Heliyon 10 (2024) e28670

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

A novel tumor mutation-related long non-coding RNA signature for predicting overall survival and immunotherapy response in lung adenocarcinoma

Wenjie Chen [a,1], Chen Liao [b,1], Xudong Xiang [a], Heng Li [a], Qiang Wu [c], Wen Li [c], Qianli Ma [a], Nan Chen [a], Benchao Chen [a], Gaofeng Li [a,*]

[a] Department of Thoracic Surgery, The Third Affiliated Hospital of Kunming Medical University, Yunnan Cancer Hospital, Kunming, China
[b] Department of Gastroenterology, The Second Affiliated Hospital of Kunming Medical University, Kunming, China
[c] Lung Cancer Center, West China Hospital, Sichuan University, Chengdu, China

ARTICLE INFO

Keywords:
Tumor mutation
Immune checkpoint inhibitor
Prognosis
Long non-coding RNA
Lung adenocarcinoma

ABSTRACT

Background: Immunotherapy has changed the treatment landscape for lung cancer. This study aims to construct a tumor mutation-related model that combines long non-coding RNA (lncRNA) expression levels and tumor mutation levels in tumor genomes to detect the possibilities of the lncRNA signature as an indicator for predicting the prognosis and response to immunotherapy in lung adenocarcinoma (LUAD).

Methods: We downloaded the tumor mutation profiles and RNA-seq expression database of LUAD from The Cancer Genome Atlas (TCGA). Differentially expressed lncRNAs were extracted based on the cumulative number of mutations. Cox regression analyses were used to identify the prognostic lncRNA signature, and the prognostic value of the five selected lncRNAs was validated by using survival analysis and the receiver operating characteristic (ROC) curve. We used qPCR to validate the expression of five selected lncRNAs between human lung epithelial and human lung adenocarcinoma cell lines. The ImmuCellAI, immunophenoscore (IPS) scores and Tumor Immune Dysfunction and Exclusion (TIDE) analyses were used to predict the response to immunotherapy for this mutation related lncRNA signature.

Results: A total of 162 lncRNAs were detected among the differentially expressed lncRNAs between the Tumor mutational burden (TMB)-high group and the TMB-low group. Then, five lncRNAs (*PLAC4*, LINC01116, LINC02163, MIR223HG, FAM83A-AS1) were identified as tumor mutation-related candidates for constructing the prognostic prediction model. Kaplan–Meier curves showed that the overall survival of the low-risk group was significantly better than that of the high-risk group, and the results of the GSE50081 set were consistent. The expression levels of PD1, PD-L1 and CTLA4 in the low-risk group were higher than those in the high-risk group. The IPS scores and TIDE scores of patients in the low-risk group were significantly higher than those in the high-risk group.

* Corresponding author.
E-mail address: ligaofeng0097@163.com (G. Li).
[1] Equal contributors.

https://doi.org/10.1016/j.heliyon.2024.e28670
Received 3 August 2023; Received in revised form 20 March 2024; Accepted 21 March 2024
Available online 22 March 2024

Conclusion: Our findings demonstrated that the five lncRNAs (*PLAC4*, LINC01116, LINC02163, *MIR223HG*, *FAM83A-AS1*) were identified as candidates for constructing the tumor mutation-related model which may serve as an indicator of tumor mutation levels and have important implications for predicting the response to immunotherapy in LUAD.

1. Background

Lung cancer is one of the most frequent causes of cancer-related deaths worldwide [1]. The histological subtype of lung cancer is classified approximately 85% as non-small cell lung cancer (NSCLC) or 15% as small cell lung cancer (SCLC). Among NSCLC classifications, lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) are the most common subtypes [2]. Traditional treatments, including surgery, radiotherapy, and chemotherapy, have limited efficacy, especially in patients with advanced NSCLC [3]. Although targeted therapies for oncogenic driver mutations have been widely used in clinical practice and have unprecedently improved outcomes, targeted therapies are ineffective in most NSCLC patients due to the lack of the oncogenic driver mutations and acquired drug resistance [4,5].

Since a large clinical study of immunotherapy for NSCLC was first reported in 2015 [6], various immunotherapy strategies have been developed and immune checkpoint inhibitors (ICIs) have been established as a standard of care for most advanced NSCLC patients [7–9]. The expression of PD-L1 was shown to be an important biomarker of patient response to ICIs. In a phase III study (KEYNOTE-024), pembrolizumab has consistently been shown to be superior to chemotherapy in patients with PD-L1 >50% of tumor cells expressing PD-L1. The median OS was 26.3 months (95% CI, 18.3 to 40.4) in the pembrolizumab group and 13.4 months (95% CI, 9.4–18.3) in the chemotherapy group, and notably 31.9% of patients in the pembrolizumab group were estimated to reach 5-year survival compared to 16.3% in the chemotherapy group [10]. Patients with PD-L1 < 50% of tumor cells expressing PD-L1 also benefited from new combination strategies of immunotherapy and chemotherapy, but better OS and PFS of the pembrolizumab group were observed regardless of PD-L1 expression [11]. PD-L1 expression for predicting immunotherapy response is imperfect and it is vital to find out biomarkers to improve response prediction.

Tumor mutational burden (TMB), defined as the level of somatic mutations per megabase that reflected cancer mutation quantity, has been identified as a predictive biomarker for ICI in many solid tumors [12,13]. Multiple studies have demonstrated that TMB was associated with long-term survival and PFS benefit in ICIs treated NSCLC patients [14,15]. In a phase II study (Checkmate-568), Nivolumab plus low-dose ipilimumab was effective as first-line treatment for advanced or metastatic NSCLC, with significantly higher PFS and ORR in patients with high TMB compared to low TMB regardless of PD-L1 expression [16]. Given that it is difficult to obtain sufficient tissue for molecular testing in advanced patients, a retrospective analysis demonstrated that blood TMB was a clinically actionable biomarker for patients treated with atezolizumab in second-line and higher NSCLC with significantly higher PFS in patients with high blood TMB compared to patients with low blood TMB [17]. Both PD-L1 expression and TMB appear to be reliable biomarkers for predicting ICI response, but there are still some controversies without consensus, such as different PD-L1 antibodies used in different clinical trials and different sizes of the target panel used for detecting TMB in different clinical trials.

long non-coding RNA (lncRNA) are defined as a type of RNA that are more than 200 nucleotides and not translated into protein [18]. In recent years, lncRNAs have been identified to play important roles in regulating the survival, proliferation and invasion of malignant cells [19].

In the immune system, lncRNAs have also been defined as critical roles in regulating immune responses, including direct and indirect interactions with chromatin, RNA, and proteins [20]. For example, the lncRNA known as 'Metastasis-Associated Lung Adenocarcinoma Transcript 1' (MALAT1) interacts with miR-195 to promote tumorigenesis and immune escape in diffuse large B cell lymphoma by up-regulating PD-L1 expression [21]. Long non-coding RNA 'NKILA' is known as a nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) interacting lncRNA that could increase the sensitivity of T cell to activation-induced cell death by inhibiting NF-κB activity [22]. Moreover, lncRNAs could be identified in cells and in biological fluids such as serum, plasma, urine, and saliva, so that lncRNAs have been used as diagnostic and prognostic biomarkers for various malignancies [23]. Although lncRNAs play crucial roles in vital cell activities, the specific functions of many lncRNAs are still not fully understood.

Numerous lncRNAs exhibit intricate regulatory functions, and gaining a comprehensive understanding of their interactions with immune-related pathways can shed light on their potential as therapeutic targets [24,25]. Elucidating how lncRNAs regulate immune responses, including the activation or suppression of immune cells and the modulation of immune checkpoint pathways, is an area requiring further exploration. The validation of candidate lncRNA biomarkers across diverse patient populations and various cancer types is imperative.

In this study, we sought to construct a signature combining lncRNA expression levels and mutation levels in tumor genomes to detect the possibilities of the lncRNA signature as an indicator of tumor mutation levels, thereby improving their prognostic value and predicting the responses to ICIs.

2. Material and methods

2.1. Data collection

All the somatic mutation information, RNA-seq expression profiles and relevant clinical data such as pathological factors and the survival outcome of the lung adenocarcinoma were downloaded from The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/repository), including 497 cancer cases, 54 cases of adjacent normal tissues and 486 cases of relevant clinical data. The standardized workflow type was HTSeq-FPKM [26]. HTSeq is a Python framework to work with high-throughput sequencing data. It provides tools to read, write, and manipulate sequencing data, and is commonly used for tasks such as sequences aligning to genes or other features in a genome. FPKM stands for "Fragments Per Kilobase of transcript per Million mapped reads". It is a method to quantify gene expression from RNA-Seq data. The mutation profile of TCGA lung adenocarcinoma samples was analyzed by using the R package 'maftool [27]'. Independent validation sets GSE50081 with 128 LUAD samples and common clinicopathological characteristics were obtained from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE50081). The annotation of both lncRNAs and mRNAs in TCGA and GEO was conducted using the Genome Reference Consortium Human Build 38 (GRCh38) assembly, and the data are available for download from the GENCODE website.

2.2. Identification of tumor mutation-related lncRNAs

To identify tumor mutation-related lncRNAs, we calculated the cumulative number of somatic mutations in each patient and ranked them in descending order, the top 25% of patients in the ranking were defined as the TMB-high group, and the last 25% of patients in the ranking were defined as the TMB-low group. The differentially expressed lncRNAs between TMB-high group and TMB-low group were screened by using R package 'limma', | Log (fold change) |$\geq$ 1.0 and p-value $<$0.05 were set as the cutoff. The R package 'heatmap' and 'ggplot2' was performed to visualize the differentially expressed lncRNAs.

2.3. Functional enrichment analysis

To predict the potential functions of tumor mutation-related lncRNAs, the correlation between the paired expression of lncRNAs and mRNAs was measured by calculating the Pearson correlation coefficient, and the top 10 mRNAs were considered as co-expressed lncRNA-related mRNAs. GO and KEGG analysis of co-expressed lncRNA-related mRNAs were performed by using the R package 'clusterProfiler', the FDR $<$0.05 and p-value $<$0.05 were set as the cutoffs.

2.4. Construction of tumor mutation-related prediction model

Significant variables identified in the univariate Cox regression analysis were included in the multivariate Cox regression analysis. This process is employed to assess the independent prognostic value of each variable while adjusting for potential confounding factors, p-value $<$0.05 were set as the cutoff [28,29]. We set the multivariate Cox risk regression coefficient as the β, and we used the formula (risk score = β_{gene} [1] $\times$ Expression $_{\text{gene}}$ [1] +β_{gene} [2] $\times$ Expression $_{\text{gene}}$ [2] + … +$\beta_{\text{gene [n]}}$ $\times$ Expression $_{\text{gene [n]}}$) to calculate the risk score of each case. After ranking the risk scores, those with a risk score higher than the median risk score was classified as the high-risk group, and the rest were classified as the low-risk group. A nomogram was constructed to predict the prognosis of TCGA lung adenocarcinoma samples, and the accuracy was accessed by using calibration curves. Overall survival is defined as the years from the date of diagnosis until the date of death from any cause [30]. Overall survival is the endpoint of the prediction model.

2.5. Validation of the prognostic value of the tumor mutation-related model

The overall survival analysis was performed by using Kaplan-Meier curves, with the difference in the survival curves assessed through the log-rank test. A significance level of p-value $<$0.05 was considered statistically significant. The accuracy of the analysis was evaluated using receiver operating characteristic (ROC) curves. External validation set GSE50081 lung adenocarcinoma data set was used to evaluate the prognostic value of the model. Then, we compared the expression levels of immune checkpoints in the high-risk score group and the low-risk score group.

2.6. Prediction of immunotherapy efficacy in the tumor mutation-related model

To estimate the abundance of infiltrating immune cells and differences between the high- and low-risk groups, we used the Immune Cell Abundance Identifier (ImmuCellAI, http://bioinfo.life.hust.edu.cn/ImmuCellAI) [31]. The prediction of the efficacy of immune checkpoint inhibitors utilized the immunophenoscore (IPS), which was a superior predictor of response to anti-CTLA4 and anti-PD-1 antibodies in two independent validation cohorts from The Cancer Immunome Atlas (https://tcia.at/) [32]. Tumor Immune Dysfunction and Exclusion (TIDE, http://tide.dfci.harvard.edu) [33,34] was used to evaluate the potential of tumor immune escape and the response to immune therapy.

Table 1
The primer sequences for target genes.

Genes	Forward primer	Reverse primer
GAPDH	5′- CCCATCACCATCTTCCAGG -3′	5′- CATCACGCCACAGTTTCCC -3′
PLAC4	5′- CTTGGGTTTTCTGTTGTTGCT -3′	5′- AGTGTGCCGTTTATGGTATTG -3′
LINC01116	5′- CTAACCTACCTGCAAGGAGAG -3′	5′- AGGACCATTAAATGGATCAAC -3′
LINC02163	5′- TGGAAATGAGGCCAGACTG -3′	5′- GCAAGAGAGAGGGAGCAAA -3′
FAM83A-AS1	5′- CTAAATTGATTTCACACCCCGC -3′	5′- TTCTTCTCTGTTGCTTTCCTGG -3′
MIR223HG	5′- TGTCAAGGGCTGGAGGAAGA -3′	5′- ATGGCTGGTTGGGGAAAGTA -3′

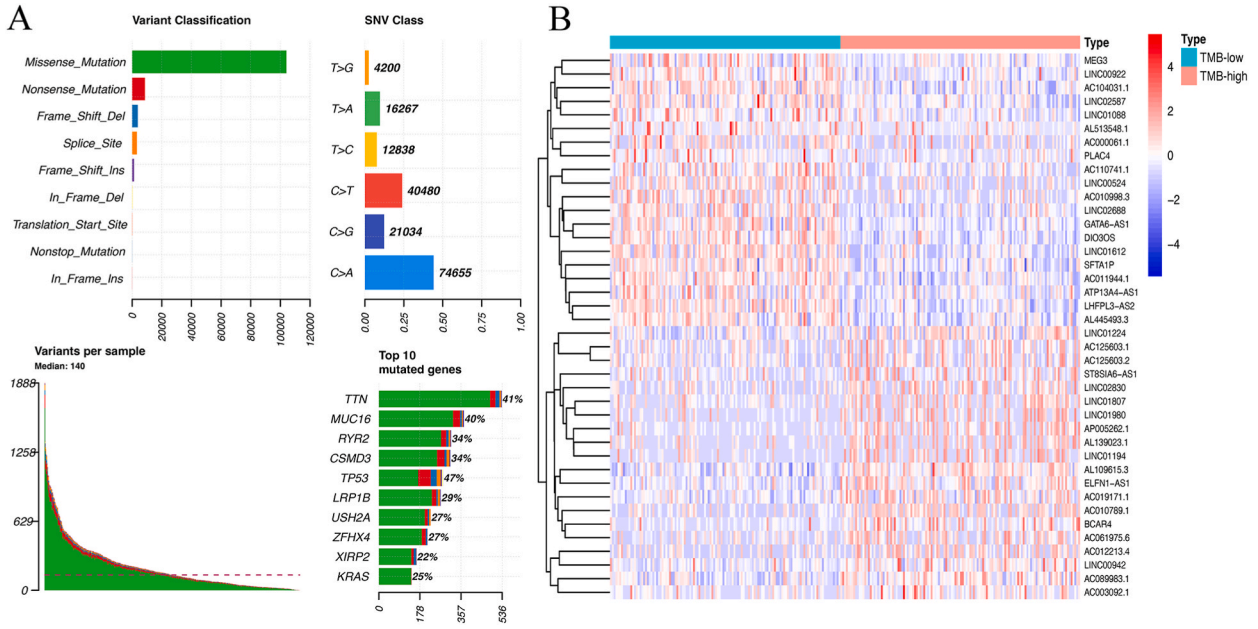


Fig. 1. Mutation profile of TCGA lung adenocarcinoma samples and differentially expressed lncRNAs between the TMB-low group and TMB-high group. (A) Mutation profile of TCGA lung adenocarcinoma samples. (B) The heatmap of the top 40 differentially expressed lncRNAs between the TMB-low group and TMB-high group (log fold change >1.0, *p*-value <0.05).

2.7. Cell culture

Human non-tumorigenic lung epithelial cell BEAS-2B as well as five human lung adenocarcinoma cancer cell lines H1299, A549, SPCA1, H358, and H1650 were obtained from the Department of Tumor Institute (Yunnan Cancer Hospital, Kunming, China) and cultured in Dulbecco's modified Eagle's medium (DMEM, HyClone), which contained 10% fetal bovine serum (FBS, Gibco). All cells were cultured at 37 °C in an incubator supplied with 5% CO2.

2.8. RNA extraction and real-time PCR assay

Total RNA was isolated from cells using TRIzol reagent (Ambion, Austin, TX, USA). Complementary DNA for reverse transcription was synthesized using SweScript RT I First Strand cDNA Synthesis Kit (Servicebio, Wuhan, China). Quantitative Real-Time PCR was performed using 2 × Universal Blue SYBR Green qPCR Master Mix (Servicebio, Wuhan, China). Human GAPDH was selected as a normalization standard. The primer sequences of the target lncRNAs and reference gene used for real-time PCR are listed in Table 1. Fold changes in the expression of the target lncRNAs between multiple samples were calculated using the two-sample *t*-test based on the $2^{-\Delta\Delta Ct}$ method, *p*-value <0.05 was considered as statistically significant. These analyses were performed by GraphPad Prism 8.0 software. Bonferroni correction method was used for multiple comparisons.

3. Result

3.1. Identification and functional enrichment analysis of differentially expressed lncRNAs associated with tumor mutations in LUAD

To identify lncRNAs associated with tumor mutation in LUAD, we obtained the mutation information and RNA-seq expression profiles of lung adenocarcinomas from TCGA. The summary of somatic mutation profile in 499 samples was showed in Fig. 1 A.

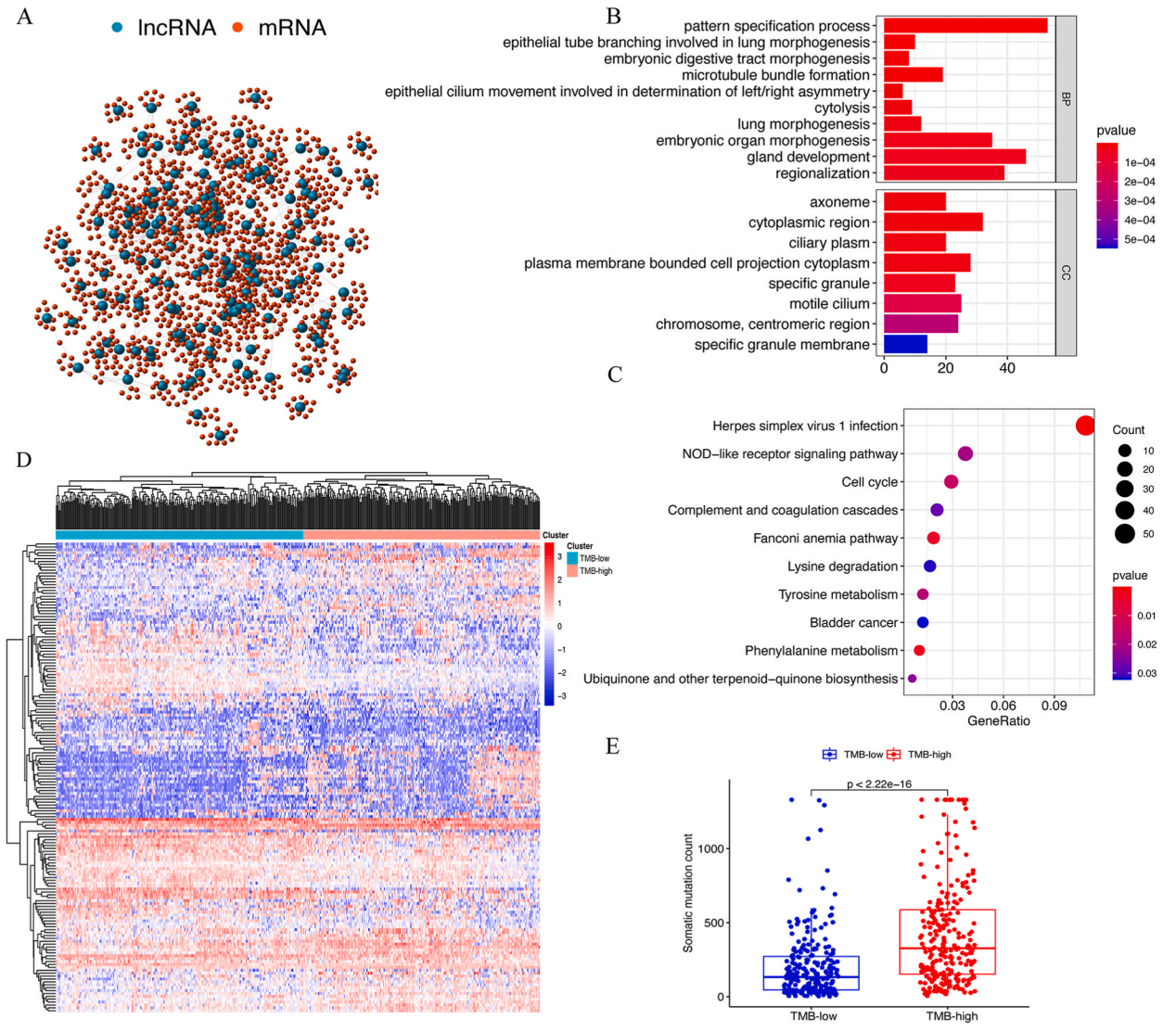


Fig. 2. Functional analysis and identification of mutation-related differentially expressed genes in TCGA lung adenocarcinoma samples. (A) Network of mutation-related lncRNAs and mRNAs based on the correlation coefficients of Pearson analysis (p-value <0.05). (B) GO enrichment analysis of mutation-related differentially expressed mRNAs. (C) KEGG enrichment analysis of mutation-related differentially expressed mRNAs. (D) Unsupervised clustering of 497 TCGA lung adenocarcinoma samples based on mutation-related differentially expressed lncRNA expression. The green cluster (left) is the TMB-low group and the red cluster is theTMB-high group. (E) The boxplot of the somatic mutation counts in TMB-low group and TMB-high group, compared by using Mann-Whitney U test. (p-value <0.05). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Ranking all samples by the cumulative number of somatic mutations in descending order, the first quartile was defined as the TMB-high group and the last quartile was defined as the TMB-LOW group. With a cutoff value | Log (fold change) |≥ 1.0 and p-value <0.05, a total of 162 differentially expressed lncRNAs were screened out using the limma method. Among them, 75 lncRNAs were up-regulated, while 87 lncRNAs were down-regulated in TMB-high group (Supplementary Table). The heatmap of the top 40 differentially expressed lncRNAs was showed in Fig. 1 B.

To explore the potential biological processes, cellular components and pathways of these 162 differentially expressed lncRNAs, we calculated the correlation coefficient between lncRNAs and mRNAs using the Pearson method, and regarded the top 10 mRNAs as co-expressed lncRNA-related mRNAs according to the correlation coefficient. Therefore, a network which consisting of co-expressed lncRNA-related mRNAs and 162 mutation-related lncRNAs was constructed (Fig. 2A). Then, we performed functional enrichment analysis with the co-expressed lncRNA-related mRNAs. GO analysis revealed co-expressed mRNAs were significantly enriched in the pattern specification process, epithelial tube branching involved in lung morphogenesis, gland development, regionalization, plasma membrane bounded cell projection cytoplasm and cytoplasmic region (Fig. 2B). KEGG analysis revealed co-expressed mRNAs were significantly enriched in the Herpes simplex virus 1 infection, NOD-like receptor signaling pathway, Cell cycle, Complement and

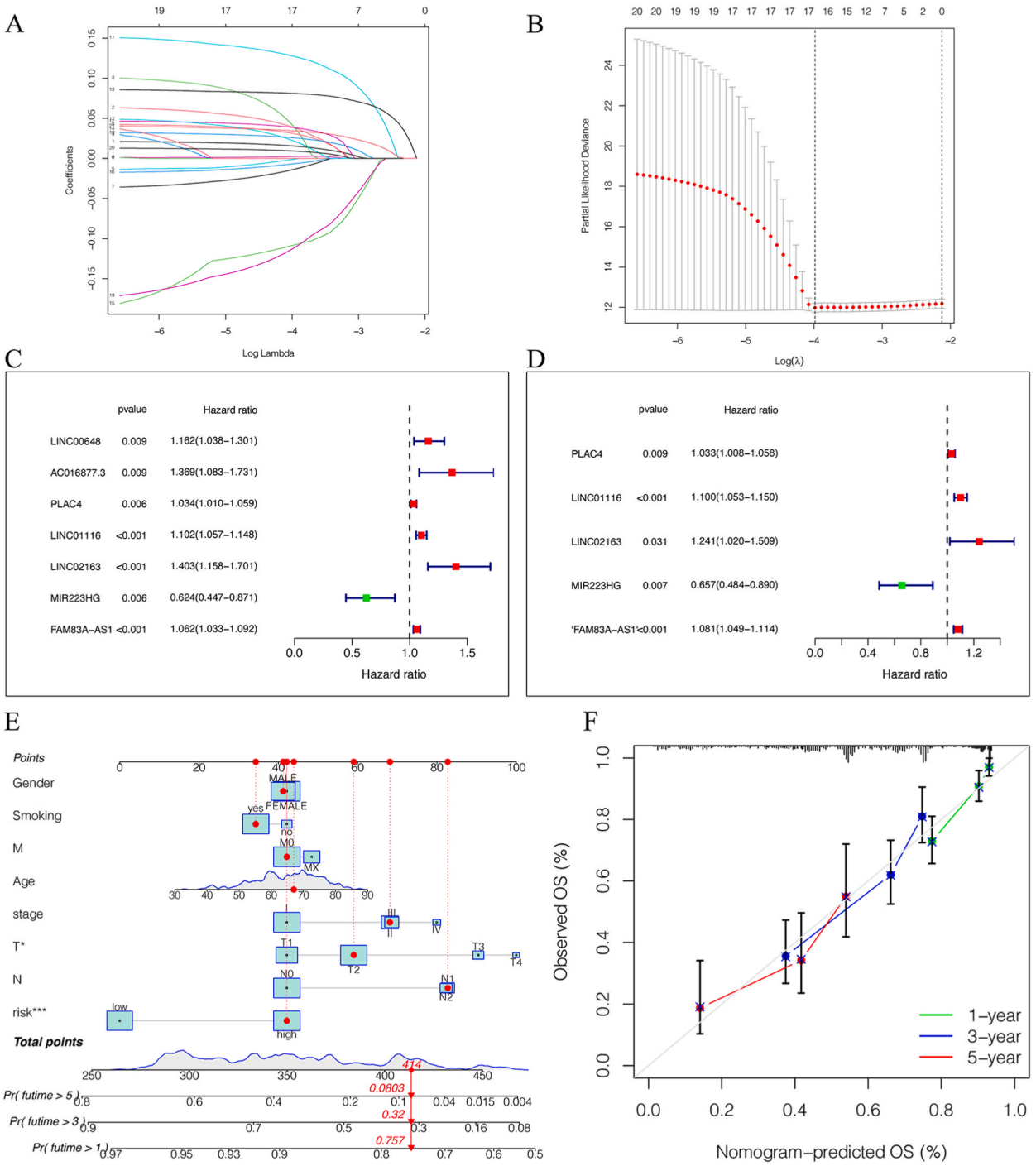


Fig. 3. Identification of mutation-related lncRNAs associated with overall survival in lung adenocarcinoma. (A–B) The Lasso regression suggested that 17 genes were essential for the prognostic prediction model. (C) The Univariate Cox regression analysis identified 7 lncRNAs (*LINC00648,* AC016877.3 *PLAC4,* LINC01116, LINC02163, *MIR223HG, FAM83A-AS1*) were as the key lncRNAs for constructing the prognostic prediction model, *p*-value <0.05. (D) The Multivariate Cox regression analysis identified 5 lncRNAs (*PLAC4,* LINC01116, LINC02163*, MIR223HG, FAM83A-AS1*) were as the key lncRNAs for constructing the prognostic prediction model, *p*-value <0.05. (E) Nomogram based on the key lncRNAs model, and (F) the calibration curves validated the discrimination.

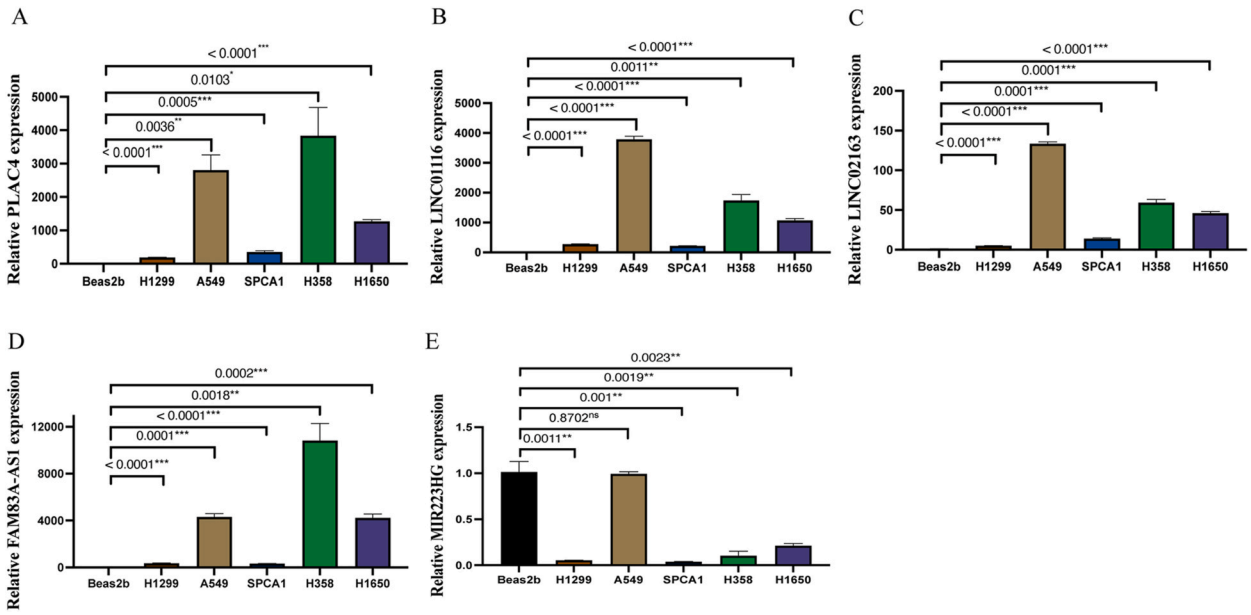


Fig. 4. (A–E) The expression of *PLAC4,* LINC01116, LINC02163, F*AM83A-AS1* and *MIR223HG* in the human non-tumorigenic lung epithelial cell line BEAS-2B and lung adenocarcinoma cell lines H1299, A549, SPCA1, H358 and H1650. *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

Table 2
Univariate and multivariate Cox regression analyses of the tumor mutation-related prediction model.

Factors	Univariate Cox			Multivariate Cox		
	HR	95% IC	*p*-value	HR	95% IC	*p*-value
Age	1.2701	(0.9289–1.7367)	0.1341	1.4674	(1.0499–2.0511)	0.0248*
Gender	1.1810	(0.8648–1.6128)	0.2953	1.0975	(0.7989–1.5077)	0.5659
Smoking history	1.0493	(0.7581–1.4522)	0.7718	1.0338	(0.7334–1.4572)	0.8495
T	1.5678	(1.2903–1.9051)	0.0000	1.3798	(1.1218–1.6971)	0.0023*
N	1.6480	(1.3765–1.9730)	0.0000	1.5397	(1.2751–1.8591)	0.0000*
M	0.9754	(0.8137–1.1692)	0.7876	1.0695	(0.8873–1.2890)	0.4806
Risk Score	1.1759	(1.0477–1.3048)	0.0000	1.1699	(1.0405–1.3001)	0.0000*

coagulation cascades, Fanconi anemia pathway, Lysine degradation, Tyrosine metabolism, Bladder cancer, Phenylalanine metabolism and Ubiquinone and other terpenoid-quinone biosynthesis (Fig. 2C).

To validate the efficiency of distinguishing high-level and low-level tumor mutations, we performed hierarchical clustering analysis on 497 LUAD samples from TCGA with 162 differentially expressed lncRNAs. All samples were divided into two groups according to the expression levels of these 162 lncRNAs, the group with higher cumulative somatic mutation was defined as the TMB-high group, and the other group was defined as the TMB-low group (Fig. 2D). By comparison, the somatic mutation count in the TMB-high group was significantly higher than that in the TMB-low group ($p < 0.05$, Fig. 2E).

3.2. Construction of the prognostic prediction model and survival analysis

To further investigate the prognostic value of 162 tumor mutation-related lncRNAs, we used Lasso regression analysis to filter out 17 lncRNAs (*LINC00648*, LINC01833, AC112721.2, P*RDM16-DT,* AC099850.4, ATP13A4-AS1, LINC01671, PLAC4, MIR193BHG, LINC02587, LINC01116, AL590226.1, COLCA1, LINC02163, MIR223HG, AC0016877.3, FAM83A-AS1) that were reliable for the prediction model (*p*-value <0.05, Fig. 3A–B). The Univariate Cox regression analysis identified 7 lncRNAs (*LINC00648,* AC016877.3 PLAC4, LINC01116, LINC02163, MIR223HG, FAM83A-A*S1*, $p < 0.05$, Fig. 3C). The Multivariate Cox regression analysis identified 5 lncRNAs (*PLAC4,* LINC01116, LINC02163, *MIR223HG, FAM83A-AS1*) were as the key lncRNAs for constructing the prognostic prediction model ($p < 0.05$, Fig. 3D). Based on the tumor mutation-related prognostic prediction model, a nomogram was constructed and a good discrimination of the model was validated by the 1-, 3- and 5-year calibration curves (Fig. 3E–F).

Then, real-time PCR was performed to further verify the expression levels of the *PLAC*4, LINC01116, LINC02163, *MIR223HG,* and *FAM83A-AS1* in human non-tumorigenic lung epithelial cell line BEAS-2B and the human lung adenocarcinoma cancer cell lines H1299, A549, SPCA1, H358, and H1650. The results indicated that PLAC4, LINC01116, L*INC02163,* and *FAM83A-AS1* were increased and *MIR223HG* was downregulated in human lung adenocarcinoma cancer, which was consistent with the predicted results (Fig. 4A–E).

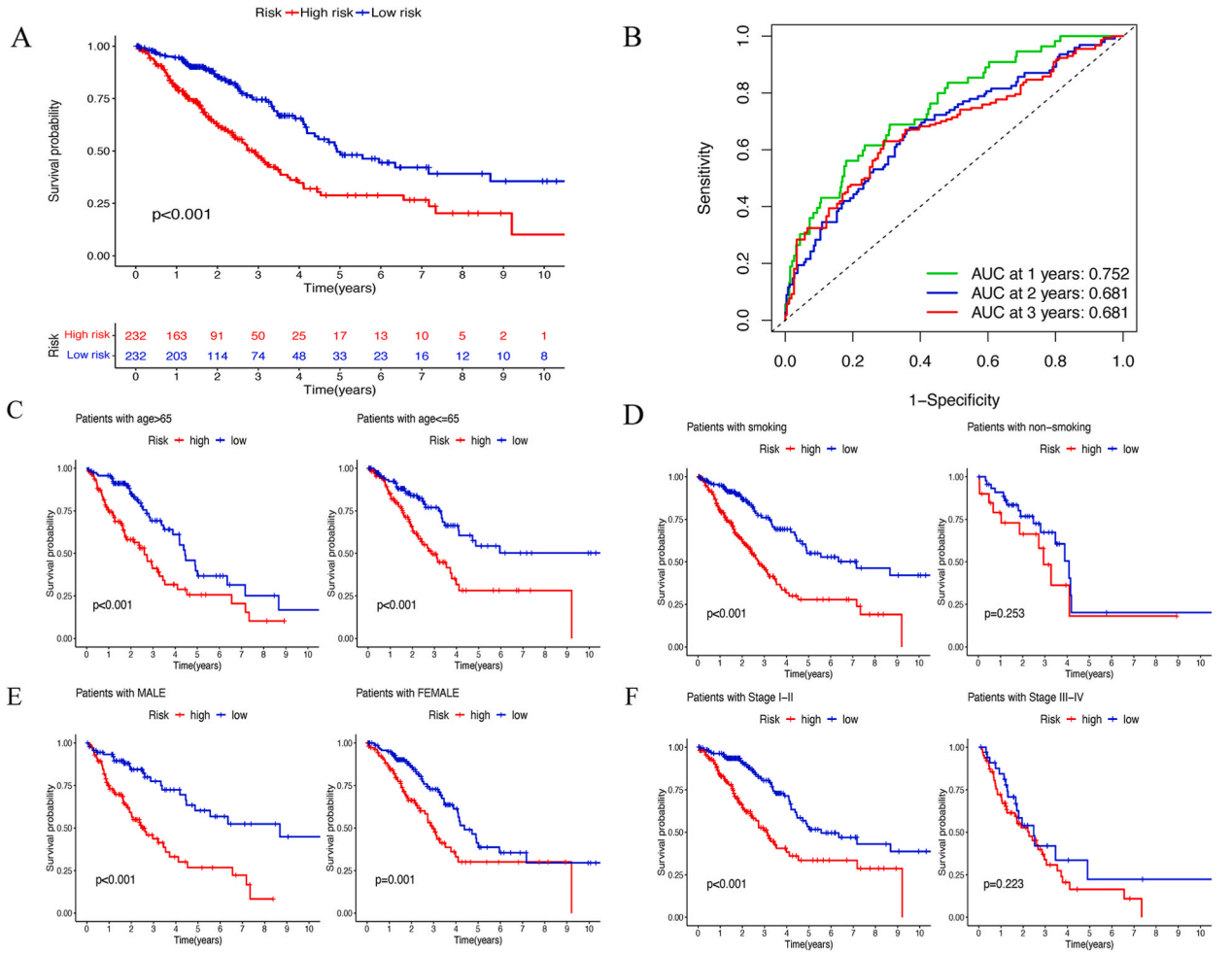
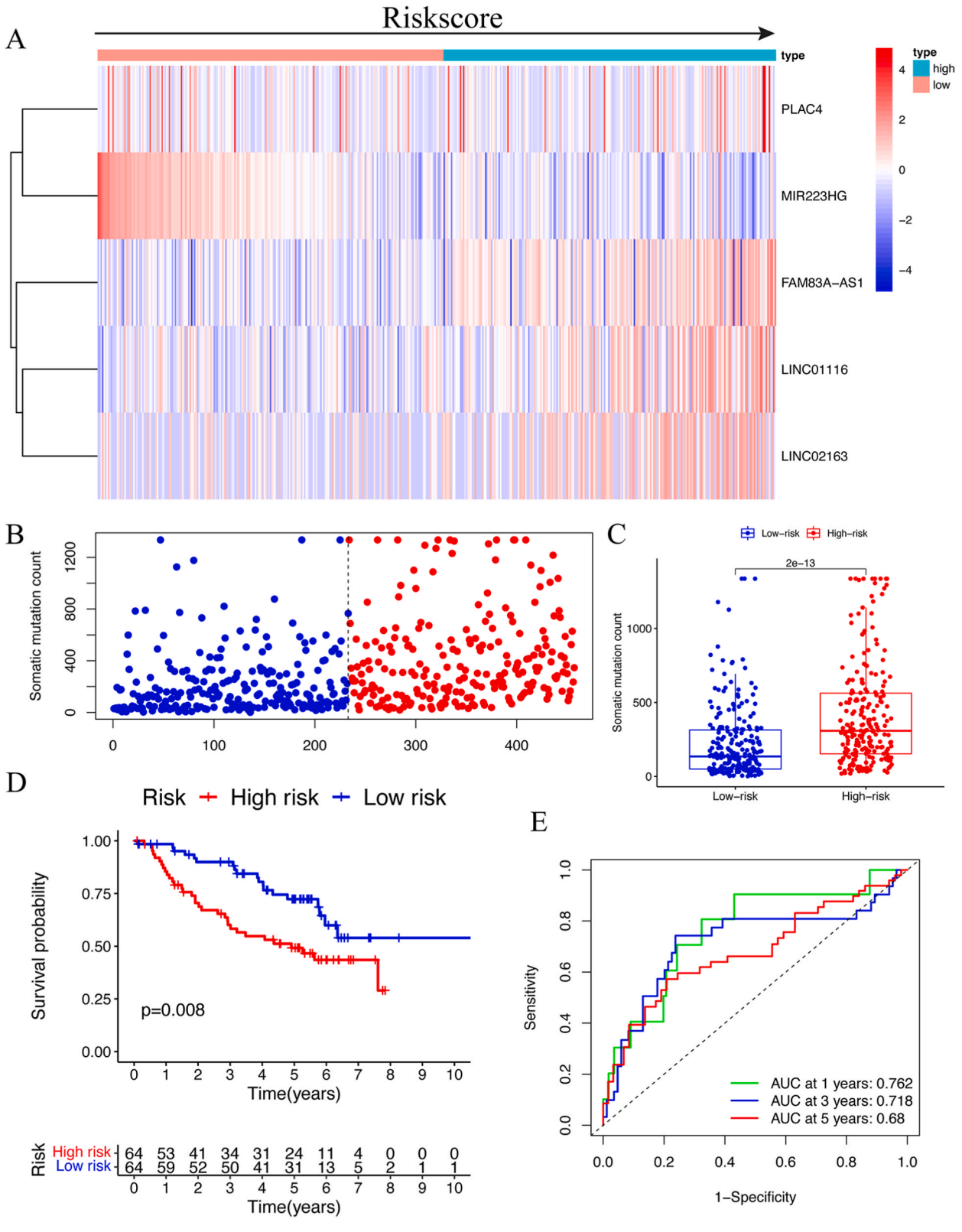


Fig. 5. Survival analysis of the prognostic lncRNA signature. (A) The Kaplan-Meier overall survival curves for risk score groups. Red curves represent the high risk group and blue curves represent the low risk group, *p*-value <0.05. (B) The ROC curves showed that the area under curve (AUC) at 1-, 3- and 5-year was 0.752, 0.681 and 0.681, respectively. Subgroup analyses classified by patients' age (C), smoking status (D), gender (E), and clinic stage (F) were conducted. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.3. Validation of the prognostic value of the constructed tumor mutation-related model

We calculated a risk score for each LUAD patient in TCGA and divided all samples into high-risk group (n = 232) and the rest into low-risk group (n = 232) based on the median risk score (0.92). Univariate and multivariate Cox regression analyses showed that the risk score was an independent prognostic factor for LUAD survival (Table 2). Kaplan-Meier curves showed that the overall survival of the low-risk group was significantly better than that of the high-risk group (*p*-value <0.05, Fig. 5A), and the ROC curves showed that the area under curve (AUC) at 1-, 3- and 5-year was 0.752, 0.681 and 0.681 (Fig. 5B). Subgroup analyses classified by patients age, smoking status, gender and clinic stage were subsequently conducted, and the results revealed that the high-risk group in this model was significantly associated with poorer prognosis in patients aged ≤65 years and those aged >65 years, smoking patients, male and female patients, and stage I-II patients (Fig. 5C–F).

The expression levels of the *PLAC4,* LINC01116, LINC02163 and *FAM83A-AS1* were up-regulated in high-risk group, the expression level of the *MIR223HG* was down-regulated in high-risk group (Fig. 6A). The somatic mutation counts increased with increasing risk score (Fig. 6B–C). To validate the prognostic value of the tumor mutation-related prediction model, we used GSE50081 lung adenocarcinoma data set. Kaplan-Meier curves showed that the overall survival of the low-risk group was significantly better than that of the high-risk group (*p*-value <0.05, Fig. 6D), and the ROC curves showed that the area under curve (AUC) at 1-, 3- and 5-year was 0.762, 0.718, and 0.68, respectively (Fig. 6E).



(caption on next page)

Fig. 6. Validation of mutation-related signature in TCGA set and the prognostic value in the GSE50081 **cohort.** (A) The key lncRNA expression pattern and somatic mutation count distribution of patients between the high-risk group and low-risk group. (B)The key lncRNA expression pattern and somatic mutation count distribution of patients between the high-risk group and low-risk group. (C) The boxplot of the somatic mutation counts in high-risk group and low-risk group, compared by using Mann-Whitney *U* test. (*p*-value <0.05). (D) The Kaplan-Meier overall survival curves of LUAD patients in the GSE50081 cohort. Red curves represent the high-risk group and blue curves represent the low-risk group, p-value <0.05. (B) The ROC curves of LUAD patients in the GSE50081 cohort showed that the areas under the curve (AUC) at 1, 3, and 5 years were 0.762, 0.718, and 0.68, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.4. Survival analysis of low-risk and high-risk groups with different expression levels of PD-L1, PD1, and CTLA4

To further investigate the immunotherapy efficacy of the constructed tumor mutation-related model, we divided all patients in TCGA into a high expression group and a low expression group according to the expression levels of the immune checkpoints. Kaplan-Meier curves showed that the PD-L1 low/High-risk type had the worst outcomes, whereas the PD-L1 high/Low-risk type had a good outcome (log-rank test $p < 0.05$ Fig. 7A), the PD1 low/High-risk type had the worst outcomes, whereas the PD1 high/Low-risk type had a good outcome (log-rank test $p < 0.05$ Fig. 7C), the CTLA4 low/High-risk type had the worst outcomes, whereas the CTLA4 high/Low-risk type had a good outcome (log-rank test $p < 0.05$ Fig. 7E). The expression levels of PD1, PD-L1 and CTLA4 in the low-risk group were higher than those in the high-risk group ($p < 0.05$, Fig. 7B–D, F).

3.5. Prediction of immunotherapy efficacy in the constructed tumor mutation-related model

We used ImmuCellAI to estimate differences in immune cell infiltration between low-risk and high-risk groups in TCGA LUAD patients, and discrepancies of in 13 types of tumor-infiltrating immune cells were identified ($p < 0.05$, Fig. 8A). Then, IPS and TIDE analyses were performed to predict the efficacies of immune checkpoint inhibitors in the low-risk group and high-risk group.

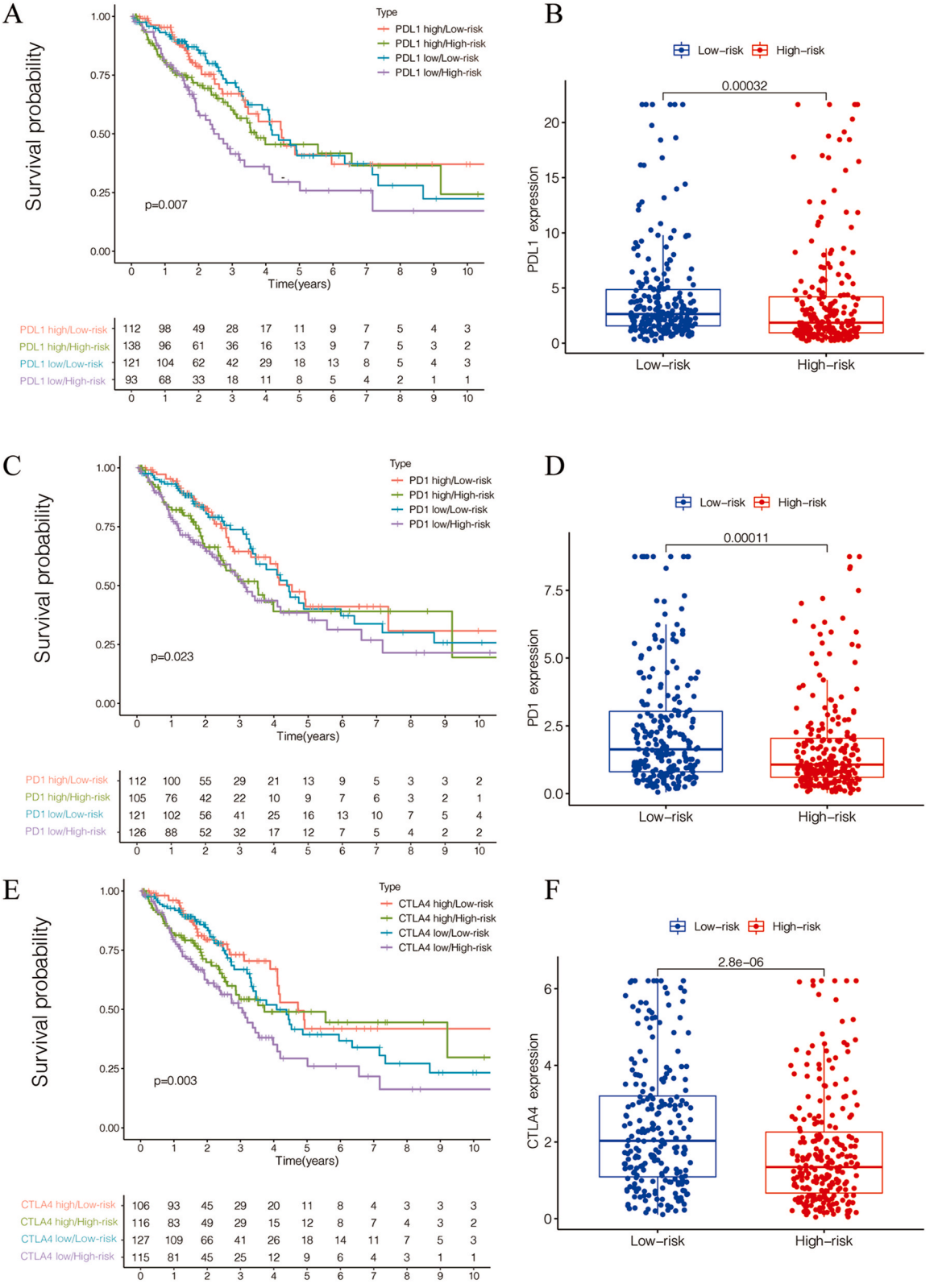
The IPS results revealed that the CTLA4_neg_PD1_neg scores of the low-risk group were significantly higher than those of the high-risk group ($p < 0.05$, Fig. 8B), the IPS CTLA4_neg_PD1_pos scores of the low-risk group were significantly higher than those of the high-risk group ($p < 0.05$, Fig. 8C), the IPS CTLA4_pos_PD1_neg scores of the low-risk group were significantly higher than those of the high-risk group ($p < 0.05$, Fig. 8D), the IPS CTLA4_pos_PD1_pos scores of the low-risk group were significantly higher than that of the high-risk group ($p < 0.05$, Fig. 8E).

The results of the TIDE algorithm revealed that TIDE scores ($p < 0.05$, Fig. 8F), T cell dysfunction scores ($p < 0.05$, Fig. 8G), Interferon gamma scores ($p < 0.05$, Fig. 8I) and Merck18 scores ($p < 0.05$, Fig. 8J) in the low-risk group were significantly higher than those in the high-risk group. Whereas, the T cell exclusion scores ($p < 0.05$, Fig. 8H) and Cancer-associated fibroblast scores ($p < 0.05$, Fig. 8K) in the low-risk group were significantly lower than those in the high-risk group.

4. Discussion

Recently, the relationship between genetic mutations and clinical outcomes has been extensively researched in malignancies. One hypothesis suggested that more tumor mutations in genomes were more likely to generate tumor-specific neoantigens that could activate immune responses [35]. Based on this concept, TMB was examined to predict the response to treatment with ICIs, and it has been shown that high TMB predicted a better response to ICIs therapy in multiple types of cancer [36]. In an ICI clinical trial, prolonged progression-free survival was found to be associated with high somatic mutation in colorectal cancer patients which mismatch repair-deficient were detected in tumor [37]. In a CTLA4 blockade clinical trial, mutational burden correlated with the degree of clinical benefit in melanoma, but alone was insufficient to predict benefit. Further research by using genome wide somatic neoepitope analysis, showed that the neoantigen counts were significantly associated with clinical benefits [38]. Similarly, in the ICIs treatment of NSCLC, multiple studies have demonstrated that high TMB which detected by whole exome sequencing was shown to be associated with better ICIs response rates [39].

In this study, we downloaded the somatic mutation profiles and RNA-seq expression database of LUAD from TCGA. Based on the cumulative number of somatic mutations, we first identified 162 tumor mutation-related lncRNAs, and then performed Lasso regression analysis and Cox hazards regression analysis to identify 5 lncRNAs (*PLAC4,* LINC01116, LINC02163, *MIR223HG, FAM83A-AS1*) as the candidates for constructing the prognostic prediction model. These candidates have been implicated in the regulation of various diseases. For instance, *PLAC4* gene which transcribed from placental cells on chromosome 21, was shown to be a potential marker for noninvasive prenatal diagnosis of fetal trisomy 21 [40]. This suggests potential implications for gene regulation in disease states and individual development, the specific mechanisms in tumor initiation and progression remain unclear. LINC02163 h*as been demonstrated to promote malignant cell progression in colorectal cancer, breast cancer, and gastric cancer* [41–43]. The novel fusion of cell surface receptors *CLEC12A* and *MIR223HG* identified in chronic myeloid leukemia by Dhungel et al. suggests a potentially significant impact on disease, the exact mode of action in LUAD remains to be elucidated [44]. Silencing LINC01116, as demonstrated by Shang et al., inhibits invasion, migration, and promotes apoptosis in LUAD cells through modulation of the AKT signaling pathway [45]. *FAM83A-AS1*, as revealed by Xiao et al. and Wang et al., plays a pivotal role in LUAD cell proliferation, migration, and the epithelial-mesenchymal transition [46]. Although it has been identified as a risk prognostic factor through the ERK signaling pathway, how FAM83A-AS1 precisely regulates the ERK signaling pathway and its downstream effects on tumor progression remains unclear. Our study sheds light on the prognostic significance of these lncRNAs, their mechanistic intricacies remain an area ripe for exploration.



(caption on next page)

Fig. 7. Kaplan-Meier curve analysis showing the overall survival of TCGA lung adenocarcinoma patients classified according to Low-risk/High-risk groups with high/low immune checkpoint gene statuses. (A) Low-risk/High-risk groups with PDL1 high/low status, (C) Low-risk/High-risk groups with PDL1 high/low status and (D) Low-risk/High-risk groups with CTLA4 high/low status. Statistical analysis was performed using the log-rank test. *p*-value <0.05. Boxplot of the (B) PD1, (D) PDL1, and (F) CTLA4 expression in the high risk-group and low-risk group, compared by using Mann-Whitney *U* test (*p*-value <0.05).

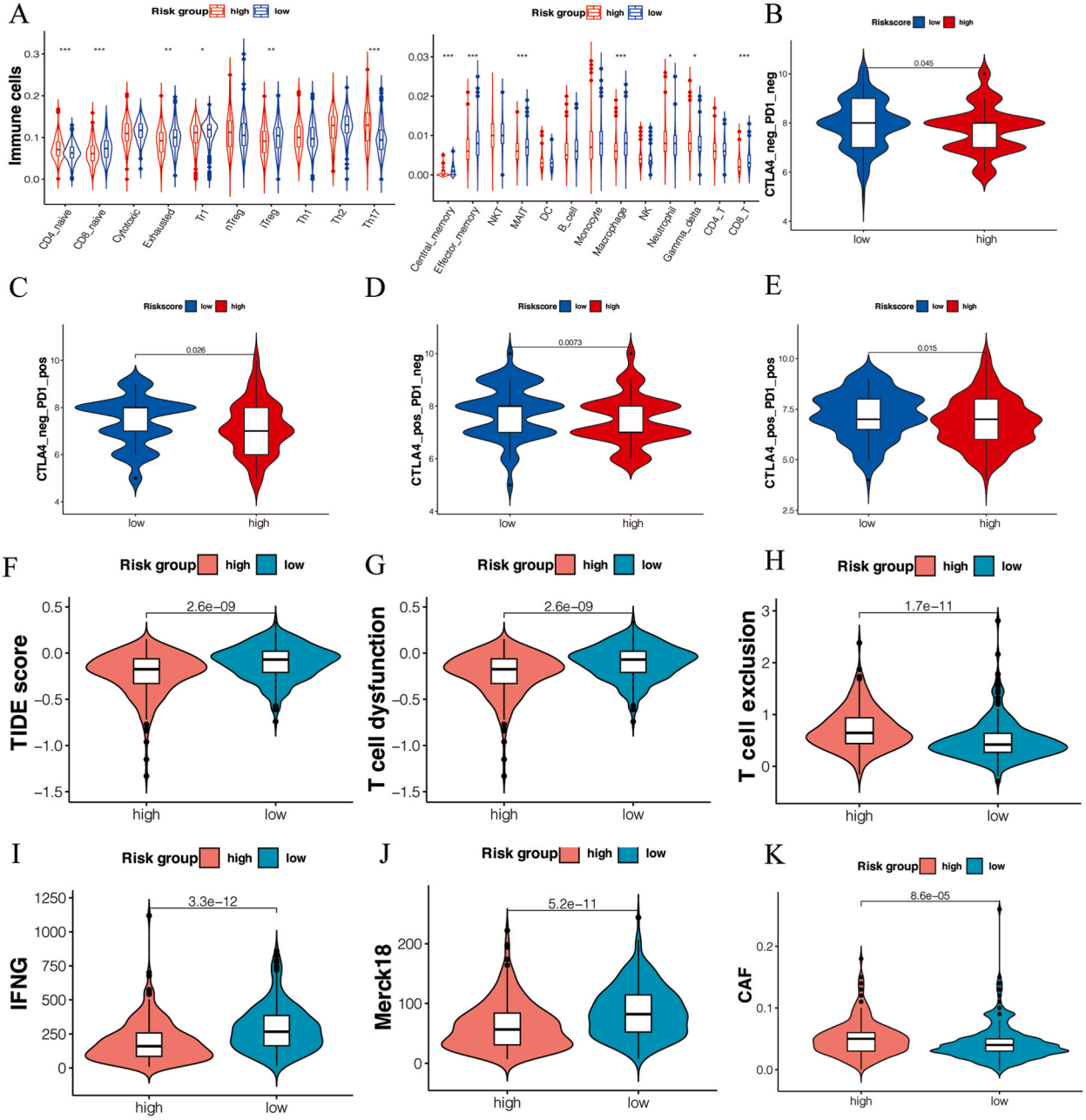


Fig. 8. Prediction of immunotherapy efficacy in the constructed tumor mutation-related model. (A) The tumor-infiltrating immune cells in the high risk-group and low-risk group, compared by using Mann-Whitney *U* test (*p*-value <0.05). (B–E) The relative probability that the low-risk group and the high-risk group will respond to ICI treatment assessed using immunophenoscore. (B) CTLA4 negative and PD1 negative group, (C) CTLA4 negative and PD1 positive group, (D) CTLA4 positive and PD1 negative group, (E) CTLA4 positive and PD1 positive group. *p*-value <0.05. (F–K) Tumor immune evasion potential and response to immunotherapy assessed using Tumor Immune Dysfunction and Exclusion. (F) TIDE score; (G) T cell dysfunction; (H) T cell exclusion; (I) Interferon gamma, IFNG; (J) Merck18; (K) Cancer-associated fibroblast, CAF. *p*-value <0.05.

We hope that our work stimulates further research initiatives to unravel the molecular underpinnings of these lncRNAs and their impact on lung adenocarcinoma prognosis.

Then, Kaplan-Meier curves showed that the tumor mutation-related signature was significantly associated with overall survival in patients with LUAD. In addition, in the smoking subgroups, the high-risk group was significantly associated with poorer overall survival, compared with non-smoking subgroups, where there was no difference in overall survival between the high-risk group and the low-risk group. Consistent with previous studies, the mutation counts in the lung cancer genome was revealed to be closely correlated with smoking status in patients and there might be a dose-response relationship between the tobacco exposure and the degree of mutational burden [47]. In the stage I-II subgroups, the high-risk group was significantly associated with poorer overall survival, compared with stage III-IV subgroups, where there was no difference in overall survival between the high-risk group and the low-risk group. However, there was currently a lack of evidence for differences in somatic mutations between early-stage LUAD and advanced-stage LUAD. Our findings indicated that the tumor mutation-related signature might be important in future assessments of the role of immunotherapy in neoadjuvant treatment for LUAD. While our study aimed to develop predictive models, we acknowledge that the AUC values, particularly on multiple occasions, were below the conventional threshold of 0.7. This suggests room for improvement in the predictive performance of our models. Future work will focus on refining our methodology to enhance model accuracy and reliability. Another limitation is that TCGA provides de-identified and aggregated data, constraining our ability to access detailed information about the specific sampling protocol employed for each case due to privacy concerns. In the future we will be working on potential avenues of research to validate and complement findings from large-scale datasets such as TCGA by collecting and analyzing our own samples with detailed sampling information.

Lastly, we predicted the efficacy of the immune checkpoint inhibitor for the tumor mutation-related signature. Kaplan-Meier curves showed that the PD-L1 low/High-risk, PD-1 low/High-risk and CTLA4 low/High-risk types had the worst outcomes, whereas the PD-L1 high/Low-risk, PD-1 high/Low-risk and CTLA4 high/Low-risk types had better outcomes. The results indicated that the tumor mutation-related signature combined with the gene expression level of the immune checkpoints might be a good biomarker for predicting the prognosis in LUAD. Furthermore, IPS and TIDE algorithm analyses showed significantly higher IPS scores and TIDE scores in the low-risk group than that in the high-risk group which indicated that the tumor mutation-related signature was associated with immunotherapy and might be useful in future clinical trials for predicting the immunotherapy response in LUAD.

Immunotherapy has revolutionized the treatment landscape for NSCLC. While both PD-L1 expression and TMB have shown promise as reliable biomarkers for predicting immunotherapy response, certain limitations persist. PD-L1 antibodies and current TMB detection methods lack standardization, are costly, and face challenges in clinical implementation. In contrast, the detection of lncRNAs, whether in liquid biopsy or tissue samples, presents a more convenient, rapid, and cost-effective approach. Coding RNAs provide essential information about protein-coding genes, lncRNAs offer complementary insights into the non-coding portion of the transcriptome. Integrating both coding and non-coding RNA information can provide a more comprehensive understanding of the molecular landscape of lung cancer, leading to improved disease indication and prediction.

In this study, we constructed a tumor mutation-related model combining the expression levels of five lncRNAs and tumor mutations to predict the prognosis of LUAD. Functionally, the signature may serve as an indicator of tumor mutation levels and have important implications for predicting the response to immunotherapy in LUAD.

Data availability

The datasets generated and analyzed during the current study are available in public TCGA database: https://portal.gdc.cancer.gov/repository, Gene Expression Omnibus (GEO) database: https://www.ncbi.nlm.nih.gov/geo, and GENCODE website: https://www.gencodegenes.org/. Data will be made available on request.

Funding

This study was supported by Yunnan Fundamental Research Projects, China (grant NO. 202301AT070153, NO.202301AU070196) and Joint Special Fundamental Research Project of Yunnan Science & Technology Department and Kunming Medical University, China (grant NO. 202201AY070001-141).

Ethics statements

No animal studies are presented in this manuscript. No human studies are presented in this manuscript. No potentially identifiable human images or data is presented in this study.

CRediT authorship contribution statement

Wenjie Chen: Writing – review & editing, Writing – original draft. **Chen Liao:** Writing – review & editing, Writing – original draft. **Xudong Xiang:** Data curation. **Heng Li:** Writing – review & editing. **Qiang Wu:** Data curation. **Wen Li:** Methodology. **Qianli Ma:** Visualization. **Nan Chen:** Visualization. **Benchao Chen:** Writing – review & editing. **Gaofeng Li:** Supervision, Project administration.

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.

Acknowledgments

The results shown here are in whole based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e28670.

References

[1] R.L. Siegel, et al., Cancer statistics, 2023, CA A Cancer J. Clin. 73 (1) (2023) 17–48.
[2] W.D. Travis, et al., Introduction to the 2015 world health organization classification of tumors of the lung, pleura, thymus, and heart, J. Thorac. Oncol. 10 (9) (2015) 1240–1242.
[3] R.S. Herbst, D. Morgensztern, C. Boshoff, The biology and management of non-small cell lung cancer, Nature 553 (7689) (2018) 446–454.
[4] D. Konig, S. Savic Prince, S.I. Rothschild, Targeted therapy in advanced and metastatic non-small cell lung cancer. An update on treatment of the most important actionable oncogenic driver alterations, Cancers 13 (4) (2021).
[5] A.C. Tan, D.S.W. Tan, Targeted therapies for lung cancer patients with oncogenic driver molecular alterations, J. Clin. Oncol. 40 (6) (2022) 611–625.
[6] N.A. Rizvi, et al., Activity and safety of nivolumab, an anti-PD-1 immune checkpoint inhibitor, for patients with advanced, refractory squamous non-small-cell lung cancer (CheckMate 063): a phase 2, single-arm trial, Lancet Oncol. 16 (3) (2015) 257–265.
[7] M. Wang, R.S. Herbst, C. Boshoff, Toward personalized treatment approaches for non-small-cell lung cancer, Nat. Med. 27 (8) (2021) 1345–1356.
[8] M. Reck, J. Remon, M.D. Hellmann, First-line immunotherapy for non-small-cell lung cancer, J. Clin. Oncol. 40 (6) (2022) 586–597.
[9] D.S. Ettinger, et al., NCCN guidelines insights: non-small cell lung cancer, version 2.2021, J. Natl. Compr. Cancer Netw. 19 (3) (2021) 254–266.
[10] M. Reck, et al., Five-year outcomes with pembrolizumab versus chemotherapy for metastatic non-small-cell lung cancer with PD-L1 tumor proportion score >/= 50, J. Clin. Oncol. 39 (21) (2021) 2339–2349.
[11] S. Gadgeel, et al., Updated analysis from KEYNOTE-189: pembrolizumab or placebo plus pemetrexed and platinum for previously untreated metastatic nonsquamous non-small-cell lung cancer, J. Clin. Oncol. 38 (14) (2020) 1505–1517.
[12] T.A. Chan, et al., Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic, Ann. Oncol. 30 (1) (2019) 44–56.
[13] D.M. Merino, et al., Establishing guidelines to harmonize tumor mutational burden (TMB): in silico assessment of variation in TMB quantification across diagnostic platforms: phase I of the Friends of Cancer Research TMB Harmonization Project, J Immunother Cancer 8 (1) (2020).
[14] N.A. Rizvi, et al., Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer, Science 348 (6230) (2015) 124–128.
[15] S.J. Klempner, et al., Tumor mutational burden as a predictive biomarker for response to immune checkpoint inhibitors: a review of current evidence, Oncol. 25 (1) (2020) e147–e159.
[16] N. Ready, et al., First-line nivolumab plus ipilimumab in advanced non–small-cell lung cancer (CheckMate 568): outcomes by programmed death ligand 1 and tumor mutational burden as biomarkers, J. Clin. Oncol. 37 (12) (2019) 992.
[17] D.R. Gandara, et al., Blood-based tumor mutational burden as a predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab, Nat. Med. 24 (9) (2018) 1441–1448.
[18] T.R. Mercer, M.E. Dinger, J.S. Mattick, Long non-coding RNAs: insights into functions, Nat. Rev. Genet. 10 (3) (2009) 155–159.
[19] A. Bhan, M. Soleimani, S.S. Mandal, Long noncoding RNA and cancer: a new paradigm, Cancer Res. 77 (15) (2017) 3965–3981.
[20] Y.G. Chen, A.T. Satpathy, H.Y. Chang, Gene regulation in the immune system by long noncoding RNAs, Nat. Immunol. 18 (9) (2017) 962–972.
[21] Q.M. Wang, et al., LncRNA MALAT1 promotes tumorigenesis and immune escape of diffuse large B cell lymphoma by sponging miR-195, Life Sci. 231 (2019) 116335.
[22] D. Huang, et al., NKILA lncRNA promotes tumor immune evasion by sensitizing T cells to activation-induced cell death, Nat. Immunol. 19 (10) (2018) 1112–1125.
[23] P. Wu, et al., Emerging role of tumor-related functional peptides encoded by lncRNA and circRNA, Mol. Cancer 19 (1) (2020) 22.
[24] R. Benacka, et al., Non-coding RNAs in human cancer and other diseases: overview of the diagnostic potential, Int. J. Mol. Sci. 24 (22) (2023) 16213.
[25] Y. Qian, L. Shi, Z. Luo, Long non-coding RNAs in cancer: implications for diagnosis, prognosis, and therapy, Front. Med. 7 (2020) 612393.
[26] C. Evans, J. Hardin, D.M. Stoebel, Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions, Briefings Bioinf. 19 (5) (2018) 776–792.
[27] A. Mayakonda, et al., Maftools: efficient and comprehensive analysis of somatic variants in cancer, Genome Res. 28 (11) (2018) 1747–1756.
[28] T. Emura, S. Matsui, H.Y. Chen, compound.Cox: univariate feature selection and compound covariate for predicting survival, Comput. Methods Progr. Biomed. 168 (2019) 21–37.
[29] C.T. Yeh, G.Y. Liao, T. Emura, Sensitivity analysis for survival prognostic prediction with gene selection: a copula method for dependent censoring, Biomedicines 11 (3) (2023) 797.
[30] J. Liu, et al., An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics, Cell 173 (2) (2018) 400–416 e11.
[31] Y.R. Miao, et al., ImmuCellAI: a unique method for comprehensive T-cell subsets abundance prediction and its application in cancer immunotherapy, Adv. Sci. 7 (7) (2020) 1902880.
[32] P. Charoentong, et al., Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade, Cell Rep. 18 (1) (2017) 248–262.
[33] J. Fu, et al., Large-scale public data reuse to model immunotherapy response and resistance, Genome Med. 12 (1) (2020) 21.
[34] P. Jiang, et al., Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response, Nat. Med. 24 (10) (2018) 1550–1558.
[35] N. Xie, et al., Neoantigens: promising targets for cancer therapy, Signal Transduct. Targeted Ther. 8 (1) (2023) 9.
[36] R.M. Samstein, et al., Tumor mutational load predicts survival after immunotherapy across multiple cancer types, Nat. Genet. 51 (2) (2019) 202–206.
[37] D.T. Le, et al., PD-1 blockade in tumors with mismatch-repair deficiency, N. Engl. J. Med. 372 (26) (2015) 2509–2520.
[38] A. Snyder, et al., Genetic basis for clinical response to CTLA-4 blockade in melanoma, N. Engl. J. Med. 371 (23) (2014) 2189–2199.
[39] J.C.F. Quintanilha, et al., Tumor mutational burden in real-world patients with pancreatic cancer: genomic alterations and predictive value for immune checkpoint inhibitor effectiveness, JCO Precis Oncol 7 (2023) e2300092.

[40] Y.M. Lo, et al., Plasma placental RNA allelic ratio permits noninvasive prenatal chromosomal aneuploidy detection, Nat. Med. 13 (2) (2007) 218–223.
[41] J. Ma, et al., LINC02163 promotes colorectal cancer progression via miR-511-3p/AKT3 axis, Artif. Cells, Nanomed. Biotechnol. 48 (1) (2020) 961–968.
[42] C. Qin, et al., Long noncoding RNA LINC02163 accelerates malignant tumor behaviors in breast cancer by regulating the MicroRNA-511-3p/HMGA2 Axis, Oncol. Res. 28 (5) (2020) 483–495.
[43] L. Dong, et al., LINC02163 regulates growth and epithelial-to-mesenchymal transition phenotype via miR-593-3p/FOXK1 axis in gastric cancer cells, Artif. Cells, Nanomed. Biotechnol. 46 (sup2) (2018) 607–615.
[44] B.P. Dhungel, et al., The fusion of CLEC12A and MIR223HG arises from a trans-splicing event in normal and transformed human cells, Int. J. Mol. Sci. 22 (22) (2021).
[45] B. Shang, et al., Silencing LINC01116 suppresses the development of lung adenocarcinoma via the AKT signaling pathway, Thorac Cancer 12 (14) (2021) 2093–2103.
[46] W. Wang, et al., LncRNA FAM83A-AS1 promotes lung adenocarcinoma progression by enhancing the pre-mRNA stability of FAM83A, Thorac Cancer 12 (10) (2021) 1495–1502.
[47] R. Govindan, et al., Genomic landscape of non-small cell lung cancer in smokers and never-smokers, Cell 150 (6) (2012) 1121–1134.