Research article

Molecular subtype construction and prognosis model for stomach adenocarcinoma characterized by metabolism-related genes

Jie Sun [a], Yuanyuan Wang [b], Kai Zhang [c], Sijia Shi [d], Xinxin Gao [e], Xianghao Jia [f], Bicong Cong [e], Chunning Zheng [f,*]

[a] Department of Gastrointestinal Surgery, Shandong Provincial Third Hospital, Jinan, 250031, China
[b] Department of Oncology and Hematology, Second Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan, 250001, China
[c] General Surgery Department, Wenshang County People's Hospital, Wenshang, 272501, China
[d] Shandong Provincial Hospital, Jinan, 250001, China
[e] Gastrointestinal Surgery, Shandong First Medical University Affiliated Provincial Hospital, Jinan, 250001, China
[f] Gastrointestinal Surgery, Shandong Provincial Hospital, Jinan, 250001, China

ARTICLE INFO

ABSTRACT

Background: Metabolic reprogramming is implicated in cancer progression. However, the impact of metabolism-associated genes in stomach adenocarcinomas (STAD) has not been thoroughly reviewed. Herein, we characterized metabolic transcription-correlated STAD subtypes and evaluated a metabolic RiskScore for evaluation survival.
Method: Genes related to metabolism were gathered from previous study and metabolic subtypes were screened using ConsensusClusterPlus in TCGA-STAD and GSE66229 dataset. The ssGSEA, MCP-Count, ESTIMATE and CIBERSORT determined the immune infiltration. A RiskScore model was established using the WGCNA and LASSO Cox regression in the TCGA-STAD queue and verified in the GSE66229 datasets. RT-qPCR was employed to measure the mRNA expressions of genes in the model.
Result: Two metabolism-related subtypes (C1 and C2) of STAD were constructed on account of the expression profiles of 113 prognostic metabolism genes with different immune outcomes and apparently distinct metabolic characteristic. The overall survival (OS) of C2 subtype was shorter than that of C1 subtype. Four metabolism-associated genes in turquoise model, which closely associated with C2 subtype, were employed to build the RiskScore (MATN3, OSBPL1A, SERPINE1, CPNE8) in TCGA-train dataset. Patients developed a poorer prognosis if they had a high RiskScore than having a low RiskScore. The promising effect of RiskScore was verified in the TCGA-test, TCGA-STAD and GSE66229 datasets. The prediction reliability of the RiskScore was validated by time-dependent receiver operating characteristic curve (ROC) and nomogram. Moreover, samples with high RiskScore had an enhanced immune status and TIDE score. Moreover, MATN3, OSBPL1A, SERPINE1 and CPNE8 mRNA levels were all elevated in SGC7901 cells. Inhibition of OSBPL1A decreased SGC7901 cells invasion numbers.
Conclusion: This work provided a new perspective into heterogeneity in metabolism and its association with immune escape in STAD. RiskScore was considered to be a strong prognostic label that could help individualize the treatment of STAD patients.

---

* Corresponding author.
  E-mail address: yinzx123@163.com (C. Zheng).

1. Introduction

Gastric cancer as a frequent gastrointestinal malignancy showed over 1 million new cases in 2020, accounting for about 5.6%, and the number of deaths is about 769,000. Accounting for about 7.7% of the population, it ranks fifth in morbidity and fourth in mortality [1–3]. Gastric cancer includes Stomach adenocarcinoma (STAD), hereditary diffuse gastric cancer, gastrointestinal stromal tumor, gastric lymphoma, neuroendocrine tumor and other rare variants, among which STAD accounts for 95% of all gastric cancer cases [4]. STAD is a malignant tumor originating from the stomach and a heterogeneous disease with various histological features and genotypes, and its occurrence and development are influenced by lifestyle and biological factors. High molecular heterogeneity results in variations in relapse and death risks [5]. Therefore, to precisely identify the risk of gastric cancer, novel prognostic factors are critically required.

The metabolic flexibility of cancer cells depends on their ability to reprogram anabolism and catabolism, with metabolic changes associated with cancer no longer seen as an indirect response to cell proliferation and survival signals [6,7]. Rapidly proliferating cancer cells are reprogrammed by autonomic metabolism to promote growth and survival, and new therapeutic strategies have begun to focus on the unique metabolic patterns observed in cancer cells. Cancer cells could promote the expression levels of glycolytic enzymes and the activity to promote aerobic glycolysis. One instance is 2DG, which suppresses HK2 by binding to it, reducing glycolysis, and leading to ROS-mediated apoptosis in various types of cancer [8,9].

Comparative studies focusing on tumors and adjacent normal tissues have revealed that dysregulated transcription of metabolic genes is widely present. Despite the important insights these studies provide into modified metabolic cellular pathways in carcinoma, tumors and healthy tissues often contain diverse cell components that could restrict the clinical applicability of these results to some extent [10]. Previous study [11] performed screened a number of metabolic pathways differentially expressed in differentiated samples based on clinical outcomes, which revealed considerable metabolic heterogeneity. A prognostic signature on the basis of metabolic gene has also been raised in head and neck cancer [12,13] and neuroblastoma [14].

This work employed The Cancer Genome Atlas (TCGA)-STAD queue to classify STAD cases on account of metabolism-related genes. We then compared prognostic and immune landscapes and pathway enrichment among various subtypes. In addition, we sought to construct RiskScore and validate them. To analyze the responses of patients showing different risks to immunotherapy, the correlation between clinical immune profiles and RiskScore was further explored.

2. Material and methods

2.1. Raw data

From TCGA database (https://portal.gdc.cancer.gov/), we collected the clinical features and expression profiles of STAD patients (including 350 STAD and 32 normal samples). Transcripts per kilobase million (TPM) was transformed by counts fragments (FPKM). The GSE66229 dataset [15] containing 300 cancer cases were acquired from Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/).

2.2. Molecular subtypes of metabolism-correlated genes

Under criteria |log2fold change (FC)|>1.5 and false discovery rate (FDR) < 0.05, differentially expressed genes (DEGs) between STAD samples and healthy samples were screened using limma package [16]. We have 2752 genes involved in metabolism that encode all known human transporter proteins and metabolic enzymes from a previous study [17]. For metabolism-related genes, genes showing expression >1 in at least half of the TCGA-STAD samples were retained and Univariate cox analysis was used for screening metabolic genes associated survival with p-value <0.05. Metabolic-related differentially expressed genes (MRDEGs) were identified by intersecting the above DEGs and prognostic metabolic genes. The k-means unsupervised clustering was conducted on MRDEGs by the "ConsensusClusterPlus" package [18] with algorithm and measure of distance spearman as well as 500 bootstraps. The k was between 2 and 10. Optimal number of categories was classified in terms of the consistency matrix and consistency cumulative distribution function.

2.3. Evaluation of immune infiltration and activity of metabolism-related pathways

SsGSEA [19] was applied on 113 metabolism-related pathways and 28 immune cells in "GSVA" package [19]. The infiltration of immune and stromal cells (ImmuneScore, StromalScore, and ESTIMATEScore) was calculated by ESTIMATE algorithm [20]. The 22 immune cell scores and 10 cell populations were analyzed using CIBERSORT [21] tool and MCP-Count method [22], respectively.

2.4. Pathway enrichment analyses

Statistical analysis of KEGG and GO was realized using the "clusterProfiler" package [23]. Moreover, GSEA was conducted based on gene set (h.all.v7.5.1.symbols.gmt).

2.5. Analysis of mutation characteristics

Molecular characteristic information (Fraction Altered, TMB, Aneuploidy, Number of Segments) in TCGA-STAD were obtained from a published pan-cancer study [24]. In the somatic mutation data processed by mutect2 software, Genes with more than 3 mutation frequencies were screened out. Fisher test screened genes showing significantly high frequency mutation in the subtypes with p < 0.05.

2.6. WGCNA

The "WGCNA" analysis package was introduced to develop a co-expression network for the genes [25]. A soft threshold is determined by analyzing the network topology. According to topological overlap matrix (TOM) converted from adjacency, clustering the genes was performed applying the average chain hierarchy clustering method. In a gene network module, the minimum gene number was 50 on the basis of the criteria of hybrid dynamic shearing tree. The eigengenes were analyzed after defining gene modules with dynamic shear method, followed by cluster analysis of the modules. Modules in relatively close distance (height = 0.25, deepSplit = 2, and minModuleSize = 50) with each other were combined into a new module. Then module correlation and molecular subtypes was analyzed to select the key gene modules for further analysis.

2.7. Construction of prognosis model

Firstly, training and test sets of samples from the TCGA-STAD were classified at a ratio of 7:3. Clinical feature differences between the training set and the test set groups were analyzed by Chi-square test. Except T.stage, significant differences were found in clinical feature groups ($p > 0.05$), showing a random and reasonable grouping (Table 1). In TCGA-train dataset, gene related to patients' survival of $p < 0.01$ in WGCNA analysis were identified by performing univariate Cox regression analysis. Then, LASSO was used to reduce metabolic genes associated with survival using glmnet package, with optimal parameter λ selected using ten-fold cross-

Table 1
clinical features of TCGA-STAD dataset.

	TCGA-STAD				
Characteristics	Train (N = 245)	Test (N = 105)	Total (N = 350)	pvalue	FDR
Age 1				0.65	1
>67	113 (32.29%)	54 (15.43%)	167 (47.71%)		
≤67	130 (37.14%)	50 (14.29%)	180 (51.43%)		
Unknow	2 (0.57%)	1 (0.29%)	3 (0.86%)		
Gender				1	1
FEMALE	87 (24.86%)	37 (10.57%)	124 (35.43%)		
MALE	158 (45.14%)	68 (19.43%)	226 (64.57%)		
T.stage				4.00E-03	0.03
T1	12 (3.43%)	4 (1.14%)	16 (4.57%)		
T2	60 (17.14%)	14 (4.00%)	74 (21.14%)		
T3	112 (32.00%)	49 (14.00%)	161 (46.00%)		
T4	61 (17.43%)	34 (9.71%)	95 (27.14%)		
Unknow	0 (0.0e+0%)	4 (1.14%)	4 (1.14%)		
N.stage				0.47	1
N0	75 (21.43%)	28 (8.00%)	103 (29.43%)		
N1	68 (19.43%)	25 (7.14%)	93 (26.57%)		
N2	51 (14.57%)	21 (6.00%)	72 (20.57%)		
N3	45 (12.86%)	26 (7.43%)	71 (20.29%)		
Unknow	6 (1.71%)	5 (1.43%)	11 (3.14%)		
M.stage				0.35	1
M0	221 (63.14%)	91 (26.00%)	312 (89.14%)		
M1	16 (4.57%)	7 (2.00%)	23 (6.57%)		
Unknow	8 (2.29%)	7 (2.00%)	15 (4.29%)		
Stage				0.35	1
I	33 (9.43%)	13 (3.71%)	46 (13.14%)		
II	83 (23.71%)	27 (7.71%)	110 (31.43%)		
III	95 (27.14%)	50 (14.29%)	145 (41.43%)		
IV	26 (7.43%)	9 (2.57%)	35 (10.00%)		
Unknow	8 (2.29%)	6 (1.71%)	14 (4.00%)		
Grade				0.69	1
G1	5 (1.43%)	4 (1.14%)	9 (2.57%)		
G2	90 (25.71%)	35 (10.00%)	125 (35.71%)		
G3	143 (40.86%)	64 (18.29%)	207 (59.14%)		
Unknow	7 (2.00%)	2 (0.57%)	9 (2.57%)		
Status				1	1
Alive	146 (41.71%)	62 (17.71%)	208 (59.43%)		
Dead	99 (28.29%)	43 (12.29%)	142 (40.57%)		

validation. Each patient was assigned with a RiskScore using formula:

$$RiskScore = \sum \beta i \times Expi$$

βi and $Expi$ present the cox regression coefficient and the expression level of i gene, respectively. According to the median of RiskScore, TCGA-train samples were classified into low and high risk group (low group and high group). Power of RiskScore to predict prognosis was assessed by KM survival curve and ROC. TCGA-test queue, entire TCGA queue and GSE66229 queue were used for validation.

2.8. Nomogram

Univariate and multivariate Cox regression analysis on clinicopathological features were conducted to further evaluate whether the RiskScore was an independent prediction factor. All independently occurring prognostic indicators were utilized to create a nomogram, which predicts OS at 1, 3, and 5 year(s) in the "rms" package. The discriminative ability of the nomogram was evaluated using calibration analysis.

Sangerbox provided assistance with this article [26].

2.9. Cell culture and transfection

DMEM containing 10% FBS (Gibco, Thermo Fisher, USA) was used to culture SGC7901 (human GC cell line) and GES-1 (human gastric mucosal epithelial cells, COBIOER, Nanjing, China) in a humid environment with at 37 °C with 5% CO_2.OSBPL1A and its
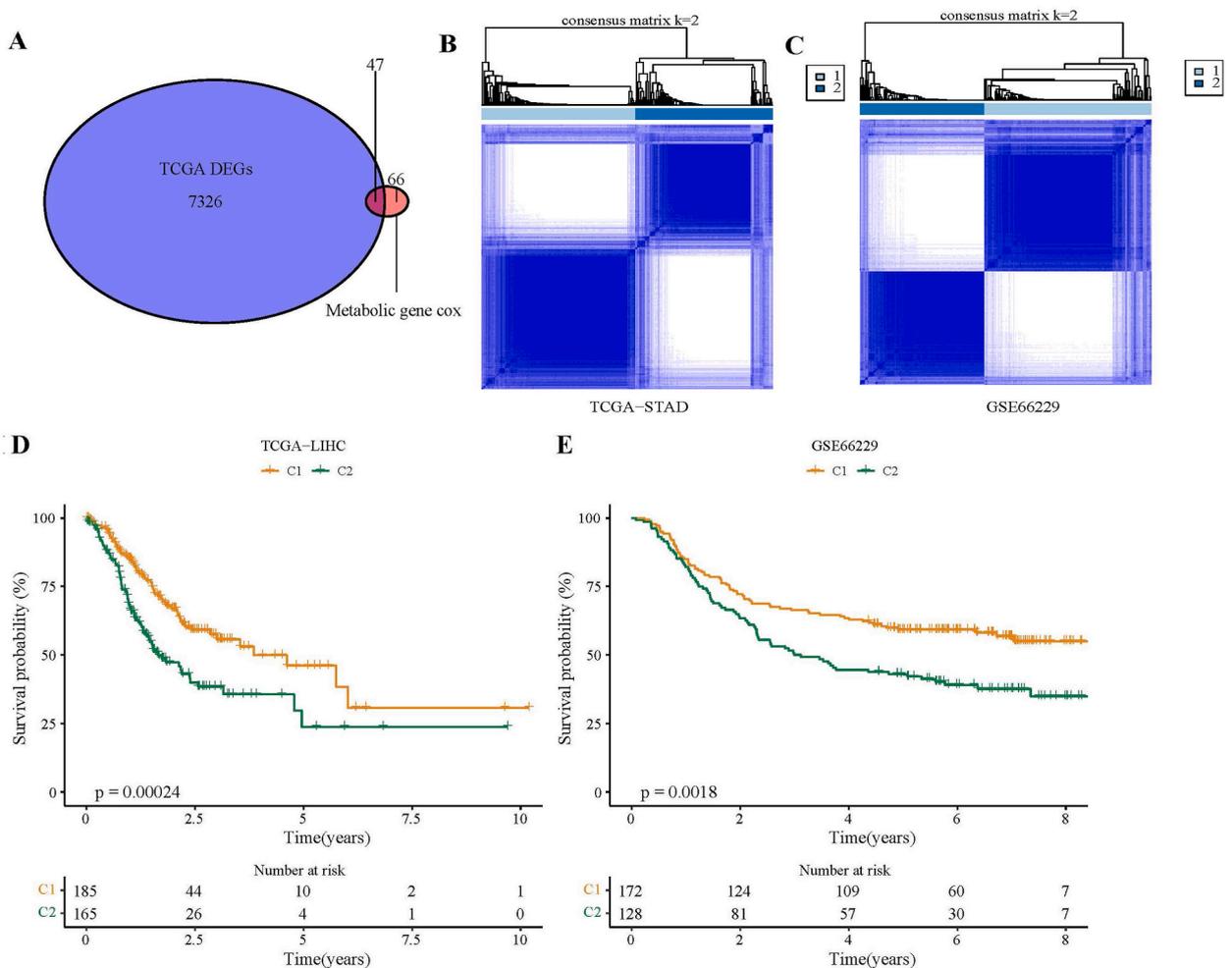
Fig. 1. 2 clusters were identified based on metabolism related genes. (A) 47 metabolic-related differentially expressed genes were screened. (B) Heatmap of clustering the samples in TCGA-STAD dataset when k = 2. (C) Heatmap of sample clustering in GSE66229 dataset when k = 2. (D) KM survival curve of 2 clusters in TCGA-STAD dataset. (E) KM survival curve of 2 clusters in GSE66229 dataset.

negative control (sh-NC, 1 μg) were synthesized by Genesee Biotechnology Co., Ltd. The OSBPL1A construct was generated using the pcDNA plasmid from Thermo Fisher Scientific. Lipofectamine 3000 (Invitrogen) was used for cell transfection following the protocol. The transfected cells were then cultured for 48 h prior to conducting the experiment. **RT-qPCR.**

Extraction of total RNA was realized using TRIzol reagent (Thermo Fisher, USA). RT-qPCR with the use of FastStart Universal SYBR Green Master (Roche, Switzerland) was performed on each sample (2 μg) on a LightCycler 480 PCR System (Roche, USA). The reaction volume of cDNA consisting of 20 μl (appropriate amount of water, 0.5 μl of forward and reverse primers, 10 μl of PCR mixture, 2 μl of cDNA template) served as a template. The PCR cycling was operated with DNA denaturation at 95 °C for 30 s (s), 45 cycles at 94 °C for 15 s, at 56 °C for 30 s, and at 72 °C for 20 s. The threshold cycle (CT) data were standardized to GAPDH by $2^{-\Delta\Delta CT}$. Table 1 listed the sequences of primer pairs for targeted genes.
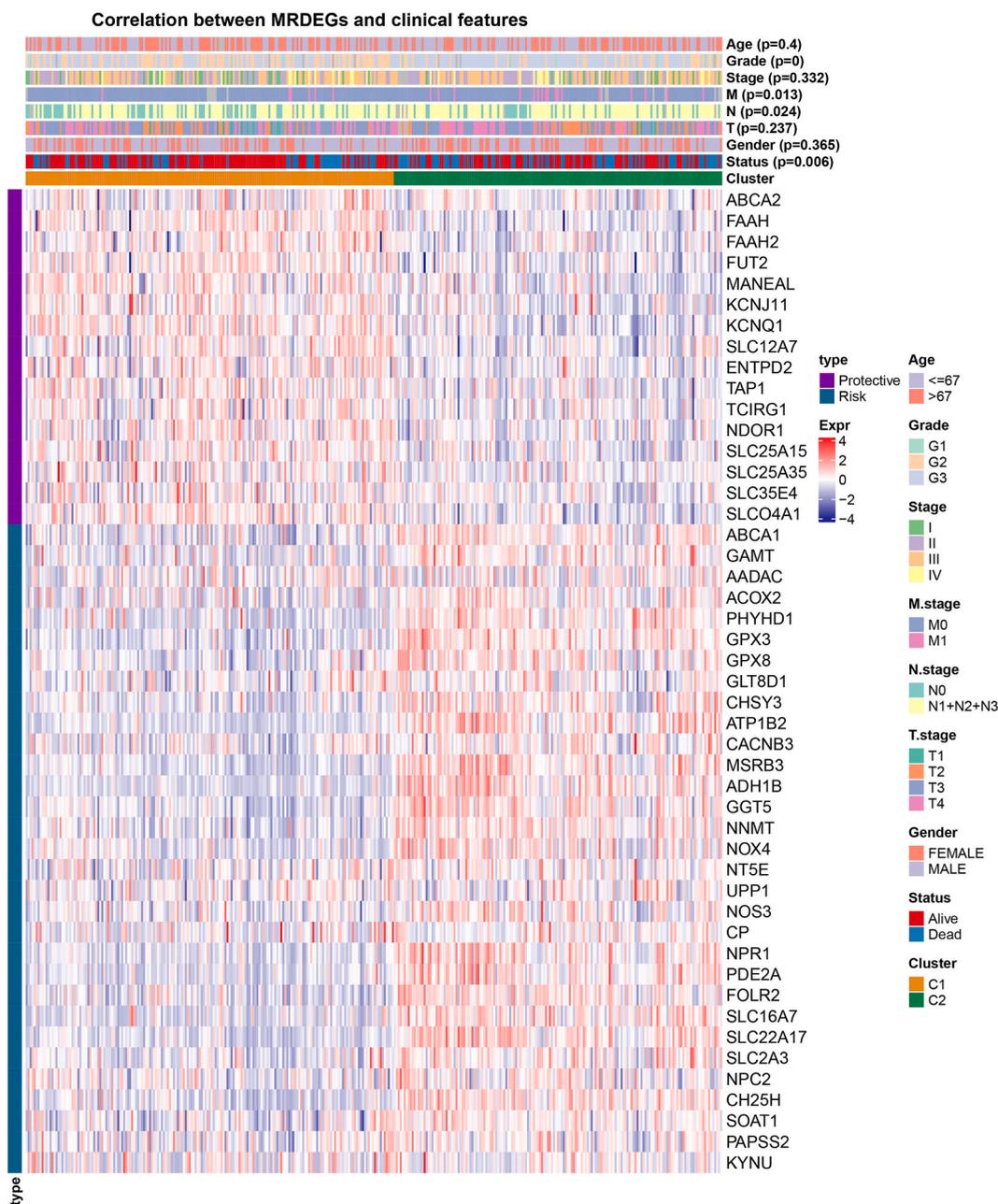
Fig. 2. The expression differences of 47 metabolic-related differentially expressed genes and distribution differences of clinical features between 2 clusters.

2.10. Transwell assays

Briefly, SGC7901 cells (5×10^4) were seeded on Matrigel-coated (BD Biosciences, USA) chambers. Complete DMEM medium and serum-free medium were respectively supplemented to the lower and upper layers. 4% paraformaldehyde was applied for cell fixation after incubation for 24 h and then the cells were dyed by crystalline violet (0.1%).

2.11. Statistical analysis

All statistics were analyzed in R software (version 3.6.0). The Wilcoxon test was applied to analyze the variances in variables between the two risk categories. Survival data were examined according to the Kaplan-Meier curve. Notably, a $P < 0.05$ was defined as a statistically significant difference.
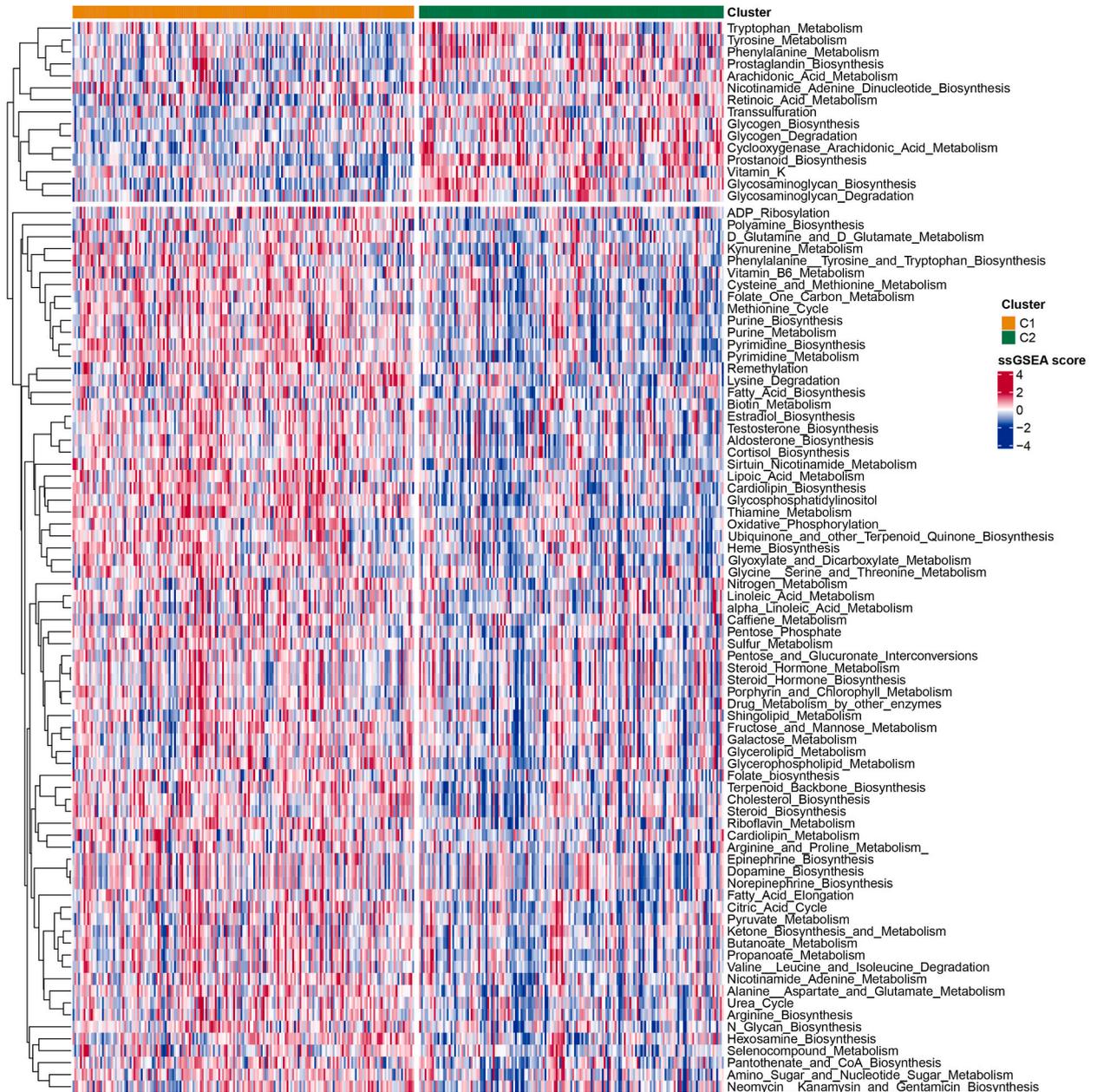
Fig. 3. 113 metabolic pathways scores differences between 2 clusters.

3. Results

3.1. Two metabolic-related molecular subtypes were identified in STAD

First, we screened DEGs between STAD and paracancerous in TCGA-STAD dataset using limma package and found 7373 DEGs. Next, 113 prognostic related metabolic genes also identified using univariate cox analysis. We obtained 47 metabolic-related differentially expressed genes (MRDEGs) in the intersection of the two types of genes (Fig. 1A) were obtained. The optimal cluster was represented by referring to cumulative density function (CDF) (fig. S1). Based on 47 MRDEGs, 2 clusters (C1 and C2) when k = 2 (Fig. 1B and C) were defined for all the samples in GSE66229 dataset and TCGA-STAD dataset. KM survival curve showed samples in C1 had a longer survival time than that in C2 both in TCGA-STAD dataset and GSE66229 dataset (Fig. 1D and E). In the distribution status of the 2 clusters among diverse clinical characteristics (Fig. 2), remarkable diversity was observed in Grade, Status, M stage, N stage in TCGA-STAD cohort study. Moreover, the difference of 47 MRDEGs expressions showed that protective genes and Risk genes were respectively higher expression in C1 and C2 (Fig. 2).

3.2. Characteristics of metabolism in C1 and C2

GSVA method was used to calculate 113 metabolic related pathways scores, and there were 74 specific metabolic characteristics in C1 and 15 specific metabolic characteristics in C2 (Fig. 3). The analysis indicated that C1 and C2 had significant differences in metabolic characteristics.
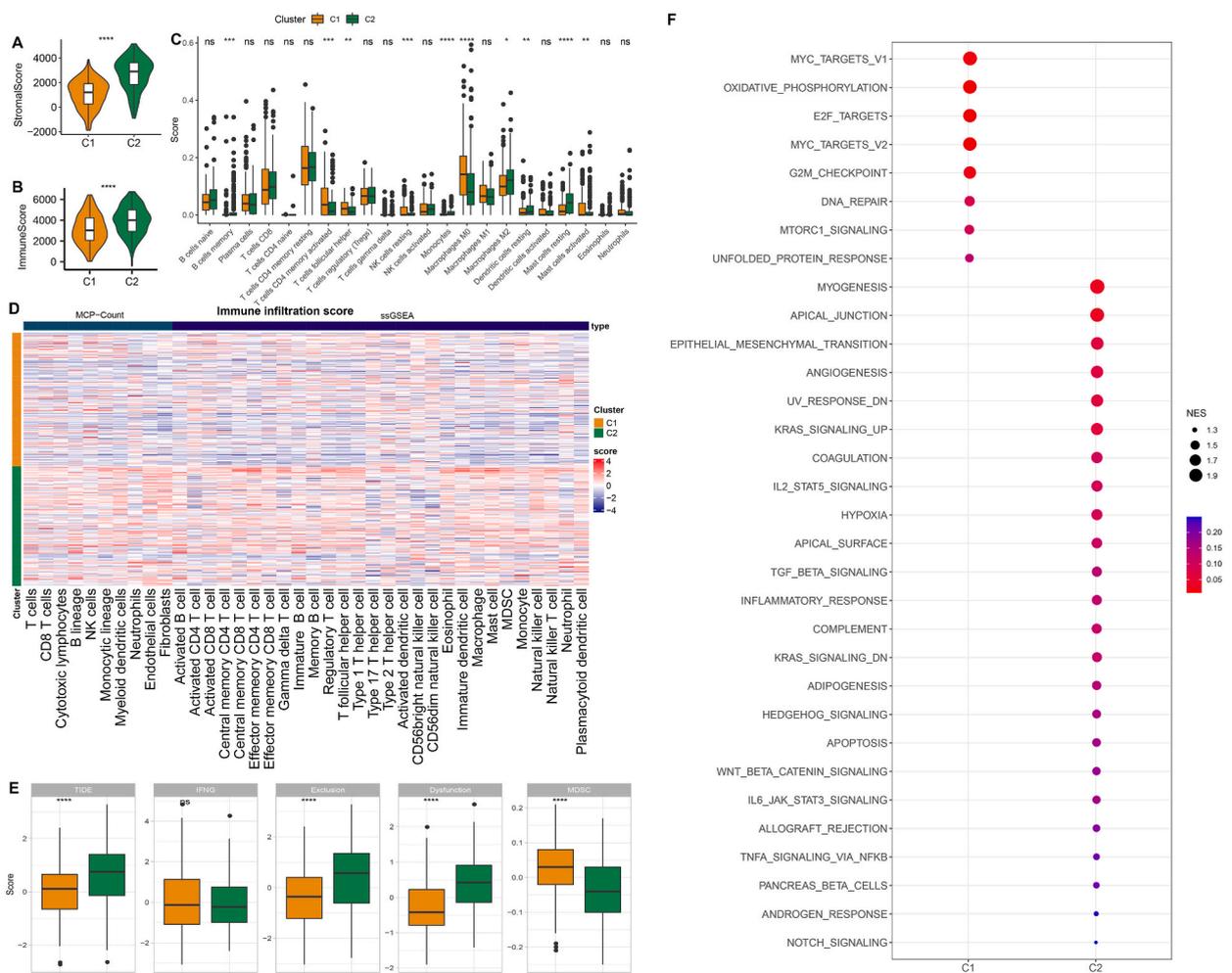
Fig. 4. Immunoinfiltration analysis between 2 clusters. (A) C2 subtype had higher StromalScore. (B) C2 subtype had enhanced ImmuneScore. (C) 22 kind immune cells score differences between 2 clusters calculated by CIBERSORT. (D) Heat map of MCP-count and ssGSEA for calculating immunity scores between 2 clusters. (E) TIDE score differences between 2 clusters. (F) GSEA analysis in 2 clusters in TCGA-STAD dataset. (ns, no significant; *p < 0.05; **p < 0.01; ***p < 0.001; and ****p < 0.0001).

3.3. Characteristics of immune in C1 and C2

In order to elucidate the immune microenvironment between C1 and C2, we first used ESTIMATE to calculate ImmuneScore and StromalScore, and the two all higher in C2 than that in C1 (Fig. 4A and B). CIBERSORT analysis on 22 kind immune cells showed that 4 immune cells (including T cells follicular helper, macrophages M0, NK cells resting, T cells CD4 memory activated) scores were higher in C1 (Fig. 4C). Heatmap of immune infiltration scores calculated by MCP-Count and ssGSEA analysis implied that most immune cells (B lineage, activated B cell, plasmacytoid dendritic cell, and so on) scores were enhanced in C2 (Fig. 4D, fig. S2). Next, the clinical effects of immune therapy on the two subtypes were evaluated using the TIDE software. TIDE score had higher in C2 in comparison to C1 indicated that a higher likelihood of escape immune and limited immunotherapy benefit in C2 (Fig. 4E). GSEA analysis demonstrated that 8 pathways and 24 pathways were enriched in C1 and C2, respectively (Fig. 4F).
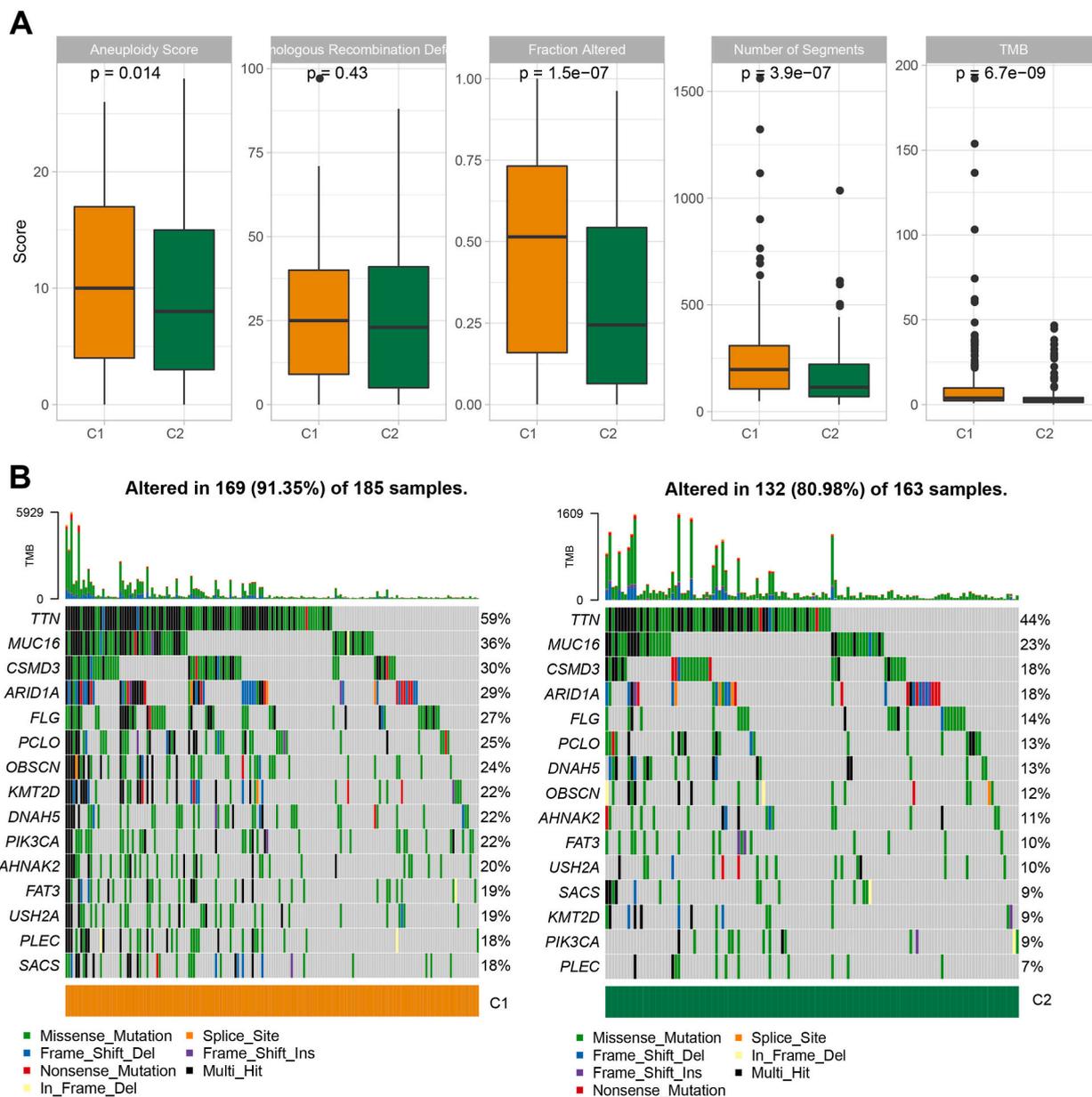
Fig. 5. Analysis of somatic mutations. (A) Comparisons on the number of Segments, Fraction Altered, TMB, and Aneuploidy Score, Homologous Recombination Defects in 2 clusters were compared. (B) Top15 gene mutation in 2 clusters.

3.4. Characteristic of mutation in C1 and C2

Immune infiltration could reflect DNA damage, including tumor mutation burden (TBM), copy number variation (CNV) burden (fraction altered and number of segments), aneuploidy, homologous recombination deficiency (HRD) [27,28]. A higher of TBM, number of segments, aneuploidy score, fraction altered were observed in C1 (Fig. 5A). A total of 1126 mutated genes showing mutation frequency greater than 3 were obtained by mutect2. Top15 genes in C1 and C2 showed that C1 had a higher percentage of the samples mutated (91.35%) than that in C2 (80.98%). Moreover, gene mutation, such as TTN (59%vs44%), MUC16 (36%vs23%), CSMD3 (30% vs18%), were much in C1 in comparison to C2 (Fig. 5B).
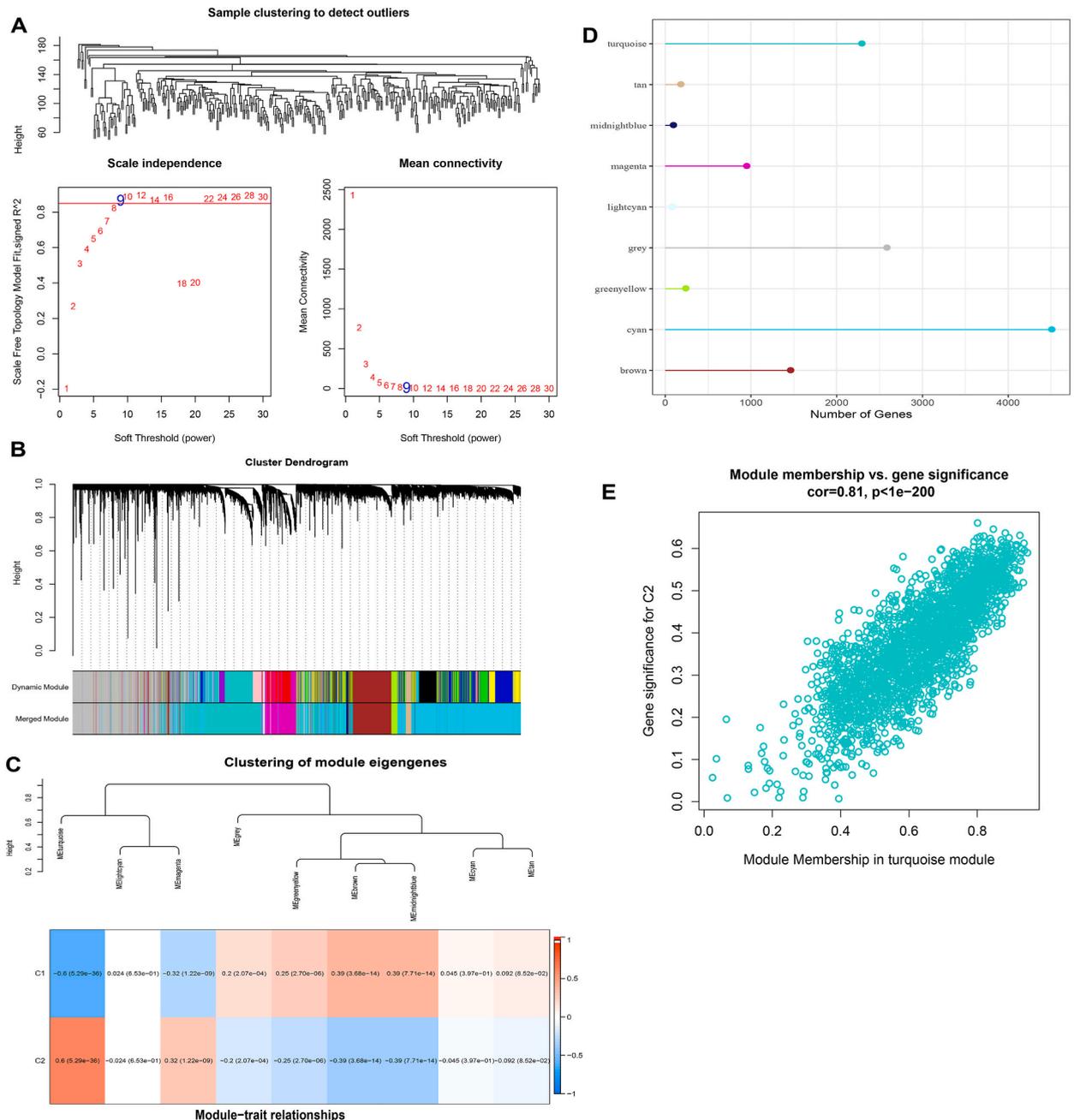
Fig. 6. Turquoise module was a hub module. (A) A cluster tree for TCGA-STAD sample. Various soft-thresholding powers calculated by analyzing the scale-free fit index (β); The mean connectivity to determine different soft-thresholding powers. (B) Based on 1-TOM, differentially expressed genes were clustered and visualized in dendrogram. (C) Correlation between 9 modules and two clusters. (D) Number of genes in 9 modules. (E) Scatter diagram was plotted based on module membership vs. gene significance for C2 in the turquoise module.
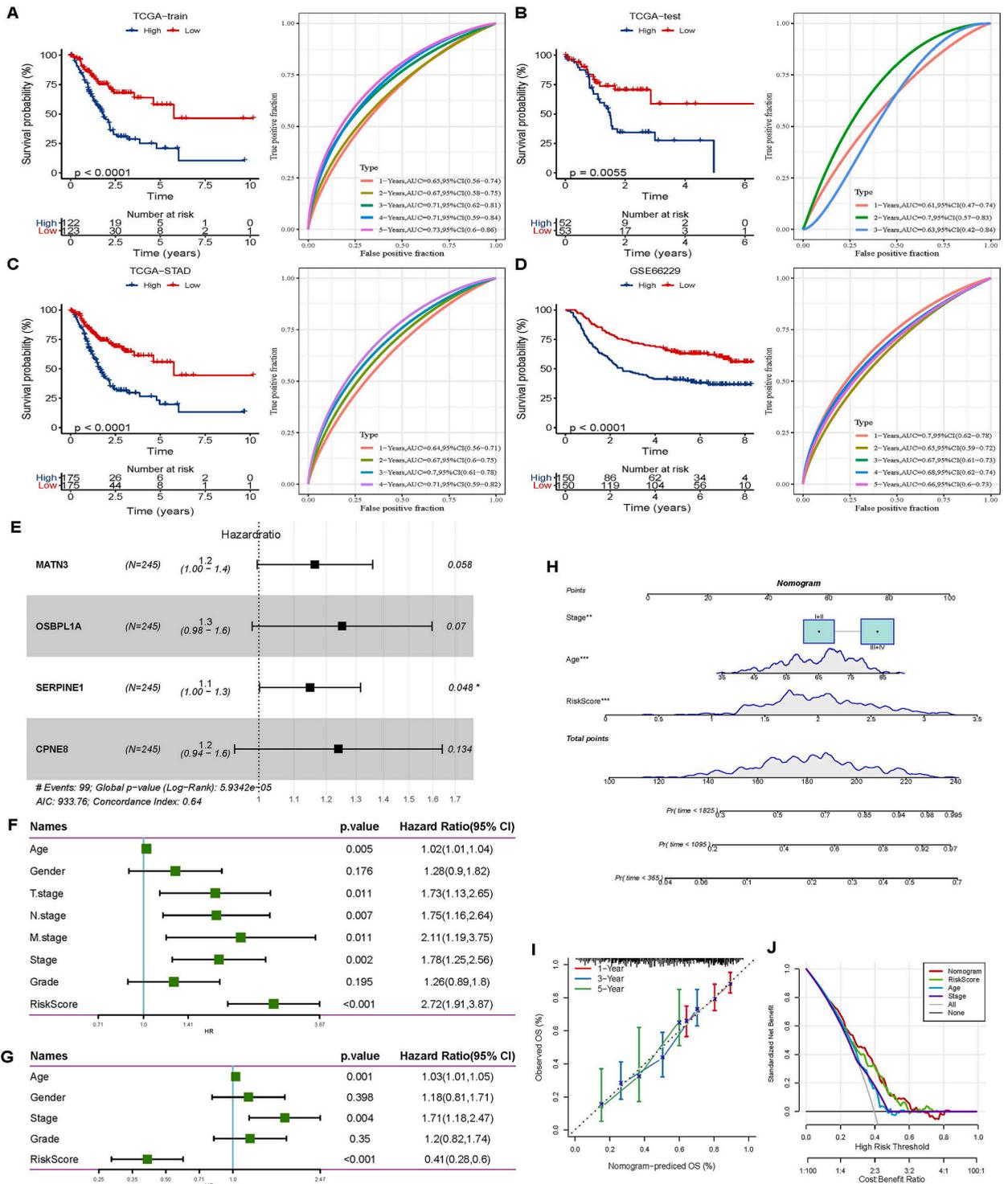
Fig. 7. Signature development using the 4 genes. (A) The RiskScore prediction in TCGA training dataset was assess by the KM and ROC curves. (B) The RiskScore prediction in TCGA test dataset was assess by the KM and ROC curves. (C) The RiskScore prediction in entire TCGA dataset was assess by the KM and ROC curves. (D) The RiskScore prediction in GSE66229 dataset was assess by the KM and ROC curves. (E) Multivariate forest map of genes in prognosis model. (F) Univariate cox analysis on the RiskScore and clinical parameters. (G) Multivariate cox analysis on the RiskScore and clinical parameters. (H) A nomogram developed by clinical parameters of independent prognosis and RiskScore. (I) 1-, 3-, 5- year calibration curve of nomogram. (J) The decision curve for the nomogram, RiskScore and independent prognostic clinical features.

3.5. Turquoise model was significantly associated with 2 clusters

Next, molecular subtype related gene model was determined by R package WGCNA. A scale-free network was developed under the soft threshold of 10 (R 2 = 0.98) (Fig. 6A). A total of 9 modules were obtained by average hierarchical clustering and dynamic tree cutting (Fig. 6B). Analysis on the correlation between model and C1/C2 showed that turquoise model was significantly negatively and positively related to C1 (r = −0.6) and C2 (r = 0.6), respectively (Fig. 6C). Number of genes in 9 model was showed in Fig. 6D. Modulemembership (MM) of genes was highly positively correlated with genesignificance (GS) in turquoise model (r = 0.81, $p < 1e$-200) (Fig. 6E). R package clusterProfiler analysis on genes in turquoise model demonstrated that those genes were enriched to cancer-correlated pathways, for instance, PI3K-Akt signaling pathway (fig. S3). Thus, turquoise model was used for next analysis.

3.6. Development and verification of prognosis model

To establish a robust risk signature for clinical practice, 30 prognostic genes ($p < 0.01$) from TCGA-train dataset were screened by univariate Cox regression analysis from genes turquoise model. Finally, we identified 4 genes analyzed by LASSO analysis. The RiskScore was follow:

RiskScore = 0.151*MATN3 expression +0.225*OSBPL1A expression +0.138*SERPINE1 expression +0.215*CPNE8 expression.

According to median of RiskScore, low group and high group of samples in TCGA train, test, entire TCGA dataset and GSE66229 dataset were divided, with the prognosis of the later group showing a worse OS than those in low group both in above datasets. In TCGA-train dataset, 1-, 2-, 3-, 4- and 5- years AUC was respectively 0.65, 0.67, 0.71, 0.71 and 0.73 (Fig. 7A). In TCGA-test dataset, 1-, 2-, and 3- years AUC was respectively 0.61, 0.7 and 0.63 (Fig. 7B). In entire TCGA dataset, 1-, 2-, 3-, and 4- AUC was respectively 0.64, 0.67, 0.7 and 0.71 (Fig. 7C). To better verify the model robustness, in GSE66229 dataset, 1-, 2-, 3-, 4- and 5- years AUC was respectively 0.7, 0.65, 0.67, 0.68 and 0.66 (Fig. 7D).

We found that the four genes were risk factors (Fig. 7E). Stage, Riskscore, and age were validated as the significant factors for prognostic prediction by T univariate and multivariate Cox regression analysis (Fig. 7F and G) and combined together to create a nomogram for assessing the risk of STAD and its prognosis. The impact of Riskscore on predicting STAD survival was the greatest (Fig. 7H). The calibration curves of predicted 1-, 3- and 5- year were close to the standard curves, suggesting an accurate prediction result by the nomogram (Fig. 7I). Decision curve (DCA) showed obviously higher efficacy of using Riskscore and Nomogram over the
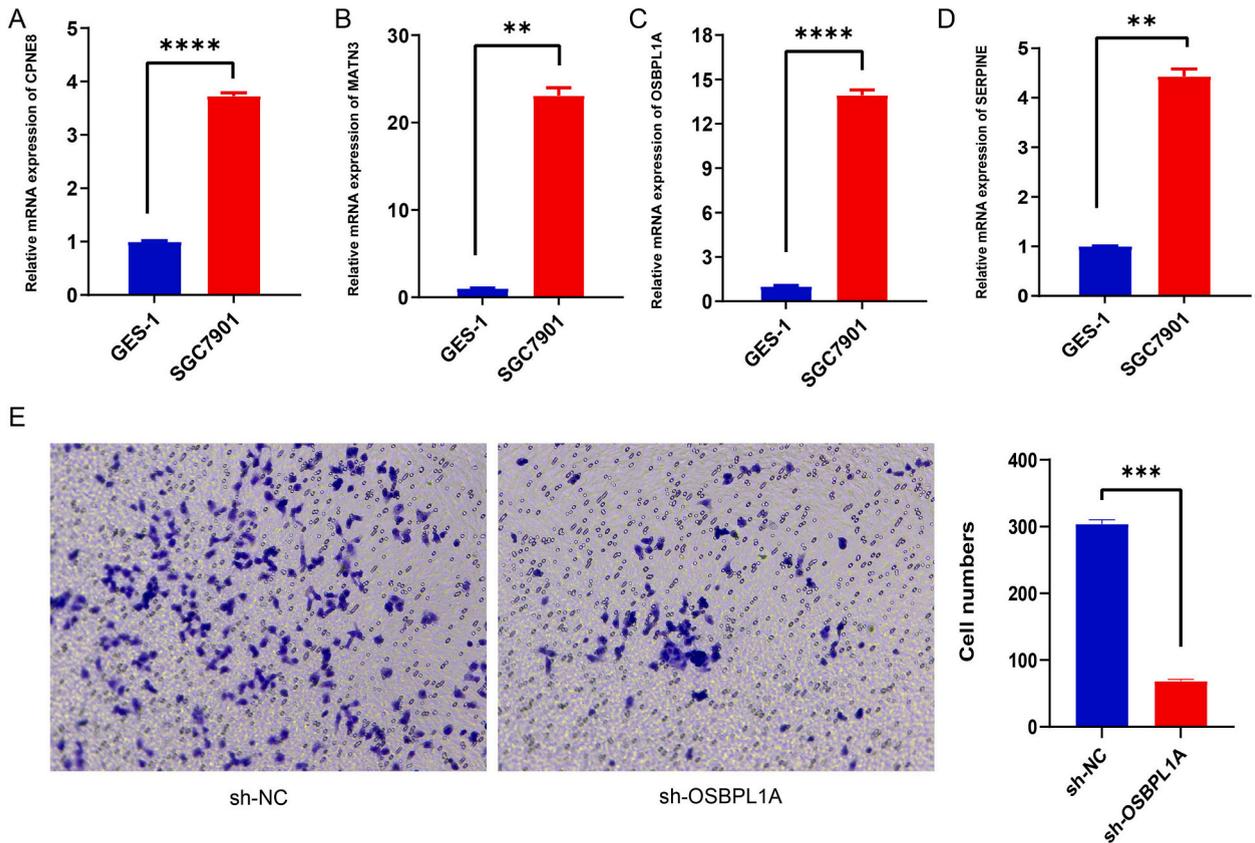
Fig. 8. Genes in model was validated. (A–D) MATN3, OSBPL1A, SERPINE1 and CPNE8 mRNA levels were elevated in SGC7901 cells than normal GSE-1 cells. (E) Inhibition of OSBPL1A decreased SGC7901 cells invasion numbers. (**p < 0.01; ***p < 0.001; and ****p < 0.0001).
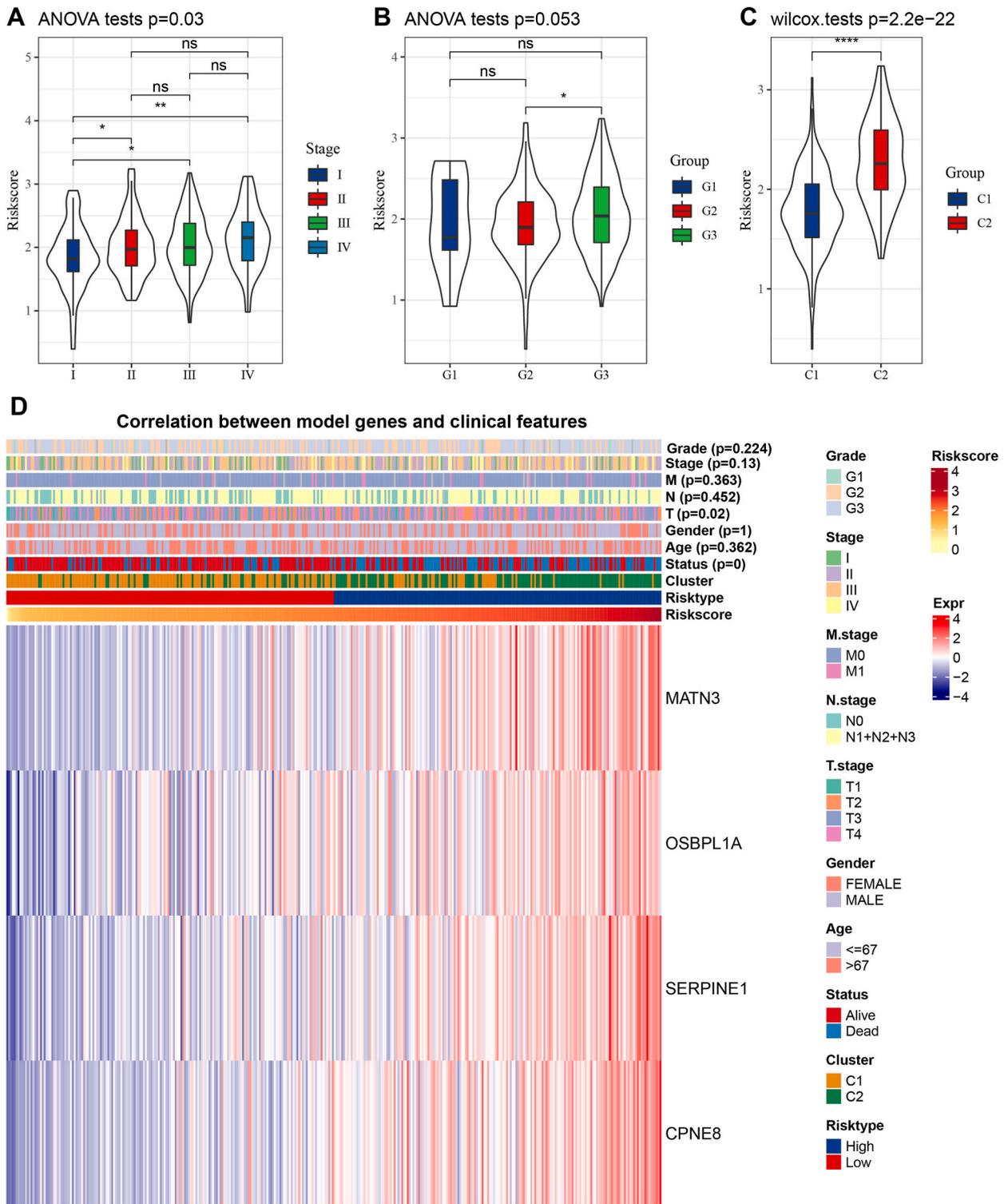
Fig. 9. Correlation analysis between RiskScore and clinical features. (A) The RiskScore differences among StageI-IV. (B) The RiskScore differences among Grade 1-3. (C) The RiskScore differences between C1 and C2. (D) the expressions of 4 genes in RiskScore and clinical feature in high- and low-group. (ns, no significant; ****p < 0.0001).

extreme curves, showing the strongest survival prediction ability (Fig. 7J).

RT-qPCR results showed that MATN3, OSBPL1A, SERPINE1 and CPNE8 mRNA levels were all elevated in SGC7901 cells (Fig. 8A–D). Of the four key genes associated with prognosis in this study, OSBPL1A was chosen for further study because it showed the largest HR value in multivariate analysis. The effect of OSBPL1A on STAD cells was confirmed to test the reliability of the prognostic genes, and inhibition of OSBPL1A was observe to be able to reduce the number of invaded SGC7901 cells (Fig. 8E).

3.7. Association between the RiskScore and clinical characteristics

Distribution of RiskScore in relation to clinical variables was investigated. Patients in advanced Stage, Grade and C2 showed higher RiskScore (Fig. 9A–C). Whether there the high- and low-RiskScore groups had different distribution of clinical features was explored. Only T stage and Status distribution had difference between two risk groups. The high group showed higher levels of the four genes (Fig. 9D).
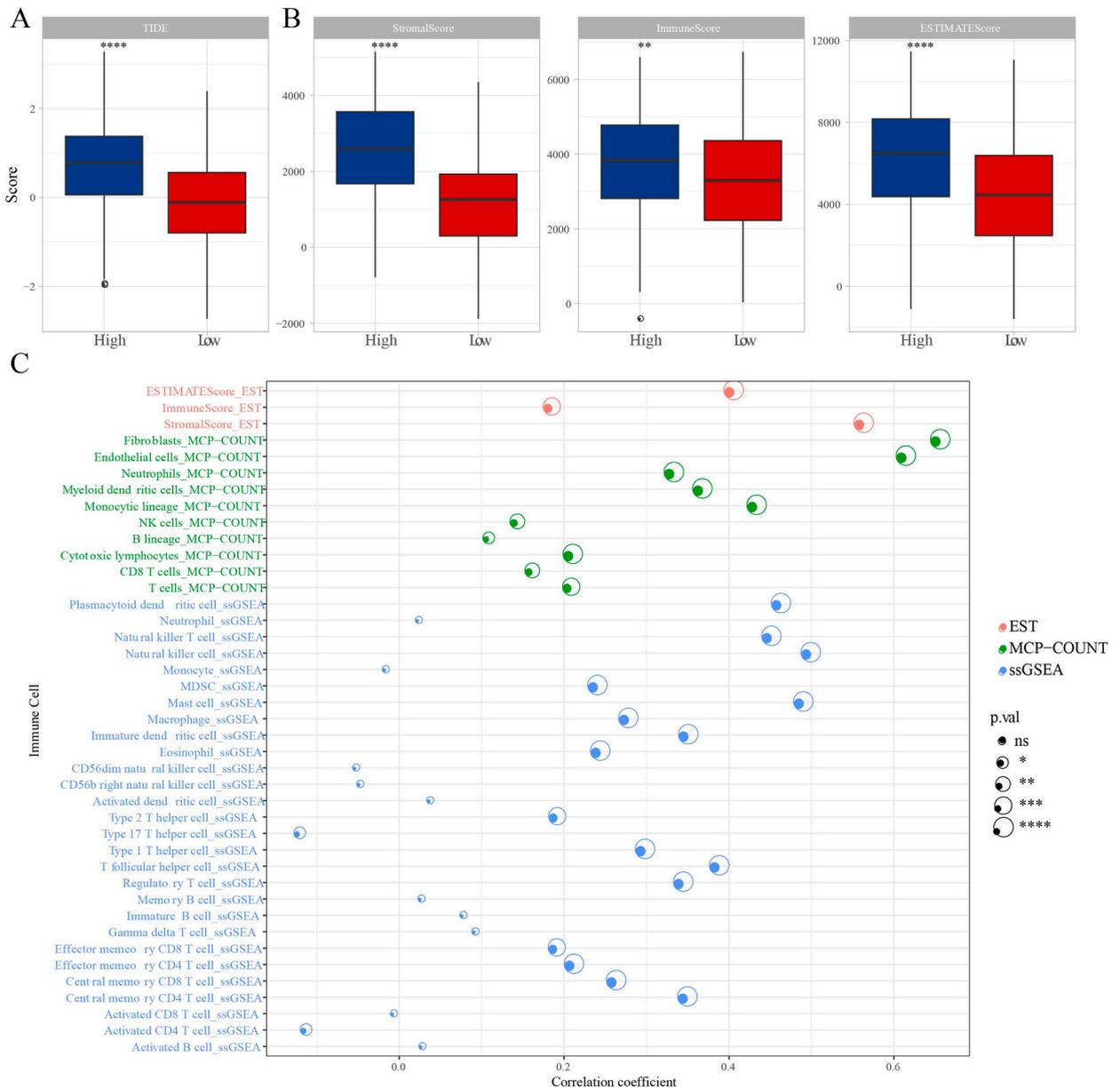
Fig. 10. Correlation analysis between RiskScore and immune cells. (A) Differences in TIDE scores between the two groups. (B) Differences in StromalScore, ImmuneScore and EstmateScore between the two groups. (C) ESTIMATE, MCP-Count and ssGSEA analysis were used to perform correlation analysis between immune cells and RiskScore. (**p < 0.01; ****p < 0.0001).

3.8. Association between the RiskScore and immune characteristics

The high group had significantly higher TIDE score, which suggested a probability of occurring immune escape and less immunotherapy benefit (Fig. 10A). Moreover, StromalScore, ImmuneScore, and ESTIMATEScore were also more enhanced in high group than low group (Fig. 10B). Subsequently, the results from the ESTIMATE, MCP-counter, and ssGSEA methods demonstrated that infiltration of immune cells such as effector memory CD8 T cells, mast cells, activated B cell, central memory CD4 T cell, macrophages, regulatory T cells was relatively high in the high group (Fig. 10C, fig. S4A). Moreover, we confirmed that high group had a high infiltration of Monocytic lineage, CD8 T cells, fibroblasts, B lineage, NK cells, Neutrophils, Cytotoxic lymphocytes, T cells, Myeloid dendritic cells, Endothelial cells (Fig. 10C, fig. S4B). Therefore, the study of genes involved in cellular energy metabolism can provide particular insights into immunotherapeutic strategies for STAD patients.

4. Discussion

The availability of high-quality data sets and the availability of sequencing techniques and tools to analyze the data sets, metabolic phenotype experiments contributes to the study of cancer metabolic complexity. Changes in cellular energy metabolisms, including glutamine metabolism, glycolysis, fatty acid metabolism, plays a critical role in the cancer initiation and development, which provide cancer cells with a growth advantage and enhance the formation of an aggressive phenotype [29]. In this study, two new metabolic classifiers were established for the STAD cohort based on metabolism-related genes. Specific differences in clinical features, metabolomics, immune features, and genomics between the two subtypes suggested the significance of metabolic heterogeneity to be taken into consideration in developing personalized treatments. Further, WGCNA and LASSO analysis was employed to establish a 4-gene prognosis system, and the results showed RiskScore and clinical features, differences in immune microenvironment that will improve the current knowledge of the metabolic heterogeneity of STAD, providing novel understanding for targeted therapies.

Cancer cells exhibit metabolic heterogenous dependencies and preferences [7,30,31]. Knowledge on metabolic variability and flexibility has important significance as it affects the way we make use of metabolic reprogramming in cancer treatment. Previously, Jin-Jia Chang et al. carried out broad energy-metabolism profiling and classified two subtypes based on 86 energy-metabolism-related genes. Among them, they found that subtypes with a higher number of younger patients exhibited more and immune and mesenchymal cell components, and showed that early-stage subtypes also had better progression-free survival [32]. Chunhua Liu et al. stratified STAD into two distinct tumor subtypes based on 13 genes involved in fatty acid metabolism [33]. In addition, 4 subgroups of pancreatic cancer were categorized based on the glycolysis-cholesterol synthesis axis [34]. The current study showed significant differences in its findings in comparison with recently published papers. Firstly, an integrative analysis on multiomics data (metabolomics, genomics, transcriptomics) were performed. Then, we proposed that unique metabolic dependencies could be tackled by focusing on metabolic-related genes. Finally, the difference of metabolic subtypes further constructed the prognostic model.

A Phase II randomized clinical trial [35] study showed an objective response rate of 11.6% and a median survival time of 5.5 months of using the PD-L1 inhibitor Pembrolizumab as advanced gastric cancer treatment. PD-L1 positive patients had 11.6% higher rates than PD-L1 negative patients (15.5% vs. 6.4%), indicating that a greater benefit from Pembrolizumab for PD-L1 positive patients. Links between immune infiltration and metabolism have been reported [36]. Reprogramming of glucose metabolism is one of the main features of TME. Tumor cells up-regulate the glycolytic pathway, carry out tumor escape, and suppress the function of immune effector cells until failure occurs [37]. C2 has a higher immunoinfiltration and a higher TIDE score, suggesting that immunotherapy in this cluster may have a smaller clinical benefit.

The present work developed a metabolism-related prognostic model with 4 genes (MATN3, OSBPL1A, SERPINE1 and CPNE8). MATN3 was reported that it was upregulated in gastric adenocarcinoma, and overexpression of MATN3 maybe be acted as an independent predictor to indicate the survival outcome of gastric adenocarcinoma [38]. In the process of reverse cholesterol transport, the first step could be influenced by the loss-of-function mutation of OSBPL1A, an intracellular lipid receptor related to a low HDL-C phenotype, indicating that mutations in OSBPL1A may lead to dyslipidemia [39]. Obviously expressed SERPINE1 is significantly correlated with an unfavorable prognostic outcome of gastric adenocarcinoma patients [40]. Feng et al. indicated by immunoassay that SERPINE1 may promote an inhibitory immune microenvironment in gastric cancer and it was positively related to the infiltration of various immune cells (e.g., neutrophils, resting macrophages M2, activated mast cells, NK cells) [41]. In breast cancer, glucose metabolism levels could be promoted by stabilizing p53-induced SERPINE1 in the miR-1185-2-3p-GOLPH3L pathway [42]. Enhanced CPNE8 expression samples was related to clinical characteristics and was predictive of unfavorable survival of patients with gastric cancer [43]. Those data supported the reliability of the current prognostic model developed using the four genes. However, we should note some limitations. In this study, research data were extracted based on public databases that were analyzed only by bioinformatics. Further in vivo validation experiments are necessary. However, further verification is required to probe into the molecular mechanisms underlying energy metabolism in STAD using cell experiments.

Cancer metabolism is a metabolic pathways consisting of intricate networks, and targeting cancer metabolism could selectively suppress cancer progression as cancer metabolisms are realized under different metabolic stress responses between normal cells and cancer cells. Therapies targeting cancer metabolism requires comprehensive explorations before clinical administration. In a sum, we identified two metabolically related molecular subtypes and constructed a 4 gene prognosis model for STAD, which provided a classifier for STAD patients and predicting prognosis.

Funding

This study received no funding.

Data availability statement

The dataset used in this study was available in GSE66229 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE66229).

Ethical statement

As the present study did not involve any human experiment, informed consent was exempted.

CRediT authorship contribution statement

Jie Sun: Writing – review & editing, Writing – original draft, Supervision, Software, Data curation, Conceptualization. **Yuanyuan Wang:** Writing – original draft, Visualization, Supervision, Project administration, Methodology, Investigation. **Kai Zhang:** Software, Resources, Project administration, Formal analysis. **Sijia Shi:** Visualization, Validation, Software, Methodology. **Xinxin Gao:** Visualization, Supervision, Methodology, Data curation. **Xianghao Jia:** Visualization, Validation, Supervision. **Bicong Cong:** Investigation, Data curation, Conceptualization. **Chunning Zheng:** Writing – review & editing, Validation, Methodology, Investigation, Conceptualization.

Declaration of competing interest

All the authors declared no personal relationships or competing financial interests that could influence the results reported by this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e28413.

Abbreviations

KEGG	Kyoto Encyclopedia of Genes and Genomes
OS	overall survival
FPKM	Fragments per kilobase of transcript per million fragments mapped
STAD	stomach adenocarcinomas
ssGSEA	Single sample gene set enrichment analysis
ROC	operating characteristic curve
TCGA	The Cancer Genome Atlas
FDR	False discovery rate
CDF	Cumulative distribution function
TPM	Transcripts per kilobase million
MRDEGs	Metabolic-related differentially expressed genes
GEO	Gene Expression Omnibus
CNV	Copy number variation
DCA	Decision curve
GO	Gene Ontology
TOM	Topological overlap matrix
CIBERSORT	Cell-type Identification using Estimating Relative Subsets of RNA Transcripts
TBM	Tumor mutation burden
HRD	Homologous recombination deficiency
ESTIMATE	Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data

References

[[1]] H. Sung, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J Clin 71 (3) (2021) 209–249.

[[2]] J. Qiu, et al., Identification of the novel prognostic biomarker SERPINH1 reveals its relationship with immunology in gastric cancer, Oncologie 25 (4) (2023) 367–379.

[[3]] H. Xu, et al., Hsa_circ_0079598 acts as a potential diagnostic and prognostic biomarker for gastric cancer, Oncologie 25 (2) (2023) 179–186.

[[4]] J.A. Ajani, et al., Gastric adenocarcinoma, Nat Rev Dis Primers 3 (2017) 17036.

[[5]] J. Zou, et al., Construction of gastric cancer patient-derived organoids and their utilization in a comparative study of clinically used paclitaxel nanoformulations, J. Nanobiotechnol. 20 (1) (2022) 233.

[[6]] W. Sun, et al., Targeting serine-glycine-one-carbon metabolism as a vulnerability in cancers, Biomark. Res. 11 (1) (2023) 48.

[[7]] Y. Zhao, et al., Biomimetic nanovesicle co-delivery system impairs energy metabolism for cancer treatment, J. Nanobiotechnol. 21 (1) (2023) 299.

[[8]] M. Stein, et al., Targeting tumor metabolism with 2-deoxyglucose in patients with castrate-resistant prostate cancer and advanced malignancies, Prostate 70 (13) (2010) 1388–1394.

[[9]] R.L. Aft, F.W. Zhang, D. Gius, Evaluation of 2-deoxy-D-glucose as a chemotherapeutic agent: mechanism of cell death, Br. J. Cancer 87 (7) (2002) 805–812.

[[10]] X. Peng, et al., Molecular characterization and clinical relevance of metabolic expression subtypes in human cancers, Cell Rep. 23 (1) (2018), 255-269.e4.

[[11]] E. Gaude, C. Frezza, Tissue-specific and convergent metabolic transformation of cancer correlates with metastatic potential and patient survival, Nat. Commun. 7 (2016) 13041.

[[12]] M.A. Prusinkiewicz, et al., Survival-associated metabolic genes in human papillomavirus-positive head and neck cancers, Cancers 12 (1) (2020).

[[13]] L. Xing, et al., A transcriptional metabolic gene-set based prognostic signature is associated with clinical and mutational features in head and neck squamous cell carcinoma, J. Cancer Res. Clin. Oncol. 146 (3) (2020) 621–630.

[[14]] X. Meng, et al., Combined analysis of RNA-sequence and microarray data reveals effective metabolism-based prognostic signature for neuroblastoma, J. Cell Mol. Med. 24 (18) (2020) 10367–10381.

[[15]] H. Xu, et al., Discovery and validation of an epithelial-mesenchymal transition-based signature in gastric cancer by genomics and prognosis analysis, BioMed Res. Int. 2021 (2021) 9026918.

[[16]] M.E. Ritchie, et al., Limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res. 43 (7) (2015) e47.

[[17]] R. Possemato, et al., Functional genomics reveal that the serine synthesis pathway is essential in breast cancer, Nature 476 (7360) (2011) 346–350.

[[18]] M.D. Wilkerson, D.N. Hayes, ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking, Bioinformatics 26 (12) (2010) 1572–1573.

[[19]] S. Hänzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data, BMC Bioinf. 14 (2013) 7.

[[20]] K. Yoshihara, et al., Inferring tumour purity and stromal and immune cell admixture from expression data, Nat. Commun. 4 (2013) 2612.

[[21]] A.J. Gentles, et al., The prognostic landscape of genes and infiltrating immune cells across human cancers, Nat Med 21 (8) (2015) 938–945.

[[22]] K. Cao, et al., Prediction of prognosis and immunotherapy response with a robust immune-related lncRNA pair signature in lung adenocarcinoma, Cancer Immunol. Immunother. 71 (6) (2022) 1295–1311.

[[23]] G. Yu, et al., clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS 16 (5) (2012) 284–287.

[[24]] V. Thorsson, et al., The immune landscape of cancer, Immunity 48 (4) (2018), 812-830.e14.

[[25]] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinf. 9 (2008) 559.

[[26]] W. Shen, et al., Sangerbox: a comprehensive, interaction-friendly clinical bioinformatics analysis platform, iMeta 1 (3) (2022) e36.

[[27]] J.J. Bednarski, B.P. Sleckman, At the intersection of DNA damage and immune responses, Nat. Rev. Immunol. 19 (4) (2019) 231–242.

[[28]] V. Queirós, et al., Effects of ifosfamide and cisplatin exposure combined with a climate change scenario on the transcriptome responses of the mussel Mytilus galloprovincialis, Sci. Total Environ. 885 (2023) 163904.

[[29]] G. Wang, et al., Targeting cellular energy metabolism- mediated ferroptosis by small molecule compounds for colorectal cancer therapy, J. Drug Target. 30 (8) (2022) 819–832.

[[30]] L. Hu, et al., A metabolism-related gene signature predicts the prognosis of breast cancer patients: combined analysis of high-throughput sequencing and gene chip data sets, Oncologie 24 (4) (2022) 803–822.

[[31]] V. Van den Bossche, et al., Microenvironment-driven intratumoral heterogeneity in head and neck cancers: clinical challenges and opportunities for precision medicine, Drug Resist Updat 60 (2022) 100806.

[[32]] J.J. Chang, et al., Comprehensive molecular characterization and identification of prognostic signature in stomach adenocarcinoma on the basis of energy-metabolism-related genes, World J. Gastrointest. Oncol. 14 (2) (2022) 478–497.

[[33]] C. Liu, et al., Classification of stomach adenocarcinoma based on fatty acid metabolism-related genes frofiling, Front. Mol. Biosci. 9 (2022) 962435.

[[34]] J.M. Karasinska, et al., Altered gene expression along the glycolysis-cholesterol synthesis Axis is associated with outcome in pancreatic cancer, Clin. Cancer Res. 26 (1) (2020) 135–146.

[[35]] C.S. Fuchs, 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. 4 (5) (2018) e180013.

[[36]] J. Fan, et al., ABC transporters affects tumor immune microenvironment to regulate cancer immunotherapy and multidrug resistance, Drug Resist Updat 66 (2023) 100905.

[[37]] I. Shevchenko, A.V. Bazhin, Metabolic checkpoints: novel avenues for immunotherapy of cancer, Front. Immunol. 9 (2018) 1816.

[[38]] P.L. Wu, et al., Martrilin-3 (MATN3) overexpression in gastric adenocarcinoma and its prognostic significance, Med Sci Monit 24 (2018) 348–355.

[[39]] M.M. Motazacker, et al., A loss-of-function variant in OSBPL1A predisposes to low plasma HDL cholesterol levels and impaired cholesterol efflux capacity, Atherosclerosis 249 (2016) 140–147.

[[40]] L. Li, et al., FN1, SPARC, and SERPINE1 are highly expressed and significantly related to a poor prognosis of gastric adenocarcinoma revealed by microarray and bioinformatics, Sci. Rep. 9 (1) (2019) 7827.

[[41]] L. Feng, et al., Cuproptosis-related gene SERPINE1 is a prognostic biomarker and correlated with immune infiltrates in gastric cancer, J. Cancer Res. Clin. Oncol. 149 (12) (2023) 10851–10865.

[[42]] Y. Xu, et al., The miR-1185-2-3p-GOLPH3L pathway promotes glucose metabolism in breast cancer by stabilizing p53-induced SERPINE1, J. Exp. Clin. Cancer Res. 40 (1) (2021) 47.

[[43]] P. Zhang, et al., CPNE8 promotes gastric cancer metastasis by modulating focal adhesion pathway and tumor microenvironment, Int. J. Biol. Sci. 18 (13) (2022) 4932–4949.