Characterization of the Immune Infiltration Landscape and Identification of Prognostic Biomarkers for Esophageal Cancer

Yuanmei Chen
Fujian Medical University cancer Hospital: Fujian Provincial Cancer Hospital
Xinyi Huang
Fujian Medical University cancer Hospital: Fujian Provincial Cancer Hospital
Haiyan Peng
Fujian Normal University - Qishan Campus
Guibin Weng
Fujian Medical University cancer Hospital: Fujian Provincial Cancer Hospital
Zhengrong Huang
Fujian Medical University cancer Hospital: Fujian Provincial Cancer Hospital
Yangfan Zhang
Fujian Normal University - Qishan Campus
Tianya Xiao
Fujian Medical University cancer Hospital: Fujian Provincial Cancer Hospital
Junqiang Chen
Fujian Medical University cancer Hospital: Fujian Provincial Cancer Hospital
Kunshou Zhu
Fujian Medical University cancer Hospital: Fujian Provincial Cancer Hospital
Yuanji Xu (xuyuanji@fjmu.edu.cn)
Fujian Medical University cancer Hospital: Fujian Provincial Cancer Hospital https://orcid.org/0000-0002-7194-0433

Primary research

Keywords: Esophageal cancer (EC), Immune cell infiltration (ICI), Prognosis, Immunotherapeutic responses, Biomarker.

Posted Date: September 21st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-919249/v1

Abstract

Background. Esophageal cancer (EC) is the 7th most common neoplasm and the 6th most common cause of cancer-related death worldwide. Immunotherapy is an effective treatment for EC patients. However, there are no dependable markers for predicting prognosis and immunotherapy responses in EC. Our study aims to explore prognostic models and markers in EC as well as predictors for immunotherapy.

Methods. The expression profiles of EC were obtained from The Cancer Genome Atlas (TCGA), the Gene Expression Omnibus (GEO), and International Cancer Genome Consortium (ICGC) databases. Cox regression analysis was performed to construct a prognostic model. Overall survival and receiver operating characteristic curve analyses were applied to verify the accuracy of the model. The CIBERSORT algorithm was conducted to quantify the infiltration of different immune cells, and EC was grouped into three immune cell infiltration (ICI) clusters. PD-1 and PD-L1 expressions were compared between the ICI clusters. Overall survival analysis between ICI score and tumor mutation burden was conducted. The immunotherapy response of patients in different ICI score clusters was also compared. The copy number variations, somatic mutations, and single nucleotide polymorphisms were analyzed. Enrichment analyses were also performed.

Results. A prognostic model was successfully constructed. Three ICI clusters were identified, and the clusters with high immune and stromal scores tended to have more immune-activated phenotypes and higher expressions of PD-1 and PDL1. The ICI score may be used as a predictor independent of tumor mutation burden. Patients with higher ICI score tended to have better immunotherapeutic responses than those with lower scores. Enrichment analyses showed that the differentially expressed genes were mostly enriched in microvillus and the KRAS and IL6/JAK/STAT3 pathways. The top eight genes with the highest mutation frequencies in EC were identified and all related to the prognosis of EC patients.

Conclusions. Our study established an effective prognostic model and identified markers for predicting the prognosis and immunotherapy response of EC patients.

Background

Esophageal cancer (EC) ranks as the seventh most common cancer and the sixth most common cause of cancer-related death worldwide [1]. Approximately 508,585 cancer-related deaths of EC were reported in 2018 [1]. Surgical removal is the conventional treatment for most types of EC. However, many patients still show local recurrence or distant metastasis within a short time after surgery [2, 3]. Although surgery, radiotherapy and chemotherapy have improved the survival rate of EC patients, the prognosis of EC remains poor, and the current 5-year overall survival (OS) rate is only 15–25% [4]. Currently, the prognosis of EC patients is mainly determined by postoperative pathological features, which relies on the Tumor Node Metastasis (TNM) staging system. However, because of problems such as stage migration, the current TNM staging system is not very accurate [5]. Therefore, the identification of accurate prognostic models and markers is urgently required to improve the prognosis of EC patients and develop individualized therapies.

Immunotherapy is a critical therapy for cancer treatment that activates the immune system to target cancer cells, thus enhancing the autoimmune function to eliminate tumors [6]. Some patients can achieve complete remission after immunotherapy [7]. However, immunotherapy only benefits a small number of patients [8]. Immunotherapy in EC patients often causes different therapeutic effects, partly because of the lack of dependable markers to predict therapeutic response [9]. Therefore, identifying new markers that predict immunotherapy response is critical to determine the subgroup of EC patients that may benefit from these treatments.

One of the mechanisms by which tumors escape immune response is by inhibiting infiltration of immune cells. Increasing evidence has demonstrated that the tumor microenvironment (TME), which includes tumor-infiltrating immune cells (TIICs), plays a vital role in tumor immune evasion and cancer cell proliferation, invasion, and metastasis [10]. The selective enhancement of anti-tumor immunity in the TME but not in non-tumor tissues enables highly effective tumor immunotherapy without side effects [11]. Better understanding of the interactions between tumor cells and the TME and the mechanisms underlying how these promote the occurrence and development of EC could offer novel insights into the future treatment strategies. To date, the broad prospects of immune cell infiltration (ICI) in the TME in EC have not been clarified.

The purpose of our study was to explore the immune infiltration landscape and potential prognostic biomarkers in EC. In our study, we analyzed the differentially expressed immune-related genes between normal esophageal and EC tumor samples in The Cancer Genome Analysis (TCGA) cohort. A prognostic model was constructed, and survival analysis and risk assessment were performed. The CIBERSORT and ESTIMATE algorithms were used to analyze the gene expressions of EC and normal samples, and an overview of the immunity in EC was obtained. We further divided EC into three subtypes based on ICI patterns. Moreover, analyses of TCGA tumor sample copy number and somatic variations were conducted. Finally, single nucleotide polymorphism (SNP) analyses of International Cancer Genome Consortium (ICGC)-UK and ICGC-CN data sets were performed. We established a prognostic model and ICI scores, which may predict the prognosis and the response to immunotherapy of EC patients.

Materials & Methods

Downloading and processing of sample source data

The expression profiles of EC were obtained from four independent cohorts, including TCGA, GSE106185, ICGC-CN, and ICGC-UK. Only data from EC tumor samples were obtained. The fragments per kilobase million value of TCGA expression profile was obtained using the TCGAbiolinks package [12], and the

annotation file of the GENCODE27 version was used to convert fragments per kilobase million to transcripts per million. Only protein-encoding transcripts were considered. Clinical and survival information of EC patients in TCGA was extracted from pan-cancer data. The clinical information included age, sex, tumor stage, and EC subtypes defined by TCGA. For evaluation of survival, only overall survival (OS) was considered. The copy number variation (CNV) data of TCGA cohort were downloaded through Firehose.

Establishment of the prognostic model

Analysis of differentially expressed genes (DEGs) between 11 normal tissues and 141 tumor tissues in TCGA cohort was conducted, and DEGs were visualized by heatmaps and volcano plots. To construct an immune-related prognosis model, the immune-related gene sets were obtained from the IMMPORT network platform. Differential analysis of genes in TCGA database was performed, and DEGs were visualized by heatmaps and volcano plots. A genetic model related to the prognosis of EC was built by univariate and multivariate Cox regression analyses, and survival analysis and risk assessment were performed.

Consensus clustering of TIICs

CIBERSORT [13] is an algorithm that can be applied to describe the cell composition of samples through gene expression values in cancers. The CIBERSORT package and LM22 were used to calculate the infiltration of 22 types of immune cells in 141 EC samples. The immune and stromal content (immune and stromal scores) of each EC sample were evaluated by the ESTIMATE algorithm. Based on the immune infiltration patterns of EC samples, unsupervised hierarchical clustering was conducted using the "ConsensusClusterPlus" package [14]. The classification was repeated 1000 times to ensure stability.

Dimensionality reduction and generation of ICI score

Unsupervised clustering was conducted to group patients in TCGA cohort based on DEGs. The DEGs that were positively and negatively associated with clustering features were called ICI gene features A and B, respectively. The dimensionality of the gene signatures A and B was reduced by the Boruta algorithm, and principal component analysis was used to extract the principal component 1 as the signature score. The final signatures A and B were obtained.

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment analysis

GO and KEGG enrichment of signatures A and B were analyzed to explore their possible functions. The clusterProfiler package [15] was used for enrichment analyses, and the outcomes were displayed in bubble charts.

ICI score and immunotherapy response

Two independent data sets (IMvigor210 and TCGA) were obtained for analysis to identify the predictive value of ICI scores for the response to immunotherapy. In the IMvigor210 cohort (cds data), the ICI score was calculated based on the principal component scores of signatures A and B, and the high and low ICI groups were distinguished accordingly.

Collection and analyses of CNV and somatic mutation data

To analyze the CNV mutations of patients in TCGA database, the mutation data were obtained from TCGA. First, the somatic mutation differences between TCGA normal and tumor samples were analyzed. The genes and mutation frequencies that mutated on each chromosome were analyzed; these are presented in circle plots. The differential genes with significant correlation were identified by comparing the correlation between CNV and mRNA expressions. These genes may be the driving genes regulating the CNVs of EC. GO enrichment analysis was then performed on the identified genes.

GISTIC2.0 analysis of the downloaded CNV fragment was performed by GenePattern5 [16] using the default settings. Genomic regions that are notably amplified or deleted in a group of samples can be identified by GISTIC. Each aberration was allocated a G score, in which the magnitude of the aberration and its frequency in the sample were taken into account. For each important area within the threshold, a "peak" is generated, which represents the part with the highest variation amplitude and frequency in the abnormal area. In addition, a leave-one-out algorithm was used to determine the "broad peak" to take the marginal error in one single sample into account. Each significantly amplified or missing sample was identified, and the genes found in each "broad peak" region were listed. To identify the mutation load of EC, the total number of non-synonymous mutations in EC was calculated. The high or low ICI score was used to evaluate the somatic changes in EC driver genes. The "maftools" package [17] was applied to explore and visualize EC driver genes.

SNP mutation analysis in the ICGC database

The ICGC database includes data of 50 cancer types (or subtypes) regarding abnormal gene expression, somatic mutations, and clinical data. The ICGC includes 89 projects in 17 administrative regions in Asia, Australia, Europe, North, and South America and includes 25,000 tumor genomes. Two datasets of EC in the ICGC database were examined, including one dataset from China (ICGC-CN) and one dataset from the United Kingdom (ICGC-UK). Gene mutation analysis was performed, and a waterfall chart was drawn through the mutation matrix. Mutation site analysis was conducted, and survival analysis was conducted based on the top four mutated genes with the most significant differences. Mutations that were significantly related to the survival and prognosis of EC were obtained.

Statistical analysis

Data processes and analyses were performed using Excel (Microsoft) and R software (version 4.0.2). To compare two groups of continuous variables, the statistical significance of normally distributed variables was analyzed by independent Student's t-test, and the difference between non-normally distributed variables was analyzed by the Mann-Whitney U test (Wilcoxon rank-sum test). The Kruskal-Wallis test was conducted to compare more than two groups, while the Wilcoxon test was applied in the comparison of two groups. The X-tile software was used to test the possible cut points iteratively. The cut-off point with the largest rank statistic was selected, and patients in each data set were divided into two categories to reduce the calculation batch effect. In each data set, a Kaplan-Meier plotter was applied to generate survival curves for the subgroups. The log-rank test was applied to evaluate whether the differences were statistically significant. The chi-square test was applied to estimate the correlation mutation frequency between ICI score subgroups and somatic cells, and Spearman analysis was conducted to calculate the correlation coefficient. A two-tailed p-value < 0.05 was regarded as statistically significant.

Other procedures

Protocols for other procedures were provided in the Supplementary Material.

Results

Construction of an immune gene prognostic model for EC

We analyzed the DEGs between normal esophageal samples and EC tumor samples in TCGA, and the DEGs are displayed with heatmaps and volcano plots in Figs. 1A and 1B. Differential analysis of immune-related genes in TCGA database was conducted to determine an immune-related prognostic model. The DEGs were also visualized by heatmaps and volcano plots (Fig. 1C–1D). A genetic model related to EC prognosis was constructed through univariate (Fig. 2A) and multivariate Cox regression analyses (Table 1). Survival analysis (Fig. 2B) and risk assessment (Fig. 2C–D) using the model were performed. Survival analysis suggested a poor prognosis in the high-risk group of the immune gene model (*HSPA6, S100A12, FABP3, CACYBP, NOS2, DKK1, and STC2*). ROC curve analysis was conducted to verify the immune gene model prediction (Fig. 2C). The AUC was 0.778, which proved that the prediction accuracy of the prognostic model was strong.

Table 1
Multivariate Cox regression analysis risk model.

id	coef	HR	HR.95L	HR.95H	pvalue
HSPA6	0.0058	1.0058	1.0013	1.0103	0.0112
S100A12	0.0024	1.0024	1.0011	1.0037	0.0003
FABP3	0.0456	1.0466	1.0133	1.0811	0.0058
CACYBP	0.0307	1.0312	1.0009	1.0624	0.0435
NOS2	0.0172	1.0174	1.0069	1.0279	0.0011
DKK1	0.0081	1.0081	1.0005	1.0158	0.0373
STC2	0.0385	1.0392	0.9996	1.0804	0.0521

The immune and stromal score and ICI subtype clustering

The CIBERSORT and ESTIMATE algorithms were applied to quantify the immune-related scores in EC tumor samples from GSE106185 and four cohorts of TCGA. The ConsensusClusterPlus package of R software was used for unsupervised clustering, and we identified three independent ICI subtypes of EC patients: A, B and C (Fig. 3B). The correlation coefficients in the three ICI subtypes were analyzed, and a correlation heat map was created to visualize the ICIs in the TME (Fig. 3A). The immune cell interaction was also visualized based on the immune score and stromal score of the three ICI clusters (Fig. 3E). The ICI cluster A was characterized by naïve B cells, activated mast cells, neutrophils, resting NK cells, CD4 memory resting T cells, and CD4 naïve T cells. The ICI cluster B was marked by memory B cells, eosinophils, M2 macrophages, resting mast cells, monocytes, plasma cells, and CD4 memory resting T cells. The ICI cluster C was characterized by resting dendritic cells (DCs), M1 and M2 macrophages, activated NK cells, CD4 memory activated T cells, CD8 T cells, and follicular helper T cells. The results show that ICI clusters B and C were related to markedly high immune scores. We further examined the expressions of two important immune checkpoint molecules, PD1 and PD-L1, in the ICI subtypes. ICI cluster C showed significantly higher PD1 and PD-L1 expressions compared with clusters B and C, as determined by Kruskal–Wallis analysis (Fig. 3C–D).

Identification of immune gene subtypes

To explore the biological features of the different immunophenotypes, we conducted differential analysis using limma to identify the genetic differences between the subtypes (Fig. 4A). To reduce redundant genes, the Boruta algorithm was applied for dimensionality reduction. GO and KEGG analyses were conducted on the identified genes (Fig. 5B–C). The genes were enriched in microvillus organization, the cluster of actin-based cell projections, cortical actin cytoskeleton, actin binding, actin filament binding, steroid hormone receptor activity, and collagen-containing extracellular matrix. In addition, the three ICI clusters suggested significant differences in the expressions of PD1 and PD-L1. The ICI gene cluster A was related to higher PD1 expression, while the ICI gene clusters A and B were related to higher PD-L1 expressions (Fig. 4B–C). Gene cluster A was marked by M2 macrophages, resting mast cells, monocytes, and Tregs, and gene cluster A was related to significantly high immune scores. Gene cluster B was characterized by resting DCs, M0 and

M1 macrophages, activated NK cells, CD4 memory activated T cells, and CD4 naïve T cells. Gene cluster C was marked by activated mast cells, CD4 memory resting T cells, and Tregs (Fig. 4D).

Construction of the ICI score

To identify potential predictors of the ICI subtype in EC patients, principal component analysis was used to calculate the ICI score A of ICI signature gene A and the ICI score B of ICI signature gene B. The ICI score A and B of every patient in this survey were calculated as the sum of individual relevant individual scores. We obtained the prognostic signature score that was regarded as the ICI score. Patients in TCGA database were divided into two groups (high and low ICI scores) according to the optimal cut-off value obtained by X-tile software. The distribution of the patients in the three gene clusters is shown in Fig. 5A.

To examine the characteristics of immune gene subtypes, we also analyzed the msigdb.v7.0.symbols gene set and c7.all.v7.0.symbols immune-related gene set in the high and low ICI score groups, and the GSVA package was applied to explore the difference between these gene sets in different samples. The limma R package was applied to conduct differential analysis on the metabolic score (|log2FC |> 0.2 and adj. P < individual 0.05). The genes were enriched in the negative regulation of mast cell activation related to the immune response, NK cells, breast cancer, cellular extravasation, and immune-related aspects. There were significant differences between groups (Fig. 5D–E, Table 2–3).

Table 2
GSVA enrichment analysis of immune sub-component type in the TCGA data set (MsigDb).

	logFC	AveExp	t	adj.P	B
GO_NEGATIVE_REGULATION_OF_MAST_CELL_ACTIVATION_INVOLVED_IN_IMMUNE_RESPONSE	0.4742	0.0042	7.0331	0.0000	14.2491
GO_NEGATIVE_REGULATION_OF_MAST_CELL_ACTIVATION	0.4455	0.0035	7.0309	0.0000	14.2382
GO_REGULATION_OF_MAST_CELL_ACTIVATION	0.3064	0.0052	6.3793	0.0000	11.0498
GO_VASCULAR_SMOOTH_MUSCLE_CONTRACTION	0.3266	0.0148	6.3681	0.0000	10.9962
WANG_IMMORTALIZED_BY_HOXA9_AND_MEIS1_DN	0.3920	0.0316	6.3429	0.0000	10.8766
WIERENGA_STAT5A_TARGETS_DN	0.2391	0.0293	6.3226	0.0000	10.7800
SMID_BREAST_CANCER_NORMAL_LIKE_UP	0.3159	0.0170	6.2782	0.0000	10.5697
ROETH_TERT_TARGETS_DN	0.4658	0.0032	6.2407	0.0000	10.3927
GO_CELLULAR_EXTRAVASATION	0.2905	0.0062	6.0878	0.0000	9.6770
GSE45365_NK_CELL_VS_CD11B_DC_MCMV_INFECTION_DN	0.2227	0.0152	6.0836	0.0000	9.6574

Table 3
GSVA enrichment analysis of immune sub-component type in the TCGA data set (immune-related gene set).

	logFC	AveExp	t	adj.P	B
GSE43863_DAY6_EFF_VS_DAY150_MEM_LY6C_INT_CXCR5POS_CD4_TCELL_DN	-0.2531	-0.0299	-6.2723	0.0000	10.9377
GSE45365_NK_CELL_VS_CD11B_DC_MCMV_INFECTION_DN	-0.2227	-0.0152	-6.1675	0.0000	10.4210
GSE8921_UNSTIM_VS_TLR1_2_STIM_MONOCYTE_24H_UP	-0.2026	-0.0430	-5.9722	0.0000	9.4735
GSE17301_ACD3_ACD28_VS_ACD3_ACD28_AND_IFNA5_STIM_CD8_TCELL_DN	-0.2278	-0.0180	-5.8408	0.0000	8.8472
GSE7852_LN_VS_THYMUS_TREG_UP	-0.2138	-0.0143	-5.7981	0.0000	8.6460
GSE20715_0H_VS_24H_OZONE_LUNG_UP	-0.2070	-0.0138	-5.7791	0.0000	8.5567
GSE24142_EARLY_THYMIC_PROGENITOR_VS_DN2_THYMOCYTE_ADULT_UP	-0.2206	-0.0244	-5.7322	0.0000	8.3371
GSE21033_CTRL_VS_POLYIC_STIM_DC_12H_UP	-0.2080	-0.0214	-5.5658	0.0001	7.5677
GSE24142_EARLY_THYMIC_PROGENITOR_VS_DN2_THYMOCYTE_FETAL_UP	-0.2151	-0.0202	-5.5438	0.0001	7.4667
GSE19888_ADENOSINE_A3R_INH_VS_ACT_IN_MAST_CELL_DN	-0.2061	0.0080	-5.5122	0.0001	7.3229

GSEA suggested that the high ICI score subgroup was enriched in E2F targets, G2M checkpoint and MYC targets, while the low ICI score subgroup was enriched in IL6/JAK/STAT3 and KRAS signalling pathway, inflammatory response and other immune-related collections (Fig. 5F–G, Table S1). Before the prognostic value of the ICI score in the TCGA database and other independent data sets was determined, the immune activity and tolerance of the groups in the TCGA database were analyzed. CD274, CTLA4, HAVCR2, IDO1, LAG3, and PDCD1 were selected as immune checkpoint–related signals, and CD8A, CXCL10, CXCL9, GZMA, GZMB, IFNG, PRF1, TBX2, and TNF were selected as immune activity–related signals (Fig. 6A). We assessed the impact of the ICI score on prognosis using Kaplan–Meier analysis. Although the OS of patients in the high ICI score group was better than that of the low ICI score group, the results were marginally significant (Fig. 6C).

Correlation of ICI score with tumor mutation burden (TMB)

Given the important clinical significance of the TMB, we explored the potential relationship between TMB and ICI scores. The TMB of patients in the high ICI and low ICI score groups were compared, and the difference was not statistically significant (Fig. 6B). The OS of patients with high TMB was slightly better than that of patients with low TMB, but the results were not statistically significant (Fig. 6D), which may be because of insufficient sample size. Comparisons of pairs of the four groups, "high TMB, high ICI score," "high TMB, low ICI score," "low TMB, high ICI score," and "low TMB, low ICI score," revealed statistically significant differences (Fig. 6E). Overall, these results indicat that the ICI score may be used as a predictor independent of TMB and may measure the response to immunotherapy.

ICI score in predicting the response of immunotherapy

We next tested the utility of the ICI score in predicting the benefit of treatment for patients. Patients in the IMvigor210 cohort who received anti-PD-L1 immunotherapy were selected and divided into two groups based on high or low ICI scores. Patients with high scores were tended to achieve better treatment response than patients with low scores (P = 0.063), a marginally significant was observed (Fig. 7A). A higher ICI score was associated with the objective response to anti-PD-1 treatment in the IMvigor210 cohort (Fig. 7B). The survival probability in the high ICI score group was slightly higher than that in the low ICI score group (P = 0.059), which is marginally significant (Fig. 7C).

Analysis of CNVs and somatic variations

We examined differences in somatic mutations between normal esophageal and EC tumor samples in TCGA cohort. The genes mutated on each chromosome and their mutation frequencies were also analyzed. Circle diagrams were used to visualize the occurrence of CNVs on 23 chromosomes (Fig. 8A). By further comparing the correlation between CNVs and mRNA expression, we identified differential genes with significant correlation. These genes may be the driving genes regulating the CNVs of EC. GO enrichment analysis was conducted on these genes, and the results indicate that the driver genes were enriched in metabolic-related functions such as enzyme inhibitor activity and phosphoric ester hydrolase activity (Fig. 8B–C, Table S2).

We also evaluated the allocation of somatic variations of EC driver genes in the low and high ICI subgroups; the EC driver genes were screened by maftools (Fig. 9A–B). The results of the driver mutations may provide new insights for studying the mechanisms of tumor ICI components and gene mutations in immune checkpoint blocking therapy. To further analyze the clinical relevance of the ICI groups, two groups of immune-related genes were selected; their clinical relevance was explored, and a heat map was drawn (Fig. 9E–F). The CNVs in the TCGA cohort was analyzed. The CNV integrated software GISTIC2.0 was used, which was mainly used to detect significantly amplified or missed genomic regions in a set of samples. This method obtains the significantly amplified and missed regions in samples through significance calculation. To identify the mutation load of EC, we calculated the total number of non-synonymous mutations in EC. The high or low ICI score was used to evaluate the somatic changes in EC driver genes. The samples were divided into the high and low ICI group of the GISTIC scoring groups; the CNV of each group was shown, and the "maftools" package was used to identify and visualize EC driver genes (Fig. 9C–D).

SNP mutation analysis in the ICGC database

Two sets of data in the ICGC database from China (ICGC-CN) and the United Kingdom (ICGC-UK) were examined. Gene mutation analysis was performed, and waterfall diagrams were drawn (Fig. 10A, 11A). The mutation sites of the mutated genes were further analyzed. These sites are all amino acid sites in the chromosome with mutations. The position and mutation frequency of these mutation sites on the chromosome were visualized (Fig. 10B, 11B). Survival analysis was carried out based on the top four most significant gene mutations. Both groups were related to the survival prognosis of EC patients, and the wild-type gene had better prognosis than the mutant type (Fig. 10C–F, 11C–F).

Discussion

In our study, we successfully constructed a prognostic model based on immune-related genes that was used to predict the prognosis of EC patients. The accuracy of the model was verified by OS and ROC analyses. CIBERSORT was used to measure the infiltration level of various types of immunocytes in EC, and EC was divided into three immune subtypes. EC patients were also grouped into three gene clusters. The clusters with high immune and stromal scores tended to have more friendly immune-activated phenotypes and higher expression levels of PD-1 and PD-L1. Using GSEA, pathways related to immune response, including IL6/JAK/STAT3 and KARS pathways, were enriched in the low ICI score groups. GO and KEGG analyses revealed that the DEGs were mostly enriched in microvillus and actin binding. GSVA suggested that DEGs were enriched in aspects associated with immune response. In addition, OS analysis suggested that the ICI score may be used as a predictor independent of TMB. Interestingly, analyses in the IMvigor210 cohort suggested that patients with high scores tended to have better immunotherapeutic responses than patients with lower scores. The CNVs, somatic variations, and SNP mutation were also explored. The top eight genes with the most significant mutation frequency were identified (*SEZ6L, ZNF560, VWC2, CCDC178, SPHKAP, FREM2, PKHD1, ZFHX4*), and all correlated with the prognosis of EC patients. Our findings demonstrate that the seven-gene prognostic model, the ICI scores, and the mutated genes are effective prognostic markers and predictors for evaluating immunotherapy response in EC patients. Integration of the ICI patterns and expression patterns of genes related to immunity may reflect a potential strategy for determining individualized treatment.

A prognostic model comprising a seven-gene signature (*HSPA6, S100A12, FABP3, CACYBP, NOS2, DKK1* and *STC2*) was constructed with good specificity and sensitivity. Dickkopf-1 (*DKK1*) is a secreted glycoprotein that blocks the Wnt/β-catenin pathway. Overexpression of *DKK1* is closely related to cancer

development and poor survival in various cancers, including EC [18, 19]. *DKK1* overexpression in primary EC tumors was highly correlated with lymph node metastasis [20], which indicates a close relationship between *DKK1* and the prognosis of patients with EC. Stanniocalcin 2 (*STC2*) is a homologue of a glycoprotein hormone and closely correlated with rectal cancer [21], lung cancer [22] and ovarian cancer [23]. *STC2* is aberrantly expressed in EC with lymphatic metastasis and was verified to be an effective predictive marker for EC patients [24]. The prognostic model based on these seven genes may offer prognosis prediction and guide treatment decisions for EC patients.

Immunotherapies, including peptide vaccine, immune checkpoint suppression, and adoptive T cell transfer, show potential for the treatment for EC patients [25]. Pembrolizumab was proposed as an antineoplastic protocol for EC patients positive for PD-L1 [25]. However, identifying patients who are suitable for immunotherapy remains a critical issue [26]. Our findings suggested ICI scores as valid prognostic markers and predictors for evaluating immunotherapy response in EC patients. EC cells are highly immunogenic and could enhance anti-tumor immunity in the early phases of EC formation [27]. The infiltration of various immune cells and PD-L1 level have represented potential markers for predicting prognosis and immunotherapeutic reactivity in the EC immune landscape [28]. In this study, the ICI scores in 141 EC samples were analyzed and the samples were divided into three immune subtypes and three ICI gene clusters. The three ICI immune-subtypes and three ICI gene clusters had notable differences in the types of immune infiltrating cells, including CD4 + T cells, CD8 + T cells, DCs, NK cells, memory B cells, and tumor-associated macrophages. Among the ICI gene clusters, ICI gene cluster C had the lowest immune score and stromal score and had the highest resting NK cells and resting CD4 + memory T cells, which indicates an immune-cold phenotype. In contrast, the ICI gene clusters A and B showed higher immune scores and immune function cell infiltration. High stromal score and immune scores were related to increased infiltration of tumor-associated macrophages and resting DCs in ICI gene cluster B, which suggests a humoral immune response in cluster B [29]. Moreover, ICI gene clusters A and B had a more friendly immune-activated phenotype with the highest infiltration of CD8 + T cells, activated CD4 + T cells, and plasma cells [30]. The high immune and stromal scores in cluster A and B were also related to higher PD1 and PDL1 expressions in the two clusters, and patients in ICI clusters A and B may have a better response in the immunotherapy. The same phenomenon was also seen in the ICI immune-subtype clusters. The integrated analysis of the ICI clusters and immune-related gene expression model may provide an effective strategy for individualized treatment. The OS analysis between low and high ICI score clusters and TMB suggested that the ICI score may be a predictive marker independent of TMB. These results suggest that differences in immune infiltrating cells may be one of the elements contributing to the differences in the patient immune response. Our analysis demonstrated that high ICI scores were associated with immunotherapy response and better prognosis. In addition, a previous immune response could inhibit the occurrence and development of cancer and has positive impacts on the response to immunotherapy and prognosis.

Exploring the ICI patterns in individual tumors is critical because of the individual differences in the immune environment. Models based on tumor sub-specific markers for prognosis prediction have been well established in head and neck squamous carcinoma and breast cancer [31, 32]. In our study, we identified markers and ICI scores to quantify the ICI clusters. The relationship between gene mutations and immunotherapy response was confirmed [33, 34]. Analysis in the IMvigor210 cohort revealed that the ICI score was higher in patients with better immunotherapy response, which demonstrates the accuracy of its predictive ability. These results indicate that anti-PD1/PDL1 treatment may be effective for patients with high ICI scores.

GO and KEGG analyses were performed on the DEGs, and the outcomes indicated that they were mostly enriched in microvillus and actin binding. Interestingly, the niche formed by microvillus might be a shelter that protects tumor elimination by the immune system, and the niches are connected with desmosomes between tumor cells and no lymphocyte infiltration [35]. Actin-binding proteins, which play roles in multiple biological activities such as cell movement, cytokinesis and other biological processes [36], are closely associated with invasion and metastasis in tumor cells, DNA repair and transcription regulation [37]. Multiple actin-binding proteins were verified to be aberrantly expressed in various tumors, and were attributed to tumor invasion and metastasis [38]. GSVA enrichment analysis of immune sub-component types showed that they were enriched in aspects related to the immune response, including negative regulation of mast cell activation, and NK cells. GSEA analysis also revealed that IL6/JAK/STAT3 and KARS signaling pathways related to immune response, were enriched in the low ICI score groups. Due to the enrichment of the IL6/JAK/STAT3 and KARS signalling pathways in the low ICI score clusters, the inhibition of IL6/JAK/STAT3 and KARS signalling combined with immune checkpoint blockade may benefit patients with low ICI scores. These findings demonstrated that EC is tightly associated with immunity.

The well-recognized relevancy between precursor chronic inflammatory diseases and high gene mutation rates with about 3000–300,000 mutations per tumor provides the basic principles for the development of immunotherapy for EC [39]. We analyzed CNVs and somatic variations in normal esophageal tissue and EC samples in TCGA dataset, and genes with CNVs were labeled. By comparing the correlation between CNVs and mRNA expression, the driver genes that may regulate CNVs in EC were identified. GO analysis suggested that the driver genes were enriched in metabolism-related functions, including serine-type endopeptidase inhibitor activity, peptidase inhibitor activity and phosphoric ester hydrolase activity. Tumor cells undergo metabolic reprogramming during tumorigenesis to satisfy the requirements for enhanced biologic energy and biological synthesis and to alleviate the oxidative stress from increased proliferation and survival of tumor cells [40]. We evaluated the distribution of somatic variations in EC driver genes between low and high ICI subsets. The results revealed that differences in ICI clusters are associated with cancer heterogeneity. These results could provide new ideas to explore the mechanisms of the ICI clusters and gene mutation in the treatment of immunological examination points.

SNP mutation analysis in the ICGC-UK and ICGC-CN data sets revealed the top eight genes with the most significant mutation frequency (*SEZ6L, ZNF560, VWC2, CCDC178, SPHKAP, FREM2, PKHD1, ZFHX4*), and all related to the prognosis of EC patients. Qing et al. conducted an integrated analysis on the RNA-sequencing data of 442 EC patients to explore new predictive markers, and *ZFHX4* was identified to be one of the aberrant genes that were related to poor prognosis [41]. The authors examined the expression of *ZFHX4* in TCGA database and discovered that the aberrant expression of *ZFHX4* was also correlated to the poor prognosis of liver cancer patients [41]. The coiled-coil domain-containing protein 178 (*CCDC178*), a member of the coiled-coil domain-containing protein family, is aberrantly expressed in hepatocellular carcinoma [42] and gastric cancer [43]. A previous study showed that

CCDC178 facilitates the metastasis of hepatocellular carcinoma cells via ERK/MAPK signaling [44]. The SNP mutation analysis in our study may help identify additional effective prognosis biomarkers. The functions of the dysregulated genes obtained from online datasets also require further exploration.

This study has several limitations. First, several cohorts were used in our analyses, and the effect of inter-batch differences on the outcomes could not be avoided. Second, some of the outcomes of statistical analysis were not significant or marginally significant, which may be because of the small sample sizes. Further study with a larger sample size or more data sets is required. In addition, the mechanisms of the identified genes are unknown, and further in vitro and in vivo experiments are needed to explore their functions in EC. Finally, the results should be validated in a larger EC cohort treated with immunotherapy.

Conclusion

In conclusion, here we constructed a prognostic model that successfully predicts the prognosis of EC patients based on immune-related genes. EC was grouped into three ICI clusters, and the ICI clusters and ICI scores could be effective prognostic markers and predictors for quantifying immunotherapy response. The top eight genes with the most significant mutation frequency (*SEZ6L, ZNF560, VWC2, CCDC178, SPHKAP, FREM2, PKHD1, ZFHX4*) all correlated with prognosis of EC patients. Therefore, this study has important significance for the prognosis prediction in EC. In addition, our findings may provide novel insights into choosing efficient options for immunotherapy.

Abbreviation

CCDC178 coiled-coil domain-containing protein 178

CNV copy number variation

CR complete response

DCs resting dendritic cells

DEGs differentially expressed genes

DKK1 Dickkopf-1

EC Esophageal cancer

GEO Gene Expression Omnibus

GO Gene Ontology

ICGC International Cancer Genome Consortium

ICGC-CN International Cancer Genome Consortium database from China

ICGC-UK International Cancer Genome Consortium database from the United Kingdom

ICI immune cell infiltration

KEGG Kyoto Encyclopedia of Genes and Genomes

OS overall survival

PD progressive disease

PR partial response

SD stable disease

SNP single nucleotide polymorphism

STC2 Stanniocalcin 2

TCGA The Cancer Genome Atlas

TIICs tumor-infiltrating immune cells

TMB tumor mutation burden

TME tumor microenvironment

TNM Tumor Node Metastasis

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the National Center for Biotechnology Information Gene Expression Omnibus [45] repository under the accession number GSE106185, The Cancer Genome Atlas (TCGA), and International Cancer Genome Consortium (ICGC) databases.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported in part by grants from the Natural Science Foundation of Fujian Province, China (#2020J011123), Fujian Provincial Health and Family Planning Commission Research Talent Training Program, China (#2020GGA017), and China Association of Gerontology and Geriatrics (#2021-CSGOR-1).

Authors' contributions

YC and YX conceived and designed the whole project. YC and XH drafted the manuscript. XH and HP analyzed the data and prepared figures. XH, HP and GW performed the analyses. ZH and YZ prepared tables. TX and JC collected the data. GW, TX and JC provided specialized expertise. YX and ZK designed the whole project and revised the manuscript. All authors read and approved the final manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank Medjaden Bioscience Limited (MJD) for the paper revision support.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424. doi:10.3322/caac.21492.
2. Blom RL, Lagarde SM, van Oudenaarde K, Klinkenbijl JHG, Hulshof MC, van Laarhoven HW, et al. Survival after recurrent esophageal carcinoma has not improved over the past 18 years. Ann Surg Oncol. 2013;20(8):2693–8. doi:10.1245/s10434-013-2936-3.
3. Cao LJ, Zhang YJ, Dong SQ, Li XZ, Tong XT, Chen D, et al. ATAD2 interacts with C/EBPβ to promote esophageal squamous cell carcinoma metastasis via TGF-β1/Smad3 signaling. J Exp Clin Cancer Res. 2021;40(1):109. doi:10.1186/s13046-021-01905-x.
4. Pennathur A, Gibson MK, Jobe BA, Luketich JD. Oesophageal carcinoma. Lancet. 381(9864), 400 – 12. 2013; doi:10.1016/S0140-6736(12)60643-6.
5. Tan Z, Ma G, Yang H, Zhang L, Rong T, LinP. Can lymph node ratio replace pn categories in the tumor-node-metastasis classification system for esophageal cancer? J Thorac Oncol. 2014;9(8):1214–21. doi:10.1097/JTO.0000000000000216.
6. Sanmamed MF, Chen L. A Paradigm Shift in Cancer Immunotherapy: From Enhancement to Normalization. Cell. 2018;175(2):313–26. doi:10.1016/j.cell.2018.09.035.
7. Koury J, Lucero M, Cato C, Chang L, Geiger J, Henry D, et al. Immunotherapies: Exploiting the Immune System for Cancer Treatment. J Immunol Res. 2018(undefined), 9585614. 2018; doi:10.1155/2018/9585614.
8. Finck A, Gill SI, June CH. Cancer immunotherapy comes of age and looks for maturity. Nat Commun. 2020;11(1):3325. doi:10.1038/s41467-020-17140-5.
9. Huang TX, Fu L. The immune landscape of esophageal cancer. Cancer Commun (Lond). 2019;39(1):79. doi:10.1186/s40880-019-0427-z.
10. Lin EW, Karakasheva TA, Hicks PD, Bass AJ, Rustgi AK. The tumor microenvironment in esophageal cancer. Oncogene. 2016;35(41):5337–49. doi:10.1038/onc.2016.34.
11. Wang J, Sun J, Liu LN, Flies DB, Nie X, Toki M, et al. Siglec-15 as an immune suppressor and potential target for normalization cancer immunotherapy. Nat Med. 2019;25(4):656–66. doi:10.1038/s41591-019-0374-x.
12. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. 2016;44(8):e71. doi:10.1093/nar/gkv1507.

13. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. Methods Mol Biol. 2018;1711(undefined):243–59. doi:10.1007/978-1-4939-7493-1_12.
14. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010;26(12):1572–3. doi:10.1093/bioinformatics/btq170.
15. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284–7. doi:10.1089/omi.2011.0118.
16. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 2011;12(4):R41. doi:10.1186/gb-2011-12-4-r41.
17. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018;28(11):1747–56. doi:10.1101/gr.239244.118.
18. Lyros O, Lamprecht AK, Nie L, Thieme R, Götzel K, Gasparri M, et al. Dickkopf-1 (DKK1) promotes tumor growth via Akt-phosphorylation and independently of Wnt-axis in Barrett's associated esophageal adenocarcinoma. Am J Cancer Res. 2019;9(2):330–46.
19. Mazon M, Masi D, Carreau M. Modulating Dickkopf-1: A Strategy to Monitor or Treat Cancer? Cancers (Basel). 8(7), undefined. 2016; doi:10.3390/cancers8070062.
20. Otto B, Koenig AM, Tolstonog GV, Jeschke A, Klaetschke K, Vashist YK, et al. Molecular changes in pre-metastatic lymph nodes of esophageal cancer patients. PLoS One. 2014;9(7):e102552. doi:10.1371/journal.pone.0102552.
21. Huang F, Li H, Qin Z, Wang A, Zhang Y, Guo J, et al. SNHG17 Serves as an Oncogenic lncRNA by Regulating the miR-361-3p/STC2 Axis in Rectal Cancer. Front Genet. 2021;12(undefined):654686. doi:10.3389/fgene.2021.654686.
22. Liu YN, Tsai MF, Wu SG, Chang TH, Tsai TH, Gow CH, et al. Acquired resistance to EGFR tyrosine kinase inhibitors is mediated by the reactivation of STC2/JUN/AXL signaling in lung cancer. Int J Cancer. 2019;145(6):1609–24. doi:10.1002/ijc.32487.
23. Wu J, Lai M, Shao C, Wang J, Wei JJ. STC2 overexpression mediated by HMGA2 is a biomarker for aggressiveness of high-grade serous ovarian cancer. Oncol Rep. 2015;34(3):1494–502. doi:10.3892/or.2015.4120.
24. Kita Y, Mimori K, Iwatsuki M, Yokobori T, Ieta K, Tanaka F, et al. STC2: a predictive marker for lymph node metastasis in esophageal squamous-cell carcinoma. Ann Surg Oncol. 2011;18(1):261–72. doi:10.1245/s10434-010-1271-1.
25. Zhao Q, Yu J, Meng X. A good start of immunotherapy in esophageal cancer. Cancer Med. 2019;8(10):4519–26. doi:10.1002/cam4.2336.
26. Strizova Z, Snajdauf M, Stakheev D, Taborska P, Vachtenheim J, Biskup J, et al. The paratumoral immune cell signature reveals the potential for the implementation of immunotherapy in esophageal carcinoma patients. J Cancer Res Clin Oncol. 2020;146(8):1979–92. doi:10.1007/s00432-020-03258-y.
27. Dhupar R, Van Der KL, Pennathur A, Schuchert MJ, Nason KS, Luketich JD, et al. Targeting Immune Checkpoints in Esophageal Cancer: A High Mutational Load Tumor. Ann Thorac Surg. 2017;103(4):1340–9. doi:10.1016/j.athoracsur.2016.12.011.
28. Duan J, Xie Y, Qu L, Wang L, Zhou S, Wang Y, et al. A nomogram-based immunoprofile predicts overall survival for previously untreated patients with esophageal squamous cell carcinoma after esophagectomy. J Immunother Cancer. 2018;6(1):100. doi:10.1186/s40425-018-0418-7.
29. Biswas SK, Mantovani A. Macrophage plasticity and interaction with lymphocyte subsets: cancer as a paradigm. Nat Immunol. 2010;11(10):889–96. doi:10.1038/ni.1937.
30. Hamanishi J, Mandai M, Iwasaki M, Okazaki T, Tanaka Y, Yamaguchi K, et al. Programmed cell death 1 ligand 1 and tumor-infiltrating CD8 + T lymphocytes are prognostic factors of human ovarian cancer. Proc Natl Acad Sci U S A. 2007;104(9):3360–5. doi:10.1073/pnas.0611533104.
31. Zhang X, Shi M, Chen T, Zhang B. Characterization of the Immune Cell Infiltration Landscape in Head and Neck Squamous Cell Carcinoma to Aid Immunotherapy. Mol Ther Nucleic Acids. 2020;22(undefined):298–309. doi:10.1016/j.omtn.2020.08.030.
32. Callari M, Cappelletti V, D'Aiuto F, Musella V, Lembo A, Petel F, et al. Subtype-Specific Metagene-Based Prediction of Outcome after Neoadjuvant and Adjuvant Treatment in Breast Cancer. Clin Cancer Res. 2016;22(2):337–45. doi:10.1158/1078-0432.CCR-15-0757.
33. George S, Miao D, Demetri GD, Adeegbe D, Rodig SJ, Shukla S, et al. Loss of PTEN Is Associated with Resistance to Anti-PD-1 Checkpoint Blockade Therapy in Metastatic Uterine Leiomyosarcoma. Immunity. 2017;46(2):197–204. doi:10.1016/j.immuni.2017.02.001.
34. Burr ML, Sparbier CE, Chan YC, Williamson JC, Woods K, Beavis PA, et al. CMTM6 maintains the expression of PD-L1 and regulates anti-tumour immunity. Nature. 2017;549(7670):101–5. doi:10.1038/nature23643.
35. Arum CJ, Anderssen E, Viset T, Kodama Y, Lundgren S, Chen D, Zhao CM. Cancer immunoediting from immunosurveillance to tumor escape in microvillus-formed niche: a study of syngeneic orthotopic rat bladder cancer model in comparison with human bladder cancer. Neoplasia. 2010;12(6):434–42. doi:10.1593/neo.91824.
36. Svitkina T. The Actin Cytoskeleton and Actin-Based Motility. Cold Spring Harb Perspect Biol. 10(1), undefined. 2018; doi:10.1101/cshperspect.a018267.
37. Izdebska M, Zielińska W, Grzanka D, Gagat M. The Role of Actin Dynamics and Actin-Binding Proteins Expression in Epithelial-to-Mesenchymal Transition and Its Association with Cancer Progression and Evaluation of Possible Therapeutic Targets. Biomed Res Int. 2018(undefined), 4578373. 2018; doi:10.1155/2018/4578373.
38. Zhang YG, Niu JT, Wu HW, Si XL, Zhang SJ, Li DH, et al. Actin-Binding Proteins as Potential Biomarkers for Chronic Inflammation-Induced Cancer Diagnosis and Therapy. Anal Cell Pathol (Amst). 2021(undefined), 6692811. 2021; doi:10.1155/2021/6692811.

39. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013;499(7457):214–8. doi:10.1038/nature12213.
40. Martínez-Reyes I, Chandel NS. Cancer metabolism: looking forward. Nat Rev Cancer undefined(undefined) undefined. 2021. doi:10.1038/s41568-021-00378-6.
41. Qing T, Zhu S, Suo C, Zhang L, Zheng Y, Shi L. Somatic mutations in ZFHX4 gene are associated with poor overall survival of Chinese esophageal squamous cell carcinoma patients. Sci Rep. 2017;7(1):4951. doi:10.1038/s41598-017-04221-7.
42. Huang J, Deng Q, Wang Q, Li KY, Dai JH, Li N, et al. Exome sequencing of hepatitis B virus-associated hepatocellular carcinoma. Nat Genet. 2012;44(10):1117–21. doi:10.1038/ng.2391.
43. Zhang J, Huang JY, Chen YN, Yuan F, Zhang H, Yan FH, et al. Whole genome and transcriptome sequencing of matched primary and peritoneal metastatic gastric carcinoma. Sci Rep. 2015;5(undefined):13750. doi:10.1038/srep13750.
44. Hu X, Zhao Y, Wei L, Zhu B, Song D, Wang J, et al. CCDC178 promotes hepatocellular carcinoma metastasis through modulation of anoikis. Oncogene. 2017;36(28):4047–59. doi:10.1038/onc.2017.10.
45. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res. 2013;41:D991-5. doi:10.1093/nar/gks1193.

Figures

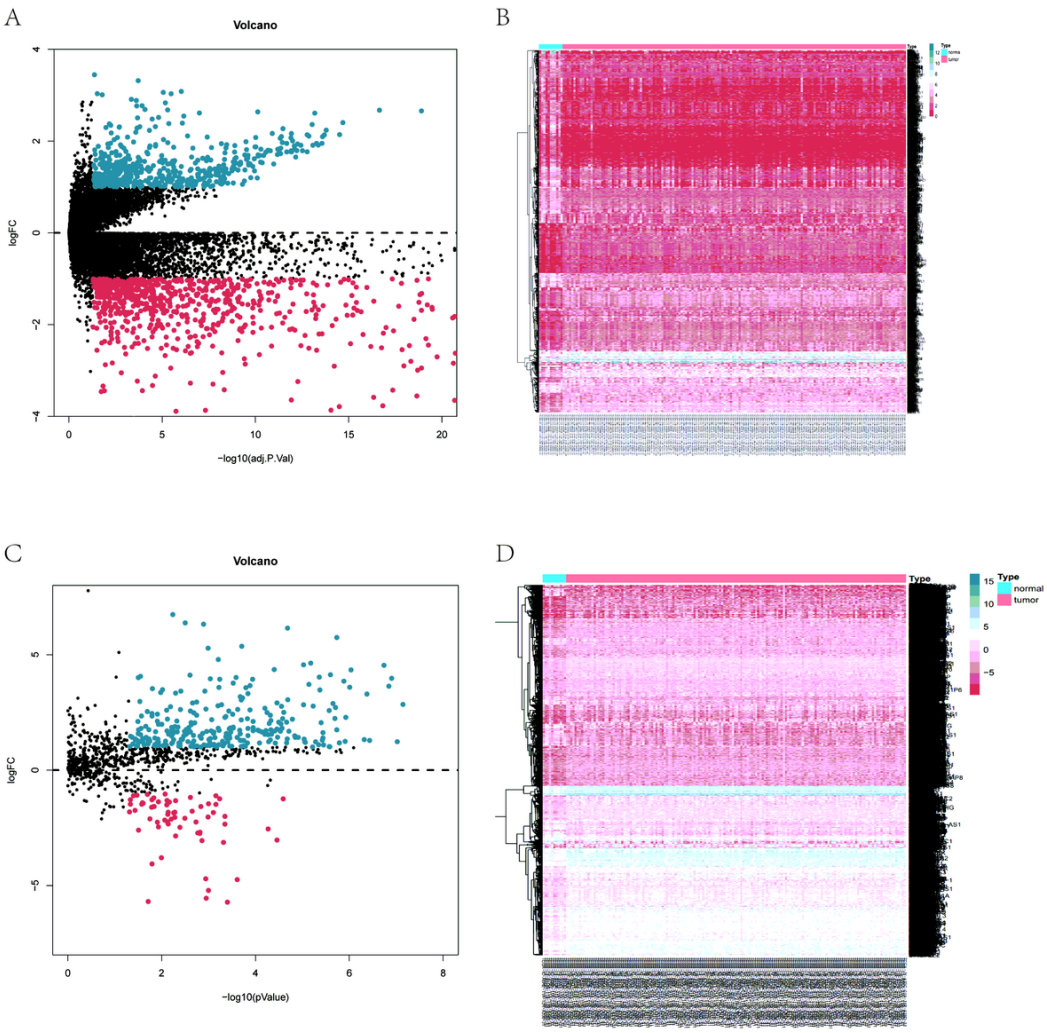

Figure 1

Differential analysis of TCGA cohort. Heatmaps and volcano plots of differential genes (A-B) and immune-related differential genes (C-D) between normal and tumor samples in the TCGA cohort.

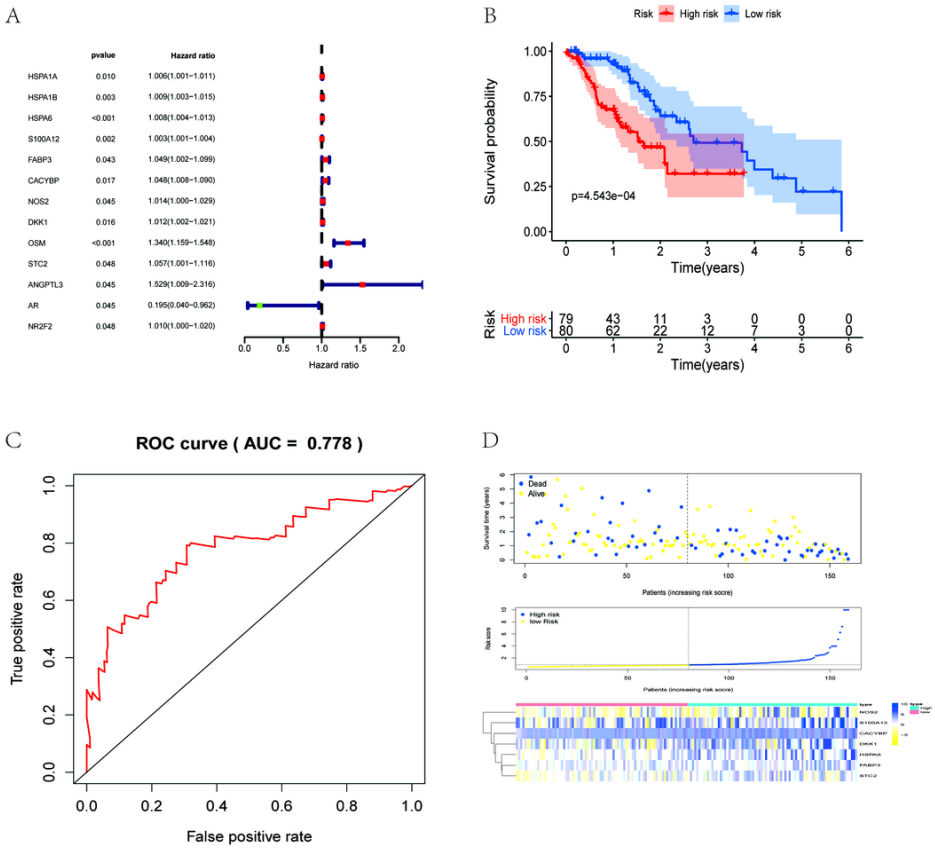

Figure 2

The prognosis of immune genes in TCGA cohort and construction of Cox risk regression model. (A) Prognostic value of genes in the prognostic medel obtained using univariate Cox analysis. (B) Kaplan-Meier survival curves of the overall survival (OS) of high- and low-risk patients. (C) ROC curve of the immune gene prognostic model in predicting the survival prognosis of EC. (D) The survival status (top), the risk curve (middle), and risk heat map (bottom) of the immune gene prognostic modal set.

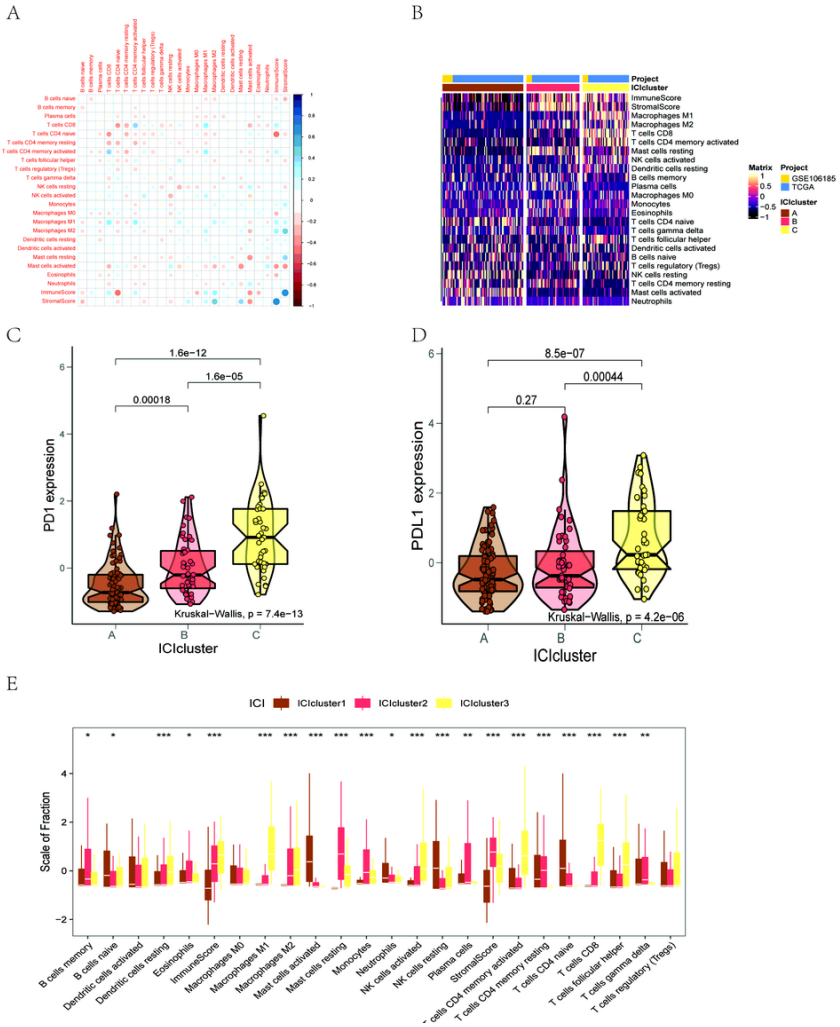

Figure 3

Infiltration of immune cells in EC's TME. (A) The proportions of tumor-infiltrating immune cells. (B) Unsupervised accumulation of tumor-infiltrating immune cells (TIICs) in 4 independent EC databases divided patients into three groups: ICI clusters A, B, and C (namely ICI clusters 1, 2, 3). Rows stand for TIICs, and columns represent samples. (C-D) PD1 (C) and PD-L1 (D) expression differences between different ICI clusters (Kruskal-Wallis test, p<0.0001). (E) The cell interaction of TIIC types, based on the immune and stromal scores of the three ICI clusters. The Kruskal-Wallis test was used. *p < 0.05; **p < 0.01; ***p < 0.001.

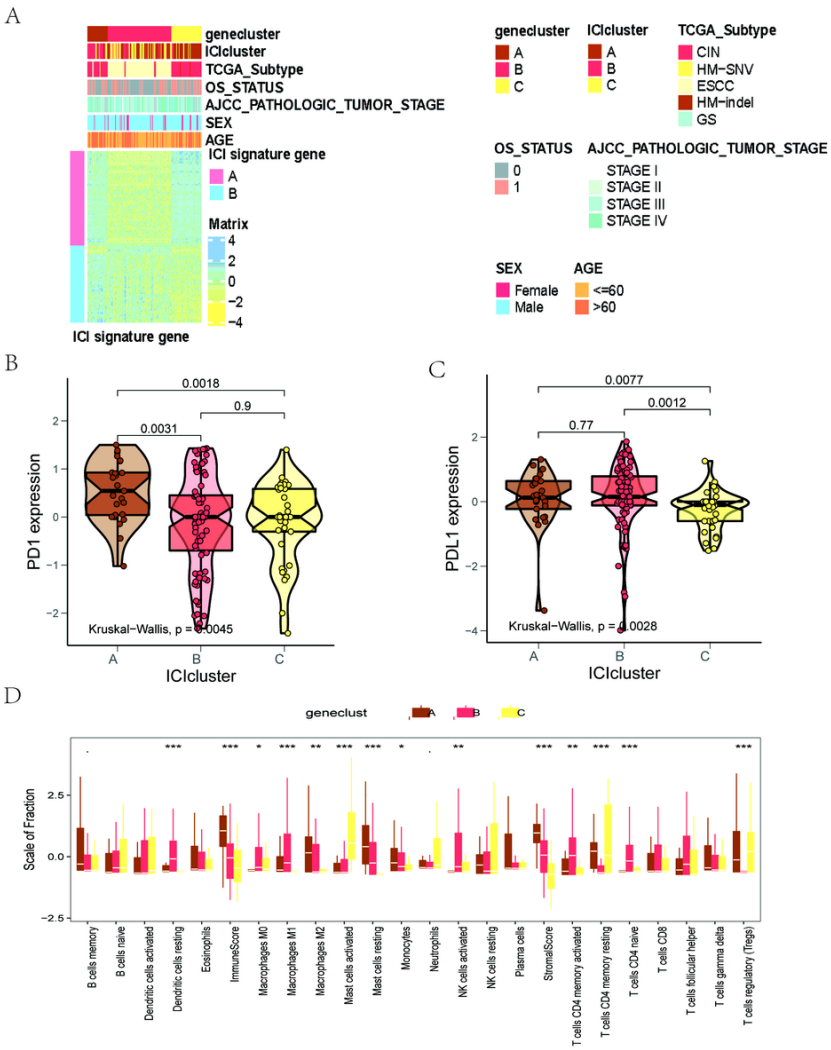

Figure 4

Identification of immunogenic gene subtypes. (A) The unsupervised clustering of DEGs in the 3 ICI clustering groups divided patients into three groups: gene clusters A, B, and C. (B-C) PD1 (B) and PD-L1 (C) expression differences among different ICI gene clusters (Kruskal-Wallis test, p<0.01). (D) The proportion of TIICs in the three gene clusters. The immune and stromal scores of the three ICI clusters were calculated. The Kruskal-Wallis test was used. *p < 0.05; **p < 0.01; ***p < 0.001.

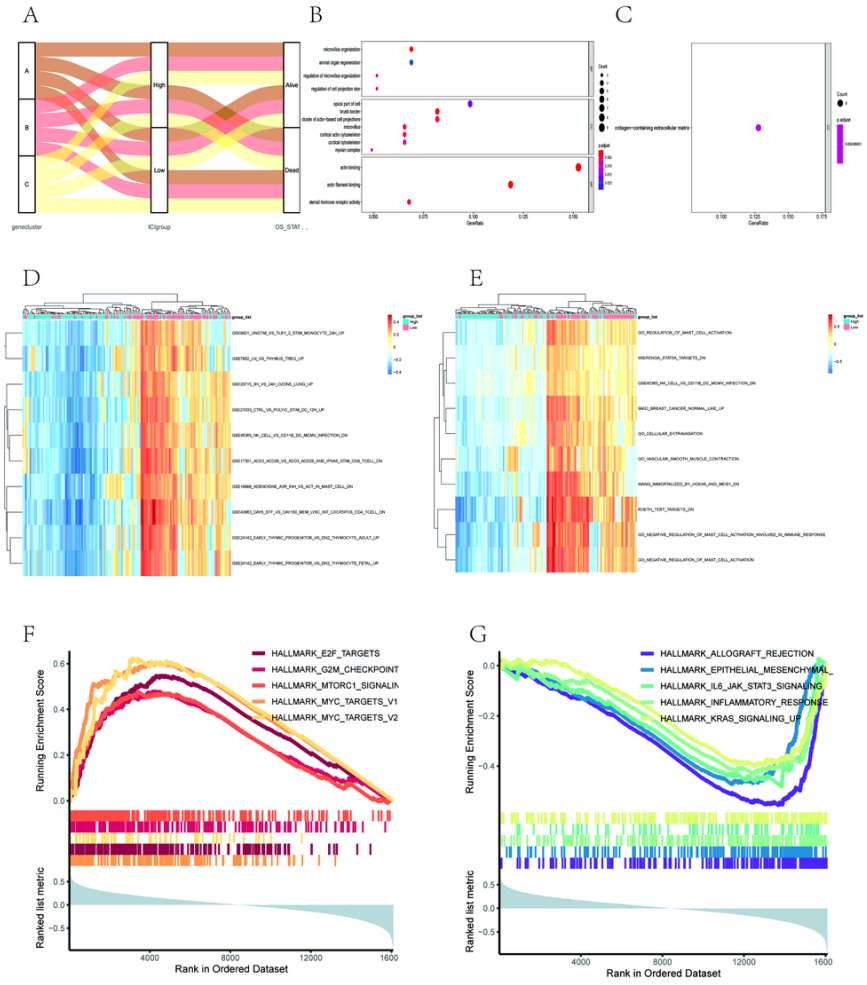

Figure 5

Construction of ICI score. (A) Alluvial plots of ICI gene cluster distribution in groups of different ICI clusters, ICI scores, and survival outcomes. (B-C) Enrichment analysis of two signature genes related to ICI: ICI signature genes A and B, including Gene Ontology (GO) analysis (B) and KEGG analysis (C). (D-E) GSVA analysis of immune sub-component type in the TCGA data set (MsigDb) (D) and immune-related gene set (E). (F)Enrichment of subgroups with high ICI scores. (G) Enrichment of subgroups with low ICI scores.

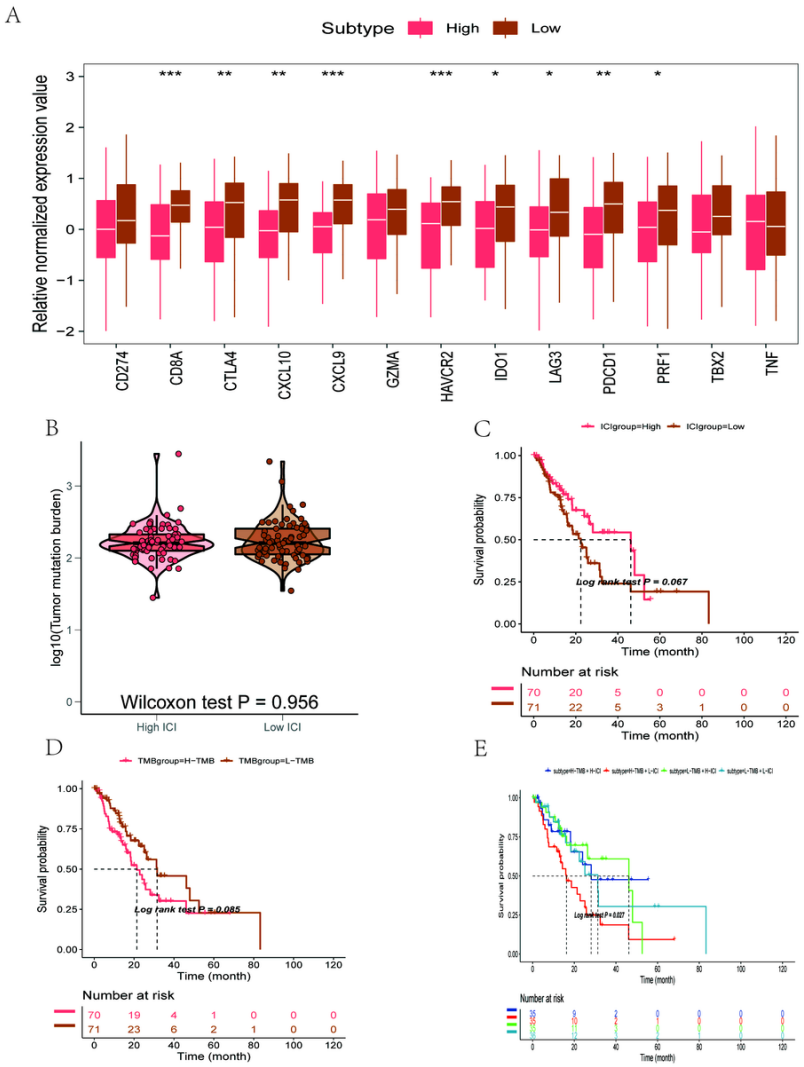

Figure 6

Analysis of ICI score subgroups. (A) The expression of immune checkpoint-related genes and immune activation-related genes in the high and low ICI score subgroups. (B) The differences in tumor mutation burden (TMB) between high and low ICI score subgroups. (C-D) Kaplan-Meier curves of the high and low ICI (C) and TMB (D) groups in the TCGA database. (E) Kaplan-Meier curve of patients in the TCGA cohort stratified by both TMB and ICI scores. Log-rank test, p = 0.027.

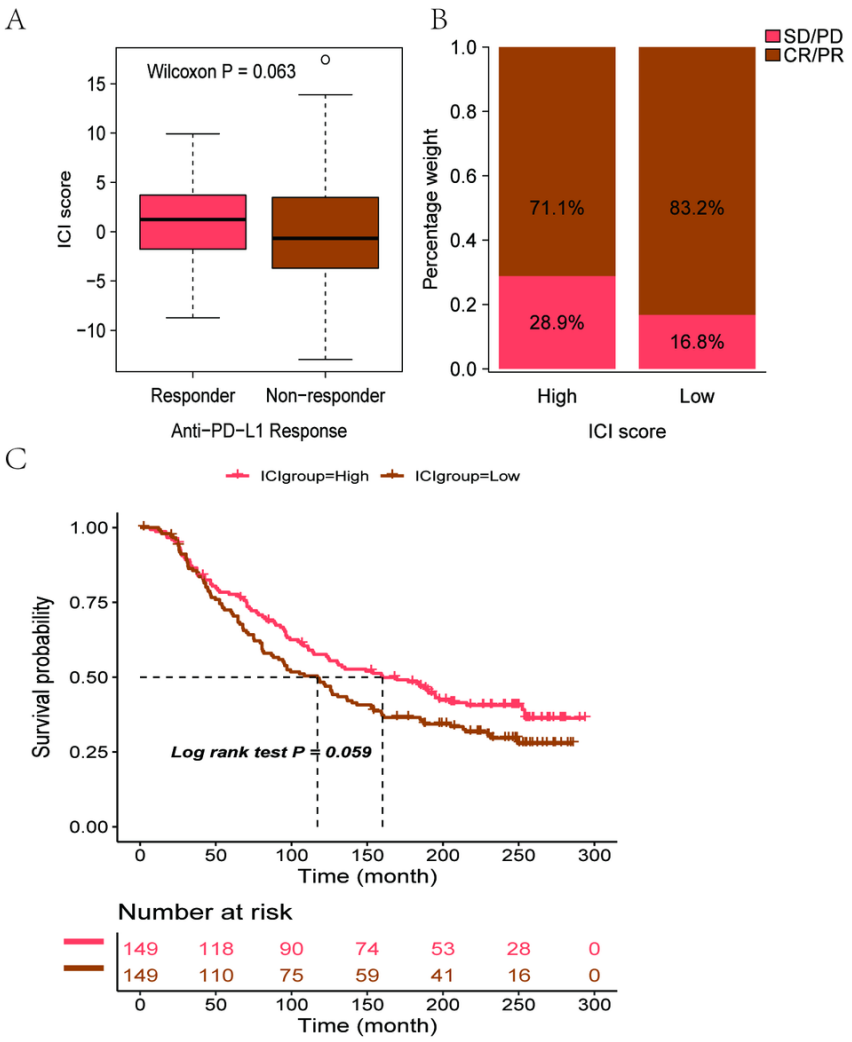

Figure 7

The role of ICI score in predicting the benefit of immunotherapy. (A) ICI scores of groups with a different anti-PD-1 response. (B) In the IMvigor210 cohort, the clinical response rate of anti-PD-L1 immunotherapy in the high or low ICI score group (complete response [CR] / partial response [PR] and stable disease [SD] / progressive disease [PD]). (C) Kaplan Meier curves of patients with high and low ICI scores in the IMvigor210 cohort.

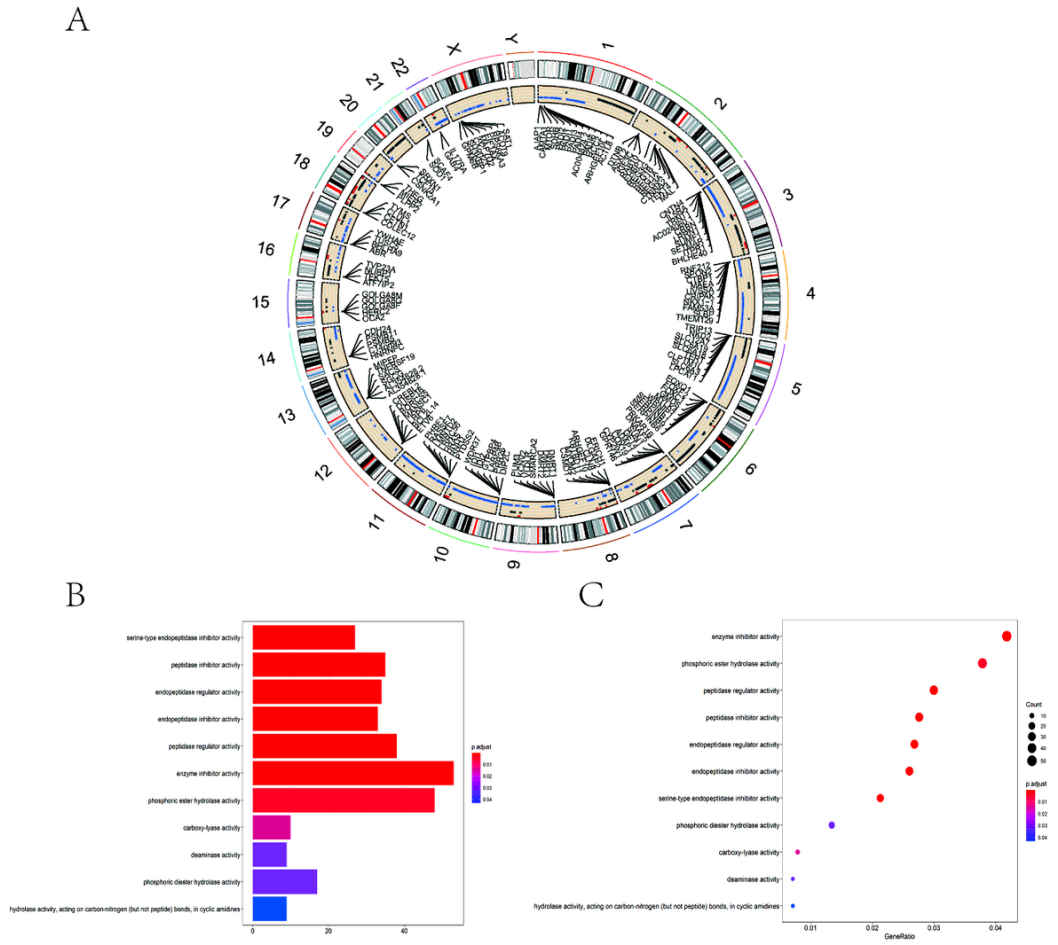

Figure 8

The TCGA somatic mutation analysis. (A) The somatic differences between normal and tumor samples in the TCGA were visualized in a circle graph. The circle graph represents the occurrence of CNV on 23 chromosomes. The black stands for the CNV increase, and the blue represents the CNV decrease. Genes with copy number variation were marked. (B) A bar graph of GO analysis of genes which regulates CNV-driven. The length of the column stands for the amount of gene enrichment. The color stands for significance, and the significance gradually increases from blue to red. (C) The bubble chart of GO analysis of genes which regulates CNV-driven. The size of the bubble represents the amount of gene enrichment. The color stands for significance, and the significance increases from blue to red.

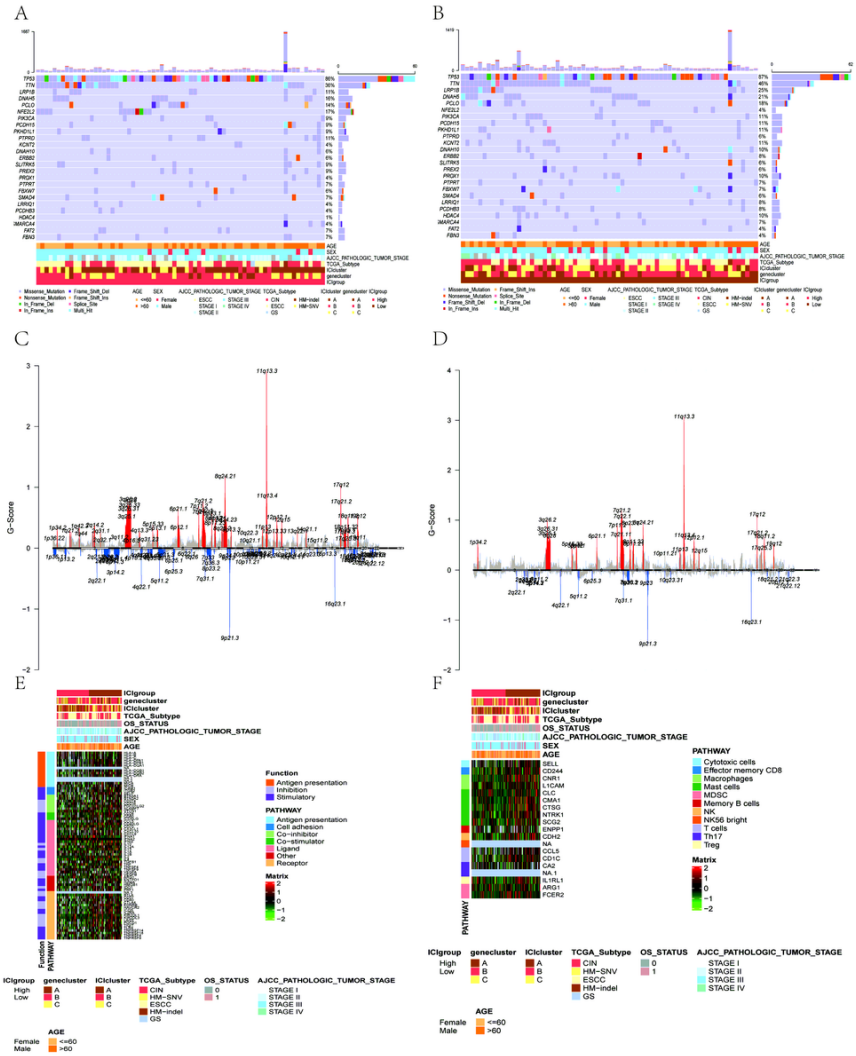

Figure 9

The analyses of TCGA tumor samples' copy number and somatic variation. (A) Waterfall chart of driver mutation in high ICI score (red) group. (B) Waterfall chart of driver mutation in low ICI score (green) group. Each column represents a sample. (C-D) The analysis of copy number change (CNV) in TCGA-BLCA cohort based on GISTIC2.0, were the GISTIC score of the high ICI group (C) and the low ICI group (D), respectively. (E-F) The clinical correlation heat map of the two groups of immune-related genes of interest in the ICI group.

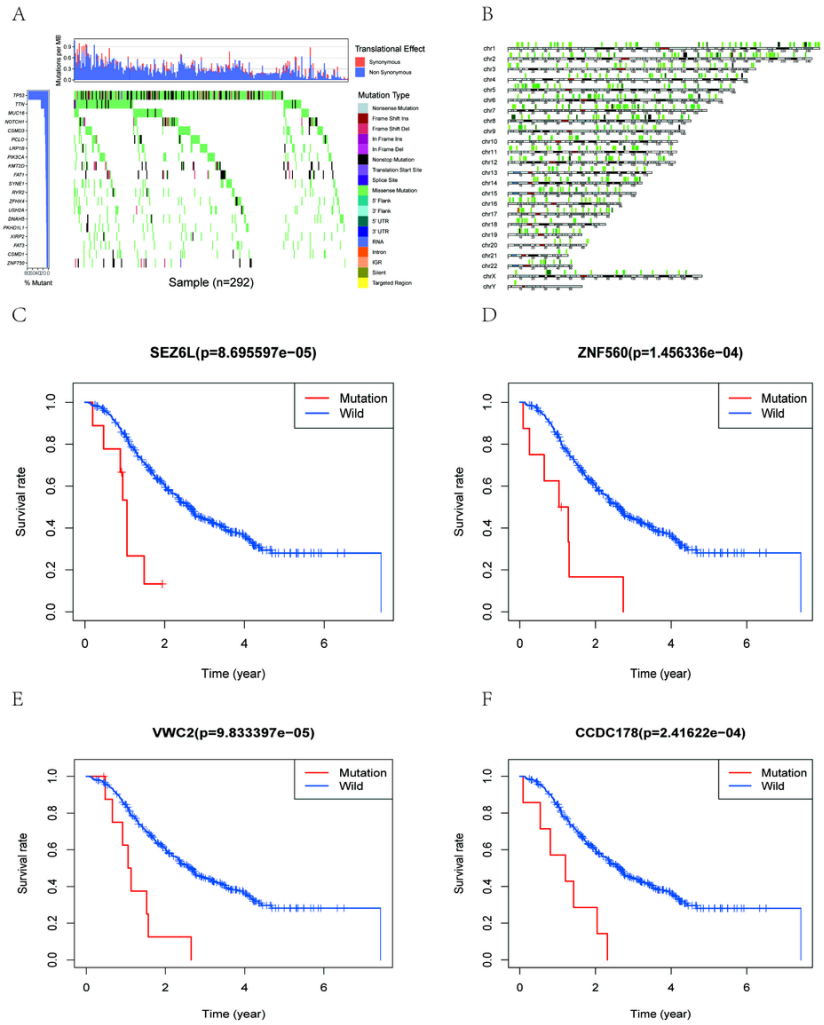

Figure 10

SNP mutation analysis of ICGC-UK data set. (A) waterfall chart of gene mutations in the ICGC-UK database, red stands for synonymous mutations, and blue represents non-synonymous mutations, and different colors stand for different types of mutations. (B) Analysis of the amino acid position of the mutant gene in the chromosome, the frequency of mutation gradually increased from light green to red. (C-F) The survival curves of the top 4 mutant genes with the most significant difference (SEZ6L, ZNF560, VWC2, CCDC178), blue for wild-type genes, and red for mutant-type genes.

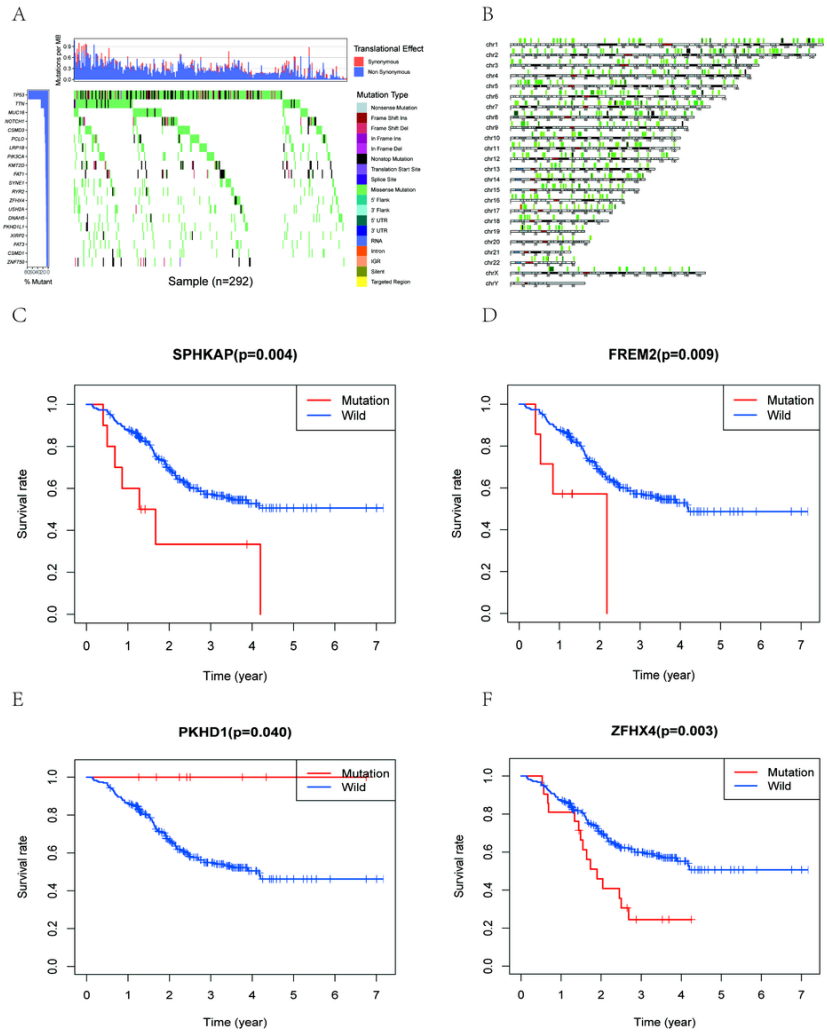

Figure 11

SNP mutation analysis of ICGC-CN data set. (A) Waterfall chart of gene mutations in the ICGC-CN dataset, red stands for synonymous mutations, and blue stands for non-synonymous mutations, and different colors represent different types of mutations. (B) Analysis of the amino acid position of the mutant gene in the chromosome, the frequency of mutation gradually increased from light green to red. (C-F) The survival curves of the top 4 mutant genes with the most significant difference (SPHKAP, FREM2, PKHD1, ZFHX4), blue for wild-type genes, and red for mutant-type genes.

Supplementary Files

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

- SupplementaryMaterial.docx