

doi: 10.3978/j.issn.2095-6959.2020.11.008

View this article at: http://dx.doi.org/10.3978/j.issn.2095-6959.2020.11.008

## HOXD10 甲基化水平与表达水平的关联及其对肺腺癌预后的影响

吴稚晖<sup>1</sup>, 袁悉奥<sup>2</sup>, 李慧源<sup>2</sup>, 龚强<sup>2</sup>, 彭千<sup>2</sup>, 吴秋<sup>2</sup>, 李文灿<sup>1</sup>

(1. 株洲市中心医院心胸外科, 湖南 株洲 412000, 2. 长沙金域医学检验实验室有限公司, 长沙 410000)

**[摘要]** 目的: 分析HOXD10基因启动子区甲基化水平与基因表达水平的关联并探索HOXD10与肺腺癌临床病理特征和预后的关系。方法: 采用R语言拓展包TCGAbiolinks进行基因差异表达分析和差异甲基化分析, 利用肿瘤基因组计划(The Cancer Genome Atlas, TCGA)项目肺腺癌(lung adenocarcinoma, LUAD)数据集中521例癌组织组织和58例癌旁组织的RNA-seq数据分析HOXD10基因差异表达, 利用同一数据集中461例癌组织组织和32例癌旁组织的甲基化芯片数据分析HOXD10启动子区差异甲基化, 并计算启动子区差异甲基化与基因表达水平的相关性, 最后分析HOXD10表达与LUAD临床病理特征和预后的关系。结果: 在TCGA项目LUAD数据集中, 521例癌组织HOXD10基因的平均表达量为30.4(normalized read counts), 低于58例癌旁组织的平均表达量170.8, 差异具有统计学意义( $P < 0.05$ ); 在癌组织和癌旁组织配对(即来源于同一患者的癌组织和癌旁组织组成配对,  $n=56$ )时HOXD10基因表达水平的平均差异值为-141.2, 差异具有统计学意义( $P < 0.05$ )。通过比较461例癌组织组织和32例癌旁组织的甲基化芯片数据, 鉴定出一个位于HOXD10基因启动子区内的差异甲基化区域, 甲基化水平癌组织高于癌旁组织( $\Delta\beta=0.3656$ ,  $fwer=0$ )。该区域位于以hg19作为参考基因组的2号染色体176980837~176982342坐标位置, 共包含12个甲基化探针, 其中cg03918304, cg10364040, cg13217260, cg20649017, cg21591742和cg25371634等6个探针的 $\beta$ 值与HOXD10基因的表达情况呈负相关( $r < -0.1$ ,  $P < 0.05$ )。250例肺腺癌患者的完整临床信息显示: HOXD10表达水平与每天抽烟根数和肿瘤大小两个临床病理指标有关( $P < 0.05$ ), HOXD10低表达组的5年中位生存时间为37.2个月, 低于高表达组的59.9个月( $P < 0.05$ )。结论: HOXD10基因在TCGA肺腺癌癌组织中的表达水平明显高于癌旁组织, 而甲基化水平明显低于癌旁组织; HOXD10的表达与肺腺癌的临床病理特征有关, 是肺腺癌预后的可能影响因素。

**[关键词]** 肺腺癌; HOXD10; 甲基化; 预后

## Association of DNA methylation and HOXD10 expression and its effect on the prognosis of lung adenocarcinoma

WU Zhihui<sup>1</sup>, YUAN Xi'ao<sup>2</sup>, LI Huiyuan<sup>2</sup>, GONG Qiang<sup>2</sup>, PENG Qian<sup>2</sup>, WU Qiu<sup>2</sup>, LI Wencan<sup>1</sup>

(1. Department of Cardiothoracic Surgery, Zhuzhou Central Hospital, Zhuzhou Hunan 412000;

2. Changsha KingMed Center For Clinical Laboratory CO., Ltd, Changsha 410000, China)

**Abstract** **Objective:** To analyze the association of DNA methylation and gene expression of HOXD10 and explore the

收稿日期 (Date of reception): 2019-12-03

通信作者 (Corresponding author): 李文灿, Email: lwccom@sina.com

基金项目 (Foundation item): 湖南省自然科学基金 (2017JJ4072). This work was supported by the Natural Science Foundation of Hunan Province, China (2017JJ4072).

relations between *HOXD10* and clinicopathological features and prognosis of lung adenocarcinoma (LUAD). **Methods:** R-language expansion package TCGAAbiolinks was used to analyze retrospectively the expression of *HOXD10* in LUAD datasets of The Cancer Genome Atlas (TCGA) project. The RNA-seq data of 521 tumor tissues and 58 adjacent tissues were collected for differentially expressed gene (DEG) analysis, and the methylation microarray data of 461 tumor tissues and 32 adjacent tissues were collected for differentially methylated region (DMR) analysis. The correlations between methylations of the promoter regions and gene expressions were calculated. The relations of *HOXD10* and clinicopathological features and prognosis were analyzed. **Results:** The average expression value of *HOXD10* in 521 tumor tissues was 30.4 (normalized read counts), lower than the value of 58 adjacent tissues (170.8), and the difference was statistically significant ( $P < 0.05$ ). The different value was  $-141.2$  ( $P < 0.05$ ) when tumor tissue and adjacent tissue from the same patient were compared ( $n = 56$ ). A DMR (chr2: 176980837–176982342) containing 12 probes was identified via differentially methylated region analysis (461 tumor tissues vs 32 adjacent tissues,  $\Delta\beta = 0.3656$ ,  $fwer = 0$ ). The  $\beta$  values of six probes (cg03918304, cg10364040, cg13217260, cg20649017, cg21591742 and cg25371634) were negatively correlated with the expressions of *HOXD10* ( $r < -0.1$ ,  $P < 0.05$ ). From clinical information of 250 LUAD patients, the expression of *HOXD10* was associated with number of cigarettes per day and tumor volume ( $P < 0.05$ ). The 5-year median survival time of *HOXD10* low-expression group was 37.2 months, lower than 59.9 months of *HOXD10* high-expression group ( $P < 0.05$ ). **Conclusion:** The *HOXD10* expressions of tumor tissues in TCGA LUAD datasets are significantly higher than adjacent tissues while the methylation  $\beta$  values are significantly lower than adjacent tissues. The expression of *HOXD10* is related to clinicopathological features of LUAD and could be a potential prognostic factor.

**Keywords** lung adenocarcinoma; *HOXD10*; methylation; prognosis

肿瘤基因组计划(The Cancer Genome Atlas, TCGA)启动于2005年,旨在运用高通量测序和生物信息技术促进人类对于癌症遗传机制的认知,从而提高人类对于癌症的防治能力。TCGA目前由美国国家癌症研究所和美国国家人类基因组研究所(the National Human Genome Research Institute)共同管理和运行,储存33种癌症相关逾2万例样本,产生了约2.5 PB数据,涵盖了基因组、表观组、转录组和蛋白质组等数据类型<sup>[1]</sup>,并持续向学术界开放数据使用权限。TCGA数据分为4个层次(level): level 1为原始数据; level 2为经过初步处理的数据; level 3为经过分割、解释的数据; level 4为热点区域或概要。其中level 1和level 2数据的获取需通过严格审批。

腺癌是肺癌最普遍的组织学类型,占手术切除肺癌的50%~70%<sup>[2]</sup>。不同于鳞状细胞肺癌,肺腺癌属于非小细胞癌,起源于支气管黏膜上皮,少数起源于大支气管黏液腺,是腺样分化或黏液分泌的恶性上皮肿瘤,表现为腺泡状、乳头状、细支气管肺泡性或实性伴有黏液的生长方式,或者

上述表现混合存在。与肺鳞状细胞癌相比,肺腺癌在分子水平的表征亦有差异,其中PD-L1基因的表达水平明显较低<sup>[3]</sup>。

DNA甲基化与基因表达关系的相关研究由来已久<sup>[4]</sup>。从19世纪末至20世纪初,二者的关联关系逐渐被癌症研究领域的科研工作者所关注。虽然DNA甲基化在正常哺乳动物发育过程中调节基因的表达,但是启动子区的过甲基化会导致重要生长发育相关基因,如肿瘤抑制基因的转录沉默,进而成为引发癌症的主要原因之一<sup>[5]</sup>。

*HOXD10*(homeobox D10)基因,位于2号染色体长臂3区1带1亚带(2q31.1),是Abd-B homeobox家族成员,含有2个外显子,编码1个含有homeobox DNA结合域的蛋白质,在细胞分化和形态发生过程中起关键作用。已有多个研究报道了*HOXD10*在结肠癌<sup>[6-7]</sup>、肝癌<sup>[8]</sup>、前列腺癌<sup>[9]</sup>、胆管细胞癌<sup>[10-11]</sup>等多种癌症中抑制肿瘤生长。亦有研究<sup>[12]</sup>采用PCR分析33名中国患者的胃部肿瘤组织-正常组织(tumor-normal pair),初步证实*HOXD10*是一个潜在的抑癌基因,其启动子过甲基化导致

基因表达下调并与胃癌发生有关联。

本研究旨在运用常规的生物信息学和统计学分析手段,对TCGA肺腺癌数据集中高通量RNA测序(RNA-seq)数据和甲基化芯片数据进行关联分析,并结合临床病理信息进一步探索HOXD10基因启动子甲基化调控基因表达水平与肺腺癌预后的关系。

## 1 资料与方法

### 1.1 数据下载

利用R语言拓展包TCGAbiolinks<sup>[13]</sup>下载存储于美国国家癌症研究所癌症数据共享系统Genomic Data Commons中的TCGA肺腺癌数据集,包括临床病理数据、RNA-seq数据和甲基化芯片数据。其中临床病理数据包含522名患者的性别、年龄、身高、体重、种族、BMI、生存状况、烟龄、每天抽烟根数、肿瘤分期、肿瘤转移情况、肿瘤形态学、放射治疗情况及化学药物治疗情况等临床信息;RNA-seq数据为原始数据经过HTSeq<sup>[14]</sup>软件处理得到的read counts数据(level 3),包含癌症组织样本521例和癌旁组织样本58例;甲基化芯片数据为Illumina Human Methylation 450k芯片β值数据(level 3),包含癌症组织样本461例和癌旁组织样本32例。

### 1.2 统计学处理

利用R语言拓展包TCGAbiolinks<sup>[15]</sup>进行基因差异表达分析和差异甲基化分析时采用的阈值分别为 $|\log FC| > 1$ 且错误发现率(false discovery rate, FDR) $< 0.01$ ,  $|\Delta\beta| > 0.3$ 且fwer $< 0.1$ 。癌症和癌旁组织基因差异表达情况和差异甲基化情况的验证采用t检验,基因表达水平与启动子区甲基化水平相关性分析采用Pearson相关系数,临床病理特征单因素分析采用Fisher检验,生存分析采用Kaplan-Meier法。以 $P < 0.05$ 为差异有统计学意义。

## 2 结果

### 2.1 HOXD10 在 TCGA 肺腺癌 (配对) 癌症组织和癌旁组织中的表达情况

TCGAbiolinks的分析结果显示:HOXD10基因在肺腺癌癌症组织的表达量低于癌旁组织,差异有统计学意义( $P < 0.05$ )。TCGA肺腺癌数据集中

RNA-seq数据中HOXD10基因表达量的验证结果显示:521例癌症组织样本和58例癌旁组织样本的平均表达量分别为30.4(normalized read counts,下同)和170.8,差异有统计学意义( $P < 0.05$ ,图1)。

配对癌症组织和癌旁组织(即来源于同一患者的癌症和癌旁组织组成配对, $n = 56$ )HOXD10基因表达量的平均差异值为-141.2,差异具有统计学意义( $P < 0.05$ )。

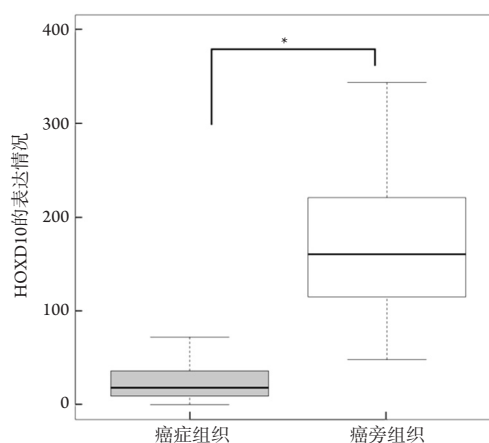


图1 TCGA肺腺癌HOXD10在癌症和癌旁组织中的表达情况  
Figure 1 Expression of HOXD10 in tumor tissues and adjacent tissues of TCGA lung adenocarcinoma

### 2.2 HOXD10 启动子区在 TCGA 肺腺癌 (配对) 癌症和癌旁组织中的甲基化水平

运用TCGAbiolinks和R语言拓展包minfi对461例癌症组织和32例癌旁组织的甲基化芯片数据进行差异甲基化分析,鉴定出一个位于HOXD10基因启动子区内的差异甲基化区域,癌症组织中的甲基化程度高于癌旁组织,差异具有统计学意义( $P < 0.05$ )。该区域位于以hg19作为参考基因组的2号染色体176980837~176982342位置,共包含12个甲基化探针,ID分别是cg02249732, cg03918304, cg05979020, cg10125317, cg10364040, cg13217260, cg17489939, cg18115040, cg20649017, cg21591742, cg25371634和cg25953239。非配对(肿瘤组织=461,癌旁组织=32)和配对( $n = 29$ )的癌症组织中12个探针甲基化β值均高于癌旁组织,差异具有统计学意义( $P < 0.05$ ,图2)。

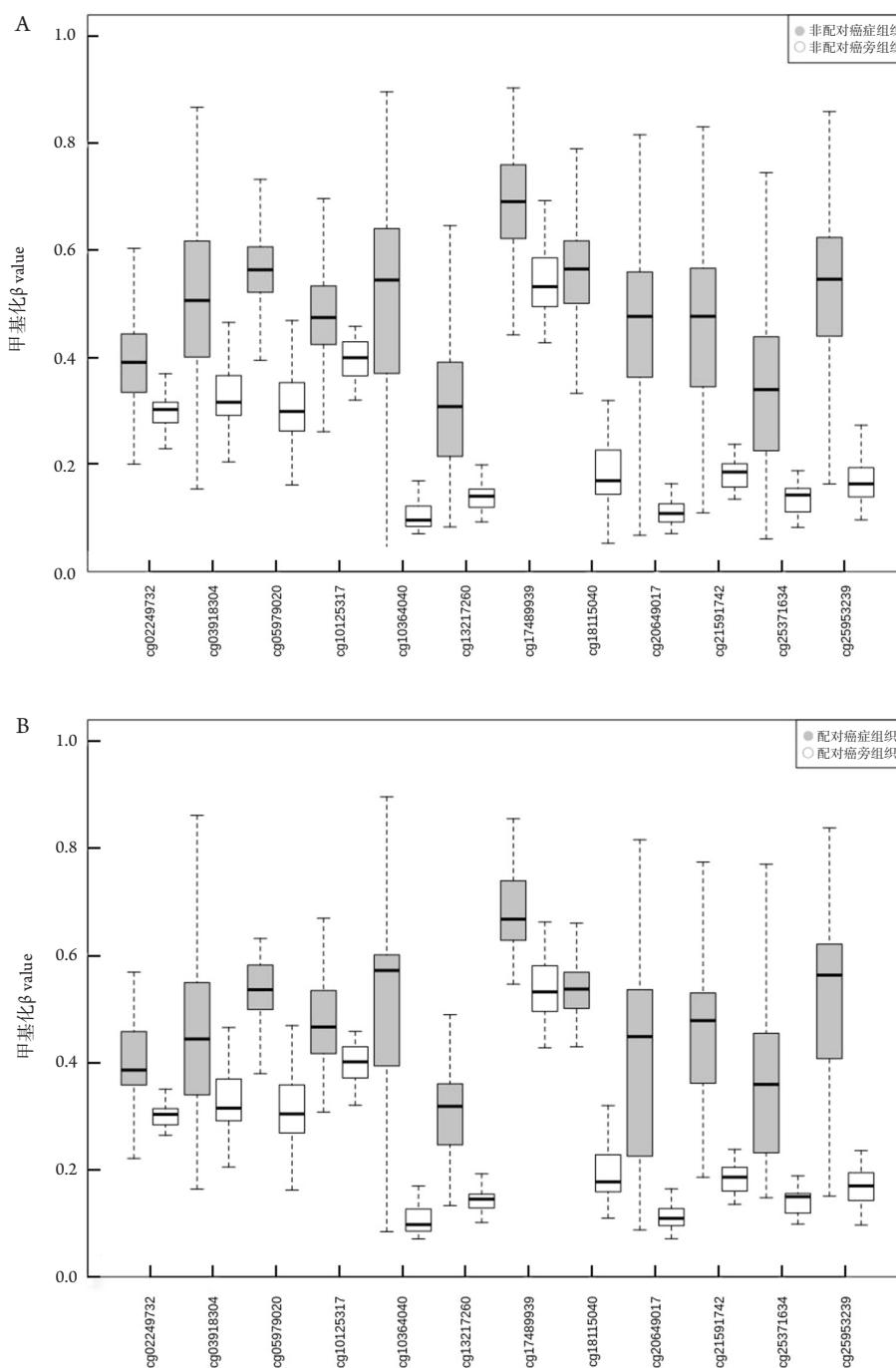


图2 TCGA肺腺癌 $HOXD10$ 启动子区非配对(A)和配对(B)癌组织和癌旁组织的甲基化情况

Figure 2 Methylation in  $HOXD10$  promoter region of unmatched (A) and matched (B) tumor tissues and adjacent tissues of TCGA lung adenocarcinoma

### 2.3 $HOXD10$ 的表达水平与启动子区甲基化水平的关系

对 $HOXD10$ 的表达水平与启动子区甲基化水平进行相关性分析, 结果显示: 12个探针中有6个探针(表1)的甲基化 $\beta$ 值与基因表达水平呈负相关( $r < -0.1$ ,  $P < 0.05$ , 图3)。

### 2.4 $HOXD10$ 的表达水平与临床病理特征的关系

TCGA肺腺癌数据集中包含522名患者的临床病理信息, 以患者群体肿瘤组织中 $HOXD10$ 表达情况的下四分位数(9)和上四分位数(36)作为阈值, 将患者群体划分为 $HOXD10$ 低表达组( $n=119$ )

和HOXD10高表达组( $n=131$ )。不同HOXD10表达水平与患者临床病理特征的关系见表2, 结果显示HOXD10表达情况与每天抽烟根数和肿瘤大小有关( $P<0.05$ )。

## 2.5 HOXD10 的表达水平与肺腺癌预后的关系

5年生存分析显示: HOXD10低表达组中位生存时间为37.2个月, 低于高表达组的59.9个月( $P<0.05$ , 图4)。

表1 与HOXD10表达水平呈负相关的甲基化探针信息

Table 1 Information of methylation probes whose value were negatively correlated with the expression of HOXD10

探针ID	物理位置	正负链	$r$	$P$
cg03918304	176981654	+	-0.252	$6.343 \times 10^{-8}$
cg10364040	176981064	+	-0.123	0.009
cg13217260	176981469	-	-0.185	$8.089 \times 10^{-5}$
cg20649017	176980837	-	-0.118	0.012
cg21591742	176980998	-	-0.103	0.028
cg25371634	176981422	+	-0.162	0.0006

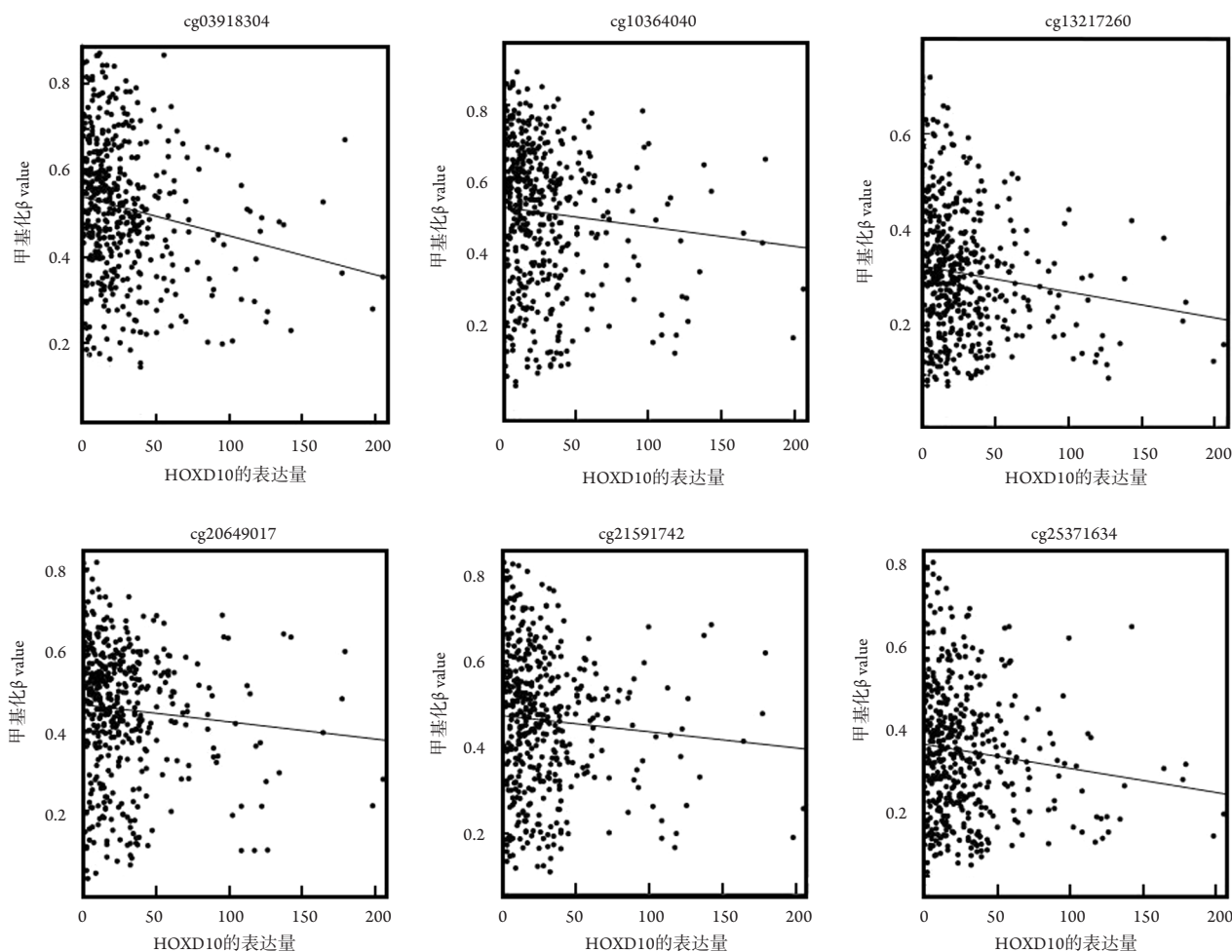


图3 HOXD10的表达量与启动子区甲基化水平的相关性

Figure 3 Correlation of HOXD10 expression and methylation in promoter region

表2 肺癌患者HOXD10表达情况与临床病理特征的关系

Table 2 Relation of expression of HOXD10 and clinicopathological features of lung adenocarcinoma patients

组别	性别		年龄/岁		每天抽烟根数		烟龄/年		肿瘤大小/cm		远处转移		邻近淋巴结转移		肿瘤分期	
	男	女	<60	≥60	<3	≥3	<20	≥20	<3	≥3	阴性	阳性	阴性	阳性	I+II	III+IV
HOXD10低表达组	57	62	10	102	64	21	6	39	32	87	76	5	74	43	90	29
HOXD10高表达组	49	82	10	115	79	12	13	34	50	79	88	5	92	35	109	19
P	0.098		0.8191		0.0257		0.123		0.0485		—		0.1382		0.0762	

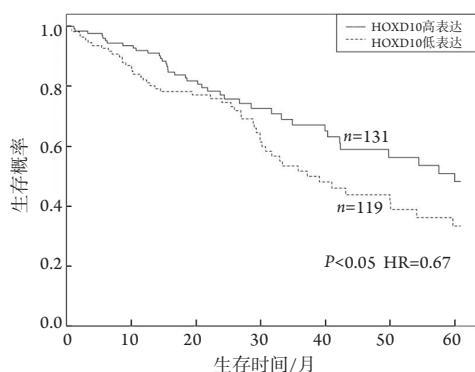


图4 HOXD10高表达组与低表达组5年生存曲线

Figure 4 Five-year survival curve of HOXD10 high-expression and low-expression group

### 3 讨论

历时13年的TCGA项目于2018年4月公布了收官之作：泛癌症图谱(Pan-Cancer Atlas)。该研究成果涉及基因组测序、转录组测序、甲基化等表观组学测序以及最终的整合分析，同时研究者也将它们与临床和影像数据相关联。泛癌症图谱将研究成果分为3个领域：肿瘤细胞起源(cell-of-origin patterns)<sup>[16]</sup>、致癌过程(oncogenic processes)<sup>[17]</sup>和癌症信号通路(signaling pathways)<sup>[18]</sup>。这些研究成果展示了过去的十多年中，TCGA项目科学家如何从分子特征角度对癌症进行研究。TCGA项目是人类疾病研究领域中大規模采样并进行实验与分析项目的典范，项目产生的数据和研究成果将持续为整个癌症领域提供大量极具参考价值的信息。

本研究通过TCGA肺癌数据集中RNA-seq数据和甲基化芯片数据进行关联分析，并辅以临床病理信息，发现无论是在非配对还是配对情况下，TCGA肺癌癌症组织中HOXD10的表达水平均低于癌旁组织，同时癌症组织中HOXD10基因

启动子区12个探针位置的甲基化β值均高于癌旁组织。但HOXD10表达水平与启动子区甲基化水平相关性分析中相关系数r值的结果未能反映出较强的负相关关系，故HOXD10表达水平与启动子区甲基化水平的负相关关系有待进一步探究。基因表达水平与预后关系分析显示：HOXD10表达是影响肺癌预后的可能因素之一，但本研究仅使用了TCGA肺癌数据集，无论是样本数量还是样本来源都存在一定局限性，所以HOXD10要作为分子标记应用于临床仍需要其他研究计划的大规模数据分析结果给以支持，此外，生物信息学相关分析结论，在应用于临床之前都需要经过严谨的多层次实验验证。

HOXD10在多种癌症中扮演抑癌基因的角色，但在不同癌种中的抑癌机制却不尽相同<sup>[6-12]</sup>，这表明HOXD10参与了多个抑癌调控通路并在整个抑癌调控网络中处于相对重要的位置。但是，该基因所处多个调控通路中上、下游的调控机制仍然有待完善。此外，关于该基因功能紊乱与癌症发展发生的直接证据尚不充足，仍需相关基础性研究，所以HOXD10相关临床应用如作为癌症治疗用药靶点的时机还远未成熟。

综上，本研究通过对TCGA项目肺癌数据集进行生物信息学和统计学分析，发现HOXD10基因的表达水平和甲基化水平在癌症和癌旁组织中均存在明显差异，基因表达在一定程度上受其基因启动子区甲基化调控，并发现HOXD10的表达可能是肺癌预后的影响因素，对今后肺癌预后相关研究有一定的理论参考价值。

### 参考文献

1. 谢龙祥, 闫中义, 党艺方, 等. TCGA数据库:海量癌症数据的源泉[J]. 河南大学学报(医学版), 2018, 37(3): 223-228.  
XIE Longxiang, YAN Zhongyi, DANG Yifang, et al. TCGA: the source

- of massive amounts of cancer data[J]. Journal of Henan University. Medical Science, 2018, 37(3): 223-228.
- Asamura H, Goya T, Koshiishi Y, et al. A Japanese lung cancer registry study: prognosis of 13010 resected lung cancers[J]. J Thorac Oncol, 2008, 3(1): 46-52.
  - 黄桃, 刘勇. 第三十届欧洲病理学大会纪要——肺病理学[J]. 临床与实验病理学杂志, 2019, 35(11): 1261-1267.  
HUANG Tao, LIU Yong. Abstracts: 30th European congress of pathology—lung pathology[J]. Chinese Journal of Clinical and Experimental Pathology, 2019, 35(11): 1261-1267.
  - Razin A, Cedar H. DNA methylation and gene expression[J]. Microbiol Rev, 1991, 55(3): 451-458.
  - Baylin SB. DNA methylation and gene silencing in cancer[J]. Nat Clin Pract Oncol, 2005, 2(Suppl 1): S4-S11.
  - Yuan YH, Wang HY, Lai Y, et al. Epigenetic inactivation of HOXD10 is associated with human colon cancer via inhibiting the RHOC/AKT/MAPK signaling pathway[J]. Cell Commun Signal, 2019, 17(1): 9.
  - 郝昌磊, 龚治林, 周启昌, 等. HOXD10基因抑制结肠直肠癌细胞生长的机制研究[J]. 中国免疫学杂志, 2018, 34(12): 50-54.  
XI Changlei, GONG Zhilin, ZHOU Qichang, et al. Mechanism of HOXD10 gene inhibiting on growth of colorectal cancer cells[J]. Chinese Journal of Immunology, 2018, 34(12): 50-54.
  - Guo Y, Peng Y, Gao D, et al. Silencing HOXD10 by promoter region hypermethylation activates ERK signaling in hepatocellular carcinoma[J]. Clin Epigenetics, 2017, 9: 116.
  - Mo RJ, Lu JM, Wan YP, et al. Decreased HoxD10 expression promotes a proliferative and aggressive phenotype in prostate cancer[J]. Curr Mol Med, 2017, 17(1): 70-78.
  - Yang H, Zhou J, Mi J, et al. HOXD10 acts as a tumor-suppressive factor via inhibition of the RHOC/AKT/MAPK pathway in human cholangiocellular carcinoma[J]. Oncol Rep, 2015, 34(4): 1681-1691.
  - 周九鹏, 王英霞. HOXD10在肝外胆管癌中的表达及其临床意义[J]. 实用医学杂志, 2018, 34(18): 30-34.  
ZHOU Jiupeng, WANG Yingxia. Expression and clinical significance of HOXD10 in extrahepatic cholangiocarcinoma[J]. The Journal of Practical Medicine, 2018, 34(18): 30-34.
  - Wang L, Chen S, Xue M, et al. Homeobox D10 gene, a candidate tumor suppressor, is downregulated through promoter hypermethylation and associated with gastric carcinogenesis[J]. Mol Med, 2012, 18: 389-400.
  - Colaprico A, Silva TC, Olsen C, et al. TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data[J]. Nucleic Acids Res, 2016, 44(8): e71.
  - Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data[J]. Bioinformatics, 2015, 31(2): 166-169.
  - Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays[J]. Bioinformatics, 2014, 30(10): 1363-1369.
  - Hoadley KA, Yau C, Hinoue T, et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer[J]. Cell, 2018, 173(2): 291-304.
  - Ding L, Bailey MH, Porta-Pardo E, et al. Perspective on oncogenic processes at the end of the beginning of cancer genomics[J]. Cell, 2018, 173(2): 305-320.
  - Sanchez-Vega F, Mina M, Armenia J, et al. Oncogenic signaling pathways in The Cancer Genome Atlas[J]. Cell, 2018, 173(2): 321-337.

本文引用: 吴稚晖, 袁悉奥, 李慧源, 龚强, 彭千, 吴秋, 李文灿. HOXD10甲基化水平与表达水平的关联及其对肺腺癌预后的影响[J]. 临床与病理杂志, 2020, 40(11): 2851-2857. doi: 10.3978/j.issn.2095-6959.2020.11.008

**Cite this article as:** WU Zhihui, YUAN Xi'ao, LI Huiyuan, GONG Qiang, PENG Qian, WU Qiu, LI Wencan. Association of DNA methylation and HOXD10 expression and its effect on the prognosis of lung adenocarcinoma[J]. Journal of Clinical and Pathological Research, 2020, 40(11): 2851-2857. doi: 10.3978/j.issn.2095-6959.2020.11.008