Heliyon 9 (2023) e21336

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Constructing an APOBEC-related gene signature with predictive value in the overall survival and therapeutic sensitivity in lung adenocarcinoma

Yu Luo [a,b,1], Huiru Wang [c,1], Jian Zhong [d], Jianrong Shi [e], Xianlin Zhang [f], Yanni Yang [g,*], Ruixin Wu [e,**]

[a] Gynecology Department of Jingmen Traditional Chinese Medicine Hospital, Jingmen, 448000, China
[b] Beijing University of Traditional Chinese Medicine Guoyitang Expert Clinic, National Medical Hall of Beijing University of Traditional Chinese Medicine, Jingmen Traditional Chinese Medicine Hospital, Jingmen, 448000, China
[c] Clinical College of Traditional Chinese Medicine, Hubei University of Traditional Chinese Medicine, Wuhan, 430014, China
[d] Department of Nephrology, Dongzhimen Hospital, Beijing University of Traditional Chinese Medicine, Beijing, 100105, China
[e] School of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, 201203, China
[f] Department of Endocrinology, Wuhan Hospital of Traditional Chinese Medicine, Wuhan Traditional Chinese Medicine Hospital, Wuhan, 430014, China
[g] Health Management Center of Jingmen Traditional Chinese Medicine Hospital, Jingmen, 448000, China

ARTICLE INFO

Keywords:
APOBEC family
APOBEC mutagenesis enrichment score
Lung adenocarcinoma
Genomic instability
Prognostic signature
Immunotherapy
Oxidative stress

ABSTRACT

Background: APOBEC family play an important role in cancer mutagenesis and tumor development. The role of APOBEC family in lung adenocarcinoma (LUAD) has not been studied comprehensively.

Materials and methods: The expression data of pan-cancer as well as LUAD was obtained from public databases. The expression level of APOBEC family genes was analyzed in different normal and cancer tissues. APOBEC mutagenesis enrichment score (AMES) was utilized to evaluate the APOBEC-induced mutations and the relation of APOBEC with genomic instability. Gene set enrichment analysis was used to identify differentially enriched pathways. Univariate Cox regression and Lasso regression were applied to screen key prognostic genes. The immune cell infiltration was estimated by CIBERSORT. RT-qPCR assay, CCK-8 and Transwell assay were conducted to explore gene expression and lung cancer cell invasion.

Results: Cancer tissues had significantly altered expression of APOBEC family genes and the expression patterns of APOBEC family were different in different cancer types. APOBEC3B was the most aberrantly expressed in most cancer types. In LUAD, we observed a significantly positive correlation of AMES with intratumor heterogeneity (ITH), tumor neoantigen burden (TNB), and tumor mutation burden (TMB). High AMES group had high mutation counts of DNA damage repair pathways, and high enrichment of cell cycle and DNA repair pathways. We identified four prognostic genes (LYPD3, ANLN, MUC5B, and FOSL1) based on AMES, and constructed an AMES-related gene signature. The expressions of four genes were enhanced and accelerated the invasion

* Corresponding author.
** Corresponding author.
E-mail addresses: 2444599248@qq.com (Y. Yang), wuanubis@163.com (R. Wu).
[1] Equal Contribution.

https://doi.org/10.1016/j.heliyon.2023.e21336
Received 19 July 2023; Received in revised form 16 October 2023; Accepted 19 October 2023
Available online 28 October 2023

ability and viability of lung cancer cells. Furthermore, we found that high group increased oxidative stress level.

Conclusions: APOBEC family was associated with genomic instability, DNA damage-related pathways, and cell cycle in LUAD. The AMES-related gene signature had a great potential to indicate the prognosis and guide immunotherapy/chemotherapy for patients suffering from LUAD.

1. Introduction

Lung cancer, which results in over 1 million deaths annually worldwide, is a major cause of cancer death [1,2]. As the most commonly diagnosed type, Lung adenocarcinoma (LUAD) has a high proportion in smokers and an unfavorable prognosis of a 5-year overall survival of around 10–20 % mainly due to the late diagnosis and treatment [3]. A series of large-scale genomic characterization studies on lung cancer have been performed, which unveiled a massive heterogeneity of somatic variations including gene mutations and copy number variations [4–6]. TP53 is a tumor suppressor gene showing the most frequent mutation in LUAD [5]. Some somatic mutations such as EGFR and ALK are associated with the targeted molecular therapies. LUAD patients who have EGFR and ALK mutations can dramatically benefit from EGFR tyrosine kinase inhibitors and ALK inhibitors respectively [7]. In addition, genome signatures such as tumor mutational burden (TMB) and tumor neoantigen burden (TNB) have been demonstrated to be associated with the efficiency of immunotherapy, for instance, immune checkpoint blockade [8]. Gao et al. found that this low mutational load is more pronounced when a particular mutant isoform with a lower TMB (G12D) is co-mutated with TP53. They also suggested that G12D/TP53 co-mutations drive immunosuppression, which may be a predictive biomarker for immune checkpoint inhibitors in LUAD patients [9]. A comprehensive exploration on the genetic characteristics helps to analyze the mechanism of LUAD development and develop new therapeutic targets and drugs [10].

Genome instability is an important hallmark of cancer caused by DNA damage and mutation. Spontaneous deamination of methylcytosine bases yielding C-to-T mutations contributes to the largest endogenous source of genome instability [11]. While APOBEC-catalyzed deamination of cytosine bases lead to the second mutation source across cancer to uracil in 5′-TCW motif (W = A or T), leading to C-T and C-G mutations [12–14]. APOBEC family comprises a total of 11 enzymes including APOBEC1, APOBCE2, APOBEC3A-H (3A, 3B, 3C, 3D, 3F, 3G, and 3H), APOBEC4, and AICD1 (also known as AID). APOBEC3B has been reported to be an only detected source of C-to-U editing activity involving in DNA damage in breast cancer [14,15]. Activated APOBEC3B expression induces cell death, DNA fragmentation, and the deviation of cell cycle in breast cancer [14]. A significant correlation between the subclonal mutation burden and APOBEC-mediated mutagenesis has been detected by intratumour heterogeneity analysis of patients with non-small cell lung cancer [16]. Chromosome instability induced by APOBEC is related to a high occurrence of recurrence or death [16], which sustains the potential prognosis importance of APOBEC in patients suffering from lung cancer.

This research explored the influence of dysregulated APOBEC in LUAD patients, and develop an APOBEC-related prognostic signature for estimating the overall survival of LUAD. The expression of APOBEC family especially APOBEC3B exhibited a significant alteration across various cancer types. APOBEC mutagenesis enrichment score (AMES) was strikingly associated with genomic instability and intratumor heterogeneity. Importantly, we established an APOBEC-related gene signature that was reliable to predict LUAD overall survival.

2. Materials and Methods

2.1. Data collection and preprocessing

The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/) pan-cancer data (sample = 10535, gene number = 60499) including RNA-sequencing (RNA-seq) data, somatic mutation data, and sample information was accessed from UCSC Xena browser (https://xenabrowser.net/). Three independent datasets, including GSE72094, GSE31210, and GSE50081, as well as the LUAD sample data, were obtained from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/).

The samples without survival time or status were removed to prepare RNA-seq data of pan-cancer samples. Ensembl ID was transferred to gene symbol. For microarray data of GSE datasets, probes were transferred to gene symbols based on the corresponding annotation data in the microarray platform. The median value of a gene expression with multiple probes was selected after removing the probes matching to multiple genes.

2.2. Definition of APOBEC mutagenesis enrichment score

APOBEC mutagenesis predominantly occurs in a TCW (C for cytosine) motif, presented as two mutation patterns (TCW-TGW and TCW-TTW), where "W" represents adenine (A) or thymine (T). APOBEC mutagenesis enrichment score (AMES) was defined by Roberts et al. which enables to evaluate the mutagenesis density of APOBEC [12].

$$\mathrm{AMES} = \frac{n(mTCW) * n(C)}{n(TCW) * n(mC)} \quad (1)$$

in this formula ①, n(mTCW) represents the total number of mutated TCW (TCW-TTW and TCW-TGW) or WGA (WGA-WAA and WGA-WCA), n(C) represents the total number of C (or G), n(TCW) represents the total number of TCW (or WGA) motifs within 20 nucleotides around mutated C (or G), and n(mC) represents the total number of mutated C (or G). Cut-off values of 1 and 2 (low AMES: ≤ 1, moderate AMES: (1, 2], high AMES: >2) were applied to classify the tumor samples.

2.3. Assessment of immunogenic features and somatic mutations

The R package "maftools" was employed for calculating TMB and intra-tumor heterogeneity (ITH) score [17]. TNB was defined as the total number of putative neoantigens, which was obtained from The Cancer Immunome Atlas (TCIA) (https://tcia.at/). The somatic mutation data from TCGA was previously processed by mutect2 and annotated in mutation annotation format (MAF) file.

2.4. Identification of driver genes in LUAD

OncodriveCLUST algorithm [18] was employed to detect driver genes in LUAD based on the somatic mutation data. One gene with false discovery rate (FDR) < 0.05 was considered as the candidate driver gene. The candidate driver genes that observed in all three AMES groups (low, moderate, and high) were considered as the driver gene.

2.5. Identification of mutagenesis signatures in LUAD

The LUAD-based somatic mutation data were analyzed using the nonnegative matrix factorization (NMF) technique [19] to develop a mutagenesis signature. When the cophenetic correlation coefficient's magnitude reduced noticeably, the optimal factorization k value was identified. By calculating cosine similarity based on the mutational catalogue in the Catalogue of Somatic Mutations in Cancer (COSMIC) database, mutagenesis signatures were developed [20].

2.6. Pathway analysis

DNA damage repair (DDR) pathways were obtained from Molecular Signatures Database (MSigDB) [21], including CRF, HRR, FA, NER, MMR, NHEJ, BER, and TLS. Hallmark pathways (h.all.v7.5.symbols.gmt) were downloaded from MSigDB database. The enrichment score of all hallmark pathways and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways was determined by performing gene set enrichment analysis (GSEA) [22].

2.7. Differential analysis between low- and high-AMES groups

We utilized limma R package [23] to screen differentially expressed genes (DEGs) between low- and high-AMES groups. ClusterProfiler R package [24] was used to annotate enriched KEGG pathways and Gene Ontology (GO) terms of upregulated and downregulated DEGs. STRING online tool [25] and Cytoscape (v3.9.1) [26] were applied to analyze the interactions of the DEGs and construct a protein-protein interaction (PPI) network. Key modules were classified by MCODE tool, and the functional analysis of the key modules was conducted by ClusterProfiler R package.

2.8. Determination of an AMES-related gene signature

The DEGs between high- and low-AMES groups were subjected to a univariate-variable Cox regression analysis. Prognostic genes were identified by screening the DEGs with $P < 0.05$. The prognostic genes were then performed with LASSO regression analysis [27] to find the minimum number of prognostic genes required to build a model. The model or gene signature was defined as: APOBEC mutagenesisrelated risk score (AMrs) $= \Sigma Exp(i)*coef(i)$, where i represents genes, exp represents the expression levels, and coef represents the LASSO coefficients.

2.9. Validation and clinical application of the gene signature

For each LUAD sample in the training dataset (TCGA-LUAD), AMrs was computed. The optimal cut-off value of AMrs for classifying samples into low-AMrs and high-AMrs groups was determined using the Survminer R program. To analyze receiver operating characteristic (ROC) curves, the TimeROC R package [28] was used. In order to compare the overall survival of the low-AMrs and high-AMrs groups, Kaplan-Meier survival analysis was employed. The validation datasets were GSE72094, GSE31210, and GSE50081. A nomogram based on AMrs and clinical traits was created using the "rms" package [29]. An examination of decision curves was performed to test the effectiveness of the nomogram.

2.10. Comparison of immune characteristics between low- and high-AMrs groups

The estimated proportion of immune cells was calculated by CIBERSORT algorithm [30]. Pearson correlation analysis was performed in "Hmisc" R package [31] to analyze the relation of AMrs with immune cell infiltration. The gene sets of T cell inflamed gene expression profile (GEP) score [32], Th1/TNF-γ [33], and cytolytic score [34] were obtained from previous studies.

2.11. Cell culture and transfection

From Zhong qiao xin zhou Biotechnology Co. (Shanghai, China), BEAS-2B, A549 and H2009 cell lines were purchased. Cell culturing was performed in Dulbecco's modified Eagle's medium (DMEM, Gibco, USA) or RIPM1640 medium supplemented with fetal bovine serum (FBS, Gibco, USA) and penicillin/streptomycin with 5 % CO_2 at the temperature of 37 °C. Utilizing Lipofectamine 2000 (Invitrogen, USA), the negative control (si NC), LYPD3 siRNA, ANLN siRNA, MUC5B siRNA, and FOSL1 siRNA (Sagon, China) were transfected into the cells.

The target sequences of siRNA are shown in Table 1. According to the manufacturer's instructions, siRNAs were transfected with Lipofectamine 2000 reagent (Invitrogen, USA).

2.12. Quantitative real-time polymerase chain reaction (qRT-PCR)

Total RNA from BEAS-2B, A549 and H2009 cell lines (Thermo Fisher, USA) was extracted using TRIzol reagent. cDNA was created from 500 ng of RNA using the HiScript II SuperMix (Vazyme, China). With the use of the SYBR Green Master Mix, qRT-PCR was carried out in ABI 7500 System (Thermo Fisher, USA). 46 cycles of 94 °C for 10 min, 94 °C for 10 s, and 60 °C for 45 s each comprised the PCR amplification conditions. GAPDH was an internal reference. Table 2 lists the sequences of primer pairs for targeted genes.

2.13. Transwell assay

The invasion of the BEAS-2B, A549, and H2009 cell lines was tested using transwell assays. To test for invasion, cells (5 × 104) were implanted into chambers covered with Matrigel (BD Biosciences, CA). In the upper layer, serum-free media was supplied, and in the lower layer, full DMEM medium was prepared. 4 % paraformaldehyde was used to fix the migrating or invading cells, which were then stained with 0.1 % crystalline violet following a 24-h incubation period. Cells were counted using a light microscope.

2.14. Cell viability

Following the instructions, cell viability was detected by performing Cell Counting Kit-8 test (Beyotime, China). Cells (1×10^3 cells per well) from various treatments were grown in 96-well plates. The solution of CCK-8 was used at the specified time intervals. The O.D 450 values of each well were determined using a microplate reader (Thermo Fisher, USA) following a 2-h incubation at 37 °C.

2.15. Statistical analysis

The statistical analyses were conducted in R software (v4.1). The bioinformatics analysis was supported by Sangerbox platform [35]. Wilcoxon rank sum test was performed to examine the significant difference between two groups. ANOVA or Kruskal-Wallis test were conducted in comparing the difference among three or over three groups.

3. Results

3.1. The expression pattern of APOBEC family in different cancer types

To comprehensively analyze the expression of APOBEC family genes in human tissues and cancer samples in order to be able to delve into the role of APOBEC genes in human cancers. We firstly evaluated the expression of 11 APOBEC family genes in various tissues or organs of normal samples. APOBEC3C and APOBEC3G were widely expressed in different tissues, while some of APOBEC family genes were specifically or highly expressed in certain tissues (Fig. 1A). For example, APOBEC1 and AICDA are specifically expressed in small intestine and blood respectively. APOBEC3A was highly expressed in blood and spleen. Heart and muscle had the most expression of APOBEC2. However, the expression pattern of APOBEC family significantly altered in cancer samples (Fig. 1B).

Table 1
Target sequences of small interfering (si) RNA.

siRNA	Target sequences
LYPD3-si	ATCAGTTTCTGGCACATAAATGC
ANLN-si	GAGAGAATCTTCAGAGAAAAATG
MUC5B-si	CGGAGACTTTGAGACGTTTGAAA
FOSL1-si	GACTGACAAACTGGAAGATGAGA

Table 2
The primers of genes.

Gene	Forward primer sequence (5′-3′)	Reverse primer sequence (5′-3′)
LYPD3	GATGCTCCCCGAACAAGATGA	CAGCGAGAATTGTCCGTGGAT
ANLN	ATCTTGCTGCAACTATTTGCTCC	TCCTGCTTAACACTGCTGCTA
MUC5B	GCCTACGAGGACTTCAACGTC	CCTTGATGACAACACGGGTGA
FOSL1	CAGGCGGAGACTGACAAACTG	TCCTTCCGGGATTTTGCAGAT
GAPDH	CTGGGCTACACTGAGCACC	AAGTGGTCGTTGAGGGCAATG

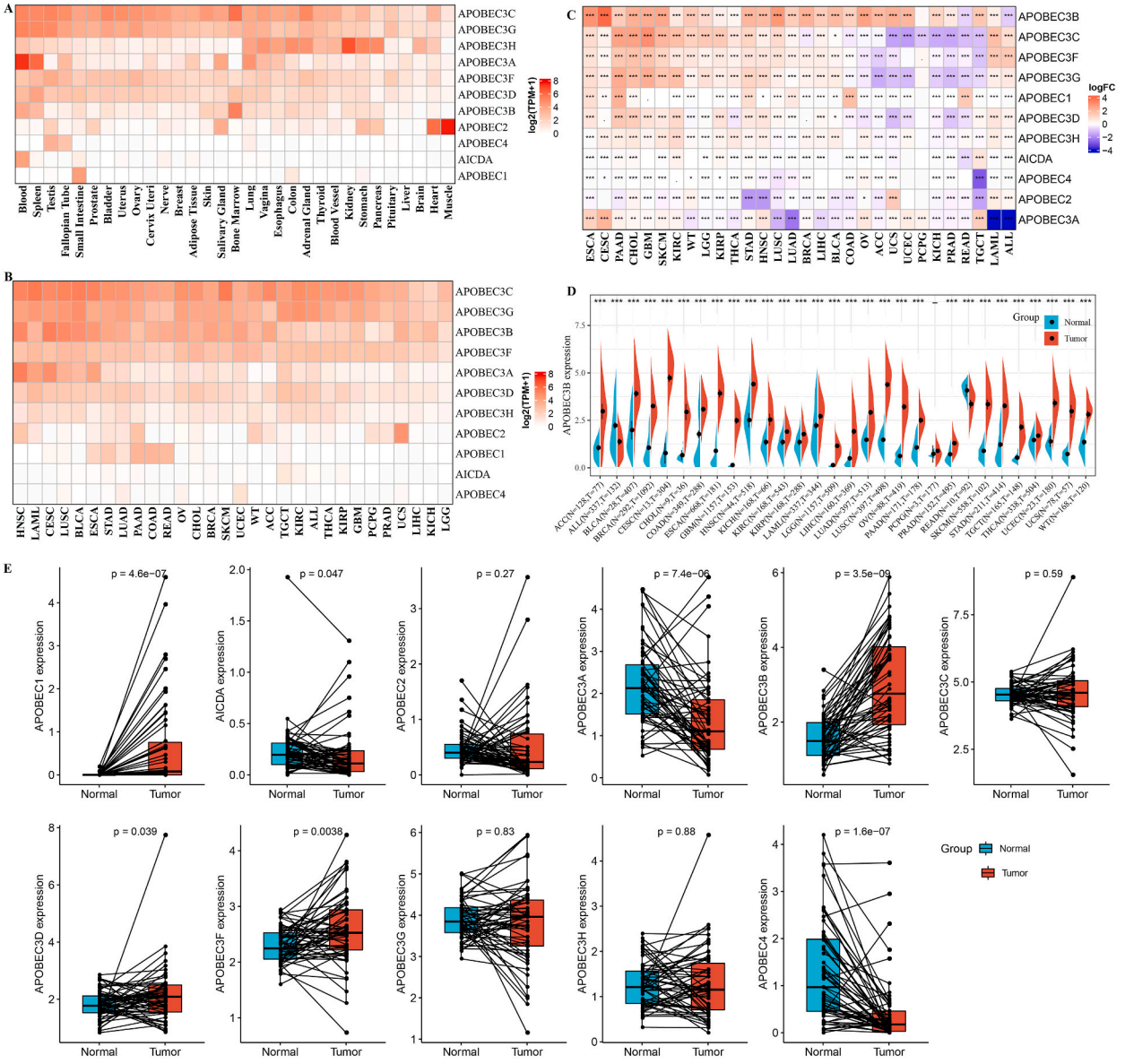
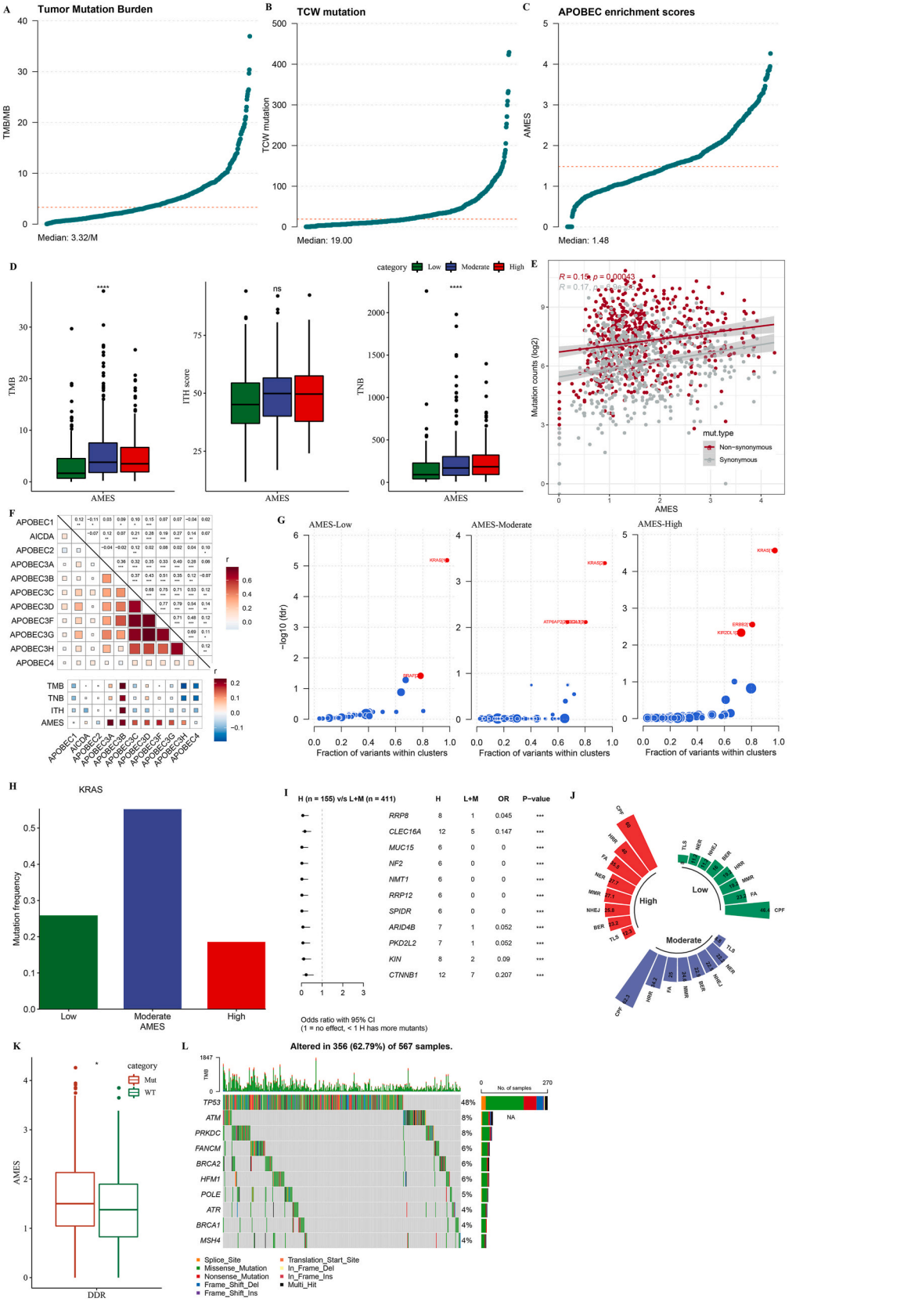


Fig. 1. The mRNA expression of 11 APOBEC family genes in different tissues and cancer types. (A) The expression of APOBEC family genes in normal tissues based on data from the GTEx database. (B) The expression pattern of APOBEC family genes across 30 solid cancer types from TCGA database. (C) A heatmap of APOBEC family genes in the comparison between tumor tissues and adjacent normal tissues. (D) The expression of APOBEC3B in 29 solid cancer types and corresponding normal tissues. (E) The expression of APOBEC family of paired LUAD samples and adjacent normal tissues in TCGA-LUAD dataset. Wilcoxon sum rank test was performed in two groups. *P < 0.05, **P < 0.01, ***P < 0.001.



(caption on next page)

Fig. 2. Mutagenesis features related to AMES in TCGA-LUAD dataset. (A–C) TMB, TCW mutation count and AMES in TCGA-LUAD dataset. Dashed red lines indicate the median values. The horizontal axis represents LUAD samples. (D) TMB, ITH, and TNB in low, moderate and high AMES groups. ANOVA was conducted. (E) Pearson correlation analysis between AMES and mutation counts of synonymous and non-synonymous groups. (F) Pearson correlation analysis of APOBEC family with TMB, TNB, ITH, and AMES. Red and blue indicate positive and negative correlation respectively. (G) Identification of driver genes in low, moderate, and high AMES groups by OncodriveCLUST. Each dot represents a gene and red dots represent the identified driver genes. (H) KRAS mutation frequency in low, moderate, and high AMES groups. (I) Significantly mutated genes between high AMES and other two groups. L, M, H represent low, moderate and high AMES groups respectively. Fisher exact test was performed. (J) The proportion of genes in eight DDR pathways in three AMEA groups. (K) Comparison of AMES between DDR-mut and DDR-WT groups. Wilcoxon test was conducted. (L) The top 10 mutated genes in eight DDR pathways. ns, not significant. *P < 0.05, ***P < 0.001, ****P < 0.0001.

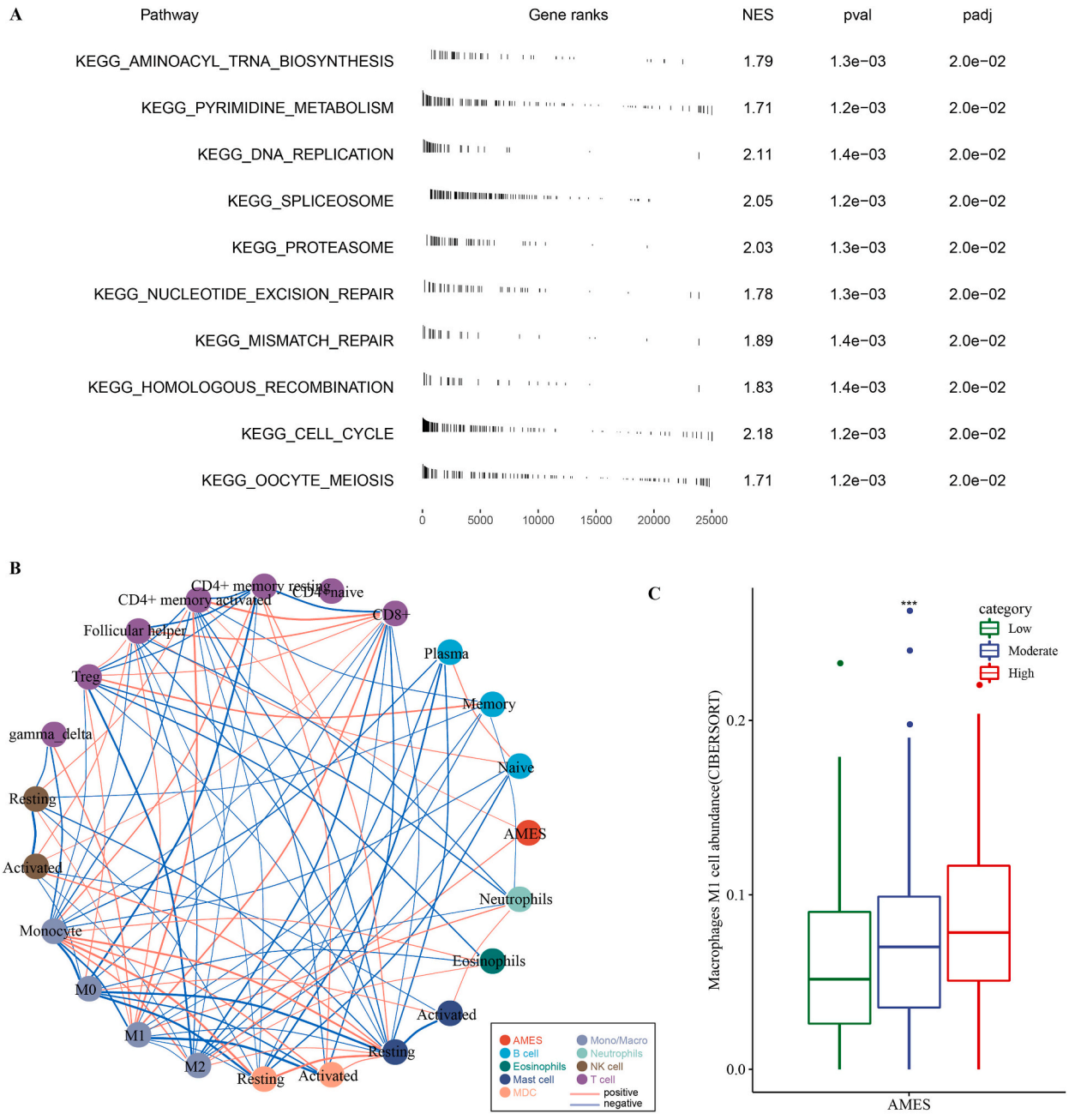


Fig. 3. Biological pathways and immune cells related to AMES. (A) Enriched KEGG pathways in high AMES group assessed by GSEA. (B) A correlation network among immune cells and AMES. The thickness of lines represents the correlation degree that thicker lines represent stronger correlation. (C) The distribution of M1 macrophages in three AMES groups. ANOVA was conducted. ***P < 0.001.

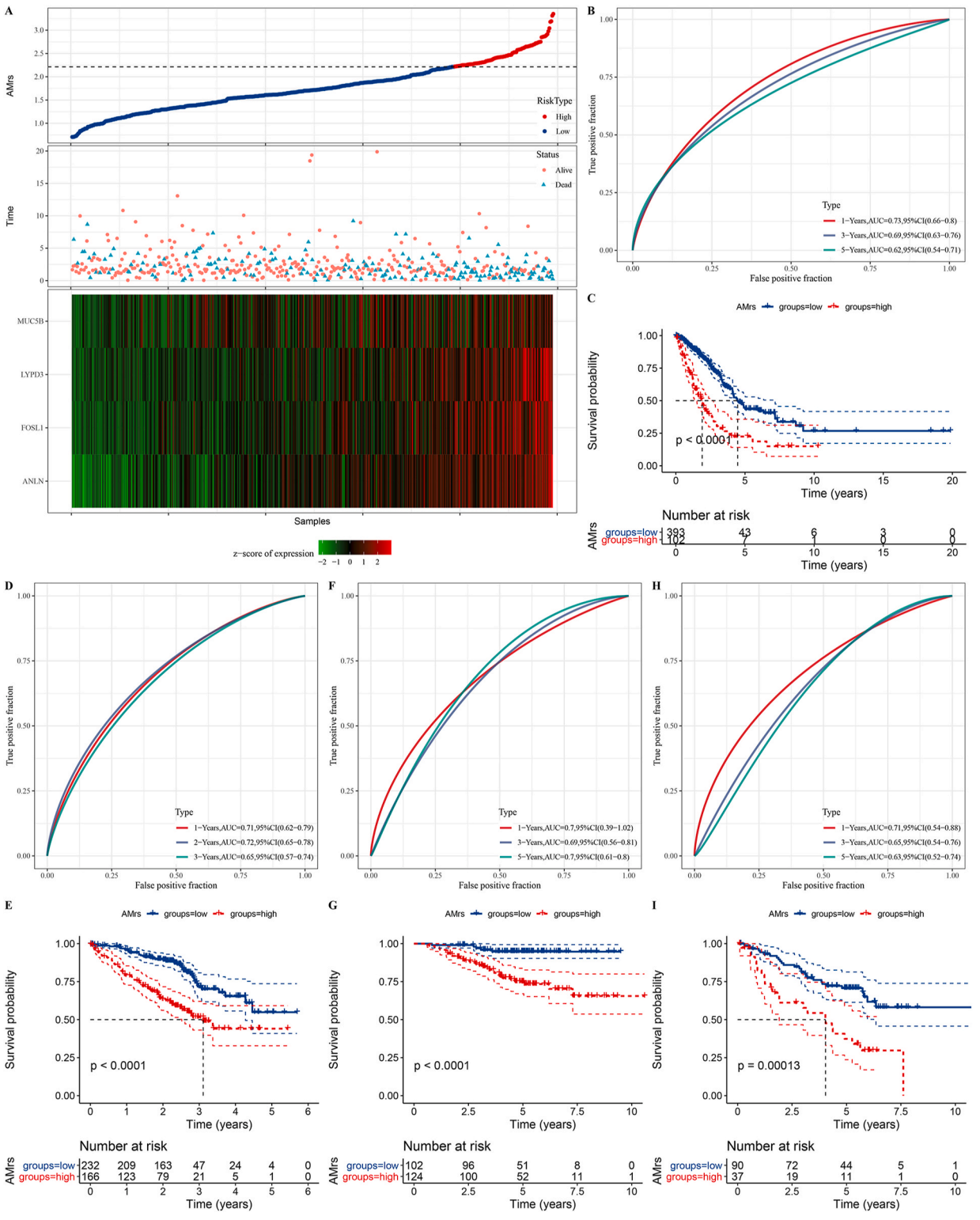


Fig. 4. Validation of the 4-gene risk signature. (A) AMrs of each LUAD corresponds to survival status and expression levels of four genes in TCGA-LUAD dataset. (B) ROC curve of the signature in predicting 1-year, 3-year and 5-year overall survival in TCGA-LUAD dataset. (C) Kaplan-Meier survival analysis on high- and low-AMrs groups in TCGA-LUAD dataset. (D, F, H) ROC curve of the signature in predicting 1-year, 3-year and 5-year overall survival in GSE72094 (D), GSE31210 (F), and GSE50081 (H) datasets. (E, G, I) Kaplan-Meier survival analysis on high- and low-AMrs groups in GSE72094 (E), GSE31210 (G), and GSE50081 (I) datasets.

APOBEC1 was highly expressed in esophageal cancer (ESCA), stomach adenocarcinoma (STAD), pancreatic adenocarcinoma (PAAD), colon adenocarcinoma (COAD), and rectal adenocarcinoma (READ). This suggests that the expression of APOBEC family members is specific across cancer types. APOBEC3C and APOBEC3G were still the two of mostly expressed genes, but the expression of APOBEC3B

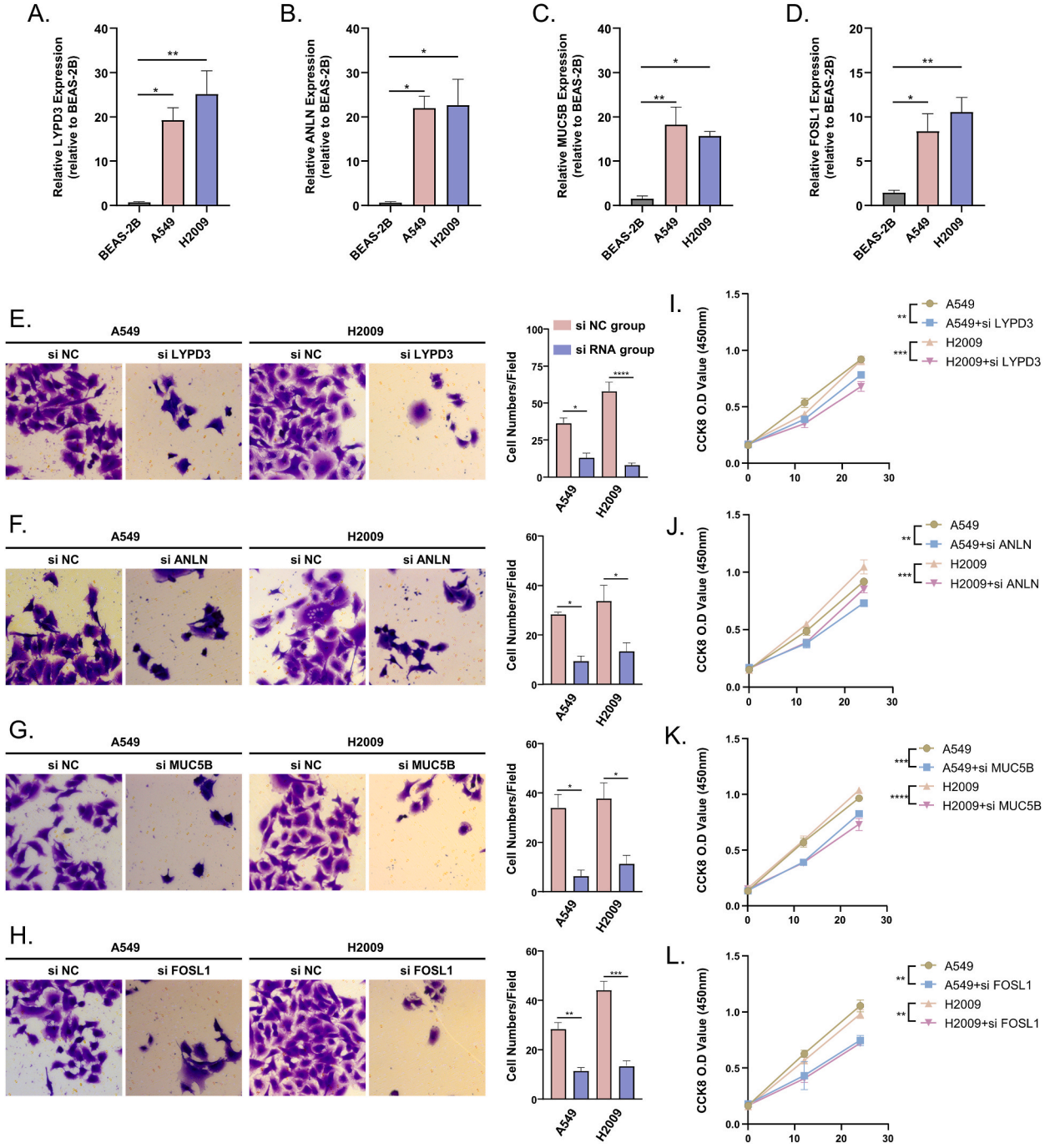


Fig. 5. LYPD3, ANLN, MUC5B, and FOSL1 promote the viability of lung cancer. (A–D) RT-qPCR assay to measure the expression levels of LYPD3, ANLN, MUC5B and FOSL1 markers in BEAS-2B, A549 and H2009 cell lines (n = 3). (E–H) After suppressing the expression of LYPD3, ANLN, MUC5B, and FOSL1, representative cell invasion detection results were obtained for A549 and H2009 cell lines (n = 3). (I–L) After suppressing the expression of LYPD3, ANLN, MUC5B, and FOSL1, representative CCK8 results were obtained for A549 and H2009 cell lines (n = 3). *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001. The results are presented as mean ± S.M.E.

seemed to be activated in more tissues in cancer samples. This expression alternation was verified by comparing cancer samples with paracancerous samples (Fig. 1C). The majority of APOBEC family genes were significantly dysregulated in different cancer types. Especially, APOBEC3B expression was drastically upregulated in most of cancer types besides READ and acute lymphoblastic leukemia (ALL) (Fig. 1C and D). A massive change of APOBEC3A expression was shown in acute myeloid leukemia (LAML) and ALL. For LUAD, the expression of APOBEC3B and APOBEC3A was dramatically altered. In addition, we specifically delineated the expression of APOBEC family genes in LUAD and paracancerous samples (Fig. 1E). Seven of genes showed significantly differential expression levels between LUAD and paracancerous samples, including APOBEC1 (P = 4.6e-07), AICDA (P = 0.047), APOBEC3A (P = 7.4e-06), APOBEC3B (P = 3.5e-09), APOBEC3D (P = 0.039), APOBEC3F, and APOBEC4 (P = 1.6e-07), where the highest expression change was observed in APOBEC3B.

3.2. *Characterizing genetic variations of APOBEC family in LUAD*

The APOBEC mutation pattern was demonstrated to link with cancer development [12]. In pan-cancer, endometrial cancer (UCEC) and COAD showed a relatively high mutation frequency of non-silent somatic mutations of APOBEC family compared to other cancer types (Fig. S1A). In 565 LUAD samples, only 27 samples had mutations of APOBEC family, and mutant (mut) group had longer overall survival than wild-type (WT) group but no statistical significance was shown (Figs. S1B and S1C). Then we evaluated the copy number variations (CNVs) of APOBEC family in LUAD, and observed that APOBEC2 and APOBEC4 had high frequencies of CNV gain while APOBEC1 and AICDA had high frequencies of CNV loss (Fig. S1D). Besides APOBEC3G and APOBEC3H, other nine APOBEC family genes had differential expression among CNV and normal groups (Fig. S1E). CNV loss (deletion) group commonly showed a lower expression level than CNV gain and diploid groups but it was not applicable in APOBEC3B. CNV loss group had extremely higher expression of APOBEC3B than diploid group as well as normal samples.

By using a calculation method that could characterize the mutagenesis of APOBEC family developed by Roberts et al. [12], we assessed AMES, as well as TMB, TCW (TCW-TTW and TCW-TGW) mutation, TNB, and ITH score for each LUAD sample. LUAD samples were classified into three groups (low, moderate, and high AMES) according to AMES. We observed a trend that TMB, ITH, and TNB

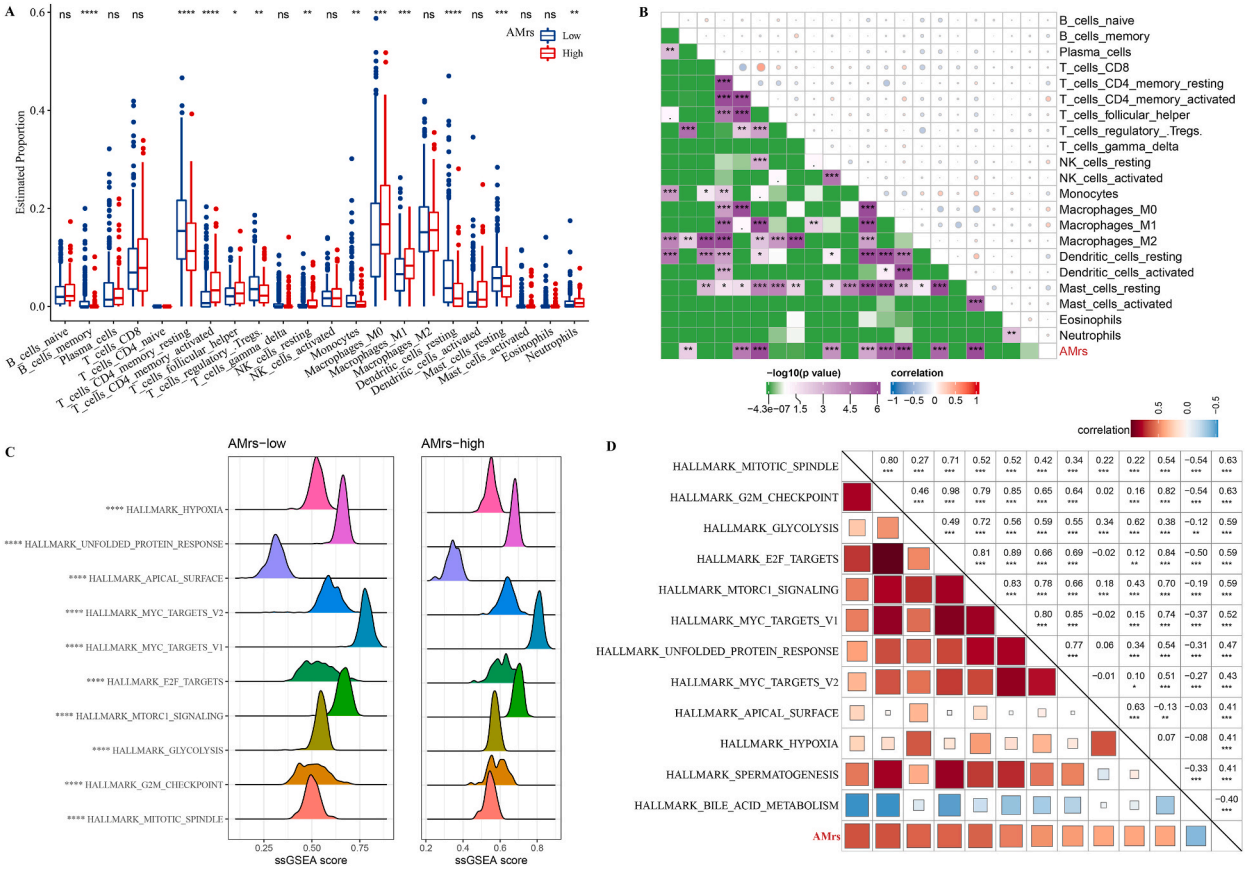


Fig. 6. The difference of immune infiltration and functional pathways between high- and low-AMrs groups in TCGA-LUAD dataset. (A) The estimated proportion of immune cells in two risk groups by CIBERSORT analysis. Wilcoxon test was conducted. (B) Pearson correlation analysis of AMrs with the enrichment of immune cells. (C) The top 10 differentially enriched hallmark pathways between high- and low-AMrs groups. (D) Pearson correlation analysis between AMrs and hallmark pathways. ns, not significant. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

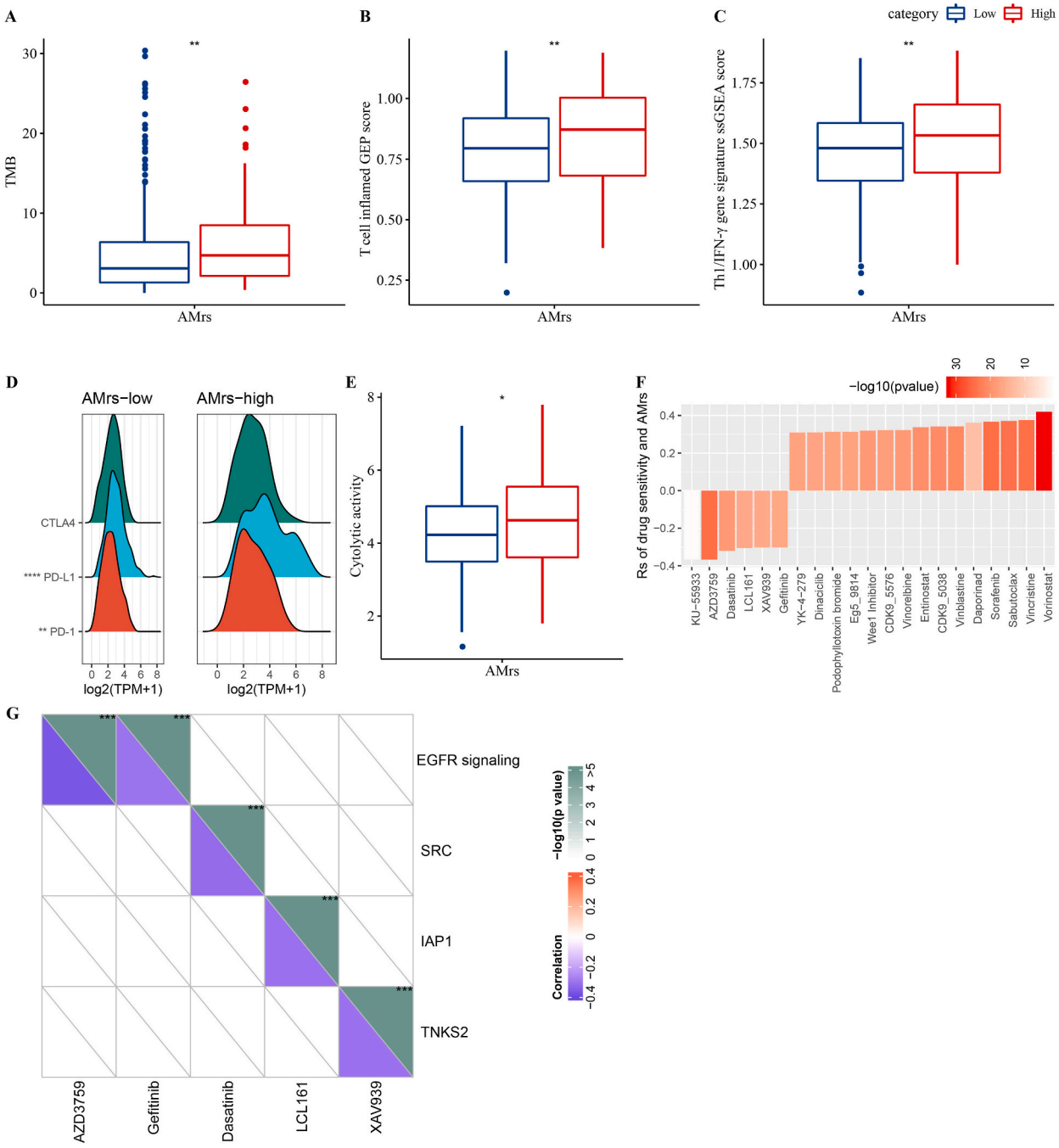


Fig. 7. The predicted response of LUAD to immunotherapy and chemotherapy in TCGA-LUAD dataset. (A–C) The score of TMB, T cell inflamed GEP, and Th1/TNF-γ signature in high- and low-AMrs groups. Wilcoxon test was conducted. (D) Comparison of CTLA4, PD-L1, and PD-1 expression between high- and low-AMrs groups. (E) The cytolytic activity in two risk groups. Wilcoxon test was conducted. (F) The Spearman correlation between AMrs and drugs. Each column represents a drug. The vertical axis indicates the correlation coefficients (Rs), where Rs > 0 represents drug resistance and Rs < 0 represents drug sensitivity. (G) The targeted signaling pathways (vertical axis) of drugs (horizontal axis). *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

scores elevated with the increasing AMES (Fig. 2A–D). In addition, AMES was positively correlated with the counts of both non-synonymous and synonymous mutation ($R = 0.15$, $P < 0.001$; $R = 0.17$, $P < 0.001$, respectively, Fig. 2E). We also performed correlation analysis between each APOBEC family gene and four mutagenesis indicators (Fig. 2F). Notably, the scores of TMB, TNB, ITH, and AMES were all strongly correlated with APOBEC3B (Fig. 2F). Compared to TMB, TNB and ITH, AMES was shown to have stronger correlation with APOBEC family genes.

To explore the cancer driver genes involved in LUAD, we applied OncodriveCLUST algorithm in low, moderate, and high AMES groups. As a result, KRAS was identified as the driver gene in all three groups (Fig. 2G). Moderate AMES group had the highest mutation rate of KRAS whereas the lowest frequency in high AMES group (Fig. 2H). Compared to low and moderate AMES groups, a fraction of genes were significantly mutated in the high AMES group, such as CLEC16A, RRP8, and MUC15 (Fig. 2I).

Previous research reported that favorable prognosis of immune checkpoint inhibition in LUAD was associated with the alterations of DDR genes [36]. Therefore, we compared the mutation frequency of eight DDR pathways in three AMES groups. High AMES group showed the most mutations of DDR pathways, where CPF pathway genes were the most frequently altered (Fig. 2J). DDR-mut group also had significantly higher AMES than DDR-WT group ($P < 0.05$, Fig. 2K). The top 10 frequently mutated genes in DDR pathways were visualized where TP53 consisted of a large mutation proportion (Fig. 2L). Furthermore, NMF methodology identified four mutational signatures in LUAD including "exposure to tobacco (smoking) mutagens", "APOBEC Cytidine Deaminase (C > T)", "defective DNA mismatch repair", and "unknown" where "APOBEC Cytidine Deaminase (C > T)" signature had the strongest correlation with AMES ($R = 0.94$, Fig. S2).

3.3. Biological pathways related to AMES in LUAD

To understand the role of AMES in LUAD, we analyzed KEGG pathways using GSEA, and identified some cell cycle-correlated pathways showing significant enrichment in high AMES group, for example, DNA replication, spliceosome, mismatch repair, cell cycle, and nucleotide excision repair (Fig. 3A). A total of 162 DEGs were identified in high AMES group with 95 upregulated and 67 downregulated (Fig. S3A). Moreover, cell cycle pathway was significantly enriched in upregulated genes, while cell communication-related pathways were enriched in downregulated genes (Fig. S3B). Moreover, we applied STRING to delineate the interaction among these DEGs, and a PPI network was constructed. Within the network, four main modules (cluster 1 to 4) were screened and cluster 1 module had the tightest interaction (Fig. S3C). Functional enrichment analysis on cluster 1 module showed that cell cycle-related terms were enriched (Fig. S3D), suggesting that the genes in cluster 1 may have an important connection with AMES.

In addition to biological pathway analysis, we also assessed the immune cell infiltration in LUAD. A correlation network among immune cells and AMES was constructed (Fig. 3B). We found that AMES was significantly and positively correlated with $CD4^+$ memory activated and M1 macrophage abundance only in the three subtypes. Therefore, we assessed the differences in the level of M1 macrophage infiltration in different AMES groups. The results shown that the infiltration of M1 macrophages increased with the AMES level ($P < 0.001$, Fig. 3C), indicating that AMES may be involved in the modulation of tumor immune microenvironment.

3.4. Development and validation of an AMES-based gene signature

In the previous section, we screened 162 DEGs between high and low AMES groups. A total of 72 prognostic DEGs (59 risk genes and 13 protective genes) were identified by univariate Cox regression analysis (Fig. S4A). To construct a risk model with the least number of genes, we conducted LASSO regression on the 72 genes. The model came to an optimal one when lambda = 0.0659, and four prognostic genes were remained including LYPD3, ANLN, MUC5B, and FOSL1 (Figs. S4B–D). Therefore, an AMES-related risk model (gene signature) was established based on the expression of LASSO coefficients of the four genes.

The APOBEC mutagenesis-related risk score (AMrs) was then determined for each LUAD sample in the TCGA-LUAD dataset. According to the ideal AMrs cut-off value, sample grouping into low- and high-risk groups was performed (Fig. 4A). The number of LUAD samples with a death status rose along with the rising AMrs. Four prognostic genes had a high level of expression in the high-risk group. The gene signature performed well in the prediction of 1-, 3-, and 5-year survival, with AUC values of 0.73, 0.69, and 0.62, respectively (Fig. 4B). The signature reliability that two risk groups had unique overall survival was further validated by Kaplan-Meier survival analysis ($P < 0.0001$, Fig. 4C). The robustness of the 4-gene signature was evaluated in another three independent datasets (GSE72094, GSE31210, and GSE50081). Expectedly, the gene signature was also effective to determine the risk level of LUAD samples in the three datasets (Fig. 4E–I). Of the AMrs in different clinical characteristics, AMrs trend corresponded to the increased AJCC stages (Fig. S5A). That was, higher stage levels had higher AMrs. Notably, higher AMrs was presented in high AMES group than low AMEA group (Figs. S5B and S5C). Besides, the prognosis of two risk groups also varied in LUAD samples with different stages, genders and ages (Fig. S5D), which further demonstrated the effectiveness of the gene signature.

3.5. LYPD3, ANLN, MUC5B, and FOSL1 promote the invasion of lung cancer

Firstly, we performed RT-qPCR experiments to detect the expression of LYPD3, ANLN, MUC5B, and FOSL1 in BEAS-2B, A549, and H2009 cell lines. The results showed a significant increase in the expression levels of LYPD3, ANLN, MUC5B, and FOSL1 in A549, and H2009 cell lines in comparison to BEAS-2B cells (Fig. 5A–D). Subsequently, we separately suppressed the expression of LYPD3, ANLN, MUC5B, and FOSL1 in A549 and H2009 cell lines. The results demonstrated a significant decrease in the invasive ability of lung cancer cell lines after the suppression of these four makers (Fig. 5E–H). Finally, we detected the vitality of lung cancer cell lines using CCK8. The results indicated a significant decrease in the vitality of lung cancer cell lines after the suppression of the aforementioned four

makers (Fig. 5I-L).

3.6. Different immune infiltration and enriched pathways between two risk groups

The characteristics of tumor immune microenvironment is a critical factor affecting the prognosis and the response to immunotherapy. Therefore, we parsed the relation between immune infiltration and AMrs in TCGA-LUAD dataset. CIBERSORT result estimated that the two risk groups had different enrichment of the 12 immune cells. Especially, high-risk group had significantly higher proportion of M0 and M1 macrophages, activated memory CD4 T cells (Fig. 6A). Consistently, Pearson correlation analysis on the relationship between AMrs and immune cells demonstrated that resting mast cells, resting memory CD4 T cells, resting dendritic cells were negatively correlated with AMrs, while the enrichment of M0 and M1 macrophages, activated memory CD4 T cells was positively correlated with AMrs (Fig. 6B). Functional analysis of two risk groups unveiled that cell cycle-related pathways such as MYC targets v1, MYC targets v2, E2F targets, and G2M checkpoint were significantly different in two groups ($P < 0.0001$, Fig. 6C). Two important tumor-related pathways, hypoxia and glycolysis, were more activated in high-AMrs group than in low-AMrs group, which may contribute to the poor prognosis of high-risk group. A positive correlation of AMrs with hypoxia and glycolysis was also exhibited (Fig. 6D).

3.7. The predictive value of AMrs for immunotherapy and chemotherapy in LUAD

Previous studies documented that high TMB and T cell inflamed GEP score were associated with positive response to immune checkpoint blockade [37]. In two risk groups, high-AMrs group had both significantly higher score of TMB and T cell inflamed GEP than low-AMrs group (Fig. 7A and B), indicating high-AMrs group may benefit much from immune checkpoint inhibitors. Th1/IFN-γ that was involved in tumor clearance and tumor escape showed different enrichment in two risk groups [38] (Fig. 7C). In addition, compared with low-AMrs group, the expression of two important immune checkpoints (PD-L1 and PD-1) and the cytolytic activity were higher in high-AMrs group (Fig. 7D and E). The above results suggested that high-AMrs group was more sensitive to immunotherapy. In the relation between AMrs and the sensitivity to chemotherapeutic drugs, we identified 21 drugs whose drug resistance (Rs

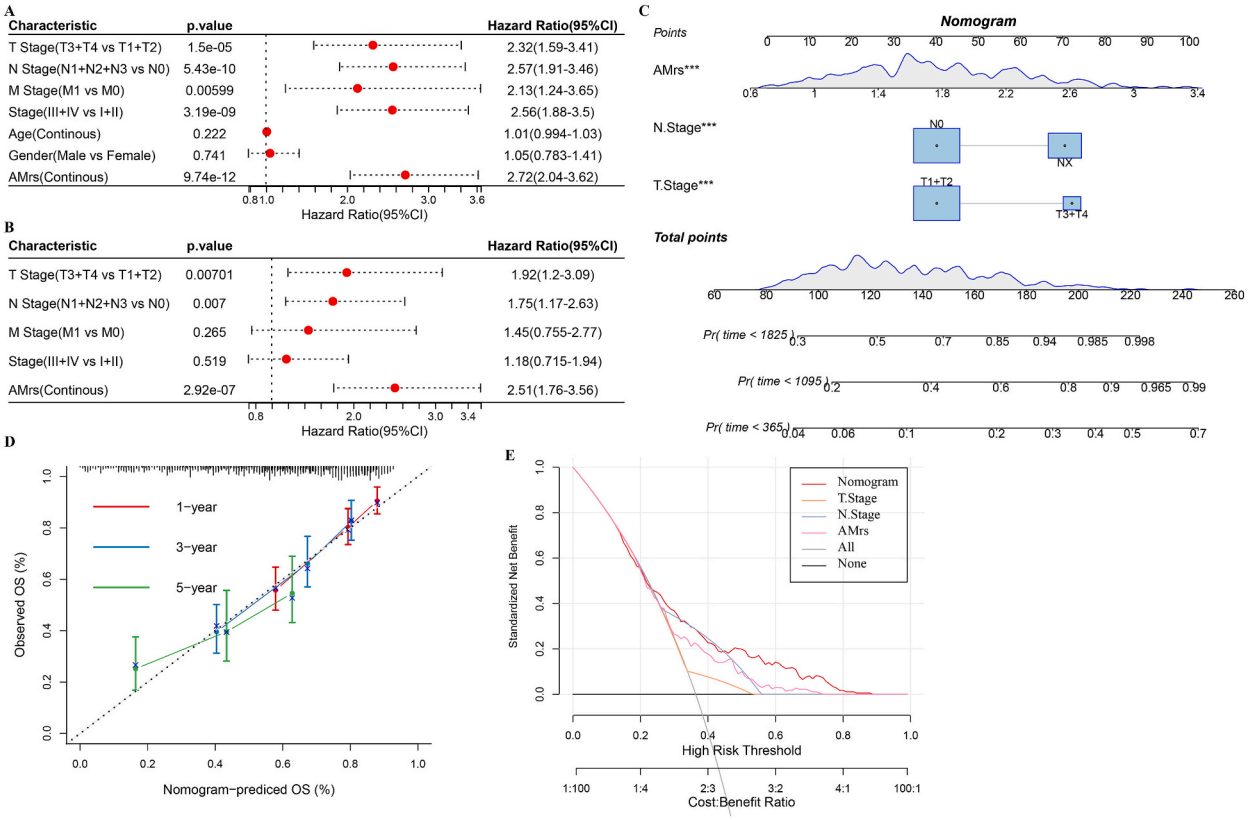


Fig. 8. Construction of a nomogram based on AMrs and clinical characteristics. (A–B) Examination of the prognostic value of AMrs and clinical characteristics by univariate (A) and multivariate (B) Cox regression analysis. Log-rank test was conducted. (C) A nomogram was constructed based on AMrs, N stage, and T stage. (D) Calibration curve for the comparison between nomogram-predicted OS and observed OS. (E) Decision analysis curve of the nomogram, T stage, N stage, and AMrs. ***$P < 0.001$.

> 0) or drug sensitivity (Rs < 0) was significantly associated with AMrs (Fig. 7F). High-AMrs group may benefit from six drugs including KU-55933, AZD3759, dasatinib, LCL161, XAV939, and gefitinib. As shown in Fig. 7G, we found that these drugs mainly targeted EGFR signaling, SRC, IAP1, and TNKS2 pathways.

3.8. Analysis of oxidative stress

Oxidative stress plays an important role in the normal regulation of a variety of physiological processes, including apoptosis, survival, etc. In particular, oxidative stress has been shown to play a key role in the progression of lung cancer [39]. To further explore the relationship between AMES and oxidative stress in the development of LUAD, we obtained oxidative stress-related biological processes (BP) from the GO database, and finally included 20 oxidative stress-related pathways. The enrichment scores of oxidative stress-related pathways in TCGA-LUAD cohort were quantified by single-sample gene set enrichment analysis (ssGSEA). Correlation analysis between oxidative stress score and clinicopathological features in TCGA-LUAD cohort showed that Age, N Stage, and Stage were significantly correlated with some oxidative stress pathways, but the correlation was weak. GOBP_CELLULAR_RESPONSE_TO_OXIDATIVE_STRESS pathway showed a significant positive correlation with Age (Fig. S6A). High-risk group had higher oxidative stress scores (Fig. S6B). AMES showed no significant difference in oxidative stress pathways (Fig. S6C). Further, we analyzed the correlation between the expression of key genes and oxidative stress pathways and the risk score. Different oxidative stress pathways were found to be significantly positively correlated with each other. Additionally, our risk score showed significant positive correlation with most oxidative stress pathways. We also observed significant positive correlations between risk genes and some oxidative stress-related pathways (Fig. S6D).

Combination of AMrs and clinical characteristics to increase the accuracy for predicting prognosis in LUAD.

AMrs was confirmed as the most significant risk factor by univariate and multivariate Cox regression analysis on clinical characteristics and AMrs (Fig. 8A and B). Besides, T stage and N stage were also the independent risk factors of LUAD prognosis. Then we included AMrs, T stage and N stage to develop a nomogram for predicting 1-year, 3-year and 5-year overall survival or death rate (Fig. 8C). The nomogram could well predict overall survival rate, which was similar to the observed one, as shown by the calibration curve (Fig. 8D) and was further verified by the decision curve analysis (Fig. 8E), implying the nomogram as an optimal tool for predicting LUAD prognosis.

4. Discussion

This research demonstrated that the expression levels of APOBEC family genes were greatly altered across 30 cancer types. Especially, APOBEC3B expression was the most aberrant in different cancer tissues compared to normal tissues, which was upregulated in most of cancer types. Previous findings have revealed that APOBEC3B acts as a driver and tumorigenic role involved in mutagenesis process in various cancer types [40], and the overexpression of APOBEC3B was found in breast cancer [14,41], gastric cancer [42], hepatocellular carcinoma [43], renal cancer [44], and lung cancer [45,46], etc. High expression of APOBEC3B was associated with a poor overall survival or increased recurrence in multiple cancers [44,47]. In non-small cell lung cancer patients, APOBEC3B expression was correlated with clinical characteristics including nodal status and TNM staging, and the outcome of NSCLC patients receiving adjuvant chemotherapy [46]. Our result of pan-cancer analysis on APOBEC family expression pattern is consistent with the previous research, which lays a basis for the further analysis of APOBEC family dysregulation in LUAD.

Lung cancer is a common malignant tumor that often leads in morbidity and mortality [48]. In the subsequent analysis, we focused on the role of APOBEC family in LUAD. Besides APOBEC3B, other six APOBEC family genes were also significantly dysregulated in LUAD, including three upregulated genes (APOBEC1, APOBEC3D, and APOBEC3F) and three downregulated genes (AICDA, APOBEC3A, and APOBEC4). AICDA was reported to derive lymphomagenesis [49] and APOBEC1 was suggest to have oncogenic potential in the early stages of esophageal adenocarcinoma [50]. APOBEC3 proteins are the most powerful to convert cytosine in single strand DNA (ssDNA) to uracil (C-to-U) through deamination process, which are the major endogenous source of promoting genomic instability and increasing TMB [51]. Qi et al. analyzed 2938 Chinese female cases (including 1374 epithelial ovarian cancer cases and 1564 healthy samples) in order to assess the association between APOBEC3 deletion and ovarian cancer risk. The results showed that patients carrying APOBEC3 deletion gene were more likely to suffer from epithelial ovarian cancer [52]. APOBEC4 was shown to enhance transcription from a broad-spectrum promoter both in mammals and viruses [53]. However, there is still a gap in research on the role of APOBEC4 in cancer.

To better understand the relation between APOBEC mutagenesis effect and genome instability, we introduced AMES calculation method previously reported by Roberts et al. [12]. By using NMF methodology, we identified “APOBEC Cytidine Deaminase (C > T)” as a significant mutagenesis signature having an extremely high correlation with AMES (R = 0.94). The result showed that on one hand, APOBEC-induced mutagenesis was widely presented in LUAD; on the other hand, AMES method was reliable to evaluate the APOBEC mutagenesis. LUAD samples were divided into three groups (AMES-low, moderate and high) according to AMES. In the relation of AMES with genome instability, we observed significantly high TMB and ITH scores in AMES-high group, as well as the positive correlation of APOBEC3B expression with TMB and ITH. The findings suggested that the genomic instability and heterogeneity in LUAD may largely attribute to APOBEC-induced mutagenesis especially APOBEC3B. To support the speculation, we also evaluated the mutagenesis of DDR pathways in three AMES groups. Not surprisingly, AMES-high group displayed the highest mutation frequency of the eight DDR pathways. Furthermore, functional analysis sustained that cell cycle and DNA repair-associated pathways were strikingly enriched in AMES-high group.

Genomic instability has been revealed to be associated with the response to cancer immunotherapy in a number of studies [54,55],

where TMB is a well-known indicator of genomic instability for predicting immunotherapy benefit [56]. Mutation-derived neoantigens are recognized as a potential target for cancer immunotherapy [57,58]. Chen et al. showed that patients with EPHA5 mutations exhibited TMB compared to EPHA5 wild-type patients and concluded that EPHA1 mutations can serve as a prognostic marker for immunotherapy in LUAD patients [59]. Specifically, it has also been found that TMB exhibits significant gender differences in the prediction of immunotherapy response. The predictive ability of TMB in female LUAD patients is significantly due to male LUAD patients [60]. Tumor neoantigens are specifically expressed in tumor cells and are strongly associated with immunogenicity and T cell recognition. In our results, AMES was also significantly correlated with the number of neoantigens and the enrichment of M1 macrophages, which may contribute to different immune microenvironment of LUAD patients with different AMES. Consequently, we considered that AMES could serve as a basis to develop a risk model for predicting LUAD prognosis and the sensitivity to immunotherapy or other therapies.

We identified a fraction of DEGs between AMES-low and AMES-high groups and cell cycle-related pathways were significantly enriched in these DEGs. Four prognostic genes were finally determined by regression analysis, including LYPD3, ANLN, MUC5B, and FOSL1, and they were used to construct a risk model. Upregulated expression of LYPD3 was reported to be associated with poor overall survival in LUAD patients [61]. LYPD3 can regulate ADRβ2 signaling involved in the metastasis in breast cancer, and LYPD3 may serve as a therapeutic target for cancer therapy [62]. ANLN and MUC5B have been abundantly reported to contribute to aggressive tumor phenotypes in multiple cancer types such as breast cancer [63,64], gastric cancer [65,66], colorectal cancer [67], NSCLC [68]. Notably, MUC5B is related to airway defence and regulate airway inflammation in the pathogenesis of airway diseases [69]. FOSL1 is also considered to play a tumorigenic role in different cancer types such as lung cancer and pancreatic cancer [70]. In this study, we used RT-qPCR assay determined the increased 4 genes expression levels in lung cancer cell lines, and knock down genes expressions inhibited lung cancer cell viability and invasion ability. High expression of the four prognostic genes was associated with poor prognosis in cancer as previously reported, which was consistent with our results that they were upregulated in high-risk group with worse prognosis.

ROC analysis and survival analysis demonstrated the reliability and effectiveness of the risk model. However, there are some limitations to our study. First, only data from the GEO and TCGA databases were collected for bioinformatics analysis in our study. For this, we need prospective studies with a large number of samples to validate these results. In addition, the accuracy and reliability of our AMrs model should still be improved with long-term clinical applications.

5. Conclusions

In conclusion, our study comprehensively evaluated the expression pattern of APOBEC family and dug out the association of APOBEC mutagenesis with genomic instability in LUAD. APOBEC family contributed as an oncogenic role possibly through regulating cell cycle and DNA damage-related pathways. Moreover, we identified the APOBEC-related prognostic genes to establish a risk model that was reliable to predict the overall survival and was potential to guide clinical treatment in selecting the patients with different sensitivity to immunotherapy or chemotherapy. The four prognostic genes were possible to serve as the prognostic biomarkers and therapeutic targets for LUAD patients.

Funding

The present study was supported by the Traditional Chinese medicine projects of Shanghai Changning District Health and Family Planning Commission (20154Y019).

Data availability statement

Data is available at Github: https://github.com/AnxinGu/raw-data.git.

Ethnical statement

Informed consent was not required for this study because it is not involved any human experiments.

CRediT authorship contribution statement

Yu Luo: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis. **Huiru Wang:** Writing – review & editing, Validation, Project administration, Methodology, Conceptualization. **Jian Zhong:** Writing – original draft, Visualization, Resources, Methodology. **Jianrong Shi:** Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Xianlin Zhang:** Writing – review & editing, Supervision, Software, Resources, Formal analysis. **Yanni Yang:** Writing – original draft, Resources, Investigation, Formal analysis, Data curation. **Ruixin Wu:** Writing – review & editing, Visualization, Supervision, Resources, Funding acquisition.

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.

Acknowledgement

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e21336.

References

[1] H. Sung, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J Clin 71 (3) (2021) 209–249.
[2] H. Wang, et al., Gene editing in non-small cell lung cancer: current application and future perspective, Oncologie 24 (1) (2022) 65–83.
[3] C. Allemani, et al., Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries, Lancet 391 (10125) (2018) 1023–1075.
[4] B.A. Weir, et al., Characterizing the cancer genome in lung adenocarcinoma, Nature 450 (7171) (2007) 893–898.
[5] C.G.A.R. Network, Comprehensive molecular profiling of lung adenocarcinoma, Nature 511 (7511) (2014) 543–550.
[6] U. Testa, E. Pelosi, G. Castelli, Molecular charcterization of lung adenocarcinoma combining whole exome sequencing, copy number analysis and gene expression profiling, Expert Rev. Mol. Diagn 22 (1) (2022) 77–100.
[7] J. Shi, et al., Advances in targeted therapy against driver mutations and epigenetic alterations in non-small cell lung cancer, Oncologie 24 (4) (2022) 613–648.
[8] P. Wang, Y. Chen, C. Wang, Beyond tumor mutation burden: tumor neoantigen burden as a biomarker for immunotherapy and other types of therapy, Front. Oncol. 11 (2021), 672677.
[9] G. Gao, et al., KRAS G12D mutation predicts lower TMB and drives immune suppression in lung adenocarcinoma, Lung Cancer 149 (2020) 41–45.
[10] L. Ding, et al., Somatic mutations affect key pathways in lung adenocarcinoma, Nature 455 (7216) (2008) 1069–1075.
[11] L.B. Alexandrov, et al., Signatures of mutational processes in human cancer, Nature 500 (7463) (2013) 415–421.
[12] S.A. Roberts, et al., An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers, Nat. Genet. 45 (9) (2013) 970–976.
[13] M.B. Burns, N.A. Temiz, R.S. Harris, Evidence for APOBEC3B mutagenesis in multiple human cancers, Nat. Genet. 45 (9) (2013) 977–983.
[14] M.B. Burns, et al., APOBEC3B is an enzymatic source of mutation in breast cancer, Nature 494 (7437) (2013) 366–370.
[15] S. Nik-Zainal, et al., Mutational processes molding the genomes of 21 breast cancers, Cell 149 (5) (2012) 979–993.
[16] M. Jamal-Hanjani, et al., Tracking the evolution of non-small-cell lung cancer, N. Engl. J. Med. 376 (22) (2017) 2109–2121.
[17] A. Mayakonda, et al., Maftools: efficient and comprehensive analysis of somatic variants in cancer, Genome Res. 28 (11) (2018) 1747–1756.
[18] D. Tamborero, A. Gonzalez-Perez, N. Lopez-Bigas, OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes, Bioinformatics 29 (18) (2013) 2238–2244.
[19] L.B. Alexandrov, et al., Deciphering signatures of mutational processes operative in human cancer, Cell Rep. 3 (1) (2013) 246–259.
[20] J.G. Tate, et al., COSMIC: the catalogue of somatic mutations in cancer, Nucleic Acids Res. 47 (D1) (2019) D941–d947.
[21] A. Liberzon, et al., The Molecular Signatures Database (MSigDB) hallmark gene set collection, Cell systems 1 (6) (2015) 417–425.
[22] A. Subramanian, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc Natl Acad Sci U S A 102 (43) (2005) 15545–15550.
[23] M.E. Ritchie, et al., Limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res. 43 (7) (2015) e47.
[24] G. Yu, et al., clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS 16 (5) (2012) 284–287.
[25] D. Szklarczyk, et al., The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets, Nucleic Acids Res. 49 (D1) (2021) D605–d612.
[26] G. Su, et al., Biological network exploration with Cytoscape 3, Curr Protoc Bioinformatics 47 (2014), 8.13.1-24.
[27] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, J Stat Softw 33 (1) (2010) 1–22.
[28] P. Blanche, J.F. Dartigues, H. Jacqmin-Gadda, Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks, Stat. Med. 32 (30) (2013) 5381–5397.
[29] Y. Liu, et al., A novel nomogram for survival prediction of patients with spinal metastasis from prostate cancer, Spine 46 (6) (2021) E364–e373.
[30] B. Chen, et al., Profiling tumor infiltrating immune cells with CIBERSORT, Methods Mol. Biol. 1711 (2018) 243–259.
[31] F.E. Harrell Jr., M.F.E. Harrell Jr., *Package 'hmisc'*. CRAN2018, 2019, pp. 235–236, 2019.
[32] M. Ayers, et al., IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade, J. Clin. Invest. 127 (8) (2017) 2930–2940.
[33] L. Danilova, et al., Programmed cell death ligand-1 (PD-L1) and CD8 expression profiling identify an immunologic subtype of pancreatic ductal adenocarcinomas with favorable survival, Cancer Immunol. Res. 7 (6) (2019) 886–895.
[34] M.S. Rooney, et al., Molecular and genetic properties of tumors associated with local immune cytolytic activity, Cell 160 (1–2) (2015) 48–61.
[35] W. Shen, et al., Sangerbox: A Comprehensive, Interaction-friendly Clinical Bioinformatics Analysis Platform, iMeta, 2022, 10.1002/imt2.36.
[36] B. Ricciuti, et al., Impact of DNA damage response and repair (DDR) gene mutations on efficacy of PD-(L)1 immune checkpoint inhibition in non-small cell lung cancer, Clin. Cancer Res. 26 (15) (2020) 4135–4142.
[37] P.A. Ott, et al., T-Cell-Inflamed gene-expression profile, programmed death ligand 1 expression, and tumor mutational burden predict efficacy in patients treated with pembrolizumab across 20 cancers: KEYNOTE-028, J. Clin. Oncol. 37 (4) (2019) 318–327.
[38] E.F. Rosloniec, K. Latham, Y.B. Guedez, Paradoxical roles of IFN-gamma in models of Th1-mediated autoimmunity, Arthritis Res. 4 (6) (2002) 333–336.
[39] E. Filaire, et al., Lung cancer: what are the links with oxidative stress, physical activity and nutrition, Lung Cancer 82 (3) (2013) 383–389.
[40] J. Zou, et al., APOBEC3B, a molecular driver of mutagenesis in human cancers, Cell Biosci. 7 (2017) 29.
[41] E. Tokunaga, et al., Expression of APOBEC3B mRNA in primary breast cancer of Japanese women, PLoS One 11 (12) (2016), e0168090.
[42] J. Zhang, et al., The roles of APOBEC3B in gastric cancer, Int. J. Clin. Exp. Pathol. 8 (5) (2015) 5089–5096.
[43] X. Luo, et al., Association of hepatitis B virus covalently closed circular DNA and human APOBEC3B in hepatitis B virus-related hepatocellular carcinoma, PLoS One 11 (6) (2016), e0157708.
[44] L. Xu, et al., High APOBEC3B expression is a predictor of recurrence in patients with low-risk clear cell renal cell carcinoma, Urol. Oncol. 33 (8) (2015) 340.e1–340.e8.
[45] H. Sasaki, et al., APOBEC3B gene overexpression in non-small-cell lung cancer, Biomed Rep 2 (3) (2014) 392–395.
[46] S. Yan, et al., Increased APOBEC3B predicts worse outcomes in lung cancer: a comprehensive retrospective study, J. Cancer 7 (6) (2016) 618–625.
[47] A.M. Sieuwerts, et al., Elevated APOBEC3B correlates with poor outcomes for estrogen-receptor-positive breast cancers, Horm Cancer 5 (6) (2014) 405–413.
[48] Y. Chen, et al., Chinese herbal prescription QYSL prevents progression of lung cancer by targeting tumor microenvironment, Oncologie 24 (2) (2022) 295–307.

[49] L. Pasqualucci, et al., AID is required for germinal center-derived lymphomagenesis, Nat. Genet. 40 (1) (2008) 108–112.
[50] G. Saraconi, et al., The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas, Genome Biol. 15 (7) (2014) 417.
[51] M. Asaoka, et al., APOBEC3-Mediated RNA editing in breast cancer is associated with heightened immune activity and improved survival, Int. J. Mol. Sci. 20 (22) (2019).
[52] G. Qi, H. Xiong, C. Zhou, APOBEC3 deletion polymorphism is associated with epithelial ovarian cancer risk among Chinese women, Tumour Biol 35 (6) (2014) 5723–5726.
[53] D. Marino, et al., APOBEC4 enhances the replication of HIV-1, PLoS One 11 (6) (2016), e0155422.
[54] M. Chen, R. Linstra, M. van Vugt, Genomic instability, inflammatory signaling and response to cancer immunotherapy, Biochim. Biophys. Acta Rev. Canc 1877 (1) (2022), 188661.
[55] E.R. Mardis, Neoantigens and genome instability: impact on immunogenomic phenotypes and immunotherapy response, Genome Med. 11 (1) (2019) 71.
[56] G. Yaghmour, et al., Role of genomic instability in immunotherapy with checkpoint inhibitors, Anticancer Res. 36 (8) (2016) 4033–4038.
[57] A.H. Pearlman, et al., Targeting public neoantigens for cancer immunotherapy, Nat Cancer 2 (5) (2021) 487–497.
[58] A.R. Aldous, J.Z. Dong, Personalized neoantigen vaccines: a new approach to cancer immunotherapy, Bioorg. Med. Chem. 26 (10) (2018) 2842–2849.
[59] Z. Chen, et al., EPHA5 mutations predict survival after immunotherapy in lung adenocarcinoma, Aging (Albany NY) 13 (1) (2020) 598–618.
[60] S. Wang, et al., The predictive power of tumor mutational burden in lung cancer immunotherapy response is influenced by patients' sex, Int. J. Cancer 145 (10) (2019) 2840–2849.
[61] P. Hu, et al., Elevated expression of LYPD3 is associated with lung adenocarcinoma carcinogenesis and poor prognosis, DNA Cell Biol. 39 (4) (2020) 522–532.
[62] M. Gruet, et al., β2-Adrenergic signalling promotes cell migration by upregulating expression of the metastasis-associated molecule LYPD3, Biology 9 (2) (2020).
[63] W. Zhou, et al., Knockdown of ANLN by lentivirus inhibits cell growth and migration in human breast cancer, Mol. Cell. Biochem. 398 (1–2) (2015) 11–19.
[64] H. Valque, et al., MUC5B leads to aggressive behavior of breast cancer MCF7 cells, PLoS One 7 (10) (2012), e46699.
[65] N.S. Pandi, et al., In silico analysis of expression pattern of a Wnt/β-catenin responsive gene ANLN in gastric cancer, Gene 545 (1) (2014) 23–29.
[66] M. Perrais, et al., Aberrant expression of human mucin gene MUC5B in gastric carcinoma and cancer cells. Identification and regulation of a distal promoter, J. Biol. Chem. 276 (18) (2001) 15386–15396.
[67] G. Wang, et al., Overexpression of Anillin (ANLN) is correlated with colorectal cancer progression and poor prognosis, Cancer Biomark 16 (3) (2016) 459–465.
[68] R. Nagashio, et al., Diagnostic and prognostic significances of MUC5B and TTF-1 expressions in resected non-small cell lung cancer, Sci. Rep. 5 (2015) 8649.
[69] M.G. Roy, et al., Muc5b is required for airway defence, Nature 505 (7483) (2014) 412–416.
[70] A. Vallejo, et al., An integrative approach unveils FOSL1 as an oncogene vulnerability in KRAS-driven lung and pancreatic cancer, Nat. Commun. 8 (2017), 14294.