

基于 TCGA 数据筛选潜在的宫颈癌 miRNA 预后标志物

苑飞艳¹, 张新民², 陈明珠¹, 杨鸣¹, 张立会^{1*}

基金项目:吉林省自然科学基金(项目编号:20200201561JC)

作者单位:1.130051 吉林 长春,吉林大学第二医院妇产科;2.130021 吉林 长春,吉林大学药学院再生医学科学研究所

作者简介:苑飞艳,毕业于吉林大学,硕士,主要研究方向为妇科肿瘤

* 通信作者,E-mail:lihui3527@sina.com

【摘要】目的 基于 TCGA 数据库中 miRNA 表达量及临床资料,寻找与宫颈癌患者生存相关的 microRNA (miRNA or miR) 生物标记物,构建模型以预测宫颈癌患者的生存期。**方法** 从 TCGA 数据库中提取出宫颈癌样本的 miRNA 表达及患者临床信息完整的数据,将样本随机分为训练集和验证集,通过统计学及生物信息学方法(Kaplan-Meier 生存曲线、单因素 Cox 回归分析、Lasso 回归、logRanK 检验,多因素 Cox 回归建模,受试者工作特征曲线检验)筛选出与宫颈癌预后相关的 miRNA。**结果** 在训练集中,通过单因素 Cox 生存分析,Lasso 回归压缩建模得出 5 个 miRNA 与宫颈癌预后有关($P < 0.05$),多因素 Cox 回归分析表明在训练集和验证集中该模型都属于独立预后因素($P < 0.05$),ROC 曲线表明该模型能够对预后做良好预判。基因本体论,通路,疾病富集分析都表明该模型中的 miRNA 标志物靶基因富集在与肿瘤相关的生物过程,通路,疾病上。**结论** 本研究构建的宫颈癌患者 miRNA 标志物预后模型,具有较好的可靠性和特异性,对 miRNA 在宫颈癌预后判断具有较高借鉴作用。

【关键词】 宫颈癌;生物信息学;miRNA;TCGA;预后标志物

【中图分类号】R 711.74 **【文献标志码】**A **【文章编号】**1674-4020(2022)12-081-05

doi:10.3969/j.issn.1674-4020.2022.12.20

Screening potential prognostic markers of cervical cancer miRNA based on TCGA data

Yuan Feiyan¹, Zhang Xinmin², Chen Mingzhu¹, Yang Ming¹, Zhang Lihui^{1*}

1. Department of Obstetrics and Gynecology, The Second Hospital of Jilin University, Changchun Jilin 130051; 2. Institute of Regenerative Medicine, College of Pharmacy, Jilin University, Changchun Jilin 130021, P. R. China

* Corresponding author, E-mail:lihui3527@sina.com

[Abstract] **Objective** To find the microRNA biomarkers related to the survival of patients with cervical cancer based on the expression of miRNA in TCGA database and clinical data, and to construct a model to predict the survival time of patients with cervical cancer. **Methods** The miRNA expression of cervical cancer samples and the data of complete clinical information of patients were extracted from TCGA database, and the samples were randomly divided into training set and verification set. The miRNA related to the prognosis of cervical cancer were screened by statistical and bioinformatics methods (Kaplan-Meier survival curve, univariate Cox regression analysis, Lasso regression, logRanK test, multivariate Cox regression modeling, receiver working characteristic curve test). **Results** In the training set, through univariate Cox survival analysis and Lasso regression compression modeling, five miRNA were related to the prognosis of cervical cancer ($P < 0.05$). Multivariate Cox regression analysis showed that the model was an independent prognostic factor in both the training group and the verification group ($P < 0.05$). ROC curve showed that the model could predict the prognosis well. Gene ontology, pathway and disease enrichment analysis all show that the target genes of miRNA markers in this model

have a significant enrichment in biological processes, pathways, and diseases related to tumors.

are enriched in tumor-related biological processes, pathways and diseases. **Conclusions** The prognostic model of miRNA markers in patients with cervical cancer constructed in this study has good credibility and specificity, and can be used as a reference for miRNA in judging the prognosis of cervical cancer.

[Key words] cervical cancer; bioinformatics; miRNA; TCGA; prognostic marker

宫颈癌是严重威胁全球女性健康和生命的恶性肿瘤,是女性癌症死亡的第四大原因,据估计,2020年全球新增病例为60.4万例,死亡人数为34.2万人^[1]。我国宫颈癌发病数约占全球总发病数的30%^[2],近年发展中国家宫颈癌的发生率和死亡率仍居高不下,且逐渐趋于年轻化。

宫颈癌主要采用手术治疗,但对于晚期宫颈癌治疗需要辅以放化疗^[3],目前癌症发生发展的分子机制尚不明确,有研究证实miRNA参与癌症的发生、进展、转移和耐药等各个阶段,将miRNA用作临床生物标记物用于诊断、预测和预后是可行的^[4]。有研究表明,致癌miRNA上调或抑癌miRNA下调可促进癌症的发生,故在癌症的靶向治疗上可有两种策略:即用miRNA模拟物用于替代、恢复下调的抑癌miRNA;或用拮抗剂减少、抑制上调的致癌miRNA^[5-6]。目前已经发现miRNA表达谱能够根据组织来源对肿瘤进行分类,且其表达变异与转移过程有关,故miRNA可作为生物标志物用于肿瘤的诊断及预测^[7]。TCGA数据收集并整理了多种肿瘤的多组学及临床数据,因此本文从TCGA数据库中获取宫颈癌样本的miRNA表达,筛选出宫颈癌预后的相关生物标记物,为临床工作及研究提供理论依据。

1 材料和方法

1.1 数据的获取及处理

从TCGA数据下载宫颈癌患者的miRNA表达数据、临床数据,去除重复数据、缺失数据,选择出肿瘤样本300个,提取样本中miRNA表达数据,将相同的成熟miRNA进行合并,合并过程区分3p和5p异构体,共有2 188种miRNA,将NA值转换为0,计算miRNA在所有样品种含量为0的样品数,如果超过25%样品为零,删除该miRNA,miRNA数量变为588个,使用miRNAMeConverter数据包将miRNA名称由TCGA数据转变成为通用名称,最终成功转化的585种miRNA。

1.2 数据分组

将筛选的300个肿瘤样本使用R语言随机分组函数(sample函数)分为训练集和验证集,各150例,训练集用于构建预测模型,验证集用于验证预测模型效能。

1.3 差异miRNA的筛选

对训练集使用R语言包Survival进行单因素Cox生存回归分析,计算每个miRNA与宫颈癌患者生存的风险比(hazard ratio, HR)和P值,从基因表达量角度分析,HR小于1表示对生存有利^[8]。选出P值小于0.05的与宫颈癌预后的相关的miRNA。

1.4 风险模型的构建

对训练集将筛选得到的miRNA使用glmnet包进行Lasso回归分析,删除回归系数为零的miRNA,筛选出与生存相关的miRNA,将每个miRNA的回归系数作为风险系数(Coef),建立风险值公式:Risk score = Coef₁ × miRNA₁Exp + Coef₂ × miRNA₂Exp + + Coef_n × miRNAnExp^[9-10],式中Coef为相应miRNA的多因素回归系数,Exp为相应miRNA的表达量,根据Risk score数值的中位数,将CC患者分为高风险评分组和低风险评分组。使用R语言包Survival, Dplyr, Survival ROC, ggplot2, survivalsvm进行数据整理,Kaplan-Meier曲线绘制,以及logRanK检验。使用多因素Cox分析风险值在训练组和验证组中是否属于独立预后因素。最后绘制风险评分和基于miRNA临床预后模型的ROC曲线并计算AUC值以评估模型的预测能力。

2 结果

2.1 样本统计

将TCGA的miRNA表达数据和临床数据中的生存时间和生存状态、种族、分期,是否经过治疗进行合并后,数据特征如表1:

表1 筛选的300个样本数据特征

临床指标	例数(%)
种族	
阿拉斯加人	8(2.67%)
亚洲人	20(6.67%)
非裔美国人	28(9.33%)
夏威夷土著民	2(0.67%)
未明的	35(11.67%)
白种人	207(69.00%)
生存状态	
生存	228(76.00%)
死亡	72(24.00%)
FIGO分期	
I-II期	233(77.67%)
III-IV期	67(22.33%)
治疗方法	
化疗	165(55.00%)
放疗	135(45.00%)

2.2 样本分组

两组数据的种族、生存状态、分期以及治疗方法,采用 χ^2 检验,两个分组中所有变量的P值都大于0.05,两

组数据中每个变量的组成,差异无统计学意义(见表2),可以用于建模和验证分析。

表 2 比较训练集和验证集差异的统计学分析

临床指标	训练集(%)	验证集(%)	χ^2 值	P 值
种族			4.773	0.44
阿拉斯加人	4(2.67%)	4(2.67%)		
亚洲人	10(6.67%)	10(6.67%)		
白种人	18(12%)	10(6.67%)		
非裔美国人	1(0.67%)	1(0.67%)		
未明的	21(14%)	14(9.33%)		
夏威夷土著民	96(64%)	111(74%)		
生存状态			0.292	0.589
生存	34(22.67%)	38(25.33%)		
死亡	116(77.33%)	112(74.67%)		
FIGO 分期			0.480	0.488
I-II 期	31(20.67%)	36(24%)		
III-IV 期	119(79.33%)	114(76%)		
治疗方法			0.121	0.728
化疗	84(56%)	81(54%)		
放疗	66(44%)	69(46%)		

2.3 差异 miRNA 的筛选

使用 Cox 单因素方差分析对每种 miRNA 表达量与生存分析,发现 70 种 miRNA 的 P 值小于 0.05,由于前 30 个 $P < 0.01$,因此选择这 30 个 miRNA 作为初步候选 miRNA,为了预后检测更加高效易行,使用 glmnet 包初筛的 miRNA 进行 Lasso 回归分析,将回归系数为零的 miRNA 逐渐剔除,最终筛选出的 5 种 miRNA 及风险系数(Coef)见表 3。

表 3 筛选的 5 种 miRNA 及对应的风险系数

miRNA	Coef
hsa-miR-101-3p	-1.96528772572106e-06
hsa-miR-532-5p	-0.000137761316051204
hsa-miR-150-3p	-0.0183375190755258
hsa-miR-378d	-0.20993920228295
hsa-miR-378a-5p	-0.00167590719238135

2.4 风险评分方程的建立及风险模型的建立

对上述 5 个 miRNA 进行多因素 Cox 回归分析,根据相应的回归系数,建立风险评分方程:Risk score = Coef₁ × hsa-miR-101-3pExp + Coef₂ × hsa-miR-532-5pExp + Coef₃ × hsa-miR-150-3pExp + Coef₄ × hsa-miR-378dExp + Coef₅ × hsa-miR-378a-5pExp, 将每个病人的这 5 种 miRNA 风险值相加按照大小排序,以中位数为界限,将训练集和验证集分为高风险组和低风险组。采用 Kaplan-Meier 生存分析,绘制 KM 曲线对预后模型进行验证,结果表明:在训练集和验证集中,高风险组的预后比低风险组差(生存时间单位为:天),P 值 < 0.05(见图 1)。

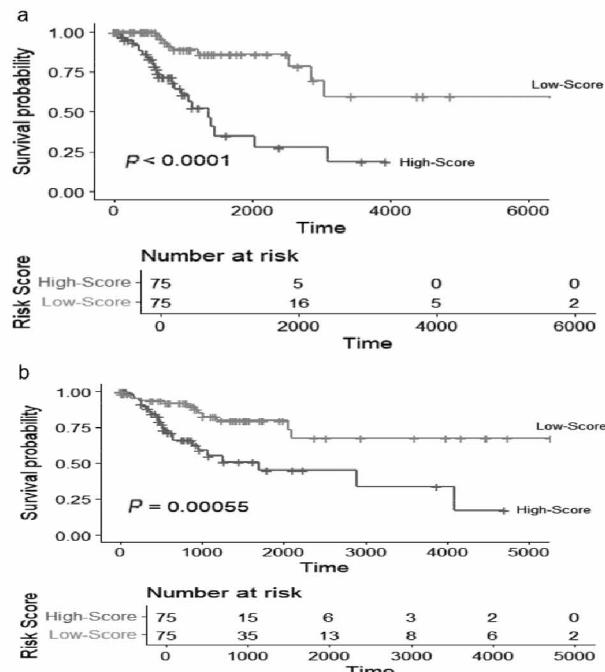


图 1 训练集(a)和验证集(b)中高低风险评分组患者的生活曲线

2.5 多因素 Cox 分析

使用多因素 Cox 分析种族、发病年龄、肿瘤分期、风险值这四个因素中,风险值在训练组和验证组中属于独立预后因素:训练集风险评分($HR = 6.123, 95\% CI: 2.714-13.814, P < 0.001$),验证集风险评分($HR = 3.230, 95\% CI: 1.578-6.610, P = 0.001$),见图 2,训练集与验证集中风险值的 P 值 < 0.05,说明该模型属于独立预后因素。

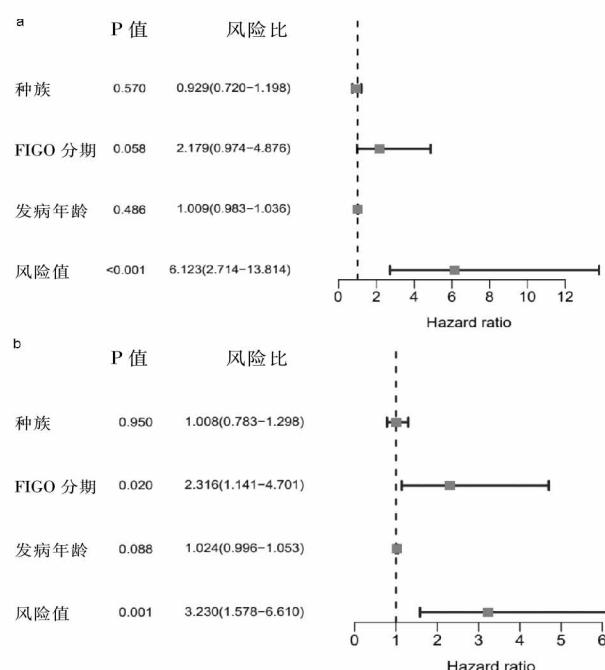


图 2 训练集(a)和验证集(b)多因素 Cox 分析结果

2.6 风险模型的评价

对风险值分组进行受试者工作特征曲线(receiver operating characteristic, ROC)检验,绘制模型的ROC曲线,计算ROC曲线下面积数值AUC以评估模型的预测效能,在训练集中该模型预测1年、3年和5年生存率的AUC分别为0.9、0.796和0.863,在验证集中该模型预测1年、3年和5年生存率的AUC分别为0.601、0.686和0.761(见图3,彩插2),训练集、验证集的生存率曲线下面积都在60%以上,说明能较准确地预测预后。

2.7 对miRNA基因使用mitarbase数据库进行预测

通过对这些miRNA对应的靶基因所在的基因通路、GO的生物过程ReactomePA通路进行富集分析发现这些靶基因主要富集于调控细胞的分化、信号的转导和凋亡等过程,例如:调节淋巴细胞和T细胞的分化,参与肽链苏氨酸磷酸化等;在KEGG通路富集分析中,靶基因主要富集在乳腺癌、肝癌、MAPK信号通路和AMPK信号通路等癌症相关通路中;另外,对疾病种类进行富集分析,发现靶基因富集与肿瘤相关疾病,其中hsa-miR-150-3p主要富集于结直肠癌、乳腺癌及多发性骨髓瘤等常见的肿瘤中。基因本体论、通路、疾病富集分析结果都表明这5种miRNA的靶基因与肿瘤密切相关,因此这5种miRNA可能是潜在的肿瘤预后标志物(见图4,彩插2)。

3 讨论

宫颈癌是世界范围内最常见的妇科恶性肿瘤之一,其发生率仅次于乳腺癌,居中国女性恶性肿瘤第二位^[11]。宫颈癌被认为是第一个病因明确的恶性肿瘤,几乎所有宫颈癌都与HPV感染有关^[12]。目前宫颈癌致病机制仍未被阐明,研究证明miRNA在生长发育、细胞增殖、凋亡等多种生物学过程中发挥重要调节作用,探索miRNA在宫颈癌发生、进展、治疗及预后等方面的作用,为寻找宫颈癌新的生物标记物提供新的研究方向。

1993年,Lee RC等^[13]在秀丽新小杆线虫发育过程中首次发现第1个miRNA,它是一类内源性的短的非编码miRNA,长度约为18~25个核苷酸,可以调节大约60%的蛋白质编码基因^[14],参与了大多数关键细胞过程中基因表达的转录后调节,如细胞增殖、分化、凋亡等的调节^[15]。miRNA的失调和许多人类疾病有关,特别是癌症,一项研究表明,超过50%的miRNA基因位于癌症相关的基因组区域或脆弱位点^[16],成熟的miRNA与mRNA通过碱基互补配对,发挥其调节靶基因表达的作用,miRNA的异常表达通常会促使癌症的进展^[17]。

近些年,随着对miRNA研究的深入,其在癌症的病因、进展、转移和预后的调节作用逐渐被人们重视,门婧睿等^[18]通过构建风险模型,最终确定hsa-miR132-3p、hsa-miR-139-5p、hsa-miR-3677-3p、hsa-miR-500A-3p可以

作为预测肝癌预后的生物标志物;miR-21已被证明与多个器官的恶性肿瘤有关,如在人类乳腺癌细胞系和组织中的表达增加,在发病的所有阶段都起着关键作用,有助于预测宫颈癌的发生^[19-20];在肾透明细胞癌中可以联合miRNA-223评估淋巴结转移、远处转移及预测肾透明细胞癌的生存^[9]。在乳腺癌的研究中还发现:miR-9、miR-10b和miR-17-5p可用于协助诊断;miR-148a和miR-335可评估乳腺癌的预后;miR-30c、miR-187和miR-339-5p则可以用来预测患者对特定治疗的反应^[21]。除此之外,miRNA在胃癌、胰腺癌、结直肠癌等中均有大量研究^[22-25]。

miRNA在宫颈癌中的研究也取得了一定的进展,有研究证明miR127可作为宫颈癌淋巴结转移的标志物,miR-199a可能是宫颈癌治疗的潜在靶点^[26];多种miRNA在宫颈癌中的表达异常可影响对化疗和放疗的敏感性^[27];本研究共筛选出5种miRNA:miR-101-3p、miR-532-5p、miR-150-3p、miR-378d、miR-378a-5p,这些miRNA在宫颈癌中已有相关研究。miR-101-3p在宫颈癌中的表达水平与肿瘤分期及中晚期宫颈癌同步放化疗抵抗相关,放疗抵抗组表达水平较放疗敏感组高^[28];miR-378已被证实在许多癌症中都有异常表达^[24,29],在宫颈疾病的发生发展中,随着宫颈病变的进展,miR-378发挥致癌基因的作用,表达水平增加,可作为早期诊断的生物标志物^[29-32];有研究证明miR-150能促进HeLa和SiHA细胞的增殖、迁移和侵袭,在癌组织中表达水平增高^[33];miR-532在肝癌和宫颈癌中均有异常表达^[34-35]。

本研究从TCGA数据库下载miRNA表达数据及患者临床信息,通过统计学及生物信息学方法单因素和多因素Cox回归分析、Lasso回归分析筛选出5种与宫颈癌预后相关的miRNA,并构建风险模型,该模型预测第1年、3年、5年生存率的AUC分别为:0.9、0.796、0.863,具有良好的预测性能。最后,本文作者对以上5种miRNA靶基因进行了预测,发现它们所在基因通路、GO、Reactome通路,均与肿瘤密切相关。

综上所述,本课题组构建的宫颈癌患者生存的风险模型,具有较好的敏感度和特异性,但本研究是基于在线数据库,预测模型还需要在实际的临床操作中进行验证,必要时需进一步通过实验的方式证实其有效性。miRNAs作为基因表达的重要调节因子,有望作为新型分子标志物应用于宫颈疾病的临床治疗及科学的研究中。

【参考文献】

- Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. CA:a Cancer Journal for Clinicians, 2021,71(3):209-249.
- 李海英,蔡竟,朱虹丽.宫颈癌的临床治疗研究进展 [J].宁夏医科大学学报,2020,42(4):419-423.

- [3] 王焜煜,王元景,孔为民.2020年美国放射肿瘤学会宫颈癌放射治疗指南解读 [J].中国临床医生杂志,2021,49(4):403-407.
- [4] Romano G, Veneziano D, Acunzo M, et al. Small non-coding RNA and cancer [J]. Carcinogenesis,2017,38(5):80.
- [5] Mollaei H, Safaralizadeh R, Rostami Z. MicroRNA replacement therapy in cancer [J]. J Cell Physiol,2019,234(8):12369-12384.
- [6] 梁文丽,刘植华.miRNA 在宫颈癌中的研究进展 [J].深圳中西医结合杂志,2020,30(21):196-198.
- [7] Acunzo M, Romano G, Wernicke D, et al. MicroRNA and cancer--a brief overview [J]. Adv Biol Regul,2015,57:1-9.
- [8] 汪海波.R 语言统计分析与应用 [M].北京:人民邮电出版社,2018.
- [9] Yu L, Xiang L, Feng J, et al. miRNA-21 and miRNA-223 expression signature as a predictor for lymph node metastasis, distant metastasis and survival in kidney renal clear cell carcinoma [J]. Journal of Cancer,2018,9(20):3651-3659.
- [10] Tibshirani R. The lasso method for variable selection in the Cox model [J]. Statistics in medicine,1997,16(4):385-395.
- [11] National Health Commission Of The People's Republic Of China. Chinese guidelines for diagnosis and treatment of cervical cancer 2018 (English version) [J]. 中国癌症研究(英文版),2019,31(2):295-305.
- [12] Schiffman M, Wentzensen N, Wacholder S, et al. Human papillomavirus testing in the prevention of cervical cancer [J]. Journal of the National Cancer Institute,2011,103(5):368-383.
- [13] Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14 [J]. Cell,1993,75(5):843-854.
- [14] Friedman RC, Farh KK, Burge CB, et al. Most mammalian mRNAs are conserved targets of microRNAs [J]. Genome research,2009,19(1):92-105.
- [15] Herranz H, Cohen SM. MicroRNAs and gene regulatory networks: managing the impact of noise in biological systems [J]. Genes & Development,2010,24(13):1339-1344.
- [16] Calin GA, Sevignani C, Dumitru CD, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. [J]. Proceedings of the National Academy of Sciences of the United States of America,2004,101(9):2999-3004.
- [17] Zhang B, Pan X, Cobb GP, et al. MicroRNAs as oncogenes and tumor suppressors. [J]. Developmental Biology, 2007, 302 (1): 1-12.
- [18] 门婧睿,谭建军,孙洪亮.肝癌预后 miRNA 风险评分模型的鉴定和分析(英文) [J].生物化学与生物物理进展,2020,47(4):344-360.
- [19] Shen S, Zhang S, Liu P, et al. Potential role of microRNAs in the treatment and diagnosis of cervical cancer [J]. Cancer Genetics, 2020, 248-249:25-30.
- [20] Park S, Eom K, Kim J, et al. MiR-9, miR-21, and miR-155 as potential biomarkers for HPV positive and negative [J]. BMC Cancer,2017,17(1):658.
- [21] 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.
- [22] 刘倩,王宗艳.胃癌患者血清 miRNA-647 水平及临床意义 [J].检验医学与临床,2021,18(7):874-875.
- [23] Zhang X, Chen H, Li H, et al. Long non-coding RNA small nucleolar RNA host gene 6 aggravates pancreatic cancer through upregulation of far upstream element binding protein 1 by sponging microRNA-26a-5p [J]. 中华医学杂志(英文版),2020,133(10):1211-1220.
- [24] Cigdem G, Hatice GA, Nihat D, et al. Novel miRNAs as potential biomarkers in stage II colon cancer: microarray analysis [J]. Molecular Biology Reports,2019,46(4):4175-4183.
- [25] Huang Q, Pan X. Prognostic lncRNAs, miRNAs, and mRNAs form a competing endogenous RNA network in colon cancer [J]. Frontiers in Oncology,2019,9:712.
- [26] Lee J, Choi CH, Choi J, et al. Altered MicroRNA expression in cervical carcinomas [J]. Clinical cancer research: an official journal of the American Association for Cancer Research,2008,14(9):2535-2542.
- [27] Phuah NH, In LLA, Azmi MN, et al. Alterations of microRNA expression patterns in human cervical carcinoma cells (Ca Ski) toward 1'-S-1-acetoxychavicol acetate and cisplatin [J]. Reproductive Sciences (Thousand Oaks, Calif.),2013,20(5):567-578.
- [28] 涂彬彬.miRNA 水平及临床病理因素预测中晚期宫颈鳞癌同步放化疗疗效的初步研究 [D].北京协和医学院,2014.
- [29] Li Z, Shen L, Li Y, et al. Clinical utility of microRNA-378 as early diagnostic biomarker of human cancers: a meta-analysis of diagnostic test [J]. Oncotarget,2016,7(36):58569-58578.
- [30] 齐文娟,许俊卿,刘俊霞.miR-378 在宫颈癌患者组织、外周血中的特征性表达及临床意义 [J].中国卫生检验杂志,2019,29(14):1724-1727.
- [31] Tan D, Zhou C, Han S, et al. MicroRNA-378 enhances migration and invasion in cervical cancer by directly targeting autophagy-related protein 12 [J]. Molecular Medicine Reports,2018,17(5):6319-6326.
- [32] Wang X, Wang H, Li Y, et al. microRNAs are biomarkers of oncogenic human papillomavirus infections [J]. Proceedings of the National Academy of Sciences of the United States of America, 2014, 111(11):4262-4267.
- [33] Zhang Z, Wang J, Li J, et al. MicroRNA-150 promotes cell proliferation, migration, and invasion of cervical cancer through targeting PDCD4 [J]. Biomedicine & Pharmacotherapy = Biomedecine & Pharmacotherapie,2018,97:511-517.
- [34] 朱慧,康天天,常卓,等.miR-532 抑制 Sema4C 逆转宫颈癌细胞上皮间质转化及增加对顺铂化疗的敏感性 [J].现代肿瘤医学,2020,28(19):3299-3305.
- [35] Yu B, Zhou S, Liang H, et al. Development and validation of a novel circulating miRNA-based diagnostic score for early detection of hepatocellular carcinoma [J]. Digestive Diseases and Sciences, 2021, 67(6):2283-2292.

(收稿日期:2022-03-17 实习编辑:牟术容)