IET Systems Biology

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

Development and validation of an immune-related gene signature for prognosis in Lung adenocarcinoma

Zehuai Guo[1] [ID] | Xiangjun Qi[1] | Zeyun Li[1] | Jianying Yang[1] | Zhe Sun[1] | Peiqin Li[1] | Ming Chen[2] | Yang Cao[2]

[1]The First Clinical School of Guangzhou University of Chinese Medicine, Guangzhou, China

[2]The First Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, China

Correspondence

Yang Cao, The First Affiliated Hospital of Guangzhou University of Chinese Medicine, No.16 Jichang Road, Baiyun District, Guangzhou 510405, China.
Email: caoyang0342@gzucm.edu.cn

Abstract

The most common type of lung cancer tissue is lung adenocarcinoma. The TCGA-LUAD cohort retrieved from the TCGA dataset was considered the internal training cohort, while GSE68465 and GSE13213 datasets from the GEO database were used as the external test cohort. The TCGA-LUAD cohort was classified into two immune subtypes using single-sample gene set enrichment analysis of the immune gene set and unsupervised clustering analysis. The ESTIMATE algorithm, the CIBERSORT algorithm, and HLA family expression levels again validated the reliability of this typing. We performed Venn analysis using immune-related genes from the immport dataset and differentially expressed genes from the subtypes to retrieve differentially expressed immune genes (DEIGs). In addition, DEIGs were used to construct a prognostic model with the least absolute shrinkage and selection operator regression analysis. A reliable risk model consisting of 11 DEIGs, including S100P, INHA, SEMA7A, INSL4, CD40LG, AGER, SERPIND1, CD1D, CX3CR1, SFTPD, and CD79A, was then built, and its reliability was further confirmed by ROC curve and calibration plot analysis. The high-risk score subgroup had a poor prognosis and a lower tumour immune dysfunction and exclusion score, indicating a greater likelihood of anti-PD-1/cytotoxic T lymphocyte antigen 4 benefit.

KEYWORDS

bioinformatics, immune-related gene, immunotherapy, lung adenocarcinoma, prognosis, ssGSEA

1 | INTRODUCTION

According to GLOBOCAN 2020 cancer report [1], lung cancer is reported to be among the most burdensome malignancies globally. Predominant cancer detected in 40%–50% of all incidents of lung malignancies is lung adenocarcinoma (LUAD) [2]. It is widely accepted that a healthy immune system's regular function may guard and inhibit the formation of malignant tumours and those persons having a genetically impaired immunological system might be more susceptible to tumours [3, 4]. Immune checkpoint inhibitor (ICI) therapy has proven to be an auspicious treatment for LUAD [5, 6], has facilitated significant progress in anticancer practice, and is now the new first-line oncological treatment choice [7]. Unlike conventional therapies, immunotherapy patients achieve

therapeutic advantages by generating a long-lasting antitumour immune response, which is dependent on immunomodulation between the cancer cells and tumour microenvironment (TME) [8]. The most prevalent targets of ICI are cytotoxic T lymphocyte antigen 4 (CTLA4) and programmed cell death 1 (PD1), which limit the stimulation and proliferation of T lymphocytes, rendering the anti-tumour responses ineffectual [9]. Despite the fact that ICI was predicted to show significant promises in the immunotherapy of LUAD patients, the clinical results and prognosis were disappointing. For instance, only 44.8% of PD-L1-positive NSCLCs respond to pembrolizumab in the first-line context [10]. Meanwhile, the heterogeneity of elevated levels of tumour-infiltrating lymphocytes and tumour mutational load in LUAD provides an extra reason for it [11]. As a result, in order to accomplish accurately individualised

decision-making for ICI therapy, a prognostic biomarker will be required to evaluate prognosis and forecast immunotherapy sensitivity in LUAD patients.

The modelling of immunological clinical prognoses is a hot area of research today. There are various methods available for identifying immune genes associated with prognosis, such as the ESTIMATE algorithm [12], the CIBERSORT algorithm [13], independent and paired sample designs [14], or direct univariate Cox regression analysis of genes from various immune databases, including the InnateDB database [14], the ImmPort database [15], and the MSigDB database combined with the ImmPort database [16]. The single-sample gene set enrichment analysis (ssGSEA) converts the gene expression profiles of individual samples into gene set enrichment profiles, a transformation that allows researchers to characterise cellular states based on the activity levels of biological processes and pathways rather than the expression levels of individual genes. Although this method of dividing tumour patients into multiple immune subtypes has been widely employed in immune-related prognostic models for numerous malignancies [17–19], LUAD patients have not yet benefited from its use.

For this research, we utilised ssGSEA to allot LUAD patients into low- and high-immune cell infiltration clusters and confirmed by ESTIMATE, GSEA, the expression level of multiple human leucocyte antigen (HLA) genes and immune cell. Subsequently, utilising univariate Cox and least absolute shrinkage and selection operator (LASSO) regression analysis, we identified an 11 differentially expressed immune genes (DEIGs) signature linked to prognosis in a cross-section of differentially expressed genes (DEGs) in subtypes and immune-related genes (IRGs) from the immport dataset. The validity of the DEIGs prognostic signature was then evaluated. Ultimately,

we discovered that this prognostic signature was not only trustworthy for predicting survival but also effective for predicting the treatment response to ICI for LUAD patients, which could allow individualised immunotherapy treatment in the future.

2 | MATERIALS AND METHOD

2.1 | Data collection and collation

We obtained the TCGA-LUAD cohort's gene expression and patient information from The Cancer Genome Atlas Programme (TCGA; https://portal.gdc.cancer.gov/repository). The mutation information for these LUAD patients was obtained from the openly accessible TCGA dataset through the GDC Data Portal. GSE68465 and GSE13213 from the GEO dataset were recruited for external validation. The expression values from the TCGA-LUAD, GSE68465, and GSE13213 datasets were all $log2(x + 1)$ transformed. Subsequently, batch effects were removed from the three datasets by the ComBat algorithm of the 'sva' package. Figure 1 shows the workflow and details of this study. 1793 IRGs were downloaded from the ImmPort Portal database (https://www.immport.org/).

2.2 | ssGSEA and clustering

To examine the associated expression pathways, activity of immunological-associated functions, and penetration levels of various immunological cells, we used 29 immune data sets (such as immunological-associated functions, immune cell types, and
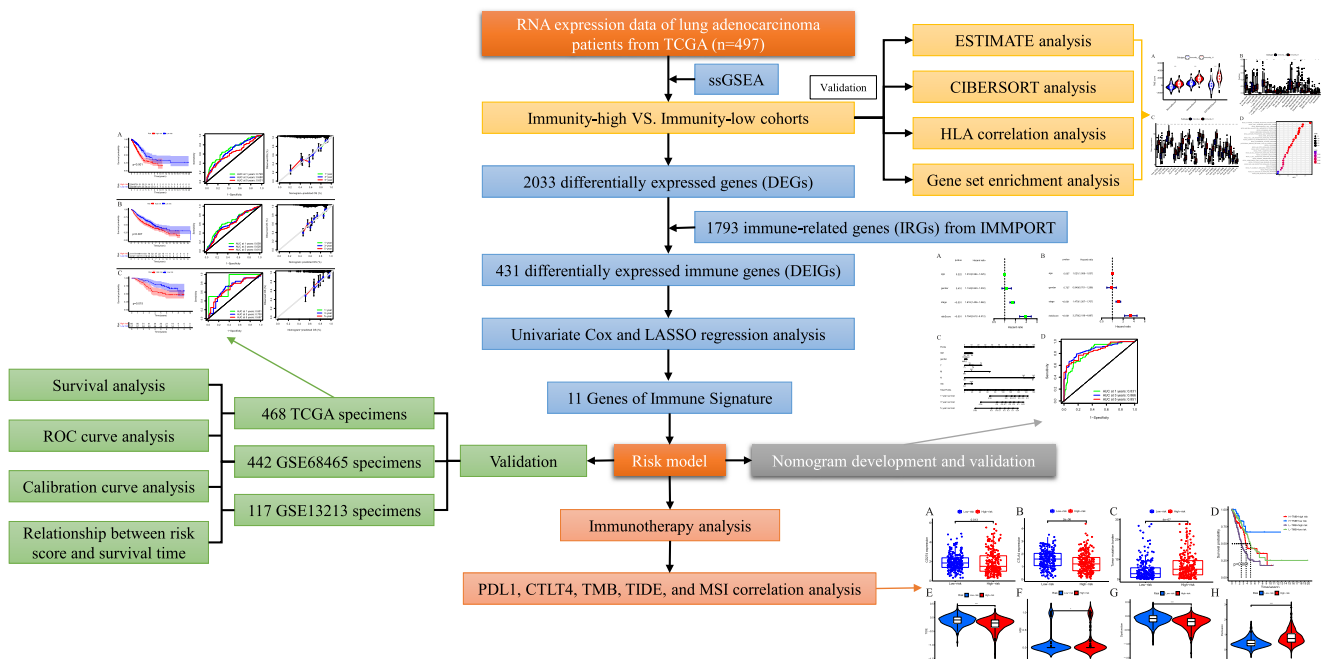

FIGURE 1 The workflow and details of this study.

immune-associated pathways) and the ssGSEA method with the R software gene set variation analysis package. After data correction of the obtained ssGSEA scores by the normalisation function 'normalise', we performed unsupervised hierarchical clustering of TCGA-LUAD to develop an immunogenomic categorisation of LUAD. We utilised k-means clustering to split these patients into two categories based on their immunological scores: Immunity-high and Immunity-low cohorts.

2.3 | Assessment of the relationship between LUAD immunogenomic subtypes and molecular features

Firstly, the immune cell infiltration levels, tumour purity, and stromal content of each LUAD specimen were computed in terms of various scores utilising the 'Estimation of STromal and Immune cells in MAlignant Tumour tissues using Expression data' (ESTIMATE) technique. Secondly, to uncover the connections between LUAD immunogenomic categorisation and immune infiltration, we analysed the scores of stromal content, immune cell infiltration, and estimation scores across various LUAD immunogenomic subtypes.

In addition, tumour-infiltrating immune cells (TIICs) in LUAD specimens were evaluated utilising the 'Cell type Identification by Estimating Relative Subsets of RNA Transcripts CIBERSOFT' (CIBERSORT) deconvolution method. The CIBERSORT platform (https://cibersortx.stanford.edu/) was used to collect the gene expression signature matrix of 22 TIICs. We utilised Mann-Whitney U-test to evaluate the percentages of immune cell subsets amongst LUAD immunogenomic subtypes and defined 100 permutations and $P < 0.05$ as the criterion for effective specimen deconvolution.

Next, for each cohort, we computed the levels of HLA genes expression. The Kolmogorov-Smirnov test was utilised to determine the levels of HLA genes expression amongst immunogenomic subtypes.

Eventually, GSEA was done utilising Java GSEA programme to examine how the immune-associated pathways vary between the two cohorts. To filter for relevant enrichment findings, the enrichment threshold was established at a false discovery rate (FDR) of <0.01. The enrichment findings were shown graphically utilising the R package 'ggplot2R.'

2.4 | Selection of immune-related prognostic genes in LUAD

The TCGA data were categorised into 2 cohorts premised on immune cell infiltration levels: high and low. We utilised the edgeR programme to evaluate DEGs in accordance with the protocols of the $p < 0.05$ and $| \log2FC | > 2$ criteria and the Venn diagram to retrieve DEIGs between the DEGs and IRGs. We utilised Univariate and LASSO regression analysis to determine the link between the overall survival (OS) of patients and the level of IRGs expression.

2.5 | Construct a prognostic model of DEIGs

The LASSO regression method's coefficients were utilised to get the risk score formula shown below: risk score = sum of coefficients * the IRGs expression level. In the TCGA-LUAD, GSE68465, and GSE13213 cohorts, the risk score from each subject was computed independently utilising this formula. Following that, subjects were split into low-risk and high-risk groups premised on the median risk score. Afterwards, the Kaplan-Meier survival curves were plotted for the 3 cohorts. To assess the model's sensitivity and specificity, we suggested calibration plots and ROC curves. We then undertook a multivariate and univariate analysis of numerous medical features of TCGA-LUAD patients to investigate the independence of the prognostic models in the absence of clinical variables.

2.6 | Assessment of variations between high- and low-risk cohorts

To begin with, we determined the levels of expression of the CTLA4 and CD274 genes in each subset. The Kolmogorov-Smirnov test was utilised to investigate the extent of CTLA4 and CD274 genes expression amongst risk categories.

Furthermore, tumour mutation burden (TMB) was described as the number of non-synonymous coding mutations per megabase (Mb). The GDC data site was utilised to obtain the mutation annotation format, which was then displayed with the maftools R programme. The median Mb score was utilised as a minimum threshold to categorise patients into high-TMB and low-TMB cohorts. Tumour immune dysfunction and exclusion (TIDE) score was thereafter computed online (HTTP://tide.dfci.harvard.edu/). Correlation analyses were performed between risk cohorts and TMB/TIDE scores.

2.7 | Development and evaluation of a predictive nomogram

Premised on the multivariate analysis findings, a nomogram was created with the rms R package to estimate the 1-, 3-, and 5-year survival chances for LUAD patients. Furthermore, the ROC curve was utilised in assessig the nomogram's predictive capacity.

3 | RESULTS

3.1 | Construction of LUAD clustering

An aggregate of 497 LUAD specimens from the TCGA was employed in this study. We utilised the ssGSEA scores to measure the levels of enrichment or activity of immunological cells, functions, or pathways in cancer specimens. The 497 specimens were then divided into two categories utilising the

unsupervised hierarchical clustering technique (Figure 2a). We designated the 2 clusters as 2 LUAD immunogenic subcategories termed Immunity High and Immunity Low premised on the heatmap of expression levels of the 29 IRG sets (Figure 2b).

The immune subtype of each sample in the TCGA cohort was shown in Supplementary Table 1.

3.2 | Validation of LUAD clustering

We utilised the ESTIMATE method to evaluate the expression signature of LUAD and computed the stromal score, ESTIMATE score, tumour purity, and immune score. According to the findings, the tumour purity of the high immune cell infiltration cohort was considerably reduced as opposed to the
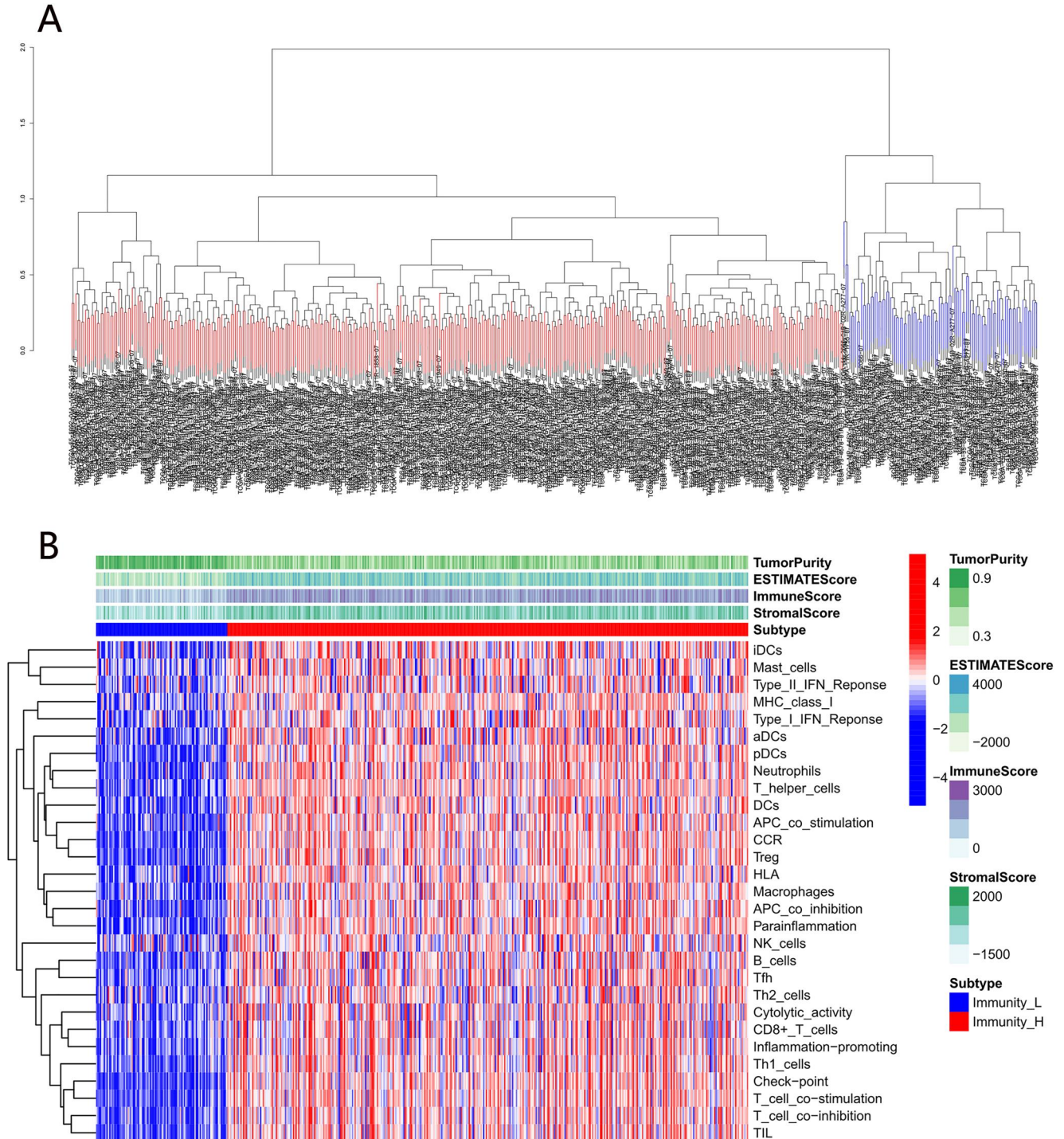

FIGURE 2 Construction of Lung adenocarcinoma (LUAD) clustering. (a) Unsupervised hierarchical clustering of 497 LUAD in TCGA. (b) The distribution of stromal score, immune score, estimate score, tumour purity, and 29 immune-related compositions of each patient in immunity-high and immunity-low cohorts.

levels of low immune cell infiltration cohort. The ESTIMATE score, immunological score, and stromal score, on the other hand, were greater in the high immune cell infiltration cohort as opposed to the low immune cell infiltration cohort (Figure 3a). The CIBERSORT technique was utilised to evaluate the aforementioned two cohorts, and it revealed that the high immune cell infiltration cohort exhibited elevated kinds of immune cells (Figure 3b). The high immune cell infiltration cohort exhibited considerably elevated proportions of dendritic cells resting and neutrophils, macrophages M1, T cells regulatory (Tregs), T cells CD8, and T cells CD4 memory activated, whereas dendritic cells activated, NK cells activated, T cells CD4 naïve, B cell naïve, and Eosinophils proportions were relatively lower. Furthermore, the HLA family was more prevalent in the high immune cell infiltration cohort (Figure 3c). To learn more about the immunogenomic subtypes' inherent genetic impacts, we executed GSEA to reveal the pathway of the immunogenomic (Figure 3d). In accordance with these outcomes, the Immunity High subtypes were predominantly enriched in natural killer cell-mediated cytotoxicity pathway, haematopoietic cell lineage, cell adhesion molecules, cytokine-cytokine receptor interaction, etc.

All things considered, our results suggest that the high- and low-immune groups constructed in this study have strong immune heterogeneity.

3.3 | Evaluation of DEIGs with low and high immune cell infiltration

Premised on the threshold of $|log2FC| > 2$ and FDR 0.05, we discovered 2033 DEGs between the high and low immunological cell infiltration cohorts (Figure 4a). We utilised the IRGs from the immport dataset and the DEGs from the low and high immune cell infiltration cohorts to perform a Venn analysis. Afterwards, we discovered 431 DEIGs (Figure 4b).

3.4 | Prognosis models of DEIGs

Following the integration of clinical data into gene expression profiles, we retrieved 468 TCGA specimens (Supplementary Table 2), 442 GSE68465 specimens (Supplementary Table 3), and 117 GSE13213 specimens (Supplementary Table 4). The
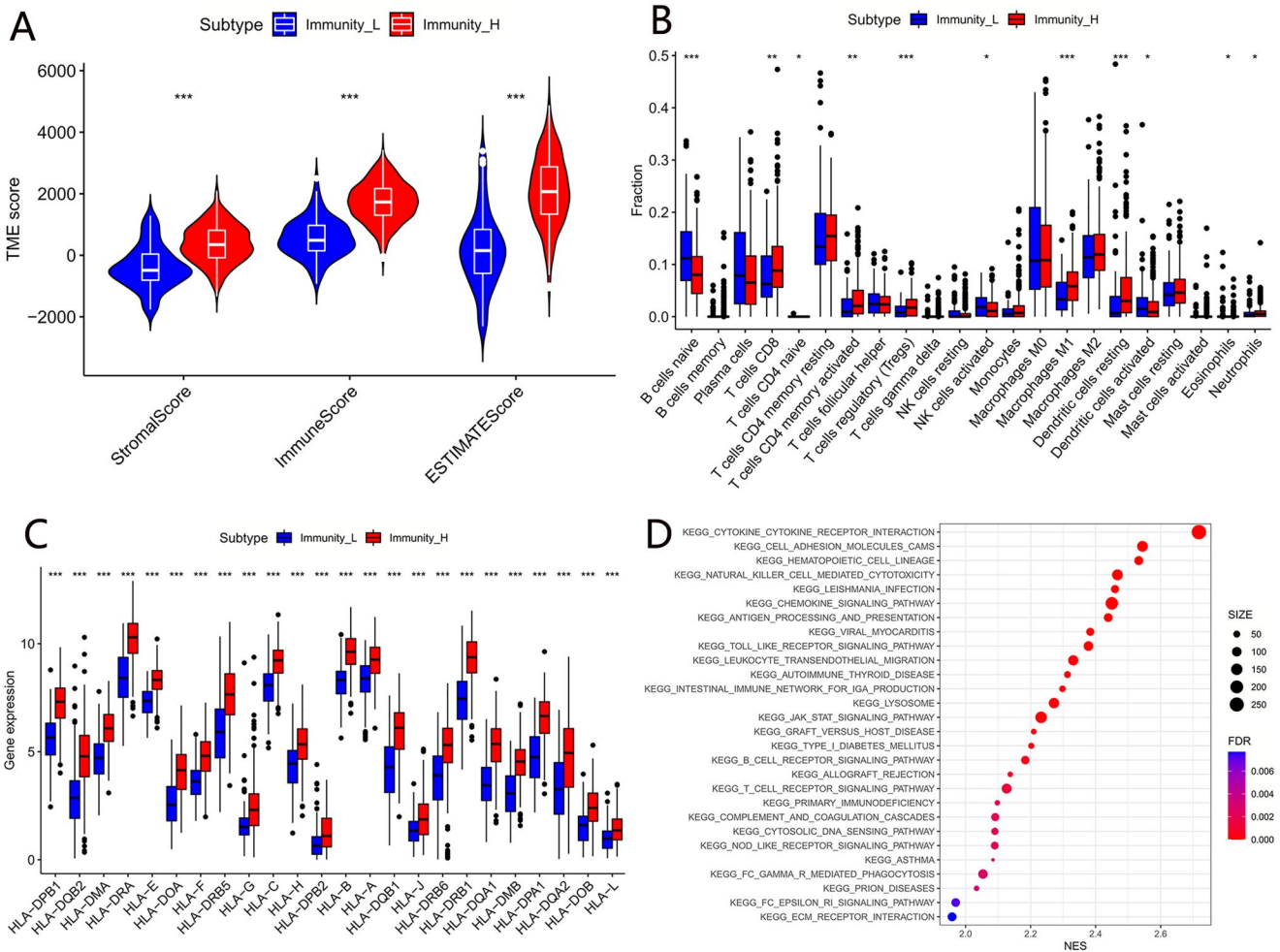

FIGURE 3 Validation of Lung adenocarcinoma (LUAD) clustering. (a) The violin plots precisely illustrated the variations in the 2 subtypes with respect to stromal cell content, immune cell infiltration level, and estimate score. (b) The variation in the percentage of each immune cell between the 2 cohorts. (c) Comparing the expression of various HLA genes in the subtypes. (d) Gene set enrichment analysis (GSEA) of the 2 subtypes. (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

training cohort consisted of TCGA specimens, whereas the test cohort consisted of GSE68465 specimens and GSE13213 specimens. Subsequently, utilising the TCGA cohort, we created a prognostic model. Premised on $p < 0.01$, univariate Cox regression analysis revealed 37 genes in the training cohort (Figure 4c). Following that, the LASSO Cox regression method was executed (Figure 4d,e). S100P, INHA, SEMA7A, INSL4, CD40LG, AGER, SERPIND1, CD1D, CX3CR1, SFTPD, and CD79A were discovered as prognose-related DEIGs, and the

mechanisms associated with these 11 genes studied are shown in Table 1. The equation below was used to compute the risk score: (S100P*0.016) + (INHA*0.024) + (SEMA7A*0.290) + (INSL4*0.082) - (CD40LG*0.051) - (AGER*0.033) - (SERPIND1*0.029) - (CD1D*0.042) - (CX3CR1*0.044) - (SFTPD*0.026) - (CD79A*0.115). TCGA-LUAD, GSE68465, and GSE13213 specimens were categorised into low- and high-risk cohorts premised on the median risk score. According to the survival analysis, low-risk individuals exhibited considerably
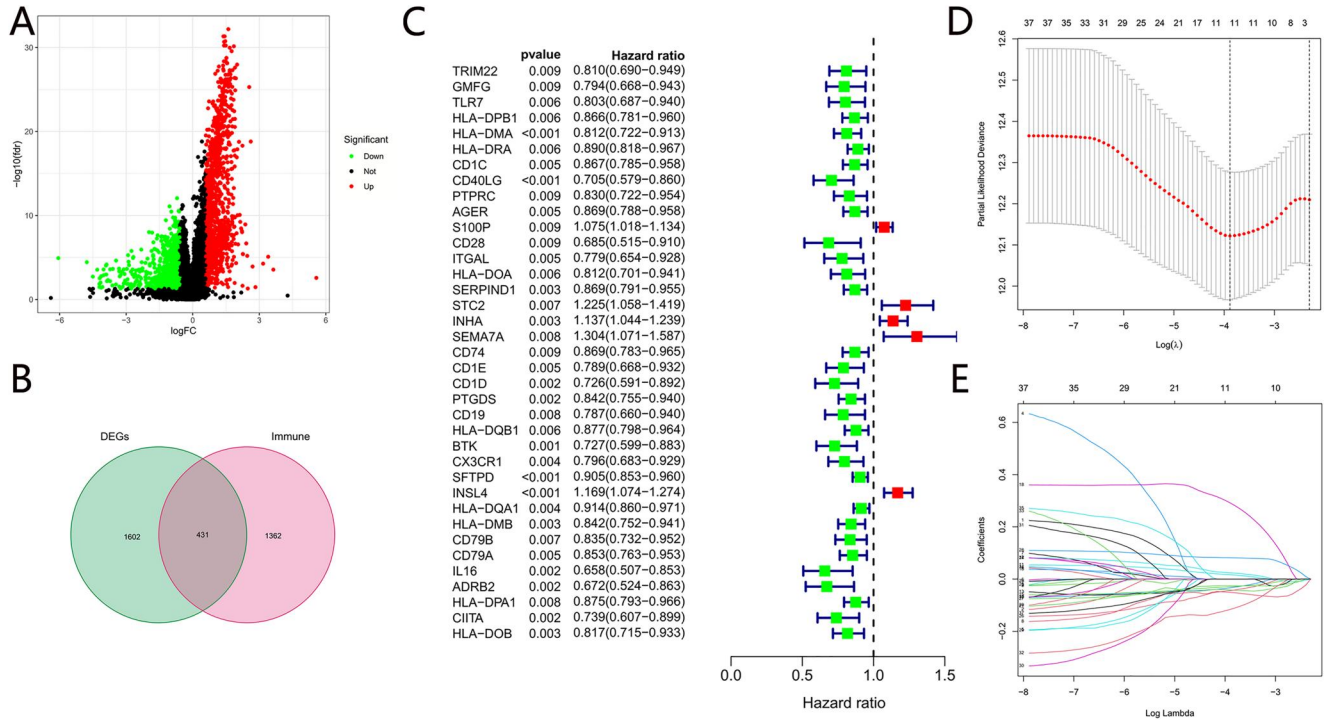

FIGURE 4 Analysis of differentially expressed immune genes (DEIGs) and prognosis model of training cohort. (a) The volcano graph depicts the distribution of differential genes between the immunity-high and immunity-low cohorts, green dots signify downregulated genes, red dots signify upregulated genes. (b) We identified 431 DEIGs by utilising the Venn diagram to identify intersection points. (c) Prognostic forest map of 37 DEIGs in TCGA. (d and e) Least absolute shrinkage and selection operator (LASSO) Cox regression analysis of TCGA cohort.

TABLE 1 Possible mechanisms for 11 genes in the immune prognostic model

Genes	Effect	State	Mechanism
S100P	Risk factor	Verified	Wnt/β-catenin [20] and PI3K/AKT signalling pathway [21].
INHA	Risk factor	Unverified	Induction of tumour angiogenesis [22].
SEMA7A	Risk factor	Unverified	Binding of β1-integrin receptor [23], induced expression of chemokines and matrix metalloproteinases [24].
INSL4	Risk factor	Verified	MAPK and AKT signalling pathways [25].
CD40LG	Protective factor	Verified	Promotion of apoptosis [26].
SFTPD	Protective factor	Verified	Oxidative stress and macrophage accumulation [27].
AGER	Protective factor	Unverified	Induction of tumour cell apoptosis [28]].
SERPIND1	Protective factor	Unverified	PI3K/AKT signalling pathway [29, 30].
CD1D	Protective factor	Unverified	Promotion of iNKT expression [31].
CX3CR1	Protective factor	Unverified	Recruitment of NK and T cells [32, 33].
CD79A	Protective factor	Unverified	Promotion of expression of IgM in B cells [34].

lengthier OS as opposed to high-risk subjects. The ROC curve analysis illustrated that the sensitivity and specificity were greatest at-risk scores of 0.720, 0.680, and 0.631 premised on 1-, 3-, and 5-year survival of the area under ROS curve (AUC) value, successively (Figure 5a). Survival analysis for the testing group also revealed that low-risk individuals exhibited considerably lengthier OS as opposed to high-risk cohort. The ROC curve analysis showed the highest sensitivity and specificity based on AUC values at 1-, 3-, and 5-year survival, with GSE68465 being 0.658, 0.628, and 0.610, and GSE132132 being 0.801, 0.7, and 0.697, respectively (Figure 5b,c). The calibration chart demonstrates that the model's efficacy was extremely good with the 45° line being the best-anticipated instance. The survival status and

risk score of TCGA-LUAD, GSE68465, and GSE13213 specimens computed utilising the prognostic model are displayed in Figure 6a–c. The survival duration for subjects in TCGA, GSE68465, and GSE13213 specimens was negatively correlated with the risk score, and their correlation coefficients were similar.

3.5 | The prognostic value of the risk subgroups in the TCGA cohort

To explore the link amongst PD-L1 (CD274)/CTLA4 and risk subgroups, we examined the differential expression of the 2
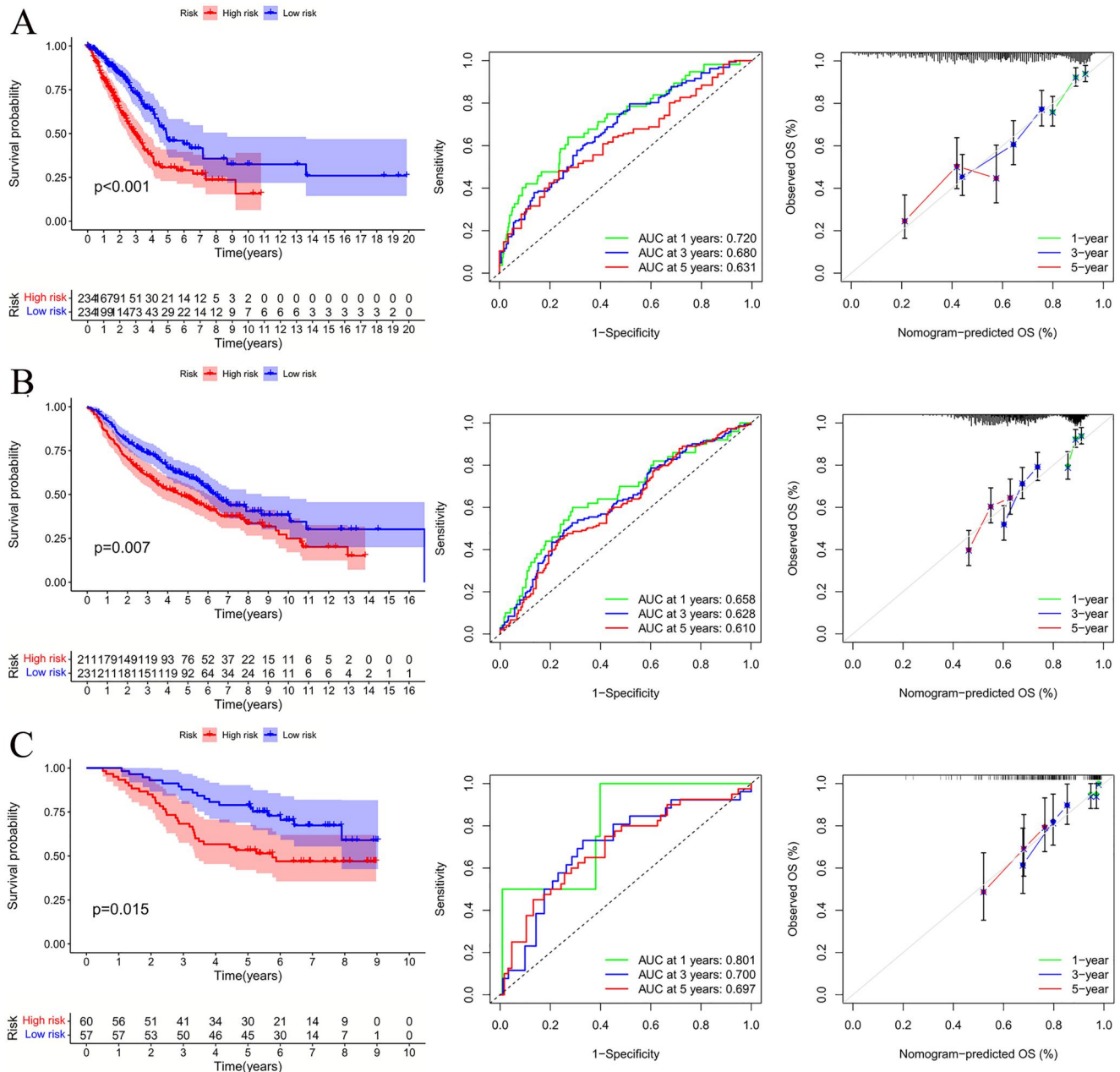

FIGURE 5 Prognosis model of TCGA-LUAD, GSE68465, and GSE13213 cohort. Survival analysis between low-risk and high-risk patients, ROC curve analysis, and calibration plots for forecasting 1-, 3-, or 5-year overall survival (OS) in the TCGA-LUAD cohort (a), the GSE68465 cohort (b), and the GSE13213 cohort (c).
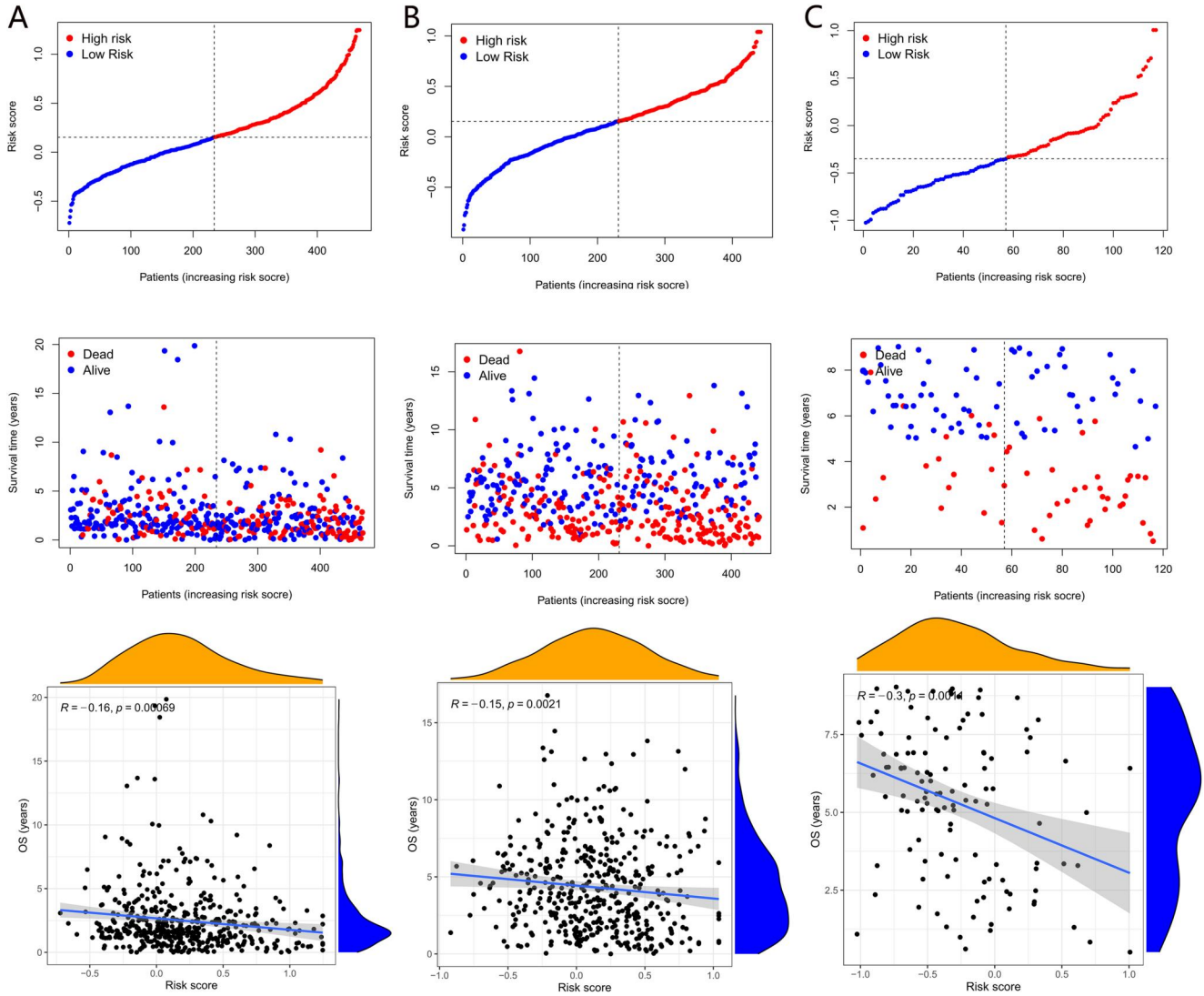
FIGURE 6 The risk score and survival time of TCGA-LUAD, GSE68465, and GSE13213 cohort. The risk score, survival status, and correlation between the risk score and overall survival (OS) in the TCGA-LUAD cohort (a), the GSE68465 cohort (b), and the GSE13213 cohort (c).

subgroups. The expression level of PD-L1/CTLA4 in the low-risk subgroup was elevated as opposed to high-risk subgroup ($p < 0.05$; Figure 7a,b). Indeed, we observed that patients in the high-risk subgroup have higher TMB than in the low-risk subgroup (Figure 7c). When the risk score was combined with the TMB level, patients with reduced TMB levels and increased immunological risk exhibited considerably shorter survival as opposed to those with elevated TMB and reduced immune risk (Figure 7d). Tumour immune dysfunction and exclusion was then utilised to examine the prospective clinical efficacy of immunotherapy in various risk categories. In our results, the high-risk subgroup exhibited a reduced TIDE score as opposed to the low-risk subgroup, implying that high-risk patients might derive more benefits from ICI treatment as opposed to low-risk patients (Figure 7e). Furthermore, we discovered that the low-risk subgroup exhibited an elevated MSI score and T cell dysfunction, whereas the high-risk subgroup exhibited an elevated T cell exclusion score (Figures 7f,h).

3.6 | Prognostic model is independent prognostic factors in patients with LUAD

To investigate if the risk score functions as an independent prognostic marker for LUAD patients, multivariate and univariate Cox regression analyses were undertaken. Following the adjustment of other clinical variables, the risk score was still found to be an independent significant predictor for LUAD prognosis [HR 3.278, 95% CI (2.199–4.887), $p < 0.001$] (Figures 8a,b).

3.7 | Predictive nomogram development and validation

Following that, we created a nomogram that combined clinical risk variables with risk subgroups to calculate the likelihood of survival in LUAD patients (Figure 8c). By adding risk
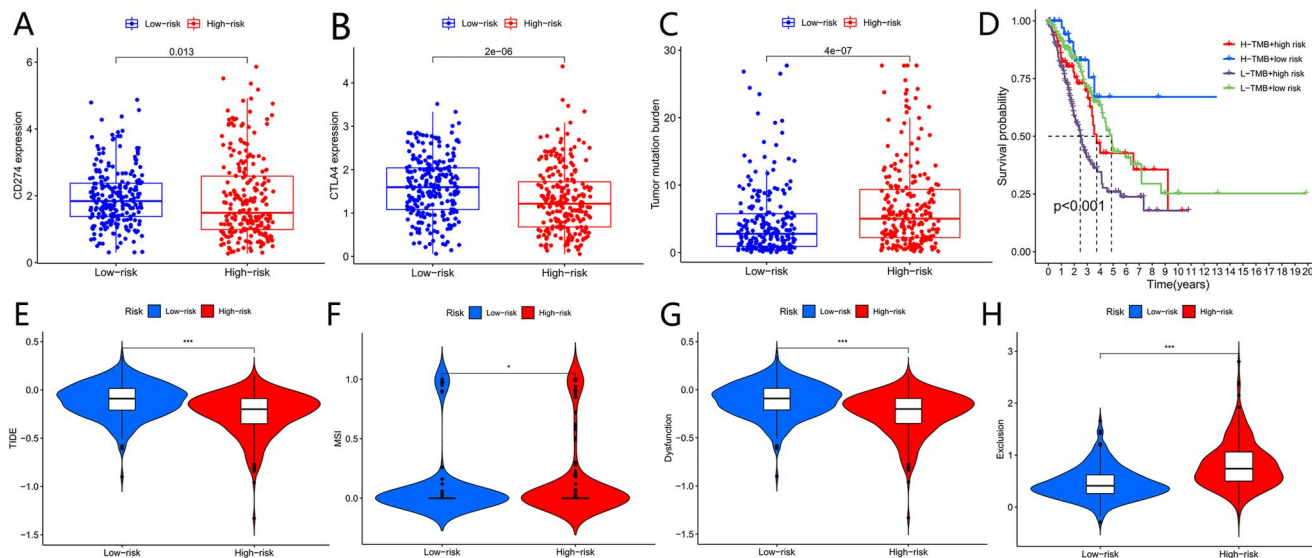
FIGURE 7 The prognostic significance of the risk subgroups in the TCGA cohort. Comparison of PD-L1 expression (a), cytotoxic T lymphocyte antigen 4 (CTLA4) expression (b), and tumour mutation burden (TMB) (c) between the high- and low-risk subgroups. (d) Kaplan-Meier (KM) curves illustrated considerably prolonged survival in patients with high TMB and low risk as opposed to patients with low TMB and high risk. Tumour immune dysfunction and exclusion (TIDE) (e), MSI (f), T cell dysfunction (g), and exclusion score (h) in the different risk subgroups. The score between the 2 subgroups was evaluated via the Wilcoxon test (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). TMB, tumour mutational burden; TIDE, tumour immune dysfunction and exclusion; MSI, microsatellite instability.

subgroups, the chance of surviving was calculated by plotting a vertical line down to the survival axis. Furthermore, the ROC curve was applied to forecast the precision of the nomogram (Figure 8d). The AUC of the nomogram for OS was 0.831, 0.866, and 0.851 at 1, 3, and 5 years, respectively. As a result, the nomogram exhibited a superior prediction value in terms of both long- and short-term survival in LUAD patients.

4 | DISCUSSION

Due to the tumour heterogeneity and complex tumourigenic process of LUAD, it is difficult to accurately predict clinical outcomes and the immune therapy response using only individual biomarkers. Therefore, this study attempts to construct a model that could both assess prognosis and predict the efficacy of immunotherapy in patients with LUAD through the lens of immune scoring and immune genes.

By intersecting the differential genes of the immune cohort with the immune genes, followed by Cox analysis and Lasso regression analysis, we identified 11 genes associated with prognosis, including S100P, INHA, SEMA7A, INSL4, CD40LG, AGER, SERPIND1, CD1D, CX3CR1, SFTPD, and CD79A. Among them, S100P, INHA, SEMA7A, and INSL4 were high-risk genes, while CD40LG, AGER, SERPIND1, CD1D, CX3CR1, SFTPD, and CD79A were low-risk genes.

As shown in Table 1, other research studies have confirmed the significance of five genes in the development of lung cancer: S100P, SERPIND1, INSL4, CD40LG, and SFTPD. In NSCLC cell lines (H1975 cells), SIX3 was found to downregulate S100P via the Wnt/β-catenin signalling pathway, thereby inhibiting cell metastasis and proliferation [20]. At the same time, studies

supported that the level of S100P mRNA was linked to the triggering status of the PI3K/AKT pathway, which is a well-recognized classical pathway that promotes migration, invasion, proliferation, and anti-medication consequence of a variety of cancers [21]. SERPIND1 can inhibit thrombin activity through interaction with heparin, also known as heparin cofactor II, and has been shown to participate in the PI3K/AKT signalling pathway and affect cell migration in both NSCLC [29] and epithelial ovarian cancer [30]. INSL4 is a member of the insulin/IGF/relaxin superfamily, and one study found that INSL4 to be an active tumour-enhancing gene in NSCLC because it stimulates the proliferation, invasion, and migration of lung cancer cells, which might be linked to improving the regulation of MAPK and AKT signalling pathways [25]. As a binding protein for the CD40L ligand of the tumour necrosis factor superfamily, CD40LG has been demonstrated to exert anti-proliferative and pro-apoptotic effects on lung cancer cells [26]. On the other hand, in mice with metastatic renal cell carcinoma, injection of CD40 in conjunction with IL-2 to promote T cell proliferation and boost dendritic cell differentiation and activity resulted in an adaptive immune response that had anticancer effects [35]. SFTPD is a pulmonary aggregate synthesised primarily by alveolar type II cells. It was found that SFTPD-negative mice showed more nitrogen oxides and macrophages in their lungs after ozone inhalation relative to SFTPD-positive mice, suggesting that SFTPD may act by enhancing innate immune defence and reducing oxidative stress [27].

At present, the role of six genes—INHA, SEMA7A, AGER, CD1D, CX3CR1, and CD79A—in lung cancer requires further validation. INHA is a protein belonging to the transforming growth factor-beta (TGF-beta) superfamily. In human ovarian
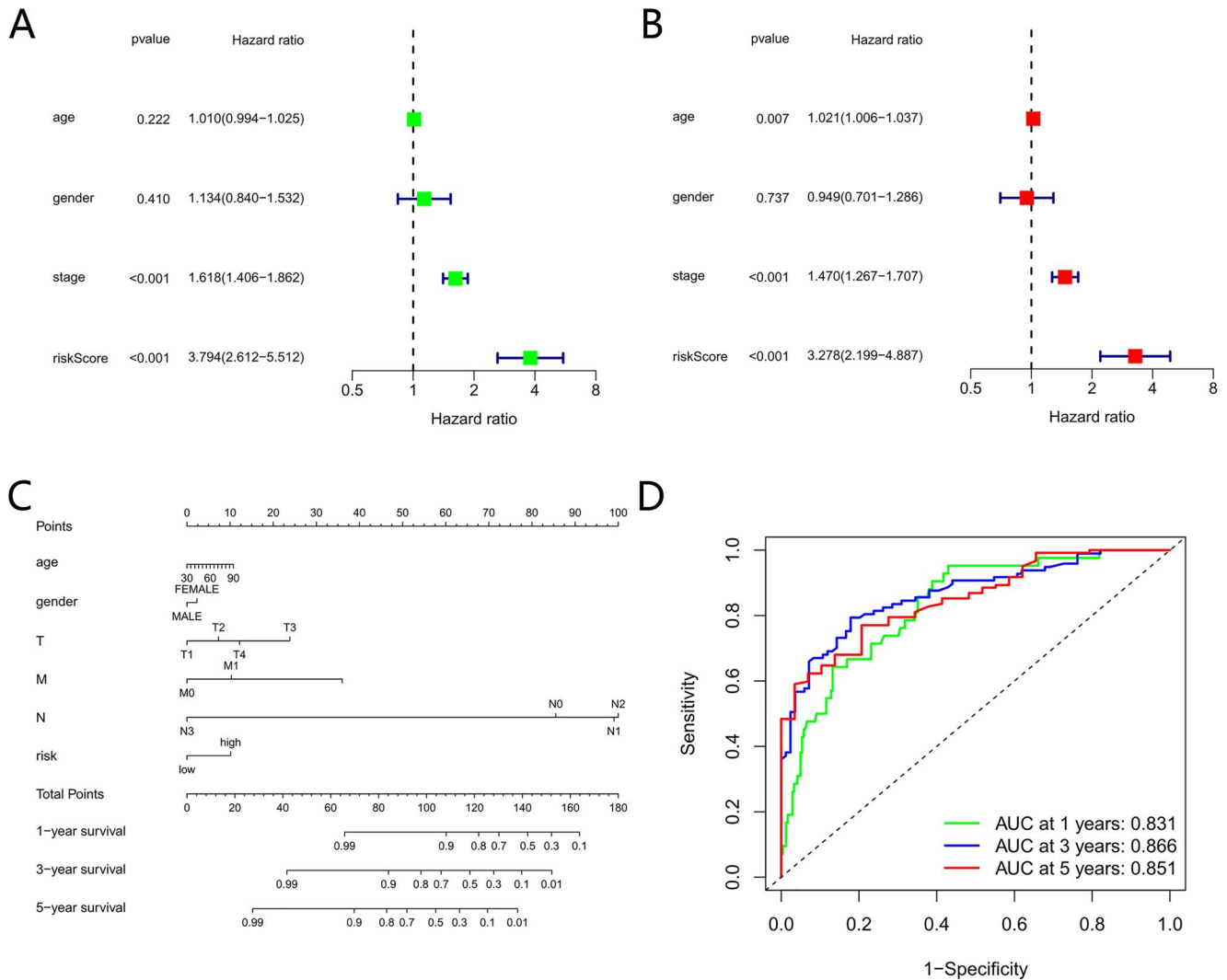
FIGURE 8 Nomogram construction and validation premised on the immune-related prognostic model. (a) Univariate and (b) multivariate Cox regression analysis of the prognostic significance of clinical variables and the risk score in TCGA cohort. (c) Nomogram combining clinical variables with the risk subgroups for forecasting 1-, 3-, and 5-year survival probability. (d) ROC curves for evaluating the predictive efficacy of the nomogram.

tissue, INHA mediates tumour angiogenesis by activating the SMAD1/5 signalling pathway and eliciting a strong paracrine response in endothelial cells [22]. Although the reports are not entirely consistent, SEMA7A has been found to facilitate metastasis in a variety of tumour types. All of the reports indicate a connection with cell adhesion mechanisms, such as activation of the epithelial-mesenchymal transition in oral cancer [36] and binding of β1-integrin in melanoma [37] and breast cancer [23]. In glioma cells, it has been demonstrated that AGER is positively linked with apoptosis, albeit the precise mechanism has not yet been thoroughly investigated [28]. Both CD1D and CX3CR1, which control NK cells, are protective factors for people with tumours since they are involved in the immune response process. [31, 32]. The homing of permetastatic T cells to CX3CL1-producing tumours is also improved by ectopic production of CX3CR1, which results in increased T cell infiltration and slowed tumour growth in tumour tissue [33]. CD79A, essential for IgM expression on the surface of human B cells, was found to enhance IgM expression by his high expression, and this facilitated expression prevented spontaneous and FasL-induced apoptosis of immune cells [34].

Overall, these 11 genes are involved in tumour cell proliferation, migration, and apoptosis as well as immune cell function, which may account for the better performance of the combined score of these 11 genes than the assessment of one immune checkpoint gene alone.

Although the discovery and application of ICI led to an unprecedented improvement in clinical response and OS, a great number of LUAD patients did not react to these treatment monoclonal antibodies targeting immunological checkpoints [38]. Contrasted with other biomarkers, the TIDE score has recently been found to be a better predictor of tumour response to immunotherapy [39]. The TIDE prediction score correlated with T cell dysfunction in tumours with elevated infiltration of cytotoxic T lymphocytes (CTL) and the inhibition of T cell infiltration in tumours having reduced levels of

CTL thus represent two different mechanisms of immune escape [40]. In this research, we discovered that the levels of CTLA4 and PD-L1 expression were more elevated and TMB was reduced in the low-risk cohort. But interestingly, contrasted to the high-risk LUAD patients, the low-risk patients exhibited elevated MSI scores, T cell dysfunction score, and TIDE score, and lower T cell exclusion score. The elevated TIDE score, the greater possibility of tumour immune escape, and the lower possibility of anti-PD-1/CTLA4 benefit, which additionally explains why some patients exhibiting elevated expression of PD-L1 or CTLA4 have poor treatment responses for ICI. Moreover, their lower ICI response might be due to immune evasion via T cell dysfunction. However, the TIDE score focussed on patient response to immunotherapy rather than patient survival time, and life expectancy was also important in making treatment decisions. In our study, the predictive value of signature was comparable with TIDE and signature might be a better predictor of OS at longer follow-up.

This research had some shortcomings that ought to be acknowledged. First, our prognostic signature's performance ought to be verified in larger LUAD datasets. Second, all of the findings were founded on publicly available datasets and ought to be validated by other actual trials.

5 | CONCLUSION

In summary, using the ssGSEA algorithm and an immunogenomics perspective, we discovered a unique predictive signature for LUAD that consists of 11 immune genes, including S100P, INHA, SEMA7A, INSL4, CD40LG, AGER, SERPIND1, CD1D, CX3CR1, SFTPD, and CD79A. The model was validated on three datasets. Furthermore, the model was able to predict prognosis and immunotherapy sensitivity in LUAD patients, which may be helpful for personalised counselling.

AUTHOR CONTRIBUTIONS
Zehuai Guo, Xiangjun Qi, and Yang Cao conceptualized the study; Zehuai Guo and Zeyun Li drafted the manuscript; Jianying Yang, Zhe Sun, Peiqin Li, and Ming Chen collected the data; Zehuai Guo, Xiangjun Qi, Zeyun Li, and Jianying Yang performed all data analysis; Yang Cao supervised the study, and Yang Cao was responsible for project administration. All authors reviewed and approved the final manuscript.

CONFLICTS OF INTEREST
The authors declare no conflicts of interest regarding the publication of this work.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are openly available in NCBI Gene Expression Omnibus at https://www.ncbi.nlm.nih.gov/geo/, The Cancer Genome Atlas Program at https://portal.gdc.cancer.gov/repository, and ImmPort Portal database at https://www.immport.org/.

CONSENT FOR PUBLICATION
All authors granted consent to publish.

ORCID
Zehuai Guo https://orcid.org/0000-0002-7085-9502

REFERENCES
1. Sung, H., et al.: Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA A Cancer J. Clin. 71(3), 209–249 (2021). https://doi.org/10.3322/caac.21660
2. Melocchi, V., et al.: Aggressive early-stage lung adenocarcinoma is characterized by epithelial cell plasticity with acquirement of stem-like traits and immune evasion phenotype. Oncogene 40(31), 4980–4991 (2021). https://doi.org/10.1038/s41388-021-01909-z
3. Finn, O.J.: Immuno-oncology: understanding the function and dysfunction of the immune system in cancer. Ann. Oncol. 23((Suppl 8)), viii6–9 (2012). https://doi.org/10.1093/annonc/mds256
4. Haas, O.A.: Primary immunodeficiency and cancer predisposition revisited: embedding two closely related concepts into an integrative conceptual framework. Front. Immunol. 9, 3136 (2018). https://doi.org/10.3389/fimmu.2018.03136
5. Doroshow, D.B., et al.: Immunotherapy in non-small cell lung cancer: facts and hopes. Clin. Cancer Res. 25(15), 4592–4602 (2019). https://doi.org/10.1158/1078-0432.ccr-18-1538
6. Rolfo, C., et al.: Immunotherapy in NSCLC: a promising and revolutionary weapon. Adv. Exp. Med. Biol. 995, 97–125 (2017)
7. Herbst, R.S., et al.: Atezolizumab for first-line treatment of PD-L1-selected patients with NSCLC. N. Engl. J. Med. 383(14), 1328–1339 (2020). https://doi.org/10.1056/nejmoa1917346
8. Lei, X., et al.: Immune cells within the tumor microenvironment: biological functions and roles in cancer immunotherapy. Cancer Lett. 470, 126–133 (2020). https://doi.org/10.1016/j.canlet.2019.11.009
9. Kalbasi, A., Ribas, A.: Tumour-intrinsic resistance to immune checkpoint blockade. Nat. Rev. Immunol. 20(1), 25–39 (2020). https://doi.org/10.1038/s41577-019-0218-4
10. Gandhi, L., et al.: Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. N. Engl. J. Med. 378(22), 2078–2092 (2018). https://doi.org/10.1056/nejmoa1801005
11. Otter, S.J., et al.: The role of biomarkers for the prediction of response to checkpoint immunotherapy and the rationale for the use of checkpoint immunotherapy in cervical cancer. Clin. Oncol. 31(12), 834–843 (2019). https://doi.org/10.1016/j.clon.2019.07.003
12. Zhang, Y., et al.: Multi-omics data analyses construct TME and identify the immune-related prognosis signatures in human LUAD. Mol. Ther. Nucleic Acids 21, 860–873 (2020). https://doi.org/10.1016/j.omtn.2020.07.024
13. Ling, B., et al.: Microenvironment analysis of prognosis and molecular signature of immune-related genes in lung adenocarcinoma. Oncol. Res. 28(6), 561–578 (2021). https://doi.org/10.3727/096504020x15907428281601
14. Wang, L., et al.: A gene expression-based immune signature for lung adenocarcinoma prognosis. Cancer Immunol. Immunother. 69(9), 1881–1890 (2020). https://doi.org/10.1007/s00262-020-02595-8
15. Guo, D., et al.: A new immune signature for survival prediction and immune checkpoint molecules in lung adenocarcinoma. J. Transl. Med. 18(1), 123 (2020). https://doi.org/10.1186/s12967-020-02286-z

16. Song, C., et al.: A prognostic nomogram combining immune-related gene signature and clinical factors predicts survival in patients with lung adenocarcinoma. Front. Oncol. 10, 1300 (2020). https://doi.org/10.3389/fonc.2020.01300

17. Jin, Y., et al.: Identification of novel subtypes based on ssGSEA in immune-related prognostic signature for tongue squamous cell carcinoma. Cancer Med. 10(23), 8693–8707 (2021). https://doi.org/10.1002/cam4.4341

18. Shen, S., et al.: Development and validation of an immune gene-set based Prognostic signature in ovarian cancer. EBioMedicine 40, 318–326 (2019). https://doi.org/10.1016/j.ebiom.2018.12.054

19. Xiao, B., et al.: Identification and verification of immune-related gene prognostic signature based on ssGSEA for osteosarcoma. Front. Oncol. 10, 607622 (2020). https://doi.org/10.3389/fonc.2020.607622

20. Liu, S., et al.: TRIM27 acts as an oncogene and regulates cell proliferation and metastasis in non-small cell lung cancer through SIX3-β-catenin signaling. Aging 12(24), 25564–25580 (2020). https://doi.org/10.18632/aging.104163

21. De Marco, C., et al.: Specific gene expression signatures induced by the multiple oncogenic alterations that occur within the PTEN/PI3K/AKT pathway in lung cancer. PLoS One 12(6), e0178865 (2017). https://doi.org/10.1371/journal.pone.0178865

22. Singh, P., et al.: Inhibin is a novel paracrine factor for tumor angiogenesis and metastasis. Cancer Res. 78(11), 2978–2989 (2018). https://doi.org/10.1158/0008-5472.can-17-2316

23. Black, S.A., et al.: Semaphorin 7a exerts pleiotropic effects to promote breast tumor progression. Oncogene 35(39), 5170–5178 (2016). https://doi.org/10.1038/onc.2016.49

24. Song, Y., et al.: The involvement of semaphorin 7A in tumorigenic and immunoinflammatory regulation. J. Cell. Physiol. 236(9), 6235–6248 (2021). Portico. https://doi.org/10.1002/jcp.30340

25. Scopetti, D., et al.: INSL4 as prognostic marker for proliferation and invasiveness in Non-Small-Cell Lung Cancer. J. Cancer 12(13), 3781–3795 (2021). https://doi.org/10.7150/jca.51332

26. Xu, W., et al.: Anti-tumor activity of gene transfer of the membrane-stable CD40L mutant into lung cancer cells. Int. J. Oncol. 37(4), 935–941 (2010). https://doi.org/10.3892/ijo_00000744

27. Groves, A.M., et al.: Prolonged injury and altered lung function after ozone inhalation in mice with chronic lung inflammation. Am. J. Respir. Cell Mol. Biol. 47(6), 776–783 (2012). https://doi.org/10.1165/rcmb.2011-0433oc

28. Jandial, R., et al.: Inhibition of GLO1 in glioblastoma multiforme increases DNA-AGEs, stimulates RAGE expression, and inhibits brain tumor growth in orthotopic mouse models. Int. J. Mol. Sci. 19(2), 406 (2018). https://doi.org/10.3390/ijms19020406

29. Liao, W.Y., et al.: Heparin co-factor II enhances cell motility and promotes metastasis in non-small cell lung cancer. J. Pathol. 235(1), 50–64 (2015). https://doi.org/10.1002/path.4421

30. Guo, Q., et al.: SERPIND1 affects the malignant biological behavior of epithelial ovarian cancer via the PI3K/AKT pathway: a mechanistic study. Front. Oncol. 9, 954 (2019). https://doi.org/10.3389/fonc.2019.00954

31. Hix, L.M., et al.: CD1d-expressing breast cancer cells modulate NKT cell-mediated antitumor immunity in a murine model of breast cancer metastasis. PLoS One 6(6), e20702 (2011). https://doi.org/10.1371/journal.pone.0020702

32. Lavergne, E., et al.: Fractalkine mediates natural killer-dependent anti-tumor responses in vivo. Cancer Res. 63(21), 7468–7474 (2003)

33. Siddiqui, I., et al.: Enhanced recruitment of genetically modified CX3CR1-positive human T cells into Fractalkine/CX3CL1 expressing tumors: importance of the chemokine gradient. J. Immunother. Cancer 4(1), 21 (2016). https://doi.org/10.1186/s40425-016-0125-1

34. Huse, K., et al.: Mechanism of CD79A and CD79B support for IgM+ B cell fitness through B cell receptor surface expression. J. Immunol. 209(10), 2042–2053 (2022). https://doi.org/10.4049/jimmunol.2200144

35. Murphy, W.J., et al.: Synergistic anti-tumor responses after administration of agonistic antibodies to CD40 and IL-2: coordination of dendritic and CD8+ cell responses. J. Immunol. 170(5), 2727–2733 (2003). https://doi.org/10.4049/jimmunol.170.5.2727

36. Liu, T.J., et al.: Semaphorin-7A contributes to growth, migration and invasion of oral tongue squamous cell carcinoma through TGF-β-mediated EMT signaling pathway. Eur. Rev. Med. Pharmacol. Sci. 22(4), 1035–1043 (2018)

37. Lazova, R., et al.: The semaphorin 7A receptor Plexin C1 is lost during melanoma metastasis. Am. J. Dermatopathol. 31(2), 177–181 (2009). https://doi.org/10.1097/dad.0b013e318196672d

38. Mazieres, J., et al.: Immune checkpoint inhibitors for patients with advanced lung cancer and oncogenic driver alterations: results from the IMMUNOTARGET registry. Ann. Oncol. 30(8), 1321–1328 (2019). https://doi.org/10.1093/annonc/mdz167

39. Pallocca, M., et al.: Combinations of immuno-checkpoint inhibitors predictive biomarkers only marginally improve their individual accuracy. J. Transl. Med. 17(1), 131 (2019). https://doi.org/10.1186/s12967-019-1865-8

40. Jiang, P., et al.: Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat. Med. 24(10), 1550–1558 (2018). https://doi.org/10.1038/s41591-018-0136-1

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.