Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Exploration of Tumor Mutation Burden Combined with Immune Infiltrates in the Prognosis of Lung Adenocarcinoma

Zhenyu Zhao
Central South University https://orcid.org/0000-0002-8697-1419
Boxue He
Central South University
Qidong Cai
Central South University
Pengfei Zhang
Central South University
Xiong Peng
Central South University
Yuqian Zhang
Central South University
Hui Xie
Central South University
Xiang Wang (wangxiang@csu.edu.cn)
Central South University https://orcid.org/0000-0002-5211-7206

Primary research

Keywords: Tumor mutation burden, immune, machine learning, lung adenocarcinoma

Posted Date: February 23rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-219997/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Background: Lung adenocarcinoma (LUAD) accounts for a majority of cancer-related deaths worldwide annually. A recent study shows that immunotherapy is an effective method of LUAD treatment, and tumor mutation burden (TMB) was associated with the immune microenvironment and affected the immunotherapy. Exploration of the gene signature associated with tumor mutation burden and immune infiltrates in predicting prognosis in lung adenocarcinoma in this study, we explored the correlation of TMB with immune infiltration and prognosis in LUAD.

Materials and Methods: In this study, we firstly got mutation data and LUAD RNA-Seq data of the LUAD from The Cancer Genome Atlas (TCGA), and according to the TMB we divided the patients into high/low-TMB levels groups. The gene ontology (GO) pathway enrichment analysis and KOBAS-Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analysis were utilized to explore the molecular function of the differentially expressed genes (DEGs) between the two groups. The function enrichment analyses of DEGs were related to the immune pathways. Then, the ESTIMATE algorithm, CIBERSORT, and ssGSEA analysis were utilized to identify the relationship between TMB subgroups and immune infiltration. According to the results, Venn analysis was utilized to select the immune-related genes in DEGs. Univariate and Lasso Cox proportional hazards regression analyses were performed to construct the signature which positively associated with the immune infiltration and affected the survival. Finally, we verified the correlation between the signature and immune infiltration.

Result: The exploration of the immune infiltration suggested that high-TMB subgroups positively associated with the high level of immune infiltration in LUAD patients. According to the TMB-related immune signature, the patients were divided into High/Low-risk groups, and the high-risk group was positively associated with poor prognostic. The results of the PCA analysis confirmed the validity of the signature. We also verified the effectiveness of the signature in GSE30219 and GSE72094 datasets. The ROC curves and C-index suggested the good clinical application of the TMB-related immune signature in LUAD prognosis. Another result suggested that the patients of the high-risk group were positively associated with higher TMB levels, PD-L1expression, and immune infiltration levels.

Conclusion: In conclusion, our signature provides potential biomarkers for studying aspects of the TMB in LUAD such as TMB affected immune microenvironment and prognosis. This signature may provide some biomarkers which could improve the biomarkers of PD-L1 immunotherapy response and were inverted for the clinical application of the TMB in LUAD. LUAD male patients with higher TMB-levels and risk scores may benefit from immunotherapy. The high-risk patients along with higher PD-L1 expression of the signature may suitable for immunotherapy and improve their survival by detecting the TMB of LUAD.

1. Background

Lung cancer (LC) is one of the most common cancers worldwide and the main cause of cancer-related mortality (Bray et al., 2018; Torre et al., 2015). In 2018, Non-small cell lung cancer (NSCLC) accounts for

85% of all LCs. The 5-year survival rate after the diagnosis of LC is 15.6% (Nanavaty, Alvarez, & Alberts, 2014). In NSCLC, Lung adenocarcinoma (LUAD) is the major histological subtype (Inamura, 2018; Sangodkar, Katz, Melville, & Narla, 2010), accounting for more than half of its morbidity and mortality (Cronin et al., 2018). Despite recent advances in surgical methods, neoadjuvant therapies, immunotherapies, and so on. But the prognosis of LUAD is still not optimistic. Recently, immunotherapy has been recognized to be an effective method for lung cancer (Remon, Vilariño, & Reguart, 2018). However, there are no biomarkers for assessing the effectiveness of immunotherapy in LUAD.

Tumor mutation burden (TMB)—the number of somatic, coding, base substitution, and indel mutations per million bases—It can be used to predict the efficacy of immune checkpoint blockade and is a useful biomarker in some cancer types to identify patients who will benefit from immunotherapy. (Samstein et al., 2019; Zehir et al., 2017). Tumor cells with high TMB may have more neoantigens, with an associated increase in cancer-fighting T cells in the tumor microenvironment and periphery. These neoantigens can be recognized by T cells, inciting an anti-tumor response (Steuer & Ramalingam, 2018). In one study, patients with high levels of TMB have a significantly better response to immunotherapy (Chan et al., 2019). In summary, TMB is suggested to be an independent predictor of immunotherapy response in various types of cancers including lung cancer (Budczies et al., 2019; Goodman et al., 2017; Hellmann et al., 2018). However, there are still limitations such as the TMB threshold, and these challenges need to be resolved before they can be widely applied to different types of tumors. In order to achieve the best utility and alignment for the TMB application in the clinical therapy, still need to have a system research process and entry point for follow-up study(Le et al., 2017; Middha et al., 2017; Topalian et al., 2012).

With the development of next-generation sequencing (NGS) technology and the establishment of the Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO), and other public datasets (Liu et al., 2019), we could analyze the TMB datasets in LUAD. In this study, we identified the relationship between the TMB value and immune infiltrates. Then, according to the DEGs of the high/low-TMB subgroups, we constructed TMB-related immune signature based on TMB by RNA sequencing data from TCGA-LUAD and verified it by GEO datasets. This signature included nine TMB-related immune genes, and high-risk patients of the signature with higher expression of the PD-L1, immune scores, and TMB value. Given the quantitative application of PD-L1 in the prediction of immunotherapy, the response is imperfect, it is necessary to improve the biomarkers of response (Fridman, Pagès, Sautès-Fridman, & Galon, 2012; Tumeh et al., 2014). Our research may provide some biomarkers which could improve the biomarkers of response and were inverted for the clinical application of the TMB in LUAD.

2. Methods

2.1 Data download and analysis

Masked somatic mutation data were obtained from the TCGA-LUAD database via the GDC data portal (https://portal.gdc.cancer.gov/). Gene expression profiles and associated clinicopathological data of LUAD patients were from the TCGA database, with samples including 497 cancer tissue samples and 54

normal tissue samples (Mayakonda, Lin, Assenov, Plass, & Koeffler, 2018). The number of obtainable clinical cases for the selected subjects was 445 after removing 52 patient samples from the study due to a lack of clinical information (such as survival time, TNM stage, and so on) or survival time less than 30 days (avoiding non-cancer-related death samples). Next, we downloaded the LUAD genes expression data and clinical data from GEO (https://www.ncbi.nlm.nih.gov/geo/), under the accession number GSE30219 and GSE72094. After checking the data consistency, 83 LUAD samples, and 393 LUAD samples were utilized to construct the testing cohort for validating the prognostic value of the TMB-related immune signature.

2.2 TMB value estimation and analysis

TMB was defined as the number of somatic mutations per million bases. First, we calculated the mutation frequency with the number of variants at the length of exons (38 million) for each sample via Perl scripts. Then, we classified LUAD samples into the low- and the high-TMB levels groups by using the median of TMB data (6.38) as the cutoff value. The 'limma' R package was utilized to identify TMB-related DEGs between low- and high-TMB groups (Ritchie et al., 2015), and all DEGs with |log FC| > 1 (|log FC| > 1 indicates multiple differences in the gene expression greater than one between the low-TMB levels group and high-TMB levels group) and False Discovery Rate (FDR) < 0.05 were exported, and the 'pheatmap' R package was used to perform heatmap. Then, we got 468 DEGs. The GO pathway enrichment analysis and KEGG pathways analysis were performed to explore the molecular function of the DEGs by 'clusterProfiler, org.Hs.eg.db, plot, ggplot2' in R package (Pathan et al., 2015), where $P < 0.05$, FDR q-value < 0.25 was considered statistically significant. The results suggested that DEGs were positively enriched in immune-related pathways.

2.3 The correlation analysis between the TMB subgroups and immune infiltrate

The abundance of 22 leukocyte subtypes in LUAD was calculated by using the "CIBERSORT" R package and the transcriptome data of the TCGA-LUAD, with a cut-off $P < 0.05$. The distributions of immune cells in each LUAD patient were shown by the "pheatmap" package. The differential abundances of immune cells between low- and high-TMB levels groups were compared by the Wilcoxon rank-sum test and exhibited with P by the "vioplot" package (Newman et al., 2015). Next, we analyzed 29 immune-related genes representing a variety of immune cell types, functions, and pathways, and we quantified enrichment levels of active immune cells, functions, or pathways between the TMB subgroups by the "ssGSEA" R package(Yi, Nissley, McCormick, & Stephens, 2020). The Stromal score, Immune score, Estimated score, and Tumor purity of TCGA-LUAD transcriptome expression data were calculated by ESTIMATE algorithm, and the correlation between TMB subgroups and immune infiltrate was verified(Montuno, Kohner, Foote, & Okun, 2013). Heatmap and statistical maps were plotted by using the "pheatmap" R package. Subsequently, the expression levels of human leukocyte antigen (HLA) and CD274 (PD-L1) were used to verify the immune infiltration levels of the TMB subgroups(Shen, Peng, & Shen, 2020).

2.4 The construction of the TMB-related immune signature

According to the GO and KEGG pathways analysis, we download 1769 immune-related genes from the Immunology Database and Analysis Portal (https://www.immport.org/shared/home) to select the TMB-related immune DEGs between the two groups through the “VennDiagram” package. After the Venn analysis, 42 TMB-related immune genes were used for subsequent research. Then, Univariate cox proportional hazards regression analyses were utilized to screen the prognosis-related genes. Finally, Lasso Cox proportional hazards regression analyses were utilized to prevent the overfitting of the signature and calculate the coefficient of the DEGs. Then, the prognosis signature was constructed by the formula: $TMBPI = \sum_{n=x}^{n} \mathrm{coef(X)} * \exp(\mathrm{X})$. (van Oest, 2019). (van Oest, 2019). R software packages "survival" and "survminer" are used to determine the optimal cut-off value for risk scores and plot Kaplan–Meier survival curves (AWH et al., 2018). In particular, depending upon the cut-off value, we divided patients into high-risk and low-risk groups. PCA analysis was performed to confirm the efficacy of the signature subgroup. The R software package "survivalROC" was used to plot the time-dependent receiver operating characteristic (ROC) curves for predicting diagnostic value (Heagerty, Lumley, & Pepe, 2000). A significant clinical diagnostic value was presented when the area under the curve (AUC) value lies in the range of 0.5–0.9, in which a bigger AUC valve is with a higher diagnostic value (Hajian-Tilaki, 2013). The concordance index (C-index) was used to evaluate the predictive ability of the risk model. In practical applications, the concordance index (C-index) has lower accuracy from 0.50–0.70; medium accuracy between 0.71–0.90; and higher accuracy from the degree more than 0.90 (Kim, Schaubel, & McCullough, 2018). Univariate and multivariate cox proportional hazards regression analysis were performed to verify the signature was independent prognosis factor in TCGA-LUAD. At last, the same were utilized to validate the performance of the signature in GSE30219 and GSE72094.

2.5 Identification of the correlation between TMB-related prognosis signature, clinical traits, and the immune microenvironment

Analysis of the relationship between TMB subgroups, and clinical characteristics (age, gender, stage grading, tumor, and TNM staging) was performed by Wilcoxon rank-sum test in the R package and plotted by “pheatmap” package. Kaplan–Meier survival analysis was performed to verify the performance of the signature of each clinical characteristic in TCGA-LUAD. “limma” package was utilized to identify the Immune score, TMB levels, and the expression of the PD-L1between the high/low-risk groups. “ggpubr” package was performed to plot the results. At last, the “CIBERSORT” algorithm was utilized to calculate the content of immune cells in each TCGA-LUAD patients, and Spearman analysis was utilized to explore the correlation between the immune cells and high/low-risk groups. The R package “ggplot2, ggpubr, and ggExtra” were used to plot the results (Newman et al., 2019; Warne & Burningham, 2019).

2.6 Statistical validation

The univariate and multivariate Cox proportional-hazard models were used to evaluate the hazard ratios of prognostic factors. Subgroup differences were analyzed by the Wilcoxon rank-sum test, Spearman analysis, or Kruskal test. All statistical analysis was performed using the R software (Version 3.6.1). $P < 0.05$ was thought to be significant.

3. Results

3.1 Construction of TMB grouping for TCGA-LUAD and identification of the DEGs between the TMB subgroups.

The workflow of the study was shown in Fig. 1. In our study, 445 LUAD patients of the TCGA data set were assigned to the training sample cohort. 83 LUAD patients of the GSE30219 dataset and 393 LUAD patients of the GSE72094 dataset were assigned to the testing sample cohort for batch processing. According to the median value of the TMB (6.38), the TCGA-LUAD patients were divided into two groups, including the high-TMB levels group (n = 240) and low-TMB levels group (n = 205) (*Table. S1*). After the differential analysis, 468 DEGs with |Log FC | > 1 were used for subsequent analysis (*Figure. 2A, Table. S2*). For those TMB-related genes, KEGG analysis demonstrated that these DEGs primarily participated in Hematopoietic cell lineage and Neuroactive ligand – receptor interaction (*Figure. 2B, Table. S3*), and GO analysis revealed that these DEGs were mostly enriched in leukocyte mediated cytotoxicity, external side of the plasma membrane, and receptor-ligand activity among Biological Process (BP), Cellular Component (CC), and Molecular Function (MF), respectively (*Figure. 2C, Table. S4*). These results suggested that DEGs were positively enriched in immune-related pathways.

3.2 identification of the immune infiltrates TMB subgroups.

After performed the GO and KEGG analysis. Twenty-nine immune-related terms were included to eliminate the richness of multiple immune cell types in TMB subgroups (*Figure. 3A*). Then, based on the expression profile of TCGA-LUAD, an estimation algorithm was utilized to calculate tumor purity, estimated score, immune score, and stromal score to verify the feasibility of the above grouping strategy (*Table S5*). Compared with the low-TMB levels group, the high-TMB levels group had lower tumor purity, but a higher estimated score, immune score, and stroma score (*Figure. 3A*). Thus, the violin plot suggested that the high-TMB levels group was significantly positively correlated with the ESTIMATE Score, Immune Score, and Stromal Score, while the low-TMB levels group was negatively correlated with the tumor purity ($P < 0.05$) (*Figure. 3C-F*). At the same time, we found that the expression of the HLA family and PD-L1 was significantly higher in the group with higher TMB levels than in the group with lower TMB levels ($P < 0.05$) (*Figure. 3G-H*). “CIBERSORT” algorithm and Wilcoxon rank-sum test were utilized to perform the analysis of the immune cells, 205 low-TMB levels samples and 240 high-TMB levels samples were thought to be significant (*Table. S6*). The results were revealed that the infiltration levels of CD8 + T cell, Activated CD4 + memory T cell, Follicular helper T cells, Resting NK cells, M1 Macrophages in the high-TMB levels group were higher compared with those in the low-TMB levels group, whereas Resting CD4 + memory T cell, Regulatory T cells (Tregs), Monocytes, Resting Dendritic cells, Activated Dendritic cells, and Resting Mast

cells in the high-TMB levels group were lower compared with that in the low-TMB levels group (*Figure. 3B*). These results suggested that the TMB subgroups of TCGA-LUAD were positively associated with the immune infiltrate.

3.3 Screening of the TMB-related genes and construction of the prognosis signature.

Through the Venn analysis of 1769 immune genes (https://www.immport.org/resources) and 468 DEGs, 42 TMB-related immune genes for subsequent analysis (*Figure. 4A, Table. S6*). Through univariate and lasso Cox proportional hazards regression, nine TMB-related immune DEGs were utilized to construct the prognosis signature (*Figure.4B-D*, Table 1). The results of the PCA analysis identified the validity of the signature (*Figure. 4F*). The high-risk group was significantly associated with a poor prognosis (*Figure. 4F, $P < 0.001$*). The AUC of the ROC curves was 0.707, 0.700, and 0.660 for the 1, 3, and 5-year OS, separately (*Figure. 4G*). To test the robustness of the prognostic signature, according to the risk score cut-off of the training cohort, GSE72094 and GSE30219 were structured into a high-risk group and a low-risk group. In the GSE72094 cohort, the Kaplan–Meier survival curves of the prognostic signature have a statistically significant difference in the two predicted risk groups (*Figure. 4H, $P < 0.001$*). The AUC was 0.723, 0.727, and 0.732 for the 1, 3, and 5-year OS, separately (*Figure. 4I*). And in the GSE30219 cohort, Kaplan–Meier survival curves of the prognostic signature had a statistically significant difference in the two predicted risk groups (*Figure. 4J, $P < 0.001$*). The AUC was 0.707, 0.713, and 0.710 for the 1, 3, and 5-year OS, separately (Figure. 4K). The C-index was also utilized to verify the good performance of the signature in three cohorts (*Table. 2*). The Kaplan–Meier curve showed that the survival of patients was significantly poorer in the high-risk group in clinical subgroups of the TCGA-LUAD cohort (*Figure. 5*). Although some results of clinical subgroups were not statistically significant, the patients in high-risk groups showed poor prognosis than the low-risk group. The result of the patients with the M1 stage was not considered due to a too low number of samples. Univariate and multivariate Cox regression analysis confirmed the independent prognostic value of risk models in the LUAD cohort (*Figure. 6A-F*).

Table 1
The 9-gene signature screened by their coefficients.

Id	Coefficient	HR	HR.95L	HR.95H	P-value
CD1B	-2.36E-02	8.68E-01	7.59E-01	9.93E-01	3.86E-02
CD1E	-3.11E-02	9.36E-01	8.78E-01	9.97E-01	4.12E-02
INSL4	1.20E-02	1.02E + 00	1.01E + 00	1.03E + 00	3.15E-03
KLRC2	1.95E-01	1.33E + 00	1.19E + 00	1.50E + 00	1.81E-06
NR1I2	1.35E-01	1.15E + 00	1.02E + 00	1.30E + 00	2.36E-02
SCGB3A1	-1.52E-04	1.00E + 00	9.99E-01	1.00E + 00	1.60E-02
SFTPA1	-5.66E-05	1.00E + 00	1.00E + 00	1.00E + 00	1.84E-02
SFTPA2	-3.53E-05	1.00E + 00	1.00E + 00	1.00E + 00	1.55E-02
SLC10A2	4.09E-02	1.04E + 00	1.01E + 00	1.06E + 00	2.79E-03
HR, hazard ratio					

Table 2
The C-index of the TMB-related immune signature in the TCGA-LUAD training cohort, GSE30219 testing cohort, and GSE72094 testing cohort.

	Risk model		
	C-index	HR (95% CI)	*P*
TCGA-LUAD	0.631	0.587–0.675	5.93E-09
GSE72094	0.656	0.584–0.689	3.54E-07
GSE30219	0.638	0.555–0.641	7.19E-06
HR: Hazard Ratio; CI: Confidence Interval; TMB, tumor mutation burden.			

Taken together, these results indicated a better predictive performance about our prognostic signature.

3.4 The correlation of the TMB-related immune signature and clinical traits.

According to the Fig. 7, the results showed that the risk scores of the high-TMB levels patients were higher than the low-TMB level patients, the results suggested that the prognosis of the high-TMB levels patients were poor than the low-TMB level patients (*Figure. 7C, P = 6.70e-11*). Higher TNM stage and N stage were also positively with higher risk scores (*Figure. 7D, P = 0.028; Figure. 7F, P = 0.011*). Notably, when compared to female patients, male patients with the higher risk scores. Although correlation analysis between the risk scores and age, T stage, and M stage was not statistically significant, higher patients' age, T stage, and M stage may increase risk following risk scores (*Figure. 7H-J*). The expression of the PD-L1 was higher in the high-risk group.

3.5 The correlation of the TMB-related immune genes signature and immune infiltrates.

To explore the correlation between the immune-related lncRNA risk model and the level of immune infiltration, we calculated the correlation between the signature and immune cells, the expression of the PD-L1, and immune scores by Spearman correlation analysis. The results suggested that patients with higher PD-L1 expression and immune scores were positively associated with higher risk scores. We found that T cells CD8, T cells CD4 memory activated, NK cells resting, Mast cells activated, Macrophages M1, and Macrophages M0 were positively correlated with risk scores. And the correlation values of them were 0.21, 0.33, 0.20, 0.11, 0.24, and 0.21, respectively. While Dendritic cells activated, Monocytes, Mast cells resting, Dendritic cells resting, T cells CD4 memory resting, and Macrophages M2 are negatively correlated with a risk score, the correlation values of the Plasma cells and Mast cell resting are − 0.19, -0.32, -0.36, -0.47, -0.29 and − 0.19 (*Figure, 8*). In summary, these results indicated that the prognostic signature in LUAD was related to the infiltration of these immune cell subtypes.

Taken together, these results indicated that the signature was positively associated with tumor immune environment and LUAD patient prognosis. These nine TMB-related immune genes signature could perform as prognosis factor.

Table 3

Univariate and Multivariate Cox regression analysis in the TCGA-LUAD training cohort, GSE30219 testing cohort, and GSE72094 testing cohort.

Variables	Univariate analysis				Multivariate analysis			
	HR	HR.95L	HR.95H	P	HR	HR.95L	HR.95H	P
TCGA training cohort(n = 445)								
Age(≥ 65/<65)	1.00E + 00	9.82E-01	1.02E + 00	9.85E-01	1.01E + 00	9.91E-01	1.03E + 00	3.08E-01
Gender	1.05E + 00	7.43E-01	1.49E + 00	7.72E-01	9.23E-01	6.44E-01	1.32E + 00	6.63E-01
TNM stage	1.57E + 00	1.34E + 00	1.84E + 00	3.35E-08	1.30E + 00	8.57E-01	1.97E + 00	2.17E-01
T	1.57E + 00	1.29E + 00	1.92E + 00	7.88E-06	1.19E + 00	9.41E-01	1.51E + 00	1.45E-01
N	1.69E + 00	1.39E + 00	2.06E + 00	1.55E-07	1.26E + 00	8.79E-01	1.82E + 00	2.07E-01
M	1.90E + 00	1.07E + 00	3.38E + 00	2.91E-02	1.06E + 00	3.75E-01	3.01E + 00	9.09E-01
Risk score	2.33E + 00	1.77E + 00	3.09E + 00	2.62E-09	1.93E + 00	1.44E + 00	2.59E + 00	1.13E-05
GSE30219 testing cohort(n = 83)								
Age(≥ 65/<65)	1.04E + 00	1.02E + 00	1.05E + 00	4.50E-07	1.04E + 00	1.02E + 00	1.05E + 00	5.20E-06
Gender	1.83E + 00	1.14E + 00	2.94E + 00	1.25E-02	1.50E + 00	9.25E-01	2.42E + 00	1.01E-01
TNM stage	1.77E + 00	1.53E + 00	2.05E + 00	3.55E-14	1.16E + 00	7.69E-01	1.74E + 00	4.83E-01
T	1.67E + 00	1.46E + 00	1.92E + 00	1.55E-13	1.27E + 00	9.67E-01	1.66E + 00	8.61E-02
N	1.79E + 00	1.53E + 00	2.10E + 00	7.59E-13	1.25E + 00	9.37E-01	1.66E + 00	1.30E-01
M	3.17E + 00	1.48E + 00	6.77E + 00	2.97E-03	1.24E + 00	4.36E-01	3.56E + 00	6.82E-01
Risk score	1.82E + 00	1.56E + 00	2.12E + 00	4.23E-14	1.50E + 00	1.22E + 00	1.83E + 00	8.90E-05

HR: Hazard Ratio

Variables	Univariate analysis				Multivariate analysis			
	HR	HR.95L	HR.95H	P	HR	HR.95L	HR.95H	P
GSE72094 testing cohort(n = 393)								
Age(≥ 65/<65)	1.01E + 00	9.88E-01	1.03E + 00	4.79E-01	1.01E + 00	9.88E-01	1.03E + 00	4.38E-01
Gender	1.55E + 00	1.07E + 00	2.25E + 00	2.19E-02	1.71E + 00	1.17E + 00	2.50E + 00	6.04E-03
TNM stage	1.62E + 00	1.36E + 00	1.94E + 00	8.88E-08	1.67E + 00	1.38E + 00	2.00E + 00	6.41E-08
Risk score	1.58E + 00	1.39E + 00	1.80E + 00	1.46E-12	1.60E + 00	1.40E + 00	1.83E + 00	8.04E-12
HR: Hazard Ratio								

4. Discussion

With the increasing exploration of lung cancer immunotherapy, various kinds of immune checkpoint blockade (ICB) have emerged for several cancers, such as targeting PD1, and PDL1. ICB s have made a significant breakthrough in lung cancer therapy and revolutionized the management of lung cancer (Melosky et al., 2019; van den Ende et al., 2020). Identification of immune-related gene preference and dependence mechanisms in lung cancer regulation become more interesting (Chen et al., 2019; Maung, Ergin, Javed, Inga, & Khan, 2020). TMB was a powerful factor in tumorigenesis, and it has been used to reveal the process of mutation accumulation in tumors and genetic alterations interacting with immune cells in the tumor microenvironment (Goodman et al., 2018; Hsu et al., 2019). TCGA and GEO databases already have a large number of RNA-seq data of tumor samples in multiple cancers. Biomarkers of LUAD have been developed utilizing the TCGA or GEO data (Shang et al., 2017). Therefore, we investigated the relationship between tumor TMB value and the immune microenvironment of LUAD.

In the present study, we analyzed the correlation between TMB value and the immune microenvironment in TCGA-LUAD samples. The findings revealed that there were statistically significant differences in tumor purity, estimate score, immune score, and stromal score between the groups with high and low TMB value subgroups ($P < 0.05$). Besides, the heterogeneity of the immune microenvironment about TMB subgroups was verified by the expression of HLA and PD-L1 and the CIBERSORT algorithm. These results identified the correlation between TMB value and the immune microenvironment in TCGA-LUAD. According to existing studies, higher TMB level patients are suitable for immune checkpoint blockade therapy. In lung cancer, approximately 200 nonsynonymous cell mutations have been identified by whole-exome sequencing to predict TMB thresholds. However, disease-specific TMB thresholds for the effective prediction of responses in lung cancer have not been well established. (Bins et al., 2018; Chan et al., 2019;

Chhabra et al., 2018; Zhang et al., 2017). Besides, the quantitative detection of PD-L1 immunotherapy response is an imperfect biomarker due to its heterogeneous and dynamic nature, diagnostic reproducibility and definition are somewhat difficult, and negative predictive value is insufficient (Cristescu et al., 2018).

Focus on problem-solving, we construct a TMB-related immune signature which may improve the biomarkers for PD-L1 immunotherapy response. This signature was constructed by the immune-related DEGs of the high/low-TMB levels groups. The results identified that the prognosis signature was positively associated with the OS of the TCGA-LUAD patients, and the clinical performance of the signature was verified by ROC analysis and C-index in the GSE30219, and GSE72094 cohorts. Moreover, the prognosis signature was closely correlated with TMB value, immune score, and the PD-L1 expression. In the signature, high-risk group patients increase the risk with higher TMB levels. The expression of the PD-L1 and immune scores were also higher in the high-risk group. These results suggested that patients in high-risk groups may benefit from immunotherapy. Notably, the correlation between clinical features and prognosis signature suggests that gender is related to the high-risk group, and the high-risk patients are more male. Concerning the latest research progress, compared with male patients, immune checkpoint inhibitors (ICIs) are often more effective in female cancer patients. Tumors are usually more antigenic in males, which can be reflected by the increase in tumor mutations and cancer germline antigens. TMB could regard as a biomarker reflecting tumor antigenicity, which was more effective than male patients in predicting the immunotherapy response of female lung cancer patients (Wang, Cowley, & Liu, 2019; Wang, Zhang, He, Wu, & Liu, 2019). According to our research, the results suggested that TMB also occurred in LUAD male patients and the prognosis of LUAD male patients would be poor. It may improve the prognosis of the LUAD male patients. Besides, the higher TNM stage and N stage were also positively with higher risk scores, which will contribute towards the effectiveness and usability of the prognosis signature.

At last, we surveyed background information about the DEGs of the signature. Current studies have shown that CD1family genes are related to the tumor microenvironment. CD1B was significantly positively correlated with checkpoint molecules HAVCR2 and CTLA4 (Yuan et al., 2020). Another study showed that in LUAD, a higher CD1B expression showed significantly increased fractions of Resting Mast cells, Monocytes, Memory B cells, and Activated Resting Dendritic cells, but significantly decreased fractions of Follicular Helper T cells, Resting NK cells, M0 Macrophages, and Naive B cells (Guo et al., 2020; Pereira et al., 2019). SCGB3A1 was a significant independent predictor of poor clinical outcome in early-stage NSCLC (Marchetti et al., 2004). SCGB3A1 also was an inhibitor of tumor cell growth, invasion, and AKT activation (Krop et al., 2005). According to the new study, SCGB3A1 was the target gene of the HIF-2αand repressed tumor cell growth through ATK pathways in $Kras^{G12D}$-driven NSCLC cells (Mazumdar et al., 2010). Recent studies confirm that KLRC2 is a novel checkpoint inhibitor that is expressed on NK cells and activated CD8 + T cell subsets. KLRC2 blocking antibodies release the reactivity of these effector cells, thereby inhibiting tumor progression. Monalizumab is inhibiting this checkpoint in humans, and future clinical trials will reveal how effective it is in combination with other cancer treatments(André

et al., 2018; van Hall et al., 2019). In addition, SFTPA2 and SFTPA2 variants differentially affect lung function and exhibit sex-specific differences (Lv & Huang, 2020). Studies related to the immune microenvironment of INSL4, SLC10A2, and NR1I2 were scarce and need to be further studied.

Compared to other studies, we first used TMB-related immune genes to build a prognosis model from TCGA-LUAD and validated it in two GEO datasets. Our genes signature also had a better prediction ability in three datasets. This signature perhaps provides potential biomarkers for studying the relationship of TMB, immune microenvironmental diseases, immunotherapy, and therapeutic responses, which may be helpful for the diagnosis, treatment, and prognosis of LUAD. And this signature may also improve the biomarkers for PD-L1 immunotherapy response, and makes up for the instability of quantitative detection of PD-L1 immunotherapy response.

However, our research still has some limitations: first, lack of basic experiment to validate the association between prognosis signature and tumor cell immune infiltrates; Second, lack of large clinical sample to verify the prognostic effect of prognosis signature and its potential function mechanism with immune infiltrates. Relevant variants and big sample clinical trials are needed in the future to validate the performance of the prognosis signature combined with the PD-L1 in LUAD.

Conclusions

In conclusion, our signature provides potential biomarkers for studying aspects of the TMB in LUAD such as TMB affected immune microenvironment and prognosis. This signature may provide some biomarkers which could improve the biomarkers of PD-L1 immunotherapy response and were inverted for the clinical application of the TMB in LUAD. LUAD male patients with higher TMB-levels and risk scores may benefit from immunotherapy. The high-risk patients along with higher PD-L1 expression of the signature may suitable for immunotherapy and improve their survival by detecting the TMB of LUAD.

Abbreviations

LUAD: Lung adenocarcinoma; TMB: tumor mutation burden; TCGA: The Cancer Genome Atlas; GEO: Gene Expression Omnibus; ROC: receiver operating characteristic; OS: overall survival; LC: Lung cancer; NSCLC: Non-small cell lung cancer; NGS: next-generation sequencing; K-W test: Kruskal-Wallis test; FC: Fold change; FDR: False Discovery Rate; GO: gene ontology; KEGG: KOBAS-Kyoto Encyclopedia of Genes and Genomes; PPI: protein-protein interaction; TMBPI: tumor mutation burden prognostic index; AUC: area under the curve; TNM: Tumor-node metastasis; C-index: concordance index; ICIs: immune checkpoint inhibitors; ICB: immune checkpoint blockade; BP: Biological Process; CC: Cellular Component; MF: Molecular Function; MHC: major histocompatibility complex; TCRs: T-cell receptors.

Declarations

Ethics approval and consent to participate

There were no ethics associated with this work because the study didn't involve animal experiments or human specimens.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (81672308, X. Wang) and the Hunan Provincial Key Area R&D Programmes (2019SK2253, X. Wang).

Authors' contributions

XW conceived and designed the work. ZYZ and QDC carried out the software coding and data analysis. BXH and PFZ formatted the tables and figures. ZYZ, BXH, and QDC wrote the manuscript. HX and YQZ critically reviewed the codes and the manuscript. All authors read and approved the final manuscript.

Acknowledgments

The authors thank Professor Yongguang Tao for his comments and suggestions throughout the writing process.

References

André, P., Denis, C., Soulas, C., Bourbon-Caillet, C., Lopez, J., Arnoux, T., . . . Vivier, E. (2018). Anti-NKG2A mAb Is a Checkpoint Inhibitor that Promotes Anti-tumor Immunity by Unleashing Both T and NK Cells. *Cell, 175*(7), 1731-1743.e1713. doi:10.1016/j.cell.2018.10.014

AWH, C., J, Z., S, B., H, T., A, C., K, S., . . . PJ, J. (2018). Development of pre and post-operative models to predict early recurrence of hepatocellular carcinoma after surgical resection. *Journal of hepatology, 69*(6), 1284-1293. doi:10.1016/j.jhep.2018.08.027

Bins, S., Basak, E. A., El Bouazzaoui, S., Koolen, S. L. W., Oomen-de Hoop, E., van der Leest, C. H., . . . Mathijssen, R. H. J. (2018). Association between single-nucleotide polymorphisms and adverse events in

nivolumab-treated non-small cell lung cancer patients. *Br J Cancer, 118*(10), 1296-1301. doi:10.1038/s41416-018-0074-1

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin, 68*(6), 394-424. doi:10.3322/caac.21492

Budczies, J., Allgäuer, M., Litchfield, K., Rempel, E., Christopoulos, P., Kazdal, D., . . . Stenzinger, A. (2019). Optimizing panel-based tumor mutational burden (TMB) measurement. *Ann Oncol, 30*(9), 1496-1506. doi:10.1093/annonc/mdz205

Chan, T. A., Yarchoan, M., Jaffee, E., Swanton, C., Quezada, S. A., Stenzinger, A., & Peters, S. (2019). Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann Oncol, 30*(1), 44-56. doi:10.1093/annonc/mdy495

Chen, P. H., Cai, L., Huffman, K., Yang, C., Kim, J., Faubert, B., . . . DeBerardinis, R. J. (2019). Metabolic Diversity in Human Non-Small Cell Lung Cancer Cells. *Mol Cell, 76*(5), 838-851.e835. doi:10.1016/j.molcel.2019.08.028

Chhabra, Y., Wong, H. Y., Nikolajsen, L. F., Steinocher, H., Papadopulos, A., Tunny, K. A., . . . Waters, M. J. (2018). A growth hormone receptor SNP promotes lung cancer by impairment of SOCS2-mediated degradation. *Oncogene, 37*(4), 489-501. doi:10.1038/onc.2017.352

Cristescu, R., Mogg, R., Ayers, M., Albright, A., Murphy, E., Yearley, J., . . . Kaufman, D. (2018). Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science, 362*(6411). doi:10.1126/science.aar3593

Cronin, K. A., Lake, A. J., Scott, S., Sherman, R. L., Noone, A. M., Howlader, N., . . . Jemal, A. (2018). Annual Report to the Nation on the Status of Cancer, part I: National cancer statistics. *Cancer, 124*(13), 2785-2800. doi:10.1002/cncr.31551

Fridman, W. H., Pagès, F., Sautès-Fridman, C., & Galon, J. (2012). The immune contexture in human tumours: impact on clinical outcome. *Nat Rev Cancer, 12*(4), 298-306. doi:10.1038/nrc3245

Goodman, A. M., Kato, S., Bazhenova, L., Patel, S. P., Frampton, G. M., Miller, V., . . . Kurzrock, R. (2017). Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. *Mol Cancer Ther, 16*(11), 2598-2608. doi:10.1158/1535-7163.mct-17-0386

Goodman, A. M., Piccioni, D., Kato, S., Boichard, A., Wang, H. Y., Frampton, G., . . . Kurzrock, R. (2018). Prevalence of PDL1 Amplification and Preliminary Response to Immune Checkpoint Blockade in Solid Tumors. *JAMA Oncol, 4*(9), 1237-1244. doi:10.1001/jamaoncol.2018.1701

Guo, J., Jin, H., Xi, Y., Guo, J., Jin, Y., & Jiang, D. (2020). The miR-582/CD1B Axis Is Involved in Regulation of Dendritic Cells and Is Associated with Clinical Outcomes in Advanced Lung Adenocarcinoma. *Biomed*

Res Int, 2020, 4360930. doi:10.1155/2020/4360930

Hajian-Tilaki, K. (2013). Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian J Intern Med, 4*(2), 627-635.

Heagerty, P. J., Lumley, T., & Pepe, M. S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics, 56*(2), 337-344. doi:10.1111/j.0006-341x.2000.00337.x

Hellmann, M. D., Nathanson, T., Rizvi, H., Creelan, B. C., Sanchez-Vega, F., Ahuja, A., . . . Wolchok, J. D. (2018). Genomic Features of Response to Combination Immunotherapy in Patients with Advanced Non-Small-Cell Lung Cancer. *Cancer Cell, 33*(5), 843-852.e844. doi:10.1016/j.ccell.2018.03.018

Hsu, Y. C., Chang, Y. H., Chang, G. C., Ho, B. C., Yuan, S. S., Li, Y. C., . . . Chen, H. Y. (2019). Tumor mutation burden and recurrent tumors in hereditary lung cancer. *Cancer Med, 8*(5), 2179-2187. doi:10.1002/cam4.2120

Inamura, K. (2018). Clinicopathological Characteristics and Mutations Driving Development of Early Lung Adenocarcinoma: Tumor Initiation and Progression. *Int J Mol Sci, 19*(4). doi:10.3390/ijms19041259

Kim, S., Schaubel, D. E., & McCullough, K. P. (2018). A C-index for recurrent event data: Application to hospitalizations among dialysis patients. *Biometrics, 74*(2), 734-743. doi:10.1111/biom.12761

Krop, I., Parker, M. T., Bloushtain-Qimron, N., Porter, D., Gelman, R., Sasaki, H., . . . Polyak, K. (2005). HIN-1, an inhibitor of cell growth, invasion, and AKT activation. *Cancer Res, 65*(21), 9659-9669. doi:10.1158/0008-5472.can-05-1663

Le, D. T., Durham, J. N., Smith, K. N., Wang, H., Bartlett, B. R., Aulakh, L. K., . . . Diaz, L. A., Jr. (2017). Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science, 357*(6349), 409-413. doi:10.1126/science.aan6733

Liu, M., Zhou, S., Wang, J., Zhang, Q., Yang, S., Feng, J., . . . Zhong, S. (2019). Identification of genes associated with survival of breast cancer patients. *Breast Cancer, 26*(3), 317-325. doi:10.1007/s12282-018-0926-9

Lv, C., & Huang, L. (2020). Xenobiotic receptors in mediating the effect of sepsis on drug metabolism. *Acta Pharm Sin B, 10*(1), 33-41. doi:10.1016/j.apsb.2019.12.003

Marchetti, A., Barassi, F., Martella, C., Chella, A., Salvatore, S., Castrataro, A., . . . Buttitta, F. (2004). Down regulation of high in normal-1 (HIN-1) is a frequent event in stage I non-small cell lung cancer and correlates with poor clinical outcome. *Clin Cancer Res, 10*(4), 1338-1343. doi:10.1158/1078-0432.ccr-1174-03

Maung, T. Z., Ergin, H. E., Javed, M., Inga, E. E., & Khan, S. (2020). Immune Checkpoint Inhibitors in Lung Cancer: Role of Biomarkers and Combination Therapies. *Cureus, 12*(5), e8095. doi:10.7759/cureus.8095

Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C., & Koeffler, H. P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res, 28*(11), 1747-1756. doi:10.1101/gr.239244.118

Mazumdar, J., Hickey, M. M., Pant, D. K., Durham, A. C., Sweet-Cordero, A., Vachani, A., . . . Keith, B. (2010). HIF-2alpha deletion promotes Kras-driven lung tumor development. *Proc Natl Acad Sci U S A, 107*(32), 14182-14187. doi:10.1073/pnas.1001296107

Melosky, B., Juergens, R., McLeod, D., Leighl, N., Brade, A., Card, P. B., & Chu, Q. (2019). Immune checkpoint-inhibitors and chemoradiation in stage III unresectable non-small cell lung cancer. *Lung Cancer, 134*, 259-267. doi:10.1016/j.lungcan.2019.05.027

Middha, S., Zhang, L., Nafa, K., Jayakumaran, G., Wong, D., Kim, H. R., . . . Hechtman, J. F. (2017). Reliable Pan-Cancer Microsatellite Instability Assessment by Using Targeted Next-Generation Sequencing Data. *JCO Precis Oncol, 2017*. doi:10.1200/po.17.00084

Montuno, M. A., Kohner, A. B., Foote, K. D., & Okun, M. S. (2013). An algorithm for management of deep brain stimulation battery replacements: devising a web-based battery estimator and clinical symptom approach. *Neuromodulation, 16*(2), 147-153. doi:10.1111/j.1525-1403.2012.00457.x

Nanavaty, P., Alvarez, M. S., & Alberts, W. M. (2014). Lung cancer screening: advantages, controversies, and applications. *Cancer Control, 21*(1), 9-14. doi:10.1177/107327481402100102

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., . . . Alizadeh, A. A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods, 12*(5), 453-457. doi:10.1038/nmeth.3337

Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., . . . Alizadeh, A. A. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol, 37*(7), 773-782. doi:10.1038/s41587-019-0114-2

Pathan, M., Keerthikumar, S., Ang, C. S., Gangoda, L., Quek, C. Y., Williamson, N. A., . . . Mathivanan, S. (2015). FunRich: An open access standalone functional enrichment and interaction network analysis tool. *Proteomics, 15*(15), 2597-2601. doi:10.1002/pmic.201400515

Pereira, C. S., Pérez-Cabezas, B., Ribeiro, H., Maia, M. L., Cardoso, M. T., Dias, A. F., . . . Macedo, M. F. (2019). Lipid Antigen Presentation by CD1b and CD1d in Lysosomal Storage Disease Patients. *Front Immunol, 10*, 1264. doi:10.3389/fimmu.2019.01264

Remon, J., Vilariño, N., & Reguart, N. (2018). Immune checkpoint inhibitors in non-small cell lung cancer (NSCLC): Approaches on special subgroups and unresolved burning questions. *Cancer Treat Rev, 64*, 21-29. doi:10.1016/j.ctrv.2018.02.002

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res, 43*(7), e47. doi:10.1093/nar/gkv007

Samstein, R. M., Lee, C. H., Shoushtari, A. N., Hellmann, M. D., Shen, R., Janjigian, Y. Y., . . . Morris, L. G. T. (2019). Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet, 51*(2), 202-206. doi:10.1038/s41588-018-0312-8

Sangodkar, J., Katz, S., Melville, H., & Narla, G. (2010). Lung adenocarcinoma: lessons in translation from bench to bedside. *Mt Sinai J Med, 77*(6), 597-605. doi:10.1002/msj.20226

Shang, J., Song, Q., Yang, Z., Li, D., Chen, W., Luo, L., . . . Li, S. (2017). Identification of lung adenocarcinoma specific dysregulated genes with diagnostic and prognostic value across 27 TCGA cancer types. *Oncotarget, 8*(50), 87292-87306. doi:10.18632/oncotarget.19823

Shen, Y., Peng, X., & Shen, C. (2020). Identification and validation of immune-related lncRNA prognostic signature for breast cancer. *Genomics, 112*(3), 2640-2646. doi:10.1016/j.ygeno.2020.02.015

Steuer, C. E., & Ramalingam, S. S. (2018). Tumor Mutation Burden: Leading Immunotherapy to the Era of Precision Medicine? *J Clin Oncol, 36*(7), 631-632. doi:10.1200/jco.2017.76.8770

Topalian, S. L., Hodi, F. S., Brahmer, J. R., Gettinger, S. N., Smith, D. C., McDermott, D. F., . . . Sznol, M. (2012). Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med, 366*(26), 2443-2454. doi:10.1056/NEJMoa1200690

Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., & Jemal, A. (2015). Global cancer statistics, 2012. *CA Cancer J Clin, 65*(2), 87-108. doi:10.3322/caac.21262

Tumeh, P. C., Harview, C. L., Yearley, J. H., Shintaku, I. P., Taylor, E. J., Robert, L., . . . Ribas, A. (2014). PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature, 515*(7528), 568-571. doi:10.1038/nature13954

van den Ende, T., van den Boorn, H. G., Hoonhout, N. M., van Etten-Jamaludin, F. S., Meijer, S. L., Derks, S., . . . van Laarhoven, H. W. M. (2020). Priming the tumor immune microenvironment with chemo(radio)therapy: A systematic review across tumor types. *Biochim Biophys Acta Rev Cancer, 1874*(1), 188386. doi:10.1016/j.bbcan.2020.188386

van Hall, T., André, P., Horowitz, A., Ruan, D. F., Borst, L., Zerbib, R., . . . Vivier, E. (2019). Monalizumab: inhibiting the novel immune checkpoint NKG2A. *J Immunother Cancer, 7*(1), 263. doi:10.1186/s40425-019-0761-3

van Oest, R. (2019). A new coefficient of interrater agreement: The challenge of highly unequal category proportions. *Psychol Methods, 24*(4), 439-451. doi:10.1037/met0000183

Wang, S., Cowley, L. A., & Liu, X. S. (2019). Sex Differences in Cancer Immunotherapy Efficacy, Biomarkers, and Therapeutic Strategy. *Molecules, 24*(18). doi:10.3390/molecules24183214

Wang, S., Zhang, J., He, Z., Wu, K., & Liu, X. S. (2019). The predictive power of tumor mutational burden in lung cancer immunotherapy response is influenced by patients' sex. *Int J Cancer, 145*(10), 2840-2849. doi:10.1002/ijc.32327

Warne, R. T., & Burningham, C. (2019). Spearman's g found in 31 non-Western nations: Strong evidence that g is a universal phenomenon. *Psychol Bull, 145*(3), 237-272. doi:10.1037/bul0000184

Yi, M., Nissley, D. V., McCormick, F., & Stephens, R. M. (2020). ssGSEA score-based Ras dependency indexes derived from gene expression data reveal potential Ras addiction mechanisms with possible clinical implications. *Sci Rep, 10*(1), 10258. doi:10.1038/s41598-020-66986-8

Yuan, J., Yuan, B., Zeng, L., Liu, B., Chen, Y., Meng, X., . . . Yang, S. (2020). Identification and validation of tumor microenvironment-related genes of prognostic value in lung adenocarcinoma. *Oncol Lett, 20*(2), 1772-1780. doi:10.3892/ol.2020.11735

Zehir, A., Benayed, R., Shah, R. H., Syed, A., Middha, S., Kim, H. R., . . . Berger, M. F. (2017). Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med, 23*(6), 703-713. doi:10.1038/nm.4333

Zhang, Y., Wang, D. C., Shi, L., Zhu, B., Min, Z., & Jin, J. (2017). Genome analyses identify the genetic modification of lung cancer subtypes. *Semin Cancer Biol, 42*, 20-30. doi:10.1016/j.semcancer.2016.11.005

Figures

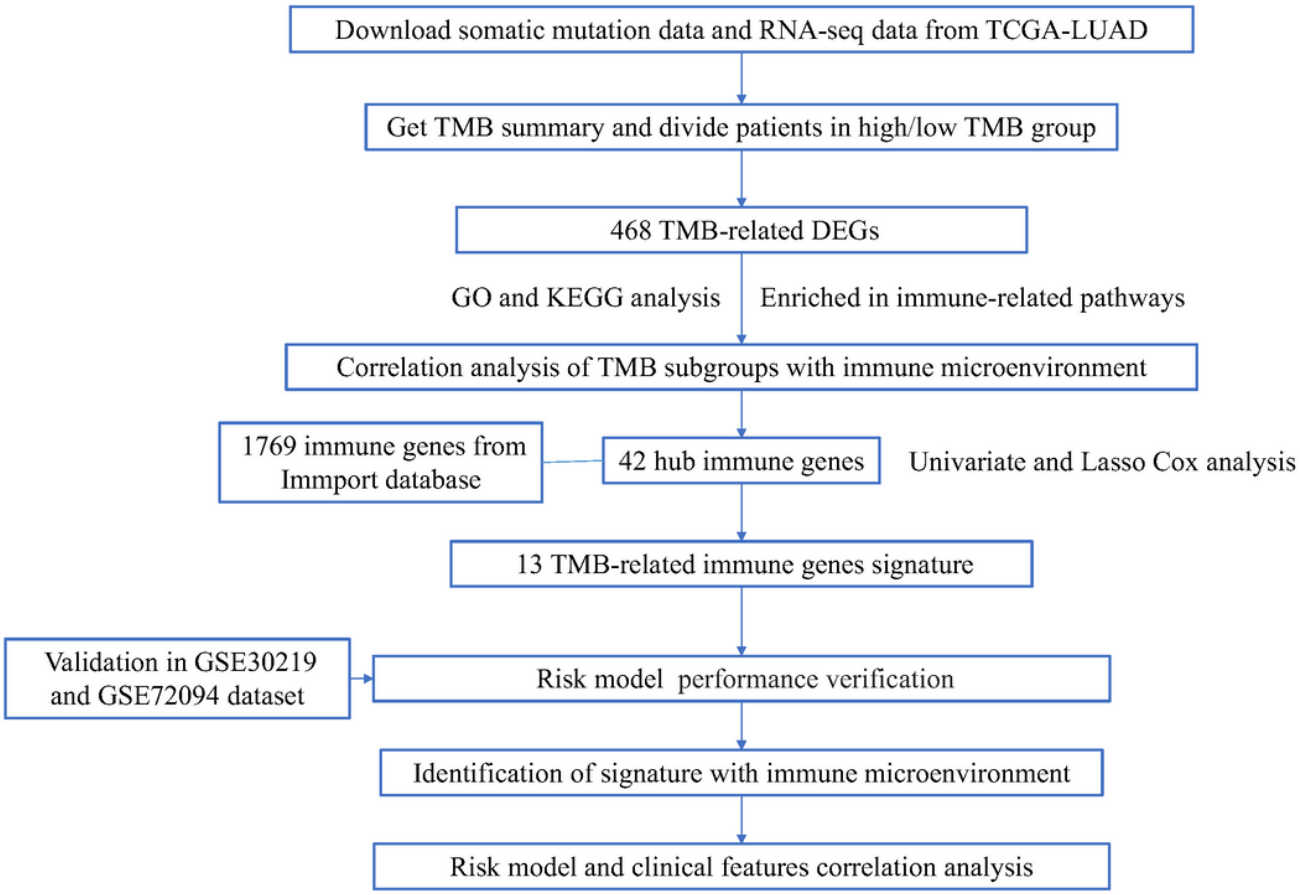

Figure 1

workflow of the study.

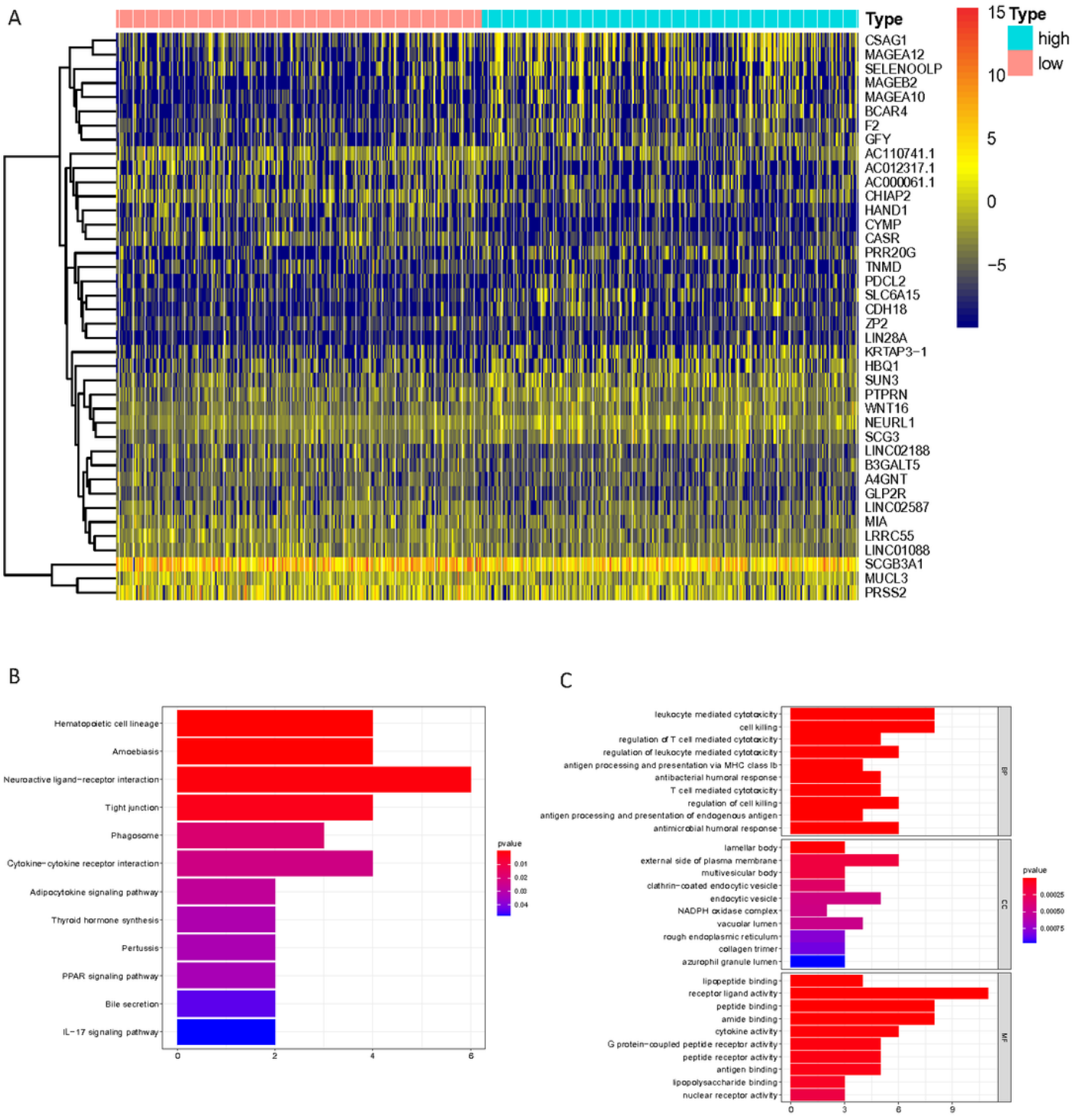

Figure 2

(A) The heatmap plot of top 40 DEGs in low- and high-TMB samples; (B) KEGG revealed that these genes were involved in immune-related pathways; (C) GO results revealed that these genes were involved in immune-related pathways. BP, CC, and MF; BP, Biological Process; CC, Cellular Component; MF, Molecular Function; DEG, differentially expressed gene; TMB, tumor mutation burden.

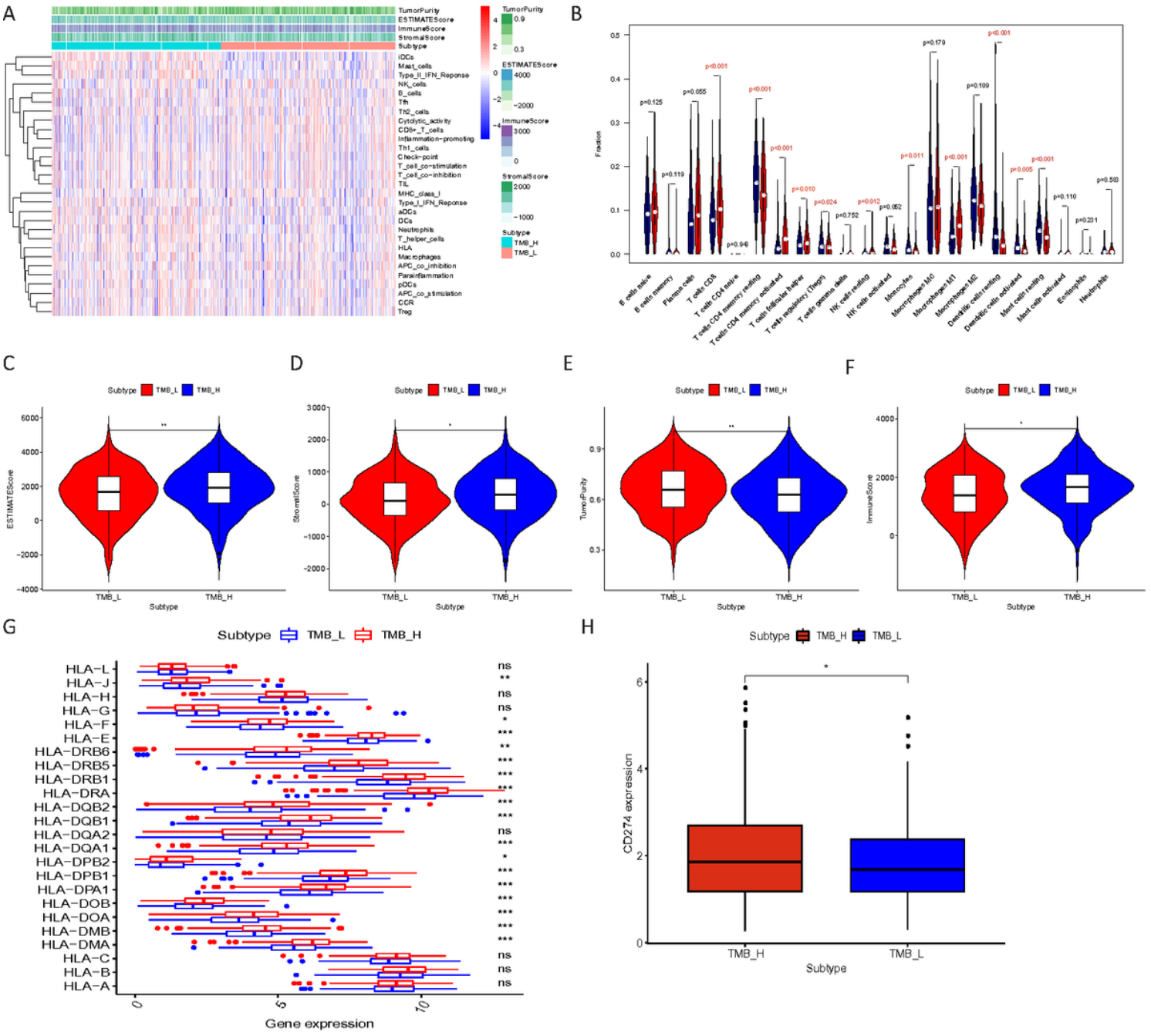

Figure 3

The correlation analysis between TMB subgroups and infiltration (A) The expression of the immune cells expressed were higher in high-TMB levels than low-TMB levels group. Using ESTIMATE's algorithm, the Tumor Purity, ESTIMATE Score, Immune Score and Stromal Score of each sample were displayed together with the grouping information; (B) Wilcoxon rank-sum test revealed the infiltration levels of 22 immune fractions in low-TMB levels group (in blue) and in high-TMB levels group (in red); (C-F) The box-plot showed that there was a statistical difference in Tumor Purity, ESTIMATE Score, Immune Score and Stromal Score between the TMB subgroups ($P < 0.05$); (G) The expression of HLA family genes in high-TMB levels group (red) were significantly higher than that in low-TMB levels group (blue); (H) The

expression of PD-L1 in high-TMB levels group (red) were significantly higher than that in low-TMB levels group (blue). TMB, tumor mutation burden.

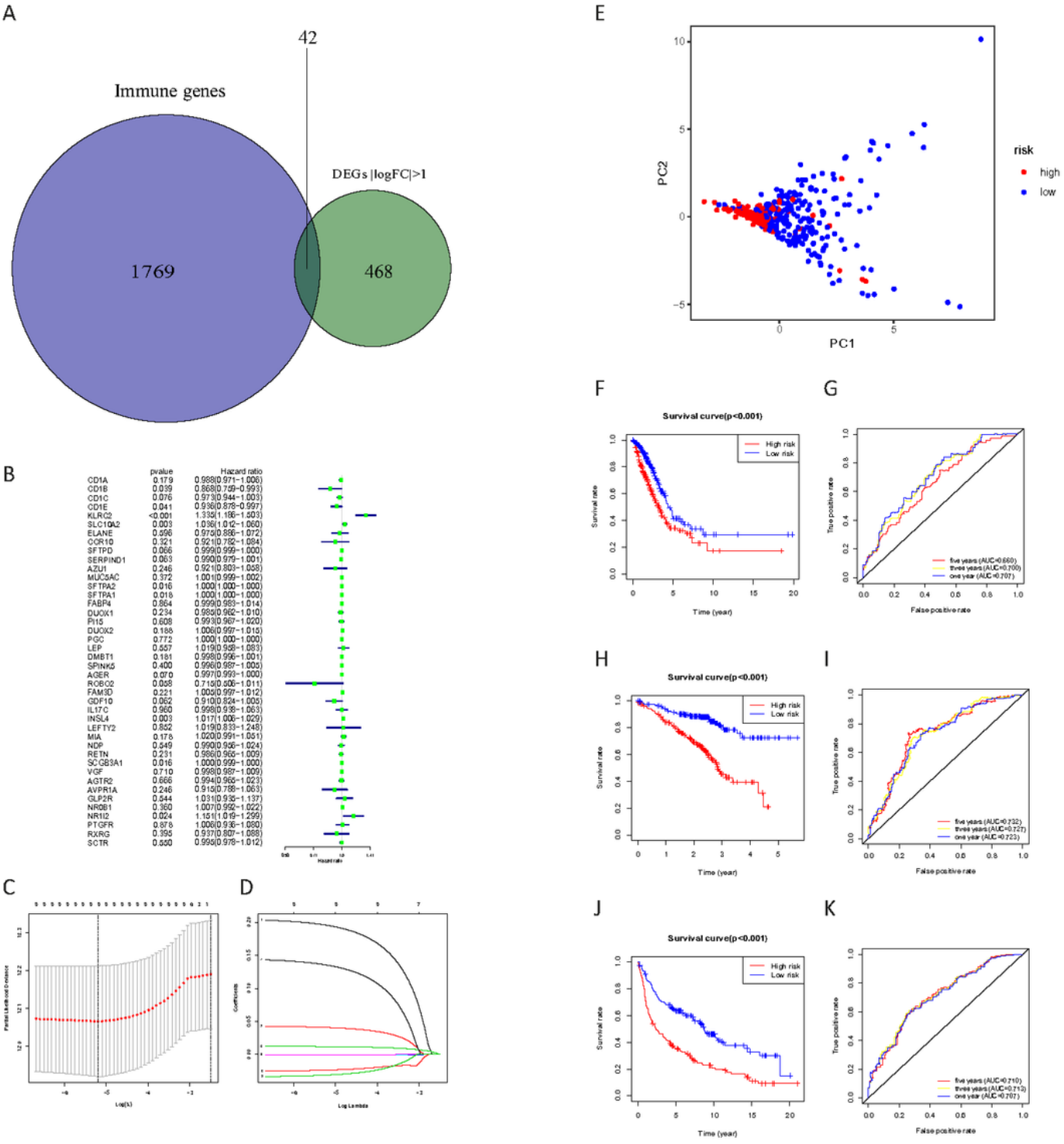

Figure 4

(A) Identification of 42 TMB-related immune genes; (B) The univariable Cox HR regression analysis of selected genes in the TMB-related immune genes (P < 0.05); (C-D) The LASSO Cox analysis identified 9 TMB-related immune genes most correlated with prognostics and the optimal values of the penalty

parameter were determined by 10-round cross-validation; (E) The PCA analysis identified the performance of the grouping; (F-G) Patients in the high-risk group (red) exhibited worse overall survival than those in the low-risk group (blue) in TCGA-LUAD cohort; Time-dependent ROC analysis of the prognosis signature in TCGA-LUAD cohort; (H-I) Patients in the high-risk group (red) exhibited worse overall survival than those in the low-risk group (blue) in GSE72094 cohort; Time-dependent ROC analysis of the prognosis signature in GSE72094 cohort; (H-I) Patients in the high-risk group (red) exhibited worse overall survival than those in the low-risk group (blue) in GSE30219 cohort; Time-dependent ROC analysis of the prognosis signature in GSE30219 cohort. HR, hazard ratio; DEG, differentially expressed gene; TMB, tumor mutation burden.

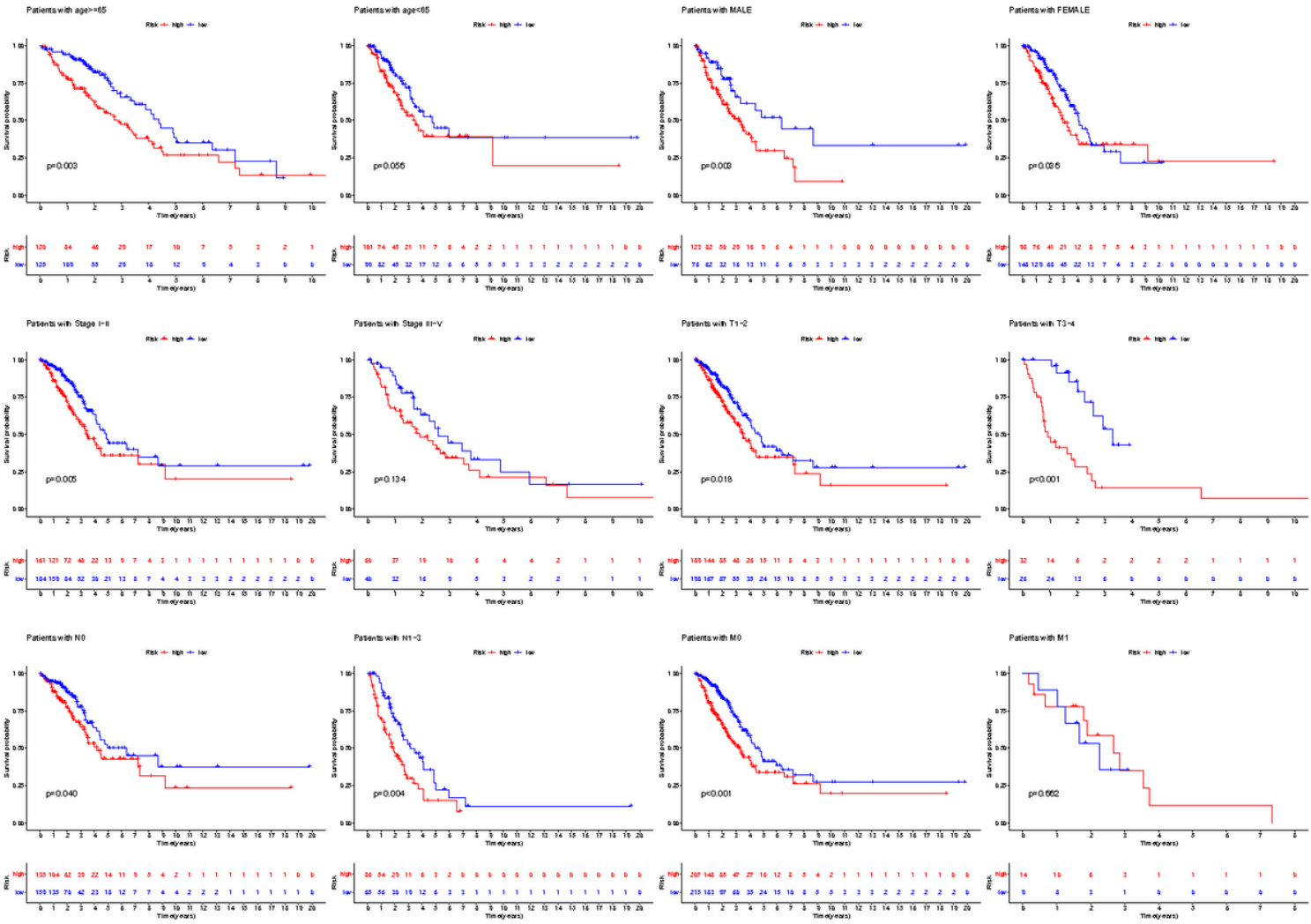

Figure 5

The Kaplan–Meier curve showed that the survival of patients was significantly poorer in the high-risk group in clinical subgroups of the TCGA-LUAD cohort ($P < 0.05$).

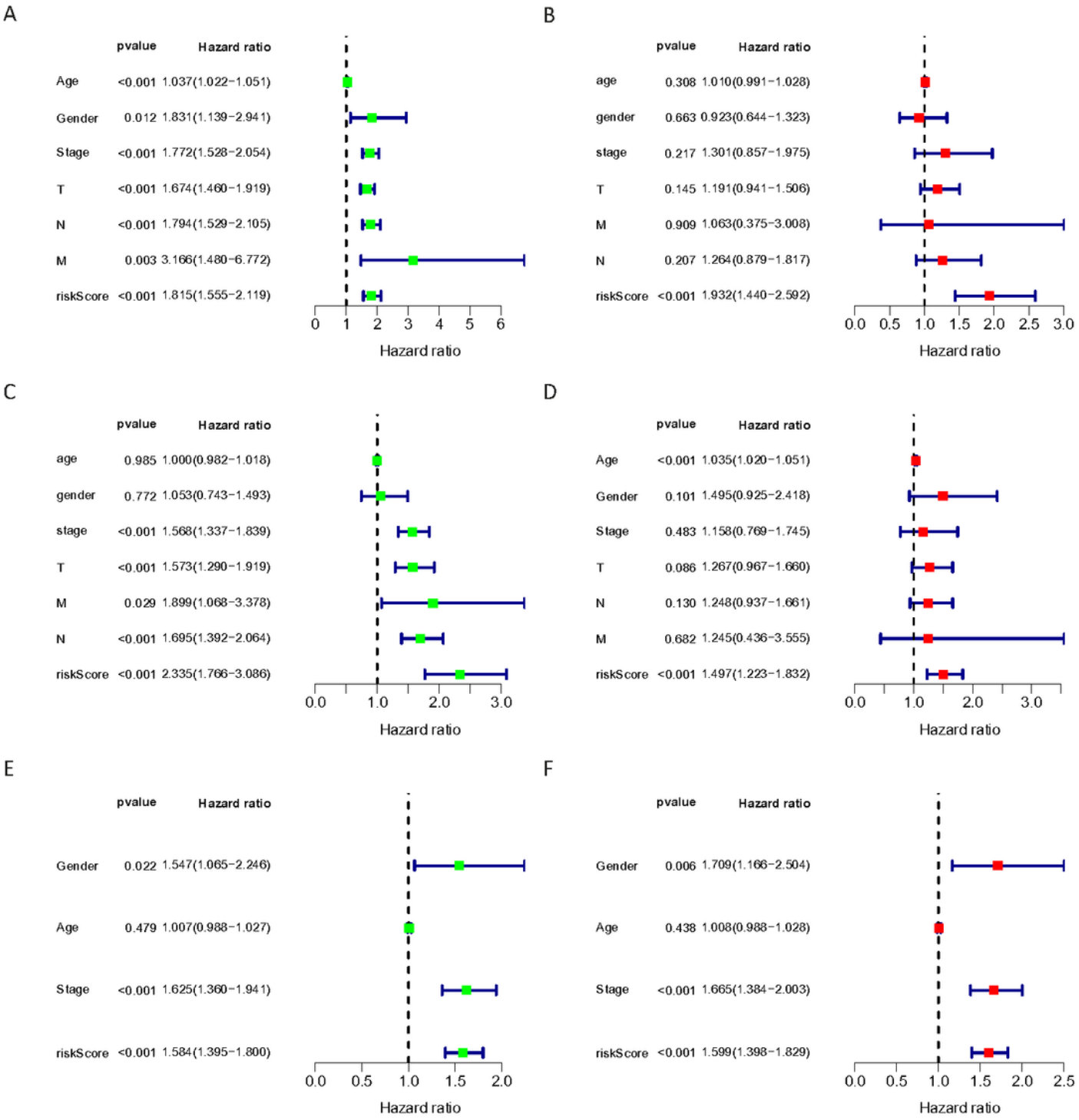

Figure 6

The Cox regression analysis for evaluating the independent prognostic value of the risk score. (A-B) The univariate (A) and multivariate (B) Cox regression analysis of risk score, age, gender, and TNM stage in TCGA-LUAD cohort; (C-D) The univariate (C) and multivariate (D) Cox regression analysis of risk score, age, gender, and TNM stage in GSE30219 cohort; (E-F) The univariate (E) and multivariate (F) Cox regression analysis of risk score, age, gender, and TNM stage in GSE72094 cohort.

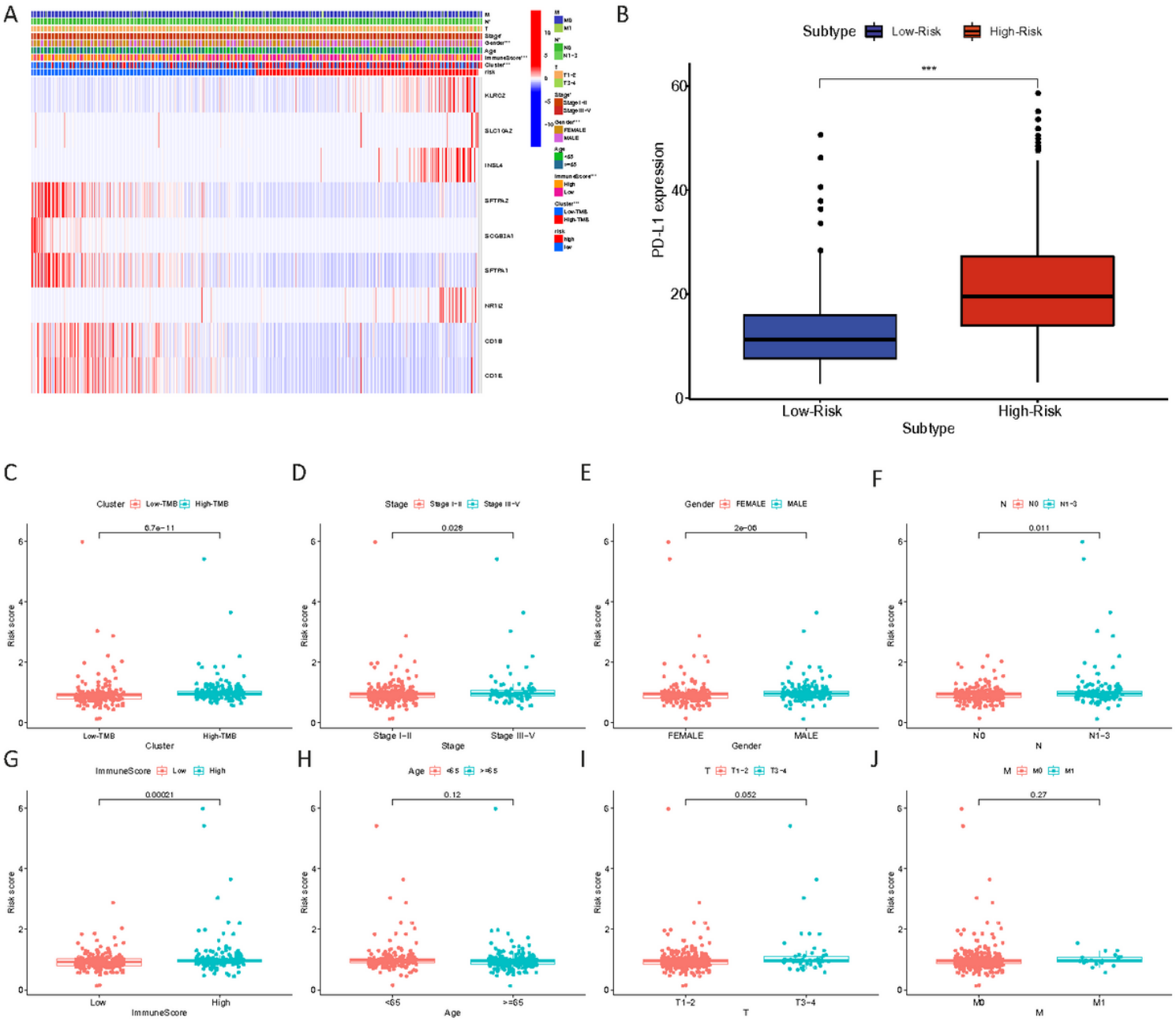

Figure 7

(A) The heatmap showed the TNM stage, immune scores, TMB subgroups, gender, and N stage were significantly associated with risk scores, and the expression of the nine genes were difference in high/low-risk group; (B) The expression of PD-L1 in high-risk group (red) were significantly higher than that in low-risk group (blue); (C-J) The Spearman analysis of the clinicopathological features, immune scores, and TMB subgroups with the risk scores. *** means P < 0.001, ** means P < 0.01, * means P < 0.05. TMB, tumor mutation burden.

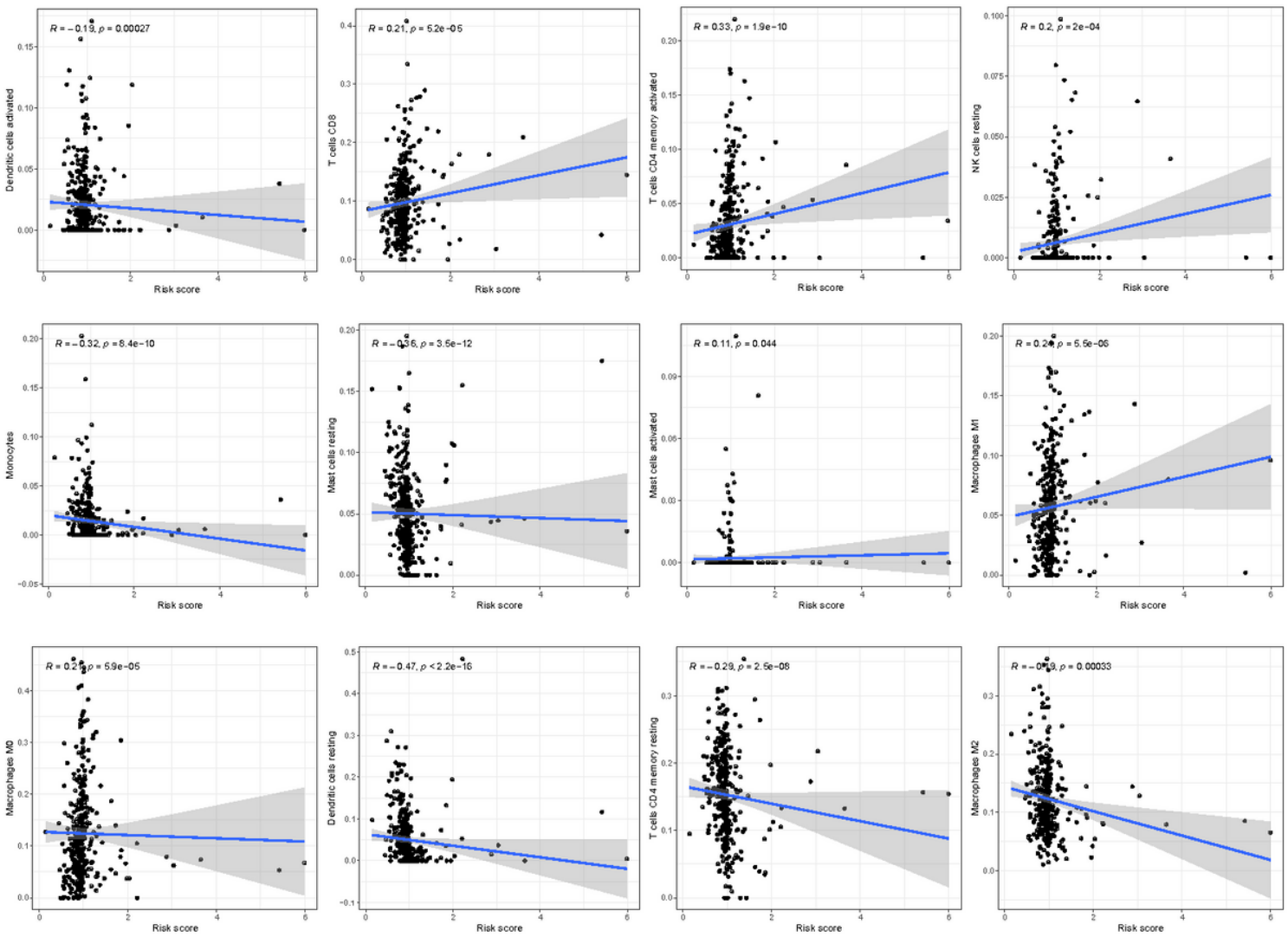

Figure 8

Correlation between the 9 TMB-related immune prognostic signatures for LUAD and the infiltration of immune cell subtypes. TMB, tumor mutation burden.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.xls
- SupplementaryTable2.xls
- SupplementaryTable3.xls
- SupplementaryTable4.xls
- SupplementaryTable5.xls
- SupplementaryTable6.xls

- SupplementaryTable7.xls