基于 TCGA 和 GEO 数据库建立了肝内胆管癌的预后风险模型及验证分析

毛 俊，沈秀芬，马 润，何 薇，瞿巧莉，胡 莹（昆明医科大学第二附属医院检验科，昆明 650101）

摘 要: 目的 基于TCGA（the cancer genome atlas）和GEO（gene expression omnibus）数据库构建肝内胆管癌（intrahepatic cholangiocarcinoma, ICCA）预后风险模型，筛选ICCA预后相关基因。**方法** TCGA数据库31例ICCA组织及9例癌旁组织数据作为训练集，GEO数据库30例ICCA组织及27例癌旁组织数据作为验证集，R软件“DESeq2”包过滤表达有差异的基因，过滤条件：差异倍数绝对值＞2，校正*P*值＜0.05。单因素COX回归分析筛选两组数据预后差异均有统计学意义的基因，通过LASSO回归分析构建ICCA的预后风险模型。计算训练集及验证集风险分数，并根据中值分为高、低风险组，绘制Kaplan-Meier生存曲线图和时间依赖性受试者工作特征（receiver operating characteristic, ROC）曲线。将风险分数与临床病理信息进行单、多因素COX回归分析，并绘制列线图展示，综合评价及验证模型效能。利用基因本体论（gene ontology, GO）、京都基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG)、基因集富集分析（Gene Set Enrichment Analysis, GSEA）和单样本基因集富集分析（Single Sample Gene Set Enrichment Analysis, ssGSEA）分析造成高低风险组预后差异的原因。**结果** TCGA数据共筛选出2 922个差异表达基因，GEO数据共筛选出3 075个（均*P*<0.05）。经单因素COX回归分析，TCGA筛选出68个基因（HR=0.13 ~ 7.2，均*P*<0.05），GEO筛选出413个基因（HR=0.17 ~ 215.1，均*P*＜0.05），两组数据预后差异均有统计学意义的有9个基因：GOLGA7B，MTFR2，TPM2，PIWIL4，EPHX4，PRICKLE1，DIO2，FUT4和COL4A3（其中TCGA数据库HR=0.506 ~ 2.760, GEO数据库HR=0.428 ~ 1.992，均*P*<0.05）。LASSO回归成功构建6基因预后风险模型，模型风险分数=0.464×表达量$_{MTFR2}$＋0.550×表达量$_{TPM2}$−0.511×表达量$_{PIWIL4}$−0.097×表达量$_{PRICKLE1}$＋0.215×表达量$_{DIO2}$−0.313×表达量$_{COL4A3}$，训练集中风险分数中值为1.43。Kaplan-Meier生存分析表明在总生存率上，高风险组低于低风险组（*P*<0.001）。ROC曲线提示，1，3，5年AUC分别为0.971（cutoff=0.22），0.921（cutoff=2.33）和0.701（cutoff=1.52），模型预测能力良好。单因素COX回归风险分数HR=5.18(95%CI:2.15 ~ 12.49), *P*<0.001，多因素COX回归风险分数HR=72.5(95%CI:4.52 ~ 1 162.9), *P*=0.002。验证集中模型风险分数中值为2.48。Kaplan-Meier生存分析表明，高风险组生存率低于低风险组（*P*=0.004）。ROC结果显示1，3，5年AUC分别为0.908（cutoff=3.23），0.851（cutoff=1.02）和0.752（cutoff=2.70），单因素COX回归风险分数HR=2.76(95%CI:1.65 ~ 4.60), *P*<0.001，多因素COX回归风险分数HR=4.68(95%CI:2.13 ~ 10.3)，*P*<0.001，风险模型效能得到验证。GO，KEGG，GSEA和ssGSEA分析结果表明造成高低风险组预后差异的原因可能与机体免疫反应的抑制有关（均*P*<0.05）。**结论** 此次构建的预后风险模型在评估ICCA患者预后上具有一定的价值，为临床诊疗提供参考。

关键词：肝内胆管癌；生物信息学；生存分析；风险分数；风险模型
中图分类号：R735.8；R730.43 **文献标识码：**A **文章编号**:1671-7414（2023）03-040-08
doi:10.3969/**j.issn.**1671-7414.2023.03.008

Establishment and Verification of Prognostic Risk Model of Intrahepatic Cholangiocarcinoma Based on TCGA and GEO Database

MAO Jun, SHEN Xiu-fen, MA Run, HE Wei, QU Qiao-li, HU Ying,（*Department of Clinical Laboratory*，*the Second Affiliated Hospital of Kunming Medical University*, *Kunming* 650101, *China*）

Abstract: Objective to construct a prognostic risk model of intrahepatic cholangiocarcinoma (ICCA) based on TCGA（the cancer genome atlas）and GEO（gene expression omnibus）database, and to screen ICCA prognostic related genes. **Methods** The data of 31 cases of ICCA tissues and 9 cases of para-carcinoma tissues in TCGA database were used as training set, and the data of 30 cases of ICCA tissues and 27 cases of para-carcinoma tissues in GEO database were used as verification set. The differentially expressed genes were filtered by R software “DESeq2” package. The filtering conditions were as follows: the

基金项目：云南省高层次卫生健康技术人才培养支持计划（D-2018041）；昆明医科大学硕士研究生创新基金（2022S253）：胆管癌中CPT2表达下调提高顺铂抗性并通过ROS/NFkapaB通路促进肿瘤生长研究。
作者简介：毛俊（1996–），男，硕士研究生，检验技师，研究方向：肿瘤分子生物学，E-mail:m1604777034@163.com。
通讯作者：胡莹（1977–），女，博士研究生，副主任医师，研究方向：临床分子生物学及微生物学，E-mail：hy2002@126.com。

absolute value of difference multiple was more than 2, and the correction $P< 0.05$. Univariate COX regression analysis was used to screen the genes with statistically significant prognosis differences in both groups. LASSO regression analysis was used to construct the prognostic risk model of ICCA. The risk scores of training set and verification set were calculated and divided into high risk group and low risk group according to the median. Kaplan-Meier survival curve and time-dependent receiver operating characteristic (ROC) curve were drawn. The risk score and clinicopathological information were analyzed by univariate and multivariate COX regression analysis, and a line chart was drawn to comprehensively evaluate and verify the effectiveness of the model. Gene Ontology (GO), Kyoto Encyclopedia of Gene and Genomes (KEGG), Gene Set Enrichment Analysis (GSEA) and Single Sample Gene Set Enrichment Analysis (ssGSEA) were used to analyze the reasons for the difference in prognosis between high and low risk groups. **Results** A total of 2 922 differentially expressed genes were screened by TCGA data and 3 075 genes were screened by GEO data (all $P<0.05$). Univariate COX regression analysis showed that 68 genes were screened by TCGA(HR=0.13 ~ 7.2, all $P<0.05$) and 413 genes were screened by GEO (HR=0.17 ~ 215.1, all $P<0.05$). There were 9 genes with significant prognosis in both groups: GOLGA7B，MTFR2，TPM2，PIWIL4，EPHX4，PRICKLE1，DIO2，FUT4 and COL4A3 (TCGA-HR=0.506 ~ 2.760, GEO-HR=0.428 ~ 1.992, all $P<0.05$). A six-gene prognostic risk model was successfully constructed by LASSO regression. The model Risk Score=$0.464 \times EXP_{MTFR2}+0.550 \times EXP_{TPM2}-0.511 \times EXP_{PIWIL4}-0.097 \times EXP_{PRICKLE1}+0.215 \times EXP_{DIO2}-0.313 \times EXP_{COL4A3}$.In training set, the median of risk score was 1.43. Kaplan-Meier survival analysis showed that the overall survival rate in the high risk group was lower than that in the low risk group ($P < 0.001$). The ROC curve showed that the AUC of 1，3 and 5 years were 0.971 (cutoff value=0.22), 0.921 (cutoff value=2.33) and 0.701 (cutoff value=1.52), indicating that the model had good predictive ability. Univariate COX regression risk score HR=5.18 (95%CI:2.15 ~ 12.49, $P < 0.001$), multivariate COX regression risk score HR=72.5 (95%CI:4.52 ~ 1162.9, $P=0.002$). In the verification set, the median risk score of the model was 2.48. Kaplan-Meier survival analysis showed that the survival rate in the high risk group was lower than that in the low risk group($P=0.004$). The results of ROC showed that the AUC values of 1，3 and 5 years were 0.908(cutoff value=3.23), 0.851(cutoff value=1.02) and 0.752(cutoff value=2.7), the univariate COX regression risk score HR=2.76 (95%CI:1.65 ~ 4.60, $P<0.001$), and the multivariate COX regression risk score HR=4.68 (95%CI:2.13 ~ 10.3, $P<0.001$). The effectiveness of the risk model was verified. The results of GO，KEGG，GSEA and ssGSEA analysis showed that the reason for the difference in prognosis between high and low risk groups might be related to the inhibition of immune response(all $P<0.05$). **Conclusion** The prognostic risk model constructed this time has a certain value in evaluating the prognosis of patients with ICCA and provides reference for clinical diagnosis and treatment.

Keywords: intrahepatic cholangiocarcinoma；bioinformatics；survival analysis；risk score；risk model

肝内胆管癌（intrahepatic cholangiocarcinoma，ICCA）是仅次于肝细胞癌的第二常见原发性肝癌，约占 20%[1]。ICCA 现在有望治愈的唯一途径仍然是手术治疗，而能进行手术治疗的患者数量不到三分之一 [1]，并且目前诊断 ICCA 通常采取实验室检查、影像学检查和组织活检等 [2]，缺乏早期特异性标志物，往往错过手术的最佳时期。且经手术后的大多数患者，其 5 年生存率仅 20% ~ 35%[3]。因不能手术而去接受放、化疗等治疗的 ICCA 患者中位生存期仅 12.9 个月 [1]。因此，确定更精准的胆管癌预后指标对临床制定个体精准医疗方案及评估患者生存预后具有重要意义。本研究基于 TCGA 和 GEO 数据库，从 TCGA 数据库下载 TCGA-CHOL 数据作为训练集，GEO 数据库下载 GSE107943 数据作为验证集，通过差异表达基因筛选，单因素 COX 和 LASSO 回归等生信分析建立了 ICCA 的预后风险模型，并在验证集中进行验证。以期能够更好地应用于临床，辅助评估 ICCA 患者的预后，制定针对性的治疗策略。

1 材料与方法

1.1 研究对象 从 TCGA 数据库下载胆管癌（TCGA-CHOL）相关数据，包括表达谱数据、临床病理特征和生存信息。最终获得 36 例癌组织和 9 例癌旁组织，利用 perl 脚本进行整理合并，得到行为样本名称，列为基因名的基因表达矩阵。选择 ICCA 作为研究对象，共有 31 例 ICCA 癌组织及 9 例癌旁组织的相关数据，其中男性 13 例，女性 18 例，平均年龄 62.45 ± 13.23 岁；15 例报告死亡，16 例报告存活，平均存活时间 2.34 ± 1.45 年；26 例白种人，3 例黄种人，2 例黑种人；M 分期：26 例 M0 期，3 例 M1 期，2 例 M 分期未知（MX）；N 分期：24 例 N0 期，4 例 N1 期，3 例 N 分期未知 (NX)；T 分期：17 例 T1 期，9 例 T2 期，5 例 T3 期。临床病理分期：17 例 Stage Ⅰ期，8 例 Stage Ⅱ期，1 例 Stage Ⅲ期，5 例 Stage Ⅳ期，将其作为训练集。从 GEO 数据库下载 GSE107943，获取表达谱数据、临床病理特征和生存信息，共获得 30 例 ICCA 癌组织及 27 例癌旁组织，其中男性 24 例，女性 6 例，

平均年龄 65.6 ± 8.59 岁；17 例报告死亡，13 例报告存活，平均存活时间 2.32 ± 1.71 年；乙型肝炎 4 例，非乙型肝炎 26 例；肿瘤大小：6.13 ± 3.21cm；有血管侵入 12 例，血管未侵入 18 例，临床病理分期：15 例 Stage Ⅰ期，6 例 Stage Ⅱ期，1 例 Stage Ⅲ期，8 例 Stage Ⅳ期，作为验证集。

纳入标准：将肿瘤切除部位为肝内胆管的病例资料作为纳入标准；排除标准：①其他肿瘤切除部位包括肝外胆管、胆囊、肝脏等；②生存时间不足 30 天的患者。

1.2 *方法*

1.2.1 筛选差异表达基因：利用 R 软件“DESeq2”包对两组数据的癌组织和癌旁组织进行差异表达分析，校正 *P* 值（adjust *P* value，padj）< 0.05，|log2 Fold change|>2 被认为差异有统计学意义。

1.2.2 单因素 COX 回归分析：利用“survival”包 [4] 将得到的差异表达基因纳入单因素 COX 回归分析，筛选出预后相关的有统计学差异的表达基因。将两组数据中有预后统计学意义的基因取其交集，得到在 GEO,TCGA 数据中均有预后统计学意义的基因。

1.2.3 构建预后模型：在 TCGA 数据中利用“survival”包 glmnet 函数将在两组数据均有预后意义的差异表达基因进行 10 折交叉验证的 LASSO 回归分析，构建 ICCA 的预后风险模型。并计算每个样本的风险分数，公式如下：

$$风险评分 = \sum \beta_x \times Exp_x$$

β_x 表示 LASSO 回归分析筛选出的各个基因的系数，Exp_x 表示这些基因的表达水平。以风险分数中位数作为截断值将训练集分成高、低风险组。

1.2.4 绘制预后模型的 Kaplan-Meier 曲线和受试者工作特征（receiver operating characteristic,ROC）曲线：对高低风险组利用“survival”包绘制 Kaplan-Meier 曲线进行生存分析，并利用“timeROC”包绘制时间依赖的 ROC 曲线评估模型 1，3，5 年生存的预测效能。

1.2.5 利用 GEO 数据集验证：按上述公式以相同的系数计算验证集各个样本的风险分数并同样以中位数为截断值将验证集划分为高、低风险组，并绘制 Kaplan-Meier 曲线和 ROC 曲线进行验证。

1.2.6 独立预后因子的筛选：将各个样本病理资料与风险分数进行单、多因素 COX 回归分析，以明确影响 ICCA 总体生存率的独立危险因素有哪些，并绘制列线图展示。

1.2.7 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) 分析：高、低风险组间差异表达基因将通过“DESeq2”包来获取，padj < 0.05 且 |log2 Fold change|>2 被认为差异有统计学意义，并进行 GO 和 KEGG 富集分析（“clusterProfiler”包），以 *P*<0.05 为差异有统计学意义。

1.2.8 基因集富集分析（Gene Set Enrichment Analysis，GSEA）：利用“clusterProfiler”包对高低风险表达组进行 GSEA 富集分析，同时下载 MSigDB 数据库中的“c2.cp.kegg.v7.5.1.entrez.gmt”基因集。以此基因集中的通路和基因作为参考，显著富集的通路以 $P < 0.05$, FDR < 0.25 为判断标志。

1.2.9 单样本基因集富集分析（Single Sample Gene Set Enrichment Analysis, ssGSEA）：读入文献定义的 28 种免疫细胞的参考基因集 [5]，利用“GSVA”包 ssGSEA[6] 对高低风险组进行免疫浸润分析，分析高低风险组之间的免疫浸润差异。以 $P < 0.05$ 为差异有统计学意义。

1.3 *统计学分析* 采用 RStudio 4.1.2 对数据进行统计分析，利用单因素 COX 回归筛选具有预后意义的差异表达基因及预后相关病理信息，多因素 COX 回归用以确定影响总体生存率的独立危险因素。LASSO 回归分析选择变量和调整模型复杂度，从而避免过度拟合，并降维构建预后风险模型。Kaplan-Meier 曲线用以评价高低风险组的总体生存率，ROC 曲线评价模型预测效能。以 *P*<0.05 为差异有统计学意义。

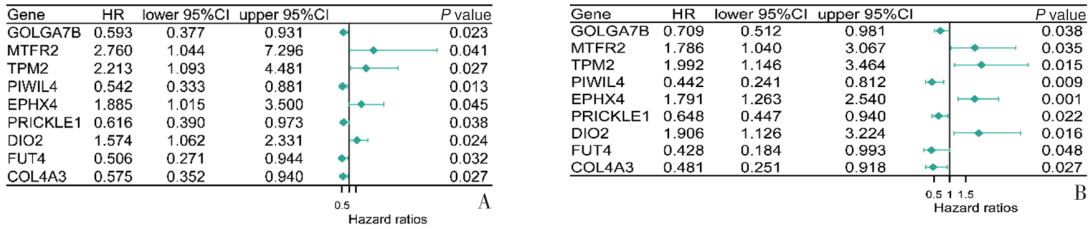
2 结果

2.1 *TCGA 与 GEO 的差异表达基因* TCGA 数据集癌组织和癌旁组织进行差异表达分析后，共获得 2 922 个基因，其中 1 753 个基因表达量升高，1 169 个基因表达量降低。而 GEO 数据集差异表达分析共筛选出 3 075 个差异表达基因，其中 1 920 个表达上调基因，1 155 个表达下调基因（均 *P*<0.05）。

2.2 *预后相关基因* TCGA 的差异表达基因经单因素 COX 回归分析共筛选出 68 个预后相关基因（HR=0.13 ~ 7.2，均 *P*<0.05），GEO 共筛选出 413 个预后相关基因（HR=0.17 ~ 215.1，均 $P < 0.05$）。取两者重叠部分，共获得 9 个基因，见图 1。其中 GOLGA6B, PIWIL4, PRICKLE1 ,FUT4 和 COL4A3 风险比（HR）<1，可能是影响 ICCA 患者预后的保护因素。MTFR2，TPM2，EPHX4 和 DIO2 的 HR>1，可能是影响 ICCA 患者预后的危险因素。

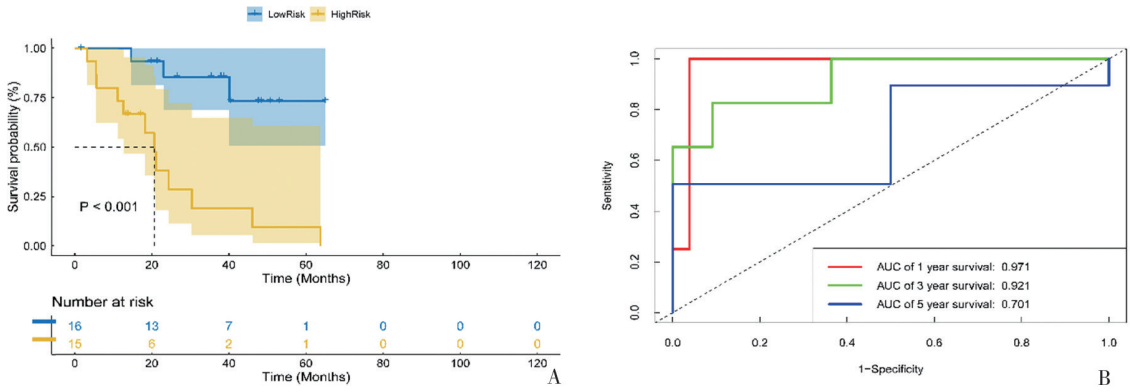
2.3 *预后风险模型及生存分析* 将 9 个交集基因纳入 LASSO 回归分析，进一步降维筛选，最后在训练集 TCGA 数据集中建立了一个 6 个基因（MTFR2，TPM2，PIWIL4，PRICKLE1，DIO2，COL4A3）的预后风险模型。风险评分（Risk Score）=0.464 × 表达量 $_{MTFR2}$ + 0.550 × 表达量 $_{TPM2}$−0.511 × 表达量 $_{PIWIL4}$−0.097 × 表达量 $_{PRICKLE1}$ + 0.215 × 表达量 $_{DIO2}$−

0.313× 表达量 $_{COL4A3}$，以此计算公式得出训练集中每个样本的风险评分，以得分的中位数 1.43 将训练集分成高、低风险组。分析发现，随着得分的增加，ICCA 患者的总体生存期逐渐减少，死亡人数逐渐增加。且 MTFR2，TPM2 和 DIO2 随着风险分数的增加表达量也逐渐增加，而 PIWIL4，PRICKLE1 和 COL4A3 表达量则显示出逐渐降低的趋势。表明 PIWIL4，PRICKLE1 和 COL4A3 可能是影响 ICCA 患者预后的保护因素，MTFR2，TPM2 和 DIO2 可能是影响 ICCA 患者预后的危险因素。Kaplan-Meier 曲线表明高风险组总体存活时间少于低风险组（P<0.001），见图 2A；ROC 曲线中预测 ICCA 患者 1，3，5 年总生存期的 AUC 值分别为 0.971（cutoff=0.22）,0.921（cutoff=2.33） 和 0.701（cutoff=1.52），见图 2B，表明构建的 6 基因模型预测能力良好。



A.TCGA 单因素 COX 回归森林图；B.GEO 单因素 COX 回归森林图

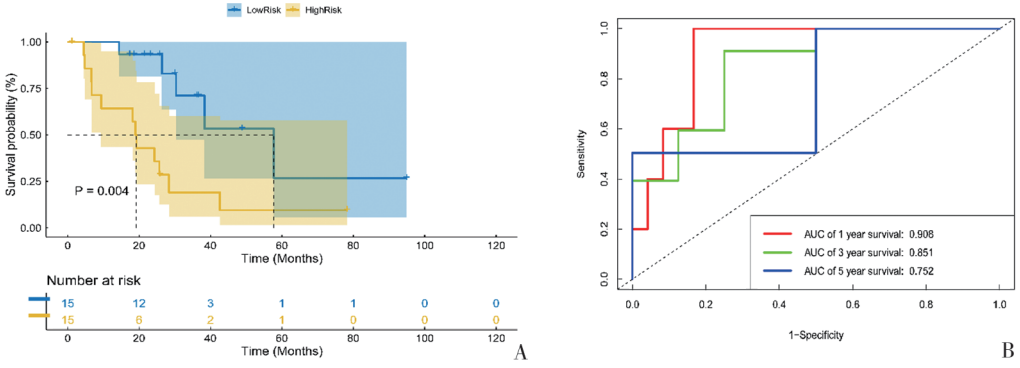
图 1　预后相关基因的筛选



A. 高低风险组 Kaplan-Meier 曲线；B. 时间依赖 ROC 曲线

图 2　训练集组 ICCA 患者生存分析

2.4　预后风险模型的验证　利用与 TCGA 训练集一样的评分计算公式得出 GEO 验证组各个样本风险分数，并采用中值 2.48 为界点使验证集分成高、低风险组，与训练集结果一致，随着风险分数的增加，ICCA 病患的总体生存时间下调，相应的死亡人数上升；且同样能够表明 PIWIL4，PRICKLE1 和 COL4A3 可能是影响 ICCA 患者预后的保护因素，MTFR2，TPM2 和 DIO2 可能是影响 ICCA 患者预后的危险因素。Kaplan-Meier 曲线同样证明两组总体生存时间差异有统计学意义 (P=0.004)，高风险组少于低风险组，见图 3A。ROC 曲线中预测 ICCA 患者 1，3，5 年总生存期的 AUC 值分别为 0.908（cutoff=3.23）,0.851（cutoff=1.02） 和 0.752（cutoff=2.70），见图 3B。表明构建的 6 基因模型在验证集中预测能力同样良好。



A. 高低风险组 Kaplan-Meier 曲线；B. 时间依赖 ROC 曲线

图 3　验证集组 ICCA 患者生存分析

2.5 ICCA 患者的独立预后因子 为了进一步评价所建立的预后风险模型的预测价值，将 TCGA 训练集风险分数与临床病理特征一同纳入单因素 COX 回归分析和多因素 COX 回归分析，单因素 COX 回归分析结果显示风险分数 HR=5.18(95%CI:2.15 ~ 12.49, *P*<0.001)，多因素 COX 回归分析结果显示风险分数 HR=72.5(95%CI:4.52 ~ 1162.9, *P*=0.002)，表明预后模型风险分数是影响 ICCA 预后的独立风险因素；将可能影响患者预后的临床病理特征与风险分数绘制列线图展示，见图 4。

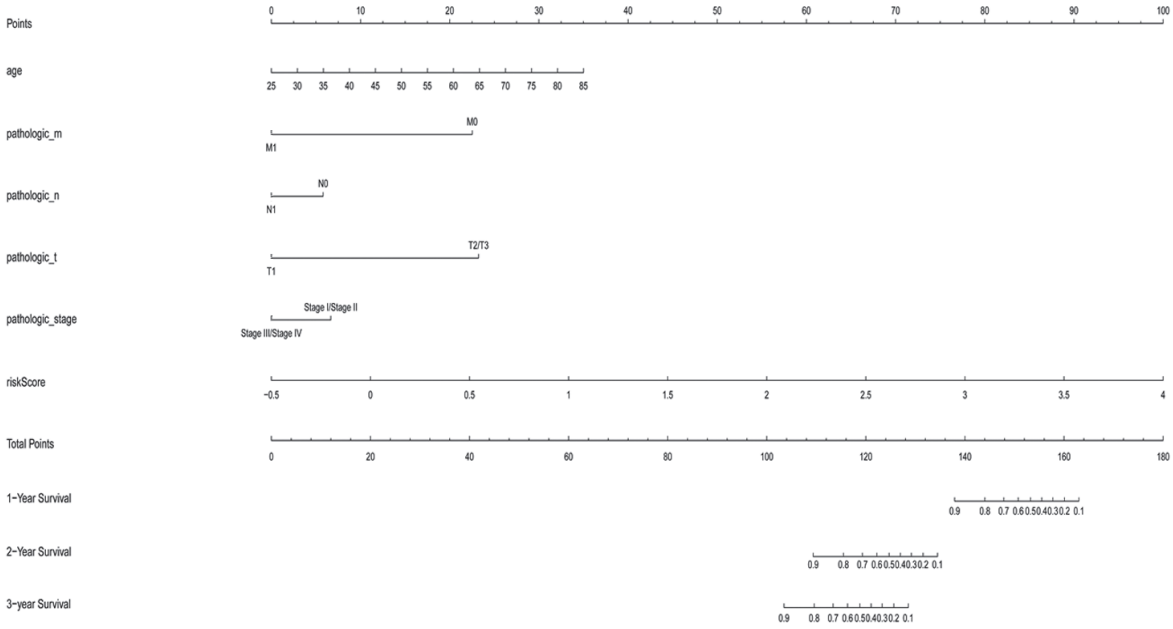


图 4 训练集组风险分数及临床病理特征列线图

在 GEO 验证集中，风险分数与临床病理特征的单因素 COX 回归分析结果显示风险分数 HR=2.76(95%CI:1.65 ~ 4.60), *P*<0.001，多因素 COX 回归分析结果显示风险分数 HR=4.68(95%CI:2.13 ~ 10.3), *P*<0.001，证明风险分数在验证集中同样是独立危险因子。绘制列线图见图 5。

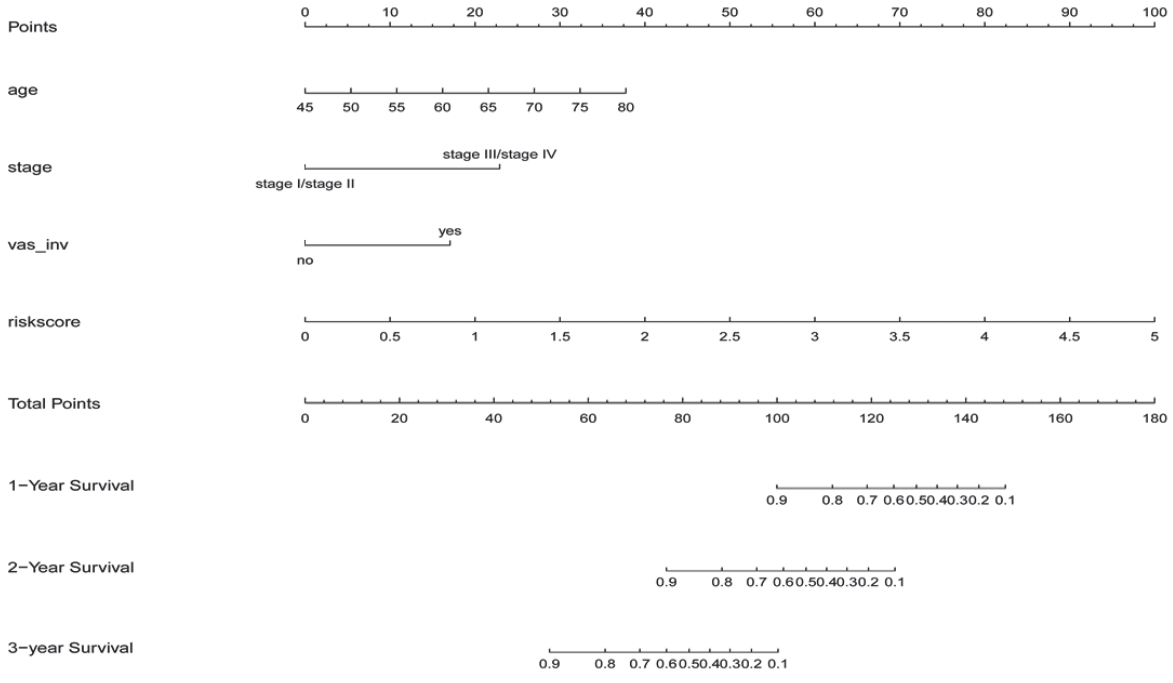


图 5 验证集组风险分数及临床病理特征列线图

2.6　GO 和 KEGG 富集分析　为了探究高、低风险组预后差异的原因，分析了高、低风险组差异基因的功能和可能参与的通路，将筛选到的 98 个在高、低风险组表达有差异的基因进行 GO,KEGG 富集分析。GO 富集分析提示高低风险差异基因功能主要富集在机体免疫反应及调节相关方面，如“体液免疫反应”（P<0.001, 校正 P=0.005, q value=0.004）、“体液免疫调节”（P<0.001, 校正 P=0.02，q value=0.02）、“免疫反应与激活细胞表面受体信号通路及信号传导”（P<0.001, 校正 P=0.02, q value=0.02）等。KEGG 结果与 GO 富集结果遥相呼应，都与机体免疫相关，包括“B 细胞受体信号通路”（P=0.006, 校正 P=0.24，q value=0.23）、“原发性免疫缺陷”（P=0.01, 校正 P=0.33, q value=0.32)等。结合 GO，KEGG 富集结果，表明高低风险组之间的差异可能是由于机体的免疫反应及免疫调节导致的。

2.7　GSEA 富集分析　为了进一步验证高、低分险组是否通过影响免疫反应及免疫调节来导致两组间的预后差异，进行了 GSEA 富集分析，结果显示高低风险组基因富集于“B 细胞受体信号通路”（富集分数 =–0.57，NES=–2.10，P<0.001, 校正 P 值 <0.001，FDR<0.001）、“原发性免疫缺陷”（富集分数 =–0.77，NES=–2.44，P<0.001, 校正 P 值 <0.001，FDR<0.001）等与 KEGG 结果一致，此外还富集在“趋化因子信号通路”（富集分数 =–0.52，NES=–2.29，P<0.001，校正 P 值 <0.001，FDR<0.001）、“T 细胞受体信号通路”（富集分数 =–0.46，NES=–1.81，P<0.001, 校正 P 值 =0.001，FDR=0.001）等免疫相关信号通路，并且这些免疫信号通路在高风险组均处于抑制状态（富集分数 <0）。这提示高分险组可能通过抑制机体的免疫反应来促进肿瘤的恶性进展，进而导致高风险组的预后不良。

2.8　ssGSEA 免疫浸润分析　在确定高、低风险组的预后差异主要由于影响机体免疫反应及免疫调节的前提下，对高、低风险组进行了 ssGSEA 免疫浸润分析，以下载的 28 种免疫细胞基因集为参照对象，评估这些细胞在高低风险组的分布及差异情况。依据 CHAROENTONG 等 [5] 人定义的基因集将 28 种免疫细胞分为三类，分别为抗肿瘤免疫细胞、促肿瘤免疫细胞和功能不明确的免疫细胞。ssGSEA 结果表明促肿瘤免疫细胞中 $CD56^-$ 自然杀伤细胞在高风险组表达量相对于低风险组更高（富集分数：高风险组 0.31 ~ 0.46，低风险组 0.31 ~ 0.42，P<0.05），而抗肿瘤细胞中 CD4 中心记忆型 T 细胞则表现出相反的结果（富集分数：高风险组 0.51 ~ 0.59，低风险组 0.44 ~ 0.62，P<0.05），此外，激活的 B 细胞，激活的 $CD8^+$T 细胞也表现出在高风险组表达抑制的状态，见图 6。这进一步验证了富集分析的高风险免疫抑制状态。

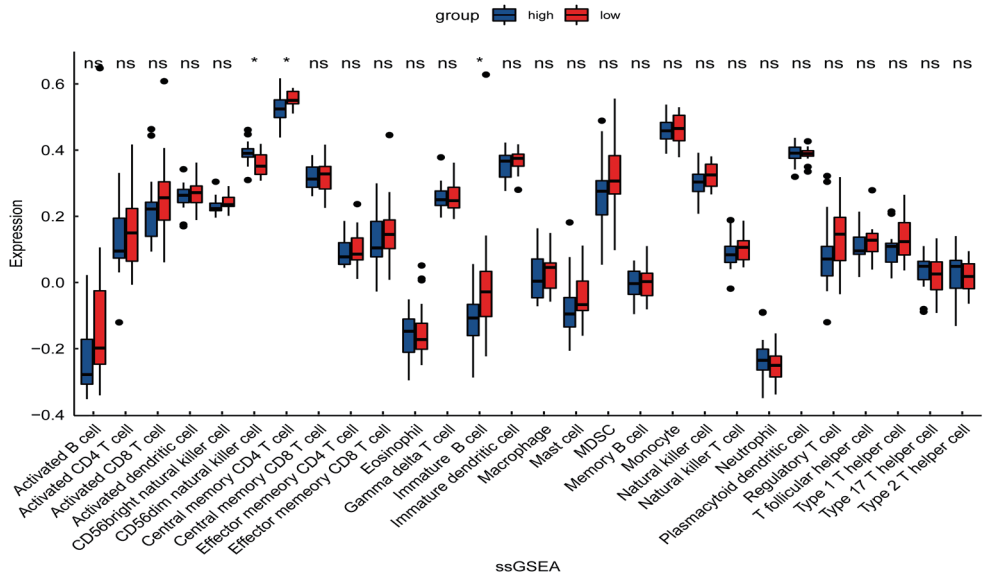


图 6　ssGSEA 免疫浸润分析

3　讨论

本研究基于 TCGA 和 GEO 数据库构建了一个 6 基因预后风险模型，包括 MTFR2，TPM2，DIO2，PIWIL4，PRICKLE1 和 COL4A3，模型的质量也在 GEO 数据集中得到了验证。其中 MTFR2 和 TPM2 的高表达与多种肿瘤的不良预后有关，包括肝细胞癌 [7]、肺腺癌 [8]、乳腺肿瘤 [9-11] 和口腔癌 [12] 等，DIO2 在浆液性卵巢癌肠道转移中过表达，表明

DIO2 可能参与肿瘤的恶性转移[19]。此外，已有研究表明，PIWIL4 能作为 ICCA 的预后生物标志物[13]，且其低表达与肾透明细胞癌患者的不良预后有关[14]。COL4A3 的表达与鼻咽癌[15]和胃癌[16]的预后呈负相关关系。PRICKLE1 的高表达减少了神经母细胞瘤的生长，并与患者的预后呈负相关关系[17]。以上研究表明这 6 个基因在 ICCA 中发挥的作用和其他肿瘤大致相同，表明研究建立的预后风险模型有望成为 ICCA 患者预测性能良好的预后生物标志物。

通过 GO，KEGG，GSEA 和 ssGSEA 分析提示 ICCA 患者的预后可能与机体免疫反应有关，这与黄建斌[18]的结论是一致的。而 MTFR2，TPM2，DIO2，PIWIL4 和 COL4A3 均与机体免疫密切相关，例如：MTFR2 的表达在胃癌中与 B 细胞、$CD8^+$T 细胞、$CD4^+$T 细胞、巨噬细胞、中性粒细胞和树突状细胞的渗透水平显著相关[26]，TPM2 与膀胱癌巨噬细胞和 NK 细胞浸润呈正相关[27]，DIO2 在炎症诱导的巨噬细胞中表达升高，并可能成为参与 COPD 机制过程的标志物[28]，PIWIL4 在 ICCA 中与静息自然杀伤细胞和激活的记忆 $CD4^+$T 细胞的富集呈正相关[20]，COL4A 家族基因的表达在肾透明细胞癌中与免疫细胞、肿瘤浸润淋巴细胞、MHC 分子、趋化因子和受体的浸润显著相关[29]。另一方面，机体免疫与肿瘤的关系主要包括免疫监视和肿瘤免疫逃逸。ssGSEA 免疫浸润分析部分指出，肿瘤可以驱动免疫抑制或调节性免疫细胞亚型的产生，也可以招募大量促肿瘤的髓系细胞来建立肿瘤微环境，从而促进肿瘤的进展。肿瘤恶性进展的关键步骤是逃避免疫破坏和启动肿瘤细胞转移，这些步骤可以通过抑制宿主的免疫系统来实现，特别是通过诱导、扩增和重新招募免疫抑制细胞[30]。由此推测 ICCA 可能通过免疫逃逸机制来实现肿瘤恶性进展，从而导致 ICCA 患者预后不良。早期研究已经表明，ICCA 能够通过 B7-H1/PD-1 通路促进 $CD8^+$ 肿瘤浸润性淋巴细胞（$CD8^+$TILs）的凋亡而参与肿瘤免疫逃逸[19]。SURIYO 等[20]的研究也表明 ICCA 能够通过 PD-L1/PD-1 轴抑制 T 细胞介导的免疫反应，使 $CD8^+$T 细胞凋亡增加，实现肿瘤细胞的免疫逃逸。而 $CD8^+$T 细胞数量的减少与 ICCA 患者的总体生存时间短密不可分[21]。此外，有学者发现 sPDL1 相对于其他类型胆管癌，在 ICCA 中表达最高，sPDL1 低水平患者表现出较长的生存时间，表明 sPDL1 高水平会增加患者死亡风险[22]。目前临床上针对免疫检查点，已经开发出免疫检查点抑制剂，如 PD-1 抑制剂和 PD-L1 抑制剂。且在 ICCA 的治疗中显示出较高的有效率[23]。因此，从肿瘤免疫出发，针对性进行治疗能使 ICCA 患者总体生存期升高，这具有重要的临床意义。

综上所述，本研究建立的风险预后模型对评估 ICCA 患者预后具有较高的临床应用价值，且对于高风险组患者的针对性疗法具有指导意义，为利用免疫检查点抑制剂来提高高风险组患者预后开创了新平台，但这仍需要大量的临床研究去论证。在未来，有望能够利用免疫抑制剂联合靶向药物进行综合治疗，这或许在胆管癌中能获得更好的治疗效果。

本研究同样存在一定的局限性。首先，本研究是通过在线公共数据库中的数据构建的预后模型，未使用本院 ICCA 患者病理标本及病理资料进行模型验证。其次，对组成预后模型的 6 个基因的作用及机制还需进行进一步的研究。

参考文献：

[1] HEWITT D B, BROWN Z J, PAWLIK T M. Surgical management of intrahepatic cholangiocarcinoma [J]. Expert Review of Anticancer Therapy, 2022, 22(1):27-38.

[2] BUCKHOLZ A P, BROWN R S. Cholangiocarcinoma: Diagnosis and Management [J]. Clinics in Liver Disease, 2020, 24(3):421-436.

[3] ZHANG X F, BEAL E W, BAGANTE F, et al. Early versus late recurrence of intrahepatic cholangiocarcinoma after resection with curative intent[J]. The British Journal of Surgery, 2018, 105(7): 848-856.

[4] MORENO-BETANCUR M, CARLIN J B, BRILLEMAN S L, et al. Survival analysis with time-dependent covariates subject to missing data or measurement error: Multiple Imputation for Joint Modeling (MIJM)[J]. Biostatistics, 2018, 19(4): 479-496.

[5] CHAROENTONG P, FINOTELLO F, ANGELOVA M, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade[J]. Cell Reports, 2017, 18(1): 248-262.

[6] BARBIE D A, TAMAYO P, BOEHM J S, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1[J]. Nature, 2009, 462(7269): 108-112.

[7] LI Dan, JI Yanmei, GUO Jialong, et al. Upregulated expression of MTFR2 as a novel biomarker predicts poor prognosis in hepatocellular carcinoma by bioinformatics analysis[J]. Future Oncology, 2021, 17(24): 3187-3201.

[8] CHEN Cheng, TANG Yang, QU Wendong, et al. Evaluation of clinical value and potential mechanism of MTFR2 in lung adenocarcinoma via bioinformatics[J]. BMC Cancer, 2021, 21(1): 619.

（下转第 64 页）

2016, 56(4): 570-577.
[4] PAPROCKA P, DURNAŚ B, MAŃKOWSKA A, et al. *Pseudomonas aeruginosa* infections in cancer patients[J]. Pathogens，2022, 11(6): 679.
[5] SEELY K D, MORGAN A D, HAGENSTEIN L D, et al. Bacterial involvement in progression and metastasis of colorectal neoplasia[J]. Cancers (Basel), 2022, 14(4): 1019.
[6] BESSER J, CARLETON H A, GERNER-SMIDT P, et al. Next-generation sequencing technologies and their application to the study and control of bacterial infections[J]. Clinical Microbiology and Infection, 2018, 24(4):335-341.
[7] RAPOPORT B L. Management of the cancer patient with infection and neutropenia[J]. Seminars in Oncology, 2011, 38(3): 424-430.
[8] GHONEUM A, ALMOUSA S, WARREN B, et al. Exploring the clinical value of tumor microenvironment in platinum-resistant ovarian cancer[J]. Seminars in Cancer Biology, 2021, 77: 83-98.

收稿日期：2022-09-20
修回日期：2022-12-08

（上接第 46 页）

[9] LU Guanming, LAI Yuanhui, WANG Tian-tian, et al. Mitochondrial fission regulator 2 (MTFR2) promotes growth, migration, invasion and tumour progression in breast cancer cells[J]. Aging, 2019, 11(22): 10203-10219.
[10] LU Wenjie, ZANG Rukun, DU Yuanna, et al. Overexpression of MTFR2 predicts poor prognosis of breast cancer[J]. Cancer Management and Research, 2020, 12: 11095-11102.
[11] ZHANG Jinfeng, ZHANG Jian, XU Shouping, et al. Hypoxia-induced TPM2 methylation is associated with chemoresistance and poor prognosis in breast cancer[J]. Cellular Physiology and Biochemistry, 2018, 45(2): 692-705.
[12] ZHAO Xiaotong, ZHU Yan, ZHOU Jiefu, et al. Development of a novel 7 immune-related genes prognostic model for oral cancer: A study based on TCGA database[J]. Oral Oncology, 2021, 112: 105088.
[13] ZOU Wenbo, WANG Zizheng, ZHANG Xiuping, et al. PIWIL4 and SUPT5H combine to predict prognosis and immune landscape in intrahepatic cholangiocarcinoma[J]. Cancer Cell International, 2021, 21(1): 657.
[14] ILIEV R, STANIK M, FEDORKO M, et al. Decreased expression levels of PIWIL1, PIWIL2, and PIWIL4 are associated with worse survival in renal cell carcinoma patients[J]. OncoTargets and Therapy, 2016, 9: 217-222.
[15] YANG Xiting, WU Qiuji, WU Fengyang, et al. Differential expression of COL4A3 and collagen in upward and downward progressing types of nasopharyngeal carcinoma[J]. Oncology Letters, 2021, 21(3): 223.
[16] NIE Xiaocui, WANG Jianping, ZHU Wan, et al. COL4A3 expression correlates with pathogenesis, pathologic behaviors, and prognosis of gastric carcinomas[J]. Human Pathology, 2013, 44(1): 77-86.
[17] DYBERG C, PAPACHRISTOU P, HAUG B H, et al. Planar cell polarity gene expression correlates with tumor cell viability and prognostic outcome in neuroblastoma[J]. BMC Cancer, 2016, 16(1): 259.
[18] 黄健斌. 基于 TCGA 数据库肝内胆管细胞癌的 4-mRNA 预后风险模型构建 [D]. 广州：南方医科大学 , 2021.
HUANG Jianbin. Identification of prognostic four-mRNA signature model in intrahepatic cholangiocarcinoma based on TCGA database [D]. Guangzhou：Southern Medical University, 2021.
[19] YE Yufu, ZHOU Lin, XIE Xiajun, et al. Interaction of B7-H1 on intrahepatic cholangiocarcinoma cells with PD-1 on tumor-infiltrating T cells as a mechanism of immune evasion[J]. Journal of Surgical Oncology, 2009, 100(6): 500-504.
[20] SURIYO T, FUANGTHONG M, ARTPRADIT C, et al. Inhibition of T-cell-mediated immune response via the PD-1/ PD-L1 axis in cholangiocarcinoma cells[J]. European Journal of Pharmacology, 2021, 897: 173960.
[21] KITANO Y, OKABE H, YAMASHITA Y I, et al. Tumour-infiltrating inflammatory and immune cells in patients with extrahepatic cholangiocarcinoma[J]. British Journal of Cancer, 2018, 118(2): 171-180.
[22] 遆振宇 , 高小鹏 , 千东维 , 等. 血清 sPDL1 水平和外周血 NLR 在判断晚期胆管癌患者生存预后中的意义 [J]. 现代检验医学杂志 , 2019, 34(6): 41-46.
TI Zhenyu, GAO Xiaopeng, QIAN Dongwei, et al. Soluble programmed death-ligand 1 (sPDL1) and Neutrophilto-Lymphocyte ratio (NLR) predicts prognostic survival in advanced biliary tract cancer patients treated with palliative chemotherapy[J]. Journal of Modern Laboratory Medicine, 2019, 34(6): 41-46.
[23] ZENG Fanli, CHEN Jingfang. Application of immune checkpoint inhibitors in the treatment of cholangiocarcinoma[J]. Technology in Cancer Research & Treatment, 2021, 20: 15330338211039952.

收稿日期：2022-11-16
修回日期：2023-01-19