Development of an immune-related gene signature applying Ridge method for improving immunotherapy responses and clinical outcomes in lung adenocarcinoma

Zhen Chen[1] and Yongjun Zhang[2]

[1] Department of Cardiothoracic Surgery, The First College of Clinical Medical Science, China Three Gorges University, Yichang, China
[2] Department of Cardiothoracic Surgery, Xiangyang Central Hospital, Xiangyang, China

ABSTRACT

Background: Lung adenocarcinoma (LUAD) is a major cause of cancer mortality. Considering the critical role of tumor infiltrating lymphocytes in effective immunotherapy, this study was designed to screen molecular markers related to tumor infiltrating cells in LUAD, aiming to improve immunotherapy response during LUAD therapy.

Methods: The ConsensusClusterPlus method was used for clustering immune molecular subtypes of LUAD. Immune cell infiltration and immunotherapeutic potential in each subtype was evaluated employing single-sample gene set enrichment analysis (ssGSEA), Tumor Immune Dysfunction and Exclusion (TIDE), and Immunophenoscore (IPS). Immune-related co-expression modules were classified by weighted gene co-expression network analysis (WGCNA) analysis. The sequencing data of immune-related genes were comprehensively analyzed by introducing a new computational framework and 10 machine learning algorithms (a total of 101 combinations) to determine the prognostic genes, which were further combined to develop an immune prognostic signature (IMMPS) using the stepCox and Ridge methods. The expression of the signature genes was validated by quantitative real-time PCR (qRT-PCR).

Results: Samples from The Cancer Genome Atlas dataset (TCGA-LUAD) were divided into two subtypes (immunosuppressive subgroup C1 and immune-activated subgroup C2); notably, the C2 subgroup was more likely to benefit from immunotherapy ($p < 0.05$). An IMMPS developed based on seven immune infiltrating cell-related genes (*SEMA7A, EFHD2, CHST11, SLC24A4, MAL, JCHAIN*, and *SCARF1*) could accurately predict the overall survival of LUAD in five LUAD cohorts, with an average C-index higher than 0.69. LUAD patients with a low IMMPS value had a higher immune cell infiltration ($p < 0.05$). In addition, the IMMPS exhibited better prediction performance in comparison to 154 published gene signatures, suggesting that the IMMPS was an independent prognostic risk factor for evaluating the overall survival of LUAD patients. Since *BTNL9* was the most relevant immune checkpoint gene, *in vitro* experiment showed that the expression of the seven key genes (*SEMA7A, EFHD2, CHST11, SLC24A4, MAL, JCHAIN*, and *SCARF1*) in LUAD cell lines was consistent with that in normal lung epithelial cells after inhibiting *BTNL9* expression ($p < 0.05$).

Submitted 13 November 2024
Accepted 17 February 2025
Published 8 May 2025

Corresponding author
Yongjun Zhang, hybswl@163.com

Academic editor
Fanglin Guan

Additional Information and Declarations can be found on page 21

DOI **10.7717/peerj.19121**

Conclusions: Our results contributed to a better understanding of immunological characteristics of LUAD. The IMMPS could serve as a promising tool for improving the clinical outcome of patients suffering from LUAD.

Subjects Cell Biology, Computational Biology, Cardiology, Immunology, Oncology
Keywords Lung adenocarcinoma, Immunotherapy, Molecular biology, Immunology, Computational biology and bioinformatics

INTRODUCTION

Lung cancer is one of the most lethal and widely diagnosed cancers in the world (*Siegel, Miller & Jemal, 2019*; *Ding, Lv & Hua, 2022*; *Rice et al., 2016*). Non-small cell lung cancer (NSCLC), which accounts for approximately 80% of all the cancer cases in the lung, could be mainly divided into lung squamous cell carcinoma (LUSC) and lung adenocarcinoma (LUAD) (*Qian et al., 2023*; *Guo et al., 2021*). Noticeably, the two types of NSCLC are increasingly regarded as separate clinical entities since they share distinctively different molecular features and prognosis (*Tian, 2017*). As indicated by previously published studies, the tumor microenvironment (TME) of LUAD is enriched with different types of immune cells related to clinical outcomes (*Dai et al., 2019*; *Zheng, Hu & Yao, 2017*). Immunotherapies based on immune checkpoint inhibitors (ICIs) for treating lung cancer has been proven to be effective (*Chae et al., 2018*; *Zhou & Gao, 2022*). Several clinical trials of neoadjuvant ICI therapy also demonstrated their efficacy in resectable lung cancer (*Gao et al., 2020*; *Forde et al., 2018*). Additionally, a variety of ICIs, including atezolizumab targeting PD-L1 and navulizumab targeting PD-1 (*Malhotra, Jabbour & Aisner, 2017*), have been approved by the Food and Drug Administration (FDA) as second-line therapies for NSCLC treatment. However, influenced by various biological and molecular characteristics of different NSCLC subtypes, only around 15% of NSCLC patients can benefit from taking ICI, whereas many NSCLC patients have poor clinical outcomes (*Schoenfeld & Hellmann, 2020*). Currently, we face a lack of accurately prognostic biomarkers for NSCLC, especially prognostic markers related to immune. Hence, analysis of the TME in LUAD to improve immunotherapy strategies has great clinical significance.

Immune infiltration in TME has been widely found to be related to cancer prognosis, including LUAD (*Dickerson et al., 2023*; *Okcu et al., 2023*; *Tang et al., 2023*). For instance, counts of two lymphocyte populations (CD8/CD45RO, CD3/CD45RO, and CD3/CD8) in the tumor core (CT) and invasive margins (IM) can serve as prognostic markers for colorectal cancer (*Pagès et al., 2009*). An internationally recognized approach for the risk evaluation of colon cancer is to determine the immune score for patients using the four density percentiles of $CD3^+$ and $CD8^+$ T cells in CT and IM (*Pagès et al., 2018*). These findings pointed to the potential to facilitate clinical decision-making process based on cancer immune infiltration (*Fridman et al., 2012*). Currently, characterization of molecular analyses have been employed to evaluate immune infiltration in cancer patients, such as ESTIMATE (*Yoshihara et al., 2013*), CIBERSORT (*Chen et al., 2018*), and TIMER analyses (*Li et al., 2020*), laying a solid foundation for further study of clinical cancer features and immune infiltration. TME subtypes have been previously classified (*Bagaev et al., 2021*) by

transcriptome analysis, and four unique TME subtypes conserved in 20 different tumors were discovered. TME subtypes are also connected with patients' response to immunotherapy in a range of cancers. Recently, potential immunotherapy benefit for colon cancer patients in different subtypes has been successfully predicted utilizing immune-infiltrating cell scores, and accordingly a consensus immune-associated lncRNA signature for predicting the clinical outcomes of colon cancer was developed (*Liu et al., 2022*). The above findings supported the efficacy of developing a TME-based molecular signature to optimize precision therapies for cancer patients.

Tumor heterogeneity with differences ranging from genotype to phenotype between individual patients is an important feature of cancers. Ideally, biomarkers are expected to reflect a broad spectrum of gene expressions in tumor tissues, which therefore require a combination of multiple genes to address the problem of heterogeneity (*Koncina et al., 2020*). A variety of prognostic signatures have been developed applying high-throughput data and bioinformatics algorithms (*Zhao et al., 2023*; *Scortegagna et al., 2023*; *Pan et al., 2022*). In the last 5 years, more than 154 prognostic gene models have been created and validated as candidate biomarkers for LUAD. However, limitations such as a lack of rigorous validation in different cohorts or inappropriate modelling also hinder the application of these features in clinical setting.

The current work developed a novel computational framework to determine tumor infiltration in LUAD and to discover immune-correlated genes by conducting comprehensive analysis based on immune-infiltrating cells in the TME. The potential importance of these genes as predictive biomarkers for LUAD prognosis and immunotherapy was also analyzed.

MATERIALS AND METHODS

The acquisition of LUAD patient datasets

The transcriptional profiles of patients with LUAD and their clinical information were collected from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov) and Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) under the following criteria: (1) Complete data on overall survival (OS) and survival status, stage, gender, age; (2) IlluminaHiSeq platform or Affymetrix HG-U133_Plus 2.0 platform; and (3) Sample size larger than 200. After filtering, 1,327 LUAD patients from five datasets were included in this study. Specifically, GSE30219 dataset contained 83 patients (*Rousseaux et al., 2013*); GSE31210 dataset contained 226 patients (*Okayama et al., 2012*); GSE50081 dataset contained 127 patients from (*Der et al., 2014*); GSE72094 dataset contained 398 patients (*Schabath et al., 2016*); 493 patients were collected from TCGA-LUAD (https://xenabrowser.net/datapages/?cohort=GDC%20TCGA%20Lung%20Adenocarcinoma%20(LUAD)&removeHub=https%3A%2F%2Fxena.treehouse.gi.ucsc.edu%3A443). The TCGA-LUAD dataset served as a training dataset for analyzing the features of immune genes, while the other four datasets were independent test datasets for validation. Detailed clinical information of the patients from these five groups was shown in Table 1.

Gene expression profiling data processing

Background correction, quantile normalization, and log-2 transformation of GEO microarray datasets were processed in the R package oligo (*Carvalho & Irizarry, 2010*), and the Robust Multi-array Average algorithm was used for normalizing all the microarray datasets. According to the annotation files of the microarrays, the median value of multiple probes mapping to the same gene was taken, while a probe was deleted when it was mapped to multiple genes. Subsequently, transcripts per kilobase million (TPM) was converted from the RNA-seq read count in TCGA-LUAD.

Assessment of immune infiltrating cells

Marker genes for 28 types of highly reliable immune-infiltrating cells were collected from a previous study (*Charoentong et al., 2017*). Single-sample gene set enrichment analysis (ssGSEA) in the R package GSVA (*Hänzelmann, Castelo & Guinney, 2013*) was conducted for calculating the proportion of immune-infiltrating cells in each patient, followed by verifying the results by ESTIMATE (*Yoshihara et al., 2013*), CIBERPSORT (*Newman et al., 2015*), TIMER (*Li et al., 2016*), EPIC (https://gfellerlab.shinyapps.io/EPIC_1-1/) (*Racle et al., 2017*), and MCP-counter (*Becht et al., 2016*) algorithms.

Classification of immune-related molecular subtypes

Patients in the TCGA-LUAD cohort were classified applying a clustering method of unsupervised resampling based on the components of 28 types of immune infiltrating cells. The R package ConsensusClusterPlus (*Wilkerson & Hayes, 2010*) was used to calculate immune infiltrating cell scores for individual samples as a data matrix. Specifically, 80% of the samples were randomly selected for each iteration, and the distance between samples was measured by Pearson correlation coefficient. This was process was repeated 1,000 times, where samples were partitioned into up to k (maximum k = 10) clusters using the partioning around medoids (PAM) algorithm. Subsequently, the optimal number of clusters was determined according to consensus score matrix, cumulative distribution function (CDF) curve and proportion of ambiguous clustering (*Pagès et al., 2018*) score. The significance of the clustering results was assessed using sigclust (*Harrison et al., 2003*).

Screening tumor-infiltrating immune-related gene markers

A novel computational framework for a comprehensive analysis of immune-infiltrating cells in the TME was created to identify immune genes related to tumor infiltration. Multiple machine learning methods were used to establish an immunity prognostic signature (IMMPS). The median absolute deviation (MAD) (*Okayama et al., 2012*) of all genome-wide protein-coding genes was computed, and potentially deregulated genes were defined as having a MAD greater than top 50%. Weighted gene co-expression network analysis (WGCNA) (*Langfelder & Horvath, 2008*) was constructed and the appropriate soft threshold β was calculated to satisfy a scale-free network. Additionally, the weighted adjacency matrix was converted to topological overlap matrix (TOM) and modules were identified by the dynamic tree cutting method. Next, immune-related modules were recognized based on the correlation of each module with immune clusters. The

Table 1 Information on clinicopathologic characteristics of patients in five datasets (GSE30219, GSE31210, GSE50081, GSE72094, and TCGA-LUAD).

Variable	GSE30219, *N* = 83[1]	GSE31210, *N* = 226[1]	GSE50081, *N* = 127[1]	GSE72094, *N* = 398[1]	TCGA, *N* = 493[1]
Age	60 (55, 69)	61 (55, 65)	70 (63, 76)	70 (64, 76)	66 (59, 72)
Gender					
Female	18 (22%)	121 (54%)	62 (49%)	222 (56%)	264 (54%)
Male	65 (78%)	105 (46%)	65 (51%)	176 (44%)	229 (46%)
Smoker					
Current	0 (NA%)	0 (0%)	36 (31%)	0 (0%)	116 (24%)
Ever	0 (NA%)	111 (49%)	56 (49%)	300 (91%)	294 (61%)
Never	0 (NA%)	115 (51%)	23 (20%)	31 (9.4%)	69 (14%)
Unknown	83	0	12	67	14
T Stage					
T1	69 (83%)	0 (NA%)	43 (34%)	0 (NA%)	164 (33%)
T2	12 (14%)	0 (NA%)	82 (65%)	0 (NA%)	265 (54%)
T3	2 (2.4%)	0 (NA%)	2 (1.6%)	0 (NA%)	43 (8.8%)
T4	0 (0%)	0 (NA%)	0 (0%)	0 (NA%)	18 (3.7%)
Unknown	0	226	0	398	3
N Stage					
N0	80 (96%)	0 (NA%)	94 (74%)	0 (NA%)	319 (66%)
N1	3 (3.6%)	0 (NA%)	33 (26%)	0 (NA%)	93 (19%)
N2	0 (0%)	0 (NA%)	0 (0%)	0 (NA%)	68 (14%)
N3	0 (0%)	0 (NA%)	0 (0%)	0 (NA%)	2 (0.4%)
Unknown	0	226	0	398	11
M Stage					
M0	83 (100%)	0 (NA%)	127 (100%)	0 (NA%)	327 (93%)
M1	0 (0%)	0 (NA%)	0 (0%)	0 (NA%)	24 (6.8%)
Unknown	0	226	0	398	142
Stage					
I	0 (0%)	0 (0%)	0 (0%)	0 (0%)	265 (55%)
II	0 (0%)	0 (0%)	0 (0%)	0 (0%)	116 (24%)
III	0 (0%)	0 (0%)	0 (0%)	0 (0%)	79 (16%)
IV	0 (0%)	0 (0%)	0 (0%)	0 (0%)	25 (5.2%)
Stage I	69 (83%)	168 (74%)	92 (72%)	254 (65%)	0 (0%)
Stage II	13 (16%)	58 (26%)	35 (28%)	67 (17%)	0 (0%)
Stage III	1 (1.2%)	0 (0%)	0 (0%)	57 (15%)	0 (0%)
Stage IV	0 (0%)	0 (0%)	0 (0%)	15 (3.8%)	0 (0%)
Unknown	0	0	0	5	8
Status					
Alive	40 (48%)	191 (85%)	76 (60%)	285 (72%)	315 (64%)
Dead	43 (52%)	35 (15%)	51 (40%)	113 (28%)	178 (36%)
OS.time	2,070 (855, 3,435)	1,744 (1,246, 2,050)	1,595 (697, 2,097)	824 (540, 1,012)	669 (434, 1,148)

Note:
[1] Median (IQR) or Frequency (%).

relationship between the prognosis of LUAD and genes in the immune-related modules was analyzed by univariate Cox regression analysis to filter significantly prognostically relevant genes as a hub gene set.

In addition, a total of 101 combinations of 10 machine learning algorithms, including Least absolute shrinkage and selection operator (LASSO), CoxBoost, Ridge, generalized boosted regression modeling (GBM), partial least squares regression for Cox (plsRcox), supervised principal components (SuperPC), elastic network (Enet), stepwise Cox, random survival forest (RSF), survival support vector machine (survival-SVM) together with 10-fold cross-validation were introduced to select candidate genes with the highest C-index for IMMPS. The stepCox (both) and Ridge were used to develop the IMMPS model. Specifically, the former was applied to filter the most valuable TIICGs, while the later was used to fit the most reliable model. After the Ridge regression, the IMMPS formula for each variable was as follow:

$$IMMPS = \sum_{k=1}^{n} x^k a^k$$

where n denoted the number of feature genes, a^k denoted the ridge regression coefficient of the k^{th} feature gene, and x^k denoted the expression level of the k^{th} feature gene.

Enrichment analysis

Gene oncology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of specific gene modules were performed using the R package "clusterProfiler" (*Yu et al., 2012*).

Immunotherapy response prediction

GSE126044 (*Cho et al., 2020*) and GSE135222 (*Kim, Choi & Jung, 2020*) were two cohorts containing follow-up information of each NSCLC patient's response to anti-PD-1/PD-L1 therapy. Subclass mapping was conducted using the SubMap (*Hoshida et al., 2007*) algorithm to evaluate the response of immune molecular subtypes to immunotherapy, and potential treatment benefit was determined by Tumor Immune Dysfunction and Exclusion (TIDE) algorithm.

Cell culture and transfection

Human normal lung epithelial cells BEAS-2B (BNCC359274) and human LUAD cells lines PC-9 (BNCC340767) and H1395 (BNCC100270) were purchased from Bena Biotechnology Co. (Beijing, China). Penicillin/streptomycin and Dulbecco's modified Eagle's medium (A1896701; DMEM, Gibco, Grand Island, NY, USA) with 10% fetal bovine serum (FBS, Gibco, USA) was used for cell culturing at 37 °C in 5% CO_2.

BTNL9 plays a critical role in immune regulation, especially in TME that may affect the activity of immune infiltrating cells (*Zhang et al., 2023*). In this study, to investigate the regulatory effects of *BTNL9* on the expression of the seven immune-related genes, human LUAD cell lines (PC9 and H1395) and normal lung epithelial cells (BEAS-2B) were

Table 2 Sequence of primers.

Gene	Forward primer sequence (5′-3′)	Reverse primer sequence (5′-3′)
SEMA7A	TTCAGCCCGGACGAGAACT	GAACCGAGGGATCTTCCCAT
EFHD2	GGATGAGGACTTTGACAGCAAGC	TTGACACCCTCACTGGAGACGT
CHST11	CACAAGCCGTAAGCGGAGG	CATGGGGTCGCTGTACTTCC
SLC24A4	ATGGCCCCAGTGAATGGGA	CCAGCCACATCTTCGCTCAG
MAL	ACCGCTGCCCTCTTTTACC	GAAGCCGTCTTGCATCGTGAT
JCHAIN	TCCTGGCGGTTTTTATTAAGGC	AGTAATCCGGGCACACTTACAT
SCARF1	CCGATCAGACCTCAAGGACAG	CCCAGGGTAGCTTGTGGGA
BTNL9	GGACCTGTTCAGTCTGGAAAC	TCTGGACCACCAACTCTTTCT
GAPDH	CTGGGCTACACTGAGCACC	AAGTGGTCGTTGAGGGCAATG

transfected with *BTNL9*-specific siRNA (5′-3′ sequence: GCCTCTAACTCCACAACAA CACT) using Lipofectamine 3000 (Thermo Fisher, Waltham, MA, USA) following the manufacturer's protocol. After 48 h, the transfection efficiency was tested by qRT-PCR.

Western blot testing

Cells were lysed using RIPA buffer supplemented with protease inhibitors. Protein samples were separated on 15% SDS-PAGE gel, subsequently transferred onto PVDF membranes (Beyotime, Shanghai, China) and blocked. The membranes were incubated overnight at 4 °C with primary antibodies against *BTNL9* (1:1,000) and GAPDH (1:1,000), and then treated with HRP-conjugated secondary antibodies for 1 h at room temperature. Finally, the protein bands were visualized utilizing the ECL system (Amersham Biosciences, Inc, Buckinghamshire, UK) and the band intensities were measured using Quantity One software (BioRad, Hercules, CA, USA).

Quantitative real-time PCR (qRT-PCR)

Total RNA was extracted from BEAS-2B, PC9 and H1395 (Thermo Fisher, Waltham, MA, USA) using TRIzol reagent. The HiScript II SuperMix (Vazyme, China) was employed for isolating cDNA from 500 ng of RNA. QRT-PCR was carried out with the use of the SYBR Green Master Mix in ABI 7500 System (Thermo Fisher Scientific, Waltham, MA, USA). PCR amplification began with 45 cycles a 94 °C for 10 min, at 94 °C for 10 seconds (s), and 60 °C for 45 s. GAPDH was an internal reference. All the primer sequences were shown in Table 2.

Statistical analyses

Heatmap was generated using the R package ComplexHeatmap (*Gu, Eils & Schlesner, 2016*). Wilcox test and Kruskal-Wallis test were employed to analyze two-group differences and comparisons between multiple groups, respectively. The correlation between two groups of samples was calculated by Pearson's correlation test. Prognostic differences among patients were assessed using Kaplan-Meier. The R package "timeROC"

was used to plot time-dependent receiver operating characteristic (ROC) curves (*Blanche, Dartigues & Jacqmin-Gadda, 2013*). All the statistical analyses were conducted in R 3.6.4 (*R Core Team, 2021*). GraphPad Prism V.8.0.2 was applied for the statistical analysis on the experimental data, which were compared using one-way analysis of variance or unpaired *t*-test. A $p < 0.05$ indicated a statistical significance.

RESULTS

Development and validation of immune infiltration molecular subtypes

Charoentong et al. (2017), *Bindea et al. (2013)* revealed that at least 28 different types of immune cell types are crucially involved in tumor infiltration. Immune cell infiltration in the TME of LUAD were analyzed by single-sample gene set enrichment analysis (ssGSEA). Patients were classified by consensus clustering. CDF curves of the consensus score matrix (Fig. 1A) and PAC results (Fig. 1B) showed that the optimal number of clusters was when k = 2 (Fig. 1C). The sigclust analysis detected significant differences (p = 1.141007e−05) between the two consensus clusters of C1 and C2 in terms of immune infiltration, with C1 showing a notably lower overall infiltration than C2 (Fig. 1D, Fig. S2A). In addition, ESTIMATE, CIBERPSORT, TIMER, EPIC, and MCP-counter results showed similar results (Fig. 1E, Figs. S2B–S2F). Hence, C1 and C2 were accordingly defined as an immunosuppressed and immune-activated subtypes. Survival analyses all demonstrated a significantly worse prognosis for C1 than C2 (Fig. 1F).

The expression patterns of immune checkpoint genes in immune-activated subtype and potential benefits from immunotherapy

After literature review, a total of 79 immune checkpoint genes (ICGs) (Table S1) mainly involved in ligand-receptor interactions were selected. These ICGs affect immune activity in different ways, including stimulation, inhibition or a combination of both. Analysis of expression patterns of these genes in the C1 and C2 subtypes demonstrated that almost all the ICGs, such as PDCD1, CD274 and CTLA4 (Fig. 2B), were upregulated in C2 (Fig. 2A). TIDE model could evaluate T cell exclusion and dysfunction signatures across 33,000 samples in 188 tumor cohorts stored in public databases, including TCGA (*Weinstein et al., 2013*), METABRIC (*Curtis et al., 2012*), and PRECOG (*Gentles et al., 2015*). Applying TIDE model, it was observed that the C1 subtype had significantly higher T cell exclusion and the C2 subtype had significantly higher T cell dysfunction, and that C2 subtype might benefit from taking immunotherapy based on combined TIDE scores (Fig. 2C). Recent studies have confirmed immunophenoscore (IPS) as an accurate predictor of ICI response, with a higher IPS predicting greater sensitivity to ICI treatment. Here, we found a significantly higher IPS score of C2 than C1 (Fig. 2D, p = 0.045), which was also validated by the cytotoxic T lymphocyte score (Fig. 2E). Further, the gene expression patterns of immune phenotypes of NSCLC patients who received ICIs were compared by SubMap analysis. Notably, the gene expression profiles of lung cancer patients who responded to anti-PD-1 immunotherapy were highly similar to those with immune-activated subtypes in both the training and validation cohorts (Fig. 2F). These results indicated that patients

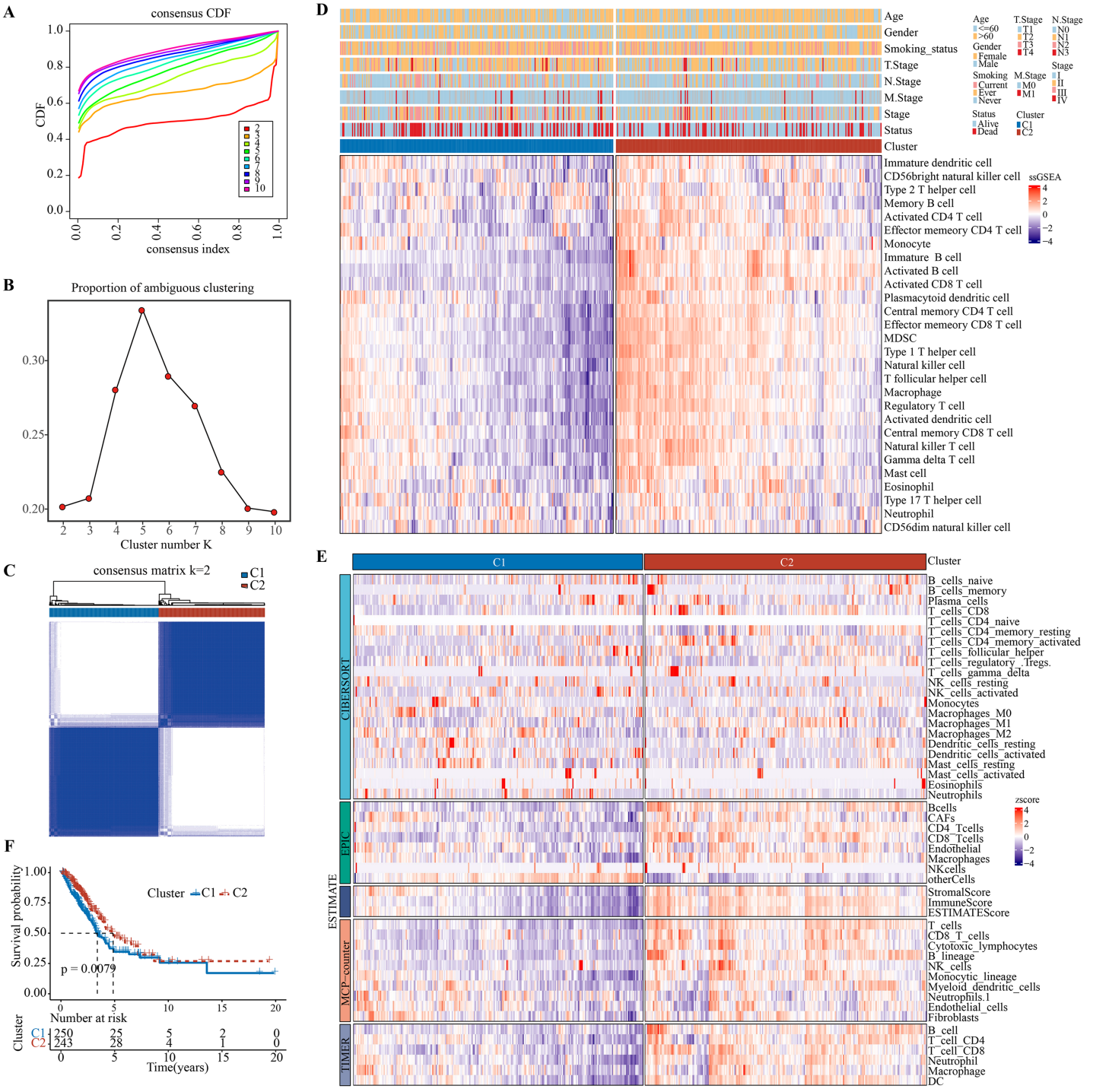

Figure 1 **Development and validation of immune infiltration consensus clusters.** (A) The CDF curves of consensus matrix for each k (indicated by colors). (B) The proportion of ambiguous clustering score, with a low PAC value indicating a flat middle segment. The optimal k (k = 2) was determined according to the lowest value of PAC. (C) When k = 2, consensus score matrix for all samples, with a higher consensus score between two samples indicating a higher chance of the two being grouped into the same cluster in different iterations. Each cell in the figure represents the consistency score between the two samples. Blue color indicates high consistency (strong similarity); white color indicates low consistency (weak similarity). (D) The infiltration abundance of 28 immune cell subsets was determined by ssGSEA for two clusters. (E) The stability and robustness of the ssGSEA results were further confirmed by the algorithms of ESTIMATE, CIBERPSORT, TIMER, MCP-counter, and EPIC. (F) Kaplan–Meier curves of OS in the C1-*vs*-C2 in TCGA-LUAD (log-rank test: $P = 0.0079$). Full-size DOI: 10.7717/peerj.19121/fig-1

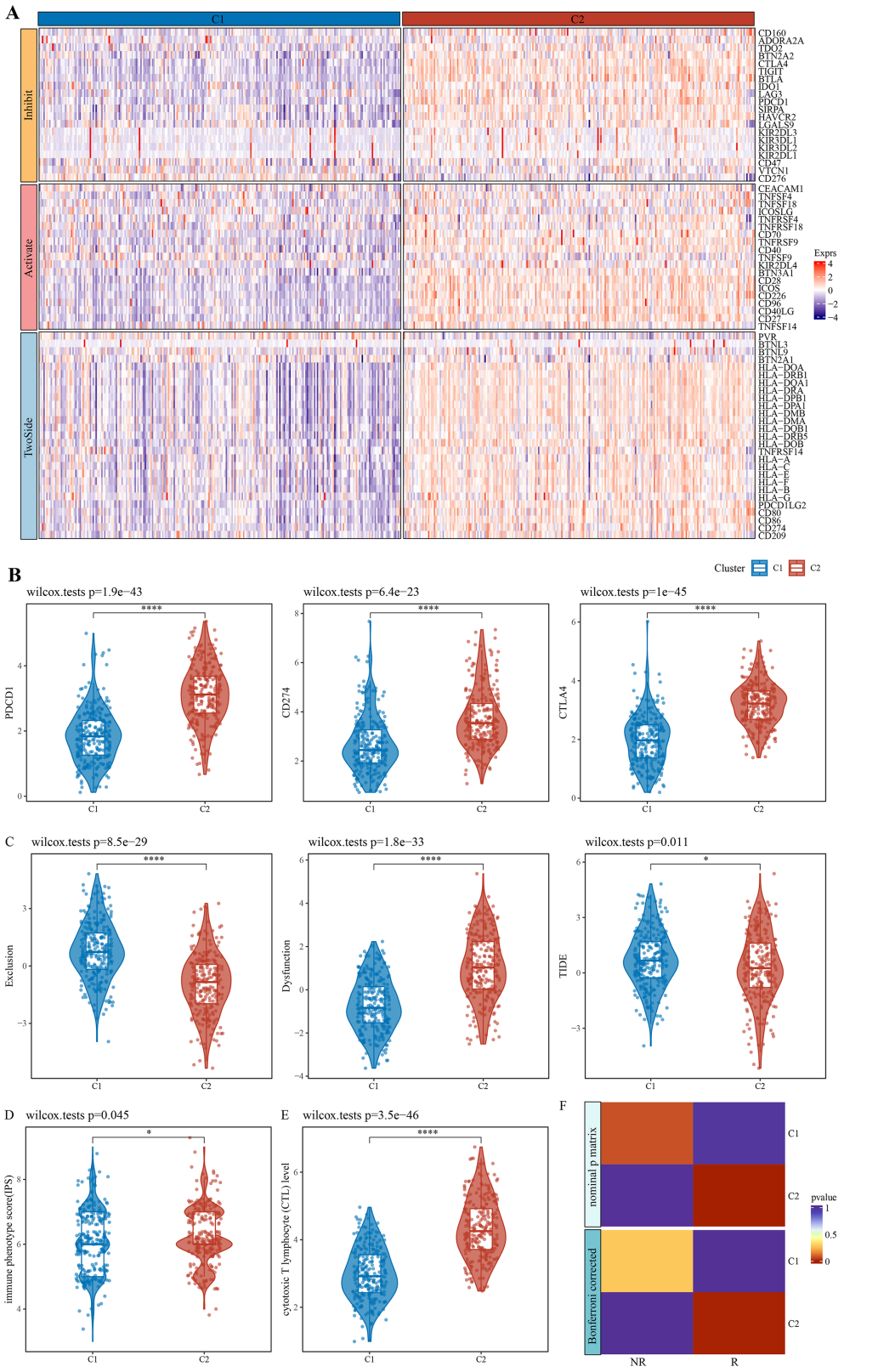

Figure 2 Expression pattern of immune checkpoint genes in immune-activated subtypes and potential benefits of immunotherapy. (A) Expression pattern of ICG in patients with two immune subtypes. (B) Expression distribution of classical immune checkpoint genes PDCD1, CD274, and CTLA4

Figure 2 (continued)
in the two immune subtypes. (C) Distribution of T cell exclusion and dysfunction in the two immune subtypes analyzed by TIDE. (D) Distribution differences of immune phenotype score in the two immune subtypes. (E) Distribution differences of the two immune subtypes at the cytotoxic T lymphocyte (CTL) level. (F) Submap analyses show that patients in the immune activation subtypes are similar to those who responded to anti- PD-1 treatment. * $p < 0.05$; **** $p < 0.0001$.
Full-size DOI: 10.7717/peerj.19121/fig-2

with immune-activated subtypes could benefit from receiving ICI immunotherapy, especially more from anti-PD-1 therapy.

Identification of immune infiltration-associated gene modules and prognostic features of the genes

The soft threshold β = 8 was selected (Figs. S2A, S2B) to ensure a scale-free nature of the network. A total of 30 gene modules were detected, as indicated by different colors, and the heatmap displayed the characteristic gene neighborhood of the modules (Fig. S2C). Additionally, the modules was weakly correlated with AJCC Stage, Age and Gender (Fig. 3A). Specifically, darkred, orangered4, and green modules were considered as the modules most closely correlated with C2 (immune activation) subgroup. Among the three modules, the correlation coefficient between module membership (MM) and gene significance (GS) of the darkred, orange4, and green modules reached 0.94, 0.94, and 0.71 (Figs. 3B–3D), respectively, indicating that the identification of these gene modules was reliable. Further functional enrichment analysis showed that the darkred module was enriched to pathways of human T−cell leukemia virus 1 infection, PD−L1 expression, PD−1 checkpoint pathway in cancer, *etc.* (Fig. 3E). The green module was significantly enriched to Cytokine-cytokine receptor interaction, B cell receptor signaling pathway, and other pathways (Fig. 3F). The orangered4 module was significantly enriched to primary immunodeficiency, MAPK signaling pathway and some other pathways (Fig. 3G). These pathways all play critical role in immune regulation. Noticeably, biological processes such as cytokine activity, cytokine receptor activity, cytokine binding, and cytokine receptor binding were enriched multiple times. Under stringent criteria, genes significantly associated with LUAD prognosis were identified by performing one-way survival analysis from at least four cohorts, and finally 26 genes were screened as potential prognostic markers with significant immune relevance in LUAD (Fig. S2E).

Development of an immune-correlated prognostic signature and assessment

Using machine learning-based integration analysis, an immune-related gene signature was developed based on the expression profiles of 26 immune-related genes. In the TCGA-LUAD dataset, 101 predictive models were developed by the global leave-one-out cross-validation (LOOCV) framework and their C-index in all the validation datasets was calculated. Interestingly, the optimal model was the combination of stepCox and Ridge, which showed the highest average C-index (0.697) in all the validation datasets (Fig. 4A). In step cox regression, when the Akaike Information Criterion (*Pagès et al., 2018*) reached

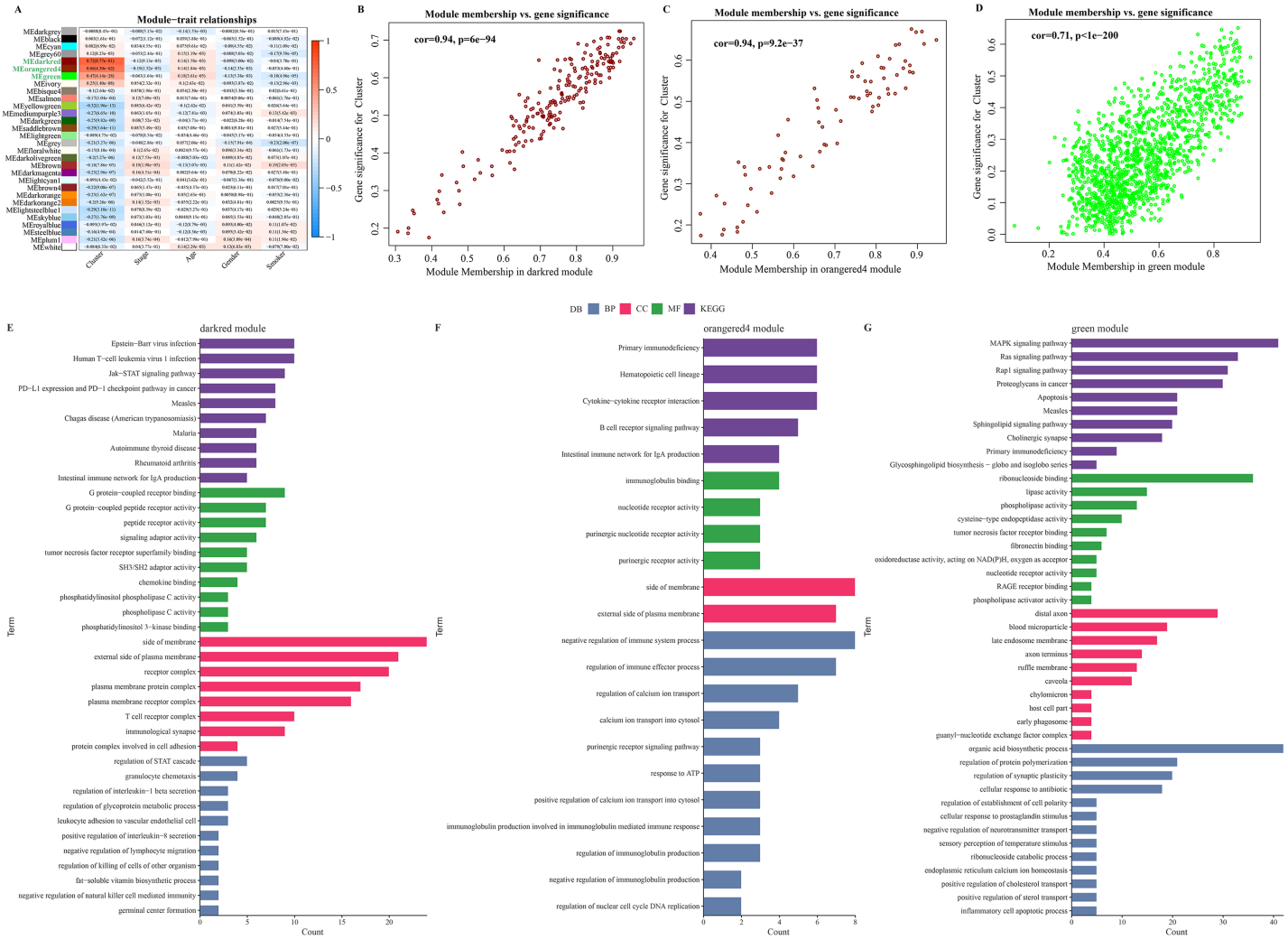

Figure 3 **Identification of immune infiltration-associated gene modules and prognostic features of the genes.** (A) Correlation analysis between gene modules and clinical features. (B–D). Close correlation between GS and MM in the darkred, orangered4 and green modules ($p < 0.0001$). Statistical tests: Pearson's correlation coefficient, two-sided unpaired t-test. (E–G). Enrichment results of genes in darkred, orangered4 and green module to KEGG Pathway, biological process (BP), cellular component (CC) and molecular function (MF).

Full-size DOI: 10.7717/peerj.19121/fig-3

its minimum value, the optimal combination of seven variable genes were identified. Therefore, the final model (IMMPS) was built with these seven genes applying regression analysis with Ridge: IMMPS = 0.093**SEMA7A* + 0.145**EFHD2* + 0.108**CHST11* − 0.502**SLC24A4* − 0.089**MAL* − 0.074**JCHAIN* − 0.128* *SCARF1*. Each patient was assigned with a risk score by the model and weighted for their regression coefficient. Based on the optimal cut-off value, all the patients were categorized into high- and low- risk groups (Figs. 4B–4F). It was observed that the OS of high-risk patients was significantly worse than the low-risk patients in the TCGA-LUAD training dataset and the five validation datasets (all $p < 0.05$). ROC analysis showed that the IMMPS reached 1-, 3-, and 5-year area under the curves (AUCs) of 0.69, 0.69, and 0.65 in TCGA-LUAD, respectively.

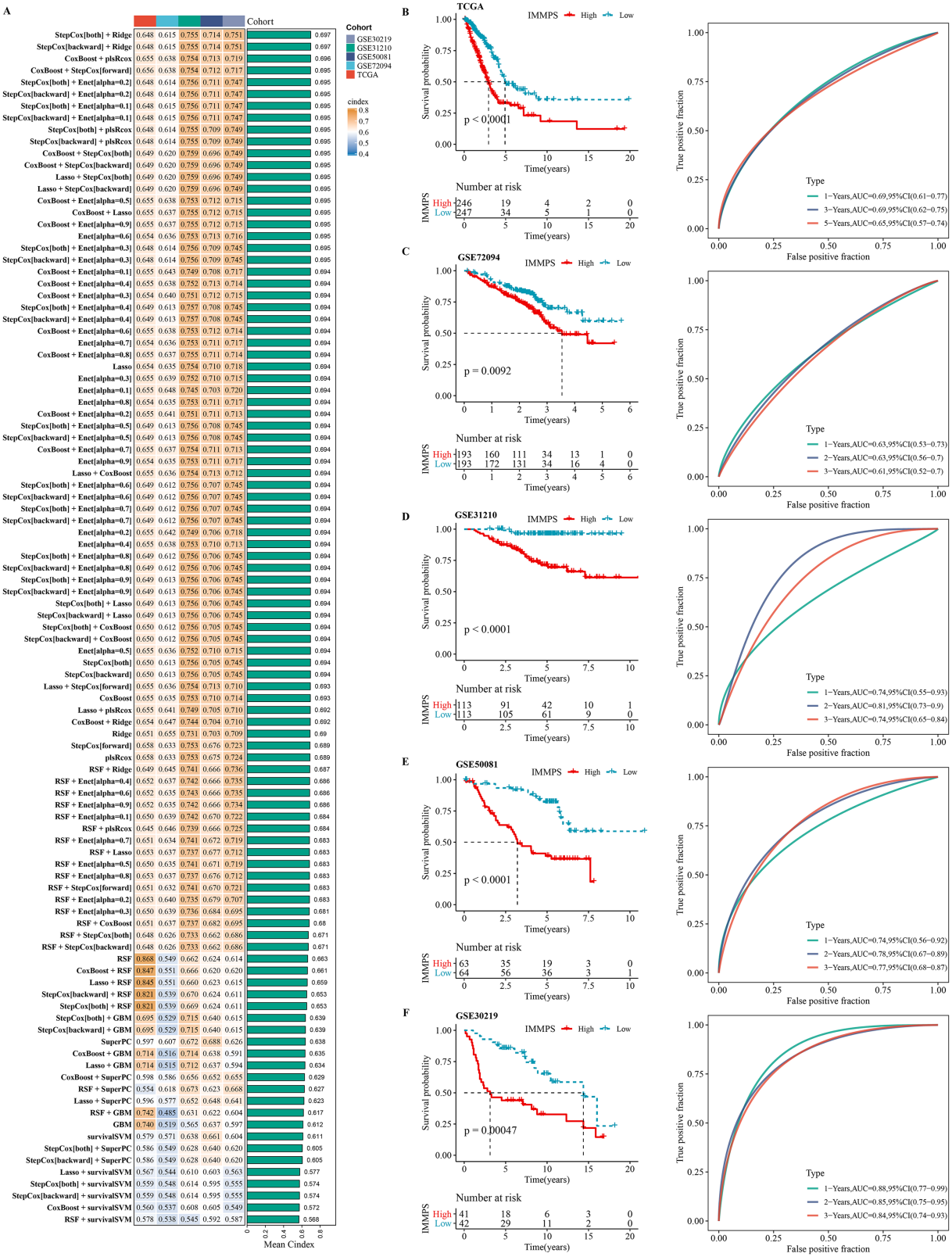

Figure 4 Construction and assessment of immune-correlated prognostic characteristics. (A) Heatmap of c-index distribution of 101 prediction models in five cohorts. (B–F) Kaplan–Meier curves and ROC of patients in low- and high-risk groups after risk assessment of patients using IMMPS in different cohorts.

Full-size DOI: 10.7717/peerj.19121/fig-4

In GSE72094, the AUC values were consistent at 0.63 for all three time points. For GSE31210, 1-, 3-, and 5-year AUCs were 0.74, 0.81, and 0.74, respectively, while in GSE50081, they were 0.74, 0.78, and 0.77, respectively. Finally, the GSE30219 cohort showed AUC values of 0.88, 0.85, and 0.84 at the corresponding time points. These data supported a stable and robust performance of IMMPS across different cohorts. Consistently, multivariate Cox regression analysis revealed that IMMPS was an independent risk factor for LUAD ($p < 0.001$, Fig. S3).

Comparison of gene expression-based prognostic features for LUAD

Next-generation sequencing and big data technology has promoted the prediction of genes by machine learning-based methods (*Ahluwalia, Kolhe & Gahlay, 2021*). This study comprehensively compared the performance of published signatures to the IMMPS. A severe lack of miRNA and lncRNA data in the validation dataset excluded the miRNA and lncRNA signatures, resulting in the collection of a total of 154 signatures (Table S2). These signatures were related to different biological processes, such as ferroptosis, WNT, stemness, autophagy, hypoxia, adipogenesis, glycolysis, epigenetics, immune response, vitamin D, aging, epithelial-mesenchymal transition, N6-methyladenosine, Toll-like receptor signaling, and drug sensitivity. The c-index of each signature in all the datasets were calculated to evaluate their predictive performance (Fig. 5). According to the c-index, ranking of the signatures in each dataset showed that one signature (0.6%) outperformed the IMMPS in three datasets. Conversely, 49 signatures (31.8%) exhibited a lower c-index than the IMMPS in three cohorts (60% of the total cohorts), 47 signatures (30.5%) showed a lower c-index than the IMMPS in four cohorts (80% of the total cohorts), and 57 signatures (37%) had a lower c-index than the IMMPS across all cohorts (Fig. S4). These results supported the stability of the IMMPS. We also noted that while most of the models performed well in their own training set and some specific external datasets, their performance was notably poor in other datasets. This discrepancy could be attributed to poor model generalization due to overfitting problem. Next, two machine learning algorithms were employed to reduce the gene numbers in our signature, which thereby manifested a stronger generalization ability. In comparison to the model developed by (*Deng et al., 2022*), the IMMPS contained only seven genes. Fewer genes with comparable predictive performance could significantly reduce the cost in clinical testing. These results confirmed that IMMPS had the advantage of higher stability but lower detection cost in comparison with previous gene signatures.

The impact of the IMMPS on ICI treatment

Comparison on the differences of the IMMPS in the two subtypes showed a significantly lower IMMPS in C2 patients than in C1 patients (Fig. 6A, $p = 0.00026$). The patients were then classified into two groups of IMMPS-High and IMMPS-Low using the IMMPS. It can be observed that immune infiltrating cell scores of patients with IMMPS-High were significantly lower (Fig. 6B). A total of 18 types of immune infiltrating cells (64.3%) were negatively correlated with the IMMPS. In particular, eosinophils had the strongest correlation with IMMPS (Fig. 6C), and patients with low IMMPS but more eosinophils

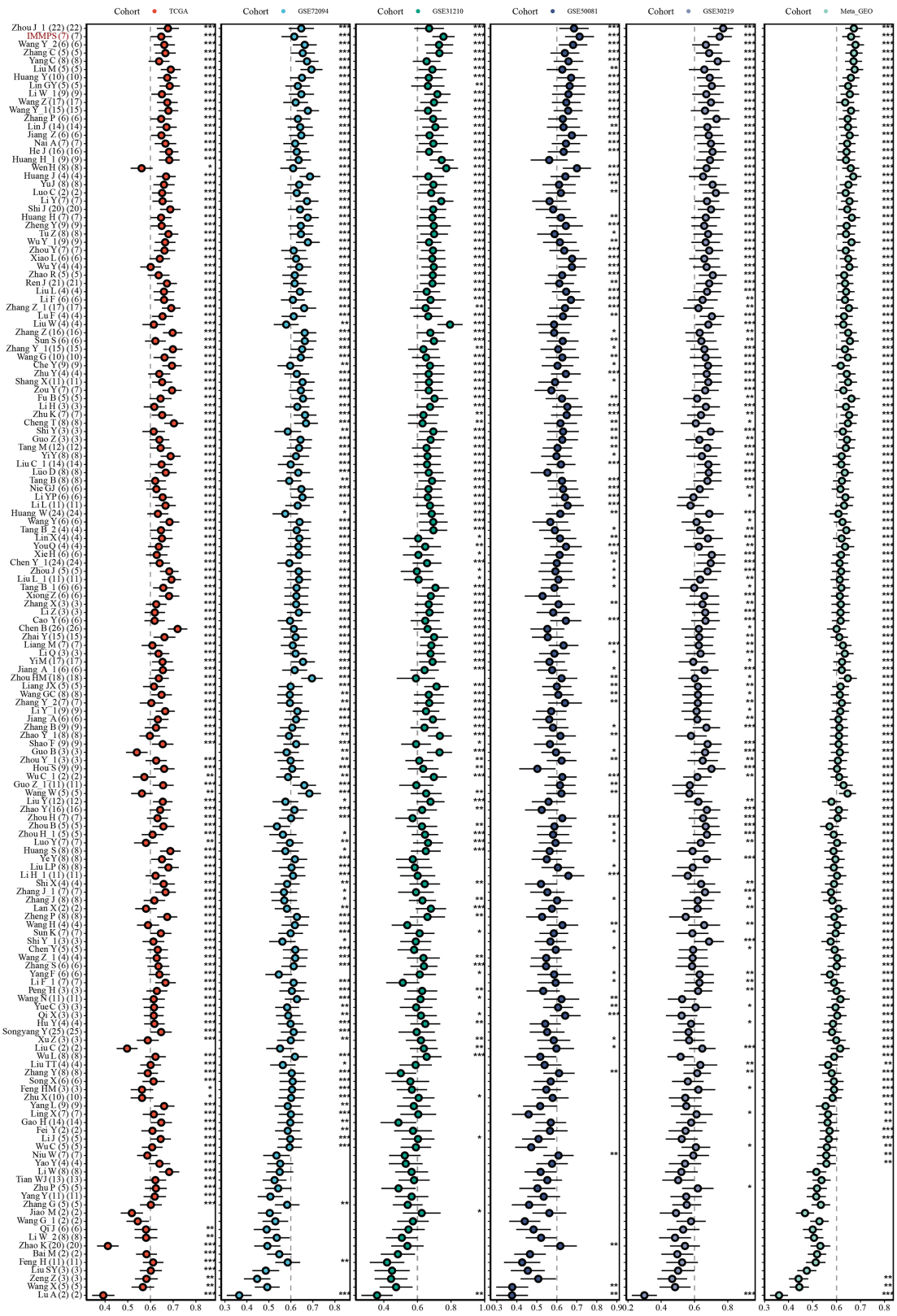

Figure 5 The c-index comparison of IMMPS with 154 gene signatures in five datasets. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; blank denotes $p > 0.05$. Each dot indicates the corresponding c-index. The horizontal lines near the dots indicates the confidence intervals. IMMPS and 154 signatures are shown vertically, and six data cohorts are shown horizontally. Full-size DOI: 10.7717/peerj.19121/fig-5

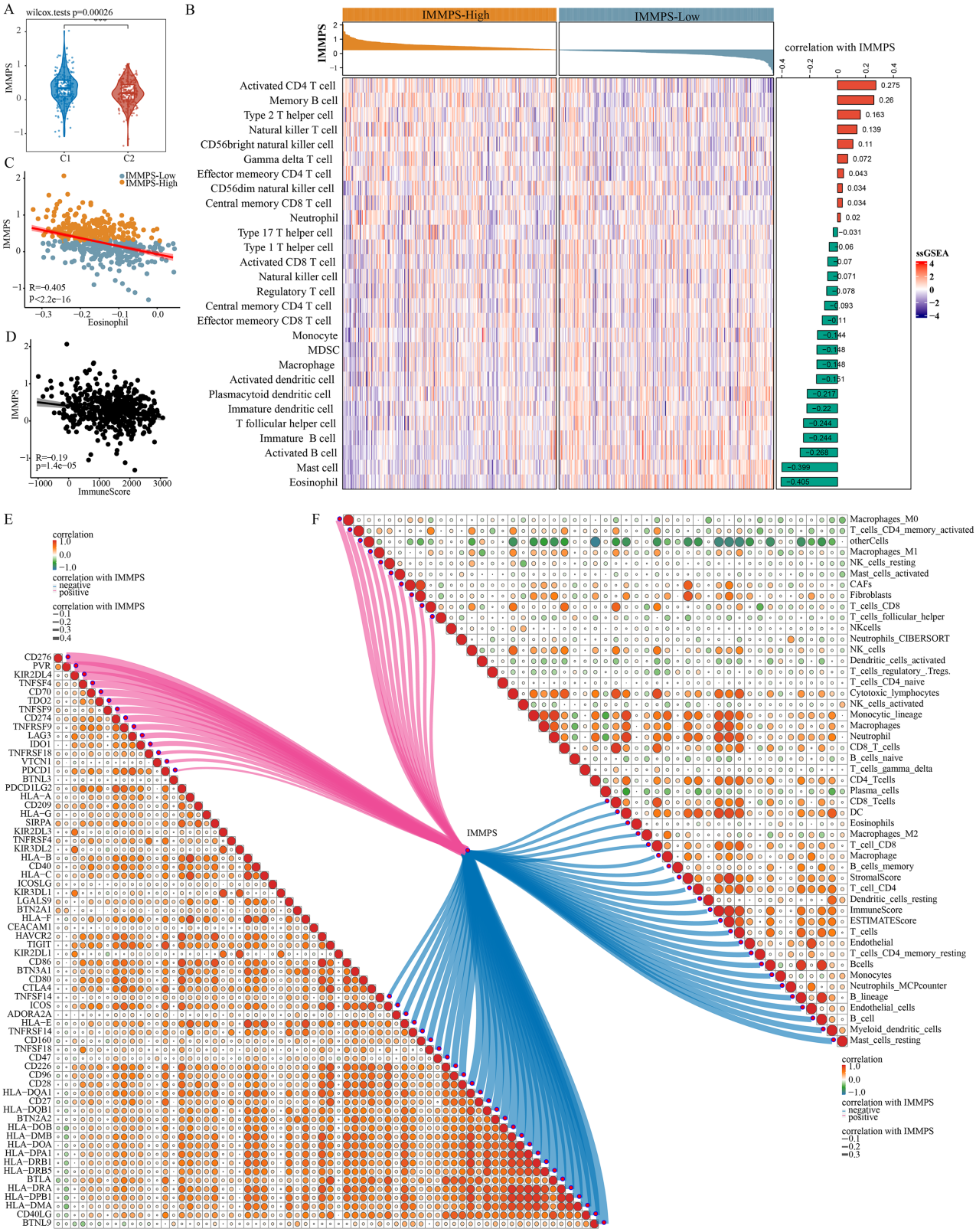

Figure 6 Impact of IMMPS on ICI treatment. (A) Differences in the distribution of IMMPS in the immunoactivated (C2) and immunosuppressed (C1) subtypes. (B) Differences in the distribution of 28 immune-infiltrating cells in patients with different IMMPS levels. The colors in the heatmap indicate the immune-cell infiltration scores after z-score. The right bar graph shows the correlation between IMMPS and the corresponding immune

Figure 6 (continued)
infiltration scores, with pink indicating positive correlation and green indicating negative correlation. The top bar graph represents the IMMPS scores of patients in the TCGA-LUAD cohort. (C, D) The correlation plot between IMMPS and eosinophil as well as ImmuneScore, each dot represents a sample. (E, F) Correlation of IMMPS with the expression of immune checkpoint genes and immune infiltrating cells obtained by different algorithms, respectively. Each dot indicated the correlation coefficient between immune checkpoint genes. Larger dots indicate higher correlation coefficients, lines indicate immune checkpoints significantly correlated with IMMPS, thicker lines indicated higher correlation, blue indicates negative correlation, red indicated positive correlation.
Full-size DOI: 10.7717/peerj.19121/fig-6

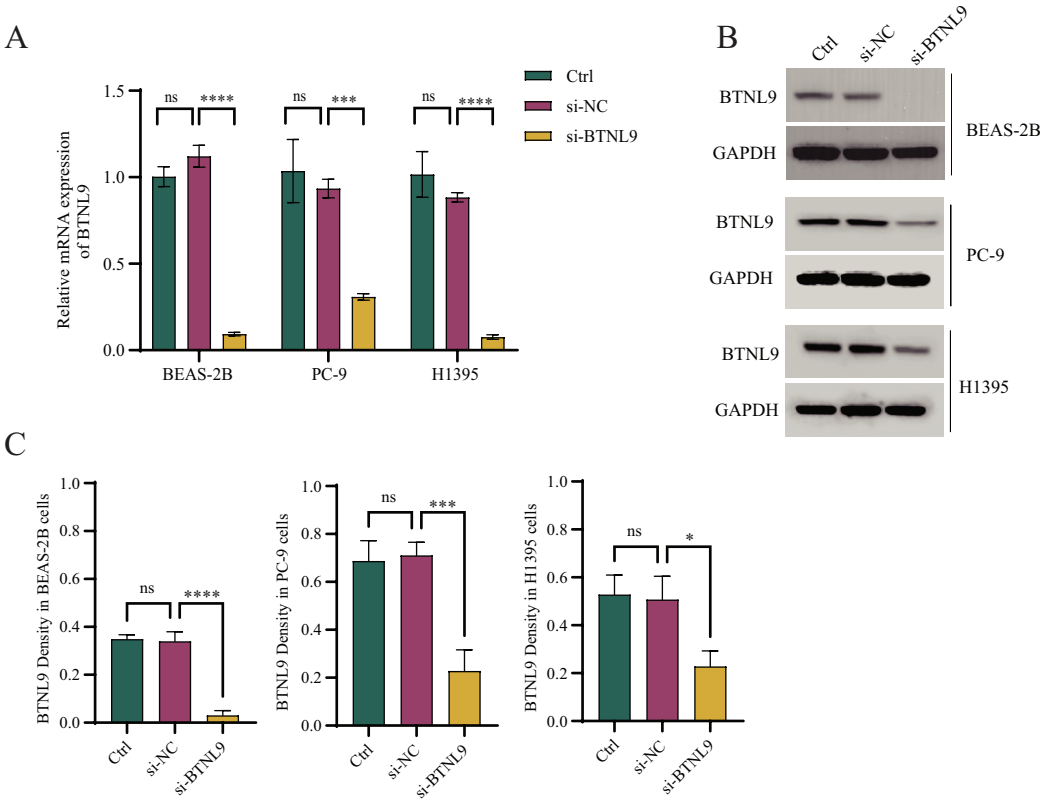

Figure 7 Based on qRT-PCR and Western blot assays to validate the knockdown efficiency of *BTNL9* in LUAD cell lines. (A) Based on qRT-PCR to validate mRNA expression levels after knockdown of *BTNL9* in LUAD cell lines. (B, C) Western blotting-based assay to validate the protein expression level and its quantitative analysis after knockdown of *BTNL9* in LUAD cell line. Ctrl refers to the blank control group; si-NC refers to the control group transfected with non-targeting siRNA; and si-BTNL9 refers to the experimental group transfected with *BTNL9*-specific siRNA. *$p < 0.05$; ***$p < 0.001$; ****$p < 0.0001$; ns, no significant difference.
Full-size DOI: 10.7717/peerj.19121/fig-7

had a lower risk of Coronavirus disease 2019 (COVID-19) and better clinical outcomes (*Xie et al., 2021*). The stability of the results was verified using five other algorithms to avoid bias from different algorithms. IMMPS-High patients showed remarkably lower immune infiltrating cell scores (Figs. S5A–S5E), which was verified by correlation analysis (Fig. 6F). These results validated a negative correlation between immune infiltration and IMMPS (Fig. 6D), specifically, patients who had low IMMPS and higher immune cell infiltration could benefit more from taking immunotherapy. Additionally, analysis on the

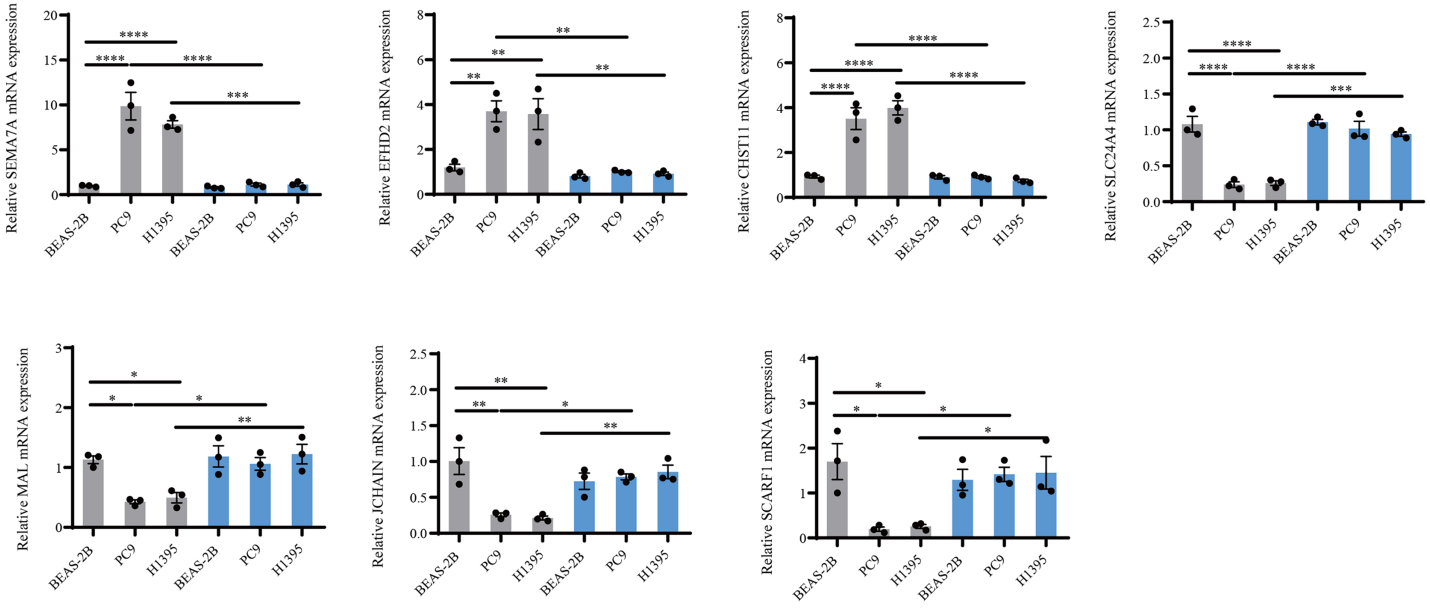

Figure 8 PCR validation of the reliability of the IMMPS model. QRT-PCR was used to detect *SEMA7A*, *EFHD2*, *CHST11*, *SLC24A4*, *MAL*, *JCHAIN* and *SCARF1* expression in BEAS-2B, PC9, H1395 cell lines after *BTNL9* silence ($n = 3$). Gray color serves as a baseline reference for normal controls to assess gene expression changes in LUAD cells; blue color indicates expression changes in mRNA levels of seven genes after knockdown of BTNL9. *$p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. The results were presented as mean ± SEM.

Full-size DOI: 10.7717/peerj.19121/fig-8

correlation between immune checkpoint genes and the IMMPS (Fig. 6E) revealed that most of the HLA Class II genes were significantly negatively related to IMMPS.

Experimental verification using cell lines

Considering that *BTNL9* was the most relevant immune checkpoint gene, *BTNL9* expression was inhibited in BEAS-2B, PC9, and H1395 cell lines and simultaneously the expression of seven genes in the IMMPS model was measured. First, the knockdown efficiency of *BTNL9* in the LUAD cell lines was detected based on qRT-PCR and Western blot assays. As shown in Fig. 7A, knockdown of *BTNL9* significantly downregulated the mRNA expression level of *BTNL9* in human normal lung epithelial cells BEAS-2B and LUAD cell lines PC9 and H1395. Similarly, *BTNL9* protein levels were significantly lowered in the si-*BTNL9* group in comparison to the blank control group (Ctrl) and the control group transfected with non-targeting siRNA (si-NC) (Figs. 7B,7C). After *BTNL9* knockdown, the mRNA expression levels of the seven genes in three cell lines (BEAS-2B, PC9, and H1395) were detected. The results showed that after knockdown of *BTNL9*, the expression of the seven genes in LUAD cell lines (PC9 and H1395) demonstrated significant differences than before the knockdown. This suggested that *BTNL9* may be an upstream regulator of these genes, and that knockdown of *BTNL9* changed the expression of these seven genes in LUAD cells, indicating that these genes may be involved in the occurrence and development of LUAD (Fig. 8).

DISCUSSION

Immunotherapy is a fast developing area of therapeutic molecularly targeted therapies for the treatment of advanced tumors, but not all patients respond actively to immunotherapy. Predictive features are urgently needed to be determined in order to accurately identify patients who can potentially benefit from immunotherapies. The TME is composed of multiple cell types interacting with each other *via* growth factors, cytokines, and chemokines (*Duan et al., 2020*), providing a basis for evaluating the effectiveness of immunotherapy (*Zemek et al., 2019*). Tumors can be classified as either immune-cold or immune-hot according to the features of the TME. Specifically, immune-cold tumors are characterized by immunosuppressive TME and insensitivity to immunotherapy, while immune-hot tumors had a high sensitivity to immunotherapy with active T-cell infiltration (*Gajewski, 2015*). Therefore, the use of practical biomarkers to differentiate the two types of tumors may be able to evaluate the response to immunotherapy.

This study determined immune-hot and immune-cold LUAD tumors using consensus clustering. Genes related to immune-hot LUAD were filtered by WGCNA, and based on the expression profiles of these genes, an integrated pipeline was developed to build the IMMPS. ROC and C-index analyses showed a high accuracy and stability of the IMMPS in one training cohort and four independent cohorts, along with a high generalization ability. These findings supported a great potential of IMMPS in clinical applications.

We collected 154 gene signatures published in the last 5 years. Despite the fact that considerable signatures have been proposed, only a few have been clinically validated or translated into clinical use (*Zeng et al., 2022*; *Wang et al., 2020*). This may be explained by a poor model generalization due to overfitting problem. Noticeably, the model proposed by *Deng et al. (2022)* showed a similar performance to the IMMPS across all the cohorts, but the applicability of their model was severely limited by a high cost to detect a total of 22 genes in the model. However, the IMMPS developed with fewer featured genes by the combination of two machine learning algorithms had a better generalization capability.

Over the past few years, checkpoint inhibitor-based immunotherapies have been substantially advanced for many cancer types, including LUAD (*Havel, Chowell & Chan, 2019*). Although immune checkpoint genes, such as PD-L1, are available biomarkers in clinical practice, their expression could not be independently used to indicate ICI response (*Lupo et al., 2018*). Our study found that immune-hot tumors had a lower IMMPS. Moreover, the IMMPS showed a significant negative correlation with a greater number of immune checkpoint genes, particularly HLA Class II genes. A study reported that HLA class II-restricted neoantigens could influence patients' responses to ICB in a way distinct and complementary to the responses mediated by HLA class I (*Shao et al., 2022*). Moreover, an early study also identified HLA class II molecule as a potential marker of immune checkpoint blockade (ICB) in NSCLC (*Mei et al., 2022*). These results indicated that LUAD patients with a low IMMPS could benefit from ICI treatment.

The IMMPS model contained seven genes, which all played an important role in the development of inflammation or cancer. *SEMA7A* has been reported to exert an

anti-inflammatory effect on the human body possibly through converting pro-inflammatory M1 macrophages into anti-inflammatory M2 macrophages (*Körner et al., 2021*). However, in the research related to atherosclerosis, *Hong et al. (2020)* found that *SEMA7A* could promote ontological to mesenchymal transition through TGF- β 2/Smad signaling. Therefore, it was speculated that *SEMA7A* can serve as a target for vascular growth inhibition in cancer treatment. *EFHD2* could promote the tolerance of NSCLC to cisplatin treatment through NOX4-ROS-ABCC1 axis (*Fan et al., 2020*). Study showed that *CHST11* is related to lung cancer (*Li et al., 2022*), consistently, our qPCR results also confirmed that this gene was high-expressed in advanced LUAD patients and may be involved in the metastasis of NSCLC through dysregulation of ceruloplasmin and intracellular iron balance (*Chang et al., 2022*). *SLC24A4* has been identified as a methylation driver gene and used to construct a riskscore model for LUAD prognosis (*Ren et al., 2020*). *MAL*, a member of MAL family of integral membrane proteins, has the potential to be considered as a biomarker for cancer (*Labat-de-Hoz et al., 2023*). *JCHAIN* is one of the marker genes of B cells in LUAD tissue (*Zhang et al., 2023*). *Ma et al. (2021)* found that LUAD patients with a high expression of *JCHAIN* have a longer survival time. Consistently, our qPCR results also revealed a low expression of *JCHAIN* in advanced LUAD patients. This indicated that *JCHAIN* may be a prognostic marker for LUAD. A recent study reported that advanced hepatocellular carcinoma patients with a high level of *SCARF1* have a favorable overall survival (*Patten et al., 2020*), suggesting an anti-tumor effect of the gene. Overall, the seven genes discovered by this study had significant biological significance and could be used as targets for LUAD treatment in the future.

However, some limitations of this research should be equally acknowledged. Firstly, all the samples analyzed in this study were retrospective, therefore future validation of the IMMPS using prospective multi-center cohorts is imperative. Secondly, some of the clinical and molecular features in the public datasets were highly underrepresented, potentially obscuring the associations between IMMPS and certain variables.

CONCLUSIONS

Applying multiple bioinformatics and machine learning algorithms, the current research developed a stable and robust gene signature for assessing the potential benefit of immunotherapy and the prognosis of LUAD patients. The IMMPS model was a promising tool to optimize clinical decision-making and a monitoring protocol for individual LUAD patient.

ABBREVIATIONS

BP	Biological Process
CC	Cellular Component
CDF	cumulative distribution function
CIBERSORT	Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts
CT	tumor core; E

net	elastic network
CTL	cytotoxic T lymphocyte
ESTIMATE	Estimation of STromal and Immune cells in MAlignant Tumours using Expression data
FDA	Food and Drug Administration
ICGs	immune checkpoint genes
ICIs	immune checkpoint inhibitors
GBM	generalized boosted regression modeling
GEO	Gene Expression Omnibus
GS	gene significance
IMMPS	immune prognostic signature
IM	invasive margins
LOOCV	leave-one-out cross-validation
LUAD	Lung adenocarcinoma
LUSC	lung squamous cell carcinoma
MAD	Median Absolute Deviation
MF	Molecular Function
MM	module membership
NSCLC	non-small cell lung cancer
PAM	partioning around medoids
PlsRcox	partial least squares regression for Cox
qRT-PCR	Quantitative real-time polymerase chain reaction
ROC	receiver operating characteristic
RSF	random survival forest
ssGSEA	Single-sample gene set enrichment analysis
SuperPC	supervised principal components
survival-SVM	survival support vector machine
TCGA	The Cancer Genome Atlas
TME	tumor microenvironment
TOM	topological overlap matrix
TPM	transcripts per kilobase million
WGCNA	Weighted gene co-expression network analysis

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Zhen Chen conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Yongjun Zhang conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The datasets generated and/or analyzed during the current study are available at NCBI: GSE30219, GSE31210, GSE50081, GSE72094, GSE126044 and GSE135222.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.19121#supplemental-information.

REFERENCES

Ahluwalia P, Kolhe R, Gahlay GK. 2021. The clinical relevance of gene expression based prognostic signatures in colorectal cancer. *Biochimica et Biophysica Acta. Reviews on Cancer* **1875(2)**:188513 DOI 10.1016/j.bbcan.2021.188513.

Bagaev A, Kotlov N, Nomie K, Svekolkin V, Gafurov A, Isaeva O, Osokin N, Kozlov I, Frenkel F, Gancharova O, Almog N, Tsiper M, Ataullakhanov R, Fowler N. 2021. Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. *Cancer Cell* **39(6)**:845–865.e7 DOI 10.1016/j.ccell.2021.04.014.

Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautès-Fridman C, Fridman WH, de Reyniès A. 2016. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biology* **17(1)**:218 DOI 10.1186/s13059-016-1070-5.

Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, Angell H, Fredriksen T, Lafontaine L, Berger A, Bruneval P, Fridman WH, Becker C, Pagès F, Speicher MR, Trajanoski Z, Galon J. 2013. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* **39(4)**:782–795 DOI 10.1016/j.immuni.2013.10.003.

Blanche P, Dartigues JF, Jacqmin-Gadda H. 2013. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Statistics in Medicine* **32(30)**:5381–5397 DOI 10.1002/sim.5958.

Carvalho BS, Irizarry RA. 2010. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26(19)**:2363–2367 DOI 10.1093/bioinformatics/btq431.

Chae YK, Arya A, Iams W, Cruz MR, Chandra S, Choi J, Giles F. 2018. Current landscape and future of dual anti-CTLA4 and PD-1/PD-L1 blockade immunotherapy in cancer; lessons learned from clinical trials with melanoma and non-small cell lung cancer (NSCLC). *Journal for Immunotherapy of Cancer* **6(1)**:39 DOI 10.1186/s40425-018-0349-3.

Chang WM, Li LJ, Chiu IA, Lai TC, Chang YC, Tsai HF, Yang CJ, Huang MS, Su CY, Lai TL, Jan YH, Hsiao M. 2022. The aberrant cancer metabolic gene carbohydrate sulfotransferase 11

promotes non-small cell lung cancer cell metastasis via dysregulation of ceruloplasmin and intracellular iron balance. *Translational Oncology* **25**:101508 DOI 10.1016/j.tranon.2022.101508.

Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z. 2017. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Reports* **18(1)**:248–262 DOI 10.1016/j.celrep.2016.12.019.

Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. 2018. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods in Molecular Biology* **1711**:243–259 DOI 10.1007/978-1-4939-7493-1.

Cho JW, Hong MH, Ha SJ, Kim YJ, Cho BC, Lee I, Kim HR. 2020. Genome-wide identification of differentially methylated promoters and enhancers associated with response to anti-PD-1 therapy in non-small cell lung cancer. *Experimental & Molecular Medicine* **52(9)**:1550–1563 DOI 10.1038/s12276-020-00493-8.

Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, Langerød A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowetz F, Murphy L, Ellis I, Purushotham A, Børresen-Dale AL, Brenton JD, Tavaré S, Caldas C, Aparicio S. 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486(7403)**:346–352 DOI 10.1038/nature10983.

Dai Q, Guo M, Duan X, Teng Z, Fu Y. 2019. Construction of complex features for computational predicting ncRNA-protein interaction. *Frontiers in Genetics* **10**:18 DOI 10.3389/fgene.2019.00018.

Deng Y, Chen X, Huang C, Song J, Feng S, Chen X, Zhou R. 2022. Screening and validation of significant genes with poor prognosis in pathologic stage-I lung adenocarcinoma. *Journal of Oncology* **2022(2)**:3794021 DOI 10.1155/2022/3794021.

Der SD, Sykes J, Pintilie M, Zhu CQ, Strumpf D, Liu N, Jurisica I, Shepherd FA, Tsao MS. 2014. Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients. *Journal of Thoracic Oncology* **9(1)**:59–64 DOI 10.1097/JTO.0000000000000042.

Dickerson LK, Carter JA, Kohli K, Pillarisetty VG. 2023. Emerging interleukin targets in the tumour microenvironment: implications for the treatment of gastrointestinal tumours. *Gut* **72(8)**:1592–1606 DOI 10.1136/gutjnl-2023-329650.

Ding Y, Lv J, Hua Y. 2022. Comprehensive metabolomic analysis of lung cancer patients treated with Fu Zheng Fang. *Current Pharmaceutical Analysis* **18(9)**:881–891 DOI 10.2174/1573412918666220822143119.

Duan Q, Zhang H, Zheng J, Zhang L. 2020. Turning cold into hot: firing up the tumor microenvironment. *Trends in Cancer* **6(7)**:605–618 DOI 10.1016/j.trecan.2020.02.022.

Fan CC, Tsai ST, Lin CY, Chang LC, Yang JC, Chen GY, Sher YP, Wang SC, Hsiao M, Chang WC. 2020. EFHD2 contributes to non-small cell lung cancer cisplatin resistance by the activation of NOX4-ROS-ABCC1 axis. *Redox Biology* **34**:101571 DOI 10.1016/j.redox.2020.101571.

Forde PM, Chaft JE, Smith KN, Anagnostou V, Cottrell TR, Hellmann MD, Zahurak M, Yang SC, Jones DR, Broderick S, Battafarano RJ, Velez MJ, Rekhtman N, Olah Z, Naidoo J, Marrone KA, Verde F, Guo H, Zhang J, Caushi JX, Chan HY, Sidhom JW, Scharpf RB, White J, Gabrielson E, Wang H, Rosner GL, Rusch V, Wolchok JD, Merghoub T, Taube JM, Velculescu VE, Topalian SL, Brahmer JR, Pardoll DM. 2018. Neoadjuvant PD-1 Blockade in

resectable lung cancer. *The New England Journal of Medicine* **378(21)**:1976–1986 DOI 10.1056/NEJMoa1716078.

Fridman WH, Pagès F, Sautès-Fridman C, Galon J. 2012. The immune contexture in human tumours: impact on clinical outcome. *Nature Reviews Cancer* **12(4)**:298–306 DOI 10.1038/nrc3245.

Gajewski TF. 2015. The next hurdle in cancer immunotherapy: overcoming the non-T-cell-inflamed tumor microenvironment. *Seminars in Oncology* **42(4)**:663–671 DOI 10.1053/j.seminoncol.2015.05.011.

Gao S, Li N, Gao S, Xue Q, Ying J, Wang S, Tao X, Zhao J, Mao Y, Wang B, Shao K, Lei W, Wang D, Lv F, Zhao L, Zhang F, Zhao Z, Su K, Tan F, Gao Y, Sun N, Wu D, Yu Y, Ling Y, Wang Z, Duan C, Tang WP, Zhang L, He S, Wu N, Wang J, He J. 2020. Neoadjuvant PD-1 inhibitor (Sintilimab) in NSCLC. *Journal of Thoracic Oncology* **15(5)**:816–826 DOI 10.1016/j.jtho.2020.01.017.

Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, Nair VS, Xu Y, Khuong A, Hoang CD, Diehn M, West RB, Plevritis SK, Alizadeh AA. 2015. The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nature Medicine* **21(8)**:938–945 DOI 10.1038/nm.3909.

Gu Z, Eils R, Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32(18)**:2847–2849 DOI 10.1093/bioinformatics/btw313.

Guo J, Yang YUN, Zhao WEI, Yan Z, Yang XIA, Yan Y, Hao R, Hu J, Jiao FEI. 2021. MiR-16-5p plays an inhibitory role in human non-small cell lung cancer through Fermitin family member 2. *Biocell* **45(3)**:627–638 DOI 10.32604/biocell.2021.013496.

Hänzelmann S, Castelo R, Guinney J. 2013. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**:7 DOI 10.1186/1471-2105-14-7.

Harrison WT, Yeates RM, Phillips ML, Nenoff TM. 2003. New framework connectivity patterns in templated networks: the creatinine zinc phosphites C4N3OH7.ZnHPO3, C4N3OH7.Zn (H2O)HPO3, and (C4N3OH7)2.ZnHPO3.H2O. *Inorganic Chemistry* **42(5)**:1493–1498 DOI 10.1021/ic020593+.

Havel JJ, Chowell D, Chan TA. 2019. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nature Reviews Cancer* **19(3)**:133–150 DOI 10.1038/s41568-019-0116-x.

Hong L, Li F, Tang C, Li L, Sun L, Li X, Zhu L. 2020. Semaphorin 7A promotes endothelial to mesenchymal transition through ATF3 mediated TGF-β2/Smad signaling. *Cell Death & Disease* **11(8)**:695 DOI 10.1038/s41419-020-02818-x.

Hoshida Y, Brunet JP, Tamayo P, Golub TR, Mesirov JP. 2007. Subclass mapping: identifying common subtypes in independent disease data sets. *PLOS ONE* **2(11)**:e1195 DOI 10.1371/journal.pone.0001195.

Kim JY, Choi JK, Jung H. 2020. Genome-wide methylation patterns predict clinical benefit of immunotherapy in lung cancer. *Clinical Epigenetics* **12(1)**:119 DOI 10.1186/s13148-020-00907-4.

Koncina E, Haan S, Rauh S, Letellier E. 2020. Prognostic and predictive molecular biomarkers for colorectal cancer: updates and challenges. *Cancers* **12(2)**:319 DOI 10.3390/cancers12020319.

Körner A, Bernard A, Fitzgerald JC, Alarcon-Barrera JC, Kostidis S, Kaussen T, Giera M, Mirakaj V. 2021. Sema7A is crucial for resolution of severe inflammation. *Proceedings of the National Academy of Sciences of the United States of America* **118(9)**:e2017527118 DOI 10.1073/pnas.2017527118.

Labat-de-Hoz L, Rubio-Ramos A, Correas I, Alonso MA. 2023. The MAL family of proteins: normal function, expression in cancer, and potential use as cancer biomarkers. *Cancers* **15(10)**:2801 DOI 10.3390/cancers15102801.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**:559 DOI 10.1186/1471-2105-9-559.

Li CH, Chan MH, Chang YC, Hsiao M. 2022. The CHST11 gene is linked to lung cancer and pulmonary fibrosis. *The Journal of Gene Medicine* **24(12)**:e3451 DOI 10.1002/jgm.3451.

Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, Li B, Liu XS. 2020. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Research* **48(W1)**:W509–w514 DOI 10.1093/nar/gkaa407.

Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, Jiang P, Shen H, Aster JC, Rodig S, Signoretti S, Liu JS, Liu XS. 2016. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biology* **17(1)**:174 DOI 10.1186/s13059-016-1028-7.

Liu Z, Liu L, Weng S, Guo C, Dang Q, Xu H, Wang L, Lu T, Zhang Y, Sun Z, Han X. 2022. Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. *Nature Communications* **13**:816 DOI 10.1038/s41467-022-28421-6.

Lupo A, Alifano M, Wislez M, Boulle G, Velut Y, Biton J, Cremer I, Goldwasser F, Leroy K, Damotte D. 2018. Biomarkers predictive of PD1/PD-L1 immunotherapy in non-small cell lung cancer. *Revue de Pneumologie Clinique* **74(5)**:339–350 DOI 10.1016/j.pneumo.2018.09.010.

Ma Q, Chen Y, Xiao F, Hao Y, Song Z, Zhang J, Okuda K, Um SW, Silva M, Shimada Y, Si C, Liang C. 2021. A signature of estimate-stromal-immune score-based genes associated with the prognosis of lung adenocarcinoma. *Translational Lung Cancer Research* **10(3)**:1484–1500 DOI 10.21037/tlcr-21-223.

Malhotra J, Jabbour SK, Aisner J. 2017. Current state of immunotherapy for non-small cell lung cancer. *Translational Lung Cancer Research* **6(2)**:196–211 DOI 10.21037/tlcr.2017.03.01.

Mei J, Jiang G, Chen Y, Xu Y, Wan Y, Chen R, Liu F, Mao W, Zheng M, Xu J. 2022. HLA class II molecule HLA-DRA identifies immuno-hot tumors and predicts the therapeutic response to anti-PD-1 immunotherapy in NSCLC. *BMC Cancer* **22**:738 DOI 10.1186/s12885-022-09840-6.

Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. 2015. Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods* **12(5)**:453–457 DOI 10.1038/nmeth.3337.

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, Takenoshita S, Gotoh N, Mizuno H, Sarai A, Kawano S, Yamaguchi R, Miyano S, Yokota J. 2012. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Research* **72(1)**:100–111 DOI 10.1158/0008-5472.CAN-11-1403.

Okcu O, Öztürk Ç, Şen B, Ayazoğlu MS, Güvendi GF, Öztürk SD, Aşkan G, Bedir R. 2023. The prognostic significance of non-lymphoid immune cells of the tumor microenvironment, including neutrophils, eosinophils, and mast cells in breast carcinomas. *Annals of Diagnostic Pathology* **65(3)**:152151 DOI 10.1016/j.anndiagpath.2023.152151.

Pagès F, Kirilovsky A, Mlecnik B, Asslaber M, Tosolini M, Bindea G, Lagorce C, Wind P, Marliot F, Bruneval P, Zatloukal K, Trajanoski Z, Berger A, Fridman WH, Galon J. 2009. In situ cytotoxic and memory T cells predict outcome in patients with early-stage colorectal cancer. *Journal of Clinical Oncology* **27(35)**:5944–5951 DOI 10.1200/JCO.2008.19.6147.

Pagès F, Mlecnik B, Marliot F, Bindea G, Ou FS, Bifulco C, Lugli A, Zlobec I, Rau TT, Berger MD, Nagtegaal ID, Vink-Börger E, Hartmann A, Geppert C, Kolwelter J, Merkel S,

Grützmann R, Van den Eynde M, Jouret-Mourin A, Kartheuser A, Léonard D, Remue C, Wang JY, Bavi P, Roehrl MHA, Ohashi PS, Nguyen LT, Han S, MacGregor HL, Hafezi-Bakhtiari S, Wouters BG, Masucci GV, Andersson EK, Zavadova E, Vocka M, Spacek J, Petruzelka L, Konopasek B, Dundr P, Skalova H, Nemejcova K, Botti G, Tatangelo F, Delrio P, Ciliberto G, Maio M, Laghi L, Grizzi F, Fredriksen T, Buttard B, Angelova M, Vasaturo A, Maby P, Church SE, Angell HK, Lafontaine L, Bruni D, El Sissy C, Haicheur N, Kirilovsky A, Berger A, Lagorce C, Meyers JP, Paustian C, Feng Z, Ballesteros-Merino C, Dijkstra J, van de Water C, van Lent-van Vliet S, Knijn N, Mușină AM, Scripcariu DV, Popivanova B, Xu M, Fujita T, Hazama S, Suzuki N, Nagano H, Okuno K, Torigoe T, Sato N, Furuhata T, Takemasa I, Itoh K, Patel PS, Vora HH, Shah B, Patel JB, Rajvik KN, Pandya SJ, Shukla SN, Wang Y, Zhang G, Kawakami Y, Marincola FM, Ascierto PA, Sargent DJ, Fox BA, Galon J. 2018. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. *Lancet* **391(10135)**:2128–2139 DOI 10.1016/S0140-6736(18)30789-X.

Pan X, Zhang C, Wang J, Wang P, Gao Y, Shang S, Guo S, Li X, Zhi H, Ning S. 2022. Epigenome signature as an immunophenotype indicator prompts durable clinical immunotherapy benefits in lung adenocarcinoma. *Briefings in Bioinformatics* **23**:514 DOI 10.1093/bib/bbab481.

Patten DA, Wilkinson AL, O'Rourke JM, Shetty S. 2020. Prognostic value and potential immunoregulatory role of SCARF1 in hepatocellular carcinoma. *Frontiers in Oncology* **10**:565950 DOI 10.3389/fonc.2020.565950.

Qian X-J, Wang J-W, Liu J-B, Yu X. 2023. The mediating role of miR-451/ETV4/MMP13 signaling axis on Epithelialmesenchymal transition in promoting non-small cell lung cancer progression. *Current Molecular Pharmacology* **17**:e210723218988 DOI 10.2174/1874467217666230721123554.

Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. 2017. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *eLife* **6**:e26476 DOI 10.7554/eLife.26476.049.

R Core Team. 2021. *R: a language and environment for statistical computing*. Version 3.6.4. Vienna: R Foundation for Statistical Computing. *Available at https://www.r-project.org*.

Ren J, Yang Y, Li C, Xie L, Hu R, Qin X, Zhang M. 2020. A novel prognostic model of early-stage lung adenocarcinoma integrating methylation and immune biomarkers. *Frontiers in Genetics* **11**:634634 DOI 10.3389/fgene.2020.634634.

Rice SJ, Miller B, Wagman M, Jamorabo DS, Liu X, Belani CP. 2016. Clinical approaches to immunotherapy in non-small cell lung cancer: current and future perspectives. *Current Molecular Pharmacology* **9(3)**:183–195 DOI 10.2174/1874467208666150716120108.

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, Timsit JF, Brambilla C, Brambilla E, Khochbin S. 2013. Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. *Science Translational Medicine* **5(186)**:186ra66 DOI 10.1126/scitranslmed.3005723.

Schabath MB, Welsh EA, Fulp WJ, Chen L, Teer JK, Thompson ZJ, Engel BE, Xie M, Berglund AE, Creelan BC, Antonia SJ, Gray JE, Eschrich SA, Chen DT, Cress WD, Haura EB, Beg AA. 2016. Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma. *Oncogene* **35(24)**:3209–3216 DOI 10.1038/onc.2015.375.

Schoenfeld AJ, Hellmann MD. 2020. Acquired resistance to immune checkpoint inhibitors. *Cancer Cell* **37(4)**:443–455 DOI 10.1016/j.ccell.2020.03.017.

Scortegagna M, Du Y, Bradley LM, Wang K, Molinolo A, Ruppin E, Murad R, Ronai ZA. 2023. Ubiquitin Ligases Siah1a/2 control alveolar macrophage functions to limit carcinogen-induced lung adenocarcinoma. *Cancer Research* **83(12)**:2016–2033 DOI 10.1158/0008-5472.CAN-23-0258.

Shao XM, Huang J, Niknafs N, Balan A, Cherry C, White J, Velculescu VE, Anagnostou V, Karchin R. 2022. HLA class II immunogenic mutation burden predicts response to immune checkpoint blockade. *Annals of Oncology* **33(7)**:728–738 DOI 10.1016/j.annonc.2022.03.013.

Siegel RL, Miller KD, Jemal A. 2019. Cancer statistics, 2019. *CA: A Cancer Journal for Clinicians* **69(1)**:7–34 DOI 10.3322/caac.21551.

Tang Z, Wang L, Wu G, Qin L, Tan Y. 2023. FGD5 as a novel prognostic biomarker and its association with immune infiltrates in lung adenocarcinoma. *Biocell* **47(11)**:2503–2516 DOI 10.32604/biocell.2023.031565.

Tian S. 2017. Classification and survival prediction for early-stage lung adenocarcinoma and squamous cell carcinoma patients. *Oncology Letters* **14(5)**:5464–5470 DOI 10.3892/ol.2017.6835.

Wang X, Yao S, Xiao Z, Gong J, Liu Z, Han B, Zhang Z. 2020. Development and validation of a survival model for lung adenocarcinoma based on autophagy-associated genes. *Journal of Translational Medicine* **18(1)**:149 DOI 10.1186/s12967-020-02321-z.

Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. 2013. The cancer genome atlas pan-cancer analysis project. *Nature Genetics* **45(10)**:1113–1120 DOI 10.1038/ng.2764.

Wilkerson MD, Hayes DN. 2010. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26(12)**:1572–1573 DOI 10.1093/bioinformatics/btq170.

Xie G, Ding F, Han L, Yin D, Lu H, Zhang M. 2021. The role of peripheral blood eosinophil counts in COVID-19 patients. *Allergy* **76(2)**:471–482 DOI 10.1111/all.14465.

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, Mills GB, Verhaak RG. 2013. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature Communications* **4**:2612 DOI 10.1038/ncomms3612.

Yu G, Wang LG, Han Y, He QY. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: A Journal of Integrative Biology* **16(5)**:284–287 DOI 10.1089/omi.2011.0118.

Zemek RM, De Jong E, Chin WL, Schuster IS, Fear VS, Casey TH, Forbes C, Dart SJ, Leslie C, Zaitouny A, Small M, Boon L, Forrest ARR, Muiri DO, Degli-Esposti MA, Millward MJ, Nowak AK, Lassmann T, Bosco A, Lake RA, Lesterhuis WJ. 2019. Sensitization to immune checkpoint blockade through activation of a STAT1/NK axis in the tumor microenvironment. *Science Translational Medicine* **11(501)**:5 DOI 10.1126/scitranslmed.aav7816.

Zeng Z, Zuo Y, Jin Y, Peng Y, Zhu X. 2022. Identification of extracellular matrix signatures as novel potential prognostic biomarkers in lung adenocarcinoma. *Frontiers in Genetics* **13**:872380 DOI 10.3389/fgene.2022.872380.

Zhang Y, Shi J, Luo J, Liu C, Zhu L. 2023. Metabolic heterogeneity in early-stage lung adenocarcinoma revealed by RNA-seq and scRNA-seq. *Clinical & Translational Oncology* **25(6)**:1844–1855 DOI 10.1007/s12094-023-03082-z.

Zhang L, Yu S, Hong S, Xiao X, Liao Z, Li Y, Xiao H. 2023. Comprehensive analysis of BTNL9 as a prognostic biomarker correlated with immune infiltrations in thyroid cancer. *BMC Medical Genomics* **16**:234 DOI 10.1186/s12920-023-01676-8.

Zhao Y, Gu S, Li L, Zhao R, Xie S, Zhang J, Zhou R, Tu L, Jiang L, Ma S, Zhang S. 2023. A novel risk signature for predicting brain metastasis in patients with lung adenocarcinoma. *Neuro-Oncology* **25(12)**:2207–2220 DOI 10.1093/neuonc/noad115.

Zheng X, Hu Y, Yao C. 2017. The paradoxical role of tumor-infiltrating immune cells in lung cancer. *Intractable & Rare Diseases Research* **6(4)**:234–241 DOI 10.5582/irdr.2017.01059.

Zhou B, Gao S. 2022. Construction and validation of a novel immune and tumor mutation burden-based prognostic model in lung adenocarcinoma. *Cancer Immunology, Immunotherapy: CII* **71**:1183–1197 DOI 10.1007/s00262-021-03066-4.