RESEARCH ARTICLE

Construction and clinical significance of prognostic risk markers based on cancer driver genes in lung adenocarcinoma

Yazhou Su[1] · Tingting Huo[2] · Yanan Wang[2] · Jingyan Li[2]

Abstract

Background Cancer driver genes (CDGs) have been reported as key factors influencing the progression of lung adenocarcinoma (LUAD). However, the role of CDGs in LUAD prognosis has not been fully elucidated.

Methods LUAD transcriptome data and CDG-related data were obtained from public databases and literature. Differentially expressed CDGs (DE-CDGs) greatly associated with LUAD survival ($P < 0.05$) were identified to establish a prognostic model. In addition, immune analysis of high-risk (HR) and low-risk (LR) groups was conducted by utilizing the CIBERSORT and single sample gene set enrichment analysis (ssGSEA) algorithms to assess immune differences. Subsequently, mutation analysis was conducted using *maftools*. Finally, candidate drugs were identified using the CellMiner database.

Results 40 DE-CDGs significantly associated with LUAD survival and 11 DE-CDGs associated with prognosis were identified through screening. Regression analysis revealed that risk score can independently predict LUAD prognosis ($P < 0.05$). Immune landscape analysis revealed that compared to the HR group, the LR group had higher immune scores and high infiltration of various immune cells such as follicular helper B cells and T cells. Mutation landscape analysis demonstrated that missense mutation was the most common mutation type in both risk groups. Drug prediction analysis revealed strong correlations of fulvestrant, S-63845, sapacitabine, lomustine, BLU-667, SR16157, motesanib, AZD-9496, XK-469, dimethylfasudil, P-529, and imatinib with the model genes, suggesting their potential as candidate drugs targeting the model genes.

Conclusion This study identified 11 effective biomarkers, DE-CDGs, which can predict LUAD prognosis and explored the biological significance of CDGs in LUAD prognosis, immunotherapy, and treatment.

Keywords Lung adenocarcinoma · Cancer driver genes · Biomarkers · Tumor microenvironment · Mutation landscape · Drug sensitivity

Introduction

As a subtype of non-small cell lung cancer (NSCLC), lung adenocarcinoma (LUAD) primarily occurs in the bronchial mucosal epithelium of the lungs [1, 2]. With a high incidence and mortality rate, it has a younger onset age [3]. Though most LUAD patients do not exhibit obvious symptoms in the early stages, as the disease progresses, patients often experience symptoms such as dry cough, hemoptysis, chest pain, dyspnea, and weight loss [4, 5]. The etiology of LUAD is not yet clear, but extensive medical research has found that smoking, age, radon exposure, and environmental pollution are risk factors for LUAD [6]. Fiberoptic bronchoscopy, serum tumor marker examination, lung X-ray, and chest CT examination are common diagnostic measures for LUAD [7–10]. Currently, although progress has been made in detection methods and targeted therapies, most cases of LUAD are diagnosed at an advanced stage, with poor prognosis and a 5 year survival rate of less than 20% [11]. Therefore, identifying more potential biomarkers may aid in improving the prognosis of LUAD.

Mutations or abnormal expression of cancer driver genes (CDGs), including oncogenes and tumor suppressor genes, can drive tumor formation [12, 13]. Studies have pointed out that mutations in oncogenes can increase the proliferation

✉ Yazhou Su
  suyazhousu@163.com

1 Department of Thoracic Surgery, The First Affiliated Hospital of Xinxiang Medical University, 88 Jiankang Road, Weihui 453100, Henan province, China

2 Department of Oncology, The First Affiliated Hospital of Xinxiang Medical University, Weihui 453100, Henan province, China

and survival ability of tumor cells, while mutations in tumor suppressor genes can lead to uncontrolled cell proliferation [14, 15]. Currently, some known tumor driver genes (such as TP53, EGFR, and HER2, etc.) have been confirmed to be linked with the malignant progression of tumors. Akira Mogi et al. [16] revealed that the tumor suppressor gene TP53 frequently mutates in human cancers, leading to a poorer prognosis and relatively stronger resistance to chemotherapy and radiotherapy in NSCLC. EGFR mutations are the second most common oncogenic driver events in NSCLC, with most EGFR-mutated NSCLC patients exhibiting exon 19 deletions or L858R substitutions, which are considered predictive of sensitivity to EGFRTKI treatment and have significant implications for the treatment and prognosis of EGFR mutation subtypes [17, 18]. Alterations in human epidermal growth factor receptor 2 (HER2, or ERBB2) have been identified as oncogenic drivers and potential therapeutic targets in various cancers, including lung cancer (LC), breast cancer (BC), and metastatic urothelial carcinoma [19–21]. M Riudavets et al. [22] found in their study on NSCLC that HER2 activation occurs through three mechanisms: gene mutations, gene amplification, and protein overexpression, each with different implications and predictive outcomes. Therefore, the study suggested adopting different treatment approaches for different types of HER2 alterations to improve patients' survival outcomes [22]. Therefore, based on these analyses, targeting CDGs has the potential to alter the cancer progression of patients to varying degrees. However, the relationship between CDGs and LUAD has not been fully illuminated at present.

This study explored the biological significance of CDGs in LUAD patients based on differentially expressed CDGs (DE-CDGs), seeking potential biomarkers for LUAD and revealing prognosis, immune, and candidate drugs for LUAD. This project can deepen the understanding of the pathogenesis of LUAD from the perspective of CDGs. The flowchart of this study is displayed in Fig. 1.

Materials and methods

Acquisition of public data

The gene expression profiles and clinical data information (age, gender, tumor grade, TMN stage) of the training set LUAD samples were obtained from The Cancer Genome Atlas database (TCGA, https://portal.gdc.cancer.gov/), including 541 cancer samples and 59 normal samples. The GSE72094 (386 cancer samples) dataset was downloaded from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/) database as the validation set. Samples with complete survival information and survival time greater than 30 days were retained in both the training set and validation set for

subsequent analysis. A total of 568 CDGs were obtained from the literature [23] (Table S1).

Identification of DE-CDGs in LUAD

In the training set, differential expression analysis was carried out by utilizing the *edgeR* package with parameters of FDR < 0.05 and |log(FC)| > 1. Finally, the intersection of the differentially expressed genes (DEGs) obtained from differential expression analysis and the CDGs obtained from the literature resulted in DE-CDGs.

Construction and validation of DE-CDGs related prognostic model

To detect genes associated with LUAD survival in DE-CDGs. We used univariate Cox regression to screen for genes with P < 0.05 for further analysis. In LASSO regression analysis, we narrowed down the range of survival-related genes and adjusted the optimal parameter λ through tenfold cross-validation. Based on the optimal cutoff values calculated by the *survminer* R package, we performed multivariate Cox regression analysis on the selected genes to establish the final prognostic model. The formula is as follows:

$$Model = \sum Coefficient(gene) * Expression\ value(gene).$$

We calculated the sample risk scores for the training set and validation set using the model formula and analyzed the survival outcomes of patients with different risk scores. According to the median risk score, we grouped the patients into high-risk (HR) and low-risk (LR) groups. We compared the survival differences between different risk groups using K-M analysis. By plotting the ROC curve, we evaluated the sensitivity and specificity of the model. Finally, we drew a heatmap to display the differential gene expression between different risk groups.
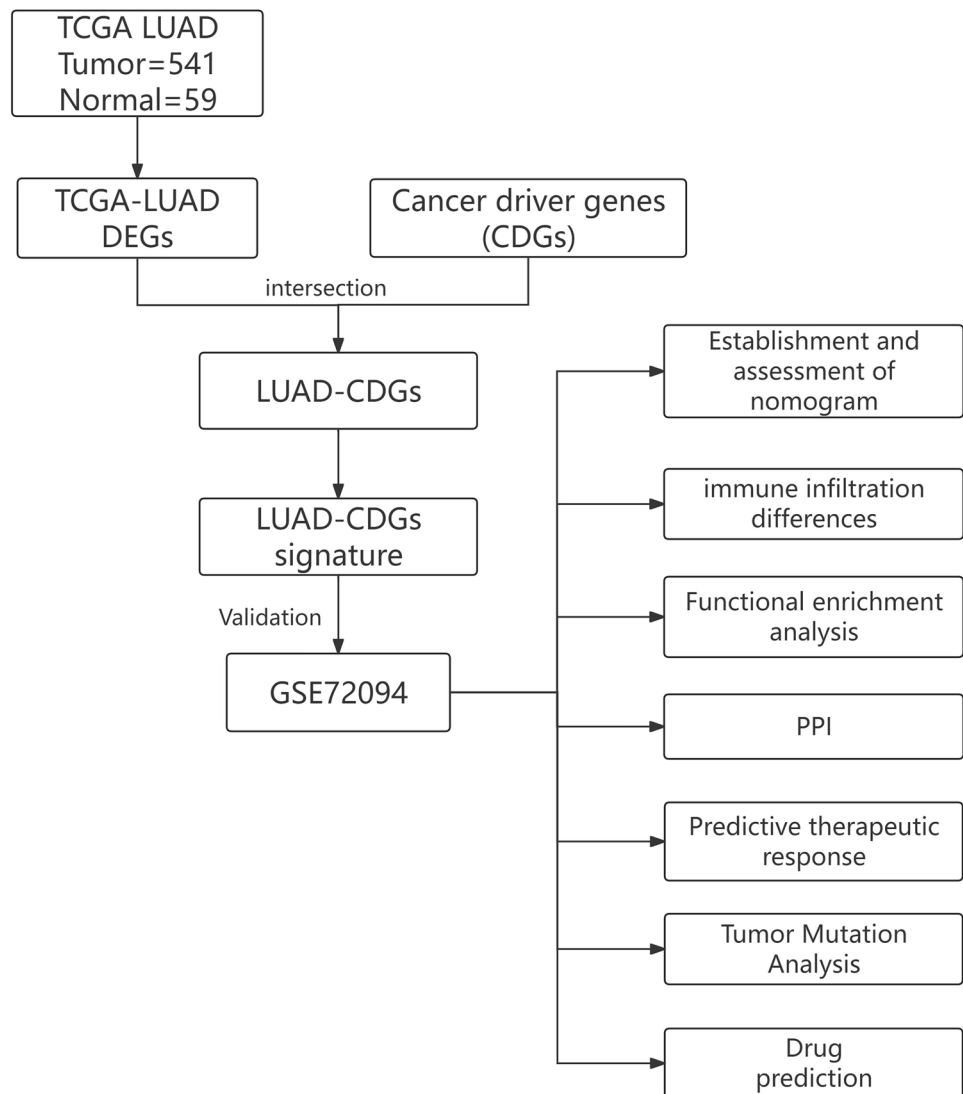
Independent prognostic analysis of prognostic model related to DE-CDGs

To select independent prognostic factors, we carried out univariate/multivariate Cox regression analyses. In addition, a prognostic nomogram was constructed to generate a calibration curve to evaluate the deviation between the nomogram and actual outcomes. Age, TNM stage, and model risk score were considered in the above analyses.

Identification and analysis of DEGs in HR and LR groups

Between the HR and LR groups, differential expression analysis was carried out by utilizing the *edgeR* package with parameters of FDR < 0.05 and |log (FC)| > 1. Subsequently,

Fig. 1 The flowchart


```
                    ┌─────────────────┐
                    │   TCGA LUAD     │
                    │   Tumor=541     │
                    │   Normal=59     │
                    └────────┬────────┘
                             │
                    ┌────────▼────────┐     ┌──────────────────┐
                    │   TCGA-LUAD     │     │ Cancer driver genes│
                    │     DEGs        │     │     (CDGs)       │
                    └─────────────────┘     └──────────────────┘
                             │  intersection
                    ┌────────▼────────┐
                    │    LUAD-CDGs    │
                    └────────┬────────┘
                             │
                    ┌────────▼────────┐
                    │    LUAD-CDGs    │
                    │    signature    │
                    └────────┬────────┘
                             │  Validation
                    ┌────────▼────────┐
                    │    GSE72094     │
                    └─────────────────┘
```

- Establishment and assessment of nomogram
- immune infiltration differences
- Functional enrichment analysis
- PPI
- Predictive therapeutic response
- Tumor Mutation Analysis
- Drug prediction

the DEGs protein–protein interaction (PPI) network was set up using the Search Tool for the Retrieval of Interacting Genes (STRING) database for the HR and LR groups. Interactions with confidence scores higher than 0.9 were selected as the basis for constructing the PPI network. KEGG and GO enrichment analyses of DEGs were performed by utilizing the *clusterProfiler* package, with *enrichplotR* package applied to visualize the results.

Immune landscape assessment and prediction of immune therapy response in the HR and LR groups

With the application of the estimate algorithm, we assessed the tumor microenvironment of each sample in the HR and LR groups, calculated the immune score, stromal score, and ESTIMATE score for each sample, performed the Wilcoxon test, and plotted violin plot for the HR and LR groups. We employed the CIBERSORT method and single sample gene set enrichment analysis (ssGSEA) method to respectively compare differences in immune-related cell infiltration and infiltration abundance related to immune function between the HR and LR groups.

To predict the immune therapy response of the HR and LR groups, the TIDE score was introduced. Furthermore, we performed the Wilcoxon test and plotted the violin plot. The immunophenoscore (IPS) of each patient was obtained from The Cancer Immunome Atlas (TCIA) database (https://tcia.at/home). We further investigated the differences in IPS between the HR and LR groups.

Tumor mutation analysis of the HR and LR groups

The *maftools* was employed to analyze and plot the mutation situation of HR and LR groups for LUAD SNV mutation data. By comparing and analyzing the similarities and

differences in mutation types, SNV class, and mutation rates between HR and LR groups, we selected the top ten genes with the highest mutation rates in the two different risk groups. Furthermore, we drew the waterfall plot displaying the mutation situation of model genes in the two groups.

Screening of potential drugs

We utilized the CellMiner database (https://discover.nci.nih.gov/cellminer/home.do) to dig out drugs related to the model genes. To visualize the results, the *ggplot2* R package was employed.

Cell culture

The cell lines used in this study mainly included the human LUAD cell line A549, Calu-3, NCI-H1975 and the human normal lung epithelial cell line (BEAS-2B). The cells were cultured in DMEM (Gibco, USA) supplemented with 10% fetal bovine serum (Gibco, USA) and 1% penicillin–streptomycin (Yeasen, China) and placed in an incubator at 37 °C and 5% CO_2.

Real-time fluorescence quantitative PCR (qRT-PCR)

First, the total RNA in the cells was extracted using the TRIzol reagent (Invitrogen, USA) according to the manufacturer's instructions. cDNA was then synthesized by reverse transcription using PrimeScript RT kit (Takara, Japan). qRT-PCR analysis was performed using SYBR Green PCR premix (Takara, Japan) on the Applied Biosystems 7500 sequence detection system (Applied Biosystem), and the relative mRNA expression was calculated using the $2^{-\Delta\Delta Ct}$ method. GAPDH was used as the internal parameter. The primer sequence is shown in Table 1.

Western blot

First, the total protein was extracted with RIPA lysis buffer (Beyotime, China) and quantified with BCA kit (Thermo Fisher Scientific, USA). The protein samples were then isolated by 10%SDS-PAGE and transferred to PVDF membranes. After sealing with 5% skim milk for 1 h, the membrane was incubated with primary antibody at 4 °C

overnight. After TBST washing for 3 times, goat anti-rabbit horseradish peroxidase secondary antibody (HRP, ab6721, 1:2000) was incubated at room temperature for 1 h. The blot was then observed with an enhanced chemiluminescence (ECL) solution. Protein expression was detected by ImageJ. Beta-actin was selected as the sample control. The primary antibodies are as follows: HOXD13 (ab19866, 1:1000, Abcam), FANCD2 (ab108928, 1:1000, Abcam), EGR2 (ab108399, 1:1000, Abcam), KLF4 (ab215036, 1:1000, Abcam).

Statistical analysis

SPSS 23.0 (USA) was used for data analysis. Data are expressed as mean ± standard deviation. The Student-t test was selected to compare the differences between the two groups. Univariate analysis of variance was used to evaluate the differences between groups. All experiments were repeated three times. $P < 0.05$ was considered statistically significant.
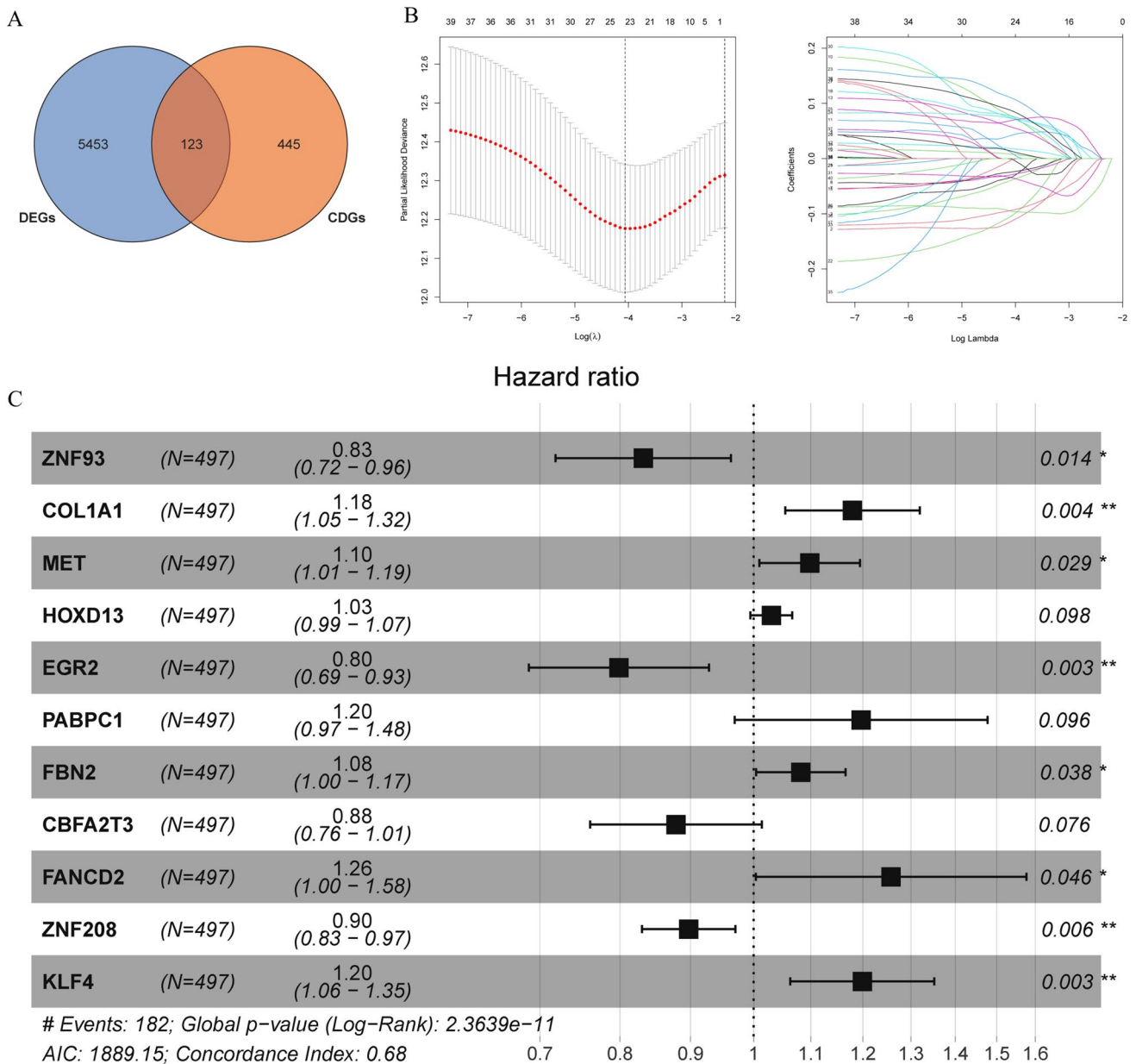
Results

Identification of DE-CDGs and construction and validation of the prognostic model

Through differential analysis, we obtained a total of 5576 DEGs related to LUAD. We intersected them with CDGs, obtaining 123 intersecting genes, namely 123 DE-CDGs (Fig. 2A). Subsequently, we obtained 40 genes strongly associated with LUAD survival through univariate Cox regression analysis (threshold: $P < 0.05$) (Table S2). Then, LASSO regression analysis was performed to eliminate genes that may have multiple collinearity (Fig. 2B). Finally, multivariate Cox regression analysis was carried out to determine the final prognostic model (Fig. 2C). The formula for the prognostic model is as follows:

Table 1 Primer sequences

| Gene	Forward	Reverse
HOXD13	5'-TCCTCTTCTGCCGTTGTAGC-3'	5'-GTAGTAGCCGTTGCCGAAGT-3'
FANCD2	5'-CGACGGCTTCTCGGAAGTAA-3'	5'-GTCGGAGGCTTGAAAGGACA-3'
EGR2	5'-AGAGAGTCAGTGGCAAATAGACAT-3'	5'-CTCCCTCGCTACCTGGAGT-3'
KLF4	5'-ATGCTCACCCCACCTTCTTC-3'	5'-GGTGGTCCGACCTGGAAAAT-3'
GAPDH	5'-AATGGGCAGCCGTTAGGAAA-3'	5'-GCGCCCAATACGACCAAATC-3'

A

B

Hazard ratio

C

ZNF93	(N=497)	0.83 (0.72 – 0.96)		0.014	*
COL1A1	(N=497)	1.18 (1.05 – 1.32)		0.004	**
MET	(N=497)	1.10 (1.01 – 1.19)		0.029	*
HOXD13	(N=497)	1.03 (0.99 – 1.07)		0.098	
EGR2	(N=497)	0.80 (0.69 – 0.93)		0.003	**
PABPC1	(N=497)	1.20 (0.97 – 1.48)		0.096	
FBN2	(N=497)	1.08 (1.00 – 1.17)		0.038	*
CBFA2T3	(N=497)	0.88 (0.76 – 1.01)		0.076	
FANCD2	(N=497)	1.26 (1.00 – 1.58)		0.046	*
ZNF208	(N=497)	0.90 (0.83 – 0.97)		0.006	**
KLF4	(N=497)	1.20 (1.06 – 1.35)		0.003	**

 # Events: 182; Global p−value (Log−Rank): 2.3639e−11
AIC: 1889.15; Concordance Index: 0.68

Fig. 2 A Venn diagram identified DE-CDGs. **B** Plot of LASSO regression analysis based on DE-CDGs. **C** Multivariate Cox regression analysis based on DE-CDGs. * $P < 0.05$, ** $P < 0.01$

$$Riskscore = -0.1842 * ZNF93 + 0.1648 * COL1A1$$
$$+ 0.0934 * MET + 0.0295 * HOXD13$$
$$- 0.2248 * EGR2 + 0.1791 * PABPC1$$
$$+ 0.0787 * FBN2 - 0.1296 * CBFA2T3$$
$$+ 0.2293 * FANCD2 - 0.1084 * ZNF208$$
$$+ 0.1812 * KLF4.$$

Based on the formula, we calculated the risk scores for all patients in the training set and validation set. Meanwhile, we recorded the clinical outcomes (survival/death) of all patients and compared the survival outcomes of patients in different risk groups, finding different risk scores for different LUAD patients in the training set and validation set. Patients with higher risk scores had higher mortality rates. The survival rate of the HR group (patients with risk scores higher than the median) was considerably lower than that of the LR group (patients with risk scores lower than the median) (Fig. 3A-D). ROC curve revealed that the prognostic model constructed in this study had good predictive performance in both the training set and validation set, with AUC values greater
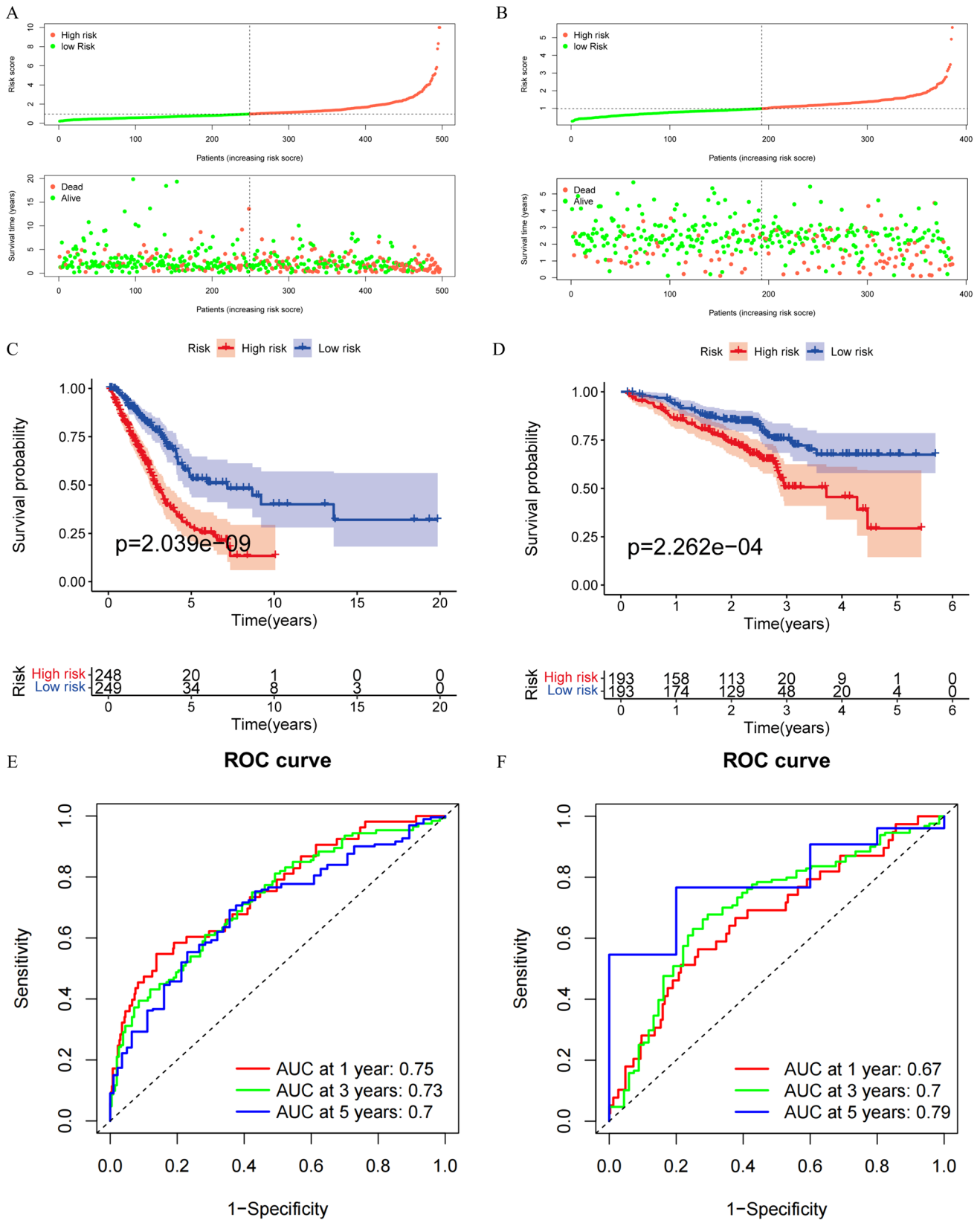
Fig. 3 **A-B** risk scores and survival outcomes of different patients in the training set **A** and the validation set **B**. **C-D** K–M curve of the training set **C** and the validation set **D**. **E–F** ROC curve of the training set **E** and the validation set **F**

than 0.67 for 1-year, 3-year, and 5-year survival (Fig. 3E-F). Finally, the results of the expression heatmap demonstrated that CBFA2T3, EGR2, ZNF93, and ZNF208 had elevated expression levels in the LR group, while KLF4, MET, COL1A1, FBN2, PABPC1, HOXD13, and FANCD2 had elevated expression levels in the HR group (Fig. 4). The good prediction performance of the model was also verified in the validation set GSE13213 (Supplementary Fig. 1).

Selection of independent prognosis factor in LUAD

The results of univariate and multivariate Cox regression analyses demonstrated that the T stage, N stage, and risk score can all serve as independent prognostic factors

for LUAD ($P < 0.05$) (Fig. 5A-B). Based on the age, gender, tumor stage, TNM stage, and risk score, the results of nomogram analysis allowed us to predict the 1-year, 3-year, and 5-year survival rates of LUAD patients (Fig. 5C). The calibration curves demonstrated a high consistency between the predicted 1-year, 3-year, and 5-year survival rates generated by our nomogram model and the actual survival rates of the patients in the dataset (Fig. 5D-F).

To further validate the predictive power of the model, we analyzed the association between the HR and LR groups and clinical features. The results showed that the risk score was significantly correlated with gender, stage, stage T, and stage N ($P < 0.05$, Fig. 6A). At the same time, to verify the predictive power of the prognostic model for patients with different clinical features, we performed subgroup survival analysis of the clinical features of the HR and LR group. Significant differences
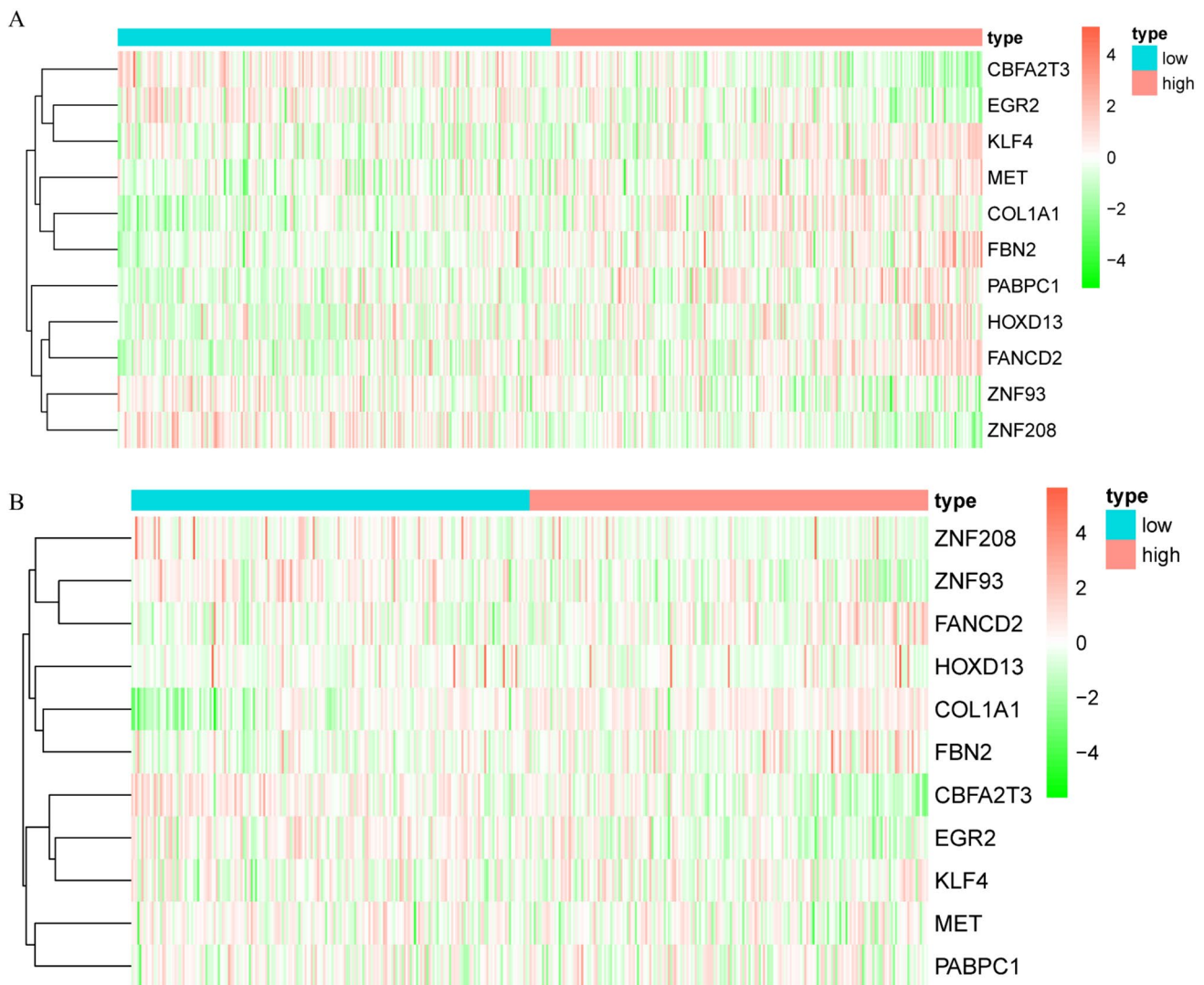
Fig. 4 **A-B** Heatmap of the gene expression levels of the model in the training set **A** and the validation set **B**
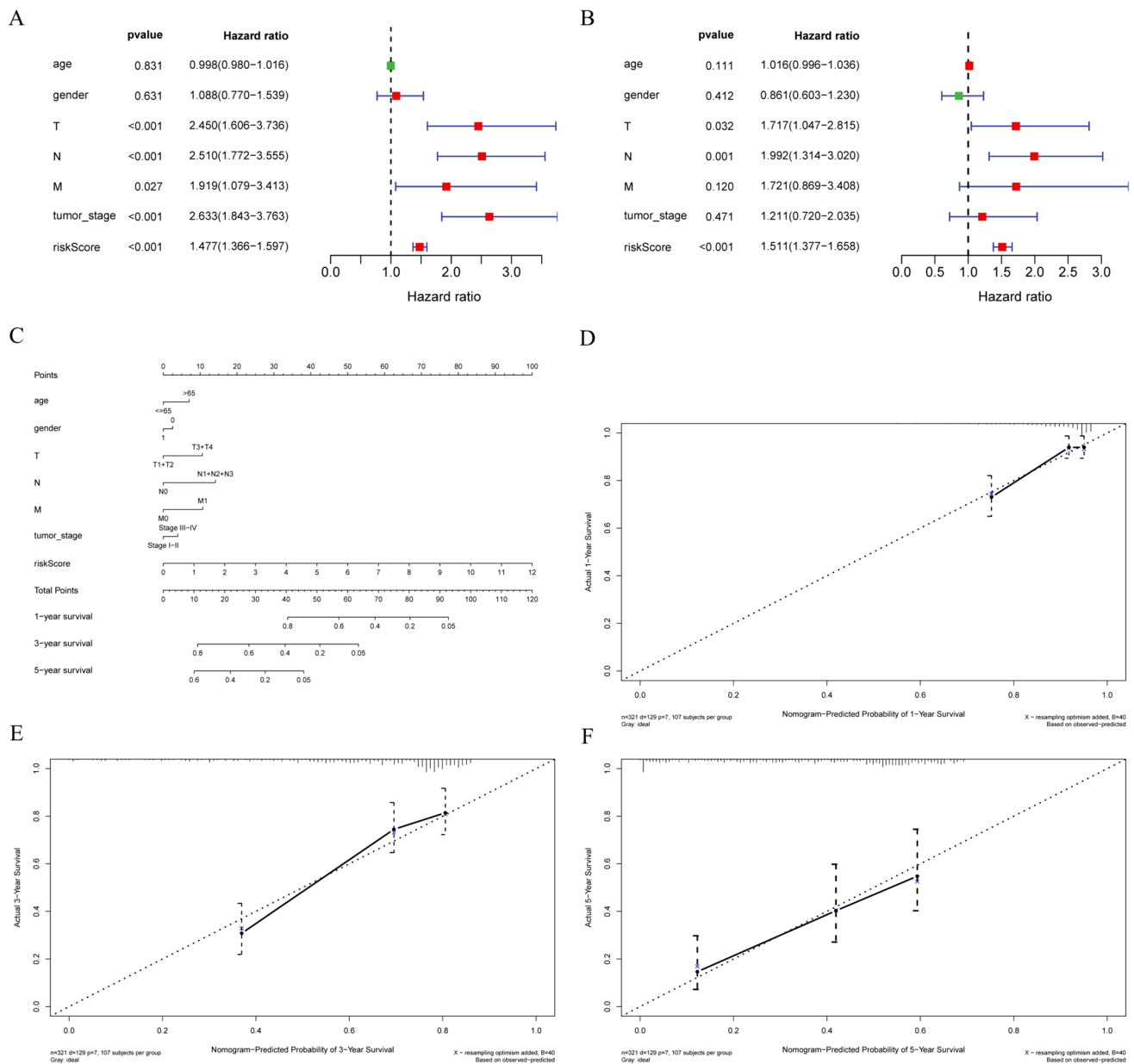
Fig. 5 **A-B** Univariate **A** and multivariate **B** Cox regression analyses identified the independent prognosis factor for LUAD. **C** Nomogram constructed based on age, gender, tumor stage, TNM staging, and risk score. **D-F** 1-year, 3-year, and 5-year calibration curves of the nomogram

in prognosis were observed for different age groups (≤ 65 years and > 65 years) (Fig. 6B), different gender groups (female and male) (6C), different stages (stage I-Π and stage Ш-IV) (Fig. 6D), different T stages (T1 + T2 and T3 + T4) (Fig. 6E), different N stages (N0 and N1) (Fig. 6F), and different M stages (M0 and M1) (Fig. 6G), which implied that the prognostic model of the present study had a good predictive and discriminatory ability in clinical practice.

Identification and analysis of DEGs between HR and LR groups

Firstly, differential expression analysis was performed on the data of the HR and LR groups using the training set data, and 481 differentially up-expressed genes and 461 differentially down-expressed genes were obtained, totaling 942 DEGs (Table S3). Based on the constructed PPI network using the STRING database, it was found that among the 942 DEGs, 938 DEGs have interactions with each other, forming 387 pairs of interacting relationships (Fig. 7A). In addition, GO
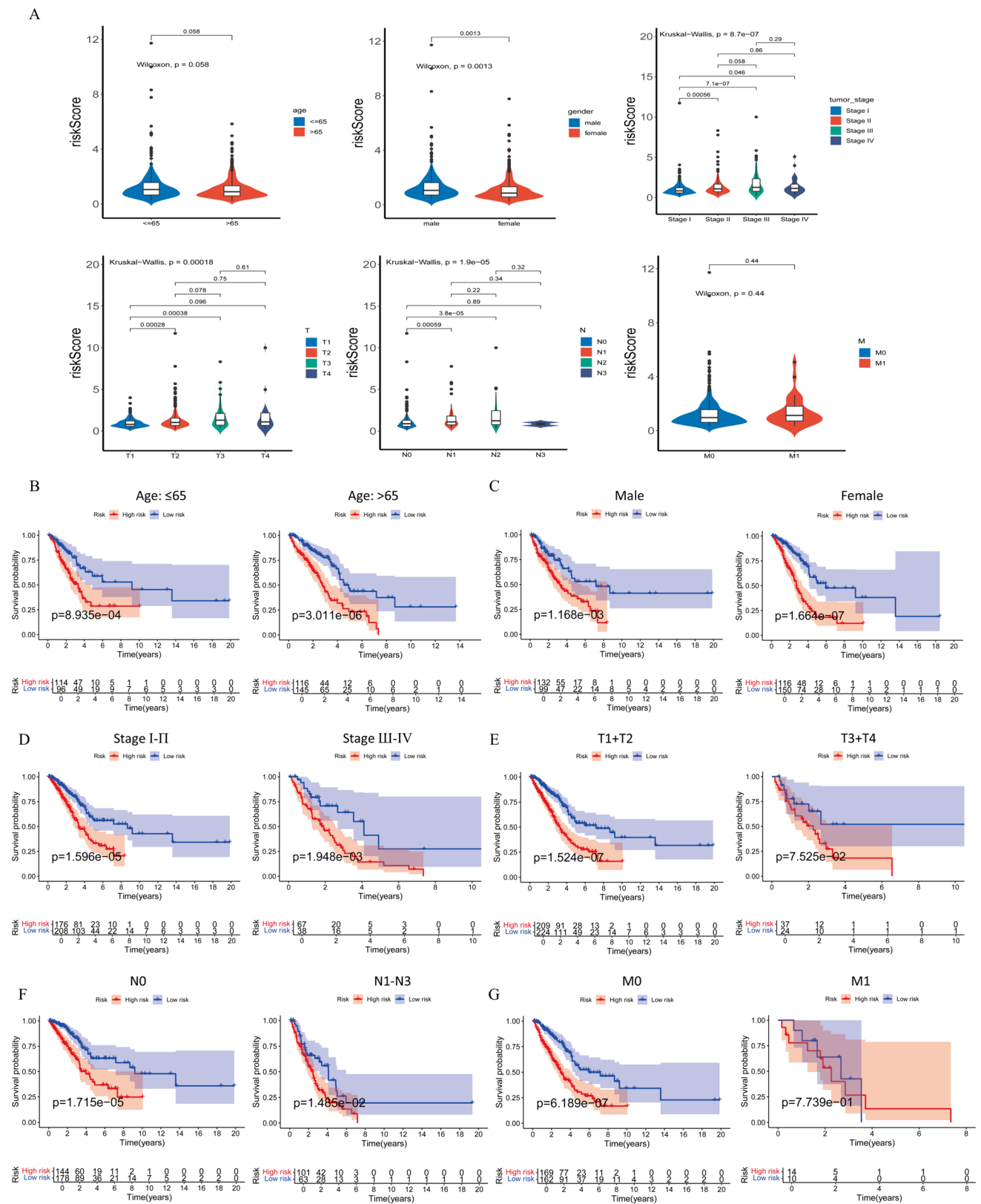
Fig. 6 **A** Correlation analysis between HR and LR groups and clinical data of LUAD patients. **B** K–M analysis for different age groups (65- and 65+years). **C** K–M analysis of gender-specific (male and female) groups. **D** K–M analysis for different stages (stage I–II and stage III–IV). **E** K–M analysis of different T-stages (T1+T2 and T3+T4). **F** K–M analysis for different N stages (N0 and N1). **G** K–M analysis of different M stages (M0 and M1)
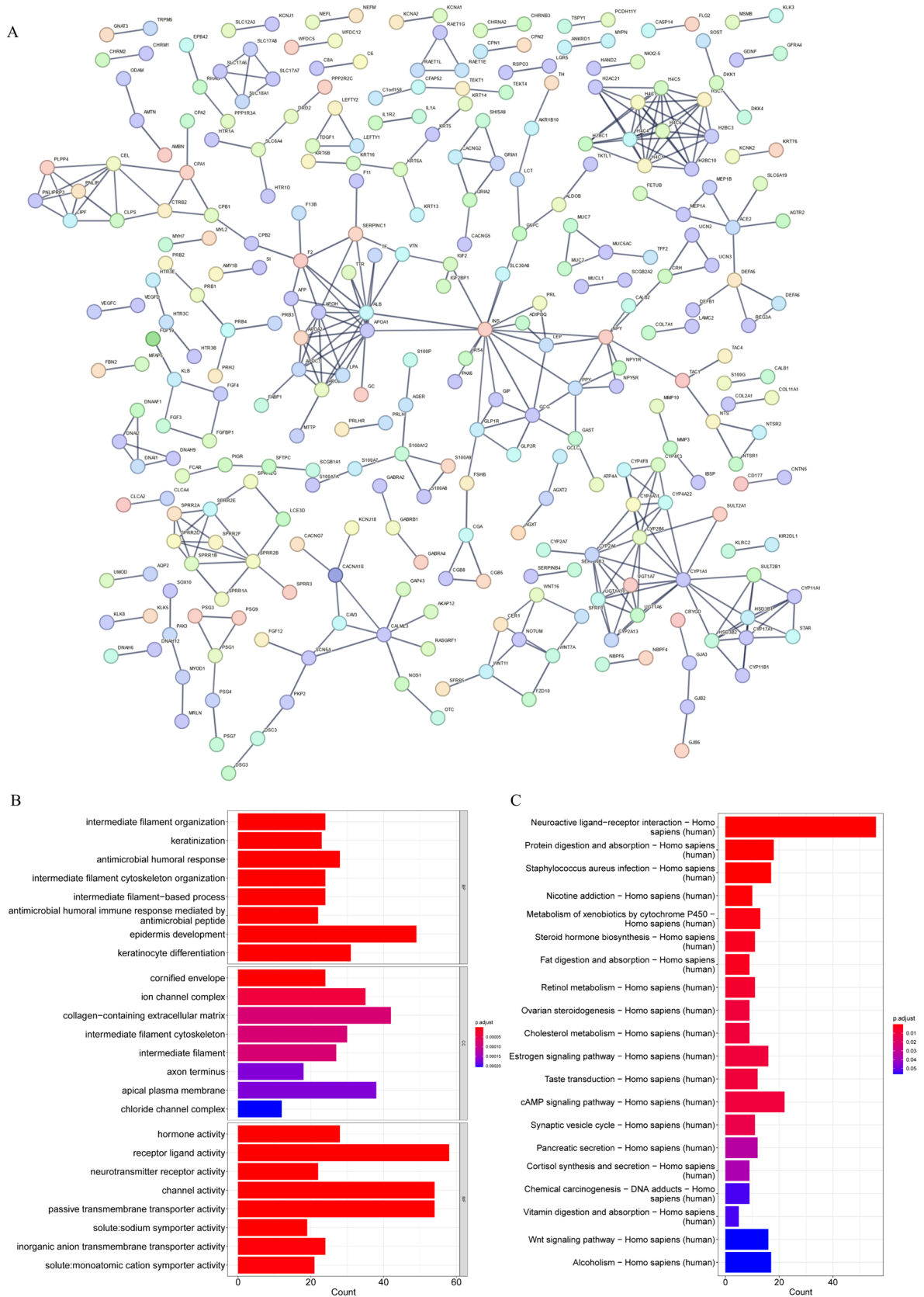
Fig. 7 **A** PPI network of DEGs between the HR and LR groups. **B**-**C** GO **B** and KEGG **C** Enrichment of DEGs between the HR and LR groups

enrichment analysis revealed that DEGs between the HR and LR groups may be associated with biological processes including antimicrobial humoral immune response mediated by antimicrobial peptides, epidermis development, and keratinocyte differentiation, as well as cellular component such as ion channel complex, collagen-containing extracellular matrix, and intermediate filament cytoskeleton. DEGs were also related to molecular functions such as hormone activity, receptor ligand activity, neurotransmitter receptor activity, channel activity, passive transmembrane transporter activity, and solute: sodium symporter activity (Fig. 7B). Furthermore, through KEGG enrichment analysis, these DEGs were revealed to be related to the regulation of signaling pathways such as neuroactive ligand–receptor interaction, Metabolism of xenobiotics by cytochrome P450, estrogen signaling pathway, cAMP signaling pathway, Wnt signaling pathway, etc. (Fig. 7C).

Evaluation of the difference in immune levels between the HR and LR groups and prediction of the possibility of immune therapy response

Based on the immune characteristic data in the training set, the score analysis of the main components in the immune microenvironment of the HR and LR groups showed that the LR group had higher immune score and ESTIMATE score than the HR group ($P < 0.05$) (Fig. 8A), indicating a potentially higher immune level. As evidenced by CIBERSORT, only CD4 memory T cells activated, macrophages M0, and neutrophils were highly infiltrated in the HR group, while other immune cells such as memory B cells, follicular helper T cells, and regulatory T cells (Tregs) were highly infiltrated in the LR group ($P < 0.05$) (Fig. 8B). ssGSEA revealed that the infiltration levels of immune cells such as dendritic cells (aDCs, iDCs, and pDCs), B cells, and mast cells were higher in the LR group. The abundance of immune functions such as APC co-stimulation, CCR, HLA, and T cell co-stimulation was also higher in the LR group, while the HR group had a lower abundance of immune cells and immune functions ($P < 0.05$) (Fig. 8C-D). In terms of IPS and TIDE scores, the LR group had higher IPS and lower TIDE scores, indicating that compared with the HR group, patients in the LR group may have a more sensitive immune response to immunotherapy and a reduced likelihood of immune escape ($P < 0.05$) (Fig. 8E-F).

Genetic mutation assessment of the HR and LR groups

Based on the training set data, analysis of the mutation landscape in the HR and LR groups revealed that missense mutation had the highest number of mutations in both groups, with the HR group having a greater number of missense

mutation than the LR group (Fig. 9A–C). Furthermore, statistical analysis of the gene mutation status in each sample of the HR and LR groups demonstrated that among the 242 samples in the LR group, 69 cases of gene mutation were observed, with FBN2, ZNF208, and COL1A1 having the highest mutation frequency (Fig. 9B). In the HR group, analysis of the 245 samples also revealed 69 cases of gene mutation, with FBN2, ZNF208, and MET being the three genes with the highest mutation frequency (Fig. 9D). In addition, based on the important role of driver genes in the development of LUAD, we further analyzed the correlation between risk score and common LUAD driver genes. The results showed that risk score was significantly negatively correlated with driver genes ROS1 and ALK, and significantly positively correlated with EGFR, MET, and NTRK2 ($P < 0.05$, Fig. 9E).

Prediction of candidate drugs and drug sensitivity analysis for LUAD

Fulvestrant, S-63845, sapacitabine, lomustine, BLU-667, SR16157, motesanib, AZD-9496, XK-469, dimethylfasudil, P-529, and imatinib are candidate drugs identified in this study that have significant correlations with model genes. Among them, CBFA2T3 is greatly positively correlated with fulvestrant (Cor = 0.587), sapacitabine (Cor = 0.568), SR16157 (Cor = 0.526), AZD-9496 (Cor = 0.517), XK-469 (Cor = 0.516), and imatinib (Cor = 0.502) ($P < 0.05$). MET is considerably negatively correlated with S-63845 (Cor = −0.570) and lomustine (Cor = −0.534) ($P < 0.05$). COL1A1 is significantly positively correlated with BLU-667 (Cor = 0.533), dimethylfasudil (Cor = 0.508), and P-529 (Cor = 0.504) ($P < 0.05$). EGR2 is significantly positively correlated with motesanib (Cor = 0.519) ($P < 0.05$) (Fig. 10). The box plot displayed that the MET high-expression group was more sensitive to S-63845 and lomustine, and the COL1A1 low-expression group was more sensitive to BLU-667 and dimethylfasudil ($P < 0.05$) (Fig. 11).

Verification of model gene expression levels

By referring to relevant literature, we selected model genes HOXD13, FANCD2, EGR2, and KLF4 to verify their expression levels. Firstly, qRT-PCR results showed that compared with the control group, the mRNA expression levels of HOXD13 and FANCD2 in LUAD cells were significantly higher ($P < 0.05$), while the expression levels of EGR2 and KLF4 were significantly lower ($P < 0.05$, Fig. 12A). Further, the trend of protein expression level shown by western blot was consistent with qRT-PCR (Fig. 12B).
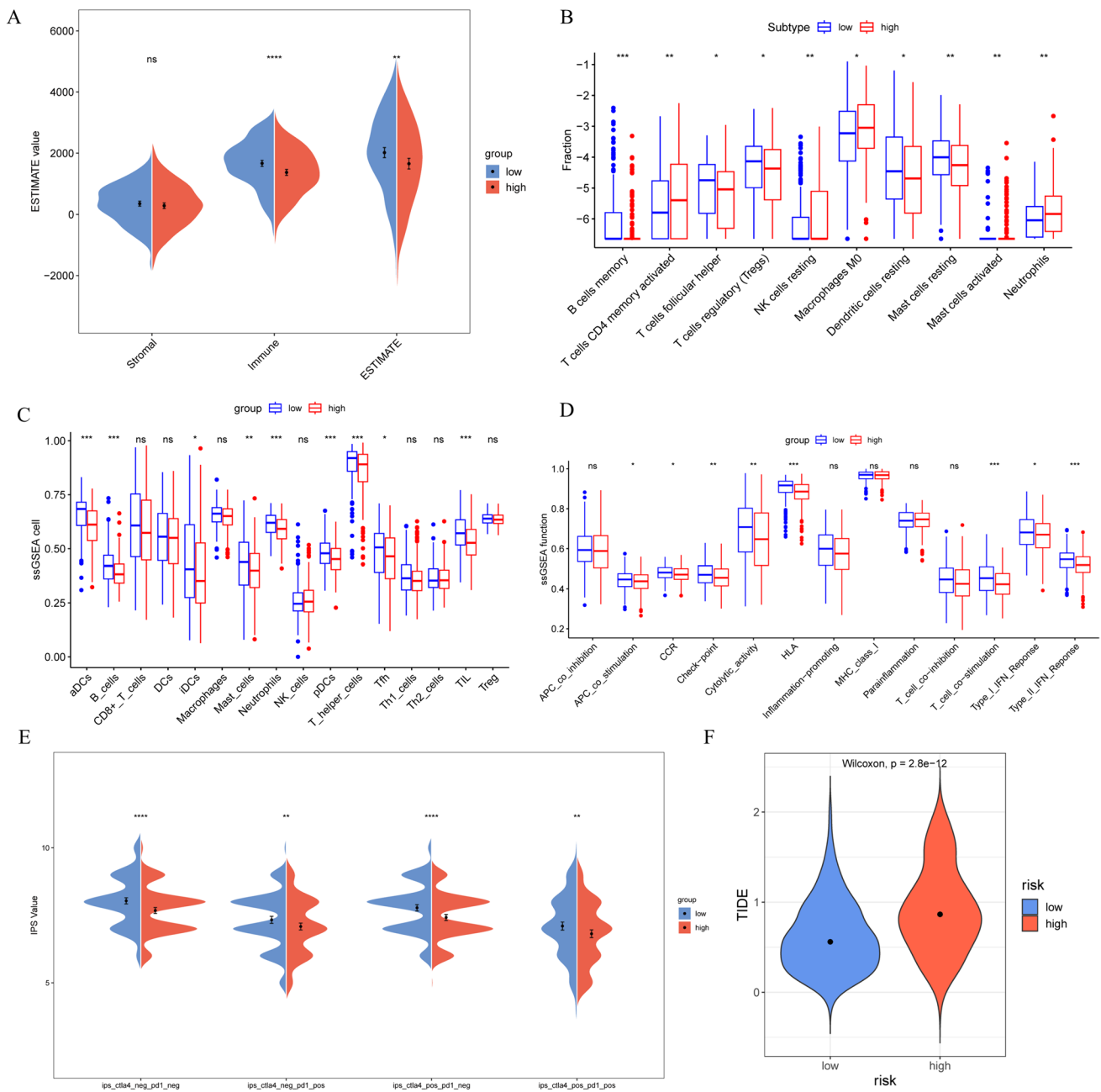
Fig. 8 **A** Scores of major components in the HR and LR group immune microenvironment. **B-D** Immune levels of the HR and LR groups shown by the CIBERSORT **B** and ssGSEA **C-D** algorithms. **E–F** IPS and TIDE scores of the HR and LR groups. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, ns $P > 0.05$

Discussion

LC is a common malignant tumor of the respiratory system, and CDGs are considered important factors affecting tumor progression and patient survival [24, 25]. Edouard Dantoing et al. [26] found that mutations in CDGs can alter tumor immune microenvironment and may promote resistance to PD1/PD-L1 in NSCLC. Therefore, targeting CDGs may be a promising alternative approach for

treating LUAD patients [26]. In addition, Zou et al. [23] also elucidated the biological role of CDGs in hepatocellular carcinoma survival and tumor immunity, resulting in multiple CDG-related biomarkers that can predict hepatocellular carcinoma prognosis, and found that these CDGs are linked with different immune cell infiltrations in the tumor microenvironment. However, the mechanism of action of CDGs in LUAD remains unexplored. Therefore, exploring the effects of CDGs on LUAD prognosis,
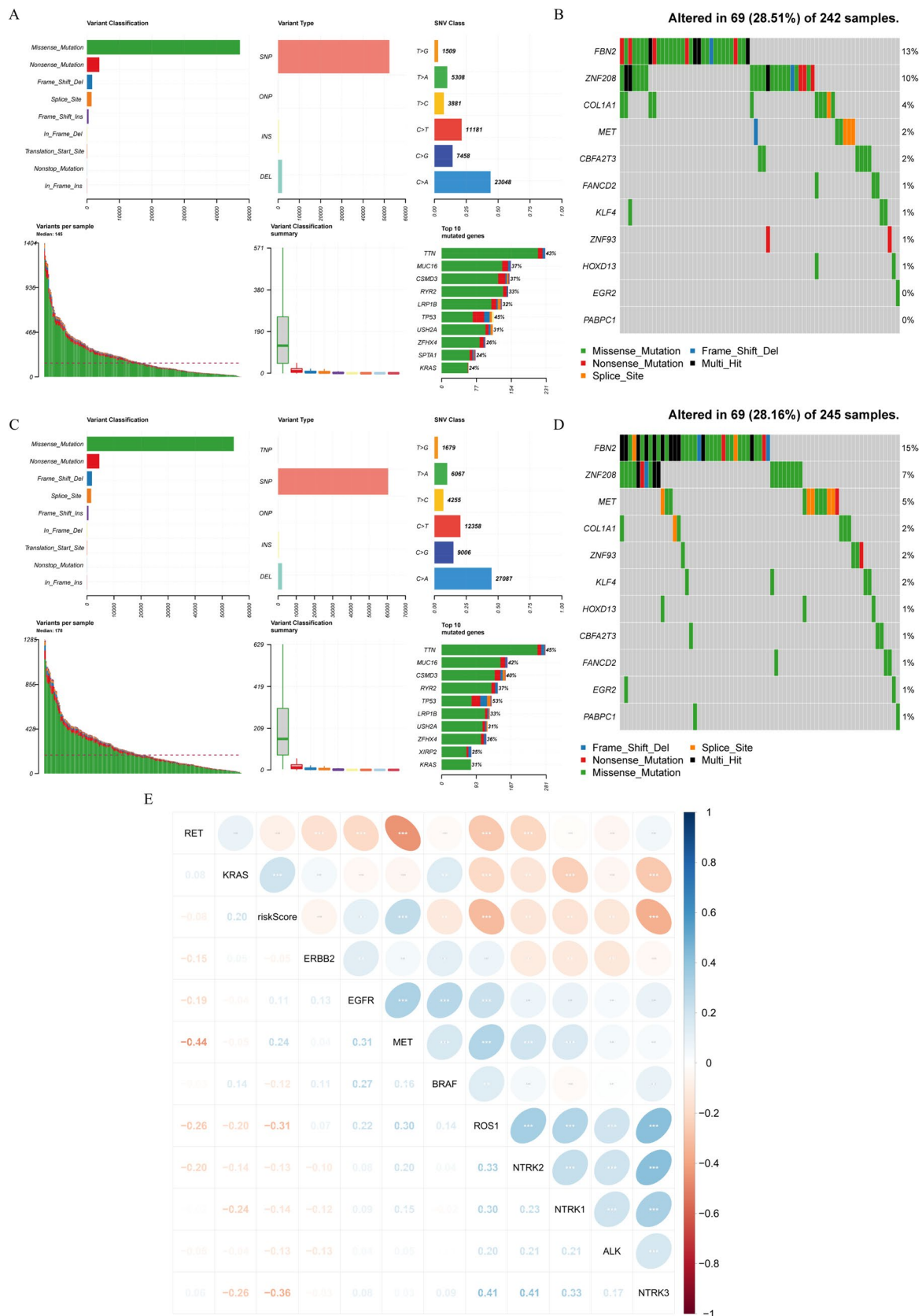
Fig. 9 **A** Mutation landscape of the LR group. **B** Gene mutation profiles of individual samples in the LR group. **C** Mutation landscape of the HR group. **D** Gene mutation profiles of individual samples in the HR group. **E** Correlation between the risk score and LUAD driver gene
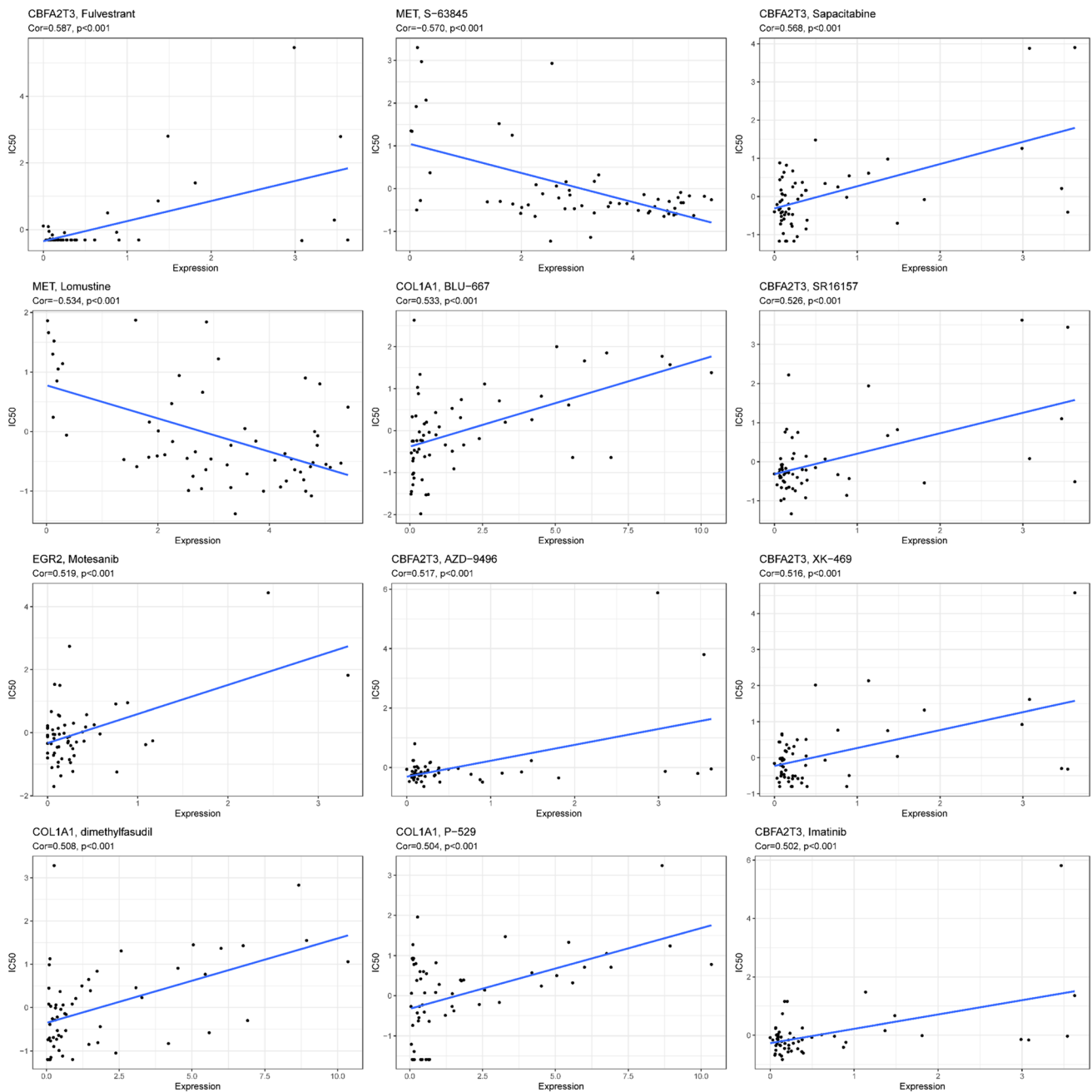
Fig. 10 Correlation analysis between model genes and drugs predicted by the CellMiner database

immunity, and treatment is of great biological significance. We herein investigate the relationship between CDGs and LUAD to reveal their association with the malignancy of LUAD and prognosis.

CBFA2T3, EGR2, ZNF93, ZNF208, KLF4, MET, COL1A1, FBN2, PABPC1, HOXD13, and FANCD2 were identified as 11 CDGs associated with LUAD prognosis in this study. Among them, CBFA2T3, EGR2, ZNF93, and ZNF208 showed elevated expression levels in the LR group, while KLF4, MET, COL1A1, FBN2, PABPC1,

HOXD13, and FANCD2 had elevated expression levels in the HR group. Previous studies have found that CBFA2T3 can facilitate the occurrence and development of tumors in different cancers. In LC, CBFA2T3 is a protective gene, with its high expression promoting the survival of LC patients [27]. In BC, research has found that CBFA2T3 acts as a transcriptional repressor when connected to the binding domain of GAL4 DNA, thus identifying it as a potential candidate gene for BC tumor suppression [28]. As a member of the zinc finger transcription factor family, EGR2
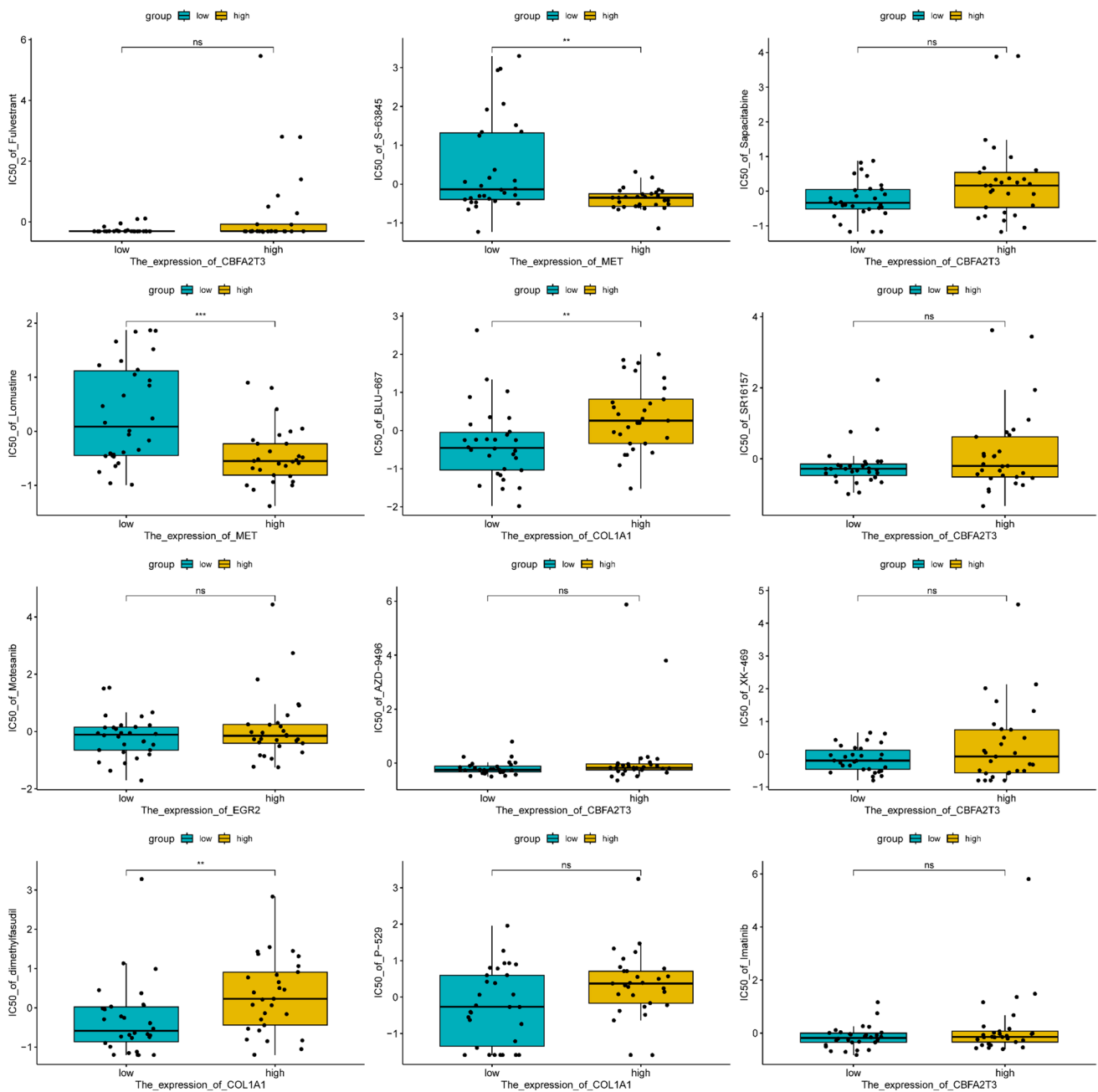
Fig. 11 Statistics of drug IC$_{50}$ values corresponding to groups with model genes high or low expressed

is considered an important regulatory factor for systemic autoimmunity [29]. Recent studies by Liu et al. [30] have unearthed that inhibiting the expression of EGR2 can promote the proliferation of NSCLC cells. ZNF93 and ZNF208 are also members of the zinc finger protein family [31, 32]. Currently, research in ovarian cancer (OC) has pointed out that ZNF93 can facilitate the migration and proliferation of OC cells. Moreover, high expression of ZNF93 is tightly linked to the clinical staging of patients, indicating poor prognosis [33]. In the exploration of LC risk factors, a strong association has been found between ZNF93 and susceptibility to LC [31]. Although the correlation of ZNF208 with LC has not been discovered yet, multiple studies have already found its close association with disease progression in various tumors such as laryngeal cancer, esophageal cancer, and pancreatic cancer [32, 34, 35]. KLF4 is a transcription factor that can regulate cell proliferation, differentiation, and self-renewal of stem cells. It mainly participates in suppressing the differentiation and proliferation of cancer cells in lung tumors, thus being considered an important

Fig. 12 The expression levels of HOXD13, FANCD2, EGR2, and KLF4 in LUAD cells were detected using **A** qRT-PCR and **B** western blot
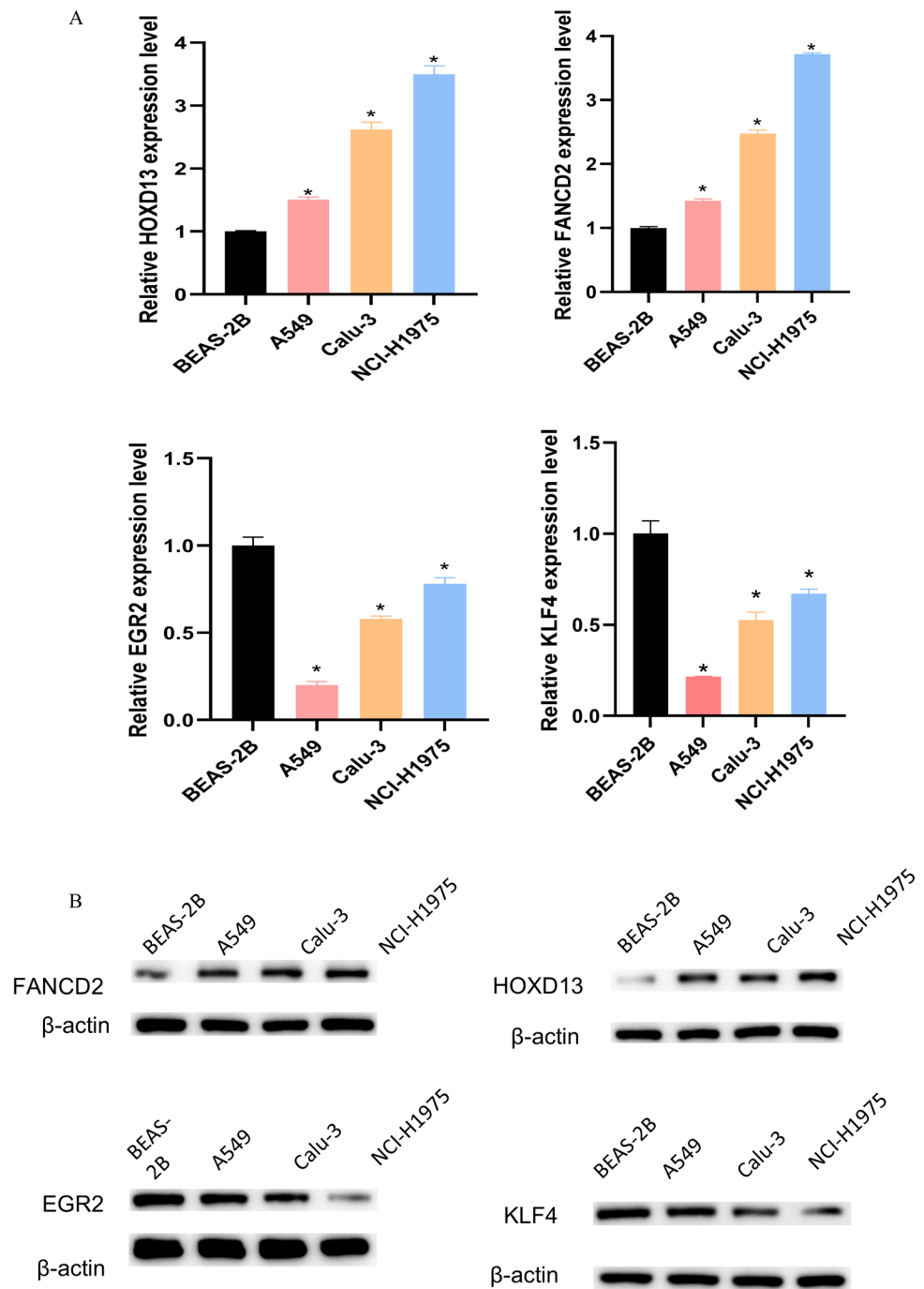
tumor suppressor [36]. Former studies have elucidated that the MET is overexpressed in many human cancers, including LC [37]. In preclinical and clinical studies of NSCLC, MET activation has been identified as a major oncogenic driver in a subset of LC, mediating malignant progression by influencing cancer cell invasion, survival, and growth [38]. Moreover, MET is also considered a secondary driver of acquired resistance to targeted therapy [38]. COL1A1 has been identified as a key biomarker and potential drug target in LC, with its expression exhibiting an obvious correlation

with overall survival (OS) and progression-free survival of LC patients [39]. Wang et al. [40] found through experimental research that elevated expression of COL1A1 can drive LUAD cells to grow, migrate, and invade. As a novel pathogenic gene, FBN2 has been found by Hong et al. [41] to be highly expressed in LC cells and to facilitate the proliferation, invasion, and migration of LC cells. More and more projects suggest that PABPC1 is aberrantly expressed in various tumor tissues and cancers, such as LC [42], gastric cancer [43], and esophageal squamous cell carcinoma [44],

etc. Li et al. [45] found in their study on LC that enhancing the ubiquitination of PABPC1 represses the proliferation of cancer cells. HOXD13 is a member of the HOX family and has the function of regulating organ development. Xu et al. [46] found that low expression of HOXD13 promotes the progression of prostate tumors. At present, there is relatively little investigation on the correlation between HOXD13 and the progression of LC, but Han et al. [47] have suggested that HOXD13 can be a promising prognosis biomarker for the diagnosis and treatment of LUAD. FANCD2 is considered a gene associated with autophagy-depended ferroptosis, with higher expression levels in LUAD patients compared to normal tissue specimens, and has been revealed to be associated with TNM staging advantages, lower chemotherapy sensitivity, and lower immune scores in LUAD [48]. Based on these analyses, 11 genes identified in this study are linked to the occurrence and progression of tumors, especially LC.

Next, through immune analysis, we identified immune landscape differences between different risk groups. The results revealed that the LR group had elevated immune scores compared to the HR group, with high infiltration of many immune cells such as follicular helper B cells and T cells. Studies have shown that both follicular helper B cells and T cells are important immune cells in the immune system, and their high infiltration in the tumor microenvironment has been proven to mediate tumor progression [49, 50]. According to the research by Tu et al. [51], the high infiltration of B cells in LUAD is linked with a higher OS of patients. In addition, Cui et al. [52] found that antigen-driven follicular helper B cells and CD4T cells collaborate to effectively facilitate the response of anti-tumor CD8$^+$ T cells. Therefore, we speculated that the highly infiltrated follicular helper B cells and T cells in the LR group may be associated with their better response to prognosis. Furthermore, the prediction of immune therapy response and immune escape revealed that the LR group has higher IPS scores and lower TIDE scores, revealing that LR patients may be more sensitive to immune therapy and less likely to experience immune escape, making them more prone to benefit from immune therapy compared to the HR group. However, due to the lack of corresponding experimental verification, in-depth research is still required regarding the actual benefits of immune therapy for the HR and LR groups.

Our study also identified some candidate drugs that are greatly related to the model genes, such as fulvestrant, S-63845, and aapacitabine. In addition, based on the comparison of IC$_{50}$ values, it is speculated that the MET high-expression group is more sensitive to S-63845 and lomustine, while the COL1A1 low-expression group is more sensitive to BLU-667 and dimethylfasudil. Despite the encouraging results of targeted therapy research in LC in recent years, most patients eventually develop resistance to targeted drugs, mainly due to changes in carcinogenic

drivers, the most common of which include epidermal growth factor receptor (EGFR), anaplastic lymphoma kinase (ALK), and TP53 mutations [53, 54]. This study is expected to explore as many driver genes as possible and to explore the prognosis, immunity, and drug candidates of LUAD by constructing a prognostic model to analyze the potential pathogenesis. In this study, we found that the risk score model was correlated with the common LUAD driver genes to varying degrees, which may provide certain reference and research direction for the drug resistance and off-target problems of targeted therapy in LUAD driver gene positive patients. For example, the model gene KLF4 is a controversial gene. As reported by Liu et al. [55], KLF4 is significantly underexpressed in cisplatin-resistant LC cell lines, and overexpression of KLF4 inhibits the viability, EMT process, and migration and invasion of drug-resistant cells, and promotes apoptosis. However, Zheng et al. [56] revealed that KLF4 overexpression caused by circUBAP2 dysregulation has a facilitating effect on NSCLC proliferation and chemotherapy resistance. In gefitinib-resistant NSCLC cells and tissues mediated by c-Met amplification, KLF4 is overexpressed and favors tumor progression [57]. At the same time, the study also showed that KLF4 can enhance gefitinib resistance by inhibiting β-catenin expression and interfering with β-catenin inhibition of c-Met phosphorylation to activate the c-Met/Akt signaling pathway. The carcinogenic or anticancer effect of KLF4 in NSCLC may be related to the subcellular localization of KLF4 [58]. For the recognized resistance driver gene MET, the TAGTON clinical trial demonstrated that osimertinib (EGFR-TKI) combined with volitinib (MET inhibitor) is a promising therapy for advanced NSCLC patients with MET-amplified/overexpressed EGFRm who have developed disease progression following previous EGFR-TKI screening. Therefore, we speculate that the combined action of model genes KLF4 and Met may be a potential mechanism leading to chemotherapy or targeted drug resistance to LUAD. The model gene COL1A1 has been reported to help solid tumors adapt to the hypoxic conditions of the tumor microenvironment, thereby promoting tumor aggressiveness and drug resistance [59]. At the same time, the drug resistance mechanism of COL1A1 in LUAD is usually carried out as a downstream regulatory gene, such as the regulatory axis miR-150/NOTCH3/COL1A1 [60], LINC00313/miR-218-5p/COL1A1 [61] and miR-29b-3p/COL1A1[62]. Therefore, the model genes and drugs screened in this study can provide not only new exploration ideas for the resistance mechanism of targeted therapy in LUAD patients, but also a reliable research direction for future drug research.

At present, certain limitations still exist in this study. First, the risk profile was constructed based on the TCGA-LUAD dataset, and clinical trials were needed to further validate the prognostic value of the risk score model and

the nomogram. Second, although the expression level of the model gene has been verified experimentally, its prognostic mechanism in LUAD remains to be further explored. Third, the prognostic model constructed in this study needs to be further explored in terms of drug resistance or off-target of LUAD. Our investigation identified the biological value of CDGs in LUAD prognosis, immune response, and treatment, proffering new prospects for future LUAD research.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest with the contents of this article.

Ethical approval Not applicable.

Consent to participate Not applicable.

References

1. Su L, Zhao J, Su H, Wang Y, Huang W, Jiang X, et al. CircRNAs in lung adenocarcinoma: diagnosis and therapy. Curr Gene Ther. 2022;22:15–22.
2. Yang L, Wang S, Gerber DE, Zhou Y, Xu F, Liu J, et al. Main bronchus location is a predictor for metastasis and prognosis in lung adenocarcinoma: a large cohort analysis. Lung Cancer. 2018;120:22–6.
3. Cronin KA, Scott S, Firth AU, Sung H, Henley SJ, Sherman RL, et al. Annual report to the nation on the status of cancer, part 1: national cancer statistics. Cancer. 2022;128:4251–84.
4. Succony L, Rassl DM, Barker AP, McCaughan FM, Rintoul RC. Adenocarcinoma spectrum lesions of the lung: detection, pathology and treatment strategies. Cancer Treat Rev. 2021;99:102237.
5. Thammakumpee K. Clinical manifestation and survival of patients with non-small cell lung cancer. J Med Assoc Thai. 2004;87:503–7.
6. de Groot P, Munden RF. Lung cancer epidemiology, risk factors, and prevention. Radiol Clin North Am. 2012;50:863–76.
7. Biciusca V, Popescu IAS, Trasca DM, Olteanu M, Stan IS, Durand P, et al. Diagnosis of lung cancer by flexible fiberoptic bronchoscopy: a descriptive study. Rom J Morphol Embryol. 2022;63:369–81.
8. Grunnet M, Sorensen JB. Carcinoembryonic antigen (CEA) as tumor marker in lung cancer. Lung Cancer. 2012;76:138–43.
9. Mirka H, Ferda J, Krakorova G, Vodicka J, Mukensnabl P, Topolcan O, et al. The use of CT pattern in differentiating non-invasive, minimally invasive and invasive variants of lung adenocarcinoma. Anticancer Res. 2021;41:4479–82.
10. Mzabi A, Benzarti W, Romdhane W, Baya W, Aissa S, Gargouri I, et al. Erythema nodosum revealing metastatic lung cancer. Eur J Case Rep Intern Med. 2021;8:002539.
11. Sun R, Hou Z, Zhang Y, Jiang B. Drug resistance mechanisms and progress in the treatment of EGFR-mutated lung adenocarcinoma. Oncol Lett. 2022;24:408.
12. Chandrashekar P, Ahmadinejad N, Wang J, Sekulic A, Egan JB, Asmann YW, et al. Somatic selection distinguishes oncogenes and tumor suppressor genes. Bioinformatics. 2020;36:1712–7.
13. Miles B, Tadi P. Genetics, Somatic Mutation. StatPearls. Treasure Island (FL) ineligible companies. Disclosure: Prasanna Tadi declares no relevant financial relationships with ineligible companies.2023.
14. Parsa N. Environmental factors inducing human cancers. Iran J Public Health. 2012;41:1–9.
15. Ghosh A, Stewart D, Matlashewski G. Regulation of human p53 activity and cell localization by alternative splicing. Mol Cell Biol. 2004;24:7987–97.
16. Mogi A, Kuwano H. TP53 mutations in nonsmall cell lung cancer. J Biomed Biotechnol. 2011;2011:583929.
17. Harrison PT, Vyse S, Huang PH. Rare epidermal growth factor receptor (EGFR) mutations in non-small cell lung cancer. Semin Cancer Biol. 2020;61:167–79.
18. Castellanos E, Feld E, Horn L. Driven by mutations: the predictive value of mutation subtype in egfr-mutated non-small cell lung cancer. J Thorac Oncol. 2017;12:612–23.
19. Li BT, Ross DS, Aisner DL, Chaft JE, Hsu M, Kako SL, et al. HER2 amplification and HER2 mutation are distinct molecular targets in lung cancers. J Thorac Oncol. 2016;11:414–9.
20. Bose R, Kavuri SM, Searleman AC, Shen W, Shen D, Koboldt DC, et al. Activating HER2 mutations in HER2 gene amplification negative breast cancer. Cancer Discov. 2013;3:224–37.
21. Patelli G, Zeppellini A, Spina F, Righetti E, Stabile S, Amatu A, et al. The evolving panorama of HER2-targeted treatments in metastatic urothelial cancer: a systematic review and future perspectives. Cancer Treat Rev. 2022;104:102351.
22. Riudavets M, Sullivan I, Abdayem P, Planchard D. Targeting HER2 in non-small-cell lung cancer (NSCLC): a glimpse of hope? an updated review on therapeutic strategies in NSCLC harbouring HER2 alterations. ESMO Open. 2021;6:100260.
23. Zou J, Qin W. Comprehensive analysis of the cancer driver genes constructs a seven-gene signature for prediction of survival and tumor immunity in hepatocellular carcinoma. Front Genet. 2022;13:937948.
24. Yi J, Wei X, Li X, Wan L, Dong J, Wang R. A genome-wide comprehensive analysis of alterations in driver genes in non-small-cell lung cancer. Anticancer Drugs. 2018;29:10–8.
25. Tan AC, Tan DSW. Targeted therapies for lung cancer patients with oncogenic driver molecular alterations. J Clin Oncol. 2022;40:611–25.
26. Dantoing E, Piton N, Salaun M, Thiberville L, Guisier F. Anti-PD1/PD-L1 immunotherapy for non-small cell lung cancer with actionable oncogenic driver mutations. Int J Mol Sci. 2021. https://doi.org/10.3390/ijms22126288.
27. Chen Y, Shen L, Chen B, Han X, Yu Y, Yuan X, et al. The predictive prognostic values of CBFA2T3, STX3, DENR, EGLN1, FUT4, and PCDH7 in lung cancer. Ann Transl Med. 2021;9:843.
28. Kochetkova M, McKenzie OL, Bais AJ, Martin JM, Secker GA, Seshadri R, et al. CBFA2T3 (MTG16) is a putative breast tumor

suppressor gene from the breast cancer loss of heterozygosity region at 16q24.3. Cancer Res. 2002;62:4599–604.

29. Morita K, Okamura T, Sumitomo S, Iwasaki Y, Fujio K, Yamamoto K. Emerging roles of Egr2 and Egr3 in the control of systemic autoimmunity. Rheumatology (Oxford). 2016;55:ii76–81.

30. Liu X, Shi H, Liu B, Li J, Liu Y, Yu B. miR-330-3p controls cell proliferation by targeting early growth response 2 in non-small-cell lung cancer. Acta Biochim Biophys Sin (Shanghai). 2015;47:431–40.

31. Liu Y, Lusk CM, Cho MH, Silverman EK, Qiao D, Zhang R, et al. Rare variants in known susceptibility loci and their contribution to risk of lung cancer. J Thorac Oncol. 2018;13:1483–95.

32. Zhang W, Shang S, Yang Y, Lu P, Wang T, Cui X, et al. Identification of DNA methylation-driven genes by integrative analysis of DNA methylation and transcriptome data in pancreatic adenocarcinoma. Exp Ther Med. 2020;19:2963–72.

33. Cui XX, Zhou C, Lu H, Han YL, Wang FM, Fan WR, et al. High expression of ZNF93 promotes proliferation and migration of ovarian cancer cells and relates to poor prognosis. Int J Clin Exp Pathol. 2020;13:944–53.

34. Wang S, Wen X, Zhao R, Bai Y. Genetic variation in the ZNF208 gene at rs8103163 and rs7248488 Is associated with laryngeal cancer in the northwestern chinese han male. Front Genet. 2022;13:813823.

35. Wang H, Yu J, Guo Y, Zhang Z, Liu G, Li J, et al. Genetic variants in the ZNF208 gene are associated with esophageal cancer in a Chinese Han population. Oncotarget. 2016;7:86829–35.

36. Yu T, Chen X, Zhang W, Liu J, Avdiushko R, Napier DL, et al. KLF4 regulates adult lung tumor-initiating cells and represses K-Ras-mediated lung cancer. Cell Death Differ. 2016;23:207–15.

37. Yang X, Liao HY, Zhang HH. Roles of MET in human cancer. Clin Chim Acta. 2022;525:69–83.

38. Drilon A, Cappuzzo F, Ou SI, Camidge DR. Targeting MET in lung cancer: Will Expectations Finally Be MET? J Thorac Oncol. 2017;12:15–26.

39. Geng Q, Shen Z, Li L, Zhao J. COL1A1 is a prognostic biomarker and correlated with immune infiltrates in lung cancer. PeerJ. 2021;9:e11145.

40. Wang Y, Mei X, Song W, Wang C, Qiu X. LncRNA LINC00511 promotes COL1A1-mediated proliferation and metastasis by sponging miR-126-5p/miR-218-5p in lung adenocarcinoma. BMC Pulm Med. 2022;22:272.

41. Hong Q, Li R, Zhang Y, Gu K. Fibrillin 2 gene knockdown inhibits invasion and migration of lung cancer cells. Cell Mol Biol (Noisy-le-grand). 2020;66:190–6.

42. Comtesse N, Keller A, Diesinger I, Bauer C, Kayser K, Huwer H, et al. Frequent overexpression of the genes FXR1, CLAPM1 and EIF4G located on amplicon 3q26-27 in squamous cell carcinoma of the lung. Int J Cancer. 2007;120:2538–44.

43. Zhu J, Ding H, Wang X, Lu Q. PABPC1 exerts carcinogenesis in gastric carcinoma by targeting miR-34c. Int J Clin Exp Pathol. 2015;8:3794–802.

44. Zhang Y, Chen C, Liu Z, Guo H, Lu W, Hu W, et al. PABPC1-induced stabilization of IFI27 mRNA promotes angiogenesis and malignant progression in esophageal squamous cell carcinoma through exosomal miRNA-21-5p. J Exp Clin Cancer Res. 2022;41:111.

45. Li K, Zheng X, Tang H, Zang YS, Zeng C, Liu X, et al. E3 ligase MKRN3 is a tumor suppressor regulating PABPC1 ubiquitination in non-small cell lung cancer. J Exp Med. 2021. https://doi.org/10.1084/jem.20210151.

46. Xu F, Shangguan X, Pan J, Yue Z, Shen K, Ji Y, et al. HOXD13 suppresses prostate cancer metastasis and BMP4-induced epithelial-mesenchymal transition by inhibiting SMAD1. Int J Cancer. 2021;148:3060–70.

47. Han L, Xu G, Xu C, Liu B, Liu D. Potential prognostic biomarkers identified by DNA methylation profiling analysis for patients with lung adenocarcinoma. Oncol Lett. 2018;15:3552–7.

48. Miao H, Ren Q, Li H, Zeng M, Chen D, Xu C, et al. Comprehensive analysis of the autophagy-dependent ferroptosis-related gene FANCD2 in lung adenocarcinoma. BMC Cancer. 2022;22:225.

49. Franchina DG, Grusdat M, Brenner D. B-Cell metabolic remodeling and cancer. Trends Cancer. 2018;4:138–50.

50. Gutierrez-Melo N, Baumjohann D. T follicular helper cells in cancer. Trends Cancer. 2023;9:309–25.

51. Tu YN, Tong WL, Callahan BM, Chobrutskiy BI, Blanck G. B-cell receptor recombinations in lung adenocarcinoma exome files correlate with a higher overall survival rate. Anticancer Res. 2020;40:2043–51.

52. Cui C, Wang J, Fagerberg E, Chen PM, Connolly KA, Damo M, et al. Neoantigen-driven B cell and CD4 T follicular helper cell collaboration promotes anti-tumor CD8 T cell responses. Cell. 2021;184(6101–18):e13.

53. O'Farrell H, Harbourne B, Kurlawala Z, Inoue Y, Nagelberg AL, Martinez VD, et al. Integrative genomic analyses identifies GGA2 as a cooperative driver of EGFR-mediated lung tumorigenesis. J Thorac Oncol. 2019;14:656–71.

54. Fan Z, Zhang Q, Feng L, Wang L, Zhou X, Han J, et al. Genomic landscape and prognosis of patients with TP53-mutated non-small cell lung cancer. Ann Transl Med. 2022;10:188.

55. Liu S, Yang H, Chen Y, He B, Chen Q. Kruppel-Like factor 4 enhances sensitivity of cisplatin to lung cancer cells and inhibits regulating epithelial-to-mesenchymal transition. Oncol Res. 2016;24:81–7.

56. Zheng G, Huang J, Chen W, You P, Ding Y, Tu P. circUBAP2 exacerbates malignant capabilities of NSCLC by targeting KLF4 through miR-3182 modulation. Aging (Albany NY). 2021;13:11083–95.

57. Feng W, Xie Q, Liu S, Ji Y, Li C, Wang C, et al. Kruppel-like factor 4 promotes c-Met amplification-mediated gefitinib resistance in non-small-cell lung cancer. Cancer Sci. 2018;109:1775–86.

58. Liu M, Li X, Peng KZ, Gao T, Cui Y, Ma N, et al. Subcellular localization of Klf4 in non-small cell lung cancer and its clinical significance. Biomed Pharmacother. 2018;99:480–5.

59. Januchowski R, Swierczewska M, Sterzynska K, Wojtowicz K, Nowicki M, Zabel M. Increased expression of several collagen genes is associated with drug resistance in ovarian cancer cell lines. J Cancer. 2016;7:1295–310.

60. Zhang Y, Chen B, Wang Y, Zhao Q, Wu W, Zhang P, et al. NOTCH3 overexpression and posttranscriptional regulation by miR-150 were associated with EGFR-TKI resistance in lung adenocarcinoma. Oncol Res. 2019;27:751–61.

61. Ding D, Xu C, Zhang J, Zhang Y, Xue L, Song J, et al. Revealing underlying regulatory mechanisms of LINC00313 in Osimertinib-resistant LUAD cells by ceRNA network analysis. Transl Oncol. 2024;43:101895.

62. Jia R, Wang C. MiR-29b-3p reverses cisplatin resistance by targeting COL1A1 in non-small-cell lung cancer A549/DDP cells. Cancer Manag Res. 2020;12:2559–66.