

## Original Article



# Identification of an immune-related risk signature and nomogram predicting the overall survival in patients with endometrial cancer

Yuan Cheng ,\* Xingchen Li ,\* Yibo Dai , Yangyang Dong , Xiao Yang , Jianliu Wang

Department of Obstetrics and Gynecology, Peking University People's Hospital, Beijing, China

## OPEN ACCESS

Received: Sep 3, 2020

Revised: Oct 31, 2020

Accepted: Dec 13, 2020

### Correspondence to

Jianliu Wang

Department of Obstetrics and Gynecology,  
Peking University People's Hospital, No. 11  
Xizhimen South Street, Xicheng Dist., Beijing,  
100044, China.

E-mail: wangjianliu@pkuph.edu.cn

\*Yuan Cheng and Xingchen Li contributed  
equally to this article.

Copyright © 2021. Asian Society of  
Gynecologic Oncology, Korean Society of  
Gynecologic Oncology, and Japan Society of  
Gynecologic Oncology

This is an Open Access article distributed  
under the terms of the Creative Commons  
Attribution Non-Commercial License ([https://  
creativecommons.org/licenses/by-nc/4.0/](https://creativecommons.org/licenses/by-nc/4.0/))  
which permits unrestricted non-commercial  
use, distribution, and reproduction in any  
medium, provided the original work is properly  
cited.

### ORCID iDs

Yuan Cheng

<https://orcid.org/0000-0002-4476-7241>

Xingchen Li

<https://orcid.org/0000-0003-3373-1670>

Yibo Dai

<https://orcid.org/0000-0002-0599-955X>

Yangyang Dong

<https://orcid.org/0000-0002-1142-9131>

Xiao Yang

<https://orcid.org/0000-0002-5150-7341>

Jianliu Wang

<https://orcid.org/0000-0001-8932-317X>

## ABSTRACT

**Objective:** Aimed to construct an immune-related risk signature and nomogram predicting endometrial cancer (EC) prognosis.

**Methods:** An immune-related risk signature in EC was constructed using the least absolute shrinkage and selection operator regression analysis based on The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases. A nomogram integrating the immune-related genes and the clinicopathological characteristics was established and validated using the Kaplan-Meier survival curve and receiver operating characteristic (ROC) curve to predict the overall survival (OS) of EC patients. The Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data (ESTIMATE) R tool was used to explore the immune and stromal scores.

**Results:** CCL17, CTLA4, GPI, HDGF, HFE2, ICOS, IFNG, IL21R, KAL1, NR3C1, S100A2, and S100A9 were used in developing an immune-related risk signature evaluation model. The Kaplan-Meier curve indicated that patients in the low-risk group had better OS ( $p < 0.001$ ). The area under the ROC curve (AUC) values of this model were 0.737, 0.764, and 0.782 for the 3-, 5-, and 7-year OS, respectively. A nomogram integrating the immune-related risk model and clinical features could accurately predict the OS (AUC=0.772, 0.786, and 0.817 at 3-, 5-, and 7-year OS, respectively). The 4 immune cell scores were lower in the high-risk group. Forkhead box P3 (FOXP3) and basic leucine zipper ATF-like transcription factor (BATF) showed a potential significant role in the immune-related risk signature.

**Conclusion:** Twelve immune-related genes signature and nomogram for assessing the OS of patients with EC had a good practical value.

**Keywords:** Endometrial Cancer; Immune Signature; Nomogram; Overall Survival; FOXP3; BATF

## INTRODUCTION

Endometrial cancer (EC) is the most common malignancy of the female genital tract and accounts for 7% of new cancer cases in women in developed countries [1]. The incidence of EC is expected to increase to 42.13 per 100,000 people in the United States by 2030 [2]. Currently, the International Federation of Gynecology and Obstetrics (FIGO) staging system, together with histological grade and type classification, is primarily used for classifying

### Funding

The National Natural Science Foundation of China (Grant no. 81802607, 81874108 and 81672571), Beijing Municipal Natural Science Foundation (Grant No.7202213) and National Key Technology R&D Program of China (Grant No. 2019YFC1005200 and 2019YFC1005201).

### Conflict of Interest

No potential conflict of interest relevant to this article was reported.

### Author Contributions

Conceptualization: C.Y., W.J.; Data curation: C.Y., L.X., D.Y.<sup>1</sup>; Formal analysis: L.X.; Funding acquisition: C.Y., W.J.; Methodology: L.X., D.Y.<sup>2</sup>; Project administration: W.J.; Resources: L.X., D.Y.<sup>2</sup>, Y.X.; Software: L.X., D.Y.<sup>2</sup>, Y.X.; Validation: C.Y., L.X., D.Y.<sup>2</sup>, Y.X.; Writing - original draft: C.Y., L.X.; Writing - review & editing: C.Y., L.X., W.J.

D.Y.<sup>1</sup>, Yibo Dai; D.Y.<sup>2</sup>, Yangyang Dong.

EC patients with different prognoses [3]. However, various factors, including genomic and clinical factors [4], play important roles in promoting EC development and determining the prognosis; the present classification systems cannot comprehensively and accurately predict the overall survival (OS) of EC patients. Hence, there is an urgent need to understand the underlying molecular mechanisms and identify the novel targets for treatment of and the appropriate interventions for patients with EC.

The progression of cancer is associated with the tumor microenvironment (TME) [5]. The TME is composed of immune cells, extracellular matrix, mesenchymal cells, and inflammatory mediators, and it has a significant influence on tumor growth, metastasis, and clinical outcomes [6]. The relationship between immune dysfunction and EC has long been reported and studied. Blocking the interaction between CD47 and SIRPα inhibited the proliferation of EC tumors and promoted the infiltration of macrophages with antitumor ability in TME [7]. High microsatellite instability (MSI-H) EC has higher immune cell infiltration than microsatellite-stable EC and showed excellent response to immunotherapy [8]. A previous study has demonstrated that some immune-related genes were associated with the prognosis of EC, such as TMEM150B, CACNA2D2, TRPM5, NOL4, CTSW, and SIGLEC1 [9]. However, a nomogram prediction model combining clinical factors and immune-related genes in EC has not been established and verified.

Currently, high-throughput data provide a comprehensive insight into the genomic and genetic changes in patients with EC [10]. The high-throughput profiles help identify possible biomarkers for predicting the survival of EC patients and their reaction to therapy. In this study, large-scale gene datasets were downloaded from the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases. The Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data (ESTIMATE) 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 [11]. The CIBERSORT algorithm is an analytical tool from the Alizadeh Lab developed by Newman et al. to provide an estimation of the abundance of member cell types in a mixed cell population, using gene expression data. Based on the immune cell-related genes, the leukocyte subsets were analyzed using the CIBERSORT algorithms [12].

After performing an integrated analysis of both TCGA and GEO, we identified 12 immune-related genes from the common differentially expressed genes (DEGs) and constructed a risk signature. Then, a nomogram that comprehensively combined the clinicopathological features and immune-related risk signature was established. The ESTIMATE and CIBERSORT algorithms were used to conduct an integrative analysis of immune scores and immune cells in different risk groups of the risk signature in EC patients. We further explored the interaction between the 12 immune-related genes and their transcription factors (TFs) to elucidate the molecular mechanism underlying the tumorigenesis of immune-induced EC.

## MATERIALS AND METHODS

### 1. Data collection and identification of differentially expressed immune-related genes

The expression data were downloaded from the TCGA and GEO databases. The mRNA sequencing datasets of 412 EC tissues and 35 adjacent normal tissues including their

corresponding clinicopathological information were downloaded from TCGA (<https://tcga-data.nci.nih.gov>) database. The gene expression profile, including 91 EC samples and 12 normal samples (GSE17025) based on the GPL570 (Affymetrix Human Genome U133 Plus 2.0 Array) platform, was collected from the GEO database (<http://www.ncbi.nlm.nih.gov/geo>). TCGA and GSE17025 are the 2 databases most often involved in the establishment of EC prediction models for large tissue sample size. Information regarding the molecular classification of TCGA was downloaded from the datasets termed Uterine Corpus Endometrial Carcinoma (TCGA, PanCancer Atlas) from cBioportal database [13].

We downloaded immune-related gene set (GO\_REGULATION\_OF\_IMMUNE\_SYSTEM\_PROCESS, Homo sapiens) from MSigDB (Molecular Signature Database) in GSEA (Gene Set Enrichment Analysis) [14]. The gene names and expression were extracted using the “Practical Extraction and Report Language.” The immune-related DEGs in EC and normal tissue samples were analyzed using the significance analysis of microarrays method with “*limma*” package in R software, following the criteria  $\log_2|FC| > 1$  and  $p < 0.05$ . The heatmap was plotted for samples and DEGs were identified with the “*heatmap*” package in R software. A website Venn diagrams tool (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) was used to identify the commonly upregulated or downregulated DEGs in the TCGA and GEO groups. These intersection genes were selected for further analysis. Open resource data were used, and no ethical issues were involved.

## 2. Enrichment analysis of the intersection genes

In the current study, the “*clusterProfiler*” package was used to conduct the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of the overlapping DEGs [15]. The GO analysis included cellular composition, molecular function, and biological process. Disease ontology (DO) annotates genes based on human diseases. DO was vital for the annotation of a human gene, translating the obtained key genes to clinical relevance. *DOSE*, an R package, is capable of analyzing semantic similarity computations of the DO terms and genes. Therefore, *DOSE* is used to determine the closeness between diseases and gene functions [16]. To investigate the underlying mechanism of these DEGs, 111 DEGs were subjected to GO, KEGG, and DO analyses using *clusterProfiler* and *DOSE* packages, and  $p$ -values  $< 0.05$  were set as the threshold values.

## 3. Construction and verification of immune-related gene risk models

The Least Absolute Shrinkage and Selector Operator (LASSO) regression analysis method was used to construct a model for predicting the prognosis of patients with high performance utilizing the “*glmnet*” R package [17]. The LASSO regression model was used to identify the most accurate predictive genes. For example, if there were 2 different prognostic related genes in parallel, LASSO would automatically filter out the secondary related one and assign the selected gene a value, which equals the regression coefficient in the classifier formulas [18].

Then, 12 immune-related hub genes predicting the prognosis of EC were analyzed in TCGA data using LASSO. The hub genes were selected for the construction of the signature, and the coefficient for each gene was obtained using a penalized regression method. The risk score based on the signature was calculated using the following formula:

$$RS = \sum_{i=1}^n Coef(i)X(i)$$

where  $n$  represents the number of modules RNAs,  $\text{Coef}(i)$  is the coefficient, and  $X(i)$  denotes the z-score-transformed relative expression level of each gene identified using the LASSO method. To compare the expression of the hub genes encoding mRNA between the EC tissues and normal tissues, the package “*pheatmap*” of R software was used.

The prognostic risk value of each patient was calculated using the aforementioned formula, and the patients were classified into high- and low-risk groups according to the median value. Pearson correlation analysis was further performed between 12 hub genes and clinicopathological features in TCGA database. Then, time-dependent ROC curve (tdROC) and Kaplan-Meier survival curve analyses were conducted to validate the signature. GSEA was also conducted to study the functions associated with different subgroups of EC. The hallmark gene set “h.all.v6.0.symbol.gmt” and “KEGG” were applied in the GSEA.

#### 4. Development and validation of a nomogram based on the risk signature

To identify the independent risk factors for OS, univariate and multivariate analyses of the clinical factors and risk score signature were performed. We determined the risk factors that were significant in the univariate analysis, and these significant factors were then included in the multivariate analysis to assess the independent risk factors. Furthermore, an OS-associated nomogram with independent risk factors from the multivariate analysis was designed using the “*rms*” and the “*Hmisc*” packages by R software. Calibration curves were drawn, and the concordance index (C-index) was obtained to assess the efficiency of the nomogram. The accuracy of the nomogram in the diagnosis of EC was analyzed by performing tdROC and survival curve analyses. The Kaplan-Meier curves were also used to distinguish the specificity of the nomogram in different subgroups in terms of age, lymph node metastasis (LNM), and recurrence.

#### 5. Calculation of immune and stromal scores

ESTIMATE is one of the algorithms developed to evaluate the cell tumor composition by calculating the immune and stromal scores using Pearson's correlation coefficient [19]. Using the “*ESTIMATE*” R package, the immune and stromal scores were calculated based on the gene expression data of EC patients. The patients were then divided into high score and low score groups according to the median score. The Kaplan-Meier survival curve was used to estimate the relationship between the 2 groups.

#### 6. Estimation of the tumor-infiltrating immune cells

RNA-sequencing data from TCGA were used to estimate the proportions of 6 types of infiltrating immune cells, including B cells, CD4 T cells, CD8 T cells, dendritic cells, macrophages, and neutrophils, using the CIBERSORT algorithm following the procedure as previously reported [20].

#### 7. Prediction of TFs for DEGs

TFs of hub genes were explored using human TF information (NetworkAnalyst, <http://www.networkanalyst.ca>), recorded using ChIP Enrichment Analysis, and visualized using the Cytoscape software [21,22].

#### 8. Statistical analysis

Continuous variables were expressed as mean±standard deviation or median; categorical variables were expressed as frequency (n) and proportion (%). The differences among variables were tested using the Student's *t*-tests, nonparametric tests, chi-square tests, or

one-way analysis of variance tests. The log-rank test was used to determine the difference in OS rate between the high-risk group and low-risk group. Statistical analyses were performed using GraphPad Prism 8 and R software, version 3.5.1. For all analyses, p-values less than 0.05 were considered significant.

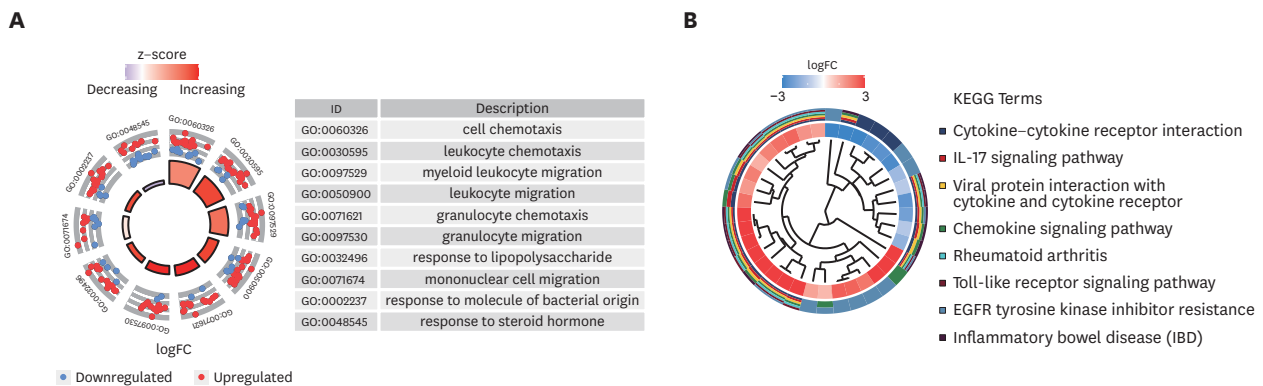
## RESULTS

### 1. Identification of overlapping immune-related DEGs

The flow diagram of the present study is shown in **Supplementary Fig. 1**. The expression of the immune-related genes is shown as a heatmap in the TCGA and GSE17025 (**Supplementary Fig. 2**). Of these 476 immune-related DEGs in the TCGA dataset, 278 were upregulated and 198 were downregulated genes (**Supplementary Fig. 3A**). In the GEO dataset, 173 immune-related DEGs were found to be dysregulated, including 108 upregulated and 65 downregulated genes (**Supplementary Fig. 3B**). The volcano plot shows the immune-related DEGs in the 2 datasets (**Supplementary Fig. 3C and D**). The Venn diagram showed that 111 genes overlapped in the 2 datasets (**Supplementary Fig. 3E**). These genes were further analyzed in the following study.

### 2. Functional analysis and immune-related risk signature construction by overlapped DEGs

To elucidate the potential function of the 111 genes, GO and KEGG analyses were carried out. As shown in **Fig. 1A**, some of GO plot significant terms were as follows: “cell chemotaxis,” “leukocyte chemotaxis,” “myeloid leukocyte migration,” and “leukocyte migration”. KEGG analysis revealed that these 111 DEGs were significantly enriched in the following pathways: “cytokine–cytokine receptor interaction,” “IL-17 signaling pathway,” “viral protein interaction with cytokine and cytokine receptor,” “chemokine signaling pathway,” and “Toll-like receptor signaling pathway” (**Fig. 1B**). These results indicated that these DEGs were distributed in immune-related processes and pathways.

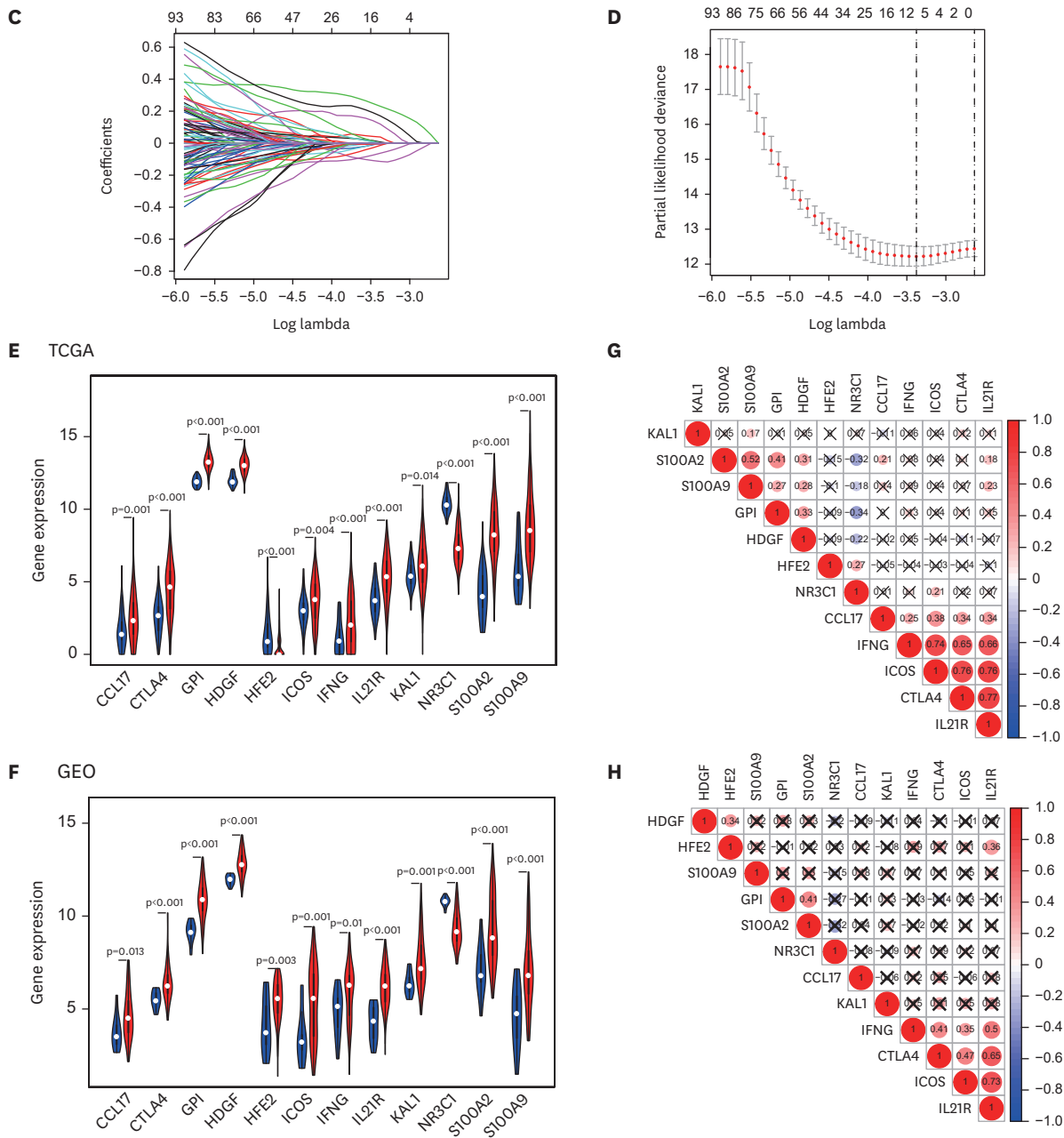


**Fig. 1.** Functional analysis and prognostic genes constructed by the overlapped immune-related DEGs. (A) Gene Ontology analysis of 111 immune-related DEGs, and only top 10 terms were shown. (B) KEGG analysis of 111 immune-related DEGs. The inner circle is composed of different genes and their expression (LogFC), while the outer circle consists of different KEGG terms. (C, D) Selection of target genes to construct the Least absolute shrinkage and selection operator risk regression model. (E) The expression of selected genes in different tissue samples in the TCGA and GEO databases. (F) Spearman correlation analysis of the selected genes in the TCGA and GEO databases. (G) The expression of selected genes in different tissue samples in the GEO database. (H) Spearman correlation analysis of the selected genes in the GEO database.

GEO, Gene Expression Omnibus; DEGs, differentially expressed genes; TCGA, The Cancer Genome Atlas.

(continued to the next page)

**Immune signature nomogram for endometrial cancer**



**Fig. 1.** (Continued) Functional analysis and prognostic genes constructed by the overlapped immune-related DEGs. (A) Gene Ontology analysis of 111 immune-related DEGs, and only top 10 terms were shown. (B) KEGG analysis of 111 immune-related DEGs. The inner circle is composed of different genes and their expression (LogFC), while the outer circle consists of different KEGG terms. (C, D) Selection of target genes to construct the Least absolute shrinkage and selection operator risk regression model. (E) The expression of selected genes in different tissue samples in the TCGA and GEO databases. (F) Spearman correlation analysis of the selected genes in the TCGA and GEO databases. (G) The expression of selected genes in different tissue samples in the GEO database. (H) Spearman correlation analysis of the selected genes in the GEO database. GEO, Gene Expression Omnibus; DEGs, differentially expressed genes; TCGA, The Cancer Genome Atlas.

To develop an immune-related risk stratification to predict the prognosis of EC in TCGA database, a LASSO regression model was conducted based on 111 genes. Twelve genes were found with their regression coefficients: CCL17, CTLA4, GPI, HDGF, HFE2, ICOS, IFNG, IL21R, KAL1, NR3C1, S100A2, and S100A9 (**Fig. 1C and D, Table 1**). The risk score formulation for this risk signature was as follows: risk signature =  $-0.06267 \times \text{CCL17} - 0.11412 \times$

**Table 1.** Twelve immune-associated genes and corresponding coefficient value

Immune-related gene name	Coefficient
CCL17	-0.06267
CTLA4	-0.11412
GPI	0.119167
HDGF	0.038758
HFE2	0.230627
ICOS	-0.00657
IFNG	-0.01385
IL21R	-0.00177
KAL1	0.007523
NR3C1	0.170435
S100A2	0.002577
S100A9	0.001639
Risk score	Low: <15.85 High: ≥15.85

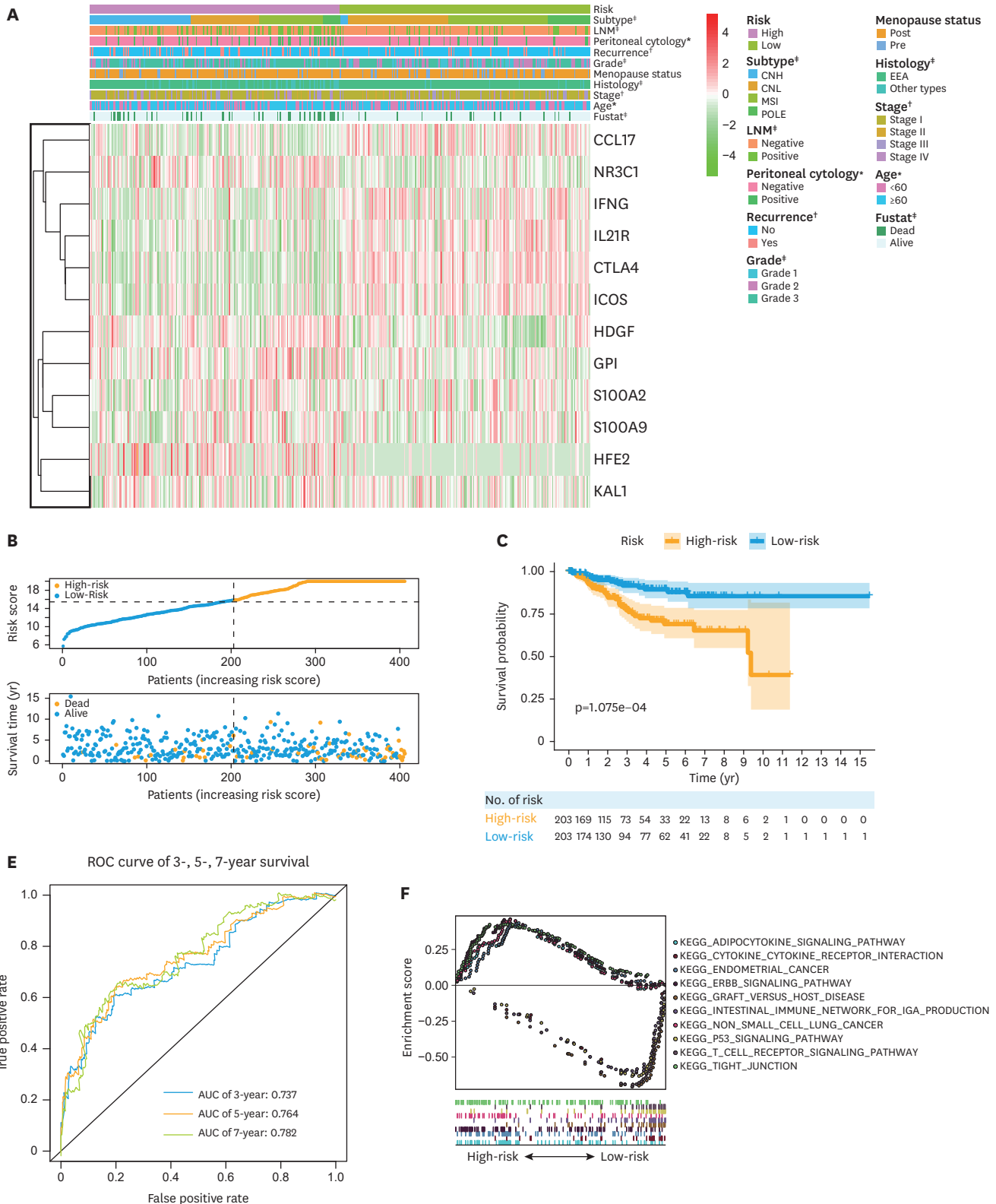
$CTLA4 + 0.119167 \times GPI + 0.038758 \times HDGF + 0.230627 \times HFE2 - 0.00657 \times ICOS - 0.01385 \times IFNG - 0.00177 \times IL21R + 0.007523 \times KAL1 + 0.170435 \times NR3C1 + 0.002577 \times S100A2 + 0.001639 \times S100A9$ .  
 The expression of these 12 genes was verified, and the results of validation suggested that the expression levels of CCL17, CTLA4, GPI, HDGF, ICOS, IFNG, IL21R, KAL1, S100A2, and S100A9 were significantly elevated in EC tissues, while those of HFE2 and NR3C1 were downregulated in EC samples in both datasets (**Fig. 1E and F**). The expression of the 12 genes was significantly correlated with each other, especially between IFNG and ICOS, ICOS and CTLA4, and ICOS and IL21R, CTLA4, and IL21R in the TCGA dataset. GEO showed similar results (**Fig. 1G and H**).

### 3. Development and verification of immune-related risk signatures

The risk score of each EC patient in TCGA database was calculated according to the abovementioned formula, and the patients were then categorized into low-risk group and high-risk group based on the median cutoff value. The number of deaths was significantly higher in the high-risk group (**Fig. 2B**). The relationship between risk score and each clinicopathological characteristic was also explored in the present study. The expression levels of the 12 genes in the high-risk group and low-risk group were presented in the heatmap (**Fig. 2A**). The results showed that there were significant differences between the low-risk group and high-risk group in terms of molecular subtype ( $p < 0.001$ ), LNM ( $p < 0.001$ ), peritoneal cytology ( $p < 0.05$ ), recurrence ( $p < 0.01$ ), grade ( $p < 0.001$ ), histology ( $p < 0.001$ ), stage ( $p < 0.01$ ), age ( $p < 0.05$ ), and living status ( $p < 0.001$ ).

The Kaplan-Meier survival analysis revealed that patients in the low-risk group had a significantly longer survival time than those in the high-risk group ( $p < 0.001$ , **Fig. 2C**). We also analyzed the Kaplan-Meier survival between the high-risk group and low-risk group in clinicopathological differences (**Supplementary Fig. 5**). The results showed that the stratification was significantly different in EC patients with all age groups, grade 3, EEA, stage I, 3 molecular subtypes, LNM and peritoneal negative group. The p-values in grade 1–2, stage III and IV were 0.0876, 0.0992 and 0.0992. The tdROC curve showed that the risk signature can accurately predict the 3-year (AUC=0.737), 5-year (AUC=0.764), and 7-year (AUC=0.782) survival (**Fig. 2D**). GSEA was conducted to explore the hallmarks and KEGG pathways of the 2 groups. As presented in **Fig. 2E**, the high-risk group was enriched in the “adipocytokine\_signaling\_pathway,” “cytokine\_cytokine\_receptor\_interaction,” “endometrial\_cancer,” “ERBB\_signaling\_pathway,” and “intestinal\_immune\_network\_for\_IGA\_production.”

Immune signature nomogram for endometrial cancer



**Fig. 2.** Differential clinicopathological factors and OS of EC in the immune-related risk signature. (A) Heatmap of clinicopathological features in high-risk and low-risk groups (\* $p < 0.05$ , † $p < 0.01$ , and ‡ $p < 0.001$ ). (B) Risk score distribution in the high-risk and low-risk groups. (C) Survival curves of the high-risk and low-risk groups. (D) Time-dependent ROC curve for prediction of 3-, 5-, and 7-year OS. (E) Gene Set Enrichment Analysis results based on the risk groups in 302 EC patients. AUC, area under the curve; EC, endometrial cancer; KEGG, Kyoto Encyclopedia of Genes and Genomes; OS, overall survival; ROC, receiver operating characteristic.



#### 4. Establishment and evaluation of a prognostic nomogram in patients with EC

To better predict the clinical outcomes of EC patients using the immune-related risk signature, univariate and multivariate Cox regression analyses were used to identify the independent risk factors for developing a nomogram in TCGA database. As shown in **Fig. 3A**, results of the multivariate analysis suggested that age (hazard ratio [HR]=1.046, 95% confidence interval [CI]=1.019–1.073,  $p<0.001$ ), stage (HR=1.495, 95% CI=1.012–2.211,  $p=0.043$ ), and risk signature (HR=2.042, 95% CI=1.090–3.827,  $p=0.026$ ) were the independent risk factors for the OS of EC. A nomogram was designed to predict the OS probability of each patient (**Fig. 3B**). In the nomogram, each predictor was assigned with a risk score. Based on the multivariate analysis results, 3 factors were integrated in the nomogram to predict the OS probability of EC patients. The calibration curve suggested that the nomogram-predicted survival rate was similar to the actual 3-, 5-, and 7-year survival rates (**Fig. 3C**).

In addition, we evenly categorized the patients into 2 subgroups according to their total points and further tested the survival assessment model using the Kaplan-Meier analysis in both the whole cohort and subgroups. Results showed that this nomogram could accurately differentiate patients across all groups (**Fig. 3D**). The AUC values of the nomogram were 0.772, 0.786, and 0.817 at 3-, 5-, and 7-year OS (**Fig. 3E**). These findings indicated that this risk signature-based nomogram had a certain reliability and specificity for evaluating prognosis.

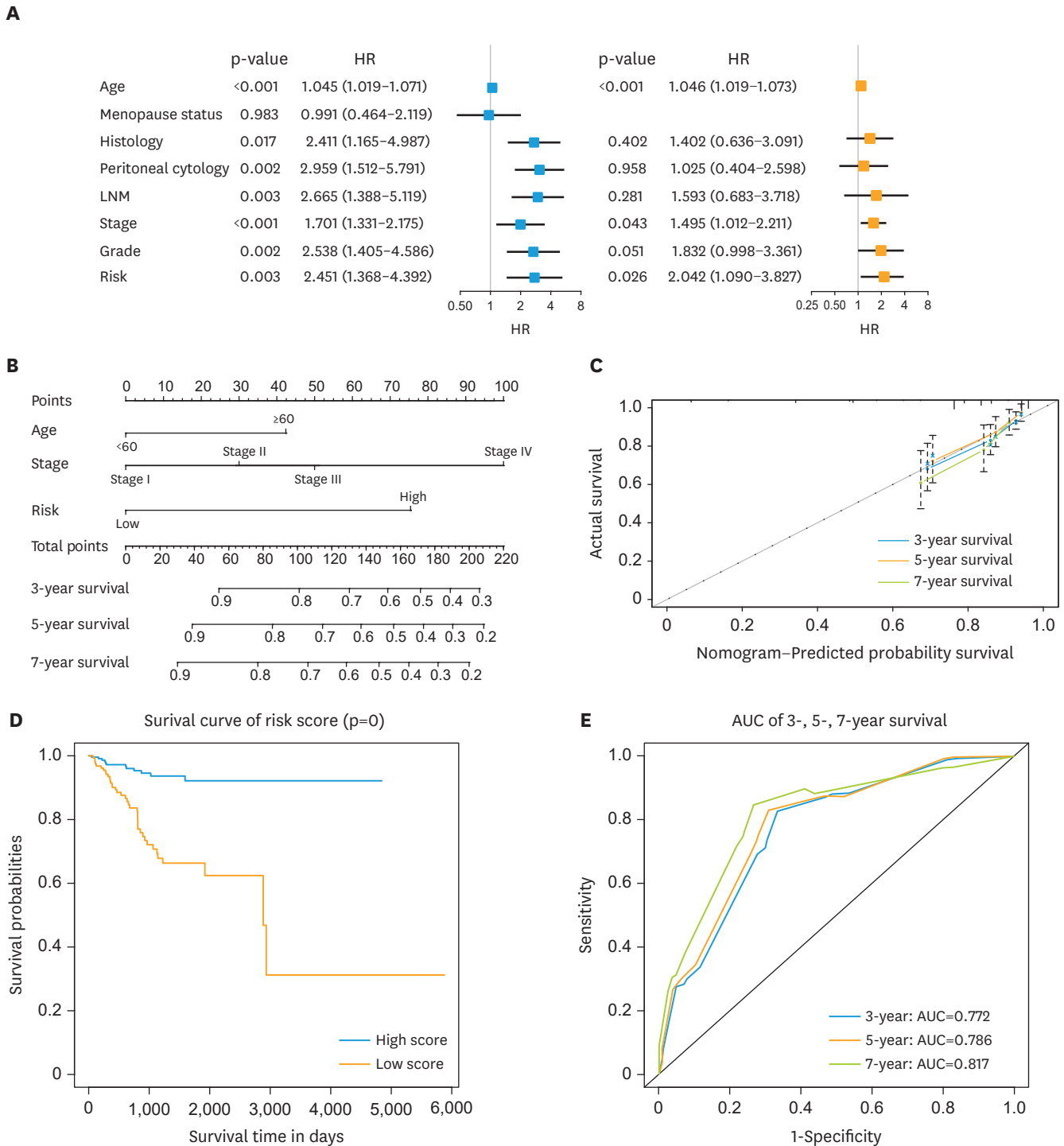
#### 5. Correlation of immune and stromal scores with different risk signature groups (cell score)

A cohort containing 412 EC patients and 20,309 available RNAs were extracted from the TCGA dataset according to the ENSEMBL Genomes and analyzed in this study. Based on the gene expression and using the ESTIMATE algorithm, the immune score, stromal score, and total score (ESTIMATE score) were calculated. The immune score, stromal score, and ESTIMATE score in the high-risk group were significantly lower than those in the low-risk group using the Wilcoxon signed-rank test (**Fig. 4A–C**). To determine the potential correlation of OS with immune score, stromal score, and ESTIMATE score, all patients were divided into top and bottom halves (high score vs. low score groups) based on their scores. The Kaplan-Meier survival curves (**Fig. 4D–F**) showed that EC patients with high immune scores had a longer OS ( $p=0.003$ , log-rank test). Consistently, patients with higher stromal scores or ESTIMATE scores showed longer median OS compared with patients with lower scores, although it was not statistically significant ( $p>0.05$ , log-rank test). We further utilized the CIBERSORT algorithm to determine the estimated fractions of 6 immune cells in each sample. The Wilcoxon rank-sum test was utilized to accurately compare the difference and found that these infiltrating immune cells, including B cells ( $p<0.001$ ), CD4+ T cells ( $p=0.002$ ), CD8+ T cells ( $p<0.001$ ), and dendritic cells ( $p<0.001$ ), had significantly lower infiltrating density in the high-risk group (**Supplementary Fig. 4**). These outcomes suggested that the prognosis of EC patients in the high immune-related risk group was correlated with lower immune scores and immune infiltrates.

#### 6. Connections between TFs and hub genes

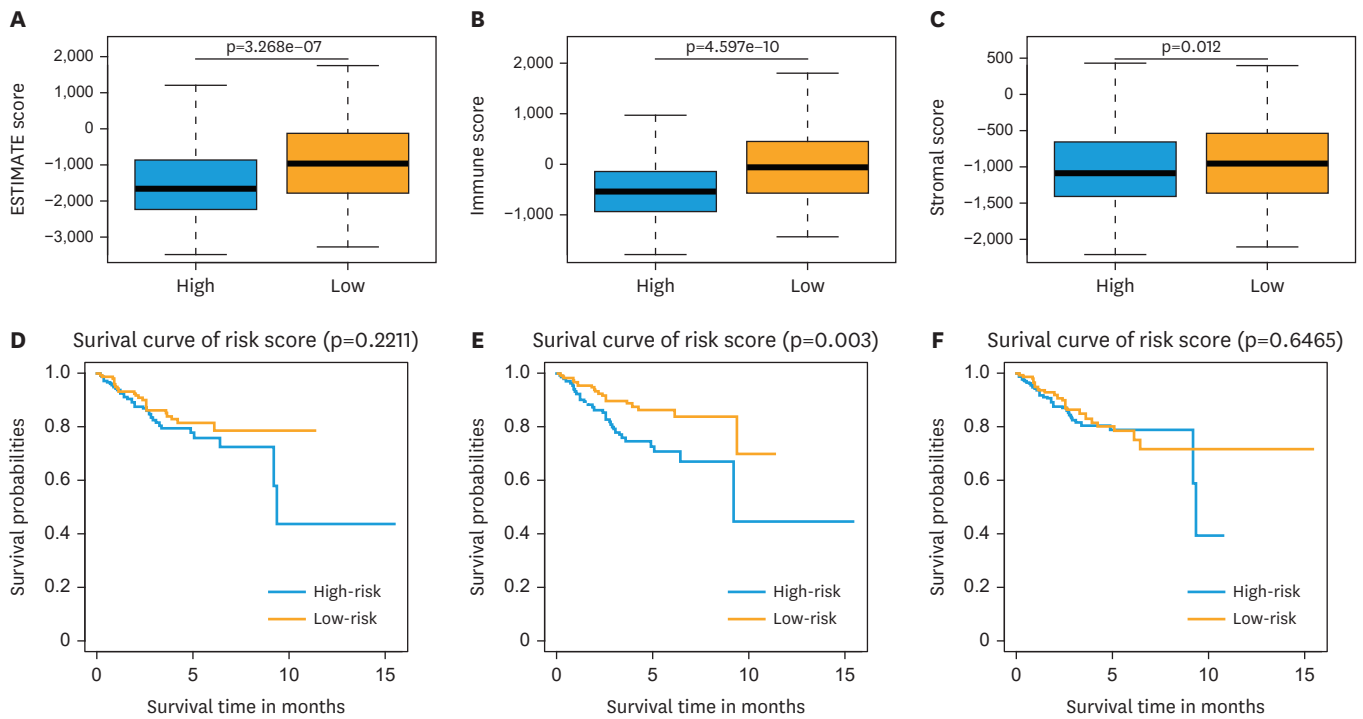
To further understand the regulatory network between TFs and these 12 hub genes, TFs with adjusted  $p<0.05$  were screened in the TCGA database. Results revealed that 104 TFs were differentially expressed in the EC samples and normal samples, including 51 upregulated TFs and 53 downregulated TFs (**Fig. 5A and C**). GPI, HDGF, HEE2, KAL1, NR3C1, S100A2, and S100A9 were high-risk immune-related genes, while CCL17, CTLA4, ICOS, IFNG, and IL21R were low-risk immune-related genes (**Fig. 5B**). The association between the hub genes and TFs through NetworkAnalyst was constructed using the Cytoscape software. As shown

**Immune signature nomogram for endometrial cancer**



**Fig. 3.** Construction and internal validation of predictive nomogram of OS for EC patients. (A) Univariate and multivariate Cox regression analyses of the association between clinicopathological factors (including the immune-related risk signature) and OS of EC patients in the TCGA datasets. (B) Prediction of OS in EC based on the nomogram. Three factors were included in this nomogram. (C) Calibration curve of the 3-, 5-, and 7-year survival. (D) Kaplan-Meier survival analysis of the 2 groups divided based on the total score of the nomogram. (E) Time-dependent ROC analysis and AUC values of the 3-, 5-, and 7-year OS in the nomogram. AUC, area under the curve; CI, confidence intervals; EC, endometrial cancer; HR, hazard ratios; LNM, lymph node metastasis; OS, overall survival; ROC, receiver operating characteristic; TCGA, The Cancer Genome Atlas.

in **Fig. 5D**, a transcription-regulated network with 51 edges and 104 nodes was obtained for hub genes. A total of 24 transcript factors were involved in the regulation of the 11 hub genes.



**Fig. 4.** ESTIMATE scores, stromal scores, and immune scores of the low-risk and high-risk groups. (A) ESTIMATE scores, (B) immune scores, (C) stroma scores in the low-risk and high-risk groups in the Kaplan-Meier analysis of OS for patients with (D) low vs. high ESTIMATE scores, (E) low vs. high immune scores, (F) low vs. high stromal scores.

ESTIMATE, the Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data; OS, overall survival.

However, no TF regulated the HFE2 gene. Forkhead box P3 (FOXP3) and basic leucine zipper ATF-like transcription factor (BATF) have been predicted to regulate IL21R, CTLA4, IFNG, ICOS, and CCL17. ETS1 was predicted to regulate IL21R, IFNG, ICOS, and NR3C1. LYL1 was predicted to regulate IL21R, CTLA4, ICOS, and CCL17. FLI1 was predicted to regulate NR3C1, CTLA4, ICOS, and HDGF. STAT1 was predicted to regulate HDGF, IFNG, ICOS, and NR3C1. Above all, we constructed a TF-regulation network of hub genes that might play an essential role in the prognosis of EC patients.

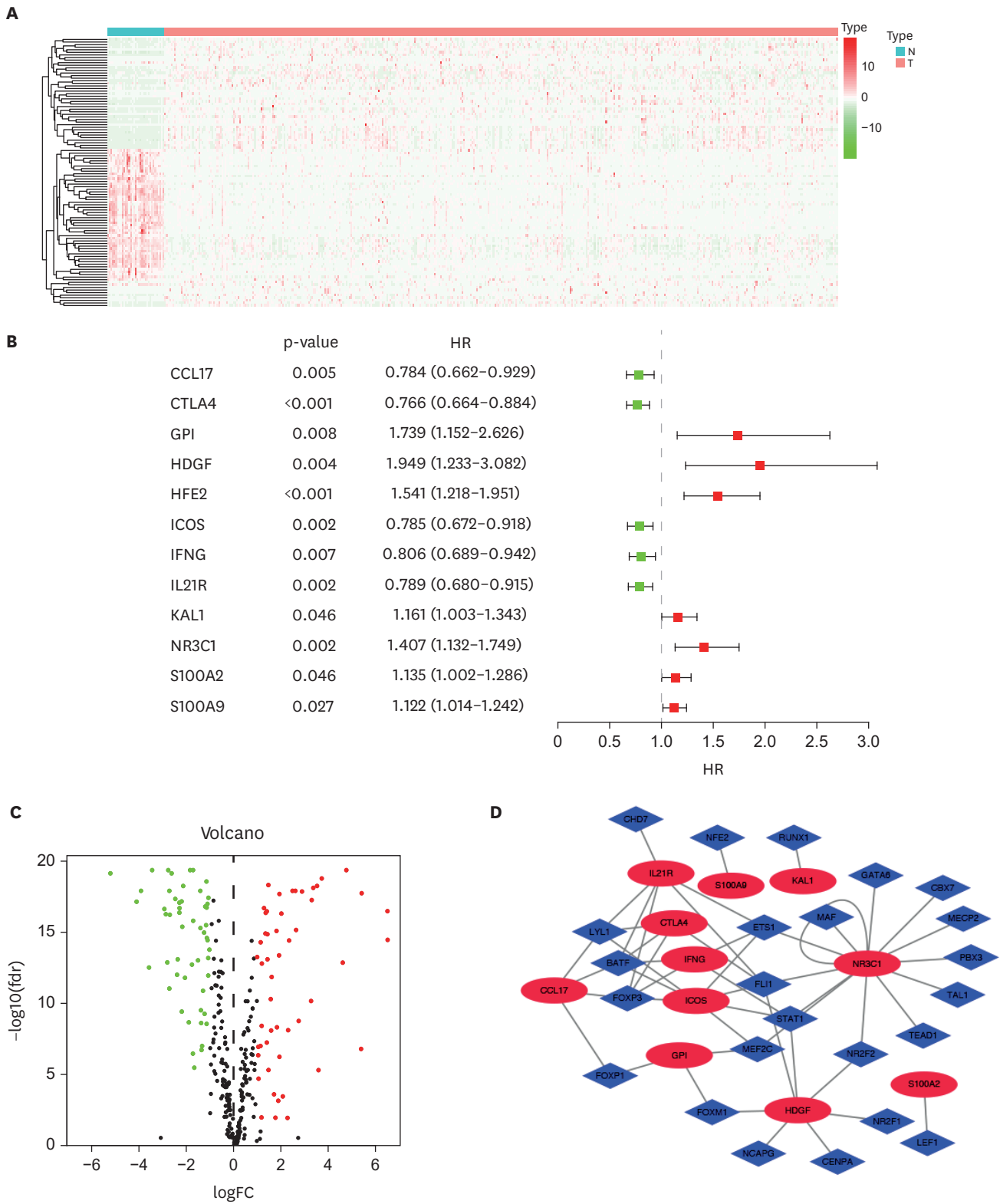
### 7. External validation of prognostic nomogram in the PKUPH cohorts

The clinicopathological characteristics of EC patients in the Peking University People's Hospital (PKUPH) cohorts were obtained from a published article (<https://pubmed.ncbi.nlm.nih.gov/31423227/>). Based on the above nomogram scores, the high-risk and low-risk groups were well divided in the PKUPH cohort (**Supplementary Fig. 6A**). The heatmap of clinicopathological features in the sequencing data of PKUPH samples showed that high-risk scores were also associated with LNM ( $p < 0.05$ ), myometrial infiltration (MI) ( $p < 0.05$ ), FIGO stage ( $p < 0.01$ ) and living status ( $p < 0.05$ ) (**Supplementary Fig. 6A**). The immune-related high-risk score and nomogram score showed worse OS (**Supplementary Fig. 6B and C**).

## DISCUSSION

EC is the fourth most common cancer in developed countries, which seriously affects women's health. EC accounts for about 76,000 deaths in women worldwide [23]. Although the prognosis of early-stage EC is good, the individual difference is extremely large.

Immune signature nomogram for endometrial cancer



**Fig. 5.** Network of transcription factors and 12 hub genes. (A) Heatmap of 104 transcription factors related to 12 hub genes in TCGA. (B) Risk factor analysis of the 12 hub genes. (C) A volcano map of 104 transcription factors related to 12 hub genes in TCGA. (D) Network of 24 transcription factors and 12 hub genes. TCGA, The Cancer Genome Atlas.

Approximately 14% of EC patients experienced recurrence and metastasis, and 78% showed initial recurrence 1–2 years after surgery [24]. In recent years, many nomogram models have been developed for prognostic prediction based on the clinicopathological factors. The nomogram according to FIGO stage, histological grade 3, primary tumor diameter  $\geq 2$  cm, and positive peritoneal cytology predicting the 3- and 5-year recurrence-free survival has been established to predict the prognosis of EC [25]. The screening of target biomolecules with predictive value based on bioinformatics analysis has been increasing in recent years. Wang et al. found that EC patients with CTSW, PCSK4, LRRC8D, TNFRSF18, IHH, and CDKN2A genes showed poor prognosis [26]. However, these studies focused only on the clinical or molecular factors, and cannot comprehensively reveal the important prognostic factors in the occurrence and prognosis of EC. In this study, we combined the genetic and clinical data to establish nomogram models for predicting the prognosis of EC patients individually.

Data from TCGA and GEO databases that sequenced the whole genome of EC patients were used to establish predictive models. In recent years, increasing attention has been paid to the gene characteristics related to the TME. Immunotherapy targeting the immune microenvironment has also been used in clinical practice. A previous study has demonstrated that some immune-related genes were related to the prognosis of EC, such as TMEM150B, CACNA2D2, TRPM5, NOL4, CTSW, and SIGLEC1 based on the TCGA database [9]. Immune-related genes (LTA, PSMC4, KAL1, TNF, SBDS, HDGF, LTB, HTR3E, NR2F1, NR3C1, PGR, and CBLC) in EC were associated with prognosis based on the TCGA database [27]. However, these studies only use one TCGA database. Hence, our study integrated the data from the TCGA and GEO databases, and the queue is larger than those used in their studies.

In this study, CCL17, CTLA4, GPI, HDGF, HFE2, ICOS, IFNG, IL21R, KAL1, NR3C1, S100A2, and S100A9 were selected to develop a risk evaluation model. The high-risk group based on 12 immune-related gene was significant correlation with molecular subtype. The data showed that the number of patients with high risk in POLE group and MSI-H were less than that in CN-low, and CN-high groups; the patients with POLE group and MSI-H was more in group of low-risk group. This further verified the correlation between immune microenvironment characteristics and molecular typing in EC. We also analyzed the Kaplan-Meier survival between the high-risk group and low-risk group in clinicopathological differences. The results suggested that it was widely applicable in EC patients with different pathological features. Then, a nomogram integrating these 12 immune-related gene risk models and clinical features (age, stage) was constructed and showed good predictive value.

GPI, HDGF, HEE2, KAL1, NR3C1, S100A2, and S100A9 were high-risk immune-related genes, while CCL17, CTLA4, ICOS, IFNG, and IL21R were low-risk immune-related genes in EC. Among them, NR3C1 and HDGF were also found in individuals with EC in a previous study [27]. The nuclear receptor subfamily 3 group C member 1 (NR3C1) gene is a glucocorticoid receptor gene associated with the development of an autoimmune blistering disorder, which is expressed in response to stress and impairment due to long-term exposure to continuous stress and social isolation as well as implicit social interaction related to religious beliefs [28,29]. A previous study found that NR3C1 was related to metastatic prostate cancer and antiandrogen resistance [30]. NR3C1 is involved in the relapse of acute lymphoblastic leukemia [31]. Accumulating evidence suggests that a hepatoma-derived growth factor (HDGF) promotes the occurrence and progression of various cancers. HDGF mRNA was modified by METTL3-mediated m<sup>6</sup>A; it promotes gastric cancer progression and has prognostic significance [32]. HDGF was positively correlated with ZEB1 expression

and showed a positive feedback loop, promoting the pathogenesis and progression of EC [33]. Overall, NR3C1 and HDGF are key regulatory immune-related genes that regulate the prognosis of EC; however, the molecular mechanism needs to be explored further.

FOXP3 and BATF might have a significant role in the immune-related prognosis of EC. Regulatory T cells (Treg cells) enhance peripheral immune tolerance by inhibiting immune responses to autoantigens and environmentally sound antigens. Foxp3, a forkhead TF, is essential for the differentiation and function of Treg cells [34]. Foxp3-expressing Treg cells can suppress the immune system and are essential in maintaining immune homeostasis and self-tolerance [35]. High expression of FoxP3+ lymphocytes was associated with poor OS in patients with pancreatic cancer [36]. BATF plays an important regulatory role in the expression of tumor-reactive CD8+ T cells [37]. BATF is involved in the progression of gastric cancer and non-small cell lung cancer (NSCLC) [38-40]. BATF knockdown inhibited the proliferation of A549 cells and promoted apoptosis [39]. BATF also promoted the immunosuppression involved in IRF4+ Tregs regulation in NSCLC [40]. These results provide important information about the mechanism underlying the pathogenesis of NSCLC. The 2 TFs of FOXP3 and BATF promote the occurrence and progression of cancer, which is consistent with our results. However, the molecular mechanisms of the 2 TFs in EC should be explored further.

Compared with the traditional study of EC biomarkers, our study analyzed several clinical samples in the public database and focused on the expression of immune-related genes in immune cells. With the study of the characteristics of cellular immune infiltration, the relationship between the regulation of immune-related genes and the incidence and prognosis of EC was investigated in depth. The use of CIBERSORT also greatly improved the efficiency of research and prevented the laborious experimental process. We can fully examine the immune inflammation patterns of EC mentioned in previous studies, to assess the specific distribution of immune cells and the changes in the corresponding biomarkers, and to discover new biological relationships among them. It provides a new therapeutic target for clinical treatment. This study has high feasibility, provides a practical method for future EC immune research, and can serve as a reference when conducting other cancer research. However, the markers obtained in this study still should be combined with those in clinical studies to determine their functions and provide a basis for future clinical research.

Taken together, our study is the first to identify and validate the 12 immune-related genes according to the nomogram scoring system; our results can further prove a connection between the nomogram score and the prognosis of EC patients, which could indicate the immune infiltration intensity in the EC microenvironment. In the external validation, a high-risk score was also associated with LNM, MI, and FIGO stage. The immune-related high-risk score group had worse OS. These selected signatures can also provide novel immune targets for EC immune-related treatment in the future.

In conclusion, we constructed a prognostic classifier aggregating 12 immune-related genes and an integrative nomogram that could accurately and effectively predict the clinical outcomes and guide the individualized anticancer treatment in patients with EC. We also explored the function of immune cell infiltration in TME for different risk groups. Although the accuracy of the risk signature still needs to be confirmed in a large-scale cohort and through more experimental studies, this is the first study to examine the relationship between aberrant immune-related gene nomogram and the prognosis of EC patients. This study will help improve our understanding of the underlying mechanisms involved in the development of EC.

## SUPPLEMENTARY MATERIALS

### Supplementary Fig. 1

Workflow diagram of the study.

[Click here to view](#)

### Supplementary Fig. 2

Expression of the immune-related genes shown by heatmap in (A) TCGA and (B) GSE17025.

[Click here to view](#)

### Supplementary Fig. 3

Distribution characteristics of immune-related DEGs in TCGA and GEO databases. (A) Number of upregulated and downregulated genes in the TCGA database. (B) Number of upregulated and downregulated genes in the GEO database. (C) Volcano plot for DEGs in tumor and normal tissues in the TCGA database. (D) Volcano plot for DEGs in tumor and normal tissues in the GEO database. (E) Venn diagram of DEGs in the 2 databases.

[Click here to view](#)

### Supplementary Fig. 4

Differential distributions of immune cells in 2 risk groups. Wilcoxon rank-sum test accurately compared the difference and indicated that 6 immune cells conferred significantly lower infiltrating density in the high-risk groups. (A) B cell. (B) CD4+ cell. (C) CD8+ cell. (D) Dendritic cell. (E) Macrophage. (F) Neutrophil.

[Click here to view](#)

### Supplementary Fig. 5

Kaplan-Meier survival between the high-risk group and low-risk group in clinicopathological differences. (A) Age <60. (B) Age ≥60. (C) Grade 1-2. (D) Grade 3. (E) EEA. (F) None-EEA. (G) Peritoneal negative. (H) Peritoneal positive. (I) Stage I. (J) Stage II. (K) Stage III. (L) Stage IV. (M) POLE. (N) MSI. (O) CNL. (P) CNH.

[Click here to view](#)

### Supplementary Fig. 6

External validation of the prediction model in the PKUPH sequencing data. (A) Based on the nomogram scores, the high-risk and low-risk groups were well divided in the PKUPH cohort. The heatmap of clinicopathological features in the sequencing data of PKUPH samples was constructed. (B) Based on 12 immune-related genes high-risk and low-risk scores, the OS was analyzed using the sequencing data of PKUPH samples. (C) Based on nomogram high-risk and low-risk scores, the OS was analyzed using the sequencing data of PKUPH samples.

[Click here to view](#)

## 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;68:394-424.  
[PUBMED](#) | [CROSSREF](#)
2. Sheikh MA, Althouse AD, Freese KE, Soisson S, Edwards RP, Welburn S, et al. USA endometrial cancer projections to 2030: should we be concerned? *Future Oncol* 2014;10:2561-8.  
[PUBMED](#) | [CROSSREF](#)
3. Amant F, Mirza MR, Koskas M, Creutzberg CL. Cancer of the corpus uteri. *Int J Gynaecol Obstet* 2018;143 Suppl 2:37-50.  
[PUBMED](#) | [CROSSREF](#)
4. Talhouk A, McConechy MK, Leung S, Yang W, Lum A, Senz J, et al. Confirmation of ProMisE: a simple, genomics-based clinical classifier for endometrial cancer. *Cancer* 2017;123:802-13.  
[PUBMED](#) | [CROSSREF](#)
5. Giraldo NA, Sanchez-Salas R, Peske JD, Vano Y, Becht E, Petitprez F, et al. The clinical role of the TME in solid cancer. *Br J Cancer* 2019;120:45-53.  
[PUBMED](#) | [CROSSREF](#)
6. Hinshaw DC, Shevde LA. The tumor microenvironment innately modulates cancer progression. *Cancer Res* 2019;79:4557-66.  
[PUBMED](#)
7. Gu S, Ni T, Wang J, Liu Y, Fan Q, Wang Y, et al. CD47 blockade inhibits tumor progression through promoting phagocytosis of tumor cells by M2 polarized macrophages in endometrial cancer. *J Immunol Res* 2018;2018:6156757.  
[PUBMED](#) | [CROSSREF](#)
8. Pakish JB, Zhang Q, Chen Z, Liang H, Chisholm GB, Yuan Y, et al. Immune microenvironment in microsatellite-unstable endometrial cancers: hereditary or sporadic origin matters. *Clin Cancer Res* 2017;23:4473-81.  
[PUBMED](#) | [CROSSREF](#)
9. Chen P, Yang Y, Zhang Y, Jiang S, Li X, Wan J. Identification of prognostic immune-related genes in the tumor microenvironment of endometrial cancer. *Aging (Albany NY)* 2020;12:3371-87.  
[PUBMED](#) | [CROSSREF](#)
10. Chiu CY, Miller SA. Clinical metagenomics. *Nat Rev Genet* 2019;20:341-55.  
[PUBMED](#) | [CROSSREF](#)
11. Jiao F, Sun H, Yang Q, Sun H, Wang Z, Liu M, et al. Association of CXCL13 and immune cell infiltration signature in clear cell renal cell carcinoma. *Int J Med Sci* 2020;17:1610-24.  
[PUBMED](#) | [CROSSREF](#)
12. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol Biol* 2018;1711:243-59.  
[PUBMED](#) | [CROSSREF](#)
13. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012;2:401-4.  
[PUBMED](#) | [CROSSREF](#)
14. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545-50.  
[PUBMED](#) | [CROSSREF](#)
15. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;16:284-7.  
[PUBMED](#) | [CROSSREF](#)
16. Yu G, Wang LG, Yan GR, He QY. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* 2015;31:608-9.  
[PUBMED](#) | [CROSSREF](#)
17. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;33:1-22.  
[PUBMED](#) | [CROSSREF](#)
18. Kim SM, Kim Y, Jeong K, Jeong H, Kim J. Logistic LASSO regression for the diagnosis of breast cancer using clinical demographic data and the BI-RADS lexicon for ultrasonography. *Ultrasonography* 2018;37:36-42.  
[PUBMED](#) | [CROSSREF](#)



19. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-García W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 2013;4:2612.  
[PUBMED](#) | [CROSSREF](#)
20. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol Biol* 2018;1711:243-59.  
[PUBMED](#) | [CROSSREF](#)
21. Lachmann A, Xu H, Krishnan J, Berger SI, Mazloom AR, Ma'ayan A. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* 2010;26:2438-44.  
[PUBMED](#) | [CROSSREF](#)
22. Xia J, Gill EE, Hancock RE. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nat Protoc* 2015;10:823-44.  
[PUBMED](#) | [CROSSREF](#)
23. Urlick ME, Bell DW. Clinical actionability of molecular targets in endometrial cancer. *Nat Rev Cancer* 2019;19:510-21.  
[PUBMED](#) | [CROSSREF](#)
24. Legge F, Restaino S, Leone L, Carone V, Ronsini C, Di Fiore GL, et al. Clinical outcome of recurrent endometrial cancer: analysis of post-relapse survival by pattern of recurrence and secondary treatment. *Int J Gynecol Cancer* 2020;30:193-200.  
[PUBMED](#) | [CROSSREF](#)
25. Cheng Y, Dong Y, Tian W, Zhang H, Li X, Wang Z, et al. Nomogram for predicting recurrence-free survival in chinese women with endometrial cancer after initial therapy: external validation. *J Oncol* 2020;2020:2363545.  
[PUBMED](#) | [CROSSREF](#)
26. Wang Y, Ren F, Chen P, Liu S, Song Z, Ma X. Identification of a six-gene signature with prognostic value for patients with endometrial carcinoma. *Cancer Med* 2018;7:5632-42.  
[PUBMED](#) | [CROSSREF](#)
27. Ding H, Fan GL, Yi YX, Zhang W, Xiong XX, Mahgoub OK. Prognostic implications of immune-related genes' (IRGs) signature models in cervical cancer and endometrial cancer. *Front Genet* 2020;11:725.  
[PUBMED](#) | [CROSSREF](#)
28. Holmes L, Chinaka C, Elmi H, Deepika K, Pelaez L, Enwere M, et al. Implication of spiritual network support system in epigenomic modulation and health trajectory. *Int J Environ Res Public Health* 2019;16:4123.  
[PUBMED](#) | [CROSSREF](#)
29. Mahmoudi H, Ebrahimi E, Daneshpazhooh M, Balighi K, Mirzazadeh A, Elikaei Behjati S, et al. Single-nucleotide polymorphisms associated with pemphigus vulgaris: potent markers for better treatment and personalized medicine. *Int J Immunogenet* 2020;47:41-9.  
[PUBMED](#) | [CROSSREF](#)
30. Zhang Z, Zhou C, Li X, Barnes SD, Deng S, Hoover E, et al. Loss of CHD1 promotes heterogeneous mechanisms of resistance to AR-targeted therapy via chromatin dysregulation. *Cancer Cell* 2020;37:584-598.e11.  
[PUBMED](#) | [CROSSREF](#)
31. Li B, Brady SW, Ma X, Shen S, Zhang Y, Li Y, et al. Therapy-induced mutations drive the genomic landscape of relapsed acute lymphoblastic leukemia. *Blood* 2020;135:41-55.  
[PUBMED](#) | [CROSSREF](#)
32. Wang Q, Chen C, Ding Q, Zhao Y, Wang Z, Chen J, et al. METTL3-mediated m<sup>6</sup>A modification of HDGF mRNA promotes gastric cancer progression and has prognostic significance. *Gut* 2020;69:1193-205.  
[PUBMED](#) | [CROSSREF](#)
33. Xiao YY, Lin L, Li YH, Jiang HP, Zhu LT, Deng YR, et al. ZEB1 promotes invasion and metastasis of endometrial cancer by interacting with HDGF and inducing its transcription. *Am J Cancer Res* 2019;9:2314-30.  
[PUBMED](#)
34. Charbonnier LM, Cui Y, Stephen-Victor E, Harb H, Lopez D, Bleesing JJ, et al. Functional reprogramming of regulatory T cells in the absence of Foxp3. *Nat Immunol* 2019;20:1208-19.  
[PUBMED](#) | [CROSSREF](#)
35. Hashemi V, Maleki LA, Esmaily M, Masjedi A, Ghalamfarsa G, Namdar A, et al. Regulatory T cells in breast cancer as a potent anti-cancer therapeutic target. *Int Immunopharmacol* 2020;78:106087.  
[PUBMED](#) | [CROSSREF](#)
36. Orhan A, Vogelsang RP, Andersen MB, Madsen MT, Hölmich ER, Raskov H, et al. The prognostic value of tumour-infiltrating lymphocytes in pancreatic cancer: a systematic review and meta-analysis. *Eur J Cancer* 2020;132:71-84.  
[PUBMED](#) | [CROSSREF](#)

37. Yang R, Cheng S, Luo N, Gao R, Yu K, Kang B, et al. Distinct epigenetic features of tumor-reactive CD8<sup>+</sup> T cells in colorectal cancer patients revealed by genome-wide DNA methylation analysis. *Genome Biol* 2019;21:2.  
[PUBMED](#) | [CROSSREF](#)
38. Kim SH, Jin H, Meng RY, Kim DY, Liu YC, Chai OH, et al. Activating hippo pathway via Rassf1 by ursolic acid suppresses the tumorigenesis of gastric cancer. *Int J Mol Sci* 2019;20:4709.  
[PUBMED](#) | [CROSSREF](#)
39. Feng Y, Pan L, Zhang B, Huang H, Ma H. BATF acts as an oncogene in non-small cell lung cancer. *Oncol Lett* 2020;19:205-10.  
[PUBMED](#)
40. Alvisi G, Brummelman J, Puccio S, Mazza EM, Tomada EP, Losurdo A, et al. IRF4 instructs effector Treg differentiation and immune suppression in human cancer. *J Clin Invest* 2020;130:3137-50.  
[PUBMED](#) | [CROSSREF](#)