四川大学学报（医学版） 2024，55（2）：279－289
J Sichuan Univ（Med Sci） doi: 10.12182/20240360106

在线全文

·骨相关研究·

‖论　著‖

生物信息学和机器学习策略识别骨关节炎炎性衰老生物标志物与临床验证*

周　巧[1,2]，刘　健[3△]，朱　艳[1]，汪　元[3]，王桂珍[3]，齐亚军[2]，胡月迪[2]

1. 安徽中医药大学第二附属医院 老年病一科（合肥 230061）；2. 安徽中医药大学第一临床医学院（合肥 230012）；
3. 安徽中医药大学第一附属医院 风湿病科（合肥 230031）

【摘要】 目的 本研究旨在识别骨关节炎（osteoarthritis, OA）中炎性衰老生物标志物。**方法** GEO（Gene Expression Omnibus）数据库获得年轻OA和老年OA微阵列基因谱，人类衰老基因组资源数据库（Human Aging Genome Resource, HAGR）获得衰老相关基因（aging-related genes, ARGs）。筛选获得年轻OA与老年OA的差异基因，再与ARGs取交集得到OA衰老相关基因。富集分析揭示OA衰老相关标志物的潜在机制。3种机器学习方法识别OA核心衰老标志物，受试者工作特征（receiver operating characteristic curve, ROC）曲线评估其诊断OA炎性衰老的能力。收集临床OA患者外周血单核细胞验证衰老相关分泌表型（senescence-associated secretory phenotype, SASP）因子和衰老标志物的表达。**结果** 总共获得45个衰老相关标志物，主要参与对细胞衰老、细胞周期、炎症反应等的调控。3种机器方法筛选得出5个核心衰老标志物（*FOXO*3、*MCL*1、*SIRT*3、*STAG*1和*S*100*A*13）。纳入20例正常组，40例OA患者，包括年轻组和老年组各20例。与年轻组相比，老年组OA中C反应蛋白（C-reactive protein, CRP）、白细胞介素（interleukin, IL）-6、IL-1β上升，IL-4水平下降（$P<0.01$）；*FOXO*3、*MCL*1、*SIRT*3 mRNA表达下降，*STAG*1和*S*100*A*13 mRNA表达上升（$P<0.01$）。Pearson相关性分析表明选定的标志物与红细胞沉降率（erythrocyte sedimentation rate, ESR）、IL-1β、IL-4、CRP、IL-6指标相关。5个核心衰老基因ROC曲线下面积均大于0.8，列线图预测模型中校正曲线的C-index为0.755，模型校准能力较好。**结论** *FOXO*3、*MCL*1、*SIRT*3、*STAG*1和*S*100*A*13可作为OA炎性衰老的新型诊断分子标志物和潜在治疗靶点。

【关键词】 骨关节炎　炎性衰老　衰老相关分泌表型　机器学习　生物标志物

Identification of Osteoarthritis Inflamm-Aging Biomarkers by Integrating Bioinformatic Analysis and Machine Learning Strategies and the Clinical Validation

ZHOU Qiao[1,2], *LIU Jian*[3△], *ZHU Yan*[1], *WANG Yuan*[3], *WANG Guizhen*[3], *QI Yajun*[2], *HU Yuedi*[2]. 1. *Department of Geriatrics, The Second Affiliated Hospital, Anhui University of Chinese Medicine*, Hefei 230061, China; 2. *First School of Clinical Medicine, Anhui University of Chinese Medicine*, Hefei 230012, China; 3. *Department of Rheumatism Immunity, The First Affiliated Hospital, Anhui University of Chinese Medicine*, Hefei 230031, China

△ Corresponding author, E-mail: liujianahzy@126.com

【Abstract】 Objective To identify inflamm-aging related biomarkers in osteoarthritis (OA). **Methods** Microarray gene profiles of young and aging OA patients were obtained from the Gene Expression Omnibus (GEO) database and aging-related genes (ARGs) were obtained from the Human Aging Genome Resource (HAGR) database. The differentially expressed genes of young OA and older OA patients were screened and then intersected with ARGs to obtain the aging-related genes of OA. Enrichment analysis was performed to reveal the potential mechanisms of aging-related markers in OA. Three machine learning methods were used to identify core senescence markers of OA and the receiver operating characteristic (ROC) curve was used to assess their diagnostic performance. Peripheral blood mononuclear cells were collected from clinical OA patients to verify the expression of senescence-associated secretory phenotype (SASP) factors and senescence markers. **Results** A total of 45 senescence-related markers were obtained, which were mainly involved in the regulation of cellular senescence, the cell cycle, inflammatory response, etc. Through

* 国家自然基金面上项目（No. 82274310）、安徽省高等学校科学研究项目（自然科学类）重点项目（No. 2022AH050449）、安徽省中医药领军人才项目（中医药发展秘〔2018〕23号）、高水平中医药重点学科建设项目——中医痹病学（No. zyyzdxk-2023100）、青年人才培养项目“杏林青秀培育计划”（No. 0500-48-65）、安徽省高校2023年自然科学重大项目（No. 2023AH040112）和安徽现代中医内科应用基础与开发研究省级实验室（No. 2016080503B041）资助

△ 通信作者，E-mail: liujianahzy@126.com

出版日期：2024-03-20

the screening with the three machine learning methods, 5 core senescence biomarkers, including *FOXO*3, *MCL*1, *SIRT*3, *STAG*1, and *S*100*A*13, were obtained. A total of 20 cases of normal controls and 40 cases of OA patients, including 20 cases in the young patient group and 20 in the elderly patient group, were enrolled. Compared with those of the young patient group, C-reactive protein (CRP), interleukin (IL)-6, and IL-1β levels increased and IL-4 levels decreased in the elderly OA patient group (P<0.01); *FOXO*3, *MCL*1, and *SIRT*3 mRNA expression decreased and *STAG*1 and *S*100*A*13 mRNA expression increased (P<0.01). Pearson correlation analysis demonstrated that the selected markers were associated with some indicators, including erythrocyte sedimentation rate (ESR), IL-1β, IL-4, CRP, and IL-6. The area under the ROC curve of the 5 core aging genes was always greater than 0.8 and the C-index of the calibration curve in the nomogram prediction model was 0.755, which suggested the good calibration ability of the model. **Conclusion** *FOXO*3, *MCL*1, *SIRT*3, *STAG*1, and *S*100*A*13 may serve as novel diagnostic biomolecular markers and potential therapeutic targets for OA inflamm-aging.

【Key words】 Osteoarthritis Inflamm-aging Senescence-associated secretory phenotype Machine learning Biomarkers

年龄是骨关节炎(osteoarthritis, OA)的首要危险因素[1]。与年龄相关的炎症状态被称为"炎性衰老"[2]。细胞衰老特征是细胞周期永久停滞和促炎因子释放到周围微环境中,称为衰老相关分泌表型(senescence-associated secretory phenotype, SASP)[3]。衰老过程中表现出共同的特征,分泌促炎细胞因子,如白细胞介素(interleukin, IL)-6、IL-1β和肿瘤坏死因子等,以及端粒侵蚀、TP53和细胞周期蛋白依赖性激酶(cell cycle protein kinase dependent, CDK)抑制剂p21和p16的表达增加[4]。在老年OA患者中发现p16增加和衰老相关β-半乳糖苷酶阳性的成纤维滑膜细胞,衰老的成纤维滑膜细胞分泌促炎因子和基质金属蛋白酶,加重滑膜炎症并导致软骨退化[5]。衰老细胞可向周围环境释放SASP,对免疫细胞产生趋化作用[6]。由免疫细胞和SASP建立的炎症微环境被认为通过降解细胞外基质(extracellular matrix, ECM)驱动软骨退变[7]。因此,炎性衰老与OA的发病密切相关。寻找OA中炎性衰老相关标志物对于OA的诊断和治疗至关重要。

目前,尚无可以延缓OA进展的疾病改善药物[8]。OA诊断主要基于临床表现和影像学检查,但这只能检测到晚期的OA。此外,结构退化的严重程度与临床表现并不完全一致[9]。因此,为OA的早期诊断和治疗寻找新的有效生物标志物至关重要[10]。随着测序的最新发展,综合生物信息学分析已被用于鉴定疾病的新基因,这些基因在疾病的不同阶段有独特表达的分子和蛋白质,被认为是诊断的重要标志物[11]。然而,目前还没有研究分析年轻OA与老年OA之间生物标志物的差异,更没有研究分析这些生物标志物与炎性衰老因子的相关性。

因此,本研究基于Gene Expression Omnibus(GEO)数据库和机器学习算法筛选OA衰老相关生物标志物,临床独立样本集进一步验证OA衰老相关生物标志物和SASP因子的表达及两者的相关性,为OA的早期诊断和治疗提供新的方向。

1 材料和方法

1.1 微阵列数据获取和处理

本研究从NCBI基因表达综合GEO数据库下载年轻OA和老年OA样本的基因表达谱芯片(GSE104113, GSE32317[12], GSE191157[13]),信息如表1所示。使用R包limma对每个数据集进行背景校正和归一化,并使用R包SVA对同一平台的3个数据集进行集成,以去除批处理效应。利用二维PCA聚类图显示样本去除批效应前后的差异。同时,从人类衰老基因组资源(Human Aging Genome Resource, HAGR, https://genomics.senescence.info/)数据库收集人衰老相关基因(aging-related genes, ARGs),包括GenAge和CellAge。

表 1 基因数据集信息

Table 1 Information on the gene datasets

Number	GEO	Platform documents	Young patient group	Elderly patient group
1	GSE104113	GPL6244	3	2
2	GSE32317	GPL570	10	9
3	GSE191157	GPL26963	4	4

1.2 与衰老相关的差异表达基因的鉴定

采用FDR法对P值进行校正,校正后的adj.P< 0.05和$|\log_2 FC(\text{fold change})| \geq 1$作为筛选获得年轻OA和老年OA差异基因(differential expressed genes, DEGs)的标准。与ARGs取交集得到OA衰老相关基因。绘制基因表达热图和火山图,将差异最显著的前20个基因可视化。采用R包Clusterprofiler对差异基因进行基因本体(GO)和京都基因与基因组百科全书(KEGG)富集分析,以校正后P< 0.05为筛选标准。

1.3 识别具有高度相关OA特征的核心衰老标志物

最小绝对收缩和选择算子(LASSO)回归[14]、支持向

量机递归特征消除(SVM-RFE)[15]和随机森林(random forest, RF)[16]是筛选和识别特征基因的方法。LASSO回归是一种常见的机器学习算法。应用R包glmnet将OA衰老相关基因纳入诊断模型,将glmnet函数的α值设置为1,通过5次交叉验证得到最佳λ,最终得到基于最佳λ的衰老特征基因。SVM-RFE是一种基于嵌入式方法的机器学习方法。应用R包e1071找到最佳变量并去除SVM生成的特征向量,从而获得衰老特征基因。RF中的递归特征消除算法(recursive feature elimination, RFE)是一种基于模型的特征选择方法。当决策树设置为500时,衰老特征基因的相对重要性大于1。最后,使用R语言中Venn包过滤3个机器算法得到的衰老特征基因的交集基因。

1.4　患者样本

收集2023年1月–2023年10月于安徽中医药大学第一附属医院风湿科住院的OA患者40例。纳入标准:①符合OA诊断,诊断标准参考中华医学会骨科学分会《骨关节炎诊治指南》的相关标准[17];②X线检查根据Kellgren-Lawrence标准[18]符合1 ~ 3级的患者;③同意参与此次调查并提交知情同意书。排除标准:①合并类风湿关节炎、干燥综合征、强直性脊柱炎、系统性红斑狼疮等其他风湿病;孕妇或哺乳期的女性患者;②合并有循环系统、呼吸系统、造血系统等严重疾病的患者;③精神病患者或不能配合治疗的患者。按照联合国世界卫生组织提出的新的年龄分段,44岁及以下为青年人,45 ~ 59岁为中年人,60岁及以上为老年人。此外,还招募了同期体检中心健康个体作为对照组,对年龄、性别进行一对一倾向性匹配,纳入20例作为对照组。安徽中医药大学附属第一医院伦理委员会对本研究进行了评审和批准(伦理号: No. 2023AH-52)。

1.5　临床炎症指标的测定

晨起采集受试者空腹静脉血5 mL,采用全自动生化分析仪(日立HITACHI7600)检测C反应蛋白(C-reactive protein, CRP)。采用魏氏法测定红细胞沉降率(erythrocyte sedimentation rate, ESR)。

1.6　ELISA法检测SASP因子

采集受试者空腹静脉血5 mL,将静脉血与5 mL磷酸盐缓冲溶液(PBS)轻轻混匀,Ficoll密度梯度离心法分离外周血单个核细胞(peripheral blood mononuclear cells, PBMC),收集各组细胞,离心后弃去细胞沉淀。每孔加100 μL上清液,37 ℃条件下孵育1.5 h,按照ELISA试剂盒说明书,检测IL-1β、IL-4、IL-6(武汉基因美科技有限公司,货号分别为JYM0083Hu、JYM0142Hu和JYM0140Hu)的表达水平。

1.7　实时定量聚合酶链反应检测核心衰老基因表达

使用Trizol(Servicebio)提取受试者外周血单个核细胞总RNA,然后使用Takara Prime Script RT预混液将总RNA逆转录为互补DNA。使用2×SYBR Green qPCR Hub Mix(不含ROX)进行实时定量聚合酶链反应(qRT-PCR)。反应条件: 95 ℃变性5 min, 60 ℃退火1 min, 70 ℃延伸12 s,共40个循环,采用$2^{-\triangle\triangle Ct}$法计算基因相对表达量。表2展示了引物序列。*GAPDH*基因被用作内部参照基因。每个生物样品均进行3次重复实验。

表 2　特异基因引物序列

Table 2　Specific gene primer sequences

Gene	Forward primer (5′→3′)	Reverse primer (5′→3′)
GAPDH	GATGAGATTGGCATGGCTTT	GTCACCTTCACCGTTCCAGT
FOXO3	TCAAGGATAAGGGCGACAGC	GGACCCGCATGAATCGACTA
MCL1	GGGCAGGATTGTGACTCTCATT	GATGCAGCTTTCTTGGTTTATGG
SIRT3	AGGCAGCGGGAGAGCTCA	CGTTTGCGACAGAAACTG
S100A13	TCCTAATGGCAGCAGAACCACTGA	TTCTTCCTGATTTCCTTGGCCAGC
STAG1	GGAATAACACACCGGAGACT	AACCTCAAAGGCAATCCTCA

GAPDH: glyceraldehyde-3-phosphate dehydrogenase; *FOXO3*: forkhead box O3; *MCL1*: BCL2 family member; *SIRT3*: sirtuin 3; *S100A13*: S100 calcium binding protein A13; *STAG1*: STAG1 cohesin complex component.

1.8　受试者工作特征曲线和预测模型的构建

比较正常组、年轻组与老年组外周血单个核细胞中核心衰老基因表达水平,使用受试者工作特征(receiver operating characteristic, ROC)曲线分析核心基因预测OA炎性衰老的效能,参数有准确度、精确度和曲线下面积(area under the curve, *AUC*)。运用R包的rms创建带有衰老特征基因的列线图,其中Points代表候选基因的分数,Total Points代表上面列出的所有基因的分数总和。列线图模型的准确性由校准曲线确定。基于拟合优度检验的校准曲线通过rms包中的calibrate和val.prob函数绘制[19],若检验结果显示$P<0.05$,则表明模型预测值和实际观测值之间存在差异,模型校准度差。C指数即一致性指数(concordance index, C-index)用来评价模型的预测能力,在0.5 ~ 1之间,数值越大可行度越高。0.5为完全随

机，说明该模型没有预测作用，1为完全一致，说明该模型预测结果与实际完全一致[20]。

1.9 统计学方法

运用SPSS 22.0(Chicago, IL, U.S.A.)统计数据，R(4.2.2版)进行机器学习。计量资料采用$\bar{x}\pm s$表示，正态性检验采用Kolmogorov-Smirnov，多组间比较采用单因素方差分析，两两比较采用LSD-t检验。相关性分析采用Pearson分析。$P<0.05$为差异有统计学意义。

2 结果

2.1 GEO数据和HAGR数据库数据集预处理

整合3个GEO数据集，进行标准化和归一化，图1是主成分分析(principal component analysis, PCA)，图1B表明校正后不同批次的样品重叠在一起，批次效应校正成功。采用R包limma和筛选标准鉴定出88个差异基因，其中39个基因在老年OA样本中上调，49个基因在老年OA样本中下调(图2A)。热图展示了上调和下调中前20个差异基因(图2B)。此外，HAGR数据库获得人类衰老相关基因，其中GenAge有307个，CellAge有279个，合并和删除重复基因后，获得543个ARGs进行后续分析。

2.2 OA衰老相关生物标志物功能富集

取GEO数据88个差异基因和HAGR数据库543个ARGs的交集基因作为OA衰老相关标志物，总共45个交集基因(图3A)。其中20个基因在老年OA样本中上调，25个基因下调。按校正后的P值即adj.P.Val从小到大进行排序，表3列出前10个差异基因。

GO分析表明，交集基因主要参与对细胞衰老、细胞周期、单核细胞增殖、成骨细胞分化、炎症反应、老化等的调控(图3B)。KEGG通路分析显示，这些交集基因富集于FOXO信号通路、细胞周期、p53信号通路、细胞凋亡、细胞衰老、破骨细胞分化过程中，且这些通路相互作用密切(图3C)。

2.3 机器学习筛选核心衰老基因

为了提高衰老相关生物标志物诊断OA的准确性，本研究使用了3种机器学习算法，LASSO(图4A、4B)、SVM-RFE(图4C、4D)和RF(图4E、4F)以筛选核心基因。LASSO回归筛选了7个候选基因，SVM-RFE筛选出14个基因，RF算法确定了18个最重要的基因(importance≥1)。综合3种算法的结果，共得到5种重叠基因，分别是*MCL1*、*SIRT3*、*S100A13*、*STAG1*和*FOXO3*(图5)，其中*S100A13*、*STAG1*在老年组OA中上调($|\log_2FC|>1$)，*MCL1*、*SIRT3*和*FOXO3*下调($|\log_2FC|<1$)。

2.4 临床患者ESR、CRP和SASP因子的变化

纳入正常组男7例，女13例，平均年龄为(57.7±5.71)岁。年轻组男7例，女13例，平均年龄为(43.5±3.18)岁。老年组男7例，女13例，平均年龄为(66.4±2.19)岁。与正常组相比，年轻组和老年组ESR、CRP、IL-6和IL-1β上升，IL-4水平下降($P<0.01$)。与年轻组相比，老年组CRP、IL-6和IL-1β均上升，IL-4水平下降($P<0.01$)(表4)。

2.5 核心衰老标志物的表达和相关性分析

与正常组相比，年轻组和老年组OA患者*FOXO3*、*MCL1*、*SIRT3* mRNA表达下降，*STAG1*和*S100A13* mRNA表达上升(均$P<0.01$)；与年轻组相比，老年组*FOXO3*、*MCL1*、*SIRT3* mRNA表达下降，*STAG1*和*S100A13* mRNA表达上升(均$P<0.01$)(图6)。相关性分析结果表明(表5)，*FOXO3*与ESR、CRP、IL-6呈负相关，与

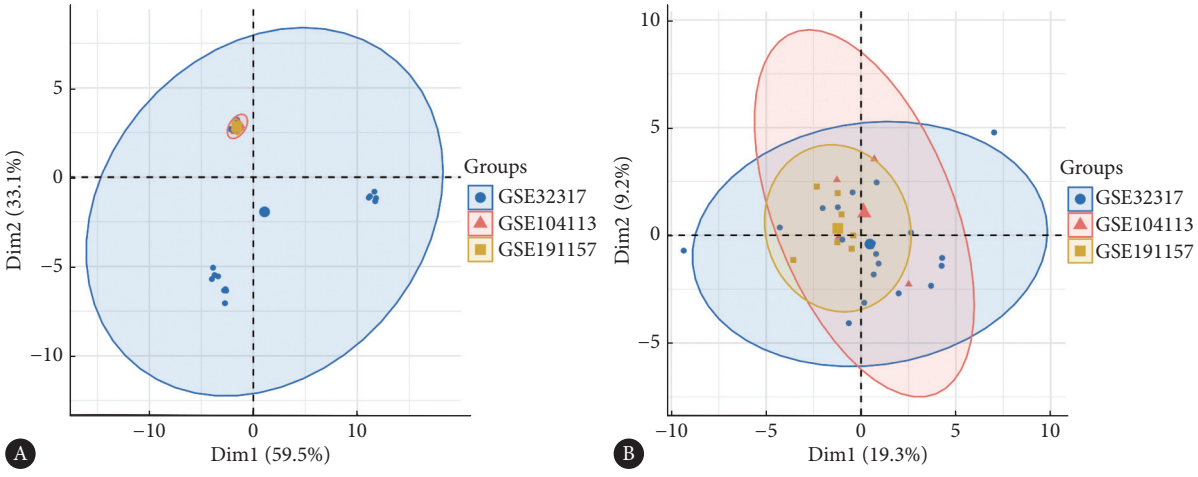

图 1 三个GEO数据集数据整合预处理

Fig 1 Data integration preprocessing for three GEO datasets

A and B are two-dimensional PCA clustering plots for each sample before (A) and after normalization (B) to remove batch effects.

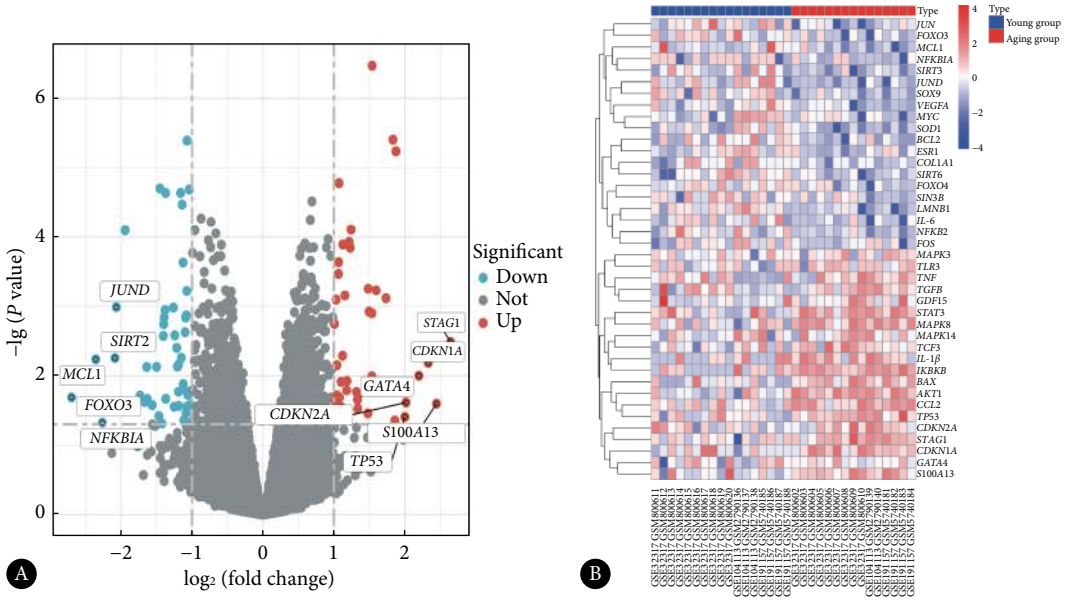

图 2　三个GEO数据集鉴定的DEGs

Fig 2　DEGs identified from the three GEO datasets

A, Volcano map. B, The heat map shows the top 20 DEGs in both up-regulated and down-regulated genes. Each row represents the intersection of genes and each column represents a different group. Red and blue represent up-regulation and down-regulation, respectively. *JUN*: Jun proto-oncogene; *NFKBIA*: NFKB inhibitor alpha; *JUND*: JunD proto-oncogene; *SOX*9: SRY-box transcription factor 9; *VEGFA*: vascular endothelial growth factor A; *MYC*: MYC proto-oncogene; *SOD*1: superoxide dismutase 1; *BCL2*: BCL2 apoptosis regulator; *ESR*1: estrogen receptor 1; *COL1A*1: collagen type Ⅰ alpha 1 chain; *SIRT6*: sirtuin 6; *FOXO*4: forkhead box O4; *SIN3B*: SIN3 transcription regulator family member B; *LMNB*1: lamin B1; *IL*-6: interleukin 6; *NFKB2*: nuclear factor kappa B subunit 2; *FOS*: Fos proto-oncogene; *MAPK*3: mitogen-activated protein kinase 3; *TLR*3: toll like receptor 3; *TNF*: tumor necrosis factor; *TGFB*: transforming growth factor beta 1; *GDF*15: growth differentiation factor 15; *STAT*3: signal transducer and activator of transcription 3; *MAPK*8: mitogen-activated protein kinase 8; *MAPK*14: mitogen-activated protein kinase 14; *TCF*3: transcription factor 3; *IL*-1*β*: interleukin 1 beta; *IKBKB*: inhibitor of nuclear factor kappa B kinase subunit beta; *BAX*: BCL2 associated X; *AKT*1: AKT serine/threonine kinase 1; *CCL2*: C-C motif chemokine ligand 2; *TP*53: tumor protein p53; *CDKN*2*A*: cyclin dependent kinase inhibitor 2A; *CDKN*1*A*: cyclin dependent kinase inhibitor 1A; *GATA*4: GATA binding protein 4.

IL-4呈正相关; *MCL*1与ESR、IL-1β呈负相关, 与IL-4呈正相关; *SIRT*3与ESR、CRP、IL-6和IL-1β呈负相关, 与IL-4呈正相关; *STAG*1与IL-4呈负相关, 与IL-6和IL-1β呈正相关, *S*100*A*13与ESR、IL-6、IL-1β呈正相关, 与IL-4呈负相关。

2.6　核心衰老基因ROC曲线和风险预测模型的构建

利用临床独立验证样本集进行ROC曲线分析(图7A), 结果表明, 5个核心基因对OA炎性衰老具有较高的诊断价值。*FOXO*3诊断价值最高(*AUC*= 0.970), 其他基因的诊断值如下: *S*100*A*13(*AUC*= 0.880)、*STAG*1(*AUC*= 0.917)、*MCL*1(*AUC*= 0.878)和*SIRT*3(*AUC*= 0.947)。基于以上5个核心基因运用R软件绘制出列线图预测模型(图7B、7C)。通过模型上方的标尺, 可以获得5个核心基因所对应的单项得分, 将各单项得分相加即为总得分, 与总得分相对应的预测概率就是预测OA衰老的风险概率。校正曲线的C-index为0.755(95%置信区间: 0.699 ~ 0.811), C-index在0.71 ~ 0.90之间为中等准确度, 说明此Nomogram预测模型准确度较高。Hosmer-Lemeshow拟合优度检验*P*= 0.331, 表明模型校准能力较好。这些结果也表明这5种核心衰老生物标志物整合在OA中也具有很高的诊断价值。

3　讨论

在OA软骨中发现衰老细胞的数量增加[21], 但其机制仍未完全清楚。SASP因子诱导的炎症环境参与软骨退化和软骨下骨重塑, 最终导致软骨丢失和OA进展[22]。细胞衰老和各种因素释放SASP因子可能是导致OA疾病进展的关键机制。生物标志物对于OA的早期诊断至关重要。基因组学、转录组学、蛋白质组学和代谢组学等多组学技术的最新进展促进了疾病生物标志物鉴定新方法的发展。GEO是最大的数据存储库, 包括微阵列芯片、二代测序和其他功能性高通量基因组学数据[23]。机器学习有强大的预测能力和并行处理能力, 可以处理大量变量, 广泛运用于精准医疗中[24]。本研究旨在整合GEO生物信息资源和机器学习策略, 识别OA炎性衰老相关的候选生物标志物, 纳入临床患者检测SASP因子和临床指标, 并验证筛选的生物标志物预测OA炎性衰老的效能, 为OA的早期诊断和治疗提供新的靶点和思路。

整合3个GEO数据集, 数据预处理和limma筛选得到

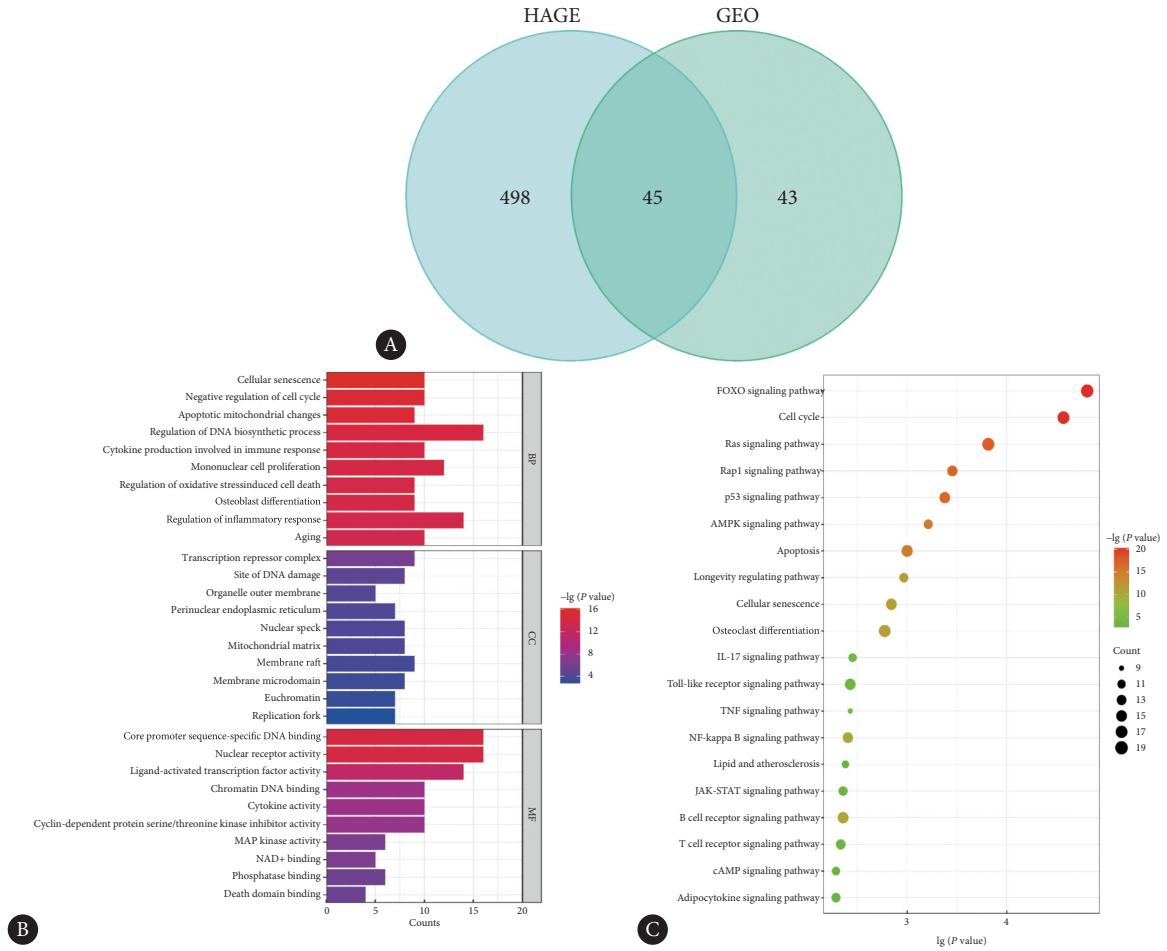

图 3 OA衰老相关标志物的富集分析

Fig 3 Enrichment analysis of markers related to OA aging

A, Venn diagram was used to screen for the intersection genes of GEO and HAGR databases; B, GO analysis; C, KEGG pathway analysis.

表 3 差异表达最显著的前10个lncRNA

Table 3 Top 10 lncRNAs with significant differential expression

Index	GEO data set	Gene	$\log_2$FC	*P* value	adj.*P*.Val
1	GSE104113	*FOXO3*	−2.348	<0.001	<0.001
	GSE32317				
	GSE191157				
2	GSE104113	*MCL1*	−2.178	0.001	0.019
	GSE32317				
	GSE191157				
3	GSE32317	*NFKBIA*	−2.132	<0.001	0.003
	GSE191157				
4	GSE104113	*SIRT3*	−2.043	<0.001	<0.001
	GSE32317				
	GSE191157				
5	GSE104113	*JUND*	−2.033	<0.001	<0.001
	GSE32317				
6	GSE104113	*STAG1*	2.322	<0.001	<0.001
	GSE32317				
	GSE191157				
7	GSE104113	*S100A13*	2.223	<0.001	0.001
	GSE32317				
	GSE191157				
8	GSE32317	*CDKN1A*	2.165	0.001	0.023
	GSE191157				
9	GSE191157	*GATA4*	2.099	<0.001	0.001
10	GSE32317	*CDKN2A*	2.008	<0.001	<0.001

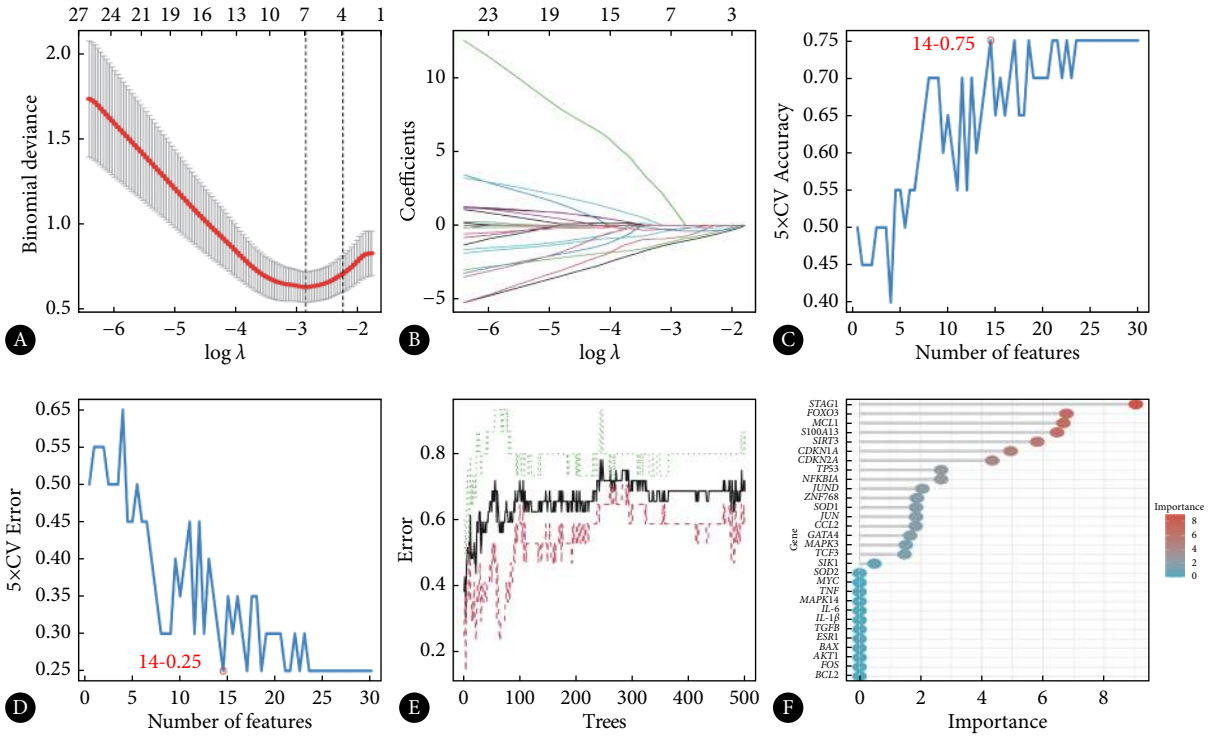

图 4 LASSO回归模型（A、B）、SVM-RFE算法（C、D）和随机森林算法（E、F）

Fig 4 LASSO regression model (A and B), SVM-RFE algorithm (C and D), and random forest algorithm (E and F)

A, The ordinate represents errors of cross-validation; the upper abscissa represents the number of variables corresponding to different λ; the lower abscissa represents log (λ), the logarithm of the lambda penalty coefficient. The parameter corresponding to the dashed line on the left side (lambda.min) with the smallest error indicates seven variables. B, The upper abscissa represents the number of nonzero coefficients in the model; the ordinate represents the value of coefficient; the lower abscissa represents normalized coefficient vector. Lines of different colors represent different variables and each curve represents the change trajectory of the coefficient of each independent variable. 5×CV indicates 5-fold cross validation. C, 14-0.75 indicates an accuracy of 0.75 for the 14 trait genes screened. D, 14-0.25 indicates an error rate of 0.25 for the 14 trait genes screened. E, Dynamic variation of random forest prediction error on the number of random trees, with the vertical coordinate error representing the error and the horizontal coordinate tree representing the tree number. F, Genes ranked by importance.

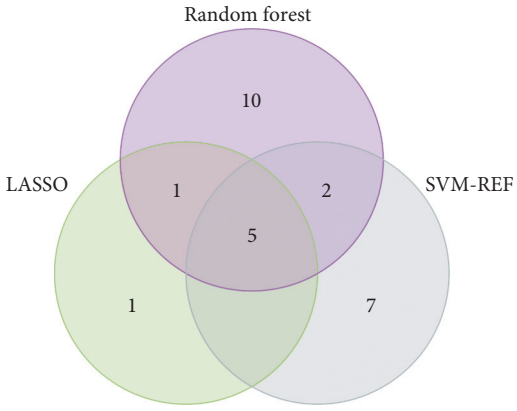

图 5 Venn图筛选3个算法的重叠基因

Fig 5 Venn diagram for screening overlapping genes of the three algorithms

88个年轻OA与老年OA的差异基因，与HAGR数据库获得的543个基因取交集，最终获得45个OA衰老相关基因。根据富集分析，这些衰老相关基因参与细胞衰老、细胞周期调节、炎症反应、转录调控和其他促进OA进展的机制。3种机器学习筛选，识别出5个OA核心衰老相关标志物（*MCL*1、*SIRT*3、*S*100*A*13、*STAG*1和*FOXO*3）。qRT-PCR结果也表明，与年轻组OA相比，老年组OA患者*FOXO*3、*MCL*1、*SIRT*3 mRNA表达下降，*STAG*1和*S*100*A*13 mRNA表达上升，与生物信息学结果一致。基于临床样本的ROC曲线结果表明，OA核心衰老标志物具有出色的预测OA炎性衰老的诊断能力，其ROC曲线下面积均大于0.85。列线图、校正曲线图显示基于5个核心衰老基因的预测模型在OA患者中具有良好的判别效度。

FOXO转录因子是细胞稳态的重要调节因子，在OA衰老过程中对半月板具有保护功能[25-26]。CHEN等[27]研究表明FOXO家族在软骨发育、软骨稳态维持和软骨细胞自噬调控中的重要作用。MCL1（MCL1凋亡调节因子，BCL-2家族成员）是BCL-2蛋白家族的抗凋亡成员，参与细胞凋亡、细胞衰老和炎症的调节，对维持细胞生存和活力至关重要[28]。在OA和衰老的软骨细胞中*MCL*1表达显著下调，miR-34a靶向抑制*MCL*1诱导软骨细胞凋亡，并促进炎症反应和ECM降解[29]。S100A13参与调节许多细胞过程，例如细胞周期进程和分化。S100A13在各种类型

表 4 临床患者ESR、CRP和SASP因子的变化

Table 4 Expression of ESR, CRP, and SASP factors in the patient groups

Indicator	Normal control group (*n*=20)	Young patient group (*n*=20)	Elderly patient group (*n*=20)
ESR/(mm/1 h)	5.50±2.24	20.93±8.22*	25.77±12.49*
CRP/(mg/L)	0.56±0.15	9.07±4.39*	13.05±3.99*,△
IL-4/(ng/L)	32.57±10.83	13.81±2.98*	11.48±0.98*,△
IL-6/(ng/L)	15.09±1.77	25.05±3.69*	34.22±10.47*,△
IL-1β/(ng/L)	17.58±4.56	25.96±4.37*	31.38±8.13*,△

ESR: erythrocyte sedimentation rate; CRP: C-reactive protein; IL: interleukin. * P<0.01, vs. the normal group; △ P<0.01, vs. the young group.

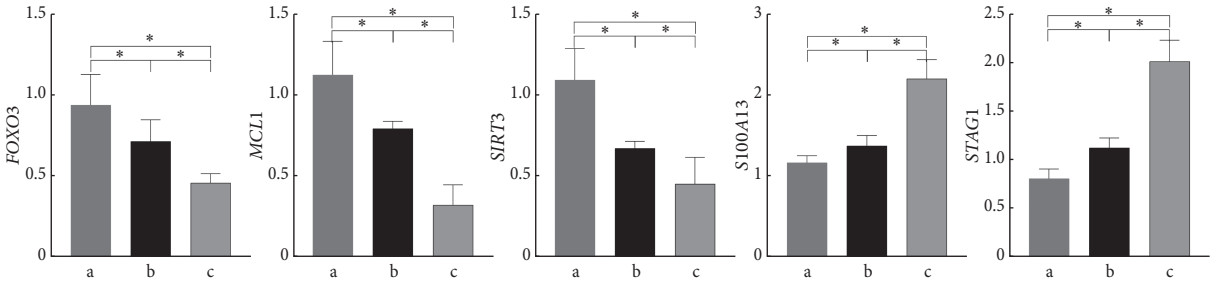

图 6 qRT-PCR检测核心衰老基因的表达

Fig 6 Expression of core senescence genes by qRT-PCR

a: The normal control group; b: the young patient group; c: the elderly patient group. * P<0.01. n=20.

表 5 相关性分析

Table 5 Correlation analysis

Indicator	*FOXO*3		*MCL*1		*SIRT*3		*STAG*1		*S*100*A*13	
	r	*P*	*r*	*P*	*r*	*P*	*r*	*P*	*r*	*P*
ESR/(mm/1 h)	−0.109	0.004	−0.305	0.001	−0.226	<0.001	−0.036	0.342	0.292	<0.001
CRP/(mg/L)	−0.144	<0.001	0.010	0.788	−0.109	0.004	−0.018	0.645	0.066	0.084
IL-4/(ng/L)	0.168	<0.001	0.306	<0.001	0.218	<0.001	−0.181	<0.001	−0.105	0.006
IL-6/(ng/L)	−0.103	0.007	0.044	0.251	−0.381	<0.001	0.084	0.027	0.243	<0.001
IL-1β/(ng/L)	0.072	0.058	−0.077	0.042	−0.120	0.002	0.204	<0.001	0.104	0.006

的细胞衰老过程中上调,增加细胞中IL-1α水平,诱导NF-κB通路表达,最终促进细胞衰老进程[30]。Sirtuin 3(SIRT3)是一种NAD依赖性脱乙酰酶,是维持骨稳态的关键调节分子[31]。SIRT3主要通过维持线粒体稳态和保持线粒体DNA的完整性和功能,在OA中发挥软骨保护作用[32]。STAG1是SCC3家族的成员之一,介导p53依赖性凋亡[33]。前期研究表明,STAG1在OA患者外周血单核细胞中高表达,参与OA免疫炎症反应[34]。综上所述,这些标志物可以调节OA炎性衰老过程,可能是OA中炎性衰老的新兴靶基因。

SASP由一系列促炎因子、趋化因子、生长因子和蛋白酶组成,当细胞衰老时,通过体内和体外多种因子的刺激而产生。滑膜细胞与衰老肌肉骨骼细胞的串扰通过放大重叠的炎症反应或触发SASP来加速OA的发病。本研究纳入正常组和OA患者,OA患者中炎症指标ESR和CRP升高(P< 0.01)。CRP和ESR是临床最常用的血清学炎性指标。CRP和ESR在OA患者体内表达升高且与OA疾病进展、预后等密切相关[35]。SASP因子IL-6和IL-1β在OA患者中升高,IL-4水平下降。细胞膜上的IL-1α可以增强NF-κB和CCAAT/增强子结合蛋白β(C/EBP-β)与DNA的结合能力,进而刺激IL-6和IL-8的转录[36]。典型的SASP因子,如IL-6和IL-1β,可以促进SASP因子分泌,形成前馈环,加速衰老。调节细胞因子如IL-4,是由多种免疫细胞产生的促合成代谢细胞因子,可以通过抑制基质金属蛋白酶调节蛋白多糖代谢抑制软骨细胞的凋亡[37]。相关性分析表明,核心衰老标志物*FOXO*3、*MCL*1、*SIRT*3与ESR、CRP、IL-6呈负相关,与IL-4呈正相关;*STAG*1、*S*100*A*13与IL-4呈负相关,与ESR、IL-6和IL-1β呈正相关。

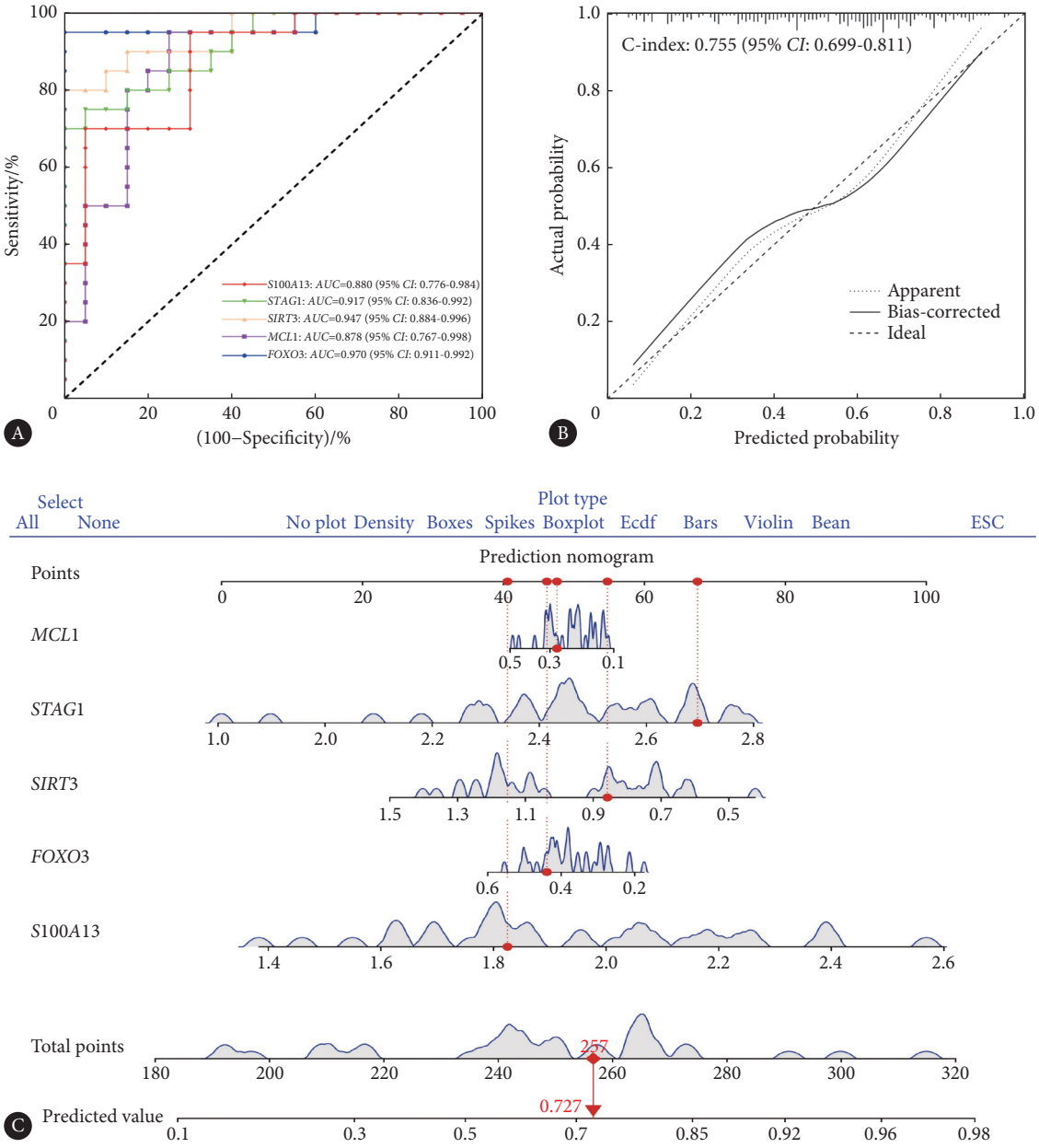

图 7　OA患者衰老核心基因的风险预测模型

Fig 7　Risk prediction model for core senescence genes in OA patients

A, ROC curve. B, Calibration curve. The horizontal coordinate is the predicted event rate (predicted probability). The vertical coordinate is the observed actual event rate (actual probability). The black line is the Apparent line for internal correction. The dotted line is a bias-corrected line of the external correction curve of the model (done by bootstrapping with 1 000 repetitions of the sample). The diagonal line is the reference line (Ideal), i.e., the case where predicted value=actual value. C, Nomogram model.

表明5个核心衰老标志物与SASP因子存在相关性, 炎症可能与衰老协同作用, 并最终促进OA的发生发展。

综上所述, 本研究基于多重生物信息学, 识别了OA中炎性衰老标志物, 可能成为OA诊断和治疗的新靶点, 为OA的诊断、监测和潜在的治疗干预提供了依据和研究方向。但本研究存在一定的局限性, 转录组数据来自公开可用的数据库, 不能排除患者群体和临床特征的复杂性。有限的样本量对结果的准确性有潜在影响, 因此需要更大的样本量和前瞻性研究设计来验证结果的准确性。

*　　*　　*

作者贡献声明　周巧负责论文构思、数据审编、正式分析、软件、验证、可视化和初稿写作, 刘健、朱艳、汪元和王桂珍负责论文构思、数据审编、监督指导和审读与编辑写作, 齐亚军和胡月迪负责正式分析、软件、验证和可视化。所有作者已经同意将文章提交给本刊, 且对将要发表的版本进行最终定稿, 并同意对工作的所有方面负责。

Author Contribution ZHOU Qiao is responsible for conceptualization, data curation, formal analysis, software, validation, visualization, and writing--original draft. LIU Jian, ZHU Yan, WANG Yuan, and WANG Guizhen are responsible for conceptualization, data curation, supervision, and writing--review and editing. QI Yajun and HU Yuedi are responsible for formal analysis, software, validation, and visualization. All authors consented to the submission of the article to the Journal. All authors approved the final version to be published and agreed to take responsibility for all aspects of the work.

利益冲突 所有作者均声明不存在利益冲突

Declaration of Conflicting Interests All authors declare no competing interests.

参 考 文 献

[1] PEAT G, KIADALIRI A, YU D. Disparities in the age at osteoarthritis diagnosis: an indicator for equity-focused prevention. Rheumatology (Oxford), 2023, 62(8): e240–e241. doi: 10.1093/rheumatology/kead080.

[2] ZHOU J, HUANG J, LI Z, *et al*. Identification of aging-related biomarkers and immune infiltration characteristics in osteoarthritis based on bioinformatics analysis and machine learning. Front Immunol, 2023, 14: 1168780. doi: 10.3389/fimmu.2023.1168780.

[3] VICTORELLI S, SALMONOWICZ H, CHAPMAN J, *et al*. Apoptotic stress causes mtDNA release during senescence and drives the SASP. Nature, 2023, 622(7983): 627–636. doi: 10.1038/s41586-023-06621-4.

[4] MAUS M, LÓPEZ-POLO V, MATEO L, *et al*. Iron accumulation drives fibrosis, senescence and the senescence-associated secretory phenotype. Nat Metab, 2023, 5(12): 2111–2130. doi: 10.1038/s42255-023-00928-2.

[5] WU C J, LIU R X, HUAN S W, *et al*. Senescent skeletal cells cross-talk with synovial cells plays a key role in the pathogenesis of osteoarthritis. Arthritis Res Ther, 2022, 24(1): 59. doi: 10.1186/s13075-022-02747-4.

[6] LU R J, WANG E K, BENAYOUN B A. Functional genomics of inflamm-aging and immunosenescence. Brief Funct Genomics, 2022, 21(1): 43–55. doi: 10.1093/bfgp/elab009.

[7] CORYELL P R, DIEKMAN B O, LOESER R F. Mechanisms and therapeutic implications of cellular senescence in osteoarthritis. Nat Rev Rheumatol, 2021, 17(1): 47–57. doi: 10.1038/s41584-020-00533-7.

[8] LI X, ROEMER F W, CICUTTINI F, *et al*. Early knee OA definition-what do we know at this stage? An imaging perspective. Ther Adv Musculoskelet Dis, 2023, 14(15): 1759720X231158204. doi: 10.1177/1759720X231158204.

[9] WANG Q, RUNHAAR J, KLOPPENBURG M, *et al*. Diagnosis of early stage knee osteoarthritis based on early clinical course: data from the CHECK cohort. Arthritis Res Ther, 2021, 23(1): 217. doi: 10.1186/s13075-021-02598-5.

[10] XUAN A, CHEN H, CHEN T, *et al*. The application of machine learning in early diagnosis of osteoarthritis: a narrative review. Ther Adv Musculoskelet Dis, 2023, 14(15): 1759720X231158198. doi: 10.1177/1759720X231158198.

[11] 周巧, 刘健, 忻凌, 等. GEO数据库联合机器学习策略识别骨关节炎特征性lncRNA分子标志物及实验验证. 四川大学学报(医学版), 2023, 54(5): 899–907. doi: 10.12182/20230960101.
ZHOU Q, LIU J, XIN L, *et al*. Identification of characteristic lncRNA molecular markers in osteoarthritis by integrating GEO database andmachine learning strategies and experimental validation. J Sichuan Univ (Med Sci), 2023, 54(5): 899–907. doi: 10.12182/20230960101.

[12] WANG Q, ROZELLE A L, LEPUS C M, *et al*. Identification of a central role for complement in osteoarthritis. Nat Med, 2011, 17(12): 1674–1679. doi: 10.1038/nm.2543.

[13] BALASKAS P, GREEN J A, HAQQI T M, *et al*. Small non-coding RNAome of ageing chondrocytes. Int J Mol Sci, 2020, 21(16): 5675. doi: 10.3390/ijms21165675.

[14] GULER H, GULER E O. Mixed Lasso estimator for stochastic restricted regression models. J Appl Stat, 2021, 48(13/14/15): 2795–2808. doi: 10.1080/02664763.2021.1922614.

[15] NAKAO H, IMAOKA M, HIDA M, *et al*. Determination of individual factors associated with hallux valgus using SVM-RFE. BMC Musculoskelet Disord, 2023, 24(1): 534. doi: 10.1186/s12891-023-06303-2.

[16] SPEISER J L, WOLF B J, CHUNG D, *et al*. BiMM forest: a random forest method for modeling clustered and longitudinal binary outcomes. ChemometrIntell Lab Syst, 2019, 185: 122–134. doi: 10.1016/j.chemolab.2019.01.002.

[17] 高宏伟, 于东旭, 韩继成, 等. 基于循证医学指南的膝关节骨关节炎非手术诊疗方案思考. 长春中医药大学学报, 2022, 38(9): 952–955. doi: 10.13463/j.cnki.cczyy.2022.09.003.
GAO H W, YU D X, HAN J C, *et al*. On the non-operative diagnosis and treatment plan for knee osteoarthritis based on the evidence-basedmedicinequideline. J Changchun Univ Chin, 2022, 38(9): 952–955. doi: 10.13463/j.cnki.cczyy.2022.09.003.

[18] OLSSON S, AKBARIAN E, LIND A, *et al*. Automating classification of osteoarthritis according to Kellgren-Lawrence in the knee using deep learning in an unfiltered adult population. BMC Musculoskelet Disord, 2021, 22(1): 844. doi: 10.1186/s12891-021-04722-7.

[19] JIN Y, KATTAN M W. Methodologic issues specific to prediction model development and evaluation. Chest, 2023, 164(5): 1281–1289. doi: 10.1016/j.chest.2023.06.038.

[20] SU W, HE B, ZHANG Y D, *et al*. C-index regression for recurrent event data. Contemp Clin Trials, 2022, 118: 106787. doi: 10.1016/j.cct.2022.106787.

[21] 段戡, 戴易, 宗亿洲, 等. 细胞衰老在骨关节炎中的研究进展. 医学研究杂志, 2023, 52(10): 6–9. doi: 10.11969/j.issn.1673-548X.2023.10.002.
DUAN K, DAI Y, ZONG Y Z, *et al*. Research progress of cellular senescence in osteoarthritis. J Med Res, 2023, 52(10): 6–9. doi: 10.11969/j.issn.1673-548X.2023.10.002.

[22] BOONE I, TUERLINGS M, COUTINHO De ALMEIDA R, *et al*. Identified senescence endotypes in aged cartilage are reflected in the blood metabolome. Geroscience, 2024, 46(2): 2359–2369. doi: 10.1007/s11357-023-01001-2.

[23] CLOUGH E, BARRETT T, WILHITE S E, *et al*. NCBI GEO: archive for

gene expression and epigenomics data sets: 23-year update. Nucleic Acids Res, 2024, 52(D1): D138–D144. doi: 10.1093/nar/gkad965.

[24] GONG Y, DING W, WANG P, *et al*. Evaluating machine learning methods of analyzing multiclass metabolomics. J Chem Inf Model, 2023, 63(24): 7628–7641. doi: 10.1021/acs.jcim.3c01525.

[25] ZHOU Q, WANG W, WU J, *et al*. Ubiquitin-specific protease 3 attenuates interleukin-1β-mediated chondrocyte senescence by deacetylating forkhead box O-3 via sirtuin-3. Bioengineered, 2022, 13(2): 2017–2027. doi: 10.1080/21655979.2021.2012552.

[26] LEE K I, CHOI S, MATSUZAKI T, *et al*. FOXO1 and FOXO3 transcription factors have unique functions in meniscus development and homeostasis during aging and osteoarthritis. Proc Natl Acad Sci U S A, 2020, 117(6): 3135–3143. doi: 10.1073/pnas.1918673117.

[27] CHEN C, XU G, CHEN J, *et al*. Decreased FoxO1 expression contributes to facet joint osteoarthritis pathogenesis by impairing chondrocyte migration and extracellular matrix synthesis. Cell Signal, 2024, 113: 110942. doi: 10.1016/j.cellsig.2023.110942.

[28] WIDDEN H, PLACZEK W J. The multiple mechanisms of Mcl1 in the regulation of cell fate. Commun Biol, 2021, 4(1): 1029. doi: 10.1038/s42003-021-02564-6.

[29] XIONG S, ZHAO Y, XU T. DNA Methyltransferase 3 beta mediates the methylation of the microrna-34a promoter and enhances chondrocyte viability in osteoarthritis. Bioengineered, 2021, 12(2): 11138–11155. doi: 10.1080/21655979.2021.2005308.

[30] SU Y, XU C, SUN Z, *et al*. S100A13 promotes senescence-associated secretory phenotype and cellular senescence via modulation of non-classical secretion of IL-1α. Aging (Albany NY), 2019, 11(2): 549–572. doi: 10.18632/aging.101760.

[31] DIAO Z, JI Q, WU Z, *et al*. SIRT3 consolidates heterochromatin and counteracts senescence. Nucleic Acids Res, 2021, 49(8): 4203–4219. doi: 10.1093/nar/gkab161.

[32] BAO J, LIN C, ZHOU X, *et al*. circFAM160A2 promotes mitochondrial stabilization and apoptosis reduction in osteoarthritis chondrocytes by targeting miR-505-3p and SIRT3. Oxid Med Cell Longev, 2021, 2021: 5712280. doi: 10.1155/2021/5712280.

[33] ALONSO-GIL D, CUADRADO A, GIMÉNEZ-LLORENTE D, *et al*. Different NIPBL requirements of cohesin-STAG1 and cohesin-STAG2. Nat Commun, 2023, 14(1): 1326. doi: 10.1038/s41467-023-36900-7.

[34] ZHOU Q, LIU J, XIN L, *et al*. The diagnostic features of peripheral blood biomarkers in identifying osteoarthritis individuals: machine learning strategies and clinical evidence. Curr Comput Aided Drug Des, 2023. doi: 10.2174/1573409920666230818092427.

[35] KRISHNA A, GARG S, GUPTA S, *et al*. C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR) trends following total hip and knee arthroplasties in an indian population--a prospective study. Malays Orthop J, 2021, 15(2): 143–150. doi: 10.5704/MOJ.2107.021.

[36] CAIADO F, KOVTONYUK L V, GONULLU N G, *et al*. Aging drives Tet2+/- clonal hematopoiesis via IL-1 signaling. Blood, 2023, 141(8): 886–903. doi: 10.1182/blood.2022016835.

[37] Van HELVOORT E M, Van Der HEIJDEN E, Van ROON J A G, *et al*. The role of interleukin-4 and interleukin-10 in osteoarthritic joint disease: a systematic narrative review. Cartilage, 2022, 13(2): 19476035221098167. doi: 10.1177/19476035221098167.

（2023 – 12 – 03收稿, 2024 – 02 – 03修回）

编辑　余　琳