

Identification of Tumor Mutation Burden, Microsatellite Instability, and Somatic Copy Number Alteration derived 9-gene Signatures to Predict Clinical Outcomes in STAD

Chuanzhi Chen

Zhejiang University

Yi Chen

Karolinska Institutet

Xin Jin

Shaoxing University Affiliated Hospital

Yongfeng Ding

Zhejiang University

Junjie Jiang

Zhejiang University

Haohao Wang

Zhejiang University

Yan Yang

Zhejiang University

Wu Lin

Zhejiang University

Xiangliu Chen

Zhejiang University

Yingying Huang

Zhejiang University

Lisong Teng (✉ lsteng@zju.edu.cn)

The First Affiliated Hospital, College of Medicine, Zhejiang University <https://orcid.org/0000-0001-6470-9017>

Primary research

Keywords: gastric cancer, immunotherapy response, predictive model, multi-omics, copy number alteration, tumor mutation burden, microsatellite instability

Posted Date: October 12th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-956254/v1>

Abstract

Background: Genomic features including tumor mutation burden (TMB), microsatellite instability (MSI) and somatic copy number alteration (SCNA), had been demonstrated to be involved with the tumor microenvironment (TME) and outcome of gastric cancer (GC).

Methods: We obtained profiles of TMB, MSI and SCNA by processing 405 GC data from The Cancer Genome Atlas (TCGA), then conducted a comprehensive analysis through “iClusterPlus”. Another independent Gene Expression Omnibus (GEO) contained specimens from 109 GC patients was designed as an external validation.

Results: Two subgroups were generated, with distinguished prognosis, somatic mutation burden, copy number changes and immune landscape. We revealed that Cluster1 was marked by a better prognosis, accompanied by higher TMB, MSIsensor score, TMEscore, and lower SCNA burden. Based on these clusters, we screened 196 differentially expressed genes (DEGs), which were subsequently projected into univariate Cox survival analysis. Thus, we constructed a 9-gene immune risk score (IRS) model using lasso penalized logistic regression. Moreover, the prognostic prediction of IRS was verified by receiver operating characteristic (ROC) curve analysis and nomogram plot.

Conclusions: Our works suggested that the 9-gene-signature prediction model, which derived from TMB, MSI, SCNA was a promising predictive tool for clinical outcome in GC patients. This novel methodology may help clinicians uncover the underlying mechanisms and guide future treatment strategies.

1. Introduction

Immune checkpoint blockade (ICB) therapy targeting programmed cell death protein 1 (PD-1) and the cytotoxic T lymphocyte antigen 4 (CTLA4) were tolerated with manageable toxicities and promising anti-tumor effect in patients with GC[1]. However, for low response rates to single agent anti-PD-1 therapy or anti-CTL4 treatment in unselected patients, single agent immunotherapy would not be an appropriate treatment of patients with operable GC[2]. Moreover, previous ICB studies have shown that response rates to immunotherapy vary widely in GC, ranging from 10–26%[3]. Thus, finding optimal biomarkers to identify potential responders to immunotherapy remains an urgent priority.

Cancer genomic characteristics had a high profile due to their key role in ICB resistance and as well as their potential in biomarker prediction. Currently, several biomarkers had been used to predict ICB responses despite some limitations. For example, tumor mutation burden (TMB) high was well-documented to contribute to therapeutic response to ICB especially in patients with melanoma and non-small cell lung cancer (NSCLC)[4, 5]. TMB high Patients had a higher chance of mobilizing the immune system to augment responding to ICB. Similarly, it was reported that microsatellite instability (MSI) high GC led to somatic mutation accumulation as well as therapy-induced immunosurveillance [6]. Besides, previous studies suggested that copy number instability score and the genome instability number, calculated based on somatic copy number alteration (SCNA) can serve as an early indicator of immune checkpoint inhibitor response versus progression[7, 8]. However, due to limitations such as measuring barriers and the absence of tumor markers, these biomarkers were rarely detected in patients with GC in practice. Early assessment of response to immunotherapy remains a current unmet clinical and scientific need to discern therapy response and tumor progression. Therefore, an

integrated approach incorporating various molecular features would be warranted to understand the unifying perspectives of the mechanisms underlying to ICB resistance, and identify subgroups of GC patients with immune microenvironments.

In the present study, integrative clustering of three genomic data sets including TMB, SCNA and MSI were used to investigate subgroups of GC through “iClusterPlus” software[9]. We further estimated the TME infiltration patterns of stomach adenocarcinoma (STAD) from TCGA and GEO data, and systematically analyzed the clusters’ relationship with genomic characteristics and clinical features in GC. We incorporated the TME infiltration evaluation into an immune risk score (IRS), to predict ICB therapeutic efficacy and survival outcomes from tumor genomic data. Depicting the immune landscape features of GC, more importantly, contribute to interpret the immunotherapy response of GC and provide new strategies for cancer treatment.

2. Material And Methods

2.1 Data source

TCGA-STAD gene expression data (n = 440), mutation annotation format (MAF) (n = 440), somatic copy number data (n = 405) and clinical data of corresponding patient (n = 444) were obtained from cBioPortal (<http://cbioportal.org/>) and UCSC Xena (<http://xena.ucsc.edu/>) website. The clinical characteristics of involved patients were displayed in Table 1. Furthermore, we conducted the transcriptome sequencing data in both in raw read counts and fragments per kilobase per million mapped reads (FPKM) values, counts data were applied to DEGs analyses, Whereas FPKM data were calculated for microsatellite instability (MSI) evaluation. The GSE26901 data from National Center for Biotechnology Information (NCBI) was derived as an independent validation cohort (n = 109). All genomic coordinates for TCGA data and GEO data in analyses of our study were based on the GRCh37 genome reference sequence[10].

Table 1
The clinicopathologic information of TCGA STAD
information

Clinicopathologic traits	Type	Patient number
Gender	Male	261
	Female	144
Age	<=60	130
	>60	275
Grade	G1	12
	G2	143
	G3	241
	Gx	9
StageT	T1	21
	T2	86
	T3	179
	T4	111
	Tx	8
StageN	N1	121
	N2	108
	N3	77
	N4	81
	Nx	18
StageM	M0	366
	M1	23
	Mx	16
Stage	I	58
	II	128
	III	174
	IV	39
	X	6

2.2 SCNA data acquisition and processing

The peak regions of recurrent DNA copy number alteration including amplification and deletion were delineated by algorithm GISTIC2[11]. Subsequently, we converted copy number alteration into binary form and defined them as SCNA genomic features. We categorized SCNA events, as reported, according to each patient's aberration status of GISTIC results: -2, homozygous loss; -1, hemizygous deletion; 0, diploid; 1, low-level gain; 2, high-level amplification. High-level amplifications and homozygous loss in peak region were defined as copy number change, with at least 50% of genes displaying an amplification or deletion[2]. To obtain binary description matrix about SCNA feature, we assigned feature changes as 1, and no feature changes as 0.

SCNA scores was calculated using masked copy number segment profiles from UCSC Xena platform, which is defined as the ratio of copy number alteration (tumor/normal) and normalized by fragment length after log2 transformation.

2.3 Modified TMB (mTMB) data acquisition and processing

We defined mTMB as the total number of unique genes with mutations in each patient. Only 7 types of mutations in this gene were considered as mTMB event: Frame_Shift_Del, Translation_Start_Site, Frame_Shift_Ins, Splice_Site, Nonstop_Mutation, Nonsense_Mutation and Missense_Mutation. After removing no functional relevance mutations, we merged the MAF data of TCGA-STAD. Then the low-frequency mutated genes were filtered through a cut-off value (a certain gene mutation occurred in 1% of the total number of samples). As a result, we extracted 1932 high-frequency mutated genes in 405 patients and binary description matrix of mTMB feature was used with subsequent calls.

TMB burden was computed by the total number of somatic mutations per Mb in each sample. Since 38 Mb is usually taken in terms of the length of human exons, the TMB burden was equal to the total mutation frequency /38[12].

2.4 MSI data acquisition and processing

MSIsensor-pro algorithm of Linux operating system was used to investigate the MSI traits of TCGA-STAD data at microsatellite transcription level[13]. We selected and calculated the most frequently altered microsatellite sites to construct the binary MSI feature based on the somatic mutation status of each sample. Then we computed MSIsensor score under the default parameters by sample matrix. We further distinguished MSI high (MSIsensor score ≥ 10) samples from MSI low or microsatellite stability (MSS) (MSIsensor score < 10) samples according to previous research[14, 15]. Finally, 1 represents MSI and 0 represents MSI low/MSS to obtain a binary matrix of MSI event.

2.5 Genomics variation data integration

We constructed a comprehensive data of 2,024 genome variant characteristics, including 54 copy number gains, 37 copy number losses, 1 MSI and 1,932 genes. Subsequently, we characterized the SCNA, mTMB and MSI traits in a binary form to delineate whether corresponding genomic alteration occurred in each patient. In detail, 1 indicated the presence of genomic changes while 0 indicated the absence of genomic changes in this data, which formed the sample matrix with three binary signatures.

2.6 Clustering and survival analysis

In our genomic variation profiles description matrix, the columns represent various samples while rows represented the corresponding genomic signatures. In total, 405 valid samples were classified by “iClusterPlus”, a comprehensive clustering method in R package[16]. With default parameters, different numbers of categories are cycled ($k = 1-5$). Finally, the optimal classification result was calculated with the highest percent of explained variation and best Bayesian information criteria (BIC), that is, $k=1$ and Cluster =2 (Figure. S1A). And the final Cluster membership was shown on table S1. We selected the top quartile features based on lasso coefficient estimates (prob = 0.75), Thus, only values greater than the upper quartile were considered to contribute significantly to the classification.

2.7 Mutation and copy-number aberrations (CNAs) analysis

Mutation analysis was performed both in Cluster1 and Cluster2 based on “maftools” package in R. The default arguments were set to analyze the MAF of the TCGA-STAD data set in each cluster. The mutation results were directly visualized by “oncoplot” function of “maftools” R package. The analysis of CNAs was performed with GISTIC2 on Linux system. The specific “-conf” parameter was set as 0.95 and top 10 significant copy number alteration area were displayed in both clusters with “gisticOncoPlot” function. G-score across all chromosomes was visualized based on the frequency of the CNAs and the average amplitude in log ratio.

2.8 Immune cell infiltration analysis

The cancer immune infiltration profiles in the two clusters were compared based on the gene expression data of TCGA-STAD. The proportion of various immune cells proportion of each sample were computed through the “Cibersort” software under the default parameters. Differences of the immune cell landscape between Cluster1 and Cluster2 were investigated by “boxplot”. Additionally, the well-established “TMEscore” was used to analyze the difference in immune efficacy between the two clusters[17].

2.9 Differentially gene expression analysis

Based on the TCGA-STAD gene expression data (counts), we used the “DESeq2” package in R to screen out the DEGs between the two clusters. The significance threshold for DEGs was set as $\text{abs}(\log_2\text{FC}) > 1$ and $p < 0.05$.

2.10 Immune risk score model construction

The normalized expression data of the DEGs were subsequently converted into binary fashion by comparing the median value of each gene in all samples. After combining with clinical data, DEGs was further selected in univariate cox analysis using “coxph” function of the “survival” package in R. With default parameters and significance ($p < 0.05$), we carried out the hub genes as independent prognostic factors. The “glmnet” package was used to perform lasso penalized regression on the samples and corresponding DEGs. The arguments used in lasso penalized regression are $\alpha = 1$, $n\lambda = 100$, and $p < 0.05$ was considered as the significance threshold. By “coef” function, we got the Y-intercept and hazard rate (HR) score of each gene. Then, HR score, Y-intercept and corresponding hub genes expression profiles were used to measure the IRS value of all sample.

2.11 Statistical Analysis

The unpaired Student's t-test was developed to estimate the comparison between two normally distributed variable. In contrast, non-normally distributed variables were measured by Wilcoxon rank-sum test. In order to compare more than two groups, Kruskal–Wallis tests and ANOVA were used as the non-parametric method and

the parametric method, respectively. Two-sided Fisher's exact tests were used to analyze contingency data. The Kaplan-Meier method was used to offer a visual representation of predicted survival curves for each clusters data with "ggplot2" and "survminer" packages. Area under the curve (AUC), the sensitivity and specificity were depicted by "pROC" package. All statistical analyses in this study were performed on R version 4.0.4 (<https://www.r-project.org/>), and p-value < 0.01 indicated a statistically significant threshold.

3. Results

3.1 Comprehensive genomic variation traits to identify two GC classifications

Integrated design workflow in our study was shown in figure 1. According to the SCNA patterns, MAF data and MSI signature of the TCGA-STAD, we meticulously characterized the three genomic status. After integrating the three genomic status (a total of 2,024), all the 405 STAD samples were divided into Cluster1 and Cluster2 based on the "iClusterPlus" (figure 2A and figure S1A). Cluster membership of all the samples were shown on table S1. It was worth noting that Cluster1 had a higher proportion of older patients (>60 years) than Cluster2. In contrast, Cluster2 was the main cluster of deaths cases, and more likely to gather higher level stageM, stage and grade on STAD, indicating that the traits of genomic change were distinctly associated with the cancer malignancy. In terms of gene mutations, the samples with more gene mutations were obviously clustered in Cluster1. Additionally, the heatmap of MSI (top panel), SCNA (middle panel), and mutation (bottom panel) for the three-cluster solution was shown in figure S1B. To further compare the clinical value of our clusters, we performed Kaplan-Meier (KM) plots to explore the outcomes of the two subgroups. Interestingly and noteworthy, we found that patients in the Cluster1 showed significantly longer overall survival (OS), disease free survival (DFS) compared to Cluster2 (figure 2B and 2C). Subsequently, we used "scatterplot3d" package to visualize MSIsensor score, SCNA burden and TMB burden across the samples, the results demonstrated that Cluster1 and Cluster2 can be well distinguished (figure 2D). Take home message of the number of samples is that the percentages of Cluster1 and Cluster2 cases is 21% and 79% respectively (Figure 2E).

3.2 Different signatures of MSI, SCNA and mTMB in two clusters

Next, the relevant quantitative indicators of three genomic characteristics including MSI, SCNA and TMB burden were analyzed between the two clusters. In terms of the SCNA burden, Cluster1 was relatively lower than Cluster2 (figure 3A). The significant differences in TMB burden and MSI burden were also shown between Cluster1 and Cluster2. Specifically, Cluster1 harbored markedly higher TMB burden and MSI burden than Cluster2 in TCGA-STAD (figure 3B and 3C). Our results suggested that Cluster1 tend to be associated with better prognosis in GC patient, with the underlying assumption that SCNA low, TMB high and MSI were more likely to benefit from ICB, which was consistent with previous reports[18, 19]. To assess the relative level of immune infiltration in the subgroups, as a commonly accepted quantifiable indicator[17, 20–22], TMEscore was conducted to compare the difference between the two groups. Intriguingly, Cluster1 were observed significantly higher TMEscore than Cluster2, which represented more abundant cancer infiltrating immune cells (figure 3D).

3.3 Differences in somatic mutations and CNAs between two clusters

To elucidate the differences between two subgroups, the somatic mutation landscapes in waterfall plot were displayed (figure 4A and 4B). Our work clarified the total mutation load in each sample and sorted the top 20 genes by mutation frequency in each subgroup. Notably, as the Y-axis represented, the mutation load of Cluster1 was distinctly higher than that of Cluster2, in accordance with the results in figure 3B. Moreover, only 7 genes in the top 20 mutant genes were shared between the two subgroups: TTN, MUC16, LRP1B, SYNE1, FAT4, PCLO, ARID1A. However, these shared genes in Cluster1 were more likely to have “multi hits mutation” instead of “missen mutation” compare to Cluster2. Meanwhile, the CNAs landscapes of the two Clusters were shown in chromosomal alterations via G-score (figure 4C and 4D). Interestingly, Cluster2 has a higher copy number variation frequency on chromosomes 6, 7 and 8 than Cluster1. On the whole, the copy number region between the two clusters demonstrated that Cluster2 had a higher genome-wide amplification and deletion than Cluster1, consistent with the results in figure 3A. In addition, heat maps were used to identify the top 10 CNAs in each cluster, in which the mainly changes of Cluster1 were deletion (green), while Cluster2 were amplification (red) (figure S2A). And the sample proportion of CNAs in Cluster1 (23% ~ 42%) is significantly lower than that in Cluster2 (57% ~ 77%). Overall, these results indicated that mutations and CNAs on both clusters were significantly different.

3.4 TME landscape of two GC subgroups

To investigate the TME between the two subgroups, Cibersort analysis was carried out to display the abundance of 22 immune cells types (figure 5A). Besides, activation status and enumeration of the 10 immune cells were significantly discrepant, as shown in figure 5B. The ratio of eosinophils, macrophages M1, neutrophils, NK cells active, NK cells resting, t cells CD4 memory active, T cells CD8 was markedly higher in Cluster1 than in Cluster2. On the other side, compared with Cluster1, the infiltration of B cells naive, mast cells resting, t cells resting were more extensive in Cluster2. Moreover, we selected seven immune cell types which were high infiltration in more than half of the samples (Figure S2B). Noteworthy, similar results were observed in Cluster2, with B cells naive and mast cells resting increased. The rejection and immune dysfunction level were further validated by TMEscore (figure 3D). The above results suggested Cluster1 was more likely to trigger antitumor immunity rather than Cluster2.

3.5 Construction of the IRS model

To further identify the differences in the transcription level between the two clusters of STAD, a total of 196 DEGs were detected by “DESeq2” package (Table 2, figure 6A). Under the cutoff of $p < 0.05$ and $|\log_2FC| > 1$, we screen 127 upregulated genes and 69 downregulated genes in Cluster2. Then, univariate cox analysis was conducted to investigate the 196 DEGs.

Finally, 10 hub genes were found to be closely correlated with OS (table 3; $p < 0.05$), in which SNORA12 were revealed to be protective factors, while TF, SLC13A5, ASF1B, RSPO4, GRIN3B, ANXA8, PPBP, ALOX15 were risk factors (figure 6B). Subsequently, we performed lasso-penalized multivariate cox modeling across 100 simulation replications, and constructed an optimal model with 9 coefficients, i.e. IRS model (figure 6B, 6C and 6E). The IRS formula used for each sample was shown in table 4. The predicted value of the model was

compared with the actual event in a boxplot (Figure S3). Then, the gene expression for the 9 genes across the all the samples was divided into high expression and low expression by the median value, and corresponding survival differences were shown on KM plots (figure 7). These 9 genes were significantly correlated with OS in STAD, indicating the reliability of our IRS model.

Table 2

The DEGs between two genomic variation clusters (top30, Cluster 1 vs. Cluster 2).

gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	change
CALB1	55.7570737	3.20704641	0.3745911	8.56145924	1.11E-17	1.41E-13	UP
FABP4	112.55467	-2.2424426	0.26338141	-8.5140503	1.68E-17	1.41E-13	DOWN
SLC6A10P	163.583951	2.65889326	0.32140159	8.27280672	1.31E-16	7.30E-13	UP
LEP	7.57683434	-1.9914675	0.25052301	-7.9492398	1.88E-15	7.86E-12	DOWN
CD300LG	5.88931295	-2.4530037	0.30967674	-7.9211752	2.35E-15	7.88E-12	DOWN
PLIN1	31.0236722	-2.0838749	0.26551582	-7.8484018	4.21E-15	1.18E-11	DOWN
WISP2	30.2571827	-1.7930524	0.25611821	-7.0008783	2.54E-12	6.09E-09	DOWN
AVPR2	12.0243545	-1.3938919	0.20080671	-6.9414605	3.88E-12	8.12E-09	DOWN
ANXA8	92.398049	2.04381023	0.32305158	6.32657553	2.51E-10	4.66E-07	UP
CHGB	154.368293	2.14618156	0.3413288	6.28772477	3.22E-10	5.40E-07	UP
TUSC5	11.0175881	-2.6739849	0.42933709	-6.2281713	4.72E-10	7.19E-07	DOWN
CAPN14	12.2466455	-1.9103772	0.30843019	-6.1938723	5.87E-10	8.19E-07	DOWN
COL11A2	66.8601521	1.69388149	0.27408034	6.18023708	6.40E-10	8.25E-07	UP
TGM1	44.3421124	1.57217566	0.25667972	6.1250483	9.07E-10	1.02E-06	UP
PPP2R2C	132.681633	1.86737944	0.30492448	6.12407195	9.12E-10	1.02E-06	UP
RHOXF2B	6.86820936	2.31485118	0.3786162	6.11397818	9.72E-10	1.02E-06	UP
KIF1A	104.350953	2.3710194	0.39752231	5.96449389	2.45E-09	2.42E-06	UP
THRSP	4.23481843	-1.8631954	0.31305487	-5.9516577	2.65E-09	2.47E-06	DOWN
NOTUM	427.531768	2.19037069	0.37084952	5.90635974	3.50E-09	3.08E-06	UP
GTSF1	42.5431785	2.08174551	0.35426946	5.87616421	4.20E-09	3.36E-06	UP

HAS1	19.2654239	-1.5712805	0.26742691	-5.8755515	4.21E-09	3.36E-06	DOWN
IL13	1.06481915	-1.4998397	0.26323412	-5.6977406	1.21E-08	9.24E-06	DOWN
3-Sep	88.6650647	1.14291951	0.20131746	5.67720026	1.37E-08	9.97E-06	UP
EEF1A2	244.421268	1.91885771	0.34016481	5.64096473	1.69E-08	1.13E-05	UP
SCG3	33.0768799	1.67738648	0.29932366	5.6039221	2.10E-08	1.35E-05	UP
PLA2G1B	11.3407016	2.28869956	0.40996224	5.58270826	2.37E-08	1.47E-05	UP
COL9A3	243.425904	1.55281499	0.27895116	5.56661956	2.60E-08	1.55E-05	UP
S100A2	244.680208	1.43386147	0.25980591	5.51897178	3.41E-08	1.97E-05	UP
EMID2	7.50106954	1.70137763	0.31126108	5.4660789	4.60E-08	2.57E-05	UP
RNF182	21.1625865	1.5567599	0.28657754	5.43224679	5.56E-08	3.00E-05	UP

Table 3

The univariate Cox results of significant DEGs.

Characteristics	HR	lower	upper	combine	p.value
TF	2.2	1.4	3.5	2.2 [1.4-3.5]	0.0011
SLC13A5	1.9	1.2	3.1	1.9 [1.2-3.1]	0.0057
ASF1B	2	1.2	3.1	2 [1.2-3.1]	0.0049
RSPO4	1.7	1.1	2.7	1.7 [1.1-2.7]	0.021
GRIN3B	1.6	1	2.6	1.6 [1-2.6]	0.033
ANXA8	1.9	1.2	3	1.9 [1.2-3]	0.0076
PPBP	1.6	1	2.6	1.6 [1-2.6]	0.04
SNORA12	0.56	0.35	0.9	0.56 [0.35-0.9]	0.017
ALOX15	1.7	1.1	2.7	1.7 [1.1-2.7]	0.023

Table 4. The IRS formula for each sample was calculated as follows:

$$\begin{aligned} \text{Immune Risk Score} = & 0.049 + 0.071 * \text{TF (Exp)} + 0.034 * \text{SLC13A5 (Exp)} + 0.074 * \text{ASF1B (Exp)} \\ & + 0.013 * \text{RSP04 (Exp)} + 0.019 * \text{GRIN3B (Exp)} + 0.066 * \text{ANXA8 (Exp)} + 0.059 * \text{PPBP (Exp)} \\ & + 0.0288 * \text{NPY (Exp)} - 0.109 * \text{SNORA12 (Exp)} + 0.056 * \text{ALOX15 (Exp)} \end{aligned}$$

3.6 Evaluating the discriminatory power of IRS model and external validation

Samples were classified into high IRS and low IRS groups based on the median value of IRS. A comprehensive heatmap was developed to display the distribution of clinical characteristics in TCGA-STAD data (figure 8A). Interestingly, the occurrence of cancer related death was comparatively enriched in the high IRS group. Besides, Cluster2 and advanced stage was gathered in the high IRS group. Furthermore, the expression level of ASF1B, SNORA12, RSP04, TF were visibly different in between high and low IRS group. To further evaluate the predictive ability of IRS, the significant differences in OS between high and low IRS groups was shown in figure 8B ($p < 0.0001$). AUC was computed to test the discriminatory powers over 1-year, 3-year, and 5-year outcome respectively (figure 8C), suggesting a promising prognostic predictive value in our training dataset. In order to measure the value of the IRS model in immunotherapy, the IRS grouping result were compared with TMEscore, and the low IRS group show a relatively high TMEscore (figure 8D). Notably, in the validation cohort (GSE26901), our IRS model suggested a distinct difference between the high and low IRS groups in clinical outcomes as well (figure 8E). And the AUC values in 1-year, 3-year, and 5-year were close to 70% (figure 8F).

3.7 Assessing predictive values and stability on IRS model

The indicative clinicopathological features of the samples, including age, gender, grade, stageM, stageN and stageT were conducted to test the stability and efficiency of ISR model. Our results indicated that the ISR had significant differences between high IRS and low IRS samples in all clinical characteristics except in high stageM (figure S4). In addition, we revealed the profile of 22 immune cells infiltration between high IRS and low IRS samples. The proportion of dendritic cells resting and Macrophages M1 were significantly higher in low IRS patients (figure S5), consisted with previous work. We subsequently generated a nomogram calibration plot to combine the clinical factors as well as Risk (IRS) to measure the clinical benefits (figure 9A). Moreover, the decision curve of this prognostic nomogram and the IRS prediction model were displayed in figure 9B. The nomogram-based 1-year, 3-year, and 5-year OS predictions for GC patients with IRS exhibited superior accuracy (figure 9C, 9D and 9E).

4. Discussion

ICB antibodies had revolutionized the therapeutic landscape in patients with various cancers[23-28], including advanced GC[29]. Notably, PD-L1 combined positive score (CPS) was widely approved as a predictive biomarker indicated efficacy of ICB in GC[30, 31]. However, these therapeutic responses occurred only in a minority of GC patients, while most GC patients were primary ICB resistance[32, 33]. Previous studies supported the idea that the clinical benefit with ICB in GC was independent of CPS positivity[29]. Thus, the combination of immunotherapy with chemotherapy and angiogenesis inhibitor had been encountering the dilemmas of lacking precise biomarkers. Extensive researches had proved the predictive ability of SCNA, mTMB and MSI to therapeutic response or resistance[4, 5, 34]. However, as predictive biomarkers individually, each one of these genomic traits is not stable enough to accurately reflect GC heterogeneity. Here, we comprehensively integrated these ICB-related genomic signatures, i.e. mTMB, MSI and SCNA to explore pertinent underlying mechanisms.

In our study, with a fully Bayesian latent variable model, we stratified TCGA-STAD into two distinct tumor subtypes according to SCNA, mTMB and MSI. Intriguingly, each subtype was correlated with a special immune profile highlighting its multidimensional relationship between intrinsic genetic characteristics and TME[35, 36]. Currently, TMEscore had been used by many researchers to predict treatment efficacy to ICB, as well as to investigate the immune suppressive mechanisms mediated by TME[37, 38]. Based on the two GC cohort, we revealed our clustering is robust in predicting OS, DFS and TMEscore (figure 3). A simple combination of SCNA, mTMB and MSI or through known benchmarking driver genes was not able to reinforce our understanding of the interplay between the cancer genomic landscape and the host specific antitumor immune response[39]. The advantage of “iClusterPlus” was its sufficient dimension reduction, with unsupervised clustering across all data types, provided the most accurate classification in clinical tumor subtypes and revealed driver omics features[40, 41]. In addition, the distribution of latent variables is more stable, since it was automatically generated by its conditional distribution of visible variables[42]. Despite the lack of user-friendliness, this approach greatly met the needs in precision medicine and help clinicians diagnose and customize treatments.

To further investigate the differences of the immune microenvironments in the two distinct genomic clusters, Cibersort was performed to assess the infiltration of 22 immune cells. It is well-established that the polarization of the macrophages to the M1 phenotype could kill the cancer cells and suppress their growth[43, 44]. On the other hand, eosinophils had been implicated as antitumor effector cells, whose tumoricidal function was mediated by TNF- α , granzymes and IL-18[45, 46]. Moreover, neutrophils, NK cells and T cells had been reported as central communicators in antitumor immunity[47-49]. Consistent with our clustering, Cluster1 tend to aggregate these immune cells activate the immune microenvironment and had a high potential for response to ICB. In order to explore the gene expression patterns of Cluster1, we screened the DEGs between Cluster1 and Cluster2, and selected the prognostic core markers to construct the prediction model. Inspiringly, our IRS model showed that patients with high IRS had a poorer prognosis and a lower proportion of macrophage M1 infiltration (figure S4 and S5). More importantly, we further using KM plot, AUC, nomogram and decision curve analysis to validate the predictive value of IRS in calculating the OS probability of GC patients. Merits of our IRS model was primarily attributed to the precise identification of TME activation based on 9-gene, particularly in predominant infiltration of M1 macrophages tumors.

Among these 9 key genes, several genes had been reported to be involved in carcinogenesis and tumor progression. For example, SLC13A5 was a sodium-coupled transporter which was proved to facilitate hepatic energy homeostasis, influence proliferation of hepatocarcinoma and resist chemotherapeutic agents in

hepatocarcinoma cells[50, 51]. RSPO4 was a member of R-spondin family. As WNT signaling activation had been found to overexpress in breast cancer, particularly in triple-negative breast cancer, the role of RSPO4 involved in GC progression remained unelucidated[52, 53]. On a similar note, ANXA8 had been revealed to be upregulated in various cancers[54-56]. The feedback loop between RA-RARA and ANXA8 fostered cancer initiation and progression[57]. More importantly, the expression levels of ASF1B was reported to be associated with TME in STAD[57]. From a mechanistic point of view, ASF1B indirectly regulate CKS1B to mediate growth, apoptosis and cell cycle progression in cancers[58].

However, due to TME complexity and tumor heterogeneity, not all patients with high IRS would benefit from immunotherapy. This research was limited by the validity of exon level transcriptomic data from immunotherapy patients. Hence, further work was needed to validate our findings in prospective cohort of GC patients receiving ICB. In the foreseen future, with the increasing availability of large-scale detection applied to GC patients treated with ICB, a systematic exploration of TME would unveil the mechanisms underlying of resistance to immunotherapy.

5. Conclusions

In summary, we comprehensively analyzed three genetic features associated with immune microenvironment, subsequently identified two distinct clusters in GC. We delineated the characteristics of both subgroups from prognosis, mutation burden, copy number changes and immune profile. Identifying their DEGs followed by screening survival-related genes, 9 hub DEGs were finally selected for downstream analysis. We proposed a 9-gene IRS that serve as a biomarker in clinical application, whose predictive value was further validated in an independent GC cohort (GSE26901). Therefore, developed a nomogram predicting the probability of a patient will survive GC for 1, 3, and 5 years. Our works provided a new approach to accelerate accurate immunotherapy, which may optimize combination therapy strategies.

Declarations

Author Contributions

Chen C and Chen Y conceived the study design. Chen C and Jin X performed RNA-seq analyses. Chen Y, Ding Y, Jiang J and Wang H performed classification analysis. Chen C drafted the manuscript. Yang Y, Wu Lin, Chen X and Huang Y participated in revision of the manuscript. Teng L was responsible for study supervision. All authors contributed to this paper and approved the final manuscript.

Funding

This research was supported by grants from the Science and Technology Project of Zhejiang Province (2021C03119), the Project of the regional diagnosis and treatment center of the Health Planning Committee (No. JBZX-201903) and Traditional Chinese Medicine Science and Technology Plan in Zhejiang Province (2020ZZ013).

Ethics approval and consent to participate

Not applicable.

Consent for publication

All contributing authors agree to publish this manuscript.

Availability of data and materials

The datasets used and analysed in this study are available from the corresponding author on reasonable request.

Competing interests

The authors declare no conflicts of interest.

Acknowledgments

We greatly thank all the staffs in the Department of Surgical Oncology, the First Affiliated Hospital, Zhejiang University School of Medicine for their support in this study, as well as the TCGA and GEO databases for the availability of the data.

References

1. Aggelis V, Cunningham D, Lordick F, Smyth EC. Peri-operative therapy for operable gastroesophageal adenocarcinoma: past, present and future. *Ann Oncol*. 2018;29(6):1377–85.
2. Xie F, Zhang J, Wang J, Reuben A, Xu W, Yi X, Varn FS, Ye Y, Cheng J, Yu M, et al. Multifactorial Deep Learning Reveals Pan-Cancer Genomic Tumor Clusters with Distinct Immunogenomic Landscape and Response to Immunotherapy. *Clin Cancer Res*. 2020;26(12):2908–20.
3. Zeng D, Wu J, Luo H, Li Y, Xiao J, Peng J, Ye Z, Zhou R, Yu Y, Wang G, et al: **Tumor microenvironment evaluation promotes precise checkpoint immunotherapy of advanced gastric cancer**. *J Immunother Cancer* 2021, 9(8).
4. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, Lee W, Yuan J, Wong P, Ho TS, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science*. 2015;348(6230):124–8.
5. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, Sucker A, Hillen U, Foppen MHG, Goldinger SM, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science*. 2015;350(6257):207–11.

6. van Velzen MJM, Derks S, van Grieken NCT, Haj Mohammad N, van Laarhoven HWM. MSI as a predictive factor for treatment outcome of gastroesophageal adenocarcinoma. *Cancer Treat Rev.* 2020;86:102024.
7. Jensen TJ, Goodman AM, Kato S, Ellison CK, Daniels GA, Kim L, Nakashe P, McCarthy E, Mazloom AR, McLennan G, et al. Genome-Wide Sequencing of Cell-Free DNA Identifies Copy-Number Alterations That Can Be Used for Monitoring Response to Immunotherapy in Cancer Patients. *Mol Cancer Ther.* 2019;18(2):448–58.
8. Weiss GJ, Beck J, Braun DP, Bornemann-Kolatzki K, Barilla H, Cubello R, Quan W Jr, Sangal A, Khemka V, Waypa J, et al. Tumor Cell-Free DNA Copy Number Instability Predicts Therapeutic Response to Immunotherapy. *Clin Cancer Res.* 2017;23(17):5074–81.
9. Mo Q, Wang S, Seshan VE, Olshen AB, Schultz N, Sander C, Powers RS, Ladanyi M, Shen R. Pattern discovery and cancer gene identification in integrated cancer genomic data. *Proc Natl Acad Sci U S A.* 2013;110(11):4245–50.
10. Jensen MA, Ferretti V, Grossman RL, Staudt LM. The NCI Genomic Data Commons as an engine for precision medicine. *Blood.* 2017;130(4):453–9.
11. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhi R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 2011;12(4):R41.
12. Schumacher TN, Kesmir C, van Buuren MM. Biomarkers in cancer immunotherapy. *Cancer Cell.* 2015;27(1):12–4.
13. Jia P, Yang X, Guo L, Liu B, Lin J, Liang H, Sun J, Zhang C, Ye K. MSIsensor-pro: Fast, Accurate, and Matched-normal-sample-free Detection of Microsatellite Instability. *Genomics Proteomics Bioinformatics.* 2020;18(1):65–71.
14. Niu B, Ye K, Zhang Q, Lu C, Xie M, McLellan MD, Wendl MC, Ding L. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics.* 2014;30(7):1015–6.
15. Abida W, Cheng ML, Armenia J, Middha S, Autio KA, Vargas HA, Rathkopf D, Morris MJ, Danila DC, Slovin SF, et al. Analysis of the Prevalence of Microsatellite Instability in Prostate Cancer and Response to Immune Checkpoint Blockade. *JAMA Oncol.* 2019;5(4):471–8.
16. Mo Q, Nikolos F, Chen F, Tramel Z, Lee YC, Hayashi K, Xiao J, Shen J, Chan KS. Prognostic Power of a Tumor Differentiation Gene Signature for Bladder Urothelial Carcinomas. *J Natl Cancer Inst.* 2018;110(5):448–59.
17. Zeng D, Li M, Zhou R, Zhang J, Sun H, Shi M, Bin J, Liao Y, Rao J, Liao W. Tumor Microenvironment Characterization in Gastric Cancer Identifies Prognostic and Immunotherapeutically Relevant Gene Signatures. *Cancer Immunol Res.* 2019;7(5):737–50.
18. Liu X, Yao J, Song L, Zhang S, Huang T, Li Y. Local and abscopal responses in advanced intrahepatic cholangiocarcinoma with low TMB, MSS, pMMR and negative PD-L1 expression following combined therapy of SBRT with PD-1 blockade. *J Immunother Cancer.* 2019;7(1):204.
19. Lu M, Zhang P, Zhang Y, Li Z, Gong J, Li J, Li J, Li Y, Zhang X, Lu Z, et al. Efficacy, Safety, and Biomarkers of Toripalimab in Patients with Recurrent or Metastatic Neuroendocrine Neoplasms: A Multiple-Center Phase Ib Trial. *Clin Cancer Res.* 2020;26(10):2337–45.

20. Huang J, Li J, Zheng S, Lu Z, Che Y, Mao S, Lei Y, Zang R, Liu C, Wang X, et al. Tumor microenvironment characterization identifies two lung adenocarcinoma subtypes with specific immune and metabolic state. *Cancer Sci.* 2020;111(6):1876–86.
21. Jiang S, Zhang Y, Zhang X, Lu B, Sun P, Wu Q, Ding X, Huang J. GARP Correlates With Tumor-Infiltrating T-Cells and Predicts the Outcome of Gastric Cancer. *Front Immunol.* 2021;12:660397.
22. Zhang H, Yang M, Wu X, Li Q, Li X, Zhao Y, Du F, Chen Y, Wu Z, Xiao Z, et al. The distinct roles of exosomes in tumor-stroma crosstalk within gastric tumor microenvironment. *Pharmacol Res.* 2021;171:105785.
23. Nishino M, Ramaiya NH, Hatabu H, Hodi FS. Monitoring immune-checkpoint blockade: response evaluation and biomarker development. *Nat Rev Clin Oncol.* 2017;14(11):655–68.
24. Rosenberg JE, Hoffman-Censits J, Powles T, van der Heijden MS, Balar AV, Necchi A, Dawson N, O'Donnell PH, Balmanoukian A, Loriot Y, et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet.* 2016;387(10031):1909–20.
25. Motzer RJ, Escudier B, McDermott DF, George S, Hammers HJ, Srinivas S, Tykodi SS, Sosman JA, Procopio G, Plimack ER, et al. Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. *N Engl J Med.* 2015;373(19):1803–13.
26. Hamanishi J, Mandai M, Ikeda T, Minami M, Kawaguchi A, Murayama T, Kanai M, Mori Y, Matsumoto S, Chikuma S, et al. Safety and Antitumor Activity of Anti-PD-1 Antibody, Nivolumab, in Patients With Platinum-Resistant Ovarian Cancer. *J Clin Oncol.* 2015;33(34):4015–22.
27. Ferris RL, Blumenschein G Jr, Fayette J, Guigay J, Colevas AD, Licitra L, Harrington K, Kasper S, Vokes EE, Even C, et al. Nivolumab for Recurrent Squamous-Cell Carcinoma of the Head and Neck. *N Engl J Med.* 2016;375(19):1856–67.
28. Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, Chow LQ, Vokes EE, Felip E, Holgado E, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. *N Engl J Med.* 2015;373(17):1627–39.
29. Kang YK, Boku N, Satoh T, Ryu MH, Chao Y, Kato K, Chung HC, Chen JS, Muro K, Kang WK, et al. Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet.* 2017;390(10111):2461–71.
30. Kim ST, Cristescu R, Bass AJ, Kim KM, Odegaard JI, Kim K, Liu XQ, Sher X, Jung H, Lee M, et al. Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. *Nat Med.* 2018;24(9):1449–58.
31. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, Kadel EE III, Koeppen H, Astarita JL, Cubas R, et al. TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature.* 2018;554(7693):544–8.
32. Fuchs CS, Doi T, Jang RW, Muro K, Satoh T, Machado M, Sun W, Jalal SI, Shah MA, Metges JP, et al. Safety and Efficacy of Pembrolizumab Monotherapy in Patients With Previously Treated Advanced Gastric and Gastroesophageal Junction Cancer: Phase 2 Clinical KEYNOTE-059 Trial. *JAMA Oncol.* 2018;4(5):e180013.
33. Roh W, Chen PL, Reuben A, Spencer CN, Prieto PA, Miller JP, Gopalakrishnan V, Wang F, Cooper ZA, Reddy SM, et al: **Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals**

- markers of response and resistance.** *Sci Transl Med* 2017, 9(379).
34. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, Lu S, Kemberling H, Wilt C, Luber BS, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science*. 2017;357(6349):409–13.
 35. Ivey JW, Bonakdar M, Kanitkar A, Davalos RV, Verbridge SS. Improving cancer therapies by targeting the physical and chemical hallmarks of the tumor microenvironment. *Cancer Lett*. 2016;380(1):330–9.
 36. Krishnan V, Chong YL, Tan TZ, Kulkarni M, Bin Rahmat MB, Tay LS, Sankar H, Jokhun DS, Ganesan A, Chuang LSH, et al. TGFbeta Promotes Genomic Instability after Loss of RUNX3. *Cancer Res*. 2018;78(1):88–102.
 37. Qiu Z, Jiang H, Ju K, Liu X. A Novel RNA-Binding Protein Signature to Predict Clinical Outcomes and Guide Clinical Therapy in Gastric Cancer. *Front Med (Lausanne)*. 2021;8:670141.
 38. Wei S, Lu J, Lou J, Shi C, Mo S, Shao Y, Ni J, Zhang W, Cheng X. Gastric Cancer Tumor Microenvironment Characterization Reveals Stromal-Related Gene Signatures Associated With Macrophage Infiltration. *Front Genet*. 2020;11:663.
 39. Zhao X, Wu S, Fang N, Sun X, Fan J. Evaluation of single-cell classifiers for single-cell RNA sequencing data sets. *Brief Bioinform*. 2020;21(5):1581–95.
 40. Mo Q, Shen R, Guo C, Vannucci M, Chan KS, Hilsenbeck SG. A fully Bayesian latent variable model for integrative clustering analysis of multi-type omics data. *Biostatistics*. 2018;19(1):71–86.
 41. Abu-Jamous B, Buffa FM, Harris AL, Nandi AK. In vitro downregulated hypoxia transcriptome is associated with poor prognosis in breast cancer. *Mol Cancer*. 2017;16(1):105.
 42. Menyhart O, Gyorffy B. Multi-omics approaches in cancer research with applications in tumor subtyping, prognosis, and diagnosis. *Comput Struct Biotechnol J*. 2021;19:949–60.
 43. Menga A, Serra M, Todisco S, Riera-Domingo C, Ammarah U, Ehling M, Palmieri EM, Di Noia MA, Gissi R, Favia M, et al. Glufosinate constrains synchronous and metachronous metastasis by promoting anti-tumor macrophages. *EMBO Mol Med*. 2020;12(10):e11210.
 44. Rao L, Wu L, Liu Z, Tian R, Yu G, Zhou Z, Yang K, Xiong HG, Zhang A, Yu GT, et al. Hybrid cellular membrane nanovesicles amplify macrophage immune responses against cancer recurrence and metastasis. *Nat Commun*. 2020;11(1):4909.
 45. Reichman H, Karo-Atar D, Munitz A. Emerging Roles for Eosinophils in the Tumor Microenvironment. *Trends Cancer*. 2016;2(11):664–75.
 46. Varricchi G, Galdiero MR, Loffredo S, Lucarini V, Marone G, Mattei F, Marone G, Schiavoni G. **Eosinophils: The unsung heroes in cancer?** *Oncoimmunology* 2018, 7(2):e1393134.
 47. Chen H, Jiang T, Lin F, Guan H, Zheng J, Liu Q, Du B, Huang Y, Lin X. PD-1 inhibitor combined with apatinib modulate the tumor microenvironment and potentiate anti-tumor effect in mice bearing gastric cancer. *Int Immunopharmacol*. 2021;99:107929.
 48. Fabian KP, Padgett MR, Donahue RN, Solocinski K, Robbins Y, Allen CT, Lee JH, Rabizadeh S, Soon-Shiong P, Schlom J, et al: **PD-L1 targeting high-affinity NK (t-haNK) cells induce direct antitumor effects and target suppressive MDSC populations.** *J Immunother Cancer* 2020, 8(1).

49. Munir H, Jones JO, Janowitz T, Hoffmann M, Euler M, Martins CP, Welsh SJ, Shields JD. Stromal-driven and Amyloid beta-dependent induction of neutrophil extracellular traps modulates tumor growth. *Nat Commun.* 2021;12(1):683.
50. Li Z, Li D, Choi EY, Lapidus R, Zhang L, Huang SM, Shapiro P, Wang H. Silencing of solute carrier family 13 member 5 disrupts energy homeostasis and inhibits proliferation of human hepatocarcinoma cells. *J Biol Chem.* 2017;292(33):13890–901.
51. Hu T, Huang W, Li Z, Kane MA, Zhang L, Huang SM, Wang H. Comparative proteomic analysis of SLC13A5 knockdown reveals elevated ketogenesis and enhanced cellular toxic response to chemotherapeutic agents in HepG2 cells. *Toxicol Appl Pharmacol.* 2020;402:115117.
52. Coussy F, Lallemand F, Vacher S, Schnitzler A, Chemlali W, Caly M, Nicolas A, Richon S, Meseure D, El Botty R, et al. Clinical value of R-spondins in triple-negative and metaplastic breast cancers. *Br J Cancer.* 2017;116(12):1595–603.
53. Park S, Cui J, Yu W, Wu L, Carmon KS, Liu QJ. Differential activities and mechanisms of the four R-spondins in potentiating Wnt/beta-catenin signaling. *J Biol Chem.* 2018;293(25):9759–69.
54. Yuan JB, Gu L, Chen L, Yin Y, Fan BY. Annexin A8 regulated by lncRNA-TUG1/miR-140-3p axis promotes bladder cancer progression and metastasis. *Mol Ther Oncolytics.* 2021;22:36–51.
55. Ma F, Li X, Fang H, Jin Y, Sun Q, Li X. Prognostic Value of ANXA8 in Gastric Carcinoma. *J Cancer.* 2020;11(12):3551–8.
56. Gou R, Zhu L, Zheng M, Guo Q, Hu Y, Li X, Liu J, Lin B. Annexin A8 can serve as potential prognostic biomarker and therapeutic target for ovarian cancer: based on the comprehensive analysis of Annexins. *J Transl Med.* 2019;17(1):275.
57. Rossetti S, Sacchi N. **Emerging Cancer Epigenetic Mechanisms Regulated by All-Trans Retinoic Acid.** *Cancers (Basel)* 2020, 12(8).
58. Zhang W, Gao Z, Guan M, Liu N, Meng F, Wang G. ASF1B Promotes Oncogenesis in Lung Adenocarcinoma and Other Cancer Types. *Front Oncol.* 2021;11:731547.

Figures

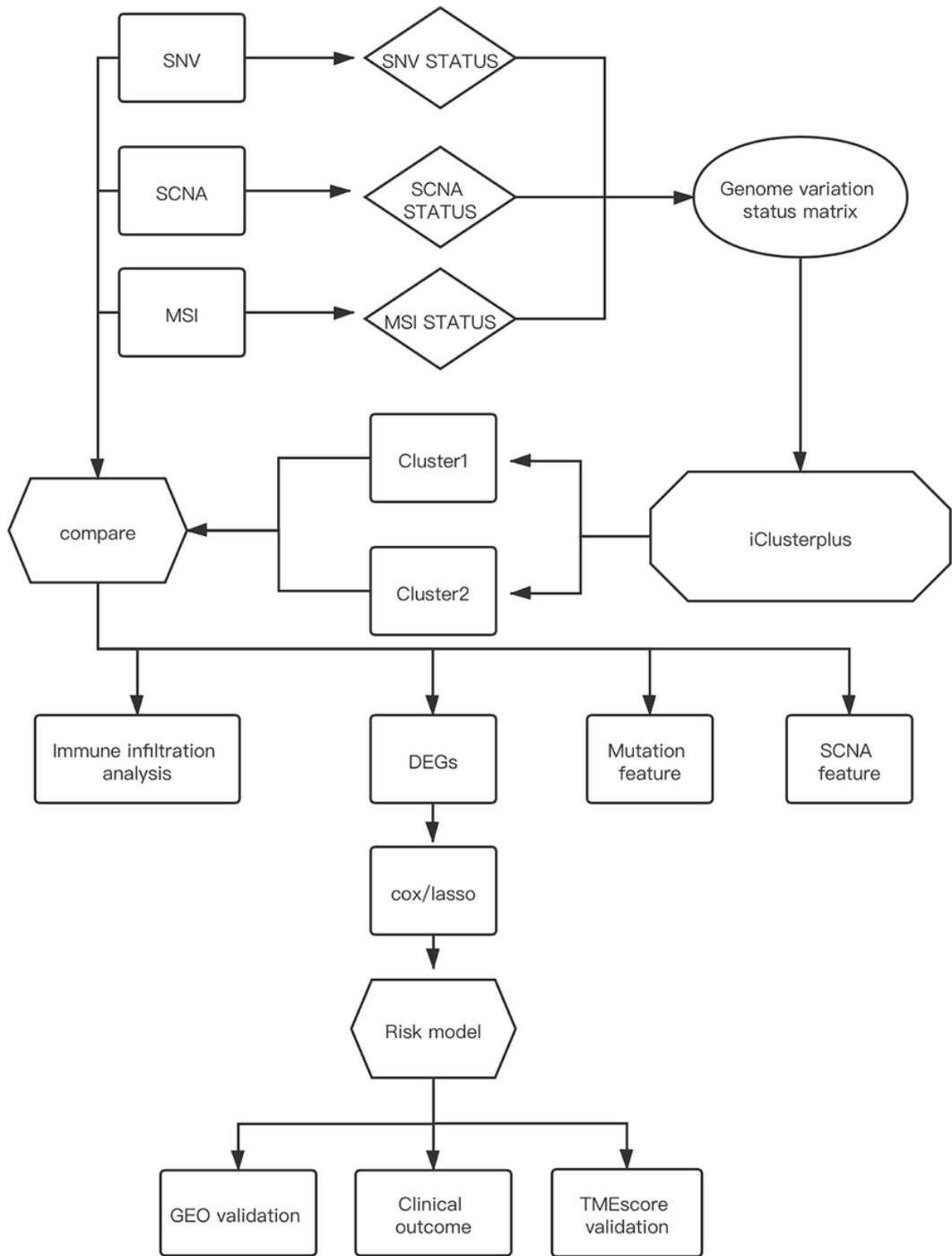


Figure 1

Comprehensive workflow of our study for constructing Risk model in GC, including downloading and processing and analysis.



Figure 2

The differences in clinicopathologic traits and genomic features between two clusters. (A) The heatmap showed clusters of 500 genomic mutation profiles. Sample annotations show the clinicopathologic

characteristics. (B) Kaplan-Meier plots showed the OS in patients with Cluster1 and Cluster2 ($p < 0.05$). (C) Kaplan-Meier plots showed the DFS in patients with Cluster1 and Cluster2 ($p < 0.05$). (D) A 3D plot showed the difference in samples between two clusters. (E) The sample percentage of the two clusters.

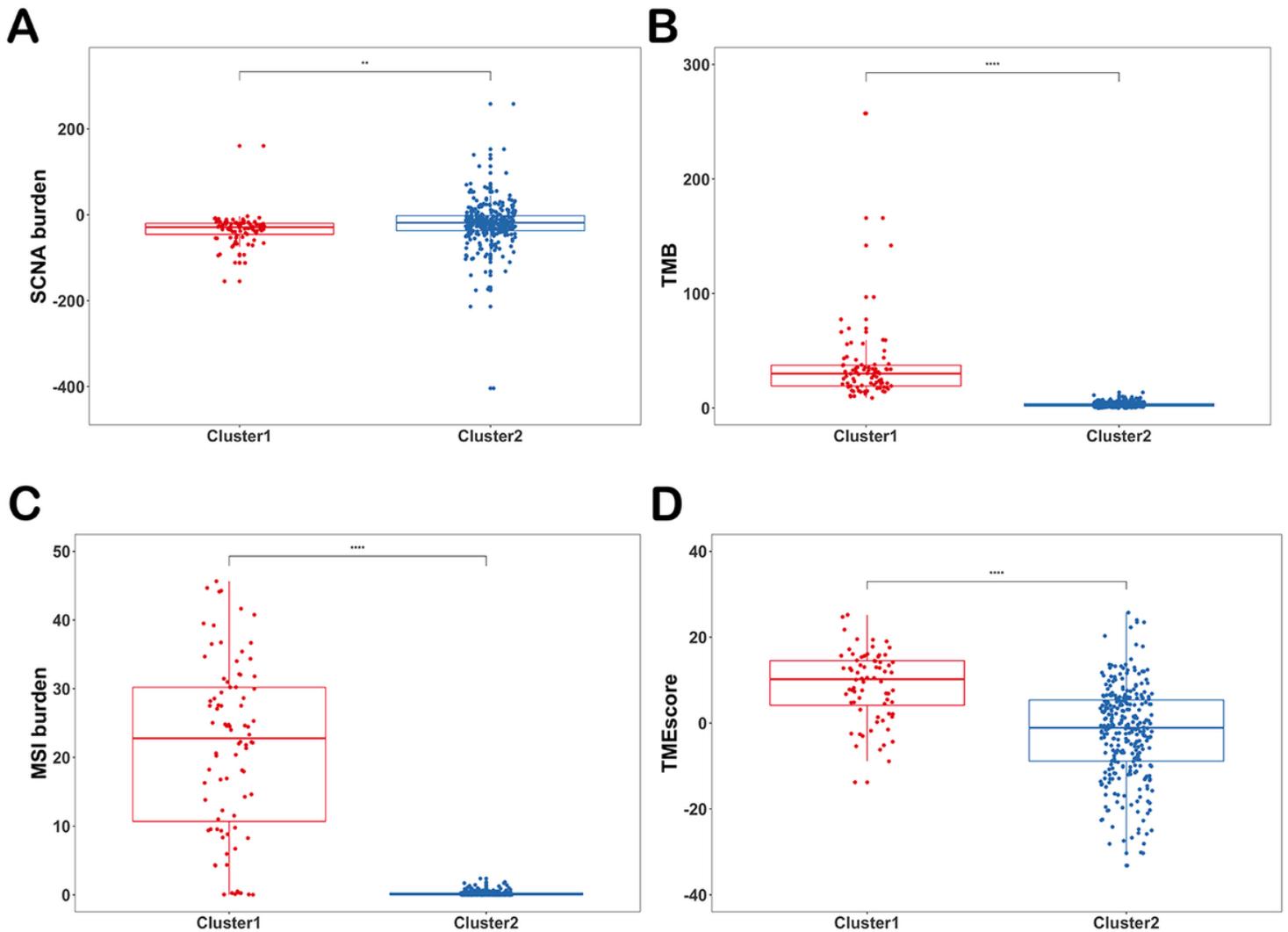


Figure 3

The differences in genomic variations and immune measurement between the two subgroups. (A) The box plots showed the SCNA burden between two clusters $**p < 0.01$. (B) The box plots showed the TMB between two clusters $****p < 0.0001$. (C) The box plots showed the MSI burden between two clusters $****p < 0.0001$. (D) The box plots showed the TMEScore between two clusters $****p < 0.0001$.

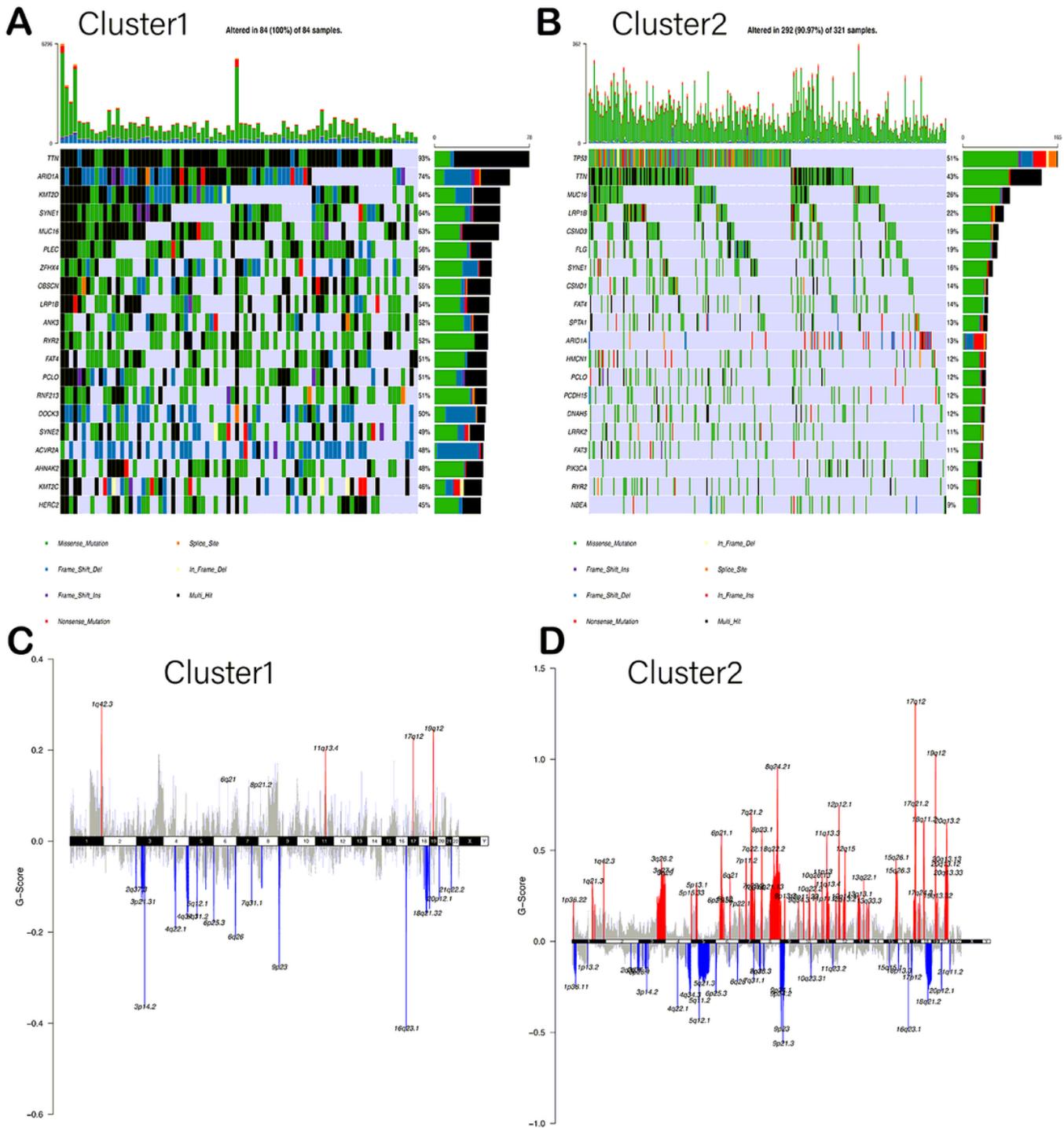


Figure 4

The differences in somatic mutation and copy number alteration between the two subgroups (A, B) The oncoplot indicated the top 25 mutated gene in both Cluster1 and Cluster2, respectively. Different colors represent different mutation types. (C, D) Cumulative CNA regions for Cluster1 and Cluster2, respectively. Amplification were represented in red color, and deletion was represented in blue color.

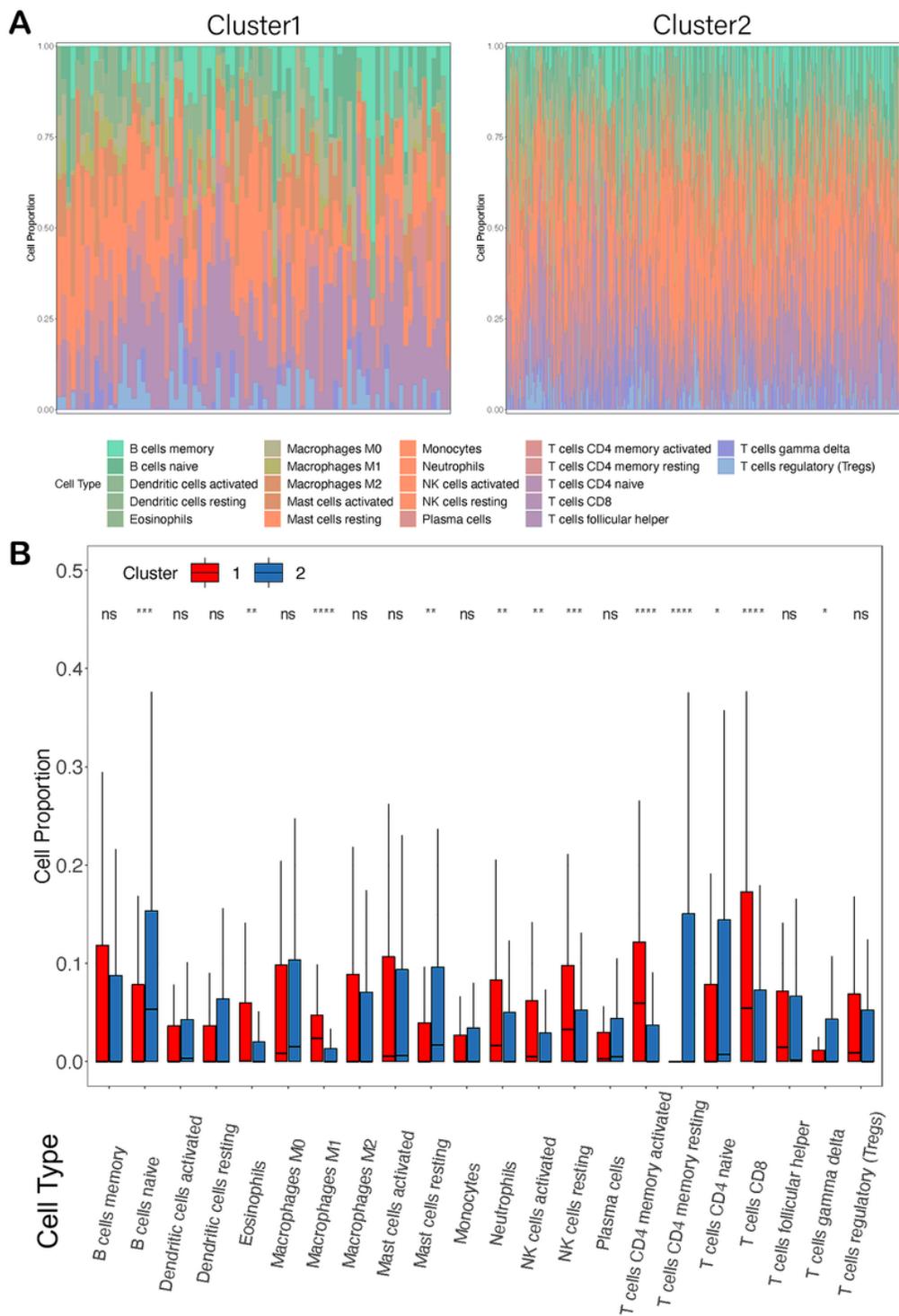


Figure 5

The differences in immune cell infiltrations between the two clusters. (A) The distribution of 22 immune cell subtypes infiltration between the two clusters in each patient. Different colors represent different immune cells types. (B) The box plots showed the infiltration of immune cells proportion between Cluster1 and Cluster2. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

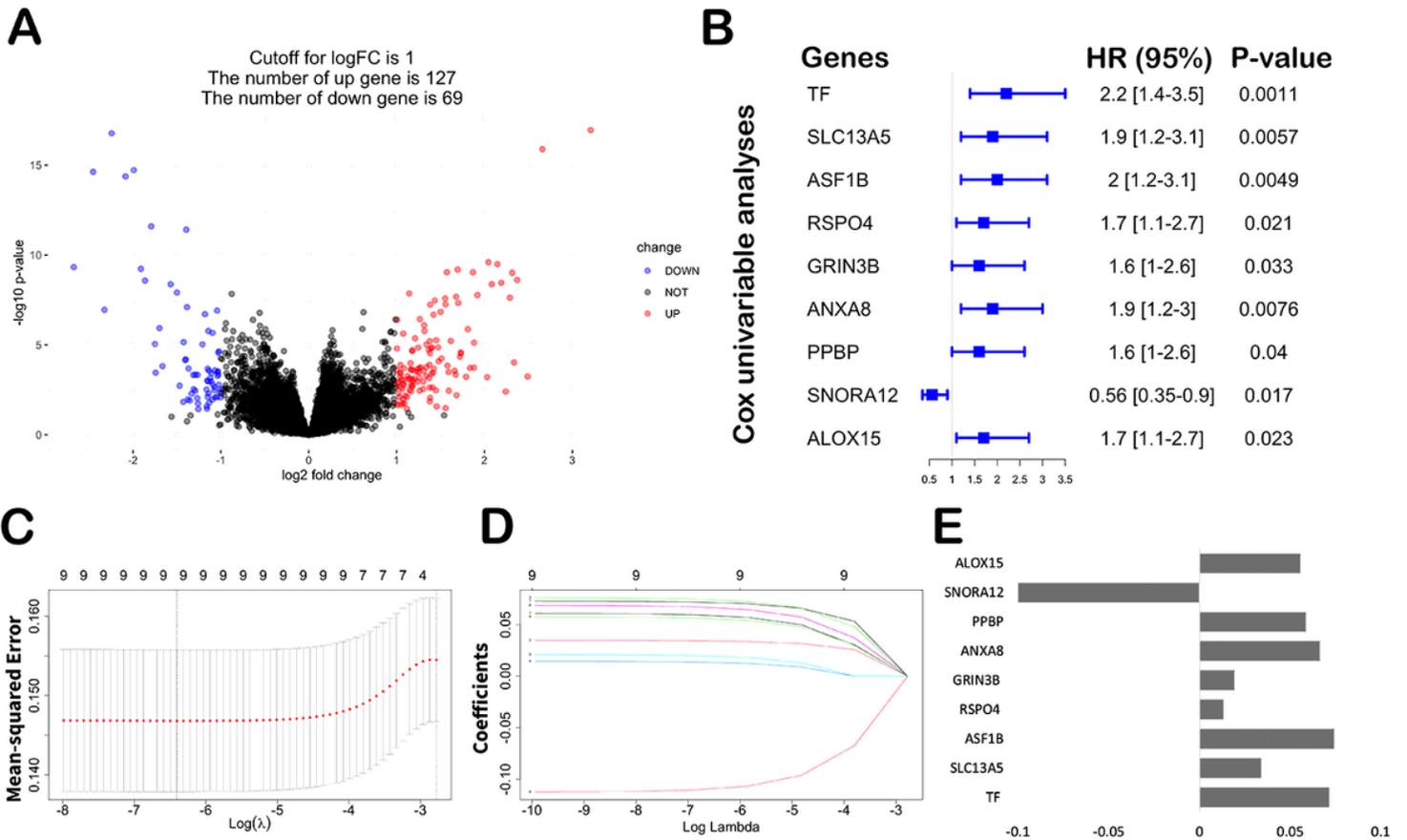


Figure 6

Construction of the Immune Risk Score model. (A) The volcano plot indicated the association between log₂ fold change and P-value in DEGs. $abs(\log_2FC) > 1, p < 0.05$ (B) Univariate analysis of 9 significant genes with OS ($p < 0.05$). (C) The correlation between $\log(\lambda)$ and the mean-squared error in the LASSO Cox regression model. (D) The correlation between $\log(\lambda)$ and the coefficients of relevant in the LASSO Cox regression model. (E) The coefficients of each independent variable in the LASSO Cox regression model.

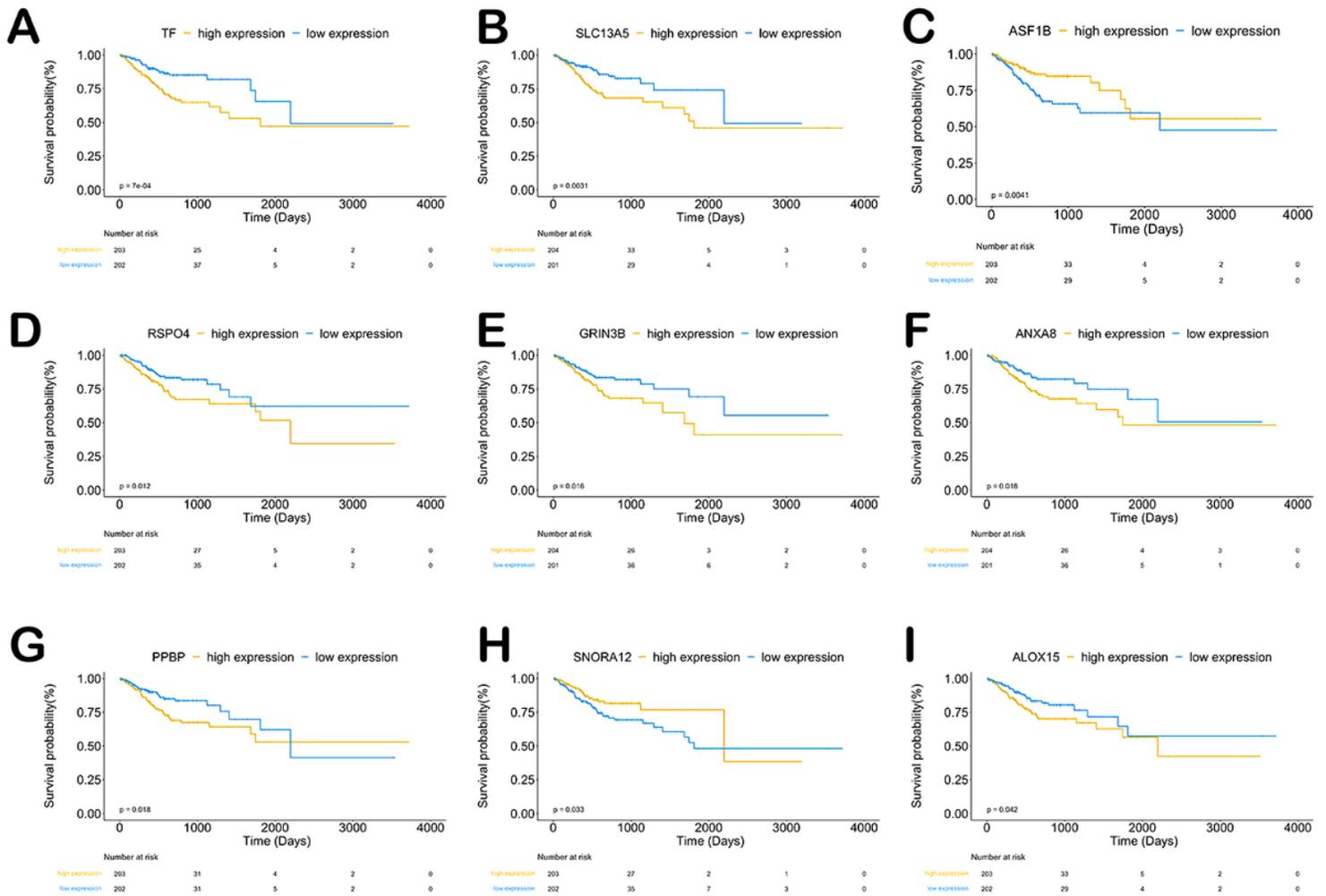


Figure 7

The KM plots showed OS in patients with high and low expression of nine significant genes, based on the median value. $p < 0.05$

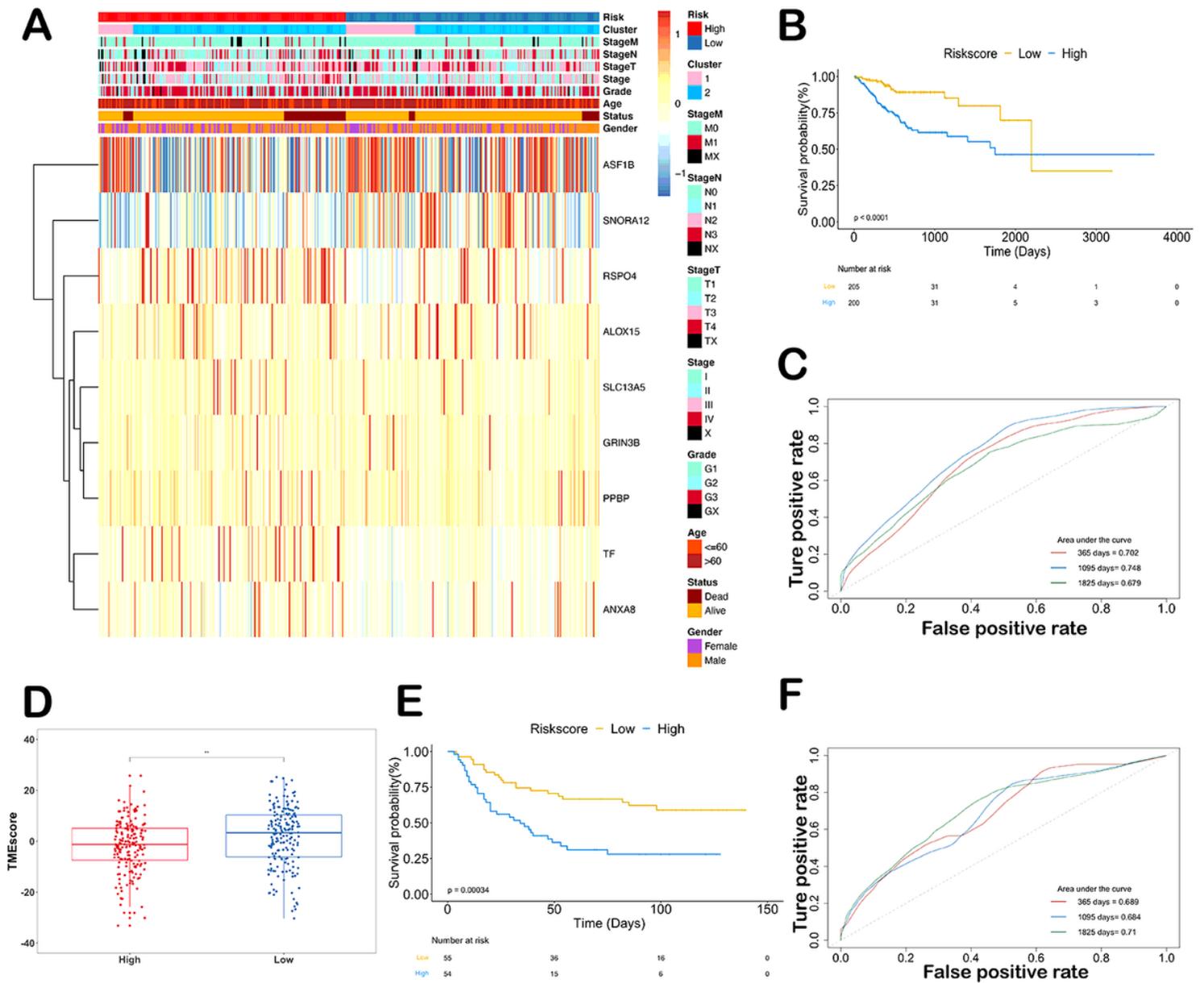


Figure 8

The predictive ability of ISR model in TCGA and GEO database. (A) The expression of 9-gene and clinicopathologic features in each patient (TCGA-STAD). The heatmap showed 9-gene expression profiles in high IRS and low IRS, based on the median value. Sample annotations show the clinicopathologic characteristics and clusters. (B) KM plots showed the OS in training group with high IRS and low IRS ($p < 0.0001$). (C) The prognostic values of IRS in training group in 1-, 3- and 5-year OS with AUC = 0.702, 0.748 and 0.679, respectively. (D) The box plots showed the difference of TMEScore between high IRS and low IRS. (E) KM plots showed the OS in validation group with high IRS and low IRS ($p < 0.001$). (F) The prognostic values of IRS in validation group in 1-, 3- and 5-year OS with AUC = 0.689, 0.684 and 0.71, respectively.

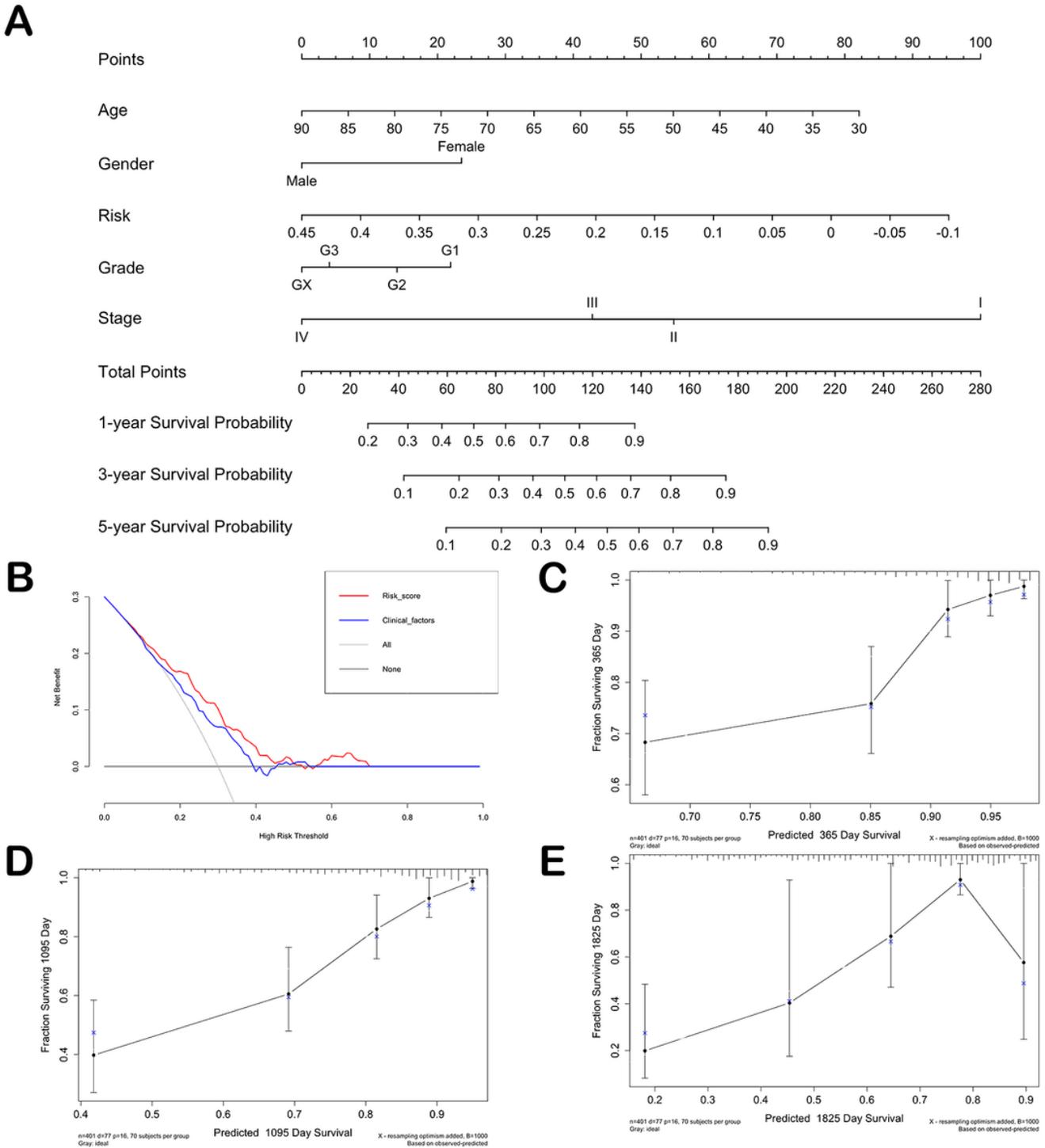


Figure 9

The nomogram of IRS and its stability evaluation. (A) Nomogram for predicting survival probability of GC patients in the training group. The total score of clinicopathologic features as well as risk score for each sample is located on the "Total points" axis, which corresponds to the survival probabilities on the 1-, 3-, and 5-year axes. (B) Decision curve for nomogram and clinical factors. (C) Calibration curves for nomogram in 1-year, 3-year, and 5-year, respectively.

Supplementary Files

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

- [figureS1.tif](#)
- [figureS2.tif](#)
- [figureS3.tif](#)
- [figureS4.tif](#)
- [figureS5.tif](#)
- [tableS1.xlsx](#)