

·生物信息学·

## 基于生物信息学分析筛选肺腺癌免疫预后相关 lncRNA

思璐,秦秋雯,沈印,罗双艳,何颖,张裕旭,刘柳

(广西壮族自治区人民医院国际医疗部,广西 南宁 530021)

**摘要:**目的 基于 TCGA 和 Molecular Signatures Database 数据,利用生物信息学方法鉴定潜在的免疫预后相关的 lncRNA(IRL)生物标志物,用于肺腺癌患者的预后预测。**方法** 从 TCGA 获取肺腺癌 IRL 数据集和相应的临床参数;从 Molecular Signatures Database v7.0 中提取免疫基因。通过构建免疫基因共表达网络鉴定免疫相关的 IRL。随后对 IRL 进行 Cox 回归分析、Lasso 回归分析筛选肺腺癌预后相关的 IRL 生物标志物。通过生存分析、ROC 曲线分析评估 IRL 预测患者生存预后的价值。**结果** 共鉴定了 1124 个 IRL,通过单因素 Cox 回归分析,初步筛选出 18 个和肺腺癌预后相关的 IRL,且通过 Lasso 回归分析对预后相关的 IRL 进行二次选择后得到 14 个 IRL 与肺腺癌的生存预后相关。Cox 分析筛选出其中 8 个 IRL 作为预后的独立危险因素用于构建风险评分模型。K-M 生存分析显示,低风险组生存时间长于高风险组( $P<0.001$ )。ROC 曲线分析显示,风险评分、年龄、性别、分期、T、M、N 对肺腺癌预后的曲线下的面积分别为 0.827、0.498、0.579、0.733、0.673、0.508、0.685。对风险模型绘制风险曲线,结果显示随着风险值的升高,患者生存时间逐渐下降,死亡人数逐渐增多。**结论** IRL 构建的 8 个风险模型可能是肺腺癌患者潜在的预测预后生物标志物。

**关键词:**肺腺癌;lncRNA;风险模型;生物标志物

中图分类号:R734.2

文献标识码:A

DOI:10.3969/j.issn.1006-1959.2023.01.001

文章编号:1006-1959(2023)01-0001-06

### Screening of lncRNA Related to Immune Prognosis of Lung Adenocarcinoma Based on Bioinformatics Analysis

SI Lu, QIN Qiu-wen, SHEN Yin, LUO Shuang-yan, HE Ying, ZHANG Yu-xu, LIU Liu

(Department of International Medical, People's Hospital of Guangxi Zhuang Autonomous Region, Nanning 530021, Guangxi, China)

**Abstract: Objective** To identify potential immunoprosthetic lncRNA (IRL) biomarkers by bioinformatics methods based on TCGA and Molecular Signatures Database data for prognosis prediction of lung adenocarcinoma patients. **Methods** The IRL dataset of lung adenocarcinoma and corresponding clinical parameters were obtained from TCGA. Immune genes were extracted from Molecular Signatures Database v7.0. Immune-related IRL was identified by constructing an immune gene co-expression network. Cox regression analysis and Lasso regression analysis were used to screen IRL biomarkers related to the prognosis of lung adenocarcinoma. The value of IRL in predicting the survival prognosis of patients was evaluated by survival analysis and ROC curve analysis. **Results** A total of 1124 IRLs were identified. Through univariate Cox regression analysis, 18 IRLs related to the prognosis of lung adenocarcinoma were initially screened, and 14 IRLs related to the survival and prognosis of lung adenocarcinoma were obtained after secondary selection of prognosis-related IRLs by Lasso regression analysis. Cox analysis screened 8 IRLs as independent risk factors for prognosis to construct a risk scoring model. K-M survival analysis showed that the survival time of the low-risk group was longer than that of the high-risk group ( $P<0.001$ ). ROC curve analysis showed that the area under the curve of risk score, age, gender, stage, T, M and N for the prognosis of lung adenocarcinoma was 0.827, 0.498, 0.579, 0.733, 0.673, 0.508 and 0.685, respectively. The risk curve was drawn for the risk model. The results showed that with the increase of the risk value, the survival time of the patients gradually decreased, and the number of deaths gradually increased. **Conclusion** Eight risk models constructed by IRL may be potential prognostic biomarkers for patients with lung adenocarcinoma.

**Key words:** Lung adenocarcinoma; lncRNA; Risk model; Biomarkers

2018 年国际癌症研究机构(IARC)的数据表明<sup>[1]</sup>, 全球绝大多数国家的肺癌发病率和死亡率均居首位,其中肺腺癌是所有肺癌中最常见的亚型,约

占有所有肺癌的 40%。手术治疗、放化疗在肺腺癌的治疗中起着关键作用<sup>[2-4]</sup>,但有 50%的肺腺癌患者死于肿瘤复发<sup>[5]</sup>,因此识别肺癌预测预后生物标记物,可为患者提供早期、有效的治疗。lncRNA(IRL)是由长度超过 200 个核苷酸的组成的 RNA,因缺少开放阅读框,不具备编码蛋白质的能力<sup>[6]</sup>,但其可通过碱基互补配对,与 DNA、RNA、蛋白质相互作用发挥生物学功能<sup>[7,8]</sup>。近期研究发现<sup>[9]</sup>,IRL 通过调控 T 细胞亚群的凋亡敏感性进而改变肿瘤微环境中 T 细胞亚群的平衡,造成肿瘤的免疫逃逸。免疫系统在癌症

基金项目:广西壮族自治区卫生和计划生育委员会自筹经费课题(编号:桂卫 Z20211383)

作者简介:思璐(1993.6-),女,广西南宁人,硕士,住院医师,主要从事肺癌及慢性阻塞性肺疾病的相关研究

通讯作者:刘柳(1982.6-),女,广西南宁人,硕士,副主任医师,主要从事 2 型糖尿病学的相关研究

的发展进程中起着关键作用<sup>[10]</sup>。多项研究表明<sup>[11,12]</sup>, 可通过鉴定免疫相关的基因以预测癌症患者的生存预后。本研究通过使用 TCGA 数据库和 Molecular Signatures Database v7.0 的数据, 利用生物信息学的方法来鉴定潜在的免疫相关的预后性 IRL 生物标志物, 以期用于肺腺癌患者的预后预测。

## 1 材料与方法

1.1 数据来源 IRL 表达数据和相应的临床数据均从 TCGA 数据库下载(<https://portal.gdc.cancer.gov>)。筛选条件如下: ①肿瘤原发位点: 肺癌; ②病理类型: 肺腺癌; ③数据类型: TCGA-基因表达定量; ④其余筛选条件默认或不选。根据肿瘤患者的生存时间来构建预后模型, 排除临床数据不完整和总体生存率 < 30 d 的肿瘤标本<sup>[13]</sup>, 最终符合要求的有 455 例肿瘤患者。Molecular Signatures Database v7.0 (<http://software.broadinstitute>) 下载免疫相关的基因集(immune system process M13664, Immune response M19817)。通过构建免疫基因共表达网络, 鉴定了 1124 个免疫相关的 IRL。

## 1.2 方法

1.2.1 数据处理及基因共表达分析 通过 strawberry-Perl(版本 5.30, 64bit)将 IRL 表达数据与相应临床数据整合。R 软件“limma”包(版本 3.6.1)鉴定免疫相关的 IRL,  $P < 0.001$  表示差异有统计学意义。

1.2.2 筛选预后相关的 IRL 通过 R 软件“survival”包进行单因素 Cox 回归分析,  $P < 0.001$  认为差异有统计学意义。通过 R 软件“glmnet”包和“survival”包对

有单因素 Cox 回归分析结果中有统计学意义的 IRL 进行 Lasso 回归分析, 筛选出影响肺腺癌预后的关键 IRL。采用多变量 Cox 回归分析构建风险模型, 根据所选 IRL 的基因表达量和系数计算每个 NSCLC 患者的风险评分, 其公式为  $\text{risk core} = G_1 \times \beta_1 G_1 + G_2 \times \beta_2 G_2 + \dots + G_n \times \beta_n G_n$ 。其中,  $n$  为纳入基因的个数,  $G$  代表基因的表达量,  $\beta$  代表基因的系数<sup>[14]</sup>。将风险评分的中位数设为截断值, 风险评分  $\geq$  中位数为高风险组, 风险评分 < 中位数为低风险组。

1.2.3 预后模型评价 运用 Kaplan-Meier(K-M)生存分析评估低风险组和高风险组的生存差异。此外, 采用 ROC 曲线比较风险评分模型与其他临床性状对预后预测的准确性。ROC 曲线分析采用 R 软件“survival ROC”包(版本 3.6.1)进行。AUC < 0.5 表示不显著, 0.5~0.7 表示准确率较低, 0.7~0.9 表示准确率中等, > 0.9 表示准确率较高<sup>[15]</sup>。

## 2 结果

2.1 筛选肺腺癌预后相关的 IRL 首先从 TCGA 数据库筛选出符合条件的肺腺癌样本数据共 455 例。通过 R 软件“limma”包(版本 3.6.1)构建免疫基因共表达, 鉴定了 1124 个 IRL,  $\text{Coef} \geq 0.4$  且  $P < 0.001$ 。通过单因素 Cox 回归分析, 初步筛选出 18 个和肺腺癌预后相关的 IRL, 见图 1。为避免单因素 Cox 回归分析过度拟合问题, 通过 Lasso 回归分析对预后相关的 IRL 进行二次选择, 得到 14 个 IRL 与肺腺癌的生存预后相关, 见图 2。

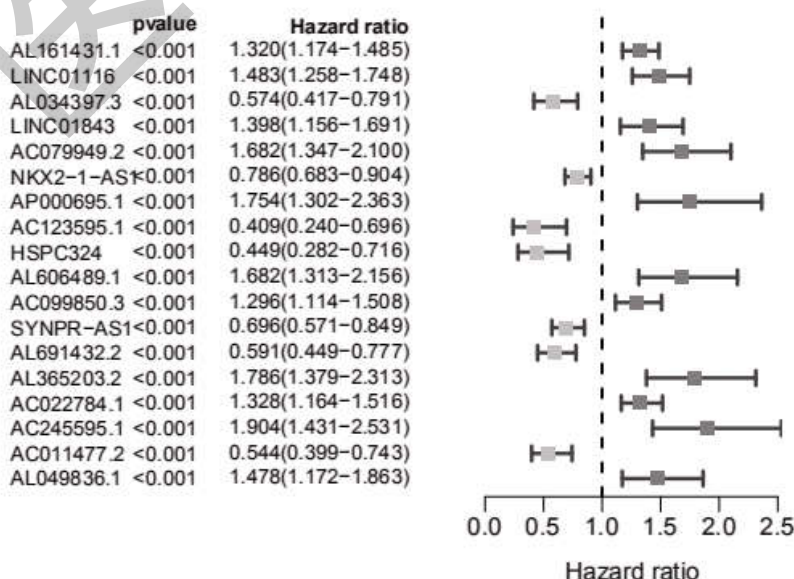


图 1 单因素 Cox 回归分析森林图

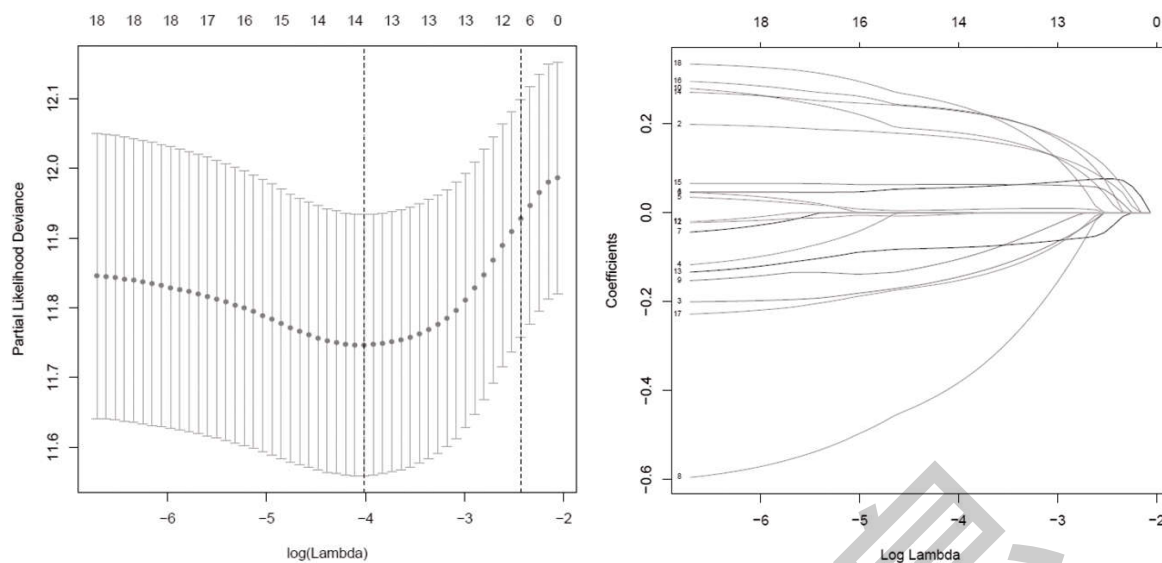


图 2 Lasso 回归分析图

## 2.2 构建 IRL 风险模型和评价

**2.2.1 构建风险模型** 将 Lasso 回归分析确定的 14 个 IRL 纳入多因素 Cox 回归分析构建风险模型,最后得到由 8 个 IRL 组成预后分险模型,见图 3。根据这 8 个 IRL 在样本中的表达量分析计算每个患者的预后风险值: 风险值=LINC01116 的表达量 $\times$ (0.213)+AL034397.3 的表达量 $\times$ (-0.248)+AC123595.1 的表达量 $\times$ (-0.601)+AL606489.1 的表达量 $\times$ (0.285)+AL365203.2 量的表达 $\times$ (0.297)+AC245595.1 的表达量 $\times$ (0.272)+AC011477.2 的表达量 $\times$ (-0.276)+AL049836.1 的表达量 $\times$ (0.3305)。

**2.2.2 评价风险模型** K-M 生存分析显示,低风险组

生存时间长于高风险组( $P<0.001$ ),见图 4。ROC 曲线分析结果显示,风险评分、年龄、性别、分期、T、M、N 对肺腺癌预后的曲线下的面积分别为 0.785、0.498、0.579、0.733、0.673、0.508、0.685,见图 5。此外,对风险模型绘制风险曲线,结果显示随着风险值的升高,患者生存时间逐渐下降,死亡人数逐渐增多,其中 AL034397.3、AC026355.1、AC011477.2 为低风险的基因随着风险值的增大,表达量逐渐下降;AL365203.2、LINC01116、AL606489.1、AC245595.1、AL049836.1 为高风险的基因随着风险值的增大,表达量逐渐升高,见图 6。

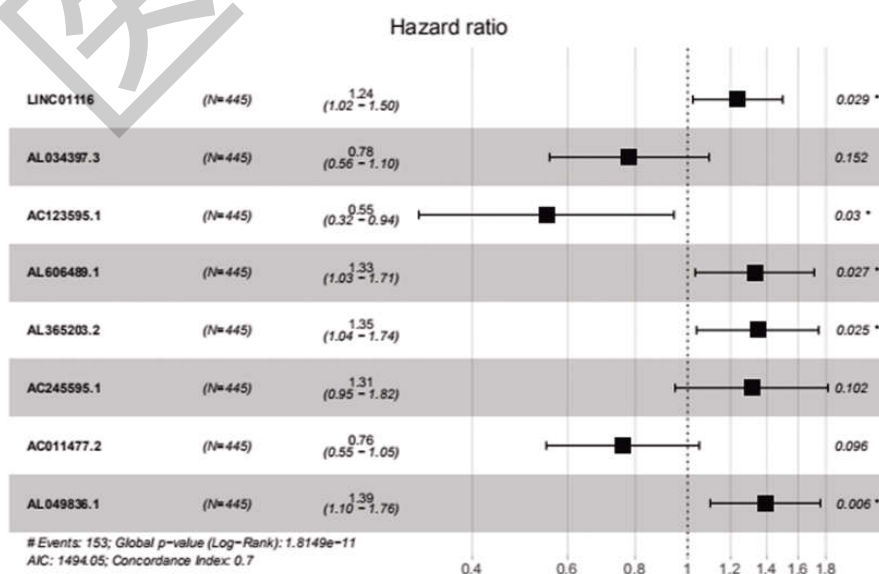


图 3 多因素 Cox 回归分析图

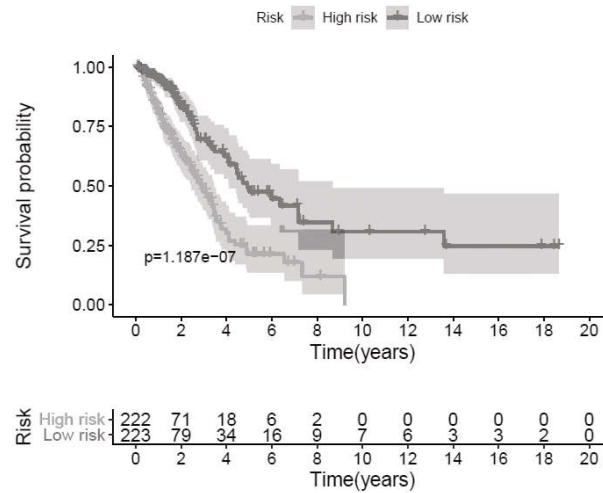


图 4 K-M 生存分析

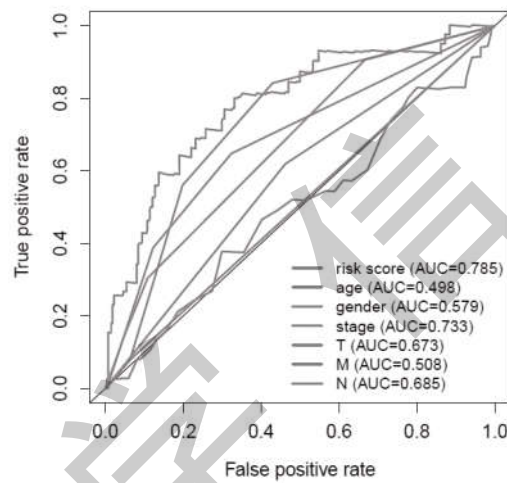
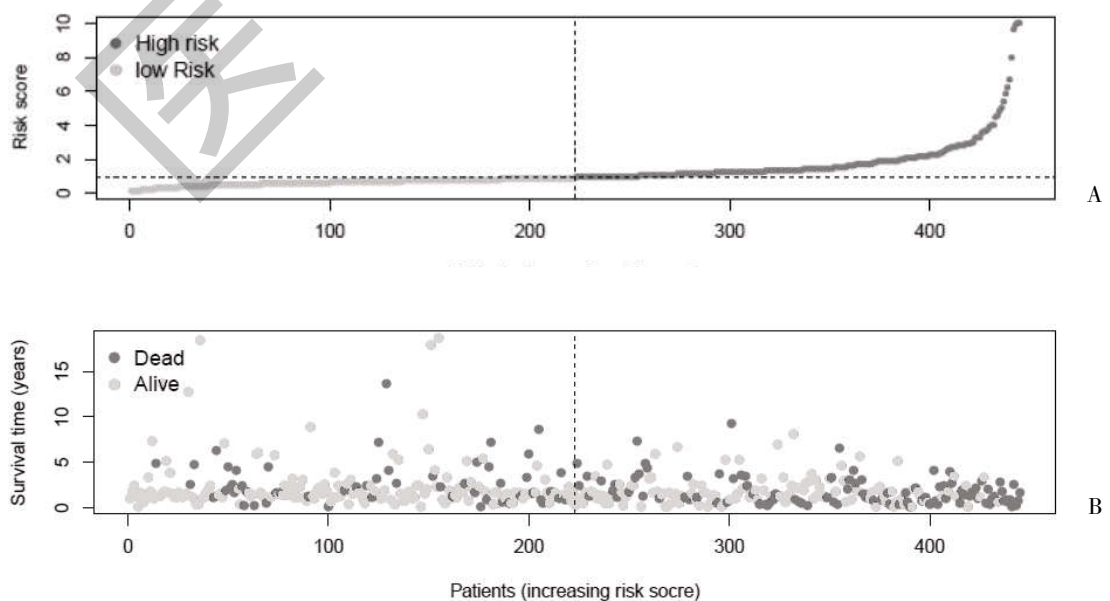
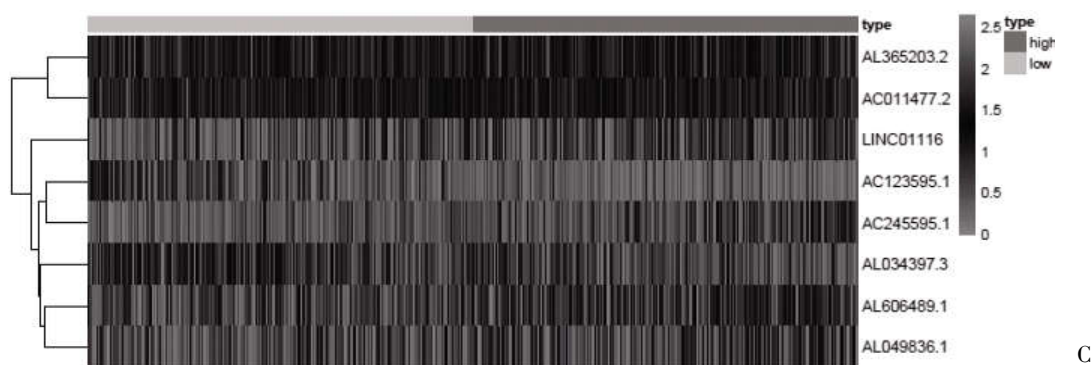


图 5 ROC 曲线图



注:A 风险评分;B:生存状况分布;C:8 个 IRL 表达热图

图 6 风险曲线图及热图



注:A 风险评分;B:生存状况分布;C:8 个 IRL 表达热图

图 6 风险曲线图及热图(续)

### 3 讨论

随着基因芯片技术及高通量测序技术的发展,基因组的表达数据越来越丰富。TCGA 数据库至今共收录了 33 类癌症的基因数据,其中包括 mRNA、lncRNA、microRNA 等<sup>[16,17]</sup>,极大的促进了癌症分子基础研究。近期研究表明<sup>[18,19]</sup>,IRL 在多种疾病的病理及生理途径中发挥重要作用,这为疾病的诊断和治疗方法提供新契机。此外,IRL 通过染色质修饰、转录和翻译过程参与基因的表达调控与肿瘤细胞的增殖、凋亡、侵袭转移、免疫<sup>[20-23]</sup>。肿瘤免疫是机体对肿瘤细胞产生的特异性免疫反应,近年来肿瘤免疫研究取得了重大突破。Huang D 等<sup>[24]</sup>研究发现,IRL 通过调控 T 细胞亚群的凋亡敏感性,从而改变肿瘤微环境中免疫激活及免疫抑制的 T 细胞亚群的平衡,造成肿瘤的免疫逃逸。Sun X 等<sup>[25]</sup>通过免疫基因共表达分析构建了一个免疫相关的 lncRNA 模型以预测卵巢癌患者的预后、药物敏感性和免疫状态,促进了卵巢癌患者的个体化治疗。

多项研究证实<sup>[26-28]</sup>,IRL 与肿瘤诊断、治疗、预后具有显著相关性。然而目前免疫相关的 IRL 在肺腺癌的预后预测少有报道。本研究通过免疫基因共表达网络共鉴定 1124 个 IRL,单因素 Cox 回归分析得出 18 个 IRL 可能与预后相关。考虑单因素 Cox 回归分析每次仅纳入一个变量,存在过度拟合的缺点,本研究在此基础上再次引入 Lasso 回归分析,通过交叉验证  $\lambda$  值,对纳入 Cox 回归分析的 lncRNA 进行二次选择。最后,通过多因素 Cox 回归得出由 LINC01116、AL034397.3、AC123595.1、AL606489.1、AL365203.2、AC245595.1、AC011477.2、AL049836.1 构建的预后风险模型与生存预后显著

相关。此外,本研究中采用中位数截断值进行分组,结果显示低风险组生存时间长于高风险组 ( $P < 0.001$ );ROC 曲线分析结果显示,风险评分、年龄、性别、分期、T、M、N 对肺腺癌预后的曲线下的面积分别为 0.785、0.498、0.579、0.733、0.673、0.508、0.685。此外,对风险模型绘制风险曲线,结果显示随着风险值的升高,患者生存时间逐渐下降,死亡人数逐渐增多,其中 AL034397.3、AC026355.1、AC011477.2 为低风险的基因随着风险值的增大,表达量逐渐下降;AL365203.2、LINC01116、AL606489.1、AC245595.1、AL049836.1 为高风险的基因随着风险值的增大,表达量逐渐升高,提示 AL034397.3、AC026355.1、AC011477.2 可作为肺腺癌预后的危险因素,而 AL365203.2、LINC01116、AL606489.1、AC245595.1、AL049836.1 可作为肺腺癌保护性的预后因素。因此,8 个 IRL 构建的风险模型可为肺腺癌患者的生存提供准确的预测和评估,同样为肺癌预后的基础研究提供更多选择的生物标志物。然而本研究存在一定的局限性,如纳入的数据来源于单一数据库,需要在较大队列中进行进一步验证;且目前 8 个 IRL 在肺腺癌的发生、发展机制以及参与的信号通路等分子生物机制尚不明确,仍需基础实验进一步验证。

综上所述,通过挖掘 TCGA 数据库构建的 8 个 IRL 肺腺癌预后模型有望成为肺腺癌的预后评估生物标志物。

### 参考文献:

- [1]Freddie B,Jacques F,Isabelle S,et al.Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries [J].CA Cancer J

- Clin,2018,68(6):394-424.
- [2]Meng W,Ye Z,Cui R,et al.MicroRNA -31 Predicts the Presence of Lymph Node Metastases and Survival in Patients with Lung Adenocarcinoma[J].Clinical Cancer Research,2013,19(19):5423-5433.
- [3]Denisenko TV,Budkevich IN,Zhivotovsky B.Cell death - based treatment of lung adenocarcinoma [J].Cell Death Dis, 2018,9(2):117.
- [4]Molina JR,Yang P,Cassivi SD,et al.Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship [J]. Mayo Clin Proc,2008,83(5):584-594.
- [5]Fry WA,Phillips JL,Menck HR.Ten-year survey of lung cancer treatment and survival in hospitals in the United States: a national cancer data base report[J].Cancer,1999,86(9):1867-1876.
- [6]Quinn JJ,Chang HY.Unique features of long non-coding RNA biogenesis and function [J].Nature Reviews Genetics, 2015,17(1):47-62.
- [7]Cesana M,Cacchiarelli D,Legnini I,et al.A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA[J].Cell,2011,147(2):358-369.
- [8]Ferrè F,Colantoni A,Helmer -Citterich M.Revealing protein-lncRNA interaction [J].Brief Bioinform,2016,17 (1):106-116.
- [9]van Dijk N,Funt SA,Blank CU,et al.The Cancer Immunogram as a Framework for Personalized Immunotherapy in Urothelial Cancer[J].Eur Urol,2019,75(3):435-444.
- [10]Desrichard A,Snyder A,Chan TA.Cancer Neoantigens and Applications for Immunotherapy[J].Clin Cancer Res,2016,22(4): 807-812.
- [11]Angell H,Galon J.From the immune contexture to the Immunoscore: the role of prognostic and predictive immune markers in cancer[J].Curr Opin Immunol,2013,25(2):261-267.
- [12]Shen Y,Peng X,Shen C.Identification and validation of immune-related lncRNA prognostic signature for breast cancer[J]. Genomics,2020,112(3):2640-2646.
- [13]Liao X,Zhu G,Huang R,et al.Identification of potential prognostic microRNA biomarkers for predicting survival in patients with hepatocellular carcinoma [J].Cancer Manag Res, 2018,10:787-803.
- [14]彭慧,秦凯,戴宇翔,等.基于 TCGA 数据库的胶质母细胞瘤 LncRNA 风险预测模型的建立 [J]. 肿瘤防治研究,2019,46(5):35-38.
- [15]杨飞龙,洪锴,赵国江,等.基于长链非编码 RNA 的生物信息学分析构建膀胱癌预后模型并确定预后生物标志物 [J].北京大学学报(医学版),2019,51(4):615-622.
- [16]Wang Z,Jensen MA,Zenkhusen JC.A Practical Guide to The Cancer Genome Atlas (TCGA)[J].Methods in Molecular Biology,2016,1418:111.
- [17]Deng M,Brägelmann J,Schultze JL,et al.Web -TCGA: an online platform for integrated analysis of molecular cancer data sets[J].BMC Bioinformatics,2016,17:72.
- [18]Skuratovskaia D,Vulf M,Komar A,et al.Promising Directions in Atherosclerosis Treatment Based on Epigenetic Regulation Using MicroRNAs and Long Noncoding RNAs [J]. Biomolecules,2019,9(6):226.
- [19]Hu G,Niu F,Humburg BA,et al.Molecular mechanisms of long noncoding RNAs and their role in disease pathogenesis[J]. Oncotarget,2018,9(26):18648-18663.
- [20]Kopp F,Mendell JT.Functional Classification and Experimental Dissection of Long Noncoding RNAs [J].Cell,2018,172(3):393-407.
- [21]Tan YT,Lin JF,Li T,et al.LncRNA-mediated posttranslational modifications and reprogramming of energy metabolism in cancer[J].Cancer Commun (Lond),2021,41(2):109-120.
- [22]Li Y,Egranov SD,Yang L,et al.Molecular mechanisms of long noncoding RNAs-mediated cancer metastasis [J].Genes, Chromosomes and Cancer,2019,58(4):200-207.
- [23]He RZ,Luo DX,Mo YY.Emerging roles of lncRNAs in the post-transcriptional regulation in cancer [J].Genes & Diseases, 2019,6(1):6-15.
- [24]Huang D,Chen J,Yang L,et al.NKILA lncRNA promotes tumor immune evasion by sensitizing T cells to activation-induced cell death[J].Nat Immunol,2018,19(10):1112-1125.
- [25]Sun X,Li S,Lv X,et al.Immune-Related Long Non-coding RNA Constructs a Prognostic Signature of Ovarian Cancer [J]. Biol Proced Online,2021,23(1):24.
- [26]Schmitt A,Chang H.Long Noncoding RNAs in Cancer Pathways[J].Cancer Cell,2016,29(4):452-463.
- [27]Bhan A,Soleimani M,Mandal SS.Long Noncoding RNA and Cancer: A New Paradigm[J].Cancer Res,2017,77(15):3965-3981.
- [28]Zhang A,Xu M,Mo YY.Role of the lncRNA-p53 regulatory network in cancer [J].Journal of Molecular Cell Biology, 2014,6(3):181-191.

收稿日期:2022-03-08;修回日期:2022-03-24

编辑/杜帆