

# Identification of prognostic immune-related genes in the tumor microenvironment of endometrial cancer

Peigen Chen<sup>1</sup>, Yuebo Yang<sup>1</sup>, Yu Zhang<sup>1</sup>, Senwei Jiang<sup>1</sup>, Xiaomao Li<sup>1</sup>, Jing Wan<sup>1</sup>

<sup>1</sup>Department of Gynecology, The Third Affiliated Hospital of Sun Yat-Sen University, Guangzhou, Guangdong Province, China

**Correspondence to:** Jing Wan, Xiaomao Li; **email:** [wanjing@mail.sysu.edu.cn](mailto:wanjing@mail.sysu.edu.cn), [lixmao@mail.sysu.edu.cn](mailto:lixmao@mail.sysu.edu.cn)

**Keywords:** endometrial cancer, tumor microenvironment, prognosis, immune score, TCGA

**Received:** October 25, 2019

**Accepted:** January 27, 2020

**Published:** February 19, 2020

**Copyright:** Chen et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Endometrial cancer (EC) is one of the most common gynecologic malignancies. To identify potential prognostic biomarkers for EC, we analyzed the relationship between the EC tumor microenvironment and gene expression profiles. Using the ESTIMATE R tool, we found that immune and stromal scores correlated with clinical data and the prognosis of EC patients. Based on the immune and stromal scores, 387 intersection differentially expressed genes were identified. Eight immune-related genes were then identified using two machine learning algorithms. Functional enrichment analysis revealed that these genes were mainly associated with T cell activation and response. Kaplan-Meier survival analysis showed that expression of TMEM150B, CACNA2D2, TRPM5, NOL4, CTSW, and SIGLEC1 significantly correlated with overall survival times of EC patients. In addition, using the TIMER algorithm, we found that expression of TMEM150B, SIGLEC1, and CTSW correlated positively with the tumor infiltration levels of B cells, CD8+ T cells, CD4+ T cells, macrophages, and dendritic cells. These findings indicate that the composition of the tumor microenvironment affects the clinical outcomes of EC patients, and suggests that it may provide a basis for development of novel prognostic biomarkers and immunotherapies for EC patients.

## INTRODUCTION

Endometrial cancer (EC) is one of the most common gynecologic malignancies, and the fourth most common cancer (about 4.8% of all cancers) in women [1]. EC affects mainly post-menopausal women [2]. The routine treatment for EC includes surgery, radiotherapy, chemotherapy, and hormonal therapy. When the disease is confined to the uterus, EC patients have a relatively good prognosis, with a 5-year survival rate of 95 %. However, when distant metastases are present at the time of diagnosis, the 5-year survival rate is only 17 %, and patients respond poorly to conventional therapies. Thus, it is critical to identify prognostic biomarkers for EC, and develop therapies that are more effective for patients with advanced forms of EC.

Tumor microenvironment (TME), the site where the tumor is located, consists of immune cells,

mesenchymal cells, endothelial cells, inflammatory mediators, and extracellular matrix (ECM) molecules [3, 4]. TME has a significant impact on tumor growth, chemoresistance, and clinical outcomes [5–9]; however, relatively little is known about the impact of TME on endometrial cancer. Infiltrating stromal and immune cells are the major components of TME, and play an essential role in cancer biology. The ESTIMATE (Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data) algorithm uses gene expression data to estimate the levels of infiltrating stromal and immune cells, and tumor purity. The predictive ability of this method has been validated in large and independent datasets [6].

Machine learning is a form of artificial intelligence that can automatically analyze patterns from sample data, and make corresponding predictions. Due to its

accuracy and predictive performance, the machine learning algorithm is used in different fields, including medical diagnostics [10]. The commonly used machine learning algorithms include SVM, KNN, LASSO, and Random forest.

Knowledge of the TME composition is critical to understand the interactions between cancer and immune cells, and the impact the immune system has on tumor behavior [11]. In this study, we used the ESTIMATE and TIMER (Tumor Immune Estimation Resource) [12] algorithms, to perform a comprehensive analysis of immune cells and genes in the TME of endometrial carcinoma, and to correlate the data to clinical outcomes and prognosis of EC patients. Our results indicate that the TME composition affects the clinical outcomes of EC patients, suggesting that it might provide a basis for development of new prognostic biomarkers and therapies, especially immunotherapies, for EC patients.

## RESULTS

### Immune and stromal scores correlate with EC clinical data and prognosis

In total, data from 521 EC patients, and 19459 RNAs extracted from RNA-seq data according to ENSEMBL Genomes (hg38), were analyzed in this study. Based on the gene expression data, immune and stromal scores were calculated using the ESTIMATE algorithm (Supplementary Table 1).

Based on the clinical data extracted from TCGA-CDR (Supplementary Table 2) and using Wilcoxon signed-rank test, we found that both immune and stromal scores of grade 3 (G3;  $n=319$ ) and high-grade ( $n=11$ ) EC were significantly lower compared to grade 1 (G1;  $n=96$ ) and grade 2 (G2;  $n=119$ ) groups ( $p=0.03$ ,  $p=0.04$ ). In addition, the scores of high grade patients were lower than the scores of grade 3 patients ( $p=0.03$ ,  $p=0.04$ ) (Figure 1A, 1B). Based on a classification by the International Society of Gynecological Pathologists, the clinical outcomes of grade 1 and grade 2 patients are better than grade 3 and high grade patients, and the high grade patients have even worse prognosis than the grade 3 patients [13]. As shown in Figure 1C and 1D, the immune and stromal scores were associated with the EC pathological subtype: endometrioid endometrial adenocarcinoma had higher immune and stromal scores than serous endometrial adenocarcinoma. In addition, when we compared the immune and stromal scores between patients with a new tumor event ( $n=115$ ) and without new tumor event ( $n=370$ ) after

initial treatment, patients without a new tumor event had higher immune and stromal scores, although this did not reach a statistical significance (Figure 1E, 1F).

According to TCGA-CDR, a progression-free interval (PFI) is defined as the time until patients develop a new tumor event, including recurrence of disease and distant metastasis [14]. To determine whether there is a correlation between the immune and stromal scores, the overall survival (OS) time, and the PFI of EC patients (Figure 2A–2D), EC patients were classified into a high score group ( $n(\text{stromal group})=243$ ,  $n(\text{immune group})=242$ ), and a low score group ( $n(\text{stromal group})=242$ ,  $n(\text{immune group})=243$ ) based on the median of scores, and Kaplan–Meier survival curve was used to analyze the correlation. We found that the high immune score positively correlated with both OS ( $p=0.01$ ) and PFI ( $p=0.04$ ) (Figure 2A, 2C), and the high stromal score positively correlated with OS ( $p=0.042$ ) (Figure 2B).

### Identification of differentially expressed genes (DEGs)

To identify the immune-related and stromal-related genes, differential analysis by using “limma” package was performed (Supplementary Table 3). 552 genes were upregulated in the high immune score group (purple circle in Figure 3A), and 690 genes were upregulated in the high stromal score group (red circle in Figure 3A). At the same time, 164 genes were downregulated in the high immune score group (purple circle in Figure 3B), and 43 genes were downregulated in the high stromal score group (red circle in Figure 3B). 387 intersection genes were selected for further analysis (overlap zone in Figure 3A, 3B).

### Enrichment analysis of intersection genes

Using the “clusterProfiler” R package, 711 Gene Ontology (GO) terms and 36 Kyoto Encyclopedia of Genes and Genomes (KEGG) terms were indicated (Supplementary Table 4). The results showed the top 10 biological processes GO terms, cellular component GO terms, molecular function GO terms (Figure 4A), and the top 20 KEGG pathway terms (Figure 4C). The correlation between the intersection genes and the top 5 biological processes, including T cell activation, regulation of lymphocyte activation, regulation of T cell activation, leukocyte adhesion, and positive regulation of cell activation is shown in Figure 4B. The KEGG analysis showed that the intersection genes were associated with immune responses, especially T cell responses.

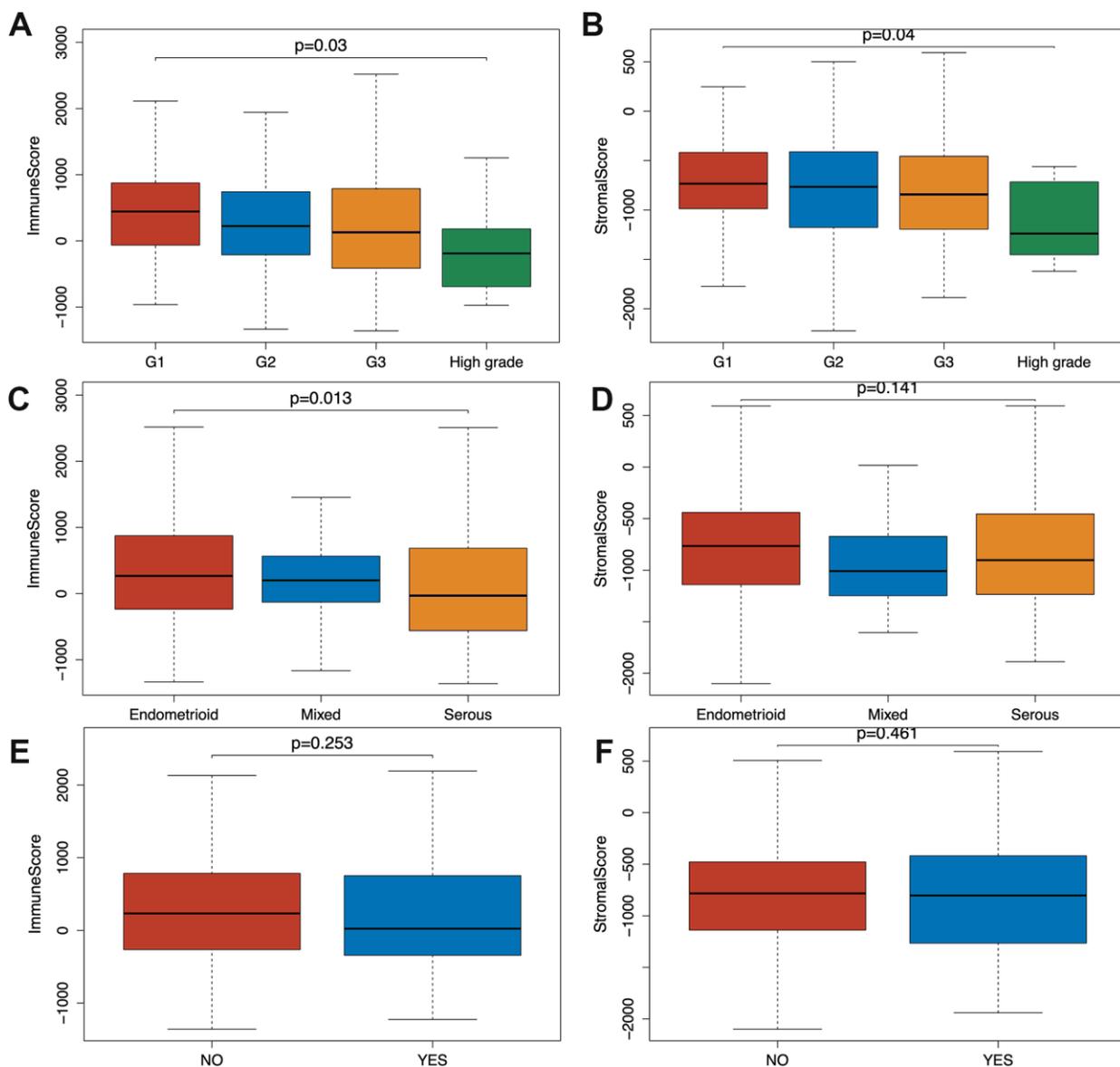
## Analysis of protein-protein interactions (PPI) among intersection genes

PPI network with 384 nodes and 1784 edges was constructed using the STRING website (Figure 5A). Using the MCODE software we found modules in the network; modules including at least 10 nodes were selected (Figure 5B, Module 1; Figure 5C, Module 2). GO and KEGG analyses of module 1 (Figure 5B) by ClueGo are shown in Supplementary Figure 2A and 2C. Likewise, GO and KEGG analyses of module 2 (Figure 5C) by ClueGo are shown in Supplementary Figure 2B and 2D. The results demonstrated that the module 1 was

mainly enriched in leukocyte migration (59.84%; Supplementary Figure 2A), Toll-like receptor signaling pathway (40%), and chemokine signaling pathway (40%; Supplementary Figure 2C). Module 2 was mainly enriched in T cell co-stimulation (72.28%; Supplementary Figure 2B) and cell adhesion molecules (40.54%; Supplementary Figure 2D).

## Identification of TME associated genes based on machine learning

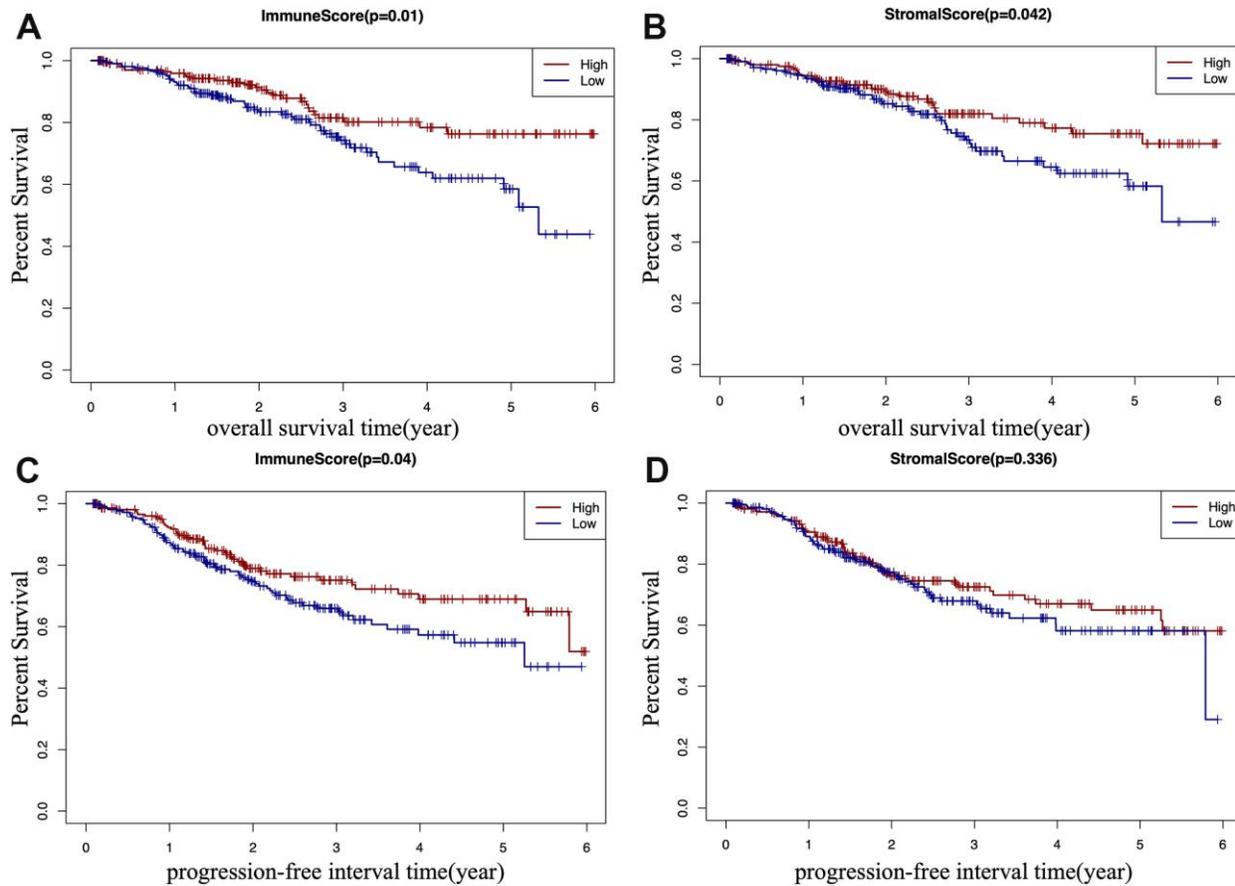
To identify the TME associated genes, we performed two different machine learning algorithms, LASSO



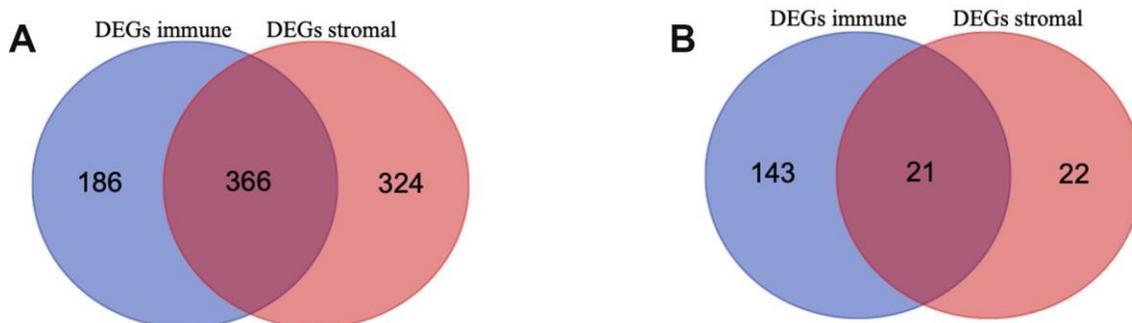
**Figure 1. Relationship between immune and stromal scores and EC clinical and pathological data.** (A, B) Distribution of immune and stromal scores of EC grades. (C, D) Distribution of immune and stromal scores of EC pathologic type, including endometrioid cancer, serous cancer and mix type. (E, F) Distribution of immune and stromal scores of new tumor event after initial treatment of EC.

algorithm and Random forest algorithm. By LASSO algorithm, 12 genes were identified (Figure 6B); by Random forest algorithm, 50 genes were identified (Figure 6A). The ROC curve of LASSO (Supplementary Figure 1A, AUC:0.753) and Random

forest test (Supplementary Figure 1B, AUC:0.960) was used. An overlap between the above two groups identified 8 TME associated genes (AQP4, ARHGAP36, CACNA2D2, CTSW, NOL4, SIGLEC1, TMEM150B and TRPM5) (Figure 6C).



**Figure 2. Kaplan-Meier (KM) survival curve of EC patients based on their immune/stromal scores.** Patients were classified into high immune/stromal scores groups and low immune/stromal scores groups. (A) The KM curve of overall survival (OS) time of high and low immune score group. (B) The KM curve of OS time of high and low stromal score group. (C) The KM curve of progression-free interval (PFI) time according to immune scores. (D) The KM curve of progression-free interval (PFI) time according to stromal scores.



**Figure 3. Differentially Expressed Genes (DEGs) selected (A, B) Venn diagram of differentially expressed genes (DEGs) base on immune and stromal score. (A) shows the commonly upregulated DEGs and (B) shows the commonly downregulated DEGs.**

## Predictive signature construction and survival analysis

Multivariate Cox regression analysis performed by “survival” R package was subsequently used to construct a predictive signature using the above 8 TME associated genes. The risk score of the prognostic signature was then calculated according to the formula:

$$\text{risk score} = \sum_{i=1}^n \beta_i * x_i$$

( $\beta$  stands for the regression coefficient) [15].

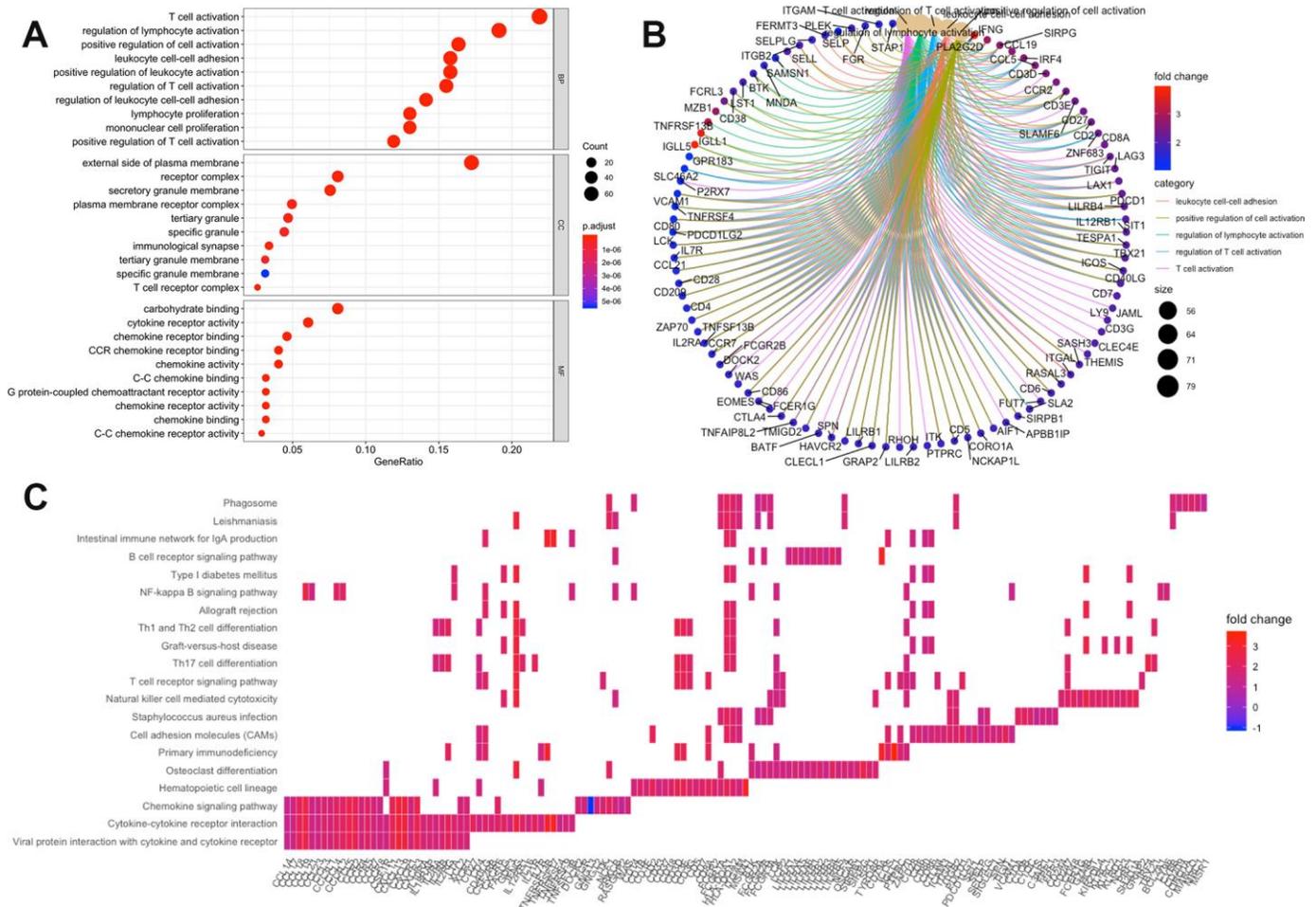
By Kaplan-Meier survival analysis, we found that this 8-gene signature was associated with OS ( $p < 0.0001$ ) (Figure 7A); this was validated by the ROC curve (AUC of 3 year: 0.756, AUC of 5 year: 0.797) (Figure 7B).

We also analyzed the association between the 8 genes and OS using the Kaplan-Meier survival analysis. We found that the high levels of TMEM150B ( $p = 0.037$ ),

CACNA2D2 ( $p = 0.0098$ ), TRPM5 ( $p = 0.031$ ) and NOLA ( $p = 0.0018$ ) negatively correlated with OS (Figure 7C–7G), while the high levels of CTSW ( $p = 0.0029$ ) and SIGLEC1 ( $p = 0.043$ ) positively correlated with OS (Figure 7H).

## Immune cells infiltration analysis

To determine whether there is a correlation between tumor infiltration with immune cells, and immune-related gene expression, the tumor infiltration with six types of immune cells (CD4 + T cells, CD8 + T cells, B cells, neutrophils, macrophages, and dendritic cells) was analyzed by TIMER (Supplementary Table 5). Figure 8 shows the correlation between the immune cell infiltration and the expression of prognostic genes. The expression of CTSW positively correlated with the infiltrating levels of B cells (partial.cor=0.586,  $p = 5.44e-28$ ), CD4+ T cells (partial.cor=0.499,  $p = 1.08e-19$ ),

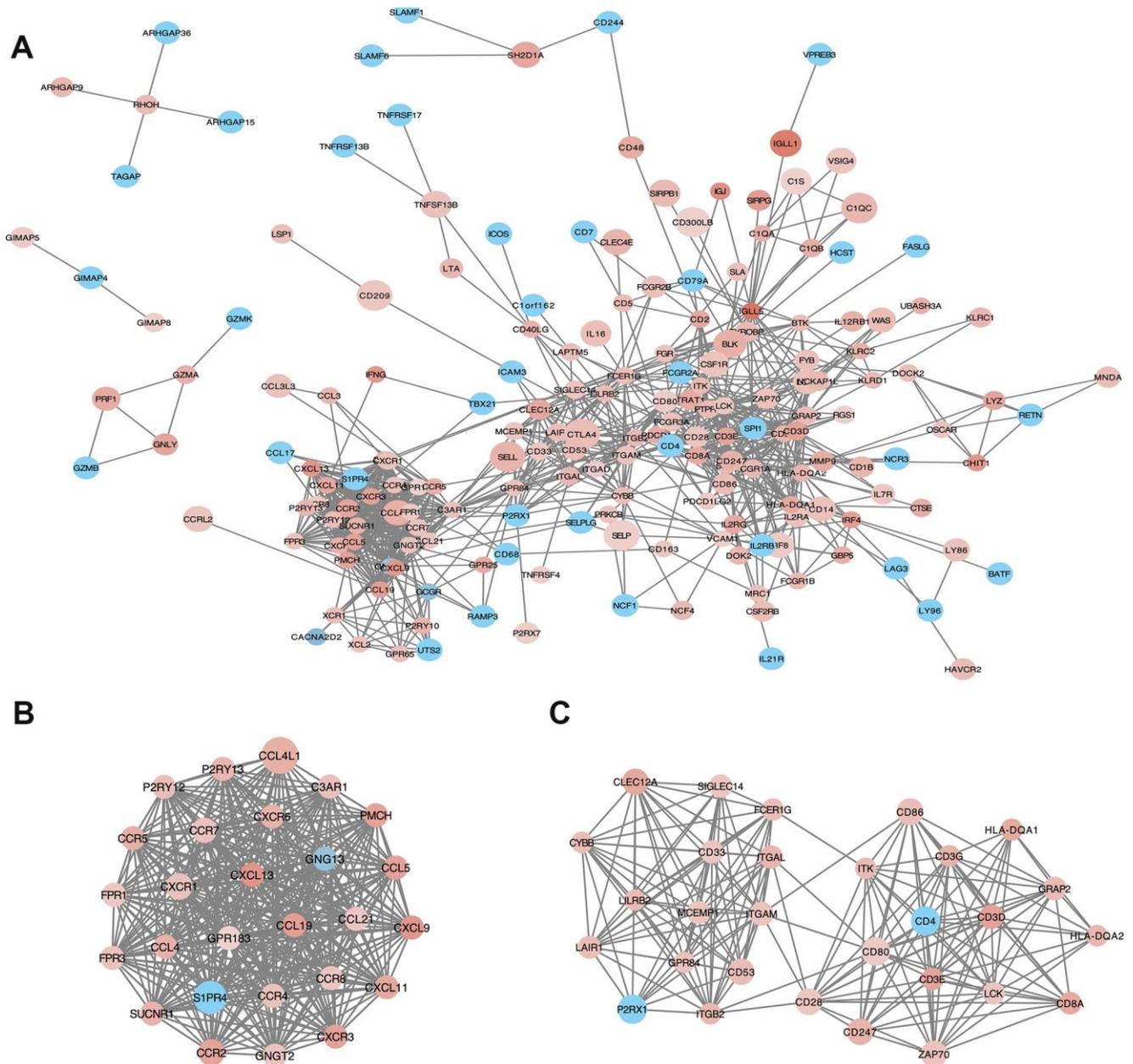


**Figure 4. Enrichment analysis of microenvironment related differentially expressed genes (DEGs).** (A) the top 10 of biological processes GO terms, cellular component GO terms, molecular function GO terms; (B) The correlation between intersection genes and top 5 biological processes GO terms; (C) KEGG (Kyoto Encyclopedia of Genes and Genomes) analysis of immune-related DEGs.

macrophages (partial.cor=0.329,  $p=8.02e-09$ ), and dendritic cells (partial.cor=0.434,  $p=7.95e-15$ ). SIGLEC1 was associated with infiltrating levels of B cells (partial.cor=0.537,  $p=5.322e-23$ ), CD4+ T cells (partial.cor=0.525,  $p=5.39e-22$ ), macrophages (partial.cor=0.364,  $p=1.43e-10$ ), neutrophils (partial.cor=0.332,  $p=5.89e-09$ ), and dendritic cells (partial.cor=0.368,  $p=8.66e-11$ ). Similarly, TMEM150B was associated with infiltrating levels of B cells (partial.cor=0.447,  $p=1.41e-15$ ), CD4+ T cells (partial.cor=0.438,  $p=4.56e-15$ ), and dendritic cells (partial.cor=0.411,  $p=2.68e-13$ ) (Figure 8).

## DISCUSSION

Tumor microenvironment (TME) plays a critical role in tumor development, progression, and responses to therapies, especially immunotherapies. However, the role of TME differs in different types of cancer. Although endometrial cancer is one of the most common gynecological cancers, the composition of the TME, and its correlation with EC prognosis remain poorly understood compared to other malignancies.



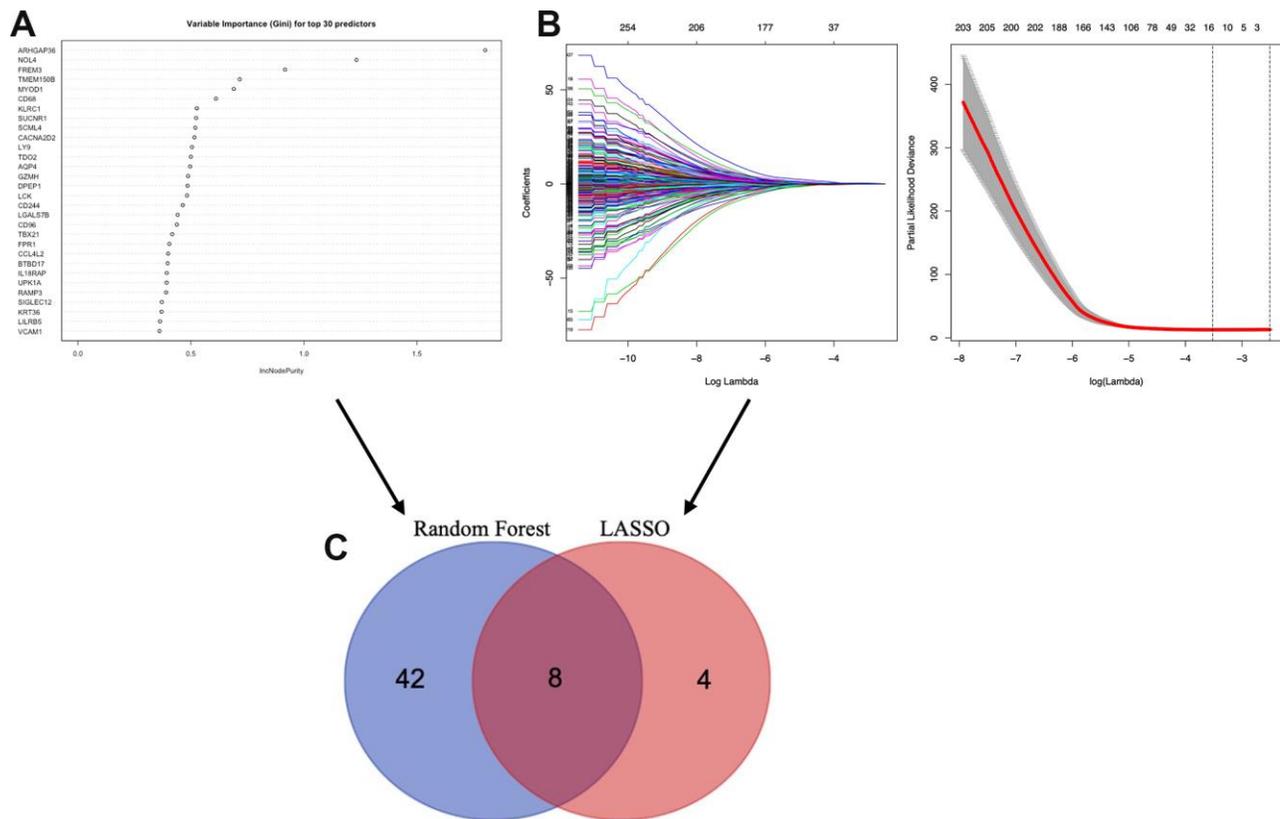
**Figure 5. Protein-protein interaction (PPI) network of microenvironment related genes.** (B) module 1 and (C) module 2 are the top two modules (>10 nodes) in the PPI network (A). The color of nodes associate with the log(FC) value, and the size reflects the combine score.

In this study, we performed a comprehensive analysis of immune cells and genes in the TME of endometrial carcinoma, and related the data to clinical outcomes and prognosis of EC patients. Using the ESTIMATE algorithm, we first analyzed the correlation between the immune/stromal scores and the clinical EC characteristics obtained from TCGA-CDR. The ESTIMATE algorithm is a widely accepted and reliable algorithm that has been used in various cancers, including glioblastoma [16], breast cancer [17], prostate cancer [18], colon cancer [19], and cutaneous melanoma [20]. Using the ESTIMATE algorithm, our data demonstrate that the immune and stromal scores positively correlate with clinical outcomes of EC patients.

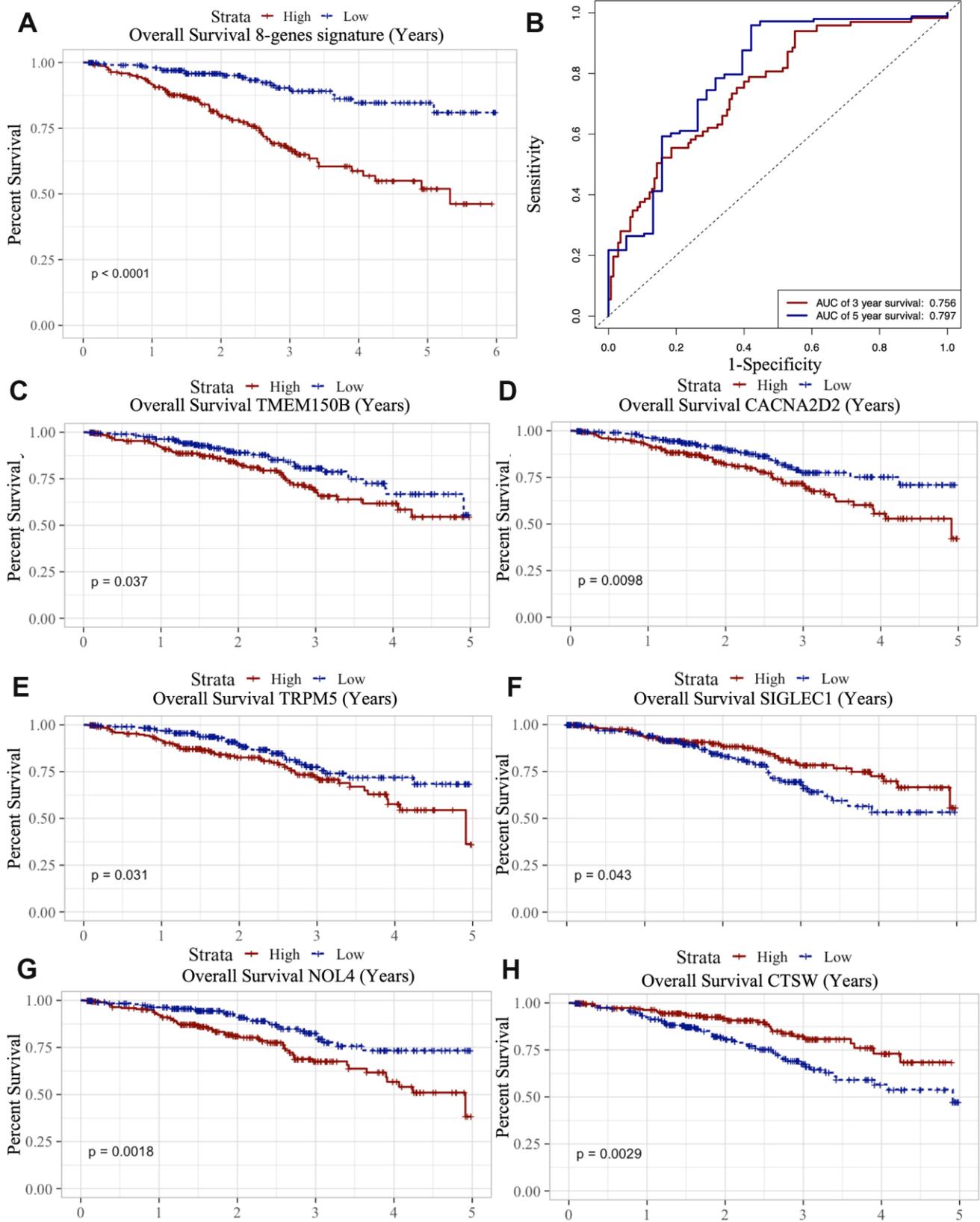
The pathogenetic types of endometrial cancer include endometrioid endometrial adenocarcinoma, serous endometrial adenocarcinoma, and mixed serous and endometrioid endometrial adenocarcinoma. Patients with serous endometrial adenocarcinoma often have a relatively poor prognosis [21–23]. In our study, we found that the immune and stromal scores of endometrioid endometrial adenocarcinoma were

significantly higher than in serous endometrial adenocarcinoma, suggesting that the high concentration of immune cells in the TME of endometrioid endometrial adenocarcinoma might represent one of the mechanisms contributing to the good prognosis of this type of EC cancer. In addition, by analyzing the correlation between the immune scores and tumor recurrence, our data show that high immune score patients have longer progression-free interval and overall survival rates, indicating that the TME composition affects the clinical outcomes of EC patients.

Next, we analyzed differentially expressed genes (DEGs) in EC, by dividing patients into high score and low score groups, based on the median immune/stromal scores. Our data show that DEGs are involved in TME, and specifically regulate T cell functions. Furthermore, analysis of the PPI network indicated enrichment clustered in T cell functions, including T cell migration, differentiation, co-stimulation, and receptor signaling. We speculate that these TME associated genes might affect the development of endometrial cancer by affecting the T cell functions.



**Figure 6. Selection of microenvironment related prognostic genes.** (A) Random forest and (B) Lasso (Least Absolute Shrinkage and Selector Operation) algorithms were performed to further select microenvironment related prognostic genes. (C) Venn diagram analysis between the genes selected by Random forest algorithm and Lasso algorithms.



**Figure 7. Survival analysis of microenvironment related prognostic genes.** (A) Kaplan-Meier (KM) survival curve of 8 microenvironment related prognostic signature. (B) ROC (receiver operating characteristic) curve of 8 microenvironment related prognostic signature. (C–G) Kaplan-Meier (KM) survival curve of microenvironment related prognostic genes.

Because of its accuracy and predictive performance, the machine learning algorithm has been used in different fields, including medical diagnostics [10].

Since complex models and highly significant relationships can be extracted from large amounts of data, the machine learning can be highly predictive for specific cancers [24, 25]. LASSO algorithm [26] and Random forest (RF) algorithm [27] can be used for classification and regression; thus, they were particularly well-suited in our study to identify the prognostic TME-related genes. Using this approach, eight immune-related genes were identified; high levels of TMEM150B, CACNA2D2, TRPM5, and NOL4 showed a negative correlation to OS, while high levels of CTSW and SIGLEC1 showed a positive correlation to OS in EC patients. These genes have been reported to be involved in carcinogenesis and development of various cancers. TMEM150B regulates autophagy and cell death by encoding an autophagy regulator [28]. CACNA2D2, the auxiliary subunit of  $\alpha_2\delta_2$ , induces cell proliferation and angiogenesis by increasing the expression of vascular endothelial growth factor to promote tumorigenesis [29]. NOL4 (nucleolar protein 4) is a novel methylation target in cervical cancer, and has been suggested as an early detection and risk prediction biomarker in cervical cancer [30]. However, most of the identified genes have not been previously linked to EC. Our data indicate that they could serve as potential prognostic biomarkers for EC.

To investigate the impact of TME infiltration with immune cells on the prognosis of EC patients, we calculated the degree of infiltration of six immune cell

types by using TIMER algorithm, and correlated the data with the expression of the identified immune genes. SIGLEC1 (CD169) is a novel biomarker of tumor-associated macrophages [24]. A previous study found that the density of CD169+ macrophages was positively associated with the abundance of CD8(+) CTL and CD57(+) NK cells in tumor tissues, and correlated with a better prognosis in EC patients [31]. This finding is consistent with the results obtained in our study. CTSW (Cathepsin W) is a novel human cysteine protease expressed in CD8+ T cells and NK cells [32], and plays an important role in cellular cytotoxicity mediated by NK cells and CD8+ T cells [33, 34]. Different T cell populations have different functions in regulating tumor grade, stage, and invasion ability in endometrial cancer [35–37]. We speculate that CTSW might be involved in the development of EC by regulating the T cell functions.

Previous studies have suggested that the tumor microenvironment in EC may have a significant prognostic value, and even play a role in resistance to treatment [38–40]. However, the tumor microenvironment is complex, and is determined by many factors. To improve the accuracy and reliability of the TME analysis, we used a large global collection of EC tissues from TCGA-UCEC, and introduced two machine learning algorithms. In addition, this study comprehensively analyzed the correlation between microenvironmental and genetic factors, and identified six potential prognostic TME-related genes (CACNA2D2, CTSW, NOL4, SIGLEC1, TMEM150B, and TRPM5). Future studies should identify the specific roles these genes play in the regulation of EC development and progression.

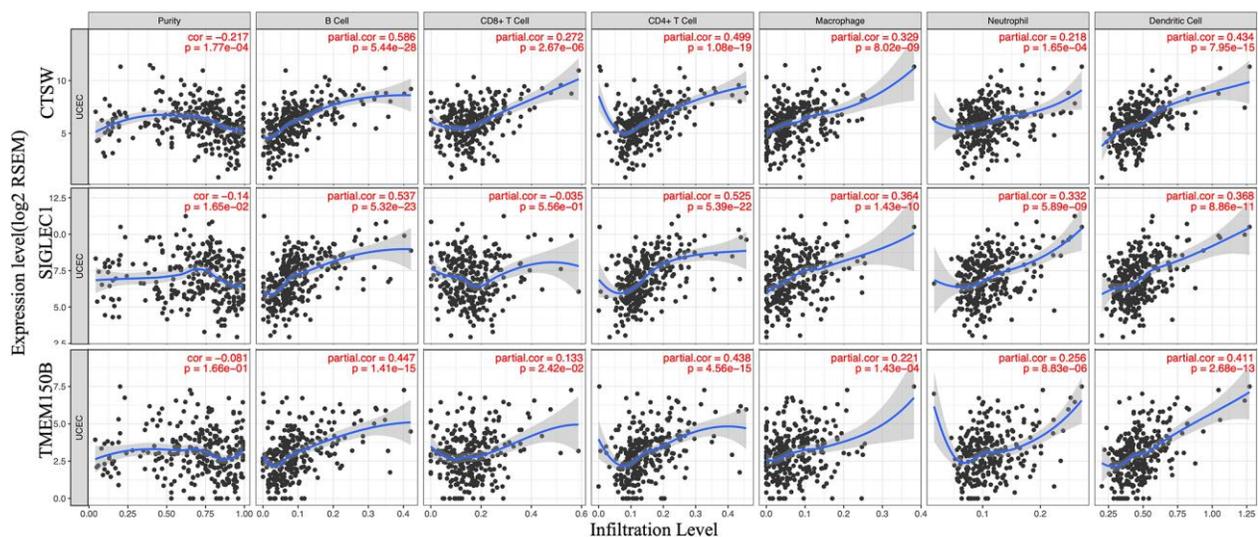


Figure 8. Correlation of microenvironment related prognostic genes' expression with immune infiltration level.

## MATERIALS AND METHODS

R software (version 3.5.1) [41] and Bioconductor [42] were used for all analyses in this study.

### Data collection and analysis

All RNA expression data were obtained from The Cancer Genome Atlas (TCGA) (Data Release 16.0 - March 26, 2019) (<https://portal.gdc.cancer.gov/>). The expression data were then normalized by Fragments Per Kilobase of transcript per Million mapped reads (FPKM). The corresponding clinical data were obtained from TCGA-CDR (TCGA Pan-cancer Clinical Data Resource) dataset [14]. Patients whose overall survival times or progression-free interval (PFI) times were less than 30 days were excluded from our study. Progression-free interval is characterized as a time without a new tumor occurrence or a death from cancer. In total, data from 521 patients were analyzed in our study, and 19459 RNAs were extracted from RNA-seq data according to ENSEMBL Genomes (hg38) (<http://ensemblgenomes.org/>). Both RNA-seq data and corresponding clinical data were publicly available.

### Calculation of immune and stromal scores

ESTIMATE (Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data) is one of the algorithms developed to evaluate the cell tumor composition by calculating the immune and stromal scores using Pearson's correlation coefficient [6]. By using "estimate" R package, the immune and stromal scores were calculated based on the gene expression data of EC patients.

### Selection of differentially expressed genes

The samples were divided into high and low immune/stromal score groups based on the medium values of the immune/stromal scores. The selection of differentially expressed genes (DEGs) was performed by using "limma" R package with p-value < 0.05 and log fold change > 1 as a filter [43]. A website Venn diagrams tool (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) was used to identify the commonly upregulated or downregulated DEGs in the immune and stromal groups. These intersection genes were selected for further analysis.

### Enrichment analysis of intersection genes

GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment analyses and visualization of intersection genes were performed by

"clusterProfiler" R package [44] and "enrichplot" R package [45] with p-value < 0.05 as the cut-off value.

### Protein-protein interaction (PPI) network construction

For understanding protein interactions, we constructed the PPI network by STRING (V11) [46] with high confidence (0.7). The information of nodes of the PPI network was then further analyzed by Cytoscape software [47]. In Cytoscape, we used Molecular COMplex DETection (MCODE) to select clusters which included 10 or more nodes [48]. ClueGo App was used to perform enrichment analysis of each cluster selected by MCODE [49].

### Identification of TME associated prognostic genes

Least Absolute Shrinkage and Selector Operation (LASSO) algorithm was used to identify candidate genes by "glmnet" R package with number of lambda = 1000 [50]. Clinical outcomes and gene expression profiles were analyzed by LASSO and Random forest algorithms. Lambda.min is the cutoff point that brings minimum mean cross-validated error. Genes with the highest lambda values were selected for further analysis. Simultaneously, Radom Forest algorithm was used for candidate genes selection by "randomForest" R package [51]. According to "randomForest" package, we set "ntree" as 10001 and "mtry" as a default value. No other options were used in machine learning algorithm.

The overlapped genes in LASSO and Random Forest algorithms were selected and then "survival" R package [52] was used to preform multivariate cox regression. The ROC (time-dependent receiver operating characteristic) curve was used to estimate the accuracy and specificity of the classification performance of the gene signature.

### Overall survival analysis

Overall survival analysis was performed by Kaplan-Meier survival analysis by using "survival" R package [52], overall survival times of EC patients (n=521), and gene expression data.

### Analysis of immune cell infiltration

The TIMER (Tumor IMmune Estimation Resource) algorithm was used to calculate the tumor abundance of six infiltrating immune cells (CD4 + T cells, CD8 + T cells, B cells, neutrophils, macrophages, and dendritic cells) based on RNA-Seq expression profiles data [12]. Compared to other calculation methods, the TIMER algorithm can eliminate bias effects by removing highly

expressed genes and eliminating collinearity between immune cells to ensure accuracy of the calculation.

The correlation between the selected prognostic genes and immune cells was calculated by Spearman's correlation analysis by TIMER. The correlation coefficient value <0.3 indicates that the correlation is negligible, while the correlation coefficient >0.3 indicates a positive/negative correlation [53].

## Abbreviations

EC: endometrial carcinoma; TME: tumor microenvironment; ECM: extracellular matrix; ESTIMATE: Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data; TIMER: Tumor Immune Estimation Resource; PFI: progression-free interval; OS: overall survival; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; KM: Kaplan-Meier; DEGs: differentially expressed genes; TCGA: The Cancer Genome Atlas; FPKM: Fragments Per Kilobase of transcript per Million mapped reads; TCGA-CDR: TCGA Pan-cancer Clinical Data Resource; MCODE: Molecular Complex Detection; LASSO: Least Absolute Shrinkage and Selector Operation; ROC: time-dependent receiver operating characteristic.

## AUTHOR CONTRIBUTIONS

Jing Wan and Peigen Chen carried out the study. Peigen Chen analyzed and interpreted the data and drafted the manuscript. Yuebo Yang and Yu Zhang collected and analyzed the data. Senwei Jiang participated in the study design and manuscript writing. Jing Wang and Xiaomao Li coordinated the study, participated in the design, and reviewed the manuscript. All authors read and approved the final manuscript.

## CONFLICTS OF INTEREST

The authors declare that there is no conflicts of interest to disclose.

## FUNDING

This study was supported by the Natural Science Foundation of Guangdong Province, China (Grant number: 2015A030310160).

## REFERENCES

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015; 136:E359–86. <https://doi.org/10.1002/ijc.29210> PMID:25220842
2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J. Cancer statistics in China, 2015. *CA Cancer J Clin*. 2016; 66:115–32. <https://doi.org/10.3322/caac.21338> PMID:26808342
3. Hanahan D, Coussens LM. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell*. 2012; 21:309–22. <https://doi.org/10.1016/j.ccr.2012.02.022> PMID:22439926
4. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000; 100:57–70. [https://doi.org/10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9) PMID:10647931
5. Galon J, Pagès F, Marincola FM, Thurin M, Trinchieri G, Fox BA, Gajewski TF, Ascierto PA. The immune score as a new possible approach for the classification of cancer. *J Transl Med*. 2012; 10:1. <https://doi.org/10.1186/1479-5876-10-1> PMID:22214470
6. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013; 4:2612. <https://doi.org/10.1038/ncomms3612> PMID:24113773
7. Şenbabaoğlu Y, Gejman RS, Winer AG, Liu M, Van Allen EM, de Velasco G, Miao D, Ostrovskaya I, Drill E, Luna A, Weinhold N, Lee W, Manley BJ, et al. Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. *Genome Biol*. 2016; 17:231. <https://doi.org/10.1186/s13059-016-1092-z> PMID:27855702
8. Curry JM, Sprandio J, Cognetti D, Luginbuhl A, Bar-ad V, Pribitkin E, Tuluc M. Tumor microenvironment in head and neck squamous cell carcinoma. *Semin Oncol*. 2014; 41:217–34. <https://doi.org/10.1053/j.seminoncol.2014.03.003> PMID:24787294
9. Cooper LA, Gutman DA, Chisolm C, Appin C, Kong J, Rong Y, Kurc T, Van Meir EG, Saltz JH, Moreno CS, Brat DJ. The tumor microenvironment strongly impacts master transcriptional regulators and gene expression class of glioblastoma. *Am J Pathol*. 2012; 180:2108–19. <https://doi.org/10.1016/j.ajpath.2012.01.040> PMID:22440258

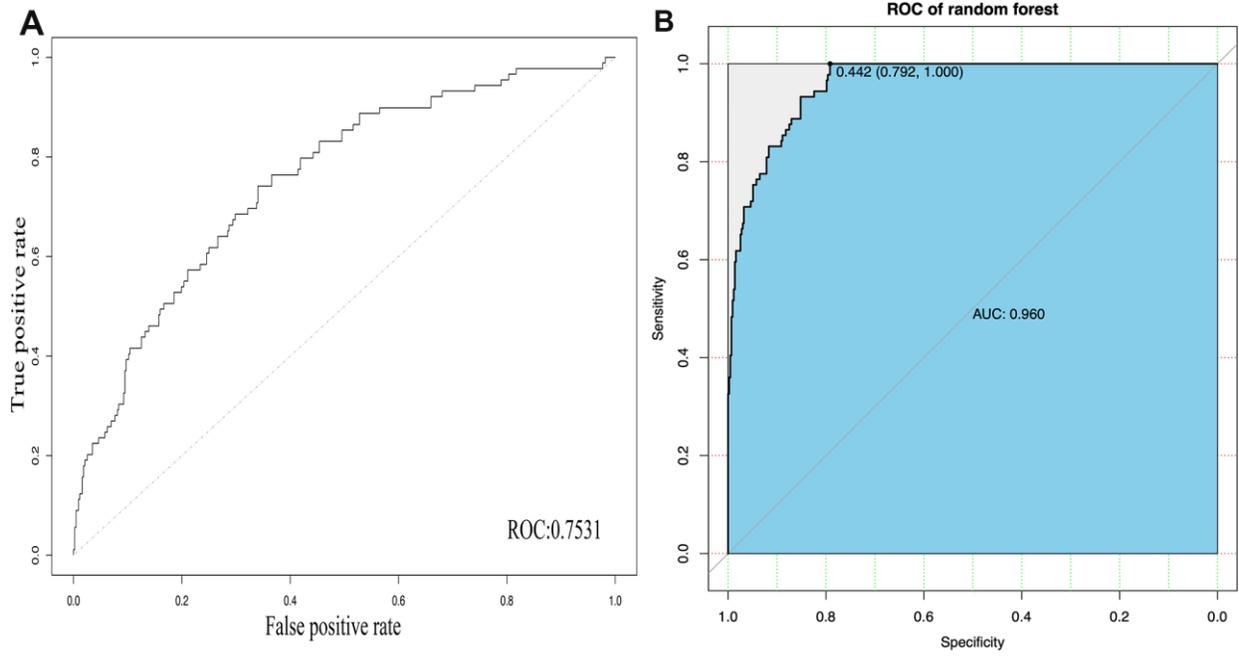
10. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med.* 2016; 375:1216–19.  
<https://doi.org/10.1056/NEJMp1606181>  
PMID:[27682033](https://pubmed.ncbi.nlm.nih.gov/27682033/)
11. Liu XS, Mardis ER. Applications of Immunogenomics to Cancer. *Cell.* 2017; 168:600–12.  
<https://doi.org/10.1016/j.cell.2017.01.014>  
PMID:[28187283](https://pubmed.ncbi.nlm.nih.gov/28187283/)
12. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, Liu XS. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* 2017; 77:e108–10.  
<https://doi.org/10.1158/0008-5472.CAN-17-0307>  
PMID:[29092952](https://pubmed.ncbi.nlm.nih.gov/29092952/)
13. Soslow RA, Tornos C, Park KJ, Malpica A, Matias-Guiu X, Oliva E, Parkash V, Carlson J, McCluggage WG, Gilks CB. Endometrial Carcinoma Diagnosis: Use of FIGO Grading and Genomic Subcategories in Clinical Practice: Recommendations of the International Society of Gynecological Pathologists. *Int J Gynecol Pathol.* 2019; 38:S64–S74.  
<https://doi.org/10.1097/PGP.0000000000000518>  
PMID:[30550484](https://pubmed.ncbi.nlm.nih.gov/30550484/)
14. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, Kovatich AJ, Benz CC, Levine DA, Lee AV, Omberg L, Wolf DM, Shriver CD, et al, Cancer Genome Atlas Research Network. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell.* 2018; 173:400–416.e11.  
<https://doi.org/10.1016/j.cell.2018.02.052>  
PMID:[29625055](https://pubmed.ncbi.nlm.nih.gov/29625055/)
15. Yang HI, Yuen MF, Chan HL, Han KH, Chen PJ, Kim DY, Ahn SH, Chen CJ, Wong VW, Seto WK, and REACH-B Working Group. Risk estimation for hepatocellular carcinoma in chronic hepatitis B (REACH-B): development and validation of a predictive score. *Lancet Oncol.* 2011; 12:568–74.  
[https://doi.org/10.1016/S1470-2045\(11\)70077-8](https://doi.org/10.1016/S1470-2045(11)70077-8)  
PMID:[21497551](https://pubmed.ncbi.nlm.nih.gov/21497551/)
16. Jia D, Li S, Li D, Xue H, Yang D, Liu Y. Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. *Aging (Albany NY).* 2018; 10:592–605.  
<https://doi.org/10.18632/aging.101415>  
PMID:[29676997](https://pubmed.ncbi.nlm.nih.gov/29676997/)
17. Vincent KM, Findlay SD, Postovit LM. Assessing breast cancer cell lines as tumour models by comparison of mRNA expression profiles. *Breast Cancer Res.* 2015; 17:114.  
<https://doi.org/10.1186/s13058-015-0613-0>  
PMID:[26289960](https://pubmed.ncbi.nlm.nih.gov/26289960/)
18. Shah N, Wang P, Wongvipat J, Karthaus WR, Abida W, Armenia J, Rockowitz S, Drier Y, Bernstein BE, Long HW, Freedman ML, Arora VK, Zheng D, Sawyers CL. Regulation of the glucocorticoid receptor via a BET-dependent enhancer drives antiandrogen resistance in prostate cancer. *eLife.* 2017; 6:6.  
<https://doi.org/10.7554/eLife.27861> PMID:[28891793](https://pubmed.ncbi.nlm.nih.gov/28891793/)
19. Alonso MH, Aussó S, Lopez-Doriga A, Cordero D, Guinó E, Solé X, Barenys M, de Oca J, Capella G, Salazar R, Sanz-Pamplona R, Moreno V. Comprehensive analysis of copy number aberrations in microsatellite stable colon cancer in view of stromal component. *Br J Cancer.* 2017; 117:421–31.  
<https://doi.org/10.1038/bjc.2017.208> PMID:[28683472](https://pubmed.ncbi.nlm.nih.gov/28683472/)
20. Yang S, Liu T, Nan H, Wang Y, Chen H, Zhang X, Zhang Y, Shen B, Qian P, Xu S, Sui J, Liang G. Comprehensive analysis of prognostic immune-related genes in the tumor microenvironment of cutaneous melanoma. *J Cell Physiol.* 2020; 235:1025–35.  
<https://doi.org/10.1002/jcp.29018> PMID:[31240705](https://pubmed.ncbi.nlm.nih.gov/31240705/)
21. Felix AS, Weissfeld JL, Stone RA, Bowser R, Chivukula M, Edwards RP, Linkov F. Factors associated with Type I and Type II endometrial cancer. *Cancer Causes Control.* 2010; 21:1851–56.  
<https://doi.org/10.1007/s10552-010-9612-8>  
PMID:[20628804](https://pubmed.ncbi.nlm.nih.gov/20628804/)
22. Bokhman JV. Two pathogenetic types of endometrial carcinoma. *Gynecol Oncol.* 1983; 15:10–17.  
[https://doi.org/10.1016/0090-8258\(83\)90111-7](https://doi.org/10.1016/0090-8258(83)90111-7)  
PMID:[6822361](https://pubmed.ncbi.nlm.nih.gov/6822361/)
23. Moore KN, Fader AN. Uterine papillary serous carcinoma. *Clin Obstet Gynecol.* 2011; 54:278–91.  
<https://doi.org/10.1097/GRF.0b013e318218c755>  
PMID:[21508697](https://pubmed.ncbi.nlm.nih.gov/21508697/)
24. Bronte V. Deciphering Macrophage and Monocyte Code to Stratify Human Breast Cancer Patients. *Cancer Cell.* 2019; 35:538–39.  
<https://doi.org/10.1016/j.ccell.2019.03.010>  
PMID:[30991022](https://pubmed.ncbi.nlm.nih.gov/30991022/)
25. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015; 521:436–44.  
<https://doi.org/10.1038/nature14539> PMID:[26017442](https://pubmed.ncbi.nlm.nih.gov/26017442/)
26. Bitton DA, Rallis C, Jeffares DC, Smith GC, Chen YY, Codlin S, Marguerat S, Bähler J. LaSSO, a strategy for genome-wide mapping of intronic lariats and branch points using RNA-seq. *Genome Res.* 2014; 24:1169–79.  
<https://doi.org/10.1101/gr.166819.113>  
PMID:[24709818](https://pubmed.ncbi.nlm.nih.gov/24709818/)
27. Degenhardt F, Seifert S, Szymczak S. Evaluation of variable selection methods for random forests and omics data sets. *Brief Bioinform.* 2019; 20:492–503.  
<https://doi.org/10.1093/bib/bbx124> PMID:[29045534](https://pubmed.ncbi.nlm.nih.gov/29045534/)

28. Mrschtik M, Ryan KM. Another DRAM involved in autophagy and cell death. *Autophagy*. 2016; 12:603–05.  
<https://doi.org/10.1080/15548627.2015.1137412>  
PMID:[27046253](https://pubmed.ncbi.nlm.nih.gov/27046253/)
29. Warnier M, Roudbaraki M, Derouiche S, Delcourt P, Bokhobza A, Prevarskaya N, Mariot P. CACNA2D2 promotes tumorigenesis by stimulating cell proliferation and angiogenesis. *Oncogene*. 2015; 34:5383–94.  
<https://doi.org/10.1038/onc.2014.467> PMID:[25619833](https://pubmed.ncbi.nlm.nih.gov/25619833/)
30. Wang SS, Smiraglia DJ, Wu YZ, Ghosh S, Rader JS, Cho KR, Bonfiglio TA, Nayar R, Plass C, Sherman ME. Identification of novel methylation markers in cervical cancer using restriction landmark genomic scanning. *Cancer Res*. 2008; 68:2489–97.  
<https://doi.org/10.1158/0008-5472.CAN-07-3194>  
PMID:[18381458](https://pubmed.ncbi.nlm.nih.gov/18381458/)
31. Ohnishi K, Yamaguchi M, Erdenebaatar C, Saito F, Tashiro H, Katabuchi H, Takeya M, Komohara Y. Prognostic significance of CD169-positive lymph node sinus macrophages in patients with endometrial carcinoma. *Cancer Sci*. 2016; 107:846–52.  
<https://doi.org/10.1111/cas.12929> PMID:[26991548](https://pubmed.ncbi.nlm.nih.gov/26991548/)
32. Linnevers C, Smeekens SP, Brömme D. Human cathepsin W, a putative cysteine protease predominantly expressed in CD8+ T-lymphocytes. *FEBS Lett*. 1997; 405:253–9.  
[https://doi.org/10.1016/s0014-5793\(97\)00118-x](https://doi.org/10.1016/s0014-5793(97)00118-x)  
PMID:[9108299](https://pubmed.ncbi.nlm.nih.gov/9108299/)
33. Oghumu S, Terrazas CA, Varikuti S, Kimble J, Vadia S, Yu L, Seveau S, Satoskar AR. CXCR3 expression defines a novel subset of innate CD8+ T cells that enhance immunity against bacterial infection and cancer upon stimulation with IL-15. *FASEB J*. 2015; 29:1019–28.  
<https://doi.org/10.1096/fj.14-264507> PMID:[25466888](https://pubmed.ncbi.nlm.nih.gov/25466888/)
34. Wex T, Wex H, Hartig R, Wilhelmsen S, Malfertheiner P. Functional involvement of cathepsin W in the cytotoxic activity of NK-92 cells. *FEBS Lett*. 2003; 552:115–19.  
[https://doi.org/10.1016/S0014-5793\(03\)00895-0](https://doi.org/10.1016/S0014-5793(03)00895-0)  
PMID:[14527671](https://pubmed.ncbi.nlm.nih.gov/14527671/)
35. Zhang W, Hou F, Zhang Y, Tian Y, Jiao J, Ma D, Kong B, Cui B. Changes of Th17/Tc17 and Th17/Treg cells in endometrial carcinoma. *Gynecol Oncol*. 2014; 132:599–605.  
<https://doi.org/10.1016/j.ygyno.2013.12.036>  
PMID:[24388919](https://pubmed.ncbi.nlm.nih.gov/24388919/)
36. Chang WC, Li CH, Huang SC, Chang DY, Chou LY, Sheu BC. Clinical significance of regulatory T cells and CD8+ effector populations in patients with human endometrial carcinoma. *Cancer*. 2010; 116:5777–88.  
<https://doi.org/10.1002/cncr.25371> PMID:[20734397](https://pubmed.ncbi.nlm.nih.gov/20734397/)
37. Yamagami W, Susumu N, Tanaka H, Hirasawa A, Banno K, Suzuki N, Tsuda H, Tsukazaki K, Aoki D. Immunofluorescence-detected infiltration of CD4+FOXP3+ regulatory T cells is relevant to the prognosis of patients with endometrial cancer. *Int J Gynecol Cancer*. 2011; 21:1628–34.  
<https://doi.org/10.1097/IGC.0b013e31822c271f>  
PMID:[21897268](https://pubmed.ncbi.nlm.nih.gov/21897268/)
38. Giannice R, Erreni M, Allavena P, Buscaglia M, Tozzi R. Chemokines mRNA expression in relation to the Macrophage Migration Inhibitory Factor (MIF) mRNA and Vascular Endothelial Growth Factor (VEGF) mRNA expression in the microenvironment of endometrial cancer tissue and normal endometrium: a pilot study. *Cytokine*. 2013; 64:509–15.  
<https://doi.org/10.1016/j.cyto.2013.07.024>  
PMID:[23985752](https://pubmed.ncbi.nlm.nih.gov/23985752/)
39. Versluis MA, Marchal S, Plat A, de Bock GH, van Hall T, de Bruyn M, Hollema H, Nijman HW. The prognostic benefit of tumour-infiltrating Natural Killer cells in endometrial cancer is dependent on concurrent overexpression of Human Leucocyte Antigen-E in the tumour microenvironment. *Eur J Cancer*. 2017; 86:285–95.  
<https://doi.org/10.1016/j.ejca.2017.09.008>  
PMID:[29059634](https://pubmed.ncbi.nlm.nih.gov/29059634/)
40. Tong H, Ke JQ, Jiang FZ, Wang XJ, Wang FY, Li YR, Lu W, Wan XP. Tumor-associated macrophage-derived CXCL8 could induce ER $\alpha$  suppression via HOXB13 in endometrial cancer. *Cancer Lett*. 2016; 376:127–36.  
<https://doi.org/10.1016/j.canlet.2016.03.036>  
PMID:[27018308](https://pubmed.ncbi.nlm.nih.gov/27018308/)
41. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2018. <https://www.r-project.org/index.html>
42. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004; 5:R80.  
<https://doi.org/10.1186/gb-2004-5-10-r80>  
PMID:[15461798](https://pubmed.ncbi.nlm.nih.gov/15461798/)
43. 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](https://pubmed.ncbi.nlm.nih.gov/25605792/)
44. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012; 16:284–87.

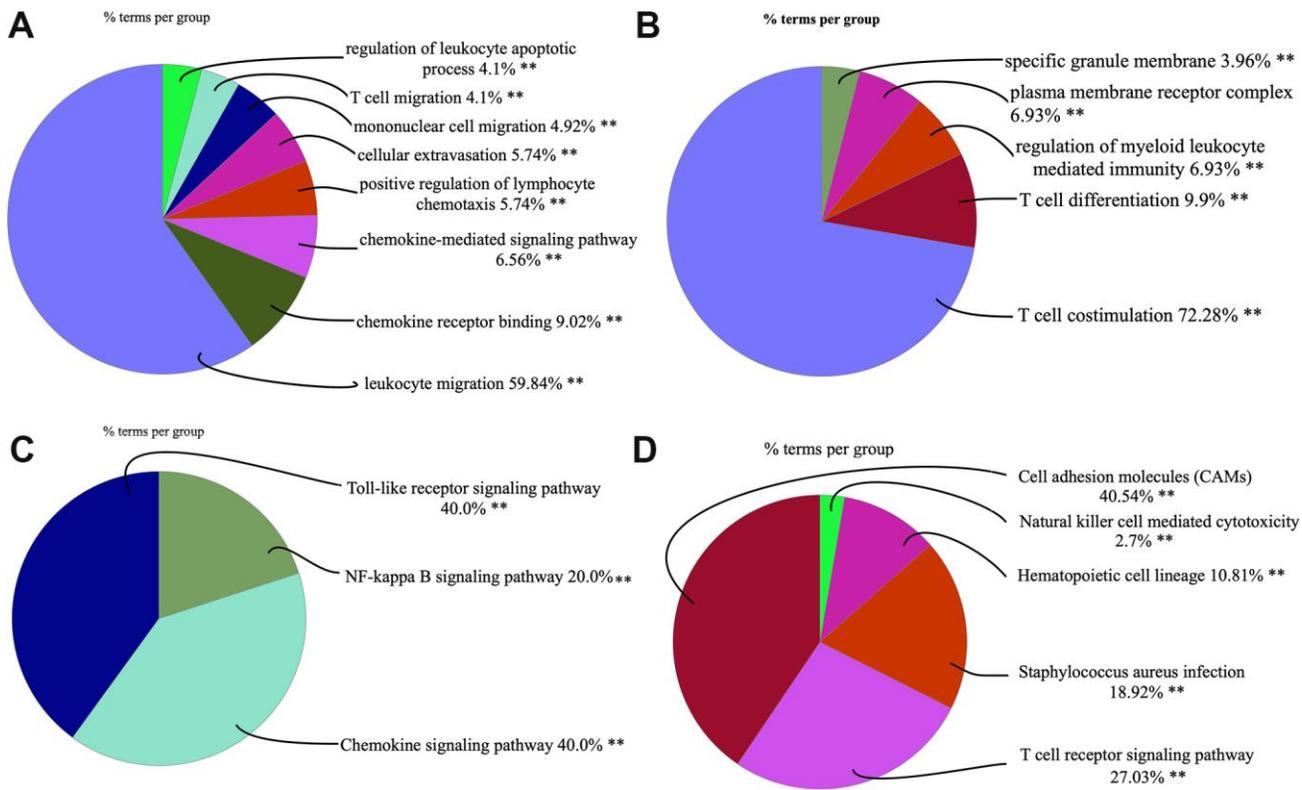
- <https://doi.org/10.1089/omi.2011.0118>  
PMID:[22455463](https://pubmed.ncbi.nlm.nih.gov/22455463/)
45. Yu G. enrichplot: Visualization of Functional Enrichment Result. R package version 1.2.0. 2018.  
<https://github.com/YuLab-SMU/enrichplot>
46. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019; 47:D607–13.  
<https://doi.org/10.1093/nar/gky1131> PMID:[30476243](https://pubmed.ncbi.nlm.nih.gov/30476243/)
47. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003; 13:2498–504.  
<https://doi.org/10.1101/gr.1239303> PMID:[14597658](https://pubmed.ncbi.nlm.nih.gov/14597658/)
48. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics.* 2003; 4:2.  
<https://doi.org/10.1186/1471-2105-4-2>  
PMID:[12525261](https://pubmed.ncbi.nlm.nih.gov/12525261/)
49. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pagès F, Trajanoski Z, Galon J. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics.* 2009; 25:1091–93.  
<https://doi.org/10.1093/bioinformatics/btp101>  
PMID:[19237447](https://pubmed.ncbi.nlm.nih.gov/19237447/)
50. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw.* 2010; 33:1–22.  
<https://doi.org/10.18637/jss.v033.i01> PMID:[20808728](https://pubmed.ncbi.nlm.nih.gov/20808728/)
51. Liaw A, Wiener M. Classification and Regression by randomForest. *R News.* 2002; 2:18–22.  
[https://www.researchgate.net/publication/228451484\\_Classification\\_and\\_Regression\\_by\\_RandomForest](https://www.researchgate.net/publication/228451484_Classification_and_Regression_by_RandomForest)
52. Therneau TM. A Package for Survival Analysis in S. version 2.38. 2015.  
<https://CRAN.R-project.org/package=survival>
53. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Med J.* 2012; 24:69–71.  
PMID:[23638278](https://pubmed.ncbi.nlm.nih.gov/23638278/)

SUPPLEMENTARY MATERIALS

Supplementary Figures



Supplementary Figure 1. ROC (receiver operating characteristic) curve of LASSO algorithm (A) and Random forest algorithm (B).



**Supplementary Figure 2. Functional enrichment of key modules of PPI network.** GO analysis (A) and KEGG (C) analysis of module 1. GO analysis (B) and KEGG (D) analysis of module 2

## **Supplementary Tables**

Please browse Full Text version to see the data of Supplementary Tables 1–5.

**Supplementary Table 1. The immune scores and stromal scores of EC.**

**Supplementary Table 2. The clinical data of EC extracted from TCGA-CDR.**

**Supplementary Table 3. Differentially expressed genes selected based on immune and stromal scores.**

**Supplementary Table 4. 711 GO terms and 36 KEGG terms of differentially expressed genes.**

**Supplementary Table 5. The infiltration of six types immune cells of EC patients.**