Clinical Medicine & Pharmacology

press

Research article

https://doi.org/10.70731/w804kb42

PBX/Knotted 1 Homeobox 2 as a Novel Prognostic Marker that Dysregulated in Lung Adenocarcinoma

Lisa Sun [a,#], **Yuxi Ren** [b,#], **Ping Xiao** [b], **Haoming Shen** [b], **Bin Qu** [b,*]

[a] *Blood Transfusion Branch, Hunan Cancer Hospital, The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha 410009, China*
[b] *Department of Laboratory Medicine, Hunan Cancer Hospital, The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha 410009, China*

KEYWORDS

PBX/Knotted 1 Homeobox 2;
Lung Adenocarcinoma;
Homeobox Genes;
Prognostic Marker

ABSTRACT

Homeobox-containing genes are evolutionary conserved regulators that play important roles in embryogenesis and tumorigenesis. We started by searching differentially expressed (DE) homeobox-containing genes in lung adenocarcinoma datasets and obtained 14 candidates from a total of 238 distinct members. Using univariant Cox and lasso regression, we created a 6 DE-homeobox-containing gene-based model that successfully predicted overall survival, disease-free time survival, and progression-free intervals in LUAD patients. Following that, we notice that *PKNOX*2 (PBX/Knotted 1 Homeobox 2) is the most potent protective factor for overall survival of LUAD patients. Its expression was significantly downregulated in LUAD and decreased as tumor pathological grade increased. *PKNOX*2 was as an independent prognostic marker in LUAD in univariate and multivariate Cox proportional and further confirmed in Kaplan-Meier survival analyses. Gene Set Enrichment Analysis (GSEA) results revealed that *PKNOX*2 related genes were enriched in immune response as well as extracellular matrix organization. Further we revealed that *PKNOX*2 was linked to tumor purity and macrophage filtrations. Since previous studies suggested that PKNOX2 play a tumor suppressive role in gastric cancer, however its pathological role in LUAD is unknown. These findings suggest that *PKNOX*2 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 furth16er 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 in-

The authors contribute equally to this study
* Corresponding author. E-mail address: qubin@hnca.org.cn

Received 15 July 2025; Accepted 27 September 2025; Published online 30 September 2025.

Table 1 | Specific information on the five lung adenocarcinoma datasets

Dataset names	Platform	Sample Size	Total Gene Numbers	Homeobox-containing gene numbers	Up-regulated Candidates	Down-regulated candidates
TCGA-LUAD	IlluminaHiSeq_RNASeqV2	Normal 59/Tumor 517	20530	227	76	30
GSE18842	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	Normal 45/Tumor 46	23520	205	42	24
GSE19188	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	Normal 65/Tumor 91	23520	205	35	22
GSE32863	GPL6884 Illumina HumanWG-6 v3.0 expression beadchip	Normal 58/Tumor 58	27520	225	22	21
GSE40419	GPL11154 Illumina HiSeq 2000 (Homo sapiens)	Normal 77/Tumor 87	22427	230	14	21

volved 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. Using lasso regression, we validated the prognostic value of a 6-differentially expressed homeobox-containing genes model. Particularly, we found that the PBX/Knotted 1 homeobox family gene *PKNOX2*, 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 *PKNOX* family genes also play an important role in lung epithelial development and morphogenesis in various tissues [16-18]. However, clinical correlation with lung adenocarcinoma of *PRKNOX2* have not yet been full explored. In this study, we discovered that PKNOX2 may play a prognostic role in lung adenocarcinoma and may play a specific role in regulating lung adenocarcinoma tumorigenesis.

Material and Methods

Human Protein-Coding Homeobox Genes Selection

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 1** contains detailed information about datasets used in this article, 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.

Survival Analysis

To improve the reliability of our study, we also removed samples with no survival time or survival time of less than 30 days. Through univariate and multivariate Cox regression analysis, we estimated the relationship between overall survival (OS) and clinical parameters.

Cox Proportional Hazards Analysis

Univariate and multivariate analyses were used to assess the relationship between the expression of the DE homeobox-containing 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 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].

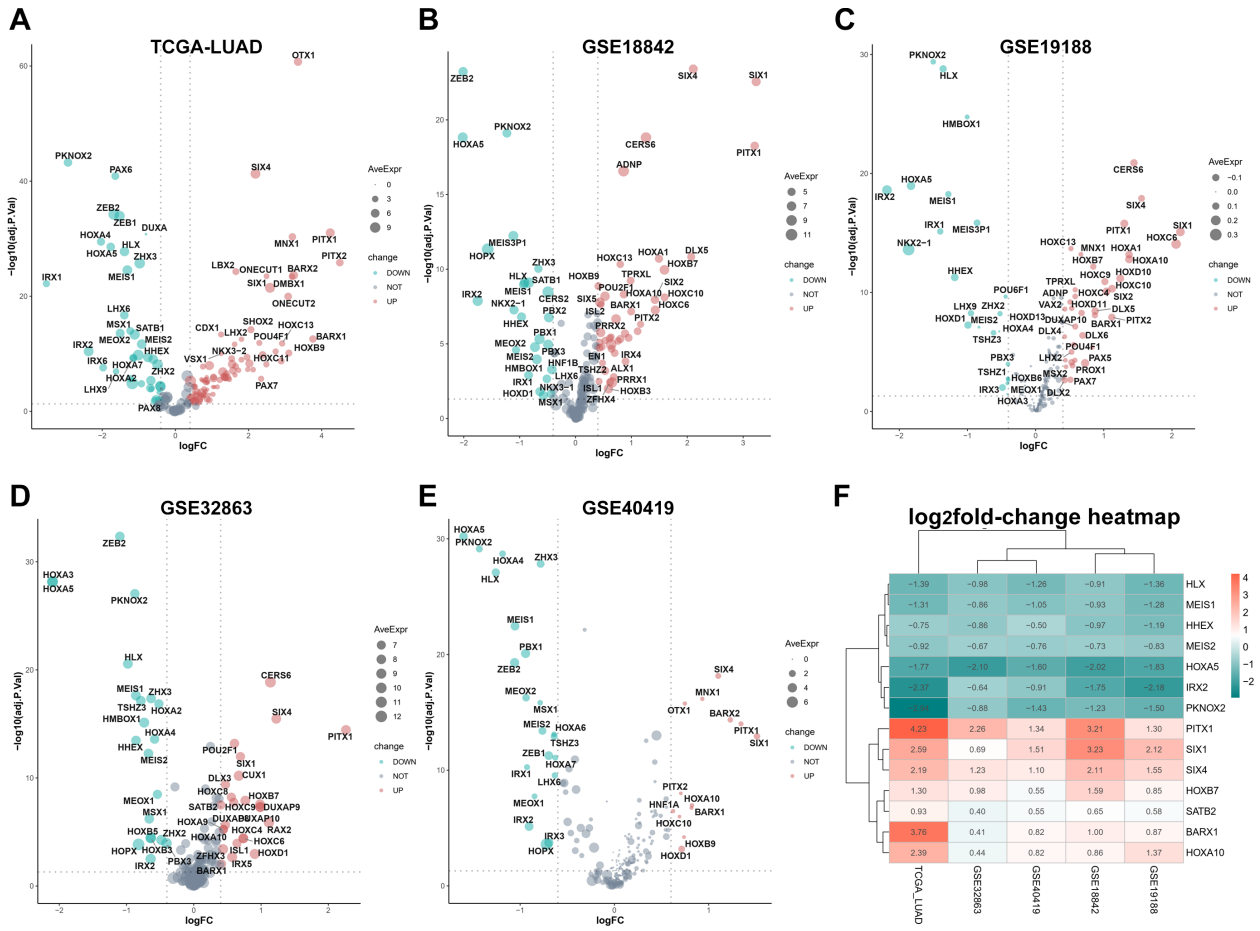

Figure 1 | Differentially expressed homeobox-containing genes in LUAD. Screening for the crucial homeobox-containing genes that dysregulated in LUAD. (A) Volcano plots of differentially expressed homeobox-containing genes in LUAD datasets. Differentially expressed genes were filtered using adjusted *P.Value* < 0.05 and $\log_2$(fold-change) > 0.40 or < -0.40 for significance in datasets including (A) TCGA-LUAD (N = 59, T = 517), (B) GSE18842 (N = 45, T = 46), (C) GSE19188 (N = 65, T = 91), (D) GSE32863 (N = 58, T = 58), and (E) GSE40419 (N = 77, T = 87). Red dot represents significantly up-regulated while green dot represents the down-regulated genes. (F) Heatmap depicted the $\log_2$fold-change value of 14 most significantly differentially expressed homeobox-containing genes (*HLX*, *MEIS*1, *HHEX*, *MEIS*2, *HOXA*5, *IRX*2, *PKNOX*2, *PITX*1, *SIX*1, *SIX*4, *HOXB*7, *SATB*2, *BARX*1, *HOXA*10) from each dataset. Lung adenocarcinoma: LUAD, N: Normal, T: Tumor.

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 PKNOX2 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

Differentially-Expressed Homeobox-Containing Genes in Lung Adenocarcinoma

The expression profiles of 238 human protein-coding homeobox-containing genes were examined in five lung adenocarcinoma (LUAD) datasets, including TCGA-LUAD, GSE18842, GSE19188, GSE32863 and GSE40419 (**Figure 1A-E**). There were two RNA-sequencing datasets (TCGA-LUAD and GSE40419), and the remaining datasets were all microarrays. Following intersection of the differentially expressed genes (DEGs) from each dataset (cutoff criteria: | $\log_2$fold-change| > 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), *HOXB*7 (Homeobox protein Hox-B) and *SATB*2 (Special AT-rich sequence-binding protein) (**Figure 1F**). Seven others were significantly

Table 2 | pecific information on the five lung adenocarcinoma datasets

Dataset	GEO ID/Source	Normal	Tumor	Platform	Protein-coding hoi	Up DEGs	Down DEGs
TCGA_LUAD_201(	TCGA-LUAD	517	59	Illumina HiSeq 200(	227	95	35
Sanchez-Palencia_:	GSE18842	45	46	GPL570	205	42	24
Hou_2010	GSE19188	65	91	GPL570	205	35	22
Selamat_2012	GSE32863	58	58	GPL6884	225	22	21
Seo_2012	GSE40419	77	87	GPL11154	230	29	30

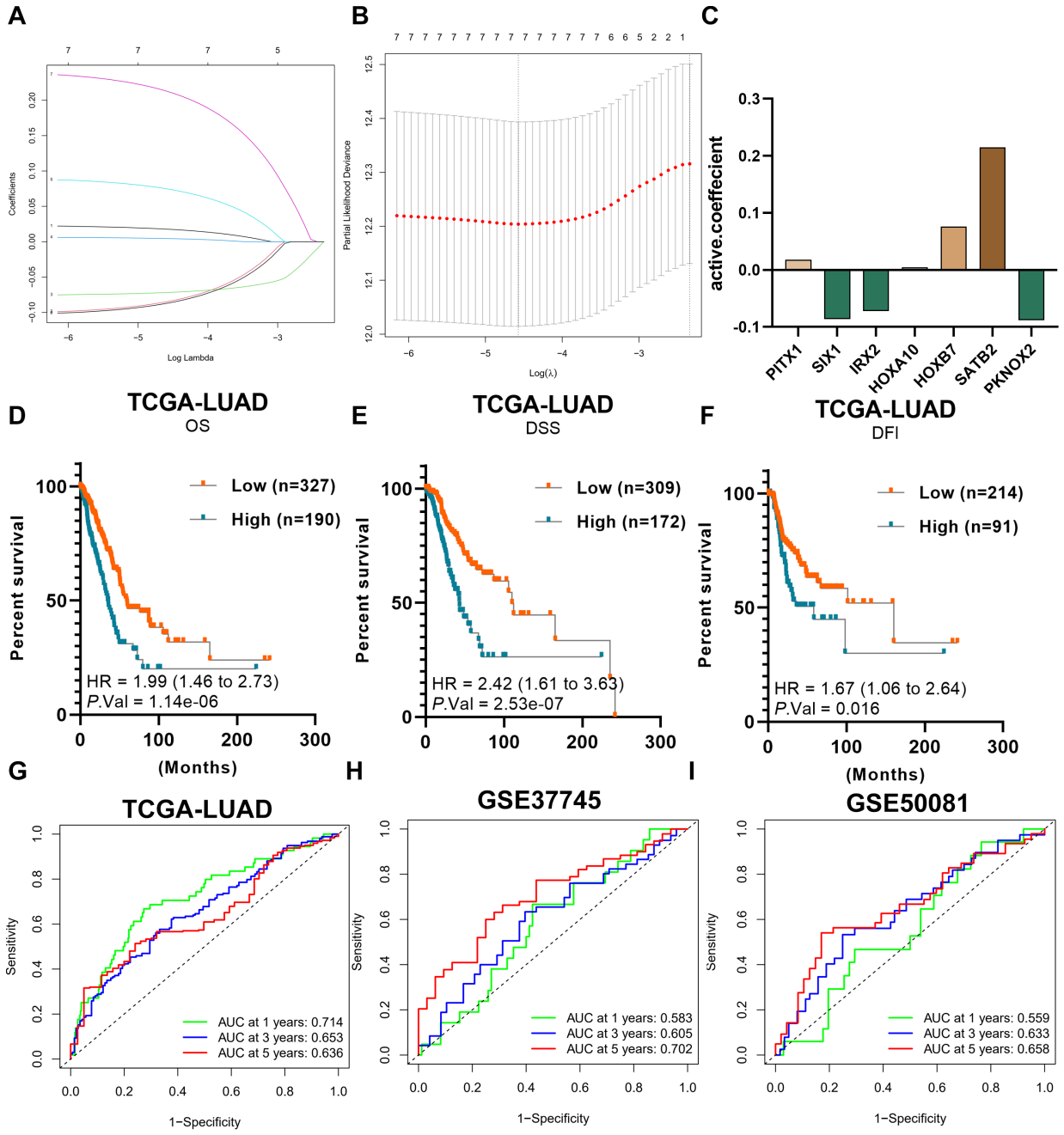

Figure 2 | Prognostic value of homeobox-containing genes in LUAD. (A, B) The results after minimum criteria calculation though LASSO regression. **(C)** The coefficients of the survival model constructed by LASSO regression. **(D-F)** Survival curves hinted that the subgroup with highly risk score had significant different OS (overall survival), PFS (progression survival survival) and PFI (progress-free interval) rates compare with the low-risk subgroup. **(G-I)** ROC curve analysis showed the predicted efficacy of seven DE homeobox-containing genes in training TCGA-LUAD and validating GSE37745, GSE50081 datasets.

down-regulated including *IRX*2 (Iroquois homeobox 2), *PKNOX*2 (PBX/knotted 1 homeobox 2), *MEIS*1/2 (Meis homeobox 1/2), *HLX* (H2.0-like homeobox), *HHEX* (Hematopoietically-expressed homeobox protein) and *HOXA5* (Figure 1B). According to the classification of homeobox-containing genes, *PITX1* belongs to the PRD (Paired Domain) class, *SIX1/4* to the SINE class, *SATB2* to the CUT class, *BARX1, HHEX, HLX, HOXA5, HOXA10, HOXB7* to the ANTP class, *IRX2, PKNOX2,* and *MEIS1/2* to the TALE (Three Amino Acid Loop Extension) class. Figure 1F depicts the log_2fold-change heatmap of 14 DEGs from each dataset. **Table 2** contains specific information on the five lung adenocarcinoma datasets.

Prognostic Significance of Homeobox-Containing Gene in Lung Adenocarcinoma

We used univariate Cox proportional hazards regression to investigate the relationship between the gene expression value of these DE homeobox genes and overall survival (OS) in TCGA-LUAD cohort. In univariate cox analysis result, we discovered that 8 genes (*PITX1*, *SIX*1, *IRX*2, *HOXA*10, *HOXB*7, *PITX*1, *SATB*2, *PKNOX*2) were significantly associated with OS (*P.Value* < 0.05) (Figure S1). *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 seven genes had coefficients that were not zero with 1000 repeats (**Figure 2A and 2B**). Risk score was calculated by the lasso coefficients: (0.0261 × *HOXA*10) + (-0.0365 × *HOXC*6) + (0.0357 × *PITX*1) + (-0.0857 × *SIX*1) + (-0.0581 × *IRX*2) + (-0.0732 × *PKNOX*2) (**Figure 2C**). The lasso Cox regression risk score results could be used to predict the OS (HR = 1.99 [1.46-2.73], *P.Value* = 1.14e-06) (**Figure 2D**), Disease-specific survival (DSS) (HR=2.42 [1.61-3.63], *P.Value* = 2.53e-07) (**Figure 2E**) and Disease-Free interval (DFI) (HR = 1.67 [1.06-2.64], *P.Value* = 0.016) (**Figure 2F**). The 1-year area under the curve (AUC) was 0.714, the 3-year AUC was 0.653, and the 5-year AUC was 0.636, according to time-dependent receiver operating characteristic (ROC) curves of the training dataset TCGA-LUAD cohort (**Figure 2G**). In the validation dataset GSE37745 and GSE50081, the 1-year AUC was 0.583 or 0.559, 3-year AUC was 0.605 or 0.633, 5-year AUC was 0.702 or 0.658 respectively (**Figure 2H and 2I**).

Prognostic Significance of PKNOX2 in Lung Adenocarcinoma

Because *PKNOX2* had the lowest HR value in univariate cox analysis result, we chose it for the follow-up study. *PKNOX2* expression was associated with OS (Overall survival) (HR = 0.60 [0.44-0.81], *P*.Value = 0.0034) (**Figure 3A**), DSS (Disease-specific survival) (HR = 0.57 [0.38-0.85], *P*.Value = 0.015) (**Figure 3B**) and DFI (Disease-free interval) (HR = 0.52 [0.33-0.81], *P*.Value = 0.015) (**Figure 3C**) in the TCGA-LUAD cohort. *PKNOX*2 expression was also found to be a protective factor in several independent LUAD microarray datasets, including GSE30219 (HR = 0.23 [0.11-0.44], *P*.Value = 0.0054) (**Figure 3D**), GSE31210 (HR = 0.31 [0.16-0.63], *P*.Value = 0.0006) (**Figure 3E**), GSE42127 (HR = 0.44 [0.23-0.82], *P*.Value = 0.022) (**Figure 3F**), GSE72094 (HR = 0.62 [0.42-0.91], *P-Value* = 0.026) (**Figure 3G**), GSE41271 (HR = 0.48 [0.29-0.78], *P*.Value = 0.0080) (Figure 3H) and GSE3141 (HR = 0.28 [0.14-0.56], *P*.Value = 0.0018) (**Figure 3I**). Kaplan-Meier survival analysis also indicated that expression levels of *PKNOX2* (222171_s_at) were significantly associated with OS (HR = 0.50 [0.39-0.64], *P*.Value = 2.1e-07) (Figure S2A), FP (Free progression) (HR = 0.51 [0.37-0.70], *P*.Value = 3.3e-05) (Figure S2B).

To investigate whether *PKNOX2* was independent prognostic factor, we carried out cox analysis. In univariate analysis, the results indicated that PKNOX2 was an independent protective factor (HR = 0.889 [0.802-0.984], P-Value = 0.024), whereas pathological stage (HR = 1.589 [1.376-1.836], *P*.Value < 0.001), T stage (HR = 0.889 [0.802-0.984], *P*.Value = 0.024), and N stage (HR = 0.889 [0.802-0.984], *P*.Value < 0.001) are independent risk factors (**Figure 4A**). In multivariate analysis, *PKNOX*2 was also an independent protective factor (HR = 0.878 [0.788-0.978], *P*.Value = 0.018), whereas pathological stage (HR = 1.358 [1.093-1.689], *P*.Value = 0.006) is independent risk factors (**Figure 4B**). Calibration plot showed that the nomogram-predicted 3-year and 5-year survival probabilities and corresponded closely to the actual observed proportions (**Figure 4C and D**).

Loss of PKNOX2 Expression in Lung Adenocarcinoma

In LUAD, *PKNOX*2 expression is frequently suppressed. For example, its expression levels were significantly lower in human lung tumor samples from two RNA-seq datasets: TCGA-LUAD (log_2FC = -2.70, *P-Value* = 2.76e-07) (**Figure 5A**) and GSE40419 (log_2FC = -1.43, *P.Value* = 7.43e-32) (**Figure 5B**). A similar situation was discovered in Kras-G12dD and p53$^{-/-}$ murine lung cancer model. When compared to normal counterparts, *Pknox2* was greatly reduced in mouse lung tumor epithelial cells (log_2FC = -2.50, *P.Value* = 1.23e-08) (E-GEOD-59831) (**Figure 5C**). In human LUAD microarray datasets, we also found significant *PKNOX*2 downregulations, including GSE10072 (log_2FC = -0.91, *P-Value* = 9.50e-30) (**Figure S3A**), GSE18842 (log_2FC = -1.28, *P*.Value = 7.52e-22) (**Figure S3B**), GSE19188 (log_2FC = -1.36, *P.Value* = 4.41e-26) (**Figure S3C**), GSE31210 (log_2FC = -1.54, *P.Value* = 5.01e-14) (**Figure S3D**), GSE32863 (log_2FC = -0.88, *P.Value* = 9.21e-30) (**Figure S3E**), GSE43458 (log_2FC = -0.79, *P.Value* = 1.47e-21) (**Figure S3F**), GSE43458 (log_2FC = -3.79, *P.Value* = 2.08e-05) (**Figure S3G**), GSE44077 (log_2FC = -0.88, *P.Value* = 4.33e-12) (**Figure S3H**), GSE116959 (log_2FC = -1.55, *P.Value* = 5.67e-11) (**Figure S3I**), GSE118370 (log_2FC = -1.71, *P.Value* = 1.22e-04) (**Figure S3J**), GSE130779 (log_2FC = -3.79, *P.Value* = 1.96e-05) (**Figure S3K**). *Pknox*2 expression was also reduced in spontaneous lung tumors in B6C3F1 mice (log_2FC = -3.82, *P.Value* = 4.96e-06) (**Figure S3L**) (GSE31013).

A decrease in *PKNOX*2 expression was linked to a more advanced pathological stage and worse prognosis. For example, in the GSE31210 cohort, *PKNOX*2 levels were higher in Stage I than in Stage II (log_2FC = -1.54, *P.Value* = 5.01e-14) (**Figure 5D**). In the TCGA cohort, a similar situation was observed. TNM (tumor-node-metastasis) staging revealed that *PKNOX*2 was reduced in samples of higher T (**Figure 5E**) (T1 vs T2: *P.Value* = 8.18e-06; T1 vs T3: *P*.Value = 0.0012; T1 vs T4: *P*.Value = 0.044) and N stages (**Figure 5F**) (N0 vs N1: *P.Value* = 3.42e-02). Meanwhile, *PKNOX*2 downregulation was positively associated with tumor progression according to the pathological stage classification (**Figure 5G**)

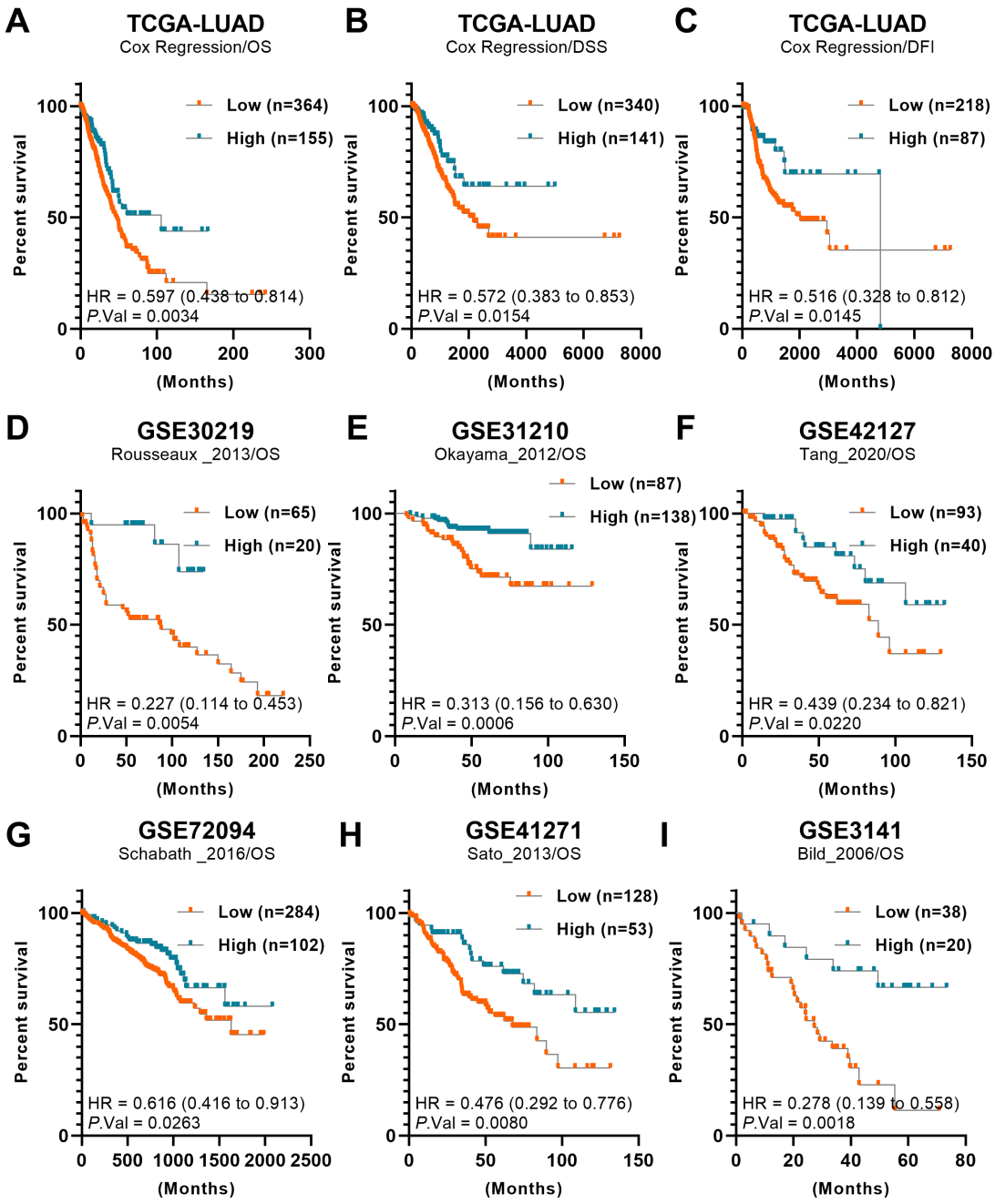

Figure 3 | Prognostic value of PKNOX2 in LUAD. (A) Univariate cox regression analysis Overall survival (OS), **(B)** Disease specific survival (DSS) and **(C)** Progression free survival (DFI) in the TCGA-LUAD cohort. The univariate cox regression analysis in various lung adenocarcinoma datasets including GSE30219 **(D)**, GSE31210 **(E)**, GSE42127 **(F),** GSE72094 **(G)**, GSE41271 **(H)** and GSE3131 **(I)** datasets.

(Stage I vs Stage II: *P.Value* = 2.74e-02). Furthermore, in the TCGA-LUAD cohort, *PKNOX2* expression was associated with treatment outcomes. Its expression was significantly higher in patients who had a complete remission/response (CR) versus those who progressed after therapy (*P.Value* = 0.049) (**Figure 5H**). *PKNOX2* mRNA levels, however, did not differ between the stable and progressive disease groups (*P.Value* = 0.51). *PKNOX2* was also found to be highly expressed in females (*P.Value* = 0.026) and older patients (age > 65) (*P.Value* = 0.030). Furthermore, *PKNOX2* mRNA levels in blood platelets from patients with non-small cell lung carcinoma (n = 60) were lower than in healthy donors (n = 59) from dataset E-GEOD-68086 (log_2FC = 2.70, *P.Value* = 2.76e-07) (**Figure 5K**).

Loss of PKNOX2 Expression in Lung Squamous Cell Carcinoma

In addition, we also found that *PKNOX2* was significantly down-regulated in various LUSC (lung squamous cell carcinoma) datasets, including TCGA-LUSC (log_2FC = -1.71, *P.Value* = 1.22e-04) (**Figure 4SA**), GSE2088 (log_2FC = -1.16, *P.Value* = 2.44e-20) (**Figure 4SB**), GSE30219 (log_2FC = -1.02, *P.Value* = 1.04e-11) (**Figure 4SC**) and GSE19188 (log_2FC = -1.55, *P.Value* = 2.01e-25) (**Figure 4SD**). These

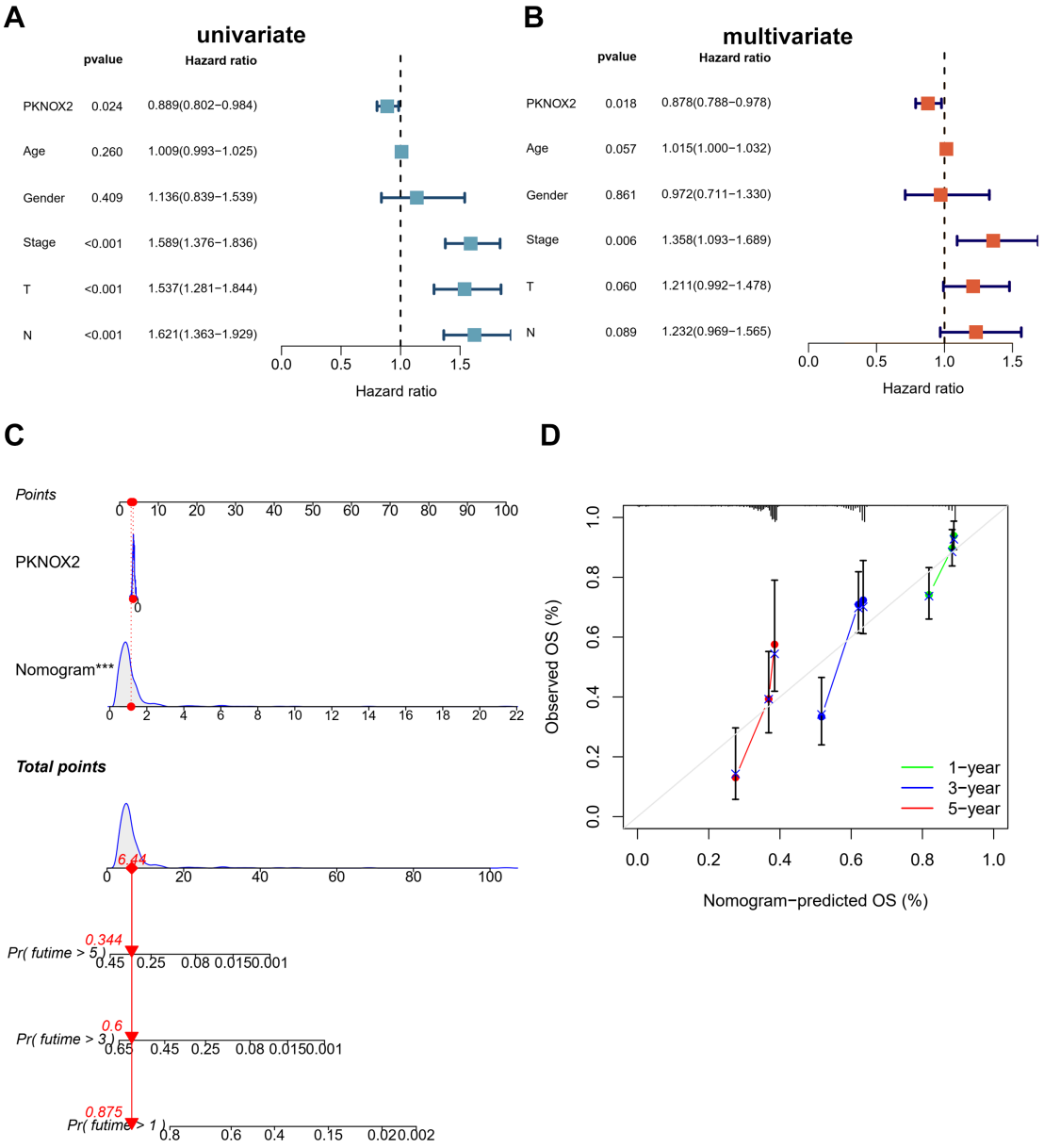

Figure 4 | Independent prognostic analysis of PKNOX2 and clinical parameters. (A) The univariate Cox regression analysis in the training cohort. **(B)** The multivariate Cox regression analysis in the training cohort. **(C)** Nomogram combining signature with clinicopathological features. **(D)** Calibration plot showing that nomogram-predicted survival probabilities corresponded closely to the actual observed proportions.

findings suggested that PKNOX2 may be frequently dysregulated in lung cancers of both the major and minor types, including LUSC and LUAD.

Gene Enrichment Analysis of PKNOX2 in Lung Adenocarcinoma

We then divided the samples into two groups based on the median expression value of *PKNOX*2 in TCGA-LUAD tumor samples, that was high (n = 238) and low expression group (239). We obtained expression profile for 38,804 genes, of which 90 were significantly differentially expressed between the two groups, with 1087 genes enriched in the *PKNOX*2 high expression group and 8 genes enriched in the *PKNOX*2 low expression group (log_2FC >1 or <-1, *P*.Value < 0.05) (**Figure 6A**). Among these were 661 protein-coding genes, 88 lincRNA genes, and 72 antisense-lncRNA genes. *ITLN*1, *VSX*2, and *PRG*4 were the most positively associated genes, while *CALML*3, *FGB*, and *FGF*5 were the most negatively associated genes. Interestingly, some positively correlated genes have been linked to tumor suppression such as *ITLN*1 and *PRG*4. There were significant differences in gender, pathological staging, T and N stage in these two groups. We loaded the differential genes into the metascape and performed Pathway & Process Enrichment assay. Pathway & Process Enrichment revealed that differential genes were significantly enriched in extracellular signal regulation and other related pathways such as chemotaxis (GO:0006935; log*P* = -23.26), Calcium signaling pathway (hsa04020; log*P* = -13.94), Extracellular matrix organization (R-HSA-1474244; log*P* = -17.73), cell-cell adhesion (GO:0098609; log*P* = -14.19) and regulation of cell adhesion (GO:0030155; log*P* = -13.01) (**Figure 6B**). The Protein-pro-

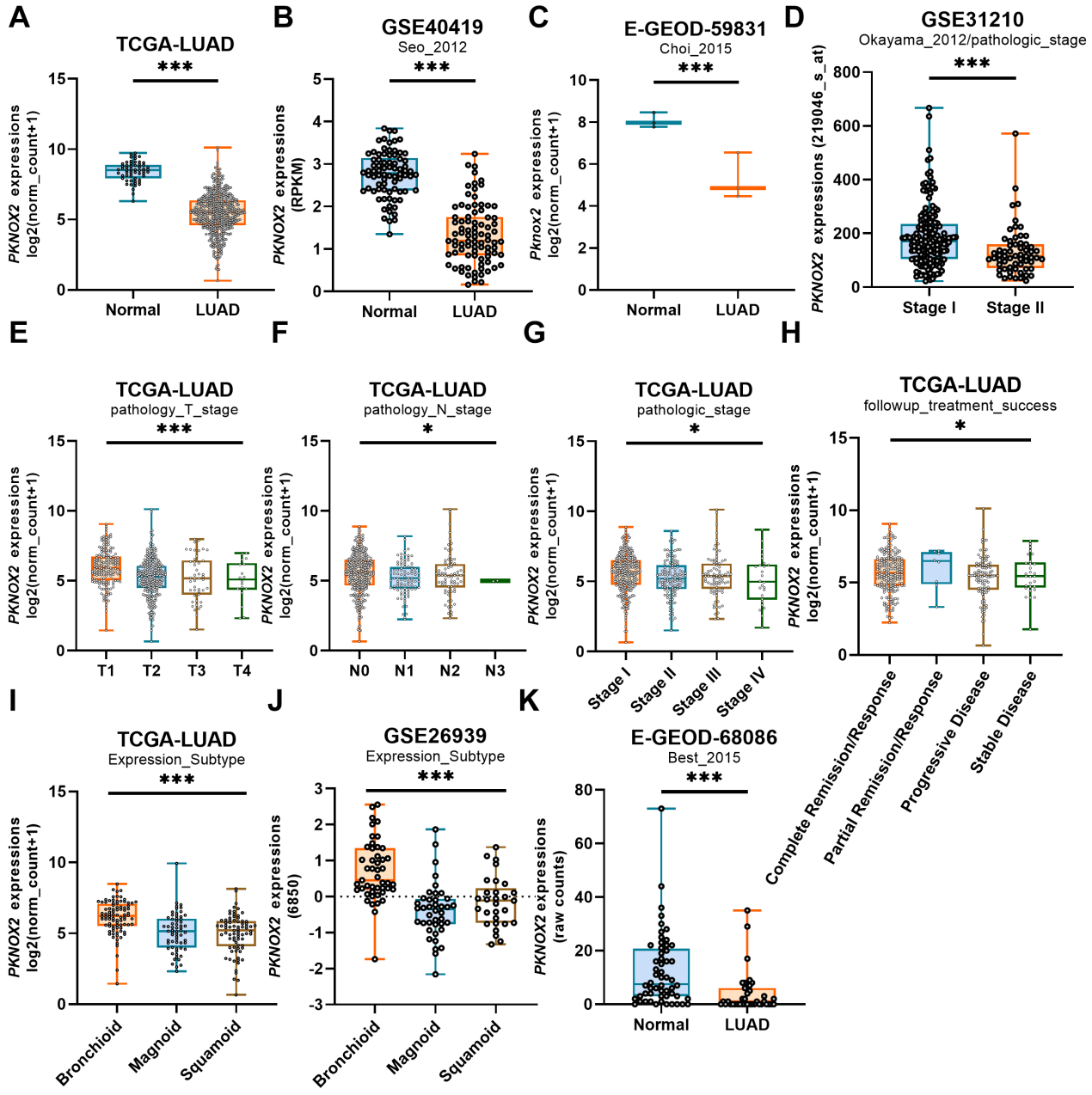

Figure 5 | Down-regulation of PKNOX2 in lung adenocarcinoma datasets Boxplots for the expression changes for *PKNOX2* in human and mice LUAD RNA-seq data (**A**) TCGA-LUAD, (**B**) GSE40419 and (**C**) E-GEOD-59831. (**D**) Boxplots for *PKNOX2* expression levels in pathological stage I-II lung adenocarcinoma tissues in GSE31210. (**E, F, G**) The *PKNOX2* expression in different pathological stages, T and N stages from samples of TCGA-LUAD. (**H**) The PKNOX2 expressions in samples with different follow-up treatment outcomes. The PKNOX2 expression in (**I**) age (>65 or <=65) and (**J**) gender groups. (K) *PKNOX2* expressions in blood platelets from LUAD and healthy donors in the RNA-seq dataset E-GEOD-68086. 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 $\log_2(\text{TPM} + 1)$.

tein Interaction Enrichment network suggested 19 MCODE clusters (**Figure 6C**). KEGG enrichment results demonstrate its key proliferative signaling pathways among Cytokine-cytokine receptor interaction (hsa04060; log*P* = -2.86) as well as their downstream, such as PI3K-Akt signaling pathway (hsa04151; log*P* = -2.63) and cAMP signaling pathway (hsa04024; log*P* = -2.09) (**Figure 6D**). These findings suggest that PKNOX2 is involved in extracellular matrix signaling pathways.

Immune Filtration of PKNOX2 in Lung Adenocarcinoma

We first used the ESTIMATE algorithm (TIMER) to score the stromal as well as the immune of tumor samples in the TCGA-LUAD dataset. *PKNOX2* high expression group possessed a significant higher stromal score, immune score and ESTIAMTE score (**Figure 7A**). Using the CIBERSORT algorithm, we scored 22 immune cells in different samples and then compared the differences in *PKNOX2* high and low

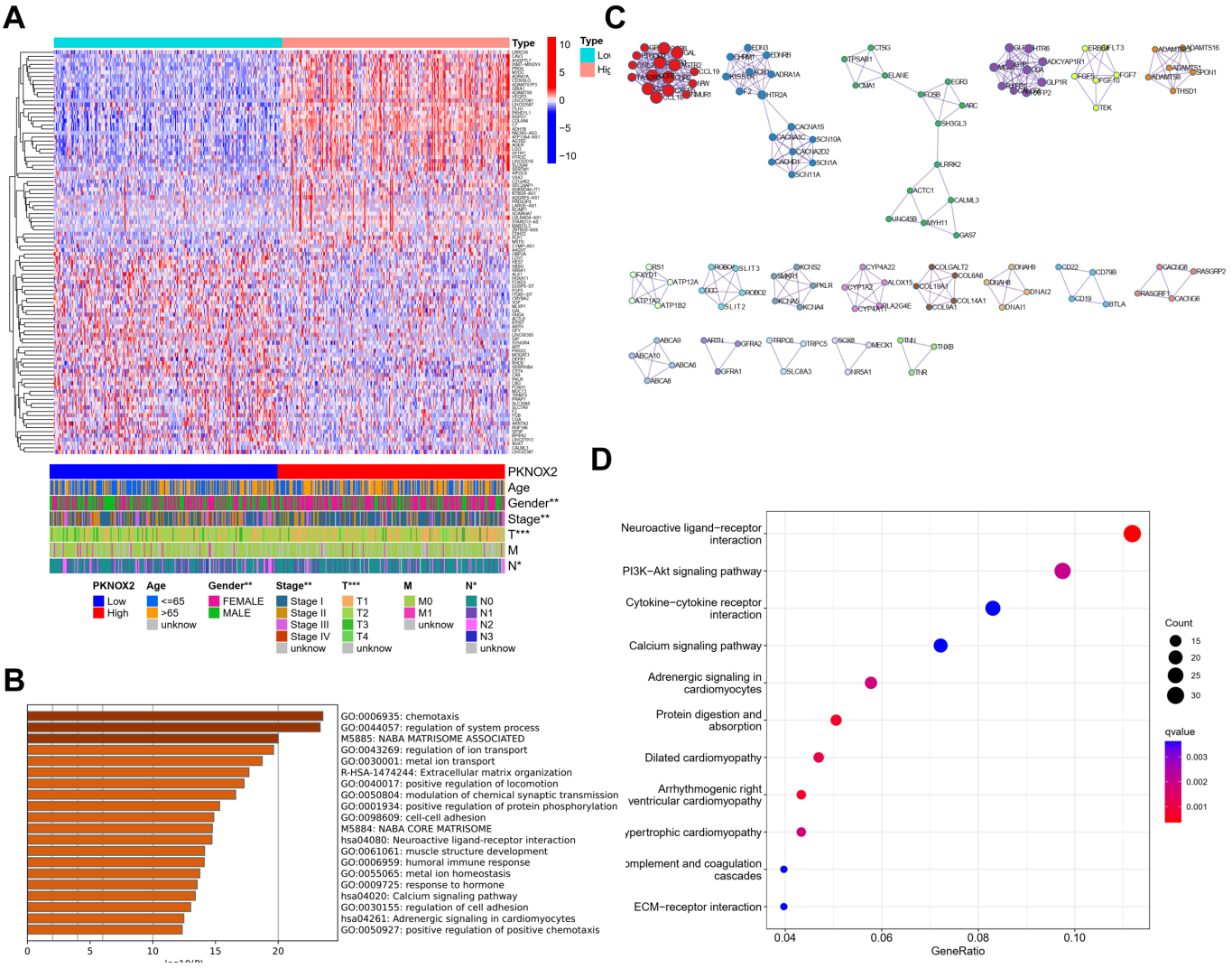

Figure 6 | Gene enrichment analysis and PKNOX2 regulatory co-expression network. (A) Heatmap of most differential expressed genes according to the expression status of PKNOX2 (high and low). The enrichment analysis was performed by Metascape. **(B)** Bar graph of enriched terms of the upregulated genes (colored by p-values). **(C)** The network of enriched terms (colored by cluster ID) of the upregulated genes. **(D)** The KEGG enrichment analysis of the most differentially expressed genes according to PKNOX2 expression level.

groups. Mast cell resting, T cell CD4 memory resting, and Monocytes are significantly higher in the *PKNOX2*-high expression group (**Figure 7B**). Meanwhile, *PKNOX2*-low expression group has a significantly increased T cells CD4 memory activated, T cell regulatory (Treg), T cells follicular helper and M0 macrophage (Figure 7B). We show scatter plots of sample *PKNOX2* expression in lung adenocarcinoma versus T cells CD4 memory activated, T cell regulatory (Treg), and T cells follicular helper (**Figure 7C**).

Tumor Suppressor FHL1 as a Potential Target of PKNOX2 in Lung Tissues

Tumor suppressor *FHL1* (Half LIM domains 1) was significantly positively associated in *PKNOX2* expression in various LUAD datasets, including TCGA (r = 0.71, *P.Value* < 2.2e-16), GSE19804 (r = 0.87, *P.Value* = 9.4e-38) (**Figure S5A**), GSE29016 (r = 0.76, *P.Value* = 4.51e-08) (**Figure S5B**), GSE40419 (r = 0.88 *P.Value* = 8.31e-53) (**Figure S5C**), GSE43458 (r = 0.87, *P.Value* = 6.69e-39) (**Figure S5D**), GSE116959 (r = 0.35, *P.Value* = 0.0035) (**Figure S5E**) and GSE130779 (r = 0.33, *P.Value* = 0.0053) (**Figure S5F**). In GTEx normal lung tissues (n = 427), *FHL1* also exhibits significant positive correlation with (r = 0.60, *P.Value* = 1.65e-42) (**Figure S5G**).

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

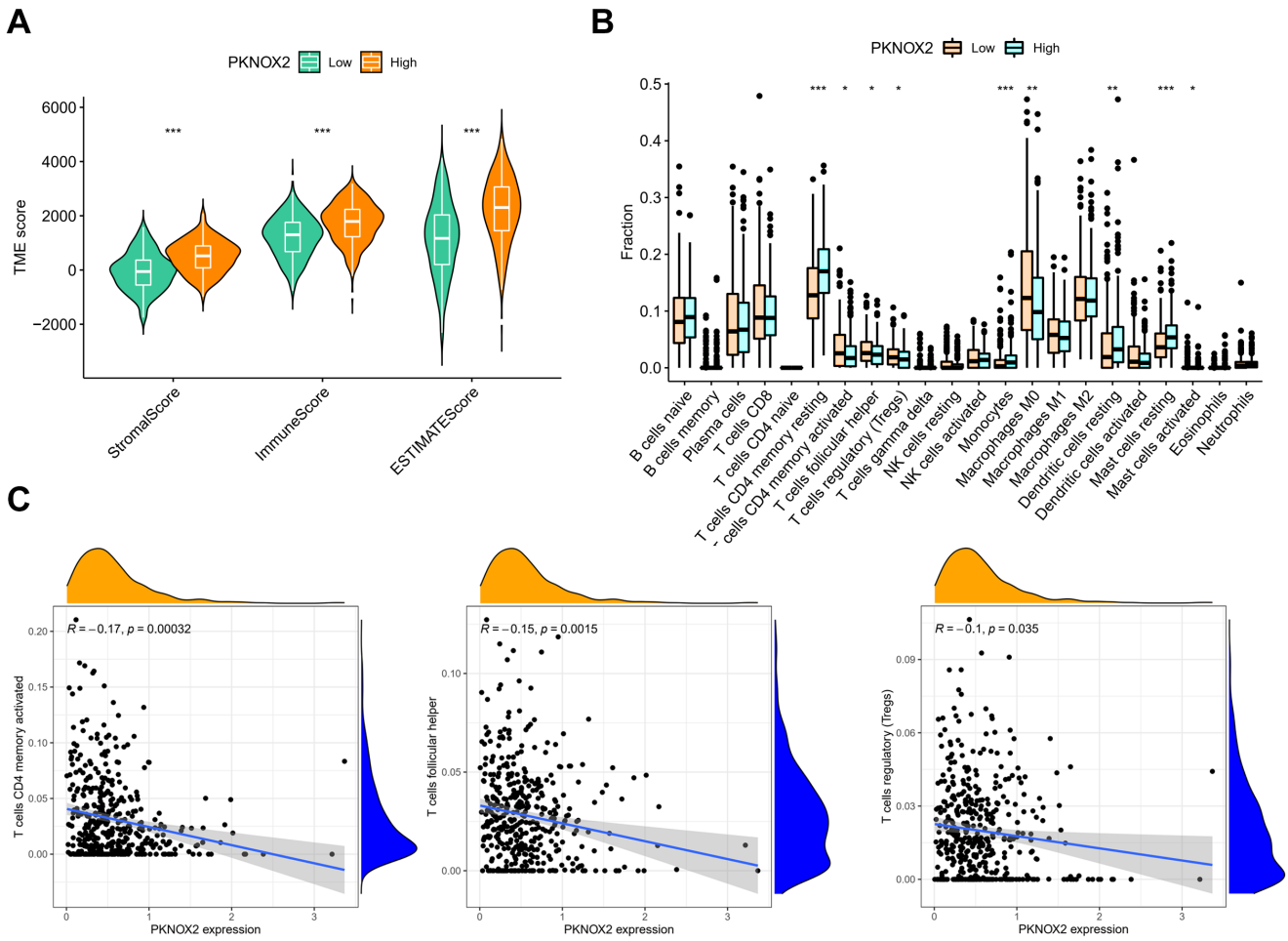

Figure 7 | Immune status associated with PKNOX2 expression in LUAD. (A) The Stromal Score, Immune Score and ESTIMATE Score were calculated by CIBERSORT and compared in different groups of PKNOX2 expression status (high vs low). **(B)** The associated of PKNOX2 with LM22 immune cells was calculated according to the expression of LM22 signature genes. (**C**) The scatter plot for the association of PKNOX2 and two specific type of immune cells including T cells CD4 memory activated and Macrophage M1.

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 first screened for the commonly dysregulated differentially expressed homeobox-containing genes in various lung adenocarcinoma datasets and obtained 14 differential genes. Through univariate followed by lasso regression, we found that the combination of 7 factors among these DEGs could well predict overall survival, disease-free progression, and disease-free interval time in lung adenocarcinoma patients. In the training (TCGA-LUAD) and validation datasets (GSE37745 and GSE50081), with 3-year ROC and 5-year ROC values all above 0.6. We particularly focused on PKNOX2, a typical homeobox genes that play critical roles in morphogenesis processes such as body segmentation during embryonic development such as limbs developing [51,52]. *PKNOX*2 expression and regulation in the bone marrow mesenchymal stem cells of Fanconi anemia patients and healthy donors. We discovered a significant loss of PKNOX2 expression and associated overall survival in various LUAD datasets in this study. Using univariate and multivariate Cox regression analysis, this evidence suggested that PKNOX2 could be used as an independent prognostic marker for LUAD. *PKNOX*2 expression was further reduced in lung adenocarcinoma with tissue malignancy progression. Previous report showed that PKNOX2 suppresses gastric cancer through the transcriptional activation of IGFBP5 and p53 [53]. However, its role in lung adenocarcinoma is still unknown. We carried out GSEA analysis and KEGG pathway enrichment in groups TGF-β pathway induced EMT and stemness characteristics are associated with epigenetic regulation in lung cancer [54]. We found *PKNOX*2-related genes enriched to extracellular matrix remodeling as well as PI3K/AKT pathway in the differential expression group of PKNOX2. It might be possible to investigate the negative regulation between PI3K/AKT pathway and PKNOX2 in LUAD.

Taken together, we showed differential expressed homeobox gene patterns in LUAD and further validated that *PKNOX2* 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 *PKNOX2* in controlling the tumorigenesis in the future.

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

Funding This work supported by the Natural Science Foundation of Hunan Province, China; No.2024JJ9266.

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 O, 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

Supplementary Materials

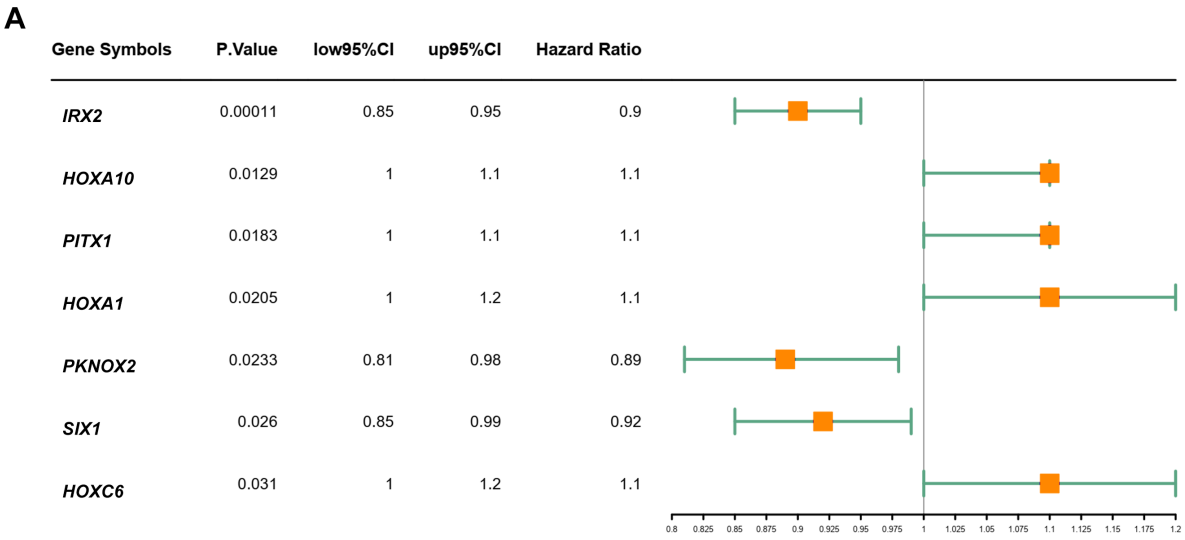

Figure S1 | Univariate Cox regression analysis identified seven DE homeobox-containing genes associated with OS in TCGA-LUAD. (**A**) Forest plot of the hazard ratios for overall survival assessed by the expression levels of each of the eight homeobox-containing genes in the TCGA-LUAD cohort using a univariate Cox regression model. Hazard ratios greater than one represented risk factors for survival, while hazard ratios less than one represented protective factors for survival.

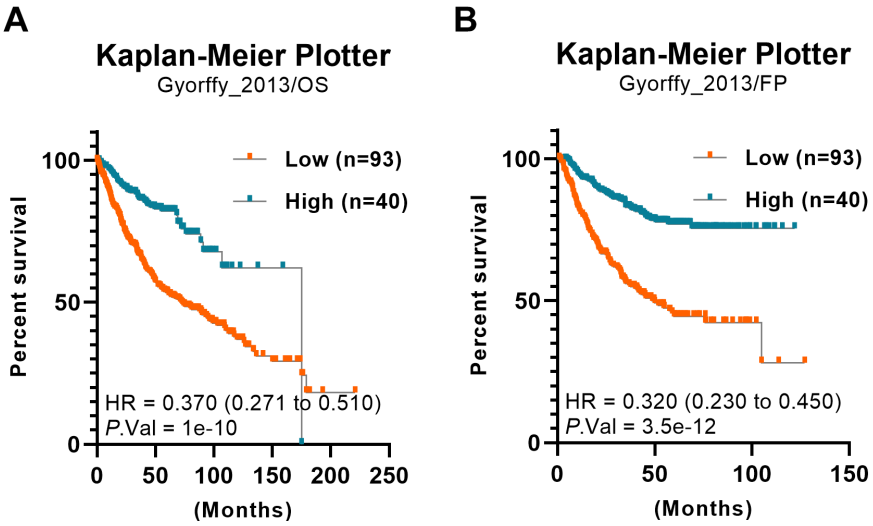

Figure S2 | PKNOX2 correlates with prognosis in the TCGA-LUAD dataset. Kaplan-meier survival analysis of PKNOX2 in lung adenocarcinoma using OS (overall survival) and FP (free progression). The data were obtained from the Kaplan-Meier plotter (www.kmplot.-com).

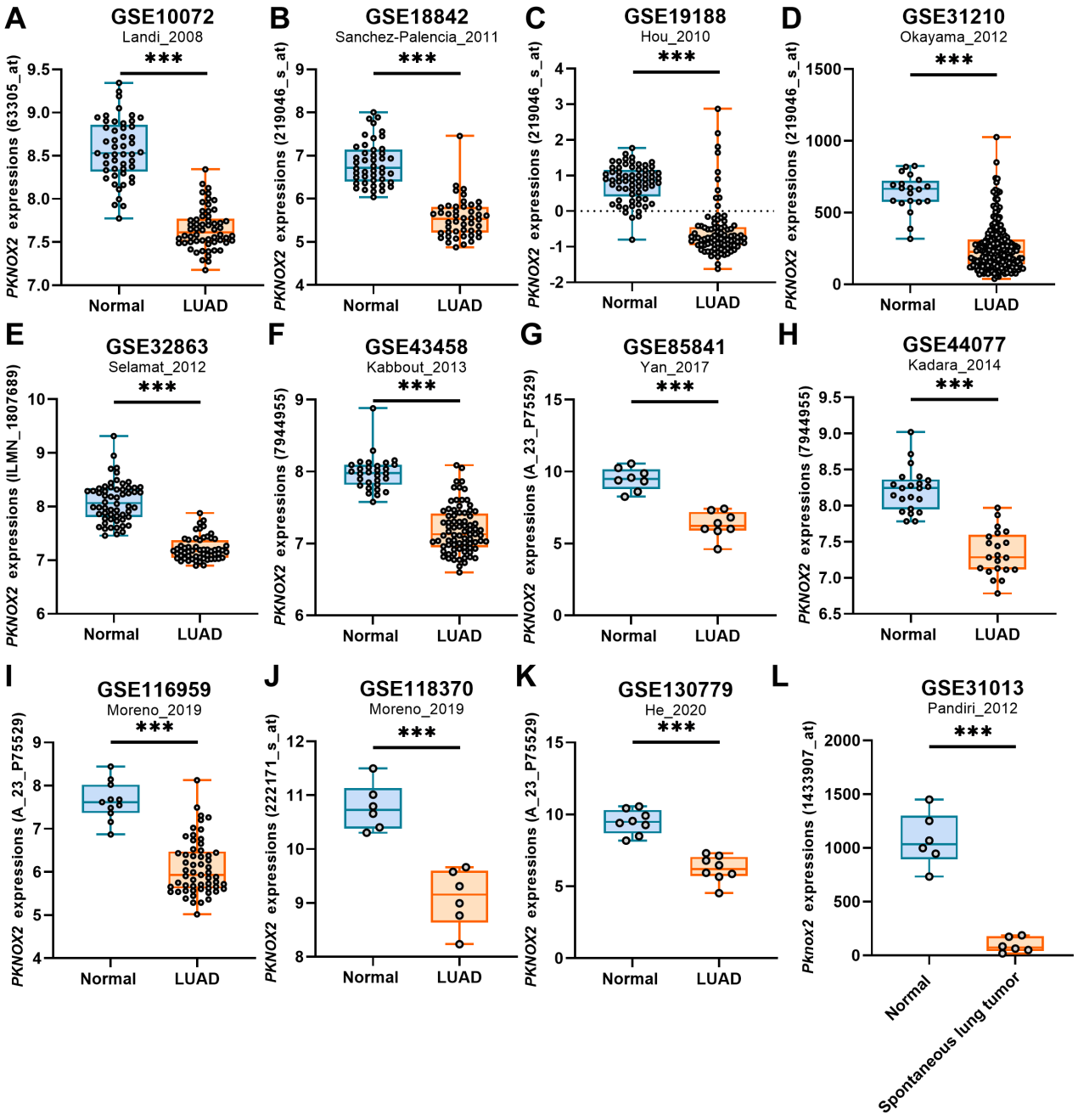

Figure S3 | Expression of PKNOX2 in lung adenocarcinoma microarray datasets. The human *PKNOX*2 expressions in datasets including (**A**) GSE10072, (**B**) GSE18842, (**C**) GSE19188, (**D**) GSE31210, (**E**) GSE32863, (**F**) GSE43458, (**G**) GSE85841, (**H**) GSE44077, (**I**) GSE116959, (**J**) GSE118370 and (**K**) GSE130779. (**L**) Dysregulation of *Pknox*2 expression in spontaneous lung tumors from B6C3F1 mice model according to the GSE31013 dataset. The specific probes used in these datasets against *PKNOX*2 are also indicated on the left side bar. The orange and blue boxes represent the cancerous and normal tissues, respectively (***p<0.01).

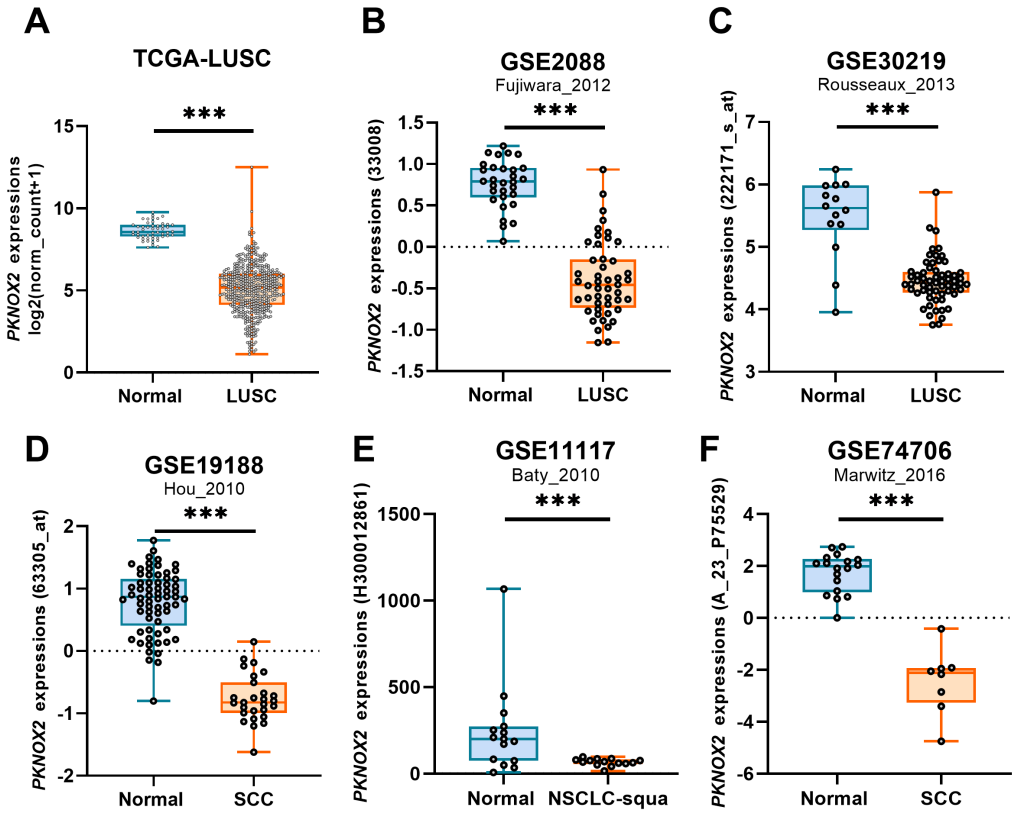

Figure S4 | Expression of PKNOX2 in lung squamous cell carcinoma datasets. The *PKNOX*2 mRNA expressions in datasets including (A) TCGA-LUSC, (**B**) GSE2088, (**C**) GSE30219, (**D**) GSE19188, (E) GSE11117 and (F) GSE74706. The orange and blue boxes represent cancerous tissues and normal tissues, respectively (***p<0.01).

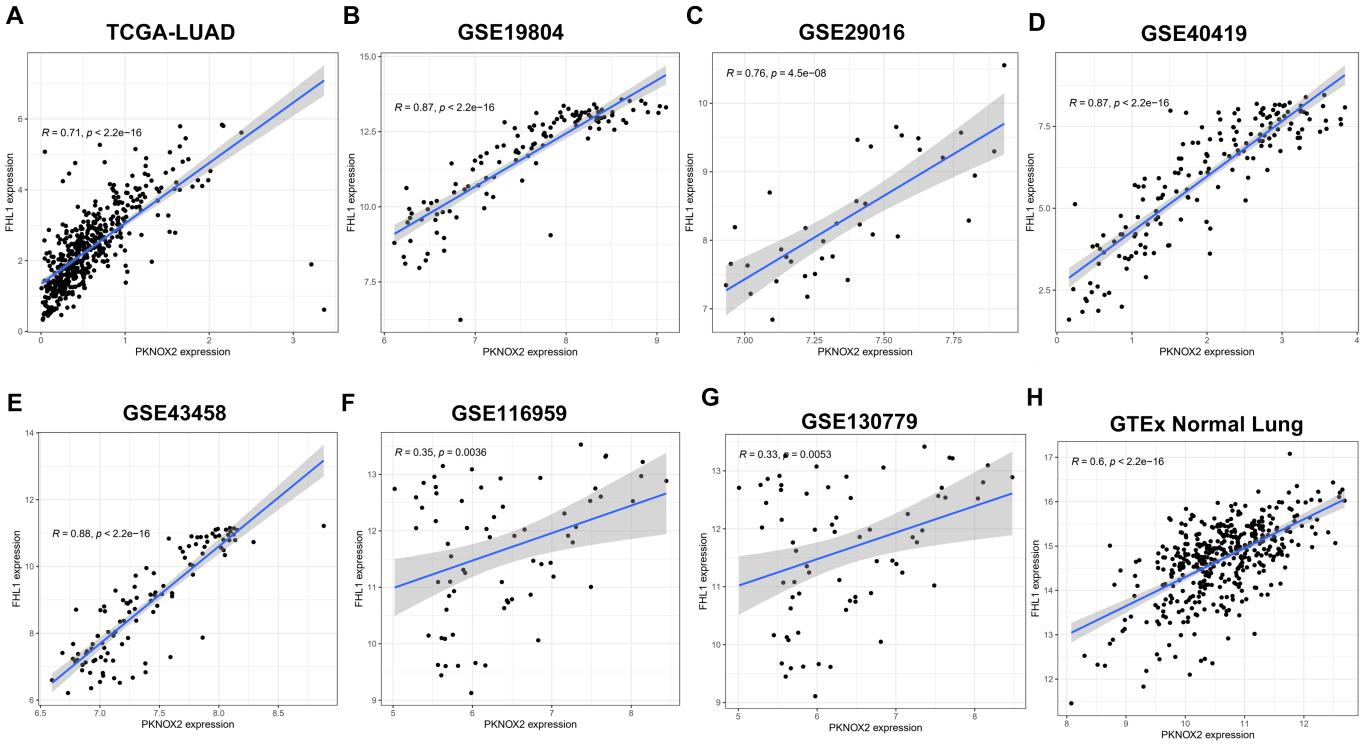

Figure S5 | Expression correlation of FHL1 and PKNOX2 in lung adenocarcinoma microarray datasets. The scatter plot showing the positive correlation of human *PKNOX*2 and *FHL*1 expressions in LUAD including **(A)** TCGA-LUAD, **(B)** GSE19804, **(C)** GSE29016, **(D)** GSE40419, **(E)** GSE43458, **(F)** GSE116959 and **(G)** GSE130779. **(H)** The expression correlation of *PKNOX*2 and *FHL*1 in 427 samples of normal lung tissue from GTEx (The Genotype-Tissue Expression) datasets.

Table S1 | The detailed information about the homeobox-containing genes in this study

gene_name	gene_id	seqnames	start	end	width	strand
PAX7	ENSG00000009709	1	18631006	18748866	117861	+
POU3F1	ENSG00000185668	1	38044611	38046794	2184	-
DMBX1	ENSG00000197587	1	46506996	46514226	7231	+
LHX8	ENSG00000162624	1	75128434	75161533	33100	+
BARHL2	ENSG00000143032	1	90711539	90717237	5699	-
ALX3	ENSG00000156150	1	110059994	110070700	10707	-
CERS2	ENSG00000143418	1	150960583	150975004	14422	-
PBX1	ENSG00000185630	1	164555584	164899296	343713	+
LMX1A	ENSG00000162761	1	165201867	165356715	154849	-
POU2F1	ENSG00000143190	1	167220829	167427345	206517	+
PRRX1	ENSG00000116132	1	170662728	170739419	76692	+
LHX4	ENSG00000121454	1	180230286	180278982	48697	+
LHX9	ENSG00000143355	1	197911902	197935478	23577	+
PROX1	ENSG00000117707	1	213983181	214041502	58322	+
HLX	ENSG00000136630	1	220879400	220885059	5660	+
MIXL1	ENSG00000185155	1	226223618	226227054	3437	+
SIX3	ENSG00000138083	2	44941898	44946077	4180	+
SIX2	ENSG00000170577	2	45005161	45009430	4270	-
OTX1	ENSG00000115507	2	63050057	63057836	7780	+
MEIS1	ENSG00000143995	2	66433452	66573869	140418	+
VAX2	ENSG00000116035	2	70900590	70933446	32857	+
EMX1	ENSG00000135638	2	72916260	72936071	19812	+
NOTO	ENSG00000214513	2	73202258	73212513	10256	+
LBX2	ENSG00000179528	2	74497517	74503316	5800	-
TLX2	ENSG00000115297	2	74513463	74517147	3685	+
POU3F3	ENSG00000198914	2	104855511	104858574	3064	+
PAX8	ENSG00000125618	2	113215997	113278950	62954	-
EN1	ENSG00000163064	2	118842171	118847678	5508	-
ZEB2	ENSG00000169554	2	144364364	144524583	160220	-
CERS6	ENSG00000172292	2	168455862	168775137	319276	+
DLX1	ENSG00000144355	2	172084740	172089677	4938	+
DLX2	ENSG00000115844	2	172099439	172102900	3462	-
EVX2	ENSG00000174279	2	176077472	176083913	6442	-
HOXD13	ENSG00000128714	2	176092891	176095938	3048	+
HOXD12	ENSG00000170178	2	176099730	176101193	1464	+
HOXD11	ENSG00000128713	2	176104216	176109754	5539	+
HOXD10	ENSG00000128710	2	176108790	176119942	11153	+
HOXD9	ENSG00000128709	2	176122720	176124937	2218	+
HOXD8	ENSG00000175879	2	176129694	176132695	3002	+
HOXD3	ENSG00000128652	2	176136612	176173102	36491	+
HOXD4	ENSG00000170166	2	176151222	176153226	2005	+
HOXD1	ENSG00000128645	2	176188579	176190907	2329	+
SATB2	ENSG00000119042	2	199269500	199471266	201767	-
PAX3	ENSG00000135903	2	222199888	222298996	99109	-
GBX2	ENSG00000168505	2	236165236	236168369	3134	-
SATB1	ENSG00000182568	3	18345387	18445588	100202	-
HESX1	ENSG00000163666	3	57197843	57226521	28679	-
POU1F1	ENSG00000064835	3	87259404	87276587	17184	-
ARGFX	ENSG00000186103	3	121570704	121586634	15931	+
SHOX2	ENSG00000168779	3	158095954	158106503	10550	-
NKX1-1	ENSG00000235608	4	1402932	1406331	3400	-
MSX1	ENSG00000163132	4	4859666	4863936	4271	+
HMX1	ENSG00000215612	4	8846076	8871817	25742	-
NKX3-2	ENSG00000109705	4	13540830	13545050	4221	-
PHOX2B	ENSG00000109132	4	41744082	41748970	4889	-
GSX2	ENSG00000180613	4	54099523	54102505	2983	+
HOPX	ENSG00000171476	4	56647988	56681899	33912	-
NKX6-1	ENSG00000163623	4	84491987	84498450	6464	-
PITX2	ENSG00000164093	4	110617423	110642123	24701	-
POU4F2	ENSG00000151615	4	146638893	146642474	3582	+
DUX4	ENSG00000260596	4	190173774	190185942	12169	+
IRX4	ENSG00000113430	5	1877413	1887236	9824	-
IRX2	ENSG00000170561	5	2745845	2751662	5818	-
IRX1	ENSG00000170549	5	3596054	3601403	5350	+
ISL1	ENSG00000016082	5	51383391	51394738	11348	+
OTP	ENSG00000171540	5	77628712	77639688	10977	-
POU5F2	ENSG00000248483	5	93733220	93741637	8418	-
PITX1	ENSG00000069011	5	135027735	135034813	7079	-
POU4F3	ENSG00000091010	5	146339024	146340520	1497	+
CDX1	ENSG00000113722	5	150166795	150184558	17764	+
TLX3	ENSG00000164438	5	171309284	171312134	2851	+
NKX2-5	ENSG00000183072	5	173232109	173235357	3249	-
MSX2	ENSG00000120149	5	174724533	174730893	6361	+
PROP1	ENSG00000175325	5	177992235	177996242	4008	-
POU5F1	ENSG00000204531	6	31164337	31180731	16395	-
PBX2	ENSG00000204304	6	32184741	32190186	5446	-
POU3F2	ENSG00000184486	6	98834592	98839470	4879	+

gene_name	gene_id	seqnames	start	end	width	strand
UNCX	ENSG00000164853	7	1232907	1237318	4412	+
MEOX2	ENSG00000106511	7	15611212	15686812	75601	-
HOXA1	ENSG00000105991	7	27092993	27095996	3004	-
HOXA2	ENSG00000105996	7	27100354	27102811	2458	-
HOXA3	ENSG00000105997	7	27106184	27152581	46398	-
HOXA4	ENSG00000197576	7	27128507	27130799	2293	-
HOXA5	ENSG00000106004	7	27141052	27143668	2617	-
HOXA6	ENSG00000106006	7	27145396	27150603	5208	-
HOXA7	ENSG00000122592	7	27153716	27157936	4221	-
HOXA9	ENSG00000078399	7	27162435	27175180	12746	-
HOXA10	ENSG00000253293	7	27170591	27180261	9671	-
HOXA11	ENSG00000005073	7	27181510	27185223	3714	-
HOXA13	ENSG00000106031	7	27193503	27200106	6604	-
EVX1	ENSG00000106038	7	27242700	27250493	7794	+
POU6F2	ENSG00000106536	7	38977998	39493095	515098	+
DLX6	ENSG00000006377	7	97005548	97011039	5492	+
DLX5	ENSG00000105880	7	97020392	97025097	4706	-
CUX1	ENSG00000257923	7	101815904	102283957	468054	+
PAX4	ENSG00000106331	7	127610292	127618114	7823	-
NOBOX	ENSG00000106410	7	144397240	144410227	12988	-
GBX1	ENSG00000164900	7	151148589	151174745	26157	-
EN2	ENSG00000164778	7	155458129	155464831	6703	+
MNX1	ENSG00000130675	7	156994051	157010651	16601	-
SHOX	ENSG00000185960	X	624344	659411	35068	+
ARX	ENSG00000004848	X	25003694	25016420	12727	-
CDX4	ENSG00000131264	X	73447254	73455245	7992	+
POU3F4	ENSG00000196767	X	83508250	83512127	3878	+
HDX	ENSG00000165259	X	84317874	84502479	184606	-
TGIF2LX	ENSG00000153779	X	89921882	89922883	1002	+
ESX1	ENSG00000123576	X	104250038	104254933	4896	-
RHOXF2B	ENSG00000203989	X	120070672	120077705	7034	-
RHOXF1	ENSG00000101883	X	120109053	120115937	6885	-
RHOXF2	ENSG00000131721	X	120158561	120165630	7070	+
NKX3-1	ENSG00000167034	8	23678693	23682927	4235	-
NKX2-6	ENSG00000180053	8	23702451	23706598	4148	-
HMBOX1	ENSG00000147421	8	28890394	29064764	174371	+
NKX6-3	ENSG00000165066	8	41645178	41650669	5492	-
ZFHX4	ENSG00000091656	8	76681219	76867285	186067	+
ZHX2	ENSG00000178764	8	122781394	122974512	193119	+
ZHX1	ENSG00000165156	8	123248451	123275541	27091	-
PAX5	ENSG00000196092	9	36833275	37034185	200911	-
BARX1	ENSG00000131668	9	93951622	93955372	3751	-
LHX6	ENSG00000106852	9	122202577	122229626	27050	-
LHX2	ENSG00000106689	9	124001670	124033301	31632	+
PBX3	ENSG00000167081	9	125747345	125967377	220033	+
LMX1B	ENSG00000136944	9	126614443	126701032	86590	+
PRRX2	ENSG00000167157	9	129665641	129722674	57034	+
BARHL1	ENSG00000125492	9	132582185	132590266	8082	+
LHX3	ENSG00000107187	9	136196250	136205109	8860	-
DBX1	ENSG00000109851	11	20156155	20160613	4459	-
PAX6	ENSG00000007372	11	31784779	31818062	33284	-
ALX4	ENSG00000052850	11	44260444	44310166	49723	-
PHOX2A	ENSG00000165462	11	72239077	72245664	6588	-
POU2F3	ENSG00000137709	11	120236640	120319944	83305	+
BSX	ENSG00000188909	11	122977570	122981720	4151	-
PKNOX2	ENSG00000165495	11	125164687	125433389	268703	+
BARX2	ENSG00000043039	11	129375940	129452279	76340	+
MKX	ENSG00000150051	10	27672875	27746060	73186	-
ZEB1	ENSG00000148516	10	31318495	31529814	211320	+
DRGX	ENSG00000165606	10	49364181	49396016	31836	-
HHEX	ENSG00000152804	10	92689951	92695646	5696	+
NKX2-3	ENSG00000119919	10	99532933	99536524	3592	+
PAX2	ENSG00000075891	10	100735603	100829941	94339	+
TLX1	ENSG00000107807	10	101130505	101137789	7285	+
LBX1	ENSG00000138136	10	101226195	101229794	3600	-
PITX3	ENSG00000107859	10	102230186	102241474	11289	-
VAX1	ENSG00000148704	10	117128521	117138301	9781	-
EMX2	ENSG00000170370	10	117542444	117549546	7103	+
HMX3	ENSG00000188620	10	123135962	123137741	1780	+
HMX2	ENSG00000188816	10	123148122	123150672	2551	+
NKX1-2	ENSG00000229544	10	124445239	124450184	4946	-
NKX6-2	ENSG00000148826	10	132783179	132786052	2874	-
VENTX	ENSG00000151650	10	133237404	133241929	4526	+
NANOGNB	ENSG00000205857	12	7765216	7774121	8906	+
NANOG	ENSG00000111704	12	7787794	7799141	11348	+
DBX2	ENSG00000185610	12	45014672	45051099	36428	-
CERS5	ENSG00000139624	12	50129306	50167533	38228	-
POU6F1	ENSG00000184271	12	51186936	51217708	30773	-
HOXC13	ENSG00000123364	12	53938765	53946544	7780	+
HOXC12	ENSG00000123407	12	53954834	53958956	4123	+
HOXC11	ENSG00000123388	12	53973126	53977643	4518	+

gene_name	gene_id	seqnames	start	end	width	strand
HOXC10	ENSG00000180818	12	53985065	53990279	5215	+
HOXC6	ENSG00000197757	12	53990624	54030823	40200	+
HOXC9	ENSG00000180806	12	53994895	54003337	8443	+
HOXC8	ENSG00000037965	12	54009106	54012362	3257	+
HOXC4	ENSG00000198353	12	54016931	54056030	39100	+
HOXC5	ENSG00000172789	12	54032853	54035358	2506	+
ALX1	ENSG00000180318	12	85280107	85301784	21678	+
CUX2	ENSG00000111249	12	111034024	111350554	316531	+
LHX5	ENSG00000089116	12	113462034	113472280	10247	-
HNF1A	ENSG00000135100	12	120978543	121002512	23970	+
GSX1	ENSG00000169840	13	27792643	27794768	2126	+
PDX1	ENSG00000139515	13	27920020	27926231	6212	+
CDX2	ENSG00000165556	13	27962137	27971139	9003	-
POU4F1	ENSG00000152192	13	78598362	78603560	5199	-
HOMEZ	ENSG00000215271	14	23272422	23299447	27026	-
ZFHX2	ENSG00000136367	14	23520855	23556192	35338	-
NKX2-1	ENSG00000136352	14	36516392	36521149	4758	-
NKX2-8	ENSG00000136327	14	36580579	36582607	2029	-
OTX2	ENSG00000165588	14	56799905	56810479	10575	-
SIX6	ENSG00000184302	14	60508951	60512850	3900	+
SIX1	ENSG00000126778	14	60643415	60658259	14845	-
SIX4	ENSG00000100625	14	60709528	60724348	14821	-
VSX2	ENSG00000119614	14	74239472	74262738	23267	+
PROX2	ENSG00000119608	14	74852871	74871940	19070	-
GSC	ENSG00000133937	14	94768216	94770230	2015	-
NANOGP8	ENSG00000255192	15	35084193	35085110	918	-
MEIS2	ENSG00000134138	15	36889204	37101299	212096	-
ONECUT1	ENSG00000169856	15	52756989	52791078	34090	-
ISL2	ENSG00000159556	15	76336724	76342476	5753	+
CERS3	ENSG00000154227	15	100400395	100544995	144601	-
IRX3	ENSG00000177508	16	54283304	54286763	3460	-
IRX5	ENSG00000176842	16	54930862	54934485	3624	+
IRX6	ENSG00000159387	16	55323760	55330760	7001	+
ZFHX3	ENSG00000140836	16	72782885	73144447	361563	-
DUXB	ENSG00000282757	16	75693929	75701459	7531	-
SEBOX	ENSG00000274529	17	28364268	28365244	977	-
LHX1	ENSG00000273706	17	36936785	36944612	7828	+
HNF1B	ENSG00000275410	17	37686432	37745247	58816	-
MEOX1	ENSG00000005102	17	43640388	43661954	21567	-
HOXB1	ENSG00000120094	17	48528526	48530997	2472	-
HOXB2	ENSG00000173917	17	48540894	48544989	4096	-
HOXB3	ENSG00000120093	17	48548870	48604912	56043	-
HOXB4	ENSG00000182742	17	48575513	48580111	4599	-
HOXB5	ENSG00000120075	17	48591257	48593961	2705	-
HOXB6	ENSG00000108511	17	48595751	48604992	9242	-
HOXB7	ENSG00000260027	17	48607227	48633572	26346	-
HOXB8	ENSG00000120068	17	48611377	48614939	3563	-
HOXB9	ENSG00000170689	17	48621159	48626356	5198	-
HOXB13	ENSG00000159184	17	48724763	48729178	4416	-
DLX4	ENSG00000108813	17	49968970	49974959	5990	+
DLX3	ENSG00000064195	17	49990005	49995224	5220	-
TGIF1	ENSG00000177426	18	3411608	3459978	48371	+
ONECUT2	ENSG00000119547	18	57435685	57491297	55613	+
RAX	ENSG00000134438	18	59267035	59274086	7052	-
TSHZ1	ENSG00000179981	18	75210755	75289950	79196	+
ADNP2	ENSG00000101544	18	80109031	80147523	38493	+
NKX2-4	ENSG00000125816	20	21395367	21398028	2662	-
NKX2-2	ENSG00000125820	20	21511010	21514026	3017	-
VSX1	ENSG00000100987	20	25070885	25082365	11481	-
TGIF2	ENSG00000118707	20	36573488	36593950	20463	+
ZHX3	ENSG00000174306	20	41178448	41317672	139225	-
ADNP	ENSG00000101126	20	50888919	50931240	42322	-
TSHZ2	ENSG00000182463	20	52972407	53495330	522924	+
ONECUT3	ENSG00000205922	19	1752373	1780988	28616	+
RAX2	ENSG00000173976	19	3769089	3772221	3133	-
CERS4	ENSG00000090661	19	8206736	8262421	55686	+
PBX4	ENSG00000105717	19	19561707	19618916	57210	-
TSHZ3	ENSG00000121297	19	31274945	31349547	74603	-
LEUTX	ENSG00000213921	19	39776595	39786167	9573	+
POU2F2	ENSG00000028277	19	42086110	42196585	110476	-
SIX5	ENSG00000177045	19	45764785	45769226	4442	-
MEIS3	ENSG00000105419	19	47403124	47419523	16400	-
TPRX1	ENSG00000178928	19	47801243	47819051	17809	-
CRX	ENSG00000105392	19	47819779	47843330	23552	+
DPRX	ENSG00000204595	19	53632056	53637009	4954	+
DUXA	ENSG00000258873	19	57154021	57167443	13423	-
TGIF2LY	ENSG00000176679	Y	3579041	3580041	1001	+
GSC2	ENSG00000063515	22	19148576	19150283	1708	-
ISX	ENSG00000175329	22	35066136	35087387	21252	+
PKNOX1	ENSG00000160199	21	42974510	43033931	59422	+