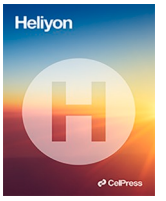
Heliyon 10 (2024) e37998

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

CD8+ T cell associated scoring model helps prognostic diagnosis and immunotherapy selection in patients with colon adenocarcinoma

Zheng Zhao [a], Mingkai Chen [b], Kuanxue Sun [a,**], Xinqi Gu [c,*]

[a] Deparment of General Surgery, Gongli Hospital, Pudong New Area, Shanghai, China
[b] Deparment of Gastroenterlogy, Zhengzhou Yihe Hospital, Zhengzhou, China
[c] Department of Gastroenterlogy, Shanghai Pudong Hospital, Pudong Medical Center of Fudan University, Shanghai, China

ARTICLE INFO

Keywords:
Colon adenocarcinoma
Immunotherapy
CD8$^+$ T cell
Prognostic signature
Multi-omics

ABSTRACT

Objective: T cell-mediated immunity plays a crucial role in the immune response against tumors, with CD 8+ T cells playing a leading role in the eradication of cancer cells.
Material and methods: A total of 5 datasets were included in this study. Single cell transcriptome data were used to discover CD8$^+$ T cell marker genes, and Bulk transcriptome data from TCGA and GEO were jointly analyzed to screen candidate prognostic genes. lasso regression was performed to construct prognostic models. Immunotherapy cohort (IMvigor 210 and GSE78220) was applied to validate the diagnostic power of markers.
Result: Single-cell transcriptome data identified 65 CD8$^+$ T cell marker genes, highlighting their importance in T cell-mediated immune responses. Among these, 11 genes were identified as CD8$^+$ T-associated differential genes through analysis of bulk data from TCGA and GEO. A prognostic model for 5 genes was identified based on Lasso regression, dividing colon adenocarcinoma (COAD) patients into high- and low-risk groups. This model exhibited higher prognostic accuracy compared to traditional clinicopathological characteristics (age, pathological stage, histological grading). Moreover, the risk score derived from this model successfully differentiated patient responses to immunotherapy, as validated in the IMvigor 210 and GSE78220 cohorts.
Conclusion: Our research introduces a novel prognostic signature based on CD8$^+$ T cell marker genes, demonstrating significant predictive power for prognosis and immunotherapy response in COAD patients. This model offers a potential tool for improving patient stratification and personalizing treatment strategies.

1. Introduction

The tissues of the colon and rectum can develop colon adenocarcinoma (COAD), a highly aggressive tumor [1]. More than 1.9 million new cases and 900,000 deaths from colorectal cancer are expected in 2020, making it the third most common cause of cancer-related deaths globally, behind lung and breast cancer [2]. In China, despite significant advances in COAD treatment strategies, the 5-year overall survival (OS) rate for COAD remains below 60 % [3]. The clinical use of immune checkpoint inhibitors in recent

* Corresponding author.
** Corresponding author.
E-mail addresses: sunkuanxue1234@163.com (K. Sun), guxinqi2023@aliyun.com (X. Gu).

https://doi.org/10.1016/j.heliyon.2024.e37998
Received 3 June 2024; Received in revised form 14 September 2024; Accepted 16 September 2024
Available online 19 September 2024

years has greatly improved the clinical benefit and changed the treatment paradigm of COAD [4,5]. However, not all patients can benefit from them [6]. Consequently, in order to forecast prognosis and treatment outcome, it is essential to create prediction models and find new biomarkers.

The tumor microenvironment (TME) is the surrounding stroma of a tumor, which consists of various cells, molecules, and blood vessels that interact with the tumor cells1. The TME plays a crucial role in tumor occurrence and progression, as it can modulate cancer cell growth, invasion, angiogenesis, metastasis, and immune evasion [7,8]. Cytotoxic T lymphocytes, or $CD8^+$ T cells, are able to recognize and eliminate tumor cells that display particular antigens. Major Histocompatibility Complex (MHC) class I molecules on antigen-presenting cells, like dendritic cells, display tumor-derived peptides that activate $CD8^+$ T cells. Then, $CD8^+$ T cells can release perforin and granzymes to kill the target tumor cells, or they can leak interferon-gamma and tumor necrosis factor-alpha to boost the antitumor immune response [9]. Furthermore, adoptive cell transfer and immune checkpoint inhibitors, in particular, are mostly administered by $CD8^+$ T cells as part of cancer immunotherapy. Antibodies known as immune checkpoint inhibitors bind to $CD8^+$ T cells' inhibitory receptors, such as cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) and programmed cell death protein 1 (PD-1), which tumor cells can use to elude immune identification [10]. By blocking these receptors, immune checkpoint inhibitors can restore the function and proliferation of $CD8^+$ T cells and enhance their antitumor activity. An example of the effect of $CD8^+$ T cells on tumors is the case of melanoma, a type of skin cancer that is highly immunogenic and responsive to immunotherapy. Melanoma patients who have high levels of tumor-infiltrating $CD8^+$ T cells have better survival and response to immunotherapy than those who have low levels [11,12]. Furthermore, $CD8^+$ T cells may interact with tumor-infiltrating neutrophils, influencing their activity and tumor growth [13]. Considering the function of $CD8^+$ T cells in immunity, earlier research has looked into the molecular traits of $CD8^+$ T cells in breast cancer [14], while a comprehensive knowledge of $CD8^+$ T cells in COAD is lacking.

According to earlier research, analyzing gene expression profiles based on immune cells' molecular characterisation obtained from scRNA-seq data may be a useful method for predicting cancer patients' prognosis and response to immunotherapy [15,16]. To identify $CD8^+$ T cell marker genes and dissect the molecular characteristics of tumor-infiltrating $CD8^+$ T cells, we first performed a comprehensive analysis of COAD scRNA-seq data in this study. After that, a $CD8^+$ T-cell marker gene signature (CTMGS) was developed for prognostic prediction of COAD patients utilizing bulk RNA-seq data. Furthermore, many immunotherapy cohorts supported the prognostic efficacy of CTMGS.

This study utilizes bulk sequencing data integrating $CD8^+$ T cell marker genes to develop prognostic and immunotherapeutic biomarkers for colorectal adenocarcinoma (COAD) patients. The predictive value of these biomarkers has been validated through TCGA and GEO databases. This study fills gaps in the fields of immunotherapy and precision medicine regarding prognosis and treatment of COAD patients, providing new directions and opportunities for future clinical practice and research.

2. Material and methods

2.1. Data download and organization

A total of 5 cohorts were included in this study. GSE146771 Smart-seq2 data (n = 12) were obtained from TISCH2 [17] (http://tisch.comp-genomics.org/). TCGA-COAD cohort data were obtained from TCGA (n = 512), and GSE39582 gene expression matrix was obtained from GEO (https://www.ncbi.nlm.nih.gov/geo/) (n = 585). Immunotherapy cohort (IMvigor 210 and GSE78220) for melanoma and Uroepithelial carcinoma PD-1 or PD-L1 immunosuppressant treatment cohort.

The relevant data was gathered from TCGA and GEO databases, quality control were performed to remove any data with missing or low-quality data. We normalized the TCGA data using TPM values. Log transformation be applied to stabilize variance and make the data more normally distributed. Finally, the normalized datasets were integrated into a single dataset for further analysis.

2.2. Identification and functional analysis of candidate genes

The marker genes of CD 8+ T cells were obtained from the "DE gene" module in TISCH with the screening criteria of $|\log FC| \geq 1$ and P value < 0.05. The TCGA-COAD were analyzed for differences by "limma" [18] package, and then the marker genes of CD 8+ T cells were intersected with the difference genes to obtain candidate prognostic genes. Furthermore, GO and KEGG function enrichment analysis was performed using the "Clusterprofiler" [19] package. $p < 0.05$ was considered significant.

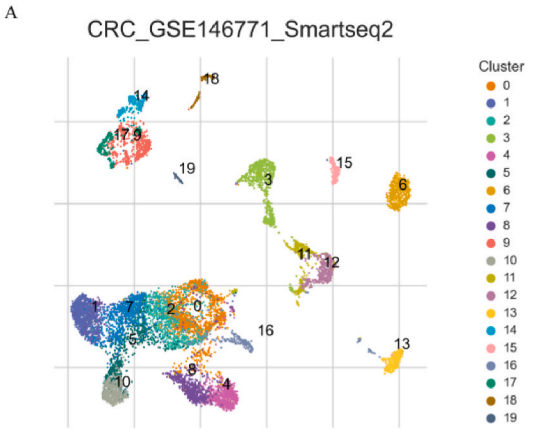
2.3. Molecular risk model construction

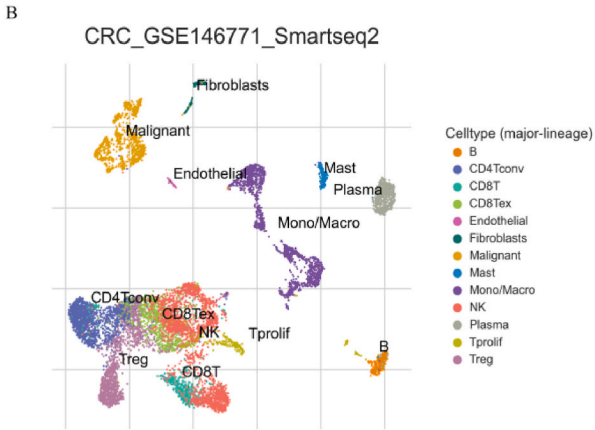
Lasso models are a type of linear regression models that perform both variable selection and regularization by adding a L1 penalty to the loss function. This penalty shrinks the coefficients of some features to zero, effectively removing them from the model. We carried out the analysis using the "glmnet" [20] package to perform lasso Cox regressions and used 10-fold cross-validation for model construction. The parameters used in the "glmnet" package were set as follows: alpha = 1 (for Lasso regression), family = "cox", and standardize = TRUE.

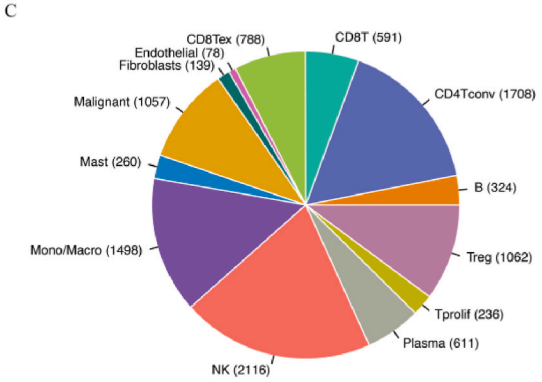
2.4. Immune cell infiltration-related analysis

The CIBERSORT algorithm is a helpful tool for obtaining the gene expression profiles and infiltration characteristics of 22 immune cell types. These can be used to profile the percentage of immune cell infiltration in high- and low-risk populations [19]. The 'estimate'

expression profile was analyzed using the Estimate algorithm (implemented in the R package), which quantifies the degree of stromal and immune cell infiltration [21] Using the TCGA-COAD cohort's data matrix, stromal scores, immune scores, and estimated scores were computed. Wilcoxon t-tests were then used to examine the variations in these scores among risk categories. In earlier publications, the immunological subgroups of TCGA-COAD patients were detailed [22].

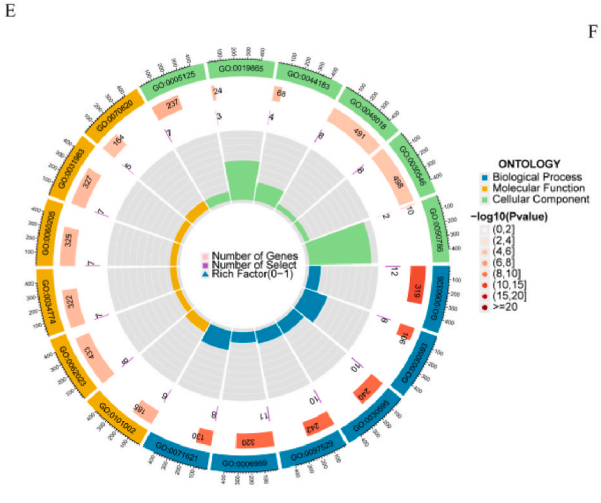






D

Cluster	Celltype	Gene	log2FC	Adjusted p-value
8	CD8T	GZMH	1.55	7.19E-198
8	CD8T	KLRG1	1.45	7.58E-216
8	CD8T	NKG7	1.29	2.54E-211
8	CD8T	FCRL6	1.24	2.32E-212
8	CD8T	PATL2	1.18	4.5E-95
8	CD8T	TPSAB1	-2.25	8.47E-21
8	CD8T	S100A9	-2.27	1.16E-10
8	CD8T	LYZ	-2.33	5.66E-25
8	CD8T	IGLL5	-2.47	6.96E-15
8	CD8T	JCHAIN	-4.73	4.76E-43



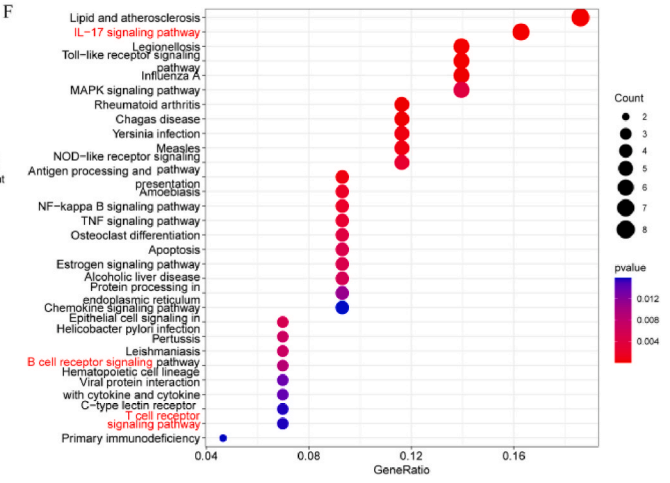


Fig. 1. Single-cell RNA sequencing analysis identifies CD8$^+$ T-cell marker genes. (A) UMAP plots from 12 COAD samples. (B) Annotation of various cell clusters in the UMAP map. (C) Proportion of cell types in COAD samples. (D) Marker genes of Top10 CD8$^+$ T cells. (E) GO and (F) KEGG analysis of marker genes of CD8$^+$ T cells.

2.5. Immunohistochemistry (IHC) assay

Tissue specimens from human or mouse were fixed with 4 % paraformaldehyde, embedded in paraffin, sectioned with 6-μm thickness, and immune stained with specific antibodies, including CXCL8 (10,760–1-AP, Proteintech), TIMP1 (16,396–1-AP, Proteintech), SPP1 (67,763–1-Ig, Proteintech), GLA (ab103590, Proteintech) and CPA3 (R&D systems, MAB3249). Images were acquired with an Olympus FSX100 microscope (Olympus, Tokyo, Japan).

2.6. Statistical analysis

R program (version: 4.2.1) was used to create all analytical graphs. The Wilcoxon T-test or chi-square test were used to compare categorical variables between various risk categories. The model diagnosis's accuracy is assessed using the area under the ROC curve. To compare the variations in survival between high- and low-risk groups, use Kaplan-Meier analysis. Statistics were deemed significant if $P < 0.05$.

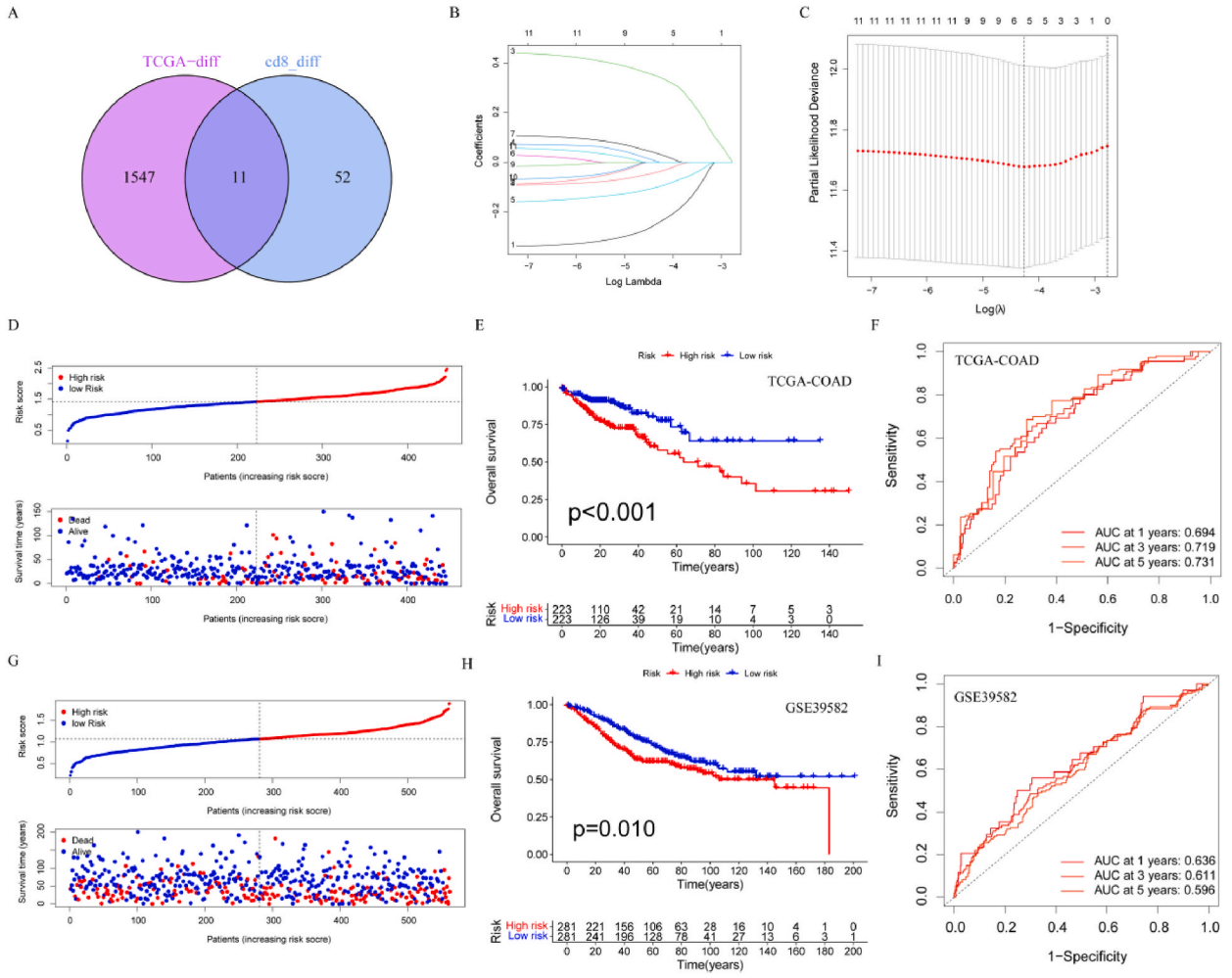


Fig. 2. Construction of CTMGS. (A) Intersection of TCGA differential genes with $CD8^+$ T cell marker genes. (B–C) lasso regression model λ selection and model parameters (D) Distribution of risk scores and survival status in the TCGA-COAD cohort. (E) Kaplan-Meier survival analysis curves for the TCGA-COAD cohort comparing the overall survival of LUAD patients in the high-risk and low-risk groups. (F) ROC curves of the TCGA-COAD cohort CTMGS predicting 1-, 3-, and 5-year risk of death. (G) Distribution of risk scores and survival status in the GEO cohort. (H) GEO cohort Kaplan-Meier survival analysis curves comparing the overall survival of patients with LUAD in the high-risk and low-risk groups. (I) ROC curves of the GEO cohort CTMGS predicting the risk of death at 1, 3 and 5 years.

3. Results

3.1. Identification of CD8+ T cell marker gene expression profiles

We acquired the gene expression profiles of 10468 cells from 12 primary COAD samples based on the scRNA-seq data of GSE146771, in order to do additional study. Using UMAP downscaling, twenty cell clusters were found (Fig. 1A). Following this, each cluster's cell identity was annotated, and cluster 8's cells were identified as $CD8^+$ T cells (Fig. 1B). A total of 591 $CD8^+$ T cells were identified, with the highest percentage of NK cells and CD+4Tconv cells in 12 samples (Fig. 1C). Genes with 65 significant differences from other clusters were identified in cluster 8, and Fig. 1D shows the Top10 genes, which were identified as COAD-associated $CD8^+$ T cell marker genes. We used GO and KEGG analysis to describe the roles of these genes. The majority of $CD8^+$ T cell marker genes, according to functional enrichment, are linked to immunological characteristics such the IL-17 signaling route, the B cell receptor signaling pathway, the T cell receptor signaling pathway, and the C-type lectin receptor signaling system (Fig. 1E–F). At the same time, we also conducted GSEA enrichment analysis and found that $CD8^+$ T cell marker genes, according to functional enrichment, are linked to cytokine signaling in immune system and innate immune system (Fig. S2D).

3.2. Establishment and validation of a 5-gene prognostic signature based on CD8+ T cell marker genes

To construct prognostic signatures of 65 $CD8^+$ T-cell marker genes, we first performed differential analysis by TCGA-COAD cohort bulk-seq data to obtain 1558 differentially expressed genes, among which 11 genes had intersection with $CD8^+$ T-cell marker genes and were considered as candidate prognostic genes (Fig. 2A). We validated the expression of these differentially expressed genes using the 10× single-cell Colon cancer dataset GSE146771, and the results were consistent (Figs. S2A–C). PPI analysis using the highly expressed DEGs indicates an interaction between TIMP1 and CXCL8 (Fig. S2E). Next, 10-fold cross-validated LASSO Cox regression analysis (Fig. 2B–C) was performed on the 11 candidate genes, and 5 genes (CXCL8, TIMP1, CPA3, SPP1, GLA) were screened for inclusion in the model. Each patient's risk score was determined using the model genes, and patients were categorized into low risk (n = 223) and high risk (n = 223) groups based on the median risk score. The distribution of risk ratings and survival status is depicted in Fig. 2D, and it is clear that the high-risk group had a higher death rate. Compared to patients with low risk scores, patients with high risk levels showed inferior survival results ($P < 0.001$; Fig. 2E).The area under the ROC curve at 1, 3, and 5 years was 0.694, 0.719, and

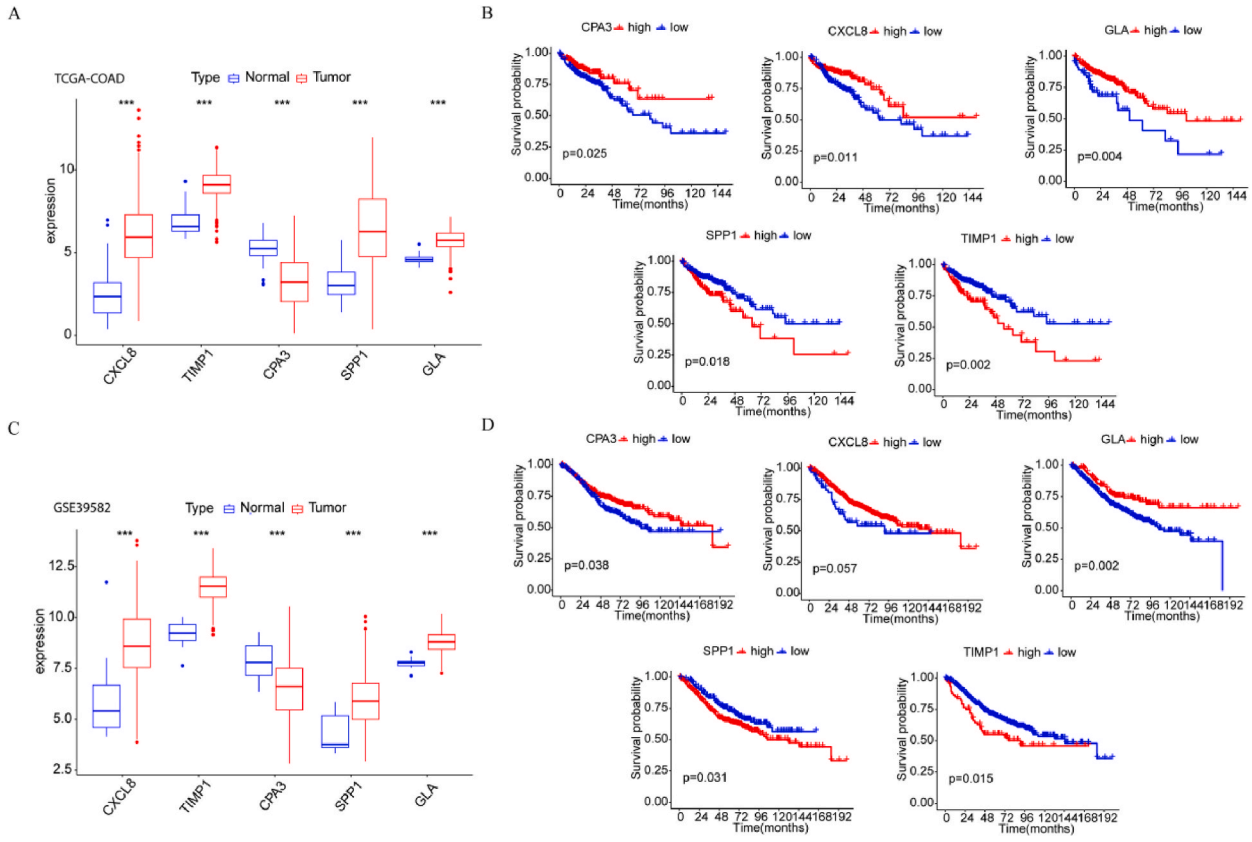


Fig. 3. Expression and prognosis of 5 model genes. Value of CD 8+ T cell-related marker genes in (A) expression and (B) prognosis in TCGA-COAD cohort. Value of CD 8+ T cell-related marker genes in (C) expression and (D) prognosis in GSE39582 cohort.

0.731, respectively, confirming the ability of Riskscore's prognostic diagnosis (Fig. 2F). GSE39582, used as a validation cohort, demonstrated a higher number of deaths in the high-risk group. Moreover, the area under the 5-year Receiver Operating Characteristic (ROC) curve reached 0.636, indicating the generalizability of the prognostic signature (see Fig. 2G–I).

3.3. Expression and prognosis of 5 model genes

Moreover, to verify the accuracy of the 5 gene features. We verified their expression and prognostic value by TCGA and GEO databases, respectively. The data showed that CXCL8, TIMP1, SPP1, GLA were up-regulated in COAD patients, while CPA3 was significantly down-regulated in COAD patients (all $P < 0.05$; Fig. 3A, C). And then, the ICH staining results also proved the same result (Fig. S1). Kaplan-meier survival curves indicated that low expression of CXCL8, TIMP1, CPA3 in the TCGA-COAD cohort was associated with worse prognosis, while high expression of SPP1, GLA represented a worse prognosis in COAD patients ($P < 0.05$; Fig. 3B). These findings were also confirmed in the GEO cohort (Fig. 3D). Thus, CD8+T-related marker genes may contribute to the development of COAD.

3.4. CTMGS correlates with clinicopathological characteristics

To determine the association of Riskscore with clinicopathological features, we first compared them with the prognostic value of patients with COAD. Age, stage, TMN staging, and Riskscore were all related to patient prognosis, according to univariate cox analysis ($P < 0.05$; Fig. 4A). Age, T-stage, and Riskscore were found to be independently correlated with the prognosis of COAD patients using multifactorial cox analysis (Fig. 4B). We performed ROC analysis and the AUC value of 5-year Riskscore was 0.731, which was comparable to the level of prognosis determined by Stage (Fig. 4D), suggesting that Riskscore could be a new valid prognostic marker. Further we calculated the proportion of clinical features in the high- and low-risk population and found a trend toward increased risk in men ($P = 0.071$), which was significantly associated with T-stage ($P < 0.05$; Fig. 4C).

3.5. Comparison of immune infiltration between high and low risk groups

ESTIMATE, CIBERSORT and ssGSEA were used to explore differences in immune function. In the estimation analysis, the high-risk population had higher stromal, immune and estimation scores compared to the low-risk population ($P < 0.05$; Fig. 5A). This may

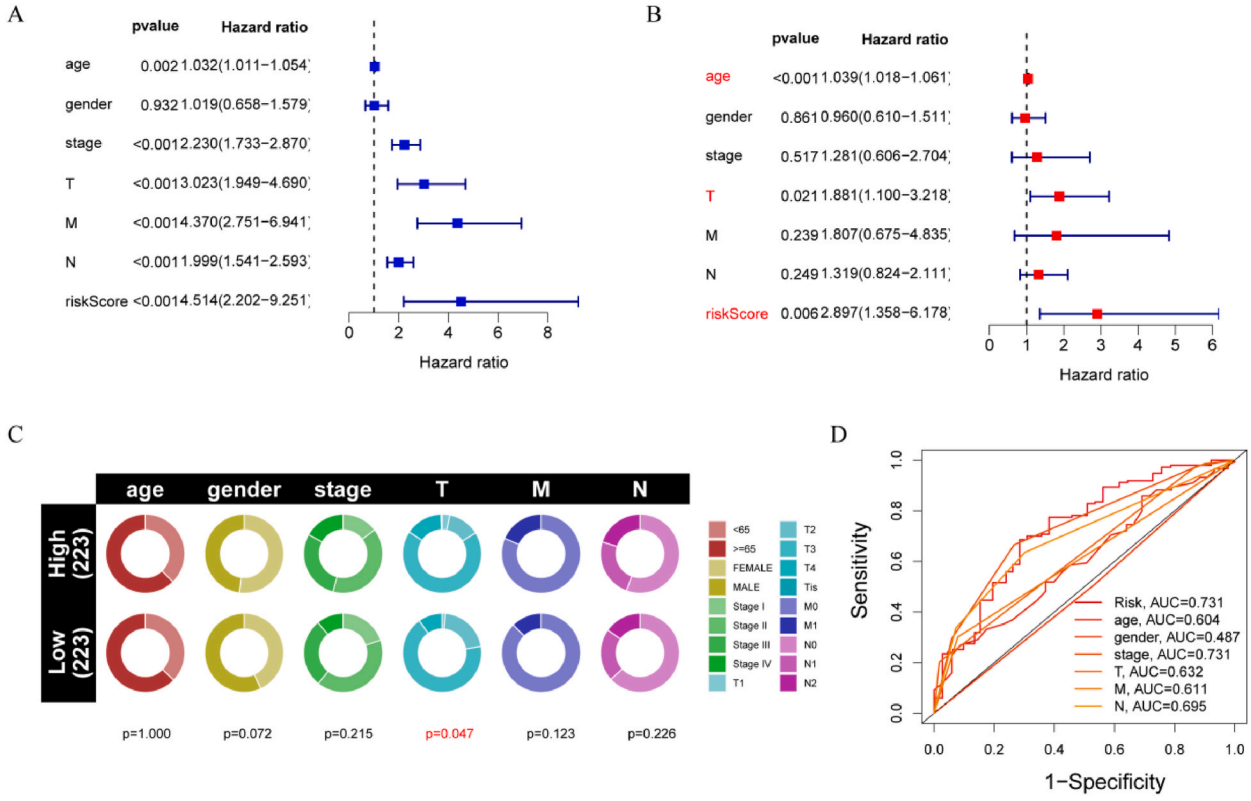


Fig. 4. CTMGS correlates with clinicopathological features. (A) Univariate and (B) multifactorial cox analysis of age, stage, TMN stage and Riskscore with patient prognosis. (C) Correlation of clinical characteristics with Riskscore. (D) Clinical Characteristics and Riskscore's ROC Curve.

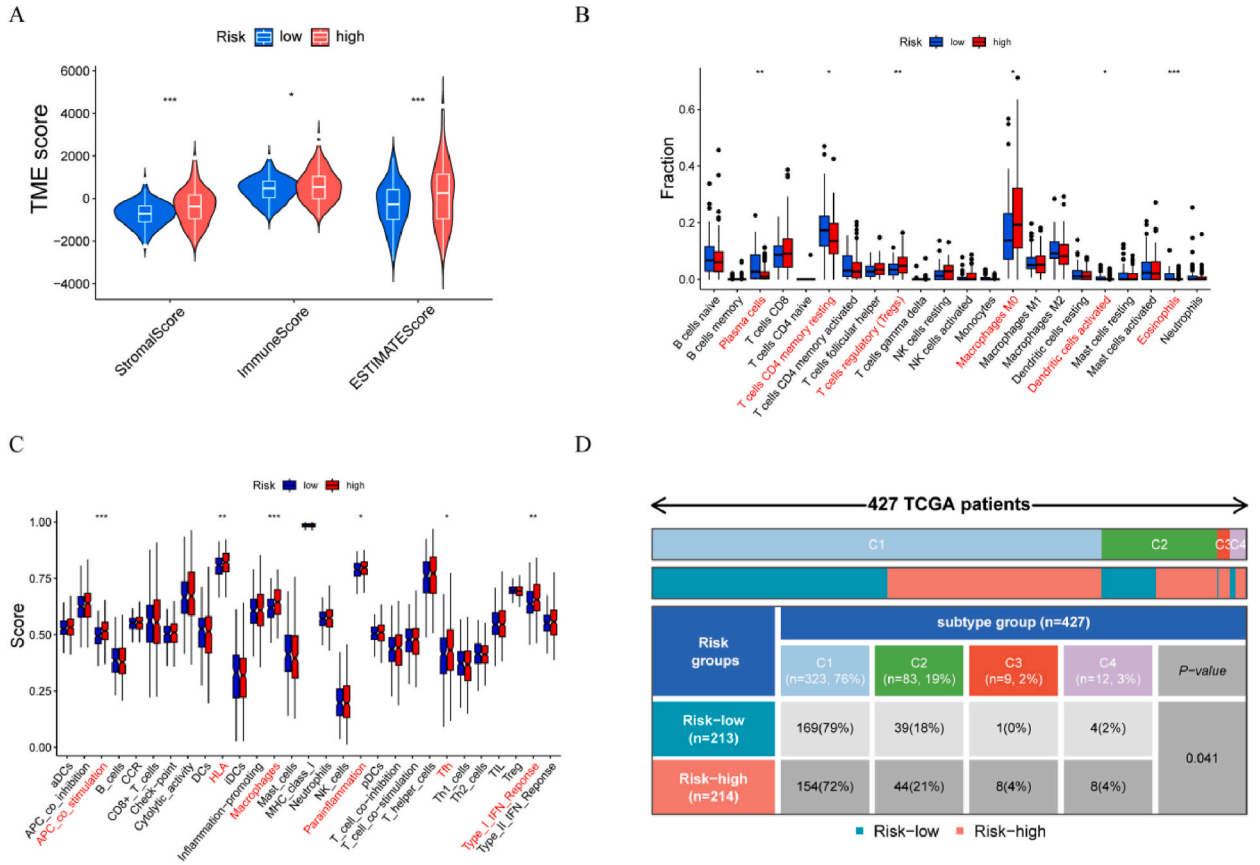


Fig. 5. Comparison of immune infiltration between high and low risk groups. (A) ESTIMATE calculates immune scoring, substrate scoring and ESTIMATE scoring for high- and low-risk patients. (B) CIBERSORT calculates the ratio of 22 immune cell types in high and low risk patients. (C) ssGSEA calculates the level of activation of immune pathways in high- and low-risk patients. (D) The relationship between C1-C6 immune status and riskscore in TCGA-COAD patients.

represent a higher immune infiltration status. The percentage of 22 immune cell species in the tumor microenvironment of COAD patients was estimated by CIBERSORT. Fig. 5B shows that patients at high risk had larger proportions of Treg cells and macrophages M0 and lower proportions of plasma cells and dendritic cells activated; Fig. 5C shows that the highrisk group had higher levels of APC co-stimulation and Type I IFN response. Tumor patients were classified into C1–C6 groups by Vesteinn Thorsson et al. based on their immunological status: Wound Healing, IFN-γ Dominant, Inflammatory, Lymphocyte Depleted, Immunologically Quiet, and TGF-β Dominant [22]. TCGA-COAD patients had only C1-C4 characteristics. Higher risk patients had a higher proportion of C2,C3,C4 population ($P = 0.041$; Fig. 5D). These results suggest that there are significant differences in immune infiltrating cells and immune function between high- and low-risk patients, which may affect the outcome of immunotherapy.

3.6. *Predictive assessment of immunotherapy response*

Immune checkpoint blockade therapy is a method of restoring immune cell attack on tumors by blocking the interaction of immune checkpoints and their ligands with antibody drugs [23]. Therefore, we first compared the correlation of riskscore with immune checkpoints including (PDCD1, CD40LG, TNFSF9). High riskscore was associated with higher PDCD1 and TNFSF9 and lower CD40LG ($P < 0.05$; Fig. 6A–C). We searched for relationships between immunotherapy responses (GSE78220 and IMvigor210) and riskscore. PDL1 is used to treat the bladder cancer cohort known as IMvigor210. A PD-1-treated melanoma cohort is GSE78220. Four groups of responses to immunotherapy were identified: disease progression (PD), stable disease (SD), partial response (PR), and complete response (CR). Patients with high risk scores in IMvigor210 had lower CR and PR ($P < 0.05$; Fig. 6E) and a worse prognosis ($P < 0.05$; Fig. 6D) than patients with low risk scores. The accuracy of the immunotherapy effect was 0.602 (Fig. 6F). Patients with high risk ratings in GSE78220 had a higher significant higher prognosis ($P < 0.05$; Fig. 6G) and higher CR and PR ($P < 0.05$; Fig. 6H) than patients with low risk scores. As shown in Fig. 6I, the area under the ROC curve was 0.692. These results suggest that CTMGS can predict response to immunotherapy, but appears to have different predictive trends for different tumors or types of immunotherapy.

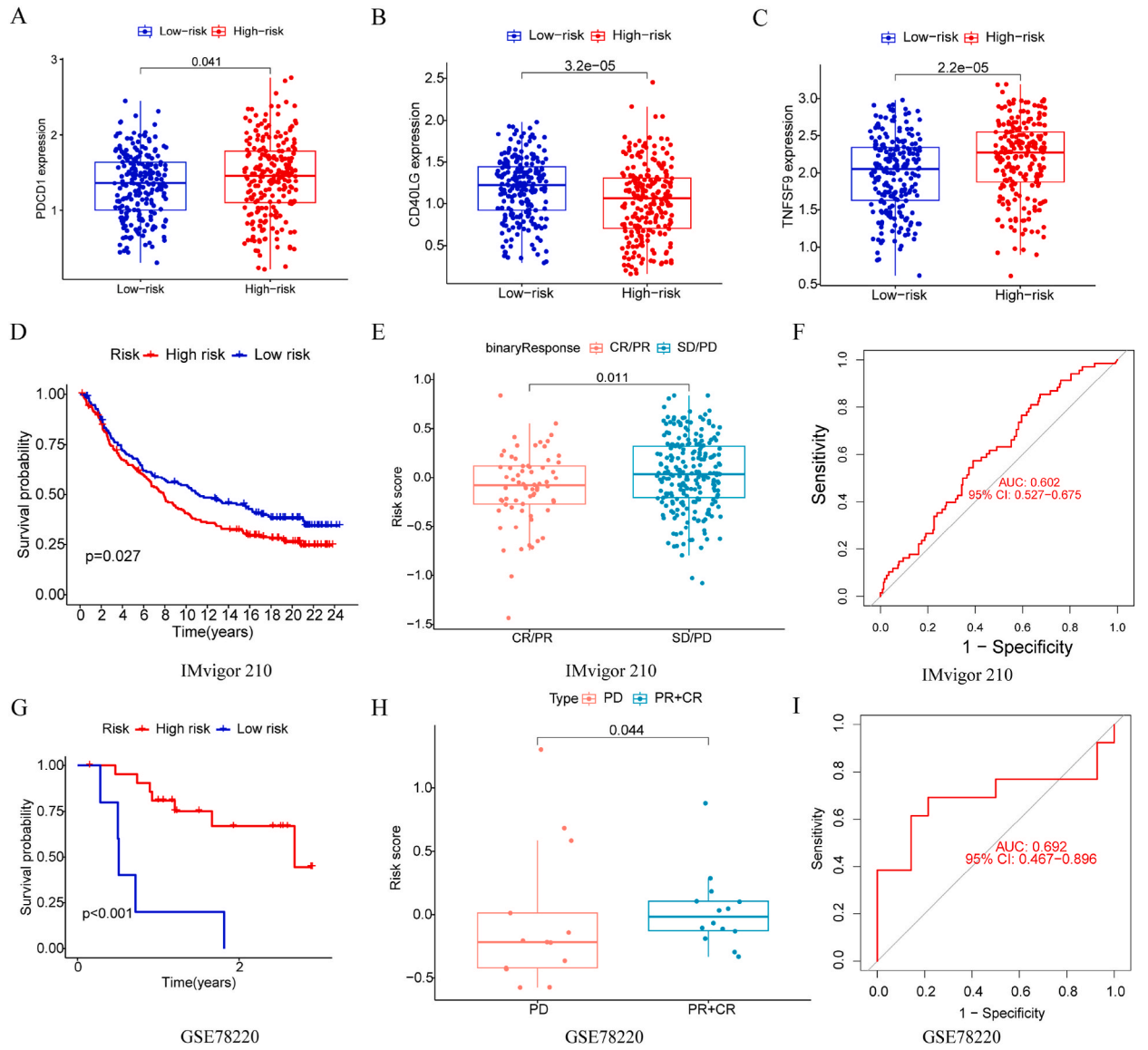


Fig. 6. Predictive Assessment of Immunotherapy Response. Expression of immune checkpoints between high and low risk patients including (A) PDCD1, (B) CD40LG, (C) TNFSF9. (D) Survival curves of high- and low-risk patients in the IMvigor 210 cohort. (E) Correlation between risk score and response status in the IMvigor 210 cohort. (F) ROC curve of risk score predicting treatment benefit rate in IMvigor 210 cohort. (G) Survival curves of high- and low-risk patients in the GSE78220. (H) Correlation between risk score and response status in the GSE78220. (I) ROC curve of risk score predicting treatment benefit rate in GSE78220.

4. Discussion

The most advanced kind of treatment for COAD, one of the deadliest cancers in the world, is anti-PD-1 immunotherapy. Finding patients who would benefit most from anti-PD-1 immunotherapy is still difficult at this time [24,25]. A range of cancers' prognosis and course can be predicted using gene signatures, which are biofunctional patterns created from the expression data of several genes [26], but their application in COAD remains to be further explored.

The degree of infiltration and activity level of $CD8^+$ T cells in tumors can reflect the immunogenicity of the tumor and the responsiveness to immunotherapy [9]. $CD8^+$ T cells, on the other hand, are subjected to a variety of inhibitory factors in the tumor microenvironment, leading to their differentiation and malfunction. Tumor cells or other immune cells, for example, can block $CD8^+$ T cell activation and proliferation by expressing immune checkpoint molecules (e.g., PD-L1). Furthermore, several metabolic stressors (e.g., hypoxia, low glucose, high lactate, etc.) and suppressive cytokines (e.g., TGF-, IL-10, etc.) in the tumor microenvironment can alter the survival and effects of $CD8^+$ T cells. As a result, the differentiation and functional status of $CD8^+$ T lymphocytes in malignancies is a dynamic process that is dependent on the balance of numerous stimulatory and inhibitory signals [25,27]. Many

histology-based strategies have been used recently to find extrinsic and intrinsic tumor components that can act as immunotherapy predictive biomarkers. First, a number of genetic variables, including TMB and neoantigens [28], and mismatch repair defects [29] have been used to predict ICB response. Response prediction has also made use of transcriptional level data, including PD-1/PD-L1 expression levels [30], immune cell infiltration analysis based on transcriptional characteristics [31] and transcriptional immune score [32], have also been used for response prediction. These studies suggest that the status of $CD8^+$ T cells and their associated marker genes may be an important immunotherapeutic target and prognostic indicator in COAD.

"This article is the first article to construct prognostic and immunotherapeutic labels for COAD patients from $CD8^+$ T cell marker genes combined with Bulk sequencing data. We first validated the predictive value in TCGA and GEO databases. Further confirmation came from IHC experiments using clinical tissues, affirming the accuracy and robustness of our constructed CTMGS as an independent prognostic indicator for COAD patients. Our study revealed some paradoxical findings: the CTMGS showed contrasting outcomes in objective remission rates between the IMvigor and GSE78220 cohorts. In the IMvigor cohort, patients in the high-risk model had a poorer prognosis and weaker remission rates, whereas in the melanoma PD-1 treatment cohort, patients in the high-risk model had a better prognosis and stronger remission rates. This is because the effectiveness of immunotherapy may depend on tumor type, immune microenvironment, molecular characteristics, treatment regimen, and resistance mechanisms of each patient [33]. Firstly, different types of tumors exhibit varying degrees of immune reactivity. For instance, melanoma and non-small cell lung cancer typically show better responses to immune checkpoint inhibitors, while some “cold” tumors like pancreatic cancer tend to respond less favorably. Secondly, the types and status of immune cells within the tumor microenvironment, such as T cells, macrophages, regulatory T cells, etc., are crucial for the response to immunotherapy. For example, a high infiltration of effector T cells often indicates better efficacy. Furthermore, the genetic and epigenetic features of tumors, such as tumor mutational burden (TMB), microsatellite instability (MSI), and PD-L1 expression, have been shown to be associated with the efficacy of immunotherapy. High TMB and MSI-H are usually correlated with better responses to immunotherapy. Different immunotherapy strategies, including standalone immune checkpoint inhibitors, combination therapies with chemotherapy or radiation, and combinations of various immune modulators, may have significantly different impacts on efficacy. Moreover, Patients may develop resistance to immunotherapy through inherent or acquired mechanisms, impacting treatment efficacy via immune escape or changes in the immune-suppressive microenvironment. This effectiveness for multiple cancers suggests a key role for our 5 model gene’s in immunotherapy and tumors.

We found that SPP1 and GLA were upregulated in COAD and correlated with poorer patient prognosis. SPP1, also known as osteopontin or secreted phosphoprotein, is a multifunctional protein that is involved in various biological processes, such as inflammation, wound healing, angiogenesis, cell adhesion, migration and survival. It can interact with different types of immune cells, such as macrophages, T cells, B cells, dendritic cells and natural killer cells, through its binding to various receptors, such as CD44, integrins and CD471. It also regulates the immune response by regulating the activation, differentiation, polarization and function of immune cells [34]. In COAD (colorectal adenocarcinoma), SPP1 has been reported to play a role in tumor progression, invasion, metastasis and resistance to therapy [35]. By affecting immune cell phenotype and invasion, SPP1 can also impact the COAD immunological microenvironment. For instance, a study discovered that SPP1 was substantially expressed in COAD and that it positively linked with dendritic cells, neutrophils, and macrophage infiltration. Additionally, SPP1 was linked with immune checkpoint expression, including CTLA-4, LAG-3, PD-L1, and PD-1 [36]. SPP1 was linked to the infiltration of B cells, $CD8^+$ T cells, $CD4^+$ T cells, and natural killer cells in COAD, according to another study. Moreover, SPP1 was associated with the expression of genes linked to T cell fatigue, cytotoxicity, and activation [37]. GLA, also known as alpha-galactosidase A, is an enzyme that catalyzes the hydrolysis of glycosphingolipids, such as globotriaosylceramide (Gb3) and globotriaosylsphingosine (lyso-Gb3) that GLA deficiency causes Fabry disease, a rare hereditary condition where Gb3 and lyso-Gb3 build up in different tissues and organs [38]. Poorly understood function in malignancies, which could be a possible COAD treatment target.

We identified 65 differentially expressed genes in CD8 T cells from the GSE146771 dataset. 11 genes were considered as candidate prognostic genes by TCGA-COAD. We then used the GSE39582 dataset to validate the prognostic analysis, which indicated that the mortality rate in the high-risk group was higher. Meanwhile, we investigated the relationships between immunotherapy responses (from the GSE78220 and IMvigor210 datasets) and risk scores. The results indicate that the CTMGS can effectively predict responses to immunotherapy; however, it seems to exhibit varying predictive trends across different tumor types and immunotherapy modalities. Although the prognostic model based on $CD8^+$ T cell marker genes in this study demonstrates high accuracy in predicting prognosis and immunotherapy response in colon adenocarcinoma (COAD) patients, its generalizability across diverse patient populations remains to be validated. Additionally, heterogeneity among different datasets may introduce biases, affecting the consistency of marker genes and the prognostic model. Future research should focus on expanding cohort diversity, standardizing data collection, validating the model’s applicability in other cancers and treatment contexts, and confirming its clinical utility through large-scale prospective clinical trials. Integrating additional biomarkers could also enhance the precision of patient stratification and personalized treatment.

Funding statement

This study was funded by the Pudong New Area Clinical Characteristic Discipline Project (Grant No. PWYts2021-11), and Supported by the Project of Key Medical Discipline of Pudong Hospital of Fudan University (Grant No. Zdxk2020-07).

Ethics approval and consent to participate

The research protocol was approved by the Medical Ethics Committee of Zhengzhou Yihe Hospital (NO. 2024010610), (Approval Date: January 6, 2024 to January 5, 2024). Written informed consent was provided from all the patients.

Consent for publication

Not applicable.

Data availability statement

All data sets used in this study are public data sets. The datasets generated analyzed during the current study are available in the GEO and TCGA repository, including GSE146771, GSE39582, GSE78220, TCGA-COAD and IMvigor210. The raw data and code supporting this article can be obtained by contacting the corresponding author with a reasonable request.

CRediT authorship contribution statement

Zheng Zhao: Writing – original draft, Data curation. **Mingkai Chen:** Software. **Kuanxue Sun:** Writing – review & editing. **Xinqi Gu:** Project administration, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e37998.

References

[1] M.F. Muller, A.E. Ibrahim, M.J. Arends, Molecular pathological classification of colorectal cancer, Virchows Arch. 469 (2) (2016) 125–134.
[2] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J Clin 71 (3) (2021) 209–249.
[3] H. Chen, B. Lu, M. Dai, Colorectal cancer screening in China: status, challenges, and prospects - China, 2022, China CDC Wkly 4 (15) (2022) 322–328.
[4] D. Ciardiello, P.P. Vitiello, C. Cardone, G. Martini, T. Troiani, E. Martinelli, F. Ciardiello, Immunotherapy of colorectal cancer: challenges for therapeutic efficacy, Cancer Treat Rev. 76 (2019) 22–32.
[5] U. Testa, E. Pelosi, G. Castelli, Colorectal cancer: genetic abnormalities, tumor progression, tumor heterogeneity, clonal evolution and tumor-initiating cells, Med. Sci. 6 (2) (2018).
[6] D.T. Le, J.N. Uram, H. Wang, B.R. Bartlett, H. Kemberling, A.D. Eyring, A.D. Skora, B.S. Luber, N.S. Azad, D. Laheru, et al., PD-1 blockade in tumors with mismatch-repair deficiency, N. Engl. J. Med. 372 (26) (2015) 2509–2520.
[7] A. Goenka, F. Khan, B. Verma, P. Sinha, C.C. Dmello, M.P. Jogalekar, P. Gangadaran, B.C. Ahn, Tumor microenvironment signaling and therapeutics in cancer progression, Cancer Commun. 43 (5) (2023) 525–561.
[8] D.C. Hinshaw, L.A. Shevde, The tumor microenvironment innately modulates cancer progression, Cancer Res. 79 (18) (2019) 4557–4566.
[9] H. Raskov, A. Orhan, J.P. Christensen, I. Gogenur, Cytotoxic CD8(+) T cells in cancer and cancer immunotherapy, Br. J. Cancer 124 (2) (2021) 359–367.
[10] B. Farhood, M. Najafi, K. Mortezaee, CD8(+) cytotoxic T lymphocytes in cancer immunotherapy: a review, J. Cell. Physiol. 234 (6) (2019) 8509–8521.
[11] Q. Wang, Y. Qin, B. Li, CD8(+) T cell exhaustion and cancer immunotherapy, Cancer Lett. 559 (2023) 216043.
[12] Y. Jin, A. Tan, J. Feng, Z. Xu, P. Wang, P. Ruan, R. Luo, Y. Weng, M. Peng, Prognostic impact of memory CD8(+) T cells on immunotherapy in human cancers: a systematic review and meta-analysis, Front. Oncol. 11 (2021) 698076.
[13] K. Brummel, A.L. Eerkens, M. de Bruyn, H.W. Nijman, Tumour-infiltrating lymphocytes: from prognosis to treatment selection, Br. J. Cancer 128 (3) (2023) 451–458.
[14] B. Virassamy, F. Caramia, P. Savas, S. Sant, J. Wang, S.N. Christo, A. Byrne, K. Clarke, E. Brown, Z.L. Teo, et al., Intratumoral CD8(+) T cells with a tissue-resident memory phenotype mediate local immunity and immune checkpoint responses in breast cancer, Cancer Cell 41 (3) (2023) 585–601 e588.
[15] P. Song, W. Li, L. Guo, J. Ying, S. Gao, J. He, Identification and validation of a novel signature based on NK cell marker genes to predict prognosis and immunotherapy response in lung adenocarcinoma by integrated analysis of single-cell and bulk RNA-sequencing, Front. Immunol. 13 (2022) 850745.
[16] Y. Li, X. Zhao, Q. Liu, Y. Liu, Bioinformatics reveal macrophages marker genes signature in breast cancer to predict prognosis, Ann. Med. 53 (1) (2021) 1019–1031.
[17] D. Sun, J. Wang, Y. Han, X. Dong, J. Ge, R. Zheng, X. Shi, B. Wang, Z. Li, P. Ren, et al., TISCH: a comprehensive web resource enabling interactive single-cell transcriptome visualization of tumor microenvironment, Nucleic Acids Res. 49 (D1) (2021) D1420–D1430.
[18] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, G.K. Smyth, Limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res. 43 (7) (2015) e47.
[19] A.M. Newman, C.L. Liu, M.R. Green, A.J. Gentles, W. Feng, Y. Xu, C.D. Hoang, M. Diehn, A.A. Alizadeh, Robust enumeration of cell subsets from tissue expression profiles, Nat. Methods 12 (5) (2015) 453–457.
[20] J.K. Tay, B. Narasimhan, T. Hastie, Elastic net regularization paths for all generalized linear models, J Stat Softw 106 (2023).
[21] K. Yoshihara, M. Shahmoradgoli, E. Martinez, R. Vegesna, H. Kim, W. Torres-Garcia, V. Trevino, H. Shen, P.W. Laird, D.A. Levine, et al., Inferring tumour purity and stromal and immune cell admixture from expression data, Nat. Commun. 4 (2013) 2612.
[22] V. Thorsson, D.L. Gibbs, S.D. Brown, D. Wolf, D.S. Bortone, T.H. Ou Yang, E. Porta-Pardo, G.F. Gao, C.L. Plaisier, J.A. Eddy, et al., The immune landscape of cancer, Immunity 48 (4) (2018) 812–830 e814.

[23] X. He, C. Xu, Immune checkpoint signaling and cancer immunotherapy, Cell Res. 30 (8) (2020) 660–669.
[24] S.A. Kovacs, J.T. Fekete, B. Gyorffy, Predictive biomarkers of immunotherapy response with pharmacological applications in solid tumors, Acta Pharmacol. Sin. 44 (9) (2023) 1879–1889.
[25] C.C. Wu, Y.A. Wang, J.A. Livingston, J. Zhang, P.A. Futreal, Prediction of biomarkers and therapeutic combinations for anti-PD-1 immunotherapy using the global gene network association, Nat. Commun. 13 (1) (2022) 42.
[26] X. Li, Z. Dai, X. Wu, N. Zhang, H. Zhang, Z. Wang, X. Zhang, X. Liang, P. Luo, J. Zhang, et al., The comprehensive analysis identified an autophagy signature for the prognosis and the immunotherapy efficiency prediction in lung adenocarcinoma, Front. Immunol. 13 (2022) 749241.
[27] P. Sharma, S. Hu-Lieskovan, J.A. Wargo, A. Ribas, Primary, adaptive, and acquired resistance to cancer immunotherapy, Cell 168 (4) (2017) 707–723.
[28] G.Y. Lyu, Y.H. Yeh, Y.C. Yeh, Y.C. Wang, Mutation load estimation model as a predictor of the response to cancer immunotherapy, NPJ Genom Med 3 (2018) 12.
[29] D.T. Le, J.N. Durham, K.N. Smith, H. Wang, B.R. Bartlett, L.K. Aulakh, S. Lu, H. Kemberling, C. Wilt, B.S. Luber, et al., Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade, Science 357 (6349) (2017) 409–413.
[30] X. Meng, Z. Huang, F. Teng, L. Xing, J. Yu, Predictive biomarkers in PD-1/PD-L1 checkpoint blockade immunotherapy, Cancer Treat Rev. 41 (10) (2015) 868–876.
[31] P. Charoentong, F. Finotello, M. Angelova, C. Mayer, M. Efremova, D. Rieder, H. Hackl, Z. Trajanoski, Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade, Cell Rep. 18 (1) (2017) 248–262.
[32] P. Jiang, S. Gu, D. Pan, J. Fu, A. Sahu, X. Hu, Z. Li, N. Traugh, X. Bu, B. Li, et al., Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response, Nat Med 24 (10) (2018) 1550–1558.
[33] R. Bai, N. Chen, L. Li, N. Du, L. Bai, Z. Lv, H. Tian, J. Cui, Mechanisms of cancer resistance to immunotherapy, Front. Oncol. 10 (2020) 1290.
[34] J. Qi, H. Sun, Y. Zhang, Z. Wang, Z. Xun, Z. Li, X. Ding, R. Bao, L. Hong, W. Jia, et al., Single-cell and spatial analysis reveal interaction of FAP(+) fibroblasts and SPP1(+) macrophages in colorectal cancer, Nat. Commun. 13 (1) (2022) 1742.
[35] A. Yim, C. Smith, A.M. Brown, Osteopontin/secreted phosphoprotein-1 harnesses glial-, immune-, and neuronal cell ligand-receptor interactions to sense and regulate acute and chronic neuroinflammation, Immunol. Rev. 311 (1) (2022) 224–233.
[36] T. Wei, G. Bi, Y. Bian, S. Ruan, G. Yuan, H. Xie, M. Zhao, R. Shen, Y. Zhu, Q. Wang, et al., The significance of secreted phosphoprotein 1 in multiple human cancers, Front. Mol. Biosci. 7 (2020) 565383.
[37] Y. Zheng, S. Hao, C. Xiang, Y. Han, Y. Shang, Q. Zhen, Y. Zhao, M. Zhang, Y. Zhang, The correlation between SPP1 and immune escape of EGFR mutant lung adenocarcinoma was explored by bioinformatics analysis, Front. Oncol. 11 (2021) 592854.
[38] J.A. Kint, Fabry's disease: alpha-galactosidase deficiency, Science 167 (3922) (1970) 1268–1269.