Iroquois homeobox gene 2 as a novel prognostic marker that dysregulated in lung adenocarcinoma

Zhen Huang
University of Electronic Science and Technology of China
Ping Xiao
Hunan Cancer Hospital, & The Affiliated Cancer Hospital of Xiangya, Central South University
Bin Qu
Hunan Cancer Hospital, & The Affiliated Cancer Hospital of Xiangya, Central South University
Lisa Sun (44028204@qq.com)
Hunan Cancer Hospital, & The Affiliated Cancer Hospital of Xiangya, Central South University

Research Article

Keywords: Iroquois homeobox 2, Lung adenocarcinoma, Homeobox genes, Prognostic marker

Posted Date: May 20th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1650352/v1

Abstract

Lung adenocarcinoma (LUAD), is one of the deadliest and most common types of cancer. Homeobox-containing genes are evolutionary conserved regulators that play important roles in embryogenesis and tumorigenesis. We discovered that seven dysregulated homeobox-containing genes with prognostic value in LUAD. Furthermore, we showed that Iroquois homeobox (IRX)2 is enriched in normal lung tissues but dysregulated in LUAD. *IRX*2 is an independent prognostic marker in LUAD in univariate and multivariate Cox proportional and further confirmed in Kaplan-Meier survival analyses. Furthermore, patients with long-term survival had significantly higher *IRX*2 expression levels in public LUAD dataset. Meanwhile, methylation level of *IRX*2 promoter region was increased in TCGA (The Cancer Genome Atlas) and was linked to a decreased overall survival rate in patients. The results of Gene Set Enrichment Analysis (GSEA) revealed that *IRX*2-related genes were enriched humoral immune and also in tumorigenesis pathways such as G2/M checkpoint and mTOR pathways. Early research suggested that IRXs play a role in lung morphogenesis, but their pathological role in LUAD is unknown. These findings suggest that IRX2 dysregulation may play a role in the development of LUAD tumors and could be used as a novel prognostic marker for LUAD patients.

Introduction

According to GLOBOCAN (Global Cancer Statistics) in 2018, lung cancer is the most commonly diagnosed cancer (11.6% of the total cases) and the leading cause of cancer death (18.4% of the total cancer deaths) globally [1]. Based on histology, lung cancer could be classified into two major groups: non-small-cell lung cancer (NSCLC) and small-cell lung cancer (SCLC) [2]. Non-small cell lung cancer accounts for approximately 85 percent of all lung cancers, with small cell lung cancer accounting for the remaining 15 percent [3]. NSCLC can be further subdivided into lung adenocarcinoma (LUAD), squamous cell carcinoma (LUSC), and large cell carcinoma. The most common histological subtype of NSCLC is LUAD, which is originated from small airway epithelial, type II alveolar cells that secrete mucus and other substances [4]. However, the tumorigenesis of LUAD is not fully elucidated.

Homeobox-containing genes are an evolutionary conserved gene family that encodes transcription factors involved in cell fate determination and tissue morphogenesis [5]. Dysregulation of these genes in tumors may result in abnormal cell proliferation and oncogenesis [6]. Homeobox-containing genes have been discovered to master lung epithelial morphogenesis and differentiation, such as *Nkx*2-1 [7–9]. In the meantime, they may act as tumor suppressors or oncogenes in lung adenocarcinoma, the most common type of NSCLC (Non-small cell lung cancer). Researchers also found that *Nkx*2-1 inhibits Kras(G12D)-driven mucinous pulmonary adenocarcinoma in a context-dependent manner in lung tumorigenesis [10]. For instance, *PITX* (Paired Like Homeodomain) are development-related transcription factors and regulates the formation and symmetry of organs, including the lung, by controlling growth control genes after the Wnt/β-catenin signaling pathway activation [11, 12]. In non-small-cell lung cancer patients, high DNA methylation of the homeobox gene *PITX*2 predicts poor outcome [13]. *PITX*1 was found to be

reduced in human lung cancer [14]. The expression balance of homeobox-containing genes may be important for LUAD tumorigenesis.

In this study, we first screened and obtained 14 differential-expressed candidates from a total of 238 distinct protein-coding homeobox genes in five different LUAD datasets using a bioinformatic approach. We found that the Iroquois homeobox gene family (IRXs), particular *IRX*2, is dysregulated in various LUAD datasets and associated with univariate overall-survival (OS), disease-specific survival (DSS) and progression free interval (PFI), and multivariate analysis in TCGA-LUAD. Previous studies showed that *Irx*s also play an important role in lung epithelial development and morphogenesis [16–18]. We also found this gene highly enriched expressed in normal lung (GTEx) and alveochar cells (HPA). Researchers showed that dysregulation of *IRX*1 and *IRX*4 in various types of cancer including LUAD [19, 20]. However, clinical correlation with lung cancer of *IRX*2 have not yet been full explored. Here in this study, we showed the family correlation with lung cancer and pave the light for the further validation as a target for lung cancer.

Material And Methods

Human homeobox genes collections

The Homeobox Database (http://homeodb.zoo.ox.ac.uk), a manually curated database for collecting and presenting homeobox genes, was used to collect 333 human homeobox genes along with related pseudogenes [21, 22]. Meanwhile, 229 unique human genes with homeobox domains were also downloaded from the curated protein family database Pfam (http://pfam.xfam.org/) using the Homeodomain annotation (PF00046) [23]. Following careful curation, 238 human protein-coding homeobox genes with their chromosomal locations were identified and detailed in Table-S1.

Data sources and differential expression analysis

The UCSC Cancer Browser (https://xena.ucsc.edu/) was used to download the pre-compiled RNA-seq expression matrix of TCGA-LUAD (The Cancer Genome Atlas lung adenocarcinoma) with 497 tumor and 54 normal tissue samples [24, 25]. To validate the expression patterns further, we searched the GEO (Gene Expression Omnibus) database for lung adenocarcinoma, and expression matrix with more than 10 LUAD clinical samples was chosen for further analysis [26]. We used TCGA-LUAD, GSE18842 [27], GSE19188 [28], GSE32863 [29] and GSE40419 [30] in the study. Table S2 contains detailed information about the datasets, such as the platform and sample size. Differential Expression Genes (DEGs) were identified using limma (version 3.46.0) for microarray [31] or the edgeR package for RNA-seq (version 3.32.1) [32] in the R statistical computing environment 4.0.0.

Cox proportional hazards analysis

Univariate and multivariate analyses were used to assess the relationship between the expression of the DE homeobox genes of interest and OS (Overall Survival), DSS (Disease-specific Survival) and PFS

(Progression free survival). To perform multivariate analysis, clinical characteristics were used as an additional variable. The statistical data were calculated using the Cox proportional hazards regression model and the survival (3.3-1; Therneau, 2021) and survminer package (0.4.9; Alboukadel Kassambara, 2021).

Kaplan-Meier survival analysis

Gene names were loaded into online database Kaplan Meier plotter to investigate the association of gene expression data with survival information of a total of 719 lung adenocarcinoma cancer patients [34].

GEPIA analysis

The differential expressions were validated in the GEPIA (Gene Expression Profiling Interactive Analysis) database, a web server for cancer and normal gene expression profiling and interactive analyses using the LUAD and LUSC RNA-seq datasets with $|Log_2FC|$ over 1 and *P*.Value less than 0.01 [35].

UALCAN analysis

The protein expression analysis was performed in UALCAN (University of Alabama Cancer Database), a portal for facilitating tumor subgroup gene expression and survival Analyses [36], using CPTAC (Clinical Proteomic Tumor Analysis Consortium) LUAD proteomic dataset [37]. The promoter methylation analysis was performed also in UALCAN using TCGA-LUAD dataset.

Human Protein Atlas analysis

The HPA (Human Protein Atlas) database was used to investigate IRX2 expression patterns in normal tissues in the tissue-based map of the human proteome and tumor samples in the pathology atlas of the human cancer transcriptome [38,39]

Linkedomics analysis

We used the Linkedomics database (http://linkedomics.org), which analyzes multi-omics data within and across 32 cancer types, to calculate the correlations among promoter methylation levels, overall-survival status and mRNA expressions of *IRX*2 in the TCGA-LUAD or CPTAC-LUAD datasets.

Results

Differential expressed homeobox-containing genes in lung adenocarcinoma

We gathered 238 human protein-coding homeobox-containing genes to investigate expression patterns in five different lung adenocarcinoma (LUAD) clinical sample datasets: TCGA-LUAD (N = 59, T = 517), GSE18842 (N = 45, T = 46), GSE19188 (N = 65, T = 91), GSE32863 (N = 58, T = 58), GSE40419 (N = 77, T = 87) (N: Normal; T: Tumor) (Figure 1A-E). Following intersection of the differentially expressed genes (DEGs) from each dataset ($log_2FC > 0.4$ or < -0.4, adjusted *P*.Value < 0.05), seven significantly up-

regulated candidates were found in tumor samples including *PITX*1(Paired-like homeodomain) 1, *SIX*1/4 (Sine oculis homeobox), *BARX*1 (Homeobox protein BarH-like), *HOXA*10(Homeobox protein Hox-A), *HOXB7*(Homeobox protein Hox-B) and *SATB*2(Special AT-rich sequence-binding protein) (Figure 1F). Seven others were significantly down-regulated including *IRX*2 (Iroquois homeobox), *PKNOX*2 (PBX/knotted 1 homeobox), *MEIS*1/2 (Meis homeobox), *HLX* (H2.0-like homeobox), *HHEX* (Hematopoietically-expressed homeobox protein) and *HOXA*5 (Figure 1B). According to the classification of homeobox genes, *PITX1* belongs to the PRD (Paired Domain) class, *SIX1/4* to the SINE class, *SATB2* to the CUT class, *BARX1, HHEX, HLX1, HOXA5, HOXA10, HOXB7* to the ANTP class, *IRX2, PKNOX2,* and *MEIS1/2* to the TALE (Three Amino Acid Loop Extension) class. Figure 1F depicted the $\log_2$FC heatmap of 14 differentially expressed homeobox genes in these datasets.

Prognostic significance of homeobox-containing gene in LUAD

The univariate Cox proportional hazards regression analysis was then used to examine the relationship between the gene expression value of these DE homeobox genes and the Overall Survival (OS) in 515 TCGA LUAD patients. In univariate cox analysis result, we discovered that 7 genes (*PITX1*, *SIX*1, *IRX*2, *HOXA*10, *HOXB7*, *PITX*1, *SATB*2, *PKNOX*2) were significantly associated with LUAD patients OS (*P-Value* < 0.05). As shown in Figure 2A, the survival prognosis forest map of univariate Cox analysis, *IRX*2, *PKNOX*2 and *SIX*1 are protective factors (Hazard Ratio <1, *P-Value* < 0.05) while others were risk factors (Hazard Ratio >1, *P-Value* < 0.05). Subsequently, we performed lasso regression and all the seven genes had coefficients that were not zero with 1000 repeats (Figure 2B and 2C). Risk score was calculated by the lasso coefficients: (0.0261 × HOXA10) + (-0.0365 × *HOXC*6) + (0.0357 × *PITX*1) + (-0.0857 × *SIX*1) + (-0.0581 × *IRX*2) + (-0.0732 × *PKNOX*2) (Figure 2D). The lasso Cox regression risk score results could be used to predict the OS (HR = 2.63 [1.80-3.81], P-Value = 5.145e-07) (Figure 2E), Disease-specific survival (DSS) (HR = 2.63 [1.8-3.8], P-Value = 1.974e-05) (Figure 2E) and Progression-Free interval (PFI) in TCGA LUAD cohort (Figure 2E-F). The ROC curves were then used to assess the prognostic accuracy of the model. The AUC was 0.714 (1-year), 0.653 (3-year), and 0.636 (5-year) in TCGA-LUAD dataset (Figure 2H). In validation dataset GSE37745, the AUC was 0.583 (1-year), 0.605 (3-year), and 0.702 (5-year) (Figure 2I).

Prognostic significance of IRX2 in LUAD

We then aimed to investigate the prognostic significance of *IRX*2 in lung adenocarcinoma patients. Cox analysis results suggested that *IRX*2 using age, gender, pathological stage, pathological M and T is also an independent risk factor in univariate (HR = 0.84 [0.770-0.932], *P-Value* < 0.001) and multivariate analysis (HR = 0.876 [0.794-0.967], *P-Value* < 0.001) (Figure 2A). Furthermore, we used Kaplan-Meier Plotter for lung cancer to confirmed the prognosis significances of *IRX*s. We found that expression levels of *IRX*2 (228462_at) were significantly associated with OS and FP (Free progression) in LUAD (HR = 0.50 [0.39-0.64], *P-Value* = 2.1e-07, HR = 0.51 [0.37-0.70], *P-Value* = 3.3e-05) in Kaplan-Meier Plotter for lung cancer (Figure 4B). Also, we showed in TCGA LUAD RNA-seq datasets, *IRX*2 expressions were also associated with OS (HR = 0.55 [0.41-0.74], *P-Value* = 5.0e-05) (Figure 4B). According to the evidence, *IRX*2 could be used as a diagnostic homeobox gene for LUAD. However, although *IRX*2 is downregulated both

in LUAD and LUSC (lung squamous cell carcinoma tumor) (Figure 4A), but it is a risk factor for LUSC (lung squamous cell carcinoma) (HR = 1.48 [1.09-2.02], *P-Value* = 0.012) in the Kaplan-Meier plotter (Figure 4A). This could be used to demonstrate the differences in molecular expression levels between LUAD and LUSC. Furthermore, we also further performed univariate Cox analysis associated with the DSS (Disease-specific survival) and PFI (Progression-free interval). *IRX*2 was the most significant DE homeobox gene associated with OS (hazard ratio = 0.90 [0.85-0.95], P-Value = 0.00011), DSS (hazard ratio = 0.88 [0.82-0.94], P-Value = 0.00028), and PFI (hazard ratio = 0.93 [0.88-0.98, P-Value = 0.0064) in LUAD (Figure 2B and 2C).

Loss of IRX2 expression in lung adenocarcinoma

We further validated the *IRX*2 down-regulations in GEPIA using TCGA-LUAD (n = 483) and LUSC (n = 486) with matched TCGA normal and GTEx data (n = 347) (Figure 4A). In the RNA-seq of blood platelets from non-small cell lung carcinoma patients (n = 60) and healthy donors (n = 59) from the E-GEOD-68086 dataset, *IRX*2 mRNA level was decreased in cancer patients (log_2FC = 2.70, *P-Value* = 2.76e-07) (Figure 4B). Meanwhile, protein level for IRX2 in LUAD was also suppressed. According to the proteomic expression profile provided by ULCAN based CPTAC (Clinical Proteomic Tumor Analysis Consortium)-LUAD dataset, the total protein level of IRX2 was decreased in tumor samples (n = 111) (*P-Value* = 8.067e-04) (Figure 4C). According to the individual cancer stage, the IRX2 protein level was significantly lower in Stage 1 (*P-Value* = 4.920e-02) and Stage 2 (*P-Value* = 8.504e-03) but not Stage 3 (*P-Value* = 5.451e-02) compared to normal tissues (Figure 4D). Furthermore, IRX2 protein was also lost in LUAD samples with higher tumor grade, Normal vs Grade 2 (*P-Value* = 3.696e-02), Normal vs Grade 3 (*P-Value* = 1.294e-04), Grade 1 vs Grade 2 (*P-Value* = 2.068e-02) and Grade 2 vs Grade 3 (*P-Value* < 1e-12) (Figure 4E). Furthermore, we used the IHC (Immunohistochemistry) data from HPA (Human Protein Atlas) database to explore the expression patterns of IRX2 protein. It was weakly expressed in macrophages but not detected in alveolar cells in normal lungs (Figure 4F, upper panel). Its expression was not detected in 7 of 11 HPA lung cancer samples (Figure 4F, lower panel). In the bulk tissue gene expression profile from GTEx dataset, however, *IRX*2 expression (ENSG00000170561.12) is enriched in normal lung (Figure S2A). In HPA single cell type specificity analysis result, the *IRX*2 expression was enriched in groups including alveolar cell type1 and type2 (Figure S2B). Taken together, these findings suggested that the IRX2 protein was enrich in normal lung but frequently reduced in LUAD.

IRX2 promoter methylation level increased in LUAD

Gene hypermethylation can silence gene expression and regulate biological processes, particularly tumor suppressor genes in cancer tissues. We calculated the methylation levels of 11 CpG sites (cg09524455, cg26504021, cg19679633, cg02135861, cg13702053, cg21093166, cg11552694, cg08204280, cg11793269, cg08235864, cg04992127) around the *IRX*2 genomic promoter region in the TCGA-LUAD dataset using the UALCAN database. The higher average methylation levels were found in primary tumor (n = 473) than normal samples (n = 32) (*P*.Value = 1.624e-12) (Figure 5A). More importantly, the *IRX*2 promoter region has a high methylation level in individual cancer stages when compared to normal

tissues (Normal vs Stage 1: *P*.Value = 1.623e -12; Stage 2: *P*.Value = 6.447e-12; Stage 3: P.Value = 2.483e-07; Stage 4 : *P-Value* = 9.143e-03) (Figure 5B). Using LinkedOmics database, we found that patients with high levels of *IRX*2 promoter methylation have a worse prognosis than those with low levels of *IRX*2 promoter methylation (HR = 1.239, *P*.Value = 0.0109) (Figure 5C). Using Pearson's correlation analysis, we discovered that *IRX*2 mRNA expression levels were negatively correlated with its methylation level in both the TCGA- (*r* = -0.673, *P*.Value = 2.339e-61) and CTPAC-LUAD datasets (*r* = -0.482, *P*.Value = 1.916e-07) (Figure 5D and 5E). These findings suggest that *IRX*2 hypermethylation is a risk factor in LUAD patients and may be linked to dysregulations of *IRX*2 mRNA expression.

Gene enrichment analysis and IRX2 regulatory co-expression network

In the LinkeDomics database, we discovered that 10530 genes have a positive correlation with IRX2, while 9668 genes have a negative correlation with IRX2 (P.Value 0.05). The top 20 most positive and negative genes associated with IRX2 are depicted as heat maps and volcano plot in Figures 6A and 6B. *IRX2* expression has a positive correlation with genes such as *C5orf38* (r = 0.889, *P*.Value = 8.17E177), *SUSD2* (r = 0.529, *P*.Value = 1.53E38), and *GPR116* (r = 0.523, *P*.Value = 1.30E37), but a negative correlation with genes such as *UCK2 (r = -0.456, P*.Value *=1.24E29), CENPA* (r= -0.434, *P*.Value =4.19E−25), *CTSL2* (r = -0.429, *P*.Value = 1.61E−24). We discovered that the mRNA expression of the tumor suppressor gene *SUSD2* (Sushi Domain Containing 2) was highly expressed in IRX2 high-group (Figure 5B). We then confirmed a positive correlation with *IRX*2 in various LUAD datasets, including TCGA (*r* = 0.49, *P*.Value = 1.9e-05), GSE18842 (*r* = 0.64, *P*.Value = 1.2e-11), GSE32863 (*r* = 0.44, *P*.Value = 8.0e-07) and GSE40419 (*r* = 0.54 *P*.Value = 1.2e-13) (Figure S3A-D). We divided the tumor samples of TCGA-LUAD into two groups (high and low) according to the expression levels of *IRX*2. A total of 842 DEGs (373 up and 469 down). Following that, we performed GO: BP (biological pathway) analysis and discovered that they were related to antimicrobial humoral immune response and cilium movement. These findings suggested that IRX2 might play a role in lung immune response and normal psychological functions (Figure 6C). The GSEA results, on the other hand, suggested that it was positively correlated with the cell cycle and some tumorigenesis pathways (Figure 6D).

Subtype expression and clinical relevance of IRX2 in LUAD

In TCGA-LUAD and GSE36471, we discovered that *IRX*2 was expressed in the bronchioid subtype and not in the magnoid or squamoid subtypes (Figure 6A and 6B). Bronchioid subtype patients have a better prognosis in both TCGA-LUAD datasets (*P*.Value = 0.0473) (Figure 7C) in GSE36471 (Wilkerson et al.). Meanwhile, in TCGA LUAD cohort, *IRX*2 expression was significantly higher in patients who had a complete response (CR) compared to those who had progressive disease after therapy (*P*.Value < 0.05) (Figure 7D). However, there was no difference in *IRX*2 mRNA levels between the groups with stable disease and those with progressive disease (*P*.Value > 0.05) (Figure 5D). The expression of *IRX*2 was found to be lower in the N3 stage when compared to the N1 stage in TCGA cohort (Figure 7E). Because the N stage in the TNM staging system stands for lymph nodes invasiveness, these findings suggested that decreased *IRX*2 expression may be related to tumor invasion. We also found additional evidence in a

lung cancer cell line. When the expression profiles of poorly invasive CL1-0 and highly invasive CL1-5 lung adenocarcinoma cell lines were compared, it was discovered that *IRX*2 was significantly decreased in the CL1-5 cell line (log_2FC = 2.13, *FDR* = 2.37e-05) from dataset GSE42407 (Figure 7F). These findings suggested that *IRX*2 may be involved in the invasion and mobility of tumor cancer cells. We further calculated the immune cell subtype correlations with IRX2 and discovered that IRX2 expression is differed in immune and stromal score (Figure S4A) and negative associated with Macrophage M1 cells (Figure S4B-D).

Discussion

Homeobox genes are master developmental controllers that regulate morphogenesis and cell differentiation in animals by acting at the top of genetic hierarchies [40]. Meanwhile, homeobox gene expression abnormalities have been found in solid tumors and have been linked to cell fate determination and carcinogenesis [41]. Homeobox genes were found to be involved in both normal lung tissue differentiation and cancerous tissue uncontrolled proliferation. PITXs are an intriguing example because they regulate lung asymmetry and control mesenchymal cell proliferation and differentiation during lung development via beta-catenin signaling [42, 43]. On the other hand, they were regarded as novel biomarkers for the diagnosis of LUAD patients [44, 45]. Six1 is required for coordination of lung epithelial, mesenchymal, and vascular development [46]. It also promotes a variety of malignant biological behaviors by activating the Notch signaling pathway in lung cancer [47].

We discovered dysregulations of *IRX*2 in LUAD in this study. IRXs are typical homeobox genes that play critical roles in morphogenesis processes such as body segmentation during embryonic development such as neural pre-patterning, tissue differentiation, neural crest development and cranial placode formation [48], embryonic heart [49, 50] and limbs developing [51, 52]. In this study, we found that loss of both mRNA and protein levels of IRX2 in LUAD. The promoter genomic region of IRX2 was hypermethylated and correlated with the OS. *IRX*1 promoter methylation may be a tumor-associated event and an independent predictor of survival advantage in patients with NSCLC. This evidence suggested that IRX2 might be used as a prognostic marker for LUAD. Previous report showed that IRX negatively regulate Dpp/TGF-β pathway activity during intestinal tumorigenesis [53]. TGF-β pathway induced EMT and stemness characteristics are associated with epigenetic regulation in lung cancer [54]. It might be possible to investigate the negative regulation between TGF-β pathway and IRX2 in LUAD.

Taken together, we showed differential expressed homeobox gene patterns in LUAD and further validated that *IRX*2 was among the most significant down-regulated members and associated with clinical prognostic significance in LUAD patients. It would be interesting to investigate the mechanisms of *IRX*2 in controlling the tumorigenesis in the future.

Declarations

Acknowledgment

We thank the Hunan Provincial Health Planning Commission for its research grant project (No. 202111002309)

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018 Nov;68(6):394–424.
2. Howlader N, Forjaz G, Mooradian MJ, Meza R, Kong CY, Cronin KA, Mariotto AB, Lowy DR, Feuer EJ. The Effect of Advances in Lung-Cancer Treatment on Population Mortality. N Engl J Med. 2020 Aug 13;383(7):640–649.
3. Sher T, Dy GK, Adjei AA. Small cell lung cancer. Mayo Clin Proc. 2008 Mar;83(3):355 – 67.
4. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong KK. Non-small-cell lung cancers: a heterogeneous set of diseases. Nat Rev Cancer. 2014 Aug;14(8):535–46.
5. Mark M, Rijli FM, Chambon P. Homeobox genes in embryogenesis and pathogenesis. Pediatr Res. 1997;42(4):421–429.
6. Cillo C, Faiella A, Cantile M, Boncinelli E. Homeobox genes and cancer. Exp Cell Res. 1999;248(1):1–9.
7. Maeda Y, Davé V, Whitsett JA. Transcriptional control of lung morphogenesis. Physiol Rev. 2007;87(1):219–244.
8. Varma S, Cao Y, Tagne JB, et al. The transcription factors Grainyhead-like 2 and NK2-homeobox 1 form a regulatory loop that coordinates lung epithelial cell morphogenesis and differentiation. J Biol Chem. 2012;287(44):37282–37295.
9. Ostrin EJ, Little DR, Gerner-Mauro KN, et al. β-Catenin maintains lung epithelial progenitors after lung specification. Development. 2018;145(5):dev160788.
10. Maeda Y, Tsuchiya T, Hao H, et al. Kras(G12D) and Nkx2-1 haploinsufficiency induce mucinous adenocarcinoma of the lung. J Clin Invest. 2012;122(12):4388–4400.
11. Nakamura T, Nakagawa M, Ichisaka T, Shiota A, Yamanaka S. Essential roles of ECAT15-2/Dppa2 in functional lung development. Mol Cell Biol. 2011;31(21):4366–4378.
12. Tran TQ, Kioussi C. Pitx genes in development and disease. Cell Mol Life Sci. 2021;78(11):4921–4938.
13. Dietrich D, Hasinger O, Liebenberg V, Field JK, Kristiansen G, Soltermann A. DNA methylation of the homeobox genes PITX2 and SHOX2 predicts outcome in non-small-cell lung cancer patients. Diagn Mol Pathol. 2012;21(2):93–104.
14. Chen Y, Knösel T, Ye F, Pacyna-Gengelbach M, Deutschmann N, Petersen I. Decreased PITX1 homeobox gene expression in human lung cancer. Lung Cancer. 2007;55(3):287–294.
15. Maeda Y, Tsuchiya T, Hao H, et al. Kras(G12D) and Nkx2-1 haploinsufficiency induce mucinous adenocarcinoma of the lung. J Clin Invest. 2012;122(12):4388–4400.

16. Yu W, Li X, Eliason S, et al. Irx1 regulates dental outer enamel epithelial and lung alveolar type II epithelial differentiation. Dev Biol. 2017;429(1):44–55.
17. Becker MB, Zülch A, Bosse A, Gruss P. Irx1 and Irx2 expression in early lung development. Mech Dev. 2001;106(1–2):155–158.
18. Doi T, Lukošiūtė A, Ruttenstock E, Dingemann J, Puri P. Expression of Iroquois genes is up-regulated during early lung development in the nitrofen-induced pulmonary hypoplasia. J Pediatr Surg. 2011;46(1):62–66
19. Küster MM, Schneider MA, Richter AM, et al. Epigenetic Inactivation of the Tumor Suppressor IRX1 Occurs Frequently in Lung Adenocarcinoma and Its Silencing Is Associated with Impaired Prognosis. Cancers (Basel). 2020;12(12):3528.
20. Chakma K, Gu Z, Abudurexiti Y, et al. Epigenetic inactivation of IRX4 is responsible for acceleration of cell growth in human pancreatic cancer. Cancer Sci. 2020;111(12):4594–4604
21. Zhong YF, Holland PW. HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. Evol Dev. 2011;13(6):567–568.
22. Zhong YF, Butts T, Holland PW. HomeoDB: a database of homeobox gene diversity. Evol Dev. 2008;10(5):516–518.
23. Mistry J, Chuguransky S, Williams L, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021;49(D1):D412-D419.
24. Goldman MJ, Craft B, Hastie M, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol. 2020;38(6):675–678.
25. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014;511(7511):543–550.
26. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res. 2013;41(Database issue):D991-D995.
27. Sanchez-Palencia A, Gomez-Morales M, Gomez-Capilla JA, et al. Gene expression profiling reveals novel biomarkers in nonsmall cell lung cancer. Int J Cancer. 2011;129(2):355–364.
28. Hou J, Aerts J, den Hamer B, et al. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. PLoS One. 2010;5(4):e10312.
29. Selamat SA, Chung BS, Girard L, et al. Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. Genome Res. 2012;22(7):1197–1211.
30. Seo JS, Ju YS, Lee WC, et al. The transcriptional landscape and mutational profile of lung adenocarcinoma. Genome Res. 2012;22(11):2109–2119.
31. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. LIMMA
32. Chen Y, Lun AT, Smyth GK. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. F1000Res. 2016;5:1438. EdgeR

33. Lánczky A, Győrffy B. Web-Based Survival Analysis Tool Tailored for Medical Research (KMplot): Development and Implementation. J Med Internet Res. 2021;23(7):e27633.
34. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017;45(W1):W98-W102.
35. Chandrashekar DS, Bashel B, Balasubramanya SAH, et al. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. Neoplasia. 2017;19(8):649–658.
36. Gillette MA, Satpathy S, Cao S, et al. Proteogenomic Characterization Reveals Therapeutic Vulnerabilities in Lung Adenocarcinoma. Cell. 2020;182(1):200–225.e35.
37. Uhlen M, Zhang C, Lee S, et al. A pathology atlas of the human cancer transcriptome. Science. 2017;357(6352):eaan2507.
38. Uhlén M, Fagerberg L, Hallström BM, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419.
39. Karlsson M, Zhang C, Méar L, et al. A single-cell type transcriptomics map of human tissues. Sci Adv. 2021;7(31):eabh2169.
40. Vasaikar SV, Straub P, Wang J, Zhang B. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. Nucleic Acids Res. 2018;46(D1):D956-D963.
41. Kolmykov S, Yevshin I, Kulyashov M, et al. GTRD: an integrated view of transcription regulation. Nucleic Acids Res. 2021;49(D1):D104-D111.
42. Otasek D, Morris JH, Bouças J, Pico AR, Demchak B. Cytoscape Automation: empowering workflow-based network analysis. Genome Biol. 2019;20(1):185.
43. Mark M, Rijli FM, Chambon P. Homeobox genes in embryogenesis and pathogenesis. Pediatr Res. 1997;42(4):421–429.
44. Abate-Shen C. Deregulated homeobox gene expression in cancer: cause or consequence?. *Nat Rev Cancer*. 2002;2(10):777–785. Liang J, Li H, Han J, et al.
45. Lin CR, Kioussi C, O'Connell S, et al. Pitx2 regulates lung asymmetry, cardiac positioning and pituitary and tooth morphogenesis. Nature. 1999;401(6750):279–282.
46. De Langhe SP, Carraro G, Tefft D, et al. Formation and differentiation of multiple mesenchymal lineages during lung development is regulated by beta-catenin signaling. PLoS One. 2008;3(1):e1516.
47. Zhang C, Chen X, Chen Y, et al. The PITX gene family as potential biomarkers and therapeutic targets in lung adenocarcinoma. Medicine (Baltimore). 2021;100(4):e23936.
48. Luo J, Yao Y, Ji S, et al. PITX2 enhances progression of lung adenocarcinoma by transcriptionally regulating WNT3A and activating Wnt/β-catenin signaling pathway. Cancer Cell Int. 2019;19:96.
49. El-Hashash AH, Al Alam D, Turcatel G, et al. Six1 transcription factor is critical for coordination of epithelial, mesenchymal and vascular morphogenesis in the mammalian lung. Dev Biol. 2011;353(2):242–258.

50. Huang S, Lin W, Wang L, et al. SIX1 Predicts Poor Prognosis and Facilitates the Progression of Non-small Lung Cancer via Activating the Notch Signaling Pathway. J Cancer. 2022;13(2):527–540.
51. Feijóo CG, Saldias MP, De la Paz JF, Gómez-Skarmeta JL, Allende ML. Formation of posterior cranial placode derivatives requires the Iroquois transcription factor irx4a. Mol Cell Neurosci. 2009;40(3):328–337.
52. Christoffels VM, Keijser AG, Houweling AC, Clout DE, Moorman AF. Patterning the embryonic heart: identification of five mouse Iroquois homeobox genes in the developing heart. Dev Biol. 2000;224(2):263–274.
53. Ban K, Wile B, Cho KW, et al. Non-genetic Purification of Ventricular Cardiomyocytes from Differentiating Embryonic Stem Cells through Molecular Beacons Targeting IRX-4. Stem Cell Reports. 2015;5(6):1239–1249.
54. Mummenhoff J, Houweling AC, Peters T, Christoffels VM, Rüther U. Expression of Irx6 during mouse morphogenesis. Mech Dev. 2001;103(1–2):193–195.
55. McDonald LA, Gerrelli D, Fok Y, Hurst LD, Tickle C. Comparison of Iroquois gene expression in limbs/fins of vertebrate embryos. J Anat. 2010;216(6):683–691.
56. Martorell Ò, Barriga FM, Merlos-Suárez A, et al. Iro/IRX transcription factors negatively regulate Dpp/TGF-β pathway activity during intestinal tumorigenesis. EMBO Rep. 2014;15(11):1210–1218.
57. Kim BN, Ahn DH, Kang N, et al. TGF-β induced EMT and stemness characteristics are associated with epigenetic regulation in lung cancer. Sci Rep. 2020;10(1):10597

Tables

Tables 1, 2 are available in the Supplementary Files section.

Figures

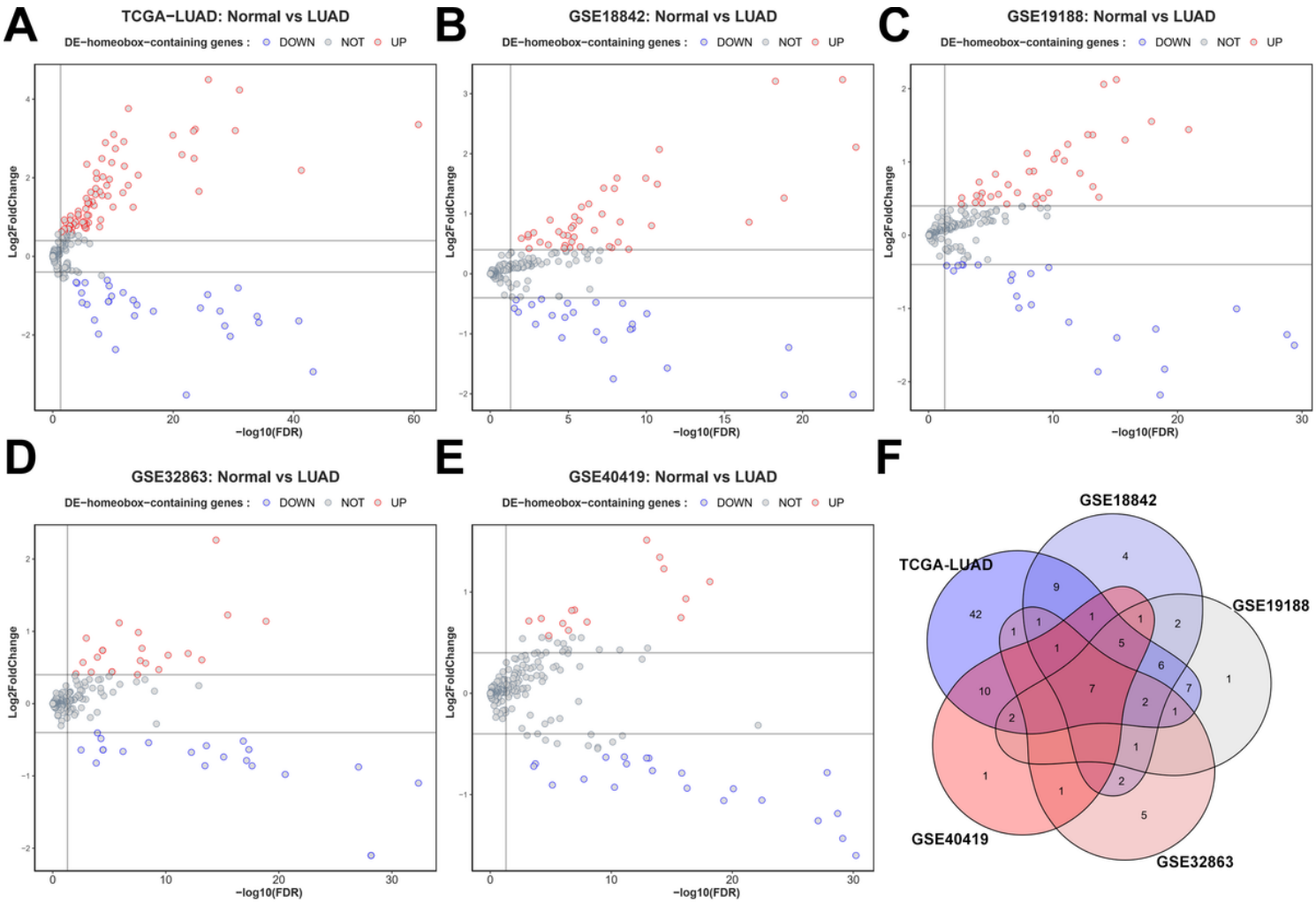

Figure 1

Differential expressed homeobox-containing genes in LUAD datasets. Volcano plot of differentially expressed homeobox-containing genes in various LUAD datasets including TCGA RNA-seq (**A**) and GSE18842 (**B**), GSE19188 (**C**), GSE32863 (**D**) and GSE40419 (**E**) microarray datasets with $|\log_2$ (fold change)$| > 0.4$ and $FDR < 0.05$. (**F**) Log2FC heatmap of 14 differentially expressed homeobox genes in these datasets, with red representing positive $\log_2$FC and blue representing negative $\log_2$FC values.

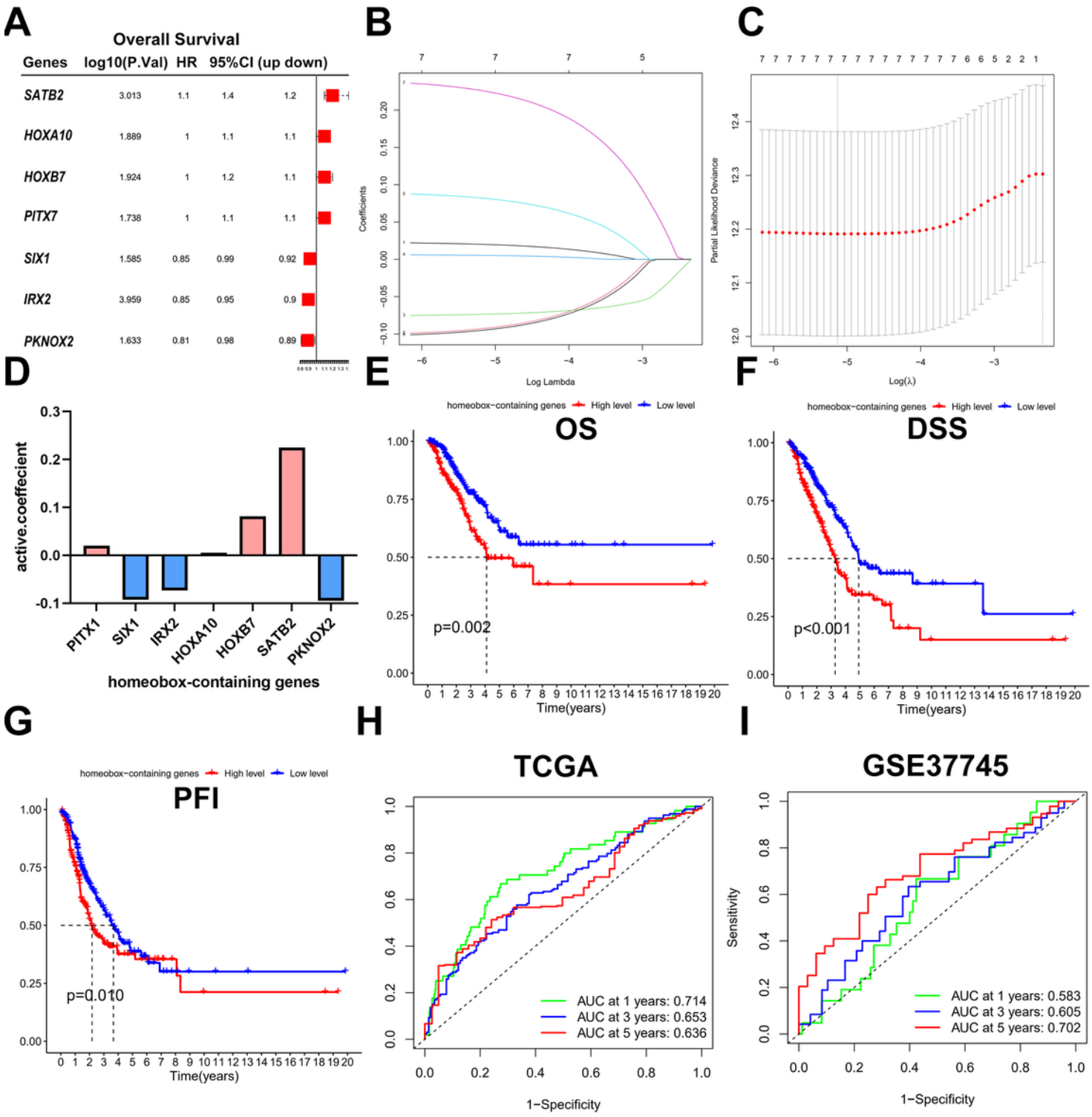

Figure 2

Prognostic value of DE-containing homeobox genes in LUAD. The results of Cox regression and LASSO regression analysis. (A) The forest plot for univariate Cox regression analysis identified seven DE homeobox-containing genes associated with OS in TCGA-LUAD. (B, C) The results after minimum criteria calculation though LASSO regression. (D) The coefficients of the survival model constructed by LASSO regression. (E-G) Survival curves hinted that the subgroup with highly risk score had significant different

OS, PFS and PFI rates compare with the low-risk subgroup. (H-I) ROC curve analysis showed the predicted efficacy of seven DE homeobox-containing genes in TCGA-LUAD and GSE37745 datasets.

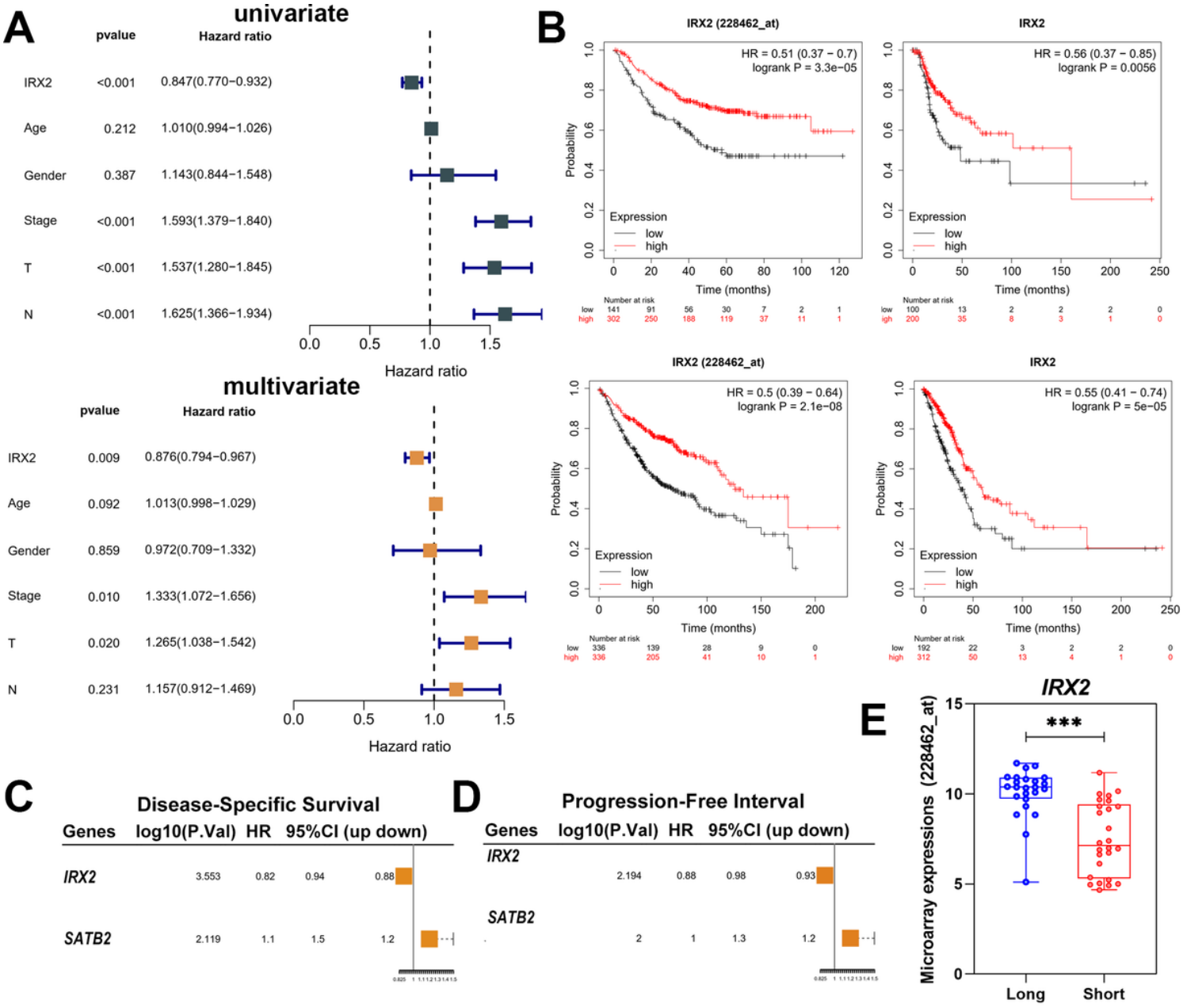

Figure 3

Prognostic value of IRX2 in LUAD. (A) Disease specific survival (DSS) and (**B**) Progression free survival (PFI) prognosis forest map based on univariate cox regression analysis of DE homeobox genes in the TCGA-LUAD cohort. The hazard ratio is represented by box in the forest plot, and the line on both sides of the point represents the 95 percent CI. (**D**) Multivariate cox regression analysis of IRX2 in TCGA-LUAD cohort. Expression levels of *IRX*2 (228462_at) association with (E) OS and (F) FP (Free progression) in various LUAD microarray datasets in Kaplan-Meier Plotter for lung cancer. (G) The Kaplan-Meier analysis showed the *IRX*2 expression association with OS in TCGA LUAD RNA-seq dataset. (H) *IRX*2 expressions in two patient populations with different survival outcomes.

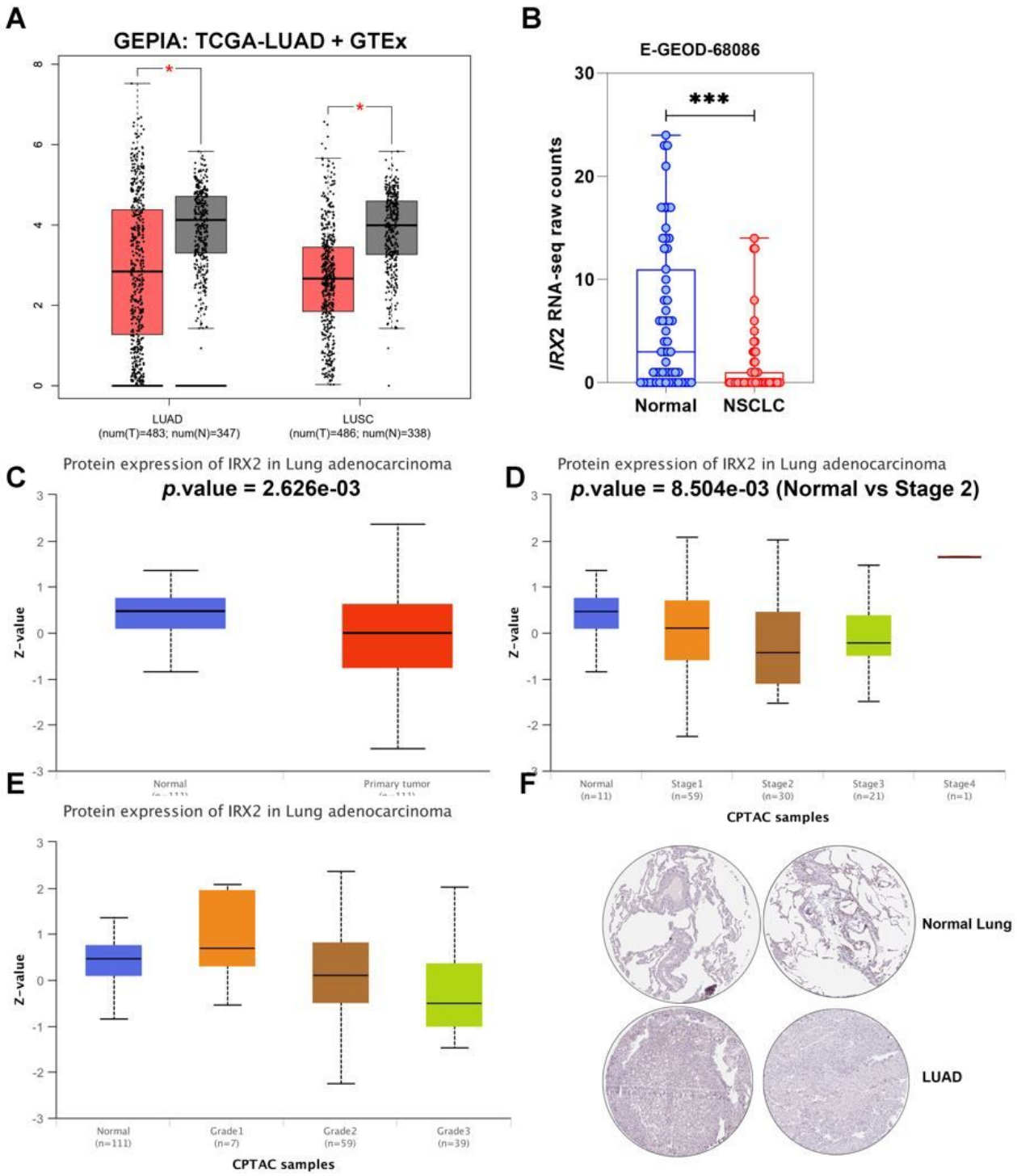

Figure 4

Frequently suppressed expressions of IRX2 in LUAD. (**A**) Boxplot of GEPIA (Gene Expression Profiling Interactive Analysis) to validate expression of *IRX*2 in TCGA LUAD and LUSC samples compared with matched normal samples and lung tissues from GTEx database. The red represented the cancer tissue group, the gray represented the normal tissue group, and the asterisk represented the *P*.Value < 0.01. The dots represented expression in each sample. The expression values are $\log_2$(TPM + 1). (B) *IRX*2 significantly reduced in blood platelets from LUAD and healthy donors in the RNA-seq dataset E-GEOD-

68086. The expression values are raw counts. The IRX2 proteomic expression profile of CPTAC- lung adenocarcinoma based on (C) Sample types (normal/primary tumor), (D) individual cancer stage and (E) Tumor grade. The protein levels were presented as Z-value standard deviations from the median across samples for LUAD using the normalized Log2 Spectral count ratio values from CPTAC. (F) Immunohistochemistry (IHC) using antibody HPA054669 in the HPA database demonstrated IRX2 protein expression in normal lung tissues and cancer samples. CPTAC: Clinical Proteomic Tumor Analysis Consortium; FC: fold change; GTEx: The Genotype-Tissue Expression; IRX: Iroquois Homeobox; LUAD: Lung Adenocarcinoma; LUSC: Lung Squamous Cell Carcinoma; N: Normal; TCGA: The Cancer Genome Atlas; T: Tumor; TPM: Transcript per million.

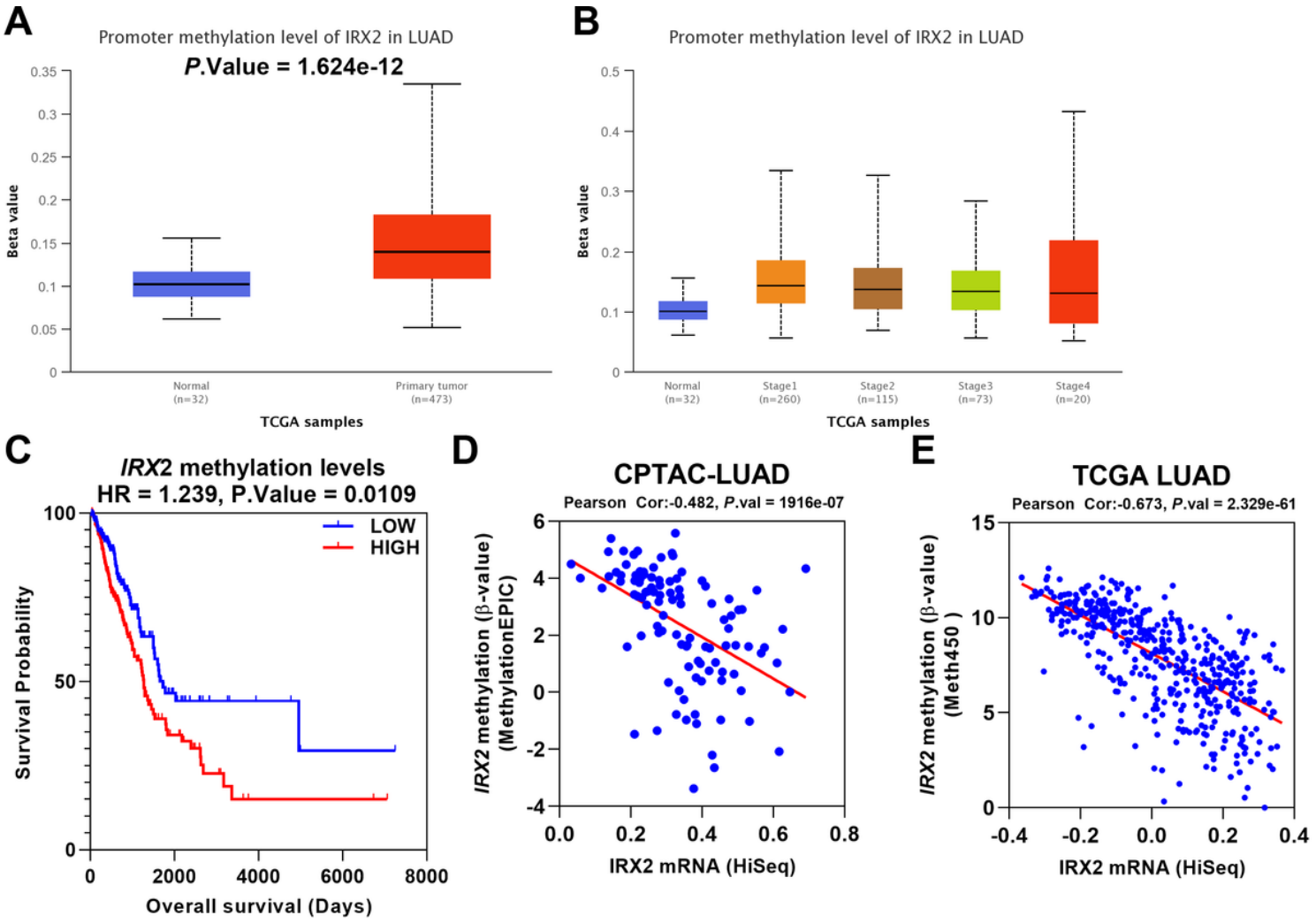

Figure 5

IRX2 promoter region methylation altered in lung adenocarcinoma. UALCAN database analysis the methylation profiles of 11 CpG islands around the *IRX*2 TSS20 and TSS1500 genomic regions in TCGA-LUAD based on (A) Sample types (normal/primary tumor), (B) individual cancer stage. The promoter methylation levels were presented as average Beta-value. (C) The associated with *IRX*2 promoter methylation levels with OS was validated in TCGA-LUAD using LinkeDomics database. The correlations of promoter region methylation levels and mRNA expression levels of *IRX*2 were validated in LinkeDomics

database using Pearson Correlation test in TCGA- (D) and CPTAC- LUAD (E) datasets. CPTAC: Clinical Proteomic Tumor Analysis Consortium; IRX: Iroquois Homeobox; OS: Overall survival; TCGA: The Cancer Genome Atlas.

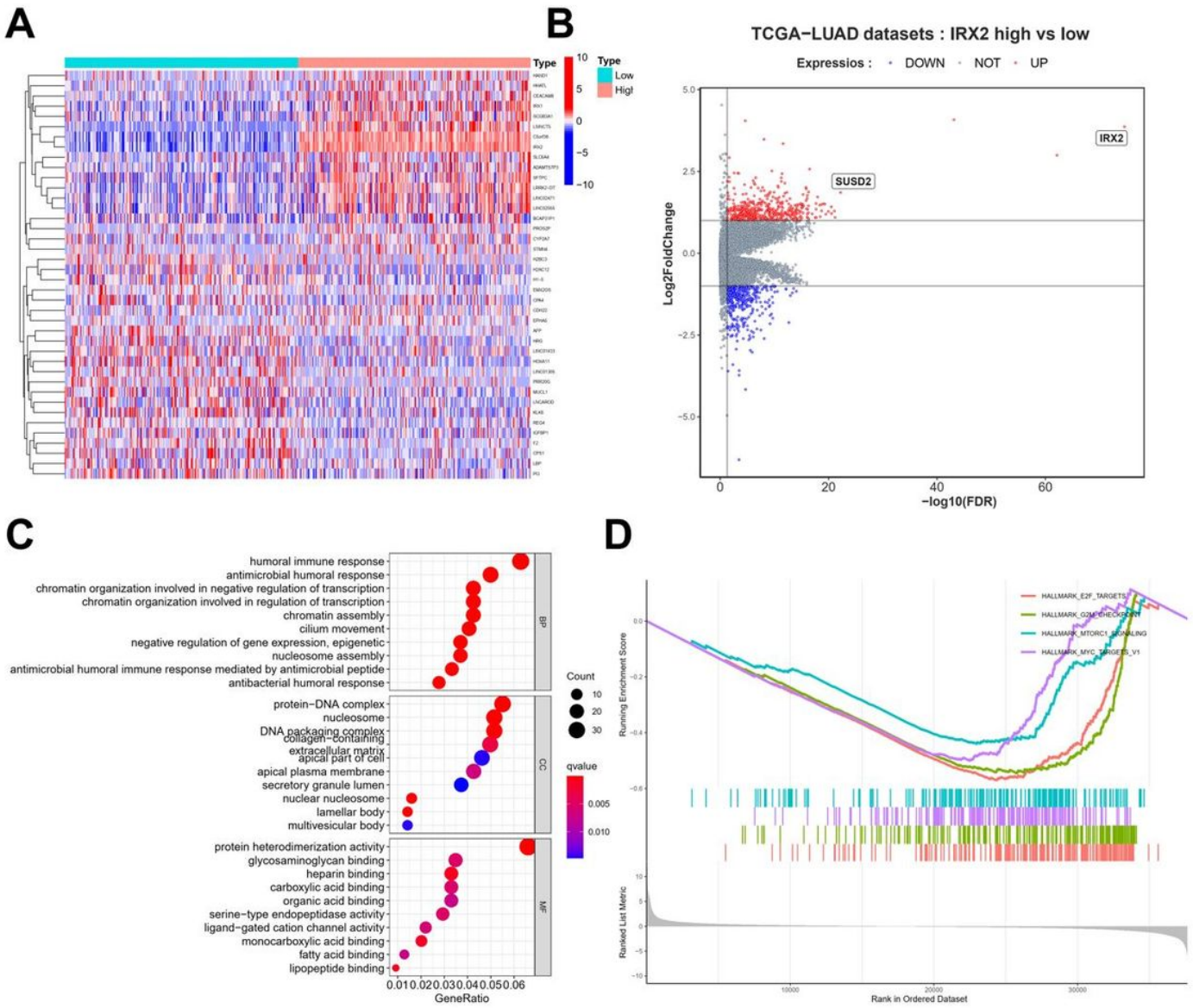

Figure 6

Gene enrichment analysis and IRX2 regulatory co-expression network. (A) Heatmap of most differential expressed genes according to the expression status of IRX2 (high and low). (**B**) The volcano plot of differentially expressed genes according to the expression status of IRX2 (high and low). (C) The KEGG enrichment analysis of the most differentially expressed genes according to IRX2 expression level. (D) The GSEA plot for the pathways that altered in two group of different expression status of IRX2.

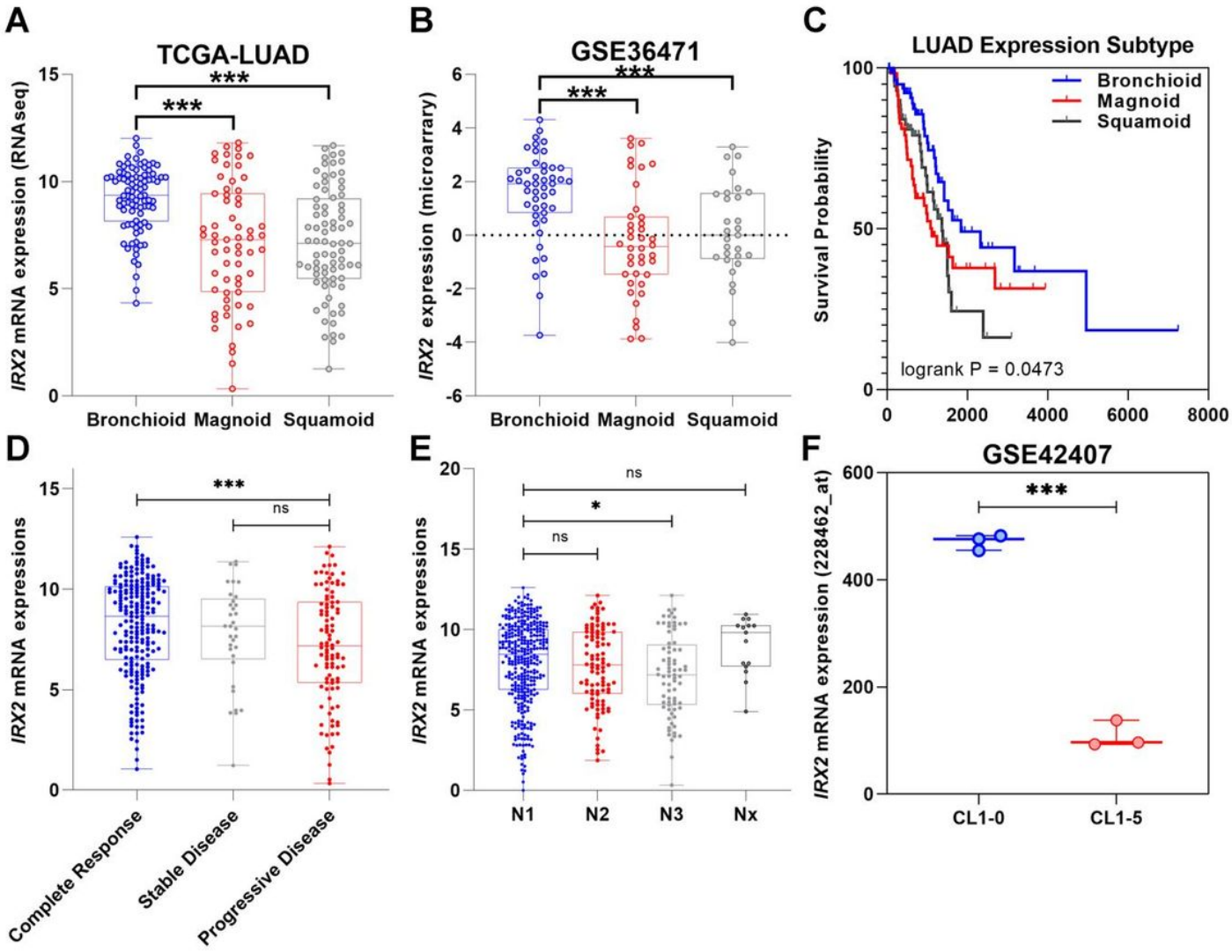

Figure 7

Subtype and clinical characteristics of IRX2. Differentially expression of IRX2 in different LUAD subtypes including in TCGA-LUAD (**A**) and GSE36471 (**B**). (**C**) The Kaplan-Meier analysis showed the different subtype with altered OS in TCGA LUAD RNA-seq dataset. (**D, E**) The expression status of IRX2 according to the different TNM stage in LUAD. (**F**) The expression of *IRX*2 (228462_at) in parent CL1-0 and its metastatic cell line CL1-5.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.csv
- TableS2.xlsx
- FigureS1.tif
- FigureS2.tif

- FigureS3.tif
- FigureS4.tif