Prediction of the diagnostic and prognostic values of homeobox family in lung cancer using bioinformatics

TINGTING HU
Traditional Chinese Medicine Hospital Affiliated to Xinjiang Medical University

PanJun Gao
Traditional Chinese Medicine Hospital Affiliated to Xinjiang Medical University

Min Jiang
Traditional Chinese Medicine Hospital Affiliated to Xinjiang Medical University

Zheng Li
Traditional Chinese Medicine Hospital Affiliated to Xinjiang Medical University

Dan Xu
Traditional Chinese Medicine Hospital Affiliated to Xinjiang Medical University

Jing Jing
Traditional Chinese Medicine Hospital Affiliated to Xinjiang Medical University

JING WANG (jingw_xj@163.com)
Traditional Chinese Medicine Hospital Affiliated to Xinjiang Medical University

Research Article

Keywords: gene signatures, lung cancer, homeobox family, prognosis, diagnosis

Posted Date: February 17th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1255514/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Background: Lung cancer is a frequently diagnosed malignancy and accounts for the main reason of cancer-related deaths worldwide due to the less of reliable diagnostic markers and the potent therapeutic targets. Therefore, exploration of the molecular mechanisms underlying the occurrence and development via systems biology approaches provides a promising option. Here, we aimed to reveal genes which could be used as the diagnostic and prognostic markers in lung cancer, aiming to find new efficient target for treatment using bioinformatics method.

Results: We adopted a series of bioinformatics methods to identify genes which are implicated in the progression of lung cancer. From the raw RNA microarray data obtained from the database of UCSC-Xena, we identified that a total of 241 differently expressed genes (DEGs), including 8 HOX (homeobox) genes in all types of lung cancer tissues (Stage I-IV) as compared to the normal lung tissues. The gene ontology (GO) analysis showed that the 241 DEGs were mainly enriched in xenobiotic glucuronidation, regionalization, GABA receptor complex, and extracellular ligand-gated ion channel activity. Gene expression profiling interactive analysis (GEPIA) demonstrated that the expression levels of HOXA11, HOXA13, HOXB9, HOXB13, HOXC11, HOXC13, HOXD11 and HOXD13 were all obviously higher in lung adenocarcinoma and squamous cell lung carcinoma tissues than in normal lung tissues. Kaplan-Meier survival analysis and operating characteristic analyses (ROC) suggested that the high expression of HOXA11/A13/B9/B13/C11/C13/D11/D13 all predicted poor prognosis of lung cancer and had high specificity and sensitivity in distinguishing lung cancer patients from healthy individuals.

Conclusion: This study using bioinformatics methods identified that 8 genes (HOXA11/A13/B9/B13/C11/C13/D11/D13) of HOX family might play an important role in lung cancer progression with high value in the diagnosis and prognostic prediction in lung cancer. However, their role in lung cancer progression still need to be verified using the fundamental experiments.

Introduction

Lung cancer is a frequently diagnosed malignancy and is the leading cause of cancer-related deaths worldwide [1, 2]. Based on the global cancer statistics, 1,824,700 new lung cancer cases were identified, causing approximately 1,589,900 deaths in 2012 [3]. Non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) are the two types of lung cancer, and NSCLC accounts for approximately 85% of all lung cancer cases [4]. Despite advancements in diagnostic and treatment methods, the 5-year overall survival rate for patients with lung cancer still remains less than 15% [5]. Therefore, development of efficient diagnosis markers and drug targets should be put on the agenda to improve the prognosis of patients with lung cancer.

The development of high-throughput molecular tools including microarrays and large-scale databases of serial analysis of gene expression and expressed sequence tags have identified that lots of genes are deregulated in lung cancer and closely implicated in the carcinogenesis [6, 7]. However, it still inevitably
encounter many obstacles in lung cancer research due to some of the limitations in microarray technology, such as small sample size, measurement error and insufficient information [8]. Therefore, it is of importance to further identify genes which are valuable in early diagnosis and treatment of lung cancer.

The homeobox (HOX) genes family is a group of transcription factors which includes 39 members in mammals. Although the expression of HOX genes is typically suppressed in adults, they are necessary for body repair and various homeostatic cellular processes [9, 10], including wound healing and vascularization [11], hematopoiesis [12], and female reproductive tract development and fertility [13], as well as cell survival and differentiation [14]. Noticeably, accumulated evidences have identified that HOX genes are deregulated in many kinds of cancers including lung cancer, where the deregulation of HOX exerts an oncogenic or a suppressive role in carcinogenesis [15-17], suggesting a vital role of HOX genes play in cancer progression.

In this study, we carried out the bioinformatics to integrate the data of lung cancer in UCSC-Xena database, aiming to further identify the genes related to the prognosis and progression of lung cancer. Our results revealed that 8 genes (HOXA11/A13/B9/B13/C11/C13/D11/D13) of HOX family might play an important role in lung cancer progression with high value in the diagnosis and prognostic prediction in lung cancer. The obtained results might be beneficial to the diagnostics and therapeutics of lung cancer.

Materials And Methods

Data collection

Raw RNA microarray data ‘TCGA Lung Cancer (LUNG)’ was obtained from the database of UCSC-Xena, which was obtained by using the sequencing platform of IlluminaHiSeq [https://xenabrowser.net/datapages/]. A total of 1006 lung cancer samples and 110 normal lung samples were included in this database, among which 521 cases were at stage I, 284 cases were at stage II, 168 cases were at stage III and 33 cases were at stage IV. The process analysis diagram was shown in Figure 1.

Identification of differentially expression genes (DEGs)

The DEGs between lung cancer tissue and normal lung tissue specimens were identified by using the R software (http://www.R project.org/version 3.4.2) limma (version 3.32.3) package [18]. Once P value < 0.05 and the absolute value of Fold Chang (FC) >1, genes were considered as DEGs. R software ggplot2 package (http://cran.r-project.org/web/packages/gplots/index.html) was used to build Volcano plot filtering and Hierarchical clustering to seek DEGs with statistical significance among atherosclerotic plaque samples and the normal arterial intima specimens [19].

Functional enrichment analysis
Gene ontology (GO) analysis, covering biological processes (GO-BP), cellular component (GO-CC), and molecular functions (GO-MF) three domains were used to evaluate the enriched function of the total DEGs using R software clusterProfiler. Fisher’s exact test was applied to detect the overlap among the DEGs and the GO annotation list beyond that which would be expected by chance. The P value < 0.05 was considered to be significantly enriched among DEGs.

Moreover, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was also executed to predict the functional enrichment of the DEGs by using Cytoscape plug-in BinGO and Database for annotation, visualization, and integration [20].

The ‘Broad Institute Web Platform for Genome Networks’ (GeNets, http://apps.broadinstitute.org/genets) was used to predict the gene-gene interaction of the DEGs.

**Gene expression profiling interactive analysis (GEPIA)**

GEPIA (http://gepia.cancer-pku.cn) [21] is a newly developed interactive web server which provides help in analyzing the RNA sequencing expression data of 9,736 tumors and 8,587 normal samples from the TCGA and the GTEx projects. GEPIA offers several customizable function analysis, such as tumor/normal differential expression analysis, survival analysis, profiling based on cancer types or pathological stages, similar gene detection and correlation analysis. As such, we applied GEPIA to analyze the different expression levels of HOX genes in different stages of lung cancer tissue, and to explore the relationship between the expression levels of HOX genes and the overall survival (OS), progression-free survival (FP), and post progression survival (PPS) for patients with lung cancer.

In detail, the prognostic value of HOXs mRNA expression was evaluated using an online database, Kaplan-Meier Plotter (www.kmplot.com), from which we can obtain the information about gene expression data and survival of 2,437 clinical lung cancer patients (http://kmplot.com/analysis/index.php?p=service&cancer=lung). To analyze the OS, FP, and PPS, lung cancer samples were divided into two groups (high vs. low expression) based on the median expression level, followed by being assessed using a Kaplan-Meier survival plot with the hazard ratio (HR) and a 95% confidence intervals (CI) and a log rank test. In addition, the receiver operating characteristic analyses (ROC) and logistic regression were carried out to assess discriminatory accuracy of lung cancer samples from the healthy individuals.

**Gene set enrichment analysis (GSEA)**

GSEA (http://www.broadinstitute.org/gsea) help investigators to interpret chip results more easily and reasonably. In the present study, GSEA was performed on the DEGs using two gene sets, KEGG gene set and the GO gene set. The false discovery rates (FDRs) <25% for the gene sets were considered as significantly enriched gene sets.

**Results**
Identification of the DEGs among the lung cancer tissues and normal lung tissues

To further search genes related to the occurrence and development of lung cancer, we first identified the DEGs among the different stages of lung cancer cases and the normal lung tissue samples. Compared with the normal lung tissues, 365 genes were found to be upregulated and 40 genes were downregulated in the stage I of lung cancer cases (Figure 2A); 404 genes were found to be upregulated and 65 genes were downregulated in the stage II of lung cancer cases (Figure 2B); 409 genes were found to be upregulated and 69 genes were downregulated in the stage III of lung cancer cases (Figure 2C); and 293 genes were highly expressed and 37 genes were lowly expressed in the stage IV of lung cancer cases (Figure 2D). The total DEGs were also demonstrated by using the clustering heat map between stage I-IV of lung cancer cases and the normal lung tissues (Figure 2E). In addition, we also applied the Venn image to integrate the DEGs among the different stages of lung cancer tissues and normal lung tissues. The results showed that a total of 241 DEGs were found between stage I-IV of lung cancer cases and the normal lung tissues (Figure 2F). In this process, we found 8 genes in the HOX family, including HOXA11, HOXA13, HOXB13, HOXB9, HOXC11, HOXC13, HOXD11 and HOXD13 were all significantly upregulated in all stages of lung cancer as compared to the normal lung tissues (Table 1), suggesting that HOX family might play an important role in the progression of lung cancer.

GO analysis of the DEGs

To predict the potential roles of the DEGs in the progression of lung cancer, GO analysis was performed on the 241 DEGs. GO analysis showed that the 241 DEGs were mainly enriched in xenobiotic glucuronidation, regionalization, pattern specification process and cellular glucuronidation in the term of biological processes (Figure 3A-B); GABA receptor complex, GABA-A receptor complex, chloride channel complex in term of cellular component (Figure 3C-D); and extracellular ligand-gated ion channel activity, GABA receptor activity and glucuronosyltransferase activity in molecular function term (Figure 3E-F). The HOX family was mainly enriched in the term of proximal promoter DNA-binding transcription activator activity, RNA polymerase II-specific, regionalization, pattern specification process, epidermis development and proximal/distal pattern formation.

KEGG and GeNets analysis of the DEGs

The KEGG analysis carried out by the cytoscape plug-in BinGO showed that the DEGs were enriched in ascorbate and aldarate metabolism, pentose and glucuronate and nicotine addiction pathways (Figure 4A). In addition, from the GeNets analysis, we found that the 241 DEGs were enriched in 11 communities, among which HOXC11 and HOXD11 played a role in community 6 while HOXA13 played its role in community 2 (Figure 4B).

GEPIA of the expression patterns of HOX family in different stages of lung cancer

Using GEPIA dataset, we compared the mRNA expression patterns of HOX factors between lung cancer tissues and normal lung tissues. The results demonstrated that the expression levels of HOXA11,
HOXA13, HOXB9, HOXB13, HOXC11, HOXC13, HOXD11 and HOXD13 were all obviously higher in lung adenocarcinoma and squamous cell lung carcinoma tissues than in lung tissues (Figure 5A-B). We also analyzed the expression profiles of HOXs in different stages of lung cancer. HOXC11 significantly varied, whereas HOXA11, HOXA13, HOXB9, HOXB13, HOXC13, HOXD11 and HOXD13 groups did not significantly differ in different stages of lung cancer (Figure 5C).

**The relationship between the expression patterns of HOXs and the prognosis of patients with lung cancer**

The Kaplan–Meier curve and log-rank test analyses revealed that the increased HOXA11/A13/B9/B13/C11/C13/D11/D13 mRNA levels were significantly associated with the shorter FP; HOXA11/A13/B9/B13/C11/D11/D13 mRNA levels were significantly associated with shorter the OS, and HOXD11 mRNA levels were significantly associated with PPS (P < 0.05) in patients with lung cancer (Figure 6). These results indicated that HOXs have a high value in predicting the prognosis of lung cancer.

**Prediction of the diagnostic value of HOXs in lung cancer**

The ROC analysis showed that the 8 HOXs used to separate different stages of lung cancer patients from healthy controls had high specificity and sensitivity (Figure 7). In detail, the area under curve (AUC) of HOXA11 to distinguish lung cancer at stage I, stage II, stage III and stage IV form healthy controls were 0.802, 0.86, 0.892 and 0.753, respectively; for HOXA13 were 0.81, 0.828, 0.803 and 0.704; for HOXB9 were 0.794, 0.827, 0.852 and 0.803; for HOXB13 were 0.841, 0.805, 0.839 and 0.628; for HOXC11 were 0.837, 0.844, 0.847 and 0.698; for HOXC13 were 0.887, 0.881, 0.892 and 0.804; for HOXD11 were 0.798, 0.844, 0.825 and 0.757; and for HOXD13 were 0.8, 0.805, 0.768 and 0.729, respectively (Table 2).

**GESA of the DEGs in stages IV of lung cancer**

Compared with the normal lung tissues, the GESA GO analysis showed that 316 gene sets were enriched at nominal P value < 1% (143 up and 173 down) in stage II lung cancer tissues, and the upregulated gene sets were enriched in platelet morphogenesis and positive regulation of autophagy, while downregulated gene sets were mainly enriched in protein ubiquitination involved in ubiquitin dependent protein catabolic process (Figure 8A-C). Compared with stage I, 191 gene sets in stage II were significantly enriched at nominal P value < 1% (75 up, 116 down), and the upregulated genes were enriched in cellular protein complex assembly and macromolecule transmembrane transporter activity, and the downregulated genes were enriched in cell division site (Figure 8D-F). Compared with stage II, 191 gene sets in stage III were significantly enriched at nominal P value < 1% (8 up, 12 down) (Figure 8G-I). Compared with stage I, 191 gene sets in stage II were significantly enriched at nominal P value < 1% (75 up, 116 down). 27 gene sets in stage IV were significantly enriched at nominal P value < 1% compared with the stage III (23 up, 4 down) (Figure 8J-L).

In addition, the GESA KEGG pathway analysis showed that 31 gene sets were significantly enriched at nominal P value < 5% in stage II of lung cancer as compared to normal lung tissues (19 up, 12 down).
(Figure 9A-C); 22 gene sets were significantly enriched at nominal P value < 5% in stage II of lung cancer as compared to stage I lung tissues (13 up, 9 down) (Figure 9D-F); 6 gene sets were significantly enriched at nominal P value < 5% in stage III of lung cancer as compared to stage II lung tissues (3 up, 3 down) (Figure 9G-I); and 5 gene sets were significantly enriched at nominal P value < 1% (4 up and 1 down) in stage IV lung cancer tissues compared to stage III lung cancer (Figure 9J-L).

**Discussion**

Up to now, the application of high-throughput techniques in lung cancer generated a larger amount of data. Based on these data, many genes have also been identified. But it remains incompletely clear that the roles of these genes play in lung cancer progression. This study integrated the database of lung cancer from UCSC-Xena, which contains 1006 lung cancer samples and 110 normal lung samples, and a total of 241 DEGs were identified between lung cancer cases (stage I-stage IV) and the normal lung tissues. Among the 241 DEGs, 8 genes from the HOX family got us attention, including HOXA11, HOXA13, HOXB13, HOXB9, HOXC11, HOXC13, HOXD11 and HOXD13, which were all significantly upregulated in all stages of lung cancer as compared to the normal lung tissues.

There are 39 HOX genes in mammals, which are divided into four linear clusters, Hox A, B, C and D with 9 to 11 genes in each cluster. The four clusters are located on different chromosomes, 7p15.3 for HOXA, 17q21.3 for HOXB, 12q13.3 for HOXC and 2q31 for HOXD in humans. In this study, we revealed two HOXA genes (HOXA11 and HOXA13), two HOXB gene (HOXB9 and HOXB13), two HOXC genes (HOXC11 and HOXC13) and two HOXD genes (HOXD11 and HOXD13) were significantly overexpressed in all stages of lung cancer tissues compared with the normal lung tissues, as demonstrated by both the UCSC-Xena and GEPIA databases. Noticeably, the majority of identified HOX genes have been known to be functionally crucial in lung cancer progression [22].

HOXA11 and HOXA13 belong to the HOXA cluster of HOX gene family. HOXA11, also known as HOX11, HOX1 or radioulnar synostosis with amegakaryocytic thrombocytopenia 1, plays a vital role in several kinds of cancers, such as gastric cancer [23], renal cell carcinoma [24] and glioblastoma [25]. Yang et al. [26] reported that HOXA11 was overexpressed in the lung adenocarcinoma samples. Zhang et al. [27] also found that HOXA11 was significantly overexpressed in lung adenocarcinoma samples as compared to the adjacent non-cancer tissues with the area under curves (AUC) of 0.955, suggesting that HOXA11 has a high clinical value in lung cancer diagnosis. The expression pattern of HOXA13 has been reported to be strongly implicated in the progression of gastric cancer [28], prostate cancer [29] and ovarian cancer [30]. Deng at al. [31] found that HOXA13 was overexpressed in lung adenocarcinoma based on data from The Cancer Genome Atlas, Oncomine and reverse transcription-quantitative polymerase chain reaction (RT-PCR). Ectopic expression of the long non-coding HOTTIP facilitated the progression of non-small cell lung cancer (NSCLC) via decreasing HOXA13 expression [32], indicating that HOXA13 might function as a tumor suppressor in NSCLC.
HOXB9, also known as a direct transcriptional target of WNT/transcription factor 4 (TCF4), is closely implicated in the modulation of several cellular activities, such as cell proliferation, differentiation and apoptosis. HOXB9 served as an oncogene in breast cancer via inducing epithelial-mesenchymal transition (EMT) [33]. HOXB9 was overexpressed in lung cancer, and the high expression level was closely associated with the poor prognosis of lung adenocarcinoma [34]. Similarly, the high expression of HOXB13 predicted a poor prognosis in lung adenocarcinoma, and upregulation of HOXB13 increased the expression of several metastasis- and drug-resistance-related genes, including ABCG1, EZH2, and Slug [35], suggesting that HOXB13 plays a crucial role in the metastasis and drug-resistance of lung cancer.

HOXC11 was found to be strongly related to the clinical prognosis of patients with renal cancer, cervical cancer, and breast cancer [36-38], and HOXC11 upregulation significantly promoted renal cancer cell proliferation [38]. In NSCLC, HOXC11 negatively regulated by miR-1197 reversed miR-1197 role in promoting cell proliferation and migration [39]. HOXC13 was reported to function as an oncogenic role in lung adenocarcinoma via modulation of the expression of CCND1 and CCNE1 [40]. Moreover, Sharpe et al. [41] reported that POU2F1 promoted the progression of head and neck squamous cell carcinoma via downregulating HOXD10 and HOXD11. However, HOXD11 role in lung cancer still remains unknown. The methylation status of HOXD13 is a potential marker and may be utilized for the diagnosis and therapy of lung adenocarcinoma [42].

In conclusion, this study identified 8 genes (HOXA11/A13/B9/B13/C11/C13/D11/D13) of HOX family overexpressed in lung cancer and they might play crucial roles in lung cancer progression with high value in the diagnosis and prognostic prediction in lung cancer. Nevertheless, their roles in lung cancer progression still need to be verified using the fundamental experiments.

Abbreviations

DEGs: Differently expressed genes; SCLC: small cell lung cancer; NSCLC: non-small cell lung cancer; GO: gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; BP: biological processes; CC: cellular component; MF: molecular functions; HOX: homeobox; GEPIA: Gene expression profiling interactive analysis; GSEA: Gene set enrichment analysis.

Declarations

Acknowledgements

Not applicable

Authors’ contributions

MIN JIANG analyzed and interpreted the PPI networks. PANJUN GAO performed the analysis of DEGs; TINGTING HU were a major contributor in writing the manuscript. All authors read and approved the final
manuscript.

**Funding**

This work was supported by Xinjiang Uygur Autonomous Region Health and Health Youth Medical Science and Technology Talent Special Project. The funding bodies did not play any roles in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

**Availability of data and materials**

Raw RNA microarray data ‘TCGA Lung Cancer (LUNG)’ was obtained from the database of UCSC-Xena, which was obtained by using the sequencing platform of IlluminaHiSeq [https://xenabrowser.net/datapages/](https://xenabrowser.net/datapages/).

**Ethics approval and consent to participate**

This article does not deal with ethics, so it does not apply.

**Declaration**

I confirm that all methods are carried out in accordance with relevant guidelines and regulations.

**Consent for publication**

Not applicable

**Competing interests**

The authors declare that they have no competing interests.

**References**

1. Sun CC, Li SJ, Li G, Hua RX, Zhou XH and Li DJ. Long Intergenic Noncoding RNA 00511 Acts as an Oncogene in Non-small-cell Lung Cancer by Binding to EZH2 and Suppressing p57. Mol Ther Nucleic Acids 2016; 5: e385.
2. Siegel RL, Miller KD and Jemal A. Cancer statistics, 2019. CA Cancer J Clin 2019; 69: 7-34.
3. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J and Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015; 65: 87-108.
4. Heist RS and Engelman JA. SnapShot: non-small cell lung cancer. Cancer Cell 2012; 21: 448 e442.
5. Sun C, Liu Z, Li S, Yang C, Xue R, Xi Y, Wang L, Wang S, He Q, Huang J, Xie S, Jiang W and Li D. Down-regulation of c-Met and Bcl2 by microRNA-206, activates apoptosis, and inhibits tumor cell proliferation, migration and colony formation. Oncotarget 2015; 6: 25533-25574.
6. Kim B, Lee HJ, Choi HY, Shin Y, Nam S, Seo G, Son DS, Jo J, Kim J, Lee J, Kim K and Lee S. Clinical validity of the lung cancer biomarkers identified by bioinformatics analysis of public expression data. Cancer Res 2007; 67: 7431-7438.

7. Li C, Yin Y, Liu X, Xi X, Xue W and Qu Y. Non-small cell lung cancer associated microRNA expression signature: integrated bioinformatics analysis, validation and clinical significance. Oncotarget 2017; 8: 24564-24578.

8. Zhang Q, Sun S, Zhu C, Zheng Y, Cai Q, Liang X, Xie H and Zhou J. Prediction and analysis of weighted genes in hepatocellular carcinoma using bioinformatics analysis. Mol Med Rep 2019; 19: 2479-2488.

9. Dunwell TL and Holland PW. Diversity of human and mouse homeobox gene expression in development and adult tissues. BMC Dev Biol 2016; 16: 40.

10. Rux DR and Wellik DM. Hox genes in the adult skeleton: Novel functions beyond embryonic development. Dev Dyn 2017; 246: 310-317.

11. Kachgal S, Mace KA and Boudreau NJ. The dual roles of homeobox genes in vascularization and wound healing. Cell Adh Migr 2012; 6: 457-470.

12. Alharbi RA, Pandha HS, Simpson GR, Pettengell R, Poterlowicz K, Thompson A, Harrington K, El-Tanani M and Morgan R. Inhibition of HOX/PBX dimer formation leads to necroptosis in acute myeloid leukemia cells. Oncotarget 2017; 8: 89566-89579.

13. Du H and Taylor HS. The Role of Hox Genes in Female Reproductive Tract Development, Adult Function, and Fertility. Cold Spring Harb Perspect Med 2015; 6: a023002.

14. Gehring WJ and Hiromi Y. Homeotic genes and the homeobox. Annu Rev Genet 1986; 20: 147-173.

15. Wang X, Sun Y, Xu T, Qian K, Huang B, Zhang K, Song Z, Qian T, Shi J and Li L. HOXB13 promotes proliferation, migration, and invasion of glioblastoma through transcriptional upregulation of lncRNA HOXC-AS3. J Cell Biochem 2019; 120: 15527-15537.

16. Chen SL, Qin ZY, Hu F, Wang Y, Dai YJ and Liang Y. The Role of the HOXA Gene Family in Acute Myeloid Leukemia. Genes (Basel) 2019; 10:

17. Ni S, Ye M and Huang T. Short stature homeobox 2 methylation as a potential noninvasive biomarker in bronchial aspirates for lung cancer diagnosis. Oncotarget 2017; 8: 61253-61263.

18. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015; 43: e47.

19. Dean CB and Nielsen JD. Generalized linear mixed models: a review and some extensions. Lifetime Data Anal 2007; 13: 497-512.

20. Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC and Lempicki RA. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 2003; 4: P3.

21. Tang Z, Li C, Kang B, Gao G and Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res 2017; 45: W98-W102.
22. Li L, Zhang X, Liu Q, Yin H, Diao Y, Zhang Z, Wang Y, Gao Y, Ren X, Li J, Cui D, Lu Y and Liu H. Emerging role of HOX genes and their related long noncoding RNAs in lung cancer. Crit Rev Oncol Hematol 2019; 139: 1-6.

23. Bai Y, Fang N, Gu T, Kang Y, Wu J, Yang D, Zhang H, Suo Z and Ji S. HOXA11 gene is hypermethylation and aberrant expression in gastric cancer. Cancer Cell Int 2014; 14: 79.

24. Wang L, Cui Y, Sheng J, Yang Y, Kuang G, Fan Y, Jin J and Zhang Q. Epigenetic inactivation of HOXA11, a novel functional tumor suppressor for renal cell carcinoma, is associated with RCC TNM classification. Oncotarget 2017; 8: 21861-21870.

25. Se YB, Kim SH, Kim JY, Kim JE, Dho YS, Kim JW, Kim YH, Woo HG, Kang SH, Kim HJ, Kim TM, Lee ST, Choi SH, Park SH, Kim IH, Kim DG and Park CK. Underexpression of HOXA11 Is Associated with Treatment Resistance and Poor Prognosis in Glioblastoma. Cancer Res Treat 2017; 49: 387-398.

26. Yang X, Deng Y, He RQ, Li XJ, Ma J, Chen G and Hu XH. Upregulation of HOXA11 during the progression of lung adenocarcinoma detected via multiple approaches. Int J Mol Med 2018; 42: 2650-2664.

27. Zhang R, Zhang TT, Zhai GQ, Guo XY, Qin Y, Gan TQ, Zhang Y, Chen G, Mo WJ and Feng ZB. Evaluation of the HOXA11 level in patients with lung squamous cancer and insights into potential molecular pathways via bioinformatics analysis. World J Surg Oncol 2018; 16: 109.

28. Wu DC, Wang SSW, Liu CJ, Wuputra K, Kato K, Lee YL, Lin YC, Tsai MH, Ku CC, Lin WH, Wang SW, Kishikawa S, Noguchi M, Wu CC, Chen YT, Chai CY, Lin CS, Kuo KK, Yang YH, Miyoshi H, Nakamura Y, Saito S, Nagata K and Yokoyama KK. Reprogramming Antagonizes the Oncogenicity of HOXA13-Long Noncoding RNA HOTTIP Axis in Gastric Cancer Cells. Stem Cells 2017; 35: 2115-2128.

29. Dong Y, Cai Y, Liu B, Jiao X, Li ZT, Guo DY, Li XW, Wang YJ and Yang DK. HOXA13 is associated with unfavorable survival and acts as a novel oncogene in prostate carcinoma. Future Oncol 2017; 13: 1505-1516.

30. Kelly Z, Moller-Levet C, McGrath S, Butler-Manuel S, Kavitha Madhuri T, Kierzek AM, Pandha H, Morgan R and Michael A. The prognostic significance of specific HOX gene expression patterns in ovarian cancer. Int J Cancer 2016; 139: 1608-1617.

31. Deng Y, He R, Zhang R, Gan B, Zhang Y, Chen G and Hu X. The expression of HOXA13 in lung adenocarcinoma and its clinical significance: A study based on The Cancer Genome Atlas, Oncomine and reverse transcription-quantitative polymerase chain reaction. Oncol Lett 2018; 15: 8556-8572.

32. Sang Y, Zhou F, Wang D, Bi X, Liu X, Hao Z, Li Q and Zhang W. Up-regulation of long non-coding HOTTIP functions as an oncogene by regulating HOXA13 in non-small cell lung cancer. Am J Transl Res 2016; 8: 2022-2032.

33. Chiba N, Comaills V, Shiotani B, Takahashi F, Shimada T, Tajima K, Winokur D, Hayashida T, Willers H, Brachtel E, Vivanco MD, Haber DA, Zou L and Maheswaran S. Homeobox B9 induces epithelial-to-mesenchymal transition-associated radioresistance by accelerating DNA damage responses. Proc Natl Acad Sci U S A 2012; 109: 2760-2765.
34. Zhan J, Wang P, Niu M, Wang Y, Zhu X, Guo Y and Zhang H. High expression of transcriptional factor HoxB9 predicts poor prognosis in patients with lung adenocarcinoma. Histopathology 2015; 66: 955-965.

35. Zhan J, Wang P, Li S, Song J, He H, Wang Y, Liu Z, Wang F, Bai H, Fang W, Du Q, Ye M, Chang Z, Wang J and Zhang H. HOXB13 networking with ABCG1/EZH2/Slug mediates metastasis and confers resistance to cisplatin in lung adenocarcinoma patients. Theranostics 2019; 9: 2084-2099.

36. Eoh KJ, Kim HJ, Lee JY, Nam EJ, Kim S, Kim SW and Kim YT. Upregulation of homeobox gene is correlated with poor survival outcomes in cervical cancer. Oncotarget 2017; 8: 84396-84402.

37. McIlroy M, McCartan D, Early S, P OG, Pennington S, Hill AD and Young LS. Interaction of developmental transcription factor HOXC11 with steroid receptor coactivator SRC-1 mediates resistance to endocrine therapy in breast cancer [corrected]. Cancer Res 2010; 70: 1585-1594.

38. Liu YJ, Zhu Y, Yuan HX, Zhang JP, Guo JM and Lin ZM. Overexpression of HOXC11 homeobox gene in clear cell renal cell carcinoma induces cellular proliferation and is associated with poor prognosis. Tumour Biol 2015; 36: 2821-2829.

39. Sun B, Hua J, Cui H, Liu H, Zhang K and Zhou H. MicroRNA-1197 downregulation inhibits proliferation and migration in human non-small cell lung cancer cells by upregulating HOXC11. Biomed Pharmacother 2019; 117: 109041.

40. Yao Y, Luo J, Sun Q, Xu T, Sun S, Chen M, Lin X, Qian Q, Zhang Y, Cao L, Zhang P and Lin Y. HOXC13 promotes proliferation of lung adenocarcinoma via modulation of CCND1 and CCNE1. Am J Cancer Res 2017; 7: 1820-1834.

41. Sharpe DJ, Orr KS, Moran M, White SJ, McQuaid S, Lappin TR, Thompson A and James JA. POU2F1 activity regulates HOXD10 and HOXD11 promoting a proliferative and invasive phenotype in head and neck cancer. Oncotarget 2014; 5: 8803-8815.

42. Han L, Xu G, Xu C, Liu B and Liu D. Potential prognostic biomarkers identified by DNA methylation profiling analysis for patients with lung adenocarcinoma. Oncol Lett 2018; 15: 3552-3557.

Tables

**Table 1** The significant changes of HOXs expression in transcription level between different stages of lung cancer and normal lung tissues.
Table 2 HOXs gene ROC AUC

| Gene   | Stage I | Stage II | Stage III | Stage IV |
|--------|---------|----------|-----------|----------|
| HOXA11 | 0.802   | 0.86     | 0.892     | 0.753    |
| HOXA13 | 0.81    | 0.828    | 0.803     | 0.704    |
| HOXB9  | 0.794   | 0.827    | 0.852     | 0.803    |
| HOXB13 | 0.841   | 0.805    | 0.839     | 0.628    |
| HOXC11 | 0.837   | 0.844    | 0.847     | 0.698    |
| HOXC13 | 0.887   | 0.881    | 0.892     | 0.804    |
| HOXD11 | 0.798   | 0.844    | 0.825     | 0.757    |
| HOXD13 | 0.8     | 0.805    | 0.768     | 0.729    |

Figures
Figure 1

The process analysis diagram of the present study.

Figure 2

Identification of the DEGs of lung cancer tissues from normal lung tissues. (A-D) Volcano plot showed the DEGs between the stage I-IV of lung cancer cases and normal lung tissues (blue, downregulated DEGs; red, upregulated DEGs). (E) The clustering Heat map displayed the DEGs between the stage I-IV of lung cancer cases and normal lung tissues. (F) Venn image demonstrated the integrated DEGs between the stage I-IV of lung cancer cases and normal lung tissues.

Figure 3

GO analysis of the differently expressed mRNAs. (A-B) GO-BP analysis and the gene regulation related to biological processes. (C-D) GO-CC analysis and the gene regulation related to cellular components. (E-F) GO-MF analysis and the gene regulation involved in molecular functions. GO, Gene ontology.
Figure 4

KEGG and GeNets analysis of the DEGs. (A) KEGG pathway analysis of the 241 DEGs. (B) CeNets analysis of the gene-gene interaction targeting the 241 DEGs.

Figure 5

GEPIA was used to assess the expression patterns of HOX family in different types and stages of lung cancer. (A-B) The expression patterns of HOX family in lung adenocarcinoma and lung squamous cell carcinoma. (C) The expression profiles of HOX family in different stages of lung cancer. LUAD: lung adenocarcinoma; LUSC: lung squamous cell carcinoma.

Figure 6

The prognostic value of the expression levels of HOXs in lung cancer patients. Kaplan-Meier plotter with log rank test was used to predict the relationship between the mRNA levels of HOXs and the PF, OS and PPS of patients with lung cancer. PF: progression-free survival; OS: overall survival; PPS: post progression survival.

Figure 7

The diagnostic value of HOXs in different stages of lung cancer. ROC curve of 8 HOXs (HOXA11/A13/B9/B13/C11/C13/D11/D13) in discrimination between (A) stage I, (B) stage II, (C) stage III, and (D) stage IV of lung patients and healthy controls.

Figure 8

GSEA GO analysis was used to predict the function of DEGs between different stages of lung cancer and normal lung tissues. (A-C) GSEA GO analysis of the different gene sets between normal lung tissues and stage II lung cancer. (D-F) GSEA GO analysis of the different gene sets between stage I and stage II lung cancer. (G-I) GSEA GO analysis of the different gene sets between stage II and stage III lung cancer. (J-L) GSEA GO analysis of the different gene sets between stage III and stage IV lung cancer.

Figure 9
GSEA KEGG pathway analysis was used to predict the function of DEGs between different stages of lung cancer and normal lung tissues. (A-C) GSEA KEGG analysis of the different gene sets between normal lung tissues and stage II lung cancer. (D-F) GSEA KEGG analysis of the different gene sets between stage I and stage II lung cancer. (G-I) GSEA KEGG analysis of the different gene sets between stage II and stage III lung cancer. (J-L) GSEA KEGG analysis of the different gene sets between stage III and stage IV lung cancer.