Identification of immunotherapy biomarkers for improving the clinical outcome of homologous recombination deficiency patients with lung adenocarcinoma

Xiang Zhou[1,*], Rongjian Xu[2,*], Tong Lu[1,*], Ran Xu[1], Chenghao Wang[1], Bo Peng[1], Xiaoyan Chang[1], Zhiping Shen[1], Kaiyu Wang[1], Jiaxin Shi[1], Jiaying Zhao[1], Lin-You Zhang[1]

[1]Department of Thoracic Surgery, The Second Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang 150084, China
[2]Department of Thoracic Surgery, The Affiliated Hospital of Qingdao University, Qingdao, Shandong 266005, China
*Equal contribution

Correspondence to: Lin-You Zhang; email: lyzhang@hrbmu.edu.cn

ABSTRACT

Homologous recombination deficiency (HRD) is a common molecular signature of genomic instability and has been shown to be a biomarker for targeted therapies. However, there is a lack of studies on the role of HRD changes in lung adenocarcinoma (LUAD) transcriptomics. HRD scores were determined using single nucleotide polymorphism (SNP) array data from LUAD patients from The Cancer Genome Atlas (TCGA) database. Transcriptional data from patients with different scores were analyzed to identify biomarkers associated with HRD. Candidate biomarkers were validated using Gene Expression Omnibus (GEO)-sourced datasets and an immunotherapy cohort. According to the bulk transcriptome and clinical characteristics of 912 LUAD patients and Single-cell RNA-seq of 9 LUAD patients from TCGA and GEO databases, we observed increased MS4A6A expression in HRD tumors; high MS4A6A expression predicted improved survival outcomes. Furthermore, a comprehensive analysis of the tumor immune microenvironment (TIME) revealed a positive correlation between *MS4A6A* expression and neoantigen loading and immune cell infiltration. Additionally, the immunotherapy cohort confirmed the possibility of using *MS4A6A* as a biomarker. Collectively, we suggest that *MS4A6A* is associated with HRD and provide a new perspective toward identifying promising biomarkers for immunotherapy.

INTRODUCTION

In recent years, tumor immunotherapy has brought about revolutionary changes in cancer treatment. Lung cancer poses one of the greatest problems in antitumor treatment, and according to cancer statistics, lung cancer ranks first in terms of incidence and mortality [1]. The emergence of immunotherapy has significantly changed the landscape of lung cancer treatments. Although immunotherapy has made breakthroughs, the objective remission rate (ORR) in NSCLC, without varying treatment populations, is approximately 20% [2, 3]. The ORR remains below 50% even in populations with more than 50% programmed death-ligand 1 (PD-L1) expression [4, 5]. Thus, effective biomarkers are essential for selecting immunotherapy populations and improving the efficacy of immunotherapy.

Based on previous studies, immunotherapy-related markers can be broadly classified into the following four categories: 1) tumor cell-related biomarkers, such as PD-1, PD-L1 expression, tumor mutational burden (TMB), DNA damage response (DDR) pathway, and neoantigens; 2) tumor microenvironment (TME)-related

markers, such as tumor-infiltrating immune cells (CD4[+] and CD8[+] T cells); 3) liquid biopsy markers, such as peripheral blood cells and circulating tumor DNA; and 4) host-related biomarkers, such as intestinal symbionts and host germline genetic characteristics. Homologous recombination deficiency (HRD) usually refers to a state of DNA repair dysfunction at the cellular level, which can be caused by many factors, such as germline or somatic mutations in homologous recombination repair (HRR)-related genes and epigenetic inactivation. HRD can affect the DDR pathway by introducing insertions/deletions in nucleic acid sequences, copy number abnormalities, and chromosomal cross-linking, resulting in genomic and chromosomal instability [6]. HRD is present in various malignancies, particularly ovarian and breast cancers [7, 8]. The status and extent of HRD have emerged as novel biomarkers for the clinical application of PARP inhibitors in patients with advanced ovarian cancer [9–11]. In a previous study, Kadouri et al. observed that HRD is a risk factor affecting the prognosis of lung adenocarcinoma (LUAD) [12]. However, few studies have reported on HRD and LUAD, and the specific mechanism by which HRD affects the prognosis of LUAD remains unclear.

To explore the association between HRD-induced genomic instability and immunotherapy biomarkers in LUAD patients, we extracted transcriptomic data and mutation data from The Cancer Genome Atlas (TCGA) database. We calculated HRD scores for each LUAD patient. By analyzing the differences in transcriptome levels between the HRD and non-HRD groups, we observed elevated *MS4A6A* expression in HRD patients and that these patients had a better prognosis. Four datasets from the Gene Expression Omnibus (GEO) database were used to validate our results. Furthermore, we revealed that *MS4A6A* expression was positively correlated with multiple infiltrated immune cells in the TME, such as CD4+ T cells, CD8+ T cells, as well as immune checkpoints (ICPs), such as PD-1 and PD-L1. IMvigor210, an immunotherapy cohort, suggested that *MS4A6A* was a better predictor than PD-1, PD-L1, or CTLA-4. Therefore, the findings of this study provide possible directions for immunotherapy biomarkers and are valuable for understanding the relationship between genomic instability and TME in LUAD patients and improving clinical outcomes in patients undergoing immunotherapy (Figure 1).

MATERIALS AND METHODS

Data collection and pre-processing

The data used in the current study are accessible from publicly available databases. Transcriptomic data and

SNP, as well as the corresponding clinical characteristics and follow-up information ($n = 485$) of LUAD, were obtained from the TCGA database. To analyze the transcriptomic data, count values and transcripts per kilobase of exon model per million mapped read (TPM) values were extracted. The "Masked Somatic Mutation" was selected as the somatic mutation data. The maftools R package [13] was used to visualize the somatic mutation landscape and calculate the TMB for each LUAD patient. Partial genomic alteration scores (percentage of chromosomal copy number altered regions outside the measured regions) and MSI-Sensor scores were obtained from the cBioPortal database (http://www.cbioportal.org). Detailed information is provided in Table 1.

Datasets containing LUAD samples and clinical follow-up information were retrieved from the GEO database, and four datasets were included: GSE11969 [14], GSE30219 [15], GSE31210 [16], and GSE37745 [17]. The scRNA-seq data were obtained from GSE189357 [18] which includes 9 LUAD samples from nine resected samples of treatment-naïve patients. The array information of these datasets is shown in Supplementary Table 1. Additionally, the IMvigor210 [19] (Table 2) and GSE126044 [20] cohorts, which contained data on immunotherapy, were included in this study. Each dataset from the GEO database sources was normalized and annotated with an ID based on platform information.

Calculation of HRD and neoantigen scores

The HRD score was defined as the unweighted sum of the loss of heterogeneity (LOH) [21], telomere allelic imbalance (TAI) [22], and massive state transition (LOS) scores [23, 24]. The neoantigen load, that is, the number of peptides predicted to bind to major histocompatibility complex (MHC) proteins, was determined based on the HLA type derived from RNA sequencing data. The neoantigen load is expressed as single nucleotide variants (SNV) and insertion and deletion (indel) mutations. The values of HRD, neoantigen load, and mutation rate (number of single-nucleotide mutations) were compiled from a pan-cancer mapping study by Thorsson et al. [25]. The detailed information is provided in Supplementary Tables 2, 3.

Identification of independent prognostic genes associated with HRD score

We used the DESeq2 R package [26] for differential analysis of the HRD and non-HRD groups, filtering out low-expression genes and selecting |log2(fold change)| > 1.5, adj. $P < 0.05$ as the criteria. log2(fold change) > 1.5 were considered as highly expressed genes and < −1.5 for lowly expressed genes in HRD. Volcano plots

were used to visualize differentially expressed genes (DEGs).

Univariate Cox regression analysis of the differentially expressed genes was performed using the survival R package (https://CRAN.R-project.org/package=survival), and a *P*-value < 0.05 was selected as the cut-off value. The screened genes were subsequently analyzed by least absolute shrinkage and selection operator (LASSO) regression to determine the maximum prediction accuracy and the balance between minimizing explanatory accuracy [27]. Finally, independent prognostic factors were determined using a multivariate Cox regression analysis. A *P*-value < 0.05 was considered an independent prognostic factor, HR >1 as a risk factor, and HR <1 as a protective factor.

Selection of immunotherapy biomarkers by scRNA-seq analysis

An analysis of 10x scRNA-seq data was conducted by R packages, including "Seurat" [28] and "SingleR" [29] we utilized the "Seurat" R package to preprocess and analyze single-cell RNA sequencing (scRNA-seq) data. The scRNA-seq data were normalized using the "NormalizeData" function of the "Seurat" package, with the normalization method set as "LogNormalize". The resulting normalized data were then converted into a
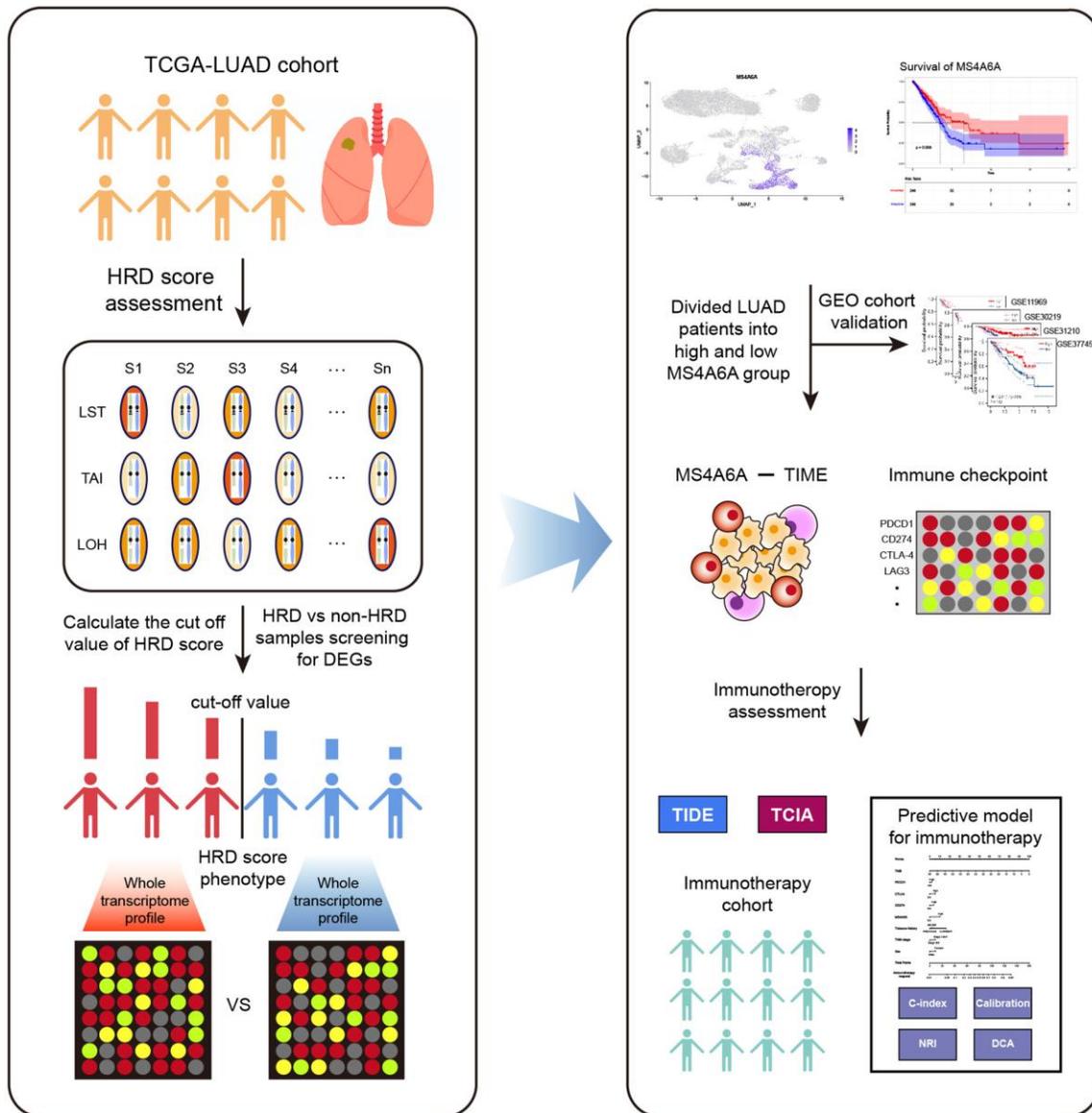

Figure 1. Workflow of this study. The analysis process of HRD-related RNA is shown on the left. HRD scores were obtained by calculating LOH, LST, and TAI for each sample of TCGA-LUAD. Patients were classified as HRD and non-HRD according to cut-off values. The screening and identification of markers are shown to the right. *MS4A6A* was identified as a potential immunotherapeutic marker by survival analysis, immune microenvironment, and immune checkpoint analysis.

Table 1. TCGA-LUAD clinical baseline information.

Characteristic	Non-HRD (*n* = 137)	HRD (*n* = 348)
Sex, *n* (%)		
Female	74 (15.4%)	183 (38.1%)
Male	60 (12.5%)	163 (34%)
Stage, *n* (%)		
Stage I	79 (16.5%)	184 (38.4%)
Stage II	27 (5.6%)	88 (18.4%)
Stage III	21 (4.4%)	55 (11.5%)
Stage IV	6 (1.3%)	19 (4%)
Age, median (IQR)	69 (61, 74)	65 (58, 72)
MSIsensor Score, median (IQR)	0 (0, 0.06)	0.02 (0, 0.13)
Mutation Count, median (IQR)	99.5 (40, 188.75)	254 (122, 451.5)
Fraction Genome Altered, median (IQR)	0.12 (0.04, 0.22)	0.31 (0.17, 0.45)
TMB (nonsynonymous), median (IQR)	3.3 (1.52, 7.13)	8.73 (4.08, 16.31)

Table 2. IMvigor210 clinical baseline information.

Characteristic	MS4A6A-low (*n* = 174)	MS4A6A-high (*n* = 174)
Bir Response, *n* (%)		
CR/PR	35 (11.7%)	33 (11.1%)
SD/PD	118 (39.6%)	112 (37.6%)
Sex, *n* (%)		
Female	39 (11.2%)	37 (10.6%)
Male	135 (38.8%)	137 (39.4%)
Immune phenotype, *n* (%)		
Desert	56 (19.7%)	20 (7%)
Excluded	67 (23.6%)	67 (23.6%)
Inflamed	22 (7.7%)	52 (18.3%)
TMB, median (IQR)	0.9 (0.53, 1.43)	0.94 (0.43, 1.87)

Seurat Object. The percentage of mitochondrial or ribosomal genes was calculated and low-quality cells were excluded to ensure the quality control (QC) [30]. We excluded samples with gene counts below 200 or above 3000, as well as those with a ribosomal RNA proportion exceeding 20%. Then, the top 3,000 genes with high variability were identified using the "FindVariableFeatures" function. To reduce the dimensionality of the scRNA-seq data, we performed principal component analysis (PCA) using the "RunPCA" function of the "Seurat" R package, based on the top 3,000 variable genes. Significant principal components (PCs) were identified through JackStraw analysis, and we selected the first 15 PCs for cell clustering analysis according to the proportion of variance explained. For cell clustering analysis, we utilized the "FindNeighbors" and "FindClusters" functions in the "Seurat" package. A k-nearest neighbor graph was constructed based on Euclidean distance in PCA using "FindNeighbors" to determine the closest neighbors of each cell. Cells were visualized using uniform manifold approximation and projection (UMAP) dimensionality reduction techniques for cell classification. To identify differentially expressed genes (DEGs) for each cluster, we used the "FindAllMarkers" function in the "Seurat" package, following Wilcoxon-Mann-Whitney tests. Marker genes for each cluster were identified using adjusted *p*-value < 0.01 and |log2 (fold change)| >1 as threshold values. We conducted a manual annotation, as described in the study by Maynard et al. [31], to identify and classify different cell types in our experimental samples. Finally, FeaturePlot and vlnPlot functions embedded in the "seurat" package were applied to visualize the cellular distribution of independent prognostic genes in the scRNA-seq dataset.

Gene set enrichment analysis

LUAD patients were divided into high and low *MS4A6A* groups according to the median *MS4A6A* expression value, and differential analysis was performed using the DESeq2 package. Gene Set Enrichment Analysis (GSEA) was performed using the clusterProfiler R package [32] to calculate the normalized enrichment score (NES) for each gene set and identify the signaling pathways enriched in the high- and low-*MS4A6A* expression groups. The selected gene set was selected as "c2.cp.v7.2. symbols" and false discovery rate (FDR) <0.25 was selected as a cut-off value.

Immune cell infiltration analysis

The microenvironmental characteristics of the tumors were assessed using the ESTIMATE R package [33]. ESTIMATE analysis quantifies immune activity (level of immune infiltration) in the tumor microenvironment based on its gene expression profile to obtain an immune score for each sample.

To investigate the abundance of immune cell infiltration in bulk tumor tissues, we used the Tumor Immune Estimation Resource (TIMER) database (https://cistrome.shinyapps.io/timer/) to predict the relative abundance of six types of infiltrating immune cells, including macrophages, dendritic cells, B cells, T cells, and neutrophils. Additionally, we extracted the expression of antigen-presentation-related genes, including those encoding MHC class I/II (I: *HLA-A*, *HLA-B*, and *HLA-C*; II: *HLA-DP*, *HLA-DM*, *HLA-DOA*, *HLA-DOB*, *HLA-DQ*, and *HLA-DR*) and antigen-binding molecules such as *B2M* and *TAP1/2*, and performed a correlation analysis between the expression of these molecules and *MS4A6A*.

Assessment of immunotherapy

Tumor immune dysfunction and exclusion (http://tide.dfci.harvard.edu/) can characterize T cell dysfunction by calculating tumor immune dysfunction and exclusion (TIDE) scores and evaluating the interaction of gene expression with the level of cytotoxic T lymphocytes (CTL) infiltration, which evaluates patient survival and response to immunotherapy [34]. Therefore, we assessed the clinical response to immunotherapy in patients with high and low *MS4A6A* expression by calculating the TIDE scores in LUAD patients. The Cancer Immunome Atlas (TCIA) was developed and maintained by the Institute of Bioinformatics [35]. This database allows querying the gene expression of specific immune-related genomes, cellular composition of immune infiltrates (characterized by genomic enrichment analysis and deconvolution), neoantigens, cancer-germline antigens,

and immunophenotype scores. Therefore, we assessed the potential immunotherapeutic effects of high and low *MS4A6A* expression levels by extracting immunophenotype scores.

In the investigation of the IMvigor210 and GSE126044 cohorts, we assessed the accuracy of *MS4A6A* against common immunotherapy-related markers, including TMB, *PD-1*, *PD-L1*, and *CTLA4*. We constructed a clinical prediction model that evaluated the effect of response to immunotherapy using logistic regression. Bootstrapping was used with 1000 iterations for re-sampling. A calibration curve was used to measure the consistency of the model. An integrated discrimination improvement (IDI) curve was used to assess its improvement, and decision curve analysis (DCA) was used to measure its clinical effect.

Statistics analysis

Data processing and analysis were performed using the R software (version 4.0.2). The statistical significance of normally distributed variables was estimated by independent Student's *t*-tests, whereas the differences between two groups of variables with non-normal distribution were assessed using the Mann–Whitney *U*-test (i.e., Wilcoxon rank-sum test). The chi-square and Fisher's exact tests were used to assess the statistical significance between the two groups of categorical variables. Kaplan–Meier (KM) survival curves were used to compare the survival rates of patients in the two groups. The log-rank test (log-rank test) was used to evaluate the significance of survival time differences between the two groups. LASSO analysis was performed using the glmnet R package. Nomograms and calibration curves were constructed using the rms package (https://CRAN.R-project.org/package=rms), and DCA was plotted using the rmda package. All statistical *P*-values were two-sided, and statistical significance was set at $P < 0.05$.

Data availability statement

The datasets and source codes used or analyzed during the current study are available from the corresponding author upon reasonable request.

RESULTS

HRD score is significantly associated with prognosis and mutational characteristics in LUAD patients

The HRD scores were calculated based on the LOU, LST, and TAI scores in the TCGA-LUAD dataset. The optimal cut-off value of the HRD score was determined by calculating the minimum *P*-value in the log-rank test. Patients with HRD scores >15 were considered to belong

to the HRD group, whereas those with HRD scores ≤15 were considered to belong to the non-HRD group. The KM curve showed that the overall survival (OS) of patients in the non-HRD group was much longer than that of patients in the HRD group (log-rank test, $P = 0.017$) (Figure 2A). The 1-, 3-, and 5-year ROC curves of survival were plotted. Their AUCs were estimated to be 0.734, 0.723, and 0.747, respectively (Figure 2B), indicating that the survival between HRD and non-HRD patients at 1, 3, and 5 years was significantly different and that the HRD score may serve as a potential prognostic biomarker. Subsequently, we investigated the relationship between HRD scores and other genomic instability features, such as MSI-sensor, genomic alteration fractions, and somatic mutation counts. The results revealed that the median somatic cumulative mutation count was significantly higher in the HRD group than in the non-HRD group (Wilcoxon signed-rank test, $P < 0.001$; Figure 2C) and was higher in the HRD group than in the non-HRD group (Wilcoxon signed-rank test, $P < 0.001$; Figure 2D). The fraction genome altered was also higher in the HRD group (Wilcoxon signed-rank test, $P < 0.0001$; Figure 2E). These findings indicate that patients with HRD have significantly higher genomic instability than those without HRD.
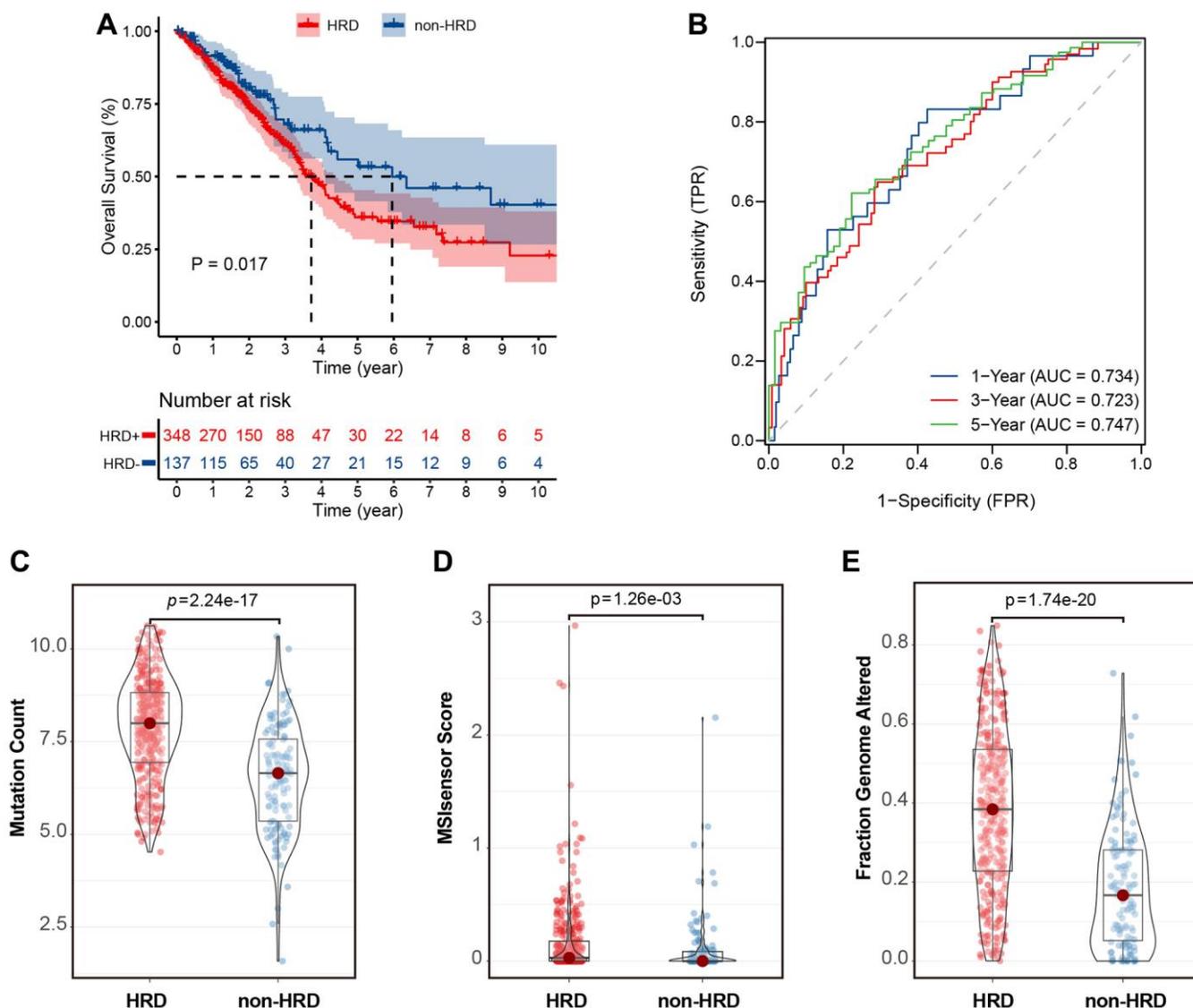

Figure 2. HRD scores were significantly associated with prognosis and mutation characteristics in the TCGA-LUAD cohort. (**A**) KM curve of overall survival of patients with HRD or non-HRD tumors in the TCGA-OSC cohort. (**B**) ROC curves of HRD scores in the TCGA-LUAD cohort. (**C**) Violin plots of somatic mutations in the HRD and non-HRD groups. Somatic mutation counts were significantly higher in the HRD group than in the non-HRD group. (**D**) Violin plots of MSI-Sensor in HRD and non-HRD groups. MSI-Sensor in the HRD group were significantly higher than those in the non-HRD group. (**E**) Violin plots of genomic alterations in the HRD group and non-HRD group. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

Differential mutation landscapes in HRD and non-HRD groups

Genomic features, such as oncogene activation (e.g., *ERBB2* amplification, *EGFR* tyrosine kinase mutations) and tumor suppressor gene inactivation (e.g., *MMR*, *BRCA1*/2), have been shown to strongly correlate with the clinical response to targeted therapies. Therefore, we compared the mutation landscape between the HRD and non-HRD groups. (Figure 3A, 3B) The results showed that the mutation landscape of the non-HRD group was significantly different from that of the HRD group. Only 11 of the top 20 genes with the highest mutation rates in the two groups overlapped (Figure 3C), and the mutation frequencies of the overlapping genes also differed significantly. For example, the *TP53* mutation frequency was 60% and 14% in the HRD and non-HRD groups, respectively. Furthermore, by screening actionable genes in the OncoKB database (https://www.oncokb.org/

actionableGenes), two of the 20 genes with the highest mutation frequency in the non-HRD group were identified as biomarkers for targeted drugs (*STK11* and *EGFR*). The mutation frequencies of these two genes in the non-HRD group were 14% and 12%, respectively. These results showed that HRD and non-HRD patients had different mutated genes. Non-HRD patients had actionable genes and drug targets suitable for targeted therapy. It further supported HRD as a potential biomarker for LUAD.

MS4A6A gene expression is positively correlated with HRD score and is an independent prognostic factor in LUAD patients

To identify mRNAs associated with HRD scores, we performed a differential analysis of RNA-seq data from HRD and non-HRD patients. A total of 326 DEGs were identified, of which 266 genes were highly expressed,
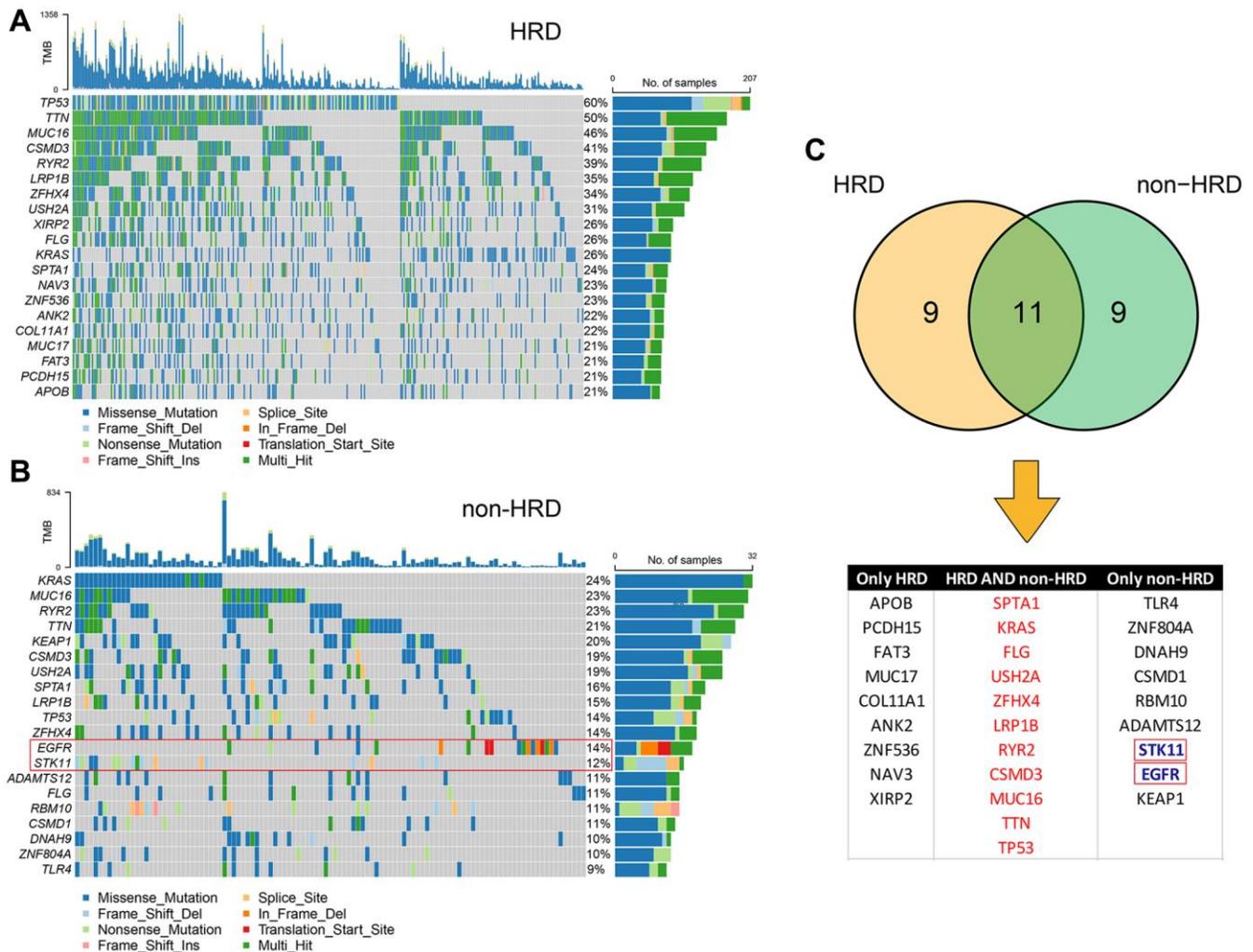

Figure 3. Mutational landscape of HRD and non-HRD patients. (**A**) Top 20 mutation landscape of HRD patients in the TCGA-LUAD cohort. (**B**) Top 20 mutation landscape of non-HRD patients in the TCGA-LUAD cohort. The genes in red boxes are actionable genes. (**C**) Overlapping information of HRD and non-HRD mutated genes; actionable genes are only in non-HRD.

and 60 genes were lowly expressed in the HRD group (Figure 4A). To identify DEGs associated with patient HRD scores and prognosis, we then conducted a univariate Cox regression analysis of the TCGA-LUAD cohort, including 326 DEGs. Univariate analysis and log-rank tests were used to identify 89 genes with prognostic potential ($P < 0.05$). LASSO-Cox proportional risk regression and 10-fold cross-validation were performed on these prognosis-related genes to screen for independent prognostic factors. LASSO coefficient profiles were generated for log lambda and the optimal λ value corresponding to the eight variables (Figure 4B, 4C). After performing stepwise multivariate Cox regression analysis, *SALL1*, *TCN1*, *RHCG*, *ANLN*, *MS4A6A*, and CIDEC were identified as independent prognostic factors. In contrast, only *MS4A6A* was a protective factor, and the rest were risk factors (Figure 4D). Then we performed scRNA-seq
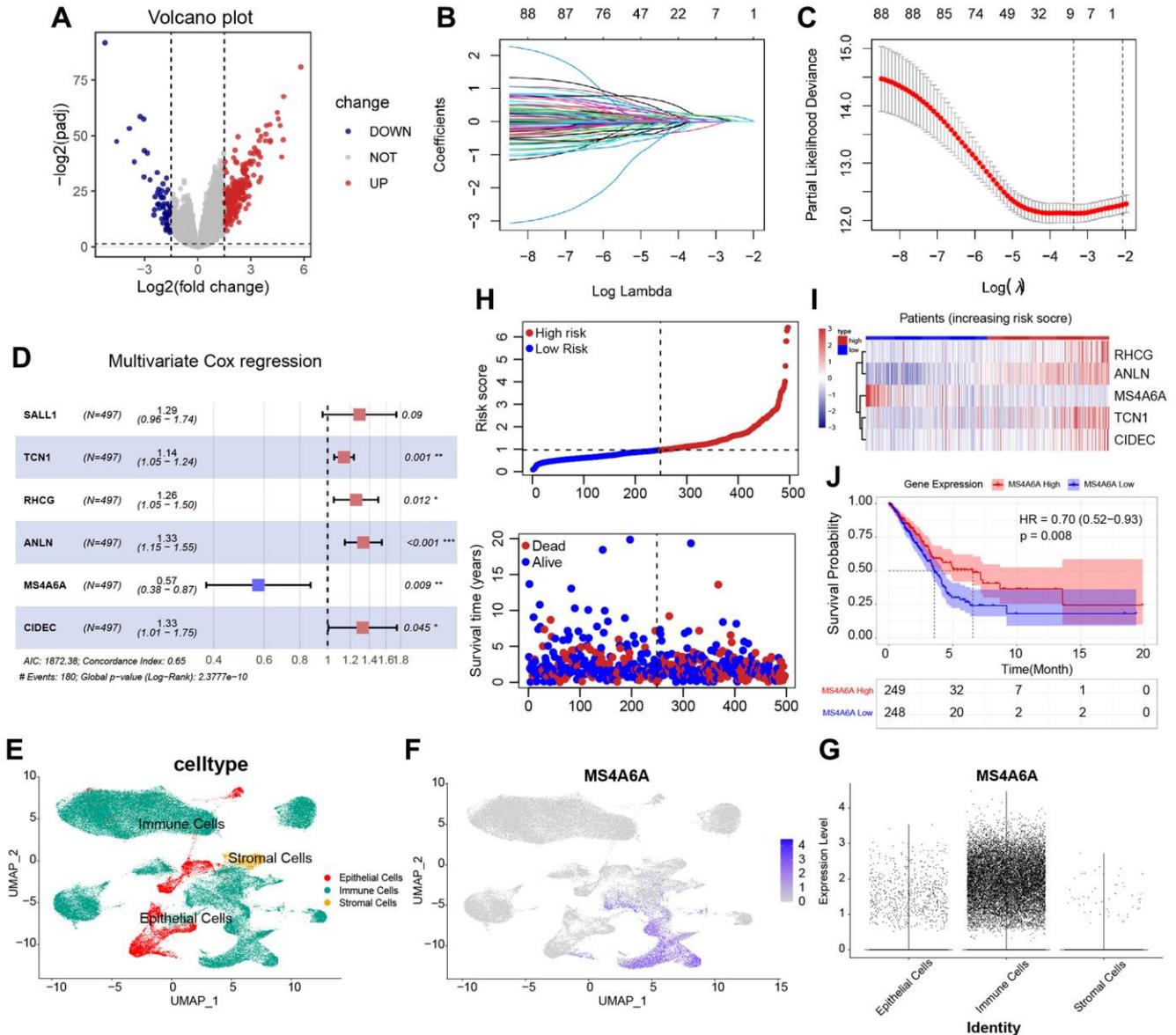

Figure 4. Identification of HRD-related prognostic RNA. (**A**) Differential analysis based on HRD vs. non-HRD patients, 326 DEGs were obtained, of which 266 genes were expressed up in HRD patients, and 60 genes were expressed down in HRD patients. (**B**, **C**) After univariate Cox regression screening, 89 prognostic genes were obtained and analyzed by LASSO regularized dimensionality reduction and eliminated the covariance between variables, after which nine genes were obtained. (**D**) Multivariate Cox regression screening of independent prognostic genes, containing five genes, of which all were risk factors except for *MS4A6A*. (**E**) Cellular distribution of 109649 cells clustered into 3 unique subsets among all merged lung adenocarcinoma tissue samples. (**F**) FeaturePlot depicting the distribution of MS4A6A. (**G**) vlnPlot showing the expression levels of MS4A6A in different cell subsets. (**H**) Risk plots of multifactorial Cox regression with the distribution of risk score in the upper layer, the distribution range of survival information in the lower layer. (**I**) The expression characteristics of five genes in high and low-risk group. (**J**) KM curves of *MS4A6A* (log-rank test).

analysis with GEO dataset: GSE189357 (Supplementary Figure 1A–1D). A total of 109,649 cells were analyzed and distinguished into epithelial cells, immune cells, and stromal cells after QC. (Figure 4E). FeaturePlot (Figure 4F) and vlnPlot (Figure 4G) visualization indicated MS4A6A has higher expression than other five independent prognostic signatures (Supplementary Figure 1E) in all kinds of cells, especially in immune cells. The assessment of immune cell signatures within the microenvironment of tumors provides crucial insights into the nature and magnitude of immune response in individual cancer patients, as well as their likelihood of responsiveness to immunotherapy [36]. Thus, focusing on immune cell signatures is more suitable for assessing the potential of immunotherapy in cancer and predicting which patients are likely to benefit from this treatment. As shown in risk plots of multi-factorial Cox regression with the distribution of risk score, we divided the patients into high risk subgroup and low risk subgroup and survival status indicated that high risk score subgroup owns worse survival outcome (Figure 4H). The expression characteristics of five genes in high and low-risk groups corresponds to the step-wise multivariate Cox regression analysis (Figure 4I). The KM curve of MS4A6A demonstrated that patients with high MS4A6A expression had a better prognosis ($P = 0.008$) (Figure 4J). To investigate whether the *MS4A6A* expression profile has a similar prognostic value in different datasets, we independently confirmed our findings in four datasets from the GEO database: GSE11969, GSE30219, GSE31210, and GSE37745. KM analysis also showed that patients with high *MS4A6A* expression had a better prognosis than those with low *MS4A6A* expression (Supplementary Figure 2A).

MS4A6A expression is positively correlated with TIME

High *MS4A6A* expression is associated with a better prognosis than low *MS4A6A* expression. Therefore, we examined the relationship between high *MS4A6A* expression and the tumor microenvironment. To investigate the relationship between MS4A6A expression and immune cell infiltration in the tumor microenvironment, we first calculated the immune score of the TIME in patients with LUAD using the ESTIMATE algorithm. We analyzed the correlation between the immune score and *MS4A6A*. As shown in Figure 5A, *MS4A6A* showed a significant positive correlation with immune score (Spearman's rank correlation coefficient, Rho = 0.83, $P < 0.001$). Accordingly, we calculated the immune score for each tumor patient in the four GEO datasets and categorized the patients according to the median value of immune infiltration. We found that *MS4A6A* expression was significantly higher in the high immune infiltration group (Supplementary Figure 2B). Furthermore, the

predicted neoantigen load was positively correlated with *MS4A6A* expression (Spearman's rank correlation coefficient, Rho = 0.266, $P < 0.05$) (Figure 5B). We used the TIMER algorithm to estimate the correlation between *MS4A6A* expression and the five types of immune cells to better understand how *MS4A6A* is related to immune cell infiltration. As shown in the scatter plot, *MS4A6A* was significantly positively correlated with macrophages (Rho = 0.722, $P < 0.001$) and dendritic cells (Rho = 0.756, $P < 0.001$), suggesting that *MS4A6A* plays a vital role in antigen presentation and processing. In addition, we found that MS4A6A expression also correlated significantly with CD8+ and CD4+ T cells, suggesting the relevance of *MS4A6A* in tumor killing (Figure 5C). To further investigate the relationship between *MS4A6A* and antigen-presentation-related genes, we analyzed the association between *MS4A6A* expression and MHC class I/II (I: *HLA-A*, *HLA-B*, *HLA-C*; II: *HLA-DP*, *HLA-DM*, *HLA-DOA*, *HLA-DOB*, *HLA-DQ*, *HLA-DR*) and key antigen-binding molecules (such as *B2M*, *TAP1/2*), and observed a significant positive correlation between them (Figure 5D). Notably, we observed that the *MS4A6A* gene was positively correlated with these immune-related genes in LUAD as well as in the other 32 cancers (Figure 5E). *MS4A6A* might be expressed more frequently on the surface of antigen-presenting cells in the tumor microenvironment.

In addition, GSEA analysis of gene expression profiles of the *MS4A6A*-positive and -negative groups revealed that the *MS4A6A* positive group was significantly enriched in DNA repair, DNA mismatch repair, and immune system-related pathways (Supplementary Figure 3).

MS4A6A expression is positively correlated with ICPs

Tumor cells activate immune checks so that antigens cannot be presented to T cells, thereby blocking the process of presenting antigens in the tumor immune link and suppressing the immune function of T cells. The use of anti-PD-1/PD-L1 in tumor treatment plays a significant role in immunotherapy [37]. We collected and analyzed 46 common genes associated with immune checkpoints to determine the relationship between *MS4A6A* and these genes [38]. Correlation analysis revealed that *MS4A6A* positively correlated with many immune checkpoint-associated genes (Figure 6A). We screened the most common immune checkpoint genes currently available, including PD-1 (*PDCD1*), PD-L1 (*CD274*), *CD48*, *CD86*, *CTLA4*, *ICOS*, *LAG3*, *PDCD1LG2*, and *TIGIT*, and compared the expression differences between the high and low *MS4A6A* expression groups. We found that all of them were highly expressed in the *MS4A6A* high expression group (Figure 6B). In addition, the four

GEO datasets showed that immune checkpoint genes are generally elevated in the high-expression MS4A6A groups (Supplementary Figure 4). It is generally accepted that key regulators of immunity function in various tissues. Therefore, we investigated *MS4A6A* expression characteristics and immune checkpoint-associated gene expression in various cancer types. Notably, we observed that MS4A6A is positively correlated with genes associated with immune checkpoints in LUAD and 32 other cancer types (Figure 6C). These results
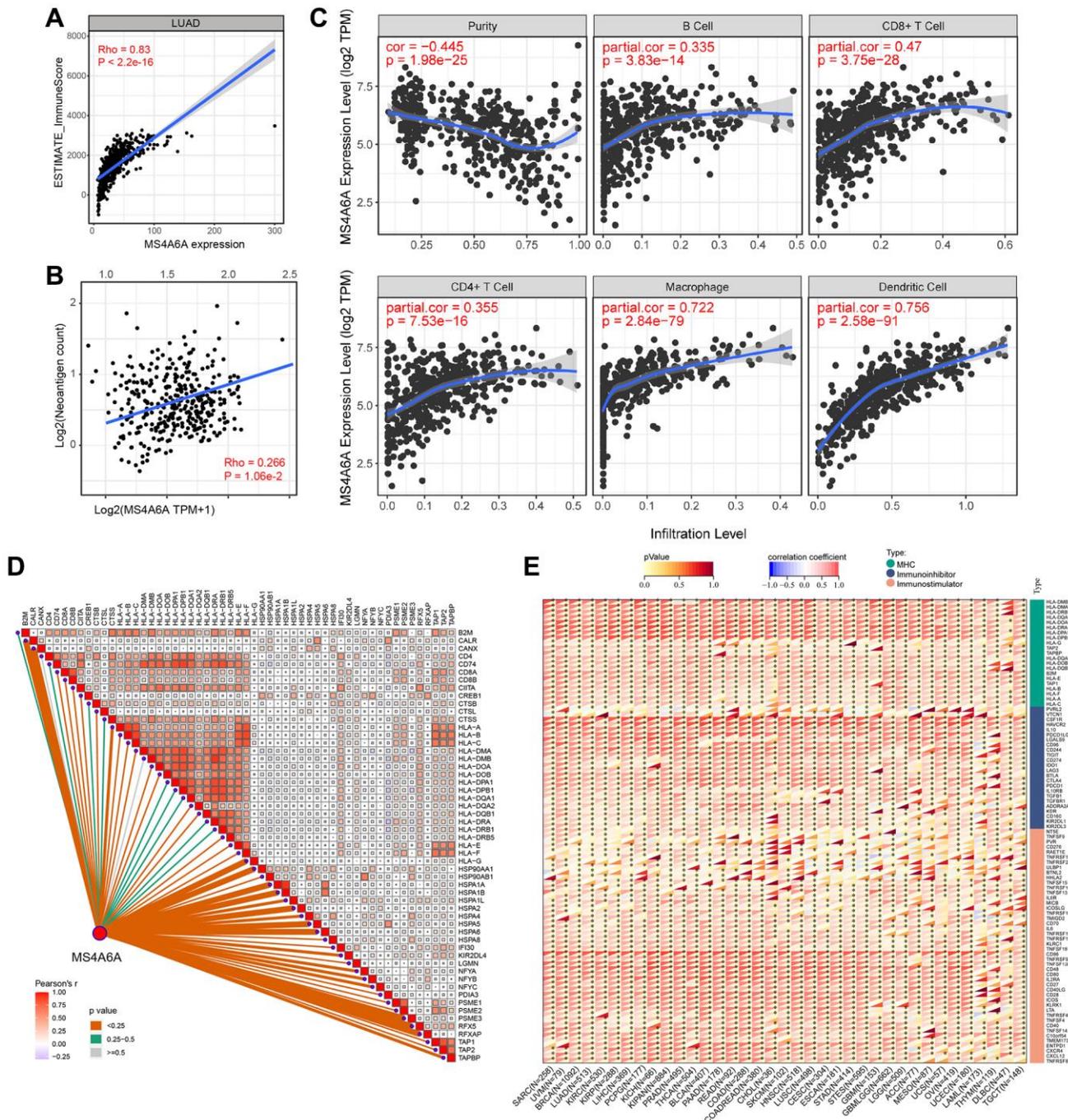

Figure 5. Relationship between *MS4A6A* expression pattern and TIME. (**A**) *MS4A6A* expression is positively correlated with ImmuneScore (Spearman rank correlation coefficient, R = 0.83, *P* ≈ 0). (**B**) MS4A6A expression is positively correlated with neoantigen load (Spearman rank correlation coefficient; *P* < 0.05). (**C**) *MS4A6A* expression was positively correlated with immune cell subpopulation (Spearman rank correlation coefficient; *P* < 0.0001). (**D**) *MS4A6A* expression is positively correlated with antigen-related genes. (**E**) Correlation of *MS4A6A* expression signature with antigen-related genes in the TCGA pan-cancer cohort.

suggest that *MS4A6A* expression may be associated with tumor immunotherapy response.

MS4A6A can be used as a potential biomarker for immunotherapy

We then focused on immunotherapy-related datasets. To further confirm the predictive value of MS4A6A for immune checkpoint blockade (ICB) treatment, we collected transcriptomic profiles and clinical information from the uroepithelial carcinoma (UC) immunotherapy cohort (IMvigor210) [39]. Patients with high *MS4A6A*

expression showed a more pronounced clinical benefit and significantly longer survival. (Figure 7A). In addition, patients with high *MS4A6A* expression had a significant treatment effect and immune response to PD-L1 blockade compared with patients with low *MS4A6A* gene expression (Figure 7B). When comparing neoantigen loads, no significant differences were found between the two groups with high or low *MS4A6A* expression (Figure 7C). However, patients with high *MS4A6A* expression had significantly higher tumor immunophenotypes and a better response to immunotherapy compared to those with lower expression (Figure 7D, 7E).
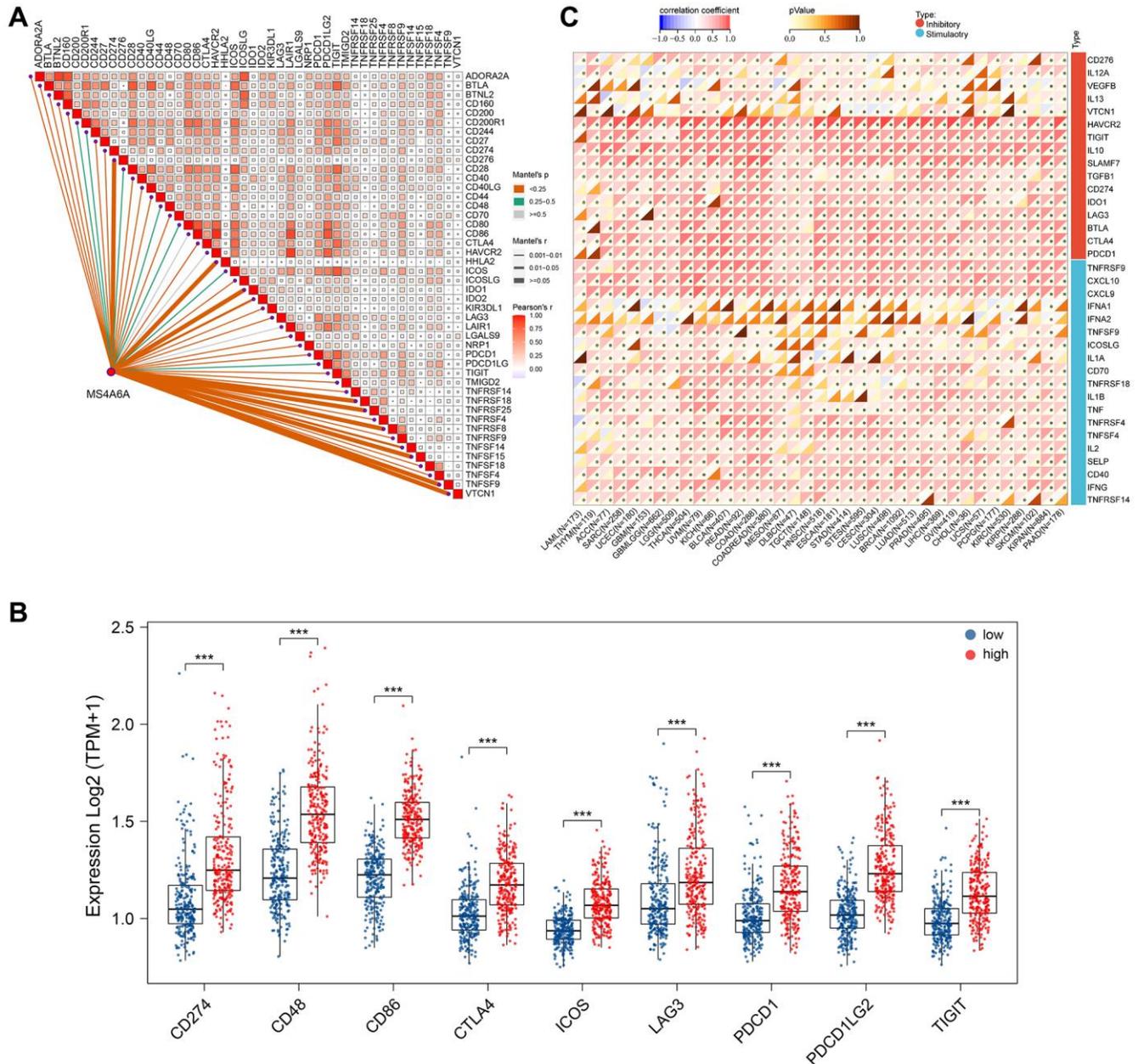
Figure 6. *MS4A6A* expression was positively correlated with ICP-related genes. (**A**) Correlation of *MS4A6A* expression with ICP-related genes. (**B**) Comparison of *MS4A6A* expression with common immune checkpoint expression in TCGA-LUAD cohort. (**C**) Correlation of *MS4A6A* expression profile with ICP-related genes in the pan-cancer cohort. $^{**}P < 0.01$, $^{***}P < 0.001$.

The TIDE and IPS scores have been widely used to predict the effects of immunotherapy. We compared the TIDE scores of patients with high and low *MS4A6A* expression in the TCGA-LUAD dataset. We observed that the TIDE scores of patients with high expression were significantly lower than those in the low expression group, suggesting that the effect of receiving ICB treatment might be better in patients in the high expression group (Figure 7F). TCIA results showed that the IPS was significantly higher in the low *MS4A6A* group ($P < 0.001$) (Figure 7G), and patients in the high *MS4A6A* expression group had a relatively higher
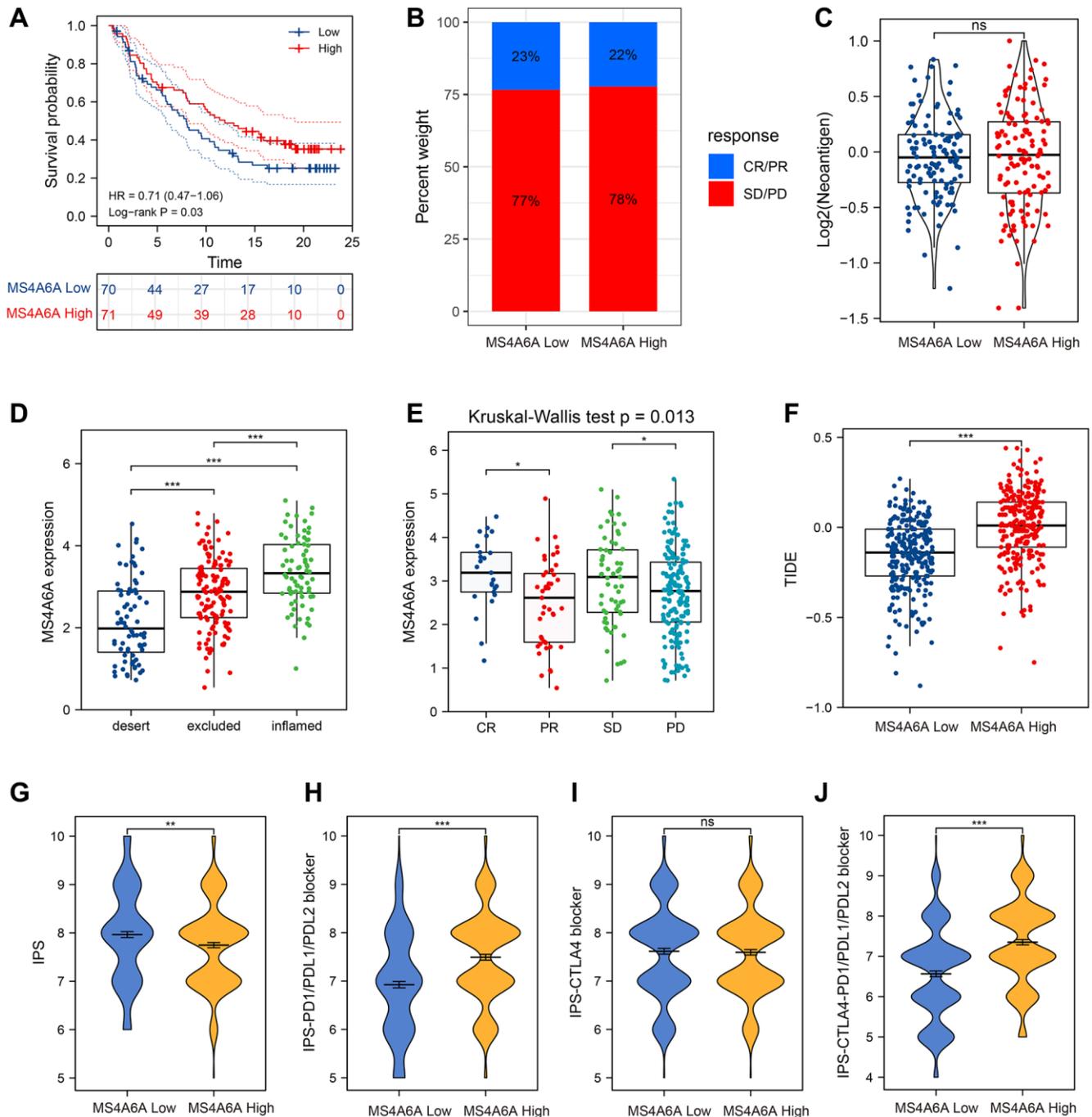
Figure 7. *MS4A6A* can be used as a potential biomarker for ICB treatment. (**A**) OS curves for high and low *MS4A6A* expression in the IMvigor210 cohort. (**B**) The proportion of immune response in *MS4A6A* high and low expression groups against ICB treatment. Abbreviations: CR: complete remission; PR: local response; SD: stable disease; PD: progressive disease. (**C**) Comparison of neoantigen load between different *MS4A6A* expression subgroups. (**D**, **E**) Comparison of *MS4A6A* expression among different immune responsive cohorts. (**F**) TIDE differences in *MS4A6A* high- and low-expression cohorts. (**G–J**) TCIA analysis of differences in IPS scores in *MS4A6A* high and low expression cohorts. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

response to anti-PD1/PDL1 treatment while there was no significant difference with anti-CTLA4 treatment ($P < 0.001$; Figure 7H–7J). These results suggest that patients with high MS4A6A expression may benefit from ICBs.

Construction of a predictive model of immunotherapy based on *MS4A6A* and clinical features

These results suggest that *MS4A6A* may be used as an immunotherapeutic marker; however, it is still unclear which markers are more effective than the expression of existing markers, such as *PD-1*, *PD-L1*, and *CTLA4*. Therefore, we plotted ROC curves based on the IMvigor210 cohort and found that the AUC values of *MS4A6A* (AUC = 0.663) were significantly higher than those of *PD-1*

(AUC = 0.553), *CTLA4* (AUC = 0.526), and *PD-L1* (AUC = 0.566), but lower than that of TMB (AUC = 0.726) (Figure 8A); the accuracy of *MS4A6A* (AUC = 0.982) was better than that of *PD-1* (AUC = 0.745), *PD-L1* (AUC = 0.782), and *CTLA4* (AUC = 0.818) in the GSE126044 cohort (Figure 8B). Therefore, *MS4A6A* may be a better biomarker for immunotherapy than *PD-1*, *PD-L1*, or *CTLA4*. Subsequently, we investigated the MS4A6A protein expression in LUAD using the Human Protein Atlas (HPA) database. We observed that it was predominantly expressed in the nucleus and was significantly elevated in LUAD tissues compared to lung tissues. (Supplementary Figure 5).

Clinical prediction models are currently essential tools for aiding clinical decision-making. Here, a model that
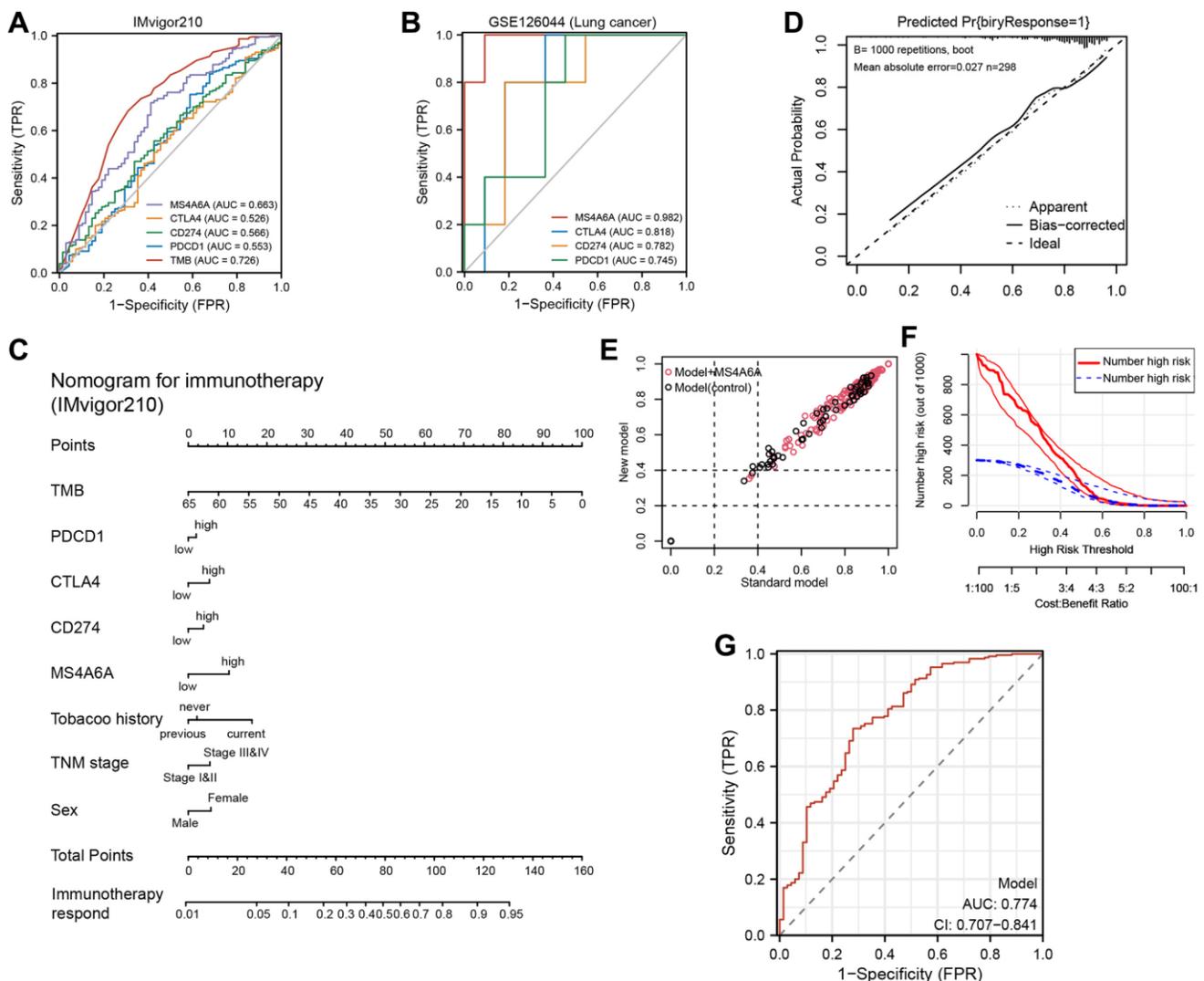
Figure 8. Construction and validation of *MS4A6A*-based immunotherapy predictive model. (A) ROC curves for *MS4A6A*, *PD-1*, *PD-L1*, and *CTLA4* based on the GSE126044 cohort. **(B)** ROC curves for TMB, *MS4A6A*, *PD-1*, *PD-L1*, and *CTLA4* based on the IMvigor210 cohort. **(C)** Immunotherapy prediction model nomogram. **(D)** Calibration curves of the immunotherapy prediction model. **(E)** Comparison of prediction accuracy of immunotherapy model with and without *MS4A6A*. **(F)** Clinical decision curves of the immunotherapy prediction model. **(G)** The ROC curve of clinical model (AUC = 0.774).

predicts immune therapy responses in patients using *MS4A6A* gene expression, as well as *PD-1*, *PD-L1*, and TMB, was developed to help clinicians make immunotherapy decisions. Using multivariate logistic regression, we constructed a prediction model for immunotherapy responders and non-responders (Figure 8C). Bootstrapping was used to assess the models, re-sampling 1000 times, and calibration curves were plotted (Figure 8D). The IDI scatter plot results showed that the accuracy of the prediction model, including *MS4A6A,* increased by 8% compared to that of the model without *MS4A6A* (Figure 8E). A DCA plot was used to reflect the net benefit of the patients from the models (Figure 8F). Then, the ROC curve of this model was plotted and the AUC value reached 0.774 (Figure 8G). Based on these results, *MS4A6A* combined with established immunotherapy markers can better predict immunotherapy outcomes.

DISCUSSION

In recent years, immunotherapy for cancer has emerged as a remarkable advance in anti-cancer research, which has revolutionized cancer treatment and changed treatment strategies. In traditional cancer treatment procedures such as chemotherapy, cancer cells and stem cells are destroyed, causing irreversible damage and even death. For patients with advanced tumors, the advent of targeted therapies has brought new hope for longer survival. For instance, EGFR-tyrosine kinase inhibitors have significantly extended the survival time of many patients with advanced NSCLC [40]. Immunotherapy differs from conventional chemotherapy and targeted therapies in one essential way by targeting immune cells rather than cancer cells. Currently, for patients with advanced melanoma with cancer metastasis and for whom all treatment options have failed, Opdivo and Keytruda could reduce or even eliminate tumors in more than 60% of patients for more than two years [41–43]. The use of immunotherapy in treating lung cancer has shown remarkable results, greatly extending the survival time of patients with advanced lung cancer [2, 44–46]. Despite the correlation between immunotherapy responsiveness and certain immunomarkers, such as PD-1, PD-L1, and TMB, not all patients with high PD-1 or TMB expression benefit from immunotherapy. Single biomarkers have limitations that affect the accuracy of screening of populations that benefit from immunotherapy.

In the present study, we analyzed the characteristics of HRD and non-HRD populations of LUAD patients. Based on their differences in transcriptome levels, we screened a set of genes strongly associated with HRD. We screened independent prognostic factors by univariate Cox regression, LASSO, and multivariate Cox regression.

Single-cell analysis was also carried out to explore the details of the independent prognostic factors at cell level. Notably, we found that *MS4A6A* gene expression was elevated in the HRD group and that patients with high *MS4A6A* expression had a better prognosis. *MS4A6A* has a higher expression in immune cells compared with other two cell types. This may suggest its association with tumor immune activity. MS4A family members play critical roles in various pathological conditions, including cancer, infectious diseases, and neurodegeneration. Also, they play a vital role in regulating immune signaling [47]. *MS4A6A*, also known as *CDA01*, *MS4A6*, *4SPAN3*, or *CD20L3*, encodes a member of the transmembrane 4A gene family. *MS4A6A* appears to be strongly associated with Alzheimer's disease [48–50]; however, the *MS4A6A* gene has not been investigated. In the TCGA-LUAD dataset, we observed that *MS4A6A* expression was positively correlated with immune cell infiltration in the tumor microenvironment, especially macrophages, dendritic cells, and multiple immune check-points, suggesting that *MS4A6A* could be a potential biomarker for ICB therapy. By analyzing immuno-therapy-related cohorts, we revealed that *MS4A6A* has a higher accuracy as a biomarker than molecules such as PD-1 and CTLA4. We constructed a predictive immunotherapy model based on the IMvigor210 dataset. The C-index and calibration curve results indicated that the model had good accuracy and consistency.

However, the present study has some limitations. Although there are some LUAD immunotherapy datasets in public databases, many are panel data instead of complete transcriptional data; thus, the expression data |of *MS4A6A* were unavailable, and only GSE126044 was eligible for inclusion in this study. In addition, IMvigor210 cohorts were obtained from patients with bladder epithelial carcinoma, containing transcriptome data and comprehensive clinical information; therefore, it has been used in several immunotherapy-related and LUAD-related studies [51]. However, its application in LUAD studies remains controversial, considering the heterogeneity of tumors. Finally, the lack of wet-lab experiments using cell line models or human lung adenocarcinoma tumor tissues limits the confidence and applicability of our results, as it does not provide experimental evidence supporting the biological relevance and functional implications of the findings. In our future work, we will consider conducting wet-lab experiments to address the limitations of this study and complement our results.

CONCLUSION

We identified *MS4A6A*, whose expression level was closely correlated with the level of HRD in LUAD and was highly accurate as an immunotherapeutic biomarker.

Furthermore, detecting MS4A6A expression in tissues or blood is more straightforward than calculating the HRD scores. Nonetheless, its practicality must be confirmed in a larger cohort and prospective studies.

AUTHOR CONTRIBUTIONS

(I) Lin-you Zhang: Conception and study design; (II) Rongjian Xu: Funding support; (III) Tong Lu, Zhiping Shen, Jiaxin Shi: Provision of study materials; (IV) Kaiyu Wang, Ran Xu, Bo Peng: Collection and assembly of data; (V) Xiang Zhou, Chenghao Wang, Xiaoyan Chang: Data analysis and interpretation; (VI) Jiaying Zhao, Rongjian Xu: Administrative support; (VII) Xiang Zhou: Manuscript drafting; (VIII) All authors: Final approval of manuscript.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

ETHICAL STATEMENT

This study utilized publicly available data from databases, and no ethical approval was required as it did not involve human subjects or animals directly.

FUNDING

REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin. 2022; 72:7–33. https://doi.org/10.3322/caac.21708 PMID:35020204

2. Garon EB, Rizvi NA, Hui R, Leighl N, Balmanoukian AS, Eder JP, Patnaik A, Aggarwal C, Gubens M, Horn L, Carcereny E, Ahn MJ, Felip E, et al, and KEYNOTE-001 Investigators. Pembrolizumab for the treatment of non-small-cell lung cancer. N Engl J Med. 2015; 372:2018–28. https://doi.org/10.1056/NEJMoa1501824 PMID:25891174

3. Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WE, Poddubskaya E, Antonia S, Pluzanski A, Vokes EE, Holgado E, Waterhouse D, Ready N, Gainor J, et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. N Engl J Med. 2015; 373:123–35. https://doi.org/10.1056/NEJMoa1504627 PMID:26028407

4. Mok TSK, Wu YL, Kudaba I, Kowalski DM, Cho BC, Turna HZ, Castro G Jr, Srimuninnimit V, Laktionov KK, Bondarenko I, Kubota K, Lubiniecki GM, Zhang J, et al, and KEYNOTE-042 Investigators. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. Lancet. 2019; 393:1819–30. https://doi.org/10.1016/S0140-6736(18)32409-7 PMID:30955977

5. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, Gottfried M, Peled N, Tafreshi A, Cuffe S, O'Brien M, Rao S, Hotta K, et al, and KEYNOTE-024 Investigators. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. N Engl J Med. 2016; 375:1823–33. https://doi.org/10.1056/NEJMoa1606774 PMID:27718847

6. Knijnenburg TA, Wang L, Zimmermann MT, Chambwe N, Gao GF, Cherniack AD, Fan H, Shen H, Way GP, Greene CS, Liu Y, Akbani R, Feng B, et al, and Cancer Genome Atlas Research Network. Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. Cell Rep. 2018; 23:239–54.e6. https://doi.org/10.1016/j.celrep.2018.03.076 PMID:29617664

7. Ledermann JA, Drew Y, Kristeleit RS. Homologous recombination deficiency and ovarian cancer. Eur J Cancer. 2016; 60:49–58. https://doi.org/10.1016/j.ejca.2016.03.005 PMID:27065456

8. Llop-Guevara A, Loibl S, Villacampa G, Vladimirova V, Schneeweiss A, Karn T, Zahm DM, Herencia-Ropero A, Jank P, van Mackelenbergh M, Fasching PA, Marmé F, Stickeler E, et al. Association of RAD51 with homologous recombination deficiency (HRD) and clinical outcomes in untreated triple-negative breast cancer (TNBC): analysis of the GeparSixto randomized clinical trial. Ann Oncol. 2021; 32:1590–6. https://doi.org/10.1016/j.annonc.2021.09.003 PMID:34520831

9. Miller RE, Leary A, Scott CL, Serra V, Lord CJ, Bowtell D, Chang DK, Garsed DW, Jonkers J, Ledermann JA, Nik-Zainal S, Ray-Coquard I, Shah SP, et al. ESMO recommendations on predictive biomarker testing for homologous recombination deficiency and PARP inhibitor benefit in ovarian cancer. Ann Oncol. 2020; 31:1606–22. https://doi.org/10.1016/j.annonc.2020.08.2102 PMID:33004253

10. Mirza MR, Monk BJ, Herrstedt J, Oza AM, Mahner S,

Redondo A, Fabbro M, Ledermann JA, Lorusso D, Vergote I, Ben-Baruch NE, Marth C, Mądry R, et al, and ENGOT-OV16/NOVA Investigators. Niraparib Maintenance Therapy in Platinum-Sensitive, Recurrent Ovarian Cancer. N Engl J Med. 2016; 375:2154–64.
https://doi.org/10.1056/NEJMoa1611310
PMID:27717299

11. Vanacker H, Harter P, Labidi-Galy SI, Banerjee S, Oaknin A, Lorusso D, Ray-Coquard I. PARP-inhibitors in epithelial ovarian cancer: Actual positioning and future expectations. Cancer Treat Rev. 2021; 99:102255.
https://doi.org/10.1016/j.ctrv.2021.102255
PMID:34332292

12. Kadouri L, Rottenberg Y, Zick A, Hamburger T, Lipson D, Peretz T, Nechushtan H. Homologous recombination in lung cancer, germline and somatic mutations, clinical and phenotype characterization. Lung Cancer. 2019; 137:48–51.
https://doi.org/10.1016/j.lungcan.2019.09.008
PMID:31542568

13. Mayakonda A, Koeffler HP. Maftools: Efficient analysis, visualization and summarization of MAF files from large-scale cohort based cancer studies. bioRxiv. 2016.
https://doi.org/10.1101/052662

14. Takeuchi T, Tomida S, Yatabe Y, Kosaka T, Osada H, Yanagisawa K, Mitsudomi T, Takahashi T. Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. J Clin Oncol. 2006; 24:1679–88.
https://doi.org/10.1200/JCO.2005.03.8224
PMID:16549822

15. Rousseaux S, Debernardi A, Jacquiau B, Vitte AL, Vesin A, Nagy-Mignotte H, Moro-Sibilot D, Brichon PY, Lantuejoul S, Hainaut P, Laffaire J, de Reyniès A, Beer DG, et al. Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. Sci Transl Med. 2013; 5:186ra66.
https://doi.org/10.1126/scitranslmed.3005723
PMID:23698379

16. Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R, Furuta K, Tsuta K, Shibata T, Yamamoto S, Watanabe S, Sakamoto H, Kumamoto K, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. Cancer Res. 2012; 72:100–11.
https://doi.org/10.1158/0008-5472.CAN-11-1403
PMID:22080568

17. Botling J, Edlund K, Lohr M, Hellwig B, Holmberg L, Lambe M, Berglund A, Ekman S, Bergqvist M, Pontén F, König A, Fernandes O, Karlsson M, et al. Biomarker discovery in non-small cell lung cancer: integrating gene expression profiling, meta-analysis, and tissue microarray validation. Clin Cancer Res. 2013; 19:194–204.
https://doi.org/10.1158/1078-0432.CCR-12-1139
PMID:23032747

18. Zhu J, Fan Y, Xiong Y, Wang W, Chen J, Xia Y, Lei J, Gong L, Sun S, Jiang T. Delineating the dynamic evolution from preneoplasia to invasive lung adenocarcinoma by integrating single-cell RNA sequencing and spatial transcriptomics. Exp Mol Med. 2022; 54:2060–76.
https://doi.org/10.1038/s12276-022-00896-9
PMID:36434043

19. Balar AV, Galsky MD, Rosenberg JE, Powles T, Petrylak DP, Bellmunt J, Loriot Y, Necchi A, Hoffman-Censits J, Perez-Gracia JL, Dawson NA, van der Heijden MS, Dreicer R, et al, and IMvigor210 Study Group. Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: a single-arm, multicentre, phase 2 trial. Lancet. 2017; 389:67–76.
https://doi.org/10.1016/S0140-6736(16)32455-2
PMID:27939400

20. Cho JW, Hong MH, Ha SJ, Kim YJ, Cho BC, Lee I, Kim HR. Genome-wide identification of differentially methylated promoters and enhancers associated with response to anti-PD-1 therapy in non-small cell lung cancer. Exp Mol Med. 2020; 52:1550–63.
https://doi.org/10.1038/s12276-020-00493-8
PMID:32879421

21. Abkevich V, Timms KM, Hennessy BT, Potter J, Carey MS, Meyer LA, Smith-McCune K, Broaddus R, Lu KH, Chen J, Tran TV, Williams D, Iliev D, et al. Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. Br J Cancer. 2012; 107:1776–82.
https://doi.org/10.1038/bjc.2012.451
PMID:23047548

22. Birkbak NJ, Wang ZC, Kim JY, Eklund AC, Li Q, Tian R, Bowman-Colin C, Li Y, Greene-Colozzi A, Iglehart JD, Tung N, Ryan PD, Garber JE, et al. Telomeric allelic imbalance indicates defective DNA repair and sensitivity to DNA-damaging agents. Cancer Discov. 2012; 2:366–75.
https://doi.org/10.1158/2159-8290.CD-11-0206
PMID:22576213

23. Popova T, Manié E, Rieunier G, Caux-Moncoutier V, Tirapo C, Dubois T, Delattre O, Sigal-Zafrani B, Bollet M, Longy M, Houdayer C, Sastre-Garau X, Vincent-Salomon A, et al. Ploidy and large-scale genomic instability consistently identify basal-like breast

carcinomas with BRCA1/2 inactivation. Cancer Res. 2012; 72:5454–62.
https://doi.org/10.1158/0008-5472.CAN-12-1470
PMID:22933060

24. Telli ML, Timms KM, Reid J, Hennessy B, Mills GB, Jensen KC, Szallasi Z, Barry WT, Winer EP, Tung NM, Isakoff SJ, Ryan PD, Greene-Colozzi A, et al. Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. Clin Cancer Res. 2016; 22:3764–73.
https://doi.org/10.1158/1078-0432.CCR-15-2477
PMID:26957554

25. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA, Ziv E, Culhane AC, Paull EO, et al, and Cancer Genome Atlas Research Network. The Immune Landscape of Cancer. Immunity. 2018; 48:812–30.e14.
https://doi.org/10.1016/j.immuni.2018.03.023
PMID:29628290

26. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15:550.
https://doi.org/10.1186/s13059-014-0550-8
PMID:25516281

27. Tibshirani R. Regression Shrinkage and Selection via the Lasso. J R Stat Soc Series B Stat Methodol. 1996; 58:267–88.

28. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015; 161:1202–14.
https://doi.org/10.1016/j.cell.2015.05.002
PMID:26000488

29. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR, Butte AJ, Bhattacharya M. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol. 2019; 20:163–72.
https://doi.org/10.1038/s41590-018-0276-y
PMID:30643263

30. Jiang A, Wang J, Liu N, Zheng X, Li Y, Ma Y, Zheng H, Chen X, Fan C, Zhang R, Fu X, Yao Y. Integration of Single-Cell RNA Sequencing and Bulk RNA Sequencing Data to Establish and Validate a Prognostic Model for Patients With Lung Adenocarcinoma. Front Genet. 2022; 13:833797.
https://doi.org/10.3389/fgene.2022.833797
PMID:35154287

31. Maynard A, McCoach CE, Rotow JK, Harris L, Haderk F, Kerr DL, Yu EA, Schenk EL, Tan W, Zee A, Tan M, Gui P, Lea T, et al. Therapy-Induced Evolution of Human Lung Cancer Revealed by Single-Cell RNA Sequencing. Cell. 2020; 182:1232–51.e22.
https://doi.org/10.1016/j.cell.2020.07.017
PMID:32822576

32. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012; 16:284–7.
https://doi.org/10.1089/omi.2011.0118
PMID:22455463

33. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013; 4:2612.
https://doi.org/10.1038/ncomms3612
PMID:24113773

34. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, Li Z, Traugh N, Bu X, Li B, Liu J, Freeman GJ, Brown MA, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med. 2018; 24:1550–8.
https://doi.org/10.1038/s41591-018-0136-1
PMID:30127393

35. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep. 2017; 18:248–62.
https://doi.org/10.1016/j.celrep.2016.12.019
PMID:28052254

36. Cristescu R, Mogg R, Ayers M, Albright A, Murphy E, Yearley J, Sher X, Liu XQ, Lu H, Nebozhyn M, Zhang C, Lunceford JK, Joe A, et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. Science. 2018; 362:eaar3593.
https://doi.org/10.1126/science.aar3593
PMID:30309915

37. Mamdani H, Matosevic S, Khalid AB, Durm G, Jalal SI. Immunotherapy in Lung Cancer: Current Landscape and Future Directions. Front Immunol. 2022; 13:823618.
https://doi.org/10.3389/fimmu.2022.823618
PMID:35222404

38. Petitprez F, Meylan M, de Reyniès A, Sautès-Fridman C, Fridman WH. The Tumor Microenvironment in the Response to Immune Checkpoint Blockade Therapies. Front Immunol. 2020; 11:784.
https://doi.org/10.3389/fimmu.2020.00784
PMID:32457745

39. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, Kadel EE III, Koeppen H, Astarita JL, Cubas R, Jhunjhunwala S, Banchereau R, Yang Y, et al. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature. 2018; 554:544–8.
https://doi.org/10.1038/nature25501
PMID:29443960

40. Wu SG, Shih JY. Management of acquired resistance to EGFR TKI-targeted therapy in advanced non-small cell lung cancer. Mol Cancer. 2018; 17:38.
https://doi.org/10.1186/s12943-018-0777-1
PMID:29455650

41. Tawbi HA, Schadendorf D, Lipson EJ, Ascierto PA, Matamala L, Castillo Gutiérrez E, Rutkowski P, Gogas HJ, Lao CD, De Menezes JJ, Dalle S, Arance A, Grob JJ, et al, and RELATIVITY-047 Investigators. Relatlimab and Nivolumab versus Nivolumab in Untreated Advanced Melanoma. N Engl J Med. 2022; 386:24–34.
https://doi.org/10.1056/NEJMoa2109970
PMID:34986285

42. Eggermont AMM, Blank CU, Mandala M, Long GV, Atkinson V, Dalle S, Haydon A, Lichinitser M, Khattak A, Carlino MS, Sandhu S, Larkin J, Puig S, et al. Adjuvant Pembrolizumab versus Placebo in Resected Stage III Melanoma. N Engl J Med. 2018; 378:1789–801.
https://doi.org/10.1056/NEJMoa1802357
PMID:29658430

43. Hamid O, Robert C, Daud A, Hodi FS, Hwu WJ, Kefford R, Wolchok JD, Hersey P, Joseph R, Weber JS, Dronca R, Mitchell TC, Patnaik A, et al. Five-year survival outcomes for patients with advanced melanoma treated with pembrolizumab in KEYNOTE-001. Ann Oncol. 2019; 30:582–8.
https://doi.org/10.1093/annonc/mdz011
PMID:30715153

44. Hellmann MD, Paz-Ares L, Bernabe Caro R, Zurawski B, Kim SW, Carcereny Costa E, Park K, Alexandru A, Lupinacci L, de la Mora Jimenez E, Sakai H, Albert I, Vergnenegre A, et al. Nivolumab plus Ipilimumab in Advanced Non-Small-Cell Lung Cancer. N Engl J Med. 2019; 381:2020–31.
https://doi.org/10.1056/NEJMoa1910231
PMID:31562796

45. Forde PM, Spicer J, Lu S, Provencio M, Mitsudomi T, Awad MM, Felip E, Broderick SR, Brahmer JR, Swanson SJ, Kerr K, Wang C, Ciuleanu TE, et al, and CheckMate 816 Investigators. Neoadjuvant Nivolumab plus Chemotherapy in Resectable Lung Cancer. N Engl J Med. 2022; 386:1973–85.
https://doi.org/10.1056/NEJMoa2202170
PMID:35403841

46. Cascone T, William WN Jr, Weissferdt A, Leung CH, Lin HY, Pataer A, Godoy MCB, Carter BW, Federico L, Reuben A, Khan MAW, Dejima H, Francisco-Cruz A, et al. Neoadjuvant nivolumab or nivolumab plus ipilimumab in operable non-small cell lung cancer: the phase 2 randomized NEOSTAR trial. Nat Med. 2021; 27:504–14.
https://doi.org/10.1038/s41591-020-01224-2
PMID:33603241

47. Mattiola I, Mantovani A, Locati M. The tetraspan MS4A family in homeostasis, immunity, and disease. Trends Immunol. 2021; 42:764–81.
https://doi.org/10.1016/j.it.2021.07.002
PMID:34384709

48. Smith AM, Davey K, Tsartsalis S, Khozoie C, Fancy N, Tang SS, Liaptsi E, Weinert M, McGarry A, Muirhead RCJ, Gentleman S, Owen DR, Matthews PM. Diverse human astrocyte and microglial transcriptional responses to Alzheimer's pathology. Acta Neuropathol. 2022; 143:75–91.
https://doi.org/10.1007/s00401-021-02372-6
PMID:34767070

49. Novikova G, Kapoor M, Tcw J, Abud EM, Efthymiou AG, Chen SX, Cheng H, Fullard JF, Bendl J, Liu Y, Roussos P, Björkegren JL, Liu Y, et al. Integration of Alzheimer's disease genetics and myeloid genomics identifies disease risk regulatory elements and genes. Nat Commun. 2021; 12:1610.
https://doi.org/10.1038/s41467-021-21823-y
PMID:33712570

50. Lacher SE, Alazizi A, Wang X, Bell DA, Pique-Regi R, Luca F, Slattery M. A hypermorphic antioxidant response element is associated with increased MS4A6A expression and Alzheimer's disease. Redox Biol. 2018; 14:686–93.
https://doi.org/10.1016/j.redox.2017.10.018
PMID:29179108

51. Lin W, Wang X, Wang Z, Shao F, Yang Y, Cao Z, Feng X, Gao Y, He J. Comprehensive Analysis Uncovers Prognostic and Immunogenic Characteristics of Cellular Senescence for Lung Adenocarcinoma. Front Cell Dev Biol. 2021; 9:780461.
https://doi.org/10.3389/fcell.2021.780461
PMID:34869385
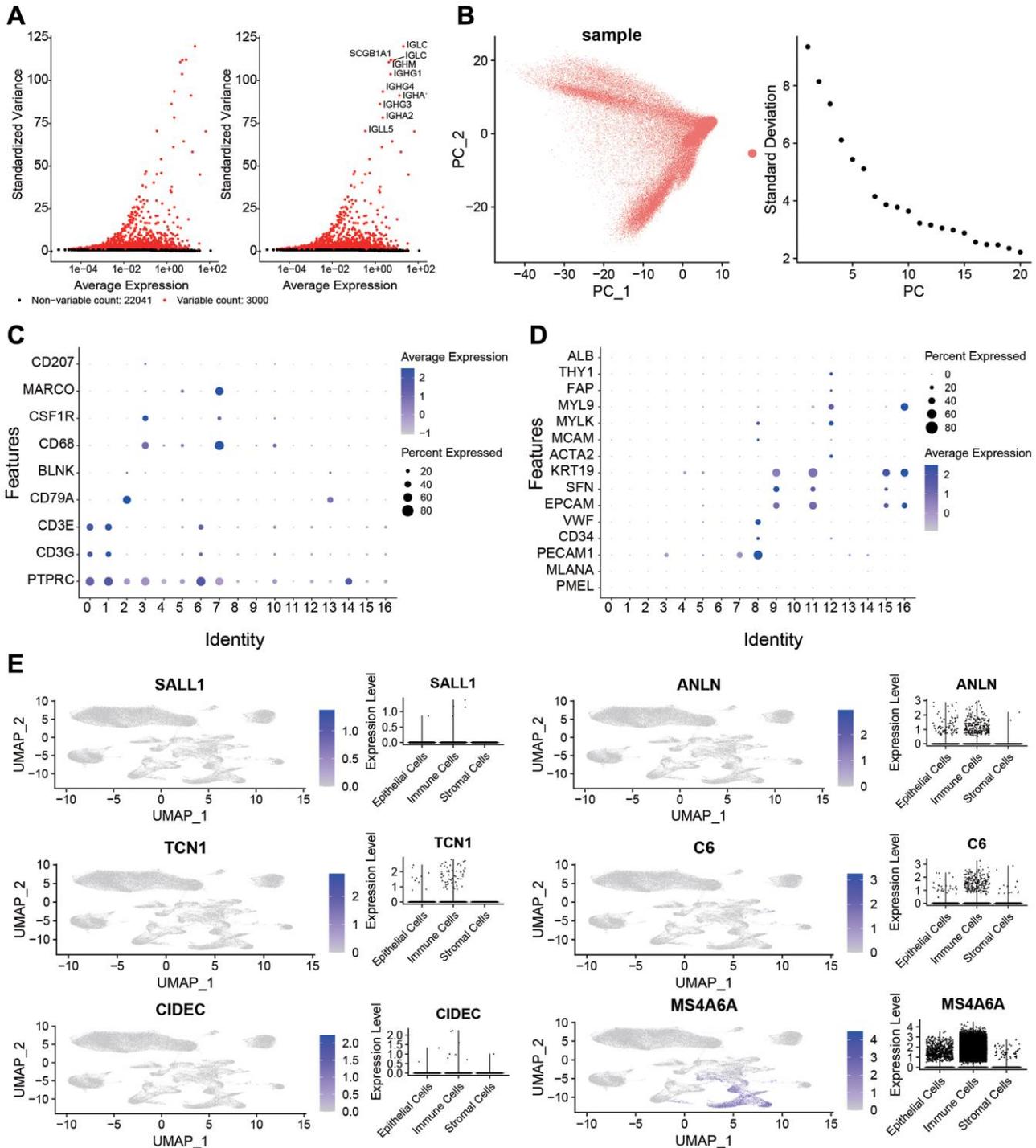
**Supplementary Figure 1. scRNA-seq analysis.** (**A**) The Variable Features plot identified the top 3000 variable genes, and the top 10 were highlighted. (**B**) PCA dimension reduction, 15 PCs were applied for cell clustering. (**C**) Bubble plot depicting the biomarkers of immune cells across 0–16 subsets. (**D**) Bubble plot depicting the biomarkers of epithelial and stromal cells across 0–16 subsets. (**E**) Distribution of 6 genes in different types of cells.
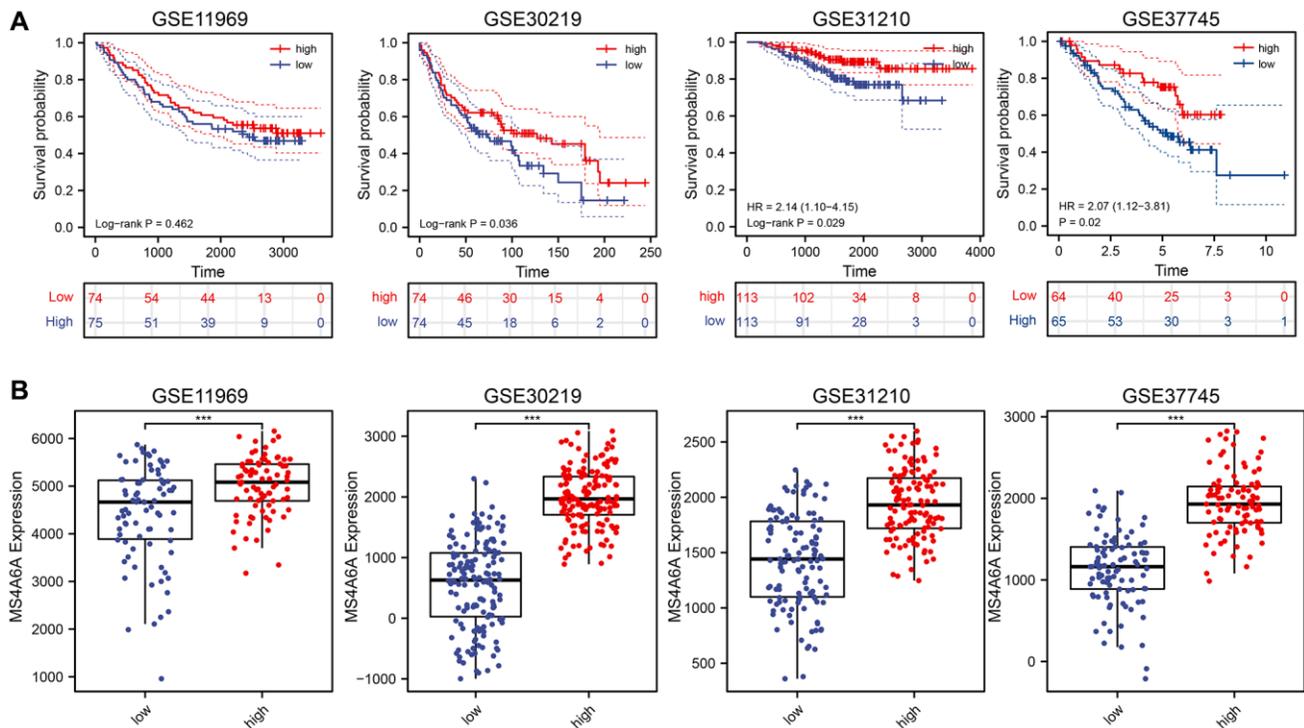
Supplementary Figure 2. Comparison of *MS4A6A* survival analysis and immune scores in GEO dataset. (**A**) KM survival curves of *MS4A6A* in the GSE11969, GSE30219, GSE31210, and GSE37745 datasets. (**B**) Differences in *MS4A6A* expression levels were compared in GSE11969, GSE30219, GSE31210, and GSE37745 in the high-low immune infiltration groups. *MS4A6A* expression was elevated in the high immune infiltration group.
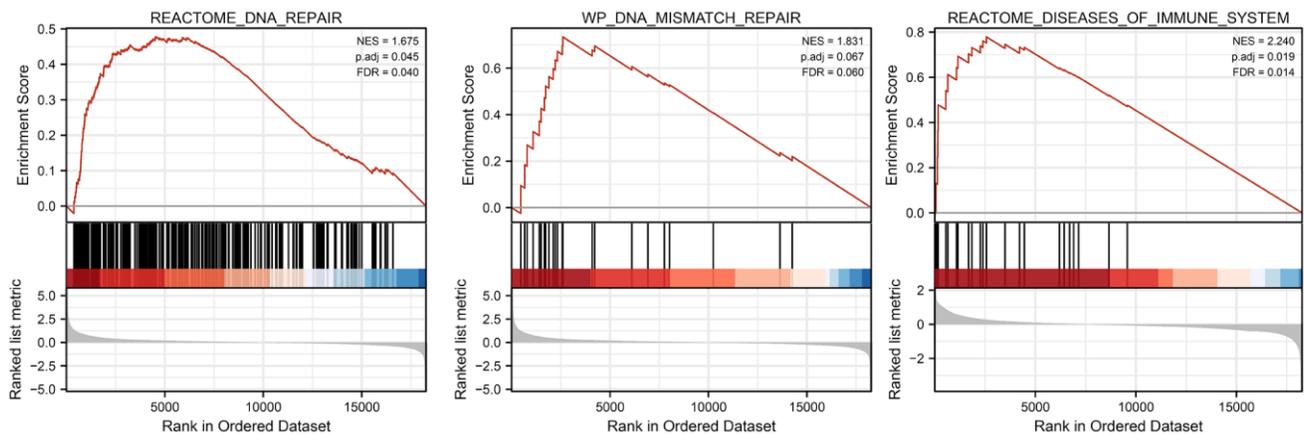

Supplementary Figure 3. GSEA of high and low expression patterns of *MS4A6A*. GSEA results revealed that the DNA repair, DNA mismatch repair, and immune system-related pathways were significantly enriched in the *MS4A6A*-positive group.
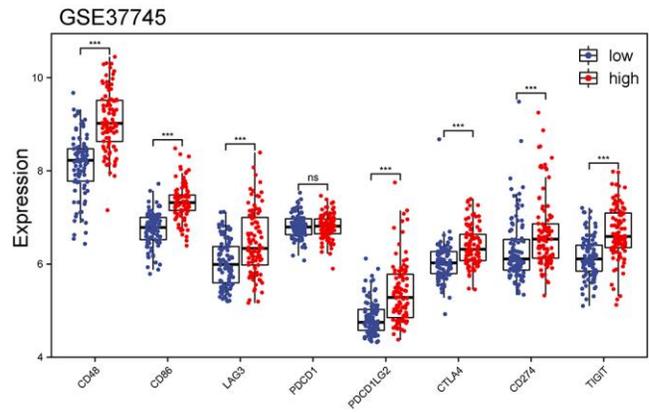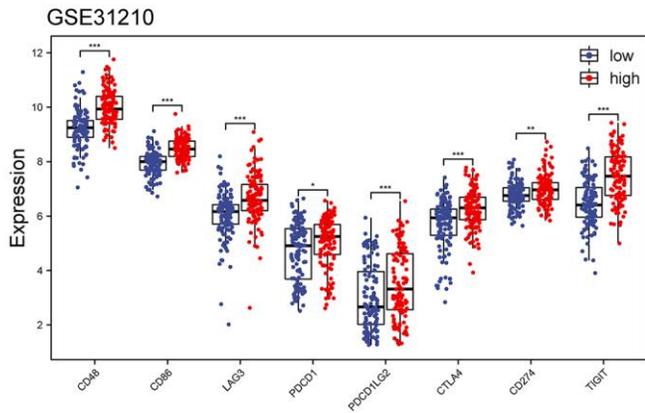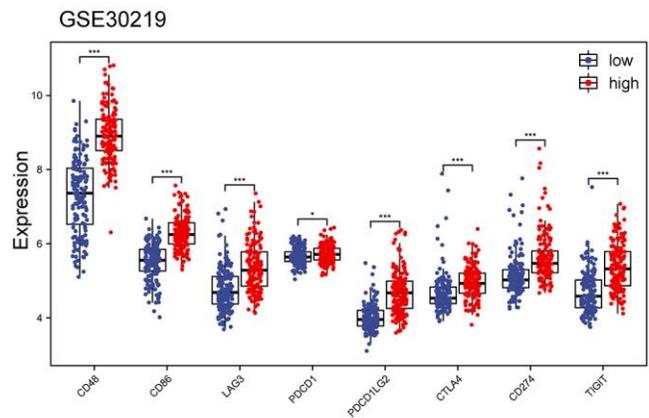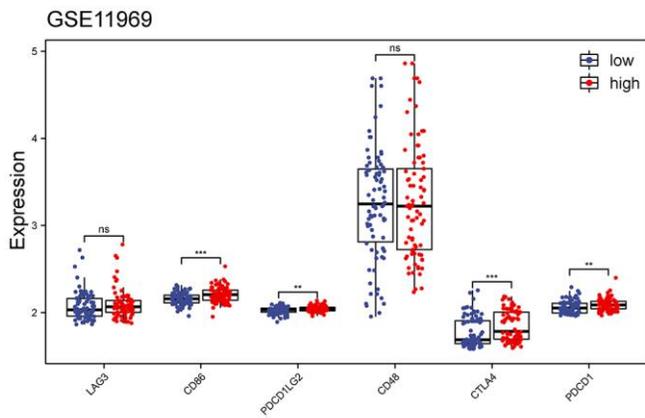
Supplementary Figure 4. Validation of relationship between *MS4A6A* gene and immune checkpoint expression in GEO dataset. Box plots demonstrate that *MS4A6A* high expression group in GSE11969, GSE30219, GSE31210, and GSE37745 datasets, as well as immune checkpoints, exhibited consistent elevated expression results with the TCGA-LUAD cohort.

**A** MS4A6A
lung adenocarcinoma id:1847
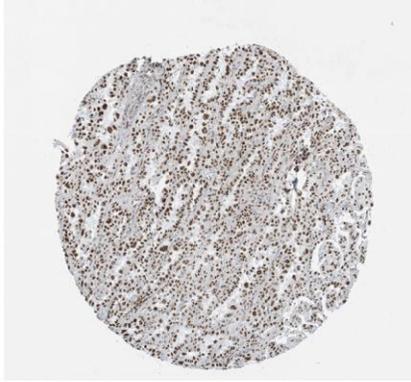
MS4A6A
lung adenocarcinoma id:3048
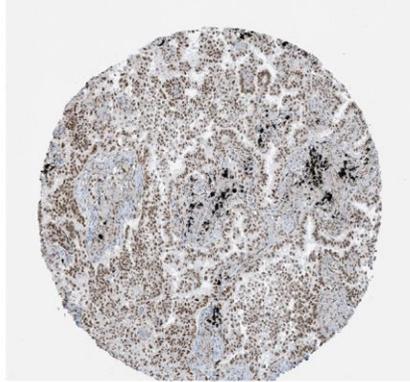
MS4A6A
lung adenocarcinoma id:3052

B MS4A6A
Lung id:2268
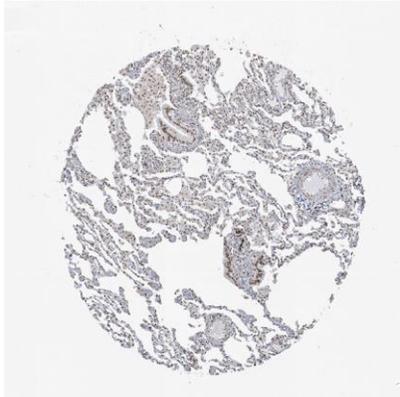
MS4A6A
Lung id:2208
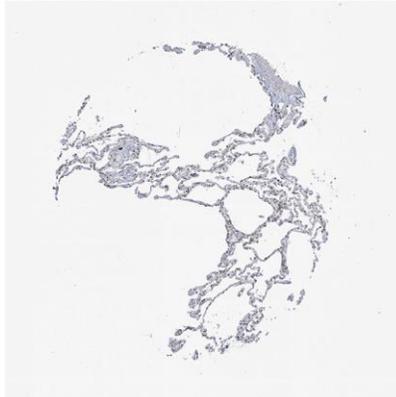
MS4A6A
Lung id:3076

Supplementary Figure 5. Immunohistochemical analysis of MS4A6A in normal lung vs. LUAD tissues. (**A**) Immunohistochemistry of MS4A6A in LUAD tissue. (**B**) Immunohistochemistry of MS4A6A in normal lung tissue.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 2 and 3.

Supplementary Table 1. Information about the GEO datasets employed.

	GSE11969	GSE30219	GSE31210	GSE37745	GSE189357
Samples	163	307	246	196	9
Normal	5	14	20	0	0
Tumor	158	293	226	196	9
Platform	GPL7015	GPL570	GPL570	GPL570	GPL24676
Data type	Expression profiling by array	Expression profiling by array	Expression profiling by array		10X genomics

Supplementary Table 2. Homologous recombination deficiency score in TCGA-LUAD patients.

Supplementary Table 3. Neoantigens score of TCGA-LUAD patients.