

基于 TCGA 数据库肺腺癌 RNAs 构建 ceRNA 网络的综合分析

唐怀慧¹, 王忠帅², 邵茜茜³, 杨青山¹

(1. 锦州医科大学附属第一医院, 辽宁 锦州 121000;

2. 鞍山市肿瘤医院, 辽宁 鞍山 114000;

3. 锦州医科大学附属第三医院, 辽宁 锦州 121000)

摘要:目的 基于 TCGA 数据库挖掘肺腺癌差异表达的 RNAs, 通过 ceRNA 网络的综合分析并预测其与患者预后的相关性。方法 利用 TCGA 数据库下载肺腺癌的表达数据, 使用 R 软件“edgeR”包初步筛选差异表达的 RNAs, 进一步构建肺腺癌的 lncRNA-miRNA-mRNA 的 ceRNA 网络, 对 ceRNA 网络中的 mRNAs 进行了 K-M 生存分析。结果 筛选的差异表达的 RNAs 中有 3 个 mRNA (ANLN、IGFBP1、TFAP2A)、4 个 lncRNA (AC015923.1、FGF12-AS2、LINC00211、MED4-AS1)、2 个 miRNA (hsa-mir-31、hsa-mir-490), 均与预后相关。其中 hsa-mir-31 预后价值最高 ($P < 0.001$), LINC00461 和 has-mir-139 为关键节点的 ceRNA 调控网络。结论 本次筛选出的 3 个 mRNA、4 个 lncRNA 和 2 个 miRNA 可作为肺腺癌的新候选预后因子, 且 LINC00461 和 has-mir-139 是肺腺癌 ceRNA 中两个重要的调控网络节点。

关键词: TCGA; 肺腺癌; ceRNA 网络; 预后

中图分类号: R734.2

文献标识码: A

DOI: 10.3969/j.issn.1006-1959.2020.07.025

文章编号: 1006-1959(2020)07-0090-04

Comprehensive Analysis of Constructing ceRNA Network Based on TCGA Database of Lung Adenocarcinoma RNAs

TANG Huai-hui¹, WANG Zhong-shuai², SHAO Qian-qian³, YANG Qing-shan¹

(1. The First Affiliated Hospital of Jinzhou Medical University, Jinzhou 121000, Liaoning, China;

2. Anshan Cancer Hospital, Anshan 114000, Liaoning, China;

3. The Third Affiliated Hospital of Jinzhou Medical University, Jinzhou 121000, Liaoning, China)

Abstract: Objective To explore the differentially expressed RNAs of lung adenocarcinoma based on TCGA database, comprehensively analyze and predict the correlation with the prognosis of patients through ceRNA network. Methods The TCGA database was used to download the expression data of lung adenocarcinoma, Use the R software “edgeR” package to preliminarily screen differentially expressed RNAs to further construct the ceRNA network of lncRNA-miRNA-mRNA of lung adenocarcinoma, K-M survival analysis was performed on the mRNAs in the ceRNA network. Results The screened differentially expressed RNAs included 3 mRNA (ANLN, IGFBP1, TFAP2A), 4 lncRNA (AC015923.1, FGF12-AS2, LINC00211, MED4-AS1), and 2 miRNA (hsa-mir-31, hsa-mir-490), both related to prognosis. Among them, hsa-mir-31 has the highest prognostic value ($P < 0.001$), and LINC00461 and has-mir-139 are key node ceRNA regulatory networks. Conclusion The 3 mRNA, 4 lncRNA and 2 miRNA selected this time can be used as new prognostic factors for lung adenocarcinoma, and LINC00461 and has-mir-139 are two important regulatory network nodes in ceRNA of lung adenocarcinoma.

Key words: TCGA; Lung adenocarcinoma; ceRNA network; Prognosis

肺腺癌(lung adenocarcinoma)是癌症死亡的重要原因,目前其发病率和死亡率仍在不断加^[1]。男性肺癌患者发病率和死亡率在所有恶性肿瘤中占第 1 位,女性占第 2 位^[2]。非小细胞肺癌(NSCLC)是肺癌中最常见的类型,其中以腺癌和鳞癌最常见,并且腺癌已经超过了鳞癌^[3]。越来越多的研究证实, lncRNA 可以作为 miRNA 和 mRNA 的竞争平台,在乳腺癌、胃癌、肝癌、肺癌及结直肠癌等多种恶性肿瘤的细胞周期和细胞死亡调控中起着重要的调节作用,影响肿瘤的侵袭和迁移,从而在肿瘤的发生、发展过程中发挥重要作用^[4]。癌症基因组图谱(TCGA)是一个公共资助的项目,该项目可提供公开的癌症数据集,以帮助改进诊断方法、治疗标准,并最终预防癌症^[5]。

本研究利用高通量测序技术和生物信息学分析方法对肺腺癌的表达数据进行分析,为肺腺癌患者筛选新候选预后因子,以期对肺腺癌预后判断提供新的思路。

1 资料与方法

1.1 资料来源 利用 TCGA 数据库(<https://portal.gdc.cancer.gov/>)获取 594 例肺腺癌患者的 mRNAs、lncRNAs 和 miRNAs 表达数据(包括癌组织 535 例,癌旁正常组织 59 例),并收集患者的临床信息。基因表达量用于差异基因分析和构建 ceRNA 网络,患者的临床信息用于生存分析。

1.2 方法

1.2.1 筛选差异基因 基于 TCGA 数据库,利用 R 软件“TCGA biolinks”包获取肺腺癌的表达数据,perl 语言提取 Ensembl 数据库鉴定了的 lncRNAs 和 mRNAs。利用 R 软件“edgeR”包,除去所有表达量低的 RNA(平均读数 ≤ 2)。设置差异倍数(fold change,

作者简介:唐怀慧(1992.5-),女,安徽宿州人,硕士研究生,主要从事放射肿瘤学研究

通讯作者:杨青山(1975.11-),男,辽宁抚顺人,博士,副主任医师,主要从事肺癌、头颈部恶性肿瘤等的放化疗和综合治疗

FC) ≥ 2 , 差异的显著性 (fdr: 调整 P 值) < 0.01 。将肺腺癌组织的表达数据与正常组织的表达数据进行差异分析, 并利用 R 软件“gplots”包绘制火山图。

1.2.2 构建 lncRNA-miRNA-mRNA 的 ceRNA 网络 提取筛选出的差异表达 RNAs, 利用 miRcode (<http://www.mircode.org/>) 数据库预测差异 lncRNAs 和差异 miRNAs 之间的相互作用。miRDB (<http://mirdb.org/>), mirTarBase (<http://www.mirbase.org/>), miRanda (<http://www.microrna.org/microrna/home.do>) 和 Target Scan (<http://www.targetscan.org/>) 数据库用于检索差异 miRNAs 的靶基因 mRNAs。利用 R 软件“VennDiagram”包绘制 Venn 图, 并利用 Cytoscape v3.7.1 构建肺腺癌 lncRNA-miRNA-mRNA 的 ceRNA 网络。

1.2.3 生存分析 基于 TCGA 数据库下载的肺腺癌患者的临床信息, 提取有效的生存状态及生存时间, 通过构建 Lasso 回归模型和 Cox 风险回归分析, 利用 R 软件“survival”包进行 K-M 生存分析, 设置 $P < 0.05$ 。筛选出的差异 RNAs 认为与预后具有相关性。

2 结果

2.1 差异表达的 RNAs 共筛选出了 2551 个差异表达的 mRNA, 其中高表达的 2033 个, 低表达 518 个; 1359 个差异表达的 lncRNA, 高表达的 1185 个, 低表达的 174 个; 99 个差异表达的 miRNA, 高表达的 78 个, 低表达 21 个。利用 R 软件“gplots”包绘制火山图, 见图 1。

2.2 肺腺癌 lncRNA-miRNA-mRNA ceRNA 网络

通过 miRcode 数据库查找与差异表达的 miRNA 相关的差异 lncRNA 有 100 个, 而与 100 个差异 lncRNA 对应的差异 miRNA 有 43 个, 筛选出四个数据库共有的基因有 4045 个, 见图 2A。并与提取出的差异表达 mRNA 取交集, 最终获得与肺腺癌有关的差异 mRNA 有 294 个, 见图 2B。对筛选出的 100 个 lncRNA、43 个 miRNA 和 294 个 mRNA 基因, 利用 Cytoscape v3.7.1 绘制 lncRNA-miRNA-mRNA 的 ceRNA 网络, 见图 3A、图 3B, 并提取了以 LINC00461 和 has-mir-139 为关键节点的核心子网络, 见图 3C、图 3D。

2.3 生存分析 ①单因素 Cox 风险回归分析共筛选出对肺腺癌预后可能具有影响作用的有 144 个 mRNA, 29 个 lncRNA, 5 个 miRNA。②采用 LASSO 回归鉴定预后相关的 RNAs, Lasso 回归模型拟合共筛选出 7 个 mRNA, 18 个 lncRNA 和 5 个 miRNA。③多因素 Cox 风险回归分析共发现 3 个 mRNA: ANLN、IGFBP1、TFAP2A; 7 个 lncRNA: AC015923.1、AC017048.3、FGF12-AS2、HOXA11-AS、LINC00211、LINC00319、MED4-AS1; 3 个 miRNA: hsa-mir-142、hsa-mir-31、hsa-mir-490, 可能与预后相关, 其中 hsa-mir-31 具有强的预后价值, 见图 4。④Kaplan-Meier 生存分析共发现 3 个 mRNAs (ANLN、IGFBP1、TFAP2A), 4 个 lncRNAs (AC015923.1、FGF12-AS2、LINC00211、MED4-AS1) 和 2 个 miRNAs (hsa-mir-31、hsa-mir-490) 具有预后性, 见图 5。

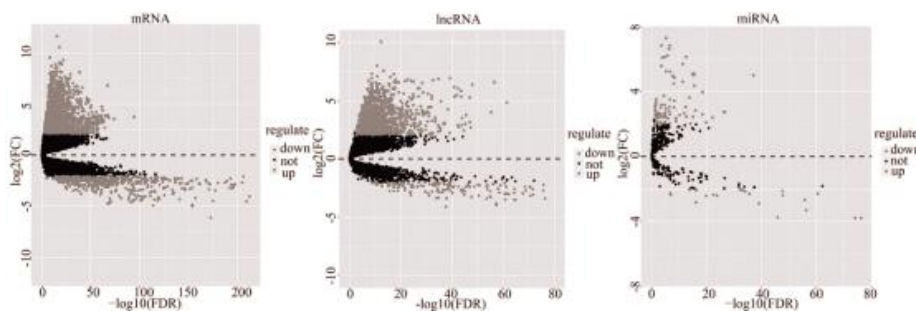
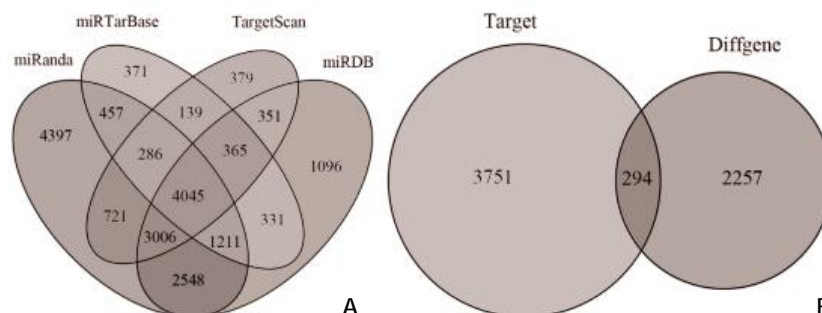
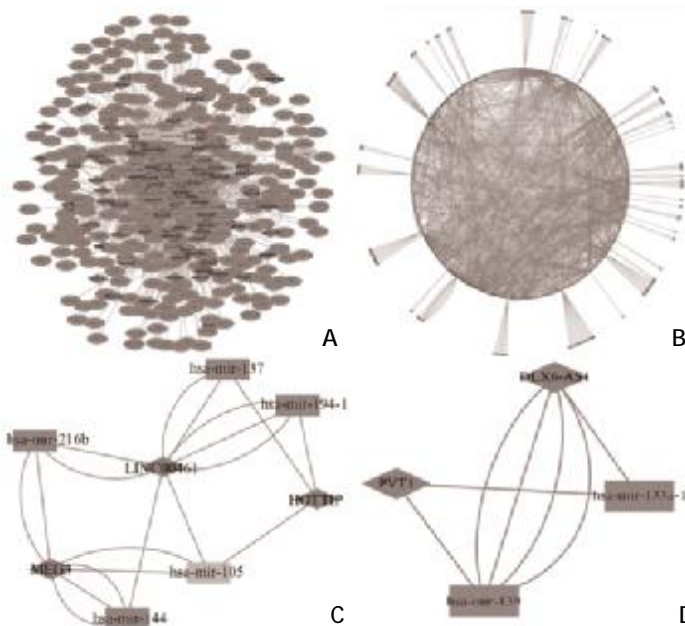


图 1 肺腺癌差异基因火山图



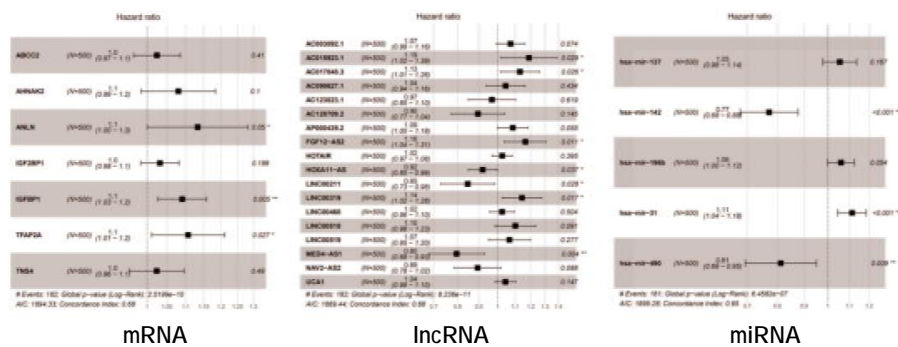
注: A: 4 个数据库预测靶基因交集的 Venn 图; B: 取 4 个数据库预测的靶基因与差异 mRNA 交集的 Venn 图, 其中 target 是指 4 个数据库交集部分的 mRNAs, diffgene 是筛选出的差异表达的 mRNAs

图 2 Venn 图



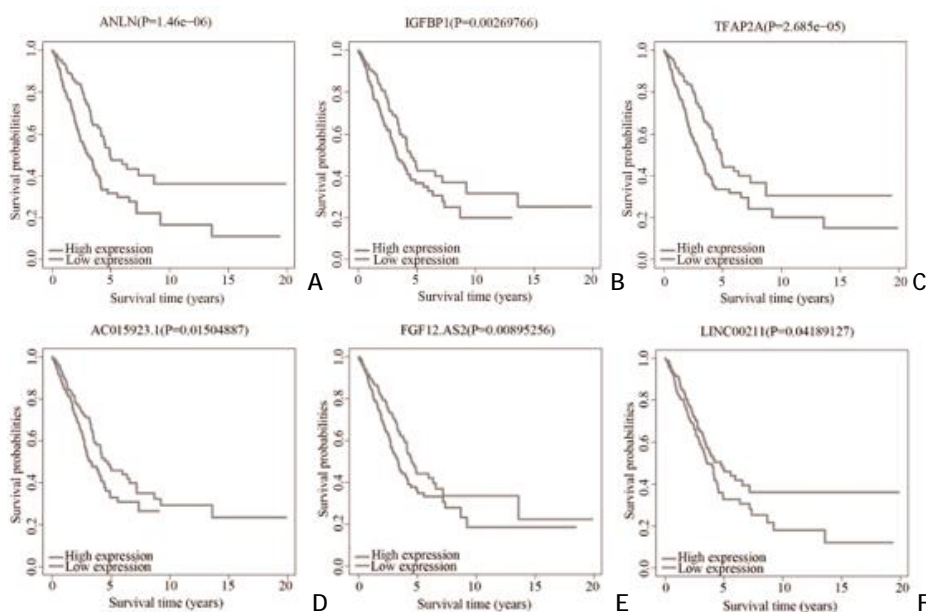
注:A:构建出的肺腺癌的 lncRNA-miRNA-mRNA 的 ceRNA 网络;B:肺腺癌 lncRNA-miRNA-mRNA 的 ceRNA 网络的圆形布局,C、D:以 LINC00461 和 has-mir-139 为关键节点的子网络

图 3 肺腺癌 lncRNA-miRNA-mRNA 的 ceRNA 网络



注:森林图中第一列表示基因名字;第二列表示纳入样本数;第三列表示 OR 点和估计 OR 的 95%CI 上下限;第四列显示 OR(95%CI);第五列表示 P-Value, $P < 0.05$ 认为有统计学意义,* 越多代表预后相关性越强

图 4 ceRNA 网络中与 OS 相关的 RNAs 的森林图



注:A、B、C:3 个 mRNAs(ANLN,IGFBP1,TFAP2A)的生存曲线;D、E、F、G:4 个 lncRNAs(AC01923.1,FGF12-AS2,LINC00211,MED4-AS1)的生存曲线;H、I:2 个 miRNAs(hsa-mir-31,hsa-mir-490)的生存曲线

图 5 与 OS 相关的 RNAs 的 K-M 生存曲线图

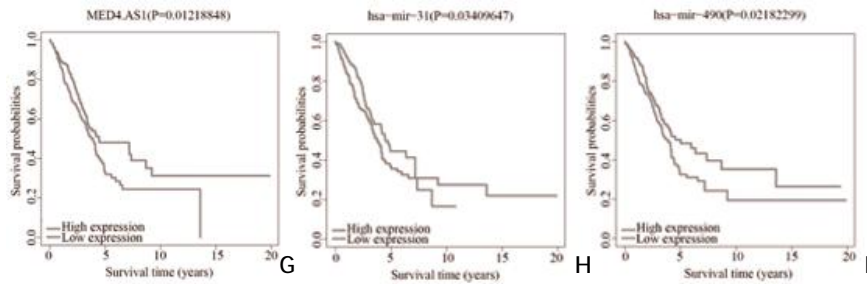


图 5(续)

3 讨论

随着医疗技术的不断发展与创新,近年来对肺癌的治疗已取得了较大进展,肺癌患者的生存率有了一定的改善,但其治疗的疗效仍不尽如人意,肺癌患者的预后仍然很差,其 5 年总体生存率仍低于 15%^[6]。因此,寻找新的肺癌治疗策略已成为当前研究的热点。近年来高通量测序技术和生物信息学分析方法的不断完善,开辟了研究肺癌的新思路和新方法。TCGA 是一个公共数据库,包含有关癌症中癌基因表达的全面数据集。本研究从 TCGA 数据库中获取了肺癌的表达数据,用以构建 lncRNA-miRNA-mRNA 的 ceRNA 网络,并分析参与 ceRNA 网络中的 RNAs 的预后相关性。

目前,越来越多的研究已经证明 ceRNA 涉及多种的肿瘤发生发展过程,其在该过程中具有重要生物学功能。如 ceRNA 网络的构建揭示了直肠癌中潜在的 lncRNA 生物标志物^[7]。FOXO1 3'UTR 通过调节 miRNA 活性在抑制乳腺癌细胞转移中起 ceRNA 的作用^[8]。长的非编码 RNA Unigene56159 通过在肝细胞癌中充当 miR-140-5p 的 ceRNA 来促进上皮-间质转化^[9]。随着对 ceRNA 网络的不断深入开发和挖掘,本研究前期发现已不能将 miRNA、lncRNA 或 mRNA 视为在肿瘤中的独立影响元素。在对 miRNA 的表达进行调控之后,由 miRNA 调节的靶基因 mRNA 的表达也随之改变,最终导致了多种疾病(包括肿瘤在内)的发生。由于癌症的发生和发展机制过于错综复杂,单基因或单一途径的作用已非常有限。因此,系统构建和分析 ceRNA 网络可以提供更有针对性的研究领域和揭示癌症潜在机制的新视角。

本研究共筛选出 2551 个差异表达的 mRNA, 1359 个差异表达的 lncRNA 和 99 个差异表达的 miRNA, 通过生物信息学数据库筛选出与差异表达的 miRNA 相关的差异 lncRNA 有 100 个, 而与 100 个差异 lncRNA 对应的差异 miRNA 有 43 个, 检索和提取出与 43 个差异 miRNA 相关的差异 mRNAs 有 294 个, 构建了肺癌的 ceRNA 网络, 发现了以 LINC00461 和 has-mir-139 为关键节点的 ceRNA 调控子网络。通过对 ceRNA 网络中涉及的差异 RNAs 进行单因素 Cox 风险回归分析、构建 LASSO

回归模型以及多因素 Cox 风险回归分析, 共发现 3 个 mRNA, 7 个 lncRNA 和 3 个 miRNA 可能具有预后相关性, 最终通过 K-M 生存分析绘制生存曲线发现 3 种 mRNA, 4 种 lncRNA, 2 种 miRNA 具有预后相关性, 且 hsa-mir-31 显示出最大的预后价值。

总之, 通过构建肺癌的 lncRNA-miRNA-mRNA 的 ceRNA 网络, 本次筛选出的 3 个 mRNA、4 个 lncRNA 和 2 个 miRNA 可作为肺癌的新候选预后因子, 且 LINC00461 和 has-mir-139 是肺癌 ceRNA 中两个重要的调控网络节点。

参考文献:

- [1]Liu H,Zhou G,Fu X,et al.Long noncoding RNA TUG1 is a diagnostic factor in lung adenocarcinoma and suppresses apoptosis via epigenetically silencing of BAX [J].Oncotarget,2017 (8): 101899-101910.
- [2]Torre LA,Bray F,Siegel RL,et al.Global Cancer Statistics,2012 [J].CA Cancer J Clin,2015,65(2):87-108.
- [3]Yang J,Lin J,Liu T,et al.Analysis of lncRNA expression profiles in non-small cell lung cancers (NSCLC) and their clinical subtypes[J].Lung Cancer,2014,85(2):110-115.
- [4]Chen X,Chen Z,Yu S,et al.Long noncoding RNA LINC01234 functions as a competing endogenous RNA to regulate CBFB expression by sponging miR-204-5p in gastric cancer[J].Clin Cancer Res,2018,24(8):2002-2014.
- [5]Tomczak K,Czerwińska P,Wiznerowicz M.Review The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge[J].Contemp Oncol(Pozn),2015,19(1A):A68-A77.
- [6]Wakeam E,Acuna SA,Leighl NB,et al.Surgery Versus Chemotherapy and Radiotherapy For Early and Locally Advanced Small Cell Lung Cancer:A Propensity-Matched Analysis of Survival[J].Lung Cancer,2017(109):78-88.
- [7]Zhang Z,Wang S,ji D,et al.Construction of a ceRNA network reveals potential lncRNA biomarkers in rectal adenocarcinoma[J].Oncology Reports,2018,39(5):2101-2113.
- [8]Yang J,Li T,Gao C,et al.FOXO1 3'UTR functions as a ceRNA in repressing the metastases of breast cancer cells via regulating miRNA activity[J].FEBS Letters,2014,588(17):3218-3224.
- [9]Lv J,Fan HX,Zhao XP,et al.Long non-coding RNA Unigene56159 promotes epithelial-mesenchymal transition by acting as a ceRNA of miR-140-5p in hepatocellular carcinoma cells[J].Cancer Letters,2016,382(2):166-175.

收稿日期:2019-12-05;修回日期:2019-12-16

编辑/成森