

ORIGINAL RESEARCH

Bioinformatics analysis identifies key genes of prognostic value in lung cancer

Dan Song^{1,*}, Li Sun¹

¹Department of Medical Oncology,
Xuzhou Central Hospital, 221009
Xuzhou, Jiangsu, China

***Correspondence**
dsong0565@163.com
(Dan Song)

Abstract

Lung cancer is the most common human malignancy worldwide and can be divided into different types of carcinomas depending on their pathological features. Advances in medical science and technology have led to the identification of some lung cancer-related marker genes, including EGFR (epidermal growth factor receptor), BRAF (B-Raf proto-oncogene), RAS (RAS proto-oncogene, GTPase) and HER2 (human epidermal growth factor receptor 2). However, the underlying biomarker and key genes associated with different types of lung cancer are still poorly understood. In this study, we analyzed a GEO (Gene Expression Omnibus) dataset and identified 28 upregulated intersection DEGs (different expression genes) and 125 downregulated intersection DEGs among AC (adenocarcinoma), PTC (primary typical carcinoid), PLCC (primary large cell carcinoma), PLCNC (primary large cell lung carcinoma) and PSCLC (primary small cell lung carcinoma). Through PPI (protein-protein interaction) network analysis, we identified 14 genes among the DEGs, namely MFAP4 (microfibril-associated protein 4), PDZD2 (PDZ domain containing 2), FBLN1 (fibulin 1), FBLN5 (fibulin 5), EFEMP1 (EGF containing fibulin extracellular matrix protein 1), KDR (kinase insert domain receptor), S1PR1 (sphingosine-1-phosphate receptor 1), CAV1 (caveolin 1), GRK5 (G protein-coupled receptor kinase 5), EDNRA (endothelin receptor type A), EDNRB (endothelin receptor type B), CALCRL (calcitonin receptor-like receptor), PTGER4 (prostaglandin E receptor 4), and ADRB1 (adrenoceptor beta 1), which were found to be downregulated in different subtypes of lung cancer and associated with poor survival outcomes. In addition, most of the screened DEGs demonstrated good predictive ability in LUAD (lung adenocarcinoma) and LUSC (lung squamous cell carcinoma). Among them, MFAP4 was found to promote cell proliferation while also suppressing cell migration and angiogenesis. In summary, we propose MFAP4, PDZD2, FBLN1, FBLN5, EFEMP1, KDR, S1PR1, CAV1, GRK5, EDNRA, EDNRB, CALCRL, PTGER4 and ADRB1 as potential prognostic markers in lung cancer patients.

Keywords

Lung cancer; PPI analysis; GEO; TCGA; Key gene

1. Introduction

Lung cancer is the leading cause of cancer-related death worldwide. In non-small cell lung cancer (NSCLC), the importance of genes as biomarkers is increasingly recognized for their pivotal role in diagnosis, treatment and prognosis estimation, gradually surpassing the significance of histological classification [1]. Comparatively, less than half of lung adenocarcinoma cases exhibit targeted driver gene mutations such as EGFR, BRAF and HER2 or rearrangements involving ALK (ALK receptor tyrosine kinase) and ROS1. However, for other NSCLC cases lacking these specific molecular markers, particularly non-squamous NSCLC, the only treatment option is conventional platinum-based dual therapy [2]. Despite significant advancements in chemotherapy, radiotherapy, surgery,

and immunotherapy for lung cancer, the 5-year survival rate for NSCLC patients remains low at 23%. To address this, comprehensive studies are necessary to uncover the molecular mechanisms underlying lung cancer and develop effective therapeutic strategies.

Significant progress has been achieved in the field of bioinformatics, particularly with advancements in microarray technology, has witnessed rapid progress, which has enabled the widespread use of high-throughput gene expression analysis platforms to identify differentially expressed genes (DEGs) involved in tumorigenesis [3, 4]. Through the application of microarray technology, an increasing number of DEGs associated with lung cancer have been discovered [5, 6]. Nevertheless, it is important to acknowledge that identifying DEGs through this technology has certain limitations regarding sample size,

TMN (Tumor, Node, Metastasis) classification, gender, ethnicity and other factors, which should be accounted when interpreting the results of DEG analyses [3].

In this study, we aimed to identify key genes associated with lung cancer by analyzing microarray data obtained from the GEO database using bioinformatics analysis. We investigated the relevant Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways specific to lung cancer and constructed networks to identify related key genes. Furthermore, by analyzing data from The Cancer Genome Atlas (TCGA) encompassing lung adenocarcinoma and lung squamous cell carcinoma, we examined the expression of these key genes in NSCLC and evaluated their impact on patients' survival. In addition, we also explored the application of these potential candidate biomarkers in the diagnosis of NSCLC, providing valuable insights into potential targets and prognostic markers for enhancing the diagnosis and treatment of lung cancer.

2. Materials and methods

2.1 PCA, volcano plot and Venn diagram analysis

PCA (principal component analysis) was performed as previously depicted [7]. The GEO dataset GSE1037, which consists data on cDNA microarray containing 40,386 elements, was used to analyze the gene expression profiles of 38 surgically resected samples of lung neuroendocrine tumors and 11 SCLC (small cell lung cancer) cell lines (accessed *via*: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1037>) with the R package (limma analysis), after applying a batching correction. PCA was used to visualize the differences between the normal group (primary normal lung tissue, $n = 19$) and the cancer group, including PTC (primary typical carcinoid, $n = 12$), AC (adenocarcinoma, $n = 12$), PLCNC (primary large cell neuroendocrine carcinoma, $n = 8$), PSCLC (primary small cell lung carcinoma, $n = 15$) and PLCC (primary large cell carcinoma, $n = 14$).

In the volcano plot, the vertical lines represent genes with an absolute value of $|\log_2(\text{fold change})| \geq 2$, indicating up-regulated or down-regulated DEGs, while the horizontal line indicates a significance level of $p < 0.05$. Significantly up-regulated genes are denoted by red dots, while significantly downregulated genes are represented by blue dots. By using a Venn diagram, we identified 28 upregulated genes and 125 downregulated genes that were differentially expressed based on the intersection of the gene sets.

2.2 GO/KEGG pathway enrichment analysis and protein-protein interaction analysis

The expression of 14 key genes in lung tumor and non-tumor tissues was analyzed using the FactoMineR package in R, and the results were visualized using heat maps. GO-KEGG analysis (<https://www.genome.jp/kegg/>) was performed as previously described [8]. The online platform Metascape (<https://metascape.org/>) and R language version 4.3.1 (Beagle Scouts, Murray Hill, NJ, USA) were used for enrichment analysis of those above-mentioned intersecting

differential genes. PPI (protein-protein interaction) analysis was performed on the STRING online platform (<https://string-db.org/>).

2.3 Correlation and cluster analysis of differential gene expression

The correlation of key gene expression between normal lung tissue and lung carcinoma tissue was analyzed using the clusterProfiler package of R language [9, 10].

2.4 GEO, GEPIA analysis and Kaplan-Meier curve analysis

The expression of 14 DEGs was analyzed by box plot from GSE1037. The GEPIA online software (<http://gepia.cancer-pku.cn/>) was used to analyze the expression of *MFAP4*, *PDZD2*, *FBLN1*, *FBLN5*, *EFEMP1*, *KDR*, *SIPRI*, *CAVI*, *GRK5*, *EDNRA*, *EDNRB*, *CALCRL*, *PTGER4* and *ADRB1* in different types of lung cancer. The prognosis of patients was assessed using the Kaplan-Meier curve tool to analyze and visualize the survival data for the investigated patients (<http://kmplot.com/analysis/>).

2.5 ROC curve analysis

ROC (Receiver Operating Characteristic) curves of differential genes were analyzed using the R language based on the TCGA-GTEX (Genotype-Tissue Expression) data. The type of data from TCGA comprised RNAseq data, while data from GEO were derived from cDNA microarrays.

2.6 Cell lines

A549 and SK-MES-1 cells were purchased from Procell (Wuhan, China), cultured in DMEM (Dulbecco's Modified Eagle Medium) with 10% FBS (fetal bovine serum) and penicillin-streptomycin, then incubated at 37 °C in a 5% CO₂ (carbon dioxide) environment, following the manufacturer's instructions.

2.7 Western blotting assay

Western blotting analysis was performed as previously depicted [11]. Briefly, proteins were extracted using RIPA lysis buffer, then denatured, electrophorized, transferred, immunoblotted and visualized using a chemiluminescent ECL reagent, followed by immunoblotting with the following antibodies: anti-mouse MFAP4 (1:1000, Santa, sc-398438, USA), anti-mouse p-ERK1/2 (1:1000, Abcam, ab278538, England), anti-mouse ERK1/2 (1:1000, Abcam, ab17942, England), anti-mouse p38 (1:1000, Santa, sc-7972, USA), anti-mouse p-p38 (1:1000, Santa, sc-166182, USA), anti-mouse p-JNK (1:1000, Santa, sc-6254, USA), anti-mouse JNK (1:1000, Santa, sc-7345, USA) and anti-mouse GAPDH (1:1000, Santa, sc-47724, USA).

2.8 Cell proliferation, migration and angiogenesis

Cell proliferation and migration assays were conducted following previously described methods [12]. The indicated cell

lines were seeded into a 96-well plate to assess cell viability at various time points (0 h, 24 h, 48 h and 72 h). Additionally, the same cell lines were seeded into a transwell chamber after incubating for 24 h, following which the number of migrated cells was counted. For angiogenesis, human umbilical vein endothelial cells (HUVECs) were cultured in a medium obtained from lung cancer cells transfected with the designated plasmids for 48 h to conduct the tubule formation experiment. The cells were washed, detached, seeded and imaged as previously described [13].

2.9 Statistical analysis

The SPSS v22.0 software (IBM, Armonk, NY, USA) was used for statistical analyses. The Student's *t*-test and one-way ANOVA (Analysis of variance) were performed, and data are presented as means \pm SEM (standard error of the mean) of three independent experiments. A *p*-value of 0.05 or less was considered significant.

3. Results

3.1 Analysis of differential genes DEGs in lung cancer based on GEO data

Bioinformatics methods were applied to analyze DEGs from GSE1037 to identify key genes related to the prognosis of lung cancer patients. After batch correction, the data were analyzed by PCA (Fig. 1A). The results showed significant differences between diverse kinds of lung cancer, including AC, PLCC, PLCNC, PSCLC and PTC, and the normal tissue group. We analyzed the DEGs between the lung cancer and normal group by volcano plot and identified 743 upregulated and 769 downregulated genes in PTC, 313 upregulated and 199 downregulated genes in AC, 454 upregulated and 501 downregulated expression genes in PLCNC, 325 upregulated and 1014 downregulated expression genes in PSCLC, and 537 upregulated and 416 downregulated genes in PLCC (Fig. 1B). From these, 28 upregulated and 125 downregulated intersection DEGs were identified among the AC, PTC, PLCC, PLCNC and PSCLC samples *via* Venn analysis (Fig. 1C).

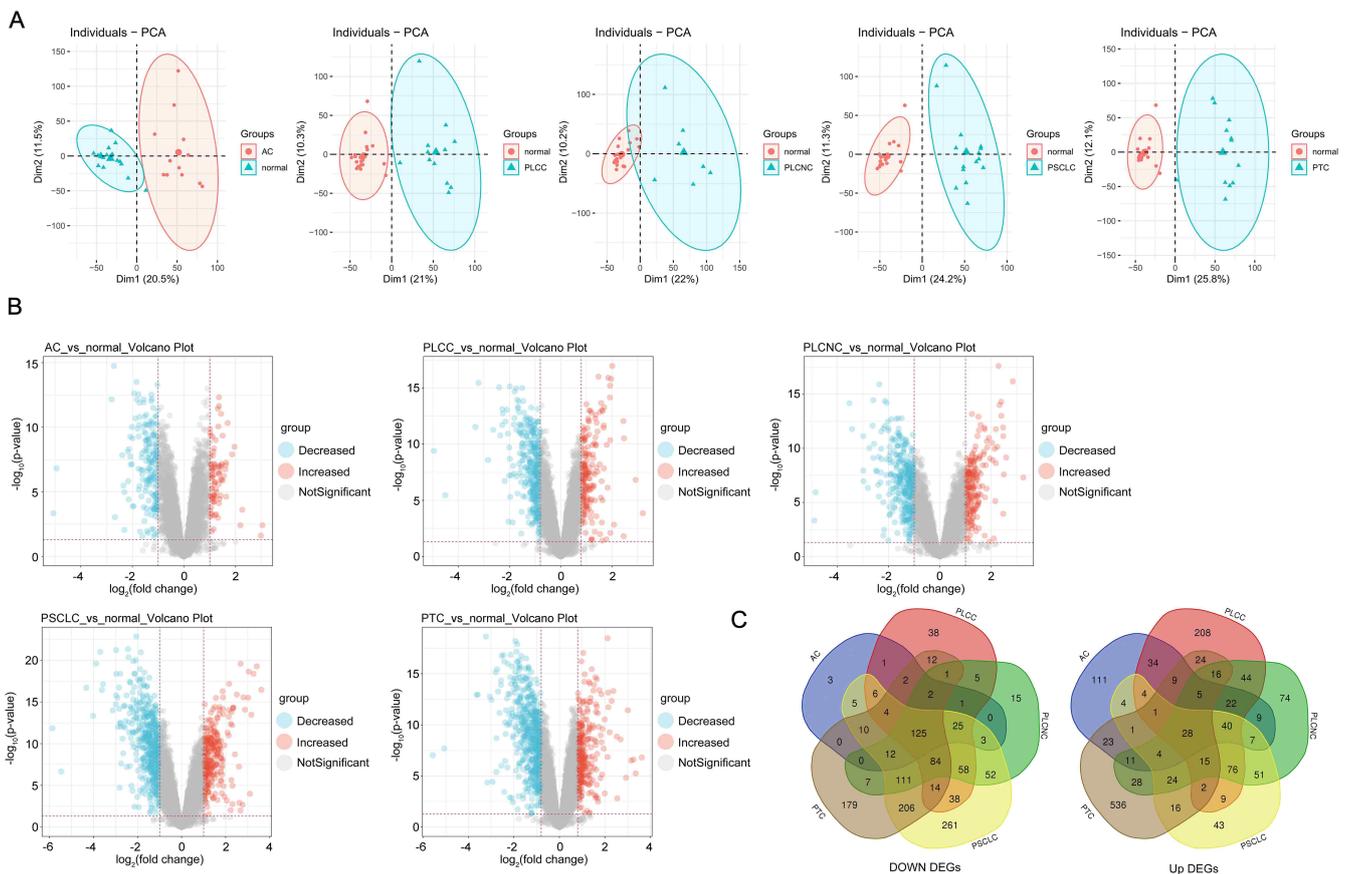


FIGURE 1. Analysis of differential genes DEGs in lung cancer based on GEO data. (A) PCA was used to analyze the difference between normal and lung cancer, including AC (adenocarcinoma), PTC (primary typical carcinoid), PLCC (primary large cell carcinoma), PLCNC (primary large cell lung carcinoma) and PSCLC (primary small cell lung carcinoma). (B) Volcano plot was used to assess DEGs between normal and different types of lung cancer. Blue: decreased expression genes, red: increased expression genes, grey: no significant change expression genes. (C) Venn plot analyzed the intersection genes among AC (adenocarcinoma), PTC (primary typical carcinoid), PLCC (primary large cell carcinoma), PLCNC (primary large cell lung carcinoma) and PSCLC (primary small cell lung carcinoma).

3.2 Differential gene enrichment analysis, cluster analysis and protein interaction network PPI analysis

Enrichment analysis was performed on the above intersection DEGs (Fig. 2A), and the results indicated that the screened DEGs were associated with angiogenesis, cell-substrate adhesion, circulatory system and locomotion process. KEGG enrichment analysis showed that the DEGs were mainly involved in the MAPK pathway. Additionally, the PPI network analysis identified 14 genes among the DEGs, namely *MFAP4*, *PDZD2*, *FBLN1*, *FBLN5*, *EFEMP1*, *KDR*, *S1PR1*, *CAV1*, *GRK5*, *EDNRA*, *EDNRB*, *CALCRL*, *PTGER4* and *ADRB1* (Fig. 2B).

Next, we analyzed the expression of the selected 14 DEGs in tumor tissues and normal tissues of lung cancer patients and found that all were downregulated expression in lung carcinoma tissues compared to normal lung tissues (Fig. 3A). Moreover, the PPI analysis (Fig. 2B) demonstrated a consistent interaction pattern among the 14 DEGs (Fig. 3B). Notably, our analysis using R revealed that nearly all of the 14 DEGs displayed significant correlations with each other (Fig. 3C).

3.3 Analysis of differential gene expression in different subtypes of lung cancer from GEO and TCGA datasets

Based on the GEO chip data GSE1037, we analyzed the expression levels of the 14 DEGs in AC, PTC, PLCC, PLCNC and PSCLC by box plot (Fig. 4) and found no significant difference in *PDZD2* between the different subtypes of lung cancer and the normal group. The expression of *MFAP4* was downregulated in all kinds of lung cancer except for PTC, and *PTGER4* was downregulated in different subtypes of lung cancer, excluding PTC and PLCNC. Additionally, *FBLN1*, *FBLN5*, *EFEMP1*, *KDR*, *S1PR1*, *CAV1*, *GRK5*, *EDNRA*, *EDNRB*, *CALCRL* and *ADRB1* were all significantly decreased in AC, PTC, PLCC, PLCNC and PSCLC compared to the normal group (Fig. 4).

Similarly, we used the GEPIA online platform to analyze the expression of 14 DEGs in LUAD (T = 483, N = 347) and LUSC (T = 486, N = 338). As shown in Fig. 5, for LUAD, which included 483 tumor cases and 347 normal cases, and LUSC, which comprised 486 tumor tissues and 338 normal tissues, all of the 14 DEGs were downregulated in the LUSC tissues compared to the normal group. Further, in LUAD tissues, all 14 DEGs showed significantly lower expression

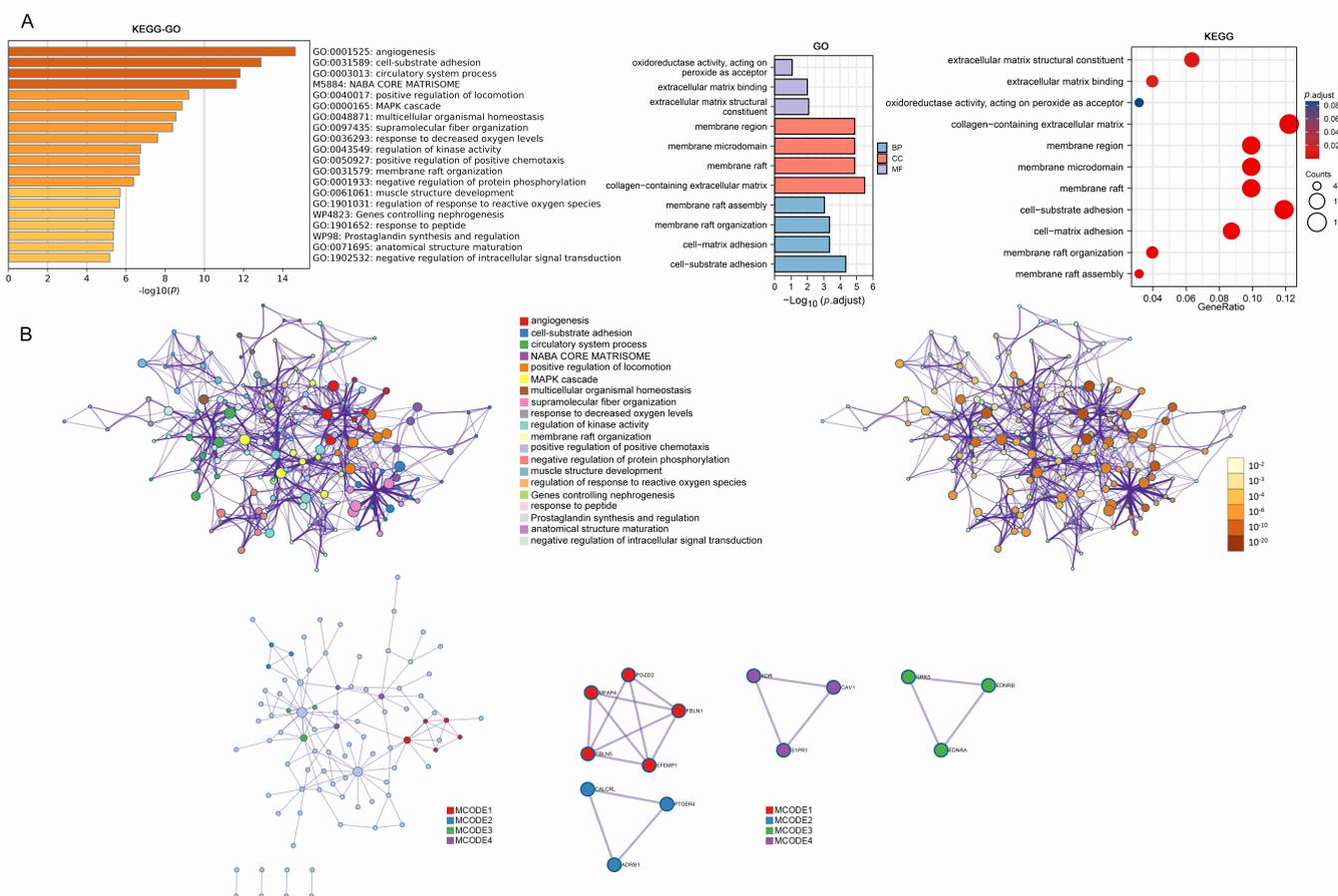


FIGURE 2. Differential gene enrichment analysis and protein interaction network PPI analysis. (A) GO-KEGG enrichment analysis of the intersection DEGs (differentially expressed genes). For GO database analysis, the functions of genes were divided into three parts: cellular component (CC), molecular function (MF), and biological process (BP). (B) PPI (protein-protein interaction) network analysis of the intersection DEGs was analyzed on the Metascape online platform (<https://metascape.org/gp/index.html#/main/step1>).

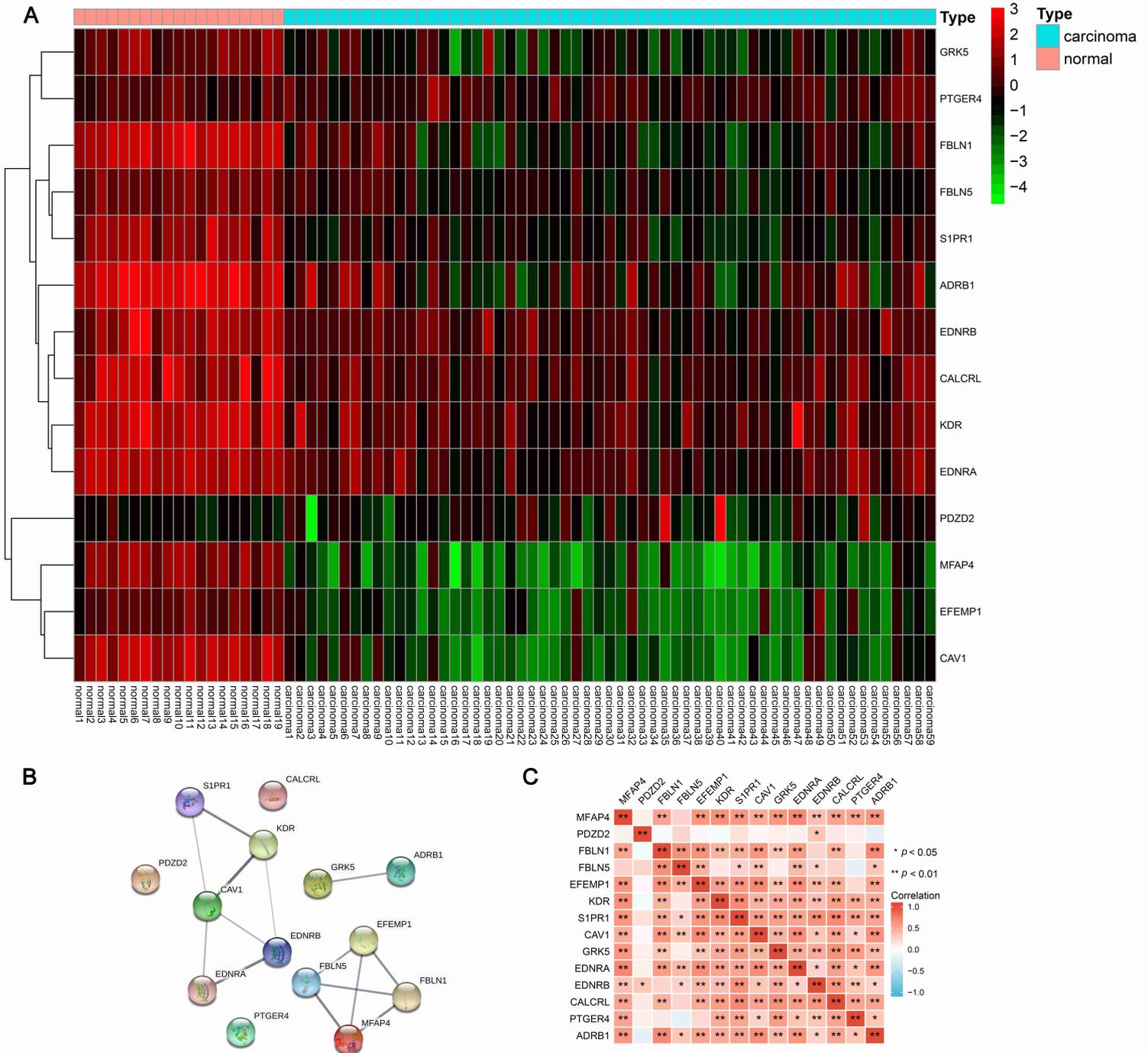


FIGURE 3. Differential gene expression correlation and cluster analysis. (A) Heatmap analysis of the expression of MFAP4 (microfibril-associated protein 4), PDZD2 (PDZ domain containing 2), FBLN1 (fibulin 1), FBLN5 (fibulin 5), EFEMP1 (EGF containing fibulin extracellular matrix protein 1), KDR (kinase insert domain receptor), S1PR1 (sphingosine-1-phosphate receptor 1), CAV1 (caveolin 1), GRK5 (G protein-coupled receptor kinase 5), EDNRA (endothelin receptor type A), EDNRB (endothelin receptor type B), CALCRL (calcitonin receptor-like receptor), PTGER4 (prostaglandin E receptor 4), and ADRB1 (adrenoceptor beta 1) between normal and carcinoma tissues. (B) PPI among MFAP4, PDZD2, FBLN1, FBLN5, EFEMP1, KDR, S1PR1, CAV1, GRK5, EDNRA, EDNRB, CALCRL, PTGER4 and ADRB1 was analyzed on the STRING online platform. (C) Correlation of MFAP4, PDZD2, FBLN1, FBLN5, EFEMP1, KDR, S1PR1, CAV1, GRK5, EDNRA, EDNRB, CALCRL, PTGER4 and ADRB1 determined by R language analysis.

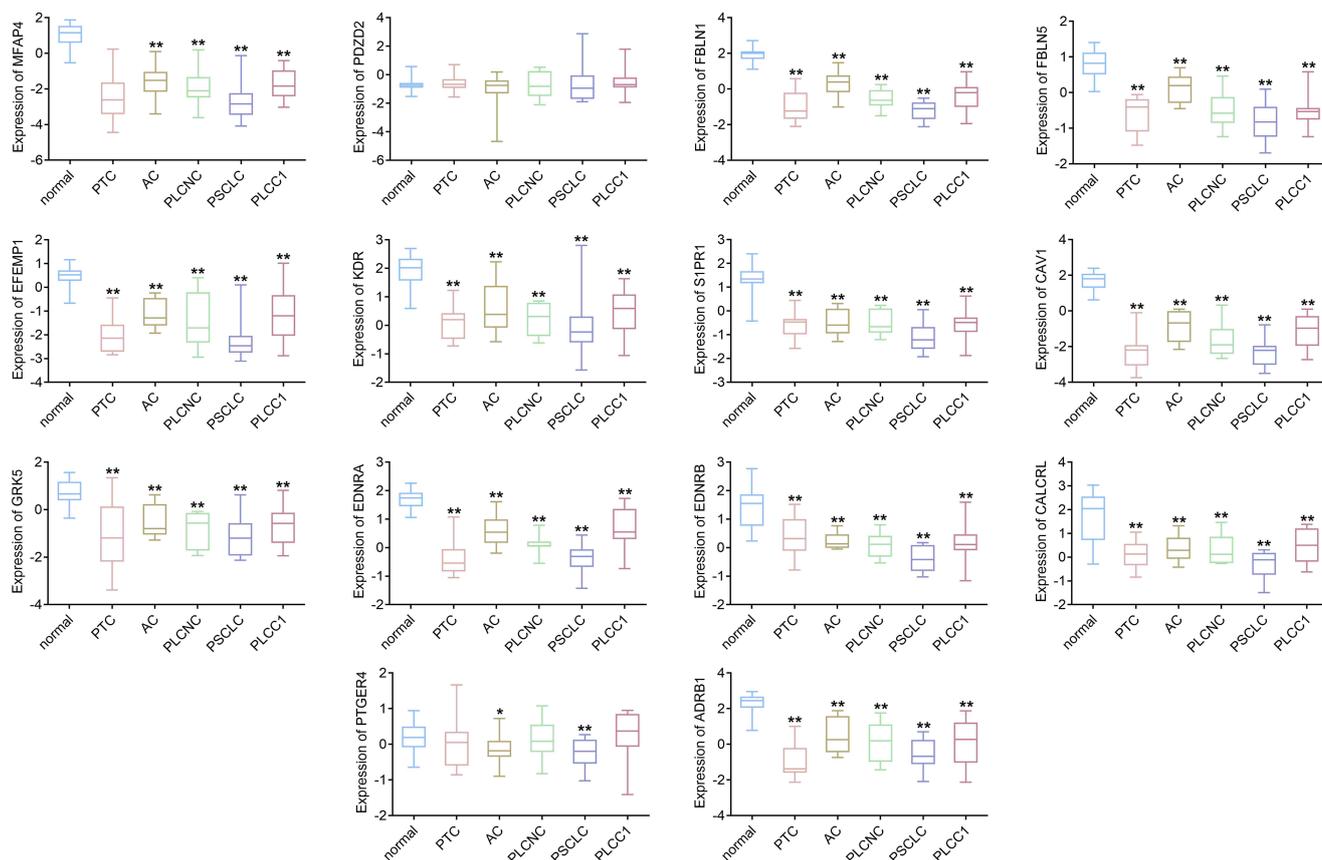


FIGURE 4. Expression of key genes in different subtypes of lung cancer. Box plot was used to analyze the expression of MFAP4 (microfibril-associated protein 4), PDZD2 (PDZ domain containing 2), FBLN1 (fibulin 1), FBLN5 (fibulin 5), EFEMP1 (EGF containing fibulin extracellular matrix protein 1), KDR (kinase insert domain receptor), S1PR1 (sphingosine-1-phosphate receptor 1), CAV1 (caveolin 1), GRK5 (G protein-coupled receptor kinase 5), EDNRA (endothelin receptor type A), EDNRB (endothelin receptor type B), CALCRL (calcitonin receptor-like receptor), PTGER4 (prostaglandin E receptor 4), and ADRB1 (adrenoceptor beta 1) in different types of lung cancer from GSE1037.

levels compared to the normal group, except for EDNRA.

Collectively, our data indicate that these 14 DEGs were downregulated in different subtypes of lung cancer.

3.4 Influence of DEGs on lung adenocarcinoma and lung squamous cell carcinoma survival

Next, we analyzed the relationship between the expression of the 14 DEGs and survival rates. As shown in Fig. 6, for LUAD, a low expression of MFAP4, PDZD2, FBLN1, FBLN5, EFEMP1, KDR, S1PR1, CAV1, GRK5, EDNRA, EDNRB, CALCRL, PTGER4 and ADRB1 were related with poor survival. Comparatively, for LUSC, only MFAP4 and PTGER4 were positively correlated with survival rate, and there was no difference between other gene expressions and survival (Fig. 6 and Supplementary Fig. 1).

3.5 Based on the TCGA database to analyze the value of differential genes in lung cancer diagnosis

The area under the ROC curve (AUC) of the nomogram of the training data set was $AUC \geq 0.9$, with a $CI > 0.9$, indicating

good predictive ability in the TCGA cohort test, as a high AUC (≥ 0.9) represents high prediction accuracy. For LUAD diagnosis, MFAP4, FBLN1, FBLN5, EFEMP1, S1PR1, CAV1, GRK5, EDNRB, CALCRL, PTGER4 and ADRB1 exerted good predictive ability in the TCGA cohort test (Fig. 7). Comparatively, for LUSC diagnosis, only MFAP4, FBLN5, EFEMP1, KDR, S1PR1, CAV1, GRK5, EDNRB, CALCRL and ADRB1 exerted good predictive ability in the TCGA cohort test (Fig. 7). Further, the ROC curves of the nomogram of the training data set for PDZD2, EDNRA and KDR in LUAD, PDZD2, PTGER4, FBLN1 and EDNRA were $AUC < 0.9$ (Supplementary Fig. 2).

3.6 Effects of upregulated MFAP4 on angiogenesis and MAPK pathway in lung cancer cells

According to the GO/KEGG enrichment analysis, differential genes were significantly enriched in angiogenesis, cell adhesion movement and MAPK pathway, further confirming the role of MFAP4 at the cell level. From the above-screened genes, MFAP4 had the most differential expression and was the most significant marker associated with patient survival and was thus selected for validation. As shown in Fig. 8A,

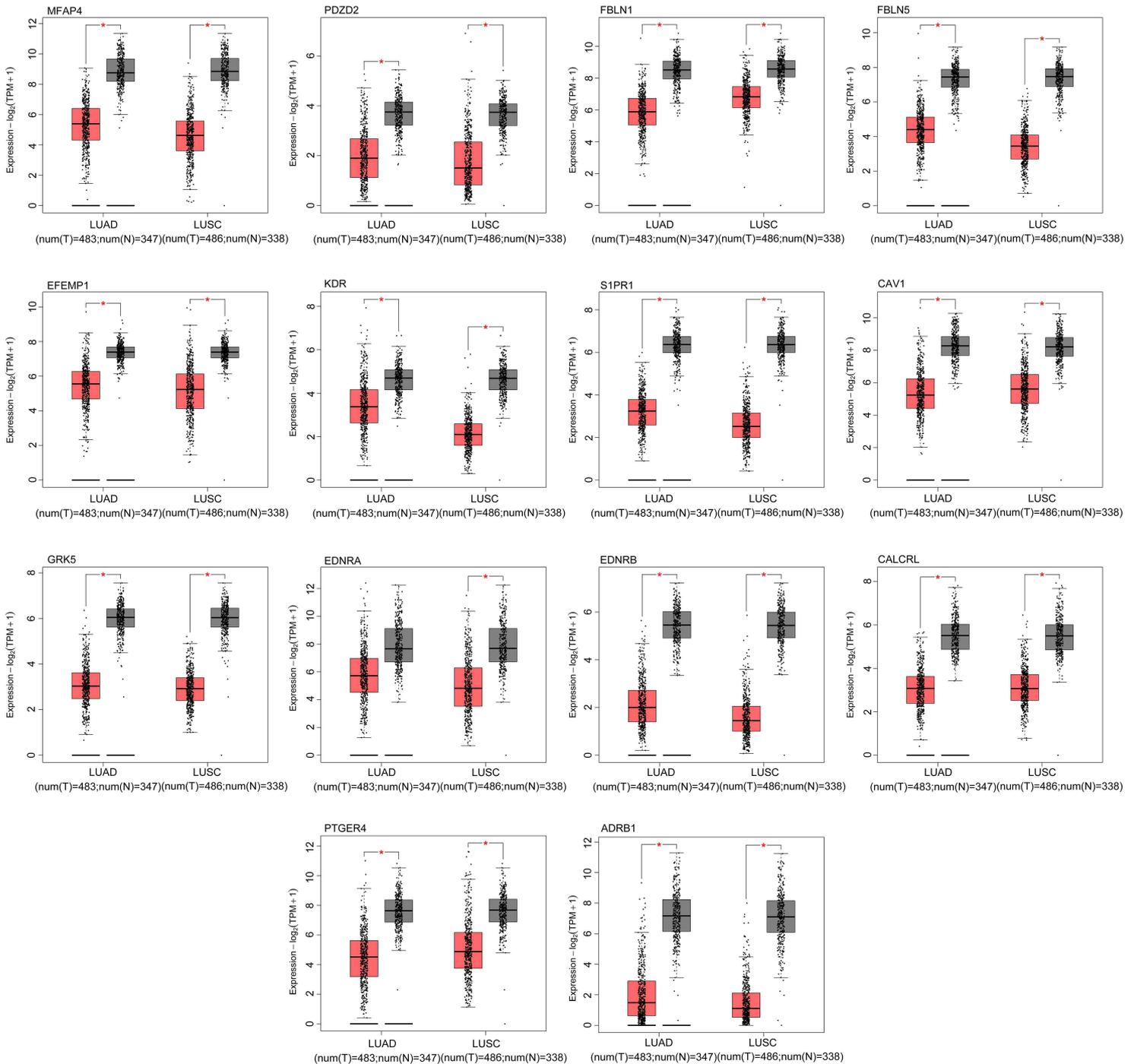


FIGURE 5. Analysis of differential gene expression in lung adenocarcinoma and lung squamous cell carcinoma based on the TCGA database. GEPIA (gene expression profiling interactive analysis) online software was used to analyze the expression of MFAP4 (microfibril-associated protein 4), PDZD2 (PDZ domain containing 2), FBLN1 (fibulin 1), FBLN5 (fibulin 5), EFEMP1 (EGF containing fibulin extracellular matrix protein 1), KDR (kinase insert domain receptor), S1PR1 (sphingosine-1-phosphate receptor 1), CAV1 (caveolin 1), GRK5 (G protein-coupled receptor kinase 5), EDNRA (endothelin receptor type A), EDNRB (endothelin receptor type B), CALCRL (calcitonin receptor-like receptor), PTGER4 (prostaglandin E receptor 4), and ADRB1 (adrenoceptor beta 1) in LUAD (lung adenocarcinoma) and LUSC (lung squamous cell carcinoma).

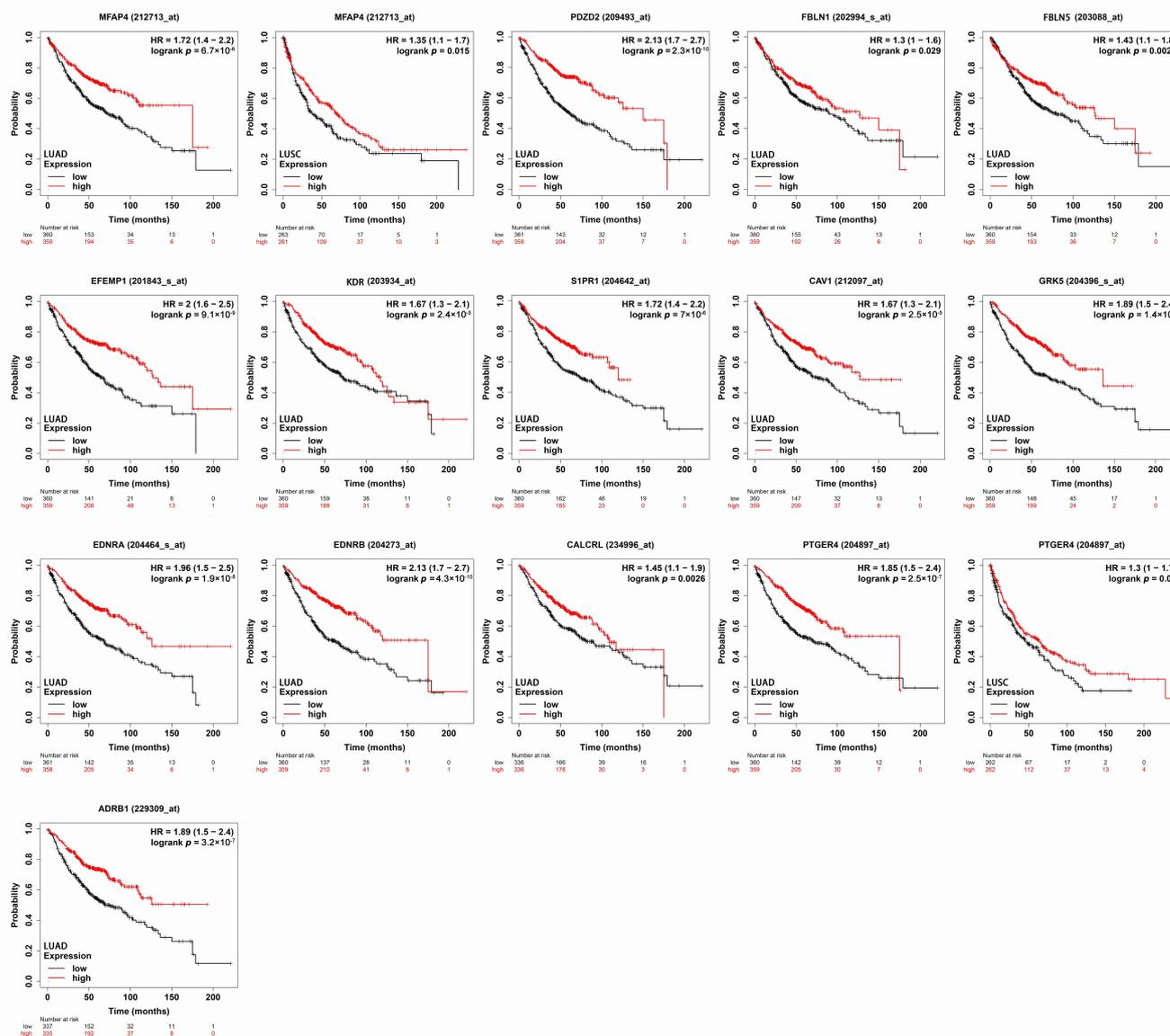


FIGURE 6. The influence of differential genes on the survival and prognosis of lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC). Kaplan-Meier plot analysis was performed to determine the relationship between key gene expression and survival in LUAD and LUSC. $p < 0.05$.

MFAP4 expression was decreased in A549 and SK-MES-1 cells compared with BEAS-2B cells. MFAP4 overexpression inhibited cell proliferation in A549 and SK-MES-1 cells compared to the control cell lines. The knockdown of MFAP4 induced cell proliferation in the indicated cell lines (Fig. 8B) [14]. Moreover, cell migration was decreased in A549 and SK-MES-1 cells following transfection with MFAP4, while it was increased in MFAP4 deficient A549 and SK-MES-1 cell lines (Fig. 8C). In addition, conditioned medium from MFAP4 overexpression reduced angiogenesis in HUVECs cell lines, and the knockdown of MFAP4 facilitated angiogenesis in HUVECs cell lines (Fig. 8D). Mechanistically, the overexpression of MFAP4 suppressed the phosphorylation of ERK1/2, p38 and JNK (c-Jun N-terminal kinases) in A549 and SK-MES-1 cells. MFAP4 deficiency induced the expression of p-ERK1/2, p-p38 and p-JNK in A549 and SK-MES-1 cell lines (Fig. 8E). Collectively, these data suggest that MFAP4 negatively regulates cell proliferation, migration and angiogenesis.

4. Discussion

Lung cancer is a highly prevalent and deadly disease worldwide, causing an estimated 1.6 million deaths annually [15]. Most lung cancer cases are classified as NSCLC, which is associated with low survival rates [16]. The complexity, histopathology, and clinical features of lung cancer make it challenging to understand the underlying mechanisms of its development and progression [17]. In this study, we used bioinformatics methods and screened 14 significant DEGs, including *MFAP4*, *PDZD2*, *FBLN1*, *FBLN5*, *EFEMP1*, *KDR*, *S1PR1*, *CAV1*, *GRK5*, *EDNRA*, *EDNRB*, *CALCRL*, *PTGER4* and *ADRB1*, in different lung cancer subtypes, which was found to be involved in angiogenesis, cell adhesion and locomotion pathways. In addition, the correlated DEGs were downregulated in lung cancer and positively associated with patient survival.

Pulmonary neuroendocrine tumors are commonly encoun-

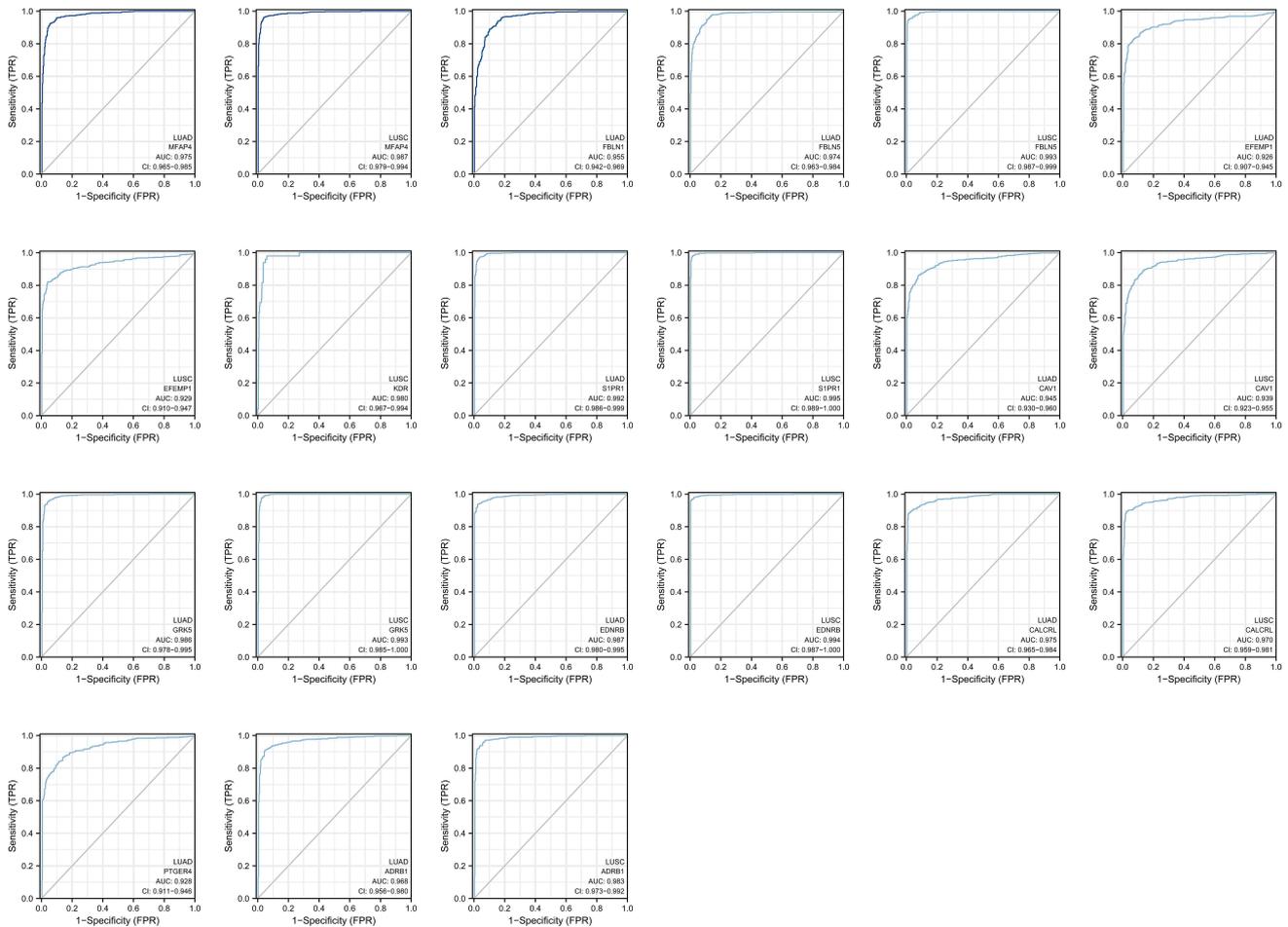


FIGURE 7. Significance of differential genes from the TCGA database in lung cancer diagnosis. ROC (receiver operating characteristic) curve analysis was used to assess the accuracy of each prognostic gene in different types of lung cancer. AUC (Area Under the ROC Curve) ≥ 0.9 , CI (confidence interval): 0.8–0.9.

tered in surgical pathology and can be classified into four types, namely primary typical carcinoid (TC), atypical carcinoid, small cell lung cancer (SCLC) and large cell neuroendocrine carcinoma (LCNEC). Pulmonary neuroendocrine tumors account for approximately 20% of all primary lung tumors, encompassing classic carcinoid, atypical carcinoid, small cell carcinoma and large cell neuroendocrine carcinoma, and approximately 25% of neuroendocrine tumors [18]. The diagnosis of pulmonary neuroendocrine tumors presents challenges due to morphological similarities with other entities. Firstly, identifying their neuroendocrine characteristics can be challenging, especially in biopsies with limited tissue samples and intraoperative frozen sections. Secondly, distinguishing true mitosis and nuclear pyknosis can be challenging, and focal necrosis may be easily overlooked, leading to difficulties in classifying tumors into distinct categories. Thirdly, clinically differentiating primary tumors from metastases may not always be evident. The classification criteria are specifically designed for primary neuroendocrine tumors of the lung, and it remains unclear how these criteria apply to metastatic tumors. In this present study, bioinformatics analysis was performed to analyze the common intersected DEGs in five subtypes of lung cancers (AC, PTC, PLCC, PSCLC, PLCNC).

Cancer is characterized by its high viability, metastatic potential and easy relapse. Therefore, research efforts often focus on understanding the role of DEGS in cell proliferation, migration, invasion and apoptosis in different kinds of cancer. Similarly, we also demonstrated that the selected DEGs were involved in angiogenesis, cell adhesion and locomotion pathways, which indicated their important role in the progression and development of lung cancer. Here, MFAP4, PDZD2, FBLN1, FBLN5, EFEMP1, KDR, S1PR1, CAV1, GRK5, EDNRB, EDNRB, CALCRL, PTGER4 and ADRB1 were associated with tumorigenesis by GO-KEGG enrichment analysis.

LUAD and LUSC are the two most common types of NSCLC, and LUSC is the main pathological type of lung malignancy [19–21]. Currently, less than 10% of patients with advanced metastatic LUSC can live more than 5 years [22]. A previous study demonstrated that CEP55 is a diagnostic marker and independent prognostic factor in LUAD and LUSC [23]. The downregulated expression of miR-126-3p was associated with the occurrence and progression of LUAD [24]. GPR115 strongly correlated with poor prognosis of LUAD but not LUSC [25]. High expression of GSDMD indicated a poor prognosis in LUAD but not in LUSC

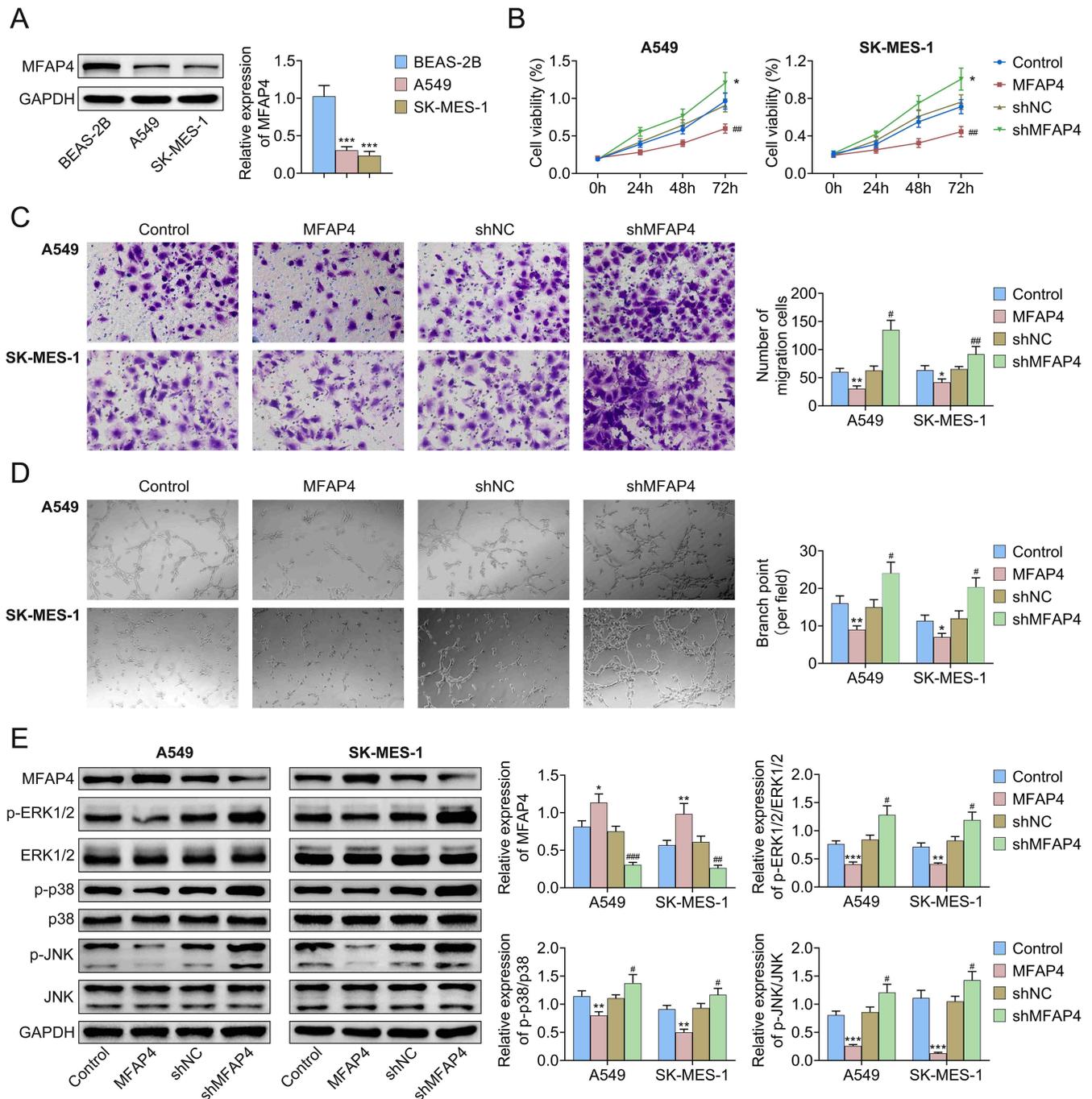


FIGURE 8. Effects of upregulated MFAP4 on angiogenesis and MAPK pathway in lung cancer cells. (A) The expression of MFAP4 was assessed in BEAS-2B, A549 and SK-MES-1 *via* western blotting. (B) A549 and SK-MES-1 were divided into four groups (Control, MFAP4, shNC and shMFAP4) depending on different plasmids transfection. Cell proliferation of the cell line was assessed by MTT assay. (C) A549 and SK-MES-1 were transfected with the indicated plasmids, and transwell assay was performed to assess cell migration with quantitative data attachment. (D) Human umbilical vein endothelial cells (HUVECs) were cultured with a transfected medium of lung cancer cells to conduct tubule formation experiments with quantitative data attachment. (E) The related protein expression was measured by western blotting. Quantitative protein expression (MFAP4, p-ERK1/2, p-p38, and p-JNK (c-Jun N-terminal kinases)) data is attached. Error bars represent data from three independent experiments (mean \pm SD). $###p < 0.01$. $**p < 0.01$. #compared to the control group. *compared to the shNC group.

[26]. Similarly, we found that MFAP4 and PTGER4 were associated with the survival rate of LUSC patients, while the expression of MFAP4, FBLN1, FBLN5, EFEMP1, KDR, S1PR1, CAV1, GRK5, EDNRB, CALCRL and ADRB1 was positively correlated with the survival of LUAD patients. However, the underlying function of these DEGs in LUSC and LUAD needs further research. Generally, MFAP4 was identified as a tumor suppressor in human cancers, with its overexpression significantly associated with better prognosis in breast cancer and reported to be a downstream target gene regulated by miR-147b, exerting anti-proliferation function in LUAD cell lines [14, 27]. Moreover, Han *et al.* [28] reported that the C8orf34-as1/miR-671-5p/MFAP4 regulatory cascade played a critical role in LUAD cell migration. Here, we also confirmed the same role of MFAP4 in LUSC and LUAD. However, the underlying molecular mechanism of MFAP4 in lung cancer progression needs more research.

5. Conclusions

Herein, we demonstrated that MFAP4, PDZD2, FBLN1, FBLN5, EFEMP1, KDR, S1PR1, CAV1, GRK5, EDNRB, EDNRB, CALCRL, PTGER4 and ADRB1 were downregulated in lung cancer and associated with poor survival, indicating their potential as hallmark genes in predicting the prognosis of different subtypes of lung cancers.

AVAILABILITY OF DATA AND MATERIALS

The data are contained within this article and supplementary material.

AUTHOR CONTRIBUTIONS

DS and LS—designed the study, supervised the data collection, analyzed the data, interpreted the data, prepared the manuscript for publication and reviewed the draft. All authors have read and approved the manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

ACKNOWLEDGMENT

Not applicable.

FUNDING

This research received no external funding.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

SUPPLEMENTARY MATERIAL

Supplementary material associated with this article can be found, in the online version, at <https://oss.jomh.org/files/article/1685899010770911232/attachment/Supplementary%20material.docx>.

REFERENCES

- [1] Nicholson AG, Tsao MS, Beasley MB, Borczuk AC, Brambilla E, Cooper WA, *et al.* The 2021 WHO classification of lung tumors: impact of advances since 2015. *Journal of Thoracic Oncology*. 2022; 17: 362–387.
- [2] Chaff JE, Shyr Y, Sepesi B, Forde PM. Preoperative and postoperative systemic therapy for operable non-small-cell lung cancer. *Journal of Clinical Oncology*. 2022; 40: 546–555.
- [3] Wang L, Qu J, Liang Y, Zhao D, Rehman FU, Qin K, *et al.* Identification and validation of key genes with prognostic value in non-small-cell lung cancer *via* integrated bioinformatics analysis. *Thoracic Cancer*. 2020; 11: 851–866.
- [4] Agapito G, Arbitrio M. Microarray data analysis protocol. *Methods in Molecular Biology*. 2022; 171: 263–271.
- [5] Dong Y, Cao H, Liang Z. A curated target gene pool assisting early disease prediction and patient-specific treatment for small cell lung cancer. *Journal of Computational Biology*. 2018; 25: 576–585.
- [6] Su C, Liu W, Wu L, Dong T, Liu J. Screening of hub gene targets for lung cancer *via* microarray data. *Combinatorial Chemistry & High Throughput Screening*. 2021; 24: 269–285.
- [7] Zhang L, Chen J, Yang H, Pan C, Li H, Luo Y, *et al.* Multiple microarray analyses identify key genes associated with the development of non-small cell lung cancer from chronic obstructive pulmonary disease. *Journal of Cancer*. 2021; 12: 996–1010.
- [8] Dong L, Yu L, Bai C, Liu L, Long H, Shi L, *et al.* USP27-mediated Cyclin E stabilization drives cell cycle progression and hepatocellular tumorigenesis. *Oncogene*. 2018; 37: 2702–2713.
- [9] Sun M, Shao X, Wang Y. Microarray data analysis for transcriptome profiling. *Methods in Molecular Biology*. 2018; 270: 17–33.
- [10] Larriba Y, Rueda C, Fernández MA, Peddada SD. Microarray data normalization and robust detection of rhythmic features. *Methods in Molecular Biology*. 2019; 99: 207–225.
- [11] Dong L, Yu L, Li H, Shi L, Luo Z, Zhao H, *et al.* An NAD⁺-Dependent Deacetylase SIRT7 Promotes HCC development through deacetylation of USP39. *IScience*. 2020; 23: 101351.
- [12] Yu L, Dong L, Wang Y, Liu L, Long H, Li H, *et al.* Reversible regulation of SATB1 ubiquitination by USP47 and SMURF2 mediates colon cancer cell proliferation and tumor progression. *Cancer Letters*. 2019; 448: 40–51.
- [13] Kelley M, Fierstein S, Purkey L, DeCicco-Skinner K. Endothelial cell tube formation assay: an *in vitro* model for angiogenesis. *Methods in Molecular Biology*. 2022; 127: 187–196.
- [14] Feng Y, Liu C, Xue Y, Chen Y, Wang Y, Wu X. MicroRNA-147b promotes lung adenocarcinoma cell aggressiveness through negatively regulating microfibril-associated glycoprotein 4 (MFAP4) and affects prognosis of lung adenocarcinoma patients. *Gene*. 2020; 730: 144316.
- [15] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. 2021; 71: 209–249.
- [16] Liu S, Zhang J, Zeng K, Wu Y. Perioperative targeted therapy for oncogene-driven NSCLC. *Lung Cancer*. 2022; 172: 160–169.
- [17] Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Lung cancer. *The Lancet*. 2021; 398: 535–554.
- [18] Cecchini MJ, Tarmey T, Ferreira A, Mangaonkar AA, Ferrer A, Patnaik MM, *et al.* Pathology, radiology, and genetics of interstitial lung disease in patients with shortened telomeres. *American Journal of Surgical Pathology*. 2021; 45: 871–884.
- [19] Bu J, Zhang P, Zhu K, Yan Y, Shi B, Wang J, *et al.* Constructing a global transcriptional regulatory landscape for early non-small cell lung cancer to identify hub genes and key pathways. *Aging*. 2020; 12: 17948–17957.

- [20] Zhang L, Chen J, Cheng T, Yang H, Li H, Pan C. Identification of the key genes and characterizations of tumor immune microenvironment in lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC). *Journal of Cancer*. 2020; 11: 4965–4979.
- [21] Chen M, Liu X, Du J, Wang X, Xia L. Differentiated regulation of immune-response related genes between LUAD and LUSC subtypes of lung cancers. *Oncotarget*. 2017; 8: 133–144.
- [22] Moin AT, Sarkar B, Ullah MA, Araf Y, Ahmed N, Rudra B. *In silico* assessment of EpCAM transcriptional expression and determination of the prognostic biomarker for human lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC). *Biochemistry and Biophysics Reports*. 2021; 27: 101074.
- [23] Fu L, Wang H, Wei D, Wang B, Zhang C, Zhu T, *et al.* The value of CEP55 gene as a diagnostic biomarker and independent prognostic factor in LUAD and LUSC. *PLOS ONE*. 2020; 15: e0233283.
- [24] Chen S, Lu H, Chen G, Yang J, Huang W, Wang X, *et al.* Downregulation of miRNA-126-3p is associated with progression of and poor prognosis for lung squamous cell carcinoma. *FEBS Open Bio*. 2020; 10: 1624–1641.
- [25] Wang Y, Shi M, Yang N, Zhou X, Xu L. GPR115 contributes to lung adenocarcinoma metastasis associated with LAMC2 and predicts a poor prognosis. *Frontiers in Oncology*. 2020; 10: 577530.
- [26] Gao J, Qiu X, Xi G, Liu H, Zhang F, Lv T, *et al.* Downregulation of GSDMD attenuates tumor proliferation *via* the intrinsic mitochondrial apoptotic pathway and inhibition of EGFR/Akt signaling and predicts a good prognosis in nonsmall cell lung cancer. *Oncology Reports*. 2018; 40: 1971–84.
- [27] Yang J, Song H, Chen L, Cao K, Zhang Y, Li Y, *et al.* Integrated analysis of microfibrillar-associated proteins reveals MFAP4 as a novel biomarker in human cancers. *Epigenomics*. 2019; 11: 1635–51.
- [28] Han P, Yang H, Li X, Wu J, Wang P, Liu D, *et al.* Identification of a novel cancer stemness-associated ceRNA Axis in lung adenocarcinoma *via* stemness indices analysis. *Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics*. 2021; 28: 715–729.

How to cite this article: Dan Song, Li Sun. Bioinformatics analysis identifies key genes of prognostic value in lung cancer. *Journal of Men's Health*. 2023; 19(7): 119-130. doi: 10.22514/jomh.2023.064.