Identification of a novel epithelial-mesenchymal transition-related gene signature for the prognosis prediction of endometrial carcinoma

Tianyuan Ruan
Third Affiliated Hospital of Sun Yat-sen University

Jing Wan
Third Affiliated Hospital of Sun Yat-sen University

Li Liu
Third Affiliated Hospital of Sun Yat-sen University

Qian Song
Third Affiliated Hospital of Sun Yat-sen University

Peigen Chen
Sixth Affiliated Hospital of Sun Yat-sen University

Xiaomao Li (lixmao@mail.sysu.edu.cn)
Third Affiliated Hospital of Sun Yat-sen University

Research Article

Keywords: Endometrial cancer, Epithelial-mesenchymal transition, Immune microenvironment, Immune checkpoint blockades, Prognostic signature

DOI: https://doi.org/10.21203/rs.3.rs-451908/v1

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

Background

Endometrial cancer (EC) is the most prevalent gynecological malignant tumor worldwide. The prognosis of EC patients with distant metastasis is poor. Epithelial-mesenchymal transition (EMT) plays a crucial role in tumor invasiveness and metastasis. The expression profile of EMT-related gene and their prognostic value and therapy target potential have not been systematically explored in endometrial cancer.

Method

The RNA-Seq expression profile and corresponding clinical data of EC patients were downloaded from The Cancer Genome Atlas database and the Gene Expression Omnibus (GEO) databases. A 9 EMT-related genes (EPHB2, TUFT1, CDKN2A, ONECUT2, RBP2, KLF8, E2F1, SIX1, ERBB2) prognosis risk model of EC was built by combining a machine learning algorithm and multivariate Cox regression. Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data (ESTIMATE) and Cibersort was applied to acquire for tumor purity and immune cell infiltration of high and low risk group. Also, response to immune checkpoint blockades (ICB) therapy and drug sensitivity were separately evaluated in TIDE and GDSC database for the EC patients.

Result

A 9 prognostic EMT-related genes (ERGs) signature was constructed according to TCGA-UCEC and then verified in the GEO dataset. EC Patients were assigned to high- and low-risk group according to the signature, and significant survival difference were discovered between the two groups. Furthermore, ERGs signature was identified as an independent prognostic factor in the multivariate Cox regression and the ROC curve revealed a high sensitivity and specificity to the model (0.78 at 2 years and 0.83 at 5 years). Furthermore, quantitative prognostic nomogram was constructed to predict survival time in EC patients. This ERGs signature was also significantly correlated with tumor purity and immune cell infiltration, which in compliance to the prediction that clinical response to ICB therapy were better in the high-risk group base on TIDE database. In addition, 8 chemical components which the high-risk group may be more sensitive to were screened out and analyzed.

Conclusions

We identified a novel EMT-related gene signature for predicting prognosis of EC patients and molecular understanding of EMT and immune microenvironment of EC. This may facilitate the decision-making during EC treatment.

Introduction
Endometrial carcinoma (EC) is one of the most prevalent gynecological malignant tumors. The incidence of EC kept increasing 1.3% per year from 2007 to 2016 worldwide. According to the latest cancer statistics, it was estimated that there would be approximately 65620 new cases diagnosed and 12590 death in the United States in 2020[1].Those patients who are diagnosed in early stage could receive appropriate treatment and achieve good prognosis, but still the prognosis of EC patients with distant metastasis was not so satisfying, with a 5-year survival rate lower than 20%[2]. It is essential to explore the underlying mechanisms of endometrial cancer metastasis and identify new biomarkers to predict the progression, prognosis, and response to treatment of EC.

Epithelial-mesenchymal transition (EMT) is a multistep process that epithelial cells develop a mesenchymal phenotype which characterized by the loss of polarity and barrier integrity, and therefore gain motility and invasive properties[3]. EMT plays a crucial role in tumor invasiveness and metastasis[4]. Accumulated evidence shows that EMT is a key process that drives cancer metastasis, enhances drug resistance, and relates to poor prognosis in multiple cancers, including endometrial cancer. A decreased level of E-cadherin was reported in endometrial cancer with an elevation of Snail and Slug nuclear expression, which significantly represented the EMT process and related to poor prognosis in endometrial cancer. However, the expression profile of EMT-related genes as prognostic factor and potential as therapeutic target have not been systematically explored.

The EMT signaling pathways could be activated by several cytokines or growth factors from the focal microenvironment. The tumor cells that undergo EMT are also associated with immune exclusion and immune deviation in the tumor microenvironment (TME)[5]. Recent pan cancer analysis also revealed a strong correlation between EMT and immune activation[6]. The crosstalk between EMT and the immune cells may bring potential therapeutic opportunities. However, the impact of EMT on the tumor immune microenvironment of endometrial cancer is still unknown.

In our study, we investigate the correlation between EMT related genes (ERGs) expression and clinical data in 543 endometrial cancer patients downloaded from The Cancer Genome Atlas (TCGA), and integrated the least absolute shrinkage and selection operator (LASSO) Cox regression model of the EMT-related genes signature. The patients with high EMT risk scores were strongly associated with lower overall survival. Then we validated this signature in GEO database. Afterwards, we established a nomogram to predict the prognosis by integrating clinical characteristics and the EMT gene signature. In the end, we conducted an integrated analysis of tumor immune microenvironment, immunotherapy response and drug sensitivity to high- and low-risk score group.

Method

Data preparation

We acquired the EMT-related genes (ERGs) list from epithelial-mesenchymal transition gene database(http://dbemt.bioinfo-minzhao.org/dbemt2.txt), which contained 1184 epithelial-mesenchymal transition genes with experimentally verified information and pre-calculated regulation information for
cancer metastasis[7]. High-throughput sequencing mRNA expression data of ERGs and corresponding clinical data of Endometrial carcinoma (UCEC) cohort were obtained from the TCGA. Somatic mutation data and copy number variation were also download from TCGA. Fragments Per Kilobase of transcript per Million method was applied to normalize the expression data. The gene expression dataset, GSE21882 based on GPL10422 SWEGENE H_v3.0.1 and GPL10423 SWEGENE H_v2.1.1 was downloaded from GEO (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21882). There were 45 type1 endometrial carcinoma samples included in the dataset. The raw data was download from GEO database as a data set soft file.

**Differentially expressed ERGs**

The “limma” package was applied to identify the differentially expressed EMT-related genes (EGRs). ERGs in TCGA-UCEC mRNA expression cohort were identified with FDR<0.05 and \(|\log FC| > 1\). Also “ggplot”, “ggrepel”, and “heatmaps” were executed in R package to perform visualization.

**Enrichment analysis of intersection genes**

GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment analyses and visualization of intersection genes were performed by “clusterProfiler”, “topGo”, “Rgraphviz” and “pathview” R package with p-value < 0.05 as the cut-off value.

**Individualized gene signature**

As we acquired the ERGs and the clinical data of 539 patients in TCGA-UCEC, univariate COX regression was performed with “survival” R packages and 101 prognostic related ERGs in TCGA-UCEC were identified with a p-value<0.05. Next, we performed a machine learning algorithm, the LASSO algorithm with a 10,000-time cross-validation. In the end, 9 genes were identified as prognostic related ERGs. With these 9 prognostic related ERGs, multivariate Cox regression analysis was performed to generate an individual predictive formula. The risk score (RS)= ( stands for the regression coefficient and x stands for the expression of selected gene) based on the formula mentioned above were calculated in TCGA-UCEC cohort and verified in GSE21882 dataset. Based on the RS, the patient in the two datasets were divided into low risk-group and high-risk group. And overall survival, disease free survival, clinical-pathological characteristics were compared between the two groups.

**Nomogram construction**

Univariate analysis has identified that age at diagnosis, FIGO stage, differentiation grade and risk score were significantly associate with overall survival. The cut-off p value was set as 0.05. All the significant prognostic variate mentioned above were enrolled in the multivariate Cox analysis. The p-value was set as 0.1 in the multivariate Cox analysis and age at diagnosis, differentiation grade, FIGO stage, risk score was included in the nomogram construction. C-index was applied to evaluate the predictive power of the constructed nomogram for the overall survival[8]. Larger C-index was regarding to a better predictive
power of the nomogram. Calibration curve was applied to determine the compliance between the predictive survival probability and the actual survival proportion after bias correction for the 3-year and 5-year overall survival.

**Tumor immune microenvironment analysis**

ESTIMATE (Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data) is one of the algorithms developed to evaluate the cell tumor composition by calculating the immune and stromal scores using Pearson's correlation coefficient. By using “estimate” R package, the immune and stromal scores were calculated based on the gene expression data of EC patients in TCGA-UCEC. The infiltration levels of 22 immune cells in each TCGA-UCEC patients were calculated through R packages “CIBERSORT” algorithm based on the gene expression levels[9].

**Immunotherapy and chemotherapy response prediction**

CTLA-4 and PD-1/PD-L1 pathway have been proven to be crucial in immune evasion, and thereby immune check point blockades therapy targeting PD-1 or CTLA-4 were developed to improve certain patients’ prognosis[10]. Tumor Immune Dysfunction and Exclusion (TIDE) algorithm were utilized to predict the clinical response to immune checkpoint blockades therapy[11]. As chemotherapy is necessary in advanced EC patients and substantial chemotherapy were still debatable after the failure of initial chemotherapy, we acquired information from the Genomics of Drug Sensitivity in Cancer (GDSC) database[12]. Clinical response to different component were predicted through R packages “pRRophetic” and the half-maximal inhibitory concentration (IC50) of the samples were predicted by ridge regression[13].

**Result**

**Identification of Differentially Expressed ERGs**

RNA-seq and clinical follow-up data from TCGA-UCEC datasets was downloaded, which contains 543 cancer samples and 35 normal samples. 1184 EMT-related genes (ERGs) are acquired from Human EMT Database. Then, the transcriptional expression of these 1184 ERGs in TCGA-UCEC RNA-seq data was extracted and compared between normal samples and UCEC samples through “limma” package in R software (|logFC| > 1, FDR < 0.05). The results indicated that there are 236 genes significantly upregulated and 169 genes significantly downregulated in UCEC (Figures 1(a) and 1(b)).

**Biological functions pathways analysis**

Biological functions pathways analysis was executed of these 405 differentially expressed ERGs. Gene ontology (GO) enrichment terms are shown in Figure 2. The result of GO enrichment shows that the screened ERGs positively regulate the DNA-binding transcription activation in UCEC. KEGG pathway enrichment of these genes are shown in Figure 2b. These results indicate that the differentially expressed
ERGs can promote cancer-related pathways, including Focal adhesion, EGFR tyrosine kinase inhibitor pathways, and may lead to transcriptional deregulation in cancer.

**Screened for the ERGs with significant prognosis**

The differentially expressed EMT-related genes (ERGs) in mRNA expression data of UCEC cohort were identified by the “limma” package in R software (FDR < 0.05, |logFC|> 1). After combining clinical data and mRNA expression data of ERGs, univariate Cox regression analyses were used to screen the differentially expressed ERGs with remarkable prognostic value ($p<0.05$). And 101 genes were screened with significant prognostic value in UCEC (Fig 3).

**Establishment and validation of the prognostic EMT-related gene signature**

In order to identify the EMT associated genes, a machine learning algorithm, LASSO algorithms, was performed and eventually 9 genes (EPHB2, TUFT1, CDKN2A, ONECUT2, RBP2, KLF8, E2F1, SIX1, ERBB2) were identified. We utilized 10,000-fold cross validation to select penalty value, lambda ($\lambda$). In our study, the optimal computed lambda ranged from 0.0413-0.0809 (Fig 4a). We set the $\lambda$ at 0.0809 as it can generate a stricter penalty model which contain much less variates. And at last, a 9-gene LASSO regression model was constructed (Fig 4b). The AUC of the constructed LASSO regression model was 0.78 when overall 5-year survival was applied and 0.78 when 2-year survival was applied. These 9 genes were selected for the next step analysis. Multivariate Cox regression analysis performed by “survival” R package was subsequently used to construct a predictive signature. The risk score of the prognostic signature was calculated and EC patients were divided into high-risk group and low-risk group. We verified this formula with overall survival data in both TCGA-UCEC cohort and GEO dataset (GSE21882). And we found that this 9-gene signature was significantly associated with OS (Figure 5a). In GSE21882, as only survival status after 5 years could be acquired, we divided the patients into non-survivors and survivors and the RS were compared between these two groups. Results showed that RS was significant higher in the non-survivors group (Fig 5b). In the TCGA-UCEC dataset, our analysis results showed that there was a significant statistical difference in the DFS between the low-risk and high-risk group (Figure 5c).

Recently, the endometrial cancer molecular classification was introduced by TCGA and promoted by NCCN which initiated a transition toward molecular-based classification with significant prognostic value and a potential impact on the treatment of EC. There are 4 molecular subgroups: p53-abnormal (p53abn), POLE-ultramutated (POLEmut), mismatch repair–deficient (MMRd), and no specific molecular profile (NSMP). In our study, we explored the above molecular profiling of the high- and low-risk group. In our study, there were higher copy number variations of BRAF, KRAS, PIK3CA, PTEN and TP53 in the high-risk group EC patients by the means of Chi-square test. Also, mutations rate of P53, PTEN and KRAS were higher in the high-risk group EC patients (Fig 5d). Coordinated to the TCGA pathology and molecular analysis, high risk group EC patients had higher P53 mutation rate and poorer prognosis.

The clinical pathological characteristics were compared between the two groups, including patients’ age, BMI, pre-menopause/ menopause status, number of pregnancies, different pathological grades, clinical-
pathological staging, clinical outcome after the initial treatment, metastasis/recurrence status. The cut-off \( p\)-value was set as 0.05. And the result showed that patients in high risk group have a higher percentage in menopause status, pregnancy no less than 2 times, G3 pathological differentiation, advanced clinical-pathological staging, and recurrence or metastasis status (Figure 6).

**Nomogram construction and validation**

After univariate and multivariate survival Cox analysis, 4 significant prognostic parameters were enrolled in the construction of nomogram (\( p<0.1 \) in multivariate cox regression), detailed result was showed in table 1. Respective points of age at diagnosis, differentiation grade, FIGO stage and risk score can be acquired through the point scale axis (Figure 7a). The calculated C-index was 0.796 (95% CI: 0.700-0.893). Calibration cures were presented in Figure 7b, which showed that the predictive probability was in good accordance with the actual survival proportion.

**Immune microenvironment analysis**

Based on the immune and stromal score which were acquired through ESTIMATE, we analyzed their relation with the prognosis data in TCGA-UCEC cohort. Patients in TCGA cohort were classified into high tumor purity and low tumor purity, significant prognostic difference of 5-year overall survival were identified, with \( p=0.012 \) (Fig 8a). Wilcoxon rank sum test was used to compared the immune score, stromal score and tumor purity between the high risk and low risk group in the TCGA-UCEC cohort. All \( p \) values were lower than 0.05, demonstrating a significant difference in immune score, stromal score and tumor purity between the two groups (Fig 8b). CTLA4 was known to negatively regulate T cell function through relatively unique and potentially non-overlapping molecular mechanisms[14]. Based on the risk score acquired through TCGA-UCEC cohort, CTLA4 was significantly up regulated in the high-risk score group (Fig 8c). The proportions of 22 tumor-immune cells were displayed in Fig 8d, and macrophage M0 and macrophage M2 were the top 2 components in EC tumor immune microenvironment. There were significantly difference between the high-risk and low-risk group regrading to the immune cells proportion, and the result were show in Figure 8e. And higher proportion of Macrophage M1 and lower Macrophage M2 were found in high risk group, which indicate a potential benefice from the immune check point blockades therapy.

**Immune checkpoint blockades therapy and drug sensitivity prediction**

Nowadays immune check point blockades targeting CTLA-4 and PD-1 has emerged as a promising strategy in various malignant tumor [10]. Clinical response to immune check point blockades were estimated, and results show that T cells dysfunction was lower in the high-risk group, which indicate a more promising response to immune checkpoint blockades targeting CTLA-4 and PD-1 (Fig 9a). To obtain a better comprehensive analysis of chemotherapy response, drug sensitivity data were acquire from the GDSC database and the IC50 were compared. Pearson correlation between the risk score and IC50 of different chemical components were conducted, and 8 chemical components were identified for a negative correlation with the risk score (\( r<-0.3, p<0.05 \)). Results were displayed in Fig 9b. Among the 8
components, A.443654 is a selective Akt inhibitor. ABT.263 and BI 2536 both can induce apoptosis. Lower IC50 of these drugs indicates that high-risk EC patients were more sensitive to the treatment compared to the low-risk EC patients.

**Discussion**

It has been widely proved that EMT is an important biological process in cancer, including endometrial cancer. Specifically, EMT is characterized by downregulation of the epithelial cell properties and activation of the mesenchymal characteristic. EMT is thought to be one of the major mechanisms that determine invasion and metastasis of cancer cells.

In this study, we collected the high-throughput data about the transcriptional expression of 543 EC samples and 35 nontumor samples with their corresponding clinical data from TCGA database. At the same time, 1184 EMT-related genes (ERGs) were acquired from Human EMT Database. Among these, we screened out the differentially expressed ERGs between EC samples and nontumor endometrium samples. Then, the differentially expressed ERGs are listed and analyzed, Biological functions pathways analysis indicated that differentially expressed ERGs can promote cancer-related pathways, including Focal adhesion, EGFR tyrosine kinase inhibitor pathways, and may lead to transcriptional misregulation in cancer. The AGE-RAGE axis is involved in the onset and progression of various chronic disorders and metabolic syndrome. Several recent researches revealed an association of the AGE-RAGE signaling with different pathological conditions involved in the progression of cancer including cell proliferation, invasion, metastasis, and angiogenesis[15]. AGEs biomarkers pentosidine malondialdehyde (MDA), a lipoxidation product closely linked to AGEs, is suggested to be found in high amount in the breast cancer and lung cancer patients[16]. Our findings also imply that AGE – RAGE signaling pathway may play a role in tumorigenesis of EC for the first time.

A novel machine learning method LASSO was used to finally identify 9 EMT-related genes, including EPHB2, TUFT1, CDKN2A, ONECUT2, RBP2, KLF8, E2F1, SIX1, ERBB2 as independent prognostic predictor, and a formula about risk score (RS) was raised through multivariate Cox regression analysis. Some of these genes, such as CDKN2A, E2F1 and ERBB2, have been characterized in endometrial cancer tumorigenesis. CDKN2A is one of the crucial defenses against cancer development in number of human cancers. CDKN2A was found to be inactivated in EC and its hypermethylation was correlated with an increased risk of EC[17, 18]. The HER2 (ERBB2) gene is amplified in 17%-33% of carcinosarcoma, uterine serous carcinoma, and a subset of high-grade endometrioid endometrial tumors[19, 20]. However, for the remaining genes, their functions have not been revealed in EC. TUFT1 was initially found to play an important role in the development and mineralization of tooth enamel, and later discovered in many cancerous tissues including pancreatic cancer, hepatocellular carcinoma. Its expression was correlates with unfavorable clinicopathologic characteristics and poor survival[21]. Recent studies have shown that TUFT1 promotes proliferation, metastasis, and epithelial mesenchymal transformation of cancer cells through the Ca2+/ PI3K/AKT pathway[22]. ONECUT2 is a member of the one cut family of transcription factors, known to be an important regulator of early retinal cell fate during development. ONECUT2 may
play a critical role in tumorigenesis pan-cancer wide[22]. RBP2 histone demethylase suppresses NOTCH signaling to sustain neuroendocrine differentiation and promote small cell lung cancer tumorigenesis[23]. Krüppel-like factor 8 (KLF8) is a critical inducer of EMT and invasion. KLF8 induces EMT primarily by repressing E-cadherin transcription. KLF8 promotes human breast cancer cell invasion and metastasis by transcriptional activation of MMP9[24]. Sine oculis-related homeobox 1 homolog (SIX1) is a transcription factor that regulates the development of many tissues and becomes reactivated or overexpressed in multiple types of human cancer. Recent research showed that SIX1 was essential for normal endometrial epithelial differentiation, and delayed DES-induced endometrial carcinogenesis by promoting basal differentiation of CK14+/18+ cells[24]. Another study also indicated that the SIX1 oncoprotein correlated with uterine cancer. SIX1 was not present in normal endometrium but was expressed in a subset of endometrial cancers in patients who were also more likely to have late-stage diseases [25]. Further studies of the function of these genes may identify new drivers and therapeutic targets for endometrial cancer.

A prognostic EMT-related gene signature for EC specimens was constructed and EC patients were divided into high-risk and low-risk groups according to the risk score in each sample. The results showed that the risk score could successfully predict the survival of EC patients, in which the high-risk group had a significantly worse OS than the low-risk group. The ROC curves also indicated that the EMT-related gene signature had favorable prognostic power. All of these findings were validated in GEO data sets. In addition, we comprehensively analyzed the relationship of the EMT related genes signature with different clinicopathological features. We found that our prognostic signature risk score was significantly correlated with clinicopathological data, including age, tumor grade and tumor stage. We also constructed the nomogram by the age, grade, stage and the EMT signature that showed a good performance in predicting the survival of EC patients.

Tumors could acquire the possibility of aggressiveness and metastases through EMT process[26]. Immune cells also participate in EMT via various pathways and in turn, cancer cell crosstalk with immune cells to release immunosuppressive substance to create an immunosuppressive microenvironment that promotes invasion and metastasis[27, 28]. Then we explored the immune microenvironments of the high- and low-risk group, and high-risk group get a lower stromal score, immune score and tumor purity. Immune infiltration analysis showed that Macrophage M2 are lower and Macrophage M1 are higher in cell fractions in the high-risk group, which indicated a proper microenvironment that enhanced T cells cytotoxic function.

There were few studies reporting immunotherapy and EMT targeted therapy in EC patients, and certain EC patients can benefit from Pembrolizumab[29]. We further evaluated the likelihood of response to immunotherapy in the high and low risk group. High-risk score group expressed higher level of CTLA4, which may indicate that they are more promising targets for therapeutic immunosuppressive treatment. T cells dysfunction score by TIDE algorithm also indicated a promising response to immune checkpoint blockades therapy targeting CTLA-4/PD-1. As second line chemotherapy was still debatable in advanced EC patients, we screened out 8 components which may be crucial in the substitute chemotherapy in the
GDSC database. A.443654 is a selective Akt inhibitor. PI3K/Akt/mTOR pathway is crucial in various tumor, including endometrium carcinoma. PTEN can regulate the tumor cells proliferation via mTOR pathway, and 30–60% EC exhibit a lack of PTEN. Reports showed that mTOR inhibitors can improve the disease-free survival in the advanced EC patients[30]. ABT.263 disrupts Bcl-2/Bcl-xl interactions with pro-death protein, leading to the initiation of apoptosis, BI 2536 is a Plk1 inhibitor, which can induce apoptosis and attenuates autophagy. and BI 2536 both can induce apoptosis. Lower IC50 of these drugs indicate that high-risk EC patients may be more sensitive to these chemo treatments.

**Conclusions**

In conclusion, we identified and validated a 9 EMT-related genes signature with certain prognostic value for EC patients. Immune microenvironment characters of high and low risk group were revealed, which may contribute to a better understanding for the molecular mechanism of EC and provide references for decision making during EC patients treatment.

**Abbreviations**

EC: Endometrial cancer

EMT: Epithelial-mesenchymal transition

TCGA: The Cancer Genome Atlas

GEO: Gene Expression Omnibus

EPHB2: EPH receptor B2

TUFT1: Tuftelin 1

CDKN2A: Cyclin dependent kinase inhibitor 2A

ONECUT2: One cut homeobox 2

RBP2: Retinol binding protein 2

KLF8: Kruppel like factor 8

E2F1: E2F transcription factor 1

SIX1: SIX homeobox 1

ERBB2: Erb-b2 receptor tyrosine kinase 2

RS: Risk score
**ESTIMATE**: Estimation of STromal and Immune cells in MAignant Tumor tissues using Expression data

**ICB**: Immune checkpoint blockade

**TIDE**: Tumor Immune Dysfunction and Exclusion

**GDSC**: Genomics of Drug Sensitivity in Cancer

**ERGs**: EMT-related genes

**UCEC**: Uterine Corpus Endometrial Carcinoma

**ROC**: Receiver operating characteristic

**TME**: Tumor microenvironment

**LASSO**: Least absolute shrinkage and selection operator

**IC50**: The half maximal inhibitory concentration

**FDR**: False discovery rate

**MMRd**: mismatch repair–deficient

**CTLA-4**: Cytotoxic T-lymphocyte antigen 4

**PD-L1**: Programmed Cell Death-Ligand 1

**FIGO**: The International Federation of Gynecology and Obstetrics

**TIDE**: Tumor Immune Dysfunction and Exclusion

**AUC**: Area under the curve

**OS**: Overall survival

**DFS**: Disease free survival

**Declarations**

**Availability of data and materials**

The datasets analyzed during the current study are available in The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO). Epithelial-Mesenchymal Transition gene list was acquired from Epithelial-Mesenchymal Transition GENE database (http://dbemt.bioinfo-minzhao.org/). Tumor Immune Estimation Resource database (TIMER) was implied for associated immune analysis. Immune infiltration was acquired from CIBERSOT databases (https://cibersort.stanford.edu/). Tumor Immune Dysfunction and
Exclusion (TIDE) and Genomics of Drug Sensitivity in Cancer (GDSC) were applied for immune checkpoint blockades therapy and drug sensitivity prediction separately.

Acknowledgements

All authors have contributed significantly. All authors are in agreement with the content of the manuscript.

Funding

Not applicable.

Contributions

JW and XL conceived and designed the overall study. TR performed the data analysis, prepared figures, and wrote the manuscript. LL and QS contributed significantly to the data analysis. PC provided important suggestions. All authors read and approved the final manuscript.

Consent for publication

Not applicable

Competing interests

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article. The authors declare there is no conflicts of interest regarding the publication of this paper.

Ethics approval and consent to participate

Not applicable.

References

1. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2020. CA Cancer J Clin 2020, 70(1):7–30.
2. Feng RM, Zong YN, Cao SM, Xu RH: Current cancer situation in China: good or bad news from the 2018 Global Cancer Statistics? Cancer Commun (Lond) 2019, 39(1):22.
3. Thiery JP, Sleeman JP: Complex networks orchestrate epithelial-mesenchymal transitions. Nat Rev Mol Cell Biol 2006, 7(2):131–142.
4. Thiery JP, Lim CT: Tumor dissemination: an EMT affair. Cancer Cell 2013, 23(3):272–273.
5. Romeo E, Caserta CA, Rumio C, Marcucci F: The Vicious Cross-Talk between Tumor Cells with an EMT Phenotype and Cells of the Immune System. Cells 2019, 8(5).
6. Mak MP, Tong P, Diao L, Cardnell RJ, Gibbons DL, William WN, Skoulidis F, Parra ER, Rodriguez-Canales J, Wistuba, II et al: A Patient-Derived, Pan-Cancer EMT Signature Identifies Global Molecular Alterations and Immune Target Enrichment Following Epithelial-to-Mesenchymal Transition. Clin Cancer Res 2016, 22(3):609–620.

7. Zhao M, Liu Y, Zheng C, Qu H: dbEMT 2.0: An updated database for epithelial-mesenchymal transition genes with experimentally verified information and precalculated regulation information for cancer metastasis. J Genet Genomics 2019, 46(12):595–597.

8. Iasonos A, Schrag D, Raj GV, Panageas KS: How to build and interpret a nomogram for cancer prognosis. J Clin Oncol 2008, 26(8):1364–1370.

9. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, Khodadoust MS, Esfahani MS, Luca BA, Steiner D et al: Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol 2019, 37(7):773–782.

10. Postow MA, Callahan MK, Wolchok JD: Immune Checkpoint Blockade in Cancer Therapy. J Clin Oncol 2015, 33(17):1974–1982.

11. Fu J, Li K, Zhang W, Wan C, Zhang J, Jiang P, Liu XS: Large-scale public data reuse to model immunotherapy response and resistance. Genome Med 2020, 12(1):21.

12. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, Bindal N, Beare D, Smith JA, Thompson IR et al: Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res 2013, 41(Database issue):D955-961.

13. Geeleher P, Cox N, Huang RS: pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. PLoS One 2014, 9(9):e107468.

14. Nirschl CJ, Drake CG: Molecular pathways: coexpression of immune checkpoint molecules: signaling pathways and implications for cancer immunotherapy. Clin Cancer Res 2013, 19(18):4917–4924.

15. Ahmad S, Khan H, Siddiqui Z, Khan MY, Rehman S, Shahab U, Godovikova T, Silnikov V, Moinuddin: AGES, RAGEs and s-RAGE; friend or foe for cancer. Semin Cancer Biol 2018, 49:44–55.

16. Waghela BN, Vaidya FU, Ranjan K, Chhipa AS, Tiwari BS, Pathak C: AGE-RAGE synergy influences programmed cell death signaling to promote cancer. Mol Cell Biochem 2020.

17. Furlan D, Carnevali I, Marcomini B, Cerutti R, Dainese E, Capella C, Riva C: The high frequency of de novo promoter methylation in synchronous primary endometrial and ovarian carcinomas. Clin Cancer Res 2006, 12(11 Pt 1):3329–3336.

18. Su L, Wang H, Miao J, Liang Y: Clinico pathological Significance and Potential Drug Target of CDKN2A/p16 in Endometrial Carcinoma. Sci Rep 2015, 5:13238.

19. Cousin S, Khalifa E, Crombe A, Laizet Y, Lucchesi C, Toulmonde M, Le Moulec S, Auzanneau C, Soubeyran I, Italiano A: Targeting ERBB2 mutations in solid tumors: biological and clinical implications. J Hematol Oncol 2018, 11(1):86.

20. Liu W, Han J, Shi S, Dai Y, He J: TUFT1 promotes metastasis and chemoresistance in triple negative breast cancer through the TUFT1/Rab5/Rac1 pathway. Cancer Cell Int 2019, 19:242.
21. Guo H, Ci X, Ahmed M, Hua JT, Soares F, Lin D, Puca L, Vosoughi A, Xue H, Li E et al: **ONECUT2 is a driver of neuroendocrine prostate cancer.** Nat Commun 2019, **10**(1):278.

22. Teng YC, Lee CF, Li YS, Chen YR, Hsiao PW, Chan MY, Lin FM, Huang HD, Chen YT, Jeng YM et al: **Histone demethylase RBP2 promotes lung tumorigenesis and cancer metastasis.** Cancer Res 2013, **73**(15):4711–4721.

23. Wang X, Lu H, Urvalek AM, Li T, Yu L, Lamar J, DiPersio CM, Feustel PJ, Zhao J: **KLF8 promotes human breast cancer cell invasion and metastasis by transcriptional activation of MMP9.** Oncogene 2011, **30**(16):1901–1911.

24. Suen AA, Jefferson WN, Wood CE, Padilla-Banks E, Bae-Jump VL, Williams CJ: **SIX1 Oncoprotein as a Biomarker in a Model of Hormonal Carcinogenesis and in Human Endometrial Cancer.** Mol Cancer Res 2016, **14**(9):849–858.

25. Suen AA, Jefferson WN, Wood CE, Williams CJ: **SIX1 Regulates Aberrant Endometrial Epithelial Cell Differentiation and Cancer Latency Following Developmental Estrogenic Chemical Exposure.** Mol Cancer Res 2019, **17**(12):2369–2382.

26. Ghatalia P, Zibelman M, Geynisman DM, Plimack ER: **Checkpoint Inhibitors for the Treatment of Renal Cell Carcinoma.** Curr Treat Options Oncol 2017, **18**(1):7.

27. Zhan HX, Zhou B, Cheng YG, Xu JW, Wang L, Zhang GY, Hu SY: **Crosstalk between stromal cells and cancer cells in pancreatic cancer: New insights into stromal biology.** Cancer Lett 2017, **392**:83–93.

28. Ma HY, Liu XZ, Liang CM: **Inflammatory microenvironment contributes to epithelial-mesenchymal transition in gastric cancer.** World J Gastroenterol 2016, **22**(29):6619–6628.

29. Makker V, Rasco D, Vogelzang NJ, Brose MS, Cohn AL, Mier J, Di Simone C, Hyman DM, Stepan DE, Dutcus CE et al: **Lenvatinib plus pembrolizumab in patients with advanced endometrial cancer: an interim analysis of a multicentre, open-label, single-arm, phase 2 trial.** Lancet Oncol 2019, **20**(5):711–718.

30. Oza AM, Pignata S, Poveda A, McCormack M, Clamp A, Schwartz B, Cheng J, Li X, Campbell K, Dodion P et al: **Randomized Phase II Trial of Ridaforolimus in Advanced Endometrial Carcinoma.** J Clin Oncol 2015, **33**(31):3576–3582.

**Tables**

Table 1: Univariate and multivariate Cox analysis
| p-value | HR    | 95%CI   | p-value | HR    | 95%CI   |
|---------|-------|---------|---------|-------|---------|
| 0.001   | 1.038 | 1.017-1.060 | 0.011   | 1.028 | 1.006-1.051 |
| 0.000   | 2.012 | 1.667-2.427 | 0.000   | 1.696 | 1.400-2.055 |
| 0.001   | 2.696 | 1.777-4.088 | 0.000   | 1.696 | 1.400-2.055 |
| 0.065   | 1.522 | 0.975-2.735 | 0.065   | 1.522 | 0.975-2.735 |
| 0.329   | 1.411 | 0.707-2.817 | 0.331   | 1.158 | 0.862-1.565 |
| 0.331   | 1.158 | 0.862-1.565 | 0.331   | 1.158 | 0.862-1.565 |
| 0.000   | 2.718 | 2.096-3.525 | 0.001   | 1.707 | 1.255-2.323 |
| 0.023   | 1.001 | 1.000-1.002 | 0.023   | 1.001 | 1.000-1.002 |
| 0.166   | 1.001 | 1.000-1.001 | 0.166   | 1.001 | 1.000-1.001 |

Figures
Figure 1

Volcano plot and heatmap. a) The volcano plot for the 1184 differently expressed EMT-related genes (ERGs) between Uterine Corpus Endometrial Carcinoma (UCEC) and normal endometrium tissues, with a FDR < 0.05, and |logFC| > 1; b) Heatmap for all the screened ERGs in UCEC.
Figure 2

Enrichment analysis of differently expressed ERGs. a) Bar plot shows the enriched biological process GO terms among ERGs with a cut-off value $P<0.05$. b) Circle plot shows the enriched KEGG terms among ERGs with a cut-off value $P<0.05$. 
Figure 3

Univariate Cox regression analysis data of the top 50 prognosis-related ERGs in UCEC-TCGA (P<0.05)
**Figure 4**

Construction of LASSO model. a) Cross validation for the penalty lambda. b) LASSO regression coefficients over different values for penalty parameters. c) AUC of the LASSO model constructed with the prognostic related ERGs in UCEC-TCAG cohort, and AUC of the 2-year survival was 0.78, AUC of the 5-year survival was 0.83.
Figure 5

Survival analysis and heatmap. a) In TCGA-UCEC cohort, the 539 patients were divided into high risk group and low risk group according to the ERGs signature risk score, and the over all survival was compared between the two groups through Kaplan-Meier method, as cut-off p-value 0.05. b) Patients in GSE21882 were divided into Non-survivors and survivors after 5-year follow up. RS was higher in the Non-survivors group; c) In TCGA-UCEC cohort, the disease free survival were compared in the high risk group and low risk group according to the gene signature risk score by Kaplan-Meier method, as cut-off p-value 0.05. d) Heat map of TCGA-UCEC patients, over view of the copy number variation and mutation status of BRAF, KRAS, PIK3CA, PTEN TP53. MSI status was also displayed in the figure.
Figure 6

Risk score and clinical parameters. According to the calculated Risk scores, the patients in TCGA-UCEC cohort were divided into low risk group and high-risk group. And different clinical characteristics were compared between the two groups with Chi-Square test. Including: a) Patients age at the time of diagnosis(p<0.001); b) Patients BMI(p=0.037). c) Different menopause status, pre-menopause status and menopause status(p<0.001); d) Number of pregnancies, 0-4(p=0.002); e) Different pathological
differentiation (p<0.001); f) Different clinical-pathological staging (p<0.001); G) Different outcome after the initial treatment, complete remission or not (p=0.105); h) Having recurrence/metastasis or not (p<0.001): The cut off p-value was set as 0.05.

Figure 7

Nomogram and Calibration plot. a) Nomogram constructed for the 3-year and 5-year predictive probability for the EC patients. b) Calibration curve for the 3-year and 5-year predictive survival probability
Immune microenvironment analysis. A) Over all survival were analyzed between high and low tumor purity group, and p=0.012. b) High risk score group were associate with lower Stromal score, Immune score and tumor purity. c) Higher CTLA-4 in the high risk group. * equals p<0.05; ** equals p<0.01. d)
immune cell infiltration landscape in TCGA-UCEC patients. e) difference in immune cell infiltration abundance between high- and low-risk group.

Figure 9

Immunotherapy and drug sensitivity prediction. a) T cell dysfunction and T cell exclusion in the high- and low-risk patients with TCGA-UCEC; b) Different chemotherapeutic response prediction in high- and low-risk patients.