·论著·一次研究·

基于生物信息学的急性髓系白血病患者预后风险模型的构建与验证

张斯娜[1]，马书杰[2]

1. 淄博市疾病预防控制中心性病艾滋病防制所（山东淄博 255020）
2. 河南理工大学医学院（河南焦作 454003）

【摘要】目的 构建并验证急性髓系白血病（acute myeloid leukemia, AML）患者预后风险模型。**方法** 基于癌症基因组图谱（The Cancer Genome Atlas, TCGA）数据库和基因型－组织表达（Genotype–Tissue Expression, GTEx）数据库，获取 AML 样本和正常骨髓样本的基因测序数据，进行基因差异分析，筛选 AML 生存相关基因，并利用随机森林模型构建基于特征基因的风险评分模型，通过基因本体（Gene Ontology, GO）富集分析探讨预测模型的生物学功能。**结果** 根据 Kaplan–Meier 法和受试者工作特征（receiver operating characteristic, ROC）曲线分析结果，存活率预测在训练集和测试集中表现出良好的敏感性和特异性。多变量 Cox 回归分析和分层分析显示，基因预后风险模型可以独立预测 AML 患者的生存情况。GO 富集分析显示，在高风险组和低风险组中，差异表达基因在与已知的 AML 相关生物学过程中富集。**结论** 利用生物信息学方法，确定了一个包含 5 个基因（*ACOT7*、*SFXN3*、*SLC29A2*、*HINT2* 和 *ZMAT1*）的 mRNA 标记，可能作为 AML 的潜在预后标记，为 AML 患者的个性化风险评估和诊疗方案制订提供参考。

【关键词】急性髓系白血病；生物信息学；癌症基因组图谱；基因型－组织表达；预后模型；随机森林

Construction and validation of a prognostic risk model for patients with acute myeloid leukemia based on bioinformatics

Si-Na ZHANG[1], Shu-Jie MA[2]

1. Institute of Sexually Transmitted Disease and AIDS Prevention and Control, Zibo Municipal Center for Disease Control and Prevention, Zibo 255020, Shandong Province, China
2. School of Medicine, Henan Polytechnic University, Jiaozuo 454003, Henan Province, China
Corresponding author: Si-Na ZHANG, Email: zasana@126.com

【Abstract】Objective To construct and validate a prognosis risk model for acute myeloid leukemia (AML) patients. **Methods** Based on The Cancer Genome Atlas (TCGA) database and Genotype-Tissue Expression (GTEx) database to obtain gene sequencing data of AML samples and normal bone marrow samples, and then perform gene differential analysis. The genes associated with AML survival were selected and a random forest model was used to construct a risk scoring model based on gene features. The biological functions of the predictive model were further explored through Gene Ontology (GO) enrichment analysis. **Results**

DOI: 10.12173/j.issn.1004-4337.202310183
通信作者：张斯娜，Email：zasana@126.com

According to the results of Kaplan-Meier analysis and receiver operating characteristic (ROC) curve analysis, the survival rate predictions demonstrated good sensitivity and specificity in both the training and testing datasets. Multivariate Cox regression analysis and stratified analysis showed that the gene prognostic risk model could independently predict the survival outcomes of patients with AML. GO enrichment analysis demonstrated that differentially expressed genes were enriched in biological processes related to known AML. **Conclusion** By using bioinformatics methods, a mRNA marker containing 5 genes (*ACOT7*, *SFXN3*, *SLC29A2*, *HINT2*, *ZMAT1*) has been identified, which may serve as a potential prognostic marker for AML and provide a basis for personalized risk assessment and the formulation of treatment plans for AML patients.

【Keywords】 Acute myeloid leukemia; Bioinformatics; The Cancer Genome Atlas; Genotype-Tissue Expression; Prognostic model; Random forest

急性髓系白血病（acute myeloid leukemia, AML）是一种未成熟髓系造血干细胞异常克隆性恶性肿瘤，具有高度异质性，其主要特征是克隆性造血干细胞或祖细胞（hematopoietic stem and progenitor cells, HSPCs）分化和凋亡障碍。AML 可导致克隆性 HSPCs 在骨髓中恶性增殖和聚集，从而影响骨髓正常造血功能[1-2]。AML 的高度异质性与多种因素有关，如基因组突变、表观遗传学异常和基因表达异常等[3]。全球范围内，AML 发病人数从 1990 年的 63 840 例增加到 2017 年的 119 570 例，发病率在过去的 28 年间增加了 87.3%[4]。化疗、生物制剂和干细胞移植是 AML 的主要治疗方式[5-7]。尽管 AML 的治疗取得了重大进展，但化疗药物毒性可能导致急性和危及生命的并发症。AML 通常预后不良，患者中位生存时间仅为 8.5 个月，2 年和 5 年总生存率（overall survival, OS）均低于 35%，仅 35%~40% 的年轻患者（＜60 岁）和 5%~15% 的老年患者（≥60 岁）在化疗后可存活 5 年以上[8-9]。随着分子生物学的迅速发展，基因组学、蛋白质组学和生物信息学分析方法已被用于开发新的个性化治疗策略，研究相关生物分子的功能，收集临床数据基因组匹配的新趋势信息是改善 AML 患者预后的有效方法[10-11]。尽管许多研究分析了 AML 的基因组变异，但基因组变异与 AML 分子机制间的关联仍不明确。因此，需要进一步探索更有效的 AML 早期诊断和预后方法。AML 的基因表达谱已被证明在诊断不同的细胞遗传学亚型、发现新的 AML 亚类和预测预后方面具有重要价值[12-13]。基因组分析是筛选潜在的疾病诊断和预后生物标志物的一种有效方法，本研究基于公共数据库的基因表达谱数据集构建并验证 AML 患者预后风险模型，识别 AML 患者的预后生物标志物。

1 资料与方法

1.1 数据来源与差异表达基因筛选

基于癌症基因组图谱（The Cancer Genome Atla, TCGA）数据库（https://gdc.xenahubs.net/download/TCGA-LAML/Xena_Matrices/TCGA-LAML.htseq_fpkm.tsv.gz），获取 AML 基因数据（151 例）作为病例组。基于基因型－组织表达（Genotype-Tissue Expression, GTEx）数据库（https://toil.xenahubs.net/download/TCGatargettex_rsem_gene_fpkm.gz），获取正常骨髓样本（70 例）作为对照组。利用 Wilcoxon 秩和检验分析 AML 样本基因数据，筛选 $|\log FC| > 2$ 且错误发现率（false discovery rate, FDR）＜ 0.05 的差异表达基因（differentially expressed genes, DEGs）。采用单因素 Cox 回归分析，筛选 DEGs 中与 AML 生存相关基因，以 $P < 0.01$ 为差异有统计学意义。

1.2 风险评分模型构建与特征基因验证

151 例 AML 样本中，剔除临床特征信息缺失和随访时间为 0 的样本后，最终纳入 132 例，将 132 例 AML 样本按照 1∶1 的比例随机分成两组，分别作为训练集和测试集。在训练集中，通过单变量 Cox 比例风险回归分析 mRNA 的表达与患者 OS 之间的关系。在构建风险预测模型前，采用随机生存森林算法对 AML 预后相关的差异基因进行筛选。在随机生存森林监督分类算法中，采用迭代过程缩小基因集。随机森林的树数（ntree）为 1 000，将每一步输入节点数的平方根设为单个分类树的每个节点随机选取的预后相关的差

异 mRNA 的大小，对袋外数据样本进行泛化误差估计。为了选择极少最重要的基因变量，可依据变量的重要性（per-mutation variable important, VIMP）对变量进行筛选，VIMP 值越大表明其预测能力越强。风险预测评分模型是基于 AML 预后相关的差异基因而构建，通过其估计回归系数加权如下：

$$RiskScore(RS) = \sum_{i=1}^{N} (\mathrm{Exp_i} * \mathrm{Coe}_i)$$

其中，N 为预后相关的差异基因数量，$\mathrm{Exp_i}$ 为预后相关的差异基因的表达值，$\mathrm{Coe_i}$ 为单变量 Cox 回归分析中预后相关差异基因的估计回归系数。模型构建后，模型中涉及的特征基因通过以下方式进一步验证：①对模型高低风险组的生存分析，最佳的截断表达值由“survminer”R 包的“surv_cutpoint”函数确定。使用 Kaplan-Meier 和 log-rank 方法在训练集中进行生存分析，再使用 AML 数据和临床信息在测试集中验证该模型。②在训练集和测试集中使用“survivalROC”R 软件包进行时间依赖的受试者工作特征（receiver operating characteristic, ROC）曲线分析，并计算 1 年、2 年和 3 年 AML 患者生存率的 ROC 曲线下面积（area under the curve, AUC）。通过决策曲线分析量化不同时间概率下的净收益，从而确定风险评分的临床实用性，决策曲线分析通过“rmda”R 软件包进行。基于所有样本的基因表达，使用“Rtsne”R 软件包进行 t-SNE 分析，探索不同组的分布。

1.3 分层分析

在全集中，对 AML 患者总生存率显著相关的临床特征数据进行分层分析，其中年龄以 60 岁为截断值，≥60 岁为老年组，<60 岁为年轻组。

1.4 功能富集分析

为了阐明与风险评分相关的生物学功能和途径，高风险组和低风险组之间的 DEGs 被用于进行基因本体（Gene Ontology, GO）富集分析。利用“clusterProfiler”R 包进行分析，以校正后 P 值小于 0.05 为差异有统计学意义。

2 结果

2.1 AML患者临床特征

训练集、测试集和全集中 AML 患者的临床特征见表 1。

表1 AML患者的临床特征

Table 1. Clinical characteristics of patients with AML

项目	训练集（n=66）	测试集（n=66）	全集（n=132）
性别（n，%）			
女性	28（42.4）	33（50.0）	61（46.2）
男性	38（57.6）	33（50.0）	71（53.8）
年龄组（n，%）			
≥60岁	32（48.5）	23（34.8）	55（41.7）
<60岁	34（51.5）	43（65.2）	77（58.3）
生存状态（n，%）			
存活	27（40.9）	25（37.9）	52（39.4）
死亡	39（59.1）	41（62.1）	80（60.6）
生存时间 [M（P_{25}，P_{75}），天]	547（205，1 216.8）	320（181，701）	366（184，822）
FLT3突变（n，%）			
阳性	16（24.2）	22（33.3）	38（28.8）
阴性	47（71.2）	43（65.2）	90（68.2）
NA	3（4.6）	1（1.5）	4（3.0）
IDH1 R132（n，%）			
阳性	7（10.6）	6（9.1）	13（9.8）
阴性	59（89.4）	59（89.4）	118（89.4）
NA	0（0.0）	1（1.5）	1（0.8）

续表1

项目	训练集（n=66）	测试集（n=66）	全集（n=132）
IDH1 R140（n，%）			
阳性	6（9.1）	5（7.6）	11（8.3）
阴性	59（89.4）	60（90.9）	119（90.2）
NA	1（1.5）	1（1.5）	2（1.5）
IDH1 R172（n，%）			
阳性	2（3.0）	2（3.0）	4（3.0）
阴性	63（95.5）	63（95.5）	126（95.5）
NA	1（1.5）	1（1.5）	2（1.5）
NPMc（n，%）			
阳性	16（24.2）	15（22.7）	31（23.5）
阴性	49（74.3）	51（77.3）	100（75.7）
NA	1（1.5）	0（0.0）	1（0.8）
RAS激活（n，%）			
阳性	2（3.0）	6（9.1）	8（6.0）
阴性	63（95.5）	60（90.9）	123（93.2）
NA	1（1.5）	0（0.0）	1（0.8）
细胞遗传学风险类别（n，%）			
非高风险	55（83.3）	48（72.7）	103（78.0）
高风险	10（15.2）	17（25.8）	27（20.5）
NA	1（1.5）	1（1.5）	2（1.5）

注：NA为缺失值。

2.2 差异表达基因鉴别

基于 AML 样本和正常骨髓样本数据，共获取 19 022 个基因的 mRNA 测序数据，通过基因差异显著性分析筛选出 2 270 个 DEGs（图 1）。通过单变量 Cox 回归分析，得到 335 个 AML 预后相关特征基因（$P < 0.01$）。

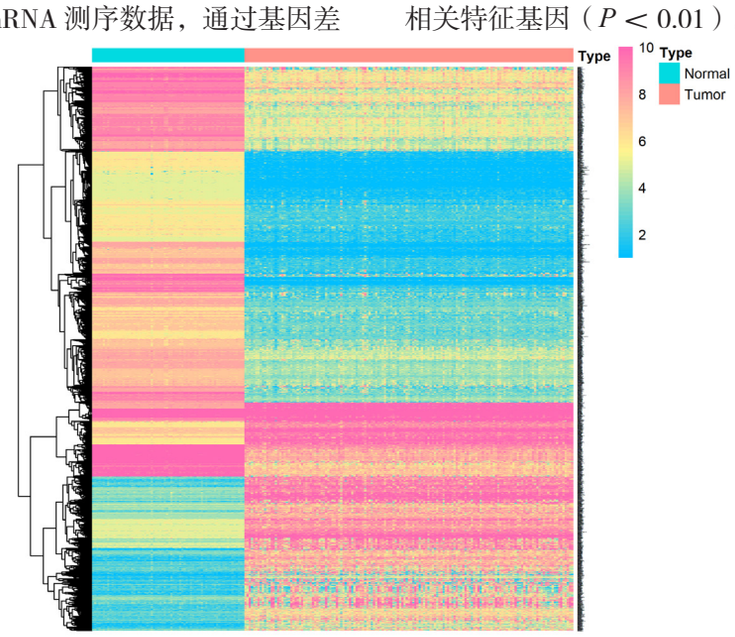


图1 差异基因表达热图

Figure 1. Heatmap of differential genes expression

注：红色表示上调基因，蓝色表示下调基因。

2.3 特征基因筛选

为了进一步缩小预后基因的数量，通过随机森林算法对 335 个与 AML 患者 OS 显著相关的 DEGs 进行分析，在每一步中，根据通过置换测试使用袋外样品对每个基因的重要分数进行估计，丢弃不重要的基因，并根据每一步中的排列重要评分，得到与 AML 患者 OS 显著相关的 5 个基因标志物（*ACOT7*、*SFXN3*、*SLC29A2*、*HINT2* 和 *ZMAT1*），见图 2–A。

风险评分模型构建如下：风险评分 =（–0.959 24 × *ZMAT1* 表达值）+（0.525 21 × *SFXN3* 表达值）+（1.060 22 × *HINT2* 表达值）+（0.669 68 × *SLC29A2* 表达值）+（0.079 19 × *ACOT7* 表达值）。基于风险评分模型计算训练集每个 AML 患者的风险评分，根据评分将患者分为高风险组（*n*=24）和低风险组（*n*=42），使用 cutoff=19.686 作为风险临界值。图 2–B、图 2–C、图 2–D 展示了 66 例 AML 患者的预后评分、生存状态和肿瘤 mRNA 表达的分布，根据 5 个特征基因的预后评分值进行排序。在 5 个基因中，4 个基因（*ACOT7*、*SFXN3*、*SLC29A2* 和 *HINT2*）与预后高风险相关 [风险比（hazard ratio, HR）> 1]，1 个基因（*ZMAT1*）为保护性因素（HR < 1）。预后评分高的基因倾向于表达高风险的 mRNA，而预后评分低的基因倾向于表达保护性 mRNA，见图 2–C 和图 2–D。Kaplan–Meier 生存分析和对数秩检验表明，高风险组和低风险组的总体生存率有显著差异，高风险组患者总生存期明显短于低风险组（中位生存期 3.79 年 vs. 0.57 年，*P* < 0.001），见图 2–E。5 个基因生存预测的时间依赖 ROC 曲线分析显示，1 年、2 年、3 年的 AUC 值分别为 0.778、0.743、0.677，曲线接近左上角，具有良好的敏感性和特异性，表明该预测模型具有较高的准确率，见图 2–F。

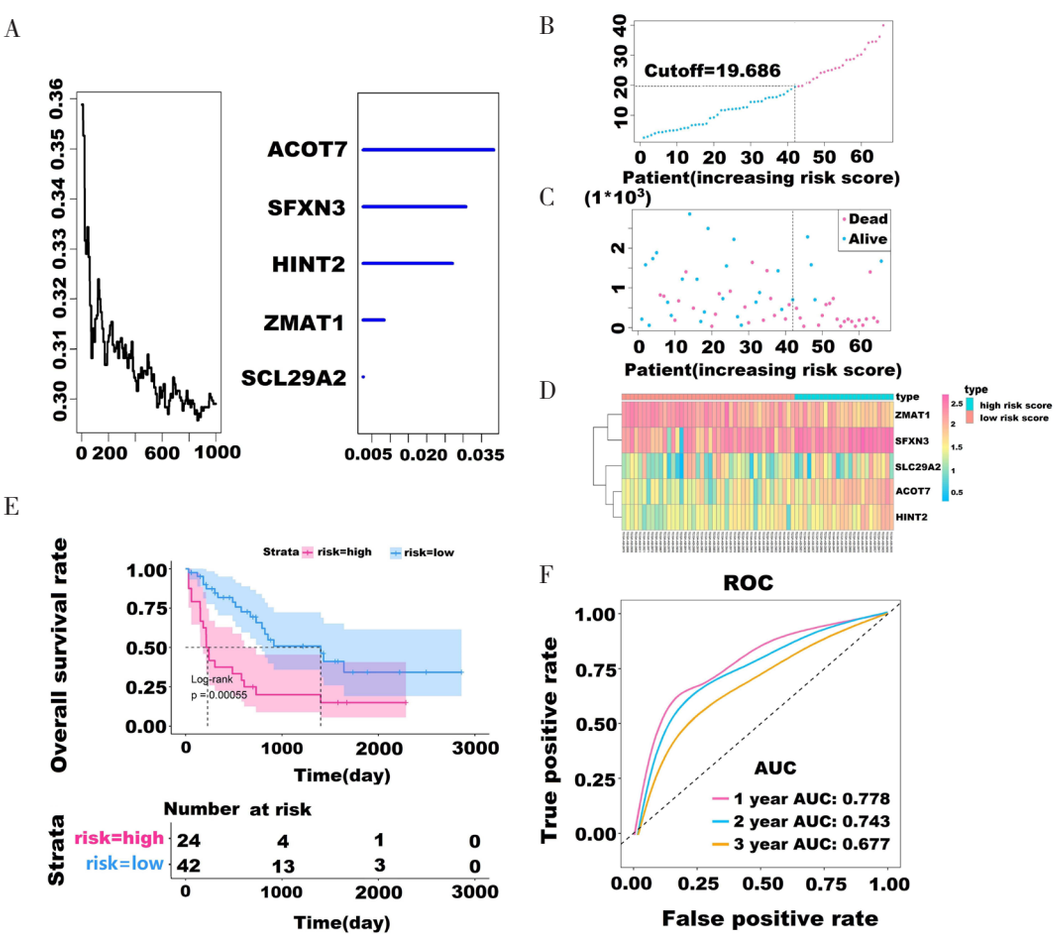


图2 特征基因风险评分分析

Figure 2. Risk scoring analysis of characteristic genes

注：A. 通过随机生存森林分析计算出的所选预后差异基因的重要分数；B. 5个特征基因风险评分分布；C. AML患者的生存状况和时间；D. AML患者中五个基因表达谱的热图；E. 根据风险临界点的总体生存结果的Kaplan–Meier生存曲线，对数秩检验的*P*值小于0.001；F. 5个特征基因的时间依赖性ROC曲线分析AML患者1年、2年和3年生存预测。

2.4 特征基因验证

为了测试 5 个特征基因在预测 AML 患者总生存期中的预后价值，在测试集中测试 5 个特征基因。通过使用相同的风险评分模型，测试集的 66 例患者被分为高风险组（n=33）和低风险组（n=33）。结果显示，高风险组患者的总生存期明显短于低风险组患者（中位生存期 0.71 年 vs. 2.5 年，$P < 0.001$），见图 3–A。时间依赖的 ROC 曲线分析对 5 个特征基因的生存预测的 AUC 值在 1 年时达到 0.775，2 年时达到 0.746，3 年时达到 0.696，见图 3–B。进一步将 5 个特征基因应用于整个数据集的患者，以验证其预测价值，来自训练集的相同风险评分模型和中位风险截止标准将全集的 132 例 AML 患者分为高风险组（n=66）和低风险组（n=66）。高风险组患者的总生存期明显短于低风险组患者（中位生存期 0.68 年 vs. 2.58 年，$P < 0.001$），见图 3–C。在 132 例 AML 患者全集中验证 5 个特征基因的 ROC 曲线在 1 年、2 年、3 年的 AUC 值分别为 0.836、0.777、0.694，见图 3–D。

2.5 特征基因的预后价值与临床变量的独立性

如表 2 所示，多变量 Cox 回归分析结果表明，经年龄、性别和各分子分析异常测试结果调整后，5 个特征基因仍与 AML 患者总生存率存在显著相关性。当控制其他临床变量时，在训练集中，高风险组相对于低风险组总生存率的风险比为 3.47 [95%CI（1.80，6.70），$P < 0.001$]，在测试集和

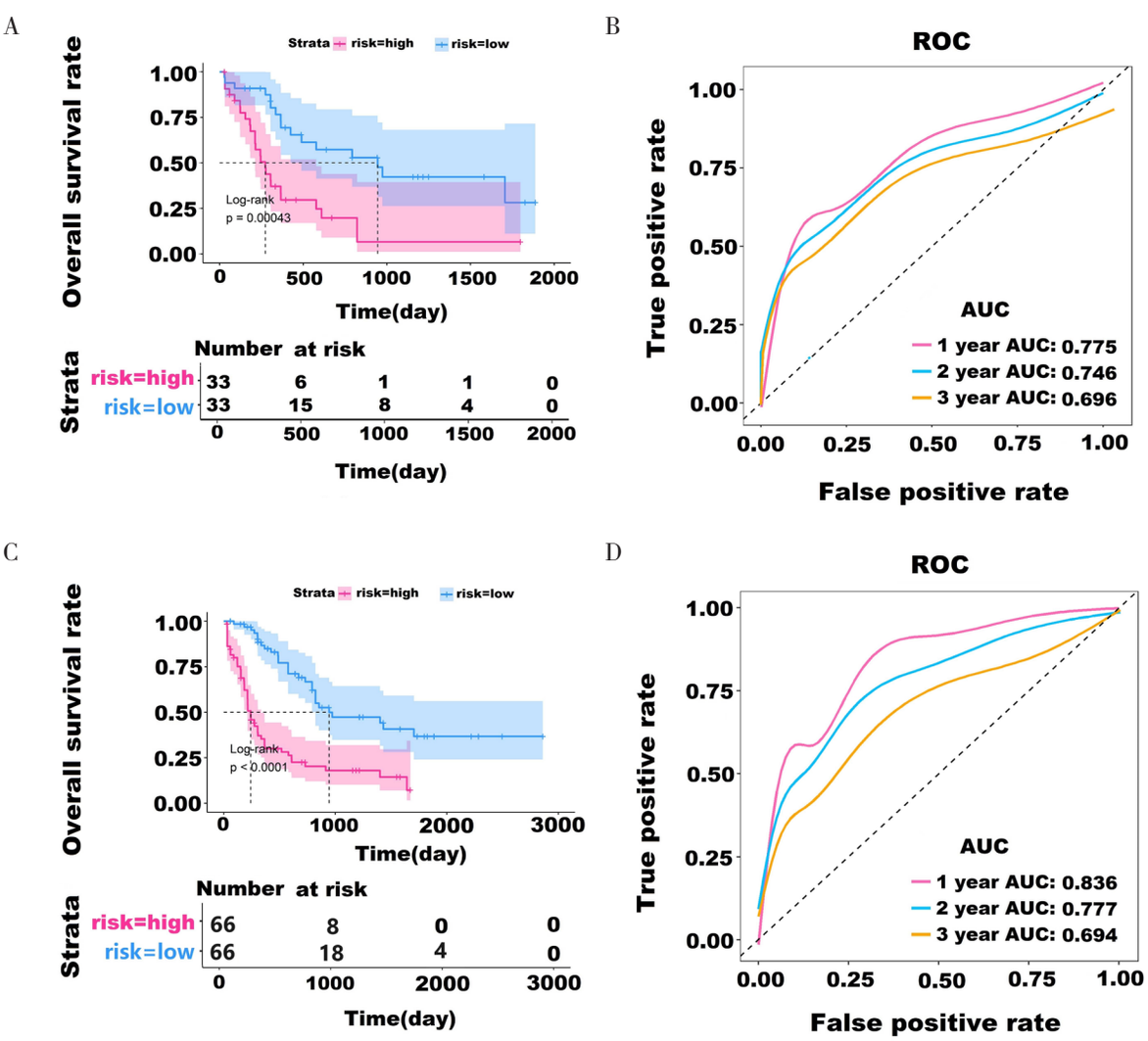


图3 5个特征基因的验证分析

Figure 3.Validation analysis of 5 characteristic genes

注：A. Kaplan–Meier估计了测试集中高风险组和低风险组之间的总体生存率；B. 在测试集中1年、2 年和3年的5个特征基因的ROC曲线；C. Kaplan–Meier估计了全集中高风险组和低风险组之间的总体生存率；D. 在全集中1年、2年和3年的5个特征基因的ROC曲线。

全集中，高风险组相对于低风险组总生存率的风险比分别为 3.25 [95%CI（1.71, 6.19），$P < 0.001$] 和 3.39 [95%CI（2.12，5.43），$P < 0.001$]。年龄与三组数据集 AML 患者的总生存率显著相关，因此根据年龄进行分层分析，所有患者被分为年轻组（n=77）和老年组（n=55），再根据风险评分模型，将年轻组和老年组患者分为高风险组和低风险组。如图 4–A 和图 4–B 所示，生存分析表明，在年轻组和老年组中，高风险组的总体生存时间明显短于低风险组（$P < 0.001$）。

表2 单变量和多变量Cox回归分析结果

Table 2. Results of univariate and multivariate Cox regression analysis

变量	组别	单变量Cox回归分析			多变量Cox回归分析		
		HR	95%CI	*P*值	HR	95%CI	*P*值
训练集							
RAS激活	阳性 vs. 阴性	3.13	（0.73，13.38）	0.124			
细胞遗传学风险类别	高风险 vs. 非高风险	1.23	（0.48，3.20）	0.665			
年龄	≥60岁 vs. <60岁	2.52	（1.32，4.81）	0.005	3.04	（1.55，5.96）	0.001
FLT3突变	阳性 vs. 阴性	0.73	（0.34，1.55）	0.408			
性别	女性 vs. 男性	1.14	（0.60，2.16）	0.689			
IDH1 R132	阳性 vs. 阴性	1.32	（0.47，3.71）	0.604			
IDH1 R140	阳性 vs. 阴性	1.80	（0.75，4.30）	0.189			
IDH1 R172	阳性 vs. 阴性	0.78	（0.11，5.68）	0.802			
NPMc	阳性 vs. 阴性	1.22	（0.62，2.41）	0.572			
风险级别	高风险 vs. 低风险	2.92	（1.55，5.49）	0.001	3.47	（1.80，6.70）	<0.001
测试集							
RAS激活	阳性 vs. 阴性	0.31	（0.07，1.32）	0.113			
细胞遗传学风险类别	高风险 vs. 非高风险	1.59	（0.83，3.05）	0.165			
年龄	≥60岁 vs. <60岁	3.01	（1.61，5.63）	0.001	3.58	（1.88，6.81）	<0.001
FLT3突变	阳性 vs. 阴性	2.44	（1.29，4.61）	0.006	2.37	（1.24，4.54）	0.009
性别	女性 vs. 男性	0.90	（0.49，1.66）	0.746			
IDH1 R132	阳性 vs. 阴性	0.47	（0.14，1.52）	0.204			
IDH1 R140	阳性 vs. 阴性	0.73	（0.22，2.38）	0.596			
IDH1 R172	阳性 vs. 阴性	0.56	（0.08，4.08）	0.567			
NPMc	阳性 vs. 阴性	1.24	（0.61，2.53）	0.551			
风险级别	高风险 vs. 低风险	3.16	（1.67，5.98）	<0.001	3.25	（1.71，6.19）	<0.001
全集							
RAS激活	阳性 vs. 阴性	0.69	（0.25，1.91）	0.479			
细胞遗传学风险类别	高风险 vs. 非高风险	1.63	（0.95，2.77）	0.074			
年龄	≥60岁 vs. <60岁	2.71	（1.74，4.24）	<0.001	2.46	（1.57，3.85）	<0.001
FLT3 突变	阳性 vs. 阴性	1.34	（0.83，2.15）	0.229			
性别	女性 vs. 男性	1.04	（0.67，1.61）	0.864			
IDH1 R132	阳性 vs. 阴性	0.77	（0.34，1.78）	0.546			
IDH1 R140	阳性 vs. 阴性	1.23	（0.61，2.48）	0.556			
IDH1 R172	阳性 vs. 阴性	0.66	（0.09，4.73）	0.676			
NPMc	阳性 vs. 阴性	1.21	（0.74，1.98）	0.443			
风险级别	高风险 vs. 低风险	3.64	（2.28，5.81）	<0.001	3.39	（2.12，5.43）	<0.001

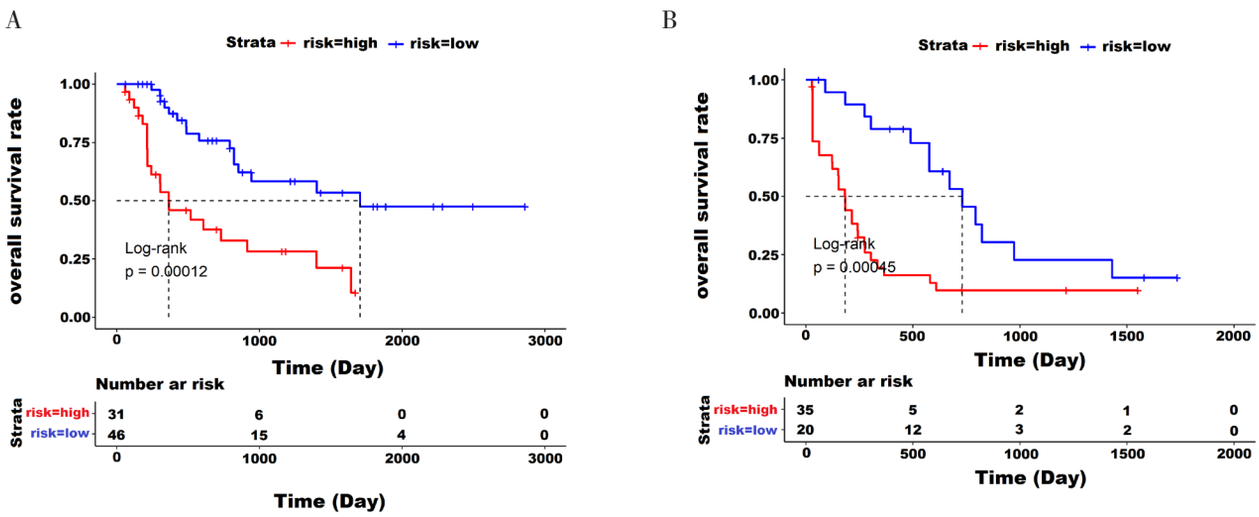


图4 两组总体生存率的估计

Figure 4. Estimation of overall survival rates for two groups

注：Kaplan–Meier估计了年轻组（A）和老年组（B）中高风险组和低风险组的总生存率。

2.6 特征基因的性能评估

t–SNE 分析表明，不同风险组的患者分布在两个方向（图 5–A）。5 个特征基因组合风险评分的决策曲线分析表明，风险评分预测 AML 患者 1 年、2 年和 3 年的生存情况表现出比极端曲线更高的净获益，而 1 年的阈值范围更大，认为 1 年风险评分表现更优，见图 5–B。

2.7 功能富集分析

如图 6–A 所示，在整个数据集中，DEGs 富集几种与代谢过程相关的途径，如固醇代谢过程、胆固醇代谢过程、类固醇代谢过程、仲醇代谢过程、脂质代谢过程和酒精代谢过程（校正后 $P < 0.05$）；在许多免疫相关的生物过程中也明显富集，如白细胞迁移、T 细胞分化和巨噬细胞迁移（校正后 $P < 0.05$）；DEGs 在凋亡相关的生物过程中也明显富集，如凋亡信号通路的负调控、线粒体凋亡途径和线粒体释放细胞色素 C 途径（校正后 $P < 0.05$）。图 6–B 展示了各途径所富集到的基因。

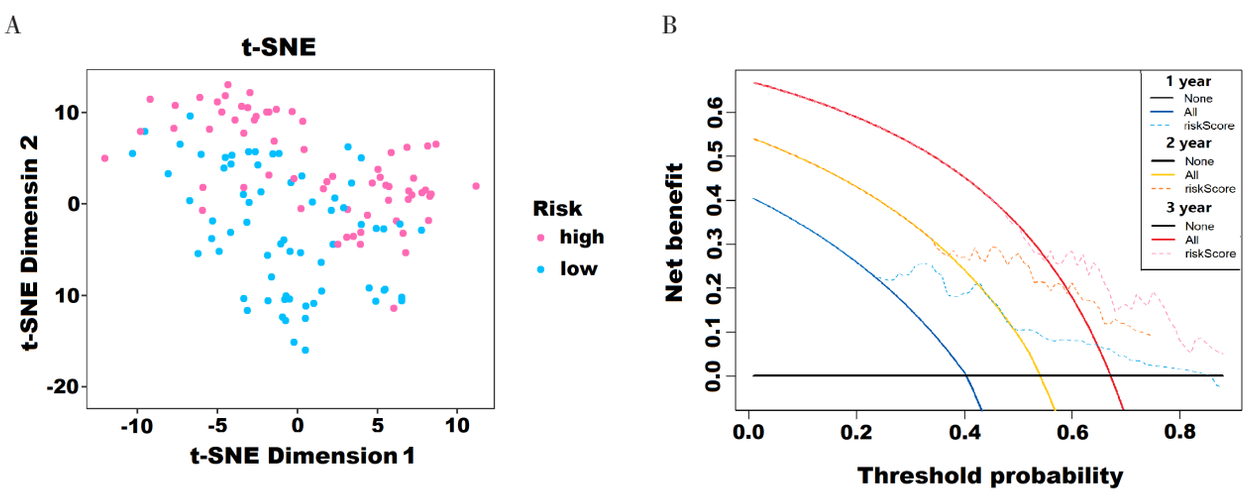


图5 特征基因的性能评估

Figure 5. Performance evaluation of characteristic genes

注：A. 整个数据集的t–SNE分析；B. 风险模型的决策曲线分析其中蓝、黄、红色分别表示AML患者1年、2年和3年的总生存期。

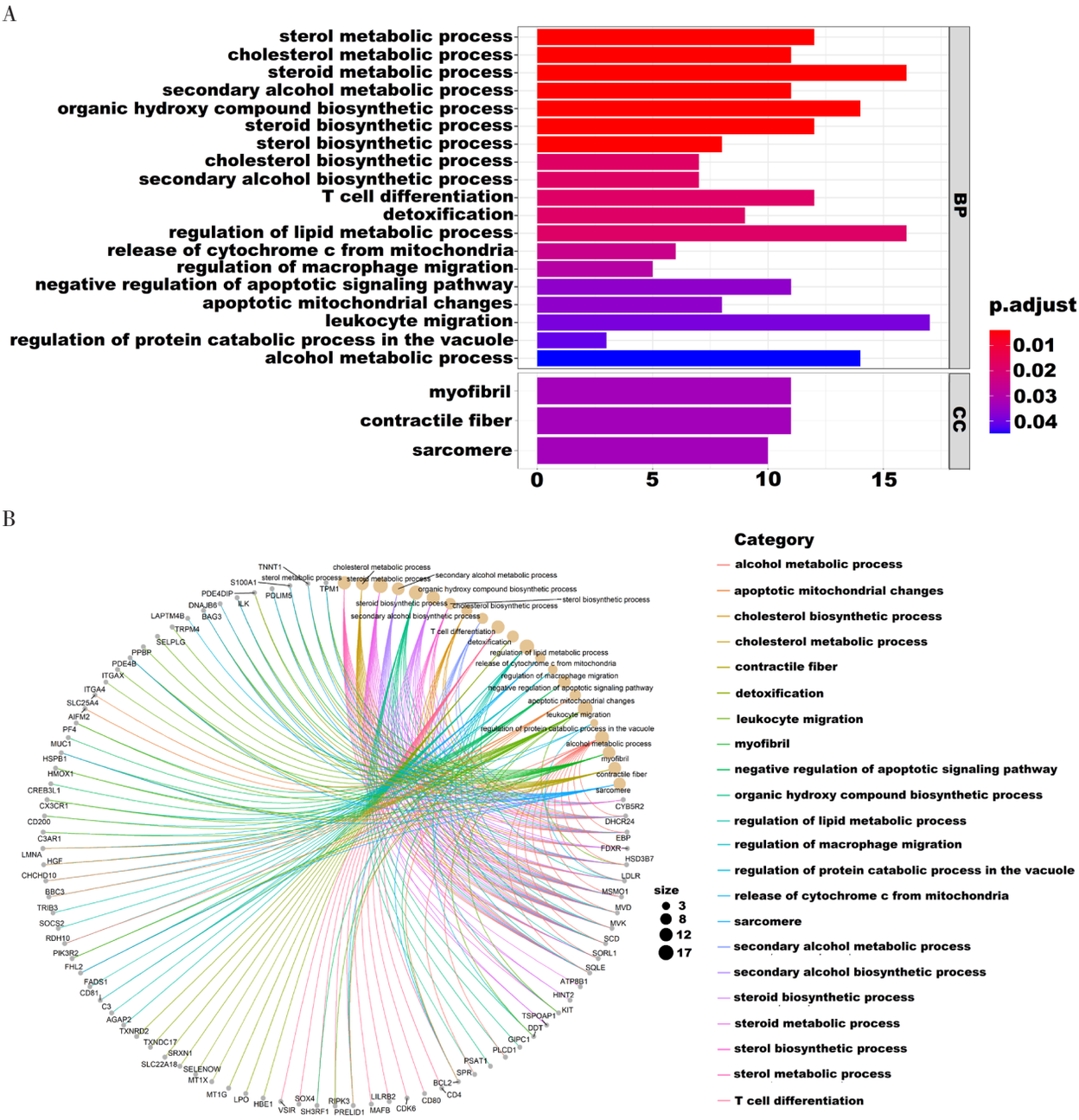


图6 GO富集分析

Figure 6. GO enrichment analysis

注：A. 整个数据集中重要的GO富集途径；B. GO富集中各途径所富集到的基因。

3 讨论

在最近的研究中，许多 AML 相关的癌基因和肿瘤抑制基因已被确定。例如，Zhou 等发现 *HSPG2* 过表达与 AML 患者的不良预后相关，AML 患者中 *HSPG2* 的表达水平显著上调，完全缓解后则下调，而在复发时再次升高[14]。Li 等研究表明，*SLC38A1* 是 AML 的重要预后和预测指标，*SLC38A1* 高表达患者的 *NPM1* 基因突变率较低，不良风险核型发生率较高，*SLC38A1* 表达水平高的患者总生存期显著缩短[15]。Wróbel 等发现 *IL-17F* 多态性与 AML 的诱导密切相关[16]。尽管发现了一些与 AML 预后和诊断相关的基因，但其在 AML 中的作用仍有待探究。AML 的预后部分由遗传因素驱动，多种基因的结合有助于提高预后预测的准确性。本研究中，AML 患者的临床

数据和基因表达数据来自 TCGA 数据库，以鉴定与预后相关的基因。通过单变量 Cox 分析和随机生存森林分析选择生存相关特征基因，最终通过多变量 Cox 回归分析构建了基于 5 个预后相关特征基因的风险评分模型。

在本研究筛选出的 5 个特征基因中，仅 *ZMAT1* 被认为是低风险且与 AML 预后相关基因。目前尚未发现 *ZMAT1* 在 AML 中的研究报告，但在胰腺导管腺癌中有过报道，研究发现 *ZMAT1* 的下调与胰腺导管腺癌不良的临床病理特征和低生存率相关，*ZMAT1* 通过诱导 *SIRT3/p53* 信号通路在胰腺导管腺癌中起到抑癌因子的作用，其中 *ZMAT1* 促进 *SIRT3* 基因的转录，随后上调 *p53* 的表达，并参与调节胰腺导管腺癌的细胞周期进程和凋亡 [17]。Zhang 等研究发现高水平的 *ACOT7* 对 AML 患者的预后有显著影响，*ACOT7* 高表达组的无事件生存期（event-free survival, EFS）和 OS 显著降低 [18]。*ACOT7* 的表达水平是影响 AML 患者预后的独立因素，*ACOT7* 的高表达水平在男性患者中更常见，且表达水平与 FAB 亚型以及细胞遗传学之间可能存在潜在相关性 [18]。*SLC29A2* 也是 *ENT2*，有研究表明肝细胞癌（hepatocellular carcinoma, HCC）组织中 *SLC29A2* 的上调与晚期、血管浸润与患者生存率不良显著相关 [19]。有研究推测 *SLC29A2* 的异常上调可能改变细胞核苷酸合成和核苷酸池平衡，从而导致癌症基因组不稳定，进而为肿瘤提供生长优势 [20]。由于 *SLC29A2* 是一种已知的腺苷转运体，而从缺氧组织中释放的腺苷刺激血管内皮生长因子的释放，再结合内皮细胞上的受体，刺激内皮细胞增殖、迁移和肿瘤血管生成 [20]。*SFXN3* 和 *HINT2* 已被证明分别在口腔鳞状细胞癌和肝细胞癌中作为癌基因 [21-22]。Murase 等认为口腔鳞状细胞癌患者血清抗 *SFXN3* 自身抗体水平显著升高，血清抗 SFXN3-autoAb 水平与原发肿瘤大小轻度相关，但在早期癌症中水平稍高，*SFXN3* 蛋白在癌巢之间的基质成纤维细胞中表达，也在鳞状上皮的基底层中表达，治疗后血清抗 *SFXN3* 自身抗体水平的变化与临床肿瘤负荷相关 [21]。Zhou 等研究表明 *HINT2* 在 HCC 组织中的表达明显低于邻近的非肿瘤组织，且 *HINT2* 低表达 HCC 患者复发的可能性更高，可作为 HCC 复发的一个预后指标 [22]。鉴于在其他类型癌症中的重要作用，这些特征基因可用于预测 AML 患者的生存率。基于 5 个特征基因的风险评分模型，低风险和高风险患者的存活率有显著差异，高风险患者比低风险患者存活时间明显更短。

基于不同风险组间的 DEGs 进行了 GO 富集分析，发现许多免疫相关的生物过程和途径得到了富集，说明 AML 可能与肿瘤免疫有密切联系的推测具有一定合理性。基于 5 个特征基因的风险评分模型初步用于 AML 患者的生存预测，并进一步证明了这些特征基因在 AML 生存中的重要作用，可作为更好地评估 AML 患者生存和预后的新方法。此外，风险评分模型在测试集及整个数据集中得到进一步验证，表明该模型在 AML 患者生存预测中具有良好的可重复性。由于流行病学的限制，本研究未能分析风险评分与 AML 患者生存率之间的具体关系，未来有待进一步深入探究。

综上所述，本研究确定了 5 个与 AML 预后密切相关的特征基因，并构建和验证了基于 5 个特征基因的风险评分模型，该模型可用于预测 AML 患者的生存期，为 AML 患者的个性化风险评估、治疗和预后提供参考。

参考文献

1 Short NJ, Rytting ME, Cortes JE. Acute myeloid leukaemia[J]. Lancet, 2018, 392(10147): 593-606. DOI: 10.1016/S0140-6736(18)31041-9.

2 Coombs CC, Tallman MS, Levine RL. Molecular therapy for acute myeloid leukaemia[J]. Nat Rev Clin Oncol, 2016, 13(5): 305-318. DOI: 10.1038/nrclinonc.2015.210.

3 Adriaens C, Standaert L, Barra J, et al. p53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity[J]. Nat Med, 2016, 22(8): 861-868. DOI: 10.1038/nm.4135.

4 Yi M, Li A, Zhou L, et al. The global burden and attributable risk factor analysis of acute myeloid leukemia in 195 countries and territories from 1990 to 2017: estimates based on the global burden of disease study 2017[J]. J Hematol Oncol, 2020, 13(1): 72. DOI: 10.1186/s13045-020-00908-z.

5 Liu L, Ma J, Qin L, et al. Interleukin-24 enhancing antitumor activity of chimeric oncolytic adenovirus for

treating acute promyelocytic leukemia cell[J]. Medicine (Baltimore), 2019, 98(22): e15875. DOI: 10.1097/MD.0000000000015875.

6 Liu Y, Cheng Z, Pang Y, et al. Role of microRNAs, circRNAs and long noncoding RNAs in acute myeloid leukemia[J]. J Hematol Oncol, 2019, 12(1): 51. DOI: 10.1186/s13045-019-0734-5.

7 Zhou JD, Li XX, Zhang TJ, et al. MicroRNA-335/ID4 dysregulation predicts clinical outcome and facilitates leukemogenesis by activating PI3K/Akt signaling pathway in acute myeloid leukemia[J]. Aging (Albany NY), 2019, 11(10): 3376-3391. DOI: 10.18632/aging.101991.

8 Shallis RM, Wang R, Davidoff A, et al. Epidemiology of acute myeloid leukemia: recent progress and enduring challenges[J]. Blood Rev, 2019, 36: 70-87. DOI: 10.1016/j.blre.2019.04.005.

9 Li Z, Weng H, Su R, et al. FTO plays an oncogenic role in acute myeloid leukemia as a N^6-Methyladenosine RNA demethylase[J]. Cancer Cell, 2017, 31(1): 127-141. DOI: 10.1016/j.ccell.2016.11.017.

10 Bret C, Viziteu E, Kassambara A, et al. Identifying high-risk adult AML patients: epigenetic and genetic risk factors and their implications for therapy[J]. Expert Rev Hematol, 2016, 9(4): 351-360. DOI: 10.1586/17474086.2016.1141673.

11 Cai SF, Levine RL. Genetic and epigenetic determinants of AML pathogenesis[J]. Semin Hematol, 2019, 56(2): 84-89. DOI: 10.1053/j.seminhematol.2018.08.001.

12 Bullinger L, Valk PJ. Gene expression profiling in acute myeloid leukemia[J]. J Clin Oncol, 2005, 23(26): 6296-6305. DOI: 10.1200/JCO.2005.05.020.

13 Li W, Gao LN, Song PP, et al. Development and validation of a RNA binding protein-associated prognostic model for lung adenocarcinoma[J]. Aging (Albany NY), 2020, 12(4): 3558-3573. DOI: 10.18632/aging.102828.

14 Zhou X, Liang S, Zhan Q, et al. *HSPG2* overexpression independently predicts poor survival in patients with acute myeloid leukemia[J]. Cell Death Dis, 2020, 11(6): 492. DOI: 10.1038/s41419-020-2694-7.

15 Li Y, Shao H, Da Z, et al. High expression of *SLC38A1* predicts poor prognosis in patients with de novo acute myeloid leukemia[J]. J Cell Physiol, 2019, 234(11): 20322-20328. DOI: 10.1002/jcp.28632.

16 Wróbel T, Gębura K, Wysoczańska B, et al. *IL-17F* gene polymorphism is associated with susceptibility to acute myeloid leukemia[J]. J Cancer Res Clin Oncol, 2014, 140(9): 1551-1555. DOI: 10.1007/s00432-014-1674-7.

17 Ma Z, Li Z, Wang S, et al. *ZMAT1* acts as a tumor suppressor in pancreatic ductal adenocarcinoma by inducing SIRT3/p53 signaling pathway[J]. J Exp Clin Cancer Res, 2022, 41(1): 130. DOI: 10.1186/s13046-022-02310-8.

18 Zhang X, Liu B, Zhang J, et al. Expression level of *ACOT7* influences the prognosis in acute myeloid leukemia patients[J]. Cancer Biomark, 2019, 26(4): 441-449. DOI: 10.3233/CBM-182287.

19 Chen CF, Hsu EC, Lin KT, et al. Overlapping high-resolution copy number alterations in cancer genomes identified putative cancer genes in hepatocellular carcinoma[J]. Hepatology, 2010, 52(5): 1690-1701. DOI: 10.1002/hep.23847.

20 Mathews CK. DNA precursor metabolism and genomic stability[J]. FASEB J, 2006, 20(9): 1300-1314. DOI: 10.1096/fj.06-5730rev.

21 Murase R, Abe Y, Takeuchi T, et al. Serum autoantibody to sideroflexin 3 as a novel tumor marker for oral squamous cell carcinoma[J]. Proteomics Clin Appl, 2008, 2(4): 517-527. DOI: 10.1002/prca.200780123.

22 Zhou DK, Qian XH, Cheng J, et al. Clinical significance of down-regulated *HINT2* in hepatocellular carcinoma[J]. Medicine (Baltimore), 2019, 98(48): e17815. DOI: 10.1097/MD.0000000000017815.

收稿日期：2023 年 10 月 30 日 修回日期：2023 年 11 月 29 日
本文编辑：张 苗 黄 笛

引用本文：张斯娜, 马书杰. 基于生物信息学的急性髓系白血病患者预后风险模型的构建与验证[J]. 数理医药学杂志, 2023, 36(12): 888-898. DOI: 10.12173/j.issn.1004-4337.202310183
Zhang SN, Ma SJ. Construction and validation of a prognostic risk model for patients with acute myeloid leukemia based on bioinformatics[J]. Journal of Mathematical Medicine, 2023, 36(12): 888-898. DOI: 10.12173/j.issn.1004-4337.202310183