

# Tumor microenvironment characterization in esophageal cancer identifies prognostic relevant immune cell subtypes and gene signatures

Yuhong Zhang<sup>1</sup>, Minqi Zhu<sup>2</sup>, Junxian Mo<sup>2</sup>, Lei Xian<sup>3</sup>

<sup>1</sup>Department of Gastroenterology, The First Affiliated Hospital of Guangxi Medical University, Nanning, China

<sup>2</sup>Guangxi Medical University, Nanning, China

<sup>3</sup>Department of Cardiothoracic Surgery, The Second Affiliated Hospital of Guangxi Medical University, Nanning, China

**Correspondence to:** Lei Xian; email: [xianlei59@163.com](mailto:xianlei59@163.com), <https://orcid.org/0000-0003-3386-158X>

**Keywords:** esophageal cancer, tumor microenvironment, prognostic model, immune cell subtypes, bioinformatics

**Received:** June 1, 2021

**Accepted:** December 8, 2021

**Published:** December 26, 2021

**Copyright:** © 2021 Zhang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Esophageal cancer (ESCA) is a common malignancy in the digestive system with a high mortality rate and poor prognosis. Tumor microenvironment (TME) plays an important role in the tumorigenesis, progression and therapy resistance of ESCA, whereas its role in predicting clinical outcomes has not been fully elucidated. In this study, we comprehensively estimated the TME infiltration patterns of 164 ESCA patients using Gene Set Variation Analysis (GSVA) and identified 4 key immune cells (natural killer T cell, immature B cell, natural killer cell, and type 1 T helper cell) associated with the prognosis of ESCA patients. Besides, two TME groups were defined based on the TME patterns with different clinical outcomes. According to the expression gene set between two TME groups, we built a model to calculate TMEScore based on the single-sample gene-set enrichment analysis (ssGSEA) algorithm. TMEScore systematically correlated the TME groups with genomic characteristics and clinicopathologic features. In conclusion, our data provide a novel TMEScore which can be regarded as a reliable index for predicting the clinical outcomes of ESCA.

## INTRODUCTION

Esophageal cancer (ESCA) is a common malignancy in the digestive system with an extremely aggressive nature and poor prognosis [1]. It is estimated that ESCA is the 8<sup>th</sup> most common cancer and the 6<sup>th</sup> leading cause of cancer-specific death [2, 3]. Although esophagectomy is still considered the cornerstone of curative treatment for locally advanced esophageal cancer, it remains associated with considerable postoperative morbidity, mortality, and recurrence rates [4]. And the eligibility of a patient for surgical resection strongly depends on the extent of the disease, as well as on the general condition of the patient [5]. Nowadays, despite recent improvements in early diagnosis and therapeutic strategy, the overall survival (OS) of

esophageal cancer patients remains lower than most solid tumors [6, 7].

As the novel therapeutic options, immune checkpoint inhibitors have been applied and revealed encouraging efficacy in ESCA [8]. Additionally, other immunogenic approaches such as adoptive T-cell therapy and peptide vaccines also exhibited promising curative effects [9, 10]. Due to the ideal results of targeting the immune microenvironment in the treatment of esophageal cancer, tumor microenvironment (TME) may play a crucial role in the tumorigenesis and progression of ESCA.

TME cells often take part in the occurrence and development of esophageal cancer.

The TME includes a complex collection of components, such as stromal cells with immunosuppressive features and several anti-tumor components, for example cytotoxic T lymphocytes, T helper type1 cells, and natural killer cells [11]. And some cells display both pro- and anti-tumor effects, during different tumor stages or in different interaction with other TME components [12]. A dynamic balance exists between the pro- and anti-tumor factors within the TME, which profoundly influences the prognosis of patients with cancer [13].

In order to identify the TME features of ESCA, RNA sequencing data and clinical information of ESCA samples were collected from The Cancer Genome Atlas (TCGA) databases. Additionally, we also identified the TME-associated differentially expressed genes (DEGs) which were related to prognosis. Based on them, the TME score was explored. Thus, our findings provide prognostic TME score and potential biomarkers which may assist oncologists in prognosis prediction.

## **MATERIALS AND METHODS**

### **Collected esophageal cancer dataset**

The RNA sequencing data (FPKM value) of gene expression in TCGA dataset were downloaded from the Genomic Data Commons (<https://portal.gdc.cancer.gov/>) using the R package TCGAbiolinks [14]. Then, FPKM values were transformed into transcripts per kilobase million (TPM) values. Data were analyzed with the R (version 3.6.1) and R Bioconductor packages.

### **Collection of clinical data of esophageal cancer dataset**

The clinical data and sample information for TCGA-ESCA and other TCGA cancer cohorts samples were obtained from the Genomic Data Commons (<https://portal.gdc.cancer.gov/>) using the R package TCGAbiolinks [14]. Overall survival (OS) information of all TCGA dataset was obtained from the supplementary data of published research. Somatic mutation data for ESCA patients were obtained from TCGA database (<https://portal.gdc.cancer.gov/projects/TCGA-ESCA>).

### **Calculation of infiltrating cells in the TME**

We used the Gene Set Variation Analysis (GSVA) [15] algorithm to analyze the gene signatures of the 28 TME cells based on the supplementary data from Jia, Q, et al. [16]. This algorithm allows for calculating the enrichment score of 28 human immune cell phenotypes,

namely, activated B cell, immature B cell, mast cell, regulatory T cell, MDSC, effector memory CD8 T cell, central memory CD4 T cell, activated dendritic cell, macrophage, type 1 T helper cell, natural killer T cell, T follicular helper cell, natural killer cell, type 2 T helper cell, effector memory CD4 T cell, CD56 bright natural killer cell, gamma delta T cell, plasmacytoid dendritic cell, activated CD4 T cell, activated CD8 T cell, neutrophil, eosinophil, CD56dim natural killer cell, immature dendritic cell, central memory CD8 T cell, type 17 T helper cell, memory B cell and monocyte.

### **Unsupervised clustering using TME infiltrating cells matrix**

Tumors with qualitatively different TME cell infiltration patterns were grouped using hierarchical agglomerative clustering (based on canberra distance and Ward's linkage (ward.D method)). Unsupervised clustering methods for dataset analysis were used to identify TME patterns and classify patients for further analysis. A consensus clustering algorithm was applied to determine the number of clusters to assess the stability of the discovered clusters. This procedure was performed using the ConsensusClusterPlus R package and was repeated 1,000 times to ensure the stability of classification [17].

### **DEGs calculation and dimension reduction of different TME groups**

To identify genes associated with TME cell enrichment score patterns, we grouped patients into TME groups based on immune-cell infiltration. DEGs among these groups were determined using the R package limma [18], which implements an empirical Bayesian approach to estimate gene-expression changes using moderated *t*-tests. DEGs among TME subgroups were determined by significance criteria (adjusted *P* value <0.05) as implemented in the R package limma. The adjusted *P* value for multiple testing was calculated using the Benjamini-Hochberg correction. An unsupervised clustering method (K-means) for analysis of DEGs was used to classify patients into several groups for further analysis. Then, the random forest classification algorithm was used to perform dimension reduction in order to reduce noise. Next, the clusterProfiler R package was adopted to annotate gene patterns [19]. After that, two TME signature gene sets (TME metagene 1 and TME metagene 2) were obtained to build the model of TME score.

### **Calculation of TME score**

GSVA was performed to calculate the signature score of different TME signature gene sets using the ssGSEA

algorithm [15]. TMEscoreA and TMEscoreB represented the GSVA score calculated from different TME metagene 1 and TME metagene 2.

$$\text{TMEscoreA} = \text{GSVA score of TME metagene 1}$$
$$\text{TMEscoreB} = \text{GSVA score of TME metagene 2}$$

The prognosis of the patients with higher TMEscoreA was poor, and patients with lower TMEscoreB was poor. The stratification of prognosis for TMEscoreA and TMEscoreB is significant. After obtaining the prognostic value of each gene signature score, we applied a method to define the TMEscore of each patient:

$$\text{TMEscore} = \text{TMEscoreB} - \text{TMEscoreA}$$

### Functional enrichment analysis

The clusterProfiler R package was performed to demonstrate functional enrichment analysis on TME signature genes to. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) terms were identified with a strict cutoff of  $P < 0.05$  [20]. We also identified functional pathways that were up and down regulated among TMEscore high and low by running a gene set enrichment analysis (GSEA) of the adjusted expression data for all transcripts. Enrichment  $P$  values were based on 1,000 permutations and subsequently adjusted for multiple testing using the Benjamini-Hochberg procedure to control the FDR. A developing R package enrichplot (<https://github.com/GuangchuangYu/enrichplot>), implements several visualization methods to help interpreting enrichment results and is adopted to visualize GSEA result of TME gene groups. The hub genes were analyzed via STRING database (<https://string-db.org/>).

### Analysis of relevant biological processes using TME signatures

To explore the correlation between the TME signature and other relevant biological processes, we used gene sets curated by Mariathasan and colleagues, including (i) CD8 T-effector signature; (ii) antigen processing machinery; (iii) immune-checkpoint; (iv) epithelial-mesenchymal transition (EMT) markers previously reported (EMT1, EMT2, EMT3); (v) pan-fibroblast TGFb response signature (Pan-F-TBRS); (vi) Angiogenesis signature previously reported; (vii) Fanconi anemia; (viii) cell cycle genes (KEGG); (ix) DNA replication (KEGG); (x) nucleotide excision repair (KEGG); (xi) DNA damage repair (KEGG); (xii) Homologous recombination (KEGG); (xiii) mismatch repair (KEGG); (xiv) WNT targets; (xv)

Cell cycle regulators (KEGG); (xvi) DNA damage repair (KEGG).

### Statistical analysis

Wilcoxon rank-sum test was used for the comparisons between two groups. The correlation coefficients were computed by Spearman and distance correlation analyses. Two-sided Fisher exact tests were used to analyze the contingency tables. The cutoff values of each dataset were evaluated based on the association between the OS of patient and TMEscore in each separate dataset using the survminer package. The R package MaxStat, which iteratively tests all possible cut points to find the one achieving the maximum rank statistic, was used to dichotomize TMEscore, and patients were then divided into low and high group. R package forestplot was used for the presentation of the results in the subgroup analysis of TMEscore in TCGA esophageal cancer dataset. To identify significant genes in the differential gene analysis, we applied the Benjamini-Hochberg method to convert the  $P$  values to FDRs. The Kaplan-Meier method was used to generate survival curves for the subgroups in each dataset, and the log-rank (Mantel-Cox) test was used to determine the statistical significance of differences. The hazard ratios for univariate analyses were calculated using a univariate Cox proportional hazards regression model. A multivariate Cox regression model was used to determine independent prognostic factors using the survminer package. All heat maps were generated by the function of heatmap (<https://github.com/raivokolde/pheatmap>). The  $P$  values were two-sided.  $P$  values of less than 0.05 were considered statistically significant.

### Data availability statement

The data used to support the findings of this study are all public and accessible at TCGA database.

## RESULTS

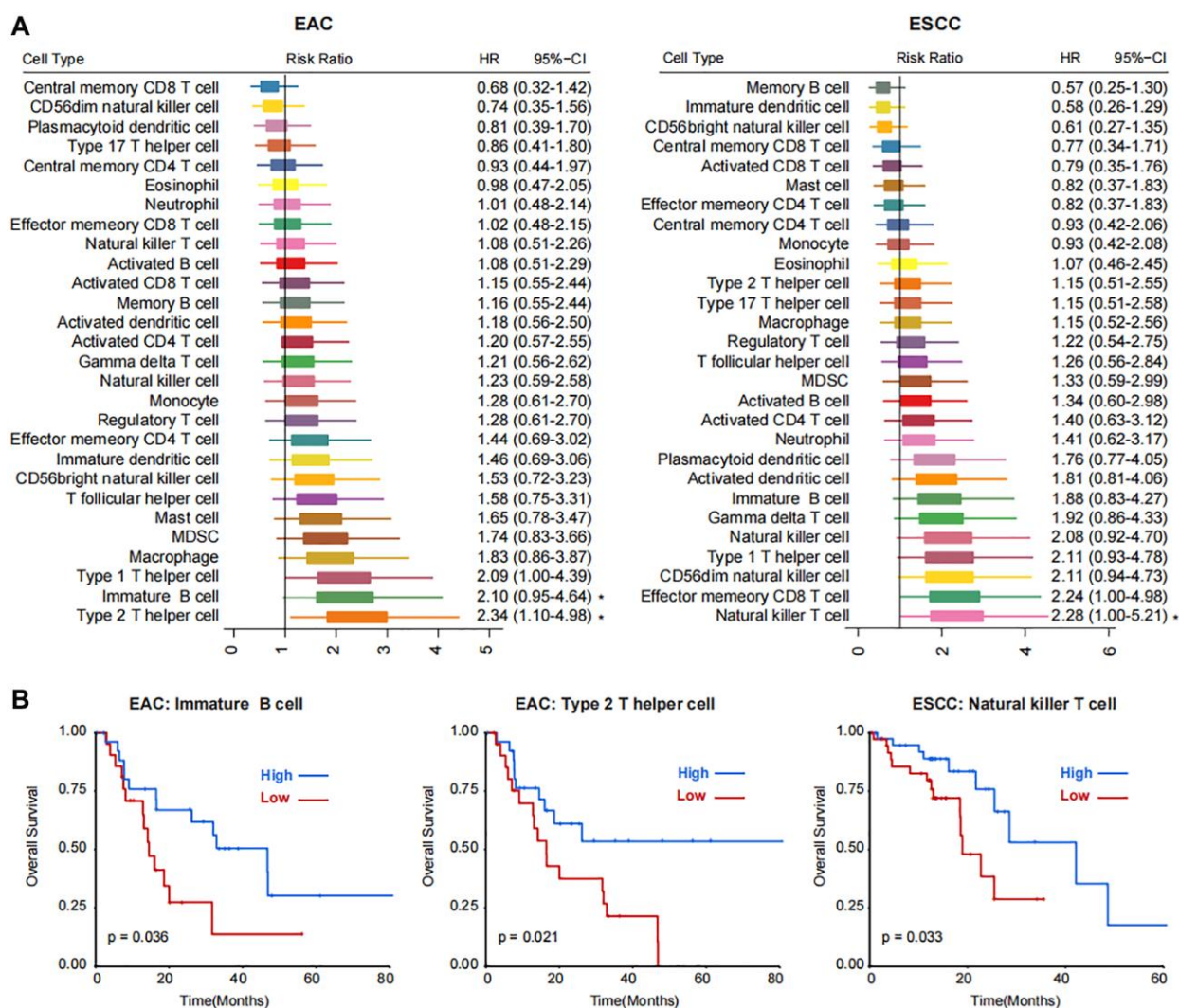
A total of 173 cases with available overall survival data were obtained from TCGA database via the association number TCGA-ESCA (Supplementary Figure 1). After removing the duplicate cases, 164 cases remained for further analysis. Based on the GSVA algorithm, 28 kinds of TME cells were calculated. The ESCA in TCGA database was divided into two subtypes, esophageal adenocarcinoma (EAC,  $n = 49$ ) and esophageal squamous-cell carcinoma (ESCC,  $n = 78$ ). The activated B cell was differentially expressed in the TME cell components between EAC and ESCC samples (Supplementary Figure 2). We next analyzed the clinical correlation of different TME cells in EAC

and ESCC subtypes. In EAC, immature B cell and type 2 T helper cell played the main roles (Figure 1A). While in ESCC, natural killer T cells might be crucial (Figure 1A). Patients with higher enrichment score in immature B cell or type 2 T helper cell might have better clinical outcomes in EAC ( $P < 0.05$ , Figure 1B). In ESCC, the clinical outcome of patients with higher natural killer T cell was favorable ( $P < 0.05$ , Figure 1B). In the comprehensive analysis of all patients with ESCA, four kinds of TME cells, i.e. natural killer T cell, immature B cell, natural killer cell, and type 1 T helper cell, were significantly correlated with prognosis (Supplementary Figure 3).

Due to little difference between TME cells in ESC and EACC, unsupervised clustering was applied in the 164 patients and two TME groups were identified

(TMEgroup1 and TMEgroup2, Figure 2A). Subsequently, we tested four cluster number parameters in the analysis from  $k = 2$  to  $k = 5$  and the results of Consensus Cluster Plus revealed that 2 was the best cluster number (Supplementary Figure 4). By comparing the two TME groups, we found the higher enrichment score of 25 kinds of TME cells in TMEgroup1 (Figure 2A and Supplementary Figure 5). Additionally, the clinical outcome of TMEgroup1 was relatively favorable ( $P = 0.03$ , Figure 2B). The prognostic stratification based on TME cluster was not affected with esophageal cancer subtypes ( $P = 0.078$ , Figure 2C).

In order to analyze the difference between TMEgroup1 and TMEgroup2, we identified the DEGs between the two TME groups. Then, the correlation between the DEGs and clinical prognosis were explored and a total

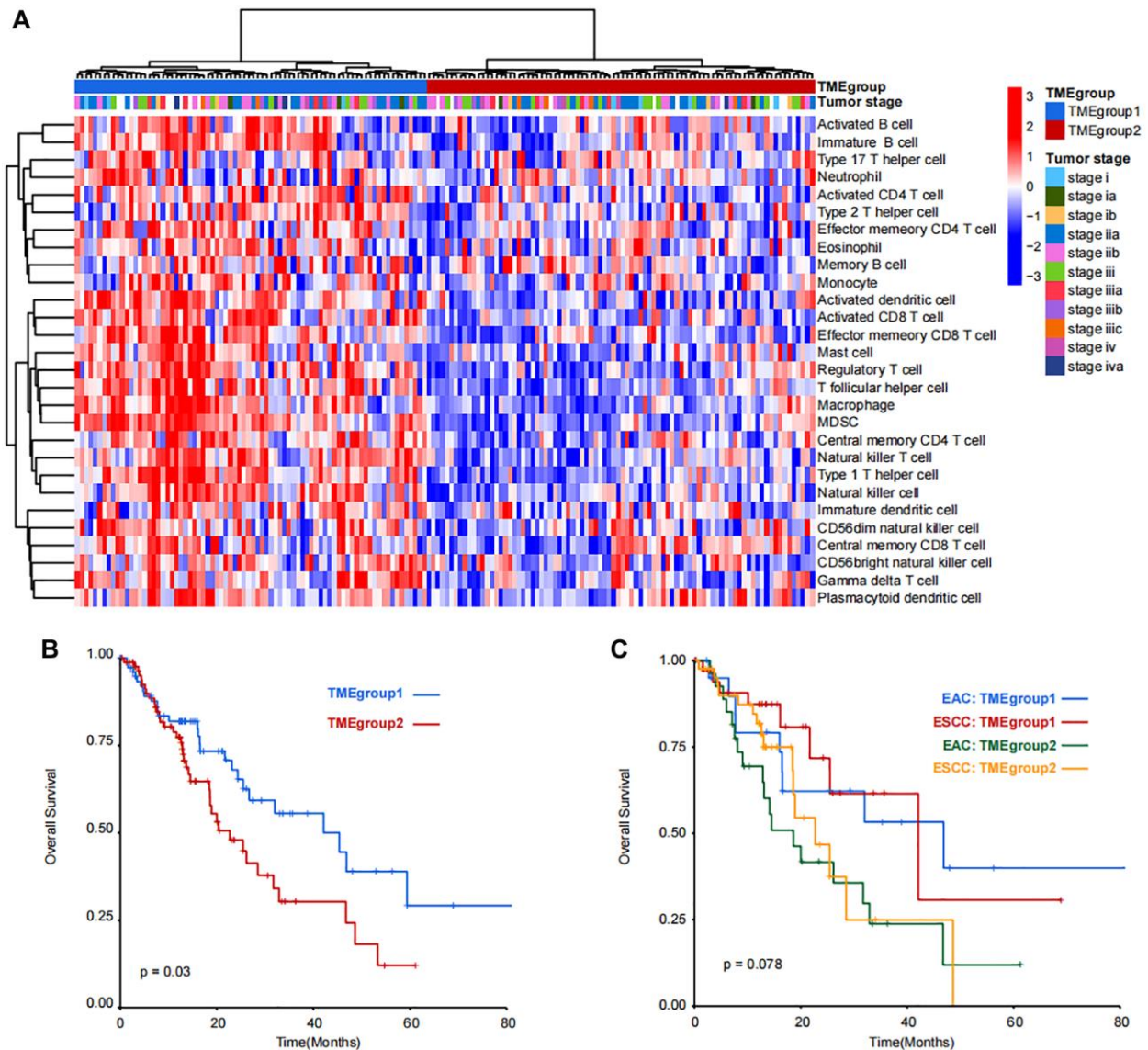


**Figure 1. Correlations between immune cell populations and the prognosis of two subtypes of esophageal cancer, EAC and ESCC. (A)** Forest plot showing the Risk Ratio and Hazard Ratio (HR) with the 95% CI of 28 kinds of TME cells in EAC and ESCC subtypes. Statistical significance is marked with asterisk at the right side. **(B)** Kaplan-Meier plots showing the survival curves of patients with high- and low- enrichment score of immature B cell and type 2 T helper cell in EAC and natural killer T cell in ESCC subtypes.



of 62 genes were finally identified (Figure 3A). Based on the expressions of these 62 genes, the patients could be divided into two groups (TMEgeneGroup1 and TMEgeneGroup2) (Figure 3B). The gene patterns between two TME gene group could also be divided into two different parts, *SS18L2*, *RPL14*, *ALG13*, *CRBN*, *CMC1*, *PIGT*, *TNNC2*, *TMEM74B*, *RBP2*, *PCK1*, *NOX1*, *MARVELD3*, *EMC10*, *LRRC45*, *RBBP7*, *TMEM106C*, *NDUFAF5*, *TASP1*, *CRNKL1*, *BCCIP*, *SNRPB*, *MRPS26*, *POLR2K*, *COX6C*, *CRIPT*, *TPRKB*, *ABRACL*, *CDKLI*, *HEBP2*, *MGST2*, *FNGR1*, *ARMT1*, *TXNLAB*, *DHODH*, *POLR2C*, *NVL*, *HIST1H1E*, *PUS10*, *CHTOP*, *EFNA1* belonged to the gene signature of TMEgeneGroup1 (TME metagene 1), and were

highly expressed in TMEgeneGroup1, while *WBP1L*, *CORO2B*, *SLIT2*, *ST6GALNAC6*, *ERAP2*, *FOS*, *CLN8*, *FAM189A2*, *RELL1*, *SEMA5A*, *XCR1*, *LSM10*, *NCDN*, *INTS5*, *GRB2*, *NUDT18*, *YAP1*, *MAP4K4*, *RNF144B*, *CDH24*, *PPIL2*, *BCR* belonged to the gene signature of TMEgeneGroup2 (TME metagene 2), and were highly expressed in TMEgeneGroup2 (Figure 3B). The clinical outcomes were also found between the two groups and TMEgeneGroup2 had a poor prognosis ( $P = 0.00016$ , Figure 3C). Most patients in TMEgroup1 were clustered into TMEgeneGroup1, and the prognostic difference between TMEgeneGroup were more obvious in that between TMEgroups (Figure 3D). The TMEgeneGroups were significantly correlated with



**Figure 2. Unsupervised clustering and prognostic analysis of immune cell populations in TCGA-ESCA dataset.** (A) Patients were divided into TMEgroup1 and TMEgroup2 by unsupervised clustering of 28 kinds of TME cells. TMEgroups and tumor stage of each patient are marked on the top of the heatmap. The color bar corresponds to the normalized enrichment score of TME cells. (B) Kaplan-Meier plot showing the overall survival of patients in TMEgroup1 (line in blue) and TMEgroup2 (line in red). (C) Kaplan-Meier plot showing the overall survival of patients in TMEgroup1 and TMEgroup2 between EAC and ESCC subtypes.

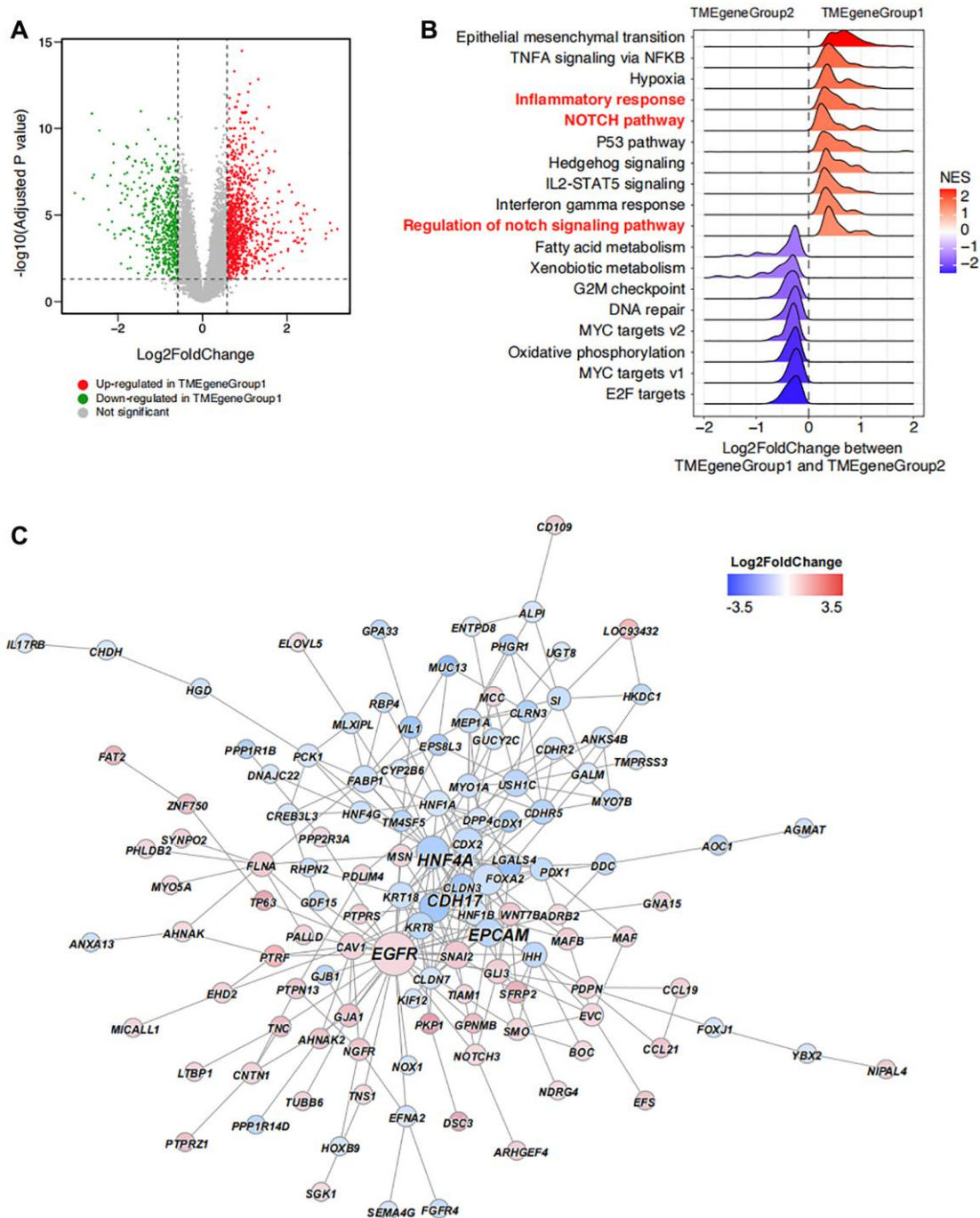




TMEGeneGroup2, and might serve as the biomarkers of TMEGeneGroup2 (Figure 4C).

Based on the 62 genes calculated from TMEGeneGroup, we built a model to increase the prediction efficiency. In addition, we also calculated TMEscoreA using TME

metagene 1, and TMEscoreB using TME metagene 2 according to ssGSEA algorithm. Through the comprehensive analysis, we found that the clinical ( $P < 0.0001$ , Supplementary Figure 7A) and the clinical outcome of patients with higher TMEscoreB was favorable ( $P < 0.0001$ , Supplementary Figure 7B).

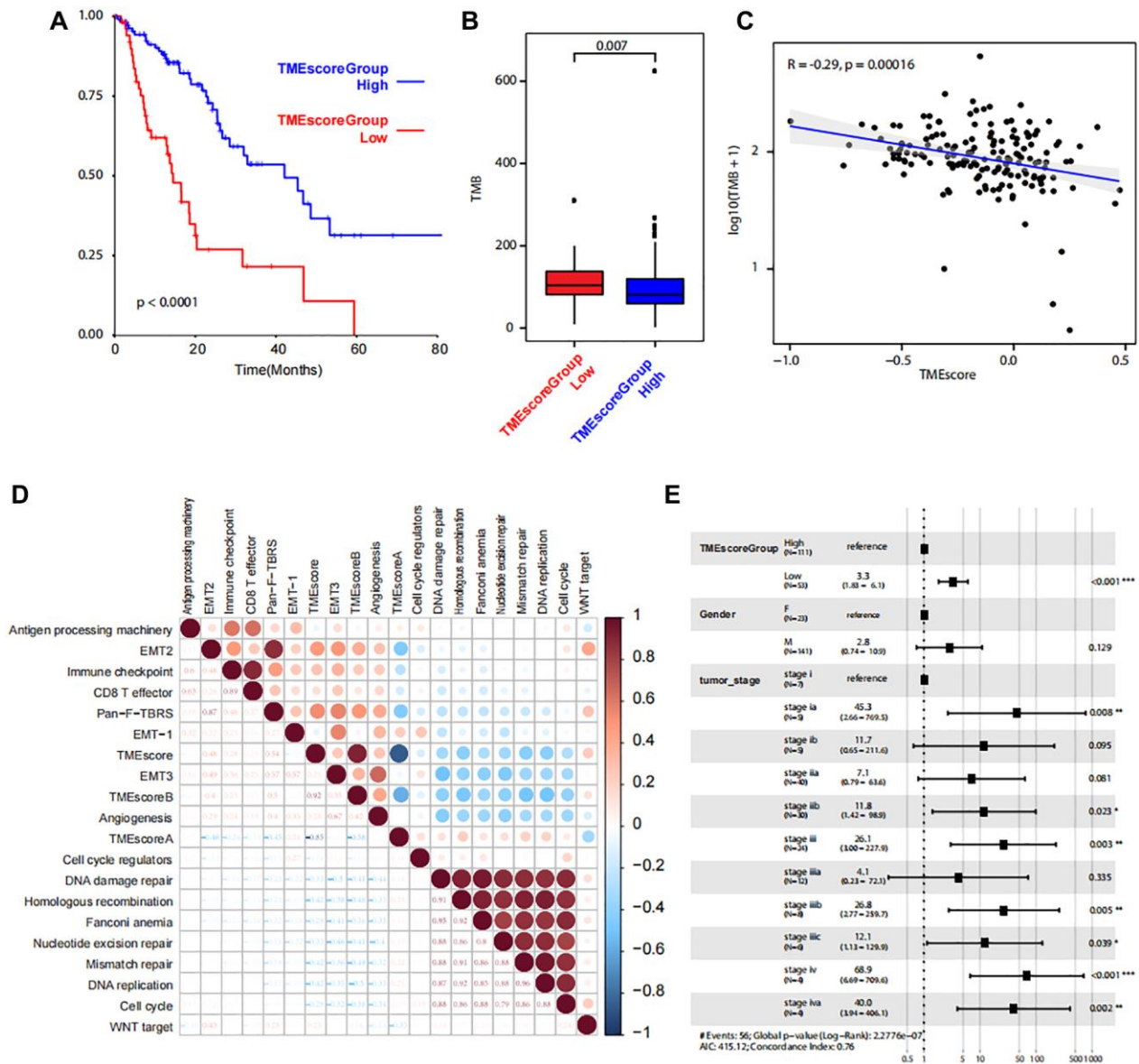


**Figure 4. Unsupervised clustering enrichment analysis of DEGs.** (A) DEGs between TMEGeneGroup1 and TMEGeneGroup2. The x axis corresponds to log2 transformed fold change value, and the y axis corresponds to  $-\log_{10}$  transformed adjusted  $P$  value. Up- and down-regulated genes in TMEGeneGroup1 are shown as red and blue dots. (B) GSEA analysis shows the top enriched pathways between TMEGeneGroup1 and TMEGeneGroup2. Pathways were ordered by NES and all pathways had  $P$  values less than 0.05. (C) PPI network of the DEGs between TMEGeneGroup1 and TMEGeneGroup2. Up- and down-regulated genes in TMEGeneGroup1 are colored in red and blue. Size of nodes and gene labels correspond to the hub score of genes.

Based on the TMEscoreA and TMEscoreB, we calculated the final TMEscore and we identified that the clinical outcome of patients with higher TMEscore was favorable ( $P < 0.0001$ , Figure 5A). Also, the tumor mutation burden (TMB) in low TME score group was higher than that in high TME score group (Figure 5B). The correlation analysis revealed a significant correlation between TMEscore and TMB (Figure 5C). After filtration of gene mutation profile between low TMEscore and high TMEscore, we found that the mutation frequency of *SYNE1* was higher in low TME score (Supplementary Figure 8). Gene expression level of 17 pathways were also analyzed to identify the

correlation between pathways and TMEscore, and the results revealed the positive correlation with EMT2, immune checkpoint, and Pan-F-TBRS, and negative correlation with DNA damage repair, homologous recombination, Fanconi anemia, nucleotide excision repair, mismatch repair, DNA replication and cell cycle (Figure 5D). Multifactor analysis showed TMEscore was an independent prognostic factor for the prognosis of patients with ESCA (Figure 5E).

To validate the prognostic impact of TMEscore model in cancer cohorts, we use 32 TCGA cancer cohorts (TCGA-ESCA was not included) and calculated



**Figure 5. Correlations between TMEscore and patient prognosis.** (A) Kaplan-Meier plot shows significant difference of the overall survival between the two groups with high and low level of TME scores. (B) Box plot shows TMB between the two groups with high and low level of TME scores. (C) Line regression shows the correlation between TMEscore and TMB. R value and P values are also labelled. (D) The correlation between cancer related pathways and TMEscore. The size and color correspond to the correlation values. (E) The multivariate Cox regression model shows the correlations between TMEscore and clinical phenotypes.



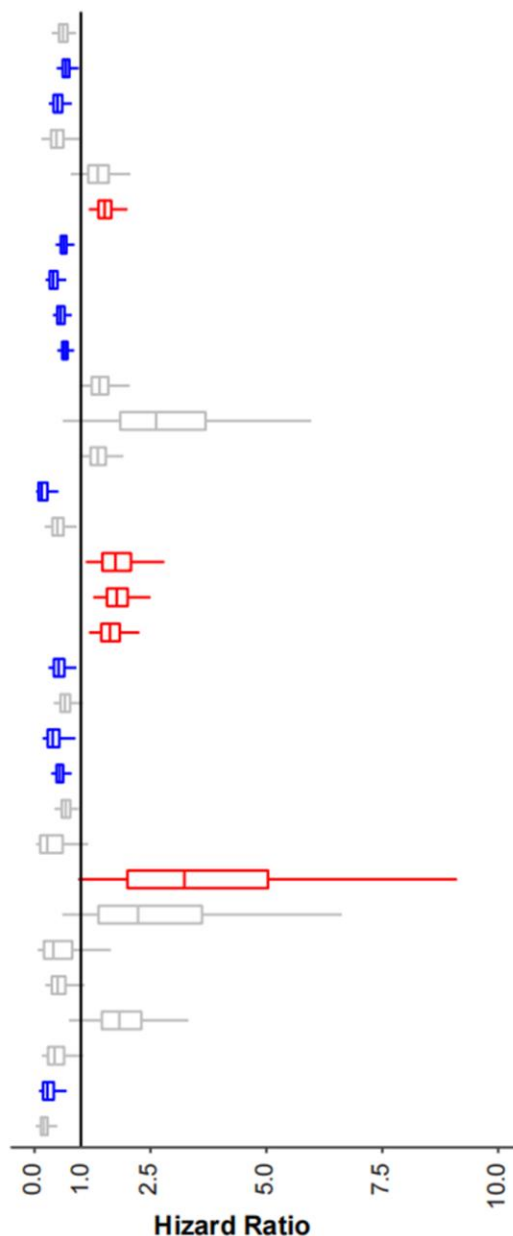
TMEscoreA and TMEscoreB of each patient. Generally, 11 (34.3%) cohorts (TCGA-PRAD, TCGA-UCS, TCGA-KIRP, TCGA-UCEC, TARGET-AML, TCGA-LIHC, TCGA-SARC, TCGA-KIRC, TCGA-LUAD, TCGA-HNSC, TCGA-BRCA) revealed the TMEscore as a good prognostic indicator, whereas 5 (15.6%) cohorts (TCGA-OV, TCGA-BLCA, TCGA-COAD, TCGA-STAD, TCGA-READ) revealed as a poor prognostic indicator (Figure 6). By validating in other TCGA cancer cohorts, we found most cohorts (21/32 of TCGA cohorts had HR less than 1 and 11/32 were significant) had a similar prognostic trend using

TMEscore, revealing TMEscore as a helpful indicator for tumorigenic microenvironment and prognosis.

## DISCUSSION

ESCA has a high mortality rate and poor prognosis, with an estimated 18,440 new cases and 16,170 deaths projected in 2020 in the United States [2]. TME plays an important role in the tumorigenesis, progression and therapy resistance of ESCA [21, 22]. Thus, there is a pressing need to explore the TME features which may assist oncologists in prognosis prediction. In this study, we

TCGA Project	P-value	Hizard Ratio
TARGET-NBL	0.052	0.610 (0.370-1.004)
TCGA-BRCA	0.025 *	0.673 (0.476-0.951)
TARGET-AML	0.004 *	0.492 (0.306-0.793)
TARGET-WT	0.206	0.471 (0.146-1.513)
TCGA-GBM	0.265	1.390 (0.779-2.482)
TCGA-OV	0.002 *	1.530 (1.169-2.004)
TCGA-LUAD	0.004 *	0.623 (0.452-0.860)
TCGA-UCEC	< 0.001 *	0.397 (0.247-0.638)
TCGA-KIRC	0.001 *	0.561 (0.393-0.799)
TCGA-HNSC	0.002 *	0.646 (0.488-0.854)
TCGA-LGG	0.071	1.411 (0.971-2.049)
TCGA-THCA	0.194	2.669 (0.606-11.748)
TCGA-LUSC	0.072	1.364 (0.972-1.915)
TCGA-PRAD	0.003 *	0.147 (0.042-0.514)
TCGA-SKCM	0.054	0.471 (0.219-1.013)
TCGA-COAD	0.016 *	1.763 (1.111-2.800)
TCGA-STAD	0.001 *	1.778 (1.263-2.502)
TCGA-BLCA	0.004 *	1.629 (1.174-2.261)
TCGA-LIHC	0.011 *	0.514 (0.308-0.858)
TCGA-CESC	0.072	0.653 (0.411-1.039)
TCGA-KIRP	0.009 *	0.384 (0.188-0.783)
TCGA-SARC	0.002 *	0.534 (0.356-0.800)
TCGA-PAAD	0.080	0.669 (0.427-1.049)
TCGA-PCPG	0.130	0.265 (0.048-1.476)
TCGA-READ	0.010 *	3.759 (1.379-10.243)
TCGA-THYM	0.228	2.262 (0.599-8.536)
TCGA-KICH	0.259	0.404 (0.084-1.947)
TCGA-ACC	0.076	0.497 (0.229-1.077)
TCGA-MESO	0.190	1.841 (0.739-4.584)
TARGET-RT	0.061	0.415 (0.165-1.042)
TCGA-UCS	0.001 *	0.247 (0.106-0.579)
TCGA-CHOL	0.110	0.192 (0.025-1.450)



**Figure 6. Validation of TMEscore on TCGA cancer cohorts.** Forest plot showing the prognosis of TMEscore in 32 tumor cohorts of TCGA. The HR of tumor types marked in blue represent the prognosis of high-TMEscore group is significantly good than that in the low TMEscore group, while those marked in red are opposite.

identified the key immune cells (natural killer T cell, immature B cell, natural killer cell, and type 1 T helper cell) associated with the prognosis and found that the TME features was significantly related to the OS of patients with ESCA. Based on the key TME genes, we explored a TMEscore which was demonstrated to be a reliable index for predicting the clinical outcome for ESCA patients.

Generally, TME cells include inflammatory cells, such as T cells, B cells, dendritic cells, natural killer cells, neutrophils, macrophages, etc. The tumorigenesis of ESCA are etiologically related to the exposure of the refluxed gastric and bile acids, which trigger chronic inflammation and the occurrence of Barrett's esophagus [23]. Other risk factors for ESCA, such as smoking and alcohol, also induce the toxic effects to increase the inflammation in the esophageal epithelium and subsequent play carcinogenic roles [24]. Thus, the ability of inflammatory responses may take part in the regulation of ESCA initiation and development. In addition, tumor cells can also recruit different immune cell populations or express immune checkpoints to suppress the anti-tumor immune response [21, 22]. Targeting immune checkpoints or supplement of the antitumor immune cells have been candidate options for anti-cancer therapy [8]. In this study, we found that natural killer T cell, immature B cell, natural killer cell and type 1 T helper cell were favorable for the prognosis of patients with ESCA.

Immune response related to the helper T cells is often impaired in ESCC patients and further damage the anti-tumor immunity [25]. Therapeutically, the double radiofrequency hyperthermia could regulate the conversion from Th2 to Th1 cells and work in the treatment of ESCA [26]. Impairment of natural killer cell activity also promotes tumor immunoevasion. T cell immunoglobulin domain and mucin domain-3 (Tim-3) was closely related to tumor invasion and distant metastasis in ESCC [27]. The mechanism was supposed to be associated with NK cell dysfunction in tumor microenvironment [28]. Besides, invariant natural killer T (iNKT) based immunotherapies showed promise for cancer patients and may benefit ESCA patients [29].

In view of the important roles of TMB in the tumorigenesis, progression and treatment, we explored a TMEscore based on the key TME genes and verified it as a reliable index for the clinical outcome and TMB. Evaluation of prognostic factors is important for oncologists in clinical decision-marking; thus many previous studies have focused on the identification of predictors for patients with ESCA. Clinical information (such as age, sex, tumor stage and grade), blood examination (such as C-reactive protein, albumin ratio), tumor characteristics (such as metabolic tumor volume

and total lesion glycolysis) and treatment methods (such as chemoradiotherapy, chemotherapy and surgery) have been regarded as prognostic factors and proved to be associated with the prognosis of patients with ESCA [30–33]. However, these prognostic factors were most clinical and did not include the molecular biomarkers. With the well-application of TCGA dataset, regulatory network of microRNA, lncRNA, mRNA and has been identified and prognosis-associated biomarkers have also been evaluated for patients with ESCA [34–36]. Unfortunately, the tumor microenvironment scores have not been explored comprehensively. Thus, this study is a good supplement to the existing research about the prognosis evaluation of patients with ESCA.

In the TMEgeneGroup1 with favorable prognosis, inflammatory response and NOTCH signaling pathway were enriched. Notch is a key modulator in regulating T-cell development, maintenance, activation and takes part in T cell-mediated immune responses [37, 38]. In addition, *EGFR* was significantly up-regulated in TMEgeneGroup1 by the hub gene analysis. As a therapeutic target, *EGFR* amplifications can be found in ESCA and the application of *EGFR* inhibitors may improve the prognosis of these ESCA patients [39, 40]. Thus, the favorable prognosis of patients in the TMEgeneGroup1 may be due to the good immune response and the application of *EGFR* inhibitors.

This study inevitably has some limitations that should be declared. Firstly, due to limited published data and fewer cases of ESCA, the amount of data analyzed in this study is limited, which may lead to potential errors or deviations. We used TCGA cohorts to validate the prognosis impact of TMEscore. And in future work, more esophageal cancer patients would be enrolled and studied to deeply explore the tumor microenvironment and help to improve the prognosis. Secondly, as a bioinformatics analysis, the roles of identified immune cells is not verified by cell and molecular experiments or clinical tissue microarray. Thirdly, all data series downloaded in this study are from America, and we are not quite sure about its applicability in European and Asian.

## CONCLUSIONS

Our data provided key immune cells (natural killer T cell, immature B cell, natural killer cell, and type 1 T helper cell) associated with the prognosis of patients with ESCA and constructed a novel prognostic model, termed TMEscore, as a reliable index for predicting the clinical outcomes of ESCA patients. And by validation of other TCGA cancer cohorts, it revealed TMEscore could be applied to predict the prognosis of other tumors.

## AUTHOR CONTRIBUTIONS

Yuhong zhang, Minqi Zhu and Junxian Mo conceived the study and wrote the paper. Yuhong zhang and Lei Xian performed analyzed the data. All authors approved the manuscript.

## CONFLICTS OF INTEREST

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

## FUNDING

This study was supported by The Natural Science Foundation of Guangxi, China (Grant No. 2018JJA140567).

## REFERENCES

1. Domper Arnal MJ, Ferrández Arenas Á, Lanás Arbeloa Á. Esophageal cancer: Risk factors, screening and endoscopic treatment in Western and Eastern countries. *World J Gastroenterol*. 2015; 21:7933–43. <https://doi.org/10.3748/wjg.v21.i26.7933> PMID:26185366
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. 2020; 70:7–30. <https://doi.org/10.3322/caac.21590> PMID:31912902
3. Pennathur A, Gibson MK, Jobe BA, Luketich JD. Oesophageal carcinoma. *Lancet*. 2013; 381:400–12. [https://doi.org/10.1016/S0140-6736\(12\)60643-6](https://doi.org/10.1016/S0140-6736(12)60643-6) PMID:23374478
4. Lagergren J, Smyth E, Cunningham D, Lagergren P. Oesophageal cancer. *Lancet*. 2017; 390:2383–96. [https://doi.org/10.1016/S0140-6736\(17\)31462-9](https://doi.org/10.1016/S0140-6736(17)31462-9) PMID:28648400
5. Borggreve AS, Kingma BF, Domrachev SA, Koshkin MA, Ruurda JP, van Hillegersberg R, Takeda FR, Goense L. Surgical treatment of esophageal cancer in the era of multimodality management. *Ann N Y Acad Sci*. 2018; 1434:192–209. <https://doi.org/10.1111/nyas.13677> PMID:29761863
6. Campbell NP, Villaflor VM. Neoadjuvant treatment of esophageal cancer. *World J Gastroenterol*. 2010; 16:3793–803. <https://doi.org/10.3748/wjg.v16.i30.3793> PMID:20698042
7. Deng HY, Zha P, Peng L, Hou L, Huang KL, Li XY. Preoperative sarcopenia is a predictor of poor prognosis of esophageal cancer after esophagectomy: a comprehensive systematic review and meta-analysis. *Dis Esophagus*. 2019; 32:doy115. <https://doi.org/10.1093/dote/doy115> PMID:30496385
8. Alsina M, Moehler M, Lorenzen S. Immunotherapy of Esophageal Cancer: Current Status, Many Trials and Innovative Strategies. *Oncol Res Treat*. 2018; 41:266–71. <https://doi.org/10.1159/000488120> PMID:29705786
9. Inuma H, Fukushima R, Inaba T, Tamura J, Inoue T, Ogawa E, Horikawa M, Ikeda Y, Matsutani N, Takeda K, Yoshida K, Tsunoda T, Ikeda T, et al. Phase I clinical study of multiple epitope peptide vaccine combined with chemoradiation therapy in esophageal cancer patients. *J Transl Med*. 2014; 12:84. <https://doi.org/10.1186/1479-5876-12-84> PMID:24708624
10. Kageyama S, Ikeda H, Miyahara Y, Imai N, Ishihara M, Saito K, Sugino S, Ueda S, Ishikawa T, Kokura S, Naota H, Ohishi K, Shiraishi T, et al. Adoptive Transfer of MAGE-A4 T-cell Receptor Gene-Transduced Lymphocytes in Patients with Recurrent Esophageal Cancer. *Clin Cancer Res*. 2015; 21:2268–77. <https://doi.org/10.1158/1078-0432.CCR-14-1559> PMID:25855804
11. Lin EW, Karakasheva TA, Hicks PD, Bass AJ, Rustgi AK. The tumor microenvironment in esophageal cancer. *Oncogene*. 2016; 35:5337–49. <https://doi.org/10.1038/onc.2016.34> PMID:26923327
12. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011; 144:646–74. <https://doi.org/10.1016/j.cell.2011.02.013> PMID:21376230
13. Smyth MJ, Ngiew SF, Ribas A, Teng MW. Combination cancer immunotherapies tailored to the tumour microenvironment. *Nat Rev Clin Oncol*. 2016; 13:143–58. <https://doi.org/10.1038/nrclinonc.2015.209> PMID:26598942
14. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Carolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, Ceccarelli M, Bontempi G, Noushmehr H. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. 2016; 44:e71. <https://doi.org/10.1093/nar/gkv1507> PMID:26704973
15. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. 2013; 14:7.



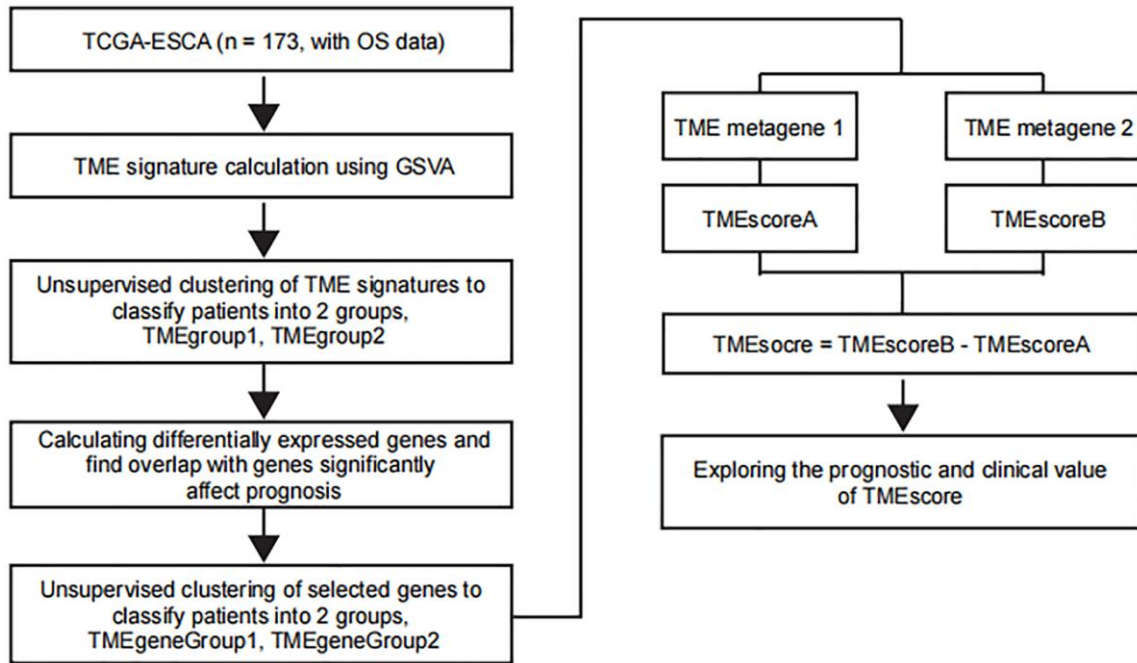
- <https://doi.org/10.1186/1471-2105-14-7>  
PMID:23323831
16. Jia Q, Wu W, Wang Y, Alexander PB, Sun C, Gong Z, Cheng JN, Sun H, Guan Y, Xia X, Yang L, Yi X, Wan YY, et al. Local mutational diversity drives intratumoral immune heterogeneity in non-small cell lung cancer. *Nat Commun.* 2018; 9:5361.  
<https://doi.org/10.1038/s41467-018-07767-w>  
PMID:30560866
17. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics.* 2010; 26:1572–3.  
<https://doi.org/10.1093/bioinformatics/btq170>  
PMID:20427518
18. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015; 43:e47.  
<https://doi.org/10.1093/nar/gkv007>  
PMID:25605792
19. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012; 16:284–7.  
<https://doi.org/10.1089/omi.2011.0118>  
PMID:22455463
20. Meng T, Huang R, Zeng Z, Huang Z, Yin H, Jiao C, Yan P, Hu P, Zhu X, Li Z, Song D, Zhang J, Cheng L. Identification of Prognostic and Metastatic Alternative Splicing Signatures in Kidney Renal Clear Cell Carcinoma. *Front Bioeng Biotechnol.* 2019; 7:270.  
<https://doi.org/10.3389/fbioe.2019.00270>  
PMID:31681747
21. Karakasheva TA, Lin EW, Tang Q, Qiao E, Waldron TJ, Soni M, Klein-Szanto AJ, Sahu V, Basu D, Ohashi S, Baba K, Giaccone ZT, Walker SR, et al. IL-6 Mediates Cross-Talk between Tumor Cells and Activated Fibroblasts in the Tumor Microenvironment. *Cancer Res.* 2018; 78:4957–70.  
<https://doi.org/10.1158/0008-5472.CAN-17-2268>  
PMID:29976575
22. Kelly RJ, Zaidi AH, Smith MA, Omstead AN, Kosovec JE, Matsui D, Martin SA, DiCarlo C, Werts ED, Silverman JF, Wang DH, Jobe BA. The Dynamic and Transient Immune Microenvironment in Locally Advanced Esophageal Adenocarcinoma Post Chemoradiation. *Ann Surg.* 2018; 268:992–9.  
<https://doi.org/10.1097/SLA.0000000000002410>  
PMID:28806299
23. Debruyne PR, Witek M, Gong L, Birbe R, Chervoneva I, Jin T, Domon-Cell C, Palazzo JP, Freund JN, Li P, Pitari GM, Schulz S, Waldman SA. Bile acids induce ectopic expression of intestinal guanylyl cyclase C Through nuclear factor-kappaB and Cdx2 in human esophageal cells. *Gastroenterology.* 2006; 130:1191–206.  
<https://doi.org/10.1053/j.gastro.2005.12.032>  
PMID:16618413
24. Abnet CC, Arnold M, Wei WQ. Epidemiology of Esophageal Squamous Cell Carcinoma. *Gastroenterology.* 2018; 154:360–73.  
<https://doi.org/10.1053/j.gastro.2017.08.023>  
PMID:28823862
25. Samiei H, Sadighi-Moghaddam B, Mohammadi S, Gharavi A, Abdolmaleki S, Khosravi A, Kokhaei P, Bazzazi H, Memarian A. Dysregulation of helper T lymphocytes in esophageal squamous cell carcinoma (ESCC) patients is highly associated with aberrant production of miR-21. *Immunol Res.* 2019; 67:212–22.  
<https://doi.org/10.1007/s12026-019-09079-7>  
PMID:31278653
26. Hong M, Jiang Z, Zhou YF. Effects of thermotherapy on Th1/Th2 cells in esophageal cancer patients treated with radiotherapy. *Asian Pac J Cancer Prev.* 2014; 15:2359–62.  
<https://doi.org/10.7314/apjcp.2014.15.5.2359>  
PMID:24716984
27. Hu S, Xie Y, Zhou N, Jin L, Tan Y, Liu D, Gong Y, Liu L, Liu J, Liu W, Chen Y, Zhang Y, Lv N. Expression of T-cell immunoglobulin- and mucin-domain-containing molecules-1 and -3 (Tim-1 and Tim-3) in Helicobacter pylori infection. *Helicobacter.* 2011; 16:373–81.  
<https://doi.org/10.1111/j.1523-5378.2011.00855.x>  
PMID:21923683
28. Zheng Y, Li Y, Lian J, Yang H, Li F, Zhao S, Qi Y, Zhang Y, Huang L. TNF- $\alpha$ -induced Tim-3 expression marks the dysfunction of infiltrating natural killer cells in human esophageal cancer. *J Transl Med.* 2019; 17:165.  
<https://doi.org/10.1186/s12967-019-1917-0>  
PMID:31109341
29. Melo AM, Conroy MJ, Foley EK, Dockry É, Breen EP, Reynolds JV, Lysaght J, Doherty DG. CD1d expression and invariant natural killer T-cell numbers are reduced in patients with upper gastrointestinal cancers and are further impaired by commonly used chemotherapies. *Cancer Immunol Immunother.* 2020; 69:969–82.  
<https://doi.org/10.1007/s00262-020-02514-x>  
PMID:32078695
30. Gupta V, Coburn N, Kidane B, Hess KR, Compton C, Ringash J, Darling G, Mahar AL. Survival prediction tools for esophageal and gastroesophageal junction cancer: A systematic review. *J Thorac Cardiovasc Surg.* 2018; 156:847–56.

<https://doi.org/10.1016/j.jtcvs.2018.03.146>  
PMID:30011772

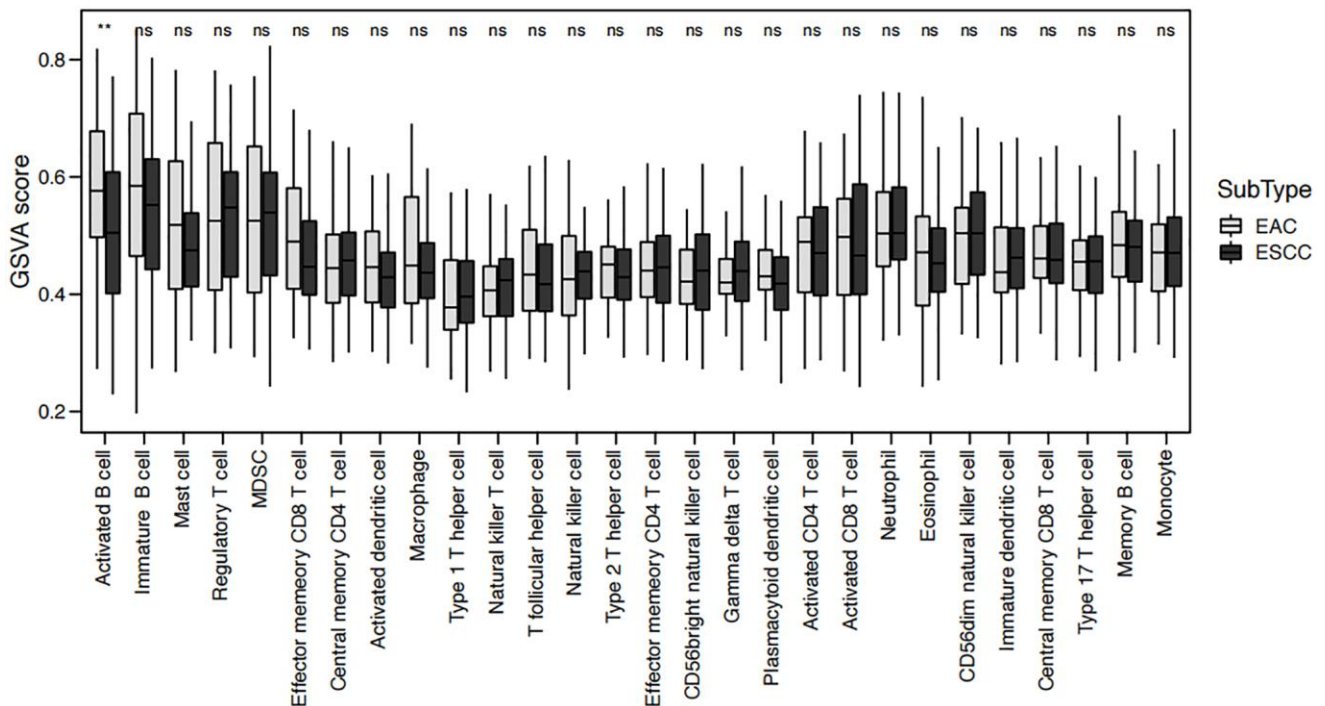
31. Ishibashi Y, Tsujimoto H, Hiraki S, Kumano I, Yaguchi Y, Horiguchi H, Nomura S, Ito N, Shinto E, Aosasa S, Yamamoto J, Ueno H. Prognostic Value of Preoperative Systemic Immunoinflammatory Measures in Patients with Esophageal Cancer. *Ann Surg Oncol*. 2018; 25:3288–99.  
<https://doi.org/10.1245/s10434-018-6651-y>  
PMID:30019304
32. Vendrely V, Launay V, Najah H, Smith D, Collet D, Gronnier C. Prognostic factors in esophageal cancer treated with curative intent. *Dig Liver Dis*. 2018; 50:991–6.  
<https://doi.org/10.1016/j.dld.2018.08.002>  
PMID:30166221
33. Yildirim BA, Torun N, Guler OC, Onal C. Prognostic value of metabolic tumor volume and total lesion glycolysis in esophageal carcinoma patients treated with definitive chemoradiotherapy. *Nucl Med Commun*. 2018; 39:553–63.  
<https://doi.org/10.1097/MNM.0000000000000837>  
PMID:29668513
34. Li CY, Zhang WW, Xiang JL, Wang XH, Li J, Wang JL. Identification of microRNAs as novel biomarkers for esophageal squamous cell carcinoma: a study based on The Cancer Genome Atlas (TCGA) and bioinformatics. *Chin Med J (Engl)*. 2019; 132:2213–22.  
<https://doi.org/10.1097/CM9.0000000000000427>  
PMID:31490264
35. Fan Q, Liu B. Identification of a RNA-Seq Based 8-Long Non-Coding RNA Signature Predicting Survival in Esophageal Cancer. *Med Sci Monit*. 2016; 22:5163–72.  
<https://doi.org/10.12659/msm.902615>  
PMID:28028307
36. Zeng JH, Xiong DD, Pang YY, Zhang Y, Tang RX, Luo DZ, Chen G. Identification of molecular targets for esophageal carcinoma diagnosis using miRNA-seq and RNA-seq data from The Cancer Genome Atlas: a study of 187 cases. *Oncotarget*. 2017; 8:35681–99.  
<https://doi.org/10.18632/oncotarget.16051>  
PMID:28415685
37. Radtke F, Fasnacht N, Macdonald HR. Notch signaling in the immune system. *Immunity*. 2010; 32:14–27.  
<https://doi.org/10.1016/j.immuni.2010.01.004>  
PMID:20152168
38. Kelliher MA, Roderick JE. NOTCH Signaling in T-Cell-Mediated Anti-Tumor Immunity and T-Cell-Based Immunotherapies. *Front Immunol*. 2018; 9:1718.  
<https://doi.org/10.3389/fimmu.2018.01718>  
PMID:30967879
39. Zhang L, Ma J, Han Y, Liu J, Zhou W, Hong L, Fan D. Targeted therapy in esophageal cancer. *Expert Rev Gastroenterol Hepatol*. 2016; 10:595–604.  
<https://doi.org/10.1586/17474124.2016.1140036>  
PMID:26895097
40. Sanchez-Vega F, Hechtman JF, Castel P, Ku GY, Tuvy Y, Won H, Fong CJ, Bouvier N, Nanjangud GJ, Soong J, Vakiani E, Schattner M, Kelsen DP, et al. *EGFR* and *MET* Amplifications Determine Response to HER2 Inhibition in *ERBB2*-Amplified Esophagogastric Cancer. *Cancer Discov*. 2019; 9:199–209.  
<https://doi.org/10.1158/2159-8290.CD-18-0598>  
PMID:30463996

SUPPLEMENTARY MATERIALS

Supplementary Figures

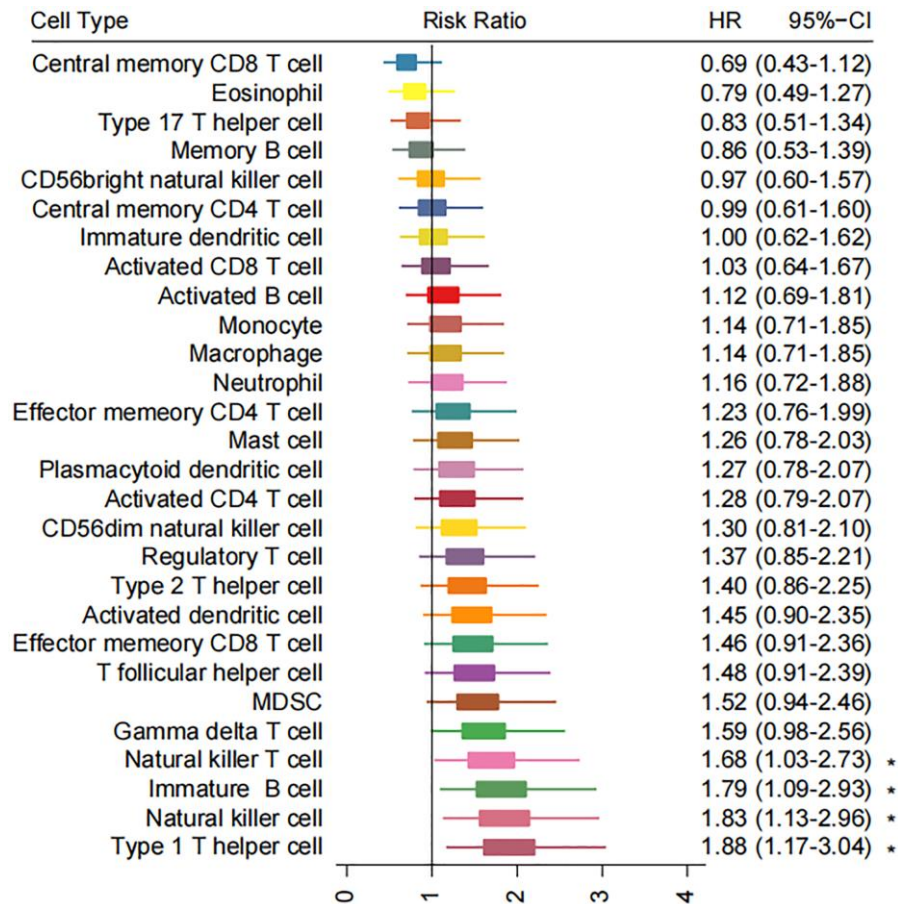


Supplementary Figure 1. The overall workflow of the study. Important analysis steps and datasets are highlighted in squares and the arrows indicates the direction of the pipeline.

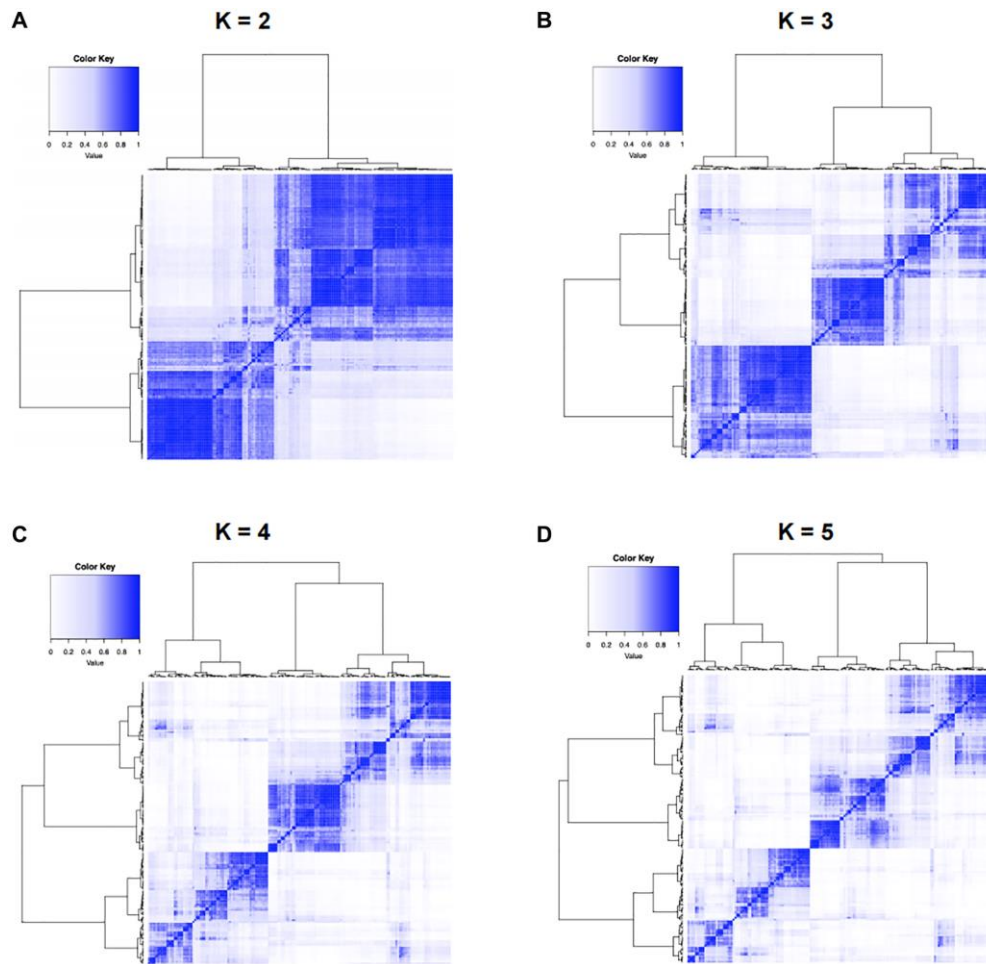


Supplementary Figure 2. Difference of TME cells between EAC and ESCC samples. Box plot shows the GSVA score of 28 kinds of TME cells. EAC and ESCC patients are colored in light grey and black. The statistical significance are also labelled. ns for not significant. \*\* for  $p < 0.01$ .

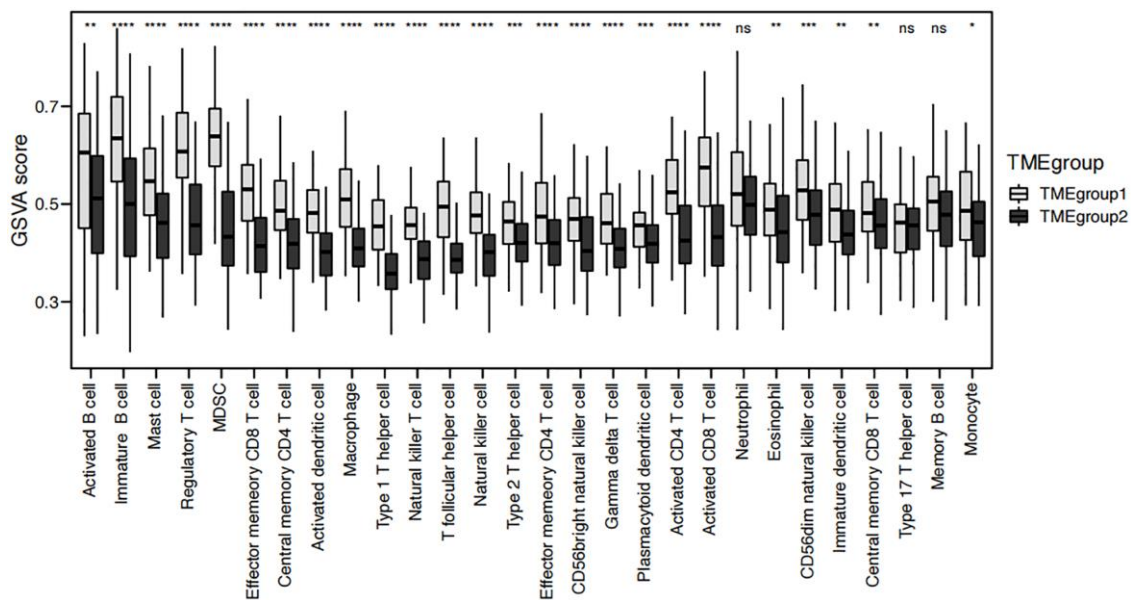




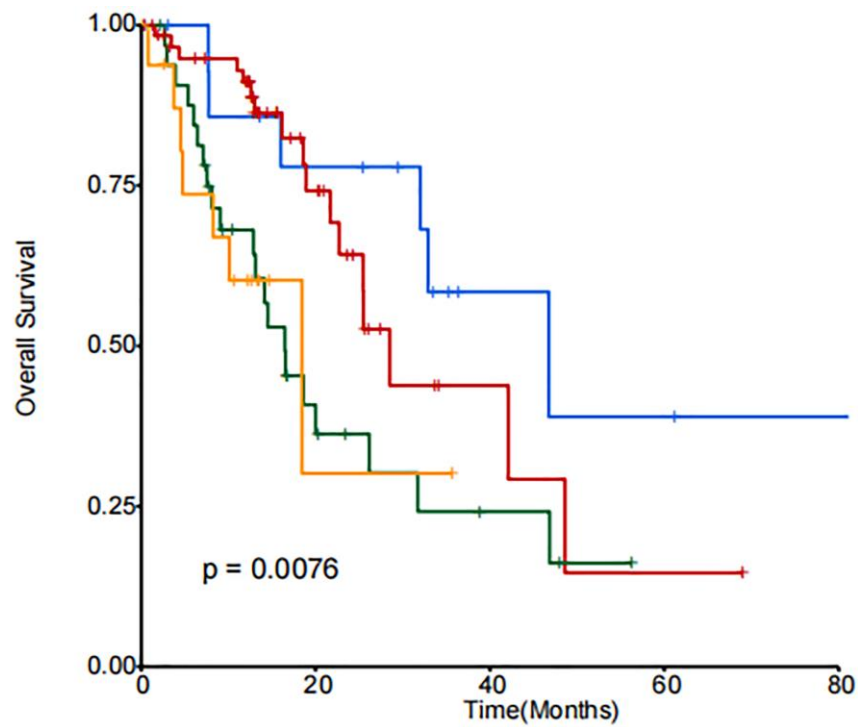
**Supplementary Figure 3. Correlations between TME cells and clinical outcomes.** Forest plot shows the Risk Ratio and Hazard Ratio with the 95% CI of 28 kinds of TME cells in all patients with ESCA. Statistical significance are marked with asterisk.



**Supplementary Figure 4. Identification of consensus clusters by TME cells.** Consensus clustering matrix for K = 2 (A), K = 3 (B), K = 4 (C), and K = 5 (D) are shown.



**Supplementary Figure 5. Difference of TME cells between TMEgroup1 and TMEgroup2.** Box plot shows the GSVA score of 28 kinds of TME cells. TMEgroup1 and ESCC TMEgroup1 are colored in light grey and black. The statistical significance are also labelled. ns for not significant, \* for  $p < 0.05$ , \*\* for  $p < 0.01$ , \*\*\* for  $p < 0.001$  and \*\*\*\* for  $p < 0.0001$ .

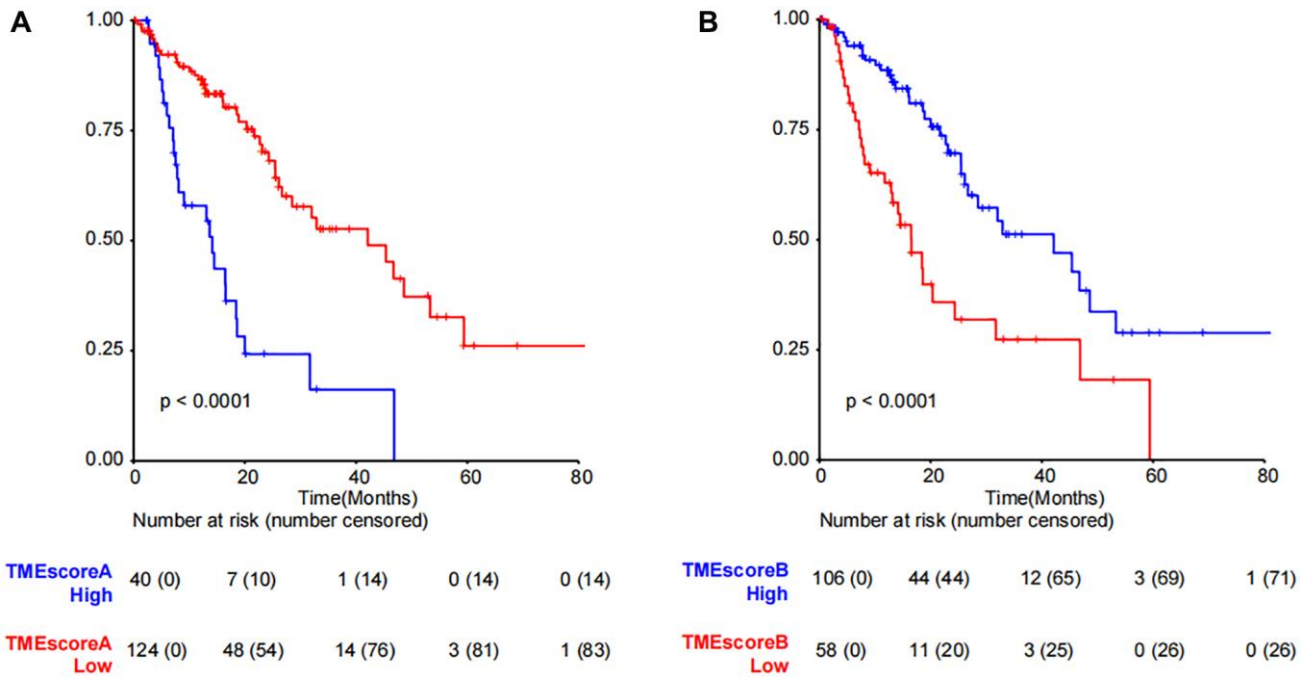


Number at risk (number censored)

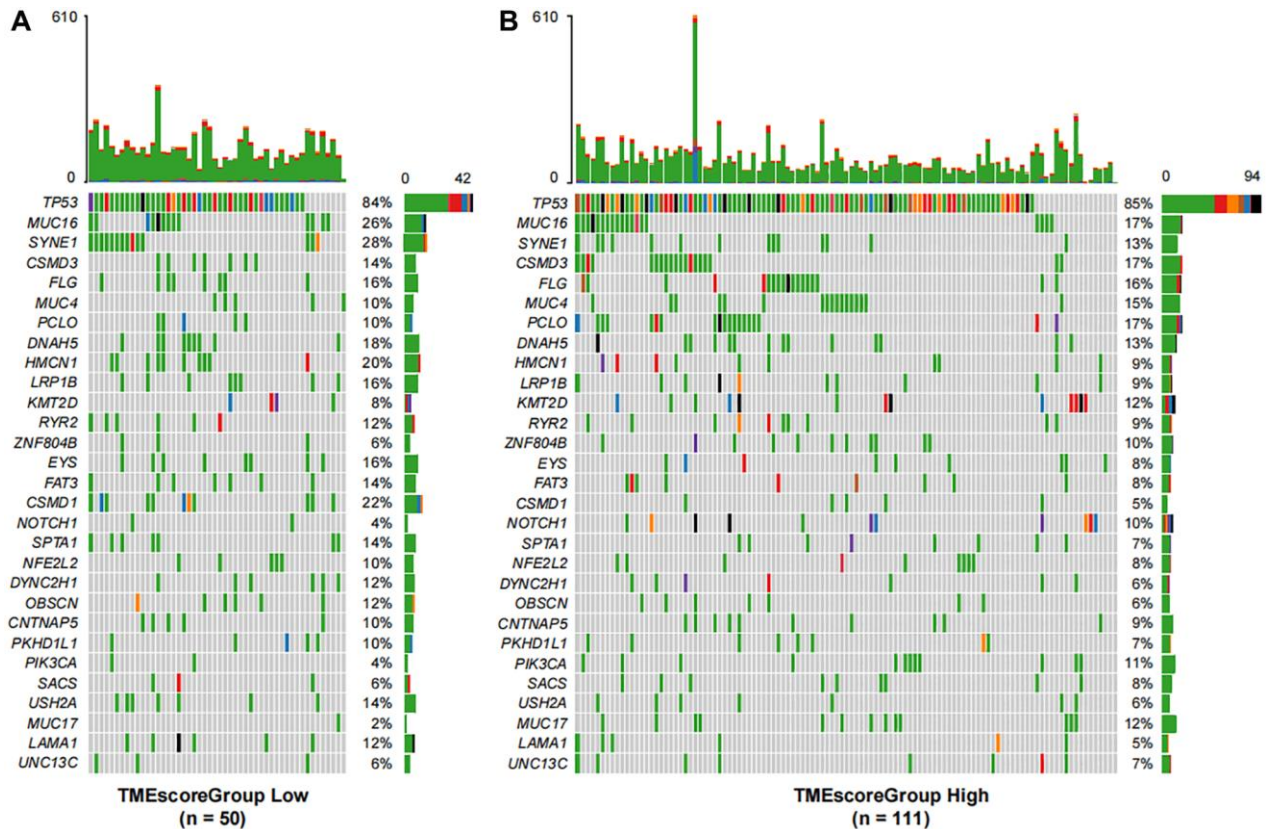
<b>EAC: TMEgeneGroup1</b>	16 (0)	10 (3)	3 (8)	2 (8)	1 (9)
<b>ESCC: TMEgeneGroup1</b>	61 (0)	18 (33)	3 (43)	1 (43)	0 (44)
<b>EAC: TMEgeneGroup2</b>	33 (0)	9 (7)	3 (10)	0 (12)	0 (12)
<b>ESCC: TMEgeneGroup2</b>	17 (0)	1 (9)	0 (10)	0 (10)	0 (10)

**Supplementary Figure 6. Survival analysis of patients between different TMEgeneGroups and ESCA subtypes.** Kaplan-Meier plot shows the overall survival among the patients within four groups of different TMEgeneGroups and ESCA subtypes. Survival curves are marked in different colors and the number of cases at risk and censored at different time points are also labelled.





**Supplementary Figure 7. Survival analysis of patients between different levels of TMEscore.** Kaplan-Meier plot shows the overall survival among the patients within high and low levels of TMEscoreA (A) and TMEscoreB (B). Survival curves with high and low levels of TMEscore are colored in blue and red. The number of cases at risk and censored at different time points are also labelled.



**Supplementary Figure 8. Gene mutation profiles of patients with high and low TMEscore.** CoMut plot shows the mutation profiles of top ranked mutated genes in TMEscoreGroup Low (A) and TMEscoreGroup High (B). The column corresponds to a patient, and the row corresponds to a gene. The bar plot on the top shows the number of total mutations within each patient, and the bar plot on the right panel shows the number of different mutations with a gene among all patients. The mutation frequencies of genes are also labelled.

## Supplementary Table

Please browse Full Text version to see the data of Supplementary Table 1.

**Supplementary Table 1. Information of each individual in TCGA-ESCA dataset.**