

论著·基础研究

采用生物信息学方法筛选与肺腺癌患者预后相关的 LncRNAs^{*}

陶 冶¹, 李云慧², 黄俊杰¹, 梁 彬^{1△}

(1. 中国医科大学生命科学学院生物信息学教研室, 辽宁沈阳 110122;

2. 北部战区总医院和平院区检验科, 辽宁沈阳 110003)

摘要:目的 采用生物信息学的方法探讨差异表达长链非编码 RNA(lncRNAs)与肺腺癌患者预后的关系。方法 从国际肿瘤数据库(TCGA)下载肺腺癌组织及癌旁组织的 RNA 测序数据和患者的临床信息,采用 R 语言程序包筛选差异表达的 lncRNAs。应用 Kaplan-Meier 生存曲线(Log-Rank 检验)及 Cox 回归模型进行预后分析。结果 筛选出差异表达 lncRNAs 共计 168 个,其中上调的 lncRNAs 有 128 个,下调的 lncRNAs 有 40 个。单因素和多因素回归分析结果显示:RP11-490M8.1($P=0.047$)、RP11-132A1.4($P=0.049$)、DRAIC($P=0.023$)、LINC00942($P=0.033$)、TMPO-AS1($P=0.006$)、Z83851.4($P=0.032$)和 LINC01133($P=0.035$)是肺腺癌患者的独立预后因素。结论 该研究筛选出 7 个 lncRNAs,可作为肺腺癌患者预后的独立标志物,为临床医生对肺腺癌患者的预后判断提供帮助。

关键词:肺腺癌; 长链非编码 RNA; 生物信息学; 国际肿瘤数据库

DOI:10.3969/j.issn.1673-4130.2019.13.002

中图法分类号:R734.2

文章编号:1673-4130(2019)13-1541-04

文献标识码:A

Identification of prognosis-related lncRNAs in lung adenocarcinoma patients using bioinformatics method^{*}

TAO Ye¹, LI Yunhui², HUANG Junjie¹, LIANG Bin^{1△}

(1. Department of Bioinformatics, Life Science College, China Medical University,

Shenyang, Liaoning 110122, China; 2. Department of Clinical Laboratory, Peace District of General

Hospital of Northern War Zone, Shenyang, Liaoning 110003, China)

Abstract: Objective To evaluate the relationship between differential expression of lncRNAs and prognosis of patients with lung adenocarcinoma by bioinformatics. **Methods** The lncRNAs sequencing data in lung adenocarcinoma (LUAD) tissues and adjacent normal tissues, and related clinical information were downloaded from the Cancer Genome Atlas (TCGA) database. The R language packages were used to screen the differentially expressed lncRNAs. The Kaplan-Meier curve (Log-Rank test) and Cox regression model were adopted for prognosis analysis. **Results** A total of 168 lncRNAs were identified, including 128 up-regulated lncRNAs and 40 down-regulated lncRNAs. Univariate and multivariate Cox regression analysis indicated that, RP11-490M8.1($P=0.047$), RP11-132A1.4($P=0.049$), DRAIC($P=0.023$), LINC00942($P=0.033$), TMPO-AS1($P=0.006$), Z83851.4($P=0.032$) and LINC01133($P=0.035$) were independent prognostic factors in LUAD patients. **Conclusion** In this study, seven lncRNAs were screened as independent prognostic markers for lung adenocarcinoma patients, which could be helpful for clinicians to judge the prognosis of lung adenocarcinoma patients.

Key words: lung adenocarcinoma; lncRNAs; bioinformatics; The Cancer Genome Atlas

据 2018 年 185 个国家的统计结果显示,肺癌占所有癌症病例的 11.6%,占癌症总死亡人数的 18.4%,位列我国癌症发病率之首^[1]。非小细胞肺癌是肺癌常见形式,包括 3 种组织病理学亚型:腺癌、鳞状细胞癌和大细胞癌。肺腺癌是最常见的变异型,约

占肺癌病例的 40.0%^[2-3]。目前,虽然在肺癌手术、化疗和分子靶向治疗技术方面取得了进展,但肺癌仍然是全球癌症死亡的主要原因^[3-4]。长链非编码 RNA(lncRNAs)是大于 200 bp 的非编码 RNA,主要通过表观遗传学、转录及转录后水平调控基因表达^[5]。研

^{*} 基金项目:国家自然科学基金项目(81301835)。

作者简介:陶冶,女,研究生,主要从事肿瘤基因组的生物信息学分析。△ 通信作者, E-mail: bliang@cmu.edu.cn。

本文引用格式:陶冶,李云慧,黄俊杰,等.采用生物信息学方法筛选与肺腺癌患者预后相关的 LncRNAs[J].国际检验医学杂志,2019,40(13):1541-1544.

究结果显示, lncRNAs 是重要的调控因子, 调控细胞增殖、分化、侵袭以及转移^[6]。本研究利用 TCGA 中肺腺癌患者的 RNA 测序数据, 采用生物信息学技术手段, 筛选肺腺癌患者差异表达的 lncRNAs, 探讨差异表达的 lncRNAs 与肺腺癌临床参数的相关性 及与肺腺癌患者预后的关系。

1 资料与方法

1.1 数据下载 从 TCGA 数据库 (https://cancergenome.nih.gov/) 下载肺腺癌患者癌组织和癌旁组织的 lncRNAs 测序数据, cBioPortal 数据库 (http://www.cbioportal.org/) 下载相应的临床资料。其中男 219 例, 年龄 38~88 岁, 平均 (65.66±9.77) 岁; 女 259 例, 年龄 40~87 岁, 平均 (65.75±9.72) 岁。样本的纳入标准: (1) 经病理证实为原发性肺腺癌, 术前未经过任何放化疗; (2) 具有完整的临床分期资料及预后信息。共计 478 例肺腺癌患者纳入分析, 临床特点见表 1。本研究中肺腺癌的分期采用 TNM 分期方法, 其中 T 分期与肿瘤浸润深度相关, 分期越高, 浸润深度越大; N 分期与淋巴结转移情况相关, 分期越高, 转移淋巴结越多; M 分期与肿瘤远处转移相关。肺腺癌患者在年龄和性别上差异无统计学意义, 具有可比性。

表 1 肺腺癌患者的临床特点		
项目	n	百分比 (%)
性别		
女	259	54.2
男	219	45.8
年龄		
<60 岁	123	25.7
≥60 岁	346	72.4
NA	9	1.9
临床分期		
I+II	368	77.0
III+IV	102	21.3
NA	8	1.7
T 分期		
T1+T2	416	87.0
T3+T4	60	12.6
TX	2	0.4
N 分期		
N0	308	64.4
N1~N3	159	33.3
NX	11	2.3
M 分期		
M0	327	97.6
M1	23	4.8
MX	128	26.8

注: NA 表示无法确定; TX 表示原发肿瘤的情况无法评估; NX 表示区域淋巴结情况无法评估; MX 表示肿瘤有无远处转移无法评估

1.2 数据处理 应用 R 语言程序包对 TCGA 数据库下载的 lncRNAs 数据信息进行差异分析, 筛选肺

腺癌组织和癌旁组织中差异表达的 lncRNAs。筛选标准: $\log_2$ | 差异倍数 (FC) ≥ 1 , $P < 0.05$ 。

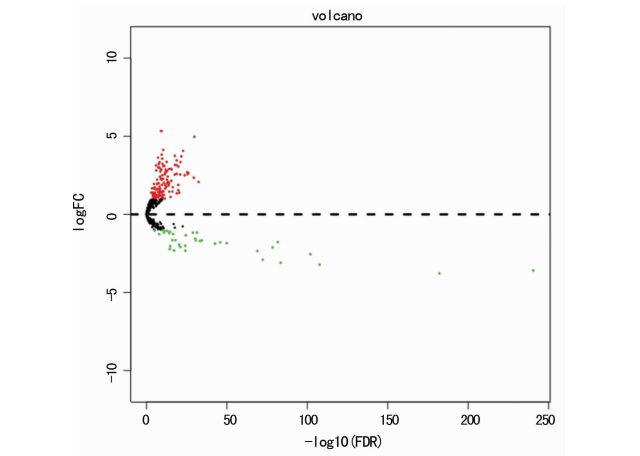
1.3 筛选与肺腺癌患者预后相关的 lncRNAs 对肺腺癌患者癌组织中多个差异表达的 lncRNAs 进行分组, 根据每个 lncRNA 的测序数据中值分成高表达组和低表达组。应用 Kaplan-Meier 生存曲线和 Log-Rank 检验方法分别对高表达和低表达的 lncRNA 与总体生存期 (OS) 的相关性进行分析, 筛选出与总体生存期显著相关的 lncRNAs。应用单因素和多因素 Cox 回归模型对与 OS 相关的 lncRNAs 行进一步分析。

1.4 差异表达 lncRNAs 与临床参数的关系 根据患者年龄、TNM 分期、总体生存状态进行分组, 分析与 OS 相关的 lncRNAs 与各临床参数的关系。

1.5 统计学处理 采用 SPSS20.0 统计学软件包进行统计学分析。计数资料采用率 (%) 表示。应用 Kaplan-Meier 曲线法进行生存分析, Log-Rank 方法检验各 lncRNA 高表达和低表达组生存率差异是否具有统计学意义。多因素预后分析采用 Cox 比例风险回归模型。lncRNA 表达和临床参数之间的关系采用 χ^2 检验。 $P < 0.05$ 表示差异有统计学意义。

2 结果

2.1 差异表达 lncRNAs 的筛选 应用 R 语言程序包筛选出癌组织和癌旁组织差异表达的 lncRNAs 有 168 个, 其中上调的 lncRNA 有 128 个, 下调的 lncRNAs 有 40 个。癌组织和癌旁组织中差异表达 lncRNAs 的分布见图 1。



注: FDR 表示错误发现率; FC 表示差异倍数; 红色表示上调 lncRNAs, 绿色表示下调 lncRNAs

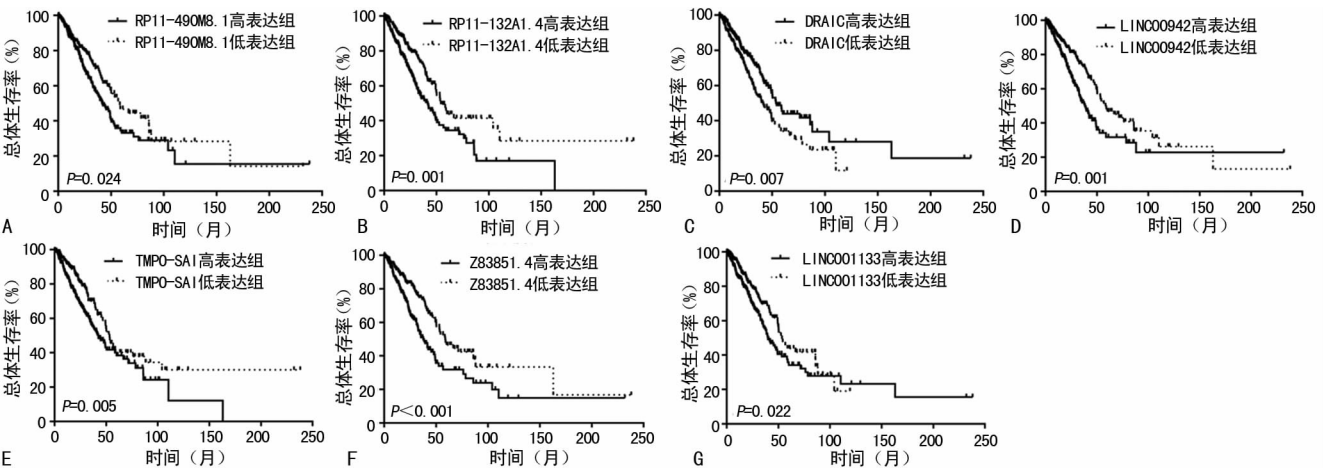
图 1 癌组织和癌旁组织差异表达的 lncRNAs 分布火山图

2.2 与 OS 相关的 lncRNAs 筛选 根据每个 lncRNA 的表达中值, 分成高表达组和低表达组。采用 Kaplan-Meier 生存曲线和 Log-Rank 检验分析 168 个差异表达的 lncRNAs 与肺腺癌患者预后的关系。结果显示, RP11-490M8.1 ($P = 0.024$)、RP11-132A1.4 ($P = 0.001$)、DRAIC ($P = 0.007$)、LINC00942 ($P = 0.001$)、TMPO-AS1 ($P = 0.005$)、Z83851.4 ($P < 0.001$) 和 LINC01133 ($P = 0.022$) 与

肺腺癌患者的预后呈显著相关。高表达 RP11-490M8. 1、RP11-132A1. 4、LINC00942、TMPO-AS1、Z83851. 4 和 LINC01133 与预后差显著相关,而低表达 DRAIC 与预后差显著相关。见图 2。

采用单因素和多因素 Cox 回归分析结果进一步证实,差异表达的 lncRNAs 与肺腺癌患者的预后相关,高表

达 RP11-490M8. 1 ($P=0.047$)、RP11-132A1. 4 ($P=0.049$)、LINC00942($P=0.033$)、TMPO-AS1($P=0.006$)、Z83851. 4($P=0.032$)、LINC01133($P=0.035$)和低表达 DRAIC($P=0.023$)是肺腺癌患者预后不良的独立预后因素。见表 2。



注:A 为 RP11-490M8. 1;B 为 RP11-132A1. 4;C 为 DRAIC;D 为 LINC00942;E 为 TMPO-AS1;F 为 Z83851. 4;G 为 LINC01133

图 2 与预后显著相关的 lncRNAs 的生存曲线。

表 2 肺腺癌患者差异表达 lncRNAs 的单因素和多因素 Cox 回归分析

参数	组别比较	单因素			多因素		
		HR	95% CI	P	HR	95% CI	P
年龄	≥60 vs. <60 岁	1.102	0.774~1.568	0.590			
性别	男 vs. 女	1.094	0.809~1.478	0.560			
远处转移	有 vs. 无	1.993	1.143~3.473	0.015			
淋巴结转移	有 vs. 无	2.588	1.909~3.509	<0.001	2.003	1.353~2.966	0.001
临床分期	Ⅲ+Ⅳ vs. Ⅰ+Ⅱ	2.593	1.885~3.564	<0.001			
T 分期	T3+T4 vs. T1+T2	2.215	1.494~3.282	<0.001	1.685	1.045~2.717	0.032
RP11-490M8. 1	高表达组 vs. 低表达组	1.421	1.047~1.930	0.024	1.478	1.005~2.176	0.047
RP11-132A1. 4	高表达组 vs. 低表达组	1.709	1.258~2.321	0.001	1.492	1.002~2.220	0.049
DRAIC	高表达组 vs. 低表达组	0.658	0.486~0.892	0.007	0.643	0.439~0.941	0.023
LINC00942	高表达组 vs. 低表达组	1.708	1.257~2.321	0.001	1.530	1.034~2.264	0.033
TMPO-AS1	高表达组 vs. 低表达组	1.548	1.139~2.105	0.005	1.742	1.170~2.594	0.006
Z83851. 4	高表达组 vs. 低表达组	1.798	1.323~2.443	<0.001	1.590	1.041~2.428	0.032
LINC01133	高表达组 vs. 低表达组	1.426	1.051~1.936	0.023	1.534	1.031~2.283	0.035

注:HR 表示风险比值

2.3 与预后显著相关的 lncRNAs 和临床参数的关系 对与预后显著相关的 7 个 lncRNAs 和各临床参数之间的关系进行分析,结果显示,RP11-132A1. 4 表达水平与年龄($P=0.016$)、淋巴结转移($P=0.016$)、临床分期($P<0.001$)和 T 分期($P=0.001$)显著相关;DRAIC 表达水平与年龄($P=0.027$)显著相关;LINC00942 表达水平与淋巴结转移($P<0.001$)和临床分期($P=0.001$)显著相关;TMPO-AS1 表达水平与临床分期($P=0.008$)显著相关;Z83851. 4 表达水

平与淋巴结转移($P=0.008$)显著相关;LINC01133 与 T 分期($P=0.049$)显著相关。见表 2。

3 讨 论

lncRNAs 是一种长度大于 200 个核苷酸,缺乏完整的特异性开放阅读框,无编码蛋白质功能的核酸分子。大量研究结果显示 lncRNAs 通过染色质重塑,组蛋白修饰和 RNA 代谢等多种生物学过程在表观遗传、转录和转录后水平调控基因表达^[7-8]。lncRNA 发挥内源性“miRNA 海绵”功能与 miRNA 相互作用,

参与靶基因的表达调控;反之,miRNA 通过 RNA 诱导沉默复合物调控 lncRNAs 发挥生物学功能,两者又可竞争结合 mRNAs,共同参与疾病的发生发展过程^[9]。

lncRNAs 与多种癌症相关,其异常表达和突变与肿瘤的增殖、凋亡、侵袭、迁移及药物敏感性密切相关。研究结果显示,在多种肿瘤患者中,某些 lncRNAs 表达发生改变,可作为癌症诊断及预后的标志物^[10]。TCGA 癌症数据库是由美国国家癌症研究所及国家人类基因组研究所联合建立,其中包含丰富的数据类型和肿瘤类型。本研究采用生物信息学技术手段,筛选肺腺癌患者差异表达的 lncRNAs 168 个进行预后分析,总体生存分析显示:RP11-490M8. 1、RP11-132A1. 4、DRAIC、LINC00942、TMPO-AS1、Z83851. 4 和 LINC01133 是肺腺癌患者独立预后的不良因素。YANG 等^[11]报道 LINC01133 通过充当 miR-106a-3p 的 ceRNA 来调控 APC 的表达和 Wnt/ β -catenin 通路,抑制胃癌的进展和转移。也有研究证明,LINC01133 低表达影响结直肠癌的转移和预后^[12]。ZANG 等^[13]报道敲除 LINC01133 减少非小细胞肺癌肿瘤细胞的增殖、迁移和侵袭以及诱导细胞凋亡发挥抑瘤作用。也有研究发现 LINC01133 是食管鳞状细胞癌预后的生物标志物和治疗靶点^[14]。HUANG 等^[15]报道,在前列腺癌细胞中 TMPO-AS1 过度表达促进细胞周期的进展、肿瘤细胞迁移和增殖,抑制细胞凋亡。也有研究证明,DRAIC/PCAT29 的表达水平与前列腺癌的发生和进展有关^[16]。其他的 lncRNAs 目前还未有报道。

此外,本研究发现 RP11-132A1. 4、LINC00942 和 Z83851. 4 与淋巴结转移显著相关,RP11-132A1. 4、LINC00942 和 TMPO-AS1 与临床分期显著相关,RP11-132A1. 4 和 LINC01133 与 T 分期显著相关,提示上述 lncRNAs 可能参与肺腺癌的发生发展过程,在肺腺癌恶性进程中发挥重要作用。

4 结 论

本研究对 TCGA 数据库中肺腺癌患者大数据进行分析,筛选出与肺腺癌预后显著相关的 lncRNAs: RP11-490M8. 1、RP11-132A1. 4、DRAIC、LINC00942、TMPO-AS1、Z83851. 4 和 LINC01133。本研究结果提示上述 lncRNAs 分别与各临床参数存在显著相关性,可能为肺腺癌机制的研究提供理论依据,为寻找肺腺癌诊断和预后新的分子标志物以及个性化治疗方案提供新的思路。

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(8): 394-424.

[2] WEI S, ZHANG Z Y, FU S L, et al. Hsa-miR-623 suppresses tumor progression in human lung adenocarcinoma [J]. Cell Death Dis, 2017, 8(5): e2829.

[3] SIEGEL R L, MILLER K D, JEMAL A. Cancer statistics, 2016[J]. CA Cancer J Clin, 2016, 66(1): 7-30.

[4] DESANTIS C E, LIN C C, MARIOTTO A B, et al. Cancer treatment and survivorship statistics [J]. 2014, CA Cancer J Clin, 2014, 64(4): 252-271.

[5] LIU Y, SHARMA S, WATABE K, et al. Roles of lncRNA in breast cancer [J]. Front Biosci (Schol Ed), 2015, 1(7): 94-108.

[6] TERASHIMA M, TANGE S, ISHIMURA A, et al. MEG3 long noncoding RNA contributes to the epigenetic regulation of epithelial-mesenchymal transition in lung cancer cell lines [J]. J Biol Chem, 2017, 292(1): 82-99.

[7] HAMILTON M J, YOUNG M D, SAUER S, et al. The interplay of long non-coding RNAs and MYC in cancer [J]. AIMS Biophys, 2015, 2(4): 794-809.

[8] HUANG F T, CHEN W Y, GU Z Q, et al. The novel long intergenic noncoding RNA UCC promotes colorectal cancer progression by sponging miR-143 [J]. Cell Death Dis, 2017, 8(5): e2778.

[9] YOON J H, ABDELMOHSEN K, SRIKANTAN S, et al. LincRNA-p21 suppresses target mRNA translation [J]. Mol Cell, 2012, 47(4): 648-655.

[10] 席姣姣, 颜景斌. 长链非编码 RNA 在肺癌中的研究进展 [J]. 生命科学, 2018, 30(1): 48-55.

[11] YANG X Z, CHENG T T, HE Q J, et al. LINC01133 as ceRNA inhibits gastric cancer progression by sponging miR-106a-3p to regulate APC expression and the Wnt/ β -catenin pathway [J]. Mol Cancer, 2018, 17(1): 126.

[12] ZHANG J H, LI A Y, WEI N. Downregulation of long non-coding RNA LINC01133 is predictive of poor prognosis in colorectal cancer patients [J]. Eur Rev Med Pharmacol Sci, 2017, 21(9): 2103-2107.

[13] ZANG C, NIE F Q, WANG Q, et al. Long non-coding RNA LINC01133 represses KLF2, P21 and E-cadherin transcription through binding with EZH2, LSD1 in non small cell lung cancer [J]. Oncotarget, 2016, 7(10): 11696-11707.

[14] YANG X Z, HE Q J, CHENG T T, et al. Predictive value of LINC01133 for unfavorable prognosis was impacted by alcohol in esophageal squamous cell carcinoma [J]. Cell Physiol Biochem, 2018, 48(1): 251-262.

[15] HUANG W, SU X, YAN W, et al. Overexpression of AR-regulated lncRNA TMPO-AS1 correlates with tumor progression and poor prognosis in prostate cancer [J]. Prostate, 2018, 78(16): 1248-1261.

[16] SAKURAI K, REON B J, ANAYA J, et al. The lncRNA DRAIC/PCAT29 locus constitutes a tumor-suppressive nexus [J]. Mol Cancer Res, 2015, 13(5): 828-838.