Identification of an Immune Gene Signature Based on Tumor Microenvironment Characteristics in Colon Adenocarcinoma

Ying Chen[1,2] and Jia Zhao[1,2] (ID)

Abstract

Tumor microenvironment (TME) changes are related to the occurrence and development of colon adenocarcinoma (COAD). This study aimed to analyze the characteristics of the immune microenvironment in CC, as well as the microenvironment's relationship with the clinical features of CC. Based on The Cancer Genome Atlas (TCGA) and GSE39582 cohorts, the scores of 22 tumor infiltrating lymphocytes (TILs) were calculated using CIBERSORT. *ConsensusClusterPlus* was used for unsupervised clustering. Three TME subtypes (TMEC1, TMEC2, and TME3) were identified based on TIL scores. TMEC2 was associated with the worst prognosis. Random forest, k-means clustering, and principal component analysis were used to construct the TME score risk signature. The median TME score was used to divide the samples into high- and low-risk groups. The prognoses of the patients with high TME scores were worse than those of the patients with low TME scores. A high TME score was an independent prognostic risk factor for patients with colon cancer. The Gene Set Enrichment Analysis (GSEA) results showed that those with high TME scores were enriched in FOCAL_ADHESION, ECM_RECEPTOR_INTERACTION, and PATHWAYS_IN_CANCER. Our findings will provide a new strategy for immunotherapy in patients with CC.

Keywords

colon adenocarcinoma, prognosis, tumor microenvironment, immunotherapy

Introduction

Colorectal cancer (CRC) is one of the most common malignant tumors of the digestive system[1,2]. In 2015, CRC was the fourth most common malignant tumor in women and the fifth most common malignant tumor in men in China[3]. Although significant progress has been made in early diagnosis, surgery, radiotherapy, and chemotherapy, the overall survival rate of CRC can still be improved[4]. The high rates of recurrence and metastasis are important reasons for the short survival time and poor prognoses of patients with colon cancer[5,6].

In the past, people focused on the treatment of a tumor. However, researchers have shown that the occurrence and development of a tumor is a dynamic process with multiple factors, multiple stages, and multiple links, it not only involves the tumor cells, but the tumor microenvironment (TME) is also closely related[7,8]. The TME is composed of tumor cells and the matrix microenvironment, which contains extracellular matrix and interstitial cells. Interstitial cells include fibroblasts, vascular structure cells, and immune cells[9]. Tumor cells, interstitial cells, and the extracellular matrix interact to produce and release various chemokines, cytokines, and other mediators. This forms an inflammatory state in the tissue, forms an immunosuppressive TME, assists tumor cells in escaping immune surveillance, and eventually leads to tumor occurrence, development, and metastasis[10,11]. Immunotherapy has shown great anticancer activity in many cancers such as colon

[1] Department of Medical Oncology, the First Hospital of China Medical University, Shenyang, China
[2] Key Laboratory of Anticancer Drugs and Biotherapy of Liaoning Province, the First Hospital of China Medical University, Shenyang, China

Submitted: February 10, 2021. Revised: February 10, 2021. Accepted: February 18, 2021.

Corresponding Author:
Jia Zhao, Department of Medical Oncology, the First Hospital of China Medical University, Shenyang, China.
Email: jiazhao84@hotmail.com

adenocarcinoma and melanoma, indicating the clinical therapeutic potential of targeting the immune microenvironment.

More and more studies have shown that changes in the immune microenvironment are closely related to the occurrence and development of CRC[12]. Clinical studies have shown that long-term use of the anti-inflammatory drug aspirin, which inhibits the inflammatory environment, can reduce the risk of cancer in patients with familial colonic polyps[13]. In recent years, it has been found that colorectal microflora are closely related to inflammation and tumorigenesis. Some intestinal microorganisms can induce intestinal immune cells to secrete interleukin (IL)-23, promoting the occurrence and development of a tumor[14]. In addition, the type, distribution location, degree of infiltration, and cytokine components of the immune cells in the immune microenvironment are closely related to the prognoses and therapeutic effects in patients with tumors[15]. A variety of cytokines in the microenvironment can promote the invasion and metastasis of CRC[16,17]. Immune cells and cytokines in the microenvironment are closely related to the occurrence, development, metastasis, and therapeutic outcomes of CRC. However, there has been no systematic analysis of the characteristics of the immune microenvironment's relation with the clinical characteristics of colon adenocarcinoma.

In this study, TME molecular subtypes were identified based on tumor infiltrating cell (TIL) scores obtained from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO). Two TME score risk signatures were identified. TME score showed a significant correlation with tumor, node, metastasis (TNM) stage and immune gene expression. Finally, we identified a group of tumor mutant genes related to TME score.

Materials and methods

Data Downloads and Processing

Clinical follow-up information was downloaded using the TCGA Genomic Data Commons (GDC) application performing interface (API), with 499 RNA sequencing (RNA-Seq) samples. The microarray gene expression profile of GSE39582, downloaded from the National Center for Biotechnology Information (NCBI) in MINiML format, contained 573 samples with clinical information. The following steps were conducted in processing the 499 RNA-Seq read count samples: (1) remove the samples with an overall survival (OS) <30 days; (2) remove normal tissue samples; (3) convert read counts to transcripts per million (TPM) with the annotation information of GENCODE v22 [in terms of data distribution, the TPM was closer to the microarray than fragments per kilobase million (FPKM)]; and (4) remove half of the genes with a TPM of 0 in the samples.

The GSE39582 cohort was preprocessed in the following manner: (1) remove normal tissue samples and samples without clinical follow-up information; (2) remove samples with an OS <30 days; and (3) map the microarray probe to

Table 1. Clinical Information of the Two Cohorts.

Characteristic		TCGA datasets (n = 427)	GSE39582 (n = 567)
Age (years)	<=60	133	160
	>60	294	406
Survival Status	Living	331	379
	Dead	96	188
Gender	female	197	254
	male	230	313
pathologic_T	T 1	10	12
	T 2	73	47
	T 3	293	367
	T 4	50	117
pathologic_N	N 0	248	304
	N 1	103	131
	N 2	76	100
	N 3	0	6
pathologic_M	M 0	315	486
	M 1	60	59
	M X	47	2
Tumor Stage	Stage I	70	37
	Stage II	162	263
	Stage III	124	205
	Stage IV	60	58

human gene SYMBOL using Bioconductor. Table 1 presents the statistical information of the preprocessed cohorts.

Calculation of the Scores of the TME-TILs

The deconvolution algorithm CIBERSORT uses a set of reference gene expression values that is considered a minimal representation for each cell type. Based on those values, the algorithm uses support vector regression to infer cell type proportions in data from bulk tumor samples with mixed cell types. CIBERSORT can be applied to distinguish 22 human immune cells including B cells, T cells, natural killer (NK) cells, macrophages, dendritic cells (DCs), and myeloid subset cells based on the high specificity and sensitivity of gene expression profiles. To quantify the proportion of immune cells in the CC samples, we used the CIBERSORT algorithm to calculate the 22 immune cell scores in the TCGA-COAD (colon adenocarcinoma) and GSE39582 cohorts by comparing them with the expression of the genes of the LM22 signature[18]. We uploaded the gene expression profile data to the CIBERSORT website (http://cibersort.stanford.edu/) and set up the expression of the LM22 signature genes with 1000 permutations to obtain the scores of the 22 immune cells.

Consensus Clustering to Obtain Molecular Subtypes Associated with the TME-TILs

Consensus clustering was performed using the *ConsensusClusterPlus* package to determine CC subgroups based on the TME-TILs[19]. We followed the methods of Zhang et al. to determine the optimal number of clusters (with k ranging

from 2 to 10)[20]. The procedure was repeated 1000 times to ensure the reproducibility of the results, which were visualized using the heat map function in R.

Differentially Expressed Genes Associated with the TME Phenotype

To identify genes associated with TME immune infiltrating patterns, we used linear models to analyze the gene expression differences between phenotype-related TME subgroups. More specifically, the limma package in R was applied to determine the differentially expressed genes (DEGs) with a false discovery rate (FDR) <0.05[21].

Re-Clustering of Phenotype-Related DEGs in the TME

Nonnegative matrix factorization (NMF) is an unsupervised clustering method that is widely used for determining tumor molecular subtypes based on genomics[22,23]. To further delve into the association between TME phenotypes and the expression of phenotype-related differentials, we re-clustered the samples using NMF to analyze the clinical characteristics based on the expression profiles of the phenotype-related DEGs. The standard "brunet" was selected when using NMF to perform 50 iterations. The number of clusters, k, was set between 2 and 10 using the NMF package in R[24]. The average contour width of the common member matrix was calculated, and the minimum number of members in each subclass was set to 10.

Dimension Reduction and Generation of TME Gene Signatures

To obtain robust TME gene signatures, we selected the DEGs with prognostic value and further evaluated the importance of these DEGs by applying the random forest algorithm. Specifically, we used the coxph function in the Survival package to conduct a univariate Cox analysis. The threshold was set at 0.05. The DEGs with prognostic value underwent random forest feature selection using the random Forest function in R. The mtry value for each segmentation was set from between 1 and 165, and the ntree value was 500. The mtry value with the lowest error rate was selected as the optimal value of the random forest algorithm. Subsequently, in accordance with the error rate of the random forest, ntree was reset to 100. Finally, each DEG was sorted based on its importance. The DEGs with a cumulative importance >95% were selected as candidate feature genes using k-means[25].

The risk coefficients of the signature scores of the categories of genes in each sample were obtained through multivariate regression. The TME score formula used for each sample is as follows:

$$\text{TMEScore} = \sum_{j=1}^{5} S_j * \beta_j$$

Here, j indicates the jth category, S_j indicates the signature score of the jth category of genes in the sample, and β_j indicates the risk regression coefficient of the signature score of the jth category of genes.

Association Between TME Scores and Clinical Features

To observe the association between TME scores and clinical phenotypes, the samples were divided into two groups based on the median TME score.

The prognosis differences between the high- and low-TME-score groups (high- and low-risk groups, respectively) were compared. Their relationship with age and sex was also analyzed.

Relationship Between TME Scores and Immune-Related Gene Expression

To examine the relationship between TME scores and immune-related gene expression, 3 immune-related gene types were collected: (1) immune activation genes, including *CXCL10*, *CXCL9*, *GZMA*, *GZMB*, *PRF1*, *IFNG*, *TBX2*, *TNF*, and *CD8A*; (2) immune checkpoint genes, including *PDCD1*, *CTLA4*, *LAG3*, *PDCD1LG2, IDO1*, *CD274*, and *HAVCR2*; and (3) TGF/EMT pathway genes, including *VIM*, *ACTA2*, *COL4A1*, *TGFBR2*, *ZEB1*, *CLDN3*, *SMAD9*, and *TWIST1*. The expression profiles of these genes were extracted to further analyze the aberrant expressions of the three gene types in the high and low TME score groups.

Correlation Between TME Scores and Genomic Mutations

To investigate the different genomic mutations in the high and low TME score samples, single-nucleotide polymorphism data were downloaded from TCGA. Intron and silent mutations were removed. The Fisher's exact test was used to analyze the differences in the mutations between these two samples. The significant difference threshold was set to $P < 0.05$.

Statistical Analysis

The Shapiro-Wilk normality test was used to determine the normality of the variables, unless otherwise stated[26]. For comparison between the two groups, the statistical significance of the normal distribution variables was estimated using the nonpaired Student's t test, and the non-normal distribution variables were analyzed using the Mann-Whitney U test. For comparisons between more than two groups, the Kruskal-Wallis test and one-way analysis of variance (ANOVA) were used as nonparametric and parametric methods, respectively[27]. The correlation coefficients were calculated using the Spearman correlation analysis. A two-sided Fisher's exact test was used to analyze the contingency table. The Benjamini-Hochberg method was used

to convert the *p*-value to the FDR. The Kaplan-Meier method was used to generate a survival curve for the subgroups in each data set, and the log rank test was used to determine the statistical significance at the level of $P < 0.05$. All analyses were performed using R 3.4.3 with the default parameters, unless otherwise specified.

Results

Calculation of TME Scores

Among the 22 TIL scores, the Spearman method is used to calculate the correlation coefficient between any two TILs, and the hierarchical clustering is used for cluster analysis, three clusters are obtained as shown in Fig. 1A, it can be seen that the three clusters have positive correlation within the class, and the correlation between classes is weak (Fig. 1A, Supplemental Table S1). Univariate Cox regression was used to analyze the relationship between the scores of the 22 TILs and prognosis. The scores of monocytes and naive T cells/CD4 cells were highly correlated with an unfavorable prognosis (log rank $P < 0.05$, hazard ratio [HR] >1). The scores of follicular helper T cells were correlated with a favorable prognosis (log rank $P < 0.05$, HR <1) (Fig. 1B, Supplemental Table S2).

Identification of Molecular Subtypes Based on the Scores of the TME-TILs

Based on the TME scores, *ConsensusClusterPlus* was used for unsupervised clustering of the TCGA samples. First, the scores of the three immune cells that were significantly correlated with prognosis were selected, and the optimal clustering number between $k = 2$ and $k = 10$ was evaluated. With 1000 repetitions of the above procedures, $k = 3$ was selected as the optimal clustering number according to the cumulative distribution function (CDF) value and delta area (Supplemental S1_FigureA-F). We defined TMEC1 to TMEC3 based on the three TME score clusters. According to the clustering results, immune cells like T cells. CD4 memory resting and activated NK cells achieved significant high scores in TMEC1. Immune cells such as M0 macrophages, NK cells resting were highly scored in TMEC2. Memory T cells/activated CD4, T cells CD8 scored highly in TMEC3 (Fig. 1C). We further analyzed the distribution differences of the scores of the 22 TILs in the three sample types; 13 (59.09%) had significant differences (Fig. 1D). The TME-TILs may be closely related to the occurrence and development of COAD. Survival analysis showed a significant difference between the TMECs in terms of progression free survival (PFS) (log rank $P = 0.0014$), as shown in Fig. 1E, with TMEC2 having the worst prognosis, TMEC3 having the best prognosis, and TMEC1 being in the middle.

We compared the pathway enrichment differences between the TME clusters (FDR < 0.1). There were no significant pathways between TMEC1 and TMEC2, and there were 28 significant pathways between TMEC3 and TMEC1, for example, NOD_LIKE_RECEPTOR_SIGNAL-ING_PATHWAY, T_CELL_RECEPTOR_SIGNALING_ PATHWAY, TOLL_LIKE_RECEPTOR_SIGNALING_ PATHWAY, APOPTOSIS, and JAK_STAT_SIGNALING_ PATHWAY (Supplemental S2_Figure A).

Two distinct pathways were found between TMEC3 and TMEC2: KEGG_PROTEASOME and KEGG_DNA_ REPLICATION (Supplemental S2_Figure B).

Identification of DEGs Among the TME Clusters

To study the differences in the gene expression patterns of each tumor microenvironment cluster (TMEC), DESeq2 was used to analyze the DEGs of the TMEC1, TMEC2, TMEC3, and normal samples. An FDR <0.05 and |Log2FC| >2 were selected. We selected 1300 DEGs that intersected the 3 groups for subsequent analysis (Fig. 2A, B). The DEGs are shown in Supplemental Table S3.

Clustering DEGs to Construct Gene Clusters in COAD

Based on the 1300 common DEGs, we first filtered 50% of the genes in the samples whose expression value was <1. We then conducted a univariate Cox analysis on the filtered 586 genes. We found 105 genes that were highly associated with prognosis ($P < 0.05$). We then used the 105 genes to conduct unsupervised clustering through NMF. The optimal clustering number was decided according to the cophenetic, dispersion, and silhouette indicators. We chose 4 as the optimal number (Fig. 2C) and defined them from GeneC1 to GeneC4. Survival analysis of the 4 gene clusters showed a significant difference in PFS (Fig. 2D). In comparing the scores of the 22 TILs in the GeneC groups, we found that GeneC1, which had the best prognosis, had significantly higher TIL scores for plasma cells and memory T cells/resting CD4 cells than the other GeneC groups (Fig. 2E).

Prognostic Risk Model Based on TME Scores

To further study the 105 DEGs with significant prognostic ability, we used *randomForest* in R to evaluate their importance. We selected ntree = 100 according to the *random Forest* plot (Supplemental S3_FigureA), and we identified 69 candidate genes (Supplemental S3_FigureB, S3_FigureC) by choosing DEGs whose accumulative importance was >95%. These genes were mainly enriched in the Gene Ontology (GO) terms relating to negative regulation of digestive system processes, as well as some Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Fig. 3A, B). There were four clusters after the k-means algorithm was used. They were defined as signature G1, signature G2, signature G3, and signature G4, which included 20, 10, 17, and 22 genes, respectively (Fig. 3C). G3 and G4 belonged to the low-expression group, G1 belonged to the high-expression
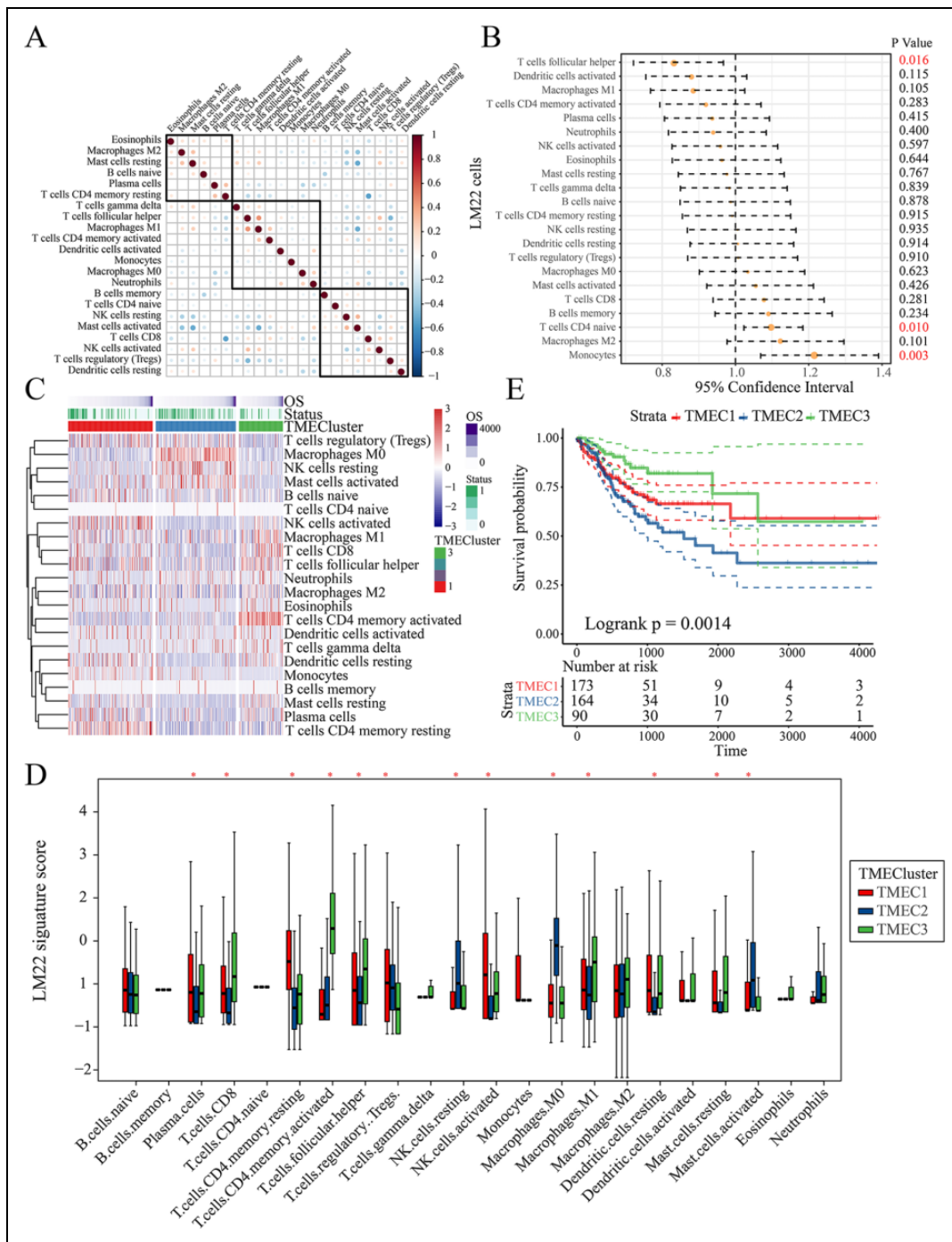
Figure 1. Identification of molecular subtypes based on the scores of the TME-TILs. (A) Correlations of the 22 TILs in the TME. The size and color of the dots represent correlation. Blue represents negative correlation, red represents positive correlation, and the blank area represents no significant correlation. The color bar represent represents the change trend of correlation, the redder the color, the stronger the positive correlation, the bluer the color, the stronger the negative correlation. (B) Forest map of the scores of the 22 immune cells in the TME. (C) Heat map of the scores of the 22 immune cells in the TME. Higher redness reflects higher scores, and lower blueness reflects lower scores. (D) Boxplot of the scores of the 22 immune cells in the three types of TMEC, with red * indicating significant differences. (E) PFS prognostic KM curve of the 3 TMEC types. The abscissa represents the survival time (days), and the ordinate represents the survival probability. TIL, tumor infiltrating lymphocytes; TMD, Tumor microenvironment; PFS, progression free survival.

group, and G2 was in middle (Fig. 3D). Principal component analysis was performed on G1, G2, and G3 using the psych R package. For each cluster, 100 iterations were performed to

obtain the optimal number of principal components (PCs). The respective PC scores were then calculated. Multivariate Cox analysis was used to establish the prognostic risk
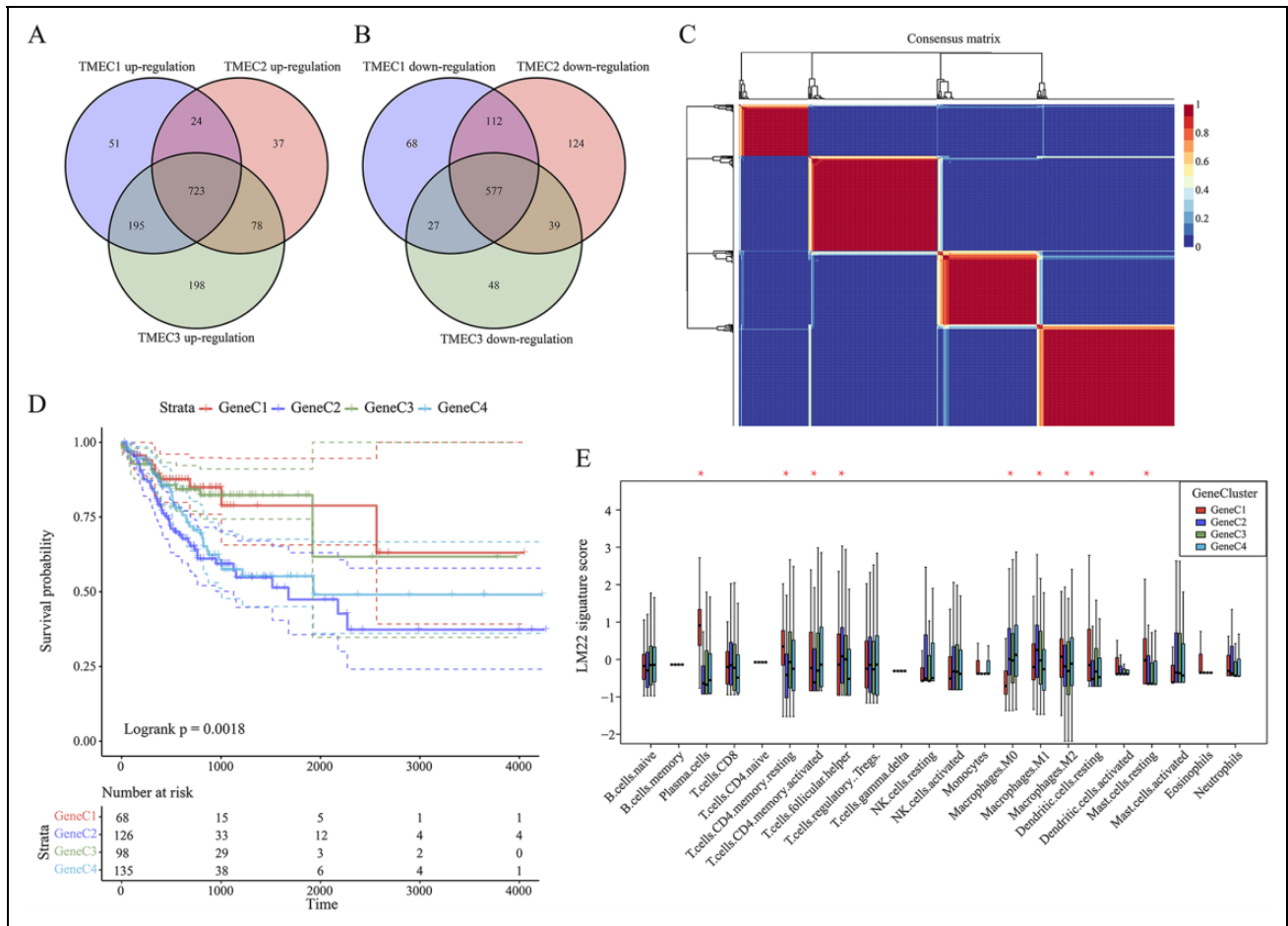
Figure 2. Identification of DEGs among the TME clusters. (A) Venn diagram of the up-regulated DEGs of TMEC1, TMEC2, and TMEC3. (B) Venn diagram of down-regulated DEGs of TMEC1, TMEC2, and TMEC3. (C) Heat map of the consistency matrix of the NMF algorithm. The color bar represent represents the change trend of correlation, the redder the color, the stronger the positive correlation, the bluer the color, the stronger the negative correlation. (D) KM curve of GeneC1, GeneC2, GeneC3, and GeneC4 in the PFS analysis, the abscissa represents the survival time (days), and the ordinate represents the survival probability. (E) Boxplot of the scores of the 22 immune cells in the GeneC1, GeneC2, GeneC3, and GeneC4 samples. The abscissa represents immune cells, and the ordinate represents immune score.* means *P*-value < 0.05. TME: Tumor microenvironment; PFS: progression free survival.

models of G1, G2, and G3, and the TME score of each sample was calculated.

We found that GeneC2 and GeneC4 had worse prognoses and much higher TME scores than GeneC1 and GeneC3, which had the best prognoses (Supplemental S4_FigureA, B). The median TME score (risk score, −0.057) was used to divide the samples into a high TME score (High Risk group) and low TME score group (Low Risk group). We found that the prognosis of the two groups were obviously different in terms of PFS (log rank *P* < 0.001, HR = 4.06) (Fig. 3E).

GSEA was used to identify the functional enrichment pathways of the high- and low-TME-score groups according to an FDR < 0.1 (Fig. 3F). All eight pathways were activated in the high-TME-score group, including BASAL_CELL_CARCINOMA, FOCAL_ADHESION, ECM_RECEPTOR_INTERACTION, PATHWAYS_IN_CANCER, and other

PATHWAYS associated with tumor proliferation and metastasis.

Relationship Between TME Scores and Clinical Features

There were significant differences in TME scores in terms of TNM and American Joint Committee on Cancer (AJCC) stages but not in terms of sex or age (Fig. 4A–F).

Univariate and Multivariate Analyses Based on the TME Scores and Clinical Features

Further analysis of the TME scores and clinical features by univariate and multivariate Cox analysis showed that N1,
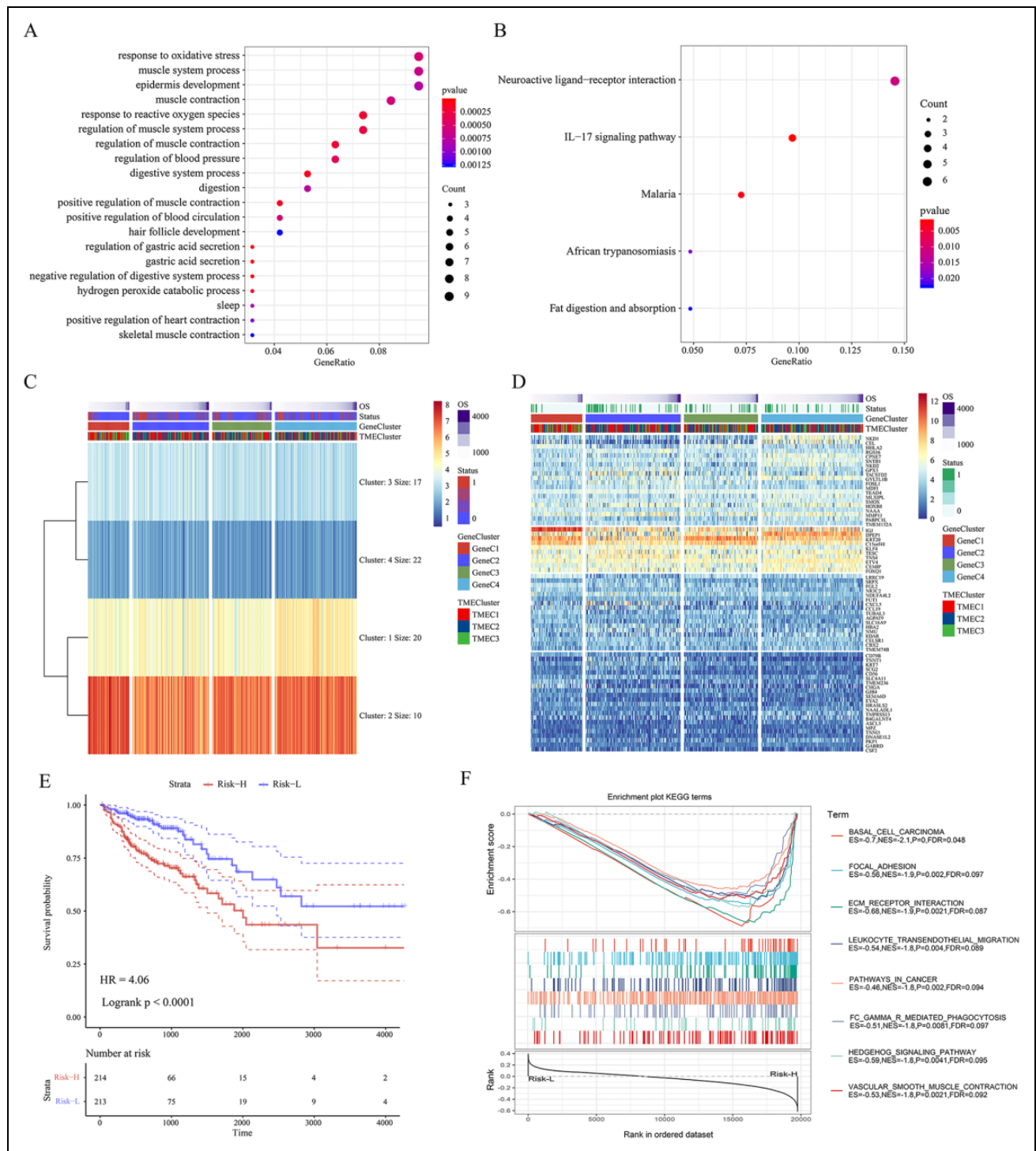
Figure 3. Construction of Prognostic risk model based on TME scores. (A) GO terms in the functional analysis of 69 genes. The abscissa represents the percentage of the number of genes in the functional term, the ordinate represents the enriched functional or pathway term, and the size of the circle represents the number of genes. The color bar represents *P*-value, and the redder the color, the more prominent it is. (B) KEGG pathway analysis of 69 genes. The abscissa represents the percentage of the number of genes in the functional term, the ordinate represents the enriched functional or pathway term, and the size of the circle represents the number of genes. The color bar represents *P*-value, and the redder the color, the more prominent it is. (C, D) Gene expression profiles of 69 genes in the different clusters. (E) KM curve of the Risk-H and Risk-L groups in the PFS analysis. The abscissa represents the survival time (days), and the ordinate represents the survival probability. (F) Differences in enrichment of KEGG pathway between high and low TMEscore groups. KEGG: Kyoto Encyclopedia of Genes and Genomes.
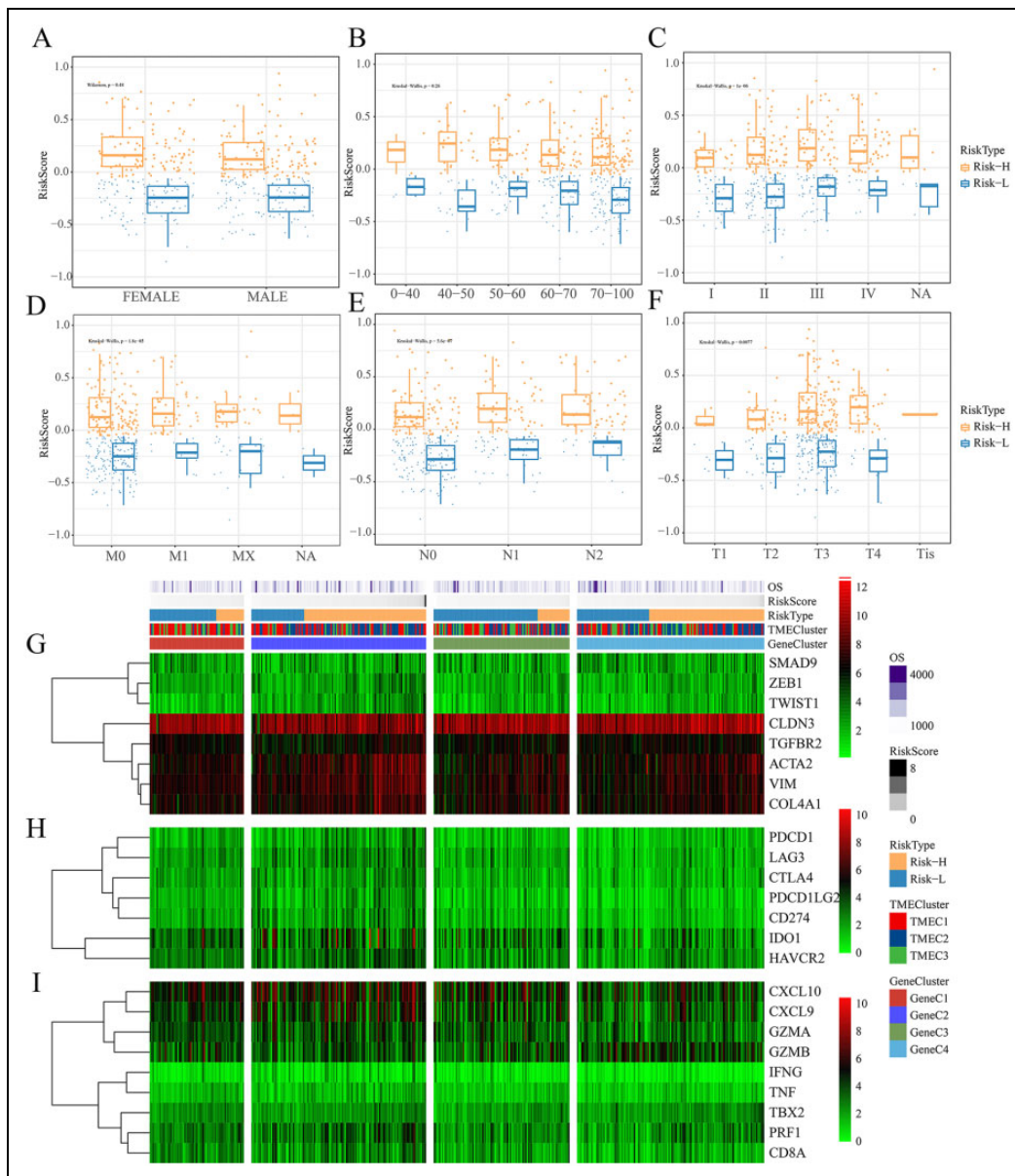
Figure 4. Relationship between TME scores and clinical features Relations between (A) gender and TME score; (B) age and TME score; (C) AJCC stage and TME score; (D) M stage and TME score; (E) N stage and TME score; and (F) T stage and TME score. The abscissa represents clinical variables, and the ordinate represents TMEscore. (G) Heat map of TGF pathway gene expression. (H) Heat map of immune checkpoint gene expression. (I) Heat map of immune activation gene expression. The color bar represent represents the change trend of correlation, the redder the color, the stronger the positive correlation.

M1, Stage III, age, and high TME scores were independent risk factors for colon cancer prognosis (Table 2).

Immune Gene Expression Patterns of the TMEs

We analyzed the expression of the TGF/EMT pathway genes (VIM, ACTA2, COL4A1, TGFBR2, ZEB1, CLDN3, SMAD9, and TWIST1) in the TMEC, GeneC, and high- and low-risk groups (Fig. 4G). The relationships between the expression of the immune checkpoint genes (PDCD1, CTLA4, LAG3, PDCD1LG2, IDO1, CD274, and HAVCR2)

and the TMEC, GeneC, and high- and low-risk groups were analyzed (Fig. 4H). The expression of the immune activation genes (CXCL10, CXCL9, GZMA, GZMB, PRF1, IFNG, TBX2, TNF, and CD8A) in the TMEC, GeneC, and high- and low-risk groups was analyzed (Fig. 4I).

Relationships Between the TME Scores and Immune Cells

We calculated the correlations between the TME scores and 22 immune cells (Table 3), and we observed that TME

Table 2. Univariate Analysis and Multivariate Analysis Based on the TME Score and Clinical Features.

Tag	Univariate cox analysis				Multivariate cox analysis			
	P-value	HR	Low 95% CI	High 95% CI	P-value	HR	Low 95% CI	High 95% CI
T2/T1	0.788434	0.740422	0.082446	6.64951	0.085299	0.065833	0.002969	1.459724
T3/T1	0.360699	2.512314	0.348494	18.11141	0.118638	0.08934	0.004301	1.855684
T4/T1	0.042337	8.024114	1.074698	59.91117	0.308283	0.204286	0.009623	4.336691
N1/N0	0.016565	1.873784	1.121118	3.131757	0.009559	0.245076	0.084607	0.709894
N2/N0	2.47E-10	4.529375	2.837032	7.231233	0.14687	0.467768	0.167579	1.305696
M1/M0	9.47E-12	4.904907	3.104477	7.749487	0.002335	163.7793	6.144636	4365.376
Stage II/ I	0.083332	2.881514	0.869774	9.546303	0.106173	13.42297	0.574952	313.3757
Stage III/ I	0.005062	5.455357	1.665799	17.86585	0.009409	77.0099	2.901863	2043.696
Stage IV/I	6.63E-06	15.28265	4.666812	50.04687	NA	NA	NA	NA
Age	0.200744	1.352589	0.851595	2.14832	0.001154	2.594517	1.460013	4.610589
Gender	0.41197	1.18459	0.790348	1.775487	0.370362	0.811356	0.513535	1.281897
TME Score	2.98E-06	2.718282	1.786924	4.135071	0.00036	2.013549	1.370871	2.957523

Table 3. Correlation Between TMEscore and Immune Cells.

Tag	R	P-value
B cells naive	0.199217973890225	0.000709875609957349
B cells memory	0.0348579878987621	1
Plasma cells	−0.111811114292333	0.354214610059808
T cells CD8	0.0942277711192144	0.775282449719287
T cells CD4 naive	0.0309884863300497	1
T cells CD4 memory resting	0.228067146791613	4.21996417212415e-05
T cells CD4 memory activated	0.143578280975776	0.0559182638855866
T cells follicular helper	0.0394924450297709	1
T cells regulatory (Tregs)	0.10499467562999	0.481023337835375
T cells gamma delta	−0.138423244849318	0.0748726574775479
NK cells resting	0.00221747458689814	1
NK cells activated	−0.0602205025871069	1
Monocytes	0.00129294704540079	1
Macrophages M0	0.159121437982563	0.0193721789134318
Macrophages M1	0.00163972460534322	1
Macrophages M2	−0.0532367111085018	1
Dendritic cells resting	−0.0710878531228252	1
Dendritic cells activated	−0.0914225309831241	0.827171858978039
Mast cells resting	−0.0068291170712964	1
Mast cells activated	0.00694292534679881	1
Eosinophils	−0.0735116628567726	1
Neutrophils	−0.0238206340464072	1

scores were significantly correlated with naïve B cells, resting memory CD4+ T cells, and M0 macrophages.

Relationship Between TME Scores and Tumor Genome Mutations

Using the TME scores to divide patients into high-risk (Risk-H) and low-risk (Risk-L) groups, we compared the relationships between TME scores and genome mutations and screened out a group of important TME score-related genes. We used the Fisher's exact test to compare the genes with significant differences in mutation frequency between the Risk-H and Risk-L groups (intron and silent mutations

were removed), and we obtained a total of 53 genes with a P-value < 0.01 (Fig. 5, Supplemental Table S4). The results showed that the mutation frequency of FAT3 in the Risk-H group was less than in the Risk-L group, which may indicate an important correlation between this gene and the TME.

TME Clusters and TME Scores Validated with the External Cohort

We downloaded the GSE39582 data set as a verification cohort. Using the same method, we divided the GSE39582 data set into three subgroups: TMEC1, TMEC2, and TMEC3. Comparing the characteristics of these three
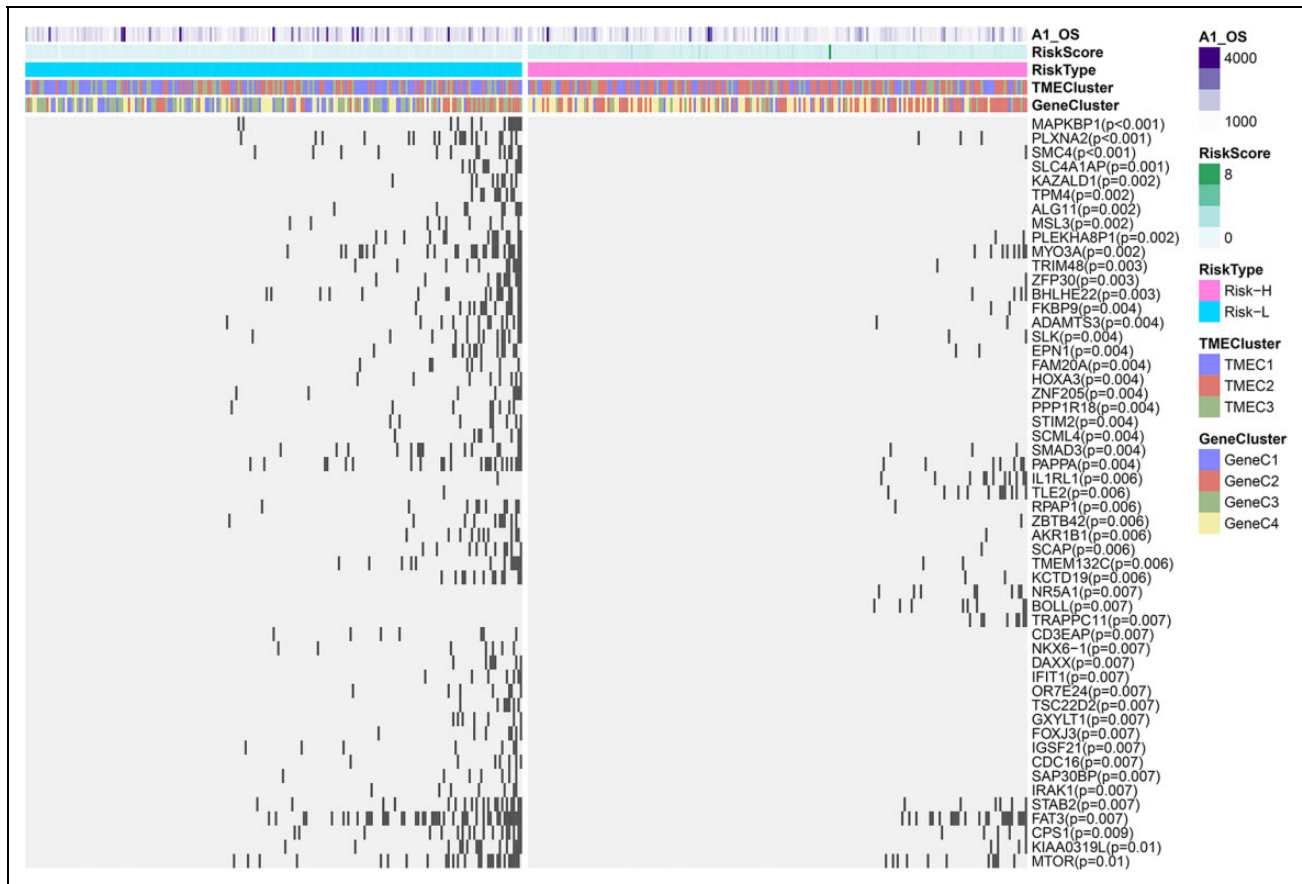
Figure 5. Relationship between TME scores and tumor genome mutations relationships between the TME and the mutation characteristics of the genomes. The horizontal axis reflects the samples, and the vertical axis reflects the genes. The black rectangle indicates mutations, and gray indicates 0 mutations. TME: Tumor microenvironment.

groups, we observed that resting memory CD4+ T cells had significantly higher scores in TMEC1, resting NK cells and activated mast cells had higher scores in TMEC2, and activated memory CD4+ T cells had higher scores in TMEC3 (Fig. 6A). The results of recurrence-free survival analysis among the TMEC groups showed that there were significant differences between the TMECs (Fig. 6B, log rank $P = 0.031$). TMEC2 had the worst prognosis, TMEC3 had the best prognosis, and TMEC1 was in the middle. These results are consistent with those of the TCGA cohort.

Furthermore, we calculated the TME scores of the patients in the validation set, and we used the same method to group the patients. The prognoses of the patients in the high-risk group were significantly worse than those of the patients in the low-risk group (Fig. 6C). These results are consistent with those of the TCGA cohort. The above results demonstrate the robustness of our model.

Analysis Flowchart

Protocol was designed to analyze the immune gene signature based on tumor microenvironment characteristics in colon cancer (Fig. 7).

Discussion

At present, the treatment mode of CC is comprehensive treatment based on surgery and chemotherapy[1]. Due to a lack of characteristic precursor symptoms of CC, most patients do not have the opportunity of surgical treatment by the time they are diagnosed. Chemotherapy drugs for patients with CC include oxaliplatin, fluorouracil, and other cytotoxic drugs. Although patients with CC in the early stage have a good response rate to chemotherapy, the application of chemotherapy is limited because of drug resistance[28]. Immunotherapy has become an important option for cancer treatment. Studies have shown that combined immunotherapy can improve the prognoses of patients. Immunotherapy is effective in patients with microsatellite unstable CC[29]. However, the incidence of microsatellite instability is only 15%–20% of all CRC[30]. Therefore, to fully realize the potential of cancer immunotherapy, a clear understanding of the characteristics of the immune microenvironment in tumors is essential to achieve the sustained clinical success of immunotherapy.

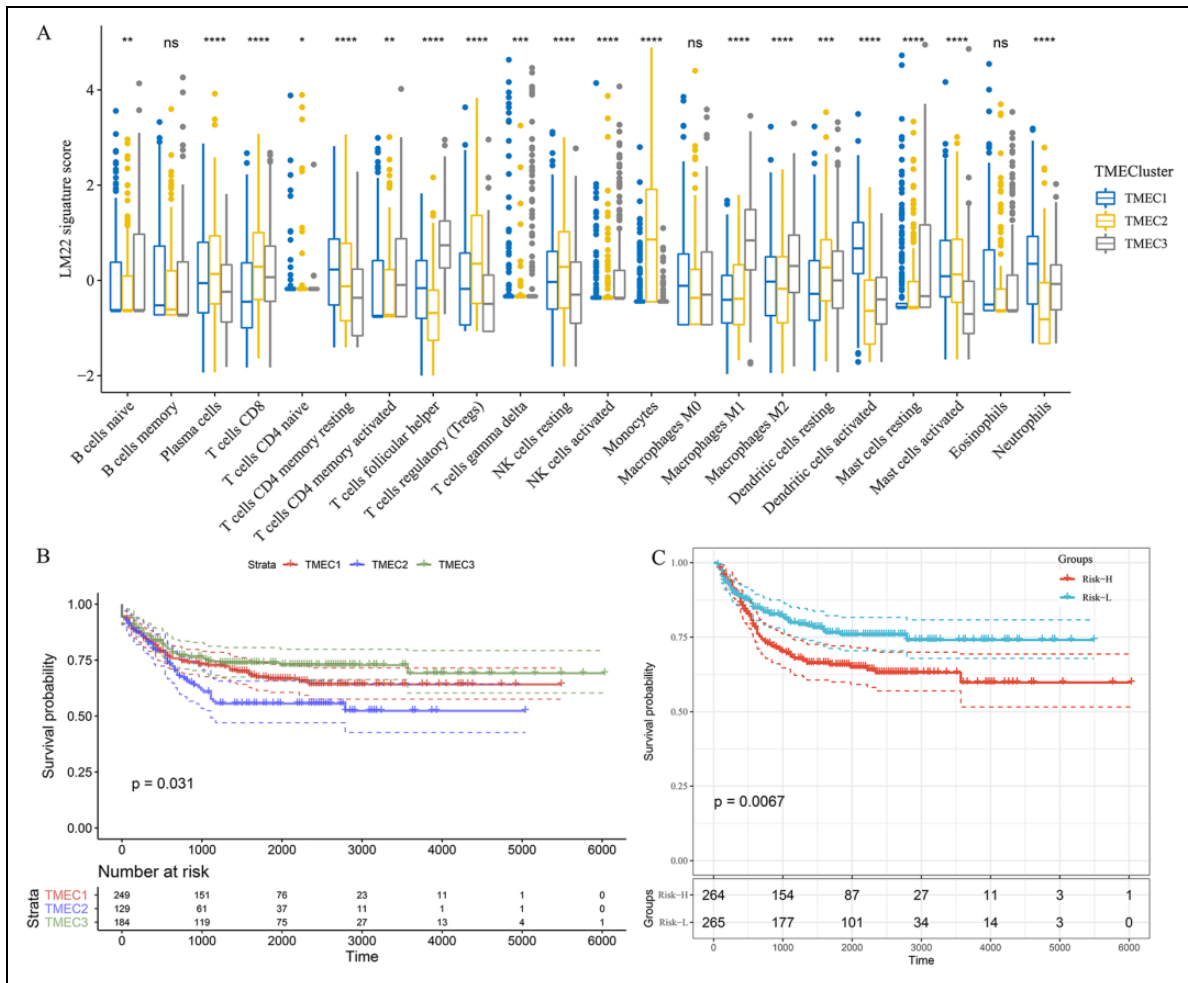The TME is very complicated. In addition to tumor cells, there are fibroblasts, endothelial cells, and immune cells[31].

Figure 6. TME clusters and TME scores validated with the external cohort. (A) Distribution of the immune cells in the different TME clusters in the validation cohort. The horizontal coordinates represent the different immune cells, and the vertical coordinates represent the different TME cluster immune scores. (B) Prognostic differences between the TME clusters in the validation set. The abscissa represents the survival time (days), and the ordinate represents the survival probability. (C) Differences in the outcomes between the TME scores, the abscissa represents the survival time (days), and the ordinate represents the survival probability. TME: Tumor microenvironment.

Immune cell infiltration is a common feature of most solid tumors. Immune tumor cell infiltration in tumors is associated with good prognosis[32]. In recent years, many studies have focused on the different immune cells in the immune microenvironment of CC, such as T cells, B cells, macrophages[33–35]. At present, there is a lack of overall and global analyses of the immune microenvironment of CC.

The TILs in the TME are complex and diverse, and their functions and effects on the prognoses of patients are different. Even the same type of immune cell can act counterintuitively in terms of the effect on prognosis[36]. For example, B lymphocyte infiltration in the TME is associated with a higher survival rate and lower recurrence rate in ovarian cancer, cervical cancer, and non-small cell lung cancer[37–39]. Paradoxically, B lymphocytes also have a role in promoting tumor progression. Some patients with advanced CRC lose B lymphocytes after treatment with rituximab, while the tumor load of patients decreases[40]. Therefore,

exploring the effect of immune cells in the immune microenvironment on the prognoses of patients may require an overall analysis.

In our research, three TME subtypes (TMEC1, TMEC2, and TMEC3) were identified based on tumor infiltrating lymphocyte (TIL) values calculated by CIBERSORT. Some immune cells, such as M0 macrophages, were highly scored in TMEC2. Resting dendritic cells scored low in TMEC2. CD8+ T cells scored high in TMEC3. The survival analysis results showed significant differences between the TMECs, with TMEC2 having the worst prognosis and TMEC3 having the best prognosis.

Many studies have shown that a variety of T lymphocytes, such as CD8+ T lymphocytes, macrophages, and dendritic cells, play an important role in colorectal cancer prognosis. The immune environment surrounding the tumor is associated with prognosis. Previous research has shown that higher levels of CD8+ T cells are usually associated with better
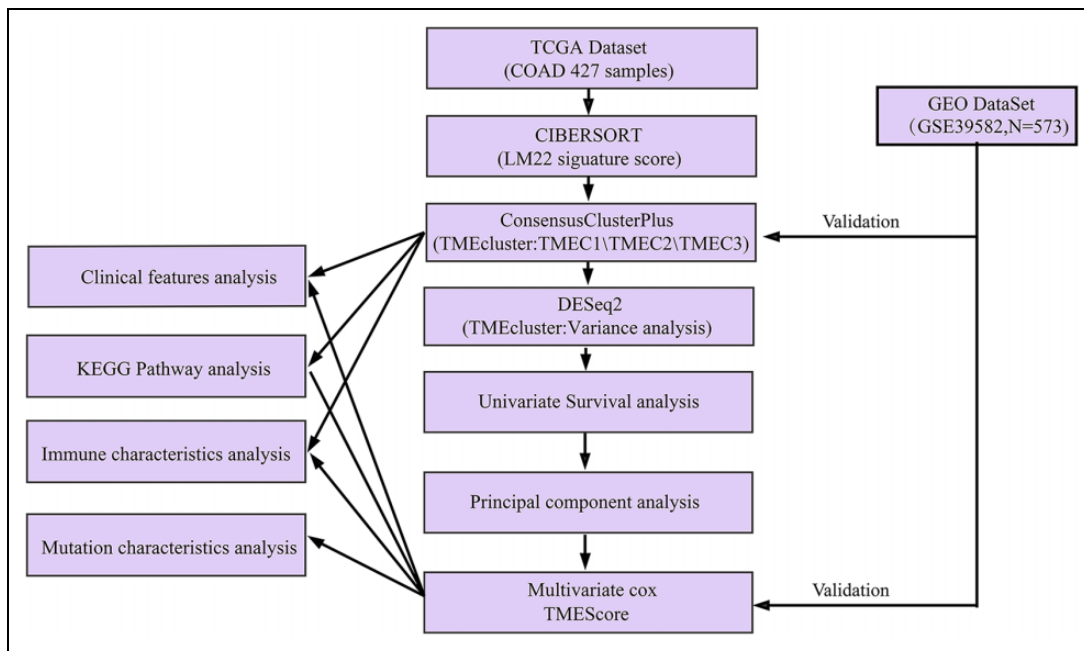
Figure 7. Analysis flowchart.

prognoses[41]. In addition, higher levels of macrophages usually mean worse prognoses[42], and lower levels of dendritic cells also mean worse prognoses[43]. Many factors affect clinical phenotypes, and immune cell infiltration is one of these indispensable aspects.

Further, we calculated the TME scores using a variety of algorithms, such as random forest, K-means clustering, and principal component analysis. The median TME score was used to divide the samples into a high-TME-score group (high-risk group) and low-TME-score group (low-risk group). The prognoses of the patients with high TME scores were worse than those of the patients with low TME scores. TME score had a significant correlation with TNM stage and immune gene expression, and a high TME score was an independent prognostic factor for patients with colon cancer. This indicates the potential guiding significance of clinical immunotherapy for patients with colon cancer.

We calculated the correlations between the TME scores and 22 immune cells. The results showed that TME scores were positively correlated with resting memory CD4+ T cells and M0 macrophages. Previous research has shown that higher CD4+ T cell and M0 macrophage infiltration levels are usually associated with worse prognoses[44,45]. This is consistent with the findings of our study, which showed that the patients with high TME scores had poor prognoses.

We identified the high/low TME score functional enrichment pathways using GSEA. These pathways were all activated in the high-TME-score group, including BASAL_CELL_CARCINOMA, FOCAL_ADHESION, ECM_RECEPTOR_INTERACTION, and PATHWAYS_IN_CANCER associated with tumor proliferation and metastasis.

Solid tumors are composed of interacting cancer cells and tumor microenvironments (TMEs), including the extracellular matrix, mesenchymal stem cells, endothelial cells, and immune cells, activating the ECM_RECEPTOR_INTERACTION pathway can cause changes in the composition of the TME and also affect tumor growth, migration, differentiation, and prognosis[46,47]. Immunotherapy has led to a change in the treatment of many advanced malignant tumors.

A key factor affecting the efficacy of immunotherapy is the TME, which contains a heterogeneous composition of immunosuppressive cells. Focal adhesion kinase (FAK), as a component of the FOCAL_ADHESION pathway, reduces the penetration of bone marrow-derived suppressor cells, tumor-associated macrophages, and regulatory T cells. In addition, FAK inhibitors have been implicated as modulators of matrix density and cancer stem cells, leading to a TME that is more conducive to anti-tumor immune responses[48].

Cancer is a genetic disease caused by mutations in the cellular genome. There is a close relationship between genome mutations and the tumor immune microenvironment[49]. We identified a group of tumor mutation genes related to the occurrence and development of the TME in CC, which indicates that there is a relatively important association between gene mutations and the TME.

Although this study analyzed the TME characteristics of CC in large cohorts, it still has some limitations. First, the population ethnicities in the TCGA database are mainly confined to White and Black people, and the extrapolation of the findings to other ethnic groups needs to be substantiated. Second, the conclusion of our research was based on a bioinformatics analysis, and it needs further validation through clinical specimens and prospective studies.

In this study, we identified the scores of 22 TILs based on TME subtypes, and a TME risk score signature was constructed and validated. TME score had a significant correlation with TNM stage, immune gene expression, and genomic mutations. A comprehensive landscape analysis of TME characteristics will provide a new strategy for immunotherapy in patients with CC.

Declaration of Conflicting Interests

Funding

ORCID iD

Jia Zhao https://orcid.org/0000-0003-0828-1081

Supplemental Material

Supplemental material for this article is available online.

References

1. Brenner H, Kloor M, Pox CP. Colorectal cancer. Lancet. 2014; 383(9927):1490–1502.

2. Kuipers EJ, Rösch T, Bretthauer M. Colorectal cancer screening–optimizing current strategies and new directions. Nat Rev Clin Oncol. 2013;10(3):130–142.

3. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J. Cancer statistics in China, 2015. CA Cancer J Clin. 2016;66(2):115–132.

4. Punt CJ, Koopman M, Vermeulen L. From tumour heterogeneity to advances in precision treatment of colorectal cancer. Nat Rev Clin Oncol. 2017;14(4):235–246.

5. Brown KGM, Solomon MJ. Progress and future direction in the management of advanced colorectal cancer. Br J surg. 2018; 105(6):615–617.

6. Lee RM, Cardona K, Russell MC. Historical perspective: two decades of progress in treating metastatic colorectal cancer. J Surg Oncol. 2019;119(5):549–563.

7. Hede K. Environmental protection: studies highlight importance of tumor microenvironment. J Nati Cancer Inst. 2004; 96(15):1120–1121.

8. Brown JM. Tumor microenvironment and the response to anticancer therapy. Cancer Biol Ther. 2002;1(5):453–458.

9. Kessenbrock K, Plaks V, Werb Z. Matrix metalloproteinases: regulators of the tumor microenvironment. Cell. 2010;141(1): 52–67.

10. Marvel D, Gabrilovich DI. Myeloid-derived suppressor cells in the tumor microenvironment: expect the unexpected. J Clin Invest. 2015;125(9):3356–3364.

11. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, Coussens LM, Gabrilovich DI, Ostrand-Rosenberg S, Hedrick CC, Vonderheide RH, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med. 2018;24(5):541–550.

12. Lasry A, Zinger A, Ben-Neriah Y. Inflammatory networks underlying colorectal cancer. Nat Immunol. 2016;17(3): 230–240.

13. Burn J, Gerdes AM, Macrae F, Mecklin JP, Moeslein G, Olschwang S, Eccles D, Evans DG, Maher ER, Bertario L, Bisgaard ML, et al. Long-term effect of aspirin on cancer risk in carriers of hereditary colorectal cancer: an analysis from the CAPP2 randomised controlled trial. Lancet. 2011;378(9809): 2081–2087.

14. Grivennikov SI, Wang K, Mucida D, Stewart CA, Schnabl B, Jauch D, Taniguchi K, Yu GY, Osterreicher CH, Hung KE, Datz C, et al. Adenoma-linked barrier defects and microbial products drive IL-23/IL-17-mediated tumour growth. Nature 2012, 491(7423):254–258.

15. Fridman WH, Pagès F, Sautès-Fridman C, Galon J. The immune contexture in human tumours: impact on clinical outcome. Nat Rev Cancer. 2012;12(4):298–306.

16. Itatani Y, Kawada K, Inamoto S, Yamamoto T, Ogawa R, Taketo MM, Sakai Y. The Role of chemokines in promoting colorectal cancer invasion/metastasis. Int J Mol Sci. 2016; 17(5):643.

17. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. Nat Med. 2013;19(11):1423–1437.

18. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12(5):453–457. doi:10.1038/nmeth.3337 (2015)

19. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, Van De Rijn M, Jeffrey SS, Thorsen T, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA. 2001;98(19):10869–10874. doi:10.1073/pnas.191367098

20. Zhang S, Wang Y, Gu Y, Zhu J, Ci C, Guo Z, Chen C, Wei Y, Lv W, Liu H, Zhang D., et al. Specific breast cancer prognosissubtype distinctions based on DNA methylation patterns. Mol Oncol. 2018;12(7):1047–1060. doi:10.1002/1878-0261.12309.

21. Ritchie ME, Phipson BA, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. doi:10.1093/nar/gkv007

22. Mirzal A. Nonparametric tikhonov regularized nmf and its application in cancer clustering. IEEE/ACM Trans Comput Biol Bioinform. 2014;11(6):1208–1217. doi:10.1109/TCBB. 2014.2328342

23. Yu, N. Gao YL, Liu JX, Shang J, Zhu R, Dai LY. Co-differential Gene selection and clustering based on graph regularized multi- view NMF in cancer Genomic Data. Genes (Basel). 2018;9(12):586. doi:10.3390/genes9120586

24. Ye C, Toyoda K., Ohtsuki T. Blind source separation on noncontact heartbeat detection by non-negative matrix factorization algorithms. IEEE Trans Biomed Eng. 2019;67(2): 482–494. doi:10.1109/TBME.2019.2915762

25. Boillaud E, Molina G. Are judgments a form of data clustering? Reexamining contrast effects with the k-means algorithm.

J Exp Psychol Hum Percept Perform. 2015;41(2):415–430. doi:10.1037/a0038896

26. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. Int J Endocrinol Metab. 2012; 10(2):486–489. doi:10.5812/ijem.3505

27. Hazra A., Gogtay N. Biostatistics series module 3: comparing groups: numerical variables. Indian J Dermatol. 2016;61(3): 251–260. doi:10.4103/0019-5154.182416

28. Dienstmann R, Salazar R, Tabernero J. Personalizing colon cancer adjuvant therapy: selecting optimal treatments for individual patients. J Clin Oncol. 2015;33(16):1787–1796.

29. Yaghoubi N, Soltani A, Ghazvini K, Hassanian SM, Hashemy SI. PD-1/ PD-L1 blockade as a novel treatment for colorectal cancer. Biomed Pharmacother. 2019;110:312–318.

30. Kim JH, Kang GH. Molecular and prognostic heterogeneity of microsatellite-unstable colorectal cancer. World J Gastroenterol 2014;20(15):4230–4243.

31. Galon J, Fridman WH, Pagès F. The adaptive immunologic microenvironment in colorectal cancer: a novel perspective. Cancer Res. 2007;67(5):1883–1886.

32. Bethmann D, Feng Z, Fox BA. Immunoprofiling as a predictor of patient's response to cancer therapy-promises and challenges. Curr Opin Immunol. 2017;45:60–72.

33. Nagarsheth N, Peng D, Kryczek I, Wu K, Li W, Zhao E, Zhao L, Wei S, Frankel T, Vatan L, Szeliga W, et al. PRC2 Epigenetically silences Th1-Type chemokines to suppress effector T-Cell trafficking in colon cancer. Cancer Res 2016;76(2): 275–282.

34. Yuan L, Tian J. LIN28B promotes the progression of colon cancer by increasing B-cell lymphoma 2 expression. Biomed Pharmacother. 2018;103:355–361.

35. Fang M, Li Y, Huang K, Qi S, Zhang J, Zgodzinski W, Majewski M, Wallner G, Gozdz S, Macek P, Kowalik A, et al. IL33 Promotes colon cancer cell stemness via jnk activation and macrophage recruitment. Cancer Res 2017;77(10): 2735–2745.

36. Gnjatic S, Bronte V, Brunet LR, Butler MO, Disis ML, Galon J, Hakansson LG, Hanks BA, Karanikas V, Khleif SN, Kirkwood JM, et al. Identifying baseline immune-related biomarkers to predict clinical outcome of immunotherapy. J Immunother Cancer. 2017;5:44.

37. Nielsen JS, Sahota RA, Milne K, Kost SE, Nesslinger NJ, Watson PH, Nelson BH. CD20+ tumor-infiltrating lymphocytes have an atypical CD27- memory phenotype and together with CD8+ T cells promote favorable prognosis in ovarian cancer. Clin Cancer Res. 2012;18(12):3281–3292.

38. Al-Shibli KI, Donnem T, Al-Saad S, Persson M, Bremnes RM, Busund LT. Prognostic effect of epithelial and stromal lymphocyte infiltration in non-small cell lung cancer. Clin Cancer Res. 2008;14(16):5220–5227.

39. Nedergaard BS, Ladekarl M, Nyengaard JR, Nielsen K. A comparative study of the cellular immune response in patients with stage IB cervical squamous cell carcinoma. low numbers of several immune cell subtypes are strongly associated with relapse of disease within 5 years. Gynecol Oncol. 2008;108(1):106–111.

40. Barbera-Guillem E, Nelson MB, Barr B, Nyhus JK, May KF, Feng L, Sampsel JW. B lymphocyte pathology in human colorectal cancer: experimental and clinical therapeutic effects of partial B cell depletion. Cancer Immunol Immunother. 2000; 48(10):541–549.

41. Tosolini M, Kirilovsky A, Mlecnik B, Fredriksen T, Mauger S, Bindea G, Berger A, Bruneval P, Fridman WH, Pagès F, Galon J. Clinical impact of different classes of infiltrating T cytotoxic and helper cells (Th1, th2, treg, th17) in patients with colorectal cancer. Cancer Res. 2011;71(4):1263–1271. doi:10.1158/0008-5472.Can-10-2907

42. Tewari M, Sahai S, Mishra RR, Shukla SK, Shukla HS. Dendritic cell therapy in advanced gastric cancer: a promising new hope? Surg Oncol. 2012;21(3):164–171. doi:10.1016/j.suronc.2012.03.003

43. Ding T, Xu J, Wang F, Shi M, Zhang Y, Li SP, Zheng L. High tumor-infiltrating macrophage density predicts poor prognosis in patients with primary hepatocellular carcinoma after resection. Hum Pathol. 2009;40(3):381–389. doi:10.1016/j.humpath.2008.08.011

44. Ohue Y, Nishikawa H.Regulatory T. (Treg) cells in cancer: can treg cells be a new therapeutic target? Cancer Sci. 2019;110(7): 2080–2089. doi:10.1111/cas.14069

45. Ahmadzadeh M, Pasetto A, Jia L, Deniger DC, Stevanović S, Robbins PF, Rosenberg SA. Tumor-infiltrating human CD4(+) regulatory T cells display a distinct TCR repertoire and exhibit tumor and neoantigen reactivity. Sci Immunol. 2019;4(31): eaao4310. doi:10.1126/sciimmunol.aao4310

46. Najafi M, Farhood B, Mortezaee K. Extracellular matrix (ECM) stiffness and degradation as cancer drivers. J Cell Biochem. 2019;120(3):2782–2790. doi:10.1002/jcb.27681. Epub 2018 Oct 15. PMID: 30321449.

47. Wang L, Zuo X, Xie K, Wei D. The Role of CD44 and cancer stem cells. Methods Mol Biol. 2018;1692:31–42. doi:10.1007/978-1-4939-7401-6_3 PMID: 28986884

48. Osipov A, Saung MT, Zheng L, Murphy AG. Small molecule immunomodulation: the tumor microenvironment and overcoming immune escape. J Immunother Cancer. 2019;7(1): 224. doi:10.1186/s40425-019-0667-0. PMID: 31439034; PMCID: PMC6704558

49. Shimizu T, Marusawa H, Endo Y, Chiba T. Inflammation-mediated genomic instability: roles of activation-induced cytidine deaminase in carcinogenesis. Cancer Sci. 2012; 103(7):1201–1206.