

## 论著·基础研究

## 胃癌患者预后相关微RNA预测模型的构建及其应用价值探讨

岳 犇, 王高明, 杨鹿笛, 崔 然, 郁丰荣

上海交通大学医学院附属仁济医院胃肠外科, 上海 200127

**[摘要]** **目的**·通过生物信息学方法建立胃癌患者预后相关微RNA (microRNA, miRNA) 的预测模型并探讨其应用价值。**方法**·收集癌症和肿瘤基因组图谱计划 (TCGA) 数据库中 397 例胃癌患者的临床病理资料, 其中 356 例临床病理资料完整。397 例患者中, 男 258 例, 女 139 例; 中位年龄为 67 岁。将 397 例患者采用随机抽样法, 按 7:3 比例分为训练集 278 例和测试集 119 例。另从基因表达数据库 (Gene Expression Omnibus, GEO) 中下载包含 20 对胃癌及对应癌旁正常组织的 miRNA 测序数据集 GSE93415, 从癌组织和癌旁正常组织差异表达的 miRNA 中筛选出候选差异表达 miRNA。基于训练集患者的信息, 利用 LASSO 回归分析, 将候选差异表达 miRNA 拟合成一个可以预测胃癌患者生存率的预后相关 miRNA 模型。对构建的预后相关 miRNA 模型的预测效能, 分别在训练集和测试集中进行验证, 以 Log-Rank 检验进行生存分析来验证模型的可靠性; 以受试者操作特征曲线下面积 (area under curve, AUC) 分析该模型的预测效能。使用基因表达数据和 pRRophetic 包中的算法预测患者对化学治疗药物的敏感性。使用 Calibration 曲线验证恩诺图的准确性。使用一致性指数 (consistency index, C-index) 检测恩诺图和其他因素建模的一致性。使用决策曲线分析 (decision curve analysis, DCA) 推测候选因素对于临床决策的帮助。**结果**·①训练集与测试集患者的临床资料和总体生存率比较, 差异均无统计学意义 (均  $P > 0.05$ )。②差异表达 miRNA 筛选结果: 从 GSE93415 测序数据集中计算得到 111 个候选差异表达 miRNA, 其中 20 个在癌组织中上调, 91 个在癌组织中下调。对 111 个候选差异表达 miRNA 进行过滤后, 得到 59 个候选差异表达 miRNA。③构建预后相关 miRNA 模型: 从 59 个候选 miRNA 中, 筛选出 5 个与生存相关的 miRNA, 分别为 let-7i-5p、let-7f-5p、miR-708-5p、miR-135b-5p、miR-100-5p, 差异表达模式 (癌组织对比癌旁组织) 均为降低, 差异表达倍数分别 2.55、2.78、2.17、3.08、3.26 倍。由 5 个与生存相关 miRNA 构建的预后表达方程为: 风险分数 =  $(-0.049 \times \text{let-7i-5p 表达量} - 0.033 \times \text{let-7f-5p 表达量} + 0.202 \times \text{miR-708-5p 表达量} - 0.088 \times \text{miR-135b-5p 表达量} + 0.016 \times \text{miR-100-5p 表达量})$ 。④预后相关 miRNA 模型的验证: 在训练集和测试集中, 高风险组患者的总体生存率低于低风险组患者, 差异均有统计学意义 (均  $P < 0.05$ )。在训练集中, 预后相关 miRNA 模型的 1 年、3 年、5 年生存时间预测概率 AUC 值分别为 0.640、0.763、0.853; 在测试集中, 则分别为 0.631、0.735、0.750。⑤影响胃癌患者预后的相关因素分析: 单因素分析结果显示年龄、肿瘤病理分期、T 分期、N 分期、M 分期和预后相关 miRNA 模型评分是胃癌患者预后的相关因素 ( $HR=1.017、1.633、1.353、1.346、2.652、15.874$ ,  $95\%CI$  为  $1.002\sim1.033、1.333\sim2.001、1.109\sim1.650、1.169\sim1.548、1.553\sim4.529、5.729\sim43.985$ ,  $P < 0.05$ )。多因素分析结果显示年龄、M 分期和预后相关 miRNA 模型评分是胃癌患者预后的独立危险因素 ( $HR=1.03、2.27、18.72$ ,  $95\%CI$  为  $1.01\sim1.05、1.09\sim4.70、5.96\sim58.77$ ,  $P < 0.05$ )。⑥预后相关 miRNA 模型与临床病理因素预测效能的比较: 在 356 例临床资料完整的胃癌患者中, 预后相关 miRNA 模型对胃癌患者 5 年生存时间预测的 AUC 值为 0.818, 高于年龄、性别、肿瘤病理分期、T 分期、N 分期、M 分期的预测效能。⑦预后相关恩诺图的评估: Calibration 验证曲线、C-index 以及 DCA 分析均提示, 胃癌患者从预后相关恩诺图中的获益程度高于年龄、性别和肿瘤病理分期。**结论**·由 5 个差异表达的 miRNA 构建了可用于预测胃癌患者生存的预后相关 miRNA 模型; 该模型对胃癌患者的生存状态和预后具有一定的区分预测能力, 可为胃癌患者的临床治疗提供参考。

**[关键词]** 胃癌; LASSO 回归模型; 微 RNA; 预测模型; 受试者操作特征曲线

**[DOI]** 10.3969/j.issn.1674-8115.2021.11.006 **[中图分类号]** R735.2 **[文献标志码]** A

## Construction and application value of prognosis-associated miRNA prediction model in gastric cancer patients

YUE Ben, WANG Gao-ming, YANG Lu-di, CUI Ran, YU Feng-rong

Department of Gastrointestinal Surgery, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200127, China

**[Abstract]** **Objective**·To construct a prognosis-associated microRNA (miRNA) prediction model in gastric cancer patients based on bioinformatics analysis and evaluate its application value. **Methods**·The clinicopathological data of gastric cancer patients were downloaded from the Cancer Genome Atlas (TCGA). There were 258 males and 139 females with a median age of 67 years. Three hundred and fifty-six of the 397 patients had complete clinicopathological data. The 397 patients were allocated into training cohort consisting of 278 patients and validation cohort consisting of 119 patients using the random sampling method, with a ratio of 7:3. A miRNA sequencing dataset GSE93415 containing 20 pairs of gastric cancer and corresponding adjacent normal tissue was downloaded from Gene Expression Omnibus database. The candidate miRNAs were selected from differentially expressed

**[基金项目]** 上海市核酸化学与纳米医学重点实验室“临床+”卓越项目 (2020ZYA008)。

**[作者简介]** 岳 犇 (1986—), 男, 住院医师, 博士; 电子信箱: yueben@163.com。

**[通信作者]** 郁丰荣, 电子信箱: rry77@126.com。

**[Funding Information]** Shanghai Key Laboratory of Nucleic Acid Chemistry and Nanomedicine "Clinical Plus" Excellence Project (2020ZYA008).

**[Corresponding Author]** YU Feng-rong, E-mail: rry77@126.com.



miRNAs in gastric cancer and adjacent tissue. A prognosis associated miRNA prediction model was constructed upon survival-associated miRNAs which were selected from candidate miRNAs through LASSO-Cox regression analysis. The performance of prognosis-associated miRNA prediction model was validated in the training cohort and validation cohort. The reliability of the model was evaluated by using Log-Rank test, and the accuracy of the model was evaluated by using the area under curve (AUC) of the receiver operating characteristic curves. Gene expression profiles and algorithms in the pRRophetic package were utilized to predict patients' sensitivity to chemotherapy drugs. Calibration curves were used to verify the accuracy of the nomogram. Consistency index (C-index) was used to check the consistency of the nomogram and other factors. Decision curve analysis (DCA) was employed to predict the contribution of candidate factors to clinical decision making. **Results** ① There was no significant difference in the baseline and overall survival between the training cohort and validation cohort ( $P>0.05$ ). ② There were 111 differentially expressed miRNAs calculated from GSE93415 dataset, of which 20 were up-regulated in tumor tissue while 91 were down-regulated. Fifty-nine miRNAs were selected as candidate miRNAs after filtration. ③ Among the 59 candidate miRNAs, 5 survival-associated miRNAs were selected, including let-7i-5p, let-7f-5p, miR-708-5p, miR-135b-5p and miR-100-5p. The differential expression patterns of gastric cancer to adjacent tissue were all down-regulation, with the fold change of 2.55, 2.78, 2.17, 3.08 and 3.26. Risk score =  $(-0.049 \times \text{let-7i-5p expression level} - 0.033 \times \text{let-7f-5p expression level} + 0.202 \times \text{miR-708-5p expression level} - 0.088 \times \text{miR-135b-5p expression level} + 0.016 \times \text{miR-100-5p expression level})$ . ④ In the training cohort and the validation cohort, the overall survival rate of patients in the high-risk group was lower, and the difference was statistically significant ( $P<0.05$ ). The AUC of the prognosis-associated miRNA model for 1-, 3- and 5-year survival prediction was 0.640, 0.763 and 0.853, and was 0.631, 0.735 and 0.750 in validation cohort. ⑤ Results of univariate analysis showed that age, tumor pathological stage, T stage, N stage, M stage and prognosis-associated miRNA model score were related factors for prognosis of gastric cancer patients ( $HR=1.017, 1.633, 1.353, 1.346, 2.652, 15.874$ ; 95%CI 1.002–1.033, 1.333–2.001, 1.109–1.650, 1.169–1.548, 1.553–4.529, 5.729–43.985;  $P<0.05$ ). Results of multivariate analysis showed that age, M stage and prognosis-associated miRNA model score were independent risk factors for prognosis of gastric cancer patients ( $HR=1.03, 2.27, 18.72$ ; 95%CI 1.01–1.05, 1.09–4.70, 5.96–58.77;  $P<0.05$ ). ⑥ The AUC of the prognosis-associated miRNA model for 5-year survival prediction was 0.818 in 356 gastric cancer patients with complete clinicopathological data, higher than that of age, gender, tumor pathological stage, T stage, N stage, M stage and merged clinical factors. ⑦ The results of Calibration curves, C-index, and DCA all indicated that patients with gastric cancer may get more net benefits from the prognostic nomogram than age, gender and tumor pathological stage. **Conclusion** A prognosis-associated miRNA prediction model that can be used to predict the survival of gastric cancer patients is constructed based on 5 miRNAs, which has a certain predictive ability to distinguish the survival status and prognosis of gastric cancer patients and can provide reference for clinical treatment.

**[Key words]** gastric cancer; LASSO regression model; microRNA; prediction model; receiver operating characteristic curve

胃癌发病率和病死率居全球第5位和第3位; 尽管手术是目前主要的治疗手段, 但半数以上行根治性切除手术的胃癌患者出现局部复发或远处转移, 影响患者的预后<sup>[1-2]</sup>。异常表达的microRNA (microRNA, miRNA) 是胃癌的特征之一<sup>[3]</sup>。miRNA是一类含有17~25个核苷酸的内源性非编码RNA, 在转录后水平调控基因的表达<sup>[4]</sup>。越来越多的研究<sup>[5-6]</sup>证明, miRNA参与肿瘤的增殖、凋亡、侵袭/转移和血管生成等生物学过程, 与肿瘤的发生、发展密切相关, 可作为肿瘤诊断和预后的潜在标志物。近年来, 随着基因检测技术的不断发展, 高通量测序技术结合生物信息学分析, 已经在临床应用中取得了重要的成果。本研究基于基因表达数据库 (Gene Expression Omnibus, GEO) 和癌症基因组图谱 (The Cancer Genome Atlas, TCGA), 回顾性分析397例胃癌患者的临床病理资料, 基于生物信息学构建胃癌患者预后相关miRNA预测模型并探讨其应用价值, 为胃癌患者个体化诊疗方案的制定提供参考。

## 1 资料与方法

### 1.1 临床资料

采用回顾性队列研究方法。利用R 4.0.3中的TCGAbiolinks包收集TCGA-STAD中397例胃癌患者的临床资料<sup>[7]</sup>, 其中男258例、女139例, 中位年龄67岁(30~88岁)。397例患者中, 356例临床病理资料完整。

将397例患者采用随机抽样法按7:3比例分为训练集(278例)和测试集(119例); 训练集用于构建预测模型, 测试集用于验证模型效能。

纳入标准: ①病理学确诊为胃癌。②具有miRNA的样本测序数据。③具有预后信息及临床资料。排除标准: ①正常胃组织样本信息。②转移性胃癌样本信息。③生存时间<10 d。

### 1.2 研究方法

**1.2.1 差异miRNA筛选** 利用美国国立生物技术信息中心 (National Center for Biotechnology Information, NCBI) 平台下的GEO数据库检索含有人源胃癌及癌旁组织样本的芯片GSE93415。该芯片共包含40例组织样本, 其中20例为癌组织, 20例为癌旁正常组织。通过GEO Datasets中GEO2R在线分析癌组织和癌旁正常组织差异表达的miRNA, 并筛选出候选差异表达miRNA<sup>[8]</sup>。

**1.2.2 预后相关miRNA模型的构建** 使用R 4.0.3中的DESeq2包对TCGA样本的基因表达数据进行标准化<sup>[9]</sup>。为了提高预测模型的准确性, 首先在候选差异表达miRNA中过滤掉低表达的miRNA, 过滤条件为在397个TCGA胃癌样本中标准化后的miRNA表达量之和>100。基于训练集患者信息, 利用R 4.0.3中的glmnet包进行LASSO回归分析, 将候选差异表达miRNA拟合成一个相对精简的预后相关miRNA模型<sup>[10-11]</sup>。

**1.2.3 预后相关 miRNA 模型的验证** 分别在训练集和测试集中对构建的预后相关 miRNA 模型的预测效能进行验证, 用生存曲线和列线图展示<sup>[12-13]</sup>。

**1.2.4 影响胃癌患者预后的相关因素分析** 选取临床病理资料完整的 356 例胃癌患者的临床病理资料, 分析年龄、性别、肿瘤病理分期、T 分期、N 分期、M 分期以及预后相关 miRNA 模型评分与胃癌患者预后之间的相关性。

**1.2.5 预后相关 miRNA 模型与临床病理因素预测效能比较** 在临床病理资料完整的数据集中比较构建的预后相关 miRNA 模型与临床病理因素对胃癌患者 5 年总体生存率的预测效能。

**1.2.6 对化学治疗药物敏感度的预测** 利用 pRRophetic 包中的算法和肿瘤样本的基因表达数据, 估算患者对于不同化学治疗 (化疗) 药物的敏感度<sup>[14]</sup>。

**1.2.7 预后相关恩诺图的构建和验证** 使用多因素 Cox 回归的方法, 将预后相关 miRNA 模型评分与患者年龄、性别以及肿瘤病理分期综合起来, 构建一个相对准确的量化的恩诺图。使用 Calibration 曲线验证恩诺图的预测效果; 其中 45°虚线为参考线, 表示预测的总体生存率与实际总体生存率完全一致; 曲线越接近参考线, 预测效果越好。

1.3 观察指标和评价标准

**1.3.1 观察指标** ①训练集和测试集患者的临床资料和生存情况。②预后相关 miRNA 模型的构建: 筛选与生存相关的 miRNA, 构建预后相关 miRNA 模型方程。③预后相关 miRNA 模型的验证: 在训练集和测试集中的预测效能。④影响胃癌患者预后的相关因素分析: 年龄、性别、肿瘤病理分期、T 分期、N 分期、M 分期、预后相关 miRNA 模型评分。⑤预后相关 miRNA 模型与临床病理因素预测效能的比较: 在 356 例临床病理资料完整的数据集中的预测效能。

**1.3.2 评价标准** 差异表达 miRNA 筛选标准为在癌组织及癌旁组织中差异表达倍数>2 且校正后的  $P$  值<0.05<sup>[15]</sup>。以 Log-rank 检验进行生存分析, 验证模型的可靠性; 以受试者操作特征曲线 (receiver operating characteristic, ROC) 曲线下面积 (area under curve, AUC) 分析模型的准确性, AUC 值越接近 1, 模型的预测效能越好<sup>[13]</sup>。

1.4 统计学分析

应用 R 4.0.3 统计软件进行数据分析。正态分布的数据以  $\bar{x} \pm s$  表示, 2 组间比较采用 Student- $t$  检验。偏态分布的数据以  $M (P_{25} \sim P_{75})$  表示, 组间比较采用 Mann-Whitney  $U$  检验。定性资料以绝对数或百分比表示, 组间

比较采用  $\chi^2$  检验。采用 Kaplan-Meier 法绘制生存曲线并计算生存率, 采用 Log-Rank 方法比较 2 组患者的生存状况。 $P < 0.05$  表示差异有统计学意义。

2 结果

2.1 患者基本资料及生存情况

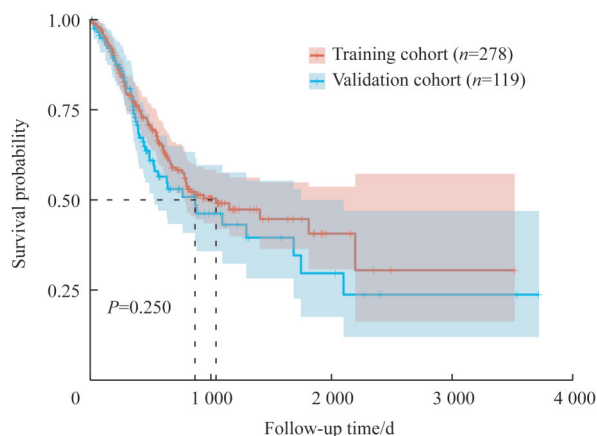
2 组患者性别、年龄、肿瘤分期、肿瘤位置、生存状态、生存时间、种族、是否接受化疗、是否接受放射治疗等一般资料比较, 差异均无统计学意义 ( $P > 0.05$ ), 具有可比性, 见表 1。

表 1 训练集与测试集患者一般资料比较  
Tab 1 Baseline of patients in the training and validation cohorts

Characteristics	Training cohort	Validation cohort	Statistic	$P$ value
Number of patients/ $n$	278	119		
Gender/ $n$				
Male	179	79	$\chi^2=0.072$	0.789
Female	99	40		
Age/year	65 $\pm$ 10	64 $\pm$ 10	$t=0.736$	0.463
Stage/ $n$				
I	30	19	$\chi^2=2.738$	0.603
II	92	34		
III	117	52		
IV	27	10		
Not report	12	4		
Race/ $n$				
Asian	61	22	$\chi^2=1.294$	0.731
Black	7	5		
White	173	76		
Not report	37	16		
Pharmaceutical treatment/ $n$				
Yes	139	49	$\chi^2=3.011$	0.222
No	121	63		
Not report	18	7		
Radiation treatment/ $n$				
Yes	62	18	$\chi^2=2.783$	0.249
No	201	95		
Not report	15	6		
Site/ $n$				
Body of stomach	67	24	$\chi^2=6.418$	0.268
Cardia	75	26		
Fundus of stomach	27	21		
Gastric antrum	99	43		
Lesser curvature of stomach	1	0		
Other	9	5		
Status/ $n$				
Alive	169	67	$\chi^2=0.523$	0.470
Dead	109	52		
Follow up/d	495 (275-810)	398 (275-657)	$W=18.013$	0.160



278例训练集患者随访时间为14~3 519 d, 中位随访时间为495 d, 1、3、5年总体生存率分别为77.5%、49.1%、40.6%。119例测试集患者随访时间为16~3 720 d, 中位随访时间为398 d, 1、3、5年总体生存率分别为73.9%、43.1%、29.7%。2组患者总体生存率比较, 差异无统计学意义 ( $P=0.250$ ), 见图1。



Note: The dotted line is the median.

图1 278例训练集和119例测试集胃癌患者的生存曲线

Fig 1 Survival curves of patients with gastric cancer in the training and validation cohorts

## 2.2 差异表达miRNA筛选结果

生物信息学分析结果显示: 从GSE93415测序数据集中计算得到111个候选差异表达miRNA, 其中20个在癌组织中上调, 91个在癌组织中下调。对111个候选差异表达miRNA进行过滤后, 得到59个候选差异表达miRNA。

## 2.3 预后相关miRNA模型的构建

对59个候选差异表达miRNA和训练集样本数据进行LASSO回归分析, 采用10折交叉验证法进行迭代分析。结果显示, 当变量数为5时, 模型均方根误差最小, 其对应的 $\lambda=0.062$ 。见图2A、B。筛选出的5个miRNA分别为let-7i-5p、let-7f-5p、miR-708-5p、miR-135b-5p、miR-100-5p, 差异表达模式(癌组织对比癌旁组织)均为降低, 差异表达倍数分别2.55、2.78、2.17、3.08、3.26倍。COX回归分析结果显示miR-708-5p为独立的危险因素, miR-135b-5p为独立的保护因素, 见图2C。通过分析这5个miRNA的表达量和胃癌患者的生存数据, 发现胃癌样本中高表达let-7f-5p、hsa-let-7i-5p和hsa-miR-135b-5p以及低表达hsa-miR-100-5p和hsa-miR-708-5p的患者的总体生存率较好, 见图2D。由上述5个miRNA构建的预后表达方程: 风险分数 =  $(-0.049 \times \text{let-7i-5p 表达量} - 0.033 \times \text{let-7f-5p 表达量} + 0.202 \times \text{miR-708-5p 表达量} - 0.088 \times \text{miR-135b-5p 表达量} + 0.016 \times \text{miR-100-5p 表达量})$ 。

利用该预后表达方程对训练集278例样本数据计算模型评分, 以风险评分的中位数为截断值, 将患者分为高风险组(预后相关miRNA模型评分>评分中位数)和低风险组(预后相关miRNA模型评分≤评分中位数), 见图3A、B。训练组患者的5个miRNA表达情况见图3C。对训练集中的高、低风险组患者进行Kaplan-Meier生存分析, 高风险组患者的总体生存率较低, 差异有统计学意义 ( $P=0.021$ ), 见图3D。利用该预后表达方程对训练集278例样本数据进行时间依赖ROC分析, 结果显示, 该预后相关miRNA模型的1年、3年、5年生存时间预测概率的AUC分别为0.640、0.763、0.853, 提示该模型对胃癌患者的预后具有一定的区分预测能力, 见图3E。

## 2.4 预后相关miRNA预测效能检验

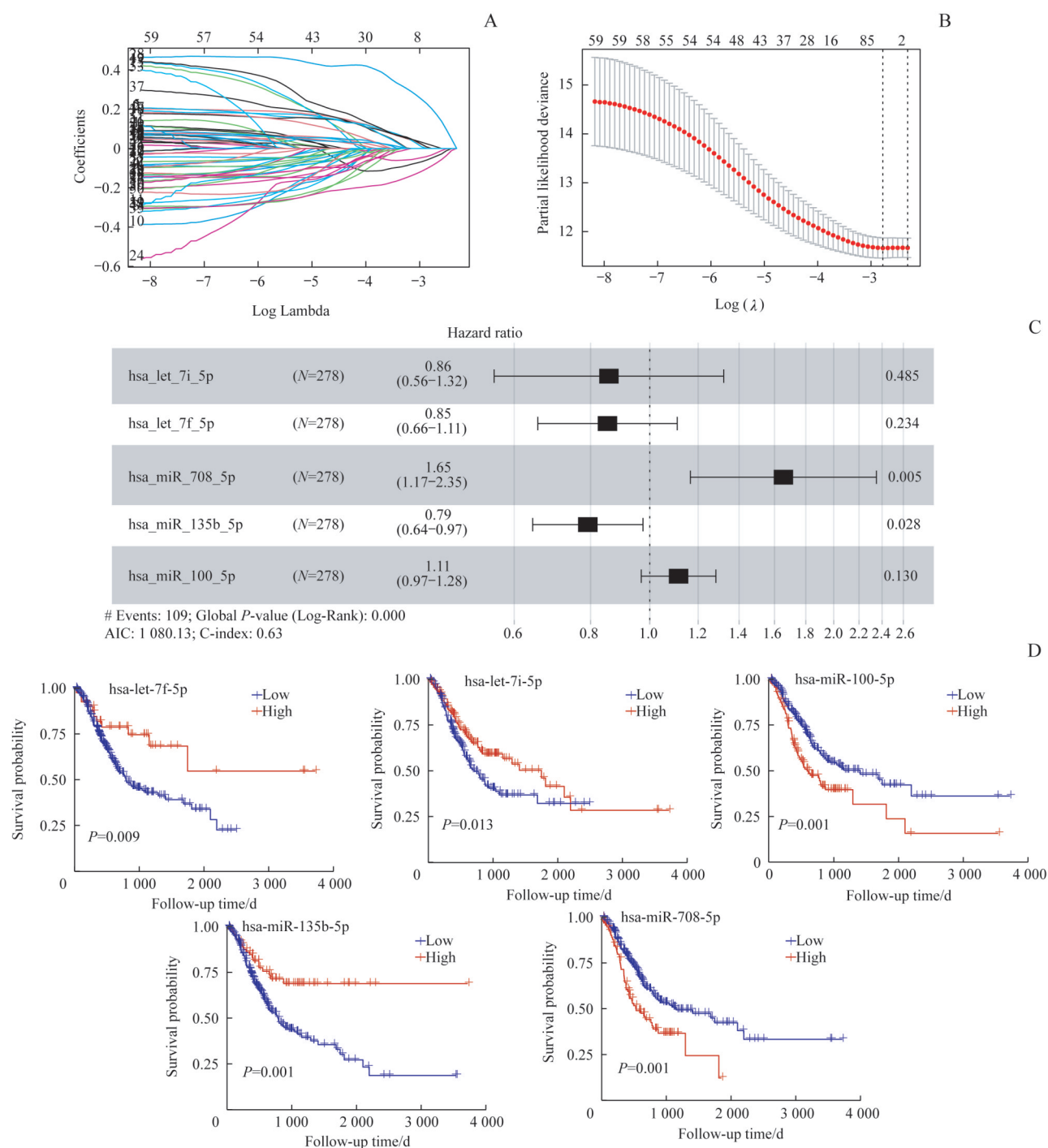
为了进一步检验上述预后相关miRNA模型的预测效能, 利用测试集的119例胃癌样本数据进行Kaplan-Meier生存分析, 结果显示高风险组患者的总体生存率较低, 差异有统计学意义 ( $P=0.016$ )。利用测试集119例胃癌样本数据进行时间依赖ROC分析, 结果显示该预后相关miRNA模型的1年、3年、5年生存时间预测概率AUC分别为0.631、0.735、0.750, 提示该模型对胃癌患者的生存状态和预后具有一定的区分预测能力(图4)。

## 2.5 影响胃癌患者预后的相关因素分析

单因素COX回归分析结果显示: 年龄、肿瘤病理分期、T分期、N分期、M分期和预后相关miRNA模型评分是胃癌患者预后的相关因素 ( $P<0.05$ ), 性别不是胃癌患者预后的相关因素 ( $P>0.05$ ), 见图5A。多因素分析结果显示: 年龄、M分期和预后相关miRNA模型评分是胃癌患者预后的独立危险因素 ( $P<0.05$ ), 见图5B。在356例临床资料完整的胃癌患者中, 预后相关miRNA模型对胃癌患者5年生存时间预测的AUC值为0.818, 高于年龄(AUC=0.619)、性别(AUC=0.500)、肿瘤病理分期(AUC=0.574)、T分期(AUC=0.726)、N分期(AUC=0.602)、M分期(AUC=0.505)的预测效能, 见图5C。

## 2.6 利用预后相关miRNA模型预测胃癌患者对于化疗药物的敏感度

利用pRRophetic包和胃癌患者的基因表达数据计算胃癌患者对于一些常规化疗药物如阿霉素、丝裂霉素C、多西他赛、顺铂和紫杉醇的敏感度。结果显示, 低风险组患者对于顺铂的敏感度明显高于高风险组患者 ( $P=0.030$ ), 见图5D。



**Note:** A. Lasso-Cox regression analysis was performed based on 59 differentially expressed miRNAs and the training cohort. Different colored lines represent different candidate miRNAs. The horizontal axis above represents the number of variables, and the vertical axis represents the coefficient of variables. B. The 10-fold cross validation method was used to determine the modeling position. The partial likelihood deviance tended to be stable with a  $\lambda$  of 0.062, so this point was selected for modeling. The upper horizontal axis represents the number of variables and the vertical axis represents the partial likelihood deviance. C. Cox analysis of 5 selected miRNAs. D. Kaplan-Meier survival curves of 5 selected miRNAs. Patients with high expression of let-7f-5p, hsa-let-7i-5p and hsa-miR-135b-5p and low expression of hsa-miR-100-5p and hsa-miR-708-5p in gastric cancer samples had better overall survival.

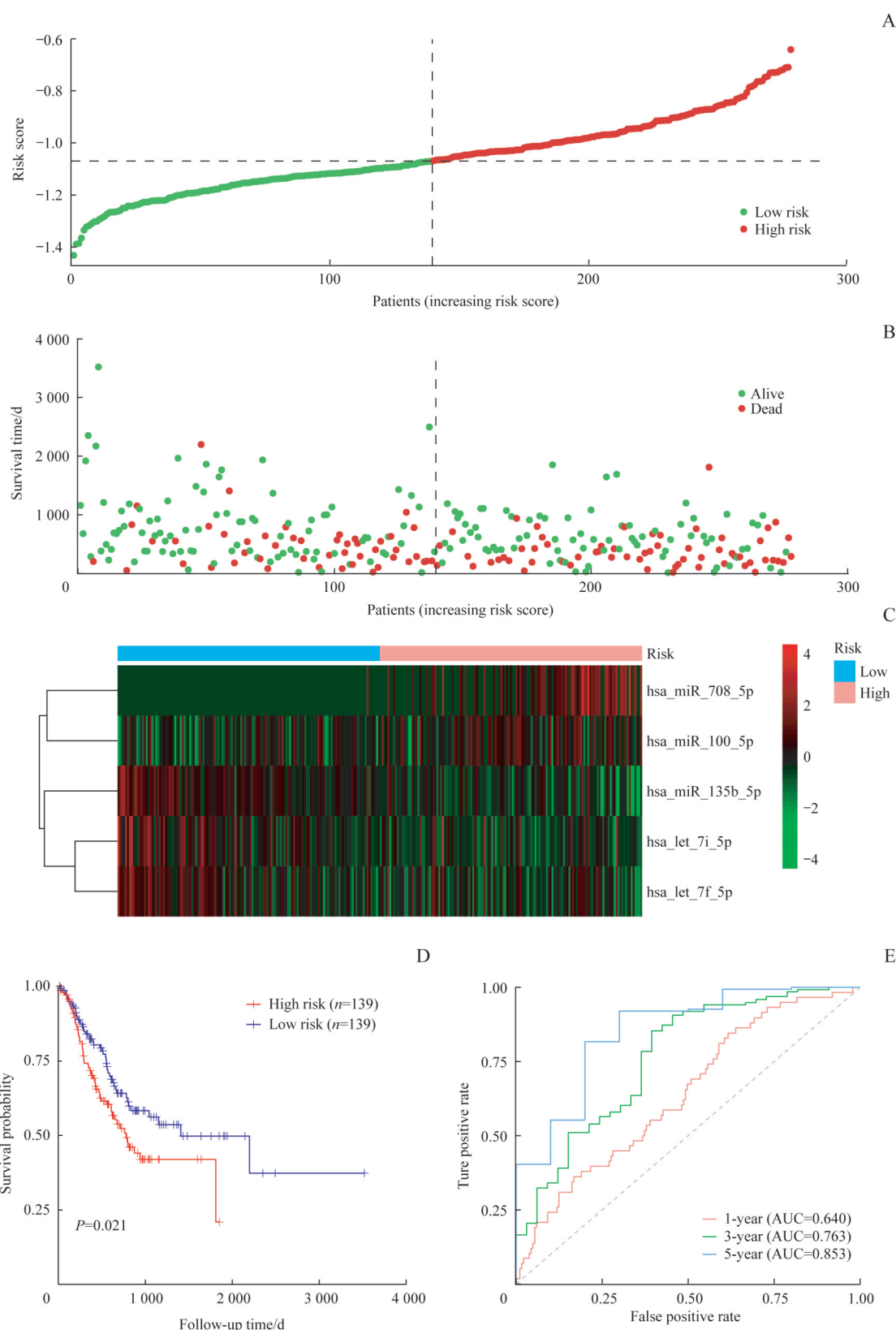
**图2 LASSO-Cox 回归分析筛选 miRNAs**

**Fig 2** miRNAs screened by Lasso-Cox regression analysis

## 2.7 建立并评估与胃癌相关的预后恩诺图

综合预后相关 miRNA 模型评分、年龄、性别和肿瘤病理分期等因素, 建立预后相关恩诺图 (图 6A)。为了评估恩诺图预测胃癌患者生存率的准确性, 采用 calibration

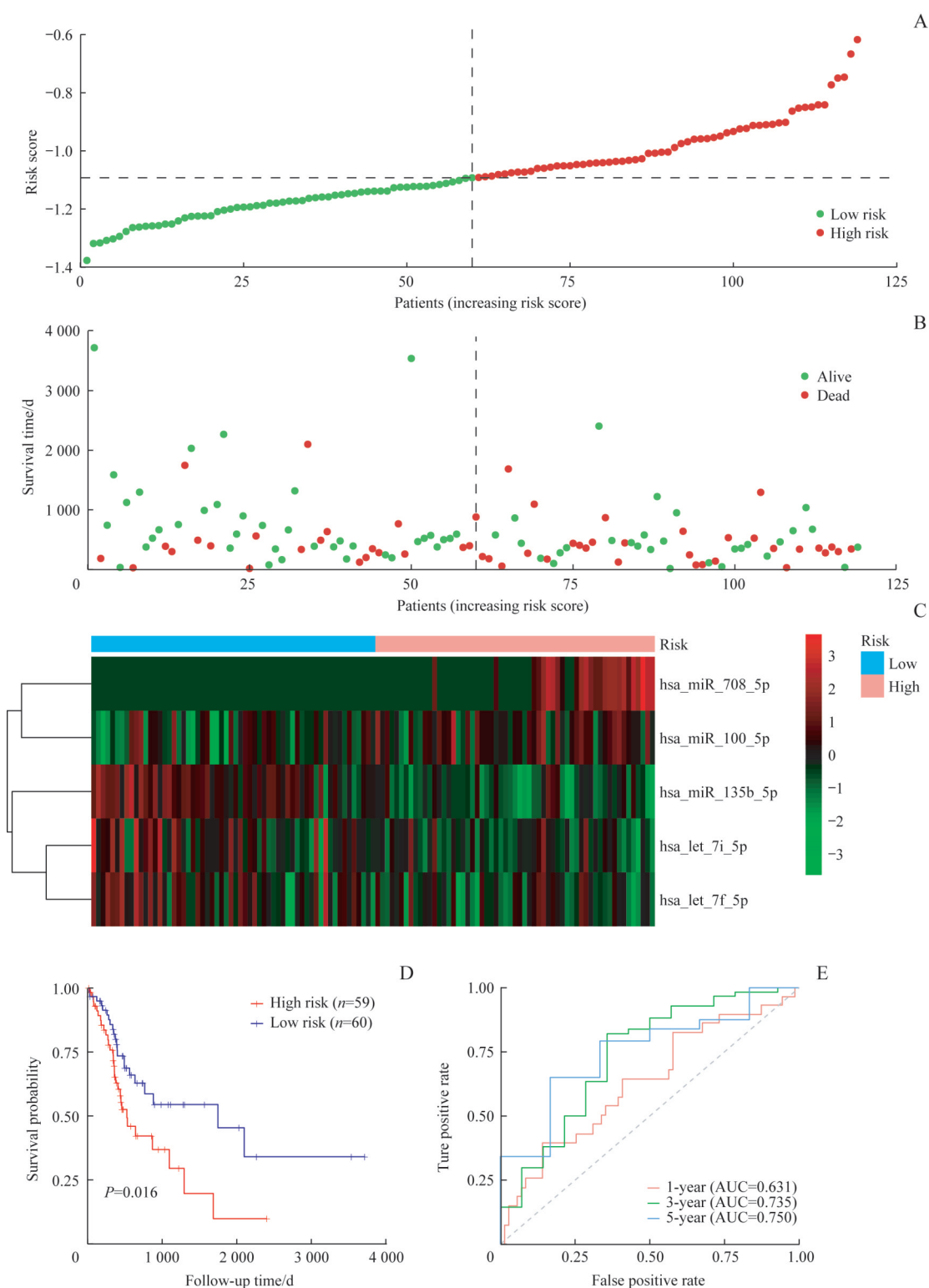
曲线进行验证 (图 6B)。图中 45°虚线为理想曲线, 表明恩诺图所预测的胃癌患者的总体生存率与患者实际的总体生存率一致。由图可知, 恩诺图所预测的胃癌患者的 1 年和 3 年总体生存率与实际总体生存率非常接近。此外,



**Note:** A. The distribution of risk score. B. The survival status and survival time of gastric cancer patients. C. Heatmap of the expression distribution of 5 selected miRNAs. D. Kaplan-Meier survival curves. E. Time-dependent ROC analysis.

图3 训练集预后相关miRNA模型评分、时间依赖ROC分析以及Kaplan-Meier生存分析

Fig 3 Prognostic miRNA model score, time-dependent ROC analysis, and Kaplan-Meier survival analysis of the training cohort



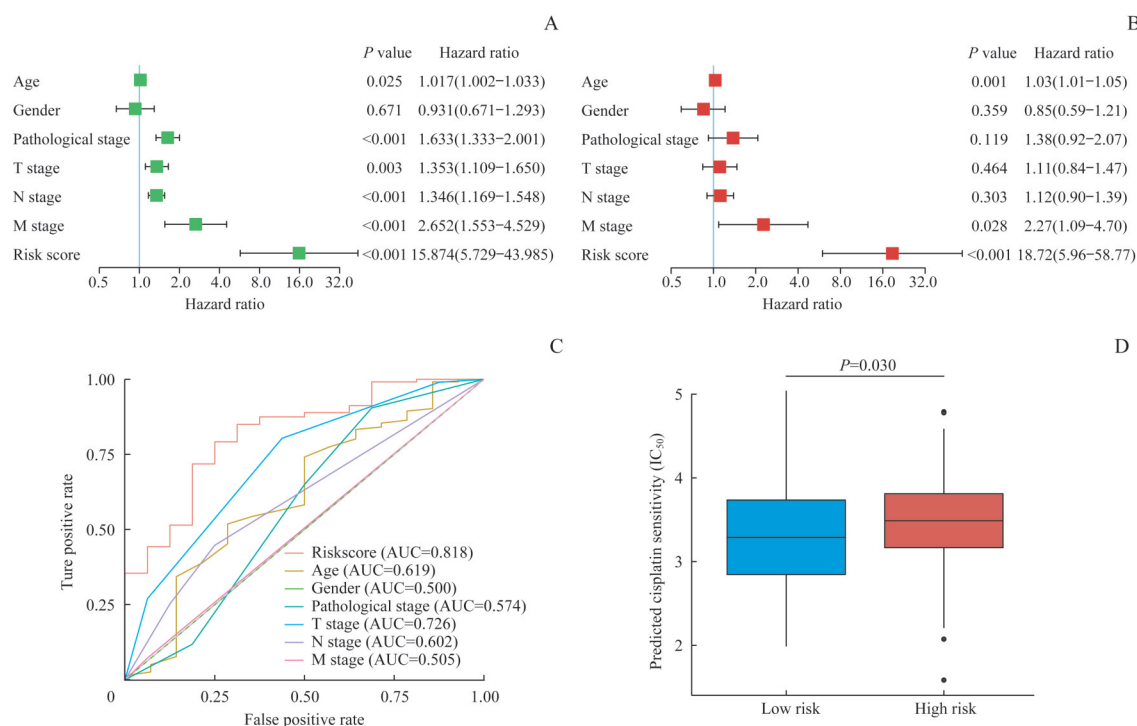
**Note:** A. The distribution of risk score. B. The survival status and survival time of gastric cancer patients. C. Heatmap of the expression distribution of 5 selected miRNAs. D. Kaplan-Meier survival curves. E. Time-dependent ROC analysis.

**图4** 测试集预后相关miRNA模型评分、时间依赖ROC分析以及Kaplan-Meier生存分析

**Fig 4** Prognostic miRNA model score, time-dependent ROC analysis, and Kaplan-Meier survival analysis of the validation cohort

计算恩诺图、患者年龄、性别和肿瘤病理分期的一致性指数 (consistency index, C-index)。通过比较,发现恩诺图的 C-index 最高 (图 6C)。经过决策曲线分析

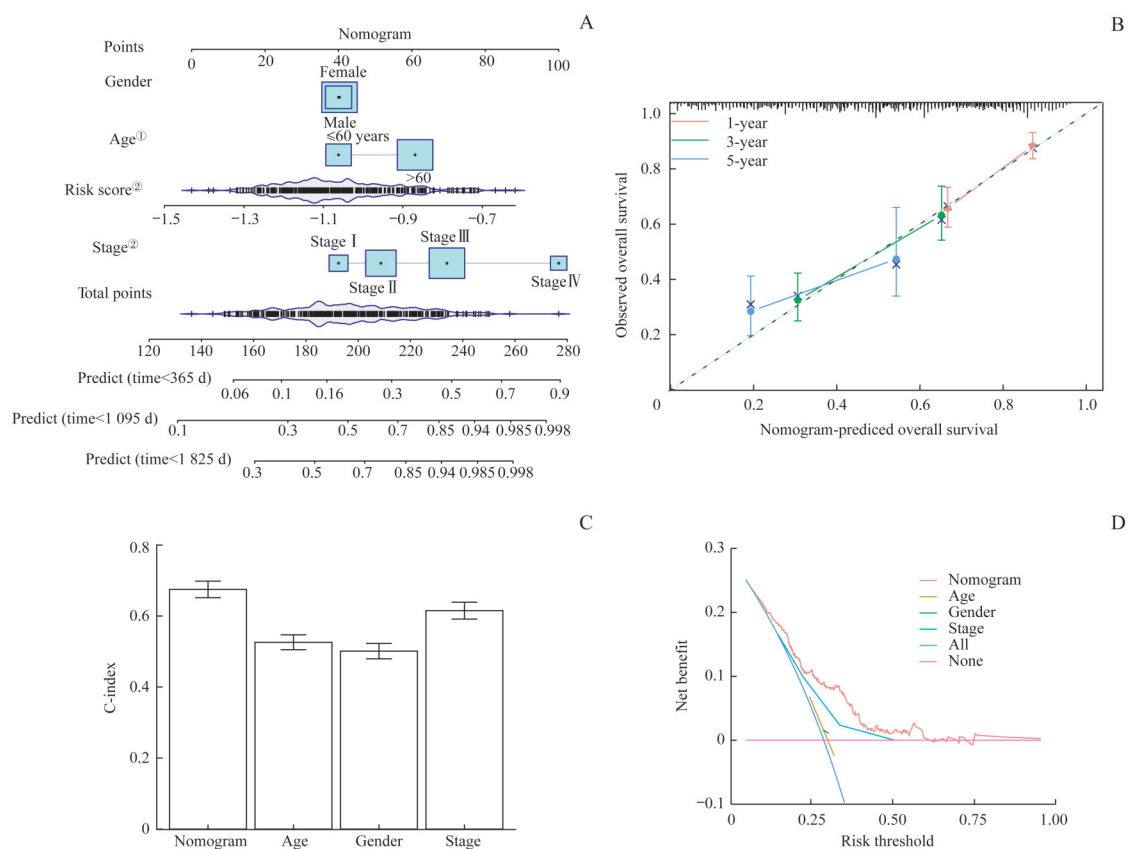
(decision curve analysis, DCA) 发现,与患者的年龄、性别和肿瘤病理分期相比,患者借助于恩诺图的辅助进行临床决策能够受益更多 (图 6D)。



**Note:** A. Univariate Cox regression analysis of prognostic factors in patients with gastric cancer. B. Multivariate Cox regression analysis of prognostic factors in patients with gastric cancer. C. ROC analysis of the risk score and other clinical factors. D. The estimated IC<sub>50</sub> of cisplatin between low- and high-risk groups.

图5 影响胃癌患者预后的相关因素分析

Fig 5 Analysis of prognostic factors in patients with gastric cancer



**Note:** A. Prognostic nomogram for patients with gastric cancer. <sup>①</sup>P<0.01; <sup>②</sup>P<0.001. B. Calibration curves to validate the accuracy of the nomogram. C. C-index of the nomogram and traditional factors. D. DCA curves.

图6 预后相关恩诺图的建立及评估

Fig 6 Construction and evaluation of the prognostic nomogram



### 3 讨论

随着高通量测序技术的兴起,基因组、转录组及蛋白质代谢组等多组学联合分析逐渐成为时下热门的生物研究领域。在转录组学中,miRNA调控网络在疾病的发生、发展中起到重要的作用。miRNA是一类非编码内源性小RNA,可在转录后水平调节靶基因的表达。miRNA表达变化或miRNA功能障碍可能影响肿瘤的发生和发展。Ke等<sup>[16]</sup>研究表明miRNA-16的低表达与肺癌的发生、发展有关,过表达miRNA-16可抑制非小细胞肺癌细胞的侵袭和迁移能力。另有研究<sup>[17-20]</sup>表明,miRNA-125家族与多种肿瘤的发生有密切关系。在结直肠癌、卵巢癌、肺癌和髓母细胞瘤等肿瘤中,miR-125表达较正常组织降低;而在前列腺癌、胰腺癌等肿瘤中,也发现其不同程度的上调。据报道<sup>[21]</sup>,胃癌组织中miRNA表达异常,提示miRNA也参与了胃癌的发生。miR-21通过作用于*PTEN*、*PDCD4*等基因,改变了细胞的增殖与侵袭能力<sup>[22]</sup>;miR-130b通过靶向*BIM*和*RUNX3*,调控细胞的凋亡,介导胃癌的发生<sup>[23]</sup>。这些发现提示,miRNA在胃癌的临床诊断、治疗选择、预后判断及复发预测等方面,有望成为新的分子标志物。

胃癌是一种上皮来源的恶性肿瘤,起源于胃壁最浅表的黏膜上皮细胞,发病率和病死率均较高。然而,胃癌的早期临床特征不明显,与胃溃疡等慢性疾病的症状相似;因此,在疾病早期易被忽略,往往很难得到有效、及时的治疗。目前,临床上广泛应用的胃癌相关的分子标志物有CEA、CA19-9、CA72-4等;然而,该肿瘤标志物的特异性较低,且不具备判断预后的能力。因此,临床上迫切需要开发新的胃癌生物标志物,有助于早期诊断和判断预后,辅助制定个性化的治疗方案。

本研究旨在通过高通量测序结合生物信息学的方法,筛选出与生存相关的miRNA,构建胃癌患者预后相关miRNA预测模型。在本研究中,通过下载分析GEO数据库的GSE93415芯片,筛选出与预后生存显著相关的5个

miRNA,分别为let-7i-5p、let-7f-5p、miR-708-5p、miR-135b-5p、miR-100-5p。虽然这5个miRNA在肿瘤组织中的表达均低于正常水平,但通过生存分析发现,胃癌样本中高表达hsa-let-7f-5p、hsa-let-7i-5p、hsa-miR-135b-5p和低表达hsa-miR-100-5p、hsa-miR-708-5p的患者总体生存状况较好。利用TCGA数据库中的患者临床及生物学信息,最终构建并验证胃癌患者预后相关miRNA预测模型。该模型的效能在测试集中得到了很好的验证,预后相关miRNA模型的1年、3年、5年生存时间预测概率AUC值分别为0.631、0.735、0.750,提示该模型具较好的敏感度和特异度。为了更好地将构建的预后相关miRNA模型应用于临床,本研究综合了预后相关miRNA模型评分以及患者年龄、性别和临床病理分期,建立了一个相对准确且量化的预后相关恩诺图。通过calibration曲线的验证和C-index的比较,发现预后相关恩诺图相对于其他因素具有较高的预测准确度。通过DCA,发现借助于预后相关恩诺图进行临床决策,患者将会受益更多。因此,本研究所构建的预后相关恩诺图具有良好的应用价值。综合以上结果,提示该模型具有一定优越性。并且,相较于TNM分期,该模型可从上游分子生物学的角度,从多维度更精准地判断预后。

本研究构建的胃癌预后预测模型对于胃癌患者的预后判断有良好的预测价值,但仍存在一定的局限性。首先,该模型的训练集及测试集均来自TCGA数据库中的同一队列。若设置独立的外部验证,尤其是中国胃癌患者的验证,可提升该模型的可信度。其次,TCGA作为常用的公共数据库,其患者以白种人为主,缺少亚洲人,尤其是中国人的临床数据。有待在今后的研究中加入中国胃癌患者的生存预后数据,对该模型进行进一步的预测验证。

综上,本研究通过筛选出的5个胃癌相关miRNA,构建了可用于胃癌患者生存预测的miRNA模型。该模型具良好的敏感度和特异度,一定程度上弥补了形态学判断预后的不足,有助于指导临床上制定个体化治疗方案。

### 参·考·文·献

- [1] Smyth EC, Nilsson M, Grabsch HI, et al. Gastric cancer[J]. Lancet, 2020, 396(10251): 635-648.
- [2] Orditura M, Galizia G, Sforza V, et al. Treatment of gastric cancer[J]. World J Gastroenterol, 2014, 20(7): 1635-1649.
- [3] Shin VY, Chu KM. MiRNA as potential biomarkers and therapeutic targets for gastric cancer[J]. World J Gastroenterol, 2014, 20(30): 10432-10439.
- [4] Lu TX, Rothenberg ME. MicroRNA[J]. J Allergy Clin Immunol, 2018, 141(4): 1202-1207.
- [5] Lee YS, Dutta A. MicroRNAs in cancer[J]. Annu Rev Pathol, 2009, 4: 199-227.
- [6] Tutar Y. miRNA and cancer; computational and experimental approaches[J]. Curr Pharm Biotechnol, 2014, 15(5): 429.
- [7] Colaprico A, Silva TC, Olsen C, et al. TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data[J]. Nucleic Acids Res, 2016, 44(8): e71.
- [8] Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets--update[J]. Nucleic Acids Res, 2013, 41(Database issue): D991-995.
- [9] Love MI, Huber W, Anders S. Moderated estimation of fold change and



- dispersion for RNA-seq data with DESeq2[J]. *Genome Biol*, 2014, 15(12): 550.
- [10] Engebretsen S, Bohlin J. Statistical predictions with glmnet[J]. *Clin Epigenetics*, 2019, 11(1): 123.
- [11] Bose E, Maganti S, Bowles KH, et al. Machine learning methods for identifying critical data elements in nursing documentation[J]. *Nurs Res*, 2019, 68(1): 65-72.
- [12] Zhou ZR, Wang WW, Li Y, et al. In-depth mining of clinical data: the construction of clinical prediction model with R[J]. *Ann Transl Med*, 2019, 7(23): 796.
- [13] Kamarudin AN, Cox T, Kolamunnage-Dona R. Time-dependent ROC curve analysis in medical research: current methods and applications[J]. *BMC Med Res Methodol*, 2017, 17(1): 53.
- [14] Geeleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels[J]. *PLoS One*, 2014, 9(9): e107468.
- [15] Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies[J]. *Nucleic Acids Res*, 2015, 43(7): e47.
- [16] Ke Y, Zhao W, Xiong J, et al. Downregulation of miR-16 promotes growth and motility by targeting HDGF in non-small cell lung cancer cells[J]. *FEBS Lett*, 2013, 587(18): 3153-3157.
- [17] Yang M, Tang X, Wang Z, et al. miR-125 inhibits colorectal cancer proliferation and invasion by targeting TAZ[J]. *Biosci Rep*, 2019, 39(12): BSR20190193.
- [18] Hu W, Lei L, Xie X, et al. Heterogeneous nuclear ribonucleoprotein L facilitates recruitment of 53BP1 and BRCA1 at the DNA break sites induced by oxaliplatin in colorectal cancer[J]. *Cell Death Dis*, 2019, 10(8): 550.
- [19] 鲁艳明, 王月, 周梦雅, 等. miR-125b靶基因鉴定及对卵巢癌细胞增殖和凋亡能力影响实验研究[J]. *中华肿瘤防治杂志*, 2018, 25(1): 8-14.
- [20] Masood N, Basharat Z, Khan T, et al. Entangling relation of micro RNA-let7, miRNA-200 and miRNA-125 with various cancers[J]. *Pathol Oncol Res*, 2017, 23(4): 707-715.
- [21] Ssrn MO, Meli LE, Dobru ED. MicroRNA modulation of host immune response and inflammation triggered by *Helicobacter pylori*[J]. *Int J Mol Sci*, 2021, 22(3): 1406.
- [22] Wang Z, Yao W, Li K, et al. Reduction of miR-21 induces SK-N-SH cell apoptosis and inhibits proliferation via PTEN/PDCD4[J]. *Oncol Lett*, 2017, 13(6): 4727-4733.
- [23] Lai KW, Koh KX, Loh M, et al. MicroRNA-130b regulates the tumour suppressor RUNX3 in gastric cancer[J]. *Eur J Cancer*, 2010, 46(8): 1456-1463.

[收稿日期] 2021-07-09

[本文编辑] 吴 洋

## 学术快讯

### 上海交通大学公共卫生学院王慧和李晓光联合团队发现良性心理应激通过SNS/ $\beta$ -ARs/CCL2轴增强抗肿瘤免疫的机制

2021年9月30日, 国际知名杂志*Nature Communications*在线发表了上海交通大学公共卫生学院王慧教授、李晓光研究员课题组题为“Environmental eustress modulates  $\beta$ -ARs/CCL2 axis to induce anti-tumor immunity and sensitize immunotherapy against liver cancer in mice”的论文。该研究发现良性心理应激通过SNS/ $\beta$ -ARs/CCL2信号通路增强抗肿瘤免疫反应, 进而抑制肝癌的发生、发展并克服PD-L1免疫治疗耐药, 为心理情绪因素在癌症预防和控制中作用的理论研究和临床指导提供了新的见解。