

Exploration of signature based on T cell-related genes in stomach adenocarcinoma by analysis of single cell sequencing data

Huimei Wang^{1,*}, Nan An^{2,*}, Aiyue Pei¹, Yongxiao Sun¹, Shuo Li¹, Si Chen³, Nan Zhang¹

¹Department of Gastroenterology, The First Hospital of Jilin University, Changchun, China

²Department of Gastric Surgery, Sun Yat-sen University Cancer Center, Guangzhou, China

³Department of Colorectal and Anal Surgery, General Surgery Center, The First Hospital of Jilin University, Changchun, China

*Equal contribution

Correspondence to: Nan Zhang, Si Chen; **email:** zhangnan@jlu.edu.cn, chen_si@jlu.edu.cn

Keywords: gastric cancer, single-cell sequencing, immune microenvironment, immunotherapy, prognosis

Received: October 10, 2023

Accepted: December 29, 2023

Published: March 25, 2024

Copyright: © 2024 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Background: Gastric cancer (GC) is a leading reason for the death of cancer around the world. The immune microenvironment counts a great deal in immunotherapy of advanced tumors, in which T cells exert an indispensable function.

Methods: Single-cell RNA sequencing data were utilized to characterize the expression profile of T cells, followed by T cell-related genes (TCRGs) to construct signature and measure differences in survival time, enrichment pathways, somatic mutation status, immune status, and immunotherapy between groups.

Results: The complex tumor microenvironment was analyzed by scRNA-seq data of GC patients. We screened for these T cell signature expression genes and the TCRGs-based signature was successfully constructed and relied on the riskscore grouping. In gene set enrichment analysis, it was shown that pro-tumor and suppressive immune pathways were more abundant in the higher risk group. We also found different infiltration of immune cells in two groups, and that the higher risk samples had a poorer response to immunotherapy.

Conclusion: Our study established a prognostic model, in which different groups had different prognosis, immune status, and enriched features. These results have provided additional insights into prognostic evaluation and the development of highly potent immunotherapies in GC.

INTRODUCTION

Gastric cancer (GC) remains one of the most prevalent malignant tumors, with the fifth highest incidence and mortality rate worldwide [1]. It is assumed that more than one million new cases are diagnosed each year and the 5-year survival rate is only 20% [2, 3]. The most general type is stomach adenocarcinoma (STAD) [4]. While endoscopic treatment or surgery can potentially cure patients in early disease stages, adjuvant therapy is recommended for late-stage patients who have undergone prior surgery and have pathological T3 or T4 lesions or positive lymph node lesions [5, 6]. However, the survival rate of

STAD patients in late stages has remained stagnant despite the limited progress made in radiotherapy, chemotherapy, and surgery [4]. For this reason, it is critical to exploit new predictive models to raise the survival rate of STAD patients.

The tumor microenvironment (TME), made up of stromal cells, immunocytes, cytokines, blood vessels, and the extracellular substratum, acts as an important part of the invasion and progress of tumors, significantly affecting the efficacy of immunotherapy [7, 8]. T cells, which recognize and assemble tumor antigen-MHC complexes via T cell receptors, have been shown to be the essential element of antitumor

immunity, with chimeric antigen receptor (CAR) T cell therapy having been clinically demonstrated to be an effective immunotherapy [9–11]. Immune checkpoint inhibitors (ICIs) have already been extensively applied, with anti-cytotoxic T-lymphocyte-associated antigen (CTLA)-4 and anti-programmed death-protein (PD)-1 being the most common. In advanced GC patients, it has been shown to be a potentially effective therapeutic measure to target the PD-1/programmed death-ligand 1 (PD-L 1) axis [12, 13]. Nevertheless, just a fraction of STAD patients react to ICIs, and some even experience adverse effects [14, 15]. However, there are relatively few well-rounded structural molecular analyses of T cells in STAD, so we took T-cell-related genes into account to explore further possibilities for gastric cancer treatment by examining the immune landscape.

Single-cell RNA sequencing (scRNA-seq) techniques can potentially supply single-cell genetic analysis data and reveal different immune subpopulations in the TME, creating unprecedented opportunities for immunotherapy and targeted therapies against cancer [16]. Through integration of scRNA-seq and batch RNA-seq data, several studies have successfully developed new prognostic models for cancer [17–19]. Our study used scRNA-seq to synthesize T cell-related genes (TCRGs) data from STAD patients and construct a prospective signature on the basis of these genes, which can serve as a reliable prognostic biomarker, guiding the development of novel chemotherapeutic and immunotherapeutic approaches. We further analyzed the differences in survival status, TME, tumor mutational load (TMB), and drug sensitivity between different risk groups. New approaches to accommodate the heterogeneity and immune landscape in the clinical management of tumors are suggested in our discovery.

RESULTS

Analysis of immune microenvironment and establishment of T cell expression profile

Data of scRNA-seq for this study were obtained from 29 gastric cancer tissue samples with strict quality control, and 117,916 cells were retained. Meanwhile, we appointed “Harmony” R package to eliminate the batch effect of scRNA-seq data, eliminating batch differences due to different patients (Figure 1B). After log-normalization and dimensionality reduction, a total of 21 single-cell subclusters were obtained (Figure 1A). Then, the “findallmarkers” function was applied to screen the highly expressed genes specific to each subcluster, and displayed the top five genes (Figure 1D). Based on the characteristic genes

screened above and the classic markers of cell subclusters, a total of 12 subclusters were identified (Figure 1C, 1E). The “featureplot” function was used to display the expression of classic T cell genes to more accurately define T cell subsets (Figure 1F). Notably, we identified that multiple cell types were in a “transformed” state, such as “pericyte–fibroblast transition”, “epithelial–mesenchymal transition” and so on. In consequential cell communication studies, we found these cells in a transformed state had complex intercellular communication with T cells in the TME (Figure 1G). By reusing the “findallmarkers” function, we identified 1343 signature genes expressed by T cells. We then examined differentially expressed genes (DEGs) from normal tissues and tumors in the TCGA-STAD cohort (Figure 2A). Altogether, 9,574 up- and 764 down-regulated genes were ascertained. Using the Venn diagram, between tumor and normal tissues we screened 332 TCRGs variably expressed (Figure 2B). We carried out Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment (Supplementary Figure 1) to probe the biological features of the genes. KEGG pathway enrichment showed involvement in apoptosis, mRNA surveillance pathway, and IL-17 signaling pathway (Figure 2C). Through GO analysis, they were found to mainly involve leukocyte cell-cell adhesion, regulation of T cell activation, and nuclear periphery (Figure 2D).

Construction of the six-gene prognostic signature based on TCRGs

Supplementary Table 1 provides patient grouping information for TCGA modeling set and GSE62254 validation set. For constructing a prognostic signature based on TCRGs, TCGA-STAD cohort is our training set. First, univariate Cox analysis was performed on 332 TCRGs, revealing 16 TCRGs significantly associated with prognosis (Supplementary Table 2), with 9 high- and 7 lower risk genes (Figure 2E). The survival curve analysis of these 16 TCRGs indicated that higher risk TCRGs delivery was adversely associated with prognosis ($P < 0.05$) (Figure 2F, 2G; Supplementary Figure 2). Next, 16 TCRGs were submitted to LASSO cox regression analysis based on the optimal lambda value, resulting in the selection of 12 genes for further analysis (CTLA4, PDE3B, ZFP36, LBH, BCL11B, TAP1, TMC6, SAMD3, CMTM3, RGS2, SPATA13, ST8SIA4) (Figure 3A, 3B). At last, six most predictive genes (Supplementary Table 3) were obtained with stepwise multivariate Cox regression analysis. Using correlation coefficients, we developed a model as follows: risk score per patient = $(-0.408 \times \text{CTLA4 expression}) + (0.245 \times \text{PDE3B expression}) + (0.177 \times \text{ZFP36 expression}) + (-0.364 \times \text{BCL11B expression}) +$

($0.472 \times \text{SAMD3 expression}$) + ($0.362 \times \text{CMTM3 expression}$). The STAD patients were grouped into high- and lower risk groups according to neutral values in the training and validation set, which was assessed by ranking high to low risk score.

Validation and independent predictive effect of the signature

We estimated the prospective meaning of the TCRGs signature. Supplementary Figure 3A, 3B demonstrates

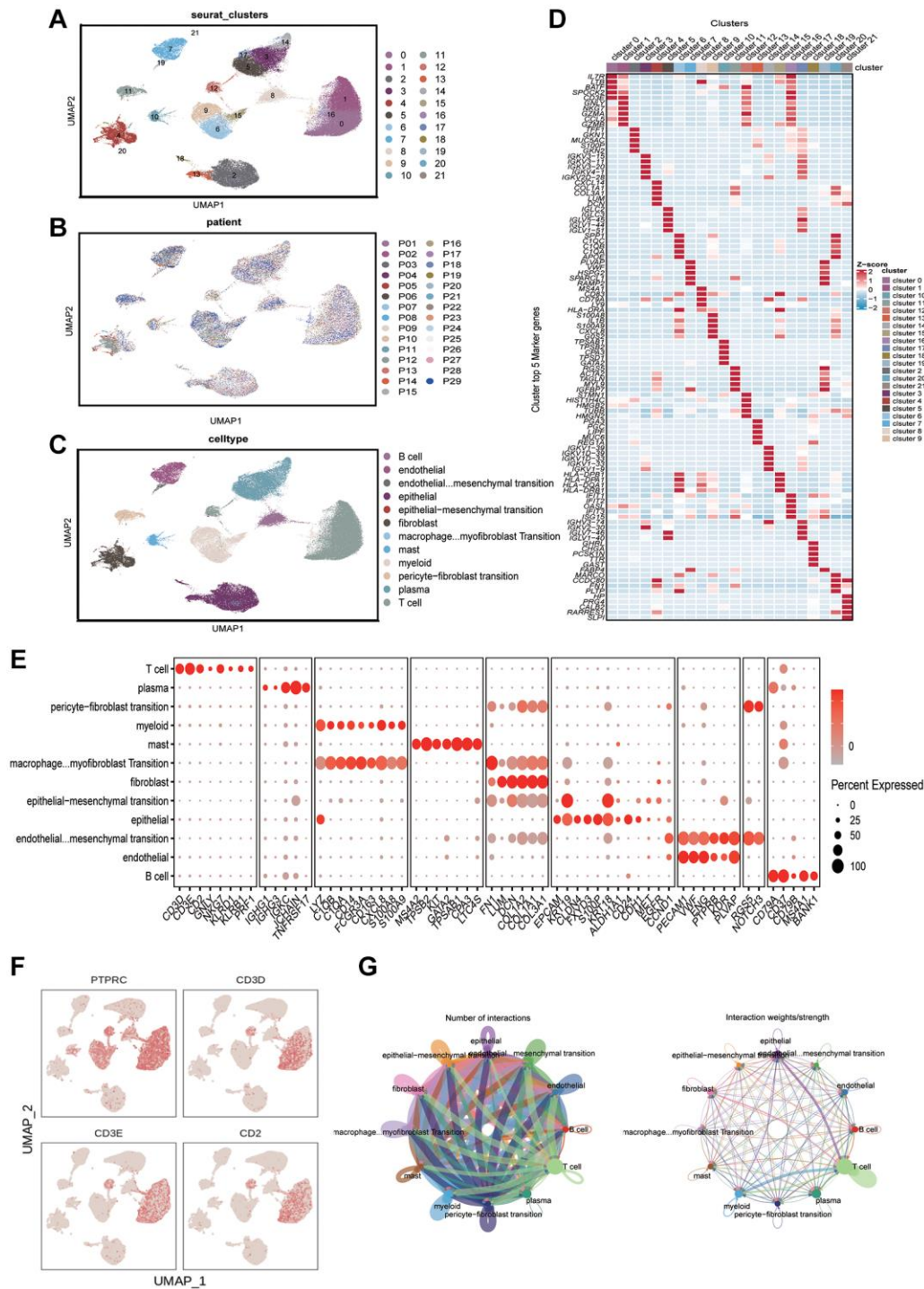


Figure 1. Analyzing the immune microenvironment. (A) Visualization of 21 subclusters based on UMAP algorithm. (B) Display of distribution of cells from different patients after removing batch effects. (C) Mark cell subclusters based on Top 5 genes and classic markers. (D) A heatmap showing the performance of the top five genes for each subclusters. (E) Display of gene expression in different subclusters based on classical cell markers. (F) The “featureplot” function describing the allocation of immune cell marker genes and T cell signature genes in the UMAP dimensionality reduction map. (G) Based on “cellchat” to display the signal strength of intercellular communication.

the six modeled gene expression levels in two groups, while Supplementary Figure 3C, 3F shows survival status and distribution. The K-M curve showed that the lower risk group also had a significantly longer survival time ($P < 0.001$) (Figure 3C, 3F). Model prediction accuracy was assessed using receiver operating characteristic (ROC) curves yielding the area under the curve (AUC) of 0.655, 0.742, and 0.739 for 1-, 3-, and 5-year survival in the training set (Figure 3D) and 0.645, 0.632, and 0.625 in validation set (Figure 3G). Obviously, a moderately accurate signature based on TCRGs was developed. The clinical characteristics and riskscore of the training group were submitted to univariate and multivariate Cox regression analyses to further explore the independent promotional value of riskscore. We discovered a significant association between riskscore and overall survival (OS) (Figure 3L) as an individual promotional factor (hazard ratio: 1.639, 95% CIs: 1.411–1.905, $P < 0.001$) (Figure 3E, 3H). This finding was confirmed in the validation group

(hazard ratio: 1.565, 95% CIs: 1.259–1.944, $P < 0.001$) (Supplementary Figure 3G, 3H).

Prediction of nomogram

We structured a nomogram to quantify clinical outcomes in STAD patients, which can predict survival by combining riskscore and clinical traits (Figure 3I). The calibration curve indicated a high concordance that exists in the predicted and respected values (Figure 3L). The decision curve analysis (DCA) curve showed the optimal clinical net benefit of the nomogram (Figure 3J), and ROC curve displayed its prediction accuracy (Figure 3K). In summary, the risk-constructed nomogram is more accurate than single factors for predicting prognosis in clinical settings.

Gene set enrichment analysis (GSEA)

We conducted GSEA to measure functional variations in prognostic markers between two groups (Figure 4A).

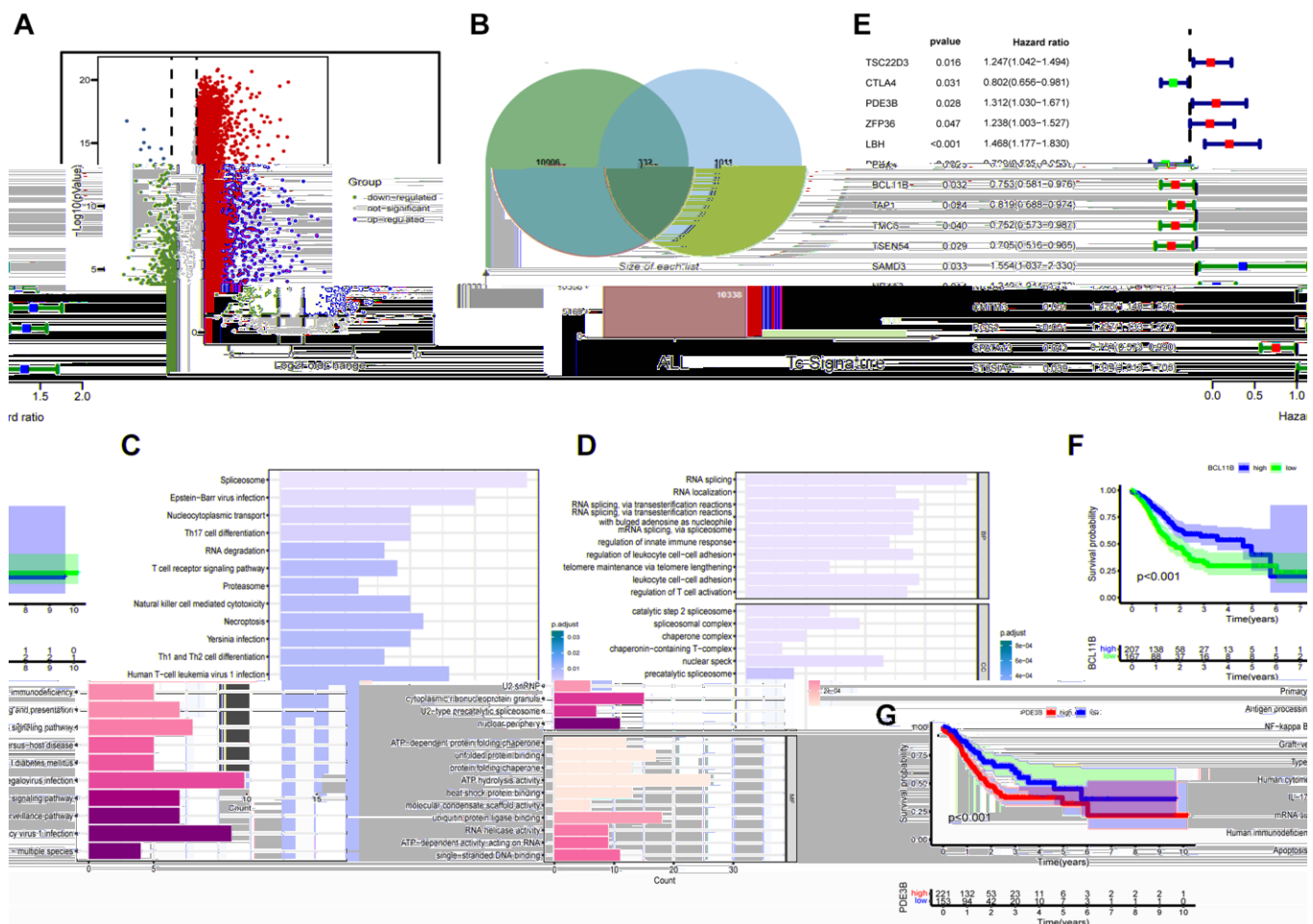


Figure 2. Functional enrichment and preliminary filters for TCRGs. (A) Volcano map of DEGs in TCGA-STAD cohort. **(B)** Venn diagram, **(C)** KEGG pathway analysis and **(D)** GO analysis of differentially expressed TCRGs. **(E)** Univariate Cox analysis of 16 prognosis-related TCRGs. K-M curves of single prognostic genes, such as **(F)** BCL11B and **(G)** PDE3B.

Six pathways, including the cancer pathways, PTEN, TGF- β and WNT signaling pathway, were more abundant in higher risk group. This observation suggests a greater abundance of pro-oncogenic pathways and immune pathway suppression in higher risk group, leading us to examine of immune related differences between the two groups further.

Evaluation of TME and immune-related genes

ESTIMATE could assess stromal and immune cell scores. The heatmap revealed differences in TME between two groups (Figure 4B), while the violin plot showed higher stromal score and ESTIMATE score in higher risk group, but no remarkable difference in

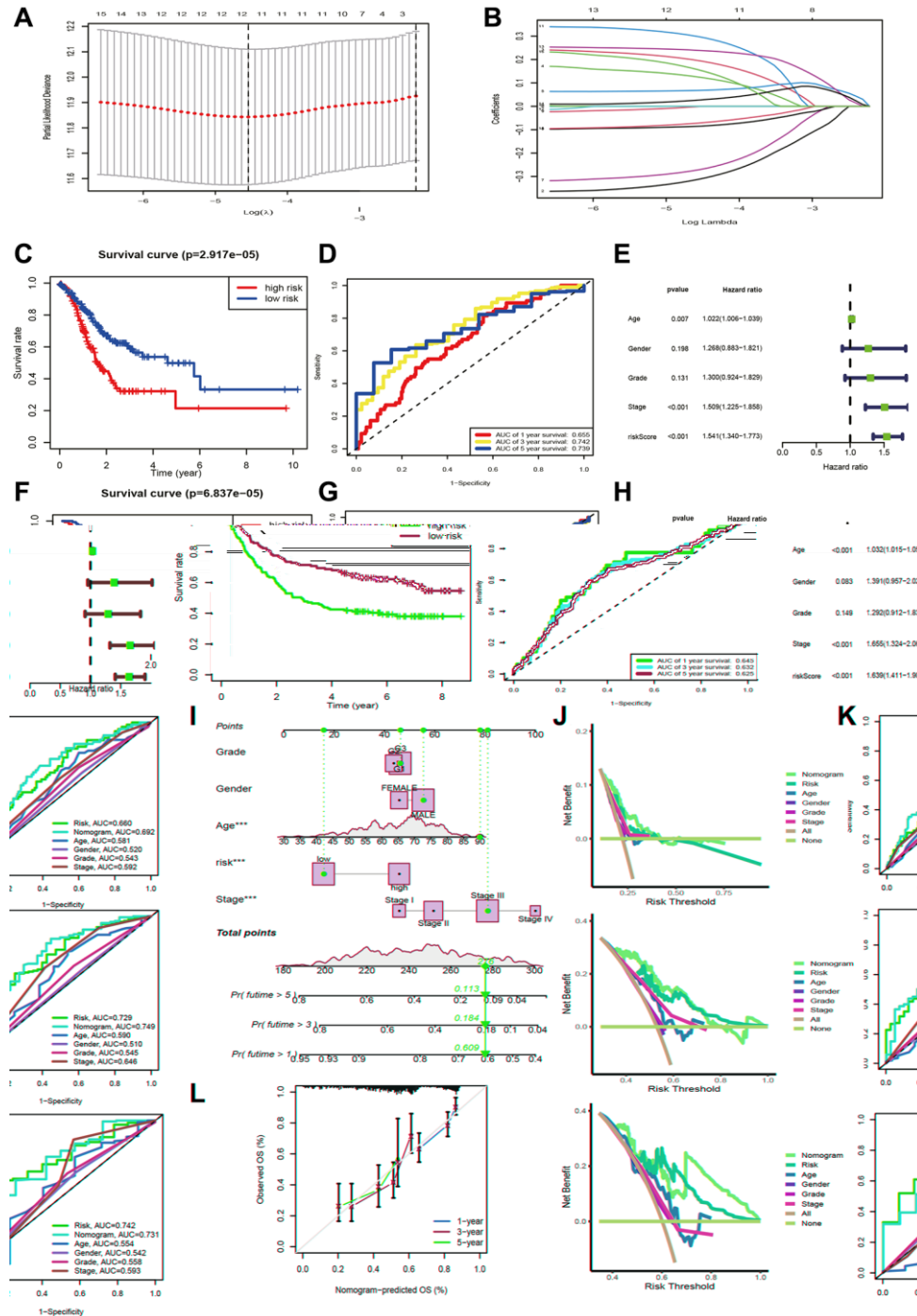


Figure 3. Prognostic signature and nomogram construction. (A, B) LASSO regression analysis. (C, F) K-M and (D, G) ROC curves for training and validation set. Clinical characteristics and riskscore of training set for (E) univariate and (H) multivariate Cox analyses. (I) Construction of nomogram. (J) DCA curves for nomogram at 1-year, 3-year, and 5-year. (K) ROC curves of nomogram, risk and clinical traits at 1-year, 3-year, and 5-year. (L) Calibration curves for nomogram.

immune score (Figure 4F). Stromal score and riskscore were found to be positively correlated by correlation analysis, but ImmuneScore was not correlated (Figure 4J). Cibersort profiling of 22 immune cells revealed higher expression of B cells memory, CD4+ T cells memory activated, Dendritic cells activated, and

Eosinophils in lower risk group, while macrophage M2 and mast cells in high-risk group (Figure 4C). Comparing the exposure levels of nine immune checkpoints, the expression of all checkpoints was higher in lower risk group except for HAVCR2, while PD-L1 and CD40 was not statistically different

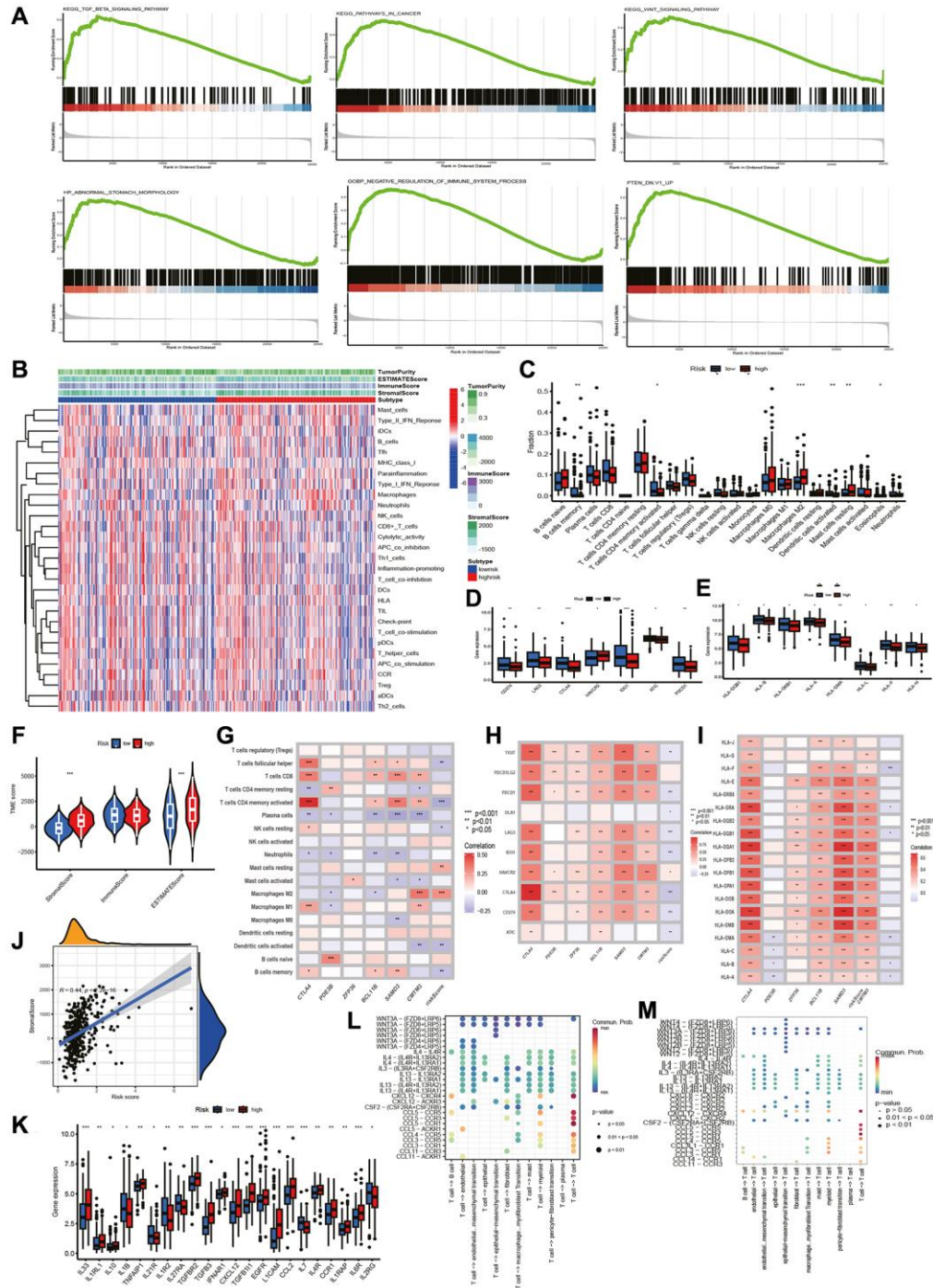


Figure 4. GSEA and TME analysis. GSEA and TME analysis. (A) Gene set enrichment analysis. (B) Heatmap shows risk and immune-related functions relationship. Differential expression of (C) 22 immune-associated cells, (D) immune checkpoint-associated genes, (E) HLA-associated genes, (F) stromal, immune and ESTIMATE scores, and (K) cytokine-related genes in two groups. Correlation of riskscore with (G) immune cells, (H) checkpoint-related genes, (I) HLA-related genes, (J) stromal score. (L) T cells initiate cytokine signal exchange between cells. (M) T cells receive cytokine signals to terminate intercellular signal communication.

(Figure 4D). Human leukocyte antigen (HLA)-related gene expression was also higher in lower risk group (Figure 4E). We then correlated riskscore and immune cells, immune checkpoints, HLA (Figure 4G–4I), and found negative correlations except for positive correlations between riskscore and mast, M2 cells, HAVCR2 immune checkpoints. We concluded that it is possible that lower risk group has superior prognosis following immunotherapy. Furthermore, we compared common cytokine expression levels, to find that IL21R, IL1R2, IL27RA, and IL2RG expressed higher levels in lower risk group (Figure 4K). Cellchat analysis based on scRNA-seq data revealed that the WNT pathway and interleukin chemokine Cxcl10 cytokines play significant roles in the complex communication between T cells and other cell types within the GC immune microenvironment (Figure 4L, 4M).

Gene mutation analysis

We analyzed the TMB of STAD patients and identified TTN, TP53, MUC16, ARID1A, and LRP1B

as top 5 mutations in both groups (Figure 5A, 5B). Lower risk group had higher TMB (Figure 5D). Figure 5F shows that TMB was negatively associated with riskscore. In addition, patients with higher TMB had a better prognosis, as shown in Figure 5C ($p = 0.017$). We also observed that the best prognosis was observed in the lower risk + higher TMB group, while the worst prognosis was observed in the higher risk + low-TMB group (Figure 5E). Further analysis revealed a significant correlation between low riskscore and MSI-L status, high riskscore and MSS status (Figure 5G, 5H).

Prediction of immunotherapy response

The tumor immune dysfunction and exclusion (TIDE) and The Cancer Immunome Atlas (TCIA) scores were used to assess immunotherapy and immune escape in TCGA-STAD cohort. For all four scores, including TIDE, a poorer response to immunotherapy was surveyed in patients in higher risk group (Figure 6A–6D). TCIA was employed to predict susceptibility to

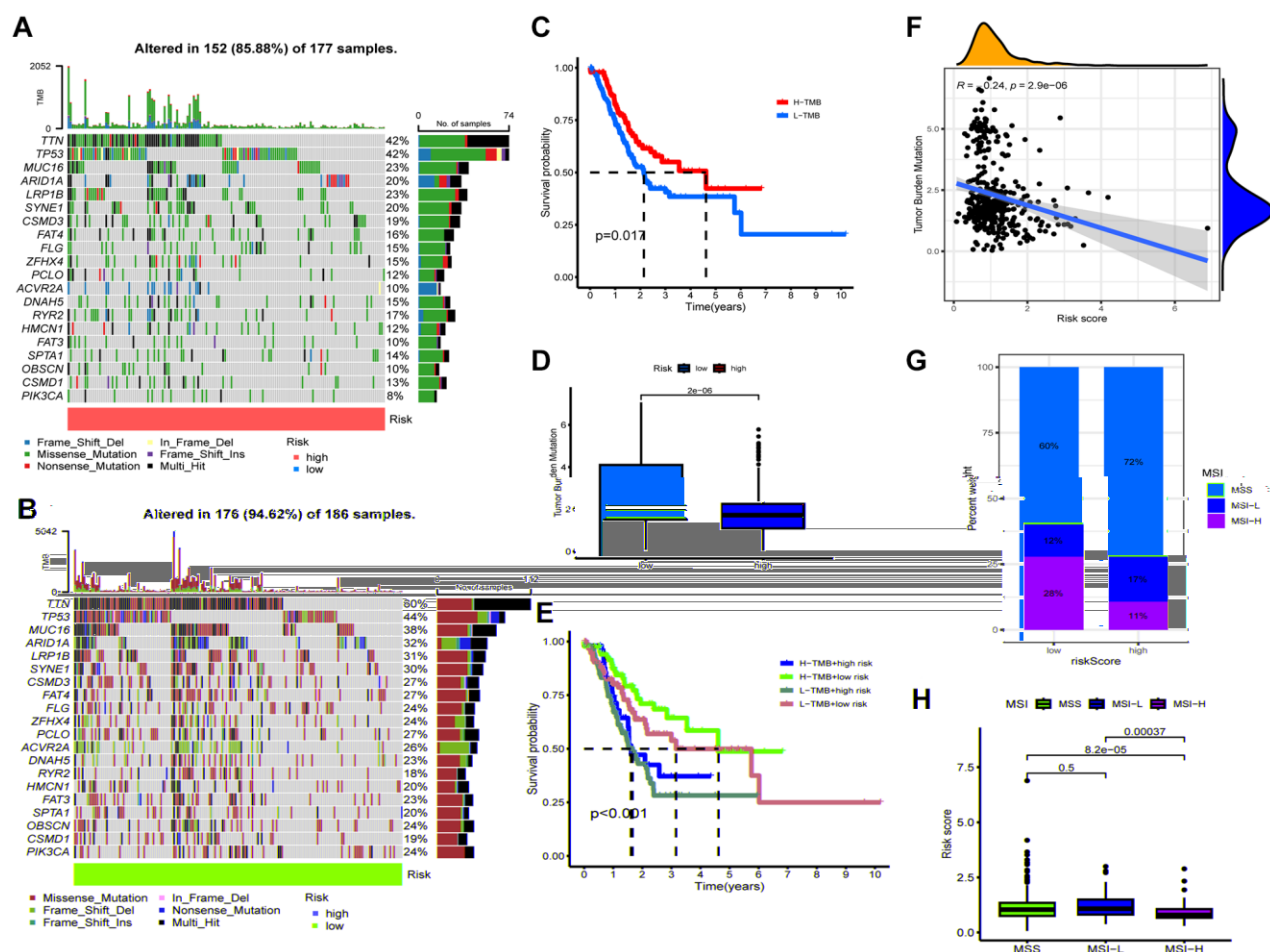


Figure 5. Characteristics of somatic mutations. Somatic mutations in different risk groups. (A) Higher risk and (B) lower risk. (C) Survival analysis of low and high tumor mutation burden (TMB). (D) Comparison of TMB levels. (E) Survival analysis of samples grouped with both risk and TMB. (F) Relevance between riskscore and TMB. (G, H) Relevance between riskscore and MSI.

immunotherapy. Outcomes revealed the expression of ips-ctla4-pos-pd1-pos, ips-ctla4-pos-pd1-neg, ips-ctla4-neg-pd1-neg, and ips-ctla4-neg-pd1-pos were higher in lower risk patients (Figure 6E). It indicated that the lower risk samples had better efficacy, regardless of CTLA 4 and PD-1 status. Then, we compared the accuracy of risk, TIDE score, and immune tumor infiltrating lymphocytes (TIL) score in predicting prognosis, and the TCRGs signature had the highest predictive accuracy (Figure 6F).

Drug susceptibility analysis

To ascertain the medical meaning of this signature, we also sought to determine the 50% maximum inhibitory concentration (IC_{50}) levels of chemotherapeutic drugs (Figure 6G). Lower risk patients were found to have a lower IC_{50} for both chemotherapy and targeted medications such as 5-Fluorouracil, Camptothecin, Cisplatin, Cyclophosphamide, Cytarabine, Epirubicin, Gemcitabine, Irinotecan, Oxaliplatin, and Vincristine. Therefore, the developed model could serve as a predictor for selecting anti-cancer drugs.

DISCUSSION

Accelerated advancement of scRNA-seq technology has resulted in an increasing number of studies exploring the structural characterization of TME [20, 21]. Immune cell regulation, particularly T cell activation and suppression, is strikingly shaping the control of diverse immune responses within TME [22, 23]. T cells occupy an important part in TME and are involved in immune escape and tumor progression [24]. Cancer immunotherapy using ICIs has become an effective therapy for tumor treatment; however, one of the reasons for its poor efficacy is the presence of immunosuppressive mechanisms in TME that diminish the effector function of CD8 tumor-infiltrating lymphocytes (TILs) [25], and T-cell up-regulation of PD-1 has become a major marker of T-cell dysfunction [26]. This study was designed to bioinformatically analyze the sc-RNA-seq spectrum of gastric cancer and identify TCRGs in gastric cancer tumor tissues. We resolved the gastric TME, analyzed multiple cell subtypes in the transformed state, anatomized the interaction of T cells with other cells in the

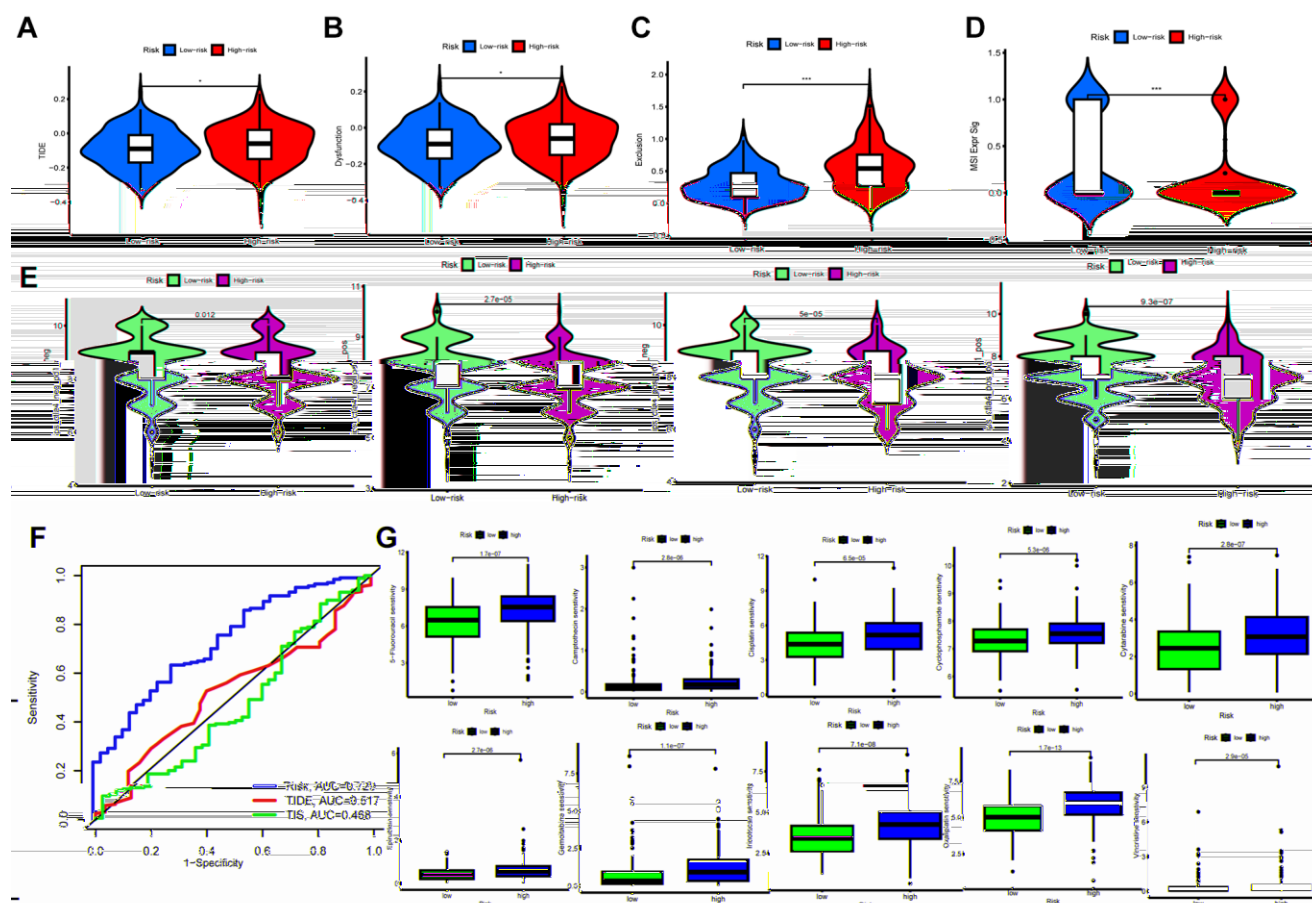


Figure 6. Immunological evaluation and drug sensitivity analysis. (A–D) TIDE, dysfunction, exclusion and MSI Expr Sig scores compared between two groups. (E) TCIA for predicting sensitivity to immunotherapy. (F) ROC curves comparing risk, TIDE, and TIS at OS. (G) Correlation between riskscore and chemotherapeutic sensitivity.

microenvironment, and identified T cell-specific gene expression. We then mined the TCGA database and developed a prognostic signature of six genes with fully validated predictive power in the GEO cohort. Subsequently, we noticed immune cell infiltration, the expressions of immune checkpoints and HLA were more abundant in the lower risk samples, and the response to immunotherapy was notably outperformed. In this research, a new prognostic model for gastric cancer was established based on T-cell-related genes in TME, which provides a theoretical basis for judging the efficacy of immunotherapy and developing new immunotherapy targets.

The prognostic signature of this study consisted of six TCRGs, including CTLA4, PDE3B, ZFP36, BCL11B, SAMD3, and CMTM3. CTLA-4 is a crucial adverse regulator in T cell responses, operating to maintain the tolerance of T cells to a self-antigen [27]. CTLA-4 limits the initiation of nascent T cells of lymphoid through B7 interactions, and Anti-CTLA-4 monoclonal antibodies (mAbs) exert immunotherapeutic effects mainly by depleting regulatory T (Treg) cells in TME and blocking transendocytosis of B7 in dendritic cells (DCs) [28, 29]. PDE3B functions importantly in regulating energy metabolism, particularly regulating compartmentalized cyclic adenosine monophosphate (cAMP)-signaling pathways [30], and its downstream enzyme, which breaks down triglycerides into fatty acids and glycerol, is associated with cAMP levels and lipolytic activity [31]. ZFP36 autonomously regulates the early activation kinetics of T cell by inhibiting the expression of activation markers, limiting T cell expansion and promoting apoptosis to inhibit enrichment and transformation of mRNA targets [32, 33]. CL11B binds directly to Id2 and represses it, a gene that encodes an E protein antagonist, to facilitate effector CD8 T cell expansion, memory formation and cytotoxic functions, and to promote and regulate NK cells differentiation [34]. SAMD3 engages in cell cycle control and also has the ability to bind to RNA and lipids, contributing to memory differentiation in CD8 T cells and is also highly expressed in NK cells [35]. CMTM3 is a neoplasm inhibitory gene that inhibits the tumorigenicity of GC cells mediated by epidermal growth factor receptor (EGFR) by increasing the activity of Rab5. CMTM3 knockdown promotes GC metastasis through the STAT3/Twist1/EMT signaling pathway [36]. The TCRGs model established in this study provides a possible molecular mechanism to further study the clinical aspects of STAD.

Further validation of the prognostic model performance built on six TCRGs was tested in the GEO cohort. Consistent results were observed across two lineups, indicating that the signal had good

robustness and repeatability. We also setup a nomogram, and used AUC curves, calibration curves, and DCA curves to visualize and test the accuracy of the predictions. Consequently, the nomogram can direct the development of personalized predictive models of STAD patients. Performing GSEA analysis on signature genes identified immune and tumor pathways can be enrolled markedly. Following this, we aimed to reveal the differential immune infiltration patterns screened by immune-related features. The relationship between riskscore and TME was investigated. First, higher risk group had higher stromal score and ESTIMATE score. Subsequently, 22 levels of immune cell infiltration demonstrated a greater proportion of activated memory CD4+ T cells, memory B cells, activated Dendritic cells, and Eosinophils in lower risk patients, indicating that a relatively active antitumor immune response may be present [37]. It was noted that higher risk group had more resting Macrophages M2 and Mast cells, and high macrophage infiltration rates can lead to poor prognosis in gastric cancer [38], and it has been shown that macrophages can significantly promote gastric cancer metastasis by enhancing the EMT process [39]. Several common immune checkpoints (CD274, CTLA-4, LAG3, PD-CD1, ATIC) were identified to be highly expressed in low-risk populations for better clinical application of ICI. HLA acts as an antigen presenting factor regulating immune response in gastric cancer [40]. Lower risk group had higher expression of HLA-related genes, indicating a more active local immune response. We conclude that lower risk group had greater likelihood of benefiting from immunotherapy. Analysis of TMB revealed a higher tolerance in lower risk group and a meaningful association between low riskscore and MSI-L status, which is consistent with our previously obtained findings. Finally, to study sensitivity to chemotherapeutic agents for clinical application, we performed pharmacosensitivity analysis in different risk groups and detected that 5-Fluorouracil, Camptothecin, Cisplatin, Cyclophosphamide, Cytarabine, Epirubicin, Gemcitabine, Irinotecan, Oxaliplatin and Vincristine had lower IC₅₀.

Although this study developed a novel and validated prognostic model, it still exists a few limitations. First, our study is a retrospective survey conducted on public database and should be prospectively and multi-center verified, in addition to validation or research with our own sequencing data. We identified six genes that were highly correlated with prognosis, and explored the biological significance, and we will further analyze their downstream mechanisms by means of immunohistochemistry and animal experiments in the future to explore how TCRGs affect TME and thus

prognosis. In addition, T-cells include exhausted T-cells, tissue-resident T-cells and other subtypes, and we plan to further subdivide the cell subtypes and explore the expression of these gene markers in them and the biological functions they play, so as to further explore the potentials of these gene markers, as well as the signature, in immunotherapy, in order to develop more targeted therapeutic agents for gastric cancer.

Collectively, we combined RNA-seq data to develop a prognostic signature based on TCRGs as well as discussing the differences of TME, TMB, and drug sensitivity in two groups, thus providing novel perspectives for further treatment of STAD clinically.

MATERIALS AND METHODS

Data collection

Data were collected from various databases. The single-cell sequencing data for 29 primary gastric adenocarcinoma samples were derived from the GEO database (GSE183904). Clinical dataset and genomic details of STAD patients were derived from 379 tumor samples and 34 normal samples in The Cancer Genome Atlas database (TCGA). After removing tumor samples without survival information, 367 tumor samples were used. Additionally, GSE62254 ($n = 300$) was downloaded in the GEO database to verify the predictive capability of the signature. We corrected for abiotic bias with “ComBat” algorithm of “sva” package between different datasets [41].

Characterizing T-cell marker genes with scRNA-seq data

We analyzed scRNA-seq data with “Seurat” R package. Initially, we used the “Seurat” R package to read $10 \times$ Genomics single cell sequencing data and create Seurat objects. We did quality check by removing cells expressing fewer than 100 genes or more than 5% mitochondrial genes. Only genes that are expressed in a minimum of three single cells were preserved. Then, we used the function “NormalizeData” to normalize the scRNA-seq results and the function “FindVariableFeatures” to identify the upper 2000 high variant genes. Next, perform principal component analysis (PCA) by applying “RunPCA” function, and the first 30 dimensions were selected to decrease the dimensionality of scRNA-seq data for the upper 2000 genes. Then we removed batch effects for 21 samples using the “RunHarmony” function. We clustered single cells into different subgroups by “FindNeighbors” and “FindClusters” functions ($\text{dim} = 50$ and $\text{resolution} = 0.5$). We performed uniform manifold approximation

and projection (UMAP) to visualize the clustering units at the two-dimensional level. After annotating by classical cell subpopulation marker genes, DEGs were counted for each cluster with the Wilcoxon-Mann-Whitney test employing the “FindAllMarkers” function in the “Seurat” package. Cut-off thresholds were adjusted for p -values < 0.01 and $|\text{Log}_2(\text{foldchange})| > 1$ to determine the marker genes.

Cell communication analysis

The “cellchat” function was applied to analyze the complex intercellular communication network in the TME. We used the major signal inputs as well as the major signal outputs of cell subpopulations in the TME from its built-in database, and the “netVisual_circle” function to visualize the strength of the intercellular communication network between the target cell cluster and other cell clusters throughout the TME.

TCGA data analysis and functional enrichment

DEGs were accessed with “limma” package [42]. The threshold used was $\text{FDR} < 0.05$ and $|\text{Log}_2(\text{foldchange})| > 1$. Venn diagrams were then used to identify TCRGs with differential expressions. For GO analysis and KEGG pathway analysis [43, 44], we employed “ClusterProfiler”, “org.Hs.eg.db”, “GOplot”, and “enrichplot” packages.

Establishment and confirmation of prognosis signature

Initially, the differential expressions of TCRGs were analyzed by single factor Cox regression ($p < 0.05$) and we drew Kaplan-Meier curves for each gene, which was gained from the “survival” and “survminer” packages. LASSO Cox regression using the “glmnet” package ($p < 0.05$) was done to prevent overfitting. Optimal λ was chosen to eliminate similar genes. Next, a signature of TCRGs was developed by stepwise multivariate Cox regression analysis, resulting in $\text{riskscore for each patient} = \beta_{\text{gene1}} \times \text{Exp}_{\text{gene1}} + \beta_{\text{gene2}} \times \text{Exp}_{\text{gene2}} + \dots + \beta_{\text{genen}} \times \text{Exp}_{\text{genen}}$. Patients were divided into high-risk and low-risk groups in terms of median riskscore. We performed survival analysis with the Kaplan-Meier method, comparing the diversity of OS between the two groups using the long rank test. For drawing ROC curves and computing the AUC, we used the “survivalminer” and “survivalROC” packages. The TCRGs signature was verified in 300 STAD cases in GSE62254, and riskscore was derived for every sample considering the identical equation. Again, to affirm the predictive power of the feature, survival analysis and ROC curves were plotted.

Establishment and verification of nomogram

Utilizing clinical features and TCRGs signature, a nomogram was created with “rms” package to estimate the 1-, 3-, and 5-year survival rate. A score for each variable was assigned by the nomogram scoring system and they were combined to obtain the total score for a single sample [45]. The signature’s precision was measured by calibration curves as well as ROC curves, while net clinical benefit was assessed using DCA curves [46].

Gene set enrichment analysis

Functional and pathway variations between risk groups were studied by GSEA [47]. GSEA software (<http://software.broadinstitute.org/gsea>) was used for loading GO and KEGG gene sets, as well as phenotype tag expression files. The threshold was $P < 0.05$ and $FDR < 0.25$ to make significance.

Assessment of tumor immune microenvironment

Evaluation of stromal and immune cells within the tumor tissue was performed with “estimation” package to determine immune score, interstitial score, estimated score and tumor purity in STAD patients [48]. Heatmap was generated utilizing the “pheatmap” package. The CIBERSORT algorithm [49] was applied in order to acquire the infiltrative characteristics in 22 types of immune cells, while we employed the single sample GSEA (ssGSEA) to determine the activity of immune cells and immune functions in each sample. The expression of immune checkpoints and HLA-related genes were compared between two groups with Wilcoxon test, and correlations between immune cells, immune checkpoints, HLA-related expression and risk scores were analyzed.

Analysis of TMB

TMB is meant to be the count of somatic mutations, insertions and deletions in the coding regions of the genome per million bases. We obtained patient somatic cell mutation data in TCGA-STAD cohort. TMB was computed for each sample, and “maftools” package was deployed for the visualization of somatic mutation patterns [50]. K-M curves were applied to assess the diversity of OS between different risk groups. The association between TMB and riskscore was also evaluated.

Immune status in risk groups and drug sensitivity

The TIDE website (<http://tide.dfci.harvard.edu>) features the TIDE score files, which were assessed

immunotherapy and immune escape in STAD patients [51]. We analyzed the immune genome and obtained the immune treatment scoring file with the TCIA online platform (<https://tcia.at/home>) [52]. The variation between different groups for these scores was then analyzed. We also used the “pRRophetic” software package [53] to calculate chemotherapy response with IC₅₀.

Statistical analysis

Statistical analyses were performed by R software version 4.2.0 (<http://www.R-project.org>). The levels of significance had the following annotations: * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$.

Data availability

Publicly available datasets were analyzed in this study. These data can be found here: <https://portal.gdc.cancer.gov/> and GEO (GSE62254) database.

AUTHOR CONTRIBUTIONS

Conceptualization, H.W. and N.A.; validation, A.P. and L.S.; writing-original draft preparation, H.W. and N.A.; writing-review and editing, N.Z. and S.C.; visualization, H.W., N.A. and Y.S.; project administration, N.Z.; funding acquisition, N.Z. and S.C.; final approval, all authors.

ACKNOWLEDGMENTS

Thanks to TCGA and GEO database for the available data.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

FUNDING

Our research manuscript receives backing from these funding sources, including the Department of Education (Grant No. JJKH20211209KJ), Jilin Provincial Science and Technology Department (Grant No. 20230203071SF), the Jilin Provincial Innovation and Entrepreneurship Talents Support Project (Grant No. 2023QN07), and the Jilin Provincial Finance Department (Grant No. JLSWSRCZX2020-083).

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer

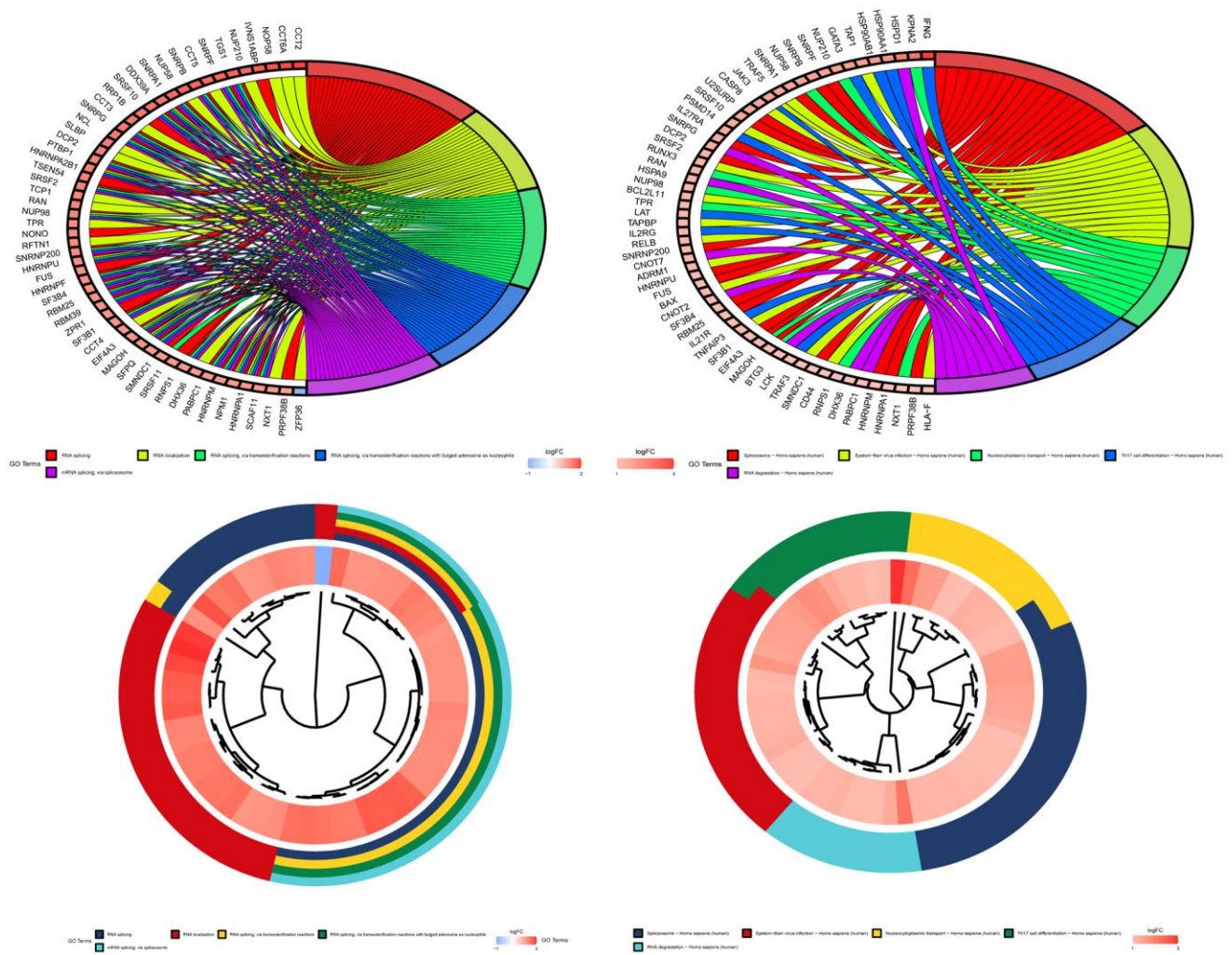
- Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021; 71:209–49.
<https://doi.org/10.3322/caac.21660>
PMID:33538338
2. Smyth EC, Nilsson M, Grabsch HI, van Grieken NC, Lordick F. Gastric cancer. *Lancet.* 2020; 396:635–48.
[https://doi.org/10.1016/S0140-6736\(20\)31288-5](https://doi.org/10.1016/S0140-6736(20)31288-5)
PMID:32861308
 3. Ilic M, Ilic I. Epidemiology of stomach cancer. *World J Gastroenterol.* 2022; 28:1187–203.
<https://doi.org/10.3748/wjg.v28.i12.1187>
PMID:35431510
 4. Vrána D, Matzenauer M, Neoral Č, Aujeský R, Vrba R, Melichar B, Rušarová N, Bartoušková M, Jankowski J. From Tumor Immunology to Immunotherapy in Gastric and Esophageal Cancer. *Int J Mol Sci.* 2018; 20:13.
<https://doi.org/10.3390/ijms20010013>
PMID:30577521
 5. Li GZ, Doherty GM, Wang J. Surgical Management of Gastric Cancer: A Review. *JAMA Surg.* 2022; 157:446–54.
<https://doi.org/10.1001/jamasurg.2022.0182>
PMID:35319717
 6. Joshi SS, Badgwell BD. Current treatment and recent progress in gastric cancer. *CA Cancer J Clin.* 2021; 71:264–79.
<https://doi.org/10.3322/caac.21657>
PMID:33592120
 7. Bejarano L, Jordão MJC, Joyce JA. Therapeutic Targeting of the Tumor Microenvironment. *Cancer Discov.* 2021; 11:933–59.
<https://doi.org/10.1158/2159-8290.CD-20-1808>
PMID:33811125
 8. Anderson NM, Simon MC. The tumor microenvironment. *Curr Biol.* 2020; 30:R921–5.
<https://doi.org/10.1016/j.cub.2020.06.081>
PMID:32810447
 9. Wang M, Gao P, Ren L, Duan J, Yang S, Wang H, Wang H, Sun J, Gao X, Li B, Li S, Su W. Profiling the peripheral blood T cell receptor repertoires of gastric cancer patients. *Front Immunol.* 2022; 13:848113.
<https://doi.org/10.3389/fimmu.2022.848113>
PMID:35967453
 10. Joyce JA, Fearon DT. T cell exclusion, immune privilege, and the tumor microenvironment. *Science.* 2015; 348:74–80.
<https://doi.org/10.1126/science.aaa6204>
PMID:25838376
 11. Kishton RJ, Sukumar M, Restifo NP. Metabolic Regulation of T Cell Longevity and Function in Tumor Immunotherapy. *Cell Metab.* 2017; 26:94–109.
<https://doi.org/10.1016/j.cmet.2017.06.016>
PMID:28683298
 12. Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. *Science.* 2018; 359:1350–5.
<https://doi.org/10.1126/science.aar4060>
PMID:29567705
 13. Kono K, Nakajima S, Mimura K. Current status of immune checkpoint inhibitors for gastric cancer. *Gastric Cancer.* 2020; 23:565–78.
<https://doi.org/10.1007/s10120-020-01090-4>
PMID:32468420
 14. Shah MA, Kennedy EB, Alarcon-Rozas AE, Alcindor T, Bartley AN, Malowany AB, Bhadkamkar NA, Deighton DC, Janjigian Y, Karipott A, Khan U, King DA, Klute K, et al. Immunotherapy and Targeted Therapy for Advanced Gastroesophageal Cancer: ASCO Guideline. *J Clin Oncol.* 2023; 41:1470–91.
<https://doi.org/10.1200/JCO.22.02331>
PMID:36603169
 15. Takei S, Kawazoe A, Shitara K. The New Era of Immunotherapy in Gastric Cancer. *Cancers (Basel).* 2022; 14:1054.
<https://doi.org/10.3390/cancers14041054>
PMID:35205802
 16. Lei Y, Tang R, Xu J, Wang W, Zhang B, Liu J, Yu X, Shi S. Applications of single-cell sequencing in cancer research: progress and perspectives. *J Hematol Oncol.* 2021; 14:91.
<https://doi.org/10.1186/s13045-021-01105-2>
PMID:34108022
 17. Ding S, Chen X, Shen K. Single-cell RNA sequencing in breast cancer: Understanding tumor heterogeneity and paving roads to individualized therapy. *Cancer Commun (Lond).* 2020; 40:329–44.
<https://doi.org/10.1002/cac2.12078>
PMID:32654419
 18. Liu Y, Shou Y, Zhu R, Qiu Z, Zhang Q, Xu J. Construction and Validation of a Ferroptosis-Related Prognostic Signature for Melanoma Based on Single-Cell RNA Sequencing. *Front Cell Dev Biol.* 2022; 10:818457.
<https://doi.org/10.3389/fcell.2022.818457>
PMID:35309911
 19. Song P, Li W, Guo L, Ying J, Gao S, He J. 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.* 2022; 13:850745.
<https://doi.org/10.3389/fimmu.2022.850745>
PMID:35757748

20. Zaitsev A, Chelushkin M, Dyikanov D, Cheremushkin I, Shpak B, Nomie K, Zyrin V, Nuzhdina E, Lozinsky Y, Zotova A, Degryse S, Kotlov N, Baisangurov A, et al. Precise reconstruction of the TME using bulk RNA-seq and a machine learning algorithm trained on artificial transcriptomes. *Cancer Cell*. 2022; 40:879–94.e16.
<https://doi.org/10.1016/j.ccell.2022.07.006>
PMID:[35944503](https://pubmed.ncbi.nlm.nih.gov/35944503/)
21. Lei S, Gao Y, Li J, Chen X, Zhou W, Wu J, Ma P, Men K, Duan X. Dual-RNA controlled delivery system inhibited tumor growth by apoptosis induction and TME activation. *J Control Release*. 2022; 344:97–112.
<https://doi.org/10.1016/j.jconrel.2022.02.022>
PMID:[35189260](https://pubmed.ncbi.nlm.nih.gov/35189260/)
22. Vesely MD, Zhang T, Chen L. Resistance Mechanisms to Anti-PD Cancer Immunotherapy. *Annu Rev Immunol*. 2022; 40:45–74.
<https://doi.org/10.1146/annurev-immunol-070621-030155>
PMID:[35471840](https://pubmed.ncbi.nlm.nih.gov/35471840/)
23. Sánchez-Paulete AR, Teijeira A, Cueto FJ, Garasa S, Pérez-Gracia JL, Sánchez-Arráez A, Sancho D, Melero I. Antigen cross-presentation and T-cell cross-priming in cancer immunology and immunotherapy. *Ann Oncol*. 2017; 28:xii44–55.
<https://doi.org/10.1093/annonc/mdx237>
PMID:[28945841](https://pubmed.ncbi.nlm.nih.gov/28945841/)
24. Yi M, Li T, Niu M, Mei Q, Zhao B, Chu Q, Dai Z, Wu K. Exploiting innate immunity for cancer immunotherapy. *Mol Cancer*. 2023; 22:187.
<https://doi.org/10.1186/s12943-023-01885-w>
PMID:[38008741](https://pubmed.ncbi.nlm.nih.gov/38008741/)
25. Xiong D, Zhang L, Sun ZJ. Targeting the epigenome to reinvigorate T cells for cancer immunotherapy. *Mil Med Res*. 2023; 10:59.
<https://doi.org/10.1186/s40779-023-00496-2>
PMID:[38044445](https://pubmed.ncbi.nlm.nih.gov/38044445/)
26. Thommen DS, Schumacher TN. T Cell Dysfunction in Cancer. *Cancer Cell*. 2018; 33:547–62.
<https://doi.org/10.1016/j.ccell.2018.03.012>
PMID:[29634943](https://pubmed.ncbi.nlm.nih.gov/29634943/)
27. Ise W, Kohyama M, Nutsch KM, Lee HM, Suri A, Unanue ER, Murphy TL, Murphy KM. CTLA-4 suppresses the pathogenicity of self antigen-specific T cells by cell-intrinsic and cell-extrinsic mechanisms. *Nat Immunol*. 2010; 11:129–35.
<https://doi.org/10.1038/ni.1835>
PMID:[20037585](https://pubmed.ncbi.nlm.nih.gov/20037585/)
28. Rowshanravan B, Halliday N, Sansom DM. CTLA-4: a moving target in immunotherapy. *Blood*. 2018; 131:58–67.
<https://doi.org/10.1182/blood-2017-06-741033>
PMID:[29118008](https://pubmed.ncbi.nlm.nih.gov/29118008/)
29. Du X, Tang F, Liu M, Su J, Zhang Y, Wu W, Devenport M, Lazarski CA, Zhang P, Wang X, Ye P, Wang C, Hwang E, et al. A reappraisal of CTLA-4 checkpoint blockade in cancer immunotherapy. *Cell Res*. 2018; 28:416–32.
<https://doi.org/10.1038/s41422-018-0011-0>
PMID:[29472691](https://pubmed.ncbi.nlm.nih.gov/29472691/)
30. Chung YW, Lagranha C, Chen Y, Sun J, Tong G, Hockman SC, Ahmad F, Esfahani SG, Bae DH, Polidovitch N, Wu J, Rhee DK, Lee BS, et al. Targeted disruption of PDE3B, but not PDE3A, protects murine heart from ischemia/reperfusion injury. *Proc Natl Acad Sci U S A*. 2015; 112:E2253–62.
<https://doi.org/10.1073/pnas.1416230112>
PMID:[25877153](https://pubmed.ncbi.nlm.nih.gov/25877153/)
31. Xue Q, Kang R, Klionsky DJ, Tang D, Liu J, Chen X. Copper metabolism in cell death and autophagy. *Autophagy*. 2023; 19:2175–95.
<https://doi.org/10.1080/15548627.2023.2200554>
PMID:[37055935](https://pubmed.ncbi.nlm.nih.gov/37055935/)
32. Cook ME, Bradstreet TR, Webber AM, Kim J, Santeford A, Harris KM, Murphy MK, Tran J, Abdalla NM, Schwarzkopf EA, Greco SC, Halabi CM, Apte RS, et al. The ZFP36 family of RNA binding proteins regulates homeostatic and autoreactive T cell responses. *Sci Immunol*. 2022; 7:eabo0981.
<https://doi.org/10.1126/sciimmunol.abo0981>
PMID:[36269839](https://pubmed.ncbi.nlm.nih.gov/36269839/)
33. Moore MJ, Blachere NE, Fak JJ, Park CY, Sawicka K, Parveen S, Zucker-Scharff I, Moltedo B, Rudensky AY, Darnell RB. ZFP36 RNA-binding proteins restrain T cell activation and anti-viral immunity. *Elife*. 2018; 7:e33057.
<https://doi.org/10.7554/eLife.33057>
PMID:[29848443](https://pubmed.ncbi.nlm.nih.gov/29848443/)
34. Holmes TD, Pandey RV, Helm EY, Schlums H, Han H, Campbell TM, Drashansky TT, Chiang S, Wu CY, Tao C, Shoukier M, Tolosa E, Von Hardenberg S, et al. The transcription factor Bcl11b promotes both canonical and adaptive NK cell differentiation. *Sci Immunol*. 2021; 6:eabc9801.
<https://doi.org/10.1126/sciimmunol.abc9801>
PMID:[33712472](https://pubmed.ncbi.nlm.nih.gov/33712472/)
35. Peters AE, Knöpper K, Grafen A, Kastenmüller W. A multifunctional mouse model to study the role of Samd3. *Eur J Immunol*. 2022; 52:328–37.
<https://doi.org/10.1002/eji.202149469>
PMID:[34626120](https://pubmed.ncbi.nlm.nih.gov/34626120/)
36. Yuan W, Liu B, Wang X, Li T, Xue H, Mo X, Yang S, Ding S, Han W. CMTM3 decreases EGFR expression and EGF-mediated tumorigenicity by promoting Rab5 activity in gastric cancer. *Cancer Lett*. 2017; 386:77–86.

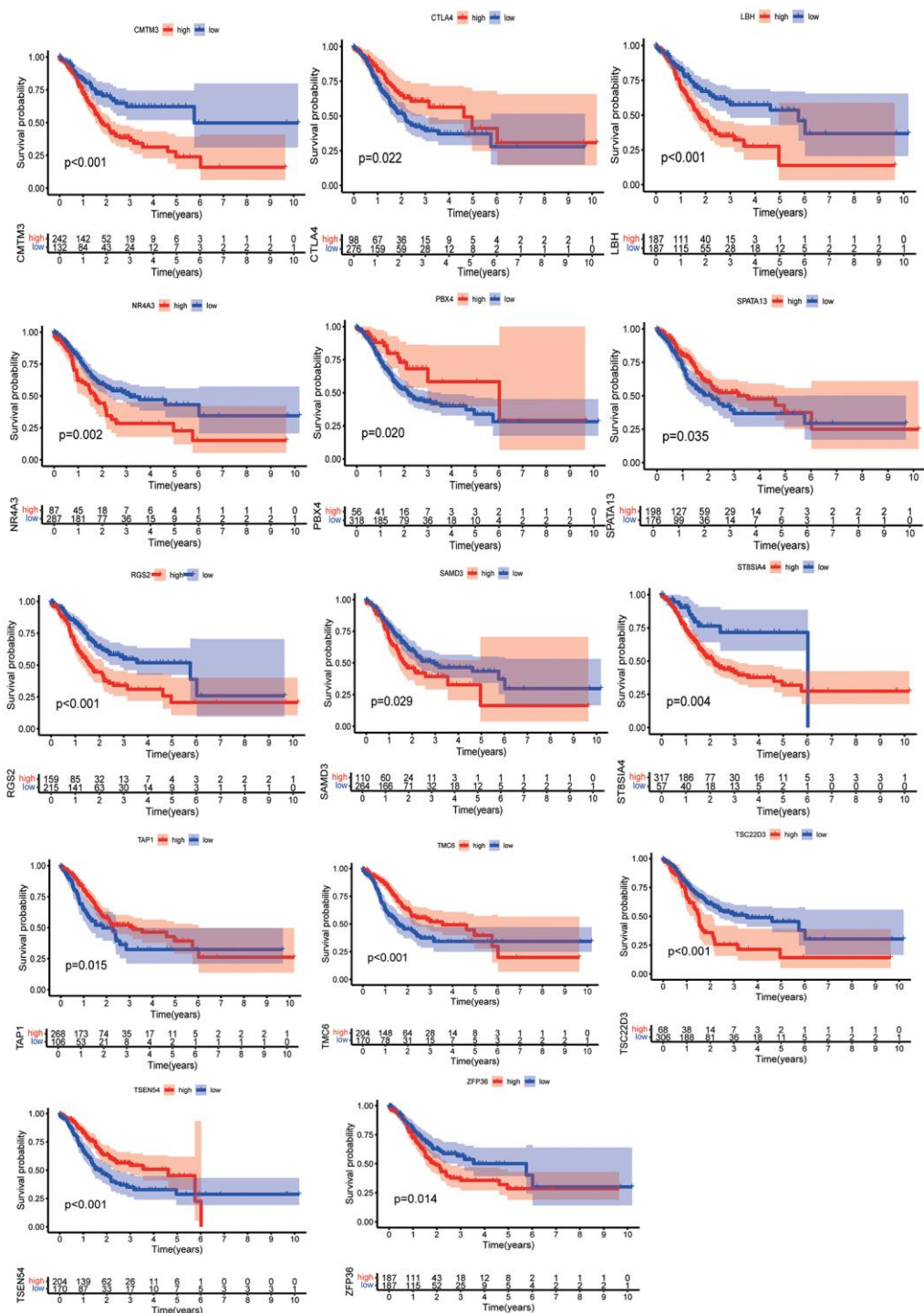
- <https://doi.org/10.1016/j.canlet.2016.11.015>
PMID: [27867015](#)
37. Li R, Liu H, Cao Y, Wang J, Chen Y, Qi Y, Lv K, Liu X, Yu K, Lin C, Zhang H, He H, Li H, et al. Identification and validation of an immunogenic subtype of gastric cancer with abundant intratumoural CD103⁺CD8⁺ T cells conferring favourable prognosis. *Br J Cancer*. 2020; 122:1525–34.
<https://doi.org/10.1038/s41416-020-0813-y>
PMID: [32205862](#)
 38. Pathria P, Louis TL, Varner JA. Targeting Tumor-Associated Macrophages in Cancer. *Trends Immunol*. 2019; 40:310–27.
<https://doi.org/10.1016/j.it.2019.02.003>
PMID: [30890304](#)
 39. Li W, Zhang X, Wu F, Zhou Y, Bao Z, Li H, Zheng P, Zhao S. Gastric cancer-derived mesenchymal stromal cells trigger M2 macrophage polarization that promotes metastasis and EMT in gastric cancer. *Cell Death Dis*. 2019; 10:918.
<https://doi.org/10.1038/s41419-019-2131-y>
PMID: [31801938](#)
 40. Xing R, Li L, Chen L, Gao Z, Wang H, Li W, Cui J, Tian G, Liang Q, Yu J, Sung JJ, Luo G, Gao H, et al. Copy number variations of HLA-I and activation of Nkp30 pathway determine the sensitivity of gastric cancer cells to the cytotoxicity of natural killer cells. *Oncogene*. 2016; 35:2584–91.
<https://doi.org/10.1038/ncr.2015.324>
PMID: [26364607](#)
 41. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012; 28:882–3.
<https://doi.org/10.1093/bioinformatics/bts034>
PMID: [22257669](#)
 42. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015; 43:e47.
<https://doi.org/10.1093/nar/gkv007>
PMID: [25605792](#)
 43. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012; 16:284–7.
<https://doi.org/10.1089/omi.2011.0118>
PMID: [22455463](#)
 44. Walter W, Sánchez-Cabo F, Ricote M. GOpot: an R package for visually combining expression data with functional analysis. *Bioinformatics*. 2015; 31:2912–4.
<https://doi.org/10.1093/bioinformatics/btv300>
PMID: [25964631](#)
 45. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol*. 2008; 26:1364–70.
<https://doi.org/10.1200/JCO.2007.12.9791>
PMID: [18323559](#)
 46. Fitzgerald M, Saville BR, Lewis RJ. Decision curve analysis. *JAMA*. 2015; 313:409–10.
<https://doi.org/10.1001/jama.2015.37>
PMID: [25626037](#)
 47. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005; 102:15545–50.
<https://doi.org/10.1073/pnas.0506580102>
PMID: [16199517](#)
 48. Yu S, Wang Y, Peng K, Lyu M, Liu F, Liu T. Establishment of a Prognostic Signature of Stromal/Immune-Related Genes for Gastric Adenocarcinoma Based on ESTIMATE Algorithm. *Front Cell Dev Biol*. 2021; 9:752023.
<https://doi.org/10.3389/fcell.2021.752023>
PMID: [34900998](#)
 49. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, Khodadoust MS, Esfahani MS, Luca BA, Steiner D, Diehn M, Alizadeh AA. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*. 2019; 37:773–82.
<https://doi.org/10.1038/s41587-019-0114-2>
PMID: [31061481](#)
 50. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. 2018; 28:1747–56.
<https://doi.org/10.1101/gr.239244.118>
PMID: [30341162](#)
 51. Killock D. Signatures IMPRES and might turn the TIDE in predicting responses. *Nat Rev Clin Oncol*. 2018; 15:654.
<https://doi.org/10.1038/s41571-018-0088-x>
PMID: [30171205](#)
 52. Kakoti S, Sato H, Laskar S, Yasuhara T, Shibata A. DNA Repair and Signaling in Immune-Related Cancer Therapy. *Front Mol Biosci*. 2020; 7:205.
<https://doi.org/10.3389/fmolb.2020.00205>
PMID: [33102516](#)
 53. Geeleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One*. 2014; 9:e107468.
<https://doi.org/10.1371/journal.pone.0107468>
PMID: [25229481](#)

SUPPLEMENTARY MATERIALS

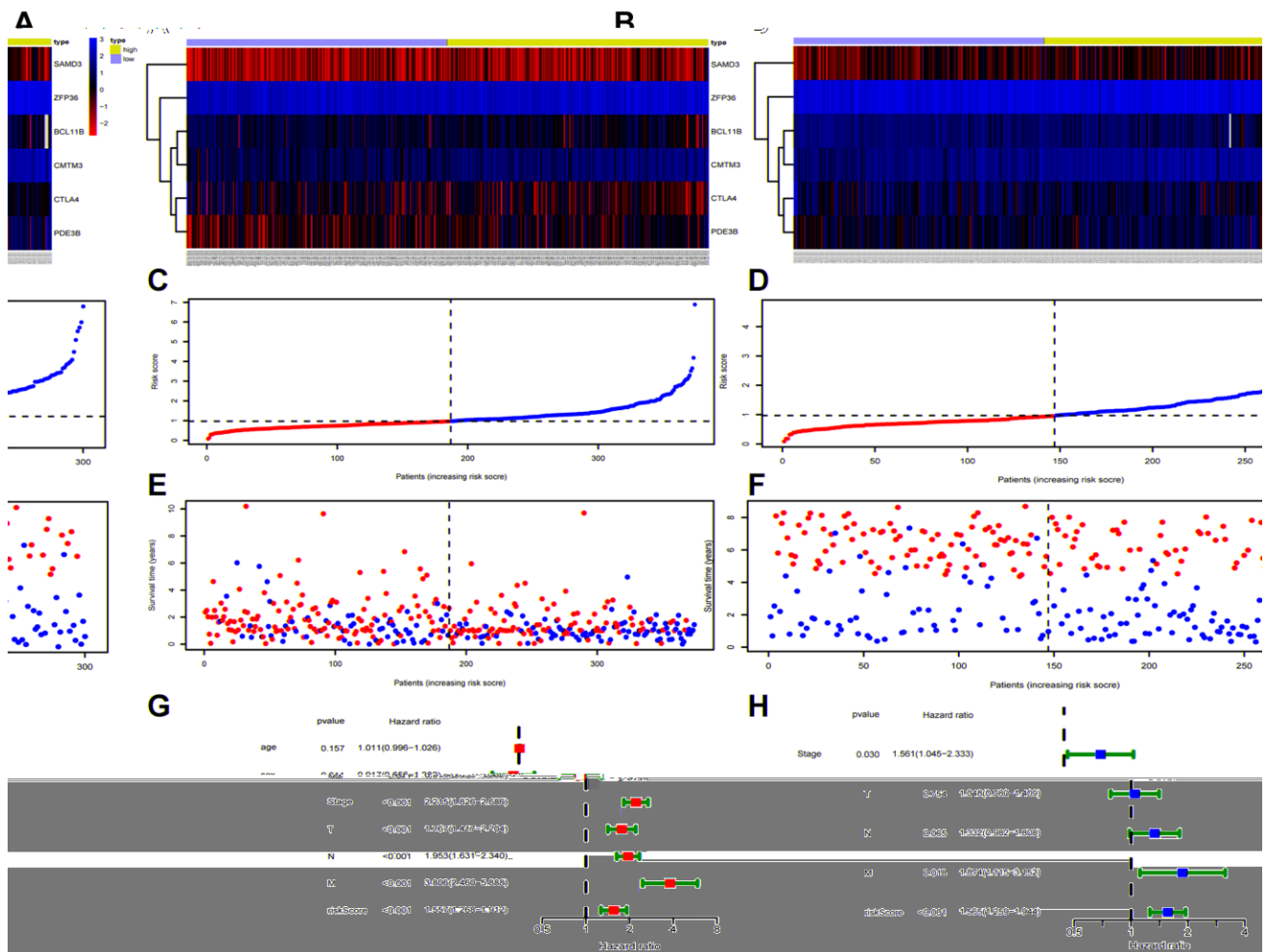
Supplementary Figures



Supplementary Figure 1. Functional enrichment for differentially expressed TCRGs. Concentric circle diagram of KEGG pathway analysis and GO analysis.



Supplementary Figure 2. K-M curves of single prognostic genes.



Supplementary Figure 3. Survival status in different groups and independent prognostic analysis in GSE62254. Heatmap for the expression characteristics of 6 modeled genes in (A) training set and (B) validation set. Distribution of risk score in (C) training set and (D) validation set. Distribution of each patient's OS in (E) training set and (F) validation set. (G) Univariate and (H) multivariate Cox analysis of clinical characteristics and risk in validation set.

Supplementary Tables

Supplementary Table 1. Clinical subgroup statistics in TCGA-STAD cohort and GSE62254 cohort.

Covariates	Type	Total	Test	Train
Age	≤65	340 (50.45%)	172 (57.33%)	168 (44.92%)
	>65	333 (49.41%)	128 (42.67%)	205 (54.81%)
	Unknown	1 (0.15%)	0 (0%)	1 (0.27%)
Gender	Female	238 (35.31%)	101 (33.67%)	137 (36.63%)
	Male	436 (64.69%)	199 (66.33%)	237 (63.37%)
Stage	Stage I	82 (12.17%)	30 (10%)	52 (13.9%)
	Stage II	210 (31.16%)	97 (32.33%)	113 (30.21%)
	Stage III	248 (36.8%)	96 (32%)	152 (40.64%)
	Stage IV	111 (16.47%)	77 (25.67%)	34 (9.09%)
	Unknow	23 (3.41%)	0 (0%)	23 (6.15%)
T	T1	21 (3.12%)	0 (0%)	21 (5.61%)
	T2	266 (39.47%)	186 (62%)	80 (21.39%)
	T3	259 (38.43%)	91 (30.33%)	168 (44.92%)
	T4	118 (17.51%)	21 (7%)	97 (25.94%)
	Unknown	10 (1.48%)	2 (0.67%)	8 (2.14%)
M	M0	609 (90.36%)	273 (91%)	336 (89.84%)
	M1	50 (7.42%)	27 (9%)	23 (6.15%)
	Unknown	15 (2.23%)	0 (0%)	15 (4.01%)
N	N0	150 (22.26%)	38 (12.67%)	112 (29.95%)
	N1	231 (34.27%)	131 (43.67%)	100 (26.74%)
	N2	153 (22.7%)	80 (26.67%)	73 (19.52%)
	N3	123 (18.25%)	51 (17%)	72 (19.25%)
	Unknown	17 (2.52%)	0 (0%)	17 (4.55%)

Supplementary Table 2. Univariate Cox analysis of differentially expressed TCRGs.

ID	HR	HR.95L	HR.95H	p-value
TSC22D3	1.247390192	1.04160897	1.49382574	0.016254411
CTLA4	0.802257011	0.656394473	0.980532802	0.031398933
PDE3B	1.312037635	1.030111424	1.671122866	0.027784017
ZFP36	1.237549084	1.003282025	1.526517667	0.046528572
LBH	1.46803763	1.177382365	1.830445697	0.000648292
PBX4	0.709013423	0.525328668	0.956924806	0.02459099
BCL11B	0.753106303	0.580945207	0.976286743	0.032258763
TAP1	0.818587069	0.688308638	0.973523725	0.023613337
TMC6	0.751788032	0.572525375	0.987179382	0.040091219
TSEN54	0.705467152	0.515530036	0.965382943	0.029249701
SAMD3	1.554138556	1.03674294	2.329744969	0.032788817
NR4A3	1.239645941	1.043627503	1.472481372	0.014436362
CMTM3	1.419781294	1.148138597	1.755693022	0.001216714
RGS2	1.337070239	1.133311226	1.577463262	0.000574322
SPATA13	0.753666261	0.573484162	0.990459492	0.042485568
ST8SIA4	1.314191148	1.013241283	1.704528233	0.039484618

Supplementary Table 3. Multivariate Cox analysis of TCRGs after LASSO analysis.

ID	coef	HR	HR.95L	HR.95H	p-value
CTLA4	−0.407822177	0.665097138	0.524167061	0.843918353	0.000788661
PDE3B	0.245124841	1.277780823	0.975172663	1.674292045	0.075462413
ZFP36	0.176830915	1.193429285	0.95400983	1.492933733	0.12165874
BCL11B	−0.363514764	0.695228463	0.527431354	0.916408575	0.009898316
SAMD3	0.472144543	1.603429132	1.011571712	2.541574611	0.044545882
CMTM3	0.362024675	1.436234381	1.152730895	1.789462923	0.0012516