



## Research article

# Identification of a prognosis-related gene signature and ceRNA regulatory networks in lung adenocarcinoma

Hong Wei, Fei Teng, XiaoLei Wang, XiuJuan Hou, HongBo Wang, Hong Wang, Hui Sun, XianLi Zhou\*

*In-Patient Ultrasound Department, The Second Affiliated Hospital of Harbin Medical University, Harbin, 150081, China*

## ARTICLE INFO

**Keywords:**

Lung adenocarcinoma  
WGCNA  
ceRNA  
Biomarker  
Signature

## ABSTRACT

The ceRNA network, consisting of both noncoding RNA and protein-coding RNA, governs the occurrence, progression, metastasis, and infiltration of lung adenocarcinoma. Signatures comprising multiple genes can effectively determine survival stratification and prognosis of patients with lung adenocarcinoma. To explore the mechanisms of lung adenocarcinoma progression and identify potential biological targets, we carried out systematic bioinformatics analyses of the genetic profiles of lung adenocarcinoma, such as weighted gene co-expression network analysis (WGCNA), differential expression (DE) assessment, univariate and multivariate Cox proportional hazard regression models, ceRNA modulatory networks generated using the ENCORI and miRcode databases, nomogram models, ROC curve assessment, and Kaplan-Meier survival curve analysis. The ceRNA network encompassed 37 nodes, comprising 12 mRNAs, 22 lncRNAs, and three miRNAs. Simultaneously, we performed integration analysis using the 12 genes from the ceRNA network. Our findings revealed that the signature established by these 12 genes serves as an adverse element in lung adenocarcinoma, contributing to unfavorable patient prognosis. To ensure the credibility of our results, we used *in vitro* experiments for further verification. In conclusion, our study delved into the potential mechanisms underlying lung adenocarcinoma via the ceRNA regulatory network, specifically focusing on the PIF1 and has-miR-125a-5p axis. Additionally, a signature comprising 12 genes was identified as a biomarker related to the prognosis of lung adenocarcinoma.

## 1. Introduction

As cancer has the highest incidence and mortality rates worldwide, lung cancer (LC) will result in 131880 deaths among 235760 patients in 2021, which has brought a huge burden to our society [1]. About 85% LC patients suffer from non-small-cell LC (NSCLC) [2], and the primary pathological form of LC is lung adenocarcinoma (LUAD; approximately 60%) [3,4]. Although many new treatments have been developed, patient prognosis remains poor [5]. Enhanced comprehension of the underlying signaling networks used in LUAD occurrence and progression is potentially beneficial in formulating highly efficacious therapies with enhanced prognosis as well as early diagnosis [6]. Therefore, in addition to common clinicopathological features, it is critical to develop novel bio-indicators for personalized hazard stratification of LUAD and prediction of patient survival.

In the human genome, non-coding RNAs (ncRNAs) are more abundant than mRNA, and ncRNAs transcriptionally and post-

\* Corresponding author. 246 Xuefu Road, Nangang District, Harbin, Heilongjiang, China.  
E-mail address: [hrbzhouxl@163.com](mailto:hrbzhouxl@163.com) (X. Zhou).

transcriptionally modulate transcript levels [7]. ncRNAs mainly function via miRNA response elements (MREs); they interact with MREs to form a competitive endogenous RNA (ceRNA) axis [8]. The ceRNA hypothesis states that ncRNAs absorb and bind miRNAs like sponges and regulate gene expression by inhibiting their ability to bind to mRNA [9]. Previous studies have reported that ceRNA networks regulate occurrence, metastasis, and outcomes of various cancers, namely LINC00665 and miR-98 networks in LC [10], OSTN-AS1 in breast cancer [11], FAS and HOTAIR in colorectal cancer [12,13], and LINC00301 in NSCLC [14]. However, their relationship to lung adenocarcinomas remains unclear. Therefore, we need to clarify their relationship to analyze the underlying signaling pathways that regulate the occurrence, development, and metastasis of lung adenocarcinoma.

In recent years, many studies have reported the significance of survival stratification and prognosis prediction in patients with LUAD based on transcriptomic and genomic data [15–17]. However, owing to limited sample quantity and lack of independent verification, the signatures reported in these studies have not yet been put into clinical practice [18,19]. Today, large public databases containing sufficient gene profile information, such as The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO), provide the possibility of developing more reliable prognostic signatures for LC.

Herein, we employed GEO and TCGA gene profile datasets to develop a ceRNA modulatory axis that potentially impacts lung adenocarcinoma patient prognosis. We further used 12 genes in the ceRNA modulatory axis to construct prognosis-related signatures to predict LUAD. The ceRNA network and signature enable clinicians to evaluate the prognosis of patients with LUAD, providing hope for individualized treatment interventions.

## 2. Materials and methods

### 2.1. Patients

All tumor and normal tissue specimens were acquired from LUAD patients at the Second Affiliated Hospital of Harbin Medical University. LUAD was confirmed in all participants by pathologists and clinicians. Following tumor surgical resection, paired tumor and adjoining normal tissues were harvested from all participants. Additionally, corresponding clinical data were also obtained from 30 patients. This study was conducted with permission from the ethical committee of the participating hospital.

### 2.2. Data downloading and processing

All necessary data were acquired from a freely available online database. LUAD gene expression information was retrieved from the GEO database. The following data were included for analysis: 1. A minimum of 20 tumor and normal tissue specimens, and 2. corresponding clinical information of patients; 3. The data has not been processed. The final datasets were GSE87340 (27 patients and 54 samples) and GSE68571 (86 patients and 96 samples). Additionally, the count, Transcript per Kilobase per Million (TPM) gene expression and clinical information of LUAD were acquired from TCGA database, and 598 RNA-seq data, 522 clinical data, 19895 mRNAs, 16784 lncRNAs, and 2197 miRNAs were obtained.

Following retrieval from the GEO database, the data was adjusted and normalized using “limma” (R 3.54.2), and data was converted to log2. Data obtained from TCGA database received gene ID annotation via the GENCODE database, and only the gene with the highest expression level was reserved for genes with duplicate gene names.

### 2.3. Screening of differentially expressed (DE) mRNA, lncRNA and miRNA

DE LUAD mRNA, lncRNA, and miRNA in TCGA were assessed via “limma” (R 3.54.2). To achieve P values, we employed false discovery rate (FDR) to adjust multi-test statistical significances. In this study, the cut-off values for significantly DE genes were  $|\log_2FC| > 1$  and  $FDR < 0.05$ .

### 2.4. Analysis of weighted gene Co-expression network (WGCNA)

WGCNA [20] is a robust biological technique that utilizes relationships between gene sets and phenotypes to identify highly related gene modules, candidate biomarker genes, and potential therapeutic targets. Herein, TCGA LUAD gene expression profiles were chosen to construct a weighted gene co-expression network (WGCN). Samples and genes were filtered and “WGCNA” (R 1.71) was employed for Pearson’s correlation computation between all gene pairs in chosen samples, and to generate the adjacency matrix. Subsequently,  $\beta = 10$  (scale-free  $R^2 = 0.90$ ) was employed as a soft cut-off to generate a scale-free axis. To further screen for functional modules within the WGCN, topological overlap measure (TOM) was computed using an adjacency matrix. Based on the TOM values, the dynamic tree-building method was used to establish gene modules and screen module eigengenes (MEs). MEs represented the genetic profiles within modules.

### 2.5. Enrichment analysis and single sample gene set enrichment assessment

“ClusterProfiler” (R 4.4.4) was employed for GO functional enrichment and KEGG pathway enrichment assessments to predict the physiological roles of target genes. We conducted single sample gene set enrichment analysis (ssGSEA) using “gsva” in “GSVA” (R 1.44.3), and calculate the standardized enrichment score of the gene set in each patient.

2.6. Generation of ceRNA modulatory axis

The miRcode database (<http://www.mircode.org/index.php>) with different lncRNA interactions of miRNAs and finding TCGA difference expression in LUAD data lncRNA overlapping lncRNA, the final lncRNA-miRNA relationship pair was obtained. The ENCORI (<https://starbase.sysu.edu.cn/index.php>) [21] database was employed to estimate the miRNAs of target genes and establish mRNA-miRNA relationship pairs.

Overlapping miRNAs shared between lncRNA-miRNA and mRNA-miRNA association pairs with DE miRNAs were used to confirm the final mRNA, lncRNA, and miRNA regulatory axis.

2.7. Cell lines and culture

Human LUAD cell lines (HCCB27, HCC-1438, NCI-H460, A549, SW 1573, COR-L23, and NCI-H810) and a normal lung epithelial cell line (BEAS-2B) were acquired from American Type Culture Collection (ATCC, Manassas, VA, USA). The LUAD cell lines were grown in Ham's F-12K medium (Jiangsu Kaiji Biotechnology Co., Ltd., Tokyo, Japan) or RPMI-1640 medium (Gibco Laboratories, Grand Island, NY, USA), and BEAS-2B cells in DMEM Medium (Gibco Laboratories).

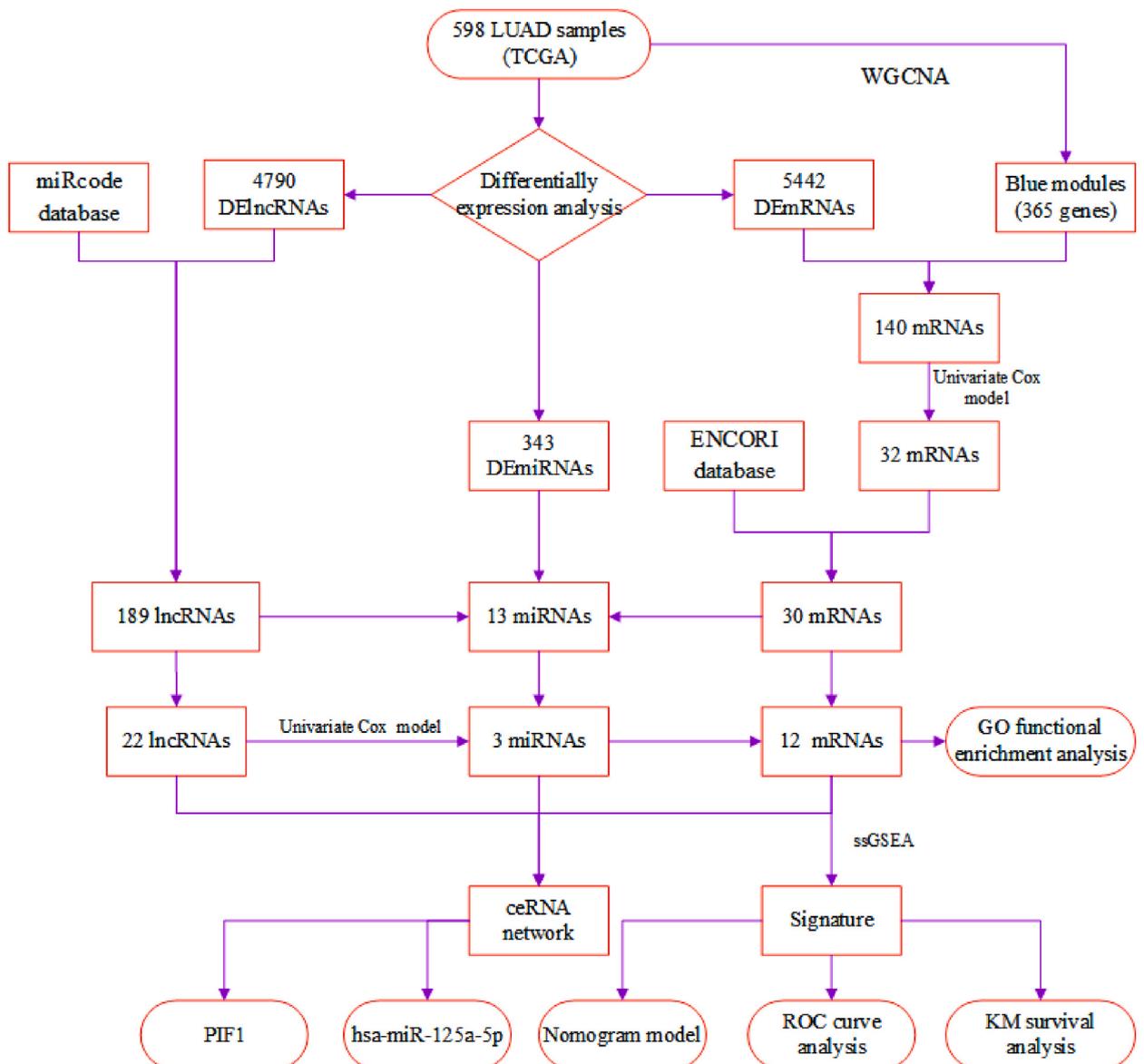


Fig. 1. The study design.

## 2.8. Total RNA Extraction and Quantitative Real-Time PCR (qRT-PCR)

Total RNA isolation from 30 matched LUAD and non-cancerous tissues employed RNA-easy Isolation Reagent (No. RC105-12, Vazyme, China). Thereafter, qRT-PCR was conducted using HiScript III first-strand immunohistochemistry (IHC) and ChamQTM Universal SYBR® qPCR Master mix (No. Q635-04, Vazyme), as per kit directions.

## 2.9. Immunohistochemistry

IHC was performed as described previously. The anti-SST rabbit polyclonal antibody (cat. No. ab187326; 1:100 dilution; Abcam, UK) was combined with tissue section, prior to an overnight incubation at 4 °C, followed by PBS wash, and a 30-min incubation in biotinylated secondary antibody (Cat. No. 111-024-005, 1:1500 dilution; Jackson ImmunoResearch, USA) at room temperature (RT), prior to a 5-min exposure to diaminobenzidine at R. To generate the negative control, the aforementioned steps were taken PBS and not primary antibody.

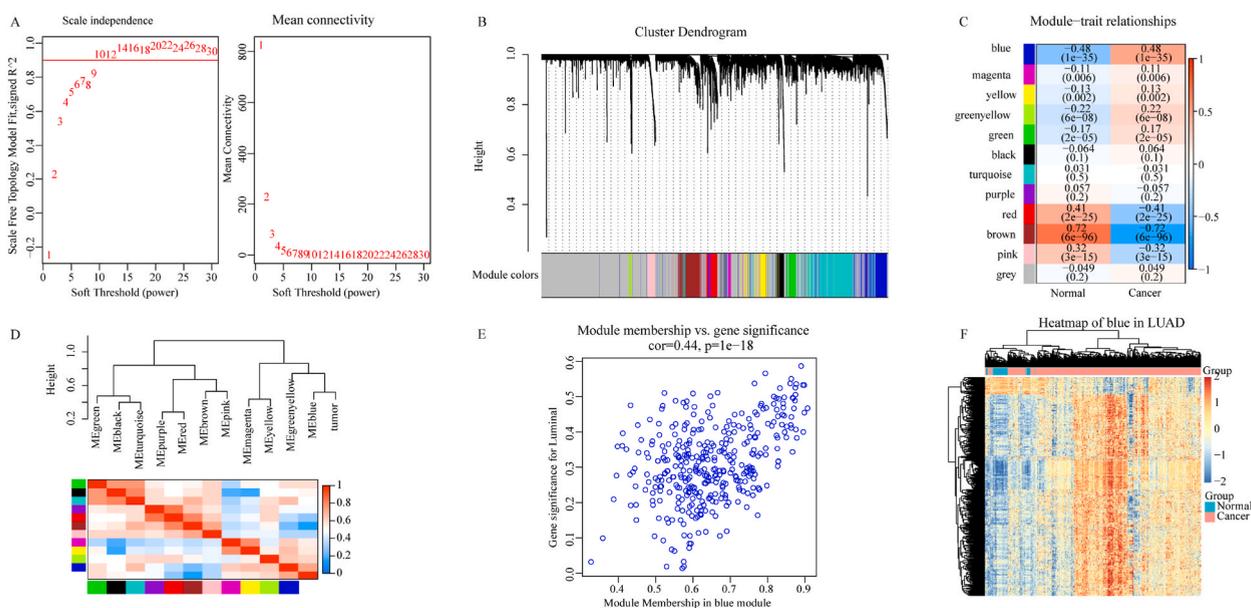
## 2.10. Western blotting assay and cell counting kit-8 (CCK-8) assay

Using Western blot, we determined target protein expression levels. The blots were incubated with an antibody against PIF1 (Cell Signaling Technology, #4903S) overnight at 4 °C. The CCK-8 assay (Beyotime, Shanghai, China) was employed for cell proliferation assessment, according to a specific protocol. The uncropped versions were provided in Supplementary Material.

## 2.11. Statistical analysis

"Surv\_cutpoint" in "survminer" (R 0.4.9) calculated the optimal truncation value, and group patients Kaplan-Meier survival curves were plotted between various patient groups using "survminer" and "survival" (R 3.3-1). Significant differences were determined using log-rank test. "pROC" (R 1.18.0) was employed for ROC curve assessment, "rms" (R 6.3-0) for nomogram model and calibration curve generation, and "ggplot2" (R 3.4.3) for drawing.

In this study, all data analyses utilized R 4.2.0 and the web-based analytical tool "Xiantao Academic." Unless specified, all statistical analyses were bilateral and significance was determined at  $P < 0.05$ .



**Fig. 2.** Identification of tumor-related genes. A, network topology scale-free fitting index acquired from soft cut-off analysis; B, hierarchical cluster assessment for co-expression clusters detection using a corresponding color code. Individual color indicates a single module in the WGCNA-generated gene co-expression network; C, correlation between genes in different modules and tumor samples or normal samples; D, characteristic gene connectivity: red and blue represent positive and negative associations, respectively; E, assessment of gene relevance and module membership in the blue module; F, gene heatmap in the blue module. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

### 3. Results

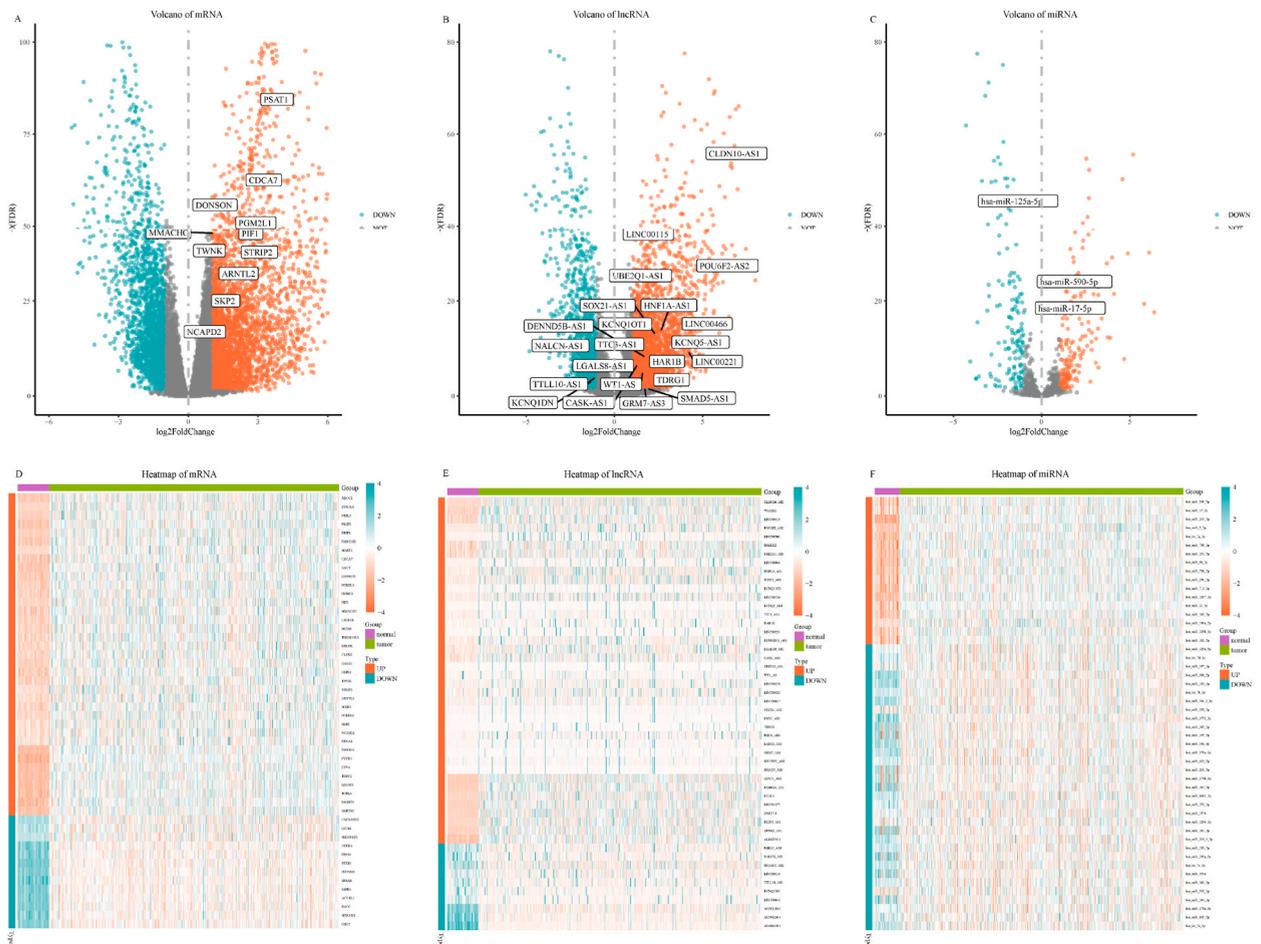
#### 3.1. Construction of WGCN

A schematic of our research design is shown in Fig. 1. To detect the functional modules of LUAD, we used the LUAD data from TCGA for WGCN. According to the data filtered by median absolute deviation (MAD), we screened 4974 genes whose expression levels were in the top 25% of MAD to construct the WGCN network. First, an appropriate soft threshold was selected according to  $R^2 = 0.9$ , and the relation matrix was established according to  $\beta = 10$  (Fig. 2a). Subsequently, the relation matrix was converted to an adjacency matrix, prior to the usage of power-exponential weighting to generate a scale-free axis. Finally, using the adjacency matrix, a TOM matrix was developed to compute the TOM dissimilarity between genes (disTOM), and a gene feature module was established. Twelve gene feature modules were obtained (Fig. 2b–d).

Correlation analysis of the gene modules with clinical phenotypes was performed, and the correlation coefficients of each module with phenotypes were calculated (Fig. 2e). We found that the blue module, containing 365 genes, had the highest correlation with the clinical phenotype compared to the other modules ( $R = 0.44$ , Fig. 2f). Hence, we chose the blue module for further assessment.

#### 3.2. DE analysis at transcriptome level

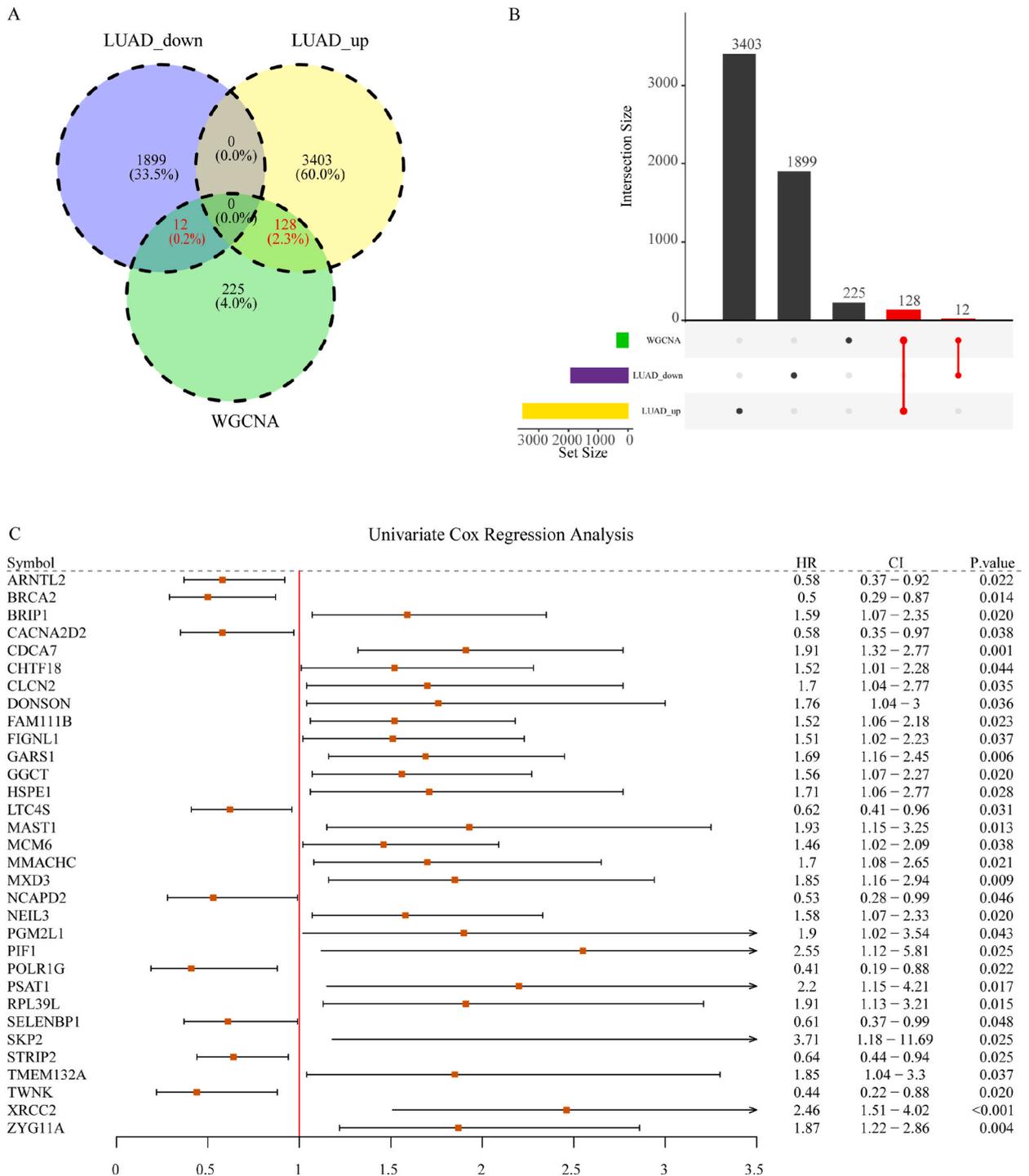
We used “DESeq2” (R) to evaluate the DE mRNA data of normal lung tissue samples and LUAD samples in TCGA. Based on the threshold criteria, we screened 1911 strongly DE downregulated mRNA and 3531 strongly DE upregulated mRNA (Fig. 3a). We mapped some DE mRNA and found that their mRNA expression levels were significantly different between normal and tumor samples (Fig. 3b).



**Fig. 3.** Identify differentially expressed (DE) genes. A-C, respectively, are the volcanic maps of DE mRNA, lncRNA, and miRNA in lung adenocarcinoma, red represents the strongly elevated gene, blue represents strongly diminished gene; D-F, are heatmaps of 100 significantly DE mRNA, lncRNA, and miRNA, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Simultaneously, the DE lncRNAs in normal lung tissue and LUAD samples in TCGA was analyzed. We found that 1100 significantly DE downregulated lncRNAs and 3690 significantly differentially expressed upregulated lncRNAs (Fig. 3c). The profiles of some DE lncRNAs in normal and tumor specimens are shown in Fig. 3d.

In addition, DE miRNA analysis in LUAD and normal tissue specimens from TCGA confirmed 158 significantly downregulated



**Fig. 4.** Identify tumor genes associated with prognosis. A, the Venn diagram shows the intersection of significantly elevated and diminished genes in lung adenocarcinoma and tumor-associated genes in the blue module; B, Upset maps showed the intersection of strongly elevated genes and strongly diminished genes with tumor-associated genes in blue modules. C, forest maps show risk ratios for tumor-related genes. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

miRNAs and 185 significantly upregulated miRNAs (Fig. 3e). We visualized the expression levels of several DE miRNAs (Fig. 3f).

### 3.3. Confirming prognosis related mRNA based on Cox proportional hazard regression model

By crossing 5442 significantly DE genes and 365 tumor-related genes in the blue module, we identified 12 significantly DE downregulated genes and 128 significantly DE upregulated genes (Fig. 4a and b).

Among tumor related DE genes, to further identify the genes associated with LUAD prognosis, we performed a univariate Cox proportional hazard regression model. Based on our analysis, the hazard ratios (HR) of the 32 genes were highly significant (Fig. 4c). Majority gene HRs were >1, suggesting the possibility of these genes as candidate prognostic indicators which enhance tumor development and progression.

### 3.4. Exploring the axis of gene regulation

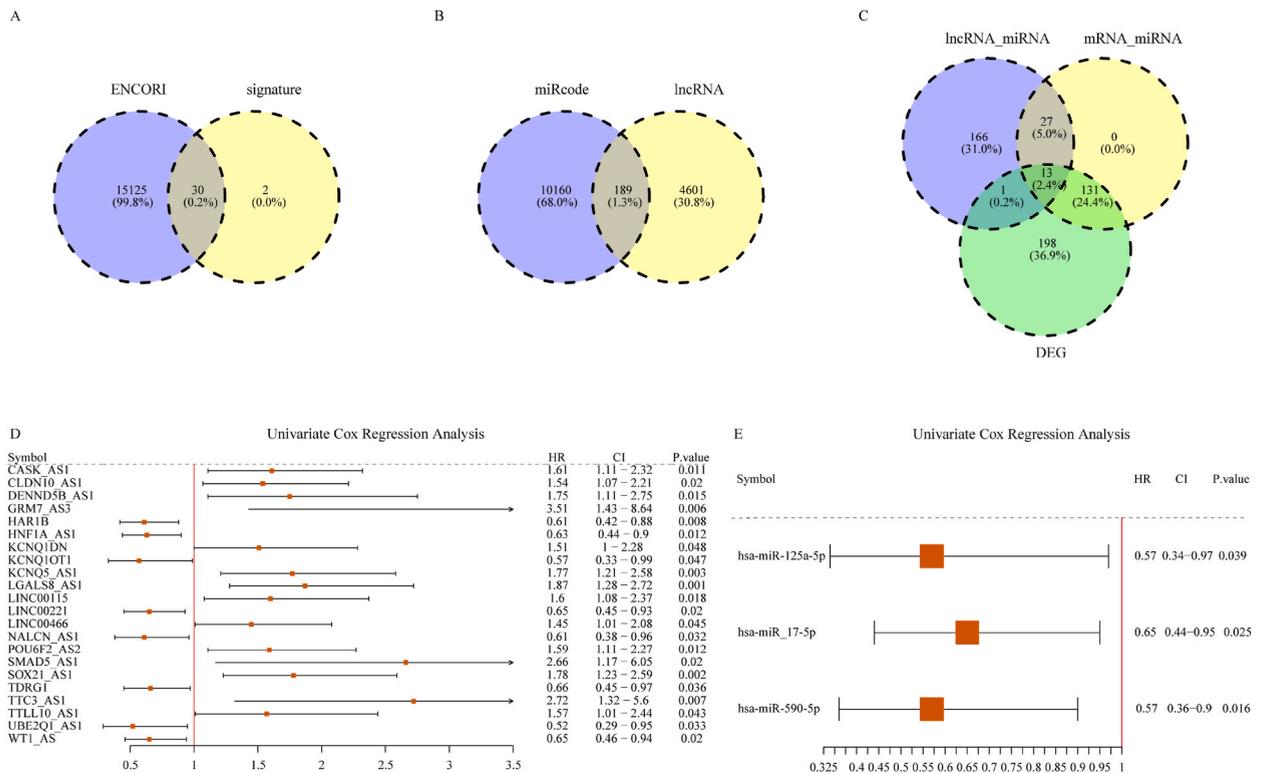
To further analyze the mechanisms of the 32 genes in the ceRNA regulatory network, we used multiple databases to predict the miRNAs and lncRNAs associated with the 32 genes. Through the intersection of 32 genes and mRNA in the ENCORI database, it was found that 30 genes are related to miRNA in the ENCORI database (Fig. 5a); Simultaneously, 4790 DE lncRNAs were predicted by miRNAs in the miRcode database, and 189 lncRNAs were found to be associated with miRNAs through the intersection of lncRNAs (Fig. 5b).

Finally, we found that 13 DE miRNAs may interact with mRNA and lncRNAs by correlating the miRNAs that existed together in the two databases and crossing them with the significantly DE miRNAs in TCGA (Fig. 5c).

The univariate risk model suggested that 22 of the 189 lncRNAs are intricately linked to LUAD prognosis (Fig. 5d); Three of the 13 miRNAs were related to LUAD prognosis. Interestingly, the univariate risk model showed that all three miRNAs were beneficial for patient prognosis (Fig. 5e).

### 3.5. Functional enrichment analysis of ceRNA network nodes

After selecting 22 lncRNAs and 3 miRNAs using the univariate Cox proportional hazard model, we found that only 12 of the 30



**Fig. 5.** Identify the nodes of the ceRNA axis. A, venn diagram shows the intersection of mRNA in the ENCORI database and 32 genes in lung adenocarcinoma; B, venn diagram revealing the intersection of lncRNA in miRcode database and lncRNA differentially expressed (DE) in lung adenocarcinoma; C, venn diagram revealing the intersection of lncRNA-miRNA axis, mRNA-miRNA, and DE miRNAs in lung adenocarcinoma; D, forest maps of 22 lncRNAs single-factor Cox proportional risk models; E, forest maps of three miRNA single-factor Cox proportional risk models.

genes were associated with the 22 lncRNAs and 3 miRNAs.

To elucidate the physiological roles of tumor-related genes, we conducted GO functional enrichment analysis of the 12 tumor-related genes. In BP, we found that these genes are related to the cell proliferation process (Fig. 6a), such as DNA replication, DNA conformity change, DNA duplex unwinding, and G2/M transition of mitotic cell cycle, including nuclear chromosome, mitochondrial nucleus, central region, and replication fork (Fig. 6b). Enrichment analysis results related to MF are also displayed (Fig. 6c), such as DNA helix activity, ATP-dependent activity, acting on DNA, cobalamin binding, and oligopeptide binding. In the GO enrichment analysis circle, we observed that PIF1 existed at the BP, CC, and MF levels (Fig. 6d-f), suggesting that this gene may strongly contribute to LUAD occurrence, development, and prognosis.

3.6. The 12 genes-based ceRNA modulatory axis

We generated a ceRNA modulatory axis using the 12 mRNA, 22 lncRNAs, and three miRNAs (Fig. 7a). The network showed their regulatory relationships in detail, including 17 mRNA-miRNA relationship pairs and 29 lncRNA-miRNA regulatory axes. In addition, the nodes of the network included 19 significantly upregulated lncRNAs, three significantly downregulated lncRNAs, one significantly downregulated miRNA, two significantly upregulated miRNAs, and 12 significantly upregulated genes (Fig. 7b-d).

Among the 12 genes, PIF1 simultaneously enriched three levels of biological function in LUAD (Fig. 6d-f). PIF1 is also highly expressed in LUAD tissue, and the Kaplan-Meier survival curve indicates that PIF1 is a risk factor (Fig. 7e); the area under the ROC curve (AUC) of this gene was 0.91 (AUC = 0.91, Fig. 7f), indicating that PIF1 has good predictive potential for the prognosis of patients with LUAD. Therefore, we investigated the mechanisms regulating the PIF1 axis. Among the three miRNAs that have a regulatory relationship with PIF1, only has-miR-125a-5p exhibited excellent predictive impact on LUAD patient prognosis (AUC = 0.90, Fig. 7g and h), and was expressed at low levels in patients with LUAD. PIF1 and has-miR-125a-5p have opposite regulatory effects on LUAD, suggesting that has-miR-125a-5p may modulate LUAD development and progression by inhibiting the function of PIF1.

3.7. Construction of the prognosis related signature

To explore the comprehensive effect of these 12 genes on the prognosis of LUAD, we further analyzed these 12 genes as signatures. Based on "GSVA," (R) we conducted ssGSEA on the signature, and calculate the normalized enrichment scores (NESs) of the signature in each patient.

First, we built a nomogram model of the NES-based signature (Fig. 8a). Unsurprisingly, the NES contents of signature was negatively linked to the patient survival. Furthermore, the best NES-based cutoff value was computed, and patients were participated into elevated- and reduced-scoring cohorts. Kaplan-Meier survival curve assessment revealed that patients with elevated scores exhibited worse overall survival (OS), indicating that enhanced NESs might be a possible risk factor for patients (Fig. 8b). To evaluate signature accuracy in estimating patient outcome, we conducted ROC curve analysis. Our findings suggested that the AUC was 0.61, demonstrating that the signature accurately predicted patient outcome (AUC = 0.61, Fig. 8c).

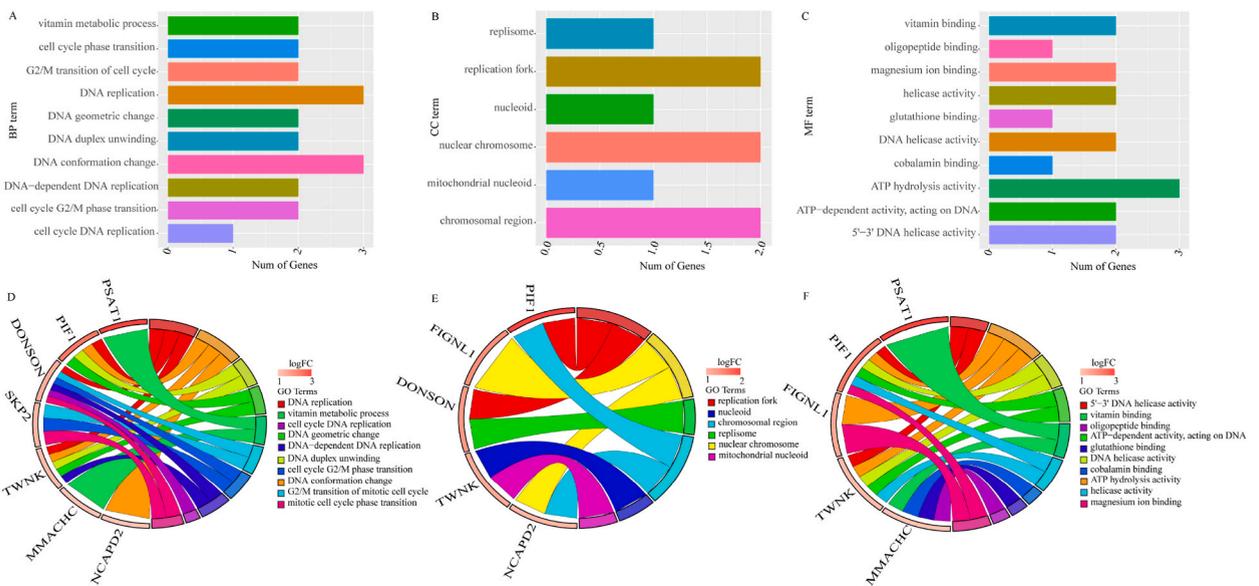
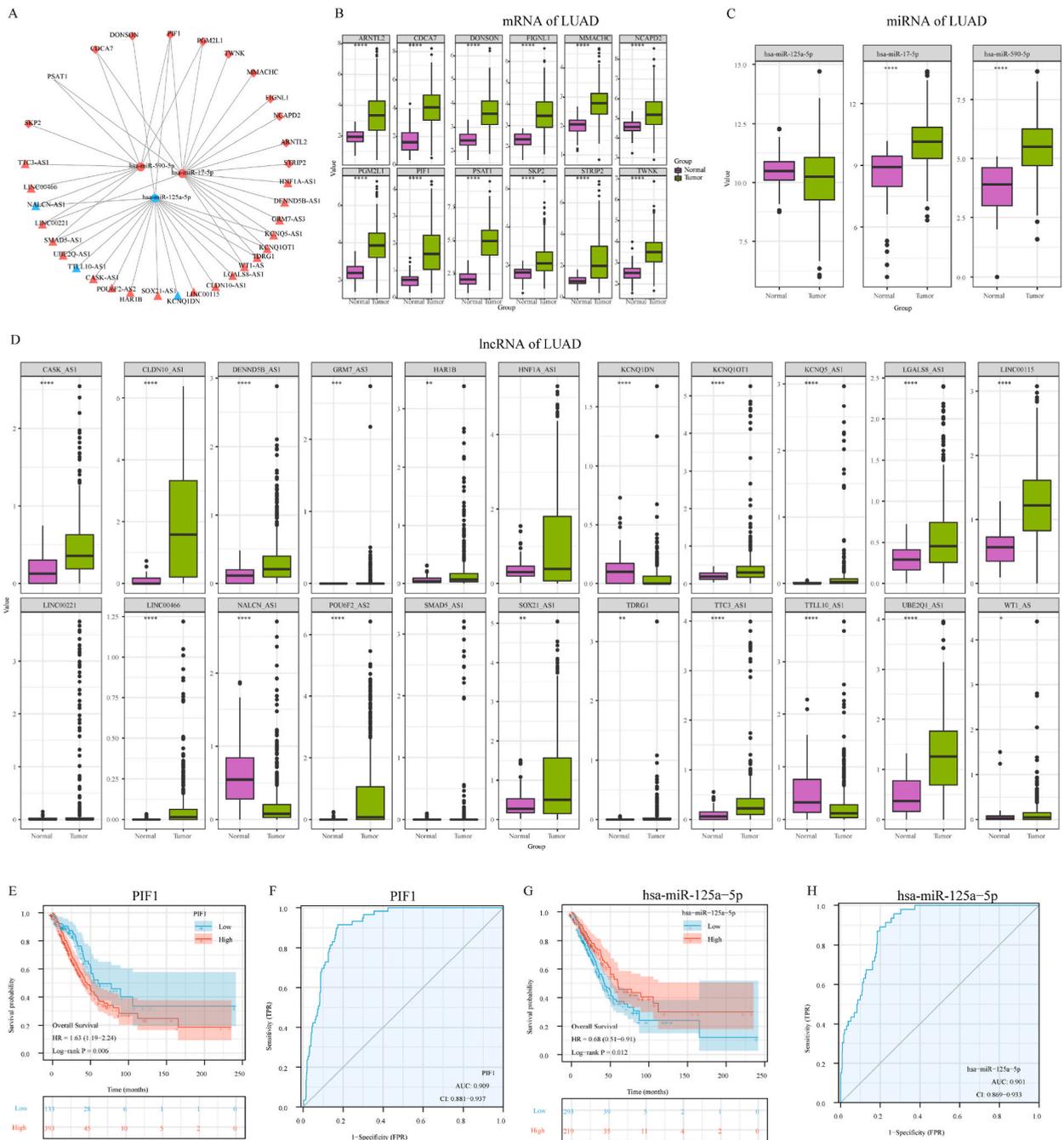


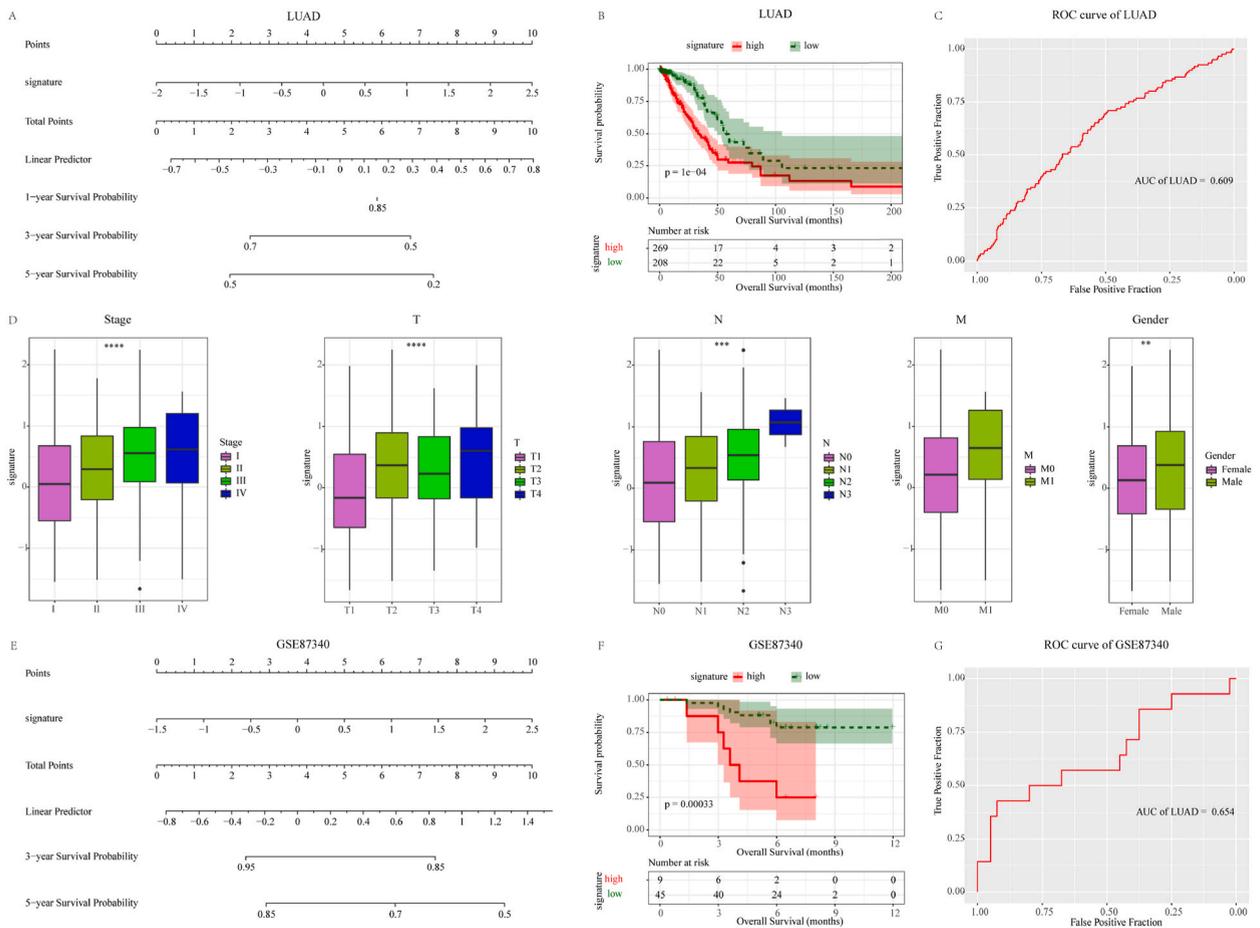
Fig. 6. GO enrichment assessment of tumor-associated genes. A-C, histograms derived from enrichment analysis of 12 tumor-associated genes, including BP, CC, and MF terms; D-F, circles of enrichment assessment of 12 tumor-associated genes including BP, CC, and MF terms. Left half of the circle represents gene name, and is categorized via logFC from bottom to top, the right half of the circle represents term name, and the line between the gene and term represents genes that exist on the term.



**Fig. 7.** Assessment of the ceRNA modulatory axis. A, ceRNA network integrating data from mRNA, lncRNA, and miRNA, the diamond indicates mRNA, triangle indicates lncRNA, ellipse indicates miRNA, red indicates enhanced RNA, and blue indicates diminished RNA; B, the boxplot shows the expression difference of 12 mRNA in the ceRNA axis; C, boxplot revealing the difference of three miRNA expressions in the ceRNA axis; D, the box plot depicting the expression difference of 22 lncRNAs in the ceRNA network; E, PIF1 survival curve analysis; F, ROC curve analysis of PIF1 ( $p = 0.0045$ ); G, has-miR-125a-5p survival curve analysis; H, ROC curve analysis of has-miR-125a-5p ( $p = 0.0017$ ). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Additionally, we assessed the association between NESs and clinicopathological features (Fig. 8d). Male patients had higher NES levels. There were also significant differences in NESs among the different pathological stages of the tumor. The NESs of stage III/IV disease patients were elevated, compared to those of stage I/II disease patients, which further confirmed the poor prognosis of stage III/IV disease patients. T, N, and M stages showed similar results.

Finally, we analyzed the signature in the GSE87340 and GSE68571 datasets to verify its universal applicability. First, we calculated



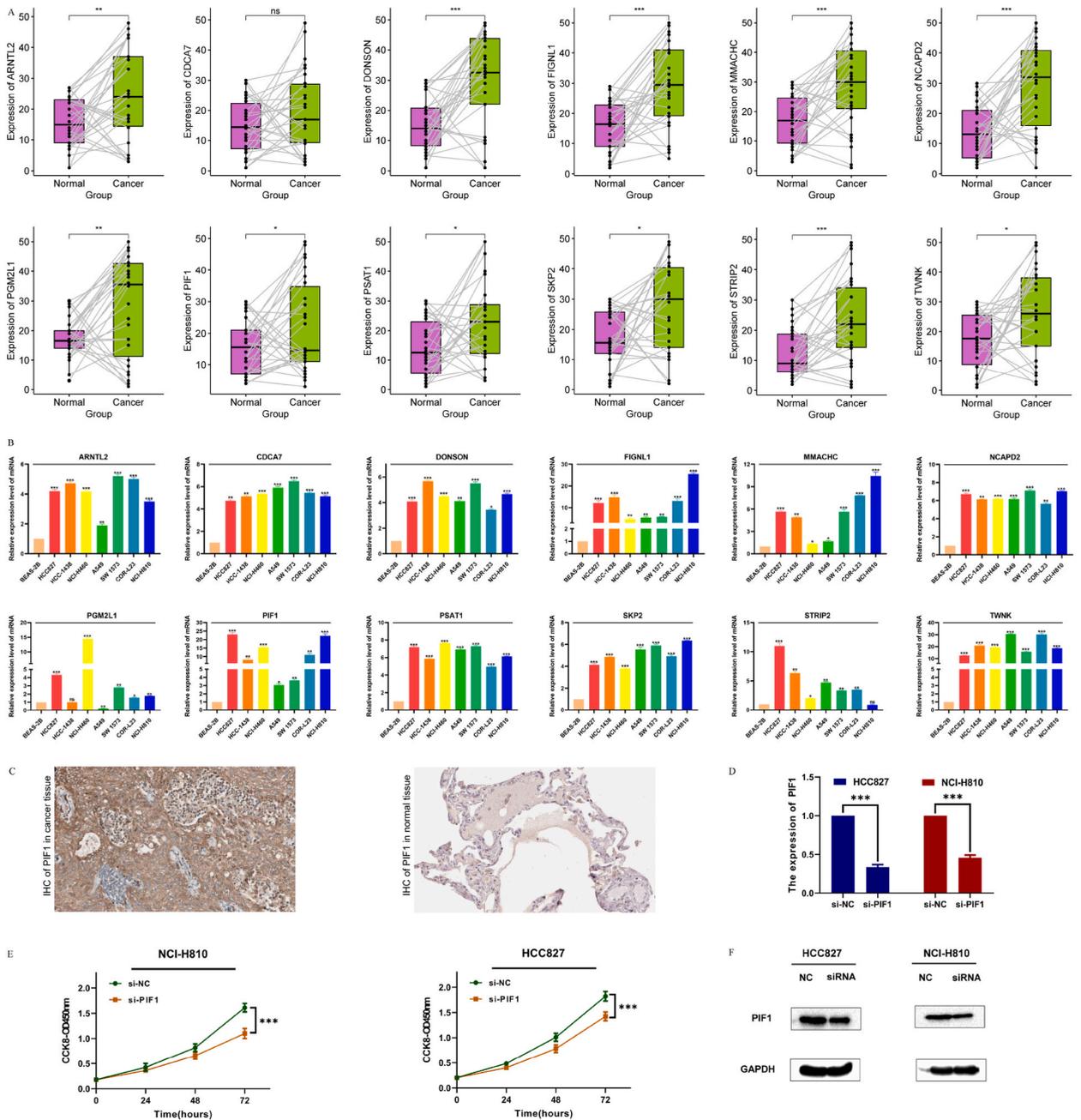
**Fig. 8.** Establishment and verification of prognostication related signatures. A, calculate the standardized signature enrichment score and generate a nomogram model; B, we separated patients into enhanced and reduced score cohorts, based on the optimal threshold value for the standardized enrichment score. Then, performed survival analysis; C, signature ROC curve analysis ( $p = 0.0006$ ); D, the boxplot shows the differences between standardized enrichment scores of different pathological stages, T stages, N stages, M stages, and gender; E-G, respectively, are the nomogram model, survival curve and the signature ROC curve in the data set GSE87340 ( $p = 0.0326$ ).

the NESs for each patient in the cohort and constructed a nomogram. It can be seen that the NESs of the signature is negatively related to patient survivability (Fig. 8e). Kaplan-Meier survival curve assessment revealed that patients with elevated scores exhibited worse OS duration, which may be a potential risk factor (Fig. 8f, Supplementary fig. 1A). To elucidate signature accuracy in estimating patient outcome, we conducted ROC curve analysis. Our findings suggested that the AUC is 0.65 and 0.64, suggesting that the signature has an excellent predictive impact on patient outcomes (Fig. 8g, Supplementary fig. 1B). In conclusion, the analysis of the GSE87340 and GSE68571 datasets showed that this signature has stable predictive efficacy and can better predict the LUAD patient prognosis.

### 3.8. Verification of signature-associated gene expressions in lung adenocarcinoma samples and cells

To further confirm signature accuracy, we obtained clinical data and corresponding tissue samples from 30 patients. Patients were stratified according to tumor location, and patient demographics is presented in Supplementary table 1. We examined the 12 signature gene expressions in clinical samples from LUAD patients using qRT-PCR (Fig. 9a). We revealed that the 12 gene expressions were higher in LUAD tissues. The 12 gene expressions were further validated in multiple LUAD cell lines and normal lung epithelial cells (Fig. 9b).

In previous studies, we found that PIF1 may be strong modulators of lung adenocarcinoma. Immunohistochemical analysis of LUAD and normal tissue samples showed that PIF1 was highly expressed in the tumor tissues. (Fig. 9c). Next, we explored the functional role of PIF1 in HCC827 and NCI-H810 cell lines. As depicted in Fig. 9d, the PIF1 expression was significantly reduced after transfection of si-PIF1 in HCC827 and NCI-H810 cell lines. The CCK-8 assay revealed that compared to the control group, the growth rate of the HCC827 and NCI-H810 cell lines transfected with si-PIF1 significantly decreased (Fig. 9e). HCC827 and NCI-H810 cells were chosen for additional assessment because PIF1 contents were highest among these cells. The reduction efficiency was assessed via knocking down PIF1 with siRNA and performing western blotting (Fig. 9f). Similar to prior data, the PIF1 contents decreased in HCC827 and



**Fig. 9.** The 12 gene expressions in human tissue specimens and cell lines. A, qRT-PCR-based evaluation of the 12 gene expressions in lung adenocarcinoma patients. B, RT-qPCR detection of mRNA expression of 12 genes in lung adenocarcinoma cell lines. C, immunohistochemistry-based evaluation of PIF1 expression in lung adenocarcinoma patients. D, PIF1 was downregulated in cell lines using siRNAs. E, the proliferation of cells transfected with siRNA against PIF1 were measured using CCK8 assays. F, western blot-based verification of PIF1 expression in cell lines, the uncropped versions were provided in supplementary material.

NCI-H810 cells transfected with siRNA.

#### 4. Discussion

Similar to majority tumors, LUAD development and progression are multi-gene, multi-variable, multi-stage processes [22,23]. Understanding the associated network underlying the occurrence and development of LUAD is essential to the establishment of efficacious intervention methods and providing a basis for its early diagnosis [24]. Herein, we employed WGCNA to identify

bioindicators related to the pathogenesis of LUAD. Using the LUAD genetic profile in TCGA, we identified 12 gene modules in the WGCNA network, among which the blue module was strongly linked to LUAD ( $p < 0.001$ ). Simultaneously, we conducted DE assessments involving mRNA, lncRNA, and miRNA expression profiles of patients with LUAD in TCGA to identify DE genes. We next constructed a univariate Cox proportional risk model based on DE tumor-related genes and identified genes related to prognosis. In addition, the ceRNA axis was generated through the association analysis of DElncRNA and DEMiRNA in the ERCORI and miRcode databases. The network had 37 nodes, including 12 mRNA (PIF1, DONSON, MMACHC, TWNK, PGM2L1, NCAPD2, ARNTL2, CDCA7, FIGNL1, STRIP2, PSAT1 and SKP2), 22 lncRNAs (SMAD5-AS1, WT1-AS, LINC00221, TDRG1, TTLL10-AS1, CLDN10-AS1, LGALS8-AS1, LINC00466, LINC00115, GRM7-AS3, SOX21-AS1, TTC3-AS1, KCNQ5-AS1, UBE2Q1-AS1, HAR1B, NALCN-AS1, CAS-K-AS1, POU6F2-AS2, KCNQ1DN, HNF1A-AS1, DENND5B-AS1, and KCNQ1OT1), and 3 miRNAs (miR-590 5p, has-miR-125a-5p and has-miR-17-5p).

Functional enrichment analyses of 12 genes suggested that PIF1 is critical for LUAD development and progression and participates in multiple carcinogenesis-related pathways, such as DNA replication, DNA duplex unwinding, ATP-dependent activity, and DNA and ATP hydrogenation activity. In our study, PIF1 was highly expressed in LUAD. The prognosis correlation analysis results indicated that PIF1 is a risk factor for LUAD, which is not conducive to patient prognosis.

The PIF1 protein family was first discovered in *Saccharomyces cerevisiae*; its main members regulates mitochondrial DNA recombination and gene stability [25]. Subsequently, PIF1 was reported to function as a helicase in nuclear DNA replication [26,27]. Recent evidences revealed that PIF1 is also a major contributor to cancer; for example, PIF1 influences cervical cancer cell proliferation and apoptosis by affecting TERT [28]. In addition, PIF1 regulates LUAD development and progression through a ceRNA modulatory axis. Ting et al. reported that CircNEIL3 upregulates PIF1 expression by inhibiting miR-1184, thereby mediating focal death and affecting radiotherapy in LUAD [29]. Previous studies reported that miR-125a-5 regulates the progression of LUAD via various ceRNA-related axes. Gao et al. found that the oncogene P35B is regulated by microRNA-125a-5p in LC [30], and miR-125a-5p affects the carcinogenic effects of NF- $\kappa$ B signaling pathway in LUAD by inhibiting the upregulation of TMPRSS4.

In this study, we established a ceRNA regulatory axis for PIF1 and has-miR-125a-5p, which regulate LUAD by mediating the function of lncRNAs. Based on the modulatory role of PIF1 and has-miR-125a-5p in LUAD, we suggest that the PIF1 and has-miR-125a-5p regulatory network strictly controls LUAD progression, invasion, and metastasis. Finally, we constructed a signature related to LUAD prognosis using the 12 genes. The results of systematic analysis and verification showed that this signature was superior at predicting patient outcome. The establishment of this signature provides clues for the prognosis, targeted molecular therapy, and clinical diagnosis of LUAD.

Our research has limitations. First, our assessment utilized previously published information. Thus, our retrospective analysis necessitates and prospective evaluation and validation. Second, to predict the LUAD prognosis using the 12-gene signature as a predictive factor, we must be cautious of potential overfitting. To ensure the accuracy of this signature in prognosis prediction, it will be crucial to address and mitigate any overfitting issues in our future research; Third, we have thus far conducted only preliminary experimental verification. Moving forward, it is imperative to devise a more robust and scientifically rigorous plan to validate the analysis results, which will help clarify the potential molecular mechanisms underlying LUAD.

## 5. Conclusion

In summary, our study identified 12 genes associated to LUAD prognosis using comprehensive bioinformatic analyses. This signature of 12 genes is a promising biomarker candidate for assessing the prognosis of patients with LUAD. Furthermore, we established a ceRNA regulatory network centered on these 12 genes. Notably, our findings highlight the significant involvement of the PIF1 and has-miR-125a-5p regulatory axis in the development, progression, and prognosis of lung adenocarcinoma.

## Ethics statement

This study was conducted in accordance with the principles of the Declaration of Helsinki. The study protocol was reviewed and approved by the Ethics Committee of the Second Affiliated Hospital of Harbin Medical University (2013-56). All the patients provided written informed consent to participate in this study.

## Funding statement

Not applicable.

## Data availability statement

The datasets analyzed in this study are available from the Gene Expression Omnibus (GEO) and Cancer Genome Atlas (TCGA) repository. The GEO web link is <https://www.ncbi.nlm.nih.gov/geo/> and the GEO datasets include GSE87340 and GSE68571. The TCGA web link is <https://portal.gdc.cancer.gov/repository>. The data generated in this study and code can be found in the Supplementary Materials.

## CRediT authorship contribution statement

**Hong Wei:** Data curation, Conceptualization. **Fei Teng:** Formal analysis. **XiaoLei Wang:** Validation. **XiuJuan Hou:** Visualization. **HongBo Wang:** Resources. **Hong Wang:** Methodology. **Hui Sun:** Formal analysis. **XianLi Zhou:** Writing – review & editing, Writing – original draft.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgment

The authors thank TCGA, PubMed, and GEO for data access.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e28084>.

## References

- [1] R.L. Siegel, K.D. Miller, H.E. Fuchs, A. Jemal, Cancer statistics, 2022, *CA Cancer J Clin* 72 (1) (2022) 7–33.
- [2] J. Remon, L.E. Hendriks, C. Cabrera, N. Reguart, B. Besse, Immunotherapy for oncogenic-driven advanced non-small cell lung cancers: is the time ripe for a change? *Cancer Treat Rev* 71 (2018) 47–58.
- [3] M. Behera, T.K. Owonikoko, A.A. Gal, C.E. Steuer, S. Kim, R.N. Pillai, F.R. Khuri, S.S. Ramalingam, G.L. Sica, Lung adenocarcinoma staging using the 2011 IASLC/ATS/ERS classification: a pooled analysis of adenocarcinoma in situ and minimally invasive adenocarcinoma, *Clin. Lung Cancer* 17 (5) (2016) e57–e64.
- [4] N. Cancer Genome Atlas Research, Comprehensive molecular profiling of lung adenocarcinoma, *Nature* 511 (7511) (2014) 543–550.
- [5] C.C. Hao, C.Y. Xu, X.Y. Zhao, J.N. Luo, G. Wang, L.H. Zhao, X. Ge, X.F. Ge, Up-regulation of VANGLI1 by IGF2BP3 and miR-29b-3p attenuates the detrimental effect of irradiation on lung adenocarcinoma, *J. Exp. Clin. Cancer Res.* 39 (1) (2020) 256.
- [6] P. Wu, Y. Zheng, Y. Wang, Y. Wang, N. Liang, Development and validation of a robust immune-related prognostic signature in early-stage lung adenocarcinoma, *J. Transl. Med.* 18 (1) (2020) 380.
- [7] E. Anastasiadou, L.S. Jacob, F.J. Slack, Non-coding RNA networks in cancer, *Nat. Rev. Cancer* 18 (1) (2018) 5–18.
- [8] F.A. Karreth, P.P. Pandolfi, ceRNA cross-talk in cancer: when ce-bling rivalries go awry, *Cancer Discov.* 3 (10) (2013) 1113–1121.
- [9] L. Salmena, L. Poliseno, Y. Tay, L. Kats, P.P. Pandolfi, A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 146 (3) (2011) 353–358.
- [10] Z. Cong, Y. Diao, Y. Xu, X. Li, Z. Jiang, C. Shao, S. Ji, Y. Shen, W. De, Y. Qiang, Long non-coding RNA linc00665 promotes lung adenocarcinoma progression and functions as ceRNA to regulate AKR1B10-ERK signaling by sponging miR-98, *Cell Death Dis.* 10 (2) (2019) 84.
- [11] Z. Liu, M. Mi, X. Li, X. Zheng, G. Wu, L. Zhang, lncRNA OSTN-AS1 may represent a novel immune-related prognostic marker for triple-negative breast cancer based on integrated analysis of a ceRNA network, *Front. Genet.* 10 (2019) 850.
- [12] Z. Chang, R. Huang, W. Fu, J. Li, G. Ji, J. Huang, W. Shi, H. Yin, W. Wang, T. Meng, et al., The construction and analysis of ceRNA network and patterns of immune infiltration in colon adenocarcinoma metastasis, *Front. Cell Dev. Biol.* 8 (2020) 688.
- [13] J. Xiang, L. Gao, H.Y. Jing, Y.X. Liu, H.F. Wang, Z.W. Chang, S.H. Liu, L. Yu, G.Y. Wang, Construction of CeRNA regulatory network based on WGCNA reveals diagnosis biomarkers for colorectal cancer, *BMC Cancer* 22 (1) (2022) 991.
- [14] C.C. Sun, W. Zhu, S.J. Li, W. Hu, J. Zhang, Y. Zhuo, H. Zhang, J. Wang, Y. Zhang, S.X. Huang, et al., FOXC1-mediated LINC00301 facilitates tumor progression and triggers an immune-suppressing microenvironment in non-small cell lung cancer by regulating the HIF1alpha pathway, *Genome Med.* 12 (1) (2020) 77.
- [15] Director's Challenge Consortium for the Molecular Classification of Lung A, K. Shedden, J.M. Taylor, S.A. Enkemann, M.S. Tsao, T.J. Yeatman, W.L. Gerald, S. Eschrich, I. Jurisica, T.J. Giordano, et al., Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study, *Nat Med* 14 (8) (2008) 822–827.
- [16] X. Li, Y. Shi, Z. Yin, X. Xue, B. Zhou, An eight-miRNA signature as a potential biomarker for predicting survival in lung adenocarcinoma, *J. Transl. Med.* 12 (2014) 159.
- [17] X. Shi, H. Tan, X. Le, H. Xian, X. Li, K. Huang, V.Y. Luo, Y. Liu, Z. Wu, H. Mo, et al., An expression signature model to predict lung adenocarcinoma-specific survival, *Cancer Manag. Res.* 10 (2018) 3717–3732.
- [18] J. Subramanian, R. Simon, Gene expression-based prognostic signatures in lung cancer: ready for clinical use? *J Natl Cancer Inst* 102 (7) (2010) 464–474.
- [19] H. Tang, S. Wang, G. Xiao, J. Schiller, V. Papadimitrakopoulou, J. Minna, I.I. Wistuba, Y. Xie, Comprehensive evaluation of published gene expression prognostic signatures for biomarker-based lung cancer clinical studies, *Ann. Oncol.* 28 (4) (2017) 733–740.
- [20] S. Das, P.K. Meher, A. Rai, L.M. Bhar, B.N. Mandal, Statistical approaches for gene selection, hub gene identification and module interaction in gene Co-expression network analysis: an application to aluminum stress in soybean (*Glycine max* L.), *PLoS One* 12 (1) (2017) e0169605.
- [21] J.H. Li, S. Liu, H. Zhou, L.H. Qu, J.H. Yang, starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data, *Nucleic Acids Res.* 42 (Database issue) (2014) D92–D97.
- [22] A. Scarpa, K. Sikora, M. Fassan, A.M. Rachiglio, R. Cappellesso, D. Antonello, E. Amato, A. Mafficini, M. Lambiase, C. Esposito, et al., Molecular typing of lung adenocarcinoma on cytological samples using a multigene next generation sequencing panel, *PLoS One* 8 (11) (2013) e80478.
- [23] M. Tessema, C.M. Yingling, Y. Liu, C.S. Tellez, L. Van Neste, S.S. Baylín, S.A. Belinsky, Genome-wide unmasking of epigenetically silenced genes in lung adenocarcinoma from smokers and never smokers, *Carcinogenesis* 35 (6) (2014) 1248–1257.
- [24] M.A. Hamidaddin, H. AlRabiah, I.A. Darwish, Development and comparative evaluation of two immunoassay platforms for bioanalysis of crizotinib: a potent drug used for the treatment of non-small cell lung cancer, *Talanta* 201 (2019) 217–225.
- [25] F. Foury, J. Kolodnynski, Pif mutation blocks recombination between mitochondrial rho+ and rho- genomes having tandemly arrayed repeat units in *Saccharomyces cerevisiae*, *Proc Natl Acad Sci U S A* 80 (17) (1983) 5345–5349.
- [26] A. Lahaye, H. Stahl, D. Thines-Sempoux, F. Foury, PIF1: a DNA helicase in yeast mitochondria, *EMBO J.* 10 (4) (1991) 997–1007.
- [27] V.P. Schulz, V.A. Zakian, The *Saccharomyces* PIF1 DNA helicase inhibits telomere elongation and de novo telomere formation, *Cell* 76 (1) (1994) 145–155.

- [28] J. Wang, X. Zhu, P. Ying, Y. Zhu, PIF1 affects the proliferation and apoptosis of cervical cancer cells by influencing TERT, *Cancer Manag. Res.* 12 (2020) 7827–7835.
- [29] T. Zhang, D.M. Wu, P.W. Luo, T. Liu, R. Han, S.H. Deng, M. He, Y.Y. Zhao, Y. Xu, CircNEIL3 mediates pyroptosis to influence lung adenocarcinoma radiotherapy by upregulating PIF1 through miR-1184 inhibition, *Cell Death Dis.* 13 (2) (2022) 167.
- [30] Y. Gao, G. Zhang, J. Liu, H. Li, Tissue-specific transplanted antigen P35B functions as an oncogene and is regulated by microRNA-125a-5p in lung cancer, *Oncol. Rep.* 45 (5) (2021).