Deciphering the immune heterogeneity dominated by RGS1+ TAMs with prognostic implications and identification of novel immunotherapeutic biomarker CD83 in lung adenocarcinoma

Jiazheng Sun
Huazhong University of Science and Technology

Huehua Guo
Huazhong University of Science and Technology

Yalan Nie
Huazhong University of Science and Technology

Sirui Zhou
Huazhong University of Science and Technology

Yulan Zeng (✉ 1989LY0551@hust.edu.cn)
Huazhong University of Science and Technology

Yalu Sun
Affiliated Hospital of Jining Medical University

Article

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

Posted Date: January 11th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3845801/v1

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

Additional Declarations: No competing interests reported.
Abstract

Background

Tumor-associated macrophages (TAMs) are a specific subset of macrophages that reside inside the tumor microenvironment (TME). 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.

Methods

The study employed scRNA-seq 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 RGS1 + TAMs and using 104 machine-learning techniques.

Results

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 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.

Conclusion

The identified RGS1 + TAMs have substantial implications for the treatment and prognosis of LUAD patients.

1. Introduction

Lung cancer ranks among the leading causes of mortality globally about cancer. The prevalence of 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 (ICI) 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 cells (1–3). The efficacy of CTLA-4 and PD-1 inhibitors has been investigated in clinical trials as a potential therapeutic approach for lung cancer (4). Nevertheless, as a result of the inherent unpredictability associated with immune escape mechanisms and the complex TME, it is important to note 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 immune checkpoint inhibitors (ICI) (5).

While previous studies have suggested that some indicators, such as tumor PD-L1 expression level (6) and tumor mutational burden (7), could potentially serve as predictors for the response to immunotherapy in LUAD (8), the practical applicability of these predictions has not been consistently reliable. Instances of hyper progressive disease subsequent to 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 elements (11). 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.

TAMs have been found to play a pivotal role inside 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 cells (12). 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), in order to promote tumor angiogenesis and assist invasive tumor growth (13).

As studies progressed, it became evident that TAMs possess anti-tumor characteristics (13). The activation of TAMs was observed to induce alterations in their secretory profile, producing anti-tumor molecules that effectively suppressed tumor development and metastasis (14). TAMs exhibit tight interactions with immune cells, hence influencing immunological responses, anti-inflammatory reactions, and the overall advancement of tumors (15). 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 signals (16). 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 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.

2. Methods

2.1 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.tcga.org) database (17). 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 clearly documented. 2. The individuals in question were diagnosed with LUAD. The GSE30219 (18), GSE3141 (19), GSE72094 (20), and GSE50081 (21) datasets, which consist of RNA-Seq data and complete clinical data (The definitive clinical features are listed in Supplementary Table 1–5), 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 GSE117570 (22).

The present study utilized immunotherapy datasets including treatment information and RNA expression data were obtained from various reputable sources, including online databases and published studies. Specifically, the datasets employed in this investigation encompassed IMvigor210 dataset (urothelial carcinoma) (23), Kim2018 dataset (metastatic gastric cancer) (24), GSE126044 (non-small cell lung cancer) (25), GSE91061 (melanoma) (26), GSE148476 (chronic lymphocytic leukemia), GSE35640 (melanoma) (27) and GSE115821 (melanoma (28). The platform documentation associated with the Bioconductor annotation program was utilized for the purpose of annotating 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.

The blood expression quantitative trait loci (eQTL) dataset utilized in this study was obtained from the eQTLGen Consortium (https://eqtlgen.org/). The dataset presented in this study encompasses cis-eQTL data for a total of 16,987 genes, which were obtained from 31,684 blood samples collected from individuals of European descent who were in good health. Significant cis-eQTL results meeting the criteria of a false discovery rate (FDR) below 0.05 and a minimum allele frequency (MAF) above 0.01 were obtained from the analysis.

2.2 Single-cell RNA sequencing analysis
In order 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 1,500 highly variable genes. Subsequently, the "RunPCA" function from the "Seurat" R package was utilized to execute principal component analysis (PCA) for the purpose of reducing the dimensionality of the scRNA-seq data, focusing on the top 1,500 genes. The JackStraw analytic method was utilized in order 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 neighbours for each element in the image. Next, the "RunTSNE" function is employed to execute t-distributed random neighbour 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 analyse 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, characterised 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 and epithelial cells underwent processing for the purpose of 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.

And then, tumor-derived macrophages were isolated and subsequently reaggregated using the "Seurat" R package. The reaggregated macrophages were then subjected to further analysis, employing specific cell-specific biomarkers. These biomarkers included Angio-TAMs (VEGFA, VCAN, FCN1, THBS1), IFN-TAMs
(ISG15, CD86), Inflam-TAMs (CXCL1, CXCL2, CXCL8, CCL3), LA-TAMs (APOC1, APOE, ACP5, FABP5), Prolif-TAMs (CDK1, HMGB1), Reg-TAMs (MRC1), and RTM-TAMs (HES1).

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. The R package "BayesPrism" was utilized to make predictions on the cellular composition and gene expression of specific cell types based on bulk RNA-seq data, with prior knowledge obtained from patient-derived single-cell scRNA-seq data.

2.3 Collection of biomarkers in cancer immunotherapy

The relationship between the mRNA level of CD83 and immune cell infiltration was investigated based on the TIMER algorithm (29), CIBERSORT algorithm (30), quantiseq algorithm (31), MCPcounter (32) and EPIC algorithm (33). 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 algorithm (34).

In addition, a total of ten immunotherapeutic biomarkers were included in the study. The "easier" package (35) was utilized for the computation of various immunological parameters, including Cytotoxic activity (CYT) (36), IFNγ signature (IFNy) (37), Roh immune score (Roh_IS) (38), chemokine signature (chemokines) (39), Davoli immune signature (Davoli_IS) (40), extended immune signature (Ayers_explIS) (37), T cell-inflamed gene expression profile (GEP) (37), immune resistance programme (RIR) (41), and tertiary lymphoid structure (TLS) (41). TIDE scores were obtained from the TIDE website (http://tide.dfci.harvard.edu/) (42).

2.4 Mendelian randomization analysis

The SNPs were well recognized as instrumental variable (IVs). To establish IVs, correlation analysis was employed to identify single nucleotide polymorphisms (SNPs) that were strongly associated with exposure factors. These SNPs were subsequently considered as instrumental variables. The filter criterion was defined as having a p-value less than 5e-08. A study on linkage disequilibrium (LD) was conducted using data from the 1000 Genomes Project. SNPs (R2 > 0.01) located within a 100kb vicinity of each gene's transcription start site (TSS) were excluded. Additionally, the SNPs associated with each expression quantitative trait locus (eQTL) were grouped into clusters, and some of which associated with potential confounding factors were removed. The Wald ratio was employed to assess each SNP in a straightforward manner. In cases where pleiotropy was absent, heterogeneity was evaluated using Cochran Q analysis. If the P-value above 0.05 and no heterogeneity was seen, the fixed-effect IVW method was deemed the primary strategy. The random-effects inverse-variance weighted (IVW) technique was utilized in cases where there was substantial heterogeneity. Two more methods, namely MR-Egger and weighted median, were subsequently utilized for the purpose of conducting sensitivity analysis in the
MR Analysis. The FDR correction method was employed to determine the statistical significance of etiological association features, with a significance threshold set at $P < 0.05$.

### 2.5 Integration of machine learning algorithms

In order 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), Lasso (44), Ridge (44), Stepwise Cox (45), CoxBoost (46), partial least squares regression for Cox (47), supervised principal components (SuperPC) (48), generalised boosted regression modelling (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 in order to produce a consensus model. A total of 104 algorithm combinations were performed in order to construct prediction models using the 10-fold cross validation technique.

### 2.6 Validation and comparison of RTMscore signature

The research conducted an extensive review of existing literature pertaining to illness prediction models specifically connected to LUAD. The study then proceeded to compare the properties of the RTMscore with those of the published models in order 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, ageing, epithelial-mesenchymal transition, acetylation, amino acid metabolism, anoikis, DNA repair, fatty acid metabolism, hypoxia, inflammatory response, N6-methyladenosine, mitochondrial homeostasis, and mTOR. (Refer to Supplemental 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.

### 2.7 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 samples were categorized into groups using the RTMscore signature cutoff value obtained by the "survminer" R programme. The R package "survival" was utilized to apply the Kaplan-Meier survival curve in order to determine the overall survival (OS) between the two groups categorized.

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, whilst one-way ANOVA testing were employed to examine statistical differences between groups. The Wilcoxon test was employed to ascertain statistical differences between groups for variables that were not normally
distributed, whereas the Kruskal-Wallis test was utilized to find statistical differences between groups. The statistical analyses were conducted using the R programme (version 4.1.2).

3. Results

3.1 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 1,500 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, 2B and Supplementary Fig. 1A). Notably, a significant proportion of immune cells and epithelial cells were found in both normal and tumor samples (Fig. 2C). 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 (Supplementary Figs. 1B-D). In order 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 (cluster 0), IL1B + TAMs (cluster 1), PLAC8 + TAMs (cluster 2), IFI27 + TAMs (cluster 3), RP11-598F7.3 + TAMs (cluster 4), ISG20 + TAMs (cluster 5), and MIR3945HG + TAMs (cluster 6) (Figs. 2D-F and Supplementary Fig. 2). Furthermore, epithelial cells derived from normal sample sources were categorized into distinct subtypes, including alveolar epithelial cells, basal cells, ciliated cells, and club cells. This classification was determined by evaluating the expression of particular markers characteristic of epithelial cells (Supplementary Figs. 1E-G).

Among the discovered cell types, it was shown that RGS1 + TAMs had high expression levels of SGK1, CCL3, and SPP1. Furthermore, these TAMs were found to be functionally enriched in regulatory pathways associated with glial cell migration and the regulation of inflammatory responses (Supplementary Fig. 2). Additionally, RGS1 + TAMs were characterized by the presence of both IFN-TAMs and Inflam-TAMs phenotypes (Fig. 2G).

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 quantities of the 21 distinct cell types present in both the GSE126044 and TCGA-LUAD cohort specimens. The study aimed to examine the correlation between the frequency of 21 specific cell types and the clinical characteristics of patients, as well as their prognosis in terms of immunological treatment, which revealed a noteworthy association
between the level of infiltration of RGS1 + TAMs and the responsiveness of immune treatment. LUAD patients who had a high level of RGS1 + TAMs infiltration experienced greater therapeutic benefits from immunotherapy (Figure. 3A).

In order to evaluate the potential differential function of RGS1 + TAMs in the presence of specific mutations, a Kaplan-Meier analysis was conducted to examine the infiltration levels of RGS1 + TAMs in patients harbouring distinct gene mutations. The results revealed that increased infiltration of RGS1 + TAMs in patients with EGFR mutations and KRAS mutations was associated with prolonged survival (Figure. 3B).

In order to gain a clearer understanding of the correlation between the characteristics of RGS1 + TAMs infiltration and other clinical variables, the study aimed to quantify the proportions of these variables in two subtypes of LUAD patients. The findings of the study revealed a statistically significant increase in the percentage of M1 patients exhibiting low levels of RGS1 + TAMs infiltration, which suggested a potential association between elevated levels of RGS1 + TAMs infiltration and a diminished propensity for metastasis in cases of LUAD. Nevertheless, there were no statistically significant differences identified for the remaining features (Figure. 3C).

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 (Figure. 3D). The GSA method was employed to compute scores for different immune-related functional signatures, including checkpoints, CCR, T cell co-inhibition, and T cell co-stimulation. Patients with LUAD who exhibited a significant presence of RGS1 + TAMs demonstrated elevated immune functional scores compared to those with a lower infiltration level, (Figure. 3E).

3.2 Causal relationship between CD83 and LUAD

3.2 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 (Figure. 4A). Two genes, CD83 and
A2M, were identified using a univariate regression analysis conducted on the hub genes (Supplementary Table 6).

Among these, CD83 had been previously identified as a distinctive marker for dendritic cells (51). 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 response (52). 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 healing (53). The study aimed to further investigate the association between CD83 and LUAD. We conducted an investigation into 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 (Figure. 4B). The GEPIA dataset (54) was utilized to acquire data (Figure. 4C), which revealed a notable elevation 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 (Figure. 4D-E).

The TME exhibited a significant association with the processes of cancer initiation, advancement, and the implementation of therapeutic strategies (55). This study aimed to examine the correlation between CD83 and immunological characteristics in order 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 (Figure. 4F).

In addition, a statistically significant negative association was indicated between the expression of CD83 and the level of tumor purity (Figure. 4G). Moreover, a strong positive correlation was observed between the expression of CD83 and ImmuneScore, StromalScore, and ESTIMATEScore (Figure. 4G). 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 (Figure. 4H). It was worth mentioning that the TCGA LUAD 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 GSE126044 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 cohort GSE148476, patients with metastatic gastric cancer in the cohort Kim2018, and patients with urothelial carcinoma in the cohort IMvigor210 (Figure 5A).

Meanwhile, in the cohorts GSE91061, GSE35640, and GSE115821, 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 (Figure 5A).

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 cohort Kim2018, GSE126044, GSE91061, GSE148476, GSE35640, and GSE115821, which yielded ROC values of 0.707, 0.875, 0.670, 0.701, 0.695, and 0.863, respectively (Figure 5B). In addition, a notable correlation was observed between the mRNA expression level of CD83 and conventional immunological checkpoint markers (Figure 5C). 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 (Figure 5D).

### 3.3 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 (Figure 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 (Figure 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 study conducted a Kaplan-Meier analysis to compare the high-RTMscore group with the low-RTMscore group (Figure 6D). The results showed a substantial correlation between the RTMscore signature and overall survival (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 (Figure 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 linked with LUAD were included in the analysis (Supplementary Table 7). The signatures examined in this study encompassed a range of biological processes, including cuproptosis, ferroptosis, autophagy, epithelial-mesenchymal transition, acetylation, amino acid metabolism, anoikis, DNA repair, fatty acid metabolism, hypoxia, inflammatory response, N6-methyladenosine modification, mitochondrial homeostasis, and mTOR signaling. 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).

4. 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 in order to facilitate informed clinical decision-making. Nevertheless, a dearth of dependable prognostic indicators exists for the purpose of identifying individuals with "high-risk" LUAD who could potentially derive advantages from immunotherapy. In order to address this knowledge gap, we conducted an investigation into 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 infiltration 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.

LUAD patients with high mRNA expression level of CD83 exhibited notable infiltration of immune cells, encompassing immunoreactive cell types such as T cells, B cells, and natural killer cells, alongside suppressive immune cell populations including regulatory T cells, TAMs, and myeloid-derived suppressor cells. It was worth noting that these immune cell subsets are significantly associated with immunotherapy. Data analysis based on clinical data also showed that patients exhibiting elevated levels of CD83 in the context of urothelial carcinoma, non-small cell lung cancer, chronic lymphocytic leukemia, metastatic gastric cancer, or 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 in order 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.

A nomogram was constructed, incorporating the variables of age, gender, stage, and RTMscore. Furthermore, the nomogram was transformed into a web-based calculator interface (https://sjz178386037.shinyapps.io/DynNomapp/) in order to streamline the evaluation of overall survival in LUAD patients. This conversion not only addressed the issue of manual computation but also enhanced the clinical utility of the nomogram.

5. 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 both overall survival and
responsiveness to immunotherapy.

**Declarations**

**Acknowledgments**

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

**Data availability**

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

**Conflict of Interest**

The authors declare no competing interests.

**Author Contributions**

JS and HG designed and conducted the entire study. JS performed the data collection, bioinformatics, and statistical data analysis. HG performed the validation experiments. YN and YS investigated the literature. YZ was responsible for the integrity of the entire study and manuscript review. All authors contributed to the article and approved the submitted version.

**Publisher’s note**

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be assessed in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

**Funding**

Not applicable.

**Ethics approval and consent to participate**

Not applicable.

**Patient consent for publication**

Not applicable.

**References**


Figures
Figure 1

The study's flowchart diagram
Figure 2

Identification of RGS1+ TAMs

(A) t-SNE plot displaying the cell clusters in the microenvironment of LUAD. (B) t-SNE plot displaying the cell composition in the microenvironment of LUAD, colored according to cell types. (C) Heatmap displaying the expression of tumor-associated macrophage type marker genes. Cell-specific marker genes were selected according to the previous study by Ma.35690521 (D) t-SNE plot displaying the TAMs clusters in the microenvironment of LUAD. (E) t-SNE plot displaying the TAMs composition in the microenvironment of LUAD, colored according to cell types. (F) Heatmap displaying the expression of macrophages M1 and M2 marker genes. Cell-specific marker genes were selected according to a previous study by Ma.35690521 (G) Barplot displaying the overall cell composition of normal and tumor samples, colored by cell types.
Figure 3

The relationship between the prevalence of these RGS1+ TAMs and the clinical characteristics and immunotherapeutic characteristics

(A) Heatmap displaying correlation between the infiltration of cell composition and immunotherapy benefit. (B) Kaplan-Meier curves of overall survival (OS) according to RGS1+TAMs infiltration in TCGA.
cohort, EGFR mutation cohort, KRAS mutation cohort and EML4–ALK mutation cohort. (C) Comparison of proportions of subsets divided by age, gender, stage, T, N, and M in different subsets. (D) GO enrichment and KEGG analysis of the DEGs between RGS1+ TAMs subtypes. (E) Heatmap displaying the correlation between the infiltration of RGS1+ TAMs and 13 immune-associated processes.

Figure 4
Identification of CD83 and analysis of the relationship between CD83 and immunological characteristics

(A) Venn diagram representing the meeting point of RGS1+ marker genes and differently expressed genes between LUAD tissues and adjacent normal tissues in TCGA cohort. (B) 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. (C) Different expression of CD83 between lung adenocarcinoma tissues and adjacent normal tissues in TCGA cohort. (D) and (E) Validation of the expression of CD83 in the normal and tumor sampes level by the Human Protein Atlas database (immunohistochemistry). (F) Heatmap displaying the correlation between the mRNA expression level of CD83 and immune infiltrating cells. (G) Box plot displaying the correlation between the mRNA expression level of CD83 and The ESTIMATE Immune Score, ImmuneScore, StromalScore, and TumorPurity. (H) Box plot displaying the correlation between the mRNA expression level of CD83 and immune modulators.
Figure 5

Predictive value of immunotherapy benefits

(A) Box plot displaying the correlation between the mRNA expression level of CD83 and immunotherapy response in the immunotherapy cohorts (IMvigor, GSE135222, GSE78220, GSE91061, GSE148476, and GSE103668). (B) ROC curves of the mRNA expression level of CD83 to predict the benefits of
immunotherapy in the immunotherapy cohorts (IMvigor, GSE135222, GSE78220, GSE91061, GSE148476, and GSE103668). (C) Heatmap displaying the correlation between the mRNA expression level of CD83 and immunological checkpoint genes. (D) ROC curves of the CD83 and PD-L1 to predict the benefits of immunotherapy.
The construction and assessment of RTMscore signature

(A) A total of 104 combinations of machine learning algorithms for the RTMscore signatures via a 10-fold cross-validation framework based on the TCGA-LUAD group. The C-index of each model was calculated across validation datasets, including the GSE30219, GSE3141, GSE72094 and GSE50081 cohorts. (B) The importance of the 4 most valuable mRNAs based on the RSF algorithm. (C) The coefficients of 3 mRNAs 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 (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

Assessment of the RTMscore signature

(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.
Supplementary Files

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

- Supplementary.pdf