Identification of a metabolism-related gene expression prognostic model in endometrial carcinoma patients

Pinping Jiang  
Jiangsu Province Hospital and Nanjing Medical University First Affiliated Hospital

Wei Sun  
Jiangsu Province Hospital and Nanjing Medical University First Affiliated Hospital

Ningmei Shen  
Jiangsu Province Hospital and Nanjing Medical University First Affiliated Hospital

Xiaohao Huang  
Jiangsu Province Hospital and Nanjing Medical University First Affiliated Hospital

Shilong Fu (✉ docfusl@163.com)  
Jiangsu Province Hospital and Nanjing Medical University First Affiliated Hospital

https://orcid.org/0000-0001-7416-3070

Research article

Keywords: Metabolism; TCGA; endometrial carcinoma; prognostic model; nomogram.

Posted Date: June 29th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-35652/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 September 7th, 2020. See the published version at https://doi.org/10.1186/s12885-020-07345-8.
Abstract

**Background:** Metabolic abnormalities have recently been widely studied in various cancer types. This study aims to explore the expression profiles of metabolism-related genes (MRGs) in endometrial cancer (EC).

**Methods:** We analyzed the expression of MRGs using The Cancer Genome Atlas (TCGA) data to screen differentially expressed MRGs (DE-MRGs) significantly correlated to EC patients’ prognosis. Functional pathway enrichment analysis of DE-MRGs were investigated. LASSO algorithm and Cox regression analysis were performed to select MRGs closely related to EC patients’ outcomes. A prognostic signature was developed and the efficacy were validated in part of and the entire TCGA EC cohort. Moreover, we developed a comprehensive nomogram including the risk model and clinical features to predict EC patients’ survival probability.

**Results:** Forty-seven differentially expressed MRGs (DE-MRGs) were significantly correlate to EC patients’ prognosis. Functional enrichment analysis showed these MRGs were highly enriched in amino acid, glycolysis, and glycerophospholipid metabolism. Nine MRGs were screened out to closely relate to EC patients’ outcomes, which are CYP4F3, CEL, GPAT3, LYPLA2, HNMT, PHGDH, CKM, UCK2 and ACACB. Based on nine DE-MRGs, we developed a prognostic signature and its efficacy in part of and the entire TCGA EC cohort was validated. The nine-MRGs signature was independent of other clinical features, and could effectively distinguish high- or low-risk EC patients and predicted patients’ OS. The nomogram showed excellent consistency between prediction and actual survival observation.

**Conclusions:** The MRG prognostic model and the comprehensive nomogram could guide for precise outcome predicting and rational therapy in clinical practice.

1. **Background**

Endometrial carcinoma (EC), one of the most common female reproductive malignancies, induced nearly 90,000 deaths worldwide each year [1]. Women with metabolic disorders, including obesity and diabetes, have a remarkably increased risk of developing endometrial cancer. While early-stage endometrial cancer predicts favorable prognosis, nearly 30% of patients still diagnosed at a late stage, and over 80% of these individuals would die in 5 years [2]. Besides, several EC patients presented a high risk of cancer progression or recurrence with insensitivity to chemotherapy, which indicated the poor outcomes [3]. Therefore, it is imperative to lay stress on molecular changes of endometrial cancer progression and develop novel predictive biomarkers to accurately estimate patients’ outcomes.

Since fundamental metabolic differences between cancer and adjacent normal cells were firstly uncovered, metabolic reprogramming has increasingly become a hot topic for cancer biology [4]. The metabolic phenotype of cancer cells is heterogeneous in various cancer types, for example, while several malignant tumors mainly rely on glycolysis, others presented a metabolic phenotype mediated by oxidative phosphorylation [5, 6]. In total, through reprogramming tumor microenvironments, catabolic and
anabolic metabolism is essential for cancer cells to sustain energy supply and biomass synthesis [7-9]. While the underlying processes and molecular alterations of metabolic programming in various cancers have been well elucidated, the expression patterns of metabolism-related genes in endometrial cancer are still unclear.

In this study, we focused on the metabolism-related genes expression alterations of the Cancer Genome Atlas (TCGA) EC patients and obtained prognostic dysregulate MRGs. Besides, we established and validated a multiple MRGs-combined expression signature predicting EC patients' outcomes. Moreover, we integrated the clinical features of patients and the MRGs model to establish a novel nomogram model that could guide for comprehensive EC therapeutic guidelines.

2. Methods

2.1 Gene expression profiles and clinical information integration

We downloaded the FPKM format mRNA expression profiles of EC from The Cancer Genome Atlas (TCGA) database (https://tcga-data.nci.nih.gov/tcga/), which contains a total of 541 cases. The corresponding clinical information was retrieved from the cBio Cancer Genomics Portal (cbioPortal, http://cbioportal.org)[10].

2.2 Extraction of metabolism-related genes from the TCGA database

Genes enriched in metabolism pathways of the KEGG database were utilized in this study as metabolic genes. The mRNA expression of metabolic genes in the TCGA database was extracted.

2.3 Identification of prognostic associated differentially expressed metabolic genes (DE-MRGs)

With the cut-off criteria as |logFC| > 1 and P-value < 0.05, we screened the DE-MRGs via "limma"R package[11]. Then, univariate Cox regression analysis was performed to identify prognostic associated DE-MRGs. Hazard Ration (HR) < 1 presents better overall survival outcomes (OS) while HR > 1 means worse OS. Genes with P < 0.05 were regarded as prognostic associated metabolic genes. The expression level of prognostic associated DE-MRGs from each patient and between cancerous and normal samples was displayed via "pheatmap" and "ggplot" R package, respectively.

2.4 Functional enrichment analysis of prognostic associated DE-MRGs

Gene ontology (GO) [12] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [13] pathway enrichment analysis were performed to explore the biological functions of such prognostic related DE-MRGs via "clusterprofiler" R package. Adjust p-value < 0.05 was set as statistically significant, and the enrichment analysis result maps were present by “ggplot2” and “GOplot” R packages.

2.5 Protein-protein interaction (PPI) network construction and hub DE-MGRs alteration analysis
The Search Tool for the Retrieval of Interacting Genes Database database (STRING, https://string-db.org/) comprised the interaction information among given proteins [14]. Based on the minimum required interaction score setting as 0.4, we utilized the STRING database to construct the PPI network which reflected the interactions among DE-MRGs. The network was present by Cytoscape software and the top 15 hub genes were selected based on the connectivity degree in string network [15]. In addition, the alteration landscapes of hub DE-MRGs in EC were visualized by cbioPortal.

2.6 Establishment of a prognostic model based on DE-MRGs

We randomly classified the total TCGA EC patients into the training and the testing cohort. The Least Absolute Shrinkage and Selection Operator (LASSO) and multivariate Cox regression analysis were then performed to selected key prognostic related DE-MRGs via "glmnet" R package. The formula of the risk score for EC patients’ prognosis prediction was as follows: risk score = the sum of each multivariate cox regression coefficient ratio of mRNA multiple each expression of mRNA. Based on the median risk score, we divided training cohort patients into high- and low-risk subgroups. Each patients’ survival status, death time, and gene expression profile in two subgroups were presented via "pheatmap" and "survival" R packages. Besides, the Kaplan-Meier curve analysis was performed and the receiver operating characteristic (ROC) curve was drawn to estimate the sensitivity and specificity of the prognostic signature.

2.7 Validation the efficacy of the prognostic DE-MRGs signature

The prognostic signature was then introduced into the testing cohort and the entire cohort. Based on the median risk score from the training cohort, patients in the testing and entire cohort were separated into high- or low-risk individuals, likewise. Kaplan-Meier curve analysis, time-dependent ROC analysis, as well as patients’ outcomes distribution were performed, either

2.8 RNA extraction of clinical samples and quantitative real-time RT-PCR (qRT-PCR) analysis.

A total of 30 RNA later-reserved EC specimens were collected from patients who underwent surgery at Jiangsu Province Women and Children Health Hospital (Nanjing, China) between September 2017 and September 2019. All samples were immediately snap frozen in liquid nitrogen and stored at -80°C until required. Total RNA was isolated using Trizol reagent (Invitrogen) from fresh-frozen tissues and transcribed into cDNA using TaqMan Reverse Transcription kit (Applied Biosystems) with random hexamer primers. QRT–PCRs were performed using 2×SYBR Green qPCR Master Mix (Selleck, Shanghai, China). The housekeeping genes GAPDH were used for normalization of qRT–PCR data before calculation using the $\Delta \Delta Ct$ method. The primers used are listed in Supplementary Table 1.

2.9 Evaluation of clinical independency and nomogram building

Next, we cut out the EC patients who lack detailed clinicopathological information. The clinicopathological characters and the MRGs expression data of remaining patients were compared and comprehensively displayed in the heatmap between high- and low-risk subgroups. Meanwhile, the clinical
indexes and risk scores were included in univariate and multivariate Cox regression analysis to validate the independence of the risk model. ROC curve regarding the signature and other clinical features was used to assess the predictive efficacy of the model. Besides, the correlation between MRGs from the risk model and the clinical index was also measured. Finally, we utilized the “rms” R package to consolidated the risk score and clinical characteristics for nomogram construction.

3. Results

3.1 Identification of a list of prognostic related DE-MRGs

The detailed flow chart for the prognostic predictive model construction in this study was shown in Figure 1. From the metabolism pathways of KEGG database, we extracted a total of 944 metabolism-related genes. After the integration of expression data regarding such MRGs expression from 552 TCGA EC cancerous and 35 non-tumor samples, we obtained 156 up-regulated and 64 down-regulated MRGs. Univariate Cox regression analysis identified 47 genes significantly correlated to EC patients’ OS (Figure 2B and Table 1). The expression pattern of 47 prognostic related MRGs was shown in the heatmap and box plot (Figure 2C-D).

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Univariate Cox regression identified 20 MR-DEGs correlated to endometrial cancer patients' OS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene ID</td>
<td>HR</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
</tr>
<tr>
<td>CYP4F3</td>
<td>1.19013465</td>
</tr>
<tr>
<td>PHGDH</td>
<td>1.02042871</td>
</tr>
<tr>
<td>GPX2</td>
<td>1.002979902</td>
</tr>
<tr>
<td>CKMT1B</td>
<td>1.223457962</td>
</tr>
<tr>
<td>CEL</td>
<td>1.008855015</td>
</tr>
<tr>
<td>DDC</td>
<td>1.098684916</td>
</tr>
<tr>
<td>KHK</td>
<td>1.189854228</td>
</tr>
<tr>
<td>LYPLA2</td>
<td>1.019470527</td>
</tr>
<tr>
<td>ACACB</td>
<td>1.834694401</td>
</tr>
<tr>
<td>ASNS</td>
<td>1.052124561</td>
</tr>
<tr>
<td>GPAT3</td>
<td>1.104677229</td>
</tr>
<tr>
<td>CKMT1A</td>
<td>1.256355384</td>
</tr>
<tr>
<td>CKM</td>
<td>1.141628713</td>
</tr>
<tr>
<td>UCK2</td>
<td>1.035172213</td>
</tr>
<tr>
<td>HNMT</td>
<td>0.861666359</td>
</tr>
<tr>
<td>SAT1</td>
<td>0.997992548</td>
</tr>
<tr>
<td>B4GALT1</td>
<td>0.995107725</td>
</tr>
<tr>
<td>GLB1</td>
<td>0.971544079</td>
</tr>
<tr>
<td>PSAT1</td>
<td>1.007191289</td>
</tr>
<tr>
<td>PFKM</td>
<td>1.087927137</td>
</tr>
</tbody>
</table>

Abbreviation: MR-DEGs: metabolism-related differentially expressed genes; OS: overall survival; HR: Hazard ratio.

3.2 Functional enrichment of the prognostic related DE-MRGs
Function annotation analysis regarding the 47 DE-ARGs was then performed. GO enrichment showed that the MRGs were mainly involved in "carboxylic acid biosynthetic process", "cofactor binding", "alpha-amino acid metabolic process" and "cellular amino acid metabolic process" (Figure 3A). The enriched GO terms and related genes expression profile were presented in Figure 3B. KEGG pathway enrichment presented that DE-MRGs were mainly involved in "Purine metabolism", "Biosynthesis of amino acids", "Glycolysis / Gluconeogenesis" and "Glycerophospholipid metabolism" (Figure 4A). The expression level of correlated genes in the enriched KEGG pathways are displayed in the heatmap (Figure 4B).

### 3.3 The PPI network of 47 prognostic related DE-MRGs and hub gene alteration analysis

Through the STRING website, we built a PPI network among 47 DE-MRGs (Supplementary figure 1). There were 47 nodes and 81 edges included in the network based on the interaction score criteria. Top 15 hub genes with the highest connectivity degrees in the string network were selected as follows: ALDH18A1, GOT1, DGAT2, GPD1, GPI, PHGDH, GPT, ENTPD3, ASNS, ADCY9, CKMT1A, LPCAT1, MTHFD2, PPAP2C and PSAT1 (Figure 5A-B). The alteration results of the hub genes showed that GPI, GPT, ADCY9, and LPCAT1 ranked as the most frequently altered genes. GPI, GPT, and LPCAT1 were frequently over-amplified among endometrial cancer patients while ADCY9 more often occurred with a missense mutation (unknown significance) (Figure 5C).

### 3.4 Identification of a nine DE-MRGs based prognostic model

Next, we randomly divided the 541 TCGA EC patients into the training cohort (n = 272) and the testing cohort (n = 269). Lasso and multivariate cox regression analysis picked out 9 genes significantly associated with prognosis, which are CYP4F3, CEL, GPAT3, LYPLA2, HNMT, PHGDH, CKM, UCK2 and ACACB (Figure 6 and Table 2). According to the results of multivariate Cox regression analysis, we constructed the prognostic model as follows: risk score = (0.110103 × expression value of CYP4F3) + (0.013456 × expression value of CEL) + (0.104444 × expression value of GPAT3) + (0.017777 × expression value of LYPLA2) + (-0.10986 × expression value of HNMT) + (0.017183 × expression value of PHGDH) + (0.200964 × expression value of CKM) + (0.051913 × expression value of UCK2) + (0.634313 × expression value of ACACB).

Based on the mean risk score from the training cohort, patients were divided into high-risk (n=136) and low-risk (n=136) subgroups. Each individuals' risk score distribution and survival status were ranked and displayed on the dot plot, which showed a significant discrepant OS (Figure 7A-B). Likewise, the Kaplan-Meier curve analysis demonstrated that the OS of the higher-risk group was significantly shorter than the low-risk group (P =1.971e-06) (Figure 7D). The expression profile of 9 prognostic DE-MRGs showed that UCK2, PHGDH, ACACB, LYPLA2, CYP3F3, GPAT3, CEL, and CKM expressed higher in the high-risk subgroup, while HNMT presented otherwise (Figure 7C). ROC curve analysis revealed the area under the ROC curve (AUC) of the prognostic MRGs model was 0.771 (Figure 7E).
Table 2: Multivariate Cox regression selected 9 MR-DEGs correlated to endometrial cancer patients' OS

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>HR</th>
<th>HR.95L</th>
<th>HR.95H</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP4F3</td>
<td>1.1163</td>
<td>0.9935</td>
<td>1.2544</td>
<td>0.0642</td>
</tr>
<tr>
<td>CEL</td>
<td>1.0135</td>
<td>1.0078</td>
<td>1.0193</td>
<td>3.6977</td>
</tr>
<tr>
<td>GPAT3</td>
<td>1.1100</td>
<td>1.0064</td>
<td>1.2243</td>
<td>0.0367</td>
</tr>
<tr>
<td>LYPLA2</td>
<td>1.0179</td>
<td>1.0039</td>
<td>1.0321</td>
<td>0.0118</td>
</tr>
<tr>
<td>HNMT</td>
<td>0.8959</td>
<td>0.7799</td>
<td>1.0293</td>
<td>0.1206</td>
</tr>
<tr>
<td>PHGDH</td>
<td>1.0173</td>
<td>1.0033</td>
<td>1.0315</td>
<td>0.0152</td>
</tr>
<tr>
<td>CKM</td>
<td>1.2226</td>
<td>1.0979</td>
<td>1.3618</td>
<td>0.0002</td>
</tr>
<tr>
<td>UCK2</td>
<td>1.0533</td>
<td>1.0240</td>
<td>1.0833</td>
<td>0.0002</td>
</tr>
<tr>
<td>ACACB</td>
<td>1.8857</td>
<td>1.1297</td>
<td>3.1475</td>
<td>0.0152</td>
</tr>
</tbody>
</table>

Abbreviation: MR-DEGs: metabolism-related differentially expressed genes; OS: overall survival; HR: Hazard ratio.

3.5 Validation the efficacy of the 9 MRGs prognostic signature

The risk model was then introduced into the testing and entire cohort, and each individuals' risk score was calculated. Based on the training cohort' cut-off risk score, we divided the testing cohort into 121 high-risk and 148 low-risk individuals. Following the results from the training cohort, the survival status (Figure 8A), survival time (Figure 8B) and KM curve analysis of high-risk subgroup presented worse outcomes compared to the low-risk subgroup in the testing cohort with shorter overall survival time (P =4.151e-04) (Figure 8D). The expression pattern of 9 MRGs was consistent with the training cohort (Figure 8C) and the AUC of the risk model in the testing cohort was 0.796 (Figure 8E). As for the entire cohort, similar results were also observed. Low-risk subgroup presented longer survival time and better survival status (Figure 9A-B). Kaplan-Meier curve analysis presented that low-risk subgroup followed with longer OS (P =1.242e-08) (Figure 9D) and ROC curve analysis showed the AUC of the model was 0.78 (Figure 9E).

3.6 Validation the expression level of 9 MRGs in clinical samples

The expression signature of 9 MRGs were subsequently explored in 30 endometrial cancer clinical specimens. Results demonstrated that UCK2, PHGDH, ACACB, LYPLA2, CYP3F3, GPAT3, CEL, and CKM overexpressed in cancerous tissues while HNMT showed otherwise, which were in accordance with above findings (Figure 10A-I).

3.7 The clinical independence and correlation estimation of the risk signature
Then, we combined the risk model with other clinical factors and performed univariate and multivariate analyses to examine the clinical independency of the model. Results showed that the model was able to serve as an independent prognostic indicator (both P<0.001) (Figure 11A-B). The AUC value of the prognostic model was 0.781, significantly higher than patients’ age (AUC= 0.535) and weight (AUC= 0.633), clinical-stage (AUC= 0.710), tumor grade (AUC= 0.656) and histology (AUC= 0.522), as well as lymph node status (AUC= 0.697) (Figure 11C). Next, we enrolled in the risk scores, the clinical features, and the nine gene expression profiles of EC patients and displayed them in the heatmap (Figure 11D). Interestingly, the clinical characteristics of patients were highly in accordance with the risk level calculated from the model. High-risk subgroup patients were accompanied by late-stage, high grade, serous carcinoma, and more metastatic lymph nodes (Figure 11D-E), which all presented worse outcomes. The correlation between each gene from the prognostic model and the patients’ clinical features were also measured. PHGDH, ACACB, HNMT, CYP4F3, and LYPLA2 were shown to be significantly associated with patients’ prognosis (Figure 11F). Other clinical features of each prognostic MRGs from the signature were presented in Figure S3.

3.8 Nomogram building and validation

Based on patients’ risk scores and clinical features, we built a comprehensive prognostic nomogram to estimate EC patients’ survival probability for 5 years based on the TCGA entire set. Seven independent prognostic parameters, including metabolic risk signature and age, grade, weight, histology, stage as well as lymph node status, were integrated into the nomogram (Figure 12A). The calibration plots show excellent consistency between the nomogram prediction and actual observation in terms of the 3- and 5-year survival rates in the TCGA cohort (Figure 12B-C).

4. Discussion

Metabolic abnormalities have recently been widely studied to serve as an important role in tumor development in various cancer types. The metabolism dysfunction in the tumor microenvironment could lead to completely diverse outcomes of patients, and metabolism-related genes can be used as prognostic markers of tumors. In this work, we thoroughly probed into the implications of metabolism-related genes in endometrial cancer progress. By analyzing the mRNA data of TCGA EC patients, we obtained 220 dysregulated MRGs, among which 47 were associated with EC patients’ OS. Functional enrichment analysis of these prognostic MRGs showed that they were closely associated with cellular amino acid, glycolysis, and glycerophospholipid metabolism. In accordance with our observation, Byrne et.al have also found that glycolysis and lipogenesis highly associated with the endometrial cancer phenotypes and the suppression of GLUT6 expression could inhibit glycolysis and survival of EC cells, underlying the crucial role of energy metabolism in tumor progression[16]. In addition to that, our results further uncovered the exact dysregulated metabolic genes of these disordered metabolism-related pathways, which may provide a new perspective on the molecular mechanism of metabolism alteration in tumor progression.
Metabolic prognostic risk signature that combined multiple metabolism-related gene expressions has been proved to serve as a powerful prognostic indicator in various malignant diseases, such as glioma, liver cancer, ovarian cancer, papillary thyroid carcinoma, etc. Zhou et al. identified a 29-energy metabolism-related gene signature, containing branched-chain amino acid transaminase 1 (BCAT1), interleukin-4 and carbohydrate sulfotransferases, to evaluate the prognosis of diffuse glioma [17]. Wang et al. enrolled 6 risks and 2 protective metabolic genes into the prognostic metabolic model which effectively predicted ovarian cancer patients' prognosis [18]. Likewise, Ma et al. developed a metabolic gene signature as a biomarker for dedifferentiated thyroid cancer[19], and Liu et al. built a four-metabolic gene signature for liver cancer patients' outcomes prediction[20].

In the present study, we performed lasso and multivariate cox regression analysis and identified a nine-gene signature including CYP4F3, CEL, GPAT3, LYPLA2, HNMT, PHGDH, CKM, UCK2, and ACACB. Among them, HNMT was considered a protective factor while others were risk factors. The diagnostic and predictive effectiveness of such prognostic genes has already been reported in other studies. Cui et al. reported a significant higher expression of Carboxyl EsterLipase (CEL) in breast cancer. The combination of CEL and other biomarkers could improve breast cancer diagnostic capability [21]. Likewise, Richard et al. found that over 70% of estrogen receptor (ER)-negative breast cancers exhibited elevated phosphoglycerate dehydrogenase (PHGDH) protein expression which is crucial for promoting the serine pathway flux[22]. In addition, Zhang et al. put forward a discovery that PHGDH could define a metabolic subtype in lung adenocarcinomas with unique metabolic dependencies [23]. In pancreatic ductal adenocarcinoma, CYP4F3, one isoform of the cytochrome P450 (CYP) superfamily, was shown to upregulated in tumor tissues and could serve as a distinguishing marker [24]. Uridine-cytidine kinase 2(UCK2) was positively correlated with early recurrence and poor prognosis in hepatocellular carcinoma. Overexpression of UCK2 increased the MMP2/9 expression and further activated the Stat3 signaling, mediated the metastasis of hepatocellular carcinoma cells [25]. As for ACAVB, Lally et al. present that humans with fatal HCC subtypes have increased acetyl-CoA carboxylase (ACC) expression and the genetic activation of ACC promoted the formation of hepatic de novo lipid and induced subsequent liver carcinogenesis [26].

Here, through bioinformatic analysis and outside validation, we innovatively reported these metabolic genes closely related to the prognosis of EC patients. In addition, the metabolic risk signature combining such genes could accurately categorize EC patients into high- or low-risk subgroup which represented patients' long-term outcomes. Last but not least, in our study, this is the first time to build a comprehensive nomogram that incorporated a metabolic-related signature with clinical features including age, stage, tumor grade and lymph node status to effectively predict the survival of EC patients. This prognostic scoring system would provide a precise method to help both physicians and patients to perform individualized survival evaluation and treatment options selection.

Conclusion
In conclusion, we identified 47 prognostic dysregulated metabolic genes in EC. The prognostic DE-MRGs were highly associated with amino acid, glycolysis, and glycerophospholipid metabolism. Top 15 hub genes in the PPI network were also selected and analyzed. We also performed lasso and multivariate cox regression to establish and validate a robust prognostic risk signature enrolling nine dysregulated MRGs. Besides, a comprehensive nomogram that combined clinical characters and the risk model was constructed and its efficacy in predicting EC patients' prognosis was also demonstrated. The 9 MRGs model and nomogram may guide rational therapy for doctors in clinical practice.

**Abbreviations**

EC: endometrial cancer; MRGs: metabolism-related genes; TCGA: The Cancer Genome Atlas; GO: gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; PPI: protein-protein interaction; STRING: Search Tool for the Retrieval of Interacting Genes Database; ROC: Receiver operating characteristic; AUC: area under the curve; LASSO: Least Absolute Shrinkage and Selection Operator; OS: overall survival; FC: fold change.

**Declarations**

**Ethics approval and consent to participate**

Ethical consent was approved by the Committees for Ethical Review of Research involving Human Subjects at Nanjing Medical University.

**Consent to participate**

Written informed consent was obtained from each patient before sample collection.

**Availability of data and material**

The expression data were deposited in the TCGA database and the clinical information was retrieved from the cBioPortal website. Besides, please contact the author for data and materials requests.

**Competing interests**

The authors declare that they have no competing interests.

**Fundings**

Not applicable.

**Authors' contributions**

Shilong Fu and Pinping Jiang designed the project. Wei Sun and Ningmei Shen contributed to data analysis and prepared the main manuscript. Xiaohao Huang and Pinping Jiang revised and submitted
the manuscript. All authors reviewed the manuscript.

Acknowledgements

Not applicable.

References


Supplementary Figure Legends

Supplementary Figure 1. GO and KEGG pathway enrichment of 220 DE-MRGs.

Supplementary Figure 2. The protein-protein network of 220 DE-MRGs.

Supplementary Figure 3. Clinical characteristics of each prognostic MRGs from the signature.

Figures

Figure 1

The flow chart of the analysis procedure in identifying a metabolism-related prognostic signature.
Figure 2

The expression profiles of prognosis associated differentially expressed metabolism-related genes (DE-MRGs) between TCGA endometrial cancer (EC) and normal tissues. (A) Volcano plot of DE-ARGs in EC and normal samples of the TCGA dataset. The vertical axis indicates the -log(P-value), and the horizontal axis indicates the log2 (fold change [FC]). The red dots and the green dots represent up- and down-regulated genes, respectively (P-value<0.05 and |log2(FC)|>1). (B) Univariate Cox regression identified 47
DE-MRGs correlated to EC patients’ outcomes; (C) Heat map of the 47 DE-MRGs in the entire CGA EC cohort. Red and green indicate higher expression and lower expression, respectively. (D) Box plot of the expression of the DE-MRGs between cancerous and normal tissues.

Figure 3

Gene ontology (GO) functional enrichment of prognostic differentially expressed MRGs. (A) GO analysis shows the biological processes, cellular component, and molecular functions involved in differential
genes. (B) Heatmap of the expression of DEGs in the enriched GO items.

Figure 4

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enrichment of prognostic differentially expressed MRGs. (A) KEGG analysis shows significantly enriched pathways of DE-MRGs. The node color changes gradually from red to blue in descending order according to the adjusted P-values. The size of the node represents the number of counts. (B) Circle plot of the enriched DEGs in the KEGG items.
Figure 5

Protein-protein network of hub DE-MRGs and alteration analysis. (A) The highest degree of hub genes was ranked; (B) The interaction network of the top 15 hub genes; (C) The gene mutation overview of 15 hub genes in TCGA EC patients.
Figure 6

Subsequent identification of prognosis related DE-MRGs using LASSO and cox regression analysis. (A) Plots of the cross-validation error rates. Each dot represents a lambda value along with error bars to give a confidence interval for the cross-validated error rate; (B) LASSO coefficient profiles of the MRGs associated with the overall survival of endometrial cancer; (C) multivariate cox regression identified 9 prognostic MRGs in the training cohort.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Hazard ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP4F3</td>
<td>$1.1$</td>
<td>0.054</td>
</tr>
<tr>
<td>CEL</td>
<td>$1.0$</td>
<td>&lt;0.001 ***</td>
</tr>
<tr>
<td>GPAT3</td>
<td>$1.1$</td>
<td>0.037 *</td>
</tr>
<tr>
<td>LYPLA2</td>
<td>$1.0$</td>
<td>0.012 *</td>
</tr>
<tr>
<td>HNMT</td>
<td>$0.9$</td>
<td>0.121</td>
</tr>
<tr>
<td>PHGDH</td>
<td>$1.0$</td>
<td>0.015 *</td>
</tr>
<tr>
<td>CKM</td>
<td>$1.2$</td>
<td>&lt;0.001 ***</td>
</tr>
<tr>
<td>UCK2</td>
<td>$1.0$</td>
<td>&lt;0.001 ***</td>
</tr>
<tr>
<td>ACACB</td>
<td>$1.0$</td>
<td>0.015 *</td>
</tr>
</tbody>
</table>

*Events: 38; Global p-value (Log-Rank): 2.1498e-06
AIC: 322.24; Concordance Index: 0.8
Figure 7
Prognostic analysis of the model in the TCGA training cohort. (A) The risk score, (B) survival status, (C) expression heatmap, (D) Kaplan-Meier survival, and (E) time-dependent ROC curves of the prognostic model for the TCGA EC training cohort.
Figure 8

Validation of the efficacy of the risk signature in the TCGA testing cohort. (A) The risk score, (B) survival status, (C) expression heatmap, (D) Kaplan-Meier survival, and (E) time-dependent ROC curves of the prognostic model for the TCGA EC testing cohort.
Figure 9

Estimation of the efficacy of the risk signature in TCGA entire EC cohort. (A) The risk score, (B) survival status, (C) expression heatmap, (D) Kaplan-Meier survival, and (E) time-dependent ROC curves of the prognostic model for the TCGA EC entire cohort.
Figure 10

Validation of the expression signature of 9 MRGs by qRT–PCR
Figure 11

Clinical characteristics of the prognostic MRGs signature. Univariate (A) and multivariate (B) regression analysis, as well as time-dependent ROC curve analysis (C) of the prognostic value between the risk model and EC patients’ OS status when compared to or combined with clinical factors; (D) Heat map showing the expression of 9 MRGs in the risk model and the clinicopathological features of patients with EC; (E) Clinicopathological significance of the prognostic signature of endometrial cancer; (F) Kaplan-Meier curve plot of prognostic MRGs from the signature.
Figure 12

Nomogram for predicting the 5-year survival probability of patients with EC. (A) Prognostic nomogram for EC patients; (B) Calibration curves for the nomogram at 3-, and 5-year.

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

- SupplementaryTable1.pdf
- FigureS1.png
- FigureS2.png
- FigureS3.png