

A 4-gene Signature Predicts Prognosis of Uterine Serous Carcinoma

Hui Chen

Guangzhou Medical University <https://orcid.org/0000-0001-8409-5980>

Lingjun Li

Third Affiliated Hospital of Guangzhou Medical College

Ping Qin

Third Affiliated Hospital of Guangzhou Medical College

Hanzhen Xiong

Third Affiliated Hospital of Guangzhou Medical College

Ruichao Chen

Third Affiliated Hospital of Guangzhou Medical College

Minfen Zhang

Third Affiliated Hospital of Guangzhou Medical College

Qingping Jiang (✉ Jiangqp2003@126.com)

Research article

Keywords: uterine serous carcinoma, biomarkers, signature, mortality, recurrence

Posted Date: July 16th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-42551/v1>

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 February 12th, 2021. See the published version at <https://doi.org/10.1186/s12885-021-07834-4>.

Abstract

Background: Uterine serous carcinoma (USC) is a rare, aggressive variant of endometrial cancer that accounts for more than 50% of endometrial cancer deaths, creating an urgent need for prognostic biomarkers.

Methods: USC RNA-Seq data and corresponding patients' clinical records were obtained from The Cancer Genome Atlas and Genotype-Tissue Expression datasets. Univariate cox, Lasso, and Multivariate cox regression analyses were conducted to forge a prognostic signature. Multivariable and univariable cox regression analysis and ROC curve evaluated the prediction efficiency both in the training and testing sets.

Results: We uncovered 1385 genes dysregulated in 110 cases of USC tissue relative to 113 cases of normal uterine tissue. Functional enrichment analysis of these genes revealed the involvement of various cancer-related pathways in USC. A novel 4-gene signature (KRT23, CXCL1, SOX9 and ABCA10) of USC prognosis was finally forged by serial regression analyses. Overall patient survival (OS) and recurrence-free survival (RFS) were significantly lower in the high-risk group relative to the low-risk group in both the training and testing sets. The area under the ROC curve of the 4-gene signature was highest among clinicopathological features in predicting OS and RFS. The 4-gene signature was found to be an independent prognostic indicator in USC and was a superior predictor of OS in early stage of USC.

Conclusions: Our findings highlight the potential of the 4-gene signature as a guide for personalized USC treatment.

1 Introduction

Endometrial cancer is the 2nd most common gynecologic malignancy worldwide (1). In China it also ranks the 2nd most common female cancer of the genital tract (2). Uterine serous carcinoma (USC/uterine serous papillary carcinoma) was first described by Hendrickson in 1982 (3). It represents a variant form of endometrial cancer whose clinicopathological and molecular features deviate from those of endometrioid carcinoma (EEC). Unlike EEC, USC tends to develop in elderly women, with low body weight and arises in the background of atrophic endometrium (4). 80–90% of USC tumors harbor TP53 mutation while retaining wildtype PTEN but losing ER/PR expression (4–8). USC accounts for almost 10% of endometrial cancers but is disproportionately responsible for poor outcomes, contributing up to 50% relapses and deaths from endometrial cancer (9). USC is characteristically aggressive, readily invading lymph-vascular space and undergoing abdominal dissemination in the early and stages even in the absence of myometrium invasion (10–15). A high proportion of USC cases present with extrauterine symptoms and adnexal, peritoneal or upper abdominal mass at diagnosis (16–18). Therefore, the clinicopathological parameters that can predict the prognosis of EEC, such as tumor size, myometrial invasion, lymph-vascular space invasion and lymph node metastasis, are not reliable indicators of USC

prognosis (4, 19–20). To the best of our knowledge, a robust system for predicting USC outcomes and recurrence is currently unavailable.

Advancements in molecular biological techniques and RNA-sequencing technology, have made it easier to identify genes that are associated with cancer initiation and progression (21). Single or multiple gene signatures exhibiting superior capacity to predict cancer outcomes relative to conventional clinicopathological features, have been developed (22–29). While similar signatures have been developed for EEC (30–32), to the best of our knowledge, none is available for USC.

Here, we carried out a genome-wide search for dysregulated genes in datasets from TCGA (The Cancer Genome Atlas) and GTEx (Genotype-Tissue Expression) and uncovered a 4-gene prognostic signature for USC. As an independent indicator of USC prognosis, this signature performs better than conventional prognostic factors.

2 Materials And Methods

2.1 Processing of TCGA-USC, GTEx datasets

Level 3 USC RNA-Seq dataset (reads FPKM with HTSeq) along with associated clinical information was downloaded from the TCGA database. Normal uterus GTEx data were downloaded from the UCSC Xena project (<http://xena.ucsc.edu/>) in October 2019. On the TCGA dataset, cases with follow-up data or overall survival (OS) of less than 30 days were excluded from the study.

2.1 Identification of dysregulated genes and functional enrichment analysis

Limma, an “R” Bioconductor package was used to identify genes that are dysregulated in USC tissues relative to normal uterine tissue by applying a threshold of $|\log_2FC| > 2$ and $FDR < 0.01$. GO (gene ontology) term analysis and KEGG (Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis) were conducted using the ClusterProfiler package on “R”. A P value = < 0.05 was considered indicative of significantly enriched functional annotations.

2.2 Construction and evaluation of the prognostic model

Half of the USC cases were randomly assigned to the training set. Cases with complete records on clinicopathological features, including OS, age, invasion, node, and stage, were assigned to the testing set. In the training set, dysregulated genes with prognostic potential were identified by univariable Cox regression analysis using the Survival package in “R”. P -value = < 0.05 was considered significant. To identify the most important prognostic genes, the least absolute shrinkage and selection operator (LASSO) regression method was executed in “R” using the Glmnet package. The prognostic signature for predicting OS was developed through multivariable Cox regression analysis using the “R” Survival package. The prognostic signature was applied in the calculation of the patients’ risk scores. The cases were then ranked into the high-risk and low-risk groups based on the median score. Kaplan-Meier survival

analysis was done using the “R” Survival package to plot the survival curves for the 2 risk groups. Receiver operating characteristic (ROC) curve analysis done using the “R” Survival ROC package to test the 4-gene signature’s accuracy in predicting OS for the high and low-risk USC cases. To validate the effectiveness of the signature, the OS risk score for each patient in the testing set was calculated using the signature, followed by Kaplan-Meier curve analysis and ROC estimation as was done in the training set. To evaluate the superiority of the 4-gene signature as a prognostic indicator, ROC curve analysis was done on other clinicopathological features, including age at diagnosis, myometrium invasion, node metastasis and stage. The process outlined above was used to test the signature’s effectiveness at predicting recurrence-free survival (RFS).

3 Results

3.1 TCGA-USC patient characteristics

A dataset of 110 UCS samples and 35 adjacent normal uterus tissue samples was downloaded from TCGA. The training and testing set consisted of data from 56 and 74 USC cases, respectively. The clinicopathological features among the 2 groups and the whole dataset did not differ significantly (P -value = > 0.05). These features were summarized in Table 1.

Table 1
Clinicopathological characteristics of USC patients in this study

Parameters		TCGA set	Training set	Testing set	P -value
		n = 110	n = 55	n = 74	
Age, years(mean \pm SD)		68.6 \pm 1.6	66.8 \pm 2.0	67.8 \pm 2.0	0.395 ^a
Myometrium invasion, percentage (mean \pm SD)		50.1 \pm 7.3	50.2 \pm 10.1	49.3 \pm 8.2	0.985 ^a
Stage, no(%)	Early (I + II)	53 (48.2)	26 (47.3)	32 (43.2)	0.797 ^b
	Late (III + IV)	57 (51.8)	29 (52.7)	42 (56.8)	
Lymph node metastasis, no(%)	YES	34 (38.6)	18 (40.0)	31 (41.9)	0.915 ^b
	NO	54 (61.4)	27 (60.0)	43 (58.1)	
Vital status, no(%)	Alive	74 (67.3)	39 (70.9)	48 (64.9)	0.769 ^b
	Dead	36 (32.7)	16 (29.1)	26 (35.1)	
^a ANOVA test					
^b Wilcox test					

3.2 Identification of dysregulated genes in USC and functional enrichment analysis

To ensure that our analysis compared equivalent numbers of USC and non-USC cases, we downloaded a dataset of normal uterus tissue samples from GTEx (n = 78), which along with the 35 in the TCGA dataset brought the total number of normal uterine cases to 113. Using Limma package in “R”, and a cutoff threshold of $|\log_2FC| > 2$, $FDR < 0.01$, 1385 genes were identified as being dysregulated in USC tissue vs the normal controls (Fig. 1A). Functional enrichment analysis revealed that the dysregulated genes are significantly associated with 717 GO term processes and 21 KEGG pathways. The most significantly enriched GO terms were extracellular matrix, mitosis, and cell adhesion, processes that might promote cancer progression (Fig. 1B). The most significantly enriched pathways are involved in cell adhesion, cell cycle, PI3K-Akt signaling pathway, cancerous microRNAs, transcriptional misregulation, and pathways involved in melanoma and bladder cancer (Fig. 1C).

3.3 Prognostic signature construction and evaluation in the training set

To identify dysregulated genes that may be associated with OS, we performed univariable Cox regression analysis and uncovered 29 genes that significantly correlated with OS (Table S1). To narrow down to the most important prognostic genes, we used LASSO regression analysis, which revealed 5 dysregulated genes as being potential critical indicators of USC survival (Fig. 2A). Next, multivariable Cox regression analysis narrowed down to a signature 4 genes, KRT23, CXCL1, SOX9 and ABCA10 (Figure. 2B) that effectively predict OS (Table 2). Among these, KRT23, CXCL1, and SOX9 exhibited positive regression coefficients, indicating a high risk of mortality. While ABCA10 showed a negative regression coefficient, implying a low mortality risk. Next, we constructed the following risk prediction formula based on the 4 prognostic genes and used it to calculate each patient’s risk score in the training set: $\text{risk score} = (0.5424 \times \text{expression level of KRT23}) + (0.2398 \times \text{expression level of CXCL1}) + (0.5398 \times \text{expression level of SOX9}) - (1.7023 \times \text{expression level of ABCA10})$. This signature was used to calculate risk scores for 56 USC cases (individually) in the training set. The risk scores were then ranked linearly and assigned as high-risk or low-risk based on whether they were higher or lower than the median risk score (Fig. 2C). The relationship between risk scores and survival time was showed in Fig. 2D. Visualization of the expression of the 4 genes in a heatmap revealed that the expression level of the 3 high-risk genes increasing with rising risk scores, while the low-risk gene showed an opposite correlation (Fig. 2E). Kaplan-Meier analysis revealed that patients in high-risk group experienced worse outcomes relative to the low-risk group (P -value = 0.003317, Figure 2F). Relative to standard clinicopathological parameters like age, myometrium invasion, node metastasis and disease stage, the 4-gene prognostic signature scored 0.855 in AUC (area under the ROC curve) analysis, indicating superior performance over conventional prognostic factors (0.213, 0.796, 0.728 and 0.564 for age, myometrium invasion, node metastasis and stage, respectively; Figure. 2G).

Table 2
Four signature genes constructed in this model

gene	coef ^a	HR	HR.95L	HR.95H	P-value ^b
ABCA10	-1.7023	5.4868	0.8160	36.8946	0.0008
KRT23	0.5424	1.7201	1.0986	2.6931	0.0040
CXCL1	0.2398	1.2709	0.9217	1.7526	0.0033
SOX9	0.5398	1.7156	0.8041	3.6602	0.0072
^a Coefficients derived from multivariable Cox regression analysis.					
^b Derived from the univariate Cox regression analysis.					

3.4 Validation of the 4-gene signature in the testing set

To assess the robustness of the 4-gene prognostic signature, risk scores for the 74 USC cases in the testing set were calculated and ranked as described in Sect. 3.3 (Figure. 3A). The relationship between risk scores and survival is shown in Fig. 3B. This analysis revealed that the expression of the 3 high-risk genes increased with rising risk scores, while the low-risk gene exhibited the opposite effect (Fig. 3C). Kaplan-Meier curve indicated that the high-risk group experienced worse outcomes relative to the low-risk group (P -value = 0.0004387, Figure 3D). The score of 0.811 for the 4-gene signature was revealed by AUC analysis was higher than for conventional prognosis indicators (0.430, 0.752, 0.808 and 0.688 for age, myometrium invasion, node metastasis and stage, respectively; Fig. 3E), consistent with observations made in the training set.

3.5 Independent prognostic value of the 4-gene signature.

To evaluate the potential of the 4-gene signature independently of conventional prognosis indicators, we used univariate and multivariate Cox regression analysis on testing set cases with reporting complete clinical features. This analysis revealed that our prognostic signature and tumor stage are both independent predictors of OS (Table 3). Next, we tested if the 4-gene signature could predict OS at different disease stages. To this end, we stratified the cases by stage into early (stage I + II) and late stage (stage III + IV). Patients in high-risk group in both early and late stage exhibited lower OS relative to those in the low-risk group (P value = 0.003306 and P value = 0.02755, respectively, Fig. 3F-G). These results indicate that the 4-gene signature has superior performance in early stage, highlighting its potential clinical application.

Table 3

Univariate and multivariate Cox regression analysis of OS in USC patients in the testing set (n = 74)

Parameters	Univariate analysis				Multivariate analysis			
	HR	HR.95L	HR.95H	<i>P</i> -value	HR	HR.95L	HR.95H	<i>P</i> -value
invasion ($\geq 50\%$ / $\leq 50\%$)	1.0203	1.0069	1.0340	0.0029	1.0080	0.9931	1.0232	0.2953
node (YES/NO)	3.6689	1.5769	8.5366	0.0026	2.0157	0.6958	5.8396	0.1965
stage (III + IV/I + II)	2.0795	1.3531	3.1959	0.0008	1.8088	1.0063	3.2510	0.0476
4-gene riskScore (high/low)	1.0511	1.0154	1.0881	0.0047	1.0511	1.0108	1.0930	0.0125

3.6 Evaluation of the 4-gene signature in predicting RFS.

To evaluate whether the 4-gene signature could predict recurrence-free survival (RFS) in USC, TCGA-USC cases with RFS data were analyzed. Cases with RFS of < 30 days were excluded and 95 cases further analyzed. Each patient's risk scores were calculated and ranked as described in Sect. 3.3 (Fig. 4A). The risk scores and recurrent time are shown in Fig. 4B. This analysis revealed that expression of the 4 genes increased with rising risk scores (Fig. 4C). Kaplan-Meier analysis revealed that the high-risk group had higher recurrence rate relative to the low-risk group (P value = 0.01198, Figure 4D). The AUC analysis of the prognostic signature revealed a score of 0.737 at RFS prediction, which was higher than the scores from conventional indicators (0.151, 0.595, 0.551 and 0.632 for age, myometrium invasion, node metastasis and stage, respectively, Fig. 4E). Univariate and multivariate Cox regression analysis revealed the prognostic signature and stage as independent prognostic factors for RFS, consistent with OS analysis (Table 4). Analysis of the effectiveness of the 4-gene signature in predicting RFS at different disease stages revealed that patients in low-risk and high-risk groups had significantly different RFS in late stage (P value = 0.003489, Fig. 4F). However, there was no difference in early stage between the two risk groups (Figure S1).

Table 4
Univariate and multivariate Cox regression analysis of RFS in USC patients in USC patients in TCGA cohort (n = 95)

Parameters	Univariate analysis				Multivariate analysis			
	HR	HR.95L	HR.95H	P-value	HR	HR.95L	HR.95H	P-value
invasion ($\geq 50\%$ / $\leq 50\%$)	1.0100	0.9981	1.0220	0.1005	1.0010	0.9885	1.0136	0.8808
node (YES/NO)	1.9975	0.9192	4.3409	0.0806	1.2448	0.5153	3.0072	0.6266
stage (III + IV/I + II)	1.9975	1.3016	3.0656	0.0015	1.9088	1.1905	3.0605	0.0073
4-gene riskScore (high/low)	1.0744	1.0248	1.1264	0.0029	1.0735	1.0212	1.1284	0.0053

4 Discussion

Here, we analyzed USC datasets from TCGA and GTEx and uncovered 1385 genes that are dysregulated USC tissues relative to normal endometrial tissue. KEGG pathway analysis revealed that these genes mainly belong to cancer-associated pathways, including melanoma and bladder cancer as well as in pathways associated with cell adhesion, cell cycle, PI3K-Akt signaling pathway, cancer-linked microRNAs and transcriptional misregulation. Disruption of cell adhesion may explain why USC tends to disseminate early, spreading to fallopian tubes or invading lymph-vascular space. The tumor suppressor, TP53 is the frequently mutated gene in USC (34). USC's high proliferative rate dysregulated cycle control may contribute to the high relapse and mortality rates in endometrial cancers. The PI3K/AKT/mTOR signaling pathway is the most frequently dysregulated pathway in EEC (34). In USC, PIK3CA mutation occurs in about 30% of cases (11, 34–35), which is consistent with the involvement of the PI3K-Akt signaling pathway seen from our analysis. Inhibition of PI3K/AKT/mTOR signaling strongly suppresses EEC progression (36–38) and clinical trials targeting PI3K/AKT/mTOR signaling in solid tumors have shown promise (39). However, the benefits of this against in endometrial cancers is controversial due to the complexity of pharmacological action and toxicity (40). Further studies are needed to better target PI3K/AKT/mTOR signaling in endometrial cancer.

The TCGA database, which offers a collection of complete transcriptomic data and associated clinical information, is publicly available for data mining (41). To identify important dysregulated genes associated with USC outcomes, we used LASSO and Cox regression analysis. LASSO is widely applied in modeling high-dimensional data and avoids overfitting risk and improves prediction accuracy (42). Our analysis generated a 4-gene signature for predicting USC OS by calculating each patient's risk score. We find that patients with high scores exhibit poor outcomes relative to those with low scores, an observation

that was validated in both the training and testing sets. ROC curve analysis revealed this signature's superiority over conventional prognostic parameters (age, myometrium invasion, node metastasis, and stage) in the training and testing sets. Our data show that both the 4-gene signature and disease stage are independent prognostic indicators OS. Patients with late-stage of the disease have an unfavorable prognosis for most malignant solid tumors. However, for USC, the early-stage disease does not necessarily correlate with good prognosis due to the tumor's propensity for shedding, spreading and invading the lymph-vascular space even when the lesion confined to the endometrium or polyps. Management of patients with early stage USC is controversial (4, 43–44). Our signature identified high-risk patients in the early stage USC group who had much poorer OS relative to low-risk patients in the same group. Our data show that this signature performed better in the early stage group than in the late stage group, highlighting its potential value in guiding the management for early stage USC.

The average recurrence rate for stage IA USC after chemotherapy, radiotherapy or surgery is 8.7%, 25% and 12.4% respectively. For stage IB/IC the corresponding recurrence rate are 10.8%, 36.6% and 37.3%, respectively (11). Our 4-gene signature predicts a higher recurrence risk in the high-risk group relative to the low-risk group. Consistently with our OS, ROC curve analysis, this 4-gene signature exhibited superior effectiveness over conventional indicators of RFS. Both the signature and disease stage were an independent prognostic factor for RFS. Our data show that the 4-gene signature is effective at RFS prediction in late stage disease but showed no difference between high and low-risk groups in early stage. This may be due to too few recurrent cases (8 cases out of 45 cases) in early stage in the TCGA cohort.

The 4 genes in the signature have been associated with various cancers. KRT23 has been implicated as an oncogene in liver cancer (46) and colorectal cancer (47). CXCL1 is overexpressed in EEC tissue relative to normal endometrium and promotes tumorigenesis by promoting neutrophil chemotaxis (48). Snail induces ovarian epithelial-mesenchymal transition via CXCL1 and CXCL2, representing an immunological therapeutic target (49). SOX9 overexpression in uterine epithelium may induce endometrial hyperplastic lesions (50), promoting endometrial cancer cell proliferation (51). ABCA10 has been proposed as a prognostic marker in ovarian carcinoma (52). Germline single nucleotide polymorphisms in ABCA10 may affect follicular lymphoma overall survival (53). So far, none of the 4 genes has been associated with USC, though CXCL1 and SOX9 are associated with EEC progression.

5 Conclusion

Here, we an analysis of USC genome-wide expression profiles in TCGA and GTEx datasets. We have identified genes that are dysregulated in USC and explored their molecular functions and pathways. More importantly, we have developed and validated a 4-gene signature that robustly predicts USC OS and RFS. This signature is an independent prognostic indicator that is more superior to conventional indicators of USC prognosis, especially when predicting OS in early stage of USC. Our findings highlight the potential of this signature as a guide for personalized USC treatment. However, more independent cohorts are

needed to validate the signature and to elucidate the molecular mechanisms of these predictive genes in USC.

Abbreviations

USC, uterine serous carcinoma; TCGA, The Cancer Genome Atlas; GTEx, Genotype-Tissue Expression; OS, overall patient survival; RFS, recurrence-free survival; EEC, endometrioid carcinoma; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; LASSO, the least absolute shrinkage and selection operator; ROC, receiver operating characteristic.

Declarations

Ethics Statement

The TCGA and GTEx datasets are publicly available, no ethical approval is required.

Data Availability

All datasets generated for this study are included in the manuscript.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

HC, LL and PQ analyzed the data and drafted the manuscript. HX helped interpret the data. RC prepared all figures. MZ edited all tables. QJ designed the study.

Acknowledgments

We would like to thank TCGA network and GTEx dataset. We also thank Dr. Jerry from Freescience Info Technology Co., Ltd., Ningbo, China, for editing the English text of a draft of this manuscript.

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68:394–424.
2. Chen W, Sun K, Zheng R, Zeng H, Zhang S, Xia C, Yang Z, Li H, Zou X, He J. Cancer incidence and mortality in China, 2014. *Chin J Cancer Res.* 2018;30:1–12.

3. Hendrickson M, Ross J, Eifel P, et al. Uterine papillary serous carcinoma: a highly malignant form of endometrial adenocarcinoma. *Am J Surg Pathol*. 1982;6:93–108.
4. Fader AN, Boruta D, Olawaiye AB, Gehrig PA. Uterine papillary serous carcinoma: epidemiology, pathogenesis and management. *Curr Opin Obstet Gynecol*. 2010;22:21–9.
5. Black JD, English DP, Roque DM, Santin AD. Targeted therapy in uterine serous carcinoma: an aggressive variant of endometrial cancer. *Womens Health (Lond)*. 2014;10:45–57.
6. Hu S, Hinson JL, Matnani R, Cibull ML, Karabakhtsian RG. Are the uterine serous carcinoma underdiagnosed? Histomorphologic and immunohistochemical correlates and clinical follow up in high-grade endometrial carcinomas initially diagnosed as high-grade endometrioid carcinoma. *Mod Pathol*. (2018) 31:358–64.
7. Murali R, Davidson B, Fadare O, et al. High-grade endometrial carcinomas: morphologic and immunohistochemical features, diagnostic challenges and recommendations. *Int J Gynecol Pathol*. 2019;38(Suppl 1):40–63.
8. Yemelyanova A, Ji H, Shih IeM, et al. Utility of p16 expression for distinction of uterine serous carcinomas from endometrial endometrioid and endocervical adenocarcinomas: immunohistochemical analysis of 201 cases. *Am J Surg Pathol*. 2009;33:1504–14.
9. del Carmen MG, Birrer M, Schorge JO. Uterine papillary serous cancer: a review of the literature. *Gynecol Oncol*. 2012;127:651–61.
10. Holman LL, Pal N, Iglesias DA. etc. Factors prognostic of survival in advanced-stage uterine serous carcinoma. *Gynecol Oncol*. 2017;146:27–33.
11. Fader AN, Santin AD, Gehrig PA. Early stage uterine serous carcinoma: management updates and genomic advances. *Gynecol Oncol*. (2013) 129:244–50.
12. Slomovitz BM, Burke TW, Eifel PJ, et al. Uterine papillary serous carcinoma (UPSC): a single institution review of 129 cases. *Gynecol Oncol*. 2003;91:463–9.
13. Hanley KZ, Fadare O, Fisher KE, et al. Clinical significant of positive pelvic washings in uterine papillary serous carcinoma confined to an endometrial polyp. *Int J Gynecol Pathol*. 2016;35:249–55.
14. Semaan A, Mert I, Munkarah AR, et al. Clinical and pathologic characteristics of serous carcinoma confined to the endometrium: a multi- institutional study. *Int J Gynecol Pathol*. 2013;32:181–7.
15. Hui P, Kelly M, O'Malley DM, et al. Minimal uterine serous carcinoma: a clinicopathological study of 40 cases. *Mod Pathol*. 2005;18:75–82.
16. Fader AN, Starks D, Gehrig PA, et al. An updated clinicopathologic study of uterine papillary serous carcinoma. *Gynecol Oncol*. 2009;115:244–8.
17. Kommos F, Faruqi A, Gilks CB, et al. Uterine serous carcinomas frequently metastasize to the fallopian tube and can mimic serous tubal intraepithelial carcinoma. *Am J Surg Pathol*. 2017;41:161–70.
18. Snyder MJ, Bentley R, Robboy SJ. Transtubal spread of serous adenocarcinoma of the endometrium: an underrecognized mechanism of metastasis. *Int J Gynecol Pathol*. 2006;25:155–60.

19. Solmaz U, Ekin A, Mat E, et al. Analysis of clinical and pathological characteristics, treatment methods, survival, and prognosis of uterine papillary serous carcinoma. *Tumori*. 2016;102:593–9.
20. Singh N, Hirschowitz L, Zaino R, et al. Pathologic prognostic factors in endometrial carcinoma (other than tumor type and grade). *Int J Gynecol Pathol*. 2019;38(Suppl 1):93–113.
21. Bradner JE, Hnisz D, Young RA. Transcriptional Addiction in Cancer. *Cell*. 2017;168:629–43.
22. Ozawa T, Kandimalla R, Gao F, et al. A microRNA signature associated with metastasis of T1 colorectal cancers to lymph nodes. *Gastroenterology*. 2018;154:844–8.
23. Beck D, Thoms JAI, Palu C, et al. A 4-gene lincRNA expression signature predicts risk in multiple cohorts of acute myeloid leukemia patients. *Leukemia*. 2018;32:263–72.
24. Bagnoli M, Canevari S, Califano D, et al. Development and validation of a microRNA-based signature (MiROvaR) to predict early relapse or progression of epithelial ovarian cancer: a cohort study. *Lancet Oncol*. 2016;17:1137–46.
25. Rini B, Goddard A, Knezevic D, et al. A 16-gene assay to predict recurrence after surgery in localized renal cell carcinoma: development and validation studies. *Lancet Oncol*. 2015;16:676–85.
26. Li J, Chen Z, Tian L, et al. LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with esophageal squamous cell carcinoma. *Gut*. 2014;63:1700–10.
27. Verhaak RG, Tamayo P, Yang JY, et al. Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. *J Clin Invest*. 2013;123:517–25.
28. Sandoval J, Mendez-Gonzalez J, Nadal E, et al. A prognostic DNA methylation signature for stage I non-small-cell lung cancer. *J Clin Oncol*. 2013;31:4140–7.
29. Li X, Zhang Y, Zhang Y, et al. Survival prediction of gastric cancer by a seven-microRNA signature. *Gut*. 2010;59:579–85.
30. Zhou M, Zhang Z, Zhao H, et al. A novel lncRNA-focus expression signature for survival prediction in endometrial carcinoma. *BMC Cancer*. 2018;18:39.
31. Farkas SA, Sorbe BG, Nilsson TK. Epigenetic changes as prognostic predictors in endometrial carcinomas. *Epigenetics*. 2017;12:19–26.
32. Torres A, Torres K, Pesci A, et al. Diagnostic and prognostic significance of miRNA signatures in tissues and plasma of endometrioid endometrial carcinoma patients. *In J Cancer*. 2013;132:1633–45.
33. Tang Z, Li C, Kang B, et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017;45:W98–102.
34. Jones NL, Xiu J, Chatterjee-Paer S, et al. Distinct molecular landscapes between endometrioid and nonendometrioid uterine carcinomas. *Int J Cancer*. 2017;140:1396–404.
35. Catusus L, D'Angelo E, Pons C, et al. Expression profiling of 22 genes involved in the PI3K-AKT pathway identifies two subgroups of high-grade endometrial carcinomas with different molecular alterations. *Mod Pathol*. 2010;23:694–702.

36. Packer LM, Geng X, Bonazzi VF, et al. PI3K inhibitors synergize with FGFR inhibitors to enhance antitumor responses in FGFR2 mutant endometrial cancers. *MOL Cancer Ther.* 2017;16:637–48.
37. Suh DS, Park SE, Jin H, et al. LRIG2 is a growth suppressor of Hec-1A and Ishikawa endometrial adenocarcinoma cells by regulating PI3K/AKT- and EGFR-mediated apoptosis and cell-cycle. *Oncogenesis.* 2018;7:3.
38. Hsu AH, Lun MA, Shim KS, et al. Crosstalk between PKC α and PI3K/AKT signaling is tumor suppressive in the endometrium. *Cell Rep.* 2018;24:655–69.
39. Bendell JC, Varghese AM, Hyman DM, et al. A first-in-human phase 1 study of LY3023414, an oral PI3K/mTOR dual inhibitor, in patients with advanced cancer. *Clin Cancer Res.* 2018;24:3253–62.
40. Barra F, Evangelisti G, Ferro Desideri L, et al. Investigational PI3K/AKT/mTOR inhibitors in development for endometrial cancer. *Expert Opin Investig Drugs.* 2019;28:131–42.
41. Ali MM, Akhade VS, Kosalai ST, et al. PAN-cancer analysis of S-phase enriched lncRNAs identifies oncogenic drivers and biomarkers. *Nat Commun.* 2018;9:883.
42. Pi L, Halabi S. Combined performance of screening and variable selection methods in ultra-high dimensional data in predicting time-to-event outcomes. *Diagn Progn Res.* 2018;2:21.
43. Wang Y, Yu M, Yang JX, et al. Clinicopathological and survival analysis of uterine papillary serous carcinoma: a single institutional review of 106 cases. *Cancer Manag Res.* 2018;10:4915–28.
44. Black C, Feng A, Bittinger S, et al. Uterine Papillary Serous Carcinoma: A Single-Institution Review of 62 Cases. *Int J Gynecol Cancer.* 2016;26:133–40.
45. Sutton G, Axelrod JH, Bundy BN, et al. Adjuvant whole abdominal irradiation in clinical stages I and II papillary serous or clear cell carcinoma of the endometrium: a Phase II study of the Gynecologic Oncology Group. *Gynecol Oncol.* 2006;100:349–54.
46. Kim D, Brocker CN, Takahashi S, et al. Keratin 23 Is a Peroxisome Proliferator-Activated Receptor Alpha-Dependent, MYC-Amplified Oncogene That Promotes Hepatocyte Proliferation. *Hepatology.* 2019;70:154–67.
47. Zhang N, Zhang R, Zou K, et al. Keratin 23 promotes telomerase reverse transcriptase expression and human colorectal cancer growth. *Cell Death Dis.* (2017) 8:e2961.
48. Wallace AE, Sales KJ, Catalano RD, et al. Prostaglandin F $_{2\alpha}$ -prostanoid receptor signaling promotes neutrophil chemotaxis via chemokine (C-X-C motif) ligand 1 in endometrial adenocarcinoma. *Cancer Res.* 2009;69:5726–33.
49. Taki M, Abiko K, Baba T, et al. Snail promotes ovarian cancer progression by recruiting myeloid-derived suppressor cell via CXCR2 ligand upregulation. *Nat Commun.* 2018;9:1685.
50. Gonzalez G, Mehra S, Wang Y, et al. Sox9 overexpression in uterine epithelia induces endometrial gland hyperplasia. *Differentiation.* 2016;92:204–15.
51. Saegusa M, Hashimura M, Suzuki E, et al. Transcriptional up-regulation of Sox9 by NF- κ B in endometrial carcinoma cells, modulating cell proliferation through alteration in the p14(ARF)/p53/p21(WAF1) pathway. *Am J Pathol.* 2012;181:684–92.

52. Wang Y, Lei L, Chi YG, et al. A comprehensive understanding of ovarian carcinoma survival prognosis by novel biomarkers. *Eur Rev Med Pharmacol Sci.* 2019;23:8257–64.
53. Baecklund F, Foo JN, Bracci P, et al. A comprehensive evaluation of the role of genetic variation in follicular lymphoma survival. *BMC Med Genet.* 2014;15:113.

Figures

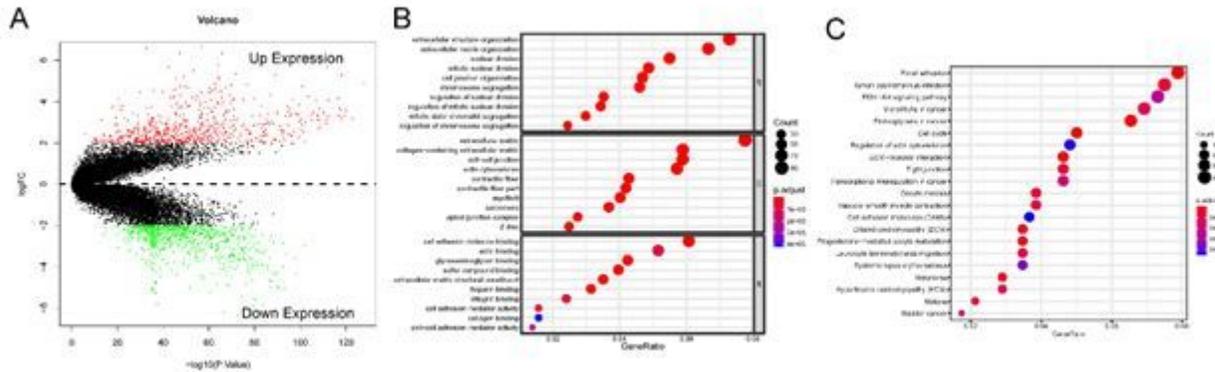


Figure 1

Dysregulated genes in the USC and functional enrichment analysis. (A) Volcano plot shows 1385 up- and down-regulated genes between 110 USCs and 113 normal uterus tissue with the threshold of $|\log_2FC| > 2$ and $FDR < 0.01$. (B) Top 30 Gene ontology (GO) biological processes of dysregulated genes. (C) Top 20 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways of dysregulated genes.

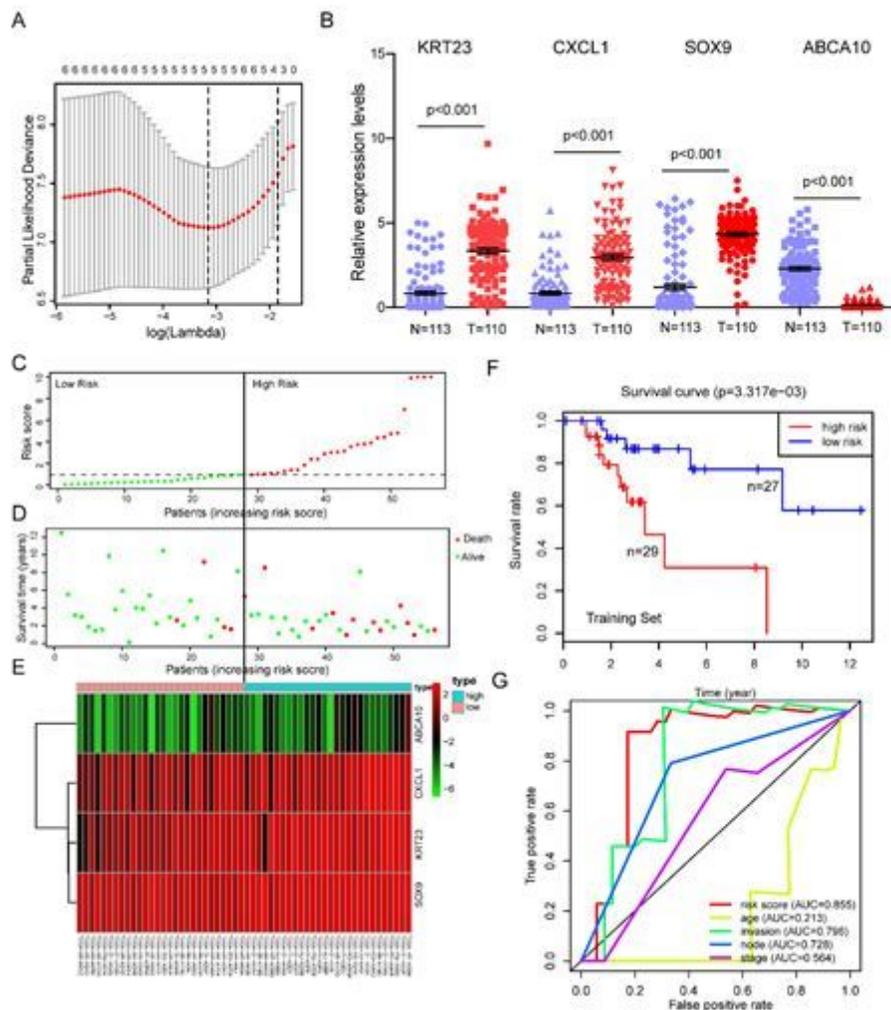


Figure 2

Identification of the 4-gene signature and prediction of overall survival (OS) in the training set. (A) LASSO regression defines 5 critical survival prognostic genes. (B) Expression differences of the 4 genes, identified by multivariable Cox regression analysis, between 110 USC patients and 113 normal endometrium tissues are analyzed by Wilcoxon test. $P < 0.05$ is considered statistically significant. (C-E) The distribution of risk score, OS, and survival status and the 4 genes expression patterns for the 56 patients in the training set. (F) Kaplan–Meier analysis to compare OS between patients in the high- and low-risk group in the training set. (G) ROC analysis of the 4-gene signature and other clinicopathological parameters (age, invasion, node metastasis and stage) for prediction OS in the training set.

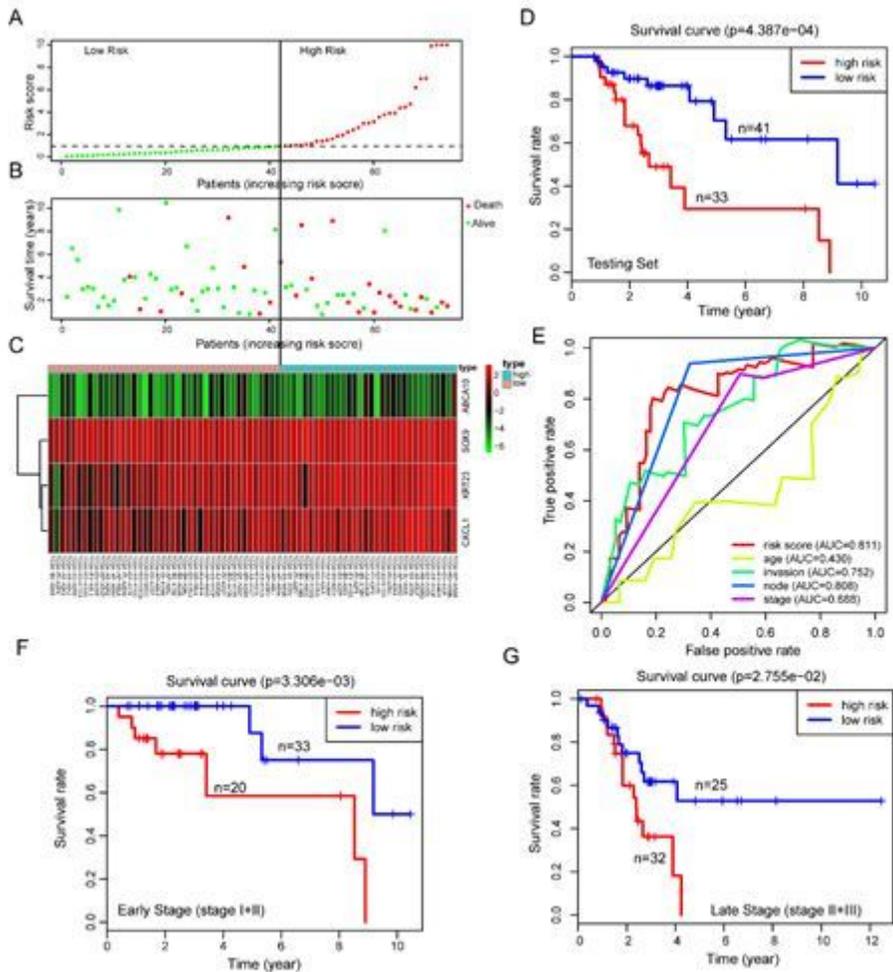


Figure 3

Validation of the 4-gene signature in predicting OS in the testing set and different stage patients. (A-C) The distribution of risk score, OS, and survival status and the 4 genes expression patterns for the 74 patients in the training set. (D) Kaplan–Meier analysis compares OS between patients in the high- and low-risk group in the testing set. (E) ROC analysis of the 4-gene signature and other clinicopathological parameters (age, invasion, node metastasis and stage) for prediction OS in the testing set. (F) Kaplan–Meier analysis compares OS between patients in the high- and low-risk group in early stage (I+II) patients. (G) Kaplan–Meier analysis compares OS between patients in the high- and low-risk group in late stage (III+IV) patients.

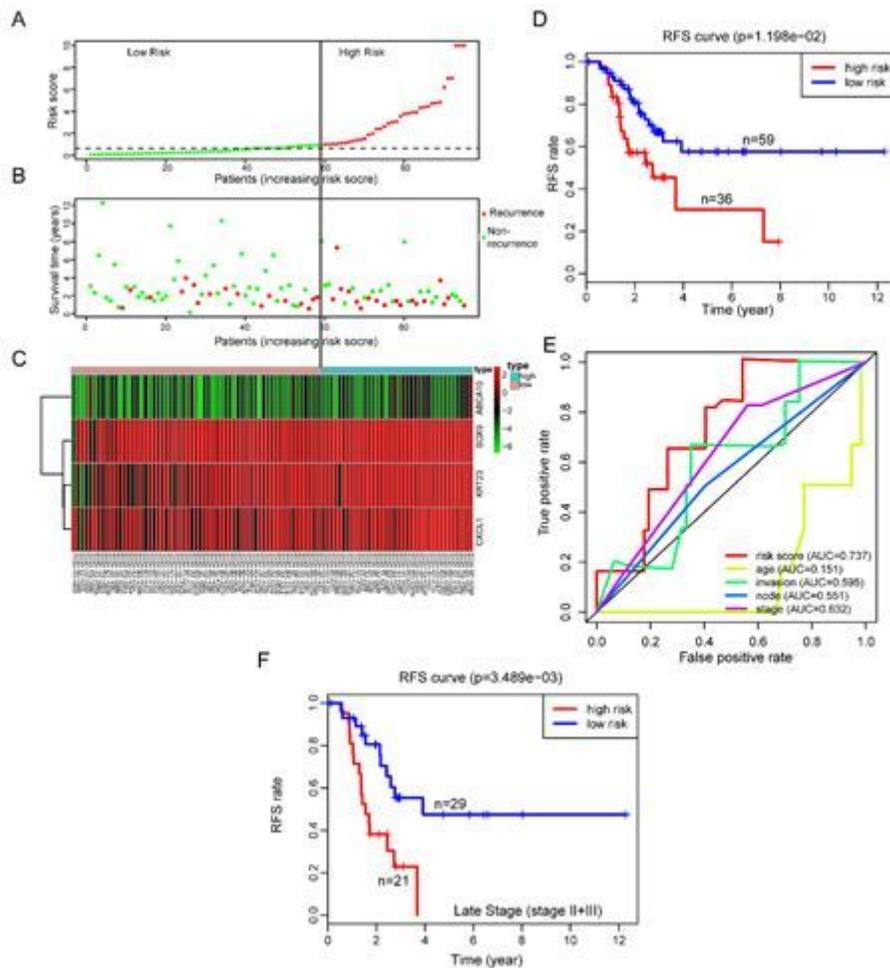


Figure 4

Validation of the 4-gene signature in predicting recurrence-free survival (RFS) in the TCGA dataset. (A-C) The distribution of risk score, RFS, and survival status and the 4 genes expression patterns for the 95 patients with RFS data. (D) Kaplan–Meier analysis compares RFS between patients in the high- and low-risk group. (E) ROC analysis of the 4-gene signature and other clinicopathological parameters (age, invasion, node metastasis and stage) for prediction RFS. (F) Kaplan-Meier analysis compares RFS between patients in the high- and low-risk group in late stage (III+IV) patients.

Supplementary Files

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

- [supplementary.docx](#)