

· 论 著 ·

肺腺癌中关键 RNA 结合蛋白预后评估模型的构建*

杨文娟, 马青梅, 马 莉, 马梦婷, 张 倩, 刘欣跃[△]

兰州大学第二医院检验医学中心, 甘肃兰州 730030

摘要:目的 基于生物信息学方法, 评估肺腺癌中关键 RNA 结合蛋白的表达及预后作用。方法 从癌症基因组图谱(TCGA)数据库中下载 526 例肺腺癌组织、59 例正常组织, 以及从基因型-组织表达(GTEX)数据库中下载 288 例正常组织的 RNA 测序数据, 筛选差异表达 RNA 结合蛋白; 单因素和多因素 Cox 回归分析筛选关键 RNA 结合蛋白并构建预后评估模型。结果 共发现 375 个差异表达 RNA 结合蛋白, 筛选出 8 个预后相关的关键 RNA 结合蛋白(WDR3、SMG9、DARS2、CARHSP1、LARP6、GAR1、INTS7 和 EXO1)。构建基于 8 个关键 RNA 结合蛋白的预后评估模型发现, 试验组和验证组模型的高风险患者较低风险患者总体生存时间更短($P < 0.05$), 两组受试者工作特征曲线下面积分别为 0.761 和 0.666, 该预后模型具有有效性和准确性。结论 该研究成功构建肺腺癌中关键 RNA 结合蛋白预后评估模型, 上述 8 个关键 RNA 结合蛋白可作为预测肺腺癌预后的分子标志物。

关键词: 肺腺癌; RNA 结合蛋白; 预后

DOI: 10.3969/j.issn.1673-4130.2021.08.009

中图法分类号: R734.2

文章编号: 1673-4130(2021)08-0932-05

文献标志码: A

Construction of key RNA binding protein prognostic evaluation model in lung adenocarcinoma*

YANG Wenjuan, MA Qingmei, MA Li, MA Mengting, ZHANG Qian, LIU Xinyue[△]

Department of Clinical Laboratory, Lanzhou University Second Hospital, Lanzhou, Gansu 730030, China

Abstract: Objective Based on bioinformatics methods, to evaluate the expression and prognosis effect of key RNA binding protein in lung adenocarcinoma. **Methods** From The Cancer Genome Atlas (TCGA), the RNA sequencing data of 526 cases of lung adenocarcinoma tissues, 59 cases of normal tissues and 288 cases of the Genotype-Tissue Expression (GTEX) normal tissues were downloaded to screen the differentially expressed RNA binding proteins. Univariate and multivariate Cox regression analysis were used to screen the key RNA binding proteins and construct the prognostic evaluation model. **Results** A total of 375 differentially expressed RNA binding proteins were found, and 8 key RNA binding proteins related to prognosis were screened (WDR3, SMG9, DARS2, CARHSP1, LARP6, GAR1, INTS7 and EXO1). The construction of a prognostic evaluation model based on 8 key RNA binding proteins found that the overall survival time of the high-risk patients of the experimental group and the validation group model was shorter than the lower-risk patients ($P < 0.05$), and the areas under the receiver operating characteristic curve of the two groups were 0.761 and 0.666, respectively, the prognostic model had validity and accuracy. **Conclusion** The study successfully constructed an RNA binding protein prognostic evaluation model in lung adenocarcinoma. The above 8 key RNA-binding proteins can be used as molecular markers for the prognosis of lung adenocarcinoma.

Key words: lung adenocarcinoma; RNA binding protein; prognosis

肺癌的发病率居各大癌症之首, 也是引起癌症患者死亡的主要原因^[1]。据世界卫生组织报道, 2018 年肺癌发病和死亡人数分别为 210 万例、180 万例, 肺腺癌占所有肺癌的 40%~55%, 发病率逐年上升^[2]。目

前, 尚缺乏对肺腺癌进行有效诊断和预后评估的分子标志物, 导致大多数患者临床确诊时已处于晚期且 5 年生存率较低。因此, 迫切需要分析肺腺癌预后的风险因素, 以及确定肺腺癌的预后分子标志物。

* 基金项目: 兰州大学第二医院“萃英科技创新”计划项目(CY2017-MS17)。

作者简介: 杨文娟, 女, 技师, 主要从事药物基因组学研究。△ 通信作者, E-mail: liuxy@lzu.edu.cn。

本文引用格式: 杨文娟, 马青梅, 马莉, 等. 肺腺癌中关键 RNA 结合蛋白预后评估模型的构建[J]. 国际检验医学杂志, 2021, 42(8): 932-936.

RNA 结合蛋白(RBPs)是一类具有 RNA 结合域的蛋白总称,包括信使 RNA(mRNA)、长链非编码 RNA(lncRNAs)、微小 RNA(miRNAs)和核糖体 RNA(rRNAs)等,参与剪接、修饰、转运、定位和翻译等一系列生物学过程。此外,RBPs 在维持细胞生理稳态,尤其是发育和应激反应等方面也发挥重要功能^[3]。GERSTBERGER 等^[4]已筛选出 1 542 种 RBPs,然而目前仅有少数研究探讨了 RBPs 的生物学功能及致病机制。研究报道称大多数肿瘤中均有 RBPs 表达失调的现象,可能与肿瘤的发展、迁移、预后等过程密切相关^[5-6]。本文对肺腺癌患者 RNA 测序数据进行生物信息学分析,挖掘差异表达 RBPs,探讨 RBPs 对肺腺癌患者生存预后的影响。

1 资料与方法

1.1 一般资料 从癌症基因组图谱(TCGA)数据库中下载 526 例肺腺癌组织和 59 例正常组织的 RNA 测序数据。为了扩大正常组织的样本数,下载基因组-组织表达(GTEX)数据库中 288 例正常组织的 RNA 测序数据。以上 TCGA 和 GTEX 数据库中肺腺癌和正常组织的测序数据均来自 UCSC 数据库(<https://xenabrowser.net/datapages/>)。

1.2 GO 功能注释和京都基因与基因组百科全书(KEGG)富集分析 为了明确差异表达 RBPs 在肺腺癌发生发展中的作用,使用 R 软件 Clusterprofiler 包对差异表达 RBPs 进行功能富集分析,包括 GO 功能注释[生物学过程、细胞学组分(CC)和分子生物学功能(MF)]和 KEGG 通路富集(BP)分析。

1.3 生存分析 收集从 UCSC 数据库中下载的 500 例肺腺癌患者的临床信息,采用 Perl 软件合并肺腺癌患者差异表达 RBPs 的表达和生存数据。通过单因素和多因素 Cox 回归分析筛选关键 RBPs。根据风险评估公式: $\text{风险评分} = \text{Exp}_1 \times \text{Coe}_1 + \text{Exp}_2 \times \text{Coe}_2 + \dots + \text{Exp}_i \times \text{Coe}_i$ (i, Exp 和 Coe 分别表示关键 RBPs 的数目、表达值和系数值),建立关键 RBPs 生存分析的预后模型。R 软件将关键 RBPs 的表达及生存数据随机分为试验组(252 例)和验证组(248 例),根据风险评分分为高风险组和低风险组,并绘制受试者工作特征曲线(ROC 曲线),计算曲线下面积(AUC)。此外,绘制验证组关键 RBPs 的 ROC 曲线并计算 AUC 值,对试验组风险评估模型进行验证。

1.4 关键 RBPs 筛选和验证 采用 Kaplan Meier-plotter 数据库(<https://kmplot.com/analysis/>)验证关键 RBPs 对肺腺癌预后总生存期(OS)的影响;采用人类蛋白质图谱(HPA)数据库(<https://www.proteinatlas.org/>)对关键 RBPs 在肺腺癌中的表达进行

验证,进一步验证构建预后模型的有效性。

1.5 统计学处理 采用 Perl 软件对下载的肺腺癌和正常组织的 RNA 测序数据进行整理及合并,得到基因表达矩阵文件。根据已知的 1 542 个 RBPs 列表,用 R 软件(3.6.1)获取肺腺癌及正常组织中 RBPs 的表达矩阵。以 $|\log_2\text{FC}| > 1$ 且错误发现率 < 0.05 为标准,R 软件 Limma 包筛选肺腺癌和正常组织中差异表达 RBPs,Ggplot2 包绘制差异表达 RBPs 的表达分布图。

2 结果

2.1 肺腺癌中差异表达 RBPs 的筛选和富集分析 本研究共纳入 526 例肺腺癌组织、347 例正常组织的 RNA 测序数据。按照 $|\log_2\text{FC}| > 1$ 且错误发现率 < 0.05 的标准,用 R 软件 Limma 包筛选出 375 个差异表达 RBPs。差异表达 RBPs 的 GO 分析结果表明,差异表达 RBPs 主要参与 RNA 代谢、lncRNA 的代谢过程和 RNA 剪切等 BP;MF 主要为 RNA 的催化活性、核糖体结构组成部分和核糖核酸酶活性;核糖体、线粒体基质和核糖体亚单位等的 CC。此外,差异表达 RBPs 参与核糖体、mRNA 的监控通路、RNA 的转录和剪切体等 KEGG 通路。见表 1。

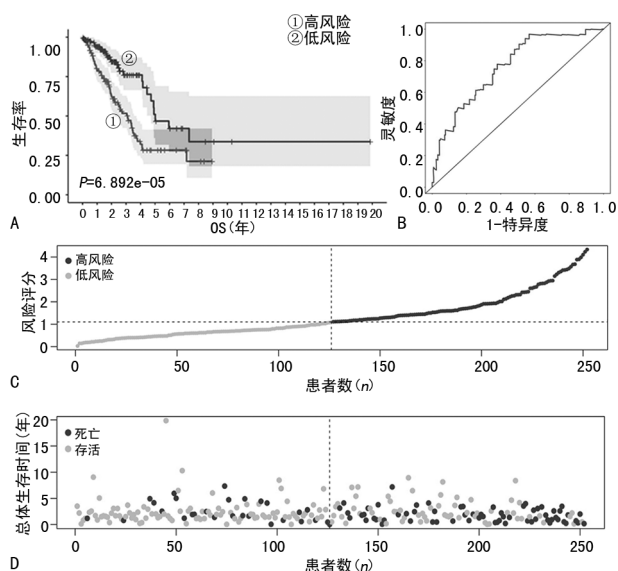
表 1 差异 RBPs 的功能富集表

差异表达 RBPs 的功能富集	名称	P	数目 (n)
GO			
BP	核糖核蛋白的生物学过程	2.72E-63	89
BP	RNA 代谢过程	2.13E-67	87
BP	lncRNA 代谢过程	7.14E-57	84
BP	RNA 剪切	3.40E-56	83
BP	mRNA 代谢过程	3.91E-58	77
CC	核糖体	7.54E-70	77
CC	核糖体亚单位	8.13E-76	72
CC	线粒体基质	1.13E-24	51
CC	大核糖体亚单位	2.22E-48	46
CC	胞浆核糖体	4.68E-47	44
MF	核糖体的结构组成	1.39E-55	62
MF	RNA 催化活性	1.67E-42	67
MF	核酸酶活性	2.43E-15	28
MF	翻译调控因子活性	1.10E-18	27
MF	核糖核酸酶活性	4.62E-20	26
KEGG	核糖体	5.19E-47	54
	RNA 的转录	1.40E-16	31
	剪切体	2.56E-15	27
	真核生物核糖体的生物合成	1.19E-11	20
	mRNA 的监控通路	5.86E-09	16

2.2 生存预后模型的构建 对 375 个差异表达 RBPs 进行单因素 Cox 回归分析, 获得 26 个 RBPs。对上述 26 个 RBPs 进行多因素 Cox 回归分析, 得到 8 个关键 RBPs (WDR3、SMG9、DARS2、CARHSP1、LARP6、GAR1、INTS7 和 EXO1), 它们是影响肺癌患者 OS 的独立预后因素。将包含 8 个关键 RBPs 的表达及生存数据的肺癌患者, 随机分为试验组和验证组。按照风险评估公式: 风险评分 = (0.696 × WDR3) + (0.764 × SMG9) + (0.591 × DARS2) + (0.675 × CARHSP1) + (0.301 × LARP6) + (0.526 × GAR1) + (0.498 × INTS7) + (-354 × EXO1), 建立肺癌患者生存预后的风险评估模型。在试验组中, 8 个关键 RBPs 的风险评分模型的 AUC 为 0.761, 且高风险肺癌患者的 OS 较低风险者短, 表明该模型对肺癌患者 OS 可进行良好的预测。在验证组中, 高风险肺癌患者 OS 较低风险者短, AUC 为 0.666。本研究成功构建了具有较高的灵敏度和特异度的肺癌 RBP 预后评估模型, 见表 2、图 1~2。

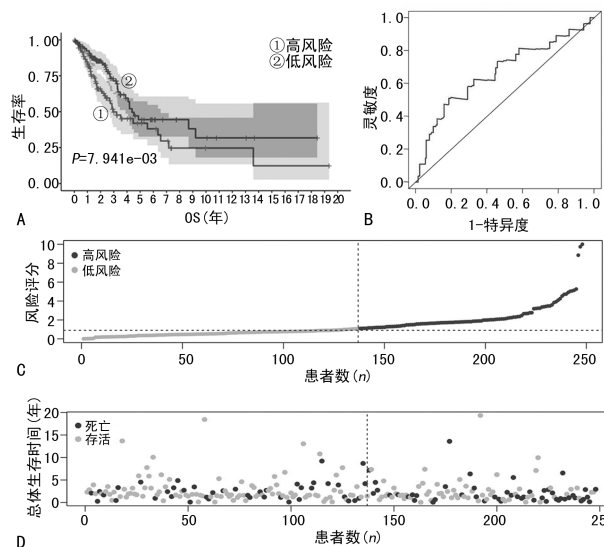
表 2 多因素 Cox 回归分析筛选关键 RBPs

关键 RBPs	P	HR(95%CI)
WDR3	0.012	2.005(1.166~3.449)
SMG9	0.005	2.146(1.262~3.651)
DARS2	0.019	1.806(1.103~2.957)
CARHSP1	<0.001	1.964(1.325~2.913)
LARP6	0.055	1.351(0.994~1.837)
GAR1	0.046	1.691(1.009~2.836)
INTS7	0.096	1.645(0.915~2.958)
EXO1	0.044	0.702(0.498~0.990)



注: A 为高、低风险患者的生存曲线图; B 为风险评分的 OS 的 ROC 曲线图; C 为风险评分分布图; D 为生存状态图。

图 1 试验组中 8 个关键 RBP 预后风险评估模型



注: A 为高、低风险患者的生存曲线图; B 为风险评分的 OS 的 ROC 曲线图; C 为风险评分分布图; D 为生存状态图。

图 2 验证组中 8 个关键 RBP 预后风险评估模型

2.3 关键 RBPs 的列线图 为了定量预测肺癌患者 1、3、5 年的生存率, 构建 8 个关键 RBPs 的列线图。在总点轴和各基因预后轴之间画 1 条垂直线, 进而估算 1、3 和 5 年的生存率, 为肺癌患者的临床决策提供依据。见图 3。

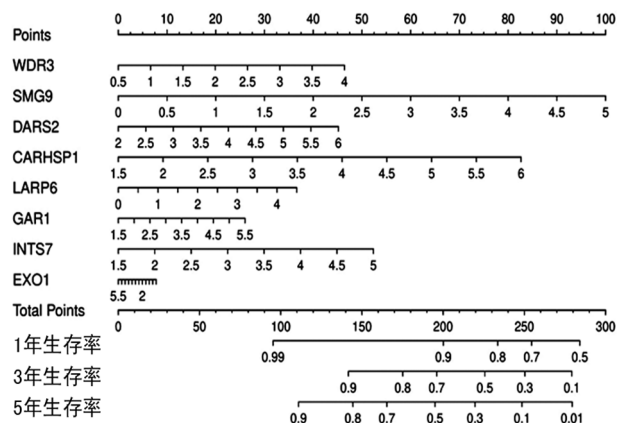
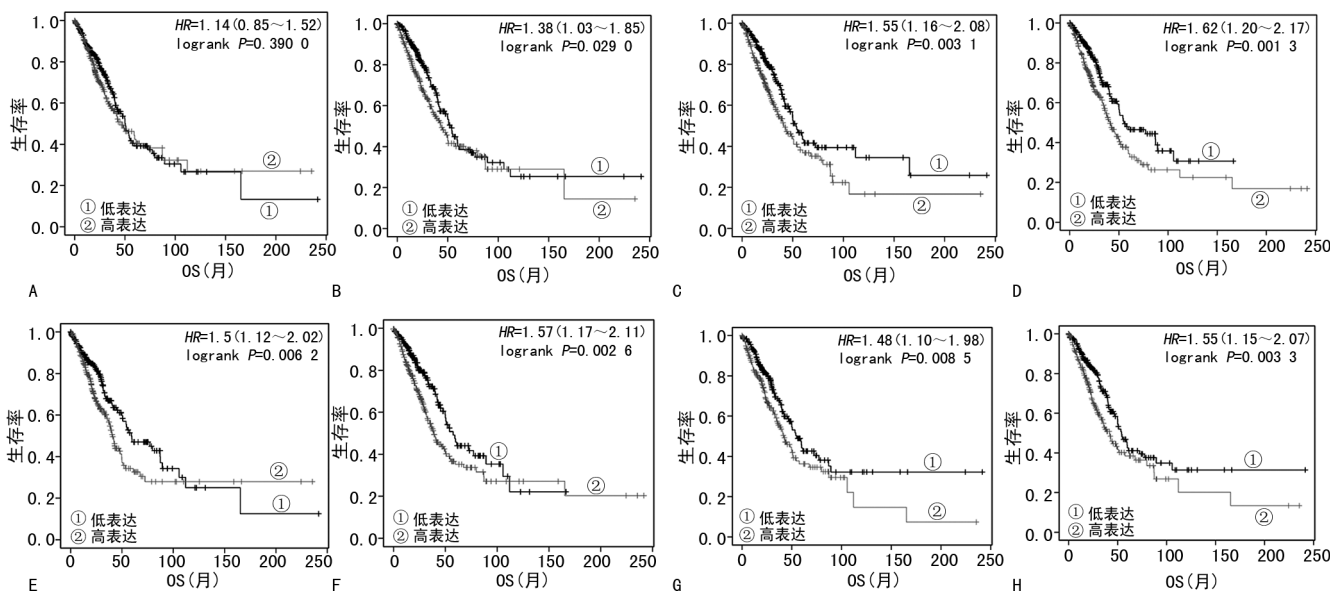


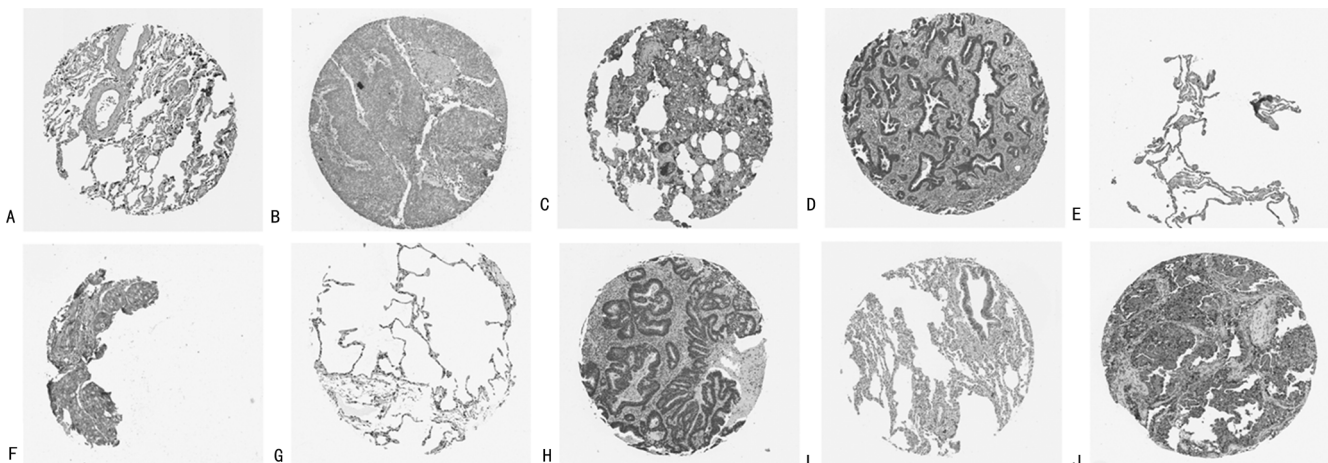
图 3 预测肺癌患者 1、3、5 年总体生存时间的列线图

2.4 关键 RBPs 的预后和表达验证 通过 Kaplan Meier-plotter 数据库明确肺癌患者中 8 个预后关键 RBPs 与 OS 的关系, 发现高表达 WDR3、DARS2、CARHSP1、LARP6、GAR1、INTS7 和 EXO1 基因的肺癌患者预后更差 ($P < 0.05$); 而 SMG9 对预后 OS 的影响无统计学意义 ($P > 0.05$)。为了进一步明确上述关键 RBPs 在肺癌中的表达情况, HPA 数据库免疫组化结果表明在肺癌组织中 SMG9、DARS2、CARHSP1、LARP6、GAR1 基因表达较正常组织高。然而, HPA 数据库尚缺乏 WDR3、INTS7 和 EXO1 基因的免疫组化表达数据。见图 4、5。



注:A~H 分别为 SMG9、WDR3、DARS2、CARHSP1、LARP6、GAR1、INTS7、EXO1 高表达者和低表达者 Kaplan Meier 生存分析图。

图 4 Kaplan Meier-plotter 数据库中肺腺癌中关键 RBPs 总体生存时间分析



注:A 为 SMG9 基因在正常组织中低表达;B 为 SMG9 基因在肺腺癌组织高表达;C 为 DARS2 基因在正常组织未检测到;D 为 DARS2 基因在肺腺癌组织高表达;E 为 CARHSP1 基因在正常组织低表达;F 为 CARHSP1 基因在肺腺癌组织中度表达;G 为 GAR1 基因在正常组织高表达 (25%~75%);H 为 GAR1 基因在肺腺癌组织高表达 (>75%);I 为 LARP6 基因在正常组织低表达;J 为 LARP6 基因在肺腺癌组织高表达。

图 5 HPA 数据库中关键 RBPs 在肺腺癌组织和正常组织中的表达验证

3 讨论

本研究从 UCSC 数据库中下载肺腺癌和正常组织的 RNA 测序数据, R 软件筛选差异表达的 RBPs, 获得 375 个差异表达 RBPs。通过单因素和多因素 Cox 回归分析得到 8 个关键 RBPs, 成功构建了肺腺癌患者 RBP 预后风险评估模型, 且验证组结果表明该模型具有较好的预后预测意义; 构建的列线图进一步直观定量地描述了肺腺癌患者 1、3、5 年生存率。Kaplan Meier-plotter 数据库表明 WDR3、DARS2、CARHSP1、LARP6、GAR1、INTS7 和 EXO1 7 个关键 RBPs 高表达的肺腺癌患者的总体生存率较低表达患者低。使用 HPA 数据库验证 SMG9、DARS2、CARHSP1、LARP6、GAR1 等基因在肺癌组织中高表达, 与本研究成果一致。上述结论均表明, 本研

究建立的关键 RBPs 预后评估模型, 对肺腺癌患者调整治疗策略、预后标志物的探索有一定的意义。

目前, 已有研究者对预后关键 RBPs 在肿瘤中的作用展开了多项研究。在西班牙人群中展开的研究发现 WDR3 基因可增加甲状腺癌的风险, 且在甲状腺癌细胞中 WDR3 表达上调^[7]。研究表明 DARS2 基因可延长肝细胞肝癌的细胞周期并抑制肝细胞肝癌细胞的凋亡^[8]; DARS2 也参与膀胱癌的发生、发展和转移等过程^[9]。胡仁旺等^[10]发现 LARP6 的异常表达可影响胃癌患者的 OS 和无进展生存期。研究表明 GAR1 在胃癌组织中尚无表达差异, 而在结肠癌组织中表达上调, 可能在结肠癌的发生发展中起着重要作用^[11]。INTS7 在转录调控中起着重要作用, 且具有高度突变性。前列腺癌的侵袭性与 INST7 的过表达

有关^[12];INTST7 被证实在胃癌、膀胱癌和肺腺癌中均为过表达^[13]。EXO1 基因异常表达可能影响前列腺癌患者的预后生存及进展^[14];ZHENG 等^[15]发现,乳腺癌患者 OS 较短,与 EXO1 基因的表达上调有关。不同肿瘤细胞的增殖、侵袭及肿瘤患者的预后与关键 RBPs 的异常表达有关,未来可通过干扰上述 RBPs 的表达,改善肺腺癌患者的预后生存情况。

综上所述,本研究成功构建了肺腺癌关键 RBPs 预后评估模型,为进一步探索 RBPs 对肺腺癌的影响提供了新思路,有助于临床开发新的靶向治疗药物,预测患者预后生存时间。然而,本研究尚存在一定的局限性,仍需要大量测序数据及相应的前瞻性临床试验进一步验证。

参考文献

[1] 中华医学会,中华医学会肿瘤学分会,中华医学会杂志社.中华医学会肺癌临床诊疗指南(2019 版)[J].中华肿瘤杂志,2020,42(4):257-287.

[2] BRAY F, FERLAY J, SOERJOMATARAM I, 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.

[3] MASUDA K, KUWANO Y. Diverse roles of RNA-binding proteins in cancer traits and their implications in gastrointestinal cancers [J]. Wiley Interdiscip Rev RNA, 2019, 10(3):e1520.

[4] GERSTBERGER S, HAFNER M, TUSCHL T. A census of human RNA-binding proteins [J]. Nat Rev Genet, 2014, 15(12):829-845.

[5] MOHIBI S, CHEN X, ZHANG J. Cancer the 'RBP' eutics-RNA-binding proteins as therapeutic targets for cancer [J]. Pharmacol Ther, 2019, 203:107390.

[6] WANG Z L, LI B, LUO Y X, et al. Comprehensive genomic characterization of RNA-binding proteins across human cancers[J]. Cell Rep, 2018, 22(1):286-298.

[7] AKDI A, GIMÉNEZ E M, GARCÍA-QUIPES W, et al. WDR3 gene haplotype is associated with thyroid cancer risk in a Spanish population [J]. Thyroid, 2010, 20(7):803-809.

[8] QIN X, LI C, GUO T, et al. Upregulation of DARS2 by HBV promotes hepatocarcinogenesis through the miR-30e-5p/MAPK/NFAT5 pathway [J]. J Exp Clin Cancer Res, 2017, 36(1):148.

[9] GUO C, SHAO T, JIANG X, et al. Comprehensive analysis of the functions and prognostic significance of RNA-binding proteins in bladder urothelial carcinoma [J]. Am J Transl Res, 2020, 12(11):7160-7173.

[10] 胡仁旺, 龚建平. RNA 结合蛋白 LARP6 影响胃癌浸润及预后 [J]. 中国普外基础与临床杂志, 2020, 27(11):1408-1412.

[11] KIM M S, KIM S S, YOO N J, et al. Expressional analysis of NOLA1, NOLA2, NOLA3 and DKC1, the core proteins in H/ACA riboproteins, in gastric and colorectal cancers [J]. Pathology, 2012, 44(6):576-577.

[12] KWON O K, HA Y S, LEE J N, et al. Comparative proteome profiling and mutant protein identification in metastatic prostate cancer cells by quantitative mass spectrometry-based proteogenomics [J]. Cancer Genomics Proteomics, 2019, 16(4):273-286.

[13] FEDERICO A, RIENZO M, ABBONDANZA C, et al. Pan-cancer mutational and transcriptional analysis of the integrator complex [J]. Int J Mol Sci, 2017, 18(5):936.

[14] HUA X, GE S, CHEN J, et al. Effects of RNA binding proteins on the prognosis and malignant progression in prostate cancer [J]. Front Genet, 2020, 11:591667.

[15] ZHENG G, ZHANG C, ZHONG C. Identification of potential prognostic biomarkers for breast cancer using WGCNA and PPI integrated techniques [J]. Ann Diagn Pathol, 2020, 50:151675.

(收稿日期:2020-09-22 修回日期:2021-02-25)

(上接第 931 页)

[13] 范智文, 赵一俏, 凌龙, 等. miR-34a 在当归补血汤抗小鼠心肌梗死后心室重塑中的表达 [J]. 实用医学杂志, 2017, 33(18):3009-3012.

[14] 罗劲勇, 万书平. 冠心病患者外周血 miR-34a、IL-18 表达与临床 SYNTAX 评分的相关性研究 [J]. 广东药科大学学报, 2018, 34(4):513-516.

[15] 靳文, 鄢文, 李冬义, 等. 肠源性毒素硫酸吡啶酚与慢性肾功能不全大鼠心肌 NLRP3 炎性小体、miR-34a 表达的相关性研究 [J]. 现代消化及介入诊疗, 2019, 24(4):343-346.

[16] 李珊, 吕安林, 王海昌, 等. 能量限制激活 Sirt1 减轻心肌

细胞缺血/再灌注损伤的作用 [J]. 心脏杂志, 2016, 28(4):405-410.

[17] 胡明珠, 周波, 盛琼, 等. PI3K/Akt/Sirt1 信号通路介导硫化氢后处理对大鼠缺血心肌的保护作用 [J]. 中国药理学通报, 2016, 32(2):268-273.

[18] FU B C, LANG J L, ZHANG D Y, et al. Suppression of miR-34a expression in the myocardium protects against ischemia-reperfusion injury via SIRT1 protective pathway [J]. Stem Cells Dev, 2017, 26(17):1270-1282.

(收稿日期:2020-09-25 修回日期:2020-12-29)