



doi:10.7659/j.issn.1005-6947.2020.10.005  
http://dx.doi.org/10.7659/j.issn.1005-6947.2020.10.005  
Chinese Journal of General Surgery, 2020, 29(10):1187-1194.

· 基础研究 ·

## 基于不同数据库来源数据的胃癌长链非编码 RNA 预后预测模型构建

周新童<sup>1,2</sup>, 党胜春<sup>2</sup>

(1. 苏州大学附属张家港医院 普通外科, 江苏 张家港 215600; 2. 江苏大学附属医院 普通外科, 江苏 镇江 212001)

### 摘要

**背景与目的:** 长链非编码 RNA (lncRNA) 对于胃癌患者的预后判断有着非常显著的影响。本研究旨在通过生物信息学的方法, 构建并验证能够准确评估胃癌患者预后的 lncRNA 预后预测模型。

**方法:** 通过癌症基因组图谱 (TCGA) 数据库、基因型-组织表达数据库 (GTEx) 获取数据作为用于构建预后模型 (建模组), 通过基因表达汇编数据库 (GEO) 获取数据用于验证 (验证组)。采用 R 软件中的 edgeR 包筛选差异表达 lncRNA; 通过单因素和多因素 Cox 回归来构建预后模型并计算风险值; 按照风险值的大小将患者分为高、低风险组, 分析风险值与临床病理参数及预后的关系。用验证组样本对建模组的结果进行验证。

**结果:** 共筛选出 288 个差异表达 lncRNA, 其中 28 个与胃癌预后有关 (均  $P < 0.05$ )。10 种 lncRNA 生物标记物 (MEG3、DNAJC9-AS1、ACTA2-AS1、C15orf54、LINC01210、OVAAL、POU6F2-AS2、ERICH3-AS1、LINC00326 及 LINC01526) 被鉴定并用于构建预后模型。高风险组的总体生存率以及无病生存率均低于低风险组 (均  $P < 0.01$ ), ROC 曲线证实该预测模型有一定的准确性 ( $AUC = 0.700$ )。单因素及多因素 Cox 回归分析显示风险值为独立的预后因子 (均  $P < 0.001$ )。风险值与胃癌 T 分期 ( $P = 0.031$ )、肿瘤分化程度 ( $P = 0.044$ ) 存在明显关系。在独立的验证组中, 高风险组的总体生存率以及无病生存率同样明显低于低风险组, 且示风险值依旧为独立的预后因子 (均  $P < 0.05$ )。

**结论:** 所构建的 10-lncRNA 模型对于胃癌患者的预后生存判断有一定的价值, 且筛选出的差异表达 lncRNA 为胃癌分子机制的深入研究提供了依据。

### 关键词

胃肿瘤; RNA, 长链非编码; 预后; 计算生物学

中图分类号: R735.9

## Construction of prognostic prediction model for gastric cancer based on aggregate data from multiple databases

ZHOU Xintong<sup>1,2</sup>, DANG Shengchun<sup>2</sup>

(1. Department of General Surgery, Zhangjiagang Hospital Affiliated to Soochow University, Zhangjiagang, Jiangsu 215600, China;  
2. Department of General Surgery, the Affiliated Hospital, Jiangsu University, Zhenjiang, Jiangsu 212001, China)

### Abstract

**Background and Aims:** Long non-coding RNAs (lncRNAs) exert significant influences on the prognosis of

**基金项目:** 江苏省镇江市重点研发计划-社会发展基金资助项目 (SH2019061)。

**收稿日期:** 2020-03-28; **修订日期:** 2020-10-16。

**作者简介:** 周新童, 苏州大学附属张家港医院住院医师, 主要从事消化道肿瘤的基础和临床方面的研究。

**通信作者:** 党胜春, Email: dscgu@163.com

gastric cancer patients. This study was designated to construct a lncRNA-based prediction model for accurately evaluating the prognosis of gastric cancer patients through bioinformatics approaches.

**Methods:** The data obtained from The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) databases were used for constructing the prognosis model (modeling group), while the data from The Gene Expression Omnibus (GEO) database were used for validation (validation group). The differentially expressed lncRNAs were screened using edgeR package in R software. Univariate and multivariate Cox regression were used to evaluate the association between lncRNA and survival time. prognostic model was created through univariate and multivariate Cox regression analyses and the risk score were calculated. The patients were divided into high-risk group and low-risk group according to their risk scores, and the relations of the risk score with clinicopathologic variables and prognosis were analyzed. The results of the modeling group were verified in the sample from validation group.

**Results:** A total of 288 differentially expressed lncRNAs were screened, and 28 of them were associated with the prognosis of gastric cancer (all  $P < 0.05$ ). Ten lncRNA biomarkers (MEG3, DNAJC9-AS1, ACTA2-AS1, C15orf54, LINC01210, OVAAL, POU6F2-AS2, ERICH3-AS1, LINC00326 and LINC01526) were identified and used to construct a prognostic model. Both overall survival rate and disease-free survival rate in high-risk group were significantly lower than those in low-risk group (both  $P < 0.01$ ). ROC curve confirmed that the prediction model had certain accuracy (AUC=0.700). The results of univariate and multivariate Cox regression analyses showed that the risk score was an independent prognostic factor (both  $P < 0.001$ ). The risk score had significant relation with T stage ( $P = 0.031$ ) and the degree of tumor differentiation ( $P = 0.044$ ). In validation cohort, the overall survival rate and disease-free survival rate in high-risk group were also lower than those in low-risk group, and the risk score remained an independent prognostic factor (all  $P < 0.05$ ).

**Conclusion:** The constructed 10-lncRNA model has certain value in predicting the prognosis of gastric cancer patients, and the screened differentially expressed lncRNAs also provide the basis for further investigating the molecular mechanism of gastric cancer.

#### Key words

Stomach Neoplasms; RNA, Long Noncoding; Prognosis; Computational Biology

**CLC number:** R735.9

胃癌是全球五大恶性肿瘤之一，其病死率在所有癌症中居第2位<sup>[1]</sup>。中国是胃癌的高发区，发病患者数大约占全世界的一半<sup>[2]</sup>。由于其发病隐匿，病情进展较快，5年生存率25%~30%<sup>[3]</sup>。随着近几年诊断及治疗水平的进步，早期患者的生存率已经得到明显提高，但进展期胃癌患者生存率依旧很低<sup>[4-6]</sup>。因此，建立良好的诊断和预后筛选评估体系，对胃癌的诊治尤为重要。

在过去的几十年里，人们在胃癌的分子机制研究方面取得了非常大的突破<sup>[7-9]</sup>。然而，目前尚无用于胃癌治疗及预后评估的分子标志物。长链非编码RNA (long non-coding RNA, lncRNA) 是一类长度超过200个核苷酸的RNA，其本身并不具备编码蛋白质的功能<sup>[10]</sup>。近年来研究显示，lncRNA在基因组中广泛分布，参与调节染色质修饰及基因表达等重要生理过程<sup>[11-13]</sup>。lncRNA在肿

瘤的早期诊断和预后评估方面的价值也逐渐的被挖掘出来<sup>[14-16]</sup>。

近几年来，越来越多的人尝试利用公共数据库，比如从癌症基因组图谱 (The Cancer Genome Atlas, TCGA) 数据库中获得大样本数据进行研究<sup>[17-19]</sup>。然而，单一的TCGA数据库使用存在着一些问题，比如正常对照样本的匮乏 (TCGA中仅有32例胃癌样本存在癌旁对照信息)。此外，笔者发现，基因型-组织表达数据库 (Genotype-Tissue Expression, GTEx) 中保存有大量可作为正常对照的样本信息。

本研究拟将TCGA数据库中胃癌患者和GTEx数据库中正常人胃黏膜的转录组数据联合起来，构建一个由10种lncRNA组成的模型 (10-lncRNA预后模型)，作为一种新的判断胃癌患者预后状态的指标。随后，通过基因表达汇编数据库

(Gene Expression Omnibus, GEO), 引入一个独立的数据集(GSE62254), 验证该预后模型的稳健性<sup>[20]</sup>。

## 1 资料与方法

### 1.1 数据下载

从TCGA官方网站(<https://portal.gdc.cancer.gov/repository>)下载获取胃癌的转录组测序数据集, 其中胃癌患者的样本375例, 对照组样本32例, 同时下载相应的临床资料。在进行预后分析及研究风险值(Risk Score)与临床资料的关系时, 剔除临床信息不全的样本, 共剩余317例。从GTEX官方网站(<https://gtexportal.org>)下载获取正常人的胃黏膜样本转录组测序数据, 共计194例。在筛选差异表达lncRNA时, 将32例对照组样本以及194例正常胃黏膜样本合并, 作为对照组与375例癌组织样本进行差异分析。从GEO官方网站(<https://ncbi.nlm.nih.gov/geo>)下载获取GSE62254数据集, 共计300例, 同获取该数据集的临床样本文件。

### 1.2 数据处理

将下载好的TCGA和GTEX的转录组数据合并, 取与GSE62254共有的lncRNA, 共计1 250个。最终获得的TCGA+GTEX矩阵有1 250行×601列(lncRNA×样本名), 此矩阵将作为建模组用于后续构建预后模型; GEO矩阵有1 250行×300列, 此矩阵将作为验证组用于对模型进行验证。

### 1.3 筛选差异表达 lncRNA

首先通过R软件中的edgeR包对数据进行标准化, 并使用sva包的combat函数对数据进行批次校正, 之后使用经验贝叶斯估计, 计算癌组织和对照组样本的差异显著性。通过Benjamini & Hochberg方法对P值进行校正, 筛选出差异倍数(fold change, FC) >4, 且校正后的P值(false discovery rate, FDR) <0.01的基因。

### 1.4 构建预后模型

将上述筛选出的lncRNA与TCGA下载的生存数据进行合并, 首先采用单因素Cox回归进行生存分析, 筛选出P值<0.05的lncRNA。随后进行多因素Cox回归分析, 分析方法为向前-向后法, 通过赤池信息量准则(Akaike information criterion, AIC), 选出AIC值最小, 即最优化的预后模型<sup>[21-22]</sup>, 并得到模型中各lncRNA的比例系数 $\beta$ , 以此

计算风险值(risk score), 计算公式为: Risk Score =  $\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$ , 其中 $\beta$ 表示各lncRNA相关系数, X表示lncRNA表达量。

根据计算出来的各个样本风险值, 以中位数为界把胃癌患者分为高风险及低风险组, 并且用R软件中的pheatmap包将风险值可视化。使用survivalROC包绘制时间依赖性ROC曲线并计算曲线下面积(area under curve, AUC)。

### 1.5 统计学处理

本研究所使用的数据处理统计软件为R软件(版本3.4.0)和SPSS(版本19.0), 各项分析均为双侧检验, 所有统计检验结果的P<0.05被认为具有统计学意义。通过单因素及多因素Cox回归分析, 筛选与胃癌预后独立相关的危险因素。预后风险值以中位数(四分位数间距)[M(IQR)]表示, 其与临床的相关性分析使用Wilcoxon秩和检验进行两组间比较, 多组间比较选用Kruskal-Wallis检验。生存分析结果通过Kaplan-Meier曲线来展示, 差异显著性采用Log-rank法检验。

## 2 结果

### 2.1 胃癌 lncRNA 表达矩阵的建立、差异分析

合并后的实验组(TCGA+GTEX)矩阵进行差异分析, 共筛选出288个差异表达lncRNA(FDR<0.01, |Log<sub>2</sub>FC|>2), 其中236个上调表达, 52个下调表达, 绘制火山图, 见图1。

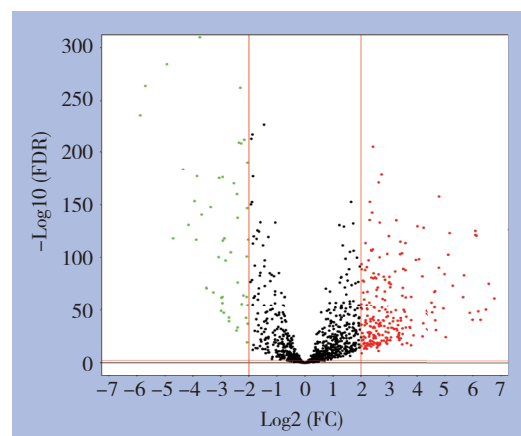


图1 建模组(TCGA+GTEX)矩阵中的差异表达lncRNA(红色代表上调lncRNA, 绿色代表下调lncRNA)

Figure 1 The differentially expressed lncRNAs in modeling cohort (TCGA+GTEX) matrix (the red nodes presenting the up-regulated lncRNAs, and the green nodes indicating down-regulated lncRNAs)

### 2.2 预后模型的构建

首先对上述的lncRNA进一步筛选，使用单因素Cox回归分析，确定与胃癌患者的总体生存相关的28个lncRNA（均 $P < 0.05$ ），用于预后模型的构建。通过多因素Cox回归分析来构建模型，当纳入的lncRNA个数为10个时，模型具有最小的AIC值：1 451.36，此时模型的拟合程度最优。这10个lncRNA分别是MEG3（ $P=0.039$ ）、DNAJC9-AS1（ $P=0.01$ ）、ACTA2-AS1（ $P=0.122$ ）、C15orf54（ $P=0.024$ ）、LINC01210（ $P < 0.01$ ）、OVAAL（ $P=0.078$ ）、POU6F2-AS2（ $P=0.07$ ）、ERICH3-AS1（ $P < 0.001$ ）、LINC00326（ $P=0.084$ ）、LINC01526（ $P=0.078$ ）。最终预后模型：风险值=（MEG3表达值 $\times 0.174$ ）+（DNAJC9-AS1表达值 $\times (-0.212)$ ）+（ACTA2-AS1表达值 $\times 0.112$ ）+（C15orf54表达值 $\times 0.140$ ）+（LINC01210表达值 $\times (-0.178)$ ）+（OVAAL表达值 $\times 0.104$ ）+（POU6F2-AS2表达值 $\times 0.076$ ）+（ERICH3-AS1表达值 $\times 0.269$ ）+（LINC00326表达值 $\times 0.098$ ）+（LINC01526表达值 $\times 0.162$ ）。

在下文验证组（GSE62254）中同样依据此公式计算风险值。

### 2.3 预后模型的风险值评估

根据公式所计算出预后风险值的中位数，将样本分为高及低风险组，并对结果进行可视化展示（图2），同时，绘制高及低风险组的Kaplan-Meier曲线，并进行Log-rank检验，分析显示，高风险组的总体生存率明显低于低风险组，高及低风险组5年总体生存率差异存在统计学意义[（ $18.6 \pm 6.2$ ）% vs.（ $57.6 \pm 6.4$ ）%， $P < 0.001$ ]（图3A）。另外，高风险组无病生存率明显低于低风险组，高及低风险组5年无病生存率差异有统计学意义[（ $34.0 \pm 8.8$ ）% vs.（ $61.4 \pm 6.6$ ）%， $P < 0.01$ ]（图3B）。绘制时间依赖性ROC曲线（图4），结果显示，该预后模型的AUC（0.700）明显大于性别（0.542）、年龄（0.586）、T分期（0.564）、N分期（0.576）、M分期（0.532）、TNM分期（0.606）、分化程度（0.563）等指标，提示该预后模型对于评估胃癌患者的预后有一定准确性。

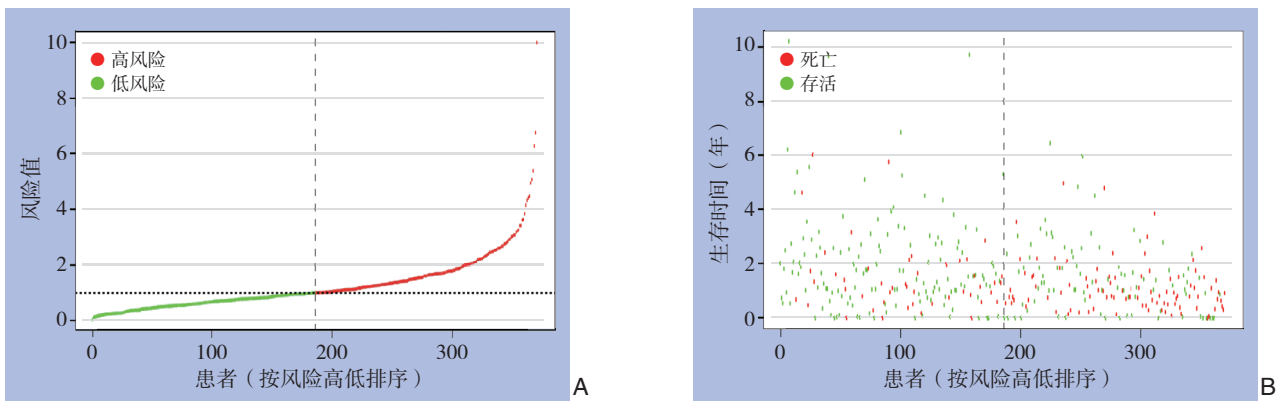


图 2 胃癌患者的预后评估 A: 风险值分布; B: 总体生存时间、生存状态

Figure 2 Prognosis evaluation of gastric cancer patients A: Distribution of risk score; B: Overall survival status and duration

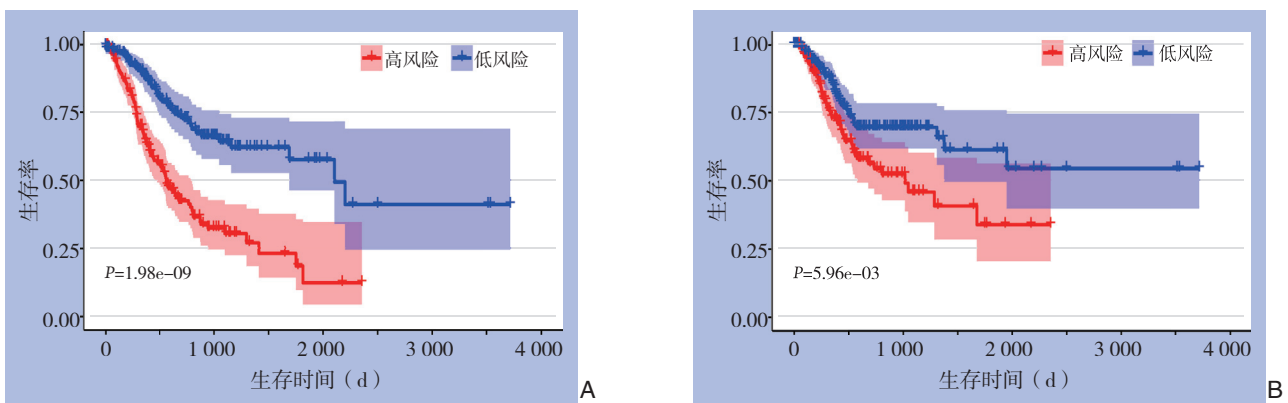


图 3 建模组中高及低风险患者 Kaplan-Meier 曲线 A: 总体生存曲线; B: 无病生存率的曲线

Figure 3 Kaplan-Meier curves for patients in high-risk and low-risk groups of modeling cohort A: Overall survival curves; B: Disease-free survival curves

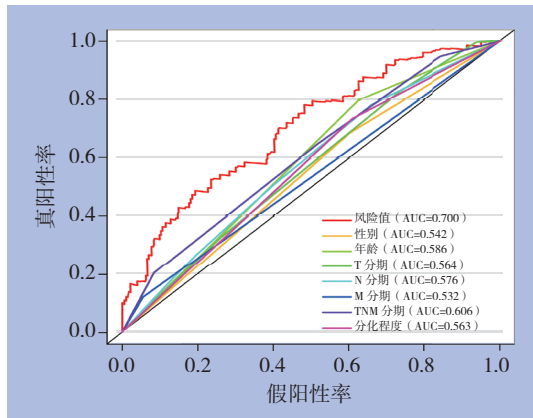


图 4 各种临床指标以及预后风险值的时间依赖性 ROC 曲线以及 AUC 值

Figure 4 Time-dependent ROC curves and AUC values for various clinical characteristics and risk score

### 2.4 筛选影响胃癌生存预后的独立危险因素

通过单因素Cox回归分析寻找与胃癌患者生存预后有关的因素, 并绘制森林图, 结果显示年龄较高 (>60岁)、肿瘤T分期较晚、N分期较晚、TNM分期较晚、风险值较高的患者预后较差 (均  $P < 0.05$ ) (图5A)。随后进一步将上述阳性指标纳入多因素Cox回归分析, 结果显示年龄和风险值是胃癌患者的独立危险因素 (均  $P < 0.001$ ) (图5B)。

### 2.5 预后模型与临床因素的关系

分析显示, 预后风险值与胃癌患者的T分期、肿瘤分化程度明显有关 (均  $P < 0.05$ ), 而与性别、年龄、N分期、M分期、TNM分期之间无明显关系 (均  $P > 0.05$ ) (表1)。

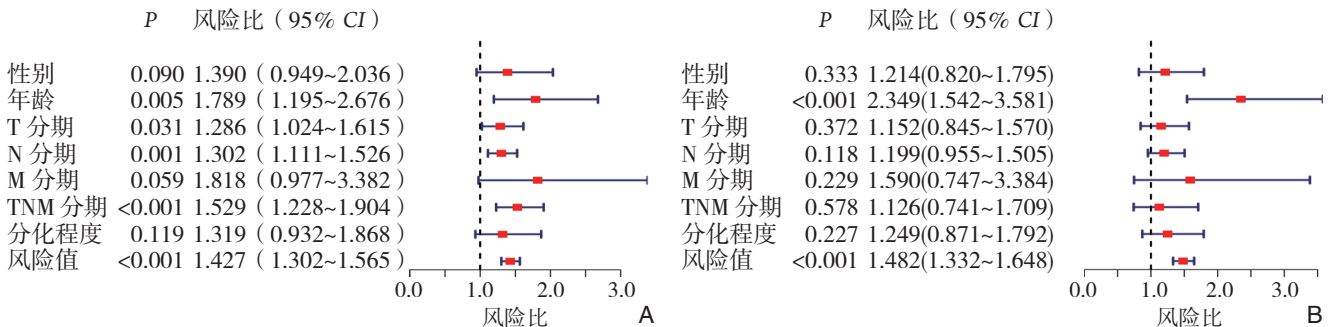


图 5 各临床指标以及预后风险值对总体生存时间影响的森林图 A: 单因素 Cox 回归分析; B: 多因素 Cox 回归分析  
Figure 5 Forest plots of the impact of clinical characteristics and risk score on overall survival time A: Univariate Cox regression analysis; B: Multivariate Cox regression analysis

表 1 胃癌的临床病理特征与风险值的关系

Table 1 Relations of clinicopathologic features with risk score in gastric cancer

因素	例数 (n)	风险值 [M (IQR)]	P	因素	例数 (n)	风险值 [M (IQR)]	P
性别				M 分期			
男	197	1.01 (0.66~1.65)	0.071	M <sub>0</sub>	295	0.97 (0.59~1.52)	0.074
女	120	0.95 (0.51~1.41)		M <sub>1</sub>	22	1.42 (0.72~1.81)	
年龄 (岁)				TNM 分期			
≤ 60	107	1.02 (0.67~1.55)	0.329	I	42	0.85 (0.39~1.23)	
> 60	210	0.96 (0.55~1.62)		II	101	0.95 (0.68~1.45)	
T 分期				III	139	1.02 (0.56~1.62)	
T <sub>1</sub>	15	0.66 (0.25~0.99)		IV	35	1.18 (0.75~1.73)	
T <sub>2</sub>	63	0.96 (0.63~1.79)	0.031	分化程度			0.044
T <sub>3</sub>	152	0.97 (0.61~1.51)		高、中分化	115	0.90 (0.56~1.42)	
T <sub>4</sub>	87	1.07 (0.61~1.64)		低分化	202	1.04 (0.61~1.67)	
N 分期							
N <sub>0</sub>	99	0.96 (0.61~1.49)	0.300				
N <sub>1</sub> ~N <sub>3</sub>	218	1.00 (0.61~1.62)					

### 2.6 预后模型的验证

按照公式计算验证组 (GSE62254) 中的风险值, 并与其临床资料合并, 绘制Kaplan-Meier 曲线, 分析显示, 高风险组的总体生存率明显低

于低风险组, 高及低风险组5年总体生存率存在明显差异 [(45.3 ± 4.0)% vs. (59.2 ± 4.0)%,  $P < 0.01$ ] (图6A)。另外, 高风险组无病生存率显著低于低风险组, 高及低风险组5年无病生存率存

在明显差异[ (42.2 ± 4.4) % vs. (60.1 ± 4.5) % ,  $P < 0.01$ ] (图6B)。随后进行的单因素 ( $P < 0.01$ ) 和多因素 ( $P = 0.019$ ) Cox回归分析表明, 风险值

仍为独立预后因素。验证结果提示, 该预后模型在不同环境下均具有良好的预测效能。

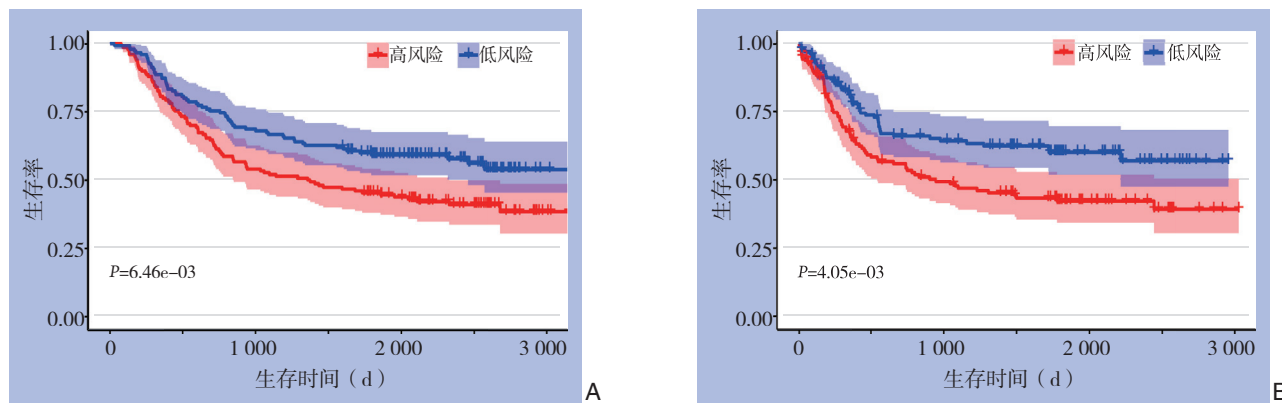


图6 验证组 (GSE62254) 中高及低风险组患者Kaplan-Meier曲线 A: 总体生存曲线; B: 无病生存曲线

Figure 6 Kaplan-Meier curves for patients in high- and low-risk groups of the validation cohort (GSE62254) A: Overall survival curves; B: Disease-free survival curves

### 3 讨论

过去, 人们一直认为lncRNA是转录过程中的废弃产物。随着研究的深入, lncRNA已经被发现参与许多基本的生物学过程, 例如调节细胞周期、细胞凋亡和DNA损伤修复<sup>[23-24]</sup>。为了探索可用于胃癌的预后风险判断的lncRNA, 我们通过挖掘高通量测序数据, 构建出能够用于评估患者生存状态的10-lncRNA预后模型。

通过使用10-lncRNA预后模型, 可以观察到高及低风险患者的生存曲线存在着明显的分离。与高风险评分患者相比, 低风险评分患者的生存时间显著延长。单因素和多因素Cox回归分析显示, 10-lncRNA预后模型与疾病预后独立相关。风险值与T分期及肿瘤分化程度有关, 而与N分期及M分期无关, 提示10-lncRNA预后模型能够用于判断患者预后的具体机制可能涉及肿瘤的生长及分化, 而非迁移及侵袭等生理学过程。

关于这10种lncRNA的特征, 其中5个 (DNAJC9-AS1、C15orf54、ERICH3-AS1、LINC00326及LINC01526) 迄今为止还没有相关的研究报道, 笔者首先报道其表达水平与预后之间关系的研究。另外5个lncRNA, 其中母体表达基因3 (maternally expressed gene 3, MEG3) 在不同的癌细胞 (例如乳腺癌、肝癌、结直肠癌及胃癌) 中通过调节主要的抑癌基因P53和Rb来发挥

抗肿瘤的作用<sup>[25]</sup>。而ACTA2-AS1的一个转录本的低表达显著促进了肝癌细胞的增殖、细胞周期进程、迁移和侵袭<sup>[26]</sup>。在卵巢癌中, LINC01210的较高表达与卵巢癌患者较差的总体生存和无病生存相关<sup>[27]</sup>。在大肠癌和黑色素瘤中, OVAAL与丝氨酸/苏氨酸蛋白激酶3 (STK3) 的相互作用增强了STK3与RAF-1之间的结合, 最终导致RAF/MEK/ERK通路的激活, 从而促进了癌细胞的增殖和存活<sup>[28]</sup>。最后一个lncRNA, POU6F2-AS2参与了电离辐射后食管癌细胞DNA损伤修复并调节细胞的存活<sup>[29]</sup>, 且在结肠癌中具有促进肿瘤增殖和耐药的作用<sup>[30]</sup>。

目前, 对于胃癌患者来说, 仍然还未有能够有效判断预后的工具。如本研究所示, 使用较少量的lncRNA (10个) 便可以预测胃癌患者的预后, 这为临床医生提供了宝贵而可行的参考。当然本研究还存在局限性, 由于高通量测序数据具有一定的误差及背景噪音<sup>[31]</sup>, 本研究虽然在分析前已对数据进行标准化及批次校正, 且通过独立验证组初步验证了模型的稳健性, 但结果仍需临床和基础实验来进一步的研究证实。

总之, 本研究构建了与胃癌患者生存相关的10-lncRNA预后模型, 并对模型的预测效能进行了验证。未来需要更多细胞、动物功能学实验来探索这些lncRNA的作用。

## 参考文献

- [1] den Hoed CM, Kuipers EJ. Gastric Cancer: How Can We Reduce the Incidence of this Disease[J]. *Curr Gastroenterol Rep*, 2016, 18(7):28–34. doi:10.1007/s11894-016-0506-0.
- [2] Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015[J]. *CA Cancer J Clin*, 2016, 66(2):115–132. doi:10.3322/caac.21338.
- [3] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016[J]. *CA Cancer J Clin*, 2016, 66(1):7–30. doi:10.3322/caac.21332.
- [4] Selim JH, Shaheen S, Sheu WC, et al. Targeted and novel therapy in advanced gastric cancer[J]. *Exp Hematol Oncol*, 2019, 8:25. doi:10.1186/s40164-019-0149-6.
- [5] Song Z, Wu Y, Yang J, et al. Progress in the treatment of advanced gastric cancer[J]. *Tumour Biol*, 2017, 39(7):101–116. doi:10.1177/1010428317714626.
- [6] Shen L, Shan YS, Hu HM, et al. Management of gastric cancer in Asia: resource-stratified guidelines[J]. *Lancet Oncol*, 2013, 14(12):e535–547. doi:10.1016/S1470-2045(13)70436-4.
- [7] Carlomagno N, Incollingo P, Tammaro V, et al. Diagnostic, Predictive, Prognostic, and Therapeutic Molecular Biomarkers in Third Millennium: A Breakthrough in Gastric Cancer[J]. *Biomed Res Int*, 2017, 2017:7869802. doi:10.1155/2017/7869802.
- [8] Ye DM, Xu G, Ma W, et al. Significant function and research progress of biomarkers in gastric cancer[J]. *Oncol Lett*, 2020, 19(1):17–29. doi:10.3892/ol.2019.11078.
- [9] Ghafouri-Fard S, Vafae R, Shoorei H, et al. MicroRNAs in gastric cancer: Biomarkers and therapeutic targets[J]. *Gene*, 2020, 757:144937. doi:10.1016/j.gene.2020.144937.
- [10] Tam C, Wong JH, Tsui SKW, et al. LncRNAs with miRNAs in regulation of gastric, liver, and colorectal cancers: updates in recent years[J]. *Appl Microbiol Biotechnol*, 2019, 103(12):4649–4677. doi:10.1007/s00253-019-09837-5.
- [11] Cao D, Ding Q, Yu W, et al. Long noncoding RNA SPRY4-IT1 promotes malignant development of colorectal cancer by targeting epithelial-mesenchymal transition[J]. *OncoTargets Ther*, 2016, 9:5417–5425. doi:10.2147/OTT.S111794.
- [12] Tariq A, Hao Q, Sun Q, et al. LncRNA-mediated regulation of SOX9 expression in basal subtype breast cancer cells[J]. *RNA*, 2020, 26(2):175–185. doi:10.1261/rna.073254.119.
- [13] Deniz E, Erman B. Long noncoding RNA(lincRNA), a new paradigm in gene expression control[J]. *Funct Integr Genomics*, 2017, 17(2–3):135–143. doi:10.1007/s10142-016-0524-x.
- [14] Fattahi S, Kosari-Monfared M, Golpour M, et al. LncRNAs as potential diagnostic and prognostic biomarkers in gastric cancer: A novel approach to personalized medicine[J]. *J Cell Physiol*, 2020, 235(4):3189–3206. doi:10.1002/jcp.29260.
- [15] 邓浩, 刘磊. 胃癌细胞中长链非编码RNA CCAT2的表达及其作用[J]. *中国普通外科杂志*, 2018, 27(04):435–441. doi:10.3978/j.issn.1005-6947.2018.04.008.
- Deng H, Liu L. Expression of long non-coding RNA CCAT2 in gastric cancer cells and its action[J]. *Chinese Journal of General Surgery*, 2018, 27(04):435–441. doi:10.3978/j.issn.1005-6947.2018.04.008.
- [16] Huang Y, Zhang J, Hou L, et al. LncRNA AK023391 promotes tumorigenesis and invasion of gastric cancer through activation of the PI3K/Akt signaling pathway[J]. *J Exp Clin Cancer Res*, 2017, 36(1):194. doi:10.1186/s13046-017-0666-2.
- [17] Cheng C, Wang Q, Zhu M, et al. Integrated analysis reveals potential long non-coding RNA biomarkers and their potential biological functions for disease free survival in gastric cancer patients[J]. *Cancer Cell Int*, 2019, 19:123. doi:10.1186/s12935-019-0846-6.
- [18] 陈圣, 赵继森, 李靖华, 等. 肝细胞癌自噬相关长链非编码RNA预后模型的建立与分析[J]. *中国普通外科杂志*, 2020, 29(7):839–848. doi:10.7659/j.issn.1005-6947.2020.07.008.
- Chen S, Zhao JS, Li JH, et al. Construction and analysis of prognostic model for hepatocellular carcinoma based on autophagy-related long non-coding RNAs[J]. *Chinese Journal of General Surgery*, 2020, 29(7):839–848. doi:10.7659/j.issn.1005-6947.2020.07.008.
- [19] Zhang Y, Li H, Zhang W, et al. LASSO-based Cox-PH model identifies an 11-lncRNA signature for prognosis prediction in gastric cancer[J]. *Mol Med Rep*, 2018, 18(6):5579–5593. doi:10.3892/mmr.2018.9567.
- [20] 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. doi:10.1093/nar/gks1193.
- [21] Liang C, Qi Z, Ge H, et al. Long non-coding RNA PCAT-1 in human cancers: A meta-analysis[J]. *Clin Chim Acta*, 2018, 480:47–55. doi:10.1016/j.cca.2018.01.043.
- [22] Ueno S, Tanabe G, Sako K, et al. Discrimination value of the new western prognostic system (CLIP score) for hepatocellular carcinoma in 662 Japanese patients. *Cancer of the Liver Italian Program*[J]. *Hepatology*, 2001, 34(3):529–534. doi:10.1053/jhep.2001.27219.
- [23] Gibb EA, Brown CJ, Lam WL. The functional role of long non-coding RNA in human carcinomas[J]. *Mol Cancer*, 2011, 10:38. doi:10.1186/1476-4598-10-38.
- [24] Gao H, Yin Y, Qian A, et al. LncRNA LINC00974 Upregulates

- CDK6 to Promote Cell Cycle Progression in Gastric Carcinoma[J]. Cancer Biother Radiopharm, 2019, 34(10): 666–670. doi:10.1089/cbr.2019.2904.
- [25] Al-Rugeebah A, Alanazi M, Parine NR. MEG3: An Oncogenic Long Non-coding RNA in Different Cancers[J]. Pathol Oncol Res, 2019, 25(3):859–874. doi:10.1007/s12253-019-00614-3.
- [26] Zhou RJ, Lv HZ. Knockdown of ACTA2AS1 promotes liver cancer cell proliferation, migration and invasion[J]. Mol Med Rep, 2019, 19(3):2263–2370. doi:10.3892/mmr.2019.9856.
- [27] Zhang C, Liu J, Zhang Y, et al. LINC01210 accelerates proliferation, invasion and migration in ovarian cancer through epigenetically downregulating KLF4[J]. Biomed Pharmacother, 2019, 119:109431. doi:10.1016/j.biopha.2019.109431.
- [28] Sang B, Zhang YY, Guo ST, et al. Dual functions for OVAAL in initiation of RAF/MEK/ERK pro-survival signals and evasion of p27-mediated cellular senescence[J]. Proc Natl Acad Sci U S A, 2018, 115(50):E11661–11670. doi:10.1073/pnas.1805950115.
- [29] Liu J, Sun X, Zhu H, et al. Long noncoding RNA POU6F2-AS2 is associated with oesophageal squamous cell carcinoma[J]. J Biochem, 2016, 160(4):195–204. doi:10.1093/jb/mvw025.
- [30] Xu G, Zhu H, Xu J, et al. Long non-coding RNA POU6F2-AS2 promotes cell proliferation and drug resistance in colon cancer by regulating miR-377/BRD4[J]. J Cell Mol Med, 2020, 24(7):4136–4149. doi:10.1111/jcmm.15070.
- [31] Laehnemann D, Borkhardt A, McHardy AC. Denoising DNA deep sequencing data-high-throughput sequencing errors and their correction[J]. Brief Bioinform, 2016, 17(1):154–179. doi:10.1093/bib/bbv029.

( 本文编辑 宋涛 )

**本文引用格式:** 周新童, 党胜春. 基于不同数据库来源数据的胃癌长链非编码RNA预后预测模型构建[J]. 中国普通外科杂志, 2020, 29(10):1187–1194. doi:10.7659/j.issn.1005-6947.2020.10.005

**Cite this article as:** Zhou XT, Dang SC. Construction of prognostic prediction model for gastric cancer based on aggregate data from multiple databases[J]. Chin J Gen Surg, 2020, 29(10):1187–1194. doi:10.7659/j.issn.1005-6947.2020.10.005

## 本刊对来稿中统计学处理的有关要求

1. 统计研究设计: 应交代统计研究设计的名称和主要做法。如调查设计(分为前瞻性、回顾性或横断面调查研究); 实验设计(应交代具体的设计类型, 如自身配对设计、成组设计、交叉设计、正交设计等); 临床试验设计(应交代属于第几期临床试验, 采用了何种盲法措施等)。主要做法应围绕4个基本原则(随机、对照、重复、均衡)概要说明, 尤其要交代如何控制重要非试验因素的干扰和影响。

2. 资料的表达与描述: 用  $\bar{x} \pm s$  表达近似服从正态分布的定量资料, 用  $M(QR)$  表达呈偏态分布的定量资料; 用统计表时, 要合理安排纵横标目, 并将数据的含义表达清楚; 用统计图时, 所用统计图的类型应与资料性质相匹配, 并使数轴上刻度值的标法符合数学原则; 用相对数时, 分母不宜小于20, 要注意区分百分率与百分比。

3. 统计分析方法的选择: 对于定量资料, 应根据所采用的设计类型、资料所具备的条件和分析目的, 选用合适的统计分析方法, 不应盲目套用  $t$  检验和单因素方差分析; 对于定性资料, 应根据所采用的设计类型、定性变量的性质和频数所具备条件以分析目的, 选用合适的统计分析方法, 不应盲目套用  $\chi^2$  检验。对于回归分析, 应结合专业知识和散布图, 选用合适的回归类型, 不应盲目套用简单直线回归分析, 对具有重复实验数据的回归分析资料, 不应简单化处理; 对于多因素、多指标资料, 要在一元分析的基础上, 尽可能运用多元统计分析方法, 以便对因素之间的交互作用和多指标之间的内在联系进行全面、合理的解释和评价。

4. 统计结果的解释和表达: 当  $P < 0.05$  (或  $P < 0.01$ ) 时, 应说明对比组之间的差异有统计学意义, 而不应说对比组之间具有显著性(或非常显著性)的差别; 应写明所用统计分析方法的具体名称(如: 成组设计资料的  $t$  检验、两因素析因设计资料的方差分析、多个均数之间两两比较的  $q$  检验等), 统计量的具体值(如  $t=3.45$ ,  $\chi^2=4.68$ ,  $F=6.79$  等)应可能给出具体的  $P$  值(如  $P=0.0238$ ); 当涉及到总体参数(如总体均数、总体率等)时, 在给出显著性检验结果的同时, 再给出95%置信区间。

中国普通外科杂志编辑部