

Identification of an immune gene signature for predicting the prognosis of patients with uterine corpus endometrial carcinoma

Cankun Zhou (✉ 13592854369@163.com)

Southern Medical University Affiliated Maternal & Child Health Hospital of Foshan
<https://orcid.org/0000-0002-7253-6686>

Chaomei Li

Southern Medical University Affiliated Maternal & Child Health Hospital of Foshan

Fangli Yan

Shaoxing People's Hospital

Yuhua Zheng

Southern Medical University Affiliated Maternal & Child Health Hospital of Foshan

Primary research

Keywords: Uterine corpus endometrial carcinoma, TCGA, Immune gene, Prognosis

Posted Date: September 3rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-16901/v3>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on November 9th, 2020. See the published version at <https://doi.org/10.1186/s12935-020-01560-w>.

Abstract

Background: Uterine corpus endometrial carcinoma (UCEC) is a frequent gynecological malignancy with a poor prognosis particularly at an advanced stage. Herein, this study aims to construct prognostic markers of UCEC based on immune-related genes to predict the prognosis of UCEC.

Methods: We analyzed expression data of 575 UCEC patients from The Cancer Genome Atlas database and immune genes from the ImmPort database, which were used for generation and validation of the signature. We constructed a transcription factor regulatory network based on Cistrome databases, and also performed functional enrichment and pathway analyses for the differentially expressed immune genes. Moreover, the prognostic value of 410 immune genes was determined using the Cox regression analysis. We then constructed and verified a prognostic signature. Finally, we performed immune infiltration analysis using TIMER-generating immune cell content.

Results: The immune cell microenvironment as well as the PI3K-Akt, and MARK signaling pathways were involved in UCEC development. The established prognostic signature revealed a ten-gene prognostic signature, comprising of *PDIA3*, *LTA*, *PSMC4*, *TNF*, *SBDS*, *HDGF*, *HTR3E*, *NR3C1*, *PGR*, and *CBLC*. This signature showed a strong prognostic ability in both the training and testing sets and thus can be used as an independent tool to predict the prognosis of UCEC. In addition, levels of B cells and neutrophils were significantly correlated with the patient's risk score, while the expression of ten genes was associated with immune cell infiltrates.

Conclusions: In summary, the ten-gene prognostic signature may guide the selection of the immunotherapy for UCEC.

Background

Uterine corpus endometrial carcinoma (UCEC) is one of the most prevalent malignant tumors in women. According to the current global cancer statistics, endometrial cancer has an incidence of about 4.4% [1], with related morbidity and mortality showing an increase annually despite the recent advances in treatment. This has been attributed primarily due to the lack of biomarkers for early diagnosis and prognosis prediction for this condition [2]. Previous studies have elucidated that the Grade, Stage, and TNM staging of UCEC are closely related to disease prognosis. However, some patients may manifest different clinical outcomes within the same stage group, implying that the clinical prognostic information generated through traditional clinical-pathological staging is insufficient. Therefore, in this context, the identification of highly accurate, reliable, and sensitive markers is critical for improving the prognosis of UCEC patients.

Recent studies have demonstrated the important role played by tumor microenvironment (TME)-stromal cells in tumor proliferation, invasion, and metastasis. These cells are closely related to the prognosis of the disease [3, 4]. In addition, host immune responses, with multiple immune cell infiltrations, are one of the main participants in TME [5, 6]. Several studies have hypothesized that UCEC may be associated with

long-term inflammatory stimuli, suggesting that the endometrium and menstrual cycles are essentially a chronic inflammatory process involving immune cells [7, 8]. Elsewhere, the effect of immune cell-derived cytokines on the survival outcomes of UCEC patients has been described [9-11]. However, the role of immune-related genes on the systematic prediction of overall survival and response to immunotherapy in UCEC remains enigmatic.

Current and emerging knowledge of tumor molecular biology has led to the development of numerous clinical therapeutic approaches for cancer treatment. Besides, attempts have been made to efficiently and accurately assess the effects of therapies, mainly through preclinical models that simulate characteristics of different types of cancers. For instance, Muhammad et al. [12] demonstrated the anti-proliferative activity of bitter melon extract (BME) in breast cancer cells using homozygous and xenograft mouse models. Furthermore, advances in molecular techniques applied to different available preclinical models have greatly increased our understanding of endometrial cancer biology [13]. Presently, the sequencing of the human genome and DNA microarray development has revealed the identification of candidate genes of prognostic or therapeutic value. For example, the Cancer Genome Atlas (TCGA) database provides comprehensive data on the molecular basis of various types of cancer [14].

In this respect, we aimed to identify prognostic immunomarkers and construct a signature for predicting UCEC. A prognostic risk scoring system was validated by testing set A and B. Specifically, we analyzed RNA-seq data from the TCGA database, as well as immune-related genes downloaded from the Immunology Database and Analysis Portal (ImmPort) databases. Subsequently, we assessed whether these immunity genes were associated with survival outcomes and clinical traits in a subgroup of UCEC patients. Lastly, we explored the relationship between risk scores in UCEC patients with immune cell infiltration, using an abundance of six immune infiltrates from the Tumor Immune Estimation Resource (TIMER) database. Therefore, this study may provide new biomarkers for prognosis and also novel immunotherapy insights for UCEC.

Materials And Methods

Publicly attainable expression datasets

We downloaded expression data from the TCGA databases. Specifically, RNA-seq and clinical data of 575 UCEC patients were downloaded from the TCGA database (<https://portal.gdc.cancer.gov>) and used as the training set. Pearson's correlation coefficient was used to eliminate outliers. Afterward, all patients were randomly assigned to the testing set A (n = 270) and the testing set B (n = 271) based on the complete TCGA data set. Collectively, the 3 sets of data were used to construct the signature.

Prognosis-related differential expression gene screening

Differential expression genes (DEGs) analysis of the training set was performed using the "limma" package, at a corrected $P < 0.05$ and $|\logFC| \geq 1$. The resulting data was used to generate heatmaps and volcano plots using R software, version 3.5.3 [15]. Sequence data for immune-related genes and tumor-associated transcription factors were retrieved from the IMMPORT (<http://www.immport.org/>) and the Cistrome (<http://cistrome.org/>) [14, 16], which were used for the identification of differentially expressed immune genes and transcription factors (TF), respectively. The resulting datasets were then used to generate heatmaps and volcano plots, as earlier described. Gene Ontology (GO), functional enrichment analyses, and Kyoto Encyclopedia of Genes and Genome (KEGG) pathway analyses were then executed for differential immune genes using "clusterProfiler, org.Hs.eg.db, plot, ggplot2" packages in R. These analyses were performed at $P < 0.05$ and $q < 0.05$ as cut-offs.

Construction of a regulatory network

Here, differentially expressed immunity genes were combined with survival time, while prognosis-related immune genes were evaluated using Cox univariate analysis. We also drew a forest map with a significance filtering standard $P < 0.01$. Thereafter, a correlation was performed with differential TFs at $|\text{cor}| > 0.4$ and $P < 0.001$ as the filtering criteria. Lastly, the resulting data were imported into Cytoscape version 3.7.1 for visualization of the regulatory network.

Development and validation of the immune prognostic signature for UCEC

The prognostic signature was constructed using the multivariate Cox regression model. Thereafter, the most significant genes concerning prognosis were determined using prognosis-related immune genes. Receiver operating characteristic (ROC) curves for assessing the sensitivity and specificity of the prognostic signature was generated using the "survivalROC" package implemented in R.

Risk scores for each patient were calculated as follows:

$$\text{Risk score} = \sum_{i=1}^n \text{exp}_i * \text{coef}_i$$

Where "n" is the number of prognostic genes, " exp_i " is the expression value of the gene i, and " coef_i " is the gene i coefficient in multivariate Cox regression analysis. Then the median risk value was used to divide the patients into high and low-risk groups, while the Kaplan-Meier curve was applied to assess the survival difference between the two groups using the log-rank test. Subsequently, a risk curve was drawn using the "pheatmap" R package.

To determine the feasibility and reliability of the ten-gene prognostic signature, we used testing set A (n = 270) and testing set B (n = 271) of the TCGA sample according to the "Publicly attainable expression

datasets" section above. All of the findings are summarized in Table A in S1 File.

Independent prognostic analysis

Combined with the risk score and clinical data of each sample for independent prognostic analysis, single-factor and multi-factor independent prognostic analysis were used to assess the prognostic value of immune-gene signature and clinical parameters. This was also used to ascertain whether the predictive power of the immune-gene signature was independent of other clinical parameters.

Relationship among clinical parameters

To assess the association between immunity genes in the prognostic signature and clinical parameters, patients were divided into two subgroups using univariate Cox regression analysis. The first group comprised of patients who were ≤ 55 whereas the second one had those who were > 55 years old. Next, clinically relevant immune genes across patients in the 2 groups were screened and mapped ($P < 0.05$) using the "beeswarm" package.

Immunohistochemistry

The Human Protein Atlas (<https://www.proteinatlas.org>) contains information on tissue and cellular distribution for all 24,000 human proteins. The database applies immunohistochemistry using specific antibodies to analyze differentially expressed proteins in normal and tumor tissues. Herein, we examined this database to analyze profiles of protein expression in ten genes across normal uterine and endometrial carcinoma tissues.

Correlation between immune cell content and the signature genes

Data on the abundance of six immune infiltrates, including B, CD4+ T, CD8+ T, and dendritic cells as well as neutrophils, and macrophages were retrieved from the TIMER official website (<https://cistrome.shinyaoos.io/timer/>). Consequently, these data were used to analyze the relationship between risk scores of UCEC patients and the aforementioned immune cells. Furthermore, we explored a correlation between the abundance of immune cells and gene expression as described in [17].

Statistical analyses

Data on survival analysis was investigated using the Kaplan-Meier curve, with statistical differences determined using the log-rank test. The area under the curve (AUC) of the ROC was used to analyze prediction accuracy of the prognostic signature, whereas effects of clinical traits on overall survival (OS) were assessed using univariate Cox and multivariate Cox regression analyses. The hazard ratio (HR) and 95% confidence interval (CI) were generated using the Cox proportional hazards model. Lastly, Univariate Cox regression analysis was employed to evaluate the correlation between immune cells and gene expression. Data with $P < 0.05$ were considered statistically significant.

Results

Identification of DEGs in UCEC

Overall, we identified 6,268 DEGs, 410 candidate prognostic immune genes, and 100 differential TFs (**Fig. 1 and 2**). The differential expression of immune genes in all endometrial cancer samples is summarized in Supplementary **File 1**. Enrichment analysis of differentially expressed immunity genes showed that biological processes (BP), mainly chemotaxis migration of anti-inflammatory cells, including leukocyte and neutrophils, were primarily enriched (**Fig. 3A**). The enriched cellular components (CC) were mainly extracellular matrix whereas the main molecular function (MF) comprised of growth factor and cytokine activity. These findings implied that most differentially expressed immunity genes were associated with UCEC development, progression, and prognosis through immune cells. The enriched top 30 KEGG pathways are given in **Fig. 3B**. Notably, several signaling pathways involved in UCEC development, including PI3K-Akt, MAPK, Ras, and JAK-STAT, were identified.

Association between immunity genes and survival rates

Univariate Cox regression analysis of differentially expressed immunity genes revealed a significant correlation between survival rates and 21 of candidate prognostic immune genes ($P < 0.01$). In particular, the genes with a significant association were *PDIA3*, *LTA*, *PSMC4*, *IL6*, *TNF*, *KCNH2*, *SYTL1*, *BACH2*, *PCSK1*, *BIRC5*, *SBDS*, *ANGPTL7*, *GPI*, *HDGF*, *ADCYAP1R1*, *HTR3E*, *NPR1*, *NR3C1*, *PGR*, *THRB*, and *CBLC*. Among them, *LTA*, *ADCYAP1R1*, *PGR*, *SYTL1*, and *PDIA3* were characterized as low-risk, while the remaining 16 were categorized as high-risk genes. Detailed information of all 21 genes is depicted in **Fig. 4A**.

Furthermore, to assess the relationship between the 21 prognosis-related immunity genes and TFs, a univariate Cox regression analysis was performed at $|\text{cor}| > 0.4$ and $P < 0.001$, where a TF regulatory network was constructed (**Fig. 4B**). Notably, the regulatory network diagram that comprised of low- (*PGR*, *SYTL1*, and *LTA*), and high-risk (*BIRC5*, *HDGF*, *HTR3E*, *THRB*, *NR3C1*, *BACH2*) genes, as well as TFs (*AR*, *BATF*, *CBX2*, *CENPA*, *E2F1*, *E2F3*, *ETS1*, *EZH2*, *FOXK1*, *FOXP3*, *GREB1*, *H2AFX*, *LMNB1*, *LYL1*, *NCAPG*, *NR3C1*, *RFX2*, *SNAI2*, *SOX17*, *SPDEF*, *SPIB*, *STAT5A*, and *WWTR1*), elucidated a positive relationship

between immunity genes and TFs. Lastly, *BIRC5* was associated with several transcription factors namely *CBX2*, *CENPA*, *E2F1*, *EZH2*, *FOXM1*, *H2AFX*, *LMNB1*, and *NCAPG*.

The prognostic prediction signature

To establish a signature for predicting the prognosis of UCEC patients, we employed a Cox regression analysis and identified a ten-gene prognostic signature based on a training set. The genes in the signature included *PDIA3*, *LTA*, *PSMC4*, *TNF*, *SBDS*, *HDGF*, *HTR3E*, *NR3C1*, *PGR*, and *CBLC* (**Table 1**). We used the prognostic signature to calculate a risk score for each patient, while the median value was used to divide the patients into a high-risk (n=270) and low-risk groups (n=271) (Supplementary **File 2** showed the risk score and immune gene expression per patient of the signature in the training set). The prediction power of the ten-gene prognostic signature for patients in training sets is outlined in **Fig. 5**, while the distribution of risk scores, gene expression levels, and patient survival status are displayed in **Fig. 5A**. Remarkably, AUC for the training set was 0.756, indicating good accuracy of the prognostic prediction-values across the ten-gene prognostic signature. From the Kaplan-Meier curve, lower overall survival rates were recorded for patients in the high-risk compared to those in the low-risk group for the training set ($P < 0.0001$) (**Fig. 5C**). Besides, 5-year OS rates of 63.1 and 89.9%, were recorded for patients in the high- and low-risk groups, respectively, whereas 9-year OS rates were 34.6 and 78.7% for patients in the high- and low-risk groups, respectively.

Validation of the ten-gene prognostic signature in UCEC

To determine the feasibility and reliability of the ten-gene prognostic signature, we validated it using testing set A (n=270) and testing set B (n=271). In the testing sets A and B, a shorter overall survival rate was noted for patients in the high risk compared to those in the low-risk groups ($P < 0.0001$) (**Fig6E/F**). The AUC for the testing set A and B were 0.706 and 0.885 (**Fig.6C/D**), respectively, suggesting that the signature strongly predicts overall survival in UCEC patients. (Supplementary **File 3 and File 4** showed the risk score and immune gene expression per patient of the signature in the testing set A and the testing set B, respectively.)

The ten-gene prognostic signature is an independent prognostic factor

To determine whether the signature risk score was an independent prognostic factor for patient survival, we employed univariate and multivariate Cox regression analyses. Results demonstrated $P < 0.05$, across both analyses, indicating that the risk score derived from the signature can be independent of other clinical traits, and thus an independent prognostic factor. In addition, univariate Cox regression analysis showed that age ($P = 0.002$, hazard ratio=1.035) and grade ($P < 0.001$, hazard ratio=2.595) were

significantly associated with prognosis. Of note, the prognosis of patients was worse with an increase in age and grade (**Fig. 7**).

Clinical parameters, immunohistochemical examination

The correlation between immune genes involved in the signature and clinical traits was assessed using Univariate Cox regression analysis. Here, patients were divided into two groups, based on clinical traits: Group 1 (comprised of patients aged <55 and ≥ 55) and Group 2 (G1 & G2 and G3). Results revealed a significant correlation between HDGF ($P < 0.001$), PGR ($P = 0.04$), PSMC4 ($P < 0.001$), TNF ($P < 0.001$), NR3C1 ($P = 0.015$), HTR3E ($P = 0.033$) and CBLC ($P = 0.003$) with age, whereas expression of HDGF, PSMC4, TNF, NR3C1, HTR3E, and CBLC increased with age. On the other hand, HDGF ($P < 0.001$), PGR ($P < 0.001$), PSMC4 ($P < 0.001$), TNF ($P < 0.001$), NR3C1 ($P < 0.001$), PDIA3 ($P < 0.001$), and SBDS ($P < 0.001$) were significantly associated with grade. Moreover, an increase in grade resulted in the upregulation of HDGF, PSMC4, TNF, NR3C1, and SBDS (**Fig. 8**). Immunohistochemical analysis based on The Human Protein Atlas database enumerated a significant upregulation of PSMC4, NR3C1, SBDS, and CBLC in endometrial cancer tissues, relative to normal tissues. On the other hand, immunohistochemical analysis of PGR and PDIA3 expression showed significant downregulation of these factors in endometrial cancer compared to normal tissues (**Fig. 9**).

Immune infiltrates analysis of the signature genes in patients with UCEC

Herein, a correlation analysis between risk scores in UCEC patients with abundance of six immune infiltrations indicated a significant positive association between B cells ($P = 3.408e-10$, $cor=0.265$) and neutrophils ($P = 0.011$, $Cor = 0.109$) with the patient's risk score (**Fig. 10**). To explain this relationship, we analyzed infiltration abundance, and noted a positive relationship between B cells and expression of LTA ($Cor=0.594$, $P = 5.55e-29$) (**Fig. 11B**), TNF ($Cor=0.117$, $P = 4.60e-02$) (**Fig. 11D**), and NR3C1 ($Cor=0.301$, $P=1.85e-07$) (**Fig. 11H**). Moreover, the infiltration abundance of neutrophils was positively correlated with expression of LTA ($Cor=0.339$, $P = 2.65e-09$) (**Fig. 11B**), PSMC4 ($Cor=0.209$, $P = 3.23e-04$) (**Fig. 11C**), TNF ($Cor=0.408$, $P = 3.56e-13$) (**Fig. 11D**), SBDS ($Cor=0.418$, $P = 7.89e-14$) (**Fig. 11E**), HDGF ($Cor=0.309$, $P = 6.50e-08$) (**Fig. 11F**), and NR3C1 ($Cor=0.48$ $P = 2.70e-18$) (**Fig. 11H**).

Discussion

Numerous reports have described the relationship between differentially expressed genes and various aspects of tumors, including tumorigenesis and prognosis [18-20]. However, a vast majority of genes implicated in playing a central role in predicting tumor prognosis are limited by certain factors, such as insufficient sample sizes. In this study, we employed a large sample size comprising of TCGA genome-wide expression data to develop a ten-gene prognostic signature for UCEC patients. The signature is

anticipated to guide the identification of potential biomarkers that can monitor the prognosis and response to immunotherapy in UCEC patients.

Our results revealed an association between differentially expressed immune genes with immune cell responses to extracellular matrix and tumor microenvironment in UCEC, which was parallel with previous studies [21]. In addition, KEGG enrichment analysis showed that UCEC may be associated with several well-known cancer-related pathways, including the PI3K-Akt, MAPK, Ras, and JAK-STAT signaling pathways. Previous studies have demonstrated the activation of the PI3K-AKT signaling pathway in UCEC patients, as well as its role in promoting tumor development [22]. Also, several studies have revealed multiple factors that activate MAPK, Ras, and JAK-STAT signaling pathways, thereby mediating proliferation, infiltration, and other biological behaviors to promote the occurrence and progression of UCEC [23-25]. Our TF-related regulatory network showed that BIRC5 was positively regulated by multiple TFs, and BIRC5 was a high-risk gene. This was consistent with previous studies demonstrating that the up-regulation of BIRC5 leads to the development and progression of many malignant tumors in humans [26]. Elsewhere, BIRC5 was noted to be overexpressed in more than 90% of UCEC [27], while another study demonstrated frequent overexpression of BIRC5 in recurrent UCEC relative to primary tumors [28].

Herein, we developed a ten-gene prognostic signature, comprising of *PDIA3*, *LTA*, *PSMC4*, *TNF*, *SBDS*, *HDGF*, *HTR3E*, *NR3C1*, *PGR*, and *CBLC*, for prediction of overall survival rates in UCEC patients. Our findings indicated that the signature effectively predicted the OS of UCEC patients, with a statistically significant correlation between the training and testing sets. These findings signify the potential of this signature as a powerful prognostic tool for the entire cohort of UCEC patients. In addition to hepatoma-derived growth factor (HDGF) and Protein disulfide-isomerase A3 (PDIA3), the remaining 8 genes have not been well validated in gynecologic oncology, especially in UCEC. The HDGF is a heparin-binding growth factor that has been purported to play a crucial role in the differentiation, growth, and division of various tissues. Several studies have demonstrated its involvement in the occurrence and development of malignant tumors, promoting proliferation and differentiation of tumor cells, as well as enhancing the metastatic ability of tumor cells through EMT [29, 30]. Besides, HDGF was noted to be an independent risk factor for the prognosis of various malignancies such as liver, gastric, cholangiocarcinoma, and non-small cell lung cancers [31]. However, in endometrial cancer, HDGF has been implicated in multiple abnormalities. For example, a higher FIGO stage mediated HDGF upregulation, a potential adverse factor for the progression and prognosis of UCEC [32].

On the other hand, PDIA3, also known as ERp57/GRP58, has been associated with malignant transformation of cells through STAT3 and Wnt signaling pathways. Also, this factor has been closely associated with the occurrence and development of various tumors[33]. Interestingly, PDIA3 has been reported to enhance the ability of cervical and ovarian cancer cells to proliferate and invade, indicating its potential as a sensitive marker for reflecting tumor prognosis during gynecologic oncology [34, 35]. In this work, these two immune genes were key DEGs ($P < 0.0001$), suggesting their possible role in the development and progression of UCEC. Notably, the overall survival rate of patients in the high-risk group was lower than those in the low-risk group, whereas the AUC values showed that a combination of the ten

immune genes exhibited a prognostic value in UCEC patients. Of note, our signature was an independent prognostic factor, because the predicted survival rates were not related to other clinical traits.

In the present study, our results revealed that age and grade were associated with the OS of UCEC patients in high-risk factors. This corresponded with previous studies showing that age, stage, and body weight are clinical prognostic factors for UCEC [36]. Conversely, age and grade were also associated with the prognosis of endometrial cancer and were also a high-risk factor for the disease. This finding agrees with the reports of previous studies that have described age, stage, and body weight as clinical prognostic factors for endometrial cancer. Further correlation analysis revealed a significant positive correlation between HDGF and PSMC4 with age and grade. This may be attributed to the fact that the up-regulation of these genes could promote tumor development [37, 38]. In terms of survival prediction, the current staging system is far from accurate at the individual level. Elsewhere, age is not a survival indicator for cancer because older people are less likely to receive adjuvant therapy [39]. Therefore, risk scores present a more reliable tool for the prognosis of UCEC patients compared to age and stage.

Currently, numerous studies have hypothesized the involvement of immune cells and related inflammatory factors in the UCEC interstitial, which is an important component of the tumor inflammatory microenvironment and generates a marked influence on the biological behavior of UCEC [40]. Consequently, we analyzed the relationship between UCEC risk-score and immune cells using immune cell infiltration abundance data from TIMER. Results indicated a significant positive correlation between B cells and neutrophils with the patient's risk-scores. Furthermore, we found a close relationship between prognostic signature genes and immune cells. Among them, neutrophils were positively correlated with the expression of several genes, including *LTA*, *PSMC4*, *TNF*, *SBDS*, *HDGF*, and *NR3C1*. This phenomenon may be attributed to the secretion of HDGF, which has been shown to promote neutrophil infiltration and induce inflammatory signals [41]. In another study by Wikberg et al. [42] demonstrated that neutrophils of the innate immune system play a significant role in acute inflammation as well as in anti-tumor immune responses. Despite the close association between neutrophil infiltration with other immune cell infiltration, studies have enumerated that neutrophil infiltration may have additional prognostic values in various cancers. For example, neutrophils persist in tissues, during chronic inflammation, causing cancer progression. Several studies have also shown that elevated numbers of neutrophils in many human cancers or a higher neutrophil/lymphocyte ratio (NLR) are associated with poor prognosis possibly because neutrophils secrete matrix metalloproteinase-9 to stimulate the angiogenic activity of cancer cells [43, 44]. Different proportions of infiltrating B cells were included in solid human tumors. Although the search for immune-related factors associated with a cancer diagnosis, prognosis, and survival has been largely limited to T cell responses, recent reports have suggested that B cells may also play critical roles in the prognosis of cancer patients. For example, findings by Schimidt et al. [45, 46] outlined that the B cell marker was the strongest prognostic factor in breast cancer and other human tumors, given the immunoglobulin kappa chain (IGKC) secreted by plasma cells. On the other hand, another work by Nielsen et al. [47] found that increase in CD20+ B cells resulted in higher survival rates of patients with advanced ovarian cancer. Hence, an increase in the risk-score is likely to elevate levels of these two immune cells and thereby influence immune escape or suppression.

Despite the important clinical value of these findings in UCEC, there were several limitations to our study. Firstly, age and grade were the only clinical traits in the TCGA database of UCEC, although related aspects such as stage and TMN may strengthen the value of the identified genes. Secondly, most of the ten-gene prognostic signature and immune cells have rarely been reported in UCEC patients. In this regard, more prospective studies are needed to validate the intrinsic relevance of these genes in the prognosis of UCEC patients.

Conclusions

In summary, this study aimed to construct immune gene-related prognostic signature and potential functions of immune genes in the signature. Here, we present a ten-gene prognostic signature that is an independent prognostic factor and might complement clinical features and facilitate personalized immunotherapy in UCEC patients.

Abbreviations

UCEC: Uterine corpus endometrial carcinoma; TME: Tumor microenvironment; BME: Bitter melon extract; TCGA: The Cancer Genome Atlas; ImmPort: Immunology Database and Analysis Portal; TIMER: Tumor Immune Estimation Resource; DEG: Differential expression genes; TF: Transcription factor; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genome; ROC: Receiver operating characteristic; AUC: Area under curve; OS: Overall survival; HR: Hazard ratio; CI: Confidence interval; BP: Biological processes; CC: Cellular components; MF: Molecular function; HDGF: Hepatoma-derived growth factor; PDIA3: Protein disulfide-isomerase A3; NLR: Neutrophil/Lymphocyte Ratio; IGKC: Immunoglobulin kappa chain;

Declarations

Acknowledgements

The information of this study is obtained by the TCGA, ImmPort, Cistrome, TIMER, and the Human Protein Atlas database. We are grateful to them for the source of data used in our study.

Author's contributions

Cankun Zhou and Yuhua Zheng performed data analysis work and aided in writing the manuscript. Cankun Zhou and Chaomei Li designed the study, assisted in writing the manuscript. Fangli Yan edited the manuscript. All authors read and approved the final manuscript.

Funding

None applicable.

Availability of data and materials

The public datasets used in our work can be found on <https://portal.gdc.cancer.gov>, <http://www.immport.org>, <http://cistrome.org/>, <https://www.proteinatlas.org>, and <https://cistrome.shinyaoos.io/timer/>.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interest.

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: a cancer journal for clinicians*. 2018, 68(6):394-424.
2. Lee YC, Lheureux S, Oza AM. Treatment strategies for endometrial cancer: current practice and perspective. *Current opinion in obstetrics & gynecology*. 2017, 29(1):47-58.
3. Pitt JM, Marabelle A, Eggermont A, Soria JC, Kroemer G, Zitvogel L. Targeting the tumor microenvironment: removing obstruction to anticancer immune responses and immunotherapy. *Annals of oncology : official journal of the European Society for Medical Oncology*. 2016, 27(8):1482-1492.
4. Najafi M, Goradel NH, Farhood B, Salehi E, Solhjoo S, Toolee H, Kharazinejad E, Mortezaee K. Tumor microenvironment: Interactions and therapy. *Journal of cellular physiology*. 2019, 234(5):5700-5721.
5. Murata T, Mizushima H, Chinen I, Moribe H, Yagi S, Hoffman RM, Kimura T, Yoshino K, Ueda Y, Enomoto T et al. HB-EGF and PDGF mediate reciprocal interactions of carcinoma cells with cancer-

- associated fibroblasts to support progression of uterine cervical cancers. *Cancer research*. 2011, 71(21):6633-6642.
6. Greten FR, Grivennikov SI. Inflammation and Cancer: Triggers, Mechanisms, and Consequences. *Immunity*. 2019, 51(1):27-41.
 7. Modugno F, Ness RB, Chen C, Weiss NS. Inflammation and endometrial cancer: a hypothesis. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2005, 14(12):2840-2847.
 8. Holub K, Biete A. New pre-treatment eosinophil-related ratios as prognostic biomarkers for survival outcomes in endometrial cancer. *BMC cancer*. 2018, 18(1):1280.
 9. Walentowicz-Sadlecka M, Sadlecki P, Bodnar M, Marszalek A, Walentowicz P, Sokup A, Wilinska-Jankowska A, Grabiec M. Stromal derived factor-1 (SDF-1) and its receptors CXCR4 and CXCR7 in endometrial cancer patients. *PloS one*. 2014, 9(1):e84629.
 10. Felix AS, Stone RA, Chivukula M, Bowser R, Parwani AV, Linkov F, Edwards RP, Weissfeld JL. Survival outcomes in endometrial cancer patients are associated with CXCL12 and estrogen receptor expression. *International journal of cancer*. 2012, 131(2):E114-121.
 11. 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. *Gynecologic oncology*. 2014, 132(3):599-605.
 12. Muhammad N, Steele R, Isbell TS, Philips N, Ray RB. Bitter melon extract inhibits breast cancer growth in preclinical model by inducing autophagic cell death. *Oncotarget*. 2017, 8(39):66226-66236.
 13. Van Nyen T, Moiola CP, Colas E, Annibali D, Amant F. Modeling Endometrial Cancer: Past, Present, and Future. *International journal of molecular sciences*. 2018, 19(8).
 14. Bhattacharya S, Dunn P, Thomas CG, Smith B, Schaefer H, Chen J, Hu Z, Zalocusky KA, Shankar RD, Shen-Orr SS et al. ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. *Scientific data*. 2018, 5:180015.
 15. 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 research*. 2015, 43(7):e47.
 16. Mei S, Qin Q, Wu Q, Sun H, Zheng R, Zang C, Zhu M, Wu J, Shi X, Taing L et al. Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. *Nucleic acids research*. 2017, 45(D1):D658-d662.
 17. 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 research*. 2017, 77(21):e108-e110.
 18. Van Gool IC, Stelloo E, Nout RA, Nijman HW, Edmondson RJ, Church DN, MacKay HJ, Leary A, Powell ME, Mileskin L et al. Prognostic significance of L1CAM expression and its association with mutant p53 expression in high-risk endometrial cancer. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*. 2016, 29(2):174-181.
 19. Zheng F, Yu H, Lu J. High expression of MUC20 drives tumorigenesis and predicts poor survival in endometrial cancer. *Journal of cellular biochemistry*. 2019, undefined(undefined):undefined.

20. Ying J, Wang Q, Xu T, Lyu J. Establishment of a nine-gene prognostic model for predicting overall survival of patients with endometrial carcinoma. *Cancer medicine*. 2018, 7(6):2601-2611.
21. Junttila MR, de Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature*. 2013, 501(7467):346-354.
22. Chen HX, Xu XX, Tan BZ, Zhang Z, Zhou XD. MicroRNA-29b Inhibits Angiogenesis by Targeting VEGFA through the MAPK/ERK and PI3K/Akt Signaling Pathways in Endometrial Carcinoma. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology*. 2017, 41(3):933-946.
23. Wang J, Sun X, Zhang H, Wang Y, Li Y. MPA influences tumor cell proliferation, migration, and invasion induced by RANKL through PRB involving the MAPK pathway in endometrial cancer. *Oncology reports*. 2015, 33(2):799-809.
24. Chiu HC, Li CJ, Yiang GT, Tsai AP, Wu MY. Epithelial to Mesenchymal Transition and Cell Biology of Molecular Regulation in Endometrial Carcinogenesis. *Journal of clinical medicine*. 2019, 8(4):undefined.
25. Subramaniam KS, Omar IS, Kwong SC, Mohamed Z, Woo YL, Mat Adenan NA, Chung I. Cancer-associated fibroblasts promote endometrial cancer growth via activation of interleukin-6/STAT-3/c-Myc pathway. *American journal of cancer research*. 2016, 6(2):200-213.
26. Garg H, Suri P, Gupta JC, Talwar GP, Dubey S. Survivin: a unique target for tumor therapy. *Cancer cell international*. 2016, 16(undefined):49.
27. Chuwa AH, Sone K, Oda K, Ikeda Y, Fukuda T, Wada-Hiraike O, Inaba K, Makii C, Takeuchi M, Oki S et al. Significance of survivin as a prognostic factor and a therapeutic target in endometrial cancer. *Gynecologic oncology*. 2016, 141(3):564-569.
28. Vanderstraeten A, Everaert T, Van Bree R, Verbist G, Luyten C, Amant F, Tuyaerts S. In Vitro Validation of Survivin as Target Tumor-associated Antigen for Immunotherapy in Uterine Cancer. *Journal of immunotherapy (Hagerstown, Md : 1997)*. 2015, 38(6):239-249.
29. Tsai HE, Liu GS, Kung ML, Liu LF, Wu JC, Tang CH, Huang CH, Chen SC, Lam HC, Wu CS et al. Downregulation of hepatoma-derived growth factor contributes to retarded lung metastasis via inhibition of epithelial-mesenchymal transition by systemic POMC gene delivery in melanoma. *Molecular cancer therapeutics*. 2013, 12(6):1016-1025.
30. Wang S, Fang W. Increased expression of hepatoma-derived growth factor correlates with poor prognosis in human nasopharyngeal carcinoma. *Histopathology*. 2011, 58(2):217-224.
31. Bao C, Wang J, Ma W, Wang X, Cheng Y. HDGF: a novel jack-of-all-trades in cancer. *Future oncology (London, England)*. 2014, 10(16):2675-2685.
32. Wang L, Jiang Q, Hua S, Zhao M, Wu Q, Fu Q, Fang W, Guo S. High nuclear expression of HDGF correlates with disease progression and poor prognosis in human endometrial carcinoma. *Disease markers*. 2014, 2014(undefined):298795.
33. Paroo Z, Ye X, Chen S, Liu Q. Phosphorylation of the human microRNA-generating complex mediates MAPK/Erk signaling. *Cell*. 2009, 139(1):112-122.

34. Kuang Y, Cai J, Li D, Han Q, Cao J, Wang Z. Repression of Dicer is associated with invasive phenotype and chemoresistance in ovarian cancer. *Oncology letters*. 2013, 5(4):1149-1154.
35. Muralidhar B, Goldstein LD, Ng G, Winder DM, Palmer RD, Gooding EL, Barbosa-Morais NL, Mukherjee G, Thorne NP, Roberts I et al. Global microRNA profiles in cervical squamous cell carcinoma depend on Drosha expression levels. *The Journal of pathology*. 2007, 212(4):368-377.
36. Morice P, Leary A, Creutzberg C, Abu-Rustum N, Darai E. Endometrial cancer. *Lancet (London, England)*. 2016, 387(10023):1094-1108.
37. Shih TC, Tien YJ, Wen CJ, Yeh TS, Yu MC, Huang CH, Lee YS, Yen TC, Hsieh SY. MicroRNA-214 downregulation contributes to tumor angiogenesis by inducing secretion of the hepatoma-derived growth factor in human hepatoma. *Journal of hepatology*. 2012, 57(3):584-591.
38. Hellwinkel OJ, Asong LE, Rogmann JP, Sültmann H, Wagner C, Schlomm T, Eichelberg C. Transcription alterations of members of the ubiquitin-proteasome network in prostate carcinoma. *Prostate cancer and prostatic diseases*. 2011, 14(1):38-45.
39. Gately L, Collins A, Murphy M, Dowling A. Age alone is not a predictor for survival in glioblastoma. *Journal of neuro-oncology*. 2016, 129(3):479-485.
40. Dossus L, Lukanova A, Rinaldi S, Allen N, Cust AE, Becker S, Tjonneland A, Hansen L, Overvad K, Chabbert-Buffet N et al. Hormonal, metabolic, and inflammatory profiles and endometrial cancer risk within the EPIC cohort—a factor analysis. *American journal of epidemiology*. 2013, 177(8):787-799.
41. Chu TH, Huang ST, Yang SF, Li CJ, Lin HW, Weng BC, Yang SM, Huang SC, Wu JC, Chang YC et al. Hepatoma-derived growth factor participates in Helicobacter Pylori-induced neutrophils recruitment, gastritis and gastric carcinogenesis. *Oncogene*. 2019, 38(37):6461-6477.
42. Wikberg ML, Ling A, Li X, Öberg Å, Edin S, Palmqvist R. Neutrophil infiltration is a favorable prognostic factor in early stages of colon cancer. *Human pathology*. 2017, 68(undefined):193-202.
43. Shimizu K, Okita R, Saisho S, Yukawa T, Maeda A, Nojima Y, Nakata M. Prognostic nutritional index before adjuvant chemotherapy predicts chemotherapy compliance and survival among patients with non-small-cell lung cancer. *Therapeutics and clinical risk management*. 2015, 11:1555-1561.
44. Kuang DM, Zhao Q, Wu Y, Peng C, Wang J, Xu Z, Yin XY, Zheng L. Peritumoral neutrophils link inflammatory response to disease progression by fostering angiogenesis in hepatocellular carcinoma. *J Hepatol*. 2011, 54(5):948-955.
45. Schmidt M, Hellwig B, Hammad S, Othman A, Lohr M, Chen Z, Boehm D, Gebhard S, Petry I, Lebrecht A et al. A comprehensive analysis of human gene expression profiles identifies stromal immunoglobulin κ C as a compatible prognostic marker in human solid tumors. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2012, 18(9):2695-2703.
46. Whiteside TL, Ferrone S. For breast cancer prognosis, immunoglobulin kappa chain surfaces to the top. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2012, 18(9):2417-2419.

47. Nielsen JS, Sahota RA, Milne K, Kost SE, Nesslinger NJ, Watson PH, Nelson BH. CD20+ tumor-infiltrating lymphocytes have an atypical CD27- memory phenotype and together with CD8+ T cells promote favorable prognosis in ovarian cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2012, 18(12):3281-3292.

Table

Table 1 The prognostic model of prognosis-related 10 immune genes

ID	Coef	HR (95% CI)	P-value
PDIA3	-0.00226	0.997739 (0.994865-1.000621)	0.124063
LTA	-1.0635	0.345246 (0.170688-0.698318)	0.003086
PSMC4	0.003143	1.003148 (1.000176-1.00613)	0.037905
TNF	0.02748	1.027862 (1.00758-1.048551)	0.006879
SBDS	0.017278	1.017428 (1.00557-1.029426)	0.00387
HDGF	0.002992	1.002996 (0.999056-1.006952)	0.136281
HTR3E	0.548394	1.730472 (1.42863-2.096087)	2.05E-08
NR3C1	0.153559	1.165977 (0.996923-1.363698)	0.054681
PGR	-0.02721	0.973152 (0.945302-1.001823)	0.066204
CBLC	0.01077	1.010828 (1.004709-1.016984)	0.000508

Abbreviation: Coef, coefficient. HR, hazard ratio. CI, confidence interval.

Figures

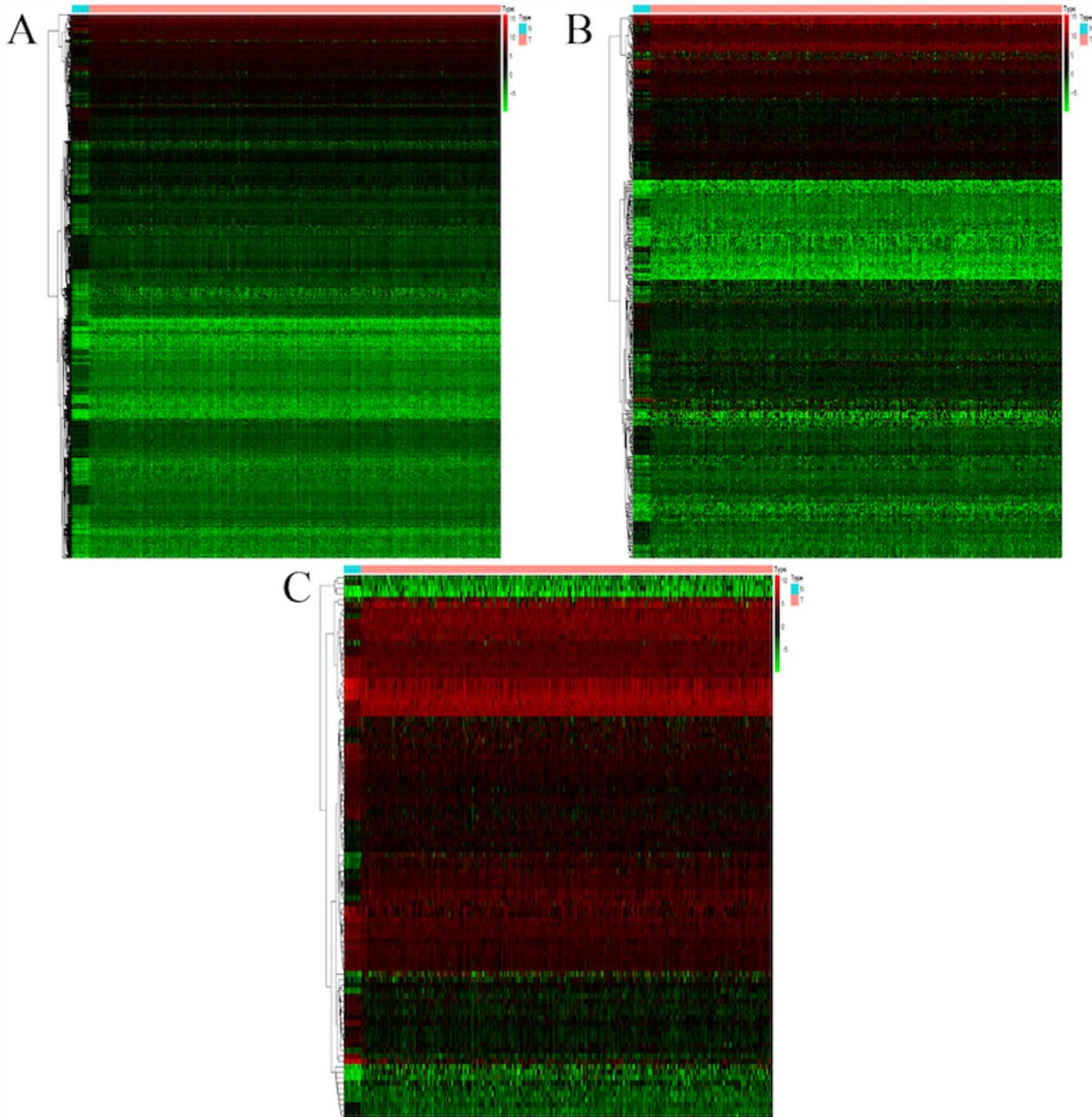


Figure 1

Hierarchical clustering heatmap of DEGs. The genes with higher expression in the heatmap are shown in red, lower expression in green, and genes with the same expression level in black. Tiffany blue represents the adjacent tissue, and the pink represents the cancer tissue. (A) DEGs of RNA-seq gene expression. (B) DEGs of immune genes. (C) DEGs of TFs. Abbreviation: DEGs, differentially expressed genes. TFs, transcription factors.

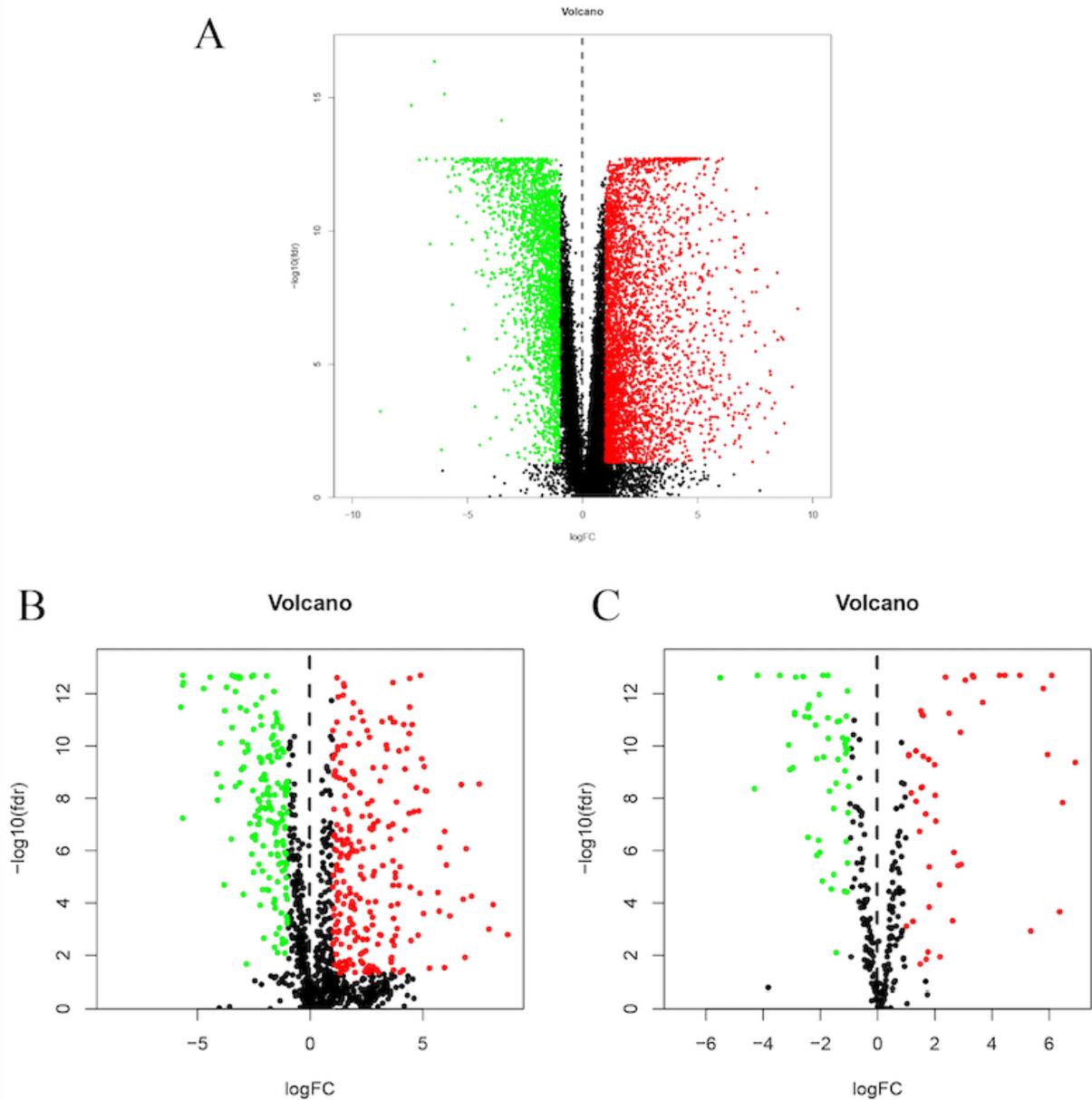


Figure 2

Volcano plot of DEGs. The red dots represent the up-regulated genes screened on the basis of corrected $P < 0.05$ and $\log_{2}FC \geq 1$. The green dots represent the down-regulated genes screened on the basis of corrected $P < 0.05$ and $\log_{2}FC \leq -1$. The black dots represent genes with no significant differences. (A) DEGs of RNA-seq gene expression. (B) DEGs of immune genes. (C) DEGs of TFs. Abbreviation: DEGs, differentially expressed genes. TFs, transcription factors. FC, fold change.

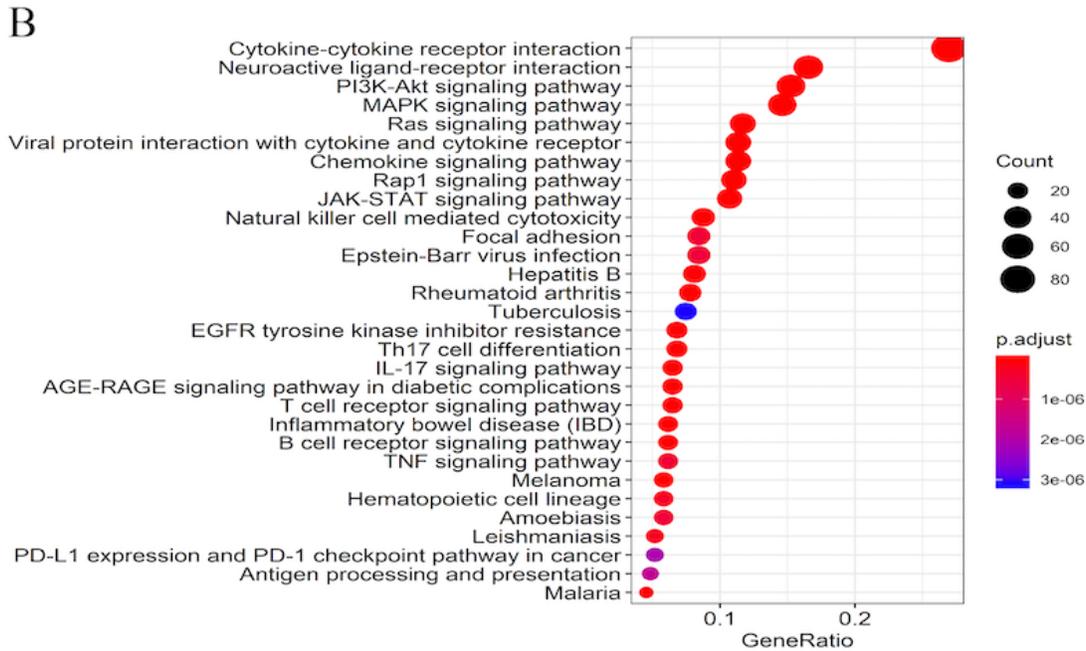
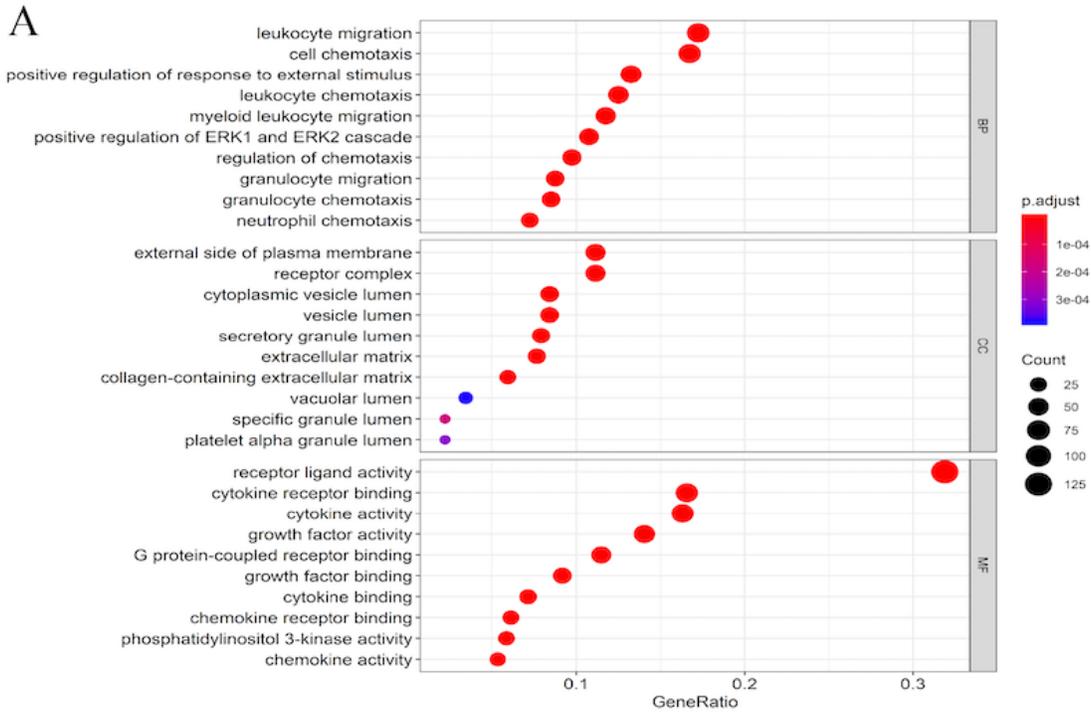


Figure 3

Functional enrichment analysis of DEGs of immune genes. (A) Biological process, Cellular composition, and Molecular function of GO enrichment analysis. (B) Enrichment analysis of KEGG pathway. The results were as follows: BP (Biological process) mainly included the chemotaxis migration of anti-inflammatory cells, including leukocyte and neutrophil, CC (Cellular composition) mainly included extracellular matrix, and MF (Molecular function) mainly included growth factor and cytokine activity. The

enriched KEGG pathway mainly included PI3K-Akt, MAPK, Ras, and JAK-STAT signaling pathways. Abbreviation: DEG, differentially expressed gene. GO, Gene Ontology. KEGG, Kyoto Encyclopedia of Genes and Genomes.

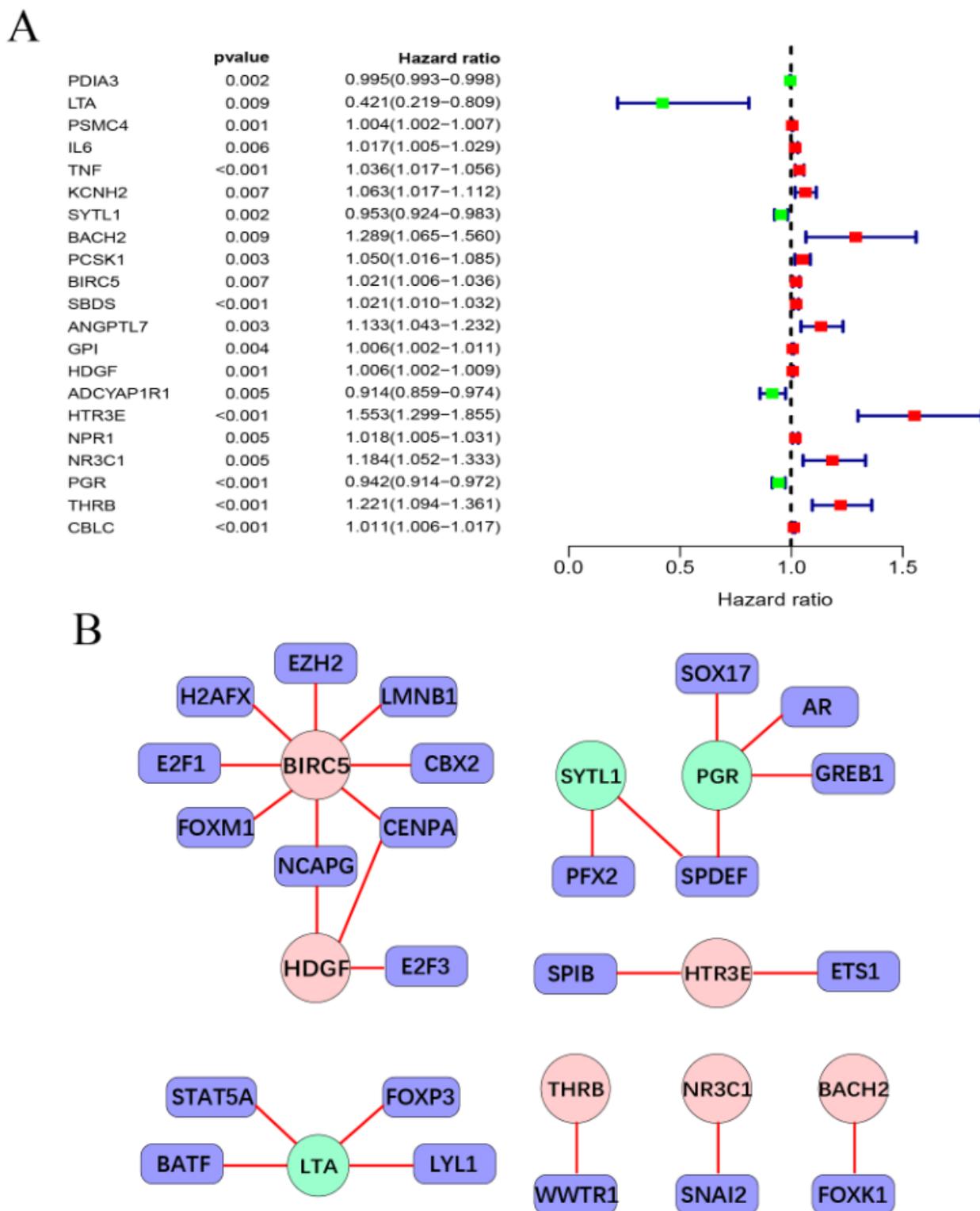


Figure 4

Correlation analysis of prognosis-related immune genes and TFs. (A) Forest of prognosis-related immune genes on the basis of $P\text{-value} \leq 0.001$. Red means high risk, green means low risk. The higher the Hazard

ratio, the higher the prognostic risk. (B) TFs and prognosis-related immune gene regulatory networks on the basis of $|\text{cor}| \geq 0.4$ and $P\text{-value} \leq 0.001$. Blue round rectangle represents TFs, dark pink and light green ovals represent high-risk and low-risk prognosis-related immune genes, respectively, and red lines represent positive regulation. Abbreviation: TFs, transcription factors.

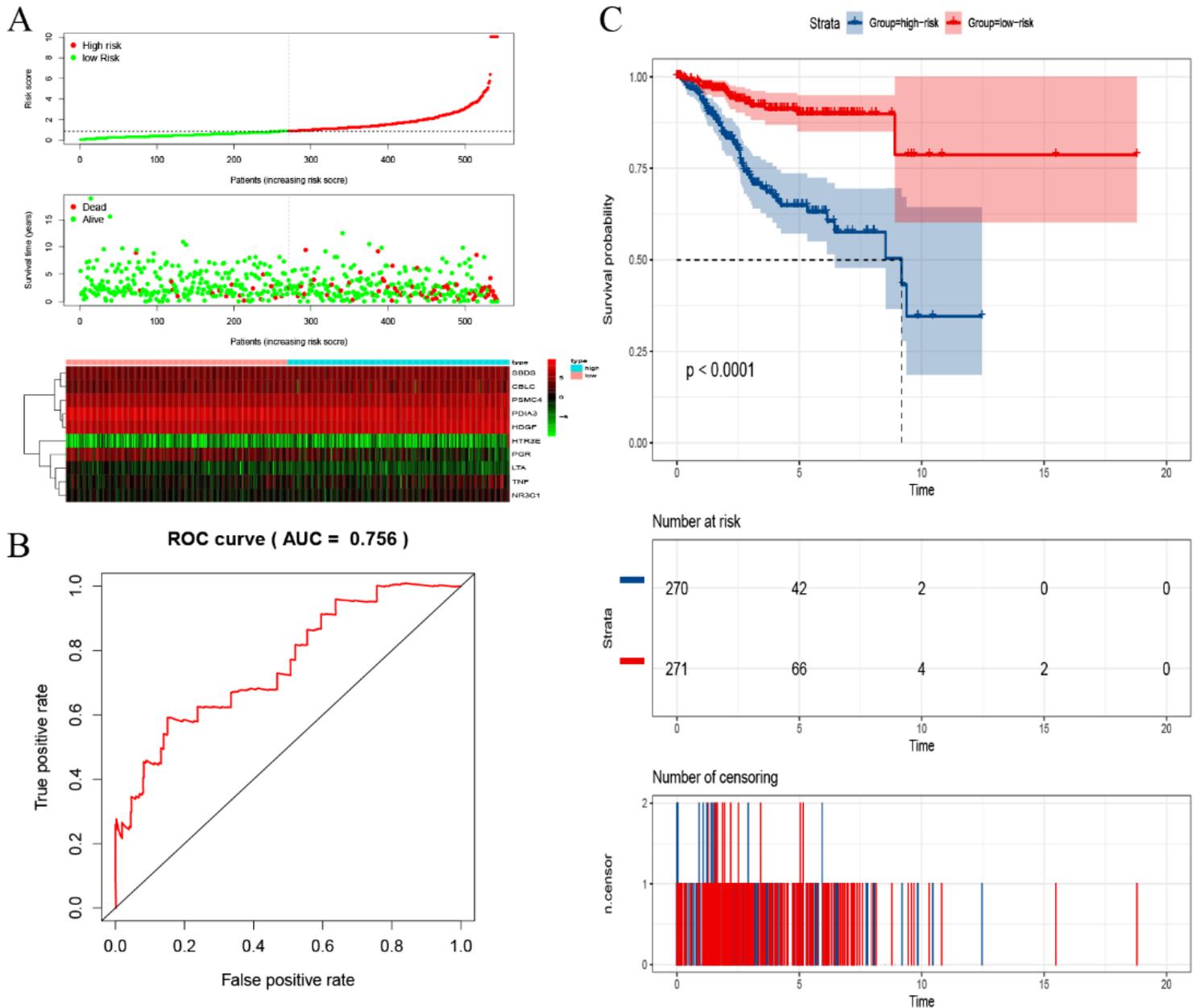


Figure 5

Correlation between the ten-gene prognostic signature and the OS of patients in the training set. (A) The distribution of risk scores, gene expression levels and patient survival status. The black dotted line represents the median cut point and divides patients into low-risk and high-risk groups. (B) ROC curve for judging the accuracy of the signature (AUC= 0.756). (C) Kaplan–Meier curves of OS of high- and low-risk groups ($P \leq 0.0001$). Abbreviation: OS, overall survival. UCEC, Uterine corpus endometrial carcinoma.

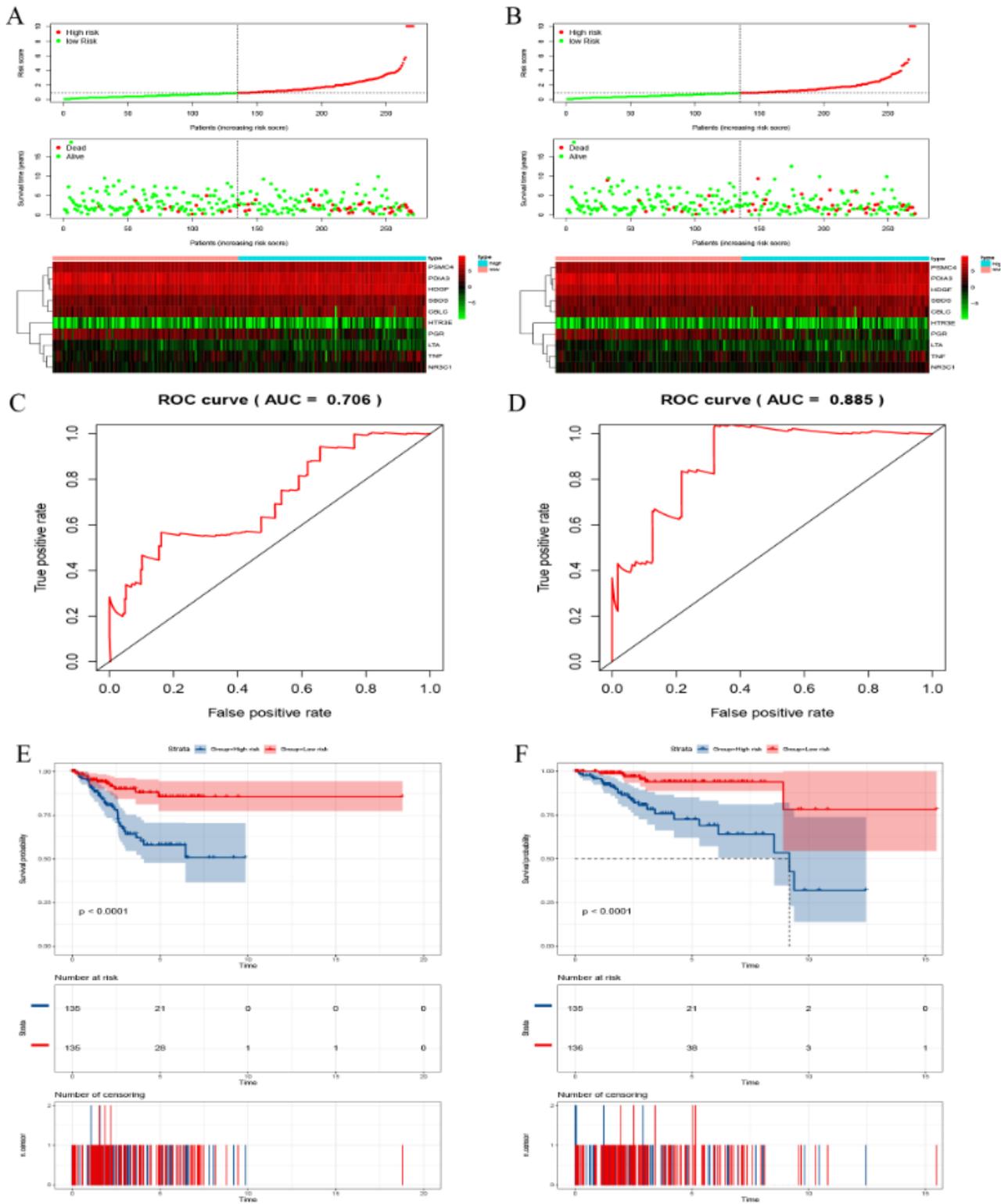


Figure 6

Validation of the ten-gene prognostic signature in the testing set A and testing set B. (A) The distribution of risk scores, gene expression levels and patient survival status in the testing set A. (B) The distribution of risk scores, gene expression levels and patient survival status in the testing set B. (C) ROC curve for judging the accuracy of the signature in the testing set A (AUC= 0.706). (D) ROC curve for judging the accuracy of the signature in the testing set B (AUC= 0.885). (E) Kaplan–Meier curves of OS of high- and

low-risk groups in the testing set A ($P=0.0001$). (F) Kaplan–Meier curves of OS of high- and low-risk groups in the testing set B ($P=0.0001$). Abbreviation: OS, overall survival.

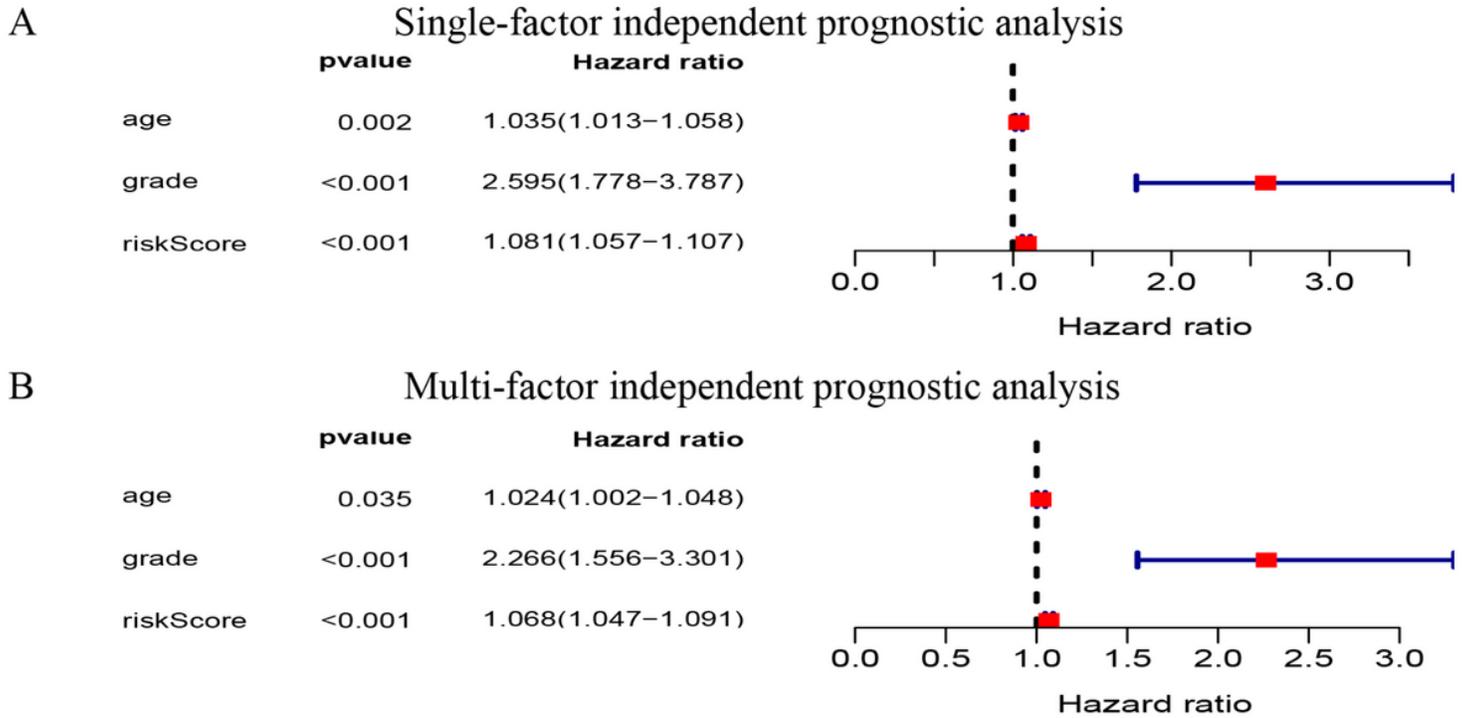


Figure 7

Univariate and multivariate Cox regression analysis of the ten-gene prognostic signature of UCEC patients in TCGA Cohort. (A) Age ($P = 0.002$), grade ($P < 0.001$), and the risk score derived from the signature ($P < 0.001$) were all significantly related to the prognosis by the Univariate Cox regression analysis. (B) Age ($P = 0.035$), grade ($P < 0.001$), and the risk score derived from the signature ($P < 0.001$) were all significantly related to the prognosis by the Multivariate Cox regression analysis. Abbreviation: UCEC, Uterine corpus endometrial carcinoma

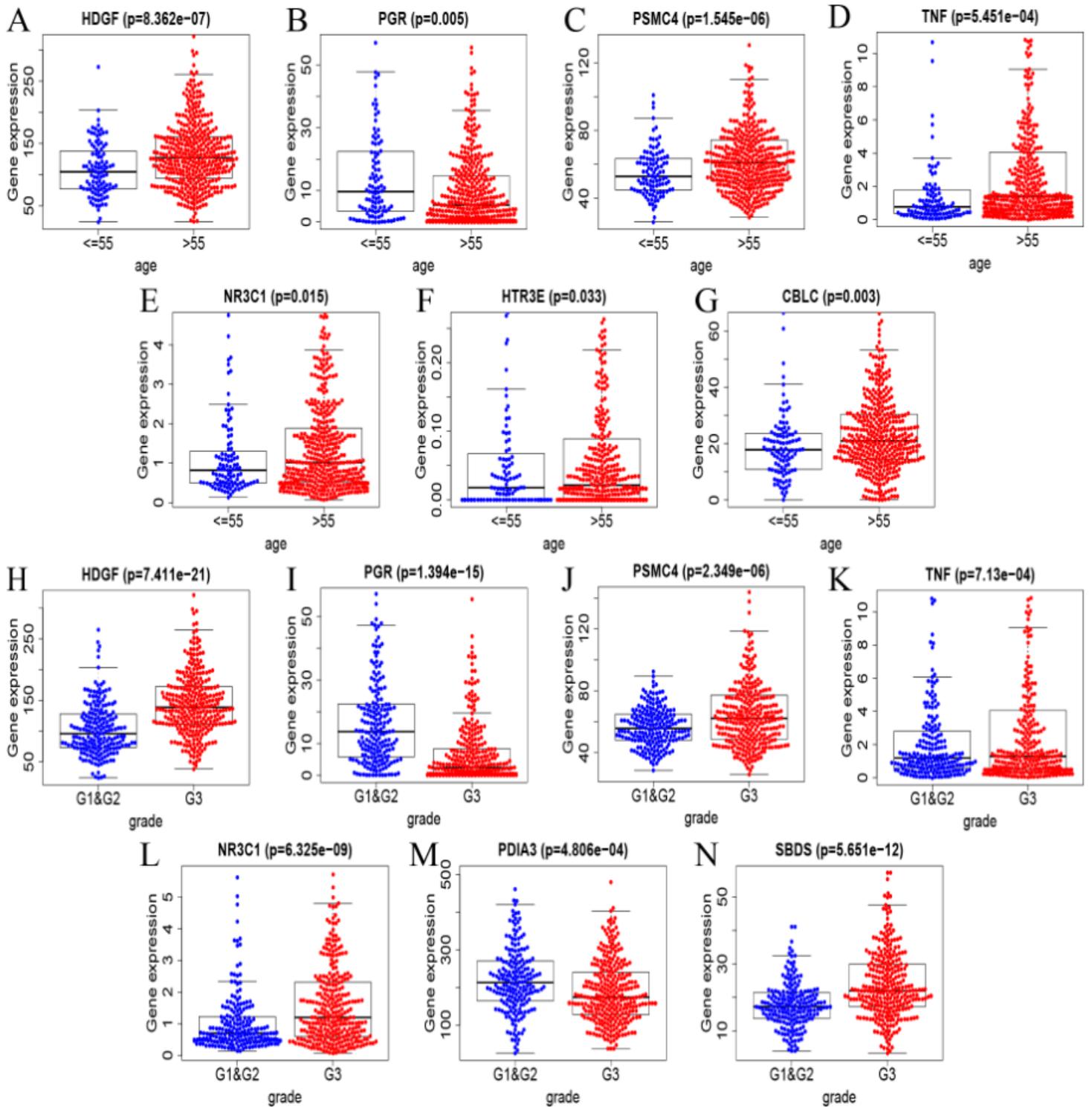


Figure 8

Gene expression levels of (A) HDGF, (B) PGR, (C) PSMC4, (D) TNF, (E) NR3C1, (F) HTR3E, and (G) CBLC between different age of UCEC. The expression of HDGF($P=5.054e-06$), PGR($P=0.04$), PSMC4($P=3.257e-04$), TNF($P=5.94e-06$), NR3C1($P=0.015$), HTR3E($P=0.033$) and CBLC($P=0.003$) was significantly correlated with age, while the expression of HDGF, PSMC4, TNF, NR3C1, HTR3E and CBLC increased with age. Gene expression levels of (H) HDGF, (I) PGR, (J) PSMC4, (K) TNF, (L) NR3C1, (M) PDIA3, and (N) SBDS between different clinical grade of UCEC. HDGF($P=7.411e-21$), PGR($P=1.394e-15$),

PSMC4(P=2.349e-06), TNF(P=7.13e-04), NR3C1(P=6.325e-09), PDIA3(P=4.806e-04), and SBDS(P=5.651e-12) was significantly associated with grade. Moreover, an increase in grade resulted in upregulation of HDGF, PSMC4, TNF, NR3C1, and SBDS. Abbreviation: UCEC, Uterine corpus endometrial carcinoma

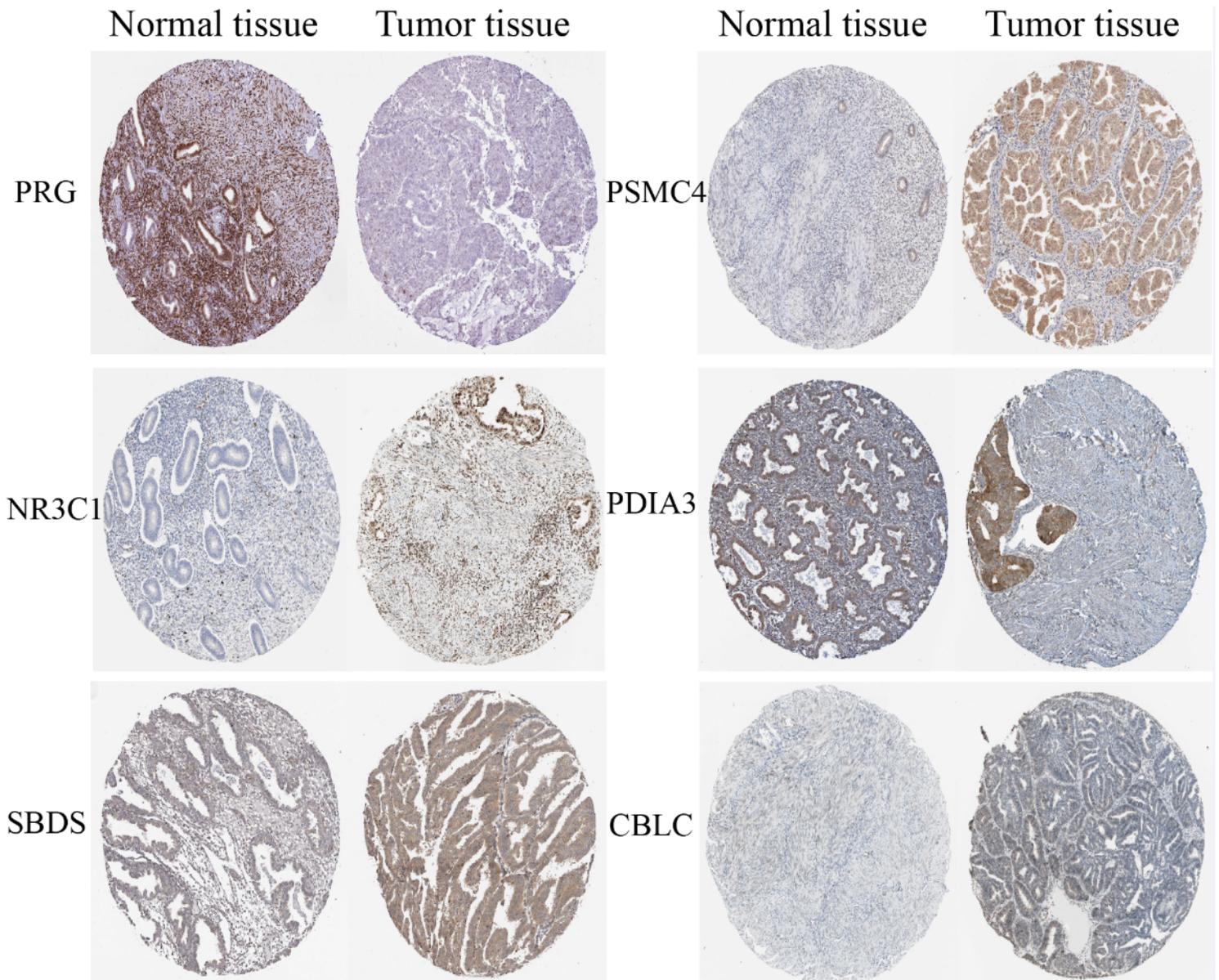


Figure 9

The level of ten genes in endometrial cancer patients in protein level (The Human Protein Atlas). Immunohistochemical examination for expression of PSMC4, NR3C1, SBDS, and CBLC were significant up-regulated in endometrial cancer tissue compared with normal tissues, while PGR and PDIA3 expression were significantly down-regulated in endometrial cancer tissues compared to normal tissues.

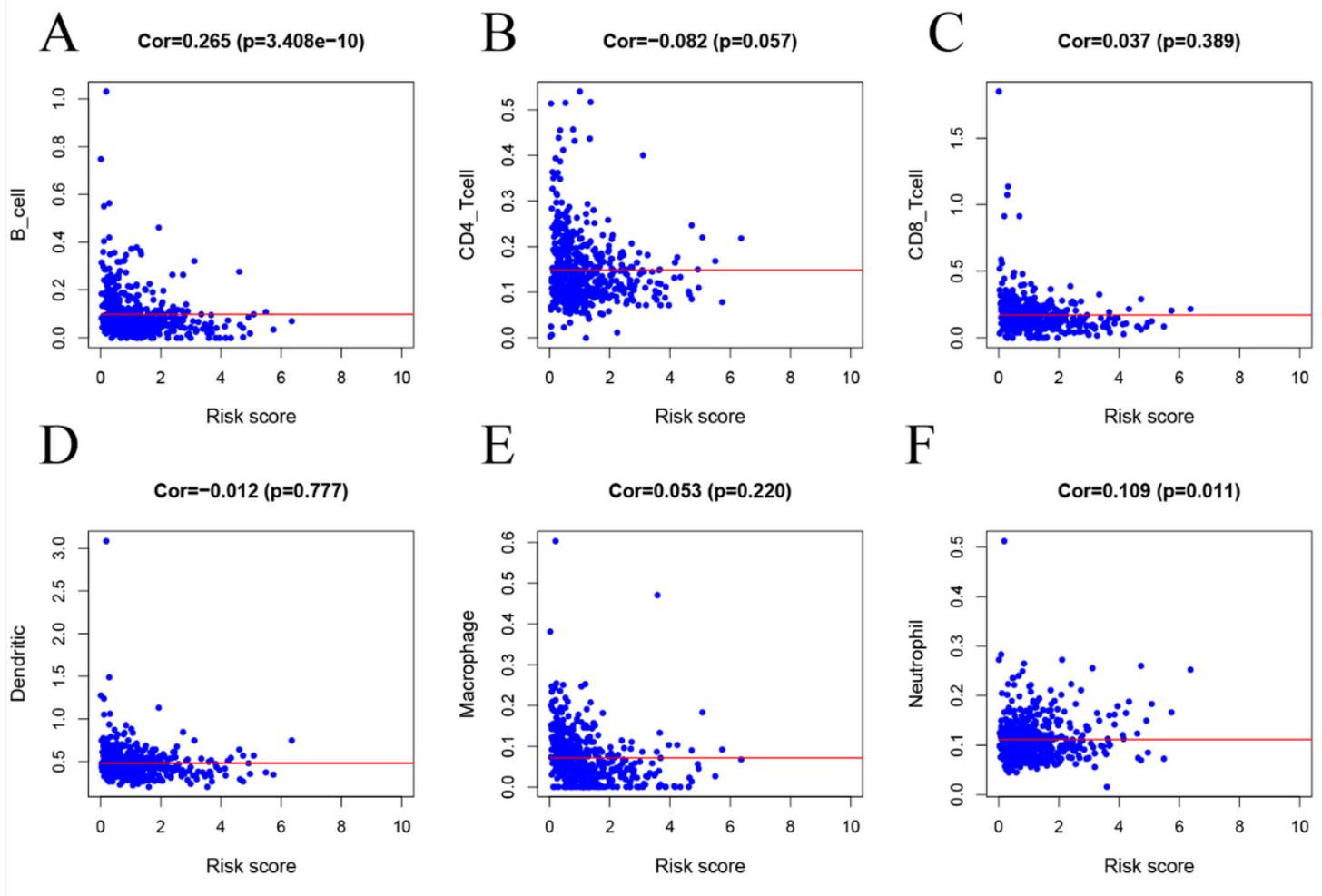


Figure 10

Association between the risk score of the ten-gene prognostic signature and the abundance of 6 immune infiltrates, where (A) B cells and (F) Neutrophils was significantly correlated with the patient's risk score, and were positively correlated.

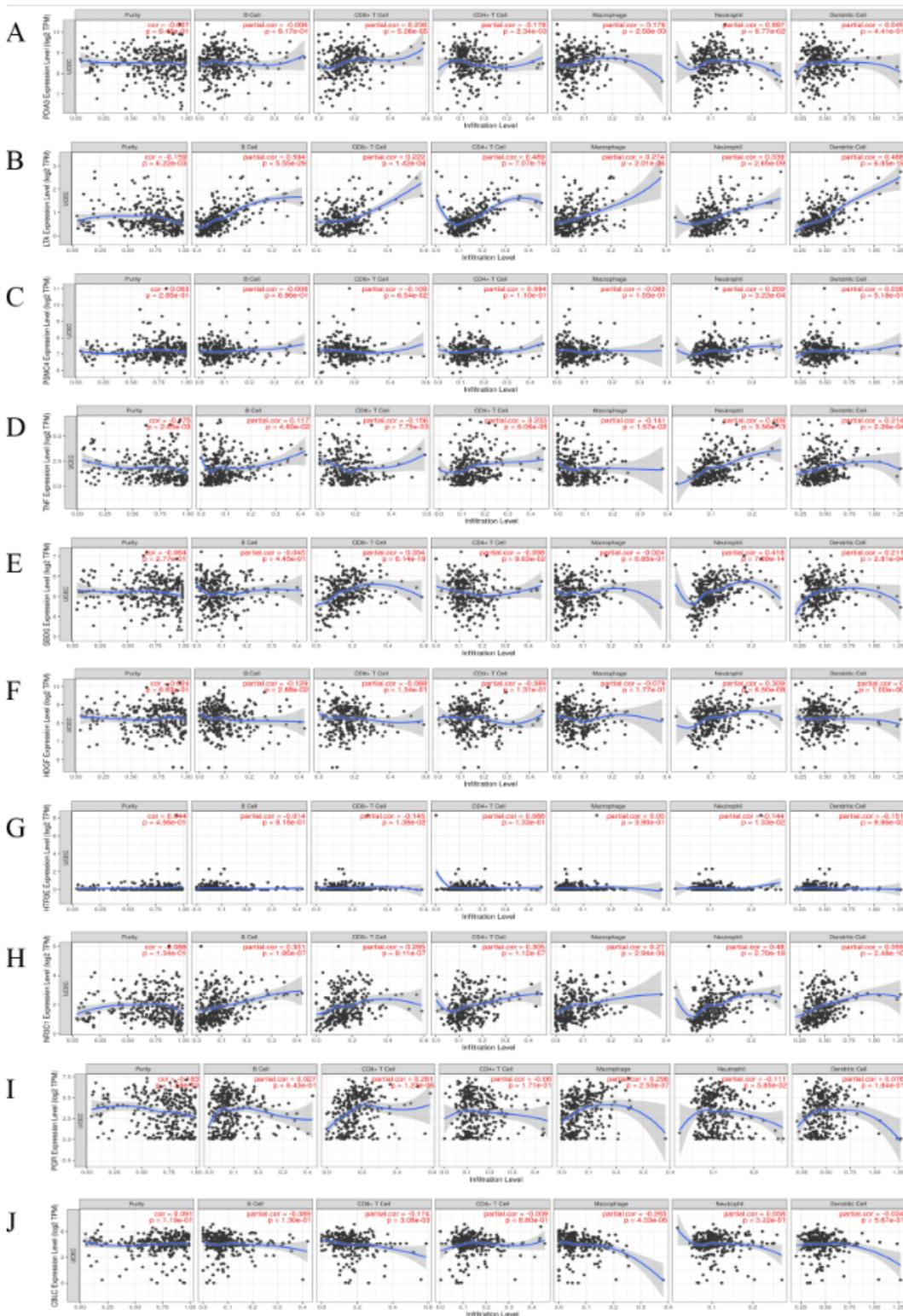


Figure 11

The correlation between (A)PDIA3, (B)LTA, (C)PSMC4, (D)TNF, (E) SBDS, (F)HDGF, (G)HTR3E, (H)NR3C1, (I)PGR, (J)CBLC and the immune infiltration level in UCEC. Abbreviation: UCEC, Uterine corpus endometrial carcinoma

Supplementary Files

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

- [File1immuneGeneExp.txt](#)
- [File2riskscoretrainingset.txt](#)
- [File3riskscoretestingsetA.txt](#)
- [File4riskscoretestingsetB.txt](#)