引用本文:渠茂林.基于GEO数据库分析非小细胞肺癌癌组织中配对盒基因8的表达及其与病人预后的关系[J].安徽医药,2021,25(6):1169-1172.**DOI:10.3969/j.issn.1009-6469.2021.06.027.**

◇临床医学◇

基于GEO数据库分析非小细胞肺癌癌组织中配对盒基因8的表达及其与病人预后的关系

渠茂林

作者单位:呼和浩特市蒙医中医医院检验科,内蒙古自治区 呼和浩特010010

摘要: **目的** 探讨基于高通量基因表达数据库(Gene expression Omnibus,GEO数据库)分析配对盒基因8(PAX8)在非小细胞肺癌(NSCLC)病人癌组织中的表达情况及与预后的关系。**方法** 利用在线工具Oncomine分析PAX8在NSCLC与正常肺组织的表达情况,基于GEO数据库下载相关芯片数据(GSE77803),收集156例NSCLC组织的基因表达谱及相关临床数据,采用中位数法将其分为PAX8低表达组(n=78)和PAX8高表达组(n=78)。分析PAX8的表达与病人性别、年龄、病理类型、T分期、N分期和术后复发情况之间的关系;采用生存分析PAX8的表达与病人预后生存之间的关系;采用基因本体(GO)注释分析PAX8基因的功能,采用KEGG PATHWAY分析PAX8可能参与的信号通路。**结果** TCGA Lung 2数据集共检测了1 537个样本,其中正常肺组织样本390例,肺癌组织样本共721例,血液样本420例,无意义样本6例。与正常肺组织相比,PAX8在肺腺癌和肺鳞癌中表达水平明显升高(P<0.05);基于GEO数据库中GSE77803(GPL570)数据集分析,NSCLC癌组织中PAX8的表达水平与病人的T分期、N分期和是否复发有关(P<0.05),生存分析显示,癌组织中PAX8低表达病人的总生存率为52.94%,中位生存期为54.6个月,显著高于PAX8高表达病人(34.35%,48.6个月)(log-rank χ^2=4.609,P=0.032)。PAX8低表达病人无病生存率为42.50%,中位生存期为51.7个月,显著高于PAX8高表达病人(28.71%,40.8个月)(log-rank χ^2=5.712,P=0.017);GO注释和KEGG分析显示,PAX8可能参与调节NSCLC发生发展相关的信号通路的表达。**结论** PAX8蛋白在肺腺癌和肺鳞癌中均存在异常高表达,PAX8低表达NSCLC病人具有更高的生存率。

关键词: 癌,非小细胞肺; 高通量基因表达数据库(GEO数据库); 配对盒基因8

Analysis on PAX8 expression in non-small cell lung cancer tissues and its relationship with prognosis based on GEO database

QU Maolin

Author Affiliation:Laboratory Department, Hohhot Mongolian Medical Hospital of Traditional Chinese Medicine, Hohhot, Inner Mongolia Autonomous Region 010010, *China*

Abstract: **Objective** To explore expression of paired box gene 8 (PAX8) in cancer tissues of non-small cell lung cancer (NSCLC) patients and its relationship with prognosis based on Gene expression Omnibus (GEO) database.**Methods** The online tool Oncomine was applied to analyze expression of PAX8 in NSCLC and normal lung tissues. Based on GEO database, related chip data (GSE77803) were downloaded. The gene expression profiles and related clinical data of 156 NSCLC tissues in 156 cases were collected and assigned into PAX8 low-expression group (n=78) and PAX8 high-expression group (n=78) by median method. The relationship between expression of PAX8 and gender, age, pathological type, T stage, N stage, postoperative recurrence was analyzed. The survival analysis was applied to analyze relationship between PAX8 expression and prognosis survival of patients. Gene Ontology (GO) annotations were applied to analyze PAX8 gene function. The signal pathways that PAX8 might participate in were analyzed by Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway.**Results** There were 1 537 samples detected by The Cancer Genome Atlas (TCGA) Lung 2 dataset, including 390 cases with normal lung tissues, 721 cases with lung cancer tissues, 420 cases with blood and 6 cases with meaningless. Compared with normal lung tissues, the level of PAX8 expression was significantly increased in lung adenocarcinoma and lung squamous carcinoma (P<0.05). Based on GSE77803 (GPL570) dataset analysis in GEO database, the level of PAX8 expression in NSCLC tissues was related with T stage, N stage and recurrence of patients (P<0.05). Survival analysis showed that overall survival rate and median survival time of patients with PAX8 low expression in cancer tissues was significantly higher than those with PAX8 high expression (52.94%, 54.6 months *vs.* 34.35%, 48.6 months) (log-rank χ^2=4.609, P=0.032). The disease-free survival rate and median survival time of patients with PAX8 low expression were significantly higher than those with PAX8 high expression (42.50%, 51.7 months *vs.* 28.71%,40.8 months) (log-rank χ^2=5.712, P=0.017).GO annotations and KEGG analysis showed that PAX8 might be involved in signaling pathways which regulated occurrence and development of NSCLC.**Conclusion** PAX8 protein is abnormally highly expressed in lung adenocarcinoma and lung squamous carcinoma.The survival rate is higher in NSCLC patients with PAX8 low expression.

Key words: Carcinoma, non-small-cell lung; Gene Expression Omnibus data base (GEO data base); Paired box gene 8

非小细胞肺癌(NSCLC)是最常见的一种肺癌,包括鳞状细胞癌(鳞癌)、腺癌、大细胞癌,约占所有肺癌的80%~85%,恶性程度高,易发生转移和扩散[1]。目前,NSCLC病人主要采用手术联合放化疗,以期延长病人的生存时间,但术后易复发转移,5年生存率很低[2]。目前,NSCLC发生发展的机制尚不清楚,从基因水平研究NSCLC的病理机制对提高早期诊断率、寻找治疗靶点、改善预后至关重要。配对盒基因8(PAX8)是一种细胞系限制性转录因子,可以与靶DNA结合,调控相关基因的转移[3]。已有研究显示,PAX8在乳腺癌、胃癌、肾癌以及肺癌等多种恶性肿瘤中均存在异常高表达[4-5],但其在NSCLC中的表达研究报道较少。高通量基因表达数据库(Gene expression Omnibus, GEO数据库)是目前最大的面向公众提供数据查询和可进行提交数据进行储存交换的芯片公共数据库之一[6]。因此本研究于2019年6月至2019年8月利用GEO数据库分析PAX8在NSCLC病人中的表达及其与病人预后的关系。

1 资料与方法

1.1 数据资料采集 利用在线工具Oncomine(https://oncomine.org/)选择TCGA Lung 2数据集,其中包括1 537个样本的基因表达谱数据,正常肺组织390例,肺癌组织共721例,血液样本420例,无意义样本6例,分析PAX8在肺癌组织与正常肺组织中的表达情况。再从GEO公共数据库(https://www.ncbi.nlm.nih.gov/gds)检索NSCLC的差异基因表达谱及相关临床数据,检索平台Platforms:GPL570,数据集:GSE77803。

1.2 数据筛选及分组 下载GSE77803(GPL570)基因表达矩阵,156例NSCLC病人的数据集包含156例NSCLC病人的基因表达谱数据和性别、年龄、病理类型、T分期、N分期和术后复发情况。根据矩阵中PAX8(121_at)表达值的中位数,将NSCLC临床样本分为PAX8低表达组(n=78)和PAX8高表达组(n=78),分析两组PAX8基因表达水平与NSCLC病人临床特征及预后之间的关系。

1.3 PAX8基因功能分析 采用基因本体(GO)注释分析(https://bioportal.bioontology.org/ontologies/GO)PAX8基因的功能,采用KEGG PATHWAY分析(https://www.kegg.jp/kegg/pathway.html)PAX8可能参与的信号通路。

1.4 统计学方法 采用SPSS 17.0进行分析。对满足正态分布的计量资料以$\bar{x} \pm s$表示,采用单因素方差分析比较多组间差异,若组间比较有差异,多组间的两两比较采用SNK-q检验。计数资料采用例(%)表示,采用χ^2检验比较组间差异。采用Kaplan-Meier绘制生存曲线,分析PAX8表达与预后生存之间的关系,总生存率表示随访终点所有病人的生存状态;无病生存率表示随访终点所有病人的无事件生存状态,比较采用log-rank χ^2检验。均以P<0.05表示差异有统计学意义。

2 结果

2.1 肺癌组织中PAX8的表达 采用Oncomine在线数据库,选择基因PAX8,设置foldchange为2,得到PAX8在不同肿瘤中的表达情况。TCGA Lung 2数据集共检测了1 537个样本,其中正常肺组织390例,肺癌组织共721例,血液样本420例,无意义样本6例。与正常肺组织(0.06±0.02)相比,PAX8在肺腺癌(0.14±0.07)和肺鳞癌(0.15±0.09)中表达水平明显升高(F=217.130,P<0.001)。

2.2 NSCLC癌组织中PAX8的表达与临床特征之间的关系 NSCLC癌组织中PAX8的表达水平与病人的T分期、N分期和是否复发有关(P<0.05),见表1。

2.3 PAX8的表达与NSCLC病人预后生存之间的关系 156例NSCLC病人的随访时间范围为6.2~64.8个月,中位随访时间为48.7个月。癌组织中PAX8低表达病人的总生存率为52.94%,中位生存期为54.6个月,显著高于PAX8高表达病人的(34.35%,48.6个月)(log-rank χ^2=4.609,P=0.032),见图1A。癌组织中PAX8低表达的无病生存率为42.50%,中位生存期为51.7个月,显著高于PAX8高表达病人的(28.71%,40.8个月)(log-rank χ^2=5.712,P=0.017),见图1B。

2.4 PAX8的GO分析 GO分析PAX8可能的功能,结果显示,在生物学过程水平上,PAX8可能参与了泌尿系统、肾脏、中枢神经系统、甲状腺等多组织器官的生理生化过程,在细胞组成水平上,PAX8主要定位与细胞核;在分子功能水平上,PAX8存在DNA结合位点,可以与靶DNA结合,发挥其转录因子的作用,见表2。

2.5 PAX8的京都基因和基因组数据库(KEGG)分析 KEGG分析PAX8可能参与的信号通路,map05216和map04918显示,PAX8蛋白可能参与调节甲状腺癌的发生发展以及甲状腺激素的合成,

表1 非小细胞肺癌癌组织中配对盒基因8的表达与临床特征之间的关系/例(%)

基本特征	例数	配对盒基因8 高表达(n=78)	配对盒基因8 低表达(n=78)	χ^2值	P值
年龄				0.650	0.420
≤55岁	87	46(52.87)	41(47.13)		
>55岁	69	32(46.38)	37(53.62)		
性别				1.713	0.191
男	94	51(54.26)	43(45.74)		
女	62	27(43.55)	35(56.45)		
病理类型				0.528	0.768
鳞癌	78	41(52.56)	37(47.44)		
腺癌	73	35(47.95)	38(52.05)		
其它	5	2(40.00)	3(60.00)		
T分期				7.878	0.005
T_1~T_2	97	40(41.24)	57(58.76)		
T_3~T_4	59	38(64.41)	21(35.59)		
N分期				6.783	0.009
N_1	92	38(41.30)	54(58.70)		
N_2	64	40(62.50)	24(37.50)		
复发				5.379	0.020
是	58	36(62.07)	22(37.93)		
否	98	42(42.86)	56(57.14)		

map05202和map05200显示,PAX8蛋白可能在肿瘤发生发展中发挥其转录因子的调节作用,见表3。

3 讨论

NSCLC约占所有肺癌的80%,是一种高度异质性疾病,探索NSCLC的发病机制,分析NSCLC的特异性分子标志物具有重要临床意义[7]。Oncomine数据库整合了GEO、TCGA和已发表的文献等来源的RNA和DNA-seq数据,是目前最常用的大型肿瘤基因芯片数据库,可用于分析差异基因的表达[8]。PAX8蛋白由450个氨基酸组成,位于2q12-14染色体的基因编码,在多种恶性肿瘤中均存在异常高表达[9]。本研究采用Oncomine基因芯片数据库检索了PAX8在肺癌中的表达情况,发现PAX8在大部分肺腺癌和肺鳞癌中均存在异常高表达,在少部分表达肺腺癌和肺鳞癌中出现异常低表达,肺腺癌和肺鳞癌均属于NSCLC,提示PAX8在大部分NSCLC中

表2 配对盒基因8(PAX8)的基因本体(GO)分析

GO分析	项目	功能
生物学过程	0001655	泌尿系统
	0001656//0001658//0001822//0001823	后肾、中肾
	0007417	中枢神经系统
	0030878//0030878//0038194	甲状腺
细胞学组成	0005634	核苷酸
	0005654	核质
	0005730	核仁
分子功能水平	0000979	RNA聚合酶Ⅱ核心启动子sequence-specific DNA结合
	0003677	DNA结合
	0003677	DNA结合
	0003700	sequence-specific DNA结合转录因子的活动
	0004996	促甲状腺激素受体活动
	0005515	蛋白质绑定
	0044212	转录调控区DNA结合

表3 配对盒基因8(PAX8)的京都基因和基因组数据库(KEGG)分析结果

项目	序列	信号通路
map05216	GO:1905332	甲状腺癌
map04918	GO:1905330	甲状腺激素合成
map05202	0007417	肿瘤的转录调节作用
map05200	0044212	肿瘤发生发展

均存在异常高表达。宋艳等[10]的研究显示,PAX8在甲状腺癌、肾细胞癌、子宫内膜腺癌等多种卫星肿瘤中存在高表达,提示PAX8在NSCLC的发生发展中具有重要意义。肿瘤GEO数据库是目前世界上最全面的储存高通量基因表达丰度数据的开源数据库,收录了多个物种,多种不同类型的高通量实验数据[11]。因此,本研究基于GEO数据库分析了PAX8的表达与NSCLC病人预后生存之间的关系。

本研究分析GEO数据库数据集(GSE77803)中PAX8的表达,发现PAX8的表达与NSCLC病人的T分期和N分期有关。T分期主要反映了肿瘤的浸润深度,N分期主要反映了肿瘤的淋巴结转移情况,是

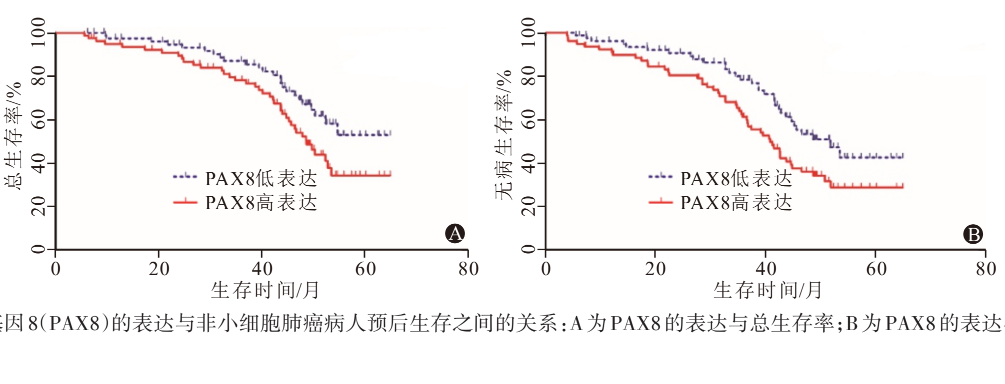

图1 配对盒基因8(PAX8)的表达与非小细胞肺癌病人预后生存之间的关系:A为PAX8的表达与总生存率;B为PAX8的表达与无病生存率

指导临床治疗的主要依据[12-13]，提示PAX8的表达与肿瘤的恶性程度有关，参与了肿瘤的发生发展。研究显示，PAX8在肥腺癌组织表达与其术后复发、转移有关，可作为评估肥腺癌根治术预后的指标，提示PAX8与NSCLC病人的临床病理特征有关，可以反映NSCLC的病情发展[14]。本研究分析GEO数据库数据集（GSE77803）中PAX8的表达，PAX8高表达NSCLC病人术后复发的比例明显高于PAX8低表达病人，提示PAX8的表达与NSCLC病人的术后复发有关。本研究发现PAX8低表达病人的总生存率和无病生存率明显高于PAX8高表达病人，提示PAX8与NSCLC病人的预后生存有关，有望作为其临床预后的分子标志物。

GO富集可以根据PAX8蛋白的序列信息，分析其可能的功能，在基因功能的分析中具有重要作用[15-16]。GO富集主要从分子功能本体论、生物学途径本体论和细胞组件本体论三个层面分析基因的功能[17]。本研究GO富集分析显示，PAX8蛋白参与机体多器官的生理生化过程，可以与相应的靶DNA结合，发挥其转录因子的调节作用。KEGG分析是一种常用的生物信息学代谢通路分析工具之一，可以分析蛋白质参与的最主要的代谢和信号转导途径[18-19]。本研究KEGG分析显示，PAX8蛋白可能参与了甲状腺癌等多种肿瘤发生发展的信号通路。Li等[20]的研究显示，PAX8蛋白在NSCLC中异常高表达，参与调节了NSCLC细胞的生长和运动，提示PAX8参与调控NSCLC的发生发展，有望成为其治疗和预后的标志物。

综上所述，PAX8蛋白可能参与调节NSCLC发生发展的信号通路，在大部分肺腺癌和肺鳞癌中均存在异常高表达，其表达与NSCLC病人的T分期、N分期以及术后复发有关，PAX8低表达病人具有更高的生存率。但PAX8蛋白作用于NSCLC发生发展的具体机制尚不清楚，有待于下一步研究。

参考文献

[1] 王永芳，张春芳，陈昊．干细胞转录因子SOX2与非小细胞肺癌的研究进展[J]．实用医学杂志，2018，34(16)：2807-2810.

[2] BRADLEY C J, YABROFF K R, MARIOTTO A B, et al. Antineoplastic treatment of advanced-stage non-small-cell lung cancer: treatment, survival, and spending (2000 to 2011)[J]. J ClinOncol, 2017, 35(5): 529-535.

[3] KAR S P, ADLER E, TYRER J, et al. Enrichment of putative PAX8 target genes at serous epithelial ovarian cancer susceptibility loci[J]. Br J Cancer, 2017, 116(4): 524-535.

[4] KILGORE MR, BOSCH DE, ADAMSON KH, et al. Unexpected PAX8 immunoreactivity in metastatic high-grade breast cancer [J]. Appl Immunohistochem Mol Morphol, 2019, 27(9): 637-643.

[5] BIE LY, LI D, WEI Y, et al. SOX13 dependent PAX8 expression promotes the proliferation of gastric carcinoma cells[J]. Artif Cells Nanomed Biotechnol, 2019, 47(1): 3180-3187.

[6] CAREY M, RAMÍREZ J C, WU S, et al. A big data pipeline: Identifying dynamic gene regulatory networks from time-course gene expression omnibus data with applications to influenza infection [J]. Stat Methods Med Res, 2018, 27(7): 1930-1955.

[7] CHUN S G, HU C, CHOY H, et al. Impact of intensity-modulated radiation therapy technique for locally advanced non-small-cell lung cancer: a secondary analysis of the NRG oncology RTOG 0617 randomized clinical trial[J]. J ClinOncol, 2017, 35(1): 56-62.

[8] YU TM, MORRISON C, GOLD EJ, et al. Multiple biomarker testing tissue consumption and completion rates with single-gene tests and investigational use of oncomine dx target test for advanced non-small-cell lung cancer: a single-center analysis[J]. Clin Lung Cancer, 2019, 20(1): 20-29.

[9] 钱萍，高宝莲．卵巢及子宫透明细胞癌、子宫内膜样癌中NapsinA、TTF-1、PAX8及CA125的表达及意义[J]．临床与实验病理学杂志，2017，33(2)：192-194.

[10] 宋艳，黄璇，沈贵华，等．配对盒基因8和配对盒基因2蛋白在上皮及肿瘤组织中的表达差异[J]．中华肿瘤杂志，2017，39(6)：424-428.

[11] LIU S, ZHANG Y, BIAN H, et al. Gene expression profiling predicts pathways and genes associated with Parkinson's disease[J]. Neurol Sci, 2016, 37(1): 73-79.

[12] IN H, SOLSKY I, PALIS B, et al. Validation of the 8th edition of the AJCC TNM staging system for gastric cancer using the national cancer database[J]. Ann Surg Oncol, 2017, 24(12): 3683-3691.

[13] TSUBOI M, HAMADA C, KATO H, et al. The effect of tegafur-uracil on survival in t categories as defined in the eighth edition of the tnm classification: an exploratory analysis of postoperative adjuvant tegafur-uracil on survival in patients with adenocarcinoma of the lung[J]. Chemotherapy, 2017, 62(6): 357-360.

[14] 翁旭，李劲松．肥腺癌组织PAX2和PAX8表达联合预测其根治术后复发、转移的价值[J]．重庆医学，2017，46(31)：4367-4369.

[15] CHEN L, ZHANG Y H, LU G, et al. Analysis of cancer-related lncRNAs using gene ontology and KEGG pathways[J]. Artif Intell Med, 2017, 76: 27-36.

[16] MI H, HUANG X, MURUGANUJAN A, et al. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements[J]. Nucleic Acids Res, 2017, 45(D1): D183-D189.

[17] PENG J, WANG T, WANG J, et al. Extending gene ontology with gene association networks[J]. Bioinformatics, 2016, 32(8): 1185-1194.

[18] KANEHISA M, SATO Y, KAWASHIMA M, et al. KEGG as a reference resource for gene and protein annotation[J]. Nucleic Acids Res, 2016, 44(D1): D457-D462.

[19] FABRIS F, FREITAS A A. New KEGG pathway-based interpretable features for classifying ageing-related mouse proteins[J]. Bioinformatics, 2016, 32(19): 2988-2995.

[20] LI J, LI H, LV X, et al. Polymorphism in lncRNA AC016683.6 and its interaction with smoking exposure on the susceptibility of lung cancer[J]. Cancer Cell Int, 2018, 18: 91.

（收稿日期：2019-09-19，修回日期：2019-10-10）