460-3. doi:10.3969/j.issn.1005-6483.2020.05.017.
- [11] Liang Y. Changes of serum IgA, serum amyloid A, chemotactic factor 5 and inflammatory cytokines in patients with anal fistula before and after operation and their relationship with anal function and efficacy[J]. *J Clin Surg*, 2020, 28(5): 460-3. doi:10.3969/j.issn.1005-6483.2020.05.017.
- [12] Greener J G, Kandathil S M, Moffat L, et al. A guide to machine learning for biologists[J]. *Nat Rev Mol Cell Biol*, 2022, 23(1): 40-55. doi:10.1038/s41580-021-00407-0.
- [13] 吴贞琳, 吴煜佳, 张悦瑶, 等. 流感病毒性肺炎线粒体相关生物标志物识别及银翘败毒散潜在作用机制[J]. *中草药*, 2024, 55(3): 882-94. doi:10.7501/j.issn.0253-2670.2024.03.019.
- [13] Wu Z L, Wu Y J, Zhang Y Y, et al. Identification of mitochondria-related biomarkers in influenza viral pneumonia and potential mechanism of Yinqiao Anti-infective Powder[J]. *Chin Tradit Herb Drugs*, 2024, 55(3): 882-94. doi:10.7501/j.issn.0253-2670.2024.03.019.
- [14] 季杰, 齐健, 洪波, 等. 基于 cfDNA 甲基化和机器学习的男性胃癌早筛预测模型的建立[J]. *安徽医科大学学报*, 2022, 57(12): 1991-6. doi:10.19405/j.cnki.issn1000-1492.2022.12.024.
- [14] Ji J, Qi J, Hong B, et al. Construction of prediction model for early screening in male patients with gastric cancer based on cell-free DNA methylation and machine learning[J]. *Acta Univ Med Anhui*, 2022, 57(12): 1991-6. doi:10.19405/j.cnki.issn1000-1492.2022.12.024.
- [15] 曹若曦, 冯文哲. 基于巨噬细胞极化探讨“煨脓长肉”法促进肛瘘术后创面愈合作用机制[J]. *辽宁中医药大学学报*, 2023, 25(6): 65-8. doi:10.13194/j.issn.1673-842x.2023.06.013.
- [15] Cao R X, Feng W Z. Investigate the mechanism of “simmering pus and forming flesh” method promoting wound healing after anal fistula surgery study based on macrophage polarization[J]. *J Liaoning Univ Tradit Chin Med*, 2023, 25(6): 65-8. doi:10.13194/j.issn.1673-842x.2023.06.013.

Screening and experimental study of potential biomarkers for anal fistula based on weighted gene co-expression network analysis and machine learning

Pang Xue

(Dept of Proctology, The First Affiliated Hospital of Shandong First Medical University, Jinan 250012)

Abstract Objective To screen for potential biomarkers of anal fistula (AF) by using weighted gene co-expression network analysis (WGCNA), machine learning, immune infiltration analysis, and animal experiments. **Methods** Download transcriptome data from the gene expression omnibus containing AF and peri-fistula tissue (PF) for differential analysis. Gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analyses on differentially expressed genes (DEGs) were performed. WGCNA results were integrated with

DEGs to screen for genes related to AF. Machine learning methods such as the least absolute shrinkage and selection operator (LASSO), support vector machine recursive feature elimination (SVM-RFE), and random forest (RF) were utilized to screen potential biomarkers for AF. Immune infiltration analysis was conducted and the AF rat model was replicated for validation. **Results** A total of 377 DEGs were obtained, mainly enriched in pathways such as B cell receptor signaling and chemokine signaling. Machine learning algorithms identified matrix metalloproteinase 13 (MMP13) as a potential biomarker for AF. In AF samples, memory B cells, plasma cells, M0 macrophages, and M1 macrophages were higher than in PF samples, while resting CD4 memory T cells and resting dendritic cells were lower than in PF samples. MMP13 showed a positive correlation with M0 macrophages, activated mast cells, and immature B cells, and a negative correlation with resting mast cells. Experimental results showed that MMP13 expression levels were higher in rat AF samples compared to the control group. **Conclusion** The onset of AF involves various immune cells and signaling pathways. MMP13 is significantly upregulated in AF tissue and correlates with multiple immune cells, making it a potential novel biomarker of AF.

Key words anal fistula; weighted gene co-expression network analysis; machine learning; biomarker; immune infiltration

Fund programs National Natural Science Foundation of China (No. 82104876); Shandong Province Medical and Health Science and Technology Development Program (No. 2017WS802); Shandong Province Traditional Chinese Medicine Science and Technology Development Program (No. 2019-0379).

Corresponding author Pang Xue, E-mail: pangxue_sdutcm@163.com

(上接第 1957 页)

different tumors and its effect on tumor staging, prognosis and immunotherapy. **Methods** The expression characteristics of CDKN3 in different cancers using data from TCGA, CCLE, ICGC, and GTEx databases were evaluated. The GEPIA2 platform and Kaplan-Meier analysis were utilized to assess the effect of CDKN3 on tumor pathological staging and survival prognosis. The TIMER platform was employed to explore the influence of CDKN3 on the tumor immune microenvironment and immunotherapy. Its effect on immune checkpoint and key immunotherapeutic predictors using bioinformatics methods was explored. The GeneMANIA tool was used to construct the protein-protein interaction (PPI) network of CDKN3. Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analyses were conducted to explore the biological processes and signaling pathways associated with CDKN3. The effect of CDKN3 on HepG2 cell proliferation, invasion, migration, and apoptosis was validated through transfection with CDKN3 siRNA. **Results** CDKN3 was found to be widely overexpressed in tumors. High expression of CDKN3 was often associated with advanced pathological staging and poor survival prognosis. CDKN3 expression was negatively correlated with most immune checkpoints and positively correlated with tumor mutation burden (TMB), microsatellite instability (MSI), and mismatch repair (MMR) genes. CDKN3 was associated with cell cycle, cellular senescence, and the p53 signaling pathway. Furthermore, EdU staining, JC-1 staining, Transwell, and Wound Healing assays confirmed that CDKN3 promoted HCC cell proliferation, invasion, and migration while inhibiting apoptosis. **Conclusion** Abnormal expression of CDKN3 is closely related to tumor staging, prognosis, and immune microenvironment characteristics, making it a potential prognostic marker and immunotherapy adjuvant target in cancer.

Key words CDKN3; tumor; pan-cancer analysis; prognostic assessment; immunotherapy; biomarkers; hepatocellular carcinoma

Fund programs National Natural Science Foundation of China(No. 82071770); Anhui Province Key Research and Development Project(No. 202004j07020044); Hefei Natural Science Foundation(No. 2022033)

Corresponding authors Shao Wei, E-mail: shaowei@ahmu.edu.cn; Gu Kangsheng, E-mail: gukangsheng@ahmu.edu.cn