Mining TCGA Data for Key Biomarkers Related to Immune Microenvironment in Endometrial Carcinoma by Immune Score and Weighted Correlation Network Analysis

Chengbin Guo

Guangzhou Women and Children's Medical Center, Guangzhou Medical University

Yuqin Tang

School of Basic Medical Sciences, Chengdu University of Traditional Chinese Medicine

Yongqiang Zhang

Sichuan University West China Hospital <https://orcid.org/0000-0002-1033-1132>

Gen Li (✉ superleegen@hotmail.com)

Guangzhou Women and Children's Medical Center, Guangzhou Medical University

Research

Keywords: Endometrial cancer, immune microenvironment, biomarker, ESTIMATE, CIBERSORT, WCGNA

Posted Date: December 16th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-126793/v1>

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

[Read Full License](#)

**Mining TCGA data for key biomarkers related to immune
microenvironment in endometrial carcinoma by immune score
and weighted correlation network analysis**

Chengbin Guo^{1†}, Yuqin Tang², Yongqiang Zhang^{3,4†*}, Gen Li^{1*}

[†]These two authors contribute equally to this article

¹Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, 510623, China

²School of Basic Medical Sciences, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China

³Molecular Medicine Center, West China Hospital, Sichuan University, Chengdu 610041, China

⁴West China School of Medicine, West China Hospital, Sichuan University, Chengdu 610041, China

Correspondence to: Gen Li, E-mail: superleegen@hotmail.com; Yongqiang Zhang, E-mail: zyq1014zyq1014@163.com.

Abstract

Background: Endometrial cancer (EC) is one of the most lethal gynecological cancer in women. It is imperative to identify the potential immune microenvironment-related biomarkers associated with the prognosis for EC.

Methods: RNA-seq data and related clinical information of EC patients were derived from The Cancer Genome Atlas (TCGA). The immune score of each EC sample was obtained by ESTIMATE algorithm. Weighted gene co-expression network analysis (WCGNA) was used to identify the interesting module and potential key genes concerning the immune score. Further, the expression patterns of the key genes were verified via the GEPIA database. Last, CIBERSORT was used to evaluate the relative abundances of 22 immune cell types in EC.

Results: Immune scores were significantly associated with tumor grade and histology of EC, and high immune scores may exert a protective influence on the survival outcome for EC. WCGNA indicated that the black module was significantly correlated with the immune score in EC. Function analysis revealed it mainly involved in those terms related to immune regulation and inflammatory response. Moreover, 11 key genes were identified from the black module, validated by the GEPIA database, and revealed strong correlations with infiltration levels of multiple immune cell types, as was the prognosis of EC.

Conclusion: In our study, 11 key genes showed abnormal expressions and strong correlations with immune cell infiltration in EC, most of which were significantly associated with the prognosis of EC. These findings made them promising therapeutic targets for the treatment of EC.

Keywords: Endometrial cancer, immune microenvironment, biomarker, ESTIMATE, CIBERSORT, WCGNA

Background

Endometrial cancer (EC) is one of the most common gynecologic malignancies and represents the leading cause of morbidity and mortality among women worldwide [1]. Importantly, the incidence of EC is rising in the United States and more than 20 other countries [2]. In 2020, it occurred 65,620 new cases and caused 12,590 deaths in the United States [3]. The 5-year overall survival rate for early stage is about 81% while that for advanced stages (IVA and IVB) is approximately 15% [4].

EC is generally classified into various histological subtypes, including Endometrioid endometrial adenocarcinoma (EEC), Serous endometrial adenocarcinoma (ESC), Mixed serous and endometrioid (MSE), clear cell, and malignant mixed Mullerian tumors (MMMT) [5, 6]. Endometrioid is the most common histology, representing about 75% of endometrial cancers, followed by serous (1–5%) and clear cell (1–5%)[7]. Though the endometrioid subtype can be high or low grade, the other histologies, especially ESC and clear cell, are generally high in grade with a worse prognosis[8]. EEC is characteristically driven with estrogen receptors(ER) and progesterone receptors(PR) and thanks to its early symptoms like abnormal uterine bleeding, EEC is usually diagnosed early [9].

Remarkable progress has been made in ovarian cancer, and in cervical cancer, thanks to the newly developing vaccination against Human Papilloma Virus (HPV). Though treatment for EC is primarily surgical with hysterectomy and bilateral salpingo-oophorectomy as the standard of care worldwide, with poor sensitivity to conventional chemotherapy, EC is still a huge threat to women's lives. Immunotherapy, based on the concept of stimulating the endogenous immune response against tumor cells, has become a dependable clinical strategy in cancer treatment. For example, the agents blocking PD1/PD-L1 have exhibited impressive effects on lung, renal cancer, and melanoma[10]. It has been found that endometrial tumor cells can activate PD-1 signaling and immunohistochemical studies reported that PD1/PD-L1 expression levels in EC (40%-80% in EEC, 10%-68% in ESC, and 23%-69% in clear cell subtypes) represent the highest expression among gynecologic cancers [11, 12]. In this study, we used several algorithms including ESTIMATE and CIBERSORT to assess the immune scores and immune infiltration in EC and identified 11 potential immune therapeutic targets involved in the regulation of the immune microenvironment of EC, providing candidate prognostic biomarkers for EC. The workflow for this study was shown in Fig. 1.

Materials and Methods

Data acquisition, immune score generation, and clinical relationship

We retrospectively collected the gene expression profilings of 545 endometrial adenocarcinoma and 35 normal tissue samples from the TCGA database (<https://portal.gdc.cancer.gov/>). The corresponding clinicopathological parameters including age, height, weight, BMI, histology, TNM stage, tumor grade, tumor burden,

and survival data were also obtained. For data preprocessing, gene names were transformed to official gene symbols with Perl language, and the only genes with non-zero expression values in at least half of the sample type were kept. The immune score of each tumor sample was calculated with the ESTIMATE algorithm using the estimate package[13] based on R language software (version 3.6.0). Afterward, the immune scores were compared between different subgroups according to clinicopathological parameters with the Wilcox test. To evaluate the prognostic association, Kaplan-Meier plots for overall survival (OS) or disease-free survival (DFS) in high- or low- immune score groups were depicted based on the optimized immune score value of each patient, with a log-rank test for statistical significance. Besides, given that the immune microenvironment might be correlated with tumor stemness[14], we also obtained the available mRNA expression-based stem index (mRNAsi) of 528 EC patients as previously reported[15], following with the exploration of the spearman correlation between immune score and mRNAsi.

Screening the differentially expressed genes (DEGs) in EC

DEGs analysis between EC and normal tissues was performed by using the “limma” package[16], with the criteria of $\text{adj.P.Val} < 0.01$ and $|\log\text{FC}| > 1$. Because the sample size of the tumor group was much larger than that of the normal group (545 vs 35), we adopted the subset-based strategy to balance the samples. Specifically, we randomly generate a subset of 50 EC tumor samples from the EC group five times without repetition, yielding a ratio of about 1.4:1 for tumor and normal samples. DEGs were screened by comparing the expression profile of each tumor subset and that of the normal tissue group. Venn diagrams were plotted to get the common DEGs by the five independent subset-based analyses.

Weighted gene co-expression network analysis

WGCNA is a systematic algorithm to cluster highly correlated genes and to identify significant modules or key genes that are associated with a certain phenotype. In the current study, we utilized the WGCNA package[17] to construct a gene co-expression network of common DEGs. In brief, sample clustering was conducted with the average linkage method to recognize and remove outlier samples, followed by the selection of the appropriate soft thresholding power (β) to achieve a scale-free topology fitting index of > 0.9 . Then the adjacency was transformed into a topological overlap matrix (TOM) and the corresponding dissimilarity matrix (1-TOM), which

was further used to implement the gene clustering dendrogram with the minimum module of 30. Highly similar dynamic modules were merged into larger ones at the cutline of 0.2. Pearson correlation analysis was carried out to evaluate the relationships between modules and the immune score. The most significant module was identified and the gene significance (GS) and module membership (MM) were calculated. Key genes were defined as those with the $GS > 0.7$ and $MM > 0.7$ in this module.

PPI Network Construction

A protein-protein interaction network of the identified module was constructed by STRING database (<https://www.string-db.org>) version 11.0 using the median confidence (combined score > 0.4) and visualized by Cytoscape software (version 3.2.1). The network topology including node degree was investigated by the cytohubba application.

Function enrichment analysis

To explore the involved biological functions and pathways of the significant module, we conducted Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis with the “clusterProfiler” R package. The significant terms were defined as those with a p.adjust value of < 0.05 .

Key genes validation

GEPIA is a widely-used web tool for data mining and visualization of the RNA sequencing expression data from the TCGA and the GTEx projects[18]. In this study, we applied GEPIA to validate the differential expression of the key genes identified by WGCNA. The default parameters ($|\log_2FC|$ Cutoff of 1, and P -value Cutoff of 0.01, and $\log_2(TPM + 1)$ for log-scale) were employed. Match TCGA normal and GTEx data were combined as the normal group ($n = 91$). Besides, the associations of the key genes and tumor grade or tumor histology were further explored, and the correlation matrix of the key genes was generated.

Survival analysis of the key genes

In order to evaluate the prognostic values of the key genes, we carried out the Kaplan-Meier survival analysis for OS ($n = 543$) and DFS ($n = 541$) with the aid of the EC samples from TCGA. For each key gene, patients were assigned to high- or low-expression groups according to the optimized immune score value of each patient.

Statistical significance was measured by the log-rank test.

Estimation of the immune cell landscape

CIBERSORT, an analytical tool providing an estimation of the abundances of member cell types via gene expression data[19], was introduced to evaluate the tumor immune infiltration levels of EC. The relative proportion of 22 tumor immune cell types including B cells naïve, Plasma cells, B cells memory, T cells CD8, T cells CD4 naïve, T cells CD4 memory activated, T cells CD4 memory resting, T cells regulatory (Tregs), T cells follicular helper, T cells gamma delta, Monocytes, NK cells activated, NK cells resting, Macrophages M0, Macrophages M1, Macrophages M2, Dendritic cells activated, Dendritic cells resting, Mast cells activated, Mast cells resting, Eosinophils and Neutrophils were computed. Moreover, the relationship between key genes and each of the immune cell types was investigated.

Results

Correlations between immune score and clinical characteristics in EC

After assessing the immune score from the tumors of EC patients, the high immune score group of EC patients showed a significantly better overall survival (OS) or disease-free survival (DFS) than that of the low immune score group ($P = 0.0061$ for OS and $P = 0.012$ for DFS, respectively) (Fig. 2A and B). To investigate the relevance to clinical features, 545 EC patients were classified by grade, Histology, stage, age, height, weight, BMI, and tumor burden. Consequently, among these clinical characteristics, no significant associations were observed between immune score and tumor stage ($P = 0.2$), age ($P = 0.24$), height ($P = 0.29$), weight ($P = 0.17$), BMI ($P = 0.11$) and tumor burden ($P = 0.1$) (Fig.2C-H). However, the immune score was significantly associated with tumor grade ($P = 0.041$) and Histology ($P = 0.034$) (Fig.2I and J). Specifically, for tumor grade, grade G3 (poorly differentiated) had a significantly lower immune score than G1 and G2. In terms of tumor histology, it seemed the ESC subtype has the lowest immune score. Besides, according to the Pearson correlation analysis, a significantly negative correlation may exist between the mRNA stemness index (mRNAsi) and the immune microenvironment in EC patients (Additional file 1: Fig. S1).

Screening of DEGs in EC

The analysis of the correlation between immune score and clinical characteristics suggested that there might be DEGs involved in forming the different immune microenvironments between EC tissues and normal tissues. Therefore, we screened the DEGs with the subset-based approach using the TCGA-UCEC dataset. First, we randomly subsampled 50 EC samples from the tumor group five times. Second, we obtained the DEGs between EC samples from each subset and the 35 normal samples, respectively (Fig. 3A-E). As the result, in subset 1, 2421 DEGs were discerned, among which 980 were up-regulated and 1441 were down-regulated; In subset 2, 2545 DEGs were discerned, among which 1105 were up-regulated and 1440 were down-regulated; In subset 3, 2516 DEGs were discerned, among which 1110 were up-regulated and 1406 were down-regulated; In subset 4, 2563 DEGs were discerned, among which 1140 were up-regulated and 1423 were down-regulated; In subset 5, 2454 DEGs were discerned, among which 1094 were up-regulated and 1359 were down-regulated. We intersected the DEGs from all 5 subsets using two Venn diagrams, including a total of 758 common up-regulated genes and 1179 common down-regulated in EC (Fig. 3F and G).

Co-expression network of DEGs in EC

WGCNA (weighted gene co-expression network analysis) was used to construct a co-expression network of DEGs we found in EC. Seven outlier samples were removed prior to network construction. The optimal soft-thresholding power of 5 (scale-free $R^2 = 0.96$) was picked to ensure the scale-free topology (Figure 4A). A total of seven modules were screened out after merging dynamic modules with the Diss Thres of 0.2 (Figure 4B). Then, we focused on the most correlated module with the immune score in EC by computing the Pearson correlation coefficient (PCC) and corresponding P -value. As Figure 4C indicated, the black module was found to be significantly positive with immune score ($PCC = 0.82$, $P = 2E-134$), including 71 DEGs in EC. With the thresholds of $GS > 0.7$ and $MM > 0.7$ to further narrow down the range of candidate key genes, 11 DEGs were finally identified for the subsequent analysis (Figure 4D).

The PPI network of the black module

A PPI network was built to analyze the black module explored above, containing 42 DEGs in EC (Fig. 5A). Results from the PPI network demonstrated that there were 8

up-regulated genes and 34 down-regulated genes in the black module. Interestingly, several members of the GTPase IMAP family, i.e., GIMAP1, GIMAP4, GIMAP6, GIMAP7, and GIMAP8 formed a close submodule. For network topology, the top 12 hub nodes in the entire PPI network with their degrees were visualized in Fig. 5B, the top 11 of which were closely connected and used to construct a sub-network (Fig. 5C). The most important protein is CXCL10 (8 edges) in EC.

The cellular functions and pathway analysis of the black module

To find out the cellular functions and pathways the genes of the black module involved in, GO and KEGG enrichment analysis were performed. GO analysis results indicated that the key genes of the black module mostly participated in the biological process (BP) of regulation of inflammatory response, and the main related molecular function (MF) terms were GTP binding, purine ribonucleoside binding, and purine nucleoside binding, but no cellular component (CC) was enriched (Fig. 6A). KEGG analysis demonstrated that the most significantly enriched pathway was complement and coagulation cascades (Fig. 6B).

Validation of key genes in the GEPIA database

Next, we verified the filtered key genes in the GEPIA database. The expression values of the 11 key genes were shown in Fig. 7, indicating a significantly lower expression level for each key gene in EC than normal tissue samples. In addition, we also showed the comparison of the expression levels of the 11 key genes in EC and normal tissues from 5 subsets of the TCGA-UCEC with boxplots (Additional file 1: Fig. S2), and as we predicted, these genes were significantly lower-expressed in EC.

Correlation analysis of the key genes

In consideration of immune score was associated with the tumor grade and histology in EC, we explored the expression levels of the 11 key genes in different grades and different histological subtypes of EC. GIMAP1, GYPC, and IFFO1 were found to significantly decline with the improved grade of EC (Fig. 8A). Only GYPC was found to be correlated with different histological subtypes of EC (Fig. 8B). Moreover, Pearson correlation analysis coupled with statistical significance were applied to evaluate the association among the key genes and revealed that these genes had strong correlations, of which the coefficient values were shown in matrices (Fig. 8C). As the matrices shown, the minimum correlation coefficient among these genes was 0.52

while the maximum was 0.93.

The key genes associate with DFS

Motivated by the markedly different expression levels between EC tissues and normal tissues, we examined the 11 key genes in perspective of DFS using Kaplan–Meier analysis and log-rank tests. We found that low expression of GIMAP1 ($P = 0.0044$), GIMAP4 ($P = 0.0001$), GIMAP6 ($P = 0.02$), GIMAP7 ($P = 0.00081$), GIMAP8 ($P = 0.0011$), GYPC ($P = 0.0011$) and IFFO1 ($P = 0.0079$) were significantly associated with worse prognosis (Fig. 9).

The key genes associate with OS

Next, we examined the 11 key genes in perspective of OS using Kaplan–Meier analysis and log-rank tests. As shown in Fig. 10, CLEC2B ($P = 0.023$), GIMAP1 ($P = 0.0072$), GIMAP4 ($P = 0.0026$), GIMAP7 ($P = 0.023$), GIMAP8 ($P = 0.0055$), GYPC ($P = 0.0077$) and IFFO1 ($P = 0.014$) were significantly associated with OS of EC patients and turned out to low expression of them with worse prognosis (Fig. 10).

The key genes were significantly associated with infiltration of immune cells in EC microenvironment

We estimated the proportions of 22 distinct immune cell types in EC samples using the CIBERSORT algorithm (Fig. 11A). Next, we explored the correlation between each immune cell type proportion and the expression levels of the 11 key genes, and the results indicated that all of these 11 key genes, respectively, were significantly associated with at least 8 kinds of immune cells (Fig. 11B). All of the 11 key genes were positively correlated to the infiltration of activated Dendritic cells, M0 macrophages, and activated mast cells and were negatively correlated to the infiltration of M1 macrophages, plasma cells, activated CD4 memory T cells, and CD8 T cells in EC samples.

Discussion

In the United States, it was reported that 65,620 new cases and 12,590 deaths of EC occurred in 2020. The mortality has increased by approximately 1.4% per year from 2005 to 2014[3, 20]. The prognosis of EC is largely based on histologic grade and clinical stage[20]. Although some reports have pointed out the presence of immune dysregulation in EC and considered immune checkpoint blockade therapy as a

potential treatment for EC patients[21], the mechanism of the dysregulation of the immune microenvironment in EC was not known deeply yet. In this study, surprisingly, we estimated the immune score for more than 500 EC samples and found that the immune score was significantly correlated with the grade and histology of EC, respectively. With this meaningful discovery, we aimed at figuring out the key genes playing pivotal roles in the constitution of the immune microenvironment in EC.

Through subset-based analysis using the TCGA-UCEC dataset, we identified 758 upregulated genes and 1179 downregulated genes in EC samples compared with normal samples. Further, we filtered the most related genes to immune regulation among the candidate genes above by WCGNA and we obtained a black module positively correlated with the immune score, which contained 71 differential immune-associated genes. Further functional analysis revealed that these 71 genes were associated with the regulation of inflammatory response. Interestingly, 5 members of GTPase of immunity-associated protein (GIMAP) family (also known as immune-associated nucleotide-binding protein (IAN)) genes were included among the key genes. Numerous studies have reported that the function of the GIMAP family in regulating T cell development, selection, and homeostasis[22-24]. Therefore, it suggested that the 5 GIMAP family genes we found might play an important part in regulating the immune microenvironment in EC. In addition, two chemokines, CXCL10/IP-10 and CCL18 were discovered. CXCL10 recruited NK cells to tumor and activated NK cells to kill tumor cells[25] and it has been reported strongly producing in tumor compared to the adjacent tissue in EC[26]. CCL18, mainly secreted by tumor-associated macrophages (TAMs) in tumors, was positively correlated with malignancy in EC[27, 28]. Our results above verified that CXCL10 and CCL18 were two important factors in constituting the immune microenvironment of EC.

Through WCGNA analysis, we chose 11 key genes in the black module that were most correlated with both module and immune for further investigation. In this study, the expressions of these key genes were significantly decreased in EC samples and verified by GEPIA database. Besides the 5 GIMAP family genes, TAGAP (T cell activation Rho GTPase activating protein) is also a GTPase related gene and an indicator of lymphocyte activation[29, 30]. These abnormally expressed genes implied the immune disorders in the endometrial tumor environment.

Considering the correlation between immune score and clinical characteristics (grade

and histology), we further tested the correlations between the key genes and the tumor grade and histology in EC, individually. The expression levels of GIMAP1 and GYPC were significantly correlated to the tumor grade and only GYPC was significantly correlated to the tumor histology in EC. In the perspective of the grade of EC, GIMAP1 showed the lowest expression in high grade while GYPC showed the lowest expression in G3. Meanwhile, GYPC showed the lowest expression in MSE in the perspective of the histology of EC. Several studies reported that GIMAP1 is critical for the development of mature lymphocytes[31-33]. We observed that higher tumor grade of EC was accompanied by lower GIMAP1 expression and it seemed that higher grade of EC might have less functional lymphocytes infiltration in EC tumor microenvironment, and then higher tumor grade of EC came with a worse prognosis. Furthermore, the survival analysis verified that low expression of GIMAP1 led to a worse prognosis in EC. At last, all of the key genes were significantly and positively correlated with each other.

Given that the key genes were downregulated in EC samples, we explored whether their expression levels were correlated to DFS and OS. As the results showed, 8 key genes (CLEC2B, GIMAP1, GIMAP4, GIMAP6, GIMAP7, GIMAP8, GYPC, and IFFO1) were positively correlated to a favorable prognosis. Interestingly, 5 members (GIMAP1, GIMAP4, GIMAP6, GIMAP7, GIMAP8) and 4 members (GIMAP1, GIMAP4, GIMAP7, and GIMAP8) of the GIMAP family genes were positively correlated with the DFS and OS, respectively in EC. Combined with the results above, GIMAP family genes hopefully had research value in EC and we will pay more attention to them in the future study. As expected, GYPC was also related to the DFS and OS, reflecting the importance of GYPC in EC. Besides, few Studies have been conducted in terms of IFFO1, the nucleoskeleton protein, which is recruited to the sites of DNA damage and promotes the repair of DNA double-strand breaks[34]. CLEC2B/AICL was reported as an important gene in NK cells and stimulated NK cell effector function such as cytotoxicity and cytokine secretion[35].

Previous studies have reported that the immune microenvironment is associated with the prognosis in EC[36-38]. Therefore, we used CIBERSORT to estimate the infiltration of 22 immune cell types in EC samples. Preliminary results showed that macrophages and T cells might be the predominant immune cell types infiltrated in EC tumor microenvironment. Moreover, the key genes we found were strongly

associated with multiple immune cell types infiltrated in EC tumor. The positive or negative correlation between the genes and immune cell types suggested lower or higher infiltration of those immune cell types in EC samples compared with normal samples. Obviously, there might be more M1 macrophages and CD8 T cells and less M0 macrophages and activated dendritic cells infiltrated in EC samples than those in normal samples.

However, limitations that should be mentioned: First of all, although we adopted the subset-based approach to balance the samples from the TCGA database, due to the small sample size of normal controls, more verifications are still required to enhance the robustness of the screened DEGs. Secondly, future biological experiments of in vitro and in vivo studies are needed to test and explore the effects of the key genes we found in EC.

Conclusion

Taken together, in this study we discovered 11 key genes, which were closely correlated with each other, abnormally expressed and associated with immune scores in EC. They might play an important role in EC immune microenvironment and function as potential biomarkers.

List of abbreviations

EC: Endometrial cancer

TCGA: The Cancer Genome Atlas

WCGNA: Weighted gene co-expression network analysis

EEC: Endometrioid endometrial adenocarcinoma

ESC: Serous endometrial adenocarcinoma

MSE: Mixed serous and endometrioid

ER: Estrogen receptors

PR: progesterone receptors

HPV: Human Papilloma Virus

OS: Overall survival

DFS: Disease free survival

DEGs: Differentially expressed genes

PPI: Protein-protein interaction

BP: Biological process

CC: Cellular component

MF: Molecular function

TAMs: Tumor-associated macrophages

GIMAP: GTPase of immunity-associated protein

IAN: immune-associated nucleotide-binding protein

TAGAP: T cell activation Rho GTPase activating protein

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The data achieved and analyzed in the current study are available in the TCGA repository, <https://portal.gdc.cancer.gov/>.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

G.L and Y.Z, designed the research; Y.T collected the data. C.G and Y.Z. analyzed the data and wrote the draft. G.L. revised the language of the paper. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References:

1. Ventriglia J, Paciolla I, Pisano C, Cecere SC, Di Napoli M, Tambaro R, Califano D, Losito S, Scognamiglio G, Setola SV, et al: Immunotherapy in ovarian, endometrial and cervical cancer: State of the art and future perspectives. *Cancer Treat Rev.* 2017; 59:109-116.
2. Lortet-Tieulent J, Ferlay J, Bray F, Jemal A: International Patterns and Trends in Endometrial Cancer Incidence, 1978-2013. *J Natl Cancer Inst.* 2018; 110:354-361.
3. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2020. *CA Cancer J Clin.* 2020; 70:7-30.
4. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2018. *CA Cancer J Clin.* 2018; 68:7-30.
5. Urick ME, Bell DW: Clinical actionability of molecular targets in endometrial cancer. *Nat Rev Cancer.* 2019; 19:510-521.
6. Gaber C, Meza R, Ruterbusch JJ, Cote ML: Endometrial Cancer Trends by Race and Histology in the USA: Projecting the Number of New Cases from 2015 to 2040. *J Racial Ethn Health Disparities.* 2016.
7. Murali R, Soslow RA, Weigelt B: Classification of endometrial carcinoma: more than two types. *Lancet Oncol.* 2014; 15:e268-278.
8. Sorosky JI: Endometrial cancer. *Obstet Gynecol.* 2012; 120:383-397.
9. Lax SF, Pizer ES, Ronnett BM, Kurman RJ: Comparison of estrogen and progesterone receptor, Ki-67, and p53 immunoreactivity in uterine endometrioid carcinoma and endometrioid carcinoma with squamous, mucinous, secretory, and ciliated cell differentiation. *Hum Pathol.* 1998; 29:924-931.
10. Terme M, Ullrich E, Aymeric L, Meinhardt K, Desbois M, Delahaye N, Viaud S, Ryffel B, Yagita H, Kaplanski G, et al: IL-18 induces PD-1-dependent immunosuppression in cancer. *Cancer Res.* 2011; 71:5393-5399.
11. Vanderstraeten A, Tuyaerts S, Amant F: The immune system in the normal endometrium and implications for endometrial cancer development. *J Reprod Immunol.* 2015; 109:7-16.
12. Brahmer JR, Tykodi SS, Chow LQ, Hwu WJ, Topalian SL, Hwu P, Drake CG, Camacho LH, Kauh J, Odunsi K, et al: Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N Engl J Med.* 2012; 366:2455-2465.
13. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA, et al: Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013; 4:2612.
14. Yi L, Huang P, Zou X, Guo L, Gu Y, Wen C, Wu G: Integrative stemness characteristics associated with prognosis and the immune microenvironment in esophageal cancer. *Pharmacol Res.* 2020; 161:105144.
15. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kaminska B, Huelsken J, Omberg L, Gevaert O, et al: Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell.* 2018; 173:338-354 e315.
16. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015; 43:e47.
17. Langfelder P, Horvath S: WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008; 9:559.
18. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z: GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 2017; 45:W98-W102.
19. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA: Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol Biol.* 2018; 1711:243-259.

20. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A: Global cancer statistics, 2012. *CA Cancer J Clin.* 2015; 65:87-108.
21. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, Skora AD, Luber BS, Azad NS, Laheru D, et al: PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med.* 2015; 372:2509-2520.
22. Schnell S, Demolliere C, van den Berk P, Jacobs H: Gimap4 accelerates T-cell death. *Blood.* 2006; 108:591-599.
23. Nitta T, Nasreen M, Seike T, Goji A, Ohigashi I, Miyazaki T, Ohta T, Kanno M, Takahama Y: IAN family critically regulates survival and development of T lymphocytes. *PLoS Biol.* 2006; 4:e103.
24. Krucken J, Epe M, Benten WP, Falkenroth N, Wunderlich F: Malaria-suppressible expression of the anti-apoptotic triple GTPase mGIMAP8. *J Cell Biochem.* 2005; 96:339-348.
25. Nagarsheth N, Wicha MS, Zou W: Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy. *Nat Rev Immunol.* 2017; 17:559-572.
26. Degos C, Heinemann M, Barrou J, Boucherit N, Lambaudie E, Savina A, Gorvel L, Olive D: Endometrial Tumor Microenvironment Alters Human NK Cell Recruitment, and Resident NK Cell Phenotype and Function. *Front Immunol.* 2019; 10:877.
27. Sakane R, Tsubamoto H, Sakata K, Inoue K, Ogino M, Shibahara H, Hao H, Hirota S: Expression of chemokine ligand 18 in stage IA low-grade endometrial cancer. *Anticancer Res.* 2014; 34:5331-5336.
28. Jing X, Peng J, Dou Y, Sun J, Ma C, Wang Q, Zhang L, Luo X, Kong B, Zhang Y, et al: Macrophage ERalpha promoted invasion of endometrial cancer cell by mTOR/KIF5B-mediated epithelial to mesenchymal transition. *Immunol Cell Biol.* 2019; 97:563-576.
29. Mao M, Biery MC, Kobayashi SV, Ward T, Schimmack G, Burchard J, Schelter JM, Dai H, He YD, Linsley PS: T lymphocyte activation gene identification by coregulated expression on DNA microarrays. *Genomics.* 2004; 83:989-999.
30. Arshad M, Bhatti A, John P, Jalil F, Borghese F, Kawalkowska JZ, Williams RO, Clanchy FIL: T cell activation Rho GTPase activating protein (TAGAP) is upregulated in clinical and experimental arthritis. *Cytokine.* 2018; 104:130-135.
31. Saunders A, Webb LM, Janas ML, Hutchings A, Pascall J, Carter C, Pugh N, Morgan G, Turner M, Butcher GW: Putative GTPase GIMAP1 is critical for the development of mature B and T lymphocytes. *Blood.* 2010; 115:3249-3257.
32. Webb LM, Datta P, Bell SE, Kitamura D, Turner M, Butcher GW: GIMAP1 Is Essential for the Survival of Naïve and Activated B Cells In Vivo. *J Immunol.* 2016; 196:207-216.
33. Datta P, Webb LM, Avdo I, Pascall J, Butcher GW: Survival of mature T cells in the periphery is intrinsically dependent on GIMAP1 in mice. *Eur J Immunol.* 2017; 47:84-93.
34. Li W, Bai X, Li J, Zhao Y, Liu J, Zhao H, Liu L, Ding M, Wang Q, Shi FY, et al: The nucleoskeleton protein IFFO1 immobilizes broken DNA and suppresses chromosome translocation during tumorigenesis. *Nat Cell Biol.* 2019; 21:1273-1285.
35. Neuss S, Bartel Y, Born C, Weil S, Koch J, Behrends C, Hoffmeister M, Steinle A: Cellular Mechanisms Controlling Surfacing of AICL Glycoproteins, Cognate Ligands of the Activating NK Receptor NKp80. *J Immunol.* 2018; 201:1275-1286.
36. Versluis MAC, Marchal S, Plat A, de Bock GH, van Hall T, de Bruyn M, Hollema H, Nijman HW: The prognostic benefit of tumour-infiltrating Natural Killer cells in endometrial cancer is dependent on concurrent overexpression of Human Leucocyte Antigen-E in the tumour microenvironment. *Eur J Cancer.* 2017; 86:285-295.

37. de Jong RA, Leffers N, Boezen HM, ten Hoor KA, van der Zee AG, Hollema H, Nijman HW: Presence of tumor-infiltrating lymphocytes is an independent prognostic factor in type I and II endometrial cancer. *Gynecol Oncol.* 2009; 114:105-110.
38. Ino K, Yamamoto E, Shibata K, Kajiyama H, Yoshida N, Terauchi M, Nawa A, Nagasaka T, Takikawa O, Kikkawa F: Inverse correlation between tumoral indoleamine 2,3-dioxygenase expression and tumor-infiltrating lymphocytes in endometrial cancer: its association with disease progression and survival. *Clin Cancer Res.* 2008; 14:2310-2317.

Figure legend

Figure 1 Workflow of this study

Figure 2 Correlation between immune score and clinical characteristics in endometrial carcinoma (EC). (A)-(B) Kaplan–Meier survival plots of immune score for OS (A) and DFS (B) for EC patients. (C)-(J) Associations of immune score and clinical characteristics including stage (C), age (D), height (E), weight (F), BMI (G), tumor burden (H), grade (I), and histology (J). OS, overall survival. DFS, disease free survival.

Figure 3 Screening of the DEGs in EC tissues compared with normal tissues with the subset-based approach using the TCGA-UCEC dataset. (A)-(E) Volcano plots showing the DEGs identified from five subsets of the TCGA-UCEC dataset. Red dots represent the upregulated genes, while blue dots represent the downregulated genes. (F)-(G) Venn diagrams showing the common upregulated genes (F) and common down-regulated genes (G) shared by the five subsets-based DEGs. DEGs, differentially expressed genes. EC, endometrial cancer.

Figure 4 Weighted gene co-expression network analysis of EC. (A) Selection of the optimal soft-thresholding power for the scale-free network. (B) DEGs dendrogram using the dissimilarity measure (1-TOM). (C) Correlation analysis of module eigengenes and immune score. Pearson's correlation coefficient and the corresponding *P*-value are shown. (D) Scatter plot of the black module showing the relationship between GS and MM. Key genes are indicated in the upper-right corner with the threshold of $GS > 0.7$ and $MM > 0.7$. TOM, topological overlap matrix. GS, gene significance. MM, module membership.

Figure 5 PPI network of the black module in endometrial cancer. (A) The PPI network consisting of 42 genes from the black module. Red nodes represent the upregulated genes, while the blue nodes represent the downregulated genes. (B) The

top 12 nodes with the highest degree in the network. (C) The significant cluster formed by the top 11 nodes with most neighbor genes. PPI, protein-protein interaction.

Figure 6 GO (A) and KEGG (B) enrichment analysis of the black module in endometrial cancer. BP, biological process. CC, cellular component. MF, molecular function.

Figure 7 Validation of the abnormal expressions of the key genes via the online database of GEPIA. * $P < 0.01$.

Figure 8 Correlation analysis of the key genes in endometrial cancer (EC). (A)-(B) Boxplots showing the correlations of the key genes with tumor grade and histology in EC. (C) Pearson correlation matrices between expression values of the key genes. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, and ns indicates no significance.

Figure 9 Kaplan–Meier survival curves of the 11 key genes characterizing DFS difference with log-rank tests in endometrial cancer. DFS, disease-free survival.

Figure 10 Kaplan–Meier survival curves of the 11 key genes characterizing OS difference with log-rank tests in endometrial cancer. OS, overall survival.

Figures

Figure 1

Workflow of this study

Figure 2

Correlation between immune score and clinical characteristics in endometrial carcinoma (EC). (A)-(B) Kaplan–Meier survival plots of immune score for OS (A) and DFS (B) for EC patients. (C)-(J) Associations of immune score and clinical characteristics including stage (C), age (D), height (E), weight (F), BMI (G), tumor burden (H), grade (I), and histology (J). OS, overall survival. DFS, disease free survival.

Figure 3

Screening of the DEGs in EC tissues compared with normal tissues with the subset-based approach using the TCGA-UCEC dataset. (A)-(E) Volcano plots showing the DEGs identified from five subsets of the TCGA-UCEC dataset. Red dots represent the upregulated genes, while blue dots represent the downregulated genes. (F)-(G) Venn diagrams showing the common upregulated genes (F) and common down-regulated genes (G) shared by the five subsets-based DEGs. DEGs, differentially expressed genes. EC, endometrial cancer.

Figure 4

Weighted gene co-expression network analysis of EC. (A) Selection of the optimal soft-thresholding power for the scale-free network. (B) DEGs dendrogram using the dissimilarity measure (1-TOM). (C) Correlation analysis of module eigengenes and immune score. Pearson's correlation coefficient and the corresponding P-value are shown. (D) Scatter plot of the black module showing the relationship between GS and MM. Key genes are indicated in the upper-right corner with the threshold of GS>0.7 and MM>0.7. TOM, topological overlap matrix. GS, gene significance. MM, module membership.

Figure 5

PPI network of the black module in endometrial cancer. (A) The PPI network consisting of 42 genes from the black module. Red nodes represent the upregulated genes, while the blue nodes represent the downregulated genes. (B) The top 12 nodes with the highest degree in the network. (C) The significant cluster formed by the top 11 nodes with most neighbor genes. PPI, protein-protein interaction.

Figure 6

GO (A) and KEGG (B) enrichment analysis of the black module in endometrial cancer. BP, biological process. CC, cellular component. MF, molecular function.

Figure 7

Validation of the abnormal expressions of the key genes via the online database of GEPIA. * $P < 0.01$.

Figure 8

Correlation analysis of the key genes in endometrial cancer (EC). (A)-(B) Boxplots showing the correlations of the key genes with tumor grade and histology in EC. (C) Pearson correlation matrices between expression values of the key genes. * P < 0.05, ** P < 0.01, *** P < 0.001, and ns indicates no significance.

Figure 9

Kaplan–Meier survival curves of the 11 key genes characterizing DFS difference with log-rank tests in endometrial cancer. DFS, disease-free survival.

Figure 10

Kaplan–Meier survival curves of the 11 key genes characterizing OS difference with log-rank tests in endometrial cancer. OS, overall survival.

Figure 11

We estimated the proportions of 22 distinct immune cell types in EC samples using the CIBERSORT algorithm (Fig. 11A). Next, we explored the correlation between each immune cell type proportion and the expression levels of the 11 key genes, and the results indicated that all of these 11 key genes, respectively, were significantly associated with at least 8 kinds of immune cells (Fig. 11B).

Supplementary Files

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

- [renamed3e8da.pdf](#)