

• 论 著 •

miRNA 作为乳腺癌潜在预后生物标志物的构建及评估*

张珂¹, 张开炯², 吴立春², 陈莉¹

四川省肿瘤医院·研究所/四川省癌症防治中心/电子科技大学医学院附属

肿瘤医院; 1. 输血科; 2. 检验科, 四川成都 610041

摘要:目的 基于癌症基因组图谱(TCGA)数据库筛选乳腺癌预后相关的微小 RNA(miRNA)生物标志物。方法 从 TCGA 数据库中获取乳腺癌 miRNA 族谱,采用生物信息学分析方法,对差异表达明显的 miRNA 采用单因素 Cox、LASSO 回归、逐步多因素 Cox 回归分析,建立预后风险评估模型,采用受试者工作特征曲线评估模型预测能力。结果 在乳腺癌 TCGA 数据库中筛选出 269 个差异表达明显的 miRNA。首先采用单因素 Cox 分析和 LASSO 回归分析,筛选出 18 个与乳腺癌预后相关的 miRNA,继而采用逐步多因素 Cox 回归分析,最终筛选出由 8 个 miRNA 组合的评分模型作为乳腺癌预后预测的生物标志物,其与单个生物标志物相比具有良好的准确性和可靠性。结论 筛选的 8 个 miRNA 组合的评分模型可作为预测乳腺癌预后的生物标志物。

关键词:乳腺癌; 癌症基因组图谱; 微小 RNA; 预后**DOI:**10.3969/j.issn.1673-4130.2021.22.005**中图法分类号:**R737.9**文章编号:**1673-4130(2021)22-2710-04**文献标志码:**A

Construction and evaluation of miRNA as potential prognostic biomarkers for breast cancer*

ZHANG Ke¹, ZHANG Kaijiong², WU Lichun², CHEN Li¹

1. Department of Blood Transfusion; 2. Department of Clinical Laboratory, Sichuan Cancer Hospital · Institute/Sichuan Cancer Center/the Affiliated Cancer Hospital School of Medicine, University of Electronic Science and Technology of China, Chengdu, Sichuan 610041, China

Abstract: **Objective** Identification of miRNA as prognostic biomarkers for breast cancer patients based on the cancer genome atlas (TCGA) database. **Methods** The miRNA expression profile and the clinical information regarding patients with breast cancer were recruited from the TCGA database. Bioinformatics analysis was used to establish prognostic risk assessment models for miRNA with obvious differential expression by univariate Cox, LASSO regression and stepwise multivariate Cox regression. Receiver operator characteristic curve was applied to determine the efficiency of survival prediction. **Results** A total of 269 differentially expressed miRNA were screened from the TCGA database of breast cancer. First, univariate Cox analysis and LASSO regression analysis were used to screen 18 miRNA associated with prognosis of breast cancer. Then, stepwise multivariate Cox regression analysis was used to screen 8 miRNA as biomarkers for prognosis of breast cancer. Compared with individual biomarkers, it has good accuracy and reliability. **Conclusion** The scoring model of the 8 miRNA combinations might be used as biomarkers to predict prognosis of breast cancer.

Key words: breast cancer; the cancer genome atlas; miRNA; prognosis

乳腺癌是全球女性最常见的恶性肿瘤,其发病率位居第一^[1]。世界卫生组织 2020 年最新癌症报告显示,乳腺癌新增人数达 226 万,正式取代肺癌成为全球第一大癌症。尽管近年来乳腺癌的治疗方法已有很大改善,但由于个体差异,预后仍然不理想^[2]。因此,有必要找到敏感而精确的生物标志物,以更好地预测乳腺癌患者的生存和预后。微小 RNA(miRNA)是一种由 21~23 个核苷酸组成的高度保守的内源性

短链非编码 RNA 分子^[3]。现已证明 miRNA 可靶向超过 60% 的人类基因^[4],几乎参与了每个生物学过程,包括发育,血管生成,细胞分化、增殖、存活、侵袭、凋亡,伤口愈合和免疫细胞功能等^[5]。既往研究证实,miRNA 在肿瘤的发生、发展、转移和患者存活中起重要作用,其能够作为肿瘤的潜在治疗工具和诊断生物标志物^[6-7]。近年来,许多集中于乳腺癌 miRNA 表达谱的研究发现,miRNA 在预测患者预后中已显

* 基金项目:四川省科技创新创业苗子工程(重点)项目(2021JDRC0152);四川省卫生健康委员会普及应用项目(20PJ115)。

作者简介:张珂,男,副主任技师,主要从事肿瘤分子诊断与输血治疗相关研究。

本文引用格式:张珂,张开炯,吴立春,等. miRNA 作为乳腺癌潜在预后生物标志物的构建及评估[J]. 国际检验医学杂志, 2021, 42(22): 2710-2713.

示出巨大希望^[8]。然而,miRNA 在乳腺癌中的生物学功能仍不为人所知,为提高乳腺癌预后的预测准确性,仍需要系统地研究多种 miRNA 的表达,利用整体的危险评分进行预后评价。因此,鉴定出更强大和实用的 miRNA 信号具有重要的临床意义。本研究通过开发和验证基于癌症基因组图谱(TCGA)数据库的 miRNA 预后模型,并进一步系统地评估该模型对乳腺癌患者总体生存的预测价值,以明确 miRNA 作为乳腺癌预后生物标志物的可靠性。

1 材料与方法

1.1 数据来源 从 TCGA 数据库 (<https://portal.gdc.cancer.gov/>) 下载 1 086 例乳腺癌组织和 804 例正常组织的 miRNA 测序数据和临床信息。

1.2 方法

1.2.1 差异表达 miRNA 筛选 采用 R 语言 edgeR 程序包筛选肿瘤组织与正常组织之间差异表达的 miRNA,以 $FDR < 0.05$ 和 $|\log_2 \text{fold change}| > 1$ 作为筛选标准。

1.2.2 预后相关 miRNA 筛选 将 1 086 例乳腺癌患者表达矩阵和临床信息进行清洗,剔除连续表达量低的数据,将清洗后剩余的 1 033 例乳腺癌患者按照 7:3 比例分为训练组(723 例)与验证组(310 例),首先使用 R 程序包 survival 进行单因素 Cox 回归分析,其次采用 glmnet 程序包进行 LASSO 回归分析,再次使用 survival 包进行逐步多因素 Cox 回归分析,以赤池信息准则(AIC)为准,选取 AIC 信息统计量最小的模型为最终模型,以 $P < 0.05$ 作为筛选标准, $P < 0.05$ 说明该变量为独立的预后影响因素。

1.2.3 miRNA 风险模型构建与验证 将单因素 Cox 回归分析与 LASSO 回归分析筛选出的 miRNA 进行逐步多因素 Cox 回归分析,建立预后风险模型。根据风险得分中值将训练组的患者分为低风险组和高风险组,使用 Kaplan-Meier 计算高低风险组患者之间的生存差异。使用 R 软件包 survival 受试者工作特征(ROC)曲线计算模型 1 年、3 年和 5 年的曲线下面积(AUC)。将训练组构建的预后风险模型纳入验证组进行验证,并计算高低风险组患者之间生存差异和模型 1 年、3 年和 5 年的 AUC。

1.3 统计学处理 采用 R 语言(版本 4.0.3)进行统计分析。对连续变量指标进行数据分布检验,正态分布且方差齐性,数据以 $\bar{x} \pm s$ 表示,组间比较采用独立样本 t 检验;否则采用 Wilcoxon 非参数检验。采用 Kaplan-Meier 生存曲线进行生存分析,采用单因素 Cox 回归、LASSO 回归和多因素 Cox 比例风险回归分析确定与生存相关的独立预后因素。以 $P < 0.05$ 为差异有统计学意义。

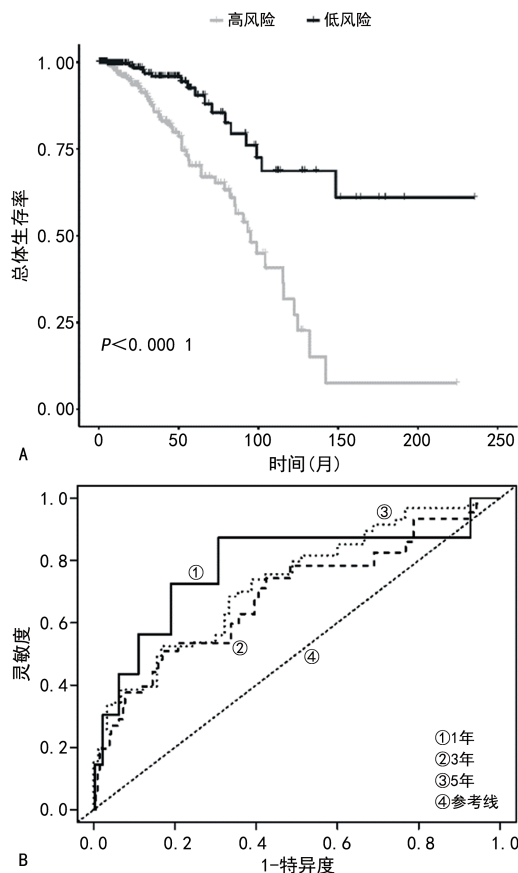
2 结果

2.1 乳腺癌差异表达 miRNA 筛选 共发现 269 个差异表达的 miRNA,其中 204 个 miRNA 表达上调,

65 个 miRNA 表达下调。

2.2 乳腺癌总体生存相关 miRNA 筛选 单因素 Cox 回归分析发现,35 个 miRNA 与乳腺癌总体生存相关。进一步通过 LASSO 回归分析发现,以最小 λ 值作为筛选准则,筛选出 18 个 miRNA 与乳腺癌患者总体生存相关。逐步多因素 Cox 回归分析发现,8 个 miRNA (hsa-miR-3923、hsa-miR-105-2、hsa-miR-3927、hsa-miR-449c、hsa-miR-106a、hsa-miR-1262、hsa-miR-605 和 hsa-miR-3929)与乳腺癌患者总体生存情况相关。

2.3 乳腺癌 miRNA 预后风险模型的构建 基于逐步多因素 Cox 回归分析筛选出的 8 个 miRNA,将每个 miRNA 的多因素 Cox 回归系数和每个 miRNA 的表达值构建了 1 个由 8 个 miRNA 组成的风险评分模型:风险得分 = $0.361 \times \text{hsa-miR-3923} + 0.097 \times \text{hsa-miR-105-2} + 0.724 \times \text{hsa-miR-3927} - 0.333 \times \text{hsa-miR-449c} - 0.232 \times \text{hsa-miR-106a} - 0.366 \times \text{hsa-miR-1262} + 0.288 \times \text{hsa-miR-605} + 0.979 \times \text{hsa-miR-3929}$,根据风险得分中位值将训练组患者分为高低风险两组(图 1A),结果显示,高风险组患者的总体生存率明显低于低风险组患者($P < 0.001$);ROC 曲线分析显示,风险模型 1 年、3 年和 5 年的 AUC 分别为 0.779、0.688 和 0.728(图 1B)。

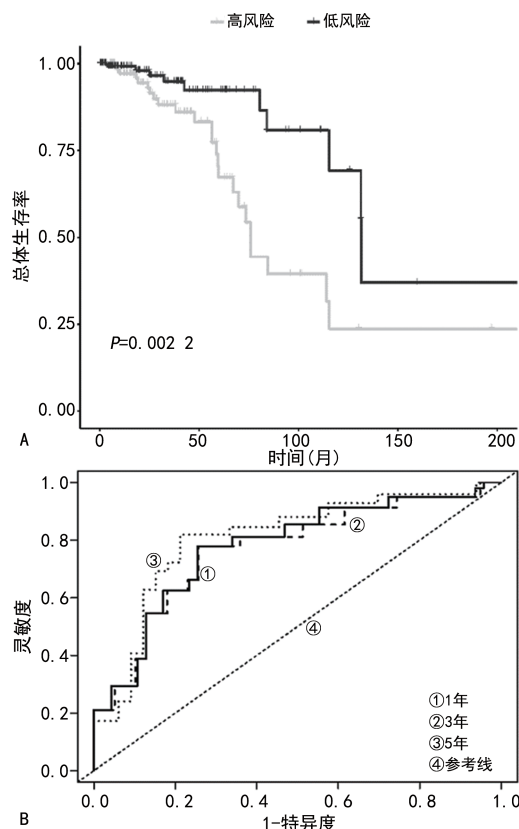


注:A 为训练组高低风险组 Kaplan-Meier 生存曲线;B 为训练组采用 ROC 曲线评估预测模型 1 年、3 年和 5 年的预测效能。

图 1 训练组 Kaplan-Meier 生存曲线和 ROC 曲线

2.4 乳腺癌 miRNA 预后风险模型的验证 在验证

组中低风险组患者的总体生存率显著高于高风险组 ($P < 0.05$, 图 2A), 风险模型 1 年、3 年和 5 年的 AUC 分别为 0.777、0.768 和 0.806 (图 2B)。



注:A 为验证组高低风险组 Kaplan-Meier 生存曲线;B 为验证组采用 ROC 曲线评估预测模型 1 年、3 年和 5 年的预测效能。

图 2 验证组 Kaplan-Meier 生存曲线和 ROC 曲线

3 讨 论

目前,越来越多的研究表明,差异表达的 miRNA 与乳腺癌的发生发展密切相关,其可作为乳腺癌诊断、预后判断和治疗的潜在生物标志物^[9-10]。与单个生物标志物相比,多个生物标志物的综合判别可以提高预测准确性。本研究先后通过单因素 Cox、LASSO 回归和多因素 Cox 比例风险回归,确定了 8 个 miRNA (hsa-miR-3923、hsa-miR-105-2、hsa-miR-3927、hsa-miR-449c、hsa-miR-106a、hsa-miR-1262、hsa-miR-605 和 hsa-miR-3929) 可作为乳腺癌独立的预后预测指标。然后,根据它们的表达谱和回归系数构建了一个由 8 个 miRNA 组成的风险模型,并用于预测乳腺癌患者的总体生存率。为了明确所构建模型预后预测价值的可靠性,本研究设置了训练组和验证组两个独立的研究队列。在训练组中,根据患者的风险评分将患者分为高风险组和低风险组,Kaplan-Meier 生存曲线表明两组患者的总体生存率差异有统计学意义 ($P < 0.001$),高风险组患者的总体生存率明显低于低风险组患者。ROC 曲线分析显示,miRNA 模型在预测乳腺癌患者存活风险方面具有较高的预测准

确性,可作为预测乳腺癌患者预后的可靠生物标志物。

在既往的研究中,研究者主要通过单个基因和通路分析探究 miRNA 作为乳腺癌或其他肿瘤的生物标志物的价值^[11-12]。例如,LI 等^[11]发现,miR-106b-5p 和 miR-93-5p 在乳腺癌细胞中表达水平增高,其高表达可抑制抑癌基因 PTEN 的表达水平,从而促进乳腺癌细胞的迁移、侵袭和增殖,同时增强乳腺癌细胞磷脂酰肌醇 3 激酶 (PI3K)/蛋白激酶 B (Akt) 途径的活性;ZHAO 等^[12]通过多变量分析发现,miR-106a 是胶质母细胞瘤 (GBM) 患者总体生存的独立预后生物标志物,可用于预测单个 GBM 患者的预后和治疗反应。BERTOLI 等^[13]研究发现,与单个 miRNA 相比,循环液体中多个 miRNA 图谱显示出更好的诊断和预后预测性能以及更高的灵敏度。因此,多基因多模型综合分析在肿瘤诊断和治疗中显得尤为重要。

近年来,亦有研究者将多基因综合分析用于肿瘤诊断和预后分析。WANG 等^[14]通过 TCGA 数据库构建了由 4 个 miRNA (hsa-miR-126、hsa-miR-3613、hsa-miR-424 和 hsa-miR-4772) 组成的用于预测胰腺腺癌 (PAAD) 总体生存率的模型,该模型 2 年生存率 AUC 为 0.789,其可作为预测 PAAD 患者生存的有效独立预后生物标志物。SIRIWARDHANA 等^[15]采用单因素 Cox 和逐步多因素 Cox 回归筛选出 5 个 miRNA (miR-375、miR-582-3p、miR-326、miR-181c-5p 和 miR-99a-5p) 用于肺腺癌预后模型的构建,其模型可作为预测肺腺癌预后的可靠工具。最近,LUO 等^[16]采用 LASSO Cox 的方法筛选出包括 miR130b、miR-18a 和 miR-223 在内的 3 个 miRNA。根据 3 个 miRNA 特征性危险评分的最佳临界值,将所有患者分为高危组和低危组。两组患者的生存率差异有统计学意义 ($P < 0.0001$),并且 3 个 miRNA 标志物具有良好的预后预测准确性,可作为肾透明细胞癌独立的预后预测因素。这些结果表明,多基因模型的构建在肿瘤预后的预测准确性方面具有重要意义。

尽管本研究构建了由 8 个 miRNA 组合的预后分子标志物,其在乳腺癌预后预测中具有良好的准确性和可靠性,显示了其可作为乳腺癌预后生物标志物的巨大潜能,但也存在一些局限性。首先,虽然本研究设置了验证组以验证所构建模型的可靠性,但没有进行临床试验;第二,本研究所构建的预后风险模型虽然能够明显区分高低风险组,但 1 年、3 年和 5 年风险得分的 AUC 仅能达到 75% 左右,其预测效能欠佳。后续研究将进一步纳入该 8 个 miRNA 进行临床验证,同时纳入其他数据库进行功能分析和探究。

综上所述,本研究构建了 1 个由 8 个 miRNA 组

成的与乳腺癌患者总体生存显著相关的预测模型,可以从高预后风险的患者中准确识别出低预后风险的患者,并在验证组中评估了该模型的准确性和可靠性。这些结果表明,整合的 8 个 miRNA 组合可以作为预测乳腺癌患者预后的潜在标志物。

参考文献

- [1] 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.
- [2] WAKS A G, WINER E P. Breast cancer treatment: a review[J]. JAMA, 2019, 321(3): 288-300.
- [3] NAHAND, J S, KARIMZADEH M R, NEZAMNIA M, et al. The role of miR-146a in viral infection[J]. IUBMB life, 2020, 72(3): 343-360.
- [4] FRIEDMAN R C, FARH K K, BURGE C B, et al. Most mammalian mRNAs are conserved targets of microRNAs[J]. Genome research, 2009, 19(1): 92-105.
- [5] ANDORFER C A, NECELA B M, THOMPSON E A, et al. MicroRNA signatures: clinical biomarkers for the diagnosis and treatment of breast cancer[J]. Trends Mol Med, 2011, 17(6): 313-319.
- [6] HAYES J, PERUZZI P P, LAWLER S. MicroRNAs in cancer: biomarkers, functions and therapy [J]. Trends Mol Med, 2014, 20(8): 460-469.
- [7] SHABANINEJAD Z, YOUSEFI F, MOVAHEDPOUR A, et al. Electrochemical-based biosensors for microRNA detection: nanotechnology comes into view[J]. Anal Biochem, 2019, 581: 113349.
- [8] KHALIFE H, SKAFI N, FAYYAD-KAZAN M, et al. MicroRNAs in breast cancer: new maestros defining the melody[J]. Cancer Genet, 2020, 246/247: 18-40.
- [9] MCGUIRE A, BROWN J A, KERIN M J. Metastatic breast cancer: the potential of miRNA for diagnosis and treatment monitoring[J]. Cancer Metastasis Rev, 2015, 34(1): 145-155.
- [10] KANDETTU A, RADHAKRISHNAN R, CHAKRABARTY S, et al. The emerging role of miRNA clusters in breast cancer progression[J]. Biochim Biophys Acta Rev Cancer, 2020, 1874(2): 188413.
- [11] LI N, MIAO Y, SHAN Y, et al. miR-106b and miR-93 regulate cell progression by suppression of PTEN via PI3K/Akt pathway in breast cancer[J]. Cell Death Dis, 2017, 8(5): e2796.
- [12] ZHAO S, YANG G, MU Y, et al. miR-106a is an independent prognostic marker in patients with glioblastoma[J]. Neuro Oncol, 2013, 15(6): 707-717.
- [13] BERTOLI G, CAVA C, CASTIGLIONI I. microRNAs: new biomarkers for diagnosis, prognosis, therapy prediction and therapeutic tools for breast cancer[J]. Theranostics, 2015, 5(10): 1122-1143.
- [14] WANG Z X, DENG T X, MA Z. Identification of a 4-miRNA signature as a potential prognostic biomarker for pancreatic adenocarcinoma[J]. J Cell Biochem, 2019, 120(10): 16416-16426.
- [15] SIRIWARDHANA C, KHADKA V S, CHEN J J, et al. Development of a miRNA-seq based prognostic signature in lung adenocarcinoma[J]. BMC cancer, 2019, 19(1): 34.
- [16] LUO Y, CHEN L, WANG G, et al. Identification of a three-miRNA signature as a novel potential prognostic biomarker in patients with clear cell renal cell carcinoma[J]. J Cell Biochem, 2019, 120(8): 13751-13764.
- [15] XUE F, HUANG F. The effect of psychological intervention on nutrient status of perioperative patients with lung cancer[J]. Iran J Public Health, 2018, 47(4): 531-537.
- [16] NEEL D R, MCCLAVE S, MARTINDALE R. Hypoalbuminaemia in the perioperative period: clinical significance and management options[J]. Best Pract Res Clin Anaesthesiol, 2011, 25(3): 395-400.
- [17] LOPATEGI A, FLORES-COSTA R, RIUS B, et al. Frontline science: specialized proresolving lipid mediators inhibit the priming and activation of the macrophage NL-PP3 inflammasome[J]. J Leukoc Biol, 2019, 105(1): 25-36.
- [18] CHENG H, CLYMER J W, HAN C B, et al. Prolonged operative duration is associated with complications: a systematic review and meta-analysis[J]. J Surg Res, 2018, 229: 134-144.
- [19] RUNGSAKULKIJ N, TANGTAWEE P, SURAGUL W, et al. Correlation of serum albumin and prognostic nutritional index with outcomes following pancreaticoduodenectomy[J]. World J Clin Cases, 2019, 7(1): 28-38.

(收稿日期: 2021-03-22 修回日期: 2021-08-10)

(收稿日期: 2021-01-12 修回日期: 2021-08-10)