

• 论 著 •

基于生物信息学分析鉴定促进非小细胞肺癌耐药的关键基因*

李佳月,付小雪,张瑞雪,孙梦瑶,杨金月,李 旺[△]

(山东航空学院生物与环境工程学院生物制药教研室,山东 滨州 256600)

[摘 要] **目的** 通过生物信息学分析基因表达综合数据库(GEO 数据库)中吉非替尼敏感和耐药非小细胞肺癌(NSCLC)细胞的基因表达谱,研究导致 NSCLC 获得性耐药的潜在关键基因,为耐药 NSCLC 患者的预后评价提供新思路。**方法** 通过 GEO2R 网络工具分析 GSE123066 和 GSE83666 数据集识别差异表达基因(DEGs)。通过韦恩图比较 2 组数据之间的 DEGs 找出关键的候选基因,然后对候选基因进行基因本体(GO)分析、京都基因与基因组百科全书(KEGG)分析及蛋白质-蛋白质相互作用(PPI)分析,并对所得到的基因进行临床预后分析、肿瘤组织及不同分期中的表达水平进行分析。**结果** 筛选到 46 个核心 DEGs 与 NSCLC 耐药显著相关,通过 Cytoscape 确定 8 个关键基因在 NSCLC 细胞的吉非替尼耐药进展中起重要作用,分别是 SOD2、GADD45A、NCF2、LOX、CYR61、SOX2、JUP 和 MUC1,最后生存分析发现 SOD2 与 NSCLC 吉非替尼耐药高度相关。**结论** SOD2 可能是导致 NSCLC 细胞对吉非替尼获得性耐药的关键基因,其高表达与 NSCLC 不良预后相关,表明 SOD2 可能是未来具有临床应用价值的潜在治疗靶点。

[关键词] 非小细胞肺癌; 生物信息学; 吉非替尼耐药; SOD2; 治疗靶点

DOI:10.3969/j.issn.1009-5519.2024.20.003 **中图法分类号:**R734.2

文章编号:1009-5519(2024)20-3432-08 **文献标识码:**A

Identification of key genes promoting drug resistance of non-small cell lung cancer based on bioinformatics analysis*

LI Jiayue, FU Xiaoxue, ZHANG Ruixue, SUN Mengyao, YANG Jinyue, LI Wang[△]

(Biopharmaceutical Teaching and Research Section, College of Biological and Environmental Engineering, Shandong University of Aeronautics, Binzhou, Shandong 256600, China)

[Abstract] **Objective** To analyze the gene expression profiles of gefitinib-sensitive and drug-resistant non-small cell lung cancer(NSCLC) cells in GEO database by bioinformatics, to study the potential key genes leading to acquired drug resistance of NSCLC, and to provide new ideas for the prognosis evaluation of patients with drug-resistant NSCLC. **Methods** The GEO2R network tool was used to analyze the GSE123066 and GSE83666 data sets to identify differentially expressed genes(DEGs). The key candidate genes were identified by comparing the DEGs between the two groups of data through the Wayne diagram, and then GO analysis, KEGG analysis and protein-protein interaction(PPI) analysis were performed on the candidate genes. The clinical prognosis analysis and expression level analysis of tumor tissues and different stages were performed on the obtained genes. **Results** A total of 46 core DEGs were significantly associated with drug-resistant NSCLC. Cytoscape identified eight key genes that play an important role in the progression of gefitinib resistance in NSCLC cells, namely SOD2, GADD45A, NCF2, LOX, CYR61, SOX2, JUP and MUC1, respectively. Among them, SOD2 was highly correlated with gefitinib resistance in NSCLC. **Conclusion** SOD2 may be the key genes leading to acquired drug resistance of NSCLC cells to gefitinib, and their high expression is associated with poor prognosis of NSCLC, indicating that SOD2 may be potential therapeutic targets with clinical application value in the future.

* 基金项目:国家自然科学基金项目(82103018);山东省自然科学基金项目(ZR2021QH029);山东省大学生创新创业训练计划项目(S202310449349);山东航空学院教学研究项目(BYJYYB202229)。

作者简介:李佳月(2002—),本科生在读,主要从事生物技术制药工作。 [△] 通信作者, E-mail:liwang19902010@163.com。

[Key words] Non-small cell lung cancer; Bioinformatics; Gefitinib resistance; SOD2; Therapeutic target

肺癌是对人群健康和生命威胁最大的恶性肿瘤之一,其发病率和死亡率增长最快^[1]。非小细胞肺癌(NSCLC)是肺癌中最常见的类型,约占肺癌病例的85%^[2]。近年来,表皮生长因子受体-酪氨酸激酶抑制剂(EGFR-TKIs)耐药性的产生已成为制约EGFR敏感突变NSCLC疗效的关键难题^[3]。本研究从基因表达综合数据库(GEO数据库)中获得NSCLC吉非替尼耐药相关基因表达数据(GSE123066和GSE83666),通过GEO2R软件鉴定差异表达基因(DEGs),并对DEGs进行基因本体(GO)分析、京都基因与基因组百科全书(KEGG)通路富集分析、蛋白质-蛋白质相互作用(PPI)分析及生存分析等,以期鉴定在吉非替尼耐药NSCLC细胞中发挥重要作用的关键基因,为耐药NSCLC患者的预后评价及发掘新的治疗方法提供新思路。

1 材料与方法

1.1 微阵列数据集 从GEO数据库下载GSE123066和GSE83666^[4],2个数据集均基于Illumina Human HT-12 V4.0表达谱芯片,展示了吉非替尼敏感和耐药NSCLC细胞的基因表达谱。2个公开数据集的生物信息学分析流程见图1。

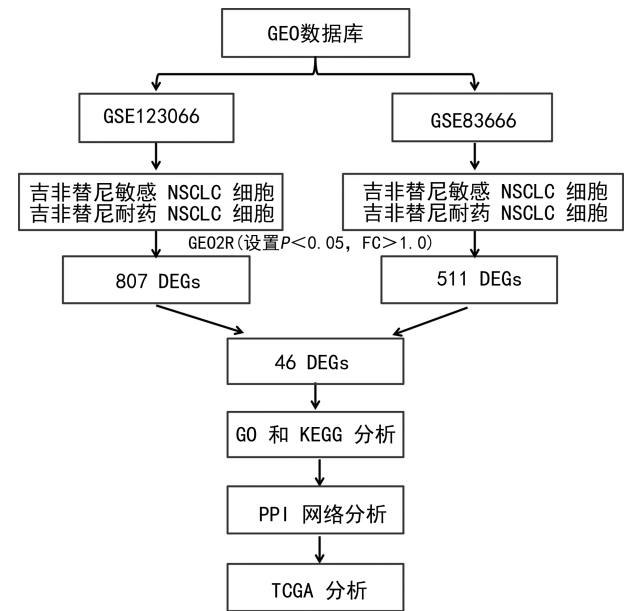


图1 本研究分析流程图

1.2 DEGs的识别 通过使用GEO2R网络工具(<http://www.ncbi.nlm.nih.gov/geo/geo2r/>)比较吉非替尼敏感和耐药NSCLC细胞鉴定DEGs。GEOquery和limma(用于微阵列分析的线性模型)来自Biocon-

ductor项目的R包。 $P<0.05$ 和 $|\log FC|>1.0$ 为差异有统计学意义。使用Venny 2.1(<https://bioinfogp.cnb.csic.es/tools/venny/index.html>)做韦恩图,提取2个数据集中重叠的DEGs。

1.3 GO和KEGG通路分析 使用DAVID(<https://david.ncifcrf.gov/>)数据库^[5]进行GO分析和KEGG通路富集分析。 $P<0.05$ 的富集被设定为显著富集的关键标准。对于GO分析,DEGs的功能分为3组:生物过程(BP)、细胞成分(CC)和分子功能(MF)。对于注释汇总结果列表,选择了GOTERM_BP_DIRECT、GOTERM_CC_DIRECT和GOTERM_MF_DIRECT。使用R语言的ggplot2包对富集结果进行可视化。

1.4 PPI网络分析 STRING(<https://string-db.org/>)数据库^[6]用于对DEGs进行PPI网络分析。 $P<0.05$ 为网络交互关系差异有统计学意义,交互得分 >0.7 分认为是高置信交互关系。应用Cytoscape 3.7.2软件进行^[7]可视化PPI网络分析,并应用CytoHubba识别PPI网络中的中心节点基因,然后将中心节点基因作为候选DEGs进行后续分析。

1.5 基于TCGA数据库的生物信息学分析 基于癌症基因组图谱数据库(TCGA),使用GEPIA2(<http://gepia2.cancer-pku.cn>)和Oncomine(<https://www.oncomine.org>)分析肺癌标本和相关临床数据中的基因表达水平。

2 结果

2.1 DEGs的识别 利用limma包从NSCLC细胞系HCC4006(GSE123066)和PC9(GSE83666)中获得DEGs。GSE123066数据集中鉴定出807个DEGs,其中包括327个上调基因和480个下调基因(图2A)。从GSE83666数据集中鉴定出511个DEGs,包含166个上调基因和345个下调基因(图2B)。使用韦恩图显示GSE123066和GSE83666数据集的基因交集。在GSE123066和GSE83666数据集中共发现46个共同DEGs(图2C)。本研究还确定了2个数据集中同时存在6个共同上调基因(FAM129A、SLC2A3、LOX、ABL1、SIRPA、TMEM45A)(图2D)和11个共同下调基因(TCN1、CYB5R2、ID1、NMNAT2、ZDHHC11、PRRG2、HSPA2、MYO5C、ARHGDIB、STAT4、

TP53INP1)(图 2E)。所有 46 个 DEGs,特别是共同上调基因和共同下调基因被认为是促进人 NSCLC 细胞吉非替尼耐药的候选基因。

2.2 46 个 DEGs 的 GO 功能富集分析 利用 DAVID 数据库对 2 个数据集获得的 46 个 DEGs 进行功能富集分析。GO 富集分析主要从 BP、CC、MF 3 个方面预测靶基因的功能。分析发现 DEGs 主要富集在 BP,包括对药物的反应、蛋白质磷酸化的正调控、细胞周期阻滞、对内质网应激的反应、凋亡过程的正调控等(图 3A)。对于 GO CC 分析,DEGs 显著富集在细胞外泌体和肌动蛋白细胞骨架(图 3B)。GO MF 分析包括烟酸核苷酸腺苷酰(基)转移酶活性和肌动蛋白丝结合(图

3C)。利用 STRING 数据库和 Cytoscape 3.7.2 软件对获得的 46 个 DGEs 构建 PPI 网络并确定关键基因。STRING 构建的 PPI 网络共有 45 个节点和 18 条边。利用 Cytoscape 软件,确定了连接度最高的前 8 个基因(图 3D),其是促进人 NSCLC 细胞吉非替尼耐药的候选基因。其中连接最多的基因是 SOD2,其次是 GADD45A、NCF2、LOX、CYR61、SOX2、JUP 和 MUC1。这 8 个中心基因包括 3 个模块: SOD2、GADD45A 和 NCF2 可形成一个模块, LOX、CYR61 和 SOX2 可形成一个模块, JUP 和 MUC1 可形成另一个模块。

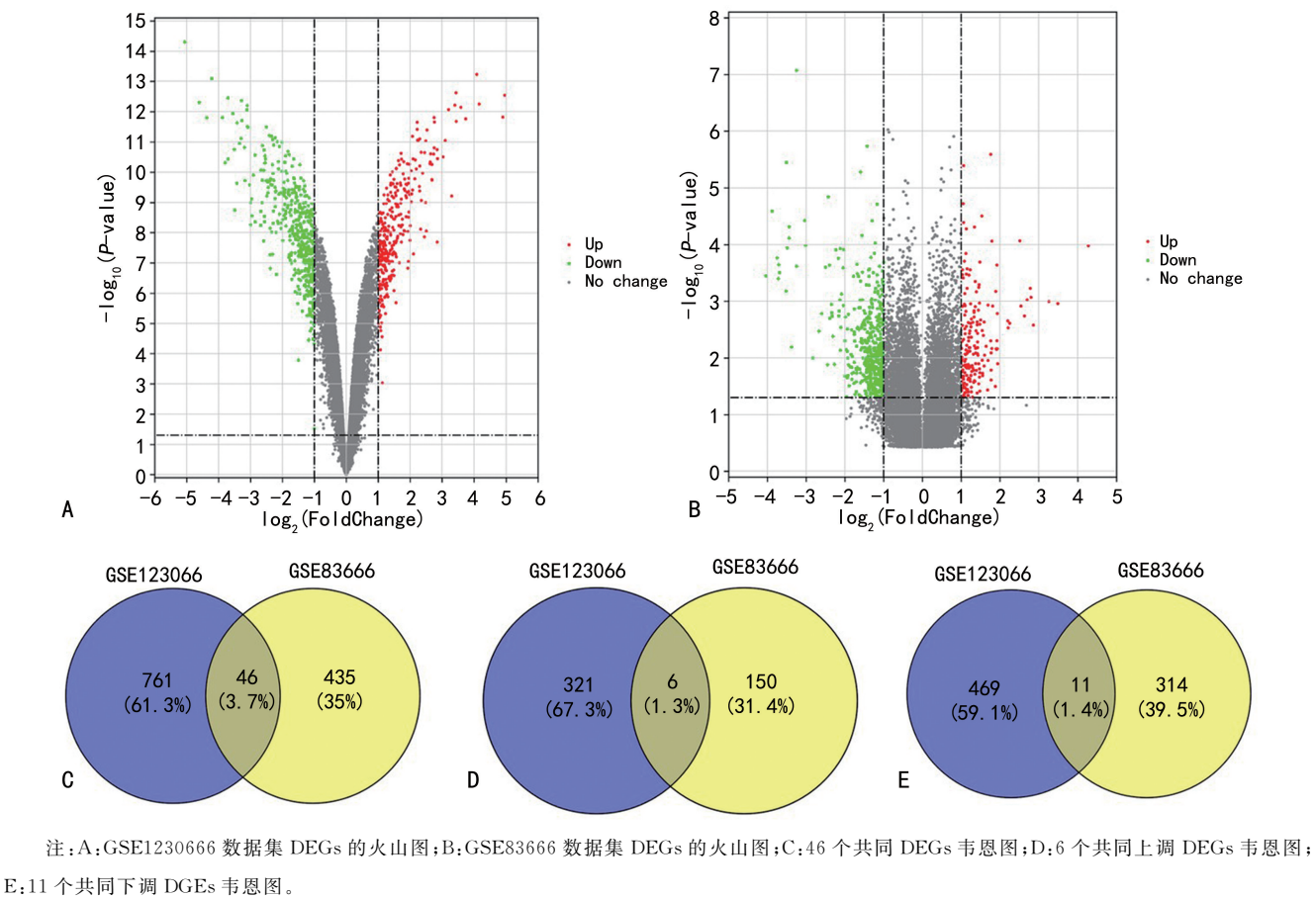
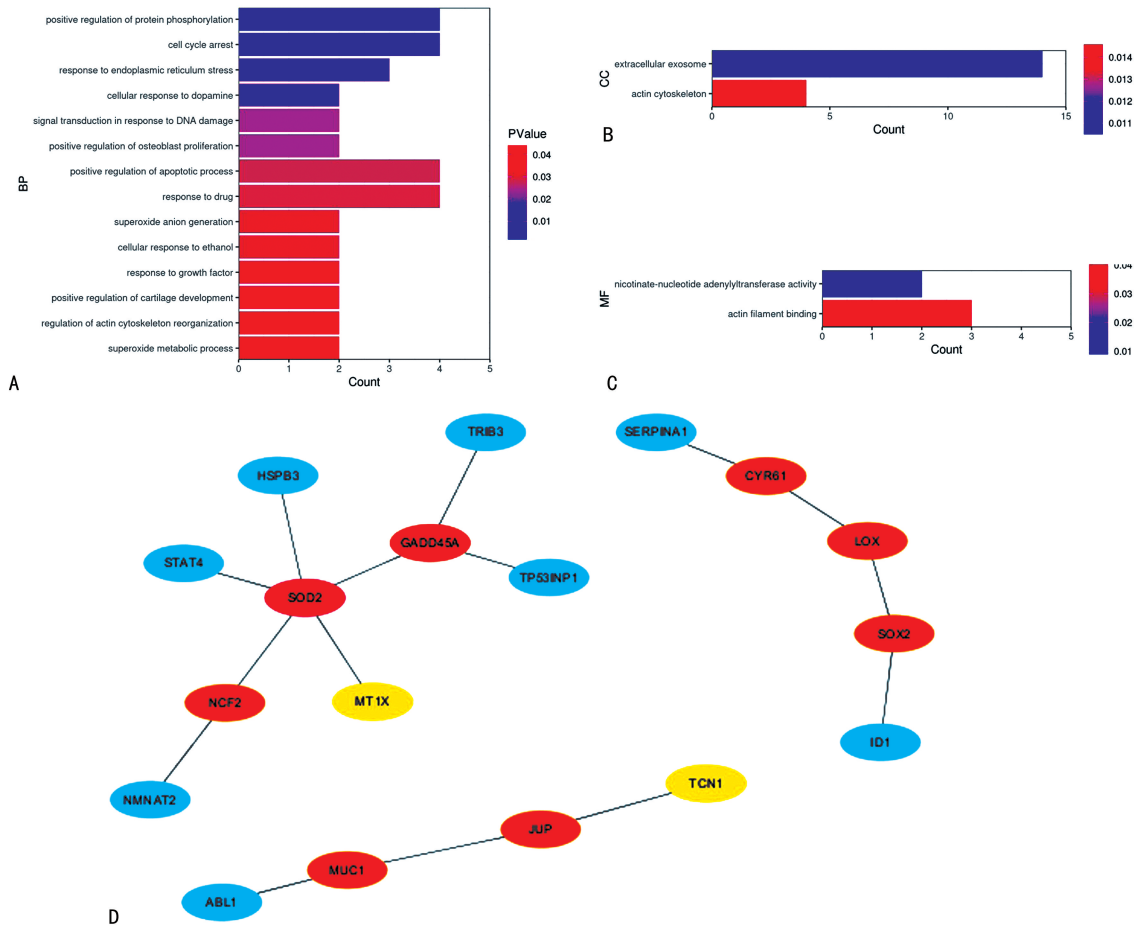


图2 GSE123066 和 GSE83666 DEGs 的鉴定

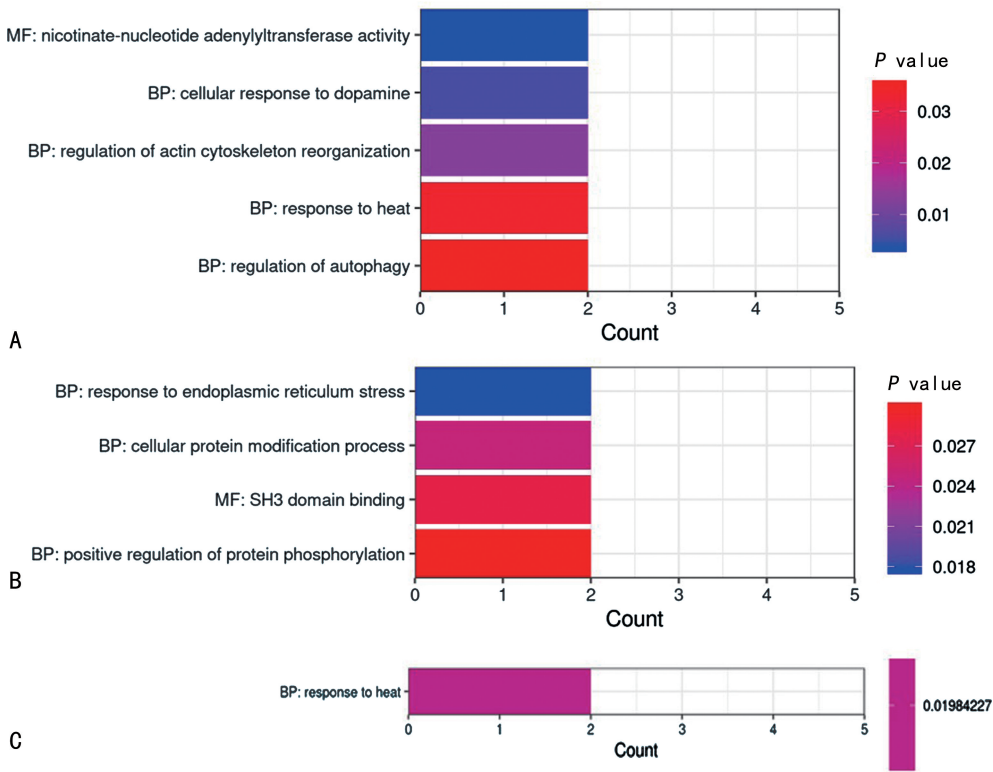
2.3 DEGs 的 GO 功能富集分析 将 6 个共同上调和 11 个共同下调的 DEGs 视为一个整体提交 DAVID 数据库进行功能富集分析。这 17 个 DEGs 主要富集在 BP 方面,包括细胞对多巴胺的反应、肌动蛋白细胞骨架重组的调节、对热的反应和自噬的调节。对于 GO CC 分析,17 个 DEGs 无显著富集。对这 17 个 DEGs 进行 GO MF 分析,发现这些 DEGs 只在烟酸核苷酸腺苷酰(基)转移酶活性上显著富集(图 4A)。

利用 DAVID 数据库对 2 个数据集获得的 6 个共同上调 DEGs 进行功能富集分析。本研究发现,这 6 个共同上调的 DEGs 在 BP 方面显著富集于响应内质网应激、细胞蛋白质修饰过程和蛋白质磷酸化的正调控,在 GO MF 分析中显著富集于 SH3 结构域结合(图 4B)。本研究还利用 DAVID 数据库分析了 11 个共同下调的 DEGs,发现只有 2 个基因(TP53INP1 和 HSPA2)在 BP 功能中显著富集,即对热的响应(图 4C)。



注: A. $P < 0.05$ 时 DEGs 的 GO_BP 分析; B. $P < 0.05$ 时 DEGs 的 GO_CC 分析; C. $P < 0.05$ 时 DEGs 的 GO_MF 分析; D. DEGs 的 PPI 网络分析。

图 3 46 个 DEGs 的 GO 功能富集分析



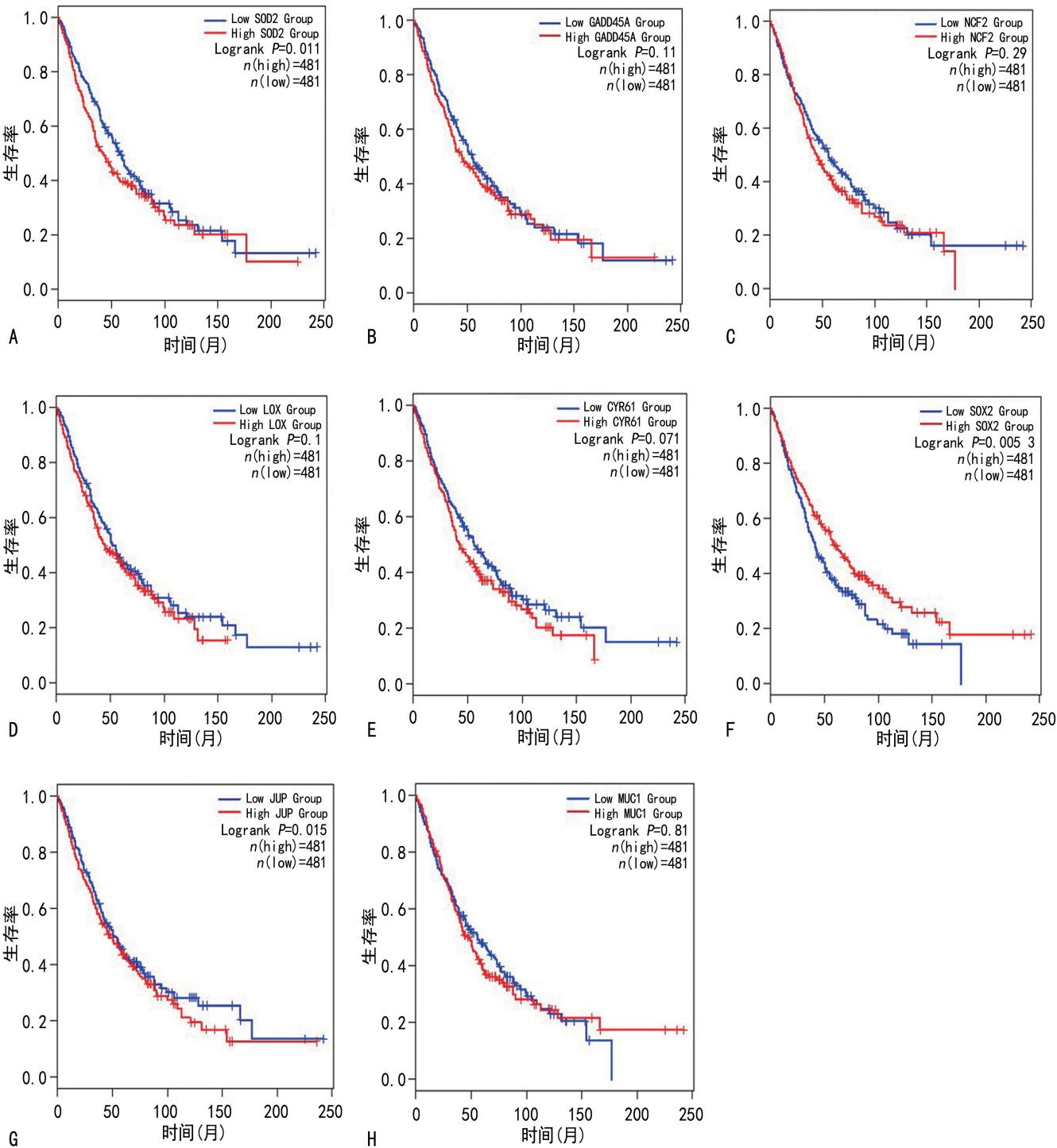
注: A. 17 个 DEGs 功能富集分析; B. 6 个共同上调 DEGs 功能富集分析; C. 11 个共同下调 DEGs 功能富集分析。

图 4 DEGs 的 GO 功能富集分析

2.4 SOD2 提示 NSCLC 临床预后不良 在鉴定的 8 个与人 NSCLC 细胞吉非替尼耐药相关的 DEGs 中,只有 SOD2 和 SOX2 的表达水平与总生存时间相关。其中,SOD2 表达越高,生存时间越短,而 SOX2 表达水平越高,生存时间越长。见图 5。

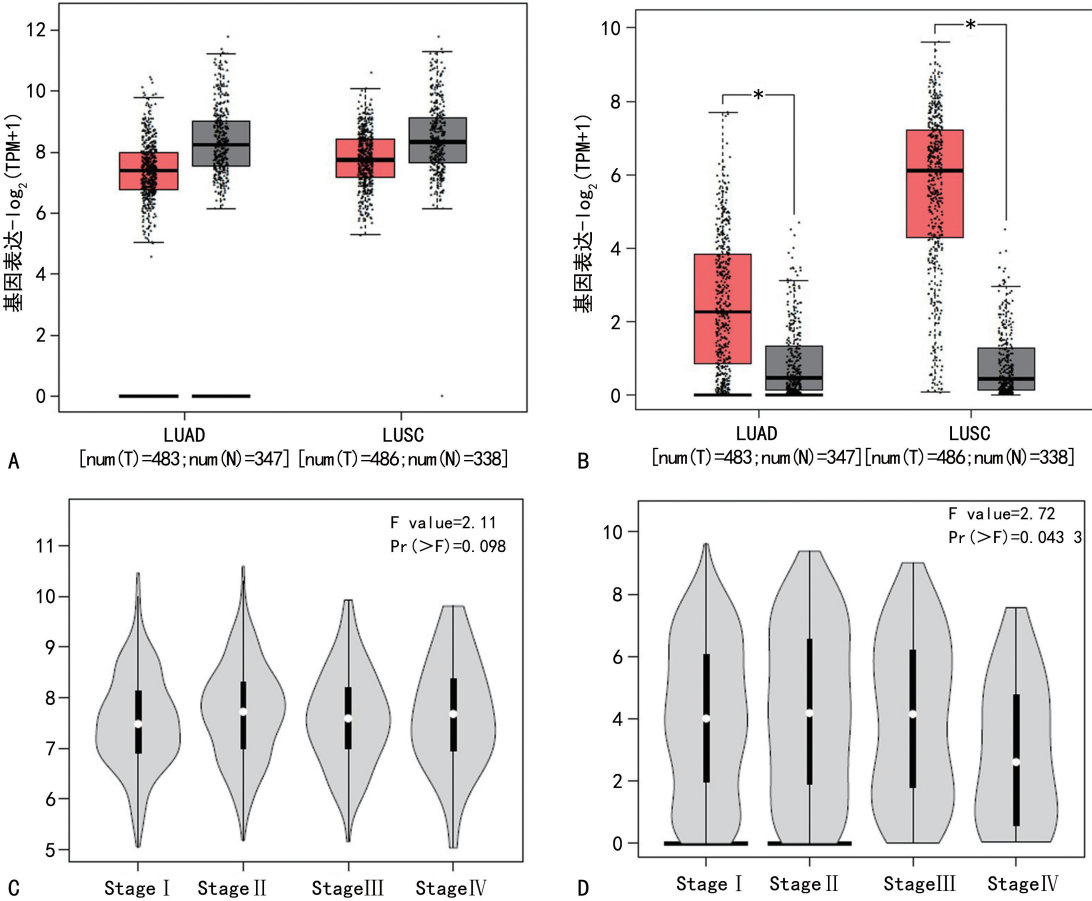
对于 SOD2,在肺腺癌(LUAD)或肺鳞癌(LUSC)

中未观察到明显变化(图 6A)。与正常组织相比,SOX2 在 LUAD 和 LUSC 中表达均上调(图 6B)。SOD2 的 mRNA 表达与患者个体癌症分期无关(图 6C),而 SOX2 的 mRNA 表达与患者个体癌症分期显著相关(图 6D)。



注: A. SOD2; B. GADD45A; C. NCF2; D. LOX; E. CYR61; F. SOX2; G. JUP; H. MUC1。

图 5 肺癌中 8 个相关 DEGs 和对总生存期的 Kaplan-Meier 曲线



注:A. SOD2 在 LUAD 或 LUSC 中的变化;B. SOX2 在 LUAD 或 LUSC 中的变化;C. SOD2 的 mRNA 表达与患者个体癌症分期关系;D. SOX2 的 mRNA 表达与患者个体癌症分期关系。

图 6 基于 TCGA 数据库数据的 SOD2 和 SOX2 在 LUAD 和 LUSC 肿瘤组织及不同分期中的表达水平

3 讨论

吉非替尼和厄洛替尼作为治疗晚期 EGFR 突变 NSCLC 的一线治疗药物,在治疗早期阶段显示出可观的疗效^[8]。然而,多数患者在几个月后对这些 EGFR-TKIs 表现出耐药,最终导致治疗失败^[9-10]。目前,已发现多种耐药机制,如 EGFR-T790M 突变、PIK3CA 突变、MET 或 FGFR1 扩增^[11-12]等基因突变。此外,有研究报道,异常表达蛋白可能是造成癌细胞耐药的关键因素,例如 EZH2 基因过表达发生在多种人类恶性肿瘤中,EZH2 高表达可导致肿瘤细胞耐药,在黑色素瘤中,UHRF1/UBE2L6/UBR4 介导的泛素化调节 EZH2 的丰度,从而调节黑色素细胞分化表型^[13]。RNF25 通过调节 ERK 和 NF- κ B 通路之间的串扰促进 EGFR 突变的 NSCLC 细胞对吉非替尼的耐药^[14]。在吉非替尼耐药的 NSCLC 细胞中,纤溶酶原激活剂尿激酶受体(PLAUR)被证明通过 EGFR/p-Akt/survivin 信号通路诱导吉非替尼耐药^[15]。肺癌,尤其是 NSCLC 的基因表达谱已经通过高通量测序平台广泛实施,基因表达谱有助于有效预

测患者最终发生 EGFR-TKI 耐药的关键因素。

本研究中,锰 SOD2 是 STRING 构建的 PPI 网络中连接最多的基因,已有研究表明 SOD2 的高表达与 NSCLC 的不良临床预后相关。SOD2 是参与活性氧(ROS)代谢的多态酶,其可以催化超氧自由基歧化成过氧化氢^[16-18]。多个共同上调的 DEGs 在 BP 方面显著富集于响应内质网应激通路,活性氧的产生被证实与内质网应激和未折叠蛋白反应有密切联系。天然倍半萜内酯类化合物山金车内酯 D(AD)诱导肝细胞癌(HCC)细胞发生内质网应激介导的胀亡,并且该过程是活性氧依赖的,内质网应激过程中 PERK-eIF2 α -ATF4-CHOP 通路的激活是 AD 诱导的胀亡的关键因子^[19]。研究表明,SOD2 表达对癌细胞发生发展至关重要,circRNA SOD2 通过作为 microRNA-2355-5p 的竞争性内源 RNA 介导钙调蛋白调节的血影蛋白相关蛋白-2,促进 NSCLC 的 EMT 进程^[20]。CHANG 等^[21]研究表明,C₈-神经酰胺作为一种强效的化疗药物可能通过调节 SOD1 和 SOD2 的开关,促进内源性 ROS 水平,最终触发 NSCLC H1299 细胞

凋亡。此外,MA 等^[22]发现,NRF2(也是本研究中心基因)通过上调 SOD2 和 GPX4 诱导 HCC827 细胞对 EGFR-TKIs 耐药,提示靶向 NRF2-GPX4/SOD2 通路可能是克服 EGFR-TKIs 耐药的潜在策略。综上所述,本研究预测活性氧介导的内质网应激反应在 NSCLC 细胞吉非替尼获得性耐药方面起到重要作用,需要进一步研究证实。

性别决定区 Y-Box2(SOX2)是一个重要的节点基因,研究表明去泛素化酶 PSMD7 通过修饰 SOX2 降解激活 Noct1 通路促进胰腺癌进展^[23]。在结直肠癌中,SOX2 通过 lncRNA AC005392.2-GLUT1 轴加速糖酵解促进血管生成拟态^[24]。在胰腺癌中,抑制 NR5A2 通过破坏 SOX2/MYC 信号传导和恢复化疗敏感性来靶向胰腺癌中的干性^[25]。有研究发现,超级强化子驱动的 SOX2 在鼻咽癌中通过染色质重组促进肿瘤形成^[26]。本研究生存分析结果显示,SOX2 表达越高,生存时间越长,这预示着在吉非替尼耐药的 NSCLC 中 SOX2 可能起到了抑癌作用。为了充分发挥 SOX2 作为 NSCLC 耐药治疗靶点的潜力,需要更全面地了解 SOX2 的所有功能。

总之,本研究提出 SOD2 可能是导致 NSCLC 细胞对吉非替尼获得性耐药的关键基因,可能是未来具有临床应用价值的潜在治疗靶点。其潜在的复杂分子机制仍需进一步探索。

参考文献

[1] BADE B C,DELA C C S. Lung cancer 2020:epidemiology, etiology, and prevention[J]. Clin Chest Med,2020,41(1):1-24.

[2] XU X L,WANG D,XU W Z,et al. NIPBL-mediated RAD21 facilitates tumorigenicity by the PI3K pathway in non-small-cell lung cancer[J]. Commun Biol,2024,7(1):206.

[3] HE J,HUANG Z,HAN L,et al. Mechanisms and management of 3rd-generation EGFR-TKI resistance in advanced non-small cell lung cancer(Review) [J]. Int J Oncol,2021,59(5):2332-239.

[4] BARRETT T,WILHITE S E,LEDOUX P,et al. NCBI GEO:archive for functional genomics data sets--update[J]. Nucleic Acids Res,2013,41:D991-D995.

[5] SHERMAN B T,HAO M,QIU J,et al. David: a web server for functional enrichment analysis and functional annotation of gene lists(2021 update)[J]. Nucleic Acids Res,2022,50(W1):W216-W221.

[6] SZKLARCZYK D,KIRSCH R,KOUTROULI M,et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest[J]. Nucleic Acids Res,2023,51(D1):D638-D646.

[7] DONCHEVA N T,MORRIS J H,HOLZE H,et al. Cytoscape string App 2. 0:analysis and visualization of heterogeneous biological networks[J]. J Proteome Res,2023,22(2):637-646.

[8] GREENHALGH J,BOLAND A,BATES V,et al. First-line treatment of advanced epidermal growth factor receptor(EGFR) mutation positive non-squamous non-small cell lung cancer [J]. Cochrane Database Syst Rev,2021,3(3):CD010383.

[9] PAN Z Z,WANG K,WANG X N,et al. Cholesterol promotes EGFR-TKIs resistance in NSCLC by inducing EGFR/Src/Erk/SP1 signaling-mediated ERR α re-expression [J]. Mol Cancer,2022,21(1):77.

[10] KOULOURIS A,TSAGKARIS C,CORRIERO A C,et al. Resistance to TKIs in EGFR-mutated non-small cell lung cancer: from mechanisms to new therapeutic strategies[J]. Cancers (Basel),2022,14(14):3337.

[11] CHEN N Y,LIN C W,LAI T Y,et al. Increased expression of SSEA-4 on TKI-resistant non-small cell lung cancer with EGFR-T790M mutation[J]. Proc Natl Acad Sci U S A,2024,121(5):e2313397121.

[12] PANG L L,GAN J D,TAN J R,et al. Efficacy and potential resistance mechanisms of afatinib in advanced non-small cell lung cancer patients with EGFR G719X/L861Q/S768I[J]. Cancer,2022,128(21):3804-3814.

[13] KUSER-ABALI G,ZHANG Y F,SZETO P,et al. UHRF1/UBE2L6/UBR4-mediated ubiquitination regulates EZH2 abundance and thereby melanocytic differentiation phenotypes in melanoma[J]. *Oncogene*,2023,42(17):1360-1373.

[14] CHO J H,YOU Y M,YEOM Y I,et al. RNF25 promotes gefitinib resistance in EGFR-mutant NSCLC cells by inducing NF-κB-mediated ERK reactivation[J]. *Cell Death Dis*, 2018, 9 (6): 587.

[15] ZHOU J,KWAK K J,WU Z R,et al. PLAUR confers resistance to gefitinib through EGFR/P-AKT/survivin signaling pathway [J]. *Cell Physiol Biochem*,2018,47(5):1909-1924.

[16] AGUIRRE J D,CULOTTA V C. Battles with Iron;manganese in oxidative stress protection [J].*J Biol Chem*,2012,287(17):13541-13548.

[17] BECUWE P, ENNEN M, KLOTZ R, et al. Manganese superoxide dismutase in breast cancer: from molecular mechanisms of gene regulation to biological and clinical significance [J]. *Free Radic Biol Med*,2014,77:139-151.

[18] COELHO D R,PALMA F R,PAVIANI V,et al. Nuclear-localized, iron-bound superoxide dismutase-2 antagonizes epithelial lineage programs to promote stemness of breast cancer cells via a histone demethylase activity[J]. *Proc Natl Acad Sci U S A*, 2022, 119 (29): e2110348119.

[19] LIN Y S,SUN Z W,SHEN L S,et al. Arnicolide D induces endoplasmic reticulum stress-mediated oncosis via ATF4 and CHOP in hepatocellular carcinoma cells[J]. *Cell Death Discov*, 2024,10(1):134.

[20] LV C S,HU Y N,ZHOU X,et al. CircRNA SOD2 motivates non-small cell lungs cancer advancement with EMT via acting as microRNA-2355-5p's competing endogenous RNA to mediate calmodulin regulated spectrin associated proteins-2[J]. *Bioengineered*,2022, 13(3): 5756-5768.

[21] CHANG Y C,FONG Y, TSAI E M, et al. Exogenous C₈-Ceramide induces apoptosis by overproduction of ROS and the Switch of superoxide dismutases SOD1 to SOD2 in human lung cancer cells[J]. *Int J Mol Sci*, 2018, 19 (10):3010.

[22] MA C S,LV Q M,ZHANG K R,et al. NRF2-GPX4/SOD2 axis imparts resistance to EGFR-tyrosine kinase inhibitors in non-small-cell lung cancer cells[J]. *Acta Pharmacol Sin*, 2021, 42 (4):613-623.

[23] LUO C,YU Y,ZHU J F,et al. Deubiquitinase PSMD7 facilitates pancreatic cancer progression through activating Nocthl pathway via modifying SOX2 degradation[J]. *Cell Biosci*, 2024,14(1):35.

[24] HUANG S M,WANG X,ZHU Y,et al. SOX2 promotes vasculogenic mimicry by accelerating glycolysis via the lncRNA AC005392. 2-GLUT1 axis in colorectal cancer[J]. *Cell Death Dis*,2023,14(12):791.

[25] ZHENG Q,TANG J J,AICHER A,et al. Inhibiting NR5A2 targets stemness in pancreatic cancer by disrupting SOX2/MYC signaling and restoring chemosensitivity[J]. *J Exp Clin Cancer Res*,2023,42(1):323.

[26] LIU S X,WANG C,LIN R B,et al. Super-enhancer driven SOX2 promotes tumor formation by chromatin re-organization in nasopharyngeal carcinoma[J]. *E Bio Med*,2023,98:104870.

(收稿日期:2024-03-10 修回日期:2024-07-12)