Contents lists available at ScienceDirect

Translational Oncology

Original Research

Typical tumor immune microenvironment status determine prognosis in lung adenocarcinoma

Caixia Ren [a],[#], Jinyu Li [b],[#], Yang Zhou [c], Shuyu Zhang [d], Qi Wang [a],[1],[*]

[a] Department of Respiratory Medicine, The Second Hospital of Dalian Medical University, Dalian, 116023, China
[b] Department of Breast Oncology, The Second Hospital of Dalian Medical University, Dalian, Liaoning, 116023, China
[c] Liaoning Clinical Research Center for Lung Cancer, The Second Hospital of Dalian Medical University, Dalian, 116023, China
[d] China National Nuclear Corporation 416 Hospital, The Second Affiliated Hospital of Chengdu Medical College, Chengdu 610051, China

ARTICLE INFO

ABSTRACT

Background: Immune cells, vital components of tumor microenvironment, regulate tumor survival and progression. Lung adenocarcinoma (LUAD), the tumor with the highest mortality rate worldwide, reconstitutes tumor immune microenvironment (TIME) to avoid immune destruction. Data have shown that TIME influences LUAD prognosis and predicts immunotherapeutic efficacy. The related information about the role of TIME's characteristics in LUAD is limited.
Methods: We performed unsupervised consensus clustering via machine-learning techniques to identify TIME clusters among 1906 patients and gathered survival data. The characteristics of TIME clusters of LUAD were visualized by multi-omics analysis, pseudo-time dynamic analysis, and enrichment analysis. TIME score model was constructed by principal component analysis. Comprehensive analysis and validation were conducted to test the prognostic efficacy and immunotherapeutic response of TIME score.
Results: TIME clusters (A, B and C) were constructed and exhibited different immune infiltration states. Multi-omics analyses included significant mutated genes (SMG), copy number variation (CNV) and cancer stemness that were significantly different among the three clusters. TIME cluster A had a lower SMG, lower CNV, and lower stemness but a higher immune infiltration level compared to TIME clusters B and C. TIME score showed that patients in low TIME score group had higher overall survival rates, higher immune infiltration level and high expression of immune checkpoints. In validation cohorts, low TIME score subgroup had better drug sensitivity and favorable immunotherapeutic response.
Conclusion: We constructed a stable model of LUAD immune microenvironment characteristics that may improve the prognostic accuracy of patients, provide improved explanations of LUAD responses to immunotherapy, and provide new strategies for LUAD treatment.

Introduction

Lung adenocarcinoma (LUAD), a common lung cancer, has the highest cancer mortality rate worldwide [1]. immunotherapy (IT) as a notable therapy has achieved clinical success over a short period of time in NSCLC [2]. Clinical trials have confirmed improvements in clinical outcomes and the safety of IT [3–5]. IT acts at key points in immune responses to accurately target tumor cells. Compared to classic chemotherapy, IT has a higher overall survival rate (OS), particularly for patients expressing high levels of PD-L1 [6,7]. Though IT has improved the health of many patients, it lacks a sensitive predictive indicator and is limited to certain qualified patients. PD-L1 expression is regarded as a biomarker for assessing the success of IT, but only half of the high PD-L1-expressing patients benefit from related IT drugs [8]. The differences among the diagnostic assays for IT limit their applications in LUAD [9]. The predictive value of immune checkpoint in immunotherapy is not as good as molecular target in targeted therapy [10–12]. The higher somatic tumor mutational burden (TMB) is believed to be

* Corresponding author.
    E-mail addresses: cxiar0306@163.com (C. Ren), lijinyu1012@qq.com (J. Li), zhouyang610@163.com (Y. Zhou), zhang.shuyu@hotmail.com (S. Zhang), wqdlmu@163.com (Q. Wang).
[#] These authors have contributed equally to this work.
[1] ORCID: https://orcid.org/0000-0003-2427-7234

associated with increased OS in IT, but there is no accepted definition of high TMB [13].

The tumor immune microenvironment (TIME) influences cancer progression and metastases and affects patient prognoses and IT outcomes [14,15]. Immune cell infiltration therapy in IT activates tumor-infiltrating lymphocytes (TLCs). Natural killer (NK) cells, dendritic cells (DCs), neutrophils, and tumor-associated macrophages (TAMs) are correlated with patient OS and relapse [16–20]. Most studies have focused on only one aspect of TIME because of limited experimental options for investigating the effect of tumor immunity in the incidence and progression of cancer. Investigating key nodes in TIME, rather than focusing on a certain aspect, is necessary.

Next-generation sequencing and machine learning are possible options for the analysis of large amounts of data from multi-center studies. We constructed a TIME score method to identify a prognostic TIME score biomarker for LUAD. Verification of the method indicated that the model could accurately predict patient responses to IT and patient outcomes. Our model provides novel insights into the TIME of LUAD patients, improves patient prognoses, and provides guidance for selecting clinical LUAD treatments.

Materials and methods

Adenocarcinoma of lung datasets collection and preprocessing

The Gene-Expression Omnibus (GEO) and the Cancer Genome Atlas (TCGA) databases were searched for LUAD gene-expression datasets reporting complete clinical annotations. 11 treatment-naive LUAD cohorts (GSE3141, GSE10245, GSE19188, GSE30219, GSE31210, GSE37745, GSE41271, GSE42127, GSE50081, GSE72094, and TCGA-LUAD) were selected. After removing cases lacking survival information, samples from 1906 patients were investigated. The backgrounds of raw data from Affymetrix and Illumina analyses of the GEO datasets were adjusted using affy and lumi software packages. For the TCGA-LUAD cohort, transcripts per kilobase million (TPM) of gene expression from mRNA sequencing data, somatic mutation profiling data, and copy number variation (CNV) data from the Genomic Data Commons (GDC) datasets were obtained using TCGAbiolinks [21]. Non-biological technical biases causing batch effects were corrected using ComBat [22]. Platforms, number of samples, baseline information, and clinical endpoints of each eligible LUAD dataset were summarized (Table S1).

Consensus clustering of TIME-infiltrating cells

Infiltration levels of distinct immune cells [23], including innate and adaptive immunity cells in each LUAD sample were quantified using single-sample gene-set enrichment analysis (ssGSEA). We used ESTIMATE [24] to calculate immune scores, stromal scores, and tumor purities that determined levels of TIME immune or matrix components. Unsupervised clustering PAM algorithms identified TIME patterns and classified patients using ConsensuClusterPlus [25]. The stability of the classification of patients was ensured by repeating ConsusClusterPlus 1000 times.

Identification of DEGs among TIME clusters

Analyses of differentially expressed genes (DEGs) in TIME clusters were performed using DESeq2 [26] and limma [27] with standard comparison parameters. Invalid genes were removed. A threshold of absolute log2 fold change > 1 and a significance threshold of adjusted $P < 0.05$ were chosen as criteria for valid DEGs. A total of 1412 qualifying DEGs were identified.

Dynamic analysis of immune landscape and multi-omics analysis of clusters

Pseudo-time dynamic analyses using the Monocle2 package [28] were performed to show the intrinsic changes and the distribution of individual patients. Immune processes were simulated in tree structures, and developmental trajectories were defined [29]. MutSigCV (v1.4) [30] was used to identify SMGs using default parameters, and ComplexHeatmap [31] was used to draw oncoprint mutation landscapes. GISTIC 2.0 [32] was used to identify significant regions that were amplified or deleted chromosomes. Gene copy number loss or gain burden was calculated as the total number of genes with copy number changes at the focal and arm levels. We used OCLR (a machine learning algorithm) [33] to calculate cancer stemness data. The mRNAsi and mDNAsi were all inferred.

Generation of and dimension reduction in TIME Score

Positively- and negatively-correlated DEG signatures in TIME gene clusters were classified as TIME gene signature A or B. The Boruta algorithm [34,35] was applied to TIME gene signatures to reduce noise, eliminate redundant genes, and connect TIME scores to the largest groups of well-and anti-correlated genes in the set. Principal component analysis (PCA) extracted principal component 1 as the signature score. The TIME score of each patient was defined using the following equation [36,37]:

$$\text{TIME score} = \sum \text{PC1A} + \sum \text{PC1B}$$

where A is the expression of TIME gene signatures A, and B is the expression of TIME gene signatures B.

Functional and pathway enrichment analyses

The GSVA R package [38] was used for GSEA of biological processes. The clusterProfiler R package [39] was used to collect information for the analyzed gene sets from GO and KEGG in the Molecular Signatures Database (MSigDB). The ssGSEA algorithm and TIP tool [40] were used to quantify the scores of immune-related pathways and cancer immunity cycle and explore correlations between the TIME scores and immune-relevant biological processes [41].

Evaluation of immune infiltration in TCGA-LUAD and tissue microarray specimens

Approximately 200 samples from different TIME score subgroups were randomly selected for immune infiltration evaluation. Two experienced pathologists examined formalin-fixed paraffin-embedded slides of the tissues provided by the TCGA data portal. Hematoxylin eosin-stained slides were evaluated and scored according to the amount of immune infiltration: 0 = absence of immune cell infiltration, 1 = minimal, 2 = mild infiltration, 3 = moderate, and 4 = severe infiltration. Samples with infiltration scores from 0 to 2 were classified as low immune infiltration, and samples with infiltration scores from 3 to 4 were classified as high immune infiltration. Poor-quality samples from tissue microarrays (TMA) (HLugA180Su04) (Outdo Biotech, Shanghai, China) were eliminated. TMAs were identified and grouped by the same pathologists.

Sample collection

From March 2014 to December 2016, 66 specimens from pre-operative LUAD patients at the Second Hospital of Dalian Medical University (SHDMU), Dalian, China, were collected and provided to us with complete clinical and prognostic information. LUAD histopathology was confirmed by two experienced pathologists. Follow-up survival

information was obtained through electronic medical record and telephone. Twenty samples were randomly selected for extracting total RNA according to the manufacturers' instructions. TransStart Tip Green qPCR SuperMix (Transgen Biotech) was utilized for real-time qRT-PCR with specific primers against CD44, CD133, ALDH1A1, PD-L1, and GAPDH, using the ABI-7900HT FAST Real-Time PCR System (Applied Biosystems, USA). GAPDH was selected for normalization. All primer sequences were shown in Table S1. This study was approved by the Institutional Ethics Committee at SHMDU and all patients signed informed consent. All the studies were designed and all test procedures were performed according to the Helsinki Declaration.

Comprehensive prediction of the TIME score

LUAD samples were categorized into high and low TIME score subgroups. The predictive accuracy of TIME scores was validated in 9 validation cohorts, including IT-related cohorts (IMvigor210, $n = 298$ [42]; TCGA-SKCM, $n = 70$ [43]; GSE135222, $n = 27$ [44]; PMID29301960, $n = 33$ [45]; and GSE91061, $n = 49$ [46], large LUAD validation cohorts (GSE72094, $n = 439$ and GSE68465, $n = 440$), TMA cohort, and SHDMU cohort. We curated gene expression profiles from pre-therapy biopsy samples and transformed them into a TPM format for analysis. The Subclass Mapping (SubMap) algorithm [47] was used to predict the possibility of anti-PD1 and anti-CTLA4 IT responses in the high and low TIME score subgroups. We obtained related annotation data from the supplementary materials reported by Lu et al. [48]. The pRRophetic R package [49] was used to estimate the therapeutic response of dimethyloxallyl glycine (DMOG) and erlotinib. The half-maximal inhibitory concentrations (IC50s) of samples were estimated using ridge regression, and the accuracy of the IC50 predictions was evaluated by 10-fold cross-validation based on the Genomics of Drug Sensitivity in Cancer (GDSC) training set [50].

Statistical analyses

All statistical analyses were conducted using R software (version 4.0.5). We used the Wilcoxon test to compare two groups and the Kruskal-Wallis test to compare more than two groups. Spearman and distance correlation analyses were used to calculate correlation coefficients. Contingency tables were analyzed using two-sided Fisher exact tests. Optimal cut-off values determined by the survminer R package were used to dichotomize continuous variables for patient survival. Survival curves for subgroups in each data set were generated using the Kaplan–Meier survival curves and significant differences were identified using log-rank tests. Hazard ratios (HRs) of the TIME scores were calculated using a univariate Cox regression model and independent prognosis factors were determined using a multivariate Cox regression model. In the validation cohort, the pROC package [51] was used on the training set to generate receiver operating characteristic (ROC) curves, calculate the area under the curve (AUC), and calculate confidence intervals to determine the diagnostic accuracy of the combination of TIME score and clinical features. The statistical significance threshold was set to 0.05 (two-tailed).

Results

Immune cell infiltration landscape in the TIME of LUAD

The flowchart of TIME cluster division and immune score construction is shown in Fig. S1A. 11 lung cancer cohorts (GEO databases: GSE3141, GSE10245, GSE19188, GSE30219, GSE31210, GSE37745, GSE50081, GSE72094, GSE41271 and GSE42127; [TCGA]-LUAD) was selected and strictly removed non-lung adenocarcinoma patients. Combined with survival information, a total of 1906 samples were analyzed by ssGSEA to identify and quantify the infiltrating level of immune cells after removing batch effects (Fig. S1B, Table S2, S3).

Unsupervised clustering was performed to divide LUAD patients with matched immune cells expression into three TIME cell clusters (Fig. S1C). We termed these clusters TIME cell cluster A, TIME cell cluster B and TIME cell cluster C. The three TIME cell clusters show distinct immune infiltration statuses, which are accompanied by differences in survival (log-rank test, $P<$ 0.001, Fig. 1A-C). Immune cell compositions were compared to assess the different prognoses among the three TIME cell clusters (Fig. 1E, Table S5). Patients in TIME cell cluster A were characterized by high expression of innate immune cells (DCs, immature DCs, activated DCs, TAMs, neutrophils, and NK CD56dim cells) and adaptive immune cells (Th1 cells, B cells, CD8$^+$T cells, T gamma delta cells, T effector memory cells, and T central memory cells). The median survival times of immune and adaptive immune cells for cell clusters A, B, and C were 3204 days, 2462 days, and 1380 days, respectively. The correlation coefficient heatmap illustrates the interaction among the immune cells in TIME (Fig. 1D, Table S4). Two central immune checkpoint proteins, PD-1 and PD-L1, were analyzed in each TIME cell cluster (Fig. 1F). The expression levels of these two immune checkpoint proteins were high in TIME cluster A and low in TIME cluster C, indicating that TIME cluster A may benefit the most from anti-PD-1 and anti-PD-L1 IT treatments. To deepen the biological and clinical differences among the three intrinsic cell clusters, we performed identical analyses in the TCGA-LUAD cohort for its large sample size, detailed multi-omics data and exhaustive clinical information. Survival analyses, immune cell composition analyses, and expression levels of PD-1 and PD-L1 among the three cell clusters were consistent with the overall trend presented above (Fig. S1D-F, Table S5).

Construction of TIME gene cluster and enrichment analysis

To clarify the underlying biological characteristics of TIME cell clusters, difference analysis identified 1412 DEGs in the TCGA-LUAD cohort for further analysis (Fig. S2A and Table S6). Unsupervised clustering identified three distinct TIME genomic clusters: TIME gene cluster A, TIME gene cluster B, and TIME gene cluster C (Fig. 2A, S2B).

The TIME gene clusters and TIME cell clusters can match well (Fig. S2C). Analysis of Kaplan-Meier curves revealed that survival times of patients in TIME gene cluster A were longer than those of patients in gene clusters B and C (Fig. 2B). A pseudo-time analysis demonstrated that the classification of gene clusters was stable (Fig. 2C). Further analysis found that clusters A, B, and C may show progressive trends in cell differentiation, which is also consistent with tumor progression [52] (Fig. S2D). Immune cell compositions showed the same trend in gene clusters and cell clusters, but the differences among gene clusters were more evident than that of cell clusters (Fig. 2D, Table S5). Estimate analysis supported the results that immune, stromal and estimate scores decreased with the classification, tumor purity increased with the classification, consistent with the deterioration of the corresponding patient's prognosis (Fig. 2E, Table S7).

Using GSEA (which included GO process and KEGG pathway identification), we investigated the biological behaviors behind differences among the three gene clusters (Fig. 2F, G, Table S8). In GO process analysis, TIME gene cluster A was mainly enriched in immune-activated processes, while TIME gene cluster C was enriched primarily on malignancy-associated biological processes. In KEGG pathway identification, TIME gene cluster A was enriched in immune activation related pathways such as B and T cell receptor signaling, and Fc gamma R-mediated phagocytosis pathways. TIME gene cluster C was enriched in base excision repair, DNA packaging complex, chromosomal region, kinetochore processes, cell cycle, oocyte meiosis, oxidative phosphorylation, pyrimidine metabolism and nucleotide excision repair pathways. GSEA results in gene cluster B revealed fewer biological processes and pathways associated with immune activation and malignancy than those for gene cluster A and gene cluster C. (Fig. S2E, Table S8). The GSEA results showed significant differences between cluster A and cluster C,
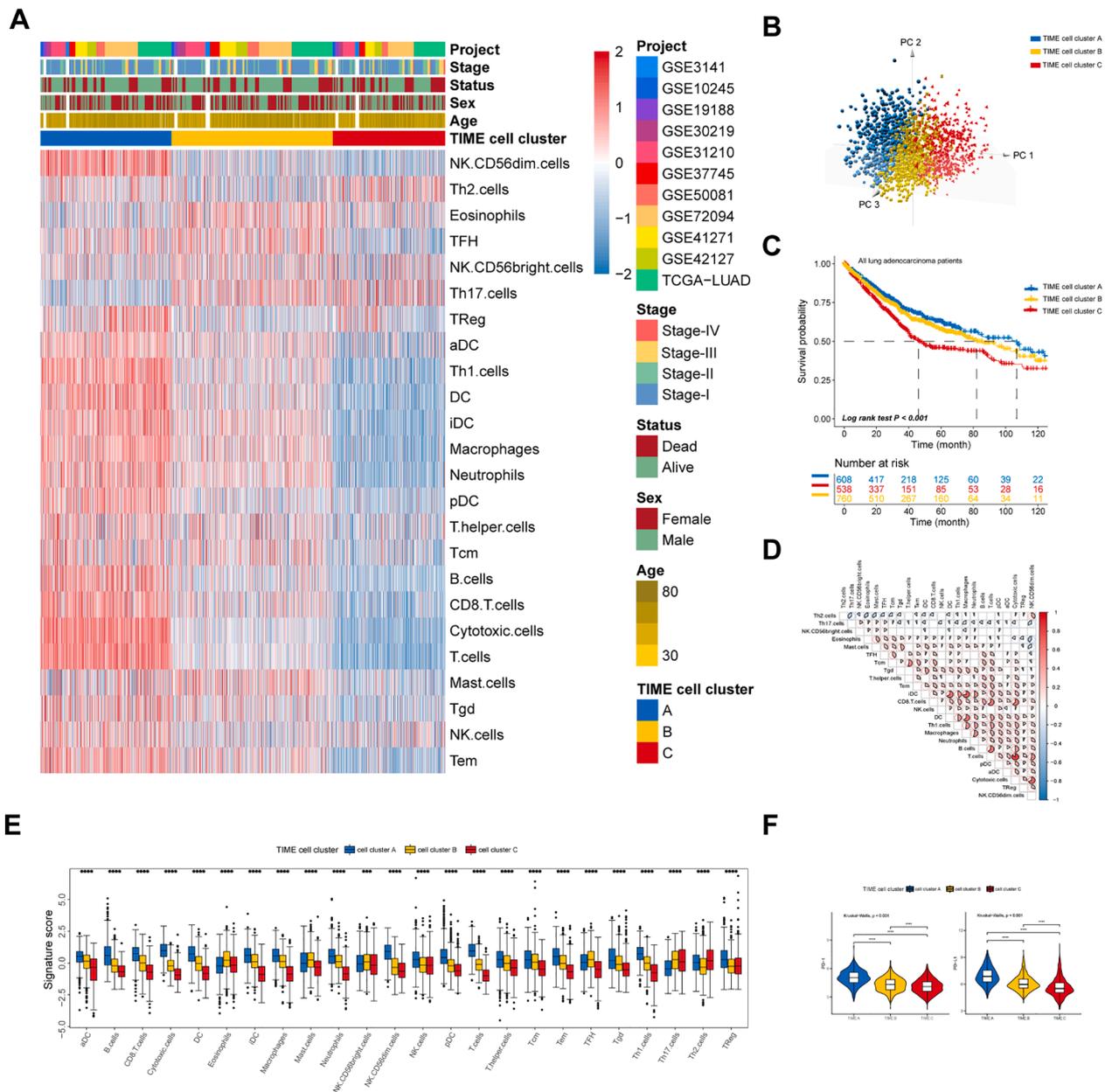
Fig. 1. Landscape of TIME in 1906 LUAD patients. (A) Heatmap of 1906 LUAD samples categorized into three TIME cell clusters via unsupervised clustering. (B) Distinguishing TIME cell clusters using three-dimensional principal component analysis. The dots of different colors correspond to patients in the three TIME cell clusters. (C) Survival curves exhibiting distinct OS of patients in the three TIME cell clusters. (D) Correlation coefficient heatmap of immune cells interaction in TIME cell clusters. (E) Boxplot showing the fraction of immune cells in three TIME cell clusters. * $P < 0.05$; **$P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$. (F) Violin plots showing PD-1/ PD-L1 expression in TIME cell clusters.

and cluster B was between the two. The significance of the differences indicates that the classification is distinct and accurate.

Characteristics of TIME gene clusters from multi-omics analysis

Numerous biological processes are involved in the complex process of tumor progression. To better understand the classification, we combined transcriptome, genome and epigenome data including gene expression, SMG, CNV and cancer stemness analysis. SMG landscapes showed that gene cluster C had a greater tumor mutation burden than clusters A and B (Fig. 3A, S3A, Table S9). TP53, KRAS, KEAP1, SMARCA4, EGFR, and BRAF expression levels were high in gene cluster C. Significant statistical differences in expression levels of TP53, KEAP1, SMARCA4, and EGFR among the gene clusters combined with known drug targets in LUAD [53] were observed, indicating that additional

analyses of TP53, EGFR, KEAP1, and BRAF were needed (Fig. 3B). Our results are consistent with the previous study [13].

The mutation rate among the three gene clusters showed an overall upward trend. High significant mutation rate was correlated with unfavorable LUAD prognoses. EGFR decreased in the three clusters, indicating that patients with high mutation rate in EGFR may be more sensitive to IT. Analysis of EGFR mutation sites in different gene clusters showed that targeted agent action sites were concentrated in gene clusters A and B [54,55] (Fig. 3C). These results indicated that it might be appropriate to administer IT combined with EGFR target drugs for LUAD patients in TIME gene cluster A. GISTIC 2.0 was used to explore CNV among the three clusters. Comparing G-scores across all chromosomes in the three clusters, we found that both frequency and amplitude gradually increased (Fig. 3D, Table S10). Amplifications (14q13.3, 12q14.1, and 12q15) and deletions (9p21.3, 9p21.3, and 9p23) within
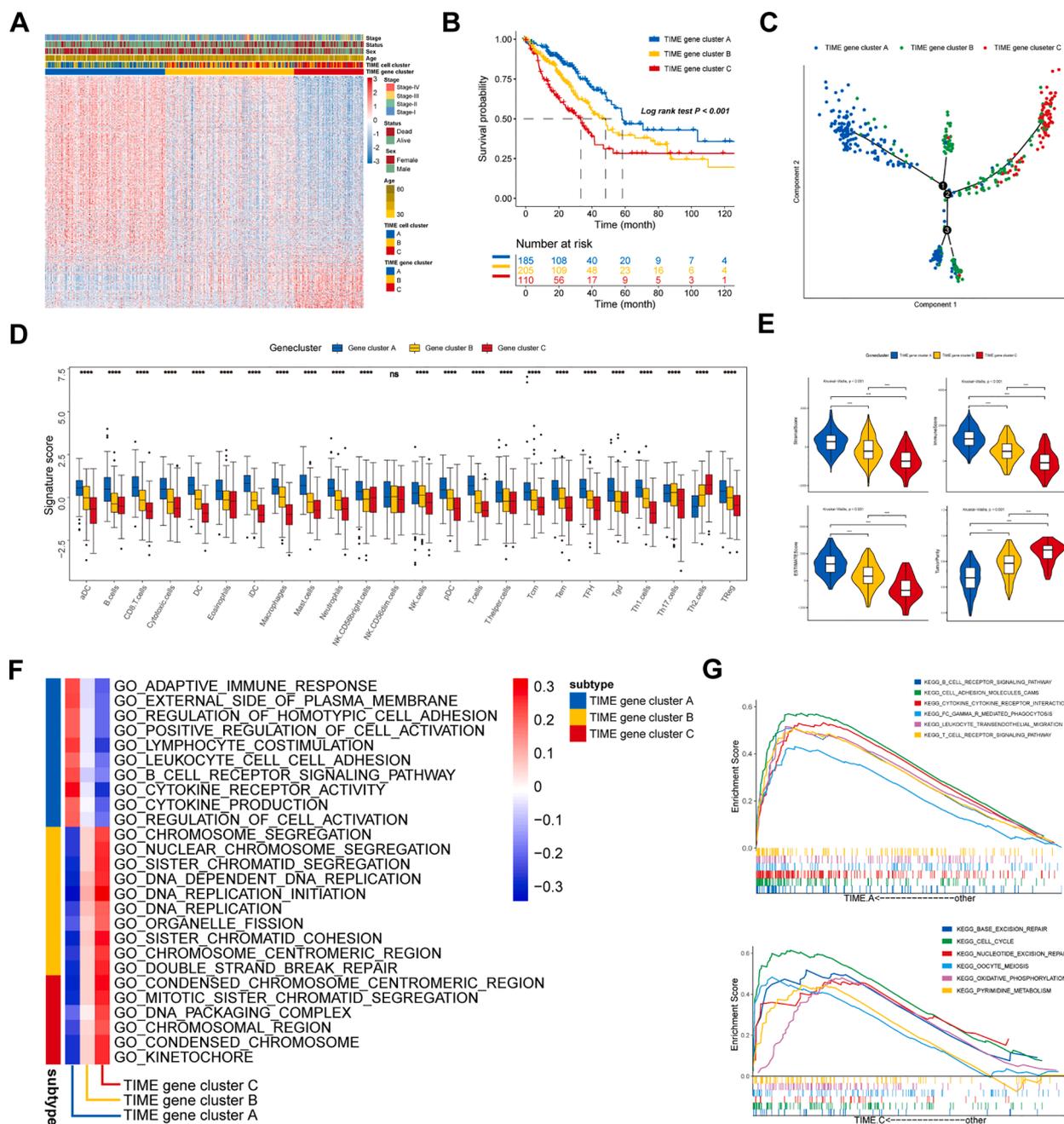
Fig. 2. Construction of the TIME gene cluster and enrichment analysis. (A) Heatmap of TCGA-LUAD samples categorized into three TIME gene clusters via unsupervised clustering. (B) Survival curves exhibiting distinct OS of patients in the three TIME gene clusters. (C) The immune landscape of LUAD in TCGA based on TIME gene clusters. Each point represents a patient. (D) Boxplot showing the fraction of immune cells in three TIME gene clusters. (E) Violin plots showing the differential expression of immune, stromal, estimate scores and tumor purity in three TIME cell clusters (TCGA-LUAD cohort). (F) Heatmap visualizing the distinct GO terms among the TIME gene clusters. Red represents activation states and blue represents inhibition states. (G) GSEA reveals the enrichment results of KEGG pathways in TIME gene cluster A and C.

chromosomal regions were detected in cluster A, amplifications (1q22, 8q24.21, and 1q21.3) and deletions (9p21.3, 9p23, and 13q12.11) within chromosomal regions were detected in cluster B, and amplifications (11q13.3, 14q13.1, and 8q24.21) and deletions (9p21.3, 4q34.3 and 22q13.31) within chromosomal regions were detected in cluster C (Fig. S3B). The copy number gain and loss burdens among the three gene clusters were investigated (Fig. 3E). Whether at the focal or arm level, the burden of CNV increased in clusters A-C and correlated with decreases in LUAD immune infiltration levels. Ranking mRNAsi and mDNAsi from low to high, we explored the correlation between stemness, classification, and corresponding clinical features (Fig. 3F, S3C,

Table S11). Gene cluster A was primarily distributed in the low mRNAsi area, and gene cluster C was primarily distributed in the high mRNAsi area. There were significant differences of mRNAsi among gene clusters ($P<$ 0.001, Fig. 3H). However, clinical features had low correlations with mRNAsi. The trends of the correlation between immune cells and mRNAsi were consistent with the trends of immune cells in TIME clusters (Fig. 3G). mRNAsi was positively correlated with typing and negatively correlated with the levels of most immune cell infiltrations. The overall trend of cancer stem cell markers in LUAD was consistent with the upward trend of mRNAsi (Fig. 3I, Table S5) [33,56,57]. We performed mDNAsi analysis in parallel with mRNAsi, and the results were
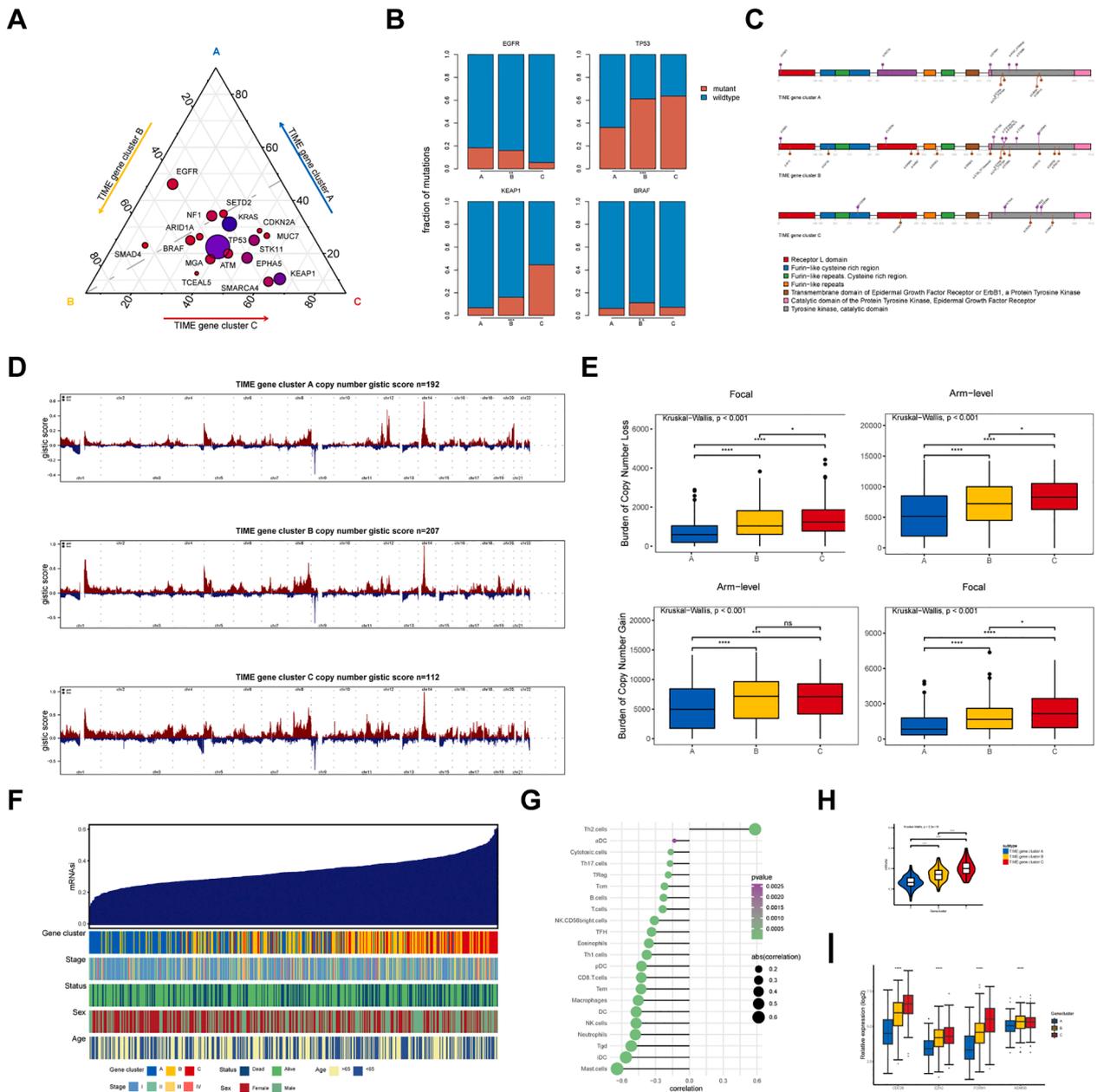
Fig. 3. Multi-omics analyses of TIME gene clusters. (A) Ternary plot showing the mutation frequency of SMGs, comparing TIME gene cluster A (top, blue), TIME gene cluster B (left, yellow), and TIME gene cluster C (right, red). The color of each node indicates relative frequency of mutations in TIME gene cluster A, B and C, whereas the node size represents their overall frequency in LUAD. (B) Rate of mutate and wildtype significant mutated genes (TP53, EGFR, KEAP1, BRAF) among TIME gene clusters. (C) Lollipop plot of different mutation spots in EGFR among TIME gene clusters. (D) Copy numbers profiles showing the amplification (crimson) and deletion (dark blue) of chromosomes from 1 to 22 in TIME gene clusters. (E) Copy numbers gain and loss burden in focal and arm-level in TIME gene clusters. (F) Association between mRNAsi and pathology features of patients. (G) Correlation analysesof immune cells and mRNAsi. The node size represents the strength of the correlation. (H) Boxplot showing the distribution of mRNAsi in different gene clusters. (I) Boxplot showing the expression of markers of cancer stemness among TIME gene clusters.

similar to those of mRNAsi (Fig. S3D, E). The data indicated that our classification is stable and reliable. SMG, CNV, and stemness gradually increased in TIME gene clusters A, B, and C. The combination of immunity, SMG, CNV, and cancer stemness can increase the understanding of cancer development and expand cancer treatment strategies.

Construction of the prognostic biomarker TIME score

The prior analyses were based on patient populations rather than individuals. To predict patient prognoses and assess immune infiltration states of individual tumors accurately, we constructed a TIME score model based on individual gene expression heterogeneity and

complexity. A total of 1412 DEGs were identified, of which 981 genes were positively associated with the classification of gene clusters, known as gene signature A and the remaining DEGs were known as gene signature B. The Boruta algorithm was used to reduce the noise or redundant genes in gene signatures A and B. The transcriptomic profiles of the 193 most abundant DEGs among the genomic clusters were delineated using heatmap (Fig. 4A, S4A, Table S12).

In this investigation, from TIME gene signatures A and B, we calculated TIME score A (TSA) and TIME score B (TSB) using the PCA algorithm, respectively. The sum of the TSA and TSB was used to quantify the TIME of individual LUAD. Thus, we acquired a prognostic model, which was termed as TIME score. TIME scores were determined and
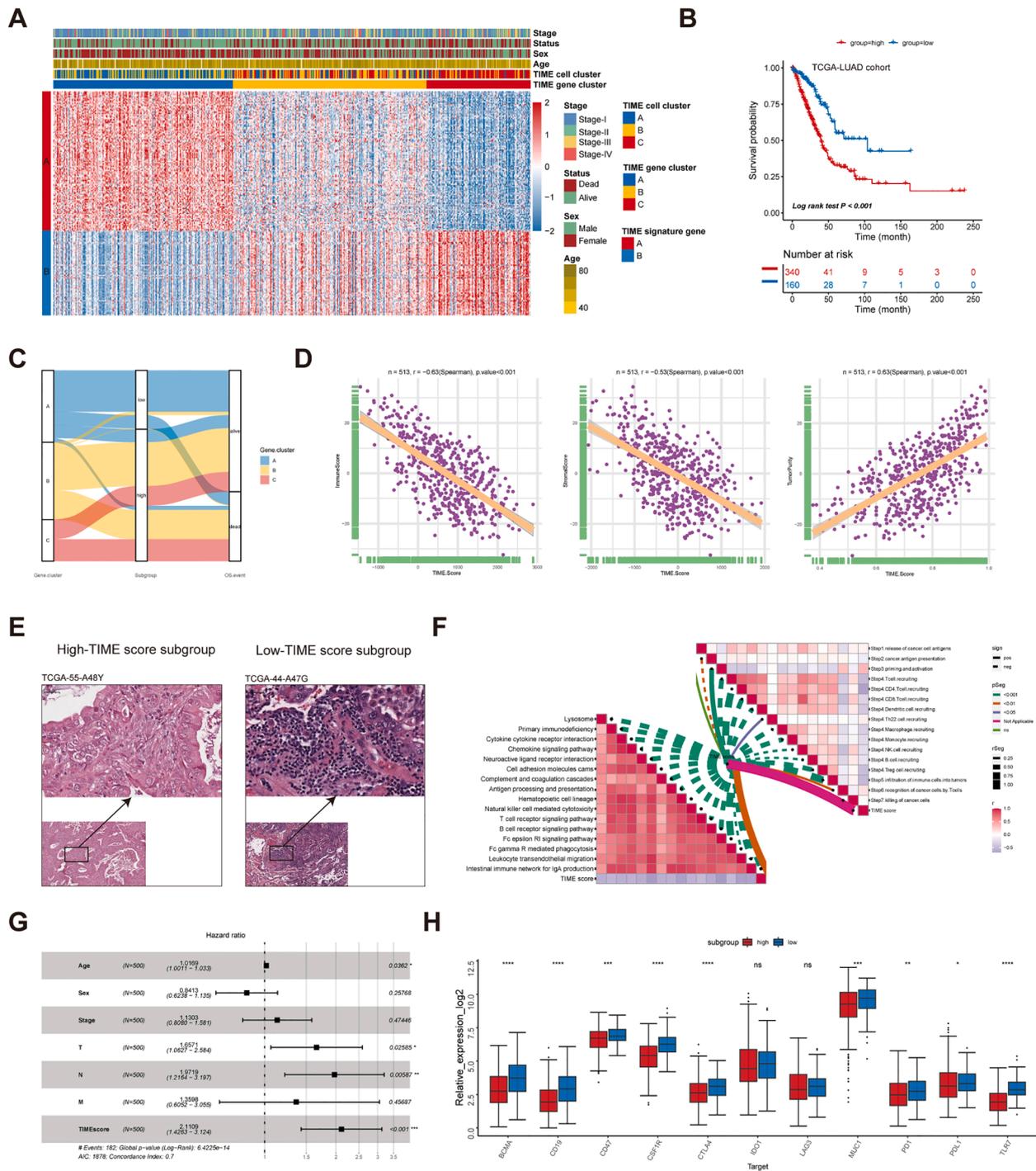
Fig. 4. Construction and characterization of TIME score model. (A) Heatmaps of the positive correlation between gene signatures A and B and immune infiltration level. (B) Survival curve for OS of TIME score subgroups. (C) Alluvial diagram of low and high TIME score subgroups with TIME gene clusters and OS event. (D) Scatterplots of the correlation between TIME scores and estimate analysis results evaluated by ESTIMATE algorithm. (E) Representative TCGA-LUAD H&E histological images in high and low TIME scores. Scale bar denotes 50 μm. (F) Correlations between TIME scores and immune enrichment pathways (left), cancer immune process (right). (G) Forest plot exhibiting multivariable independent prognostic analyses of TIME score and clinical features (age, sex, stage, TNM). (H) Boxplots showing the expression of IT targets in low- and high-TIME- score subgroups, * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$.

stratified into high or low TIME scores using optimal cutoff values calculated by the survminer package for samples in the TCGA-LUAD cohort (Table S13). Analyzing TIME scores integrated with survival information revealed that patients with low TIME scores had improved survival times (median survival time = 3117 days) compared to patients with high TIME scores (median survival time = 1195 days; log-rank test, $P < 0.001$, Fig. 4B). An alluvial diagram illustrated the construction of TIME score model (Fig. 4C). The results showed that gene cluster A was

linked primarily to low TIME scores. OS states were primarily concentrated in alive states in gene cluster A compared to gene cluster C. K-W analysis showed that the TIME scores increased in gene clusters A-C (Fig. S4B).

TIME scores reflected the immune infiltration states of the subgroups. Estimate analysis revealed that immune and stromal scores were positively correlated with TIME scores, and tumor purities were negatively correlated with TIME scores (Fig. 4D). TIME score was closely

correlated with cancer immunity, which was confirmed through pathology examination (Fig. 4E, Table S17). Levels of immune cell-infiltration in low-scoring subgroup were significantly higher than those in the high-scoring subgroup. An increased number of tertiary lymphoid structures were observed in the low-scoring subgroup compared to the high-scoring subgroup. The rate of high immune infiltration in low-scoring subgroup was higher than the rate of high immune infiltration in high-scoring subgroup (89/100 *vs.* 7/100). TIME scores were negatively correlated with the 7 steps of the Cancer Immunity Cycle [41] and could predict the entire process of cancer immunity (Fig. 4F). However, due to the complexity of the cancer microenvironment and the existence of cancer immune escape, elements of the 7 steps in the Cancer Immunity Cycle may not accurately describe cancer progression in patients. This may lead to TIME score model insignificant evaluations of low effects in the corresponding steps. Analysis of the relationship between TIME score and clinical features revealed that they were significantly correlated. (Fig. S4C). The relationship between scores and histological classifications requires further verification because of the limited sample size available in our study (Fig. S4D). We found that TIME scores could be used as effective

prognostic indicators and were supported by forest plots (Fig. 4G, S4E). TIME score was an independent prognostic indicator when combined with other indicators (HR = 2.11; concordance = 0.7). ROC analyses demonstrated that TIME score had predictive advantage of LUAD ($AUC60 = 0.74$; $AUC96 = 0.82$, Fig. S4F). We found that T-cell-targeted immunomodulators (PD-1, CTLA4, IDO1, LAG3, and PDL-1), cancer vaccine (MUC1), cell therapies (CD19 and BCMA), and other immunomodulators (TLR7, CD47, and CSF1R) were negatively correlated with TIME score (Fig. 4H, Table S5), indicating that low-score groups may benefit more from IT. We found correlations among TIME cell clusters, TIME gene clusters, and TIME scores. Scoring simplified LUAD classification criteria, which could be used for determining patient prognoses. In addition, scores could be used in combination with other indicators to guide treatment.

TIME score in the prediction of immunotherapeutic benefits

TIME score has been shown to be an independent prognostic risk factor. It is unclear if TIME score can guide IT given the relationship between TIME score and tumor immunity. ITs, including cytokines,
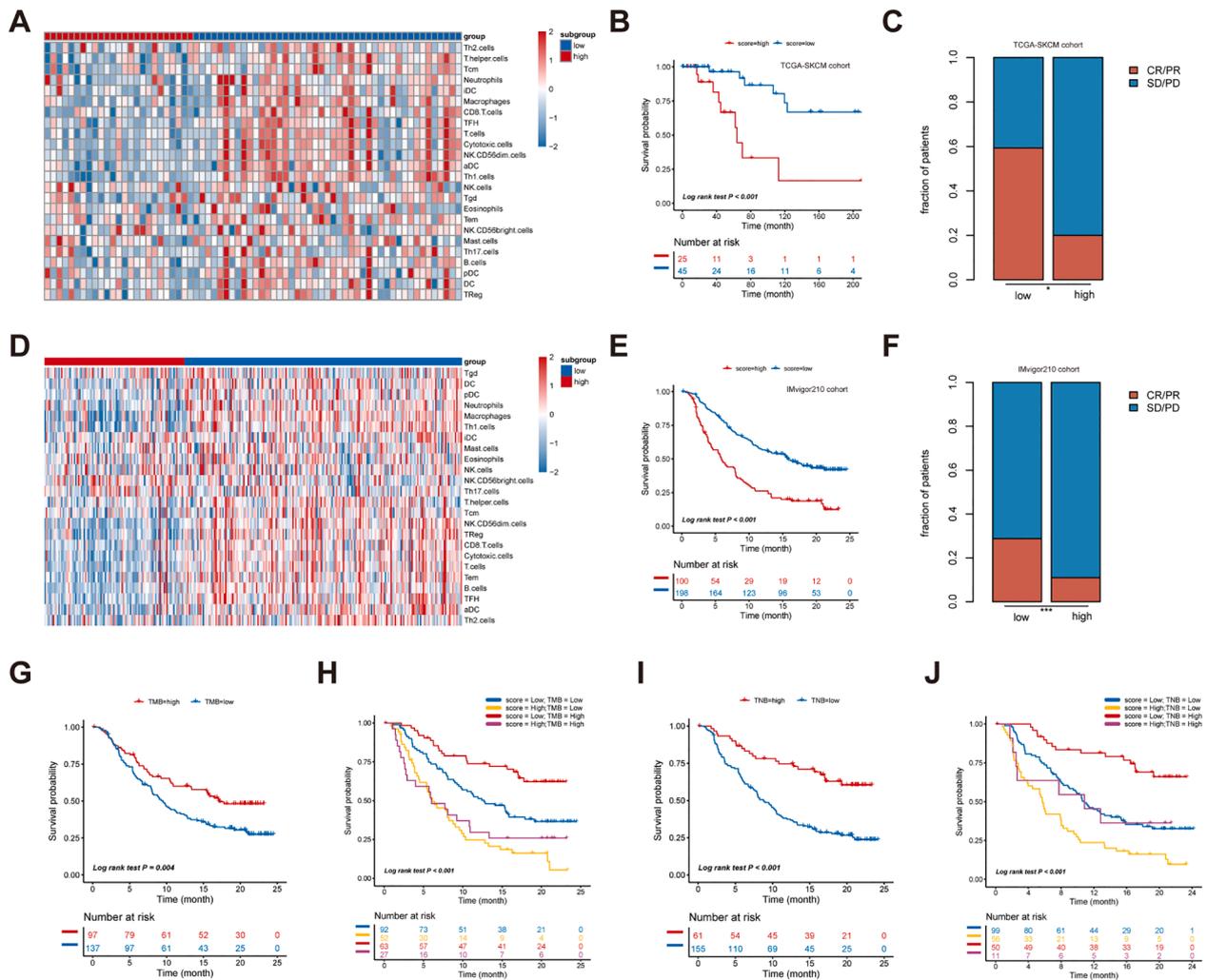
Fig. 5. Prediction of IT response in TIME score model. (A) Heatmap of immune infiltration levels in two TIME score subgroups (TCGA-SKCM, n=70, a mixed immunotherapy cohort). (B) Survival curves exhibiting distinct OS in two different TIME score subgroups (TCGA-SKCM cohort). (C) Fractions of responders (CR and PR) and non-responders (PD and SD) in TCGA-SKCM patients treated with mixed IT across two TIME score subgroups. (D) Heatmap of immune infiltration levels in two TIME score subgroups (IMvigor210, n=298, an PD-L1 inhibitor treatment cohort). (E) Survival curves exhibiting distinct OS in two TIME score subgroups (IMvigor210 cohort). (F) Fractions of responders (CR and PR) and non-responders (PD and SD) in IMvigor210 patients treated with PD-L1 blockade across two TIME score subgroups. (G) Survival curves exhibiting distinct OS in patients with high or low TMB treated with PD-L1 blockade. (H) Survival curves showing distinct OS in IMvigor210 cohort layered by TMB and TIME score. (I) Survival curves exhibiting distinct OS in patients treated with PD-L1 blockade with high or low TNB. (J) Survival curves showing distinct OS in IMvigor210 cohort layered by TNB and TIME score.

vaccines, and checkpoint blockers, are commonly used in patients and show unprecedented survival benefits. To investigate the prognostic value of the TIME score model in response to IT, we assessed data from two representative IT cohorts (TCGA-SKCM and IMvigor210) for further analyses (Table S15). We separated cohort patients into high- or low-TIME-score subgroups. The immune infiltration levels between the two subgroups were different. The low-score subgroup showed a higher infiltration status compared to the high-score subgroup (Fig. 5A, 5D). Patients with low TIME scores had better prognoses than those with high TIME scores in both TCGA-SKCM and IMvigor210 cohorts (Fig. 5B, 5E). Patients in the low-TIME-score subgroup may have improved benefits from IT. (response rate of mixed IT cohort: 59% *vs*. 20%, Fig. 5C; response rate of anti-PD-1 cohort: 29% *vs*.11%, Fig. 5F, Fig. S5A, S5B). The same results were found in the anti-PD-1/ PD-L1 cohort GSE135222, anti-PD-1/ PD-L1 cohort PMID29301960, and anti-PD-1 cohort GSE91061 (Fig. S5C-I), which confirmed the reliability and stability of our results.

TMB is significantly correlated to the efficacy of IT, and tumor neoantigen burden (TNB) may influence immune surveillance and provide targets for IT [58]. In IMvigor210 cohort, we analyzed TMB, TNB and their synergies with TIME score. The results indicated that high-TMB subgroup had a longer progression-free survival compared to the low-TMB subgroup (Fig. 5G). Low TIME score combined with high TMB had synergistic effects and provided the most favorable prognosis (Fig. 5H). The effects of TNB were similar to those of TMB. High TNB had a more favorable prognosis compared to low TNB. Low TIME score and high TNB had the best-combined effect on patient prognosis (Fig. 5I, J). Whether combined with high or low TMB, TIME scores were stable and were correlated with favorable prognoses, as was TNB. Our results indicated that TIME scores were correlated with Immunotherapeutic response. Whether in PD-1 and PD-L1 datasets or in mixed IT datasets, TIME scores served as reliable predictive indicators. When TIME scores were combined with TMB or TNB, prognoses were more favorable. Our results provide a unique opportunity to explore TMB and TNB functions in predicting IT efficacy.

Validation the performance of TIME score in LUAD cohorts

We constructed a TIME score model of LUAD. All the results above proved that patients with low TIME scores showed high immune infiltration levels, prolonged survival time, favorable responses to IT. However further validation is needed. The largest internal cohort (GSE72094, *n*= 439) and the largest external validation dataset (GSE68465, *n*= 440) were selected for the validation of TIME score model. Differences between the high-score groups and the low-score groups were found in both cohorts. Kaplan–Meier survival analyses showed that patients in low-score groups had favorable 5-year prognoses and increased overall survival (*P* < 0.001, Fig. 6A, 6C). Our TIME score model had an improved prognostic effect in the different datasets (Fig. S6A, S6B). Boxplots showed that their immune cell composition differed between high- and low-score subgroups (Fig. S6F, S6G). In addition to public data cohorts, our cohorts showed the same trend. The same pathologists determined the degree of cancer immune infiltration status in LUAD sections from the SHDMU cohort and TMA and classified them into high- or low-immune-infiltration subgroups. Samples in the high-immune-infiltration group had more tertiary lymphoid structures than those in the low-immune-infiltration group (Fig. 6E, S6D). Survival analysis showed differences between the two subgroups (Fig. S6C, S6E). Twenty lung adenocarcinoma samples of different subgroups were randomly selected from the SHDMU cohort. The expression of tumor stem cell markers and immune checkpoint molecules were verified according to the results of qRT-PCR (Fig. S6H). The results showed that the mRNA expressions of CD44, CD133, and ALDH1A1 ascended in LUAD samples in comparison to the high-immune-infiltration subgroup, while that of PD-L1 descend in LUAD samples compared to the high-immune-infiltration subgroup. The experimental results are consistent with the above analysis results. These results supported the credibility of our classification.

IT response of TIME score model illustrated that patients in the low-score groups were more likely to benefit from anti-PD-1 therapy (Fig. 6B, 6D). We investigated whether IT can work synergistically with other drugs. Estimation of therapeutic responses to the DMOG and erlotinib
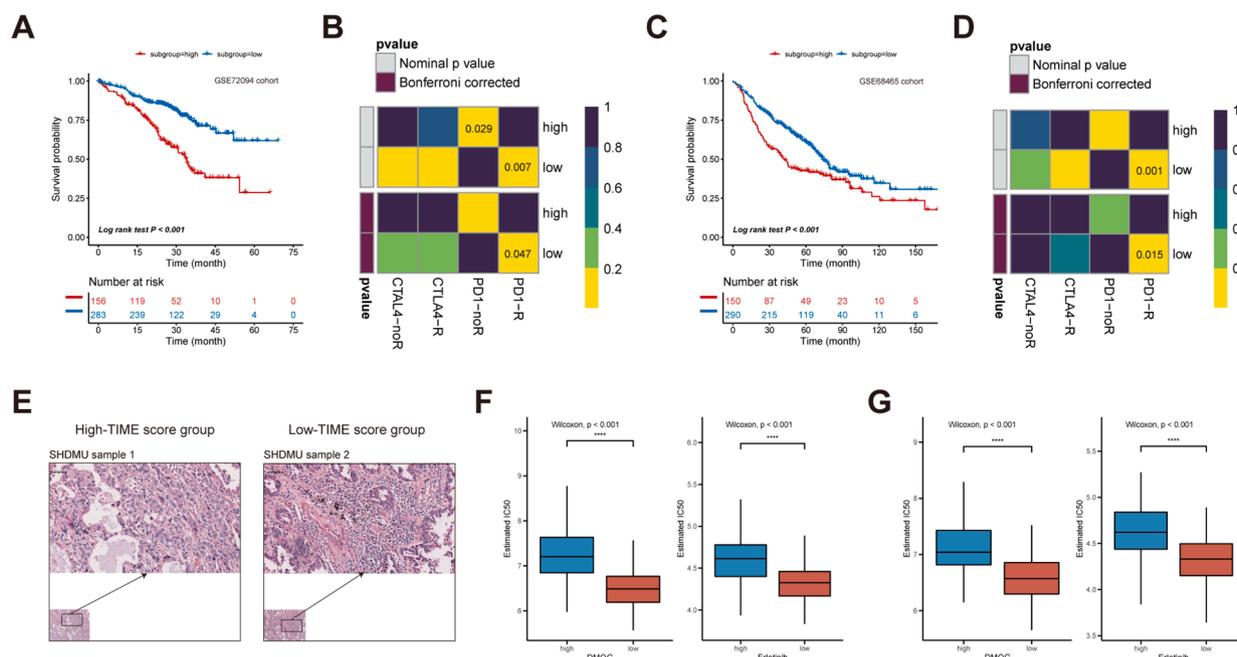
Fig. 6. Prognostic value of TIME score model and drug sensitivities in LUAD cohorts. (A) Survival curve of patients in two TIME score subgroups (internal validation dataset, GSE72094, n= 439). (B) Submap analysis for calculating the possibility of IT response (anti-PD1 and anti- CTLA4) in distinct TIME score subgroups (GSE72094 cohort). (C) Survival curve of patients in two TIME score subgroups (external validation dataset, GSE68465, n= 443). (D) Submap analysis for calculating the possibility of IT response (anti-PD1 and anti- CTLA4) in distinct TIME score subgroups (GSE72094 cohort). (E) Representative SHDMU H&E histological images in high and low TIME scores. Scale bar denotes 50 μm. (F-G) Boxplots of estimated IC50s for DMOG and Erlotinib in GSE72094 (F) and GSE68465 (G).

showed lower IC50s in patients with low TIME scores (Fig. 6F, 6G), indicating increased sensitivity to DMOG and erlotinib in low TIME score subgroup. Thus, TIME score can be used to predict the efficacy of IT. Patients with LUAD may experience increased benefits from a combination of IT with DMOG or erlotinib therapy. TIME score can simply and efficiently evaluate the immune infiltration status of LAUD patients and classify them into high- or low-TIME-score subgroup so that they can be treated separately. In addition, TIME score can effectively predict the overall survival and the response to IT of LUAD patients. The TIME score model may provide new approaches for combining IT and other treatments.

Discussion

LUAD is the most common histopathological type of lung cancer [59]. TIME is a complex ecosystem and an essential part of the tumor microenvironment that affects tumor evolution [60]. Despite much research, the understanding of prognostic factors and therapeutic targets of LUAD is limited. The development of IT has focused cancer research on TIME.

We developed a methodology to investigate the interaction between TIME and LUAD. TIME clusters were constructed based on immune cell characteristics from LUAD samples. In our study, we identified three TIME clusters A, B, and C that showed characteristic immune cell population levels. Samples in TIME cluster A, a high immune infiltration group, contained the highest numbers of immune cells, including DCs, B cells, some subtypes of T cells, and other innate immune cells. Eosinophils, mast cells, Th2 cells, and Th17 cells demonstrated different trends. DCs are responsible for the initiation of immune responses. B cells and other innate immune cells are responsible for killing tumor cells. CD8[+] T cells are believed to have anti-tumor properties [61]. CD4[+] T cells are complicated and highly heterogeneous in phenotype. Subtypes of CD4[+] T cells are classified into Th1, Th2, Th17, and Tregs [62,63]. Th1 and Tregs are involved in cancer progression. Th2, eosinophils, and mast cells are involved in combating extracellular pathogens and parasites. Th17 cells release IL-17 that modulates TIME, and promotes cancer cell growth and metastasis [64].

We used machine learning algorithms to identify DEGs in TIME clusters and used consensus clustering to classify patients into three gene clusters. A stable TIME gene cluster is the basis for TIME score construction. We verified the SMG, CNV, and cancer stemness of the three TIME gene clusters. They showed observable differences that may affect immune infiltration status and promote immune evasion. SMG mutation rates in TIME gene cluster A were lower than those in TIME gene cluster C. Mutations in lung cancer accumulate over the course of tumor evolution [65]. Our clustering results are consistent with this observation. Mutations in TP53, KEAP1, and SMARCA4 increase from TIME gene clusters A to C and are similar to previous studies [66,67]. We demonstrated that these mutations were correlated with poor prognoses and may act a pivotal part in tumor immunity. In CNV analysis, TIME with high CNV levels have more tumor-promoting and immunosuppressive properties among various cancers [68,69]. Studies have indicated that, in lung adenocarcinoma cells with high levels of CNV, tumor proliferation increased and immune infiltration decreased. Increased tumor proliferation was primarily predicted by focal CNV. Low levels of immune infiltration were primarily predicted by high levels of arm and whole-chromosome CNV [69]. Cells that escape anti-tumor surveillance may have high levels of CNV, which resulted from chromosomal instability. Our results are consistent with the observation that elevated CNV is associated with an increased hazard of mortality [70]. Stemness index, mRNAsi, and mDNAsi were used to evaluate the stemness of LUAD based on the existence of cancer stem cells (CSCs). Our high stemness index in the poor-prognosis group is consistent with the observation that CSCs can lead to tumor progression and poor prognosis [71,72]. CSCs are negatively correlated with PD-L1 expression and leukocyte fraction in cancer immunity [33]. CSCs secrete interleukin 10 and tumor necrosis

factor β that may induce more immunosuppressive phenotypes of tumor-associated macrophages [73]. In short, cancer stemness is closely related to cancer immunity. Our results showed the stability and practicality of gene clusters. These differences may be related to different immune statuses that may affect the prognoses of patients. Immune cell composition and survival analysis showed the same trend among TIME cell clusters and TIME gene clusters. These results showed that our gene clusters correlated well with TIME cell clusters, indicating that gene clusters can represent cell clusters. Because our sample size was larger than that used in previous studies, and we investigated TIME in a comprehensive manner, our classification is more accurate.

Cancer development is a dynamic process. Cancer immunity weakens as cancer progresses. Pseudo-time analysis demonstrated that our classification is consistent with the process of cell differentiation. The cancer immunity cycle contains 7 steps. Cancer immunity-related immune cells in TIME gene cluster A are highly infiltrated, consistent with the 7 steps. DCs can capture cancer antigens and present them to T cells, and T effector cells are activated. In this process, the balance between T effector cells and T regulatory cells determines the development of cancer. Activated effector T cells infiltrate the cancer microenvironment. T cell receptors bind to cancer cells and kill them specifically. Cancer cell death releases more cancer antigens that trigger the next cycle [41]. As cancer progresses, the immune response gradually weakens as the cancer microenvironment suppresses immune cell levels, which results in immune evasion [60]. Our classification result is consistent with this process.

IT has changed the treatment of multiple types of advanced cancers, especially LAUD [74–77]. However, current biomarkers, such as PD-1 and PD-L1 expression and mutation load, are not effective for predicting the efficacy of IT. There is a high level of clinical interest in biomarkers that provide accurate predictions on IT efficacy. We developed a scoring model to evaluate IT response, assess immune infiltration status, and predict patient prognosis. Five separate IT cohorts, including anti-PD-L1-treated metastatic urothelial cancer [42], IT-treated skin cutaneous melanoma [43], anti-PD-1/PD-L1-treated advanced non-small-cell lung carcinoma [44], anti-PD-1/PD-L1-treated clear cell renal cell carcinoma [45], and anti-PD-L1-treated advanced melanoma [46], were used to verify our scoring model. Patients in high-TIME-score subgroups were correlated with lower IT efficacies compared to patients in low-TIME-score subgroups. These results indicate that patients with low TIME scores might experience more benefits with early administration of single-agent IT.

We analyzed the TIME of LUAD using a comprehensive set of analytical tools and data. This resulted in the construction of a reliable model that can accurately distinguish different TIME subtypes, determine the prognoses of patients, and guide treatment. Additional research on the correlation between tumor evolution and immune evolution is needed. The bulk RNA-seq and microarray datasets we investigated lacked data on cellular proportions and heterogeneity. Investigations into deep spatial distribution [78], higher flux, and higher dimensionality are needed. Higher-resolution ratio techniques like single-cell RNA sequencing and multidimensional immunohistochemistry [60] are needed. It might be that individual differences and disease complexity do not result in greater IT benefits to all patients with low TIME scores. Large randomized controlled trials could provide results that improve the accuracy of prediction models. Our research may provide information and insights for understanding the correlation between multi-omics results and cancer immunity.

Conclusions

In conclusion, we quantified the infiltration status of tumor immune via machine-learning techniques among 1906 LUAD patients. Combined with multi-omics data analysis, pseudo-time dynamic analysis, and enrichment analysis, we developed a TIME scoring signature to evaluate immunotherapy response and predict patient prognosis. Our model of

LUAD immune microenvironment characteristics may improve the prognostic accuracy of patients, provide improved explanations of LUAD responses to immunotherapy, and provide new strategies for LUAD treatment.

Data availability statement

All data related to this study are included in the article or supplementary materials. The codes used in the study can be obtained from relevant authors upon reasonable request.

Author contributions

CXR and QW supervised the project, designed, edited, and led out the experiments of this study. CXR and JYL were involved in conceptualization, data curation, project administration, formal analysis, software, visualization and writing original draft. YZ and SYZ were involved in supervision and writing review & editing. All authors contributed to critical revision of the final manuscript and approved the final version of the manuscript.

Funding

Declaration of Competing Interest

The authors report no conflict of interest.

Acknowledgment

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tranon.2022.101367.

References

[1] J-Y Xu, C Zhang, X Wang, L Zhai, Y Ma, Y Mao, et al., Integrative proteomic characterization of human lung adenocarcinoma, Cell 182 (2020) 245–261, https://doi.org/10.1016/j.cell.2020.05.043, e17.

[2] RS Herbst, D Morgensztern, C. Boshoff, The biology and management of non-small cell lung cancer, Nature 553 (2018) 446–454, https://doi.org/10.1038/nature25183.

[3] D Schadendorf, FS Hodi, C Robert, JS Weber, K Margolin, O Hamid, et al., Pooled analysis of long-term survival data from phase II and phase III trials of ipilimumab in unresectable or metastatic melanoma, J. Clin. Oncol. 33 (2015) 1889–1894, https://doi.org/10.1200/JCO.2014.56.2736.

[4] T Okazaki, S Chikuma, Y Iwai, S Fagarasan, T. Honjo, A rheostat for immune responses: the unique properties of PD-1 and their advantages for clinical application, Nat. Immunol. 14 (2013) 1212–1218, https://doi.org/10.1038/ni.2762.

[5] W Zou, JD Wolchok, L. Chen, PD-L1 (B7-H1) and PD-1 pathway blockade for cancer therapy: Mechanisms, response biomarkers, and combinations, Sci. Transl. Med. 8 (2016), https://doi.org/10.1126/scitranslmed.aad7118, 328rv4.

[6] M Reck, D Rodríguez-Abreu, AG Robinson, R Hui, T Csőszi, A Fülöp, et al., Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer, N. Engl. J. Med. 375 (2016) 1823–1833, https://doi.org/10.1056/NEJMoa1606774.

[7] A Rittmeyer, F Barlesi, D Waterkamp, K Park, F Ciardiello, Pawel J von, et al., Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial, Lancet 389 (2017) 255–265, https://doi.org/10.1016/S0140-6736(16)32517-X.

[8] EB Garon, NA Rizvi, R Hui, N Leighl, AS Balmanoukian, JP Eder, et al., Pembrolizumab for the treatment of non-small-cell lung cancer, N. Engl. J. Med. 372 (2015) 2018–2028, https://doi.org/10.1056/NEJMoa1501824.

[9] MS Tsao, KM Kerr, M Kockx, M-B Beasley, AC Borczuk, J Botling, et al., PD-L1 Immunohistochemistry comparability study in real-life clinical samples: results of blueprint phase 2 project, J. Thorac. Oncol. 13 (2018) 1302–1311, https://doi.org/10.1016/j.jtho.2018.05.013.

[10] C-Y Yang, JC-H Yang, P-C Yang, Precision management of advanced non-small cell lung cancer, Annu. Rev. Med. 71 (2020) 117–136, https://doi.org/10.1146/annurev-med-051718-013524.

[11] R Cristescu, R Mogg, M Ayers, A Albright, E Murphy, J Yearley, et al., Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy, Science (2018) 362, https://doi.org/10.1126/science.aar3593.

[12] KC Arbour, GJ. Riely, Systemic therapy for locally advanced and metastatic non-small cell lung cancer: a review, JAMA 322 (2019) 764–774, https://doi.org/10.1001/jama.2019.11058.

[13] C Valero, M Lee, D Hoen, J Wang, Z Nadeem, N Patel, et al., The association between tumor mutational burden and prognosis is dependent on treatment context, Nat. Genet. 53 (2021) 11–15, https://doi.org/10.1038/s41588-020-00752-4.

[14] O Meurette, P. Mehlen, Notch signaling in the tumor microenvironment, Cancer Cell 34 (2018) 536–548, https://doi.org/10.1016/j.ccell.2018.07.009.

[15] R Remark, C Becker, JE Gomez, D Damotte, M-C Dieu-Nosjean, C Sautès-Fridman, et al., The non-small cell lung cancer immune contexture. A major determinant of tumor characteristics and patient outcome, Am. J. Respir. Crit. Care Med. 191 (2015) 377–390, https://doi.org/10.1164/rccm.201409-1671PP.

[16] SR Gordon, RL Maute, BW Dulken, G Hutter, BM George, MN McCracken, et al., PD-1 expression by tumour-associated macrophages inhibits phagocytosis and tumour immunity, Nature 545 (2017) 495–499, https://doi.org/10.1038/nature22396.

[17] S Jhunjhunwala, C Hammer, L. Delamarre, Antigen presentation in cancer: insights into tumour immunogenicity and immune evasion, Nat .Rev. Cancer 21 (2021) 298–312, https://doi.org/10.1038/s41568-021-00339-z.

[18] S Krishna, FJ Lowery, AR Copeland, E Bahadiroglu, R Mukherjee, L Jia, et al., Stem-like CD8 T cells mediate response of adoptive cell immunotherapy against human cancer, Science 370 (2020) 1328–1334, https://doi.org/10.1126/science.abb9847.

[19] ND Huntington, J Cursons, J. Rautela, The cancer-natural killer cell immunity cycle, Nat. Rev. Cancer 20 (2020) 437–454, https://doi.org/10.1038/s41568-020-0272-z.

[20] S Jaillon, A Ponzetta, D Di Mitri, A Santoni, R Bonecchi, A. Mantovani, Neutrophil diversity and plasticity in tumour progression and therapy, Nat. Rev. Cancer 20 (2020) 485–503, https://doi.org/10.1038/s41568-020-0281-y.

[21] A Colaprico, TC Silva, C Olsen, L Garofano, C Cava, D Garolini, et al., TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data, Nucleic Acids Res. 44 (2016) e71, https://doi.org/10.1093/nar/gkv1507.

[22] JT Leek, WE Johnson, HS Parker, AE Jaffe, JD. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments, Bioinformatics 28 (2012) 882–883, https://doi.org/10.1093/bioinformatics/bts034.

[23] G Bindea, B Mlecnik, M Tosolini, A Kirilovsky, M Waldner, AC Obenauf, et al., Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer, Immunity 39 (2013) 782–795, https://doi.org/10.1016/j.immuni.2013.10.003.

[24] K Yoshihara, M Shahmoradgoli, E Martínez, R Vegesna, H Kim, W Torres-Garcia, et al., Inferring tumour purity and stromal and immune cell admixture from expression data, Nat. Commun. 4 (2013) 2612, https://doi.org/10.1038/ncomms3612.

[25] MD Wilkerson, DN. Hayes, ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking, Bioinformatics 26 (2010) 1572–1573, https://doi.org/10.1093/bioinformatics/btq170.

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

[27] ME Ritchie, B Phipson, Di Wu, Y Hu, CW Law, W Shi, et al., limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res 43 (2015) e47, https://doi.org/10.1093/nar/gkv007.

[28] X Qiu, Q Mao, Y Tang, L Wang, R Chawla, HA Pliner, et al., Reversed graph embedding resolves complex single-cell trajectories, Nat Methods 14 (2017) 979–982, https://doi.org/10.1038/nmeth.4402.

[29] B Li, Y Cui, DK Nambiar, JB Sunwoo, R. Li, The immune subtypes and landscape of squamous cell carcinoma, Clin Cancer Res. 25 (2019) 3528–3537, https://doi.org/10.1158/1078-0432.CCR-18-4085.

[30] MS Lawrence, P Stojanov, P Polak, GV Kryukov, K Cibulskis, A Sivachenko, et al., Mutational heterogeneity in cancer and the search for new cancer-associated genes, Nature 499 (2013) 214–218, https://doi.org/10.1038/nature12213.

[31] Z Gu, R Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data, Bioinformatics 32 (2016) 2847–2849, https://doi.org/10.1093/bioinformatics/btw313.

[32] CH Mermel, SE Schumacher, B Hill, ML Meyerson, R Beroukhim, G. Getz, GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers, Genome Biol. 12 (2011) R41, https://doi.org/10.1186/gb-2011-12-4-r41.

[33] TM Malta, A Sokolov, AJ Gentles, T Burzykowski, L Poisson, JN Weinstein, et al., Machine learning identifies stemness features associated with oncogenic dedifferentiation, Cell 173 (2018) 338–354, https://doi.org/10.1016/j.cell.2018.03.034, e15.

[34] JL Speiser, ME Miller, J Tooze, E. Ip, A comparison of random forest variable selection methods for classification prediction modeling, Expert Syst. Appl. 134 (2019) 93–101, https://doi.org/10.1016/j.eswa.2019.05.028.

[35] F Degenhardt, S Seifert, S. Szymczak, Evaluation of variable selection methods for random forests and omics data sets, Brief Bioinform. 20 (2019) 492–503, https://doi.org/10.1093/bib/bbx124.

[36] B Zhang, Q Wu, B Li, D Wang, L Wang, YL. Zhou, m(6)A regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in gastric cancer, Mol Cancer 19 (2020) 53, https://doi.org/10.1186/s12943-020-01170-0.

[37] D Zeng, M Li, R Zhou, J Zhang, H Sun, M Shi, et al., Tumor microenvironment characterization in gastric cancer identifies prognostic and immunotherapeutically relevant gene signatures, Cancer Immunol. Res. 7 (2019) 737–750, https://doi.org/10.1158/2326-6066.CIR-18-0436.

[38] S Hänzelmann, R Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data, BMC Bioinform. 14 (2013) 7, https://doi.org/10.1186/1471-2105-14-7.

[39] Tianzhi Wu, Erqiang Hu, Shuangbin Xu, Meijun Chen, Pingfan Guo, Zehan Dai, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* (2021) 2:100141. doi:10.1016/j.xinn.2021.100141.

[40] L Xu, C Deng, B Pang, X Zhang, W Liu, G Liao, et al., TIP: a web server for resolving tumor immunophenotype profiling, Cancer Res. 78 (2018) 6575–6580, https://doi.org/10.1158/0008-5472.CAN-18-0689.

[41] DS Chen, I. Mellman, Oncology meets immunology: the cancer-immunity cycle, Immunity 39 (2013) 1–10, https://doi.org/10.1016/j.immuni.2013.07.012.

[42] S Mariathasan, SJ Turley, D Nickles, A Castiglioni, K Yuen, Y Wang, et al., TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells, Nature 554 (2018) 544–548, https://doi.org/10.1038/nature25501.

[43] Genomic Classification of cutaneous melanoma, Cell 161 (2015) 1681–1696, https://doi.org/10.1016/j.cell.2015.05.044.

[44] H Jung, HS Kim, JY Kim, J-M Sun, JS Ahn, M-J Ahn, et al., DNA methylation loss promotes immune evasion of tumours with high mutation and copy number load, Nat. Commun. 10 (2019) 4278, https://doi.org/10.1038/s41467-019-12159-9.

[45] D Miao, CA Margolis, W Gao, MH Voss, W Li, DJ Martini, et al., Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma, Science 359 (2018) 801–806, https://doi.org/10.1126/science.aan5951.

[46] N Riaz, JJ Havel, V Makarov, A Desrichard, WJ Urba, JS Sims, et al., Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab, Cell 171 (2017) 934–949, https://doi.org/10.1016/j.cell.2017.09.028, e16.

[47] Y Hoshida, J-P Brunet, P Tamayo, TR Golub, JP. Mesirov, Subclass mapping: identifying common subtypes in independent disease data sets, PLoS One 2 (2007) e1195, https://doi.org/10.1371/journal.pone.0001195.

[48] X Lu, L Jiang, L Zhang, Y Zhu, W Hu, J Wang, et al., Immune signature-based subtypes of cervical squamous cell carcinoma tightly associated with human papillomavirus type 16 expression, molecular features, and clinical outcome, Neoplasia 21 (2019) 591–601, https://doi.org/10.1016/j.neo.2019.04.003.

[49] P Geeleher, N Cox, RS. Huang, pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels, PLoS One 9 (2014), e107468, https://doi.org/10.1371/journal.pone.0107468.

[50] P Geeleher, NJ Cox, RS. Huang, Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines, Genome Biol. 15 (2014) R47, https://doi.org/10.1186/gb-2014-15-3-r47.

[51] X Robin, N Turck, A Hainard, N Tiberti, F Lisacek, J-C Sanchez, et al., pROC: an open-source package for R and S+ to analyze and compare ROC curves, BMC Bioinform. 12 (2011) 77, https://doi.org/10.1186/1471-2105-12-77.

[52] M Yarchoan, BA Johnson, ER Lutz, DA Laheru, EM. Jaffee, Targeting neoantigens to augment antitumour immunity, Nat. Rev. Cancer 17 (2017) 569, https://doi.org/10.1038/nrc.2017.74.

[53] J Rotow, TG. Bivona, Understanding and targeting resistance mechanisms in NSCLC, Nat. Rev. Cancer 17 (2017) 637–658, https://doi.org/10.1038/nrc.2017.84.

[54] SV Sharma, DW Bell, J Settleman, DA. Haber, Epidermal growth factor receptor mutations in lung cancer, Nat Rev Cancer 7 (2007) 169–181, https://doi.org/10.1038/nrc2088.

[55] C-S Tan, D Gilligan, S. Pacey, Treatment approaches for EGFR-inhibitor-resistant patients with non-small-cell lung cancer, Lancet Oncol. 16 (2015) e447–e459, https://doi.org/10.1016/S1470-2045(15)00246-6.

[56] L Walcher, A-K Kistenmacher, H Suo, R Kitte, S Dluczek, A Strauß, et al., Cancer stem cells-origins and biomarkers: perspectives for targeted personalized therapies, Front Immunol. 11 (2020) 1280, https://doi.org/10.3389/fimmu.2020.01280.

[57] WS Heng, R Gosens, FA. Kruyt, Lung cancer stem cells: origin, features, maintenance mechanisms and therapeutic targeting, Biochem. Pharmacol. 160 (2019) 121–133, https://doi.org/10.1016/j.bcp.2018.12.010.

[58] N McGranahan, AJ Furness, R Rosenthal, S Ramskov, R Lyngaa, SK Saini, et al., Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade, Science 351 (2016) 1463–1469, https://doi.org/10.1126/science.aaf1490.

[59] H Sung, J Ferlay, RL Siegel, M Laversanne, I Soerjomataram, A Jemal, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J. Clin. 71 (2021) 209–249, https://doi.org/10.3322/caac.21660.

[60] H Dejima, X Hu, R Chen, J Zhang, J Fujimoto, ER Parra, et al., Immune evolution from preneoplasia to invasive lung adenocarcinomas and underlying molecular features, Nat Commun 12 (2021) 2722, https://doi.org/10.1038/s41467-021-22890-x.

[61] M St Paul, PS. Ohashi, The roles of CD8(+) T cell subsets in antitumor immunity, Trends Cell Biol. 30 (2020) 695–704, https://doi.org/10.1016/j.tcb.2020.06.003.

[62] S Sakaguchi, T Yamaguchi, T Nomura, M. Ono, Regulatory T cells and immune tolerance, Cell 133 (2008) 775–787, https://doi.org/10.1016/j.cell.2008.05.009.

[63] D Jankovic, Z Liu, WC. Gause, Th1- and Th2-cell commitment during infectious disease: asymmetry in divergent pathways, Trends Immunol. 22 (2001) 450–457, https://doi.org/10.1016/s1471-4906(01)01975-5.

[64] S Hegde, VE Krisnawan, BH Herzog, C Zuo, MA Breden, BL Knolhoff, et al., Dendritic cell paucity leads to dysfunctional immune surveillance in pancreatic cancer, Cancer Cell 37 (2020) 289–307, https://doi.org/10.1016/j.ccell.2020.02.008, e9.

[65] E Lakatos, MJ Williams, RO Schenck, WC Cross, J Househam, L Zapata, et al., Evolutionary dynamics of neoantigens in growing tumors, Nat. Genet. 52 (2020) 1057–1066, https://doi.org/10.1038/s41588-020-0687-1.

[66] D Marinelli, M Mazzotta, S Scalera, I Terrenato, F Sperati, L D'Ambrosio, et al., KEAP1-driven co-mutations in lung adenocarcinoma unresponsive to immunotherapy despite high tumor mutational burden, Ann. Oncol. 31 (2020) 1746–1754, https://doi.org/10.1016/j.annonc.2020.08.2105.

[67] H Chen, J Carrot-Zhang, Y Zhao, H Hu, SS Freeman, S Yu, et al., Genomic and immune profiling of pre-invasive lung adenocarcinoma, Nat. Commun. 10 (2019) 5472, https://doi.org/10.1038/s41467-019-13460-3.

[68] W-C Lee, A Reuben, X Hu, N McGranahan, R Chen, A Jalali, et al., Multiomics profiling of primary lung cancers and distant metastases reveals immunosuppression as a common characteristic of tumor cells with metastatic plasticity, Genome Biol. 21 (2020) 271, https://doi.org/10.1186/s13059-020-02175-0.

[69] T Davoli, H Uno, EC Wooten, SJ. Elledge, Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy, Science (2017) 355, https://doi.org/10.1126/science.aaf8399.

[70] M Jamal-Hanjani, GA Wilson, N McGranahan, NJ Birkbak, TB Watkins, S Veeriah, et al., Tracking the evolution of non-small-cell lung cancer, N. Engl. J. Med. 376 (2017) 2109–2121, https://doi.org/10.1056/NEJMoa1616288.

[71] T Shibue, RA. Weinberg, EMT, CSCs, and drug resistance: the mechanistic link and clinical implications, Nat. Rev. Clin. Oncol. 14 (2017) 611–629, https://doi.org/10.1038/nrclinonc.2017.44.

[72] Y Ge, NC Gomez, RC Adam, M Nikolova, H Yang, A Verma, et al., Stem cell lineage infidelity drives wound repair and cancer, Cell 169 (2017) 636–650, https://doi.org/10.1016/j.cell.2017.03.042, e14.

[73] A Wu, J Wei, L-Y Kong, Y Wang, W Priebe, W Qiao, et al., Glioma cancer stem cells induce immunosuppressive macrophages/microglia, Neuro Oncol. 12 (2010) 1113–1125, https://doi.org/10.1093/neuonc/noq082.

[74] RS Riley, CH June, R Langer, MJ. Mitchell, Delivery technologies for cancer immunotherapy, Nat. Rev. Drug Discov. 18 (2019) 175–196, https://doi.org/10.1038/s41573-018-0006-z.

[75] K Suresh, J Naidoo, CT Lin, S. Danoff, Immune checkpoint immunotherapy for non-small cell lung cancer: benefits and pulmonary toxicities, Chest 154 (2018) 1416–1423, https://doi.org/10.1016/j.chest.2018.08.1048.

[76] EC Ko, D Raben, SC. Formenti, The integration of radiotherapy with immunotherapy for the treatment of non-small cell lung cancer, Clin. Cancer Res. 24 (2018) 5792–5806, https://doi.org/10.1158/1078-0432.CCR-17-3620.

[77] NA Rizvi, MD Hellmann, A Snyder, P Kvistborg, V Makarov, JJ Havel, et al., Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer, Science 348 (2015) 124–128, https://doi.org/10.1126/science.aaa1348.

[78] M Binnewies, EW Roberts, K Kersten, V Chan, DF Fearon, M Merad, et al., Understanding the tumor immune microenvironment (TIME) for effective therapy, Nat. Med. 24 (2018) 541–550, https://doi.org/10.1038/s41591-018-0014-x.