



OPEN

Deciphering the heterogeneity dominated by tumor-associated macrophages for survival prognostication and prediction of immunotherapy response in lung adenocarcinoma

Jiazheng Sun¹, Hehua Guo¹, Yalan Nie¹, Sirui Zhou¹, Yulan Zeng¹✉ & Yalu Sun²✉

Tumor-associated macrophages (TAMs) are a specific subset of macrophages that reside inside the tumor microenvironment. The dynamic interplay between TAMs and tumor cells plays a crucial role in the treatment response and prognosis of lung adenocarcinoma (LUAD). The study aimed to examine the association between TAMs and LUAD to advance the development of targeted strategies and immunotherapeutic approaches for treating this type of lung cancer. The study employed single-cell mRNA sequencing data to characterize the immune cell composition of LUAD and delineate distinct subpopulations of TAMs. The “BayesPrism” and “Seurat” R packages were employed to examine the association between these subgroups and immunotherapy and clinical features to identify novel immunotherapy biomarkers. Furthermore, a predictive signature was generated to forecast patient prognosis by examining the gene expression profile of immunotherapy-associated TAMs subsets and using 104 machine-learning techniques. A comprehensive investigation has shown the existence of a hitherto unidentified subgroup of TAMs known as RGS1 + TAMs, which has been found to have a strong correlation with the efficacy of immunotherapy and the occurrence of tumor metastasis in LUAD patients. CD83 was identified as a distinct biomarker for the expression of RGS1 + TAMs, showcasing its potential utility as an indicator for immunotherapeutic interventions. Furthermore, the prognostic capacity of the RTMscore signature, encompassing three specific mRNA (NR4A2, MMP14, and NPC2), demonstrated enhanced robustness when contrasted against the comprehensive collection of 104 features outlined in the published study. CD83 has potential as an immunotherapeutic biomarker. Meanwhile, The RTMscore signature established in the present study might be beneficial for survival prognostication.

Keywords Lung adenocarcinoma, Tumor-associated macrophages, Prognostic signature, Immunotherapy, Tumor microenvironment

Lung cancer ranks among the leading causes of mortality globally about cancer. The prevalence of lung adenocarcinoma (LUAD) has steadily risen, solidifying its status as one of the predominant subtypes of lung cancer in the current era. The utilization of immune checkpoint inhibitors (ICIs) in cancer therapy has recently witnessed a significant rise. The utilization of PD-1 inhibitors, combined with chemotherapy, has been employed to manage advanced or metastatic LUAD. The primary mode of action of PD-1 inhibitors is disrupting the binding between PD-1 and its corresponding ligand, PD-L1, thereby enabling the patient's immune system to selectively attack malignant cells¹⁻³. The efficacy of CTLA-4 and PD-1 inhibitors has been investigated in clinical trials as a potential therapeutic approach for lung cancer⁴. Nevertheless, as a result of the inherent unpredictability associated with immune escape mechanisms and the complex tumor microenvironment (TME), it is important to note

¹Liyuan Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. ²Affiliated Hospital of Jining Medical University, Jining, China. ✉email: 1989LY0551@hust.edu.cn; 18487344536@163.com

that only a specific subgroup of patients has the potential to achieve a curative response. On the contrary, some individuals may have acquired unforeseen resistance to ICIs⁵.

While previous studies have suggested that some indicators, such as tumor PD-L1 expression level⁶ and tumor mutational burden⁷, could potentially serve as predictors for the response to immunotherapy in LUAD⁸, the practical applicability of these predictions has not been consistently reliable. Instances of hyperprogressive disease after atezolizumab therapy have been documented, hence emphasizing the intricate and dynamic cellular interactions within the TME^{9,10}.

The TME has diverse components, such as tumor cells, immune cells, blood arteries, fibrous tissue, and other elements¹¹. The efficacy of immunotherapy was greatly influenced by the interplay between these components. A more detailed characterization of the various cell types inside the TME and their interactions will enable researchers to gain a deeper understanding of the immunological status of tumors and facilitate the identification of novel biomarkers.

Tumor-associated macrophages (TAMs) have been found to play a pivotal role in the TME. On one side, TAMs possess the ability to absorb and eliminate debris originating from tumor cells, hence facilitating the proliferation and spread of tumor cells¹². In contrast, TAMs were found to emit a considerable quantity of growth factors and cytokines, including vascular endothelial growth factor (VEGF) and tumor necrosis factor-alpha (TNF- α), to promote tumor angiogenesis and assist invasive tumor growth¹³.

As studies progressed, it became evident that TAMs possess anti-tumor characteristics¹³. The activation of TAMs was observed to induce alterations in their secretory profile, producing anti-tumor molecules that effectively suppressed tumor development and metastasis¹⁴. TAMs exhibit tight interactions with immune cells, hence influencing immunological responses, anti-inflammatory reactions, and the overall advancement of tumors¹⁵. It is noteworthy to acknowledge that the roles and phenotypes of TAMs exhibit a high degree of complexity, which is governed by multiple factors, including tumor type, cytokines present in the TME, and immunological signals¹⁶. Hence, comprehending and intervening in the mechanisms of TAMs can potentially facilitate the development of novel therapeutic approaches for combating tumors.

This study provided a comprehensive description of the immune cell composition in LUAD and characterized specific subpopulations of TAMs by analyzing single-cell mRNA sequencing (scRNA-seq) data. Additionally, the study made the novel finding of a previously unidentified subset of TAMs as RGS1 + TAMs that was strongly correlated with the therapeutic advantages of immunotherapy in patients with LUAD. The marker CD83 expressed by TAMs was identified as a means of distinguishing the appropriate population for immunotherapy. Furthermore, a predictive signature was developed to forecast the prognosis of patients. This was achieved through an analysis of the gene expression profile of RGS1 + TAMs and the use of diverse machine-learning techniques. The specific process of the study is shown in Fig. 1.

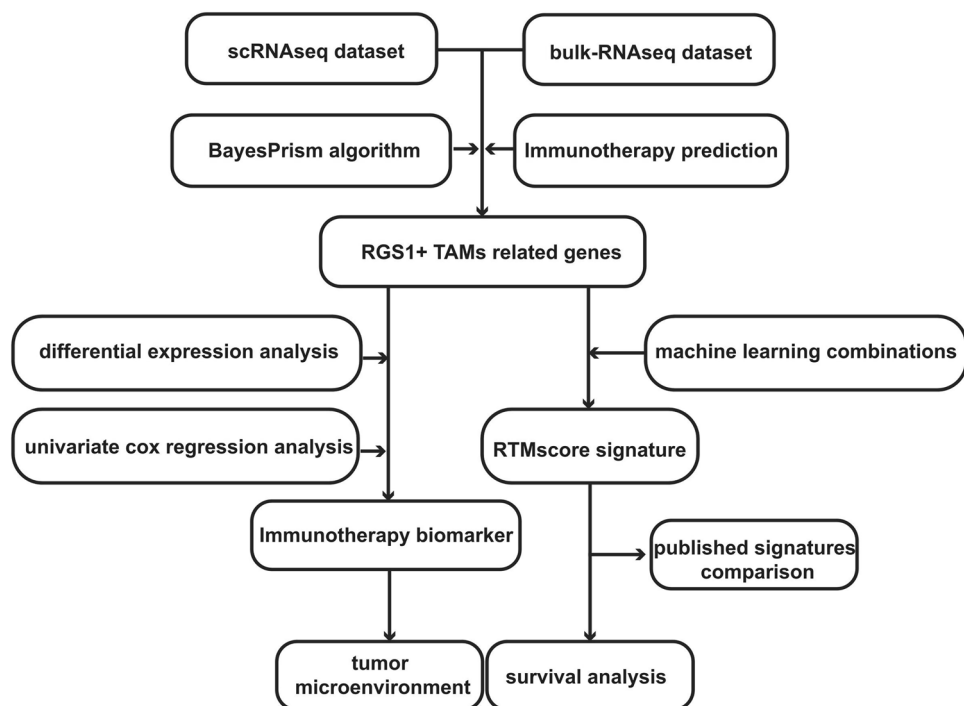


Figure 1. The study's flowchart diagram. Flow chart of the systematic identification of immunotherapy biomarker and the development of the RTMscore signature in LUAD. LUAD, lung adenocarcinoma; scRNAseq, single-cell mRNA sequencing; TAMs, tumor-associated macrophages; RTMscore, RGS1 + TAMs derived-genes score.

Methods

Data acquisition and processing

The study utilized normalized gene expression data and healthcare-related information, including healthcare status, disease diagnosis, treatment options, and medications, for a total of 493 patients diagnosed with LUAD. This data was obtained from The Cancer Genome Atlas (TCGA) (<https://www.tcgabio.org>) database¹⁷. The dataset from TCGA served as the training set for developing the RGS1 + TAMs derived genes score (RTMscore) signature. The criteria for selection were as follows: 1. The survival information of the patients was documented. 2. The individuals in question were diagnosed with LUAD. The GSE30219¹⁸, GSE3141¹⁹, GSE72094²⁰, and GSE50081²¹ datasets, which consist of RNA-Seq data and complete clinical data (The definitive clinical features are listed in Supplementary Table 1), were acquired from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>). These datasets were utilized as a validation cohort to evaluate the robustness and applicability of the RTMscore signature.

The researchers obtained single-cell RNA transcriptome data from patients with LUAD from the Gene Expression Omnibus (GEO) dataset GSE117570²².

The present study utilized immunotherapy datasets including treatment information and RNA expression data obtained from various reputable sources, including online databases and published studies. Specifically, the datasets employed in this investigation encompassed metastatic urothelial carcinoma (mUC) dataset (IMvigor210)²³, metastatic gastric cancer (mGC) dataset (Kim2018)²⁴, non-small cell lung cancer (NSCLC) dataset (GSE126044)²⁵, chronic lymphocytic leukemia (CLL) dataset (GSE148476) and melanoma datasets (melanoma1.GSE91061²⁶, melanoma2.GSE35640²⁷, and melanoma3.GSE115821²⁸) (Refer to Supplementary Table 2 for details). The platform documentation associated with the Bioconductor annotation program was utilized to annotate the GEO dataset. The TPM expression values were derived from the FPKM expression values. The process of converting Count expression values to Transcripts Per Million (TPM) expression values was performed using the “IOBR” R package.

Single-cell RNA sequencing analysis

To ensure the utilization of scRNA-seq data of superior quality, it was imperative to subject it to processing and analysis using the “Seurat” R package, incorporating meticulous filtering techniques. Cells of low quality were eliminated from the dataset through the employment of specific quality criteria, including cell quantity, gene number, and the number of unique molecular identifiers (UMIs) detected. The inclusion criteria for genes in this study required their expression to be observed in a minimum of three individual cells. Cells that exhibited expression of fewer than 50 genes were excluded from the analysis. Moreover, cells that demonstrated expression of mitochondrial genes exceeding 5% were also excluded from the dataset.

The scRNA-seq data underwent normalization using the “NormalizeData” function. Subsequently, the data was transformed into Seurat objects, and the “FindVariableFeatures” tool was employed to identify the first 1500 highly variable genes. Subsequently, the “RunPCA” function from the “Seurat” R package was utilized to execute principal component analysis (PCA) to reduce the dimensionality of the scRNA-seq data, focusing on the top 1500 genes. The JackStraw analytic method was utilized to discover principal components (PCs) that were deemed important. From this analysis, the top 15 PCs were chosen for further investigation in cell clustering. The selection of these PCs was based on their ability to explain a substantial part of the variance.

The functions “FindNeighbors” and “FindClusters” inside the “Seurat” package were employed to conduct cell clustering analysis. The construction of the k-Nearest Neighbour Graph is based on the Euclidean distance calculated in Principal Component Analysis (PCA). The “FindNeighbors” function is utilized to identify the nearest neighbors for each element in the image. Next, the “RunTSNE” function is employed to execute t-distributed random neighbor embedding (t-SNE). The process of cell aggregation was successfully proven by the use of t-SNE-1 and t-SNE-2.

The primary objective of the initial cell annotation was to analyze and categorize the prevailing cell types inside the TME. These cell types were identified based on specific markers associated with each group. The dominating cell types detected were epithelial cells, characterized by the presence of markers such as EPCAM, CDH1, KRT7, and KRT19. Immune cells were also identified using markers such as PTPRC, CD68, and JCHAIN. Additionally, stromal cells were identified based on the presence of the PECAM1 marker.

In addition, immunological cells underwent processing for extraction and reaggregation, using the established Seurat standard protocol. The annotation of clusters was conducted using reference data from the Human Cell Atlas and was further refined through manual adjustment based on specific cell-specific biomarkers. These biomarkers included CD79A for B cells, CD1C and FCER1A for dendritic cells, CD69, LYZ, LGMN, CSF1R, and CD14 for macrophages, S100A12, FCN1, and S100A9 for monocytes, NKG7, KLRD1, and KLRB1 for NK cells, JCHAIN, IGKC, and IGHG1 for plasma cells, and CD3E, IL7R, CD40LG, CD8A, and CCL5 for T cells.

Then, tumor-derived macrophages were isolated and subsequently reaggregated using the “Seurat” R package. The reaggregated macrophages were then subjected to further analysis. The differential expression of genes among cell subpopulations was investigated using the “FindAllMarkers” function from the “Seurat” R package. The Wilcoxon test was employed to detect these differences. The “ClusterGVis” R package was utilized to conduct gene expression trends analysis and functional enrichment analysis.

BayesPrism algorithm

BayesPrism algorithm utilizes RNA-seq samples from corresponding or comparable tissue types to unravel Bulk RNA-seq and spatial transcriptomics for cell type identification and gene expression analysis. It utilizes scRNA-seq data as prior knowledge to calculate the combined posterior distribution of cell type scores and cell type-specific gene expression within each batch.

BayesPrism algorithm comprises a deconvolution module and an integrated learning module. The deconvolution module creates a prior model using the cell type-specific expression profile of single-cell RNA sequencing data. It then estimates the cell type composition of RNA sequencing expression and the posterior distribution of cell type-specific gene expression for several tumor or non-tumor samples. The embedded learning module utilizes the expectation maximization (EM) technique to estimate tumor expression by combining malignant gene programs linearly. This estimation is based on the inferred expression level and the proportion of non-malignant cells calculated by the convolutional module.

Collection of biomarkers in cancer immunotherapy

The relationship between the mRNA level of CD83 and immune cell infiltration was investigated based on the TIMER algorithm²⁹, CIBERSORT algorithm³⁰, quantiseq algorithm³¹, MCPcounter³², and EPIC algorithm³³. Furthermore, the ImmuneScore, StromalScore, ESTIMATEscore, and Tumorpurity were determined from the analysis of distinct gene expression characteristics exhibited by immune and stromal cells, utilizing the ESTIMATE algorithm³⁴.

In addition, a total of ten immunotherapeutic biomarkers were included in the study. The "easier" package³⁵ was utilized for the computation of various immunological parameters, including Cytotoxic activity (CYT)³⁶, IFN γ signature (IFN γ)³⁷, Roh immune score (Roh_IS)³⁸, chemokine signature (chemokines)³⁹, Davoli immune signature (Davoli_IS)⁴⁰, extended immune signature (Ayers_expIS)³⁷, T cell-inflamed gene expression profile (GEP)³⁷, immune resistance program (RIR)⁴¹, and tertiary lymphoid structure (TLS)⁴¹. TIDE scores were obtained from the TIDE website (<http://tide.dfci.harvard.edu/>)⁴².

Integration of machine learning algorithms

To enhance the precision and consistency of the RTMscore signature, we incorporated ten machine-learning algorithms into our analysis. These algorithms encompass random survival forest (RSF)⁴³, elastic network (Enet)⁴⁴, Lasso⁴⁴, Ridge⁴⁴, Stepwise Cox⁴⁵, CoxBoost⁴⁶, partial least squares regression for Cox⁴⁷, supervised principal components (SuperPC)⁴⁸, generalized boosted regression modeling (GBM)⁴⁹, and survival support vector machine (survival-SVM)⁵⁰. Several algorithms have demonstrated the capability of doing feature selection, including Lasso, Stepwise Cox, CoxBoost, and RSF. Therefore, we integrated these algorithms to produce a consensus model. A total of 104 algorithm combinations were performed to construct prediction models using the ten-fold cross-validation technique.

Validation and comparison of RTMscore signature

The research conducted an extensive review of existing literature about illness prediction models specifically connected to LUAD. The study then proceeded to compare the properties of the RTMscore with those of the published models to evaluate the predictive capabilities of the RTMscore signature. After excluding articles that did not provide clear prediction model formulas and articles that did not have matching gene expression data in the training and validation groups, a total of 102 prediction characteristics related to LUAD were selected. These characteristics include cuproptosis, ferroptosis, autophagy, aging, epithelial-mesenchymal transition, acetylation, amino acid metabolism, anoikis, DNA repair, fatty acid metabolism, hypoxia, inflammatory response, N⁶-methyladenosine, mitochondrial homeostasis, and mTOR, etc. (Reference to Supplementary Table 3 for more details). The scores were computed utilizing the algorithms outlined in the scholarly articles, and the C-indices of all prognostic indicators were calculated based on the log₂-transformed TPM gene expression levels obtained from the TCGA database.

Statistical analysis

The R package "DEseq2" was utilized to extract the mRNAs that exhibited differential expression between lung cancer samples and normal samples in the TCGA-LUAD dataset (log₂FoldChange = 1.5 and padj = 0.05).

The prognostic value of the RTMscore signature was assessed by employing time-dependent receiver operating characteristic curves, utilizing the 'timeROC' R package. Statistical differences between groups for variables that follow a normal distribution were assessed using two-tailed t-tests. The Wilcoxon test was employed to ascertain statistical differences between groups for variables that were not normally distributed. The statistical analyses were conducted using the R program (version 4.1.2).

Results

Identification of the gene expression profiles associated with indicators of TAMs

The gene expression patterns of 5783 cells were analyzed in this study. The dataset used for analysis consisted of 4 LUAD samples and 4 normal samples of scRNA-seq dataset GSE131907. The data underwent filtration and extraction processes, resulting in the identification of 1500 variable genes. These genes were then used for reducing dimensionality and aggregating the 18 cell groups. The identification of immune, epithelial, and stromal cells was conducted using particular cell biomarkers (Figs. 2A and Supplementary Fig. 1A,B). This observation warrants further investigation.

The Human Primary Cell Atlas reference dataset was utilized to annotate a total of 4826 immune cells, along with the assessment of cell-specific markers' expression (Figs. 2B and Supplementary Fig. 1C,D). To further investigate the topic, 468 TAMs were isolated, given their significant role in the processes of carcinogenesis, development, and metastasis. A total of seven clusters were identified and subsequently labeled as follows: RGS1 + TAMs (TAM C0), IL1B + TAMs (TAM C1), PLAC8 + TAMs (TAM C2), IFI27 + TAMs (TAM C3), RP11-598F7.3 + TAMs (TAM C4), ISG20 + TAMs (TAM C5), and MIR3945HG + TAMs (TAM C6) (Fig. 2C,D).

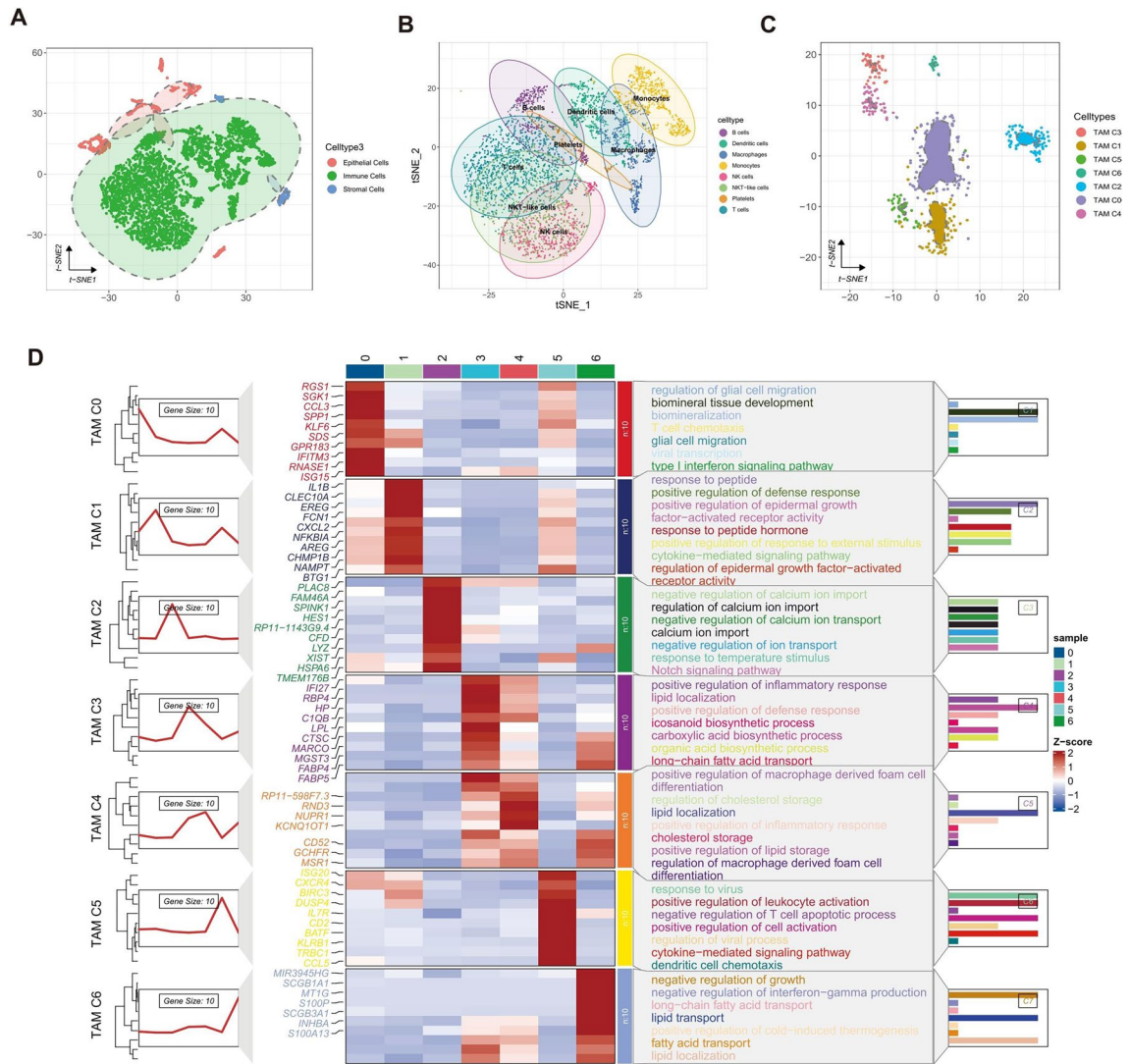


Figure 2. Deciphering of heterogeneity dominated by TAMs in LUAD. (A) t-SNE plot displaying the cell clusters in the microenvironment of LUAD. (B) t-SNE plot displaying the composition of immune cells in the microenvironment of LUAD. (C) t-SNE plot displaying the composition of TAMs in the microenvironment of LUAD, colored according to cell types. (D) The graph consists of three components. Left, lineplot displaying the expression trend of marker genes in different TAMs clusters. Middle, heatmap displaying the expression profiles of the top 10 genes ranked by LogFC of each TAMs cluster. Right, enriched GO terms for marker genes of each TAMs cluster.

Immunological infiltration of RGS1 + TAMs significantly correlates with prognosis and immunotherapy benefit in LUAD patients

The "BayesPrism" R package was utilized to ascertain the relative abundance of the 7 distinct types of TAMs present in both the GSE126044 and TCGA-LUAD cohort specimens. The study aimed to examine the correlation between the abundance of 7 distinct types of TAMs and the response to immunological treatment, which revealed a noteworthy association between the abundance of RGS1 + TAMs and the responsiveness of immune treatment (Fig. 3A).

To gain a clearer understanding of the correlation between the abundance of RGS1 + TAMs and other clinical variables, the study aimed to quantify the proportions of these variables in two subtypes of LUAD patients. The samples were categorized into two groups using the cutoff value of the abundance of RGS1 + TAMs obtained by the "survminer" R package. The results indicated that there was no notable association between the abundance of RGS1 + TAMs and clinical characteristics. (Fig. 3B).

To evaluate the potential differential function of RGS1 + TAMs in the presence of specific mutations, a Kaplan–Meier analysis was conducted to assess the survival disparities among patients with various gene mutations based on different abundance of RGS1 + TAMs. The results revealed that an increased abundance of RGS1 + TAMs in patients with EGFR mutations and KRAS mutations was associated with prolonged survival (Fig. 3C).

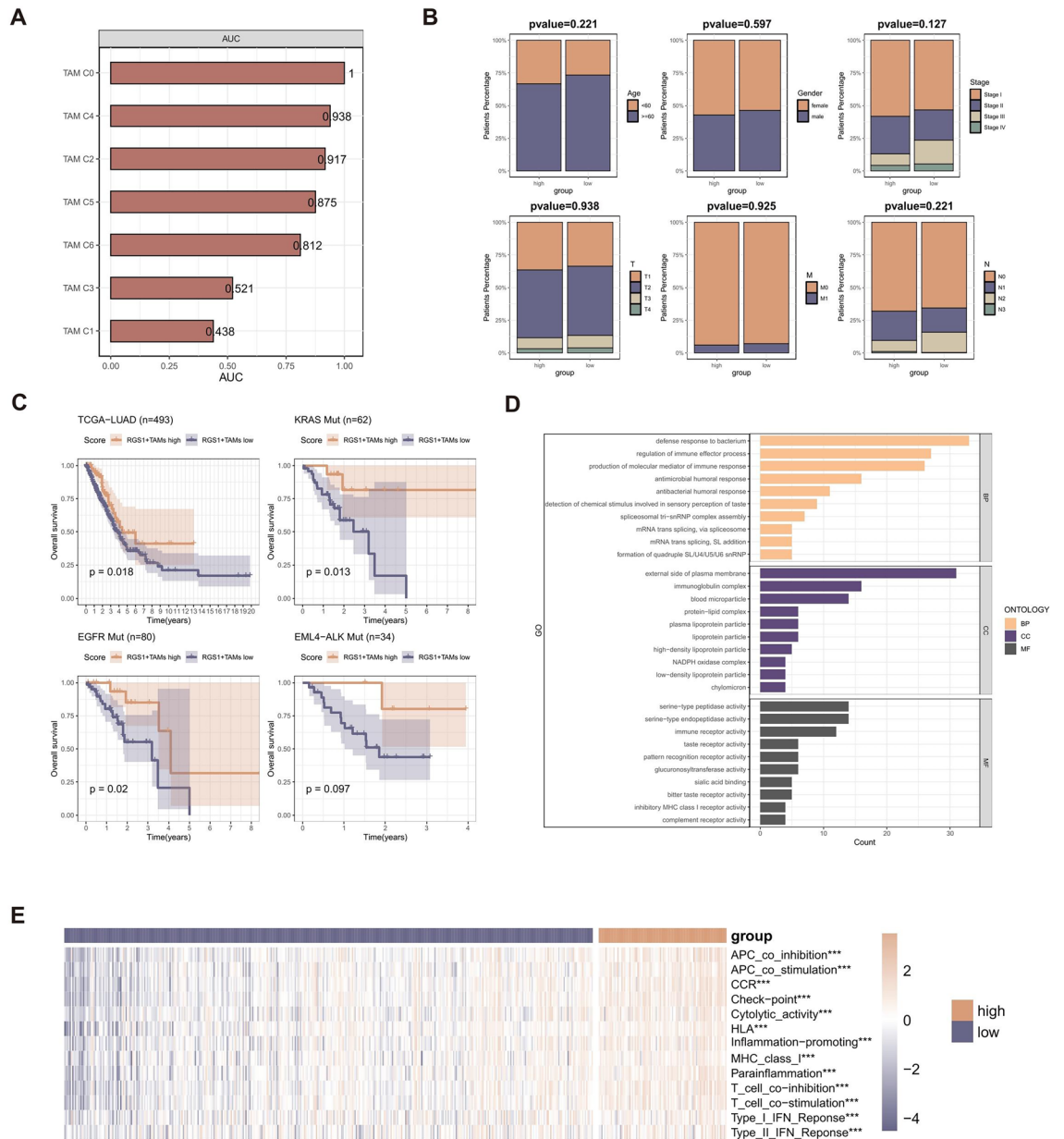


Figure 3. The relationship between the abundance of these RGS1 + TAMs and the clinical characteristics and immunotherapeutic characteristics. **(A)** Barplot displaying the area under ROC curves of the TAMs to predict the benefits of immunotherapy. **(B)** Comparison of proportions of subsets divided by age, gender, stage, T, N, and M between the group with high abundance of RGS1 + TAMs and low abundance of RGS1 + TAMs; the p-values were calculated based on the Wilcoxon rank-sum test. **(C)** Kaplan–Meier curves of OS according to the abundance of RGS1 + TAMs in the TCGA cohort (n = 493), EGFR mutation cohort (n = 80), KRAS mutation cohort (n = 62), and EML4-ALK mutation cohort (n = 34); the p values were calculated based on log-rank test. **(D)** GO enrichment analysis of the RGS1 + TAMs-related genes. MF, molecular function; CC, cellular component; BP, biological process **(E)** Heatmap displaying the correlation between the abundance of RGS1 + TAMs and 13 immune-associated processes.

In light of the distinct prognoses observed in the two subtypes of LUAD, an analysis was conducted to identify differentially expressed genes (DEGs) that could elucidate significant variations in molecular function or components of the TME between these subtypes. The differential gene expression analysis was conducted to identify enriched Gene Ontology (GO) terms. The results revealed that the most significantly enriched terms were associated with immune-related functions, including immune receptor activity and inhibitory MHC class I receptor activity. These findings suggest distinct changes in the immunological ecology within the tumor between the two subtypes (Fig. 3D). The GSVA method was employed to compute scores for different immune-related functional signatures. LUAD patients with an increased abundance of RGS1 + TAMs demonstrated elevated immune functional scores compared to those with a lower abundance (Fig. 3E).

The investigation of CD83's association with immunological features and its potential as an immunotherapeutic biomarker

The study conducted further analysis on the possibility of marker genes associated with RGS1 + TAMs as biomarkers for immunotherapy, given the strong correlation between RGS1 + TAMs and immunotherapy. With the intersection of the DEGs observed in lung cancer samples compared to normal samples, a total of 8 overlapping RGS1 + TAMs were extracted for subsequent analysis (Fig. 4A). Two genes, CD83 and A2M, were identified using a univariate regression analysis conducted on the hub genes (Supplementary Table 4).

Among these, CD83 had been previously identified as a distinctive marker for dendritic cells⁵¹. In recent years, a correlation was established between macrophages and CD83, which had been recognized as a novel immunological checkpoint in macrophages that played a role in the anti-inflammatory response⁵². Furthermore, it had been observed that CD83 had a role in the facilitation and expeditiousness of wound healing through the activation of macrophages that promoted wound healing⁵³. The study aimed to further investigate the association between CD83 and LUAD. We investigated the clinical relevance of LUAD expression in patients who were diagnosed with LUAD. The Kaplan–Meier survival curves demonstrated favorable prognostic outcomes in LUAD patients who had increased levels of CD83 mRNA (Fig. 4B). The TCGA dataset was utilized to acquire data (Fig. 4C), which revealed a notable decrease in mRNA expression levels of CD83 in LUAD samples. The immunohistochemistry (IHC) analysis of the HPA dataset revealed a comparable dysregulation in the protein level of CD83 (Fig. 4D).

The TME exhibited a significant association with the processes of cancer initiation, advancement, and the implementation of therapeutic strategies⁵⁴. This study aimed to examine the correlation between CD83 and immunological characteristics to evaluate the involvement of CD83 in LUAD TME. A significant association was discovered between the expression level of CD83 and the extent of immune cell infiltration (Fig. 5A).

In addition, a statistically significant negative association was indicated between the expression of CD83 and the level of tumor purity (Fig. 5B). Moreover, a strong positive correlation was observed between the expression of CD83 and ImmuneScore, StromalScore, and ESTIMATEScore (Fig. 5B). The study also examined the association between CD83 and established immune modulators, including CYT, TLS, IFN, Davoli_IS, Roh_IS, GEP, and Ayers_expIS, as well as chemokines, RIR and TIDE (Fig. 5C). It was worth mentioning that the TCGA-LUAD

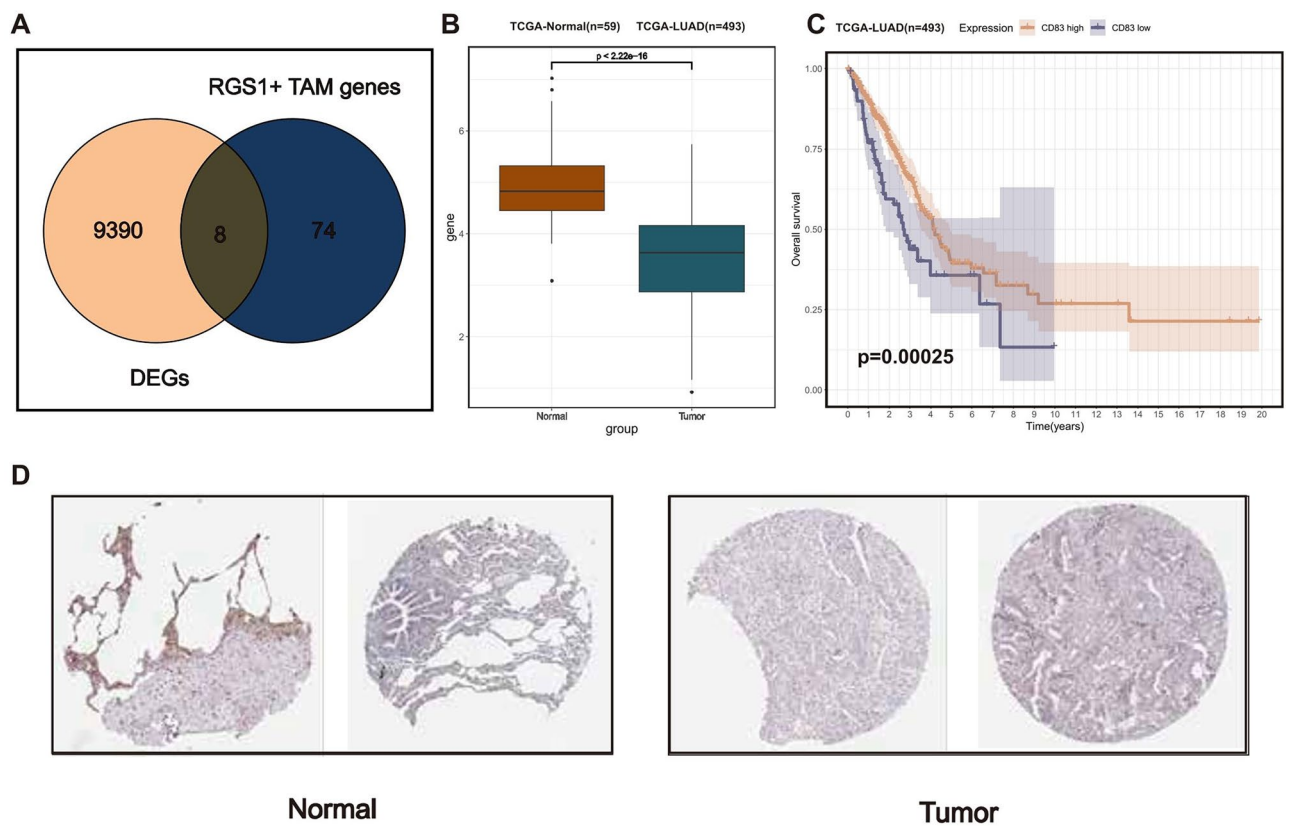


Figure 4. The identification of CD83. (A) Venn diagram representing the meeting point of RGS1 + marker genes and differently expressed genes between LUAD tissues and adjacent normal tissues in the TCGA cohort. (B) Different expression of CD83 between lung adenocarcinoma tissues and adjacent normal tissues in the TCGA cohort; the p-value was calculated based on the Wilcoxon rank-sum test. (C) Kaplan–Meier survival curve of OS between LUAD patients subjected to the high mRNA expression level of CD83 and the low mRNA expression level of CD83; the p-value was calculated based on the log-rank test. (D) Validation of the expression of CD83 in the normal and tumor samples level by the Human Protein Atlas database (immunohistochemistry).

cohort's subgroup with a high expression level of CD83 exhibited significantly elevated levels of many factors, including CYT, TLS, IFN, GEP, Davoli_IS, Roh_IS, and Ayers_expIS, which were known to be associated with possible benefits of immunotherapy. The presence of low RIR, TIDE, and chemokine levels suggests a reduced probability of immunological escape.

The findings presented in this study suggest that CD83 has potential as an immunotherapeutic biomarker. Furthermore, it was observed that measuring CD83 mRNA levels may be more feasible compared to other immunotherapeutic biomarkers that require extensive experimentation and complex mathematical analyses but are seldom utilized in clinical settings.

Subsequently, given the strong correlation observed between the mRNA expression level of CD83 and tumor-infiltrating immune cells (TIICs), pathways involved in immunotherapeutic function, expression of immune checkpoints, and predictors of immunotherapy response, cohorts undergoing immunotherapy were included in the study to confirm the predictive significance of CD83 for the response to immunotherapy.

The responsive group for NSCLC patients in the NSCLC cohort had significantly higher levels of CD83 mRNA expression than the non-responsive group. The same trend was observed in the levels of CD83 mRNA expression of patients with chronic lymphocytic leukemia in the CLL cohort, patients with metastatic gastric cancer in the mGC cohort, and patients with metastatic urothelial carcinoma in the mUC cohort (Fig. 5D).

Meanwhile, in the cohorts melanoma1, melanoma2, and melanoma3, it was shown that the mRNA expression level of CD83 exhibited a statistically significant increase in the responsive group compared to the non-responsive group among melanoma patients (Fig. 5D). In addition, a notable correlation was observed between the mRNA expression level of CD83 and conventional immunological checkpoint markers (Fig. 5E).

The receiver operating characteristic (ROC) analysis conducted in the study showed that the mRNA level of CD83 exhibited a consistent ability to predict the efficacy of immunotherapy-based treatment. This finding was further supported by the analysis of gene expression datasets, including cohort mGC, NSCLC, melanoma1, CLL, melanoma2, and melanoma3, which yielded ROC values of 0.707, 0.875, 0.670, 0.701, 0.695, and 0.863, respectively (Fig. 5F). In terms of differentiating between Nivolumab responders and non-responders, the mRNA level of CD83 exhibited a substantially higher value (AUC = 0.792) compared to PDL1 (Fig. 5G).

Construction and validation of the RTMscore signature based on the immune heterogeneity dominated by RGS1 + TAMs

A total of 104 prediction signatures were derived by utilizing ten different machine learning methods, employing data from 82 RGS1 + TAMs-associated genes. The C-index was then computed for each of these signatures across all validation groups (Fig. 6A). The findings indicated that the RSF + StepCox approach demonstrated substantial predictive capabilities. This approach involved utilizing the RSF algorithm to identify four useful mRNA molecules (Fig. 6B). Subsequently, a Stepwise Cox proportional hazards regression analysis was conducted to choose three mRNA molecules (NR4A2, MMP14, and NPC2) as the final predictive signature (Fig. 6C). The equation that has been derived is as follows:

$$\text{RTMscore} = \text{NR4A2} \times (0.1235840) + \text{MMP14} \times (0.2199992) + \text{NPC2} \times (-0.2247780).$$

The samples were categorized into groups using the cutoff value of RTMscore obtained by the "survminer" R package. The study conducted a Kaplan–Meier analysis to compare the high-RTMscore group with the low-RTMscore group (Fig. 6D). The results showed a substantial correlation between the RTMscore signature and OS in patients with LUAD from the TCGA-LUAD group. This association was further confirmed in independent cohorts, including GSE30219, GSE3141, GSE72094, and GSE50081.

The area under the receiver operating characteristic curve (AUC) values for the RTMscore signature were calculated for different time intervals (1-, 2-, 3-, 4-, and 5 years) in the TCGA-LUAD group. The estimated AUC values were 0.709, 0.657, 0.653, 0.651, and 0.545, respectively (Fig. 6E). These results indicate that the RTMscore signature has potential as a prediction tool for patients with LUAD. The validation of the model was performed in multiple cohorts, including GSE30219 (with AUC values of 0.849, 0.717, 0.72, 0.686, and 0.701), GSE3141 (with AUC values of 0.750, 0.713, 0.764, 0.683, and 0.573), GSE72094 (with AUC values of 0.733, 0.705, 0.667, 0.671, and 0.682), and GSE50081 (with AUC values of 0.682, 0.676, 0.675, 0.706, and 0.764).

Furthermore, we conducted a comparison between the predictive value of the RTMscore signature and other clinical factors (Fig. 7A). The C-index of the RTMscore signature exhibited a much greater value compared to other clinical variables, encompassing staging, age, gender, and other relevant factors.

In recent years, predictive signatures have emerged significantly in machine learning-based gene expression analysis. This advancement has enabled the prediction of disease outcomes, facilitating early disease screening and the exploration of new therapeutic approaches. A literature search was conducted to compare the RTMscore signature with previously reported signatures in studies related to the LUAD-associated disease prediction model. After excluding articles that lacked explicit prediction model formulas and did not provide corresponding gene expression data in the training and validation groups, a total of 102 predictive signatures were linked with LUAD. These signatures were identified and validated using data from the TCGA-LUAD, GSE30219, GSE3141, GSE72094, and GSE50081 cohorts. The performance of these signatures was then compared to the C-index of RTMscore. The findings indicate that the RTMscore signature exhibited superior performance compared to the majority of signatures within its respective group, as seen by the results (Fig. 7B).

Discussion

The treatment options for LUAD have become more diverse due to advancements in molecular biology and immunology, shown by the utilization of PD-1 inhibitors in conjunction with chemotherapy. The existence of a wide range of treatment options necessitates the development of more effective and individualized assessment procedures for patients to facilitate informed clinical decision-making. Nevertheless, a dearth of dependable

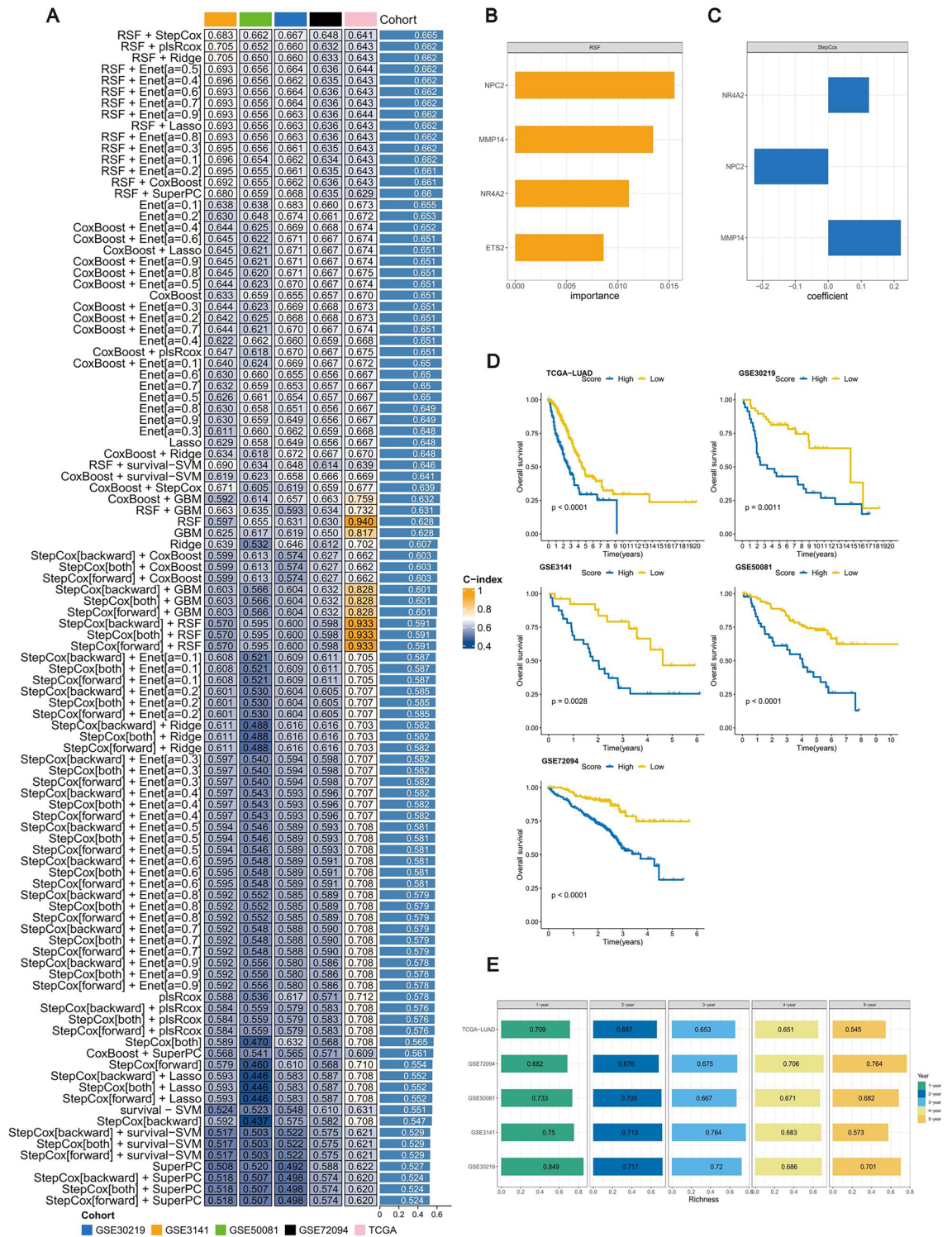


Figure 6. The construction and assessment of RTMscore signature. (A) A total of 104 combinations of machine learning algorithms for the RTMscore signatures via a tenfold cross-validation framework based on the TCGA-LUAD cohort (n = 493). The C-index of each model was calculated across validation datasets, including the GSE30219 (n = 83), GSE3141 (n = 58), GSE72094 (n = 393) and GSE50081 (n = 127) cohorts. (B) The importance of the 4 most valuable mRNAs based on the RSF algorithm. (C) The coefficients of 3 mRNAs were finally obtained in stepwise Cox regression. (D) Kaplan–Meier survival curve of OS between patients subjected to a high score of RTMscore signature and with a low score of RTMscore signature in TCGA-LUAD, GSE30219, GSE3141, GSE72094, and GSE50081 cohorts; the p-values were calculated based on log-rank test. (E) Time-dependent ROC analysis for predicting OS at 1, 2, 3, 4, and 5 years in TCGA-LUAD, GSE30219, GSE3141, GSE72094 and GSE50081 cohorts.

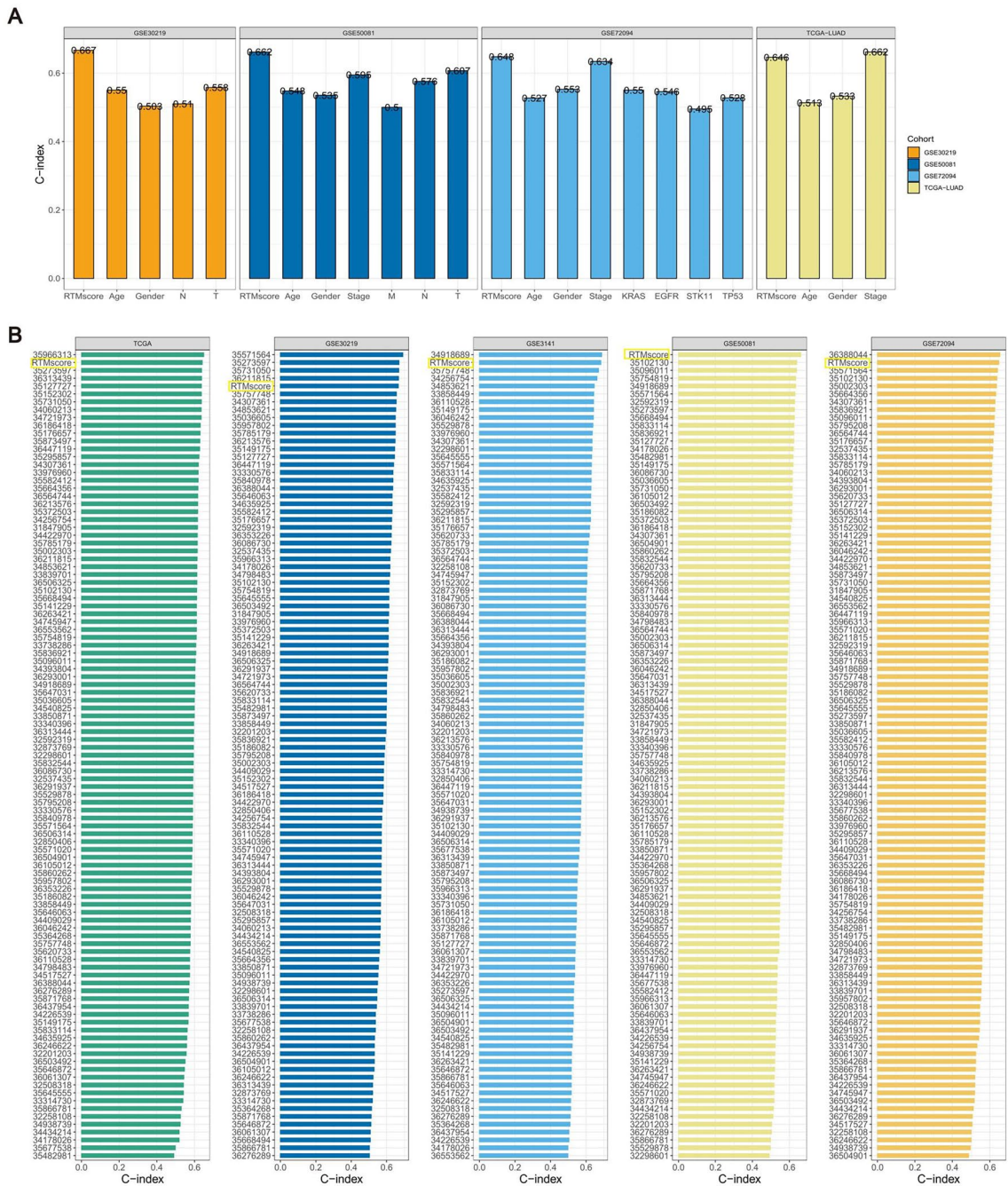


Figure 7. The comparison of the RTMScore signature with clinical characteristics and published signatures. **(A)** The C-index of the RTMScore signature and other clinical characteristics in the TCGA-LUAD, GSE30219, GSE3141, GSE72094, and GSE50081 cohorts. **(B)** The C-index of the RTMScore signature and other signatures developed in the TCGA-LUAD, GSE30219, GSE3141, GSE72094, and GSE50081 cohorts.

prognostic indicators exists to identify individuals with “high-risk” LUAD who could potentially derive advantages from immunotherapy. To address this knowledge gap, we investigated the immune cell composition of LUAD and discerned distinct subpopulations of TAMs, specifically focusing on the characterization of RGS1 + TAMs. Our findings revealed disparities in the abundance of RGS1 + TAMs between the groups that exhibited a response to treatment and those that did not.

Furthermore, pathway analysis was conducted, revealing variations in the immune ecology within the tumor among the two subtypes. Additional examination revealed genes specifically expressed in RGS1 + TAMs. The association between the expression of the CD83 gene and the response to immunotherapy was observed among the subjects.

In addition, studies have shown that several types of immune cells, such as B cells, thymic epithelial cells (TEC), T cells, dendritic cells (DC), and neutrophils, have been identified to have CD83 expression. CD83 is

crucial for activating T cells that control immunological responses in the body's periphery⁵⁵. CD83 is a marker of mature dendritic cells (matDC) due to its increased expression during the maturation of dendritic cells⁵⁶.

Data analysis based on clinical data also showed that patients exhibiting elevated levels of CD83 in the context of mUC, mGC, NSCLC, CLL, and melanoma demonstrated heightened susceptibility to immune checkpoint inhibitors. The immunomodulatory features of CD83 underscore its significant therapeutic potential.

In addition, the research effectively employed a total of 104 machine-learning algorithms to establish a robust RTMscore signature, which was created from genes associated with RGS1 + TAMs. The derivation of this signature was based on the analysis of many data sources, encompassing genetic markers, tumor characteristics, and macrophage-related aspects. The stability and durability of the RTMscore signature were ensured by leveraging the strengths of each algorithm and utilizing ensemble learning techniques. The utilization of the stable RTMscore signature has proven to be highly valuable in the domains of cancer research and precision medicine. The stability of the signature facilitated accurate prognostic predictions and enhanced comprehension of the intricate interplay between the TME and immune cells, hence leading to advancements in therapeutic approaches and eventually benefiting patient outcomes.

Furthermore, the RTMscore signature exhibited independent prognostic value, surpassing those of clinical parameters such as age, stage, and gender. In addition, it was observed that the RTMscore signature demonstrated a higher level of stability in its performance compared to a set of 102 previously published signatures when it came to predicting the prognosis. The majority of the signatures exhibited commendable performance on the training dataset; however, their performance on the validation datasets was subpar. This indicates a need for these signatures to possess greater generalizability. Simultaneously, the integration of different machine-learning techniques contributed to the robustness of our RTMscore signature.

The study undoubtedly has numerous limitations. Initially, validation of the RTMscore signature requires an independent clinical cohort, and additional data gathering is necessary to verify its predictive capacity. Besides, we integrated scRNAseq with bulk RNAseq data using the BayesPrism algorithm due to the scarcity of LUAD single-cell datasets on immunotherapy. Additional research is required to ascertain the function of RGS1 + TAMs in immunotherapy.

Conclusion

In summary, the RTMscore signature introduced in this investigation exhibits promise as a prognostic indicator for patients undergoing treatment for LUAD, with implications for overall survival. Meanwhile, CD83 has the potential as an immunotherapeutic biomarker.

Data availability

The analyzed data could be obtained from the TCGA database (<https://portal.gdc.cancer.gov/>), GEO database (<http://www.ncbi.nlm.nih.gov/geo/>), and TIDE database (<http://tide.dfc.harvard.edu/>). The code applied in the study is available from the corresponding author upon reasonable request.

Received: 8 January 2024; Accepted: 19 April 2024

Published online: 23 April 2024

References

- Zhou, C. *et al.* Sugemalimab versus placebo, in combination with platinum-based chemotherapy, as first-line treatment of metastatic non-small-cell lung cancer (GEMSTONE-302): Interim and final analyses of a double-blind, randomised, phase 3 clinical trial. *Lancet Oncol.* **23**(2), 220–233 (2022).
- Nishio, M. *et al.* Atezolizumab plus chemotherapy for first-line treatment of nonsquamous NSCLC: Results from the randomized phase 3 IMpower132 trial. *J. Thorac. Oncol.* **16**(4), 653–664 (2021).
- Jotte, R. *et al.* Atezolizumab in combination with carboplatin and nab-paclitaxel in advanced squamous NSCLC (IMpower131): Results from a randomized phase III trial. *J. Thorac. Oncol.* **15**(8), 1351–1360 (2020).
- He, Z. *et al.* LRP1B mutation is associated with tumor immune microenvironment and progression-free survival in lung adenocarcinoma treated with immune checkpoint inhibitors. *Transl. Lung Cancer Res.* **12**(3), 510–529 (2023).
- Anagnostou, V. *et al.* Evolution of neoantigen landscape during immune checkpoint blockade in non-small cell lung cancer. *Cancer Discov.* **7**(3), 264–276 (2017).
- Shen, X. & Zhao, B. Efficacy of PD-1 or PD-L1 inhibitors and PD-L1 expression status in cancer: Meta-analysis. *BMJ* **10**(362), k3529 (2018).
- Chan, T. A. *et al.* Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann. Oncol.* **30**(1), 44–56 (2019).
- Maman, S. & Witz, I. P. A history of exploring cancer in context. *Nat. Rev. Cancer* **18**(6), 359–376 (2018).
- Oguri, T. *et al.* A case of hyperprogressive disease following atezolizumab therapy for pulmonary pleomorphic carcinoma with epidermal growth factor receptor mutation. *Respir. Med. Case Rep.* **31**(33), 101405 (2021).
- Fricke, J. *et al.* Hyperprogression on immunotherapy with complete response to chemotherapy in a NSCLC patient with high PD-L1 and STK11. *Medicine (Baltimore)*. **99**(46), e22323 (2020).
- De Visser, K. E. & Joyce, J. A. The evolving tumor microenvironment: From cancer initiation to metastatic outgrowth. *Cancer Cell.* **41**(3), 374–403 (2023).
- Mantovani, A., Marchesi, F., Malesci, A., Laghi, L. & Allavena, P. Tumor-associated macrophages as treatment targets in oncology. *Nat. Rev. Clin. Oncol.* **14**(7), 399–416 (2017).
- Ruffell, B. & Coussens, L. M. Macrophages and therapeutic resistance in cancer. *Cancer Cell.* **27**(4), 462–472 (2015).
- Noy, R. & Pollard, J. W. Tumor-associated macrophages: From mechanisms to therapy. *Immunity* **41**(1), 49–61 (2014).
- Mantovani, A. & Locati, M. Tumor-associated macrophages as a paradigm of macrophage plasticity, diversity, and polarization: Lessons and open questions. *Arterioscler. Thromb. Vasc. Biol.* **33**(7), 1478–1483 (2013).
- Quatromoni, J. G. & Eruslanov, E. Tumor-associated macrophages: Function, phenotype, and link to prognosis in human lung cancer. *Am. J. Transl. Res.* **4**(4), 376–389 (2012).
- Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**(7511), 543–550 (2014).

18. Rousseaux, S. *et al.* Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. *Sci. Transl. Med.* **5**(186), 186 (2013).
19. Bild, A. H. *et al.* Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439**(7074), 353–357 (2006).
20. Schabath, M. B. *et al.* Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma. *Oncogene* **35**(24), 3209–3216 (2016).
21. Der, S. D. *et al.* Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients. *J. Thorac. Oncol.* **9**(1), 59–64 (2014).
22. Song, Q. *et al.* Dissecting intratumoral myeloid cell plasticity by single cell RNA-seq. *Cancer Med.* **8**(6), 3072–3085 (2019).
23. Mariathasan, S. *et al.* TGF β attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* **554**(7693), 544–548 (2018).
24. Kim, S. T. *et al.* Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. *Nat. Med.* **24**(9), 1449–1458 (2018).
25. Cho, J. W. *et al.* Genome-wide identification of differentially methylated promoters and enhancers associated with response to anti-PD-1 therapy in non-small cell lung cancer. *Exp. Mol. Med.* **52**(9), 1550–1563 (2020).
26. Riaz, N. *et al.* Tumor and microenvironment evolution during immunotherapy with nivolumab. *Cell* **171**(4), 934–949.e15 (2017).
27. Ulloa-Montoya, F. *et al.* Predictive gene signature in MAGE-A3 antigen-specific cancer immunotherapy. *J. Clin. Oncol.* **31**(19), 2388–2395 (2013).
28. Auslander, N. *et al.* Publisher correction: Robust prediction of response to immune checkpoint blockade therapy in metastatic melanoma. *Nat. Med.* **24**(12), 1942 (2018).
29. Li, T. *et al.* TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.* **77**(21), e108–e110 (2017).
30. Newman, A. M. *et al.* Determining cell-type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**(7), 773–782 (2019).
31. Finotello, F. & Trajanoski, Z. Quantifying tumor-infiltrating immune cells from transcriptomics data. *Cancer Immunol. Immunother.* **67**(7), 1031–1040 (2018).
32. Becht, E. *et al.* Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* **20**(17), 218 (2016).
33. Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E. & Gfeller, D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *eLife* **6**, e26476 (2017).
34. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **11**(4), 2612 (2013).
35. Lapuente-Santana, Ó., van Genderen, M., Hilbers, P. A. J., Finotello, F. & Eduati, F. Interpretable systems biomarkers predict response to immune-checkpoint inhibitors. *Patterns* **2**(8), 100293 (2021).
36. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**(1–2), 48–61 (2015).
37. Ayers, M. *et al.* IFN- γ -related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Invest.* **127**(8), 2930–2940 (2017).
38. Roh, W. *et al.* Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Sci. Transl. Med.* **9**(379), eaah3560 (2017).
39. Messina, J. L. *et al.* 12-Chemokine gene signature identifies lymph node-like structures in melanoma: potential for patient selection for immunotherapy?. *Sci. Rep.* **24**(2), 765 (2012).
40. Cabrita, R. *et al.* Author correction: Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* **580**(7801), E1–E1 (2020).
41. Jerby-Arnon, L. *et al.* A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. *Cell* **175**(4), 984–997.e24 (2018).
42. Jiang, P. *et al.* Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* **24**(10), 1550–1558 (2018).
43. Ishwaran, H. *et al.* Random survival forests for competing risks. *Biostatistics* **15**(4), 757–773 (2014).
44. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**(1), 1–22 (2010).
45. Núñez, E., Steyerberg, E. W. & Núñez, J. Regression modeling strategies. *Rev. Esp. Cardiol. Engl. Ed.* **64**(6), 501–507 (2011).
46. Binder, H., Allignol, A., Schumacher, M. & Beyersmann, J. Boosting for high-dimensional time-to-event data with competing risks. *Bioinformatics* **25**(7), 890–896 (2009).
47. Lê Cao, K. A., Boitard, S. & Besse, P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinform.* **22**(12), 253 (2011).
48. Bair, E. & Tibshirani, R. Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biol.* **2**(4), e108 (2004).
49. Guo, C. Y. & Chang, K. H. A novel algorithm to estimate the significance level of a feature interaction using the extreme gradient boosting machine. *Int. J. Environ. Res. Public Health* **19**(4), 2338 (2022).
50. Van Belle, V., Pelckmans, K., Van Huffel, S. & Suykens, J. A. K. Improved performance on high-dimensional survival data by application of survival-SVM. *Bioinformatics* **27**(1), 87–94 (2011).
51. Prechtel, A. T. & Steinkasserer, A. CD83: An update on functions and prospects of the maturation marker of dendritic cells. *Arch. Dermatol. Res.* **299**(2), 59–69 (2007).
52. Peckert-Maier, K. *et al.* CD83 expressed by macrophages is an important immune checkpoint molecule for the resolution of inflammation. *Front. Immunol.* **15**(14), 1085742 (2023).
53. Royzman, D. *et al.* Soluble CD83 improves and accelerates wound healing by the induction of pro-resolving macrophages. *Front. Immunol.* **30**(13), 1012647 (2022).
54. Locy, H. *et al.* Immunomodulation of the tumor microenvironment: Turn foe into friend. *Front. Immunol.* **11**(9), 2909 (2018).
55. Peckert-Maier, K. *et al.* Tilting the balance: Therapeutic prospects of CD83 as a checkpoint molecule controlling resolution of inflammation. *Int. J. Mol. Sci.* **23**(2), 732 (2022).
56. Zhou, L. J. & Tedder, T. F. Human blood dendritic cells selectively express CD83, a member of the immunoglobulin superfamily. *J. Immunol.* **154**(8), 3821–3835 (1995).

Acknowledgements

All authors would like to express our sincere thanks for sharing the online databases.

Author contributions

J.S. and H.G. designed and conducted the entire study. J.S. performed the data collection, bioinformatics, and statistical data analysis. Y.N., S.Z. and Y.S. investigated the literature. Y.Z. was responsible for the integrity of the entire study and manuscript review. All authors contributed to the article and approved the submitted version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-60132-4>.

Correspondence and requests for materials should be addressed to Y.Z. or Y.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024