

# Development of a 4-miRNA prognostic signature for endometrial cancer

Jiazhen Huang, MM<sup>a</sup>, Furong Du, MM<sup>b</sup>, Ning Wang, MD<sup>a,\*</sup> 

## Abstract

To develop an effective uterine corpus endometrial carcinoma (UCEC) risk assessment tool to monitor treatment outcomes. Limma package was used to analyze differentially expressed microRNAs (miRNAs) between UCEC tissues and normal tissues in the TCGA database. According to univariate Cox risk regression, least absolute shrinkage, and selection operator (LASSO) Cox analysis were performed to screen prognostic miRNAs and construct a risk scoring model. The prognostic performance of signature was evaluated by Kaplan–Meier and receiver operating characteristic. Multivariate Cox regression analysis was used to determine the independent prognostic factors of UCEC. Nomogram was constructed according to age, clinical stage, and risk score. A 4-miRNA signature based on miR-31-5p, miR-34a-5p, miR-26a-1-3p and miR-4772-3p was established. Risk scores of each patient were calculated by the 4-miRNA signature. After z-score, the patients were divided into high- and low-risk groups. The overall survival of high-risk patients was significantly shorter than that of low-risk patients, pointing to the high performance and independence of the 4-miRNA signature in predicting UCEC prognosis. The nomogram showed a high accuracy in predicting overall survival of UCEC patients. We developed a 4-miRNA signature that could effectively predict the prognosis of UCEC.

**Abbreviations:** AIC = Akaike information criterion, AUC = the area under the curve, DCA = decision curve analysis, DFS = disease free survival, DSS = disease specific survival, LASSO = least absolute shrinkage, and selection operator, miRNAs = microRNAs, OS = overall survival, PFS = progression free survival, ROC = receiver operating characteristic, TCGA = The Cancer Genome Atlas, UCEC = uterine corpus endometrial carcinoma.

**Keywords:** miRNA, prognosis, risk score, uterine corpus endometrial carcinoma

## 1. Introduction

Uterine corpus endometrial carcinoma (UCEC), which is 1 of the most common malignant tumors in female reproductive tract, is a heterogeneous disease involving a variety of tissue types, each of which constitutes a disease entity.<sup>[1,2]</sup> The most common histological subtype is endometrioid adenocarcinoma.<sup>[3]</sup> The standard treatment for UCEC is total hysterectomy and bilateral salpingectomy with or without pelvic and para-aortic lymph node dissection.<sup>[4]</sup> Although the chance of surgical cure is relatively high, there are few appropriate treatment options available for patients with advanced or postoperative recurrence.<sup>[5]</sup> Developing effective risk assessment tools for UCEC patients helps monitor UCEC patients with high risk of recurrence and metastasis.

The rapid development of whole genome technology has greatly promoted tumor biomarker research, leading to the discovery of a large number of biomarkers related to tumor progression, but many reports have only focused on the study of individual genes.<sup>[6,7]</sup> Due to the genetic heterogeneity of UCEC,

data about the efficacy and availability of biomarkers are limited.<sup>[8]</sup> Several studies indicated that polygenic systems are more accurate than using a single gene in the assessment of tumor risk.<sup>[9,10]</sup> At present, researchers have identified polygenic signatures related to prognosis of different types of cancer. In hepatocellular carcinoma, a 4-gene prognostic model based on CENPA, SPP1, MAGEB6 and HOXD9 accurately predicts the survival of patients with hepatocellular carcinoma.<sup>[11]</sup> Another study established a model based on 4 immune-related genes, the signature effectively predicts dedifferentiation and immune exhaustion of thyroid cancer.<sup>[12]</sup> Jun et al developed a 6-gene signature for evaluating the recurrence-free survival and castration resistance of prostate cancer, and the signature is generally applicable in most clinical practice.<sup>[13]</sup> A retrospective study of gynecological cancer summarized the importance of microRNAs (miRNAs) in evaluating the prognosis of gynecological malignant tumors (ovarian cancer, cervical cancer and endometrial cancer), and pointed out that miRNA signature can be used as a biomarker for the diagnosis and prognosis of gynecological cancer.<sup>[14]</sup> So

This work is supported by Scientific Research Funding of Liaoning Education Department (2020) [grant number: LZ2020063].

The authors have no consent to disclose.

The authors have no conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are publicly available. The data that support the findings of this study are available from the first author and corresponding author upon reasonable request.

The data used in our study were accessed from the TCGA database, a freely available database, thus ethical approval of The Second Hospital of Dalian Medical University was unnecessary for our study.

<sup>a</sup> Department of Obstetrics and Gynecology, The Second Hospital of Dalian Medical University, Dalian, Liaoning, China, <sup>b</sup> State Development of Key Laboratory of Translational Medicine and Innovative Drug, Jiangsu Simcere Diagnostics Co., Ltd., Nanjing, Jiangsu, China.

\*Correspondence: Ning Wang, Department of Obstetrics and Gynecology, The Second Hospital of Dalian Medical University, Dalian 116000, Liaoning, China (e-mail: nonaware@sina.com).

Copyright © 2022 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Huang J, Du F, Wang N. Development of a 4-miRNA prognostic signature for endometrial cancer. *Medicine* 2022;101:41(e30974).

Received: 15 July 2022 / Received in final form: 2 September 2022 / Accepted: 6 September 2022

<http://dx.doi.org/10.1097/MD.00000000000030974>

far, there have been few studies investigating the prognostic miRNA signature of endometrial carcinoma, which undoubtedly requires further development and research.

In this study, based on the clinical information and miRNA expression data of UCEC patients acquired from The Cancer Genome Atlas (TCGA), we constructed a miRNA-based signature capable of predicting the prognosis of UCEC patients, and evaluated the accuracy of the model in the test set, verification set and the whole data set. The 4-miRNA signature provides molecular insights into the risk assessment of UCEC patients and contributes to the clinical treatment.

## 2. Materials and Methods

### 2.1. Data acquisition and processing

The clinical follow-up information of UCEC patients and miRNA expression data in tumor tissues were downloaded from TCGA database (<https://portal.gdc.cancer.gov/>). After excluding the samples without clinical follow-up information, survival time or status, a total of 569 samples were retained, of which 536 were tumor samples and 33 were normal samples. The sample clinical information statistics can be found in Table 1.

The data used in our study were accessed from the TCGA database, a freely available database, thus ethical approval of The Second Hospital of Dalian Medical University was unnecessary for our study.

### 2.2. Differential miRNA expression analysis of UCEC

Limma package was used to analyze the differentially expressed miRNAs between normal samples and UCEC tumor samples.<sup>[15]</sup> The threshold was set to false discovery rate < 0.05 and  $|\log_2(\text{fold change (FC)})| > 1.5$ , miRNAs meeting the standard were defined as differentially expressed miRNAs.

### 2.3. Patient grouping

A total of 536 samples in TCGA data set were divided into training set and verification set. To avoid random assignment deviation in affecting the stability of subsequent modeling, all the samples were randomly grouped 100 times in advance, in which the proportion of grouping sampling was training set: verification set = 1:1. The most suitable training set and verification set were selected according to the criteria of (a) a similar distribution of age, sex, follow-up time and proportion of death between the 2 groups, and (b) similar number of samples in the 2 randomly grouped data sets after gene expression profile clustering. Based on the criteria, 268 samples were assigned to the training set and the remaining 268 samples were in the verification set. The sample information of training set and test set was shown in Table 2, and there was no significant difference in clinical characteristics between the 2 groups.

### 2.4. Construction of prognostic risk model based on miRNA genes

The R-packet survival coxph function was used to analyze the differential miRNA in the training set by univariate Cox proportional hazard regression analysis.<sup>[16]</sup> MiRNAs meeting the threshold of  $P < .05$  were considered as miRNAs significantly related to the prognosis of UCEC. To reduce the number of miRNAs, R software (version 4.0.2, the R Foundation) package glmnet for conducting LASSO cox regression analysis to further screen prognostic miRNA(s).<sup>[17,18]</sup> Five-fold cross validation was carried out to identify the optimal lambda value from the minimum partial likelihood deviance. We used Akaike information criterion (AIC) for stepwise regression to avoid

**Table 1**

**Statistical table of the sample clinical information.**

Clinical Features	TCGA-UCEC
Type	
Normal	536
Tumor	33
OS	
0	448
1	88
PFS	
0	413
1	123
DSS	
0	475
1	59
Unknown	2
DFS	
0	361
1	57
Unknown	118
Stage	
I	338
II	49
III	121
IV	28
Grade	
G1	98
G2	121
G3	306
G4	11
Age, yr	
≤65	303
>65	231
Unknown	2

DFS = disease free survival, DSS = disease specific survival, OS = overall survival, PFS = progression free survival, TCGA = The Cancer Genome Atlas, UCEC = uterine corpus endometrial carcinoma.

**Table 2**

**Sample information of the TCGA training set and verification set.**

Clinical features	TCGA-train	TCGA-test	P
OS			
0	230	218	.1996
1	38	50	
PFS			
0	211	202	.4112
1	57	66	
DSS			
0	240	235	.332
1	28	31	
Unknown	0	2	
DFS			
0	183	178	.5078
1	31	26	
Unknown	54	64	
Stage			
I	172	166	.9536
II	24	25	
III	58	63	
IV	14	14	
Grade			
G1	45	53	.5874
G2	64	57	
G3	155	151	
G4	4	7	
Age, yr			
≤65	151	152	.3665
>65	115	116	
Unknown	2	0	

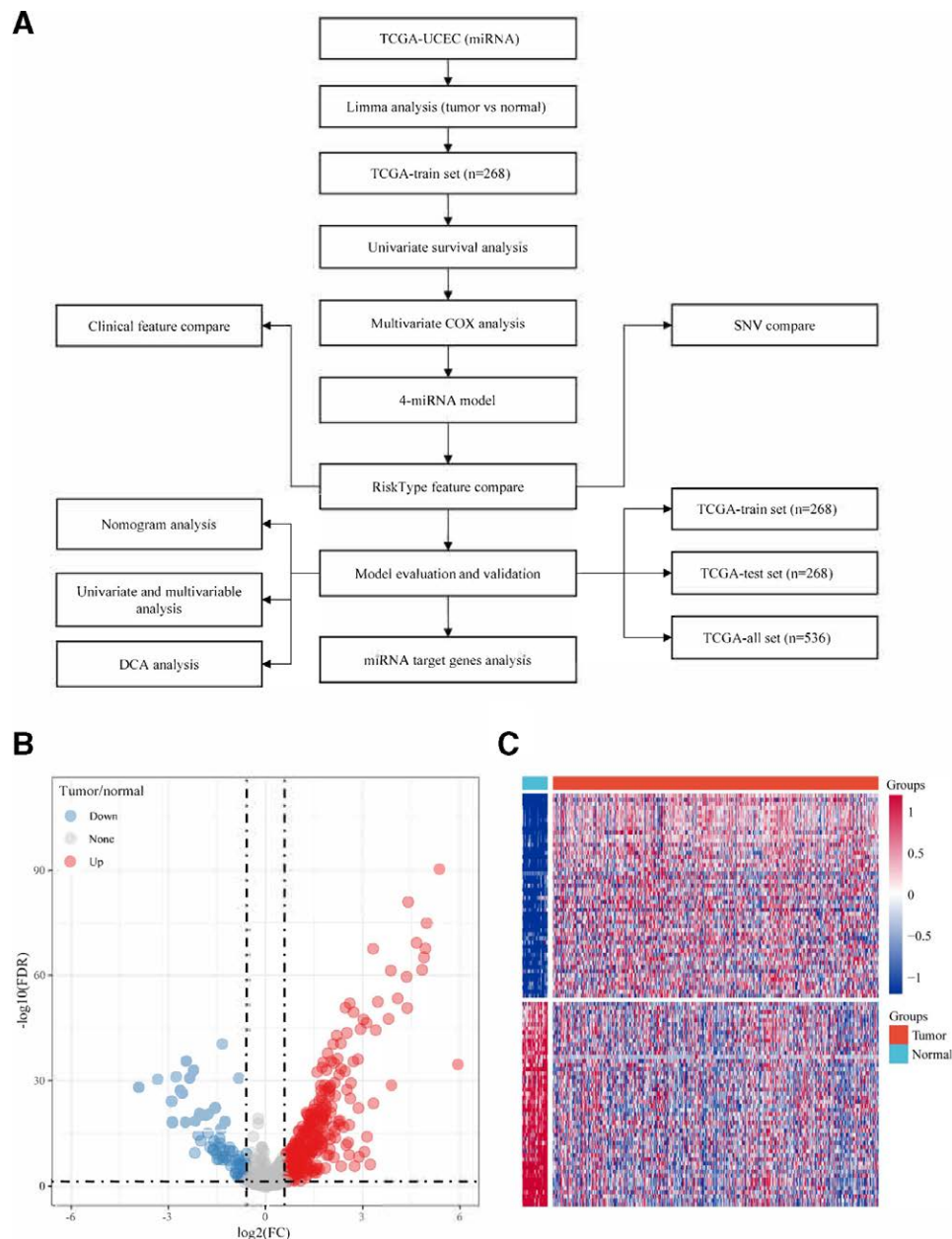
DFS = disease free survival, DSS = disease specific survival, OS = overall survival, PFS = progression free survival, TCGA = The Cancer Genome Atlas.

adverse effects of over-fitting. The stepAIC method in the MASS package starts with the most complex model and deletes a variable to reduce AIC, with a smaller value indicating a model of sufficient degree of fit fewer parameters.<sup>[19]</sup> After stepwise regression, the number of miRNAs with the optimal prognosis was reduced. Then, a prognostic miRNA model was established according to the coefficient of LASSO-Cox regression model multiplied by the level of miRNA expression.

**2.5. Evaluation and verification of risk model**

The risk score for each patient was evaluated according to miRNA model, and the risk score was standardized by z-score. Samples with risk score > 0 after z-score were divided into high-risk group, while those with risk score < 0 were divided into low risk-group. The logarithmic rank test in the training

and verification set was used for survival analysis to confirm the survival difference between the high-risk group and the low-risk group. Receiver operating characteristic (ROC) analysis was performed using time ROC<sup>[20]</sup> in the R software package to compare the sensitivity of 1-year, 3-year and 5-year survival predictions, and the area under the curve (AUC) was calculated. In addition, to further assess the reliability of the miRNA model in other clinical times, we also analyzed the survival situation of the high- and low-risk groups under the time and state of progression free survival (PFS), disease free survival (DFS), and disease specific survival (DSS), and drew the ROC curve of the risk score model under the time and state of PFS, DFS and DSS. In addition, the clinical practicability of the miRNA model was evaluated and the differences of different clinical features (survival rate, age, pathological grade, clinical stage) and gene mutation frequency were compared in high- and low-risk groups.



**Figure 1.** Screening of differentially expressed miRNAs. (A) The research flow chart of UCEC prognosis miRNA-related signature. (B) MiRNA Volcano Diagram of the difference between normal samples and UCEC samples. (C) The difference of miRNA heatmap between normal samples and UCEC samples. miRNAs = microRNAs, UCEC = uterine corpus endometrial carcinoma.

### 2.6. Construction and verification of nomogram

Univariate and multivariate Cox regression analysis were performed to investigate the independence of risk model in predicting the prognosis of UCEC. The nomogram model was constructed based on age, pathological grade, clinical stage and risk score factors,<sup>[21]</sup> and the performance of the nomogram was evaluated by C index, ROC analysis and decision curve analysis (DCA).

### 2.7. Analysis of miRNA target genes

The regulation mode of miRNA in the risk scoring model was analyzed. Firstly, the target gene of miRNA was identified by miRWalk database.<sup>[22]</sup> Then, the R software package WebGestaltR was used to examine the KEGG pathway and GO functional enrichment of the target genes of miRNAs.<sup>[23]</sup>

### 2.8. Statistical analysis

All statistical analyses were carried out using R software (<http://www.Rproject.org>). Chi-square test was employed to analyze the differences in clinical characteristics between training set and validation set. The overall survival (OS) rate was determined by Kaplan–Meier method.  $P < .05$  was considered to be statistically significant.

## 3. Results

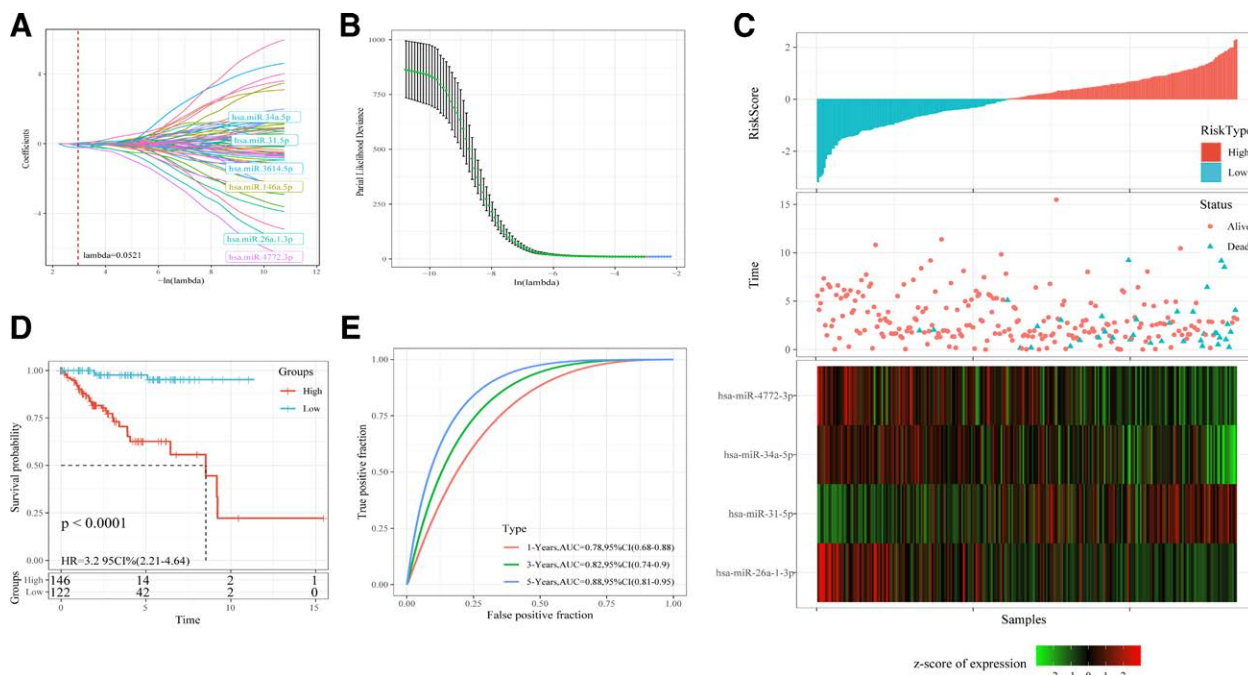
### 3.1. Identification of differentially expressed miRNA

A flow chart was drew (Fig. 1A) based on the overview of the analysis steps to explain our workflow more clearly. Difference analysis showed that 388 miRNAs were differentially expressed in UCEC tissues compared with normal tissues. Among them, 318 were differentially upregulated in UCEC, and the other 70

were differentially low-expressed miRNAs (Fig. 1B). The heatmap of these 388 differentially expressed miRNAs were also displayed in Fig. 1C.

### 3.2. The training set was used to construct a prognostic risk model based on miRNA genes

Through univariate Cox proportional hazard regression analysis, 70 miRNAs selected in the training set were detected to be significantly correlated with the OS of UCEC patients. LASSO cox analysis further filtered 70 prognostic miRNAs. By analyzing the changing trajectory of each independent variable, it can be found that the number of independent variable coefficients closing to zero increased with the increase of  $\lambda$  (Fig. 2A). According to the  $\lambda$  value (0.0521), when the model reached the optimal value, the optimal number of miRNAs was determined to be 6 (Fig. 2B). The number of miRNAs was reduced to 4 (miR-31-5p, miR-34a-5p, miR-26a-1-3p and miR-4772-3p) by stepAIC. By multiplying the coefficient of LASSO-Cox regression model by the expression level of each miRNA, risk score = 0.235\* (miR-31-5p)-0.3\* (miR-34a-5p)-0.46\* (miR-26a-1-3p)-0.587\* (miR-4772-3p). The risk score of each UCEC sample in the training set was calculated by the risk score formula, and the risk score was standardized by z-score. The standardized samples greater than 0 were divided into high-risk group (n = 146), while those lower than 0 were divided into low-risk group (n = 122). The survival status of different UCEC samples and the heatmap of 4 prognostic miRNAs were displayed in Figure 2C. Survival analysis demonstrated that the OS of the low-risk group was notably longer than that of the high-risk group (Fig. 2D). ROC analysis showed that the AUC values of 1-, 3- and 5-year survival rates were 0.78, 0.82, and 0.88, respectively (Fig. 2E). This indicated that the prognostic model had a better performance in the prediction of long-term survival.



**Figure 2.** The training set was used to construct a prognostic risk model based on miRNA genes. (A) There are 70 miRNA coefficients of LASSO, and the horizontal axis represents the log value of the independent variable  $\lambda$ . (B) The confidence interval under each lambda. (C) From top to bottom is the risk score based on 4-miRNA model (top), the survival status of different UCEC samples and the heatmap of 4 prognostic miRNAs. (D) The time-dependent ROC curve of UCEC patients in the training set. (E) Survival analysis of UCEC high-risk group and low-risk group in training set. LASSO = least absolute shrinkage, and selection operator, miRNAs = microRNAs, ROC = receiver operating characteristic, UCEC = uterine corpus endometrial carcinoma.

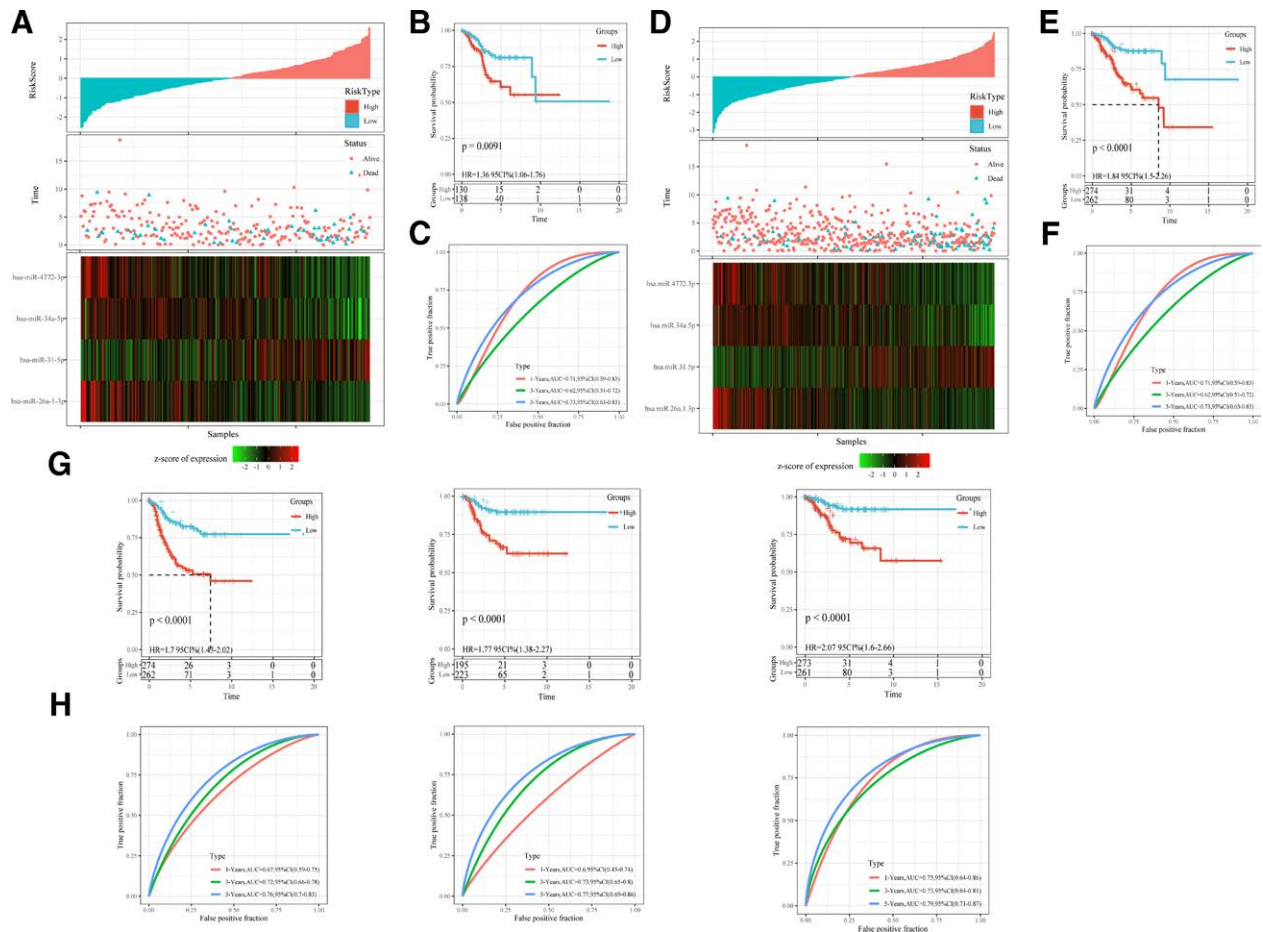
**3.3. The miRNAs signature related to UCEC prognosis were validated in the validation set and entire dataset**

To evaluate the robustness of the 4-miRNA model, the predictive ability of the prognostic signature was verified in the verification set (n = 268) and the whole data set (n = 536). The risk score of the patients in the training set was obtained based on the 4-miRNA model, and the risk score of these samples was transformed into z-score score. A total of 130 samples with z-score > 0 were divided into high-risk group, whereas 138 samples with z-score < 0 were divided into low-risk group. The risk score distribution of the samples in the training cohort, the expression profile of the 4 prognostic miRNAs and the survival status of the patients also showed similar results to the training set (Fig. 3A and B). ROC analysis manifested that 4-miRNA model showed a better performance in predicting 1-year, 3-year and 5-year OS (Fig. 3C). Throughout the data set, we grouped using the same method as the training set and validation set, and consistent with the results of the training set, the high-risk group had a worse OS (Fig. 3D and E) than the low-risk group. The AUC values of 1-, 3- and 5-year survival in the whole data set were 0.74, 0.69, and 0.78, respectively (Fig. 3F). Kaplan–Meier analysis of PFS, DFS and DSS of UCEC patients showed that

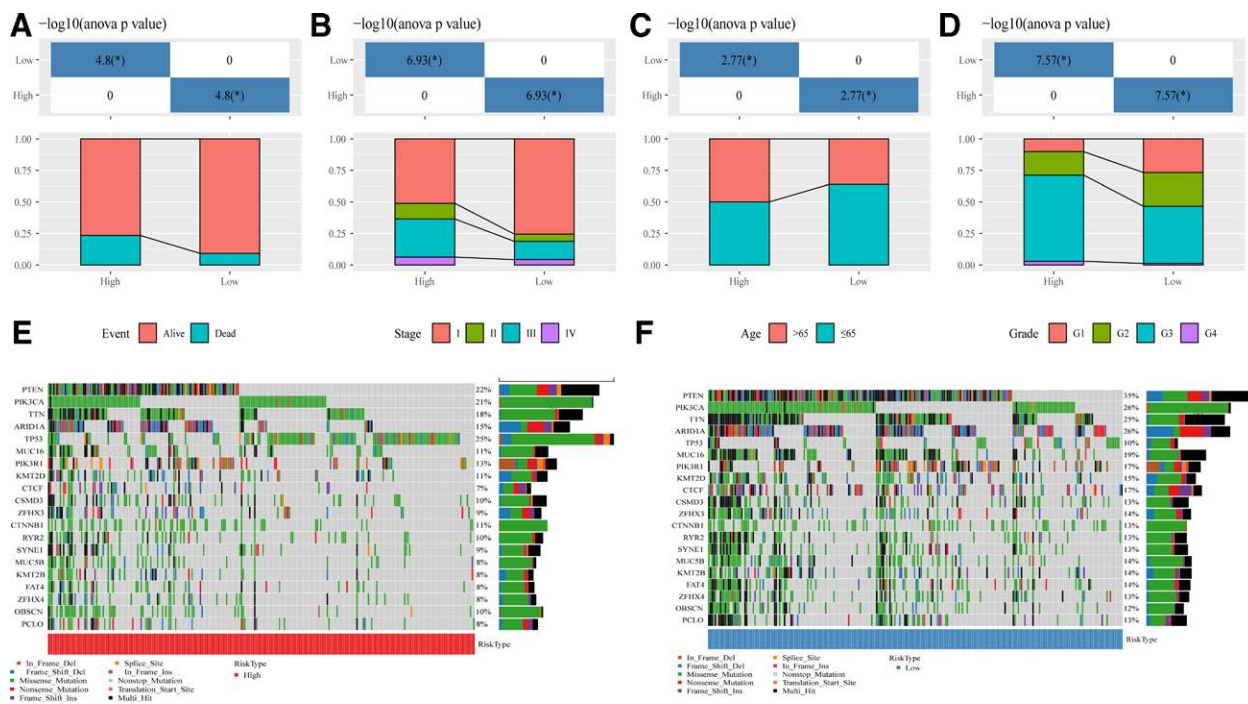
PFS, DFS and DSS of low-risk patients were significantly better than those of high-risk patients (Fig. 3G). From the time-dependent ROC curve, it can be seen that the 4-miRNA signature also had a good performance in predicting PFS, DFS and DSS in patients with UCEC (Fig. 3H). Therefore, the 4-miRNA signature developed in the current study was verified to have a high accuracy and robustness in predicting the prognosis of UCEC.

**3.4. Comparison of clinical characteristics and analysis of gene mutation between high- and low-risk groups**

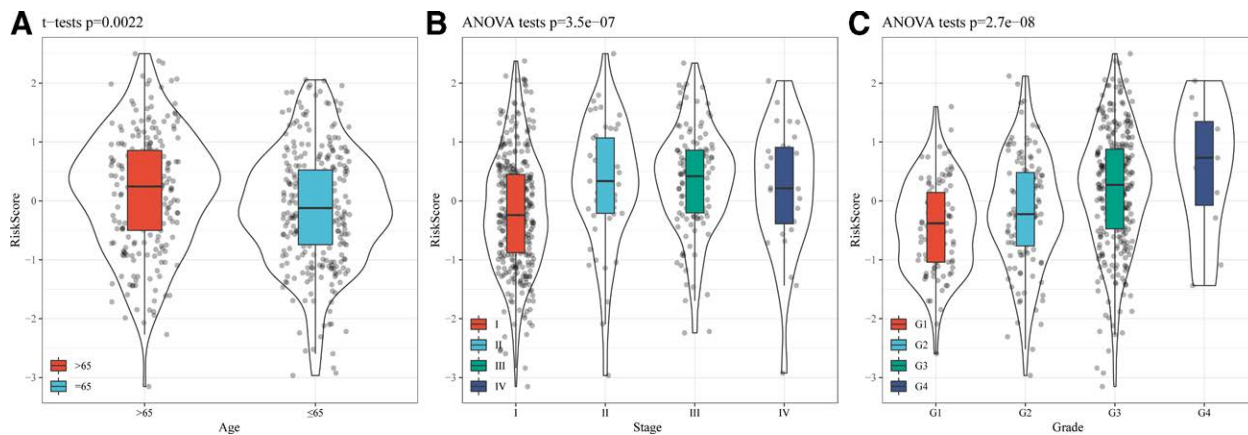
After comparing the differences of different clinical features and gene mutation frequency in high- and low-risk groups, we found that there were significant differences in survival rate, proportion of stage, grade and age in the 2 groups. Compared with low-risk patients, high-risk patients with poor prognosis had lower survival rates, higher ratios of II, III and IV, and higher proportions of G2, G3 and G4, and more patients elder than 65 years old (Fig. 4A–D). In the analysis of gene mutations, TP53 showed a high mutation in the high-risk group of UCEC patients, and in the low-risk group, PTEN had a high mutation (Fig. 4E and F).



**Figure 3.** The miRNA signature associated with UCEC outcomes was validated in the validation set and across the dataset. (A) From top to bottom, the heatmap of the risk score, survival status and 4 prognostic miRNAs of UCEC patients based on 4-miRNA model in the verification set. (B) To verify the distribution of survival curve of patients in high-risk group and low-risk group. (C) The time-dependent ROC curve of UCEC patients in the validation set. (D) From top to bottom, the risk score of UCEC patients based on 4-miRNA model, survival status and heatmap of 4 prognostic miRNAs in the whole data. (E) Kaplan–Meier curve of high-risk and low-risk UCEC patients in the whole data set. (F) Time-dependent ROC analysis compared the accuracy of our 4-miRNA signature in predicting 1-year, 3-year, and 5-year OS in patients with UCEC. (G) The Kaplan–Meier curves of PFS, DFS and DSS of UCEC patients in high-risk group and low-risk group. (H) 4-miRNA signature was used to predict the time-dependent ROC curves of PFS, DFS and DSS in patients with UCEC. DFS = disease free survival, DSS = disease specific survival, miRNAs = microRNAs, PFS = progression free survival, ROC = receiver operating characteristic, UCEC = uterine corpus endometrial carcinoma.



**Figure 4.** Comparison of clinical characteristics and molecular mutation analysis between high- and low-risk groups. (A) Comparison of survival ratio between high- and low-risk groups. (B) Comparison of the proportion of stage between high- and low-risk groups. (C) Analysis of the proportion of grade in patients with high and low risk. (D) Analysis of age distribution in high- and low-risk groups. (E) Gene mutation analysis in high- and low-risk groups.



**Figure 5.** Clinical practicability of 4-miRNA signature. The relationship between the risk score and age (A), in clinical stage (B), pathological grade (C). miRNAs = microRNAs.

### 3.5. Clinical applicability of the 4-miRNA signature

We also analyzed the relationship between the 4-miRNA prognostic marker and age and clinical stage, pathological grade. From the results of the analysis, it can be seen that there was a significant difference in the risk score between patients elder than 65 years old and patients younger than or equal to 65 years old (Fig. 5A). In addition, there were significant differences in risk scores between different clinical stage and pathological grades (Fig. 5B and C). The risk score increased with the increase of pathological grade. The above results prove that the 4-miRNA signature was related to the progression of UCEC.

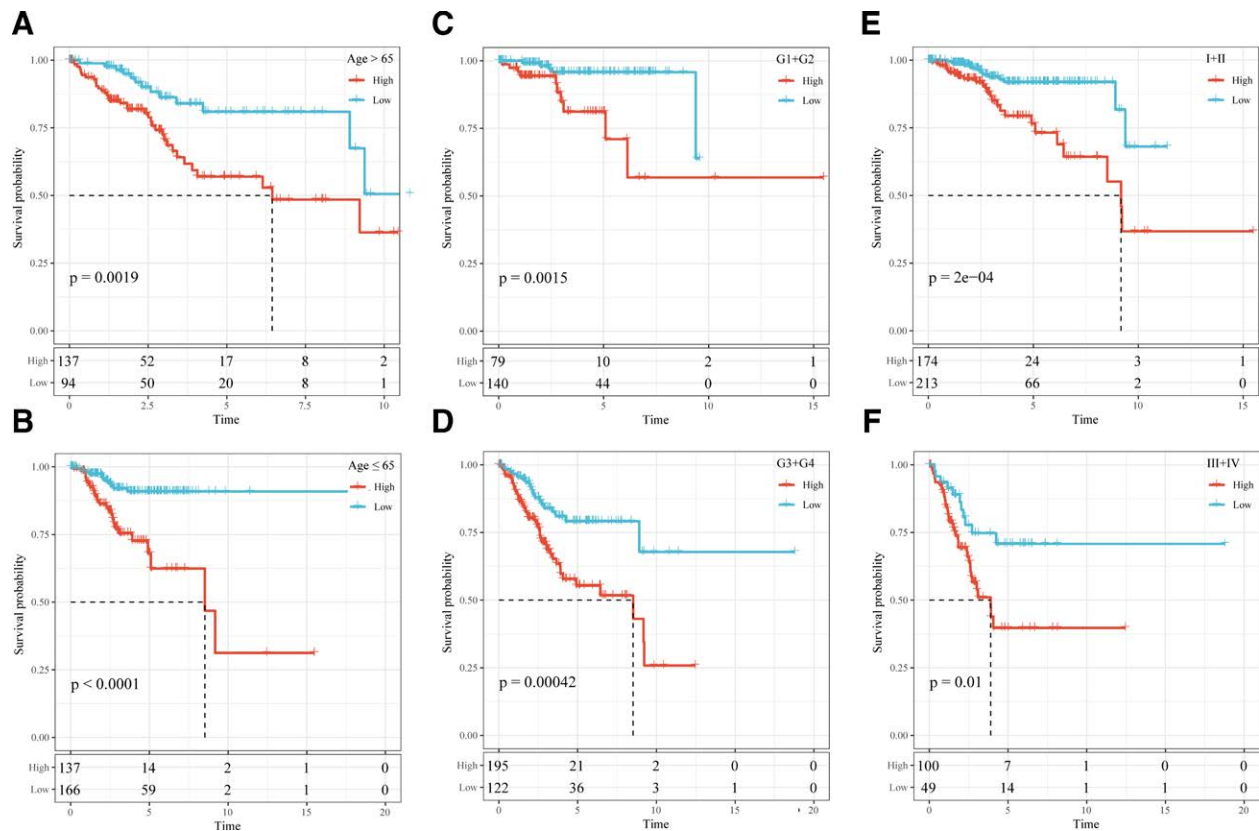
### 3.6. The model based on the 4-miRNA signature can predict different clinical features

We also conducted hierarchical survival analysis to investigate whether the 4-miRNAs models could predict survival outcomes

in subgroups with different clinicopathologies. The results showed that 4-miRNAs signature could distinguish the risk of recurrence in patients with different age (Fig. 6A and B), pathological grade (Fig. 6C and D) or clinical stage (Fig. 6E and F). This further indicates that 4-miRNAs signature is reliable in evaluating the prognosis of UCEC.

### 3.7. The 4-miRNA signature was an independent prognostic factor for UCEC

To verify the independence of the 4-miRNA model in the prognosis of UCEC, univariate and multivariate COX regression analysis was carried out in the whole data set. Univariate COX regression analysis demonstrated that age, pathological grade, clinical stage, and risk type were prominently correlated with OS in patients with UCEC (Fig. 7A). Subsequent multivariate COX analysis revealed that pathological grade,



**Figure 6.** Stratified analysis of different clinicopathological subgroups in TCGA dataset. (A) Age > 65; (B) Age ≤ 65; (C) G1 + G2; (D) G3 + G4; (E) I + II; (F) III + IV. TCGA = The Cancer Genome Atlas.

clinical stage and risk type were independent prognostic factors of UCEC (Fig. 7B). Next, the line chart constructed by age, pathological grade, clinical stage and risk score showed that the risk score had the greatest impact on survival prediction (Fig. 7C). From the calibration chart of the survival rate of the nomogram, the nomogram can accurately predict 1-year, 3-year and 5-year OS (Fig. 7D). In addition, the DCA of age, clinical stag, grade, pathological grade, risk score and nomogram confirmed that the model has good practicability (Fig. 7E).

### 3.8. Prediction of target genes and analysis of pathway enrichment

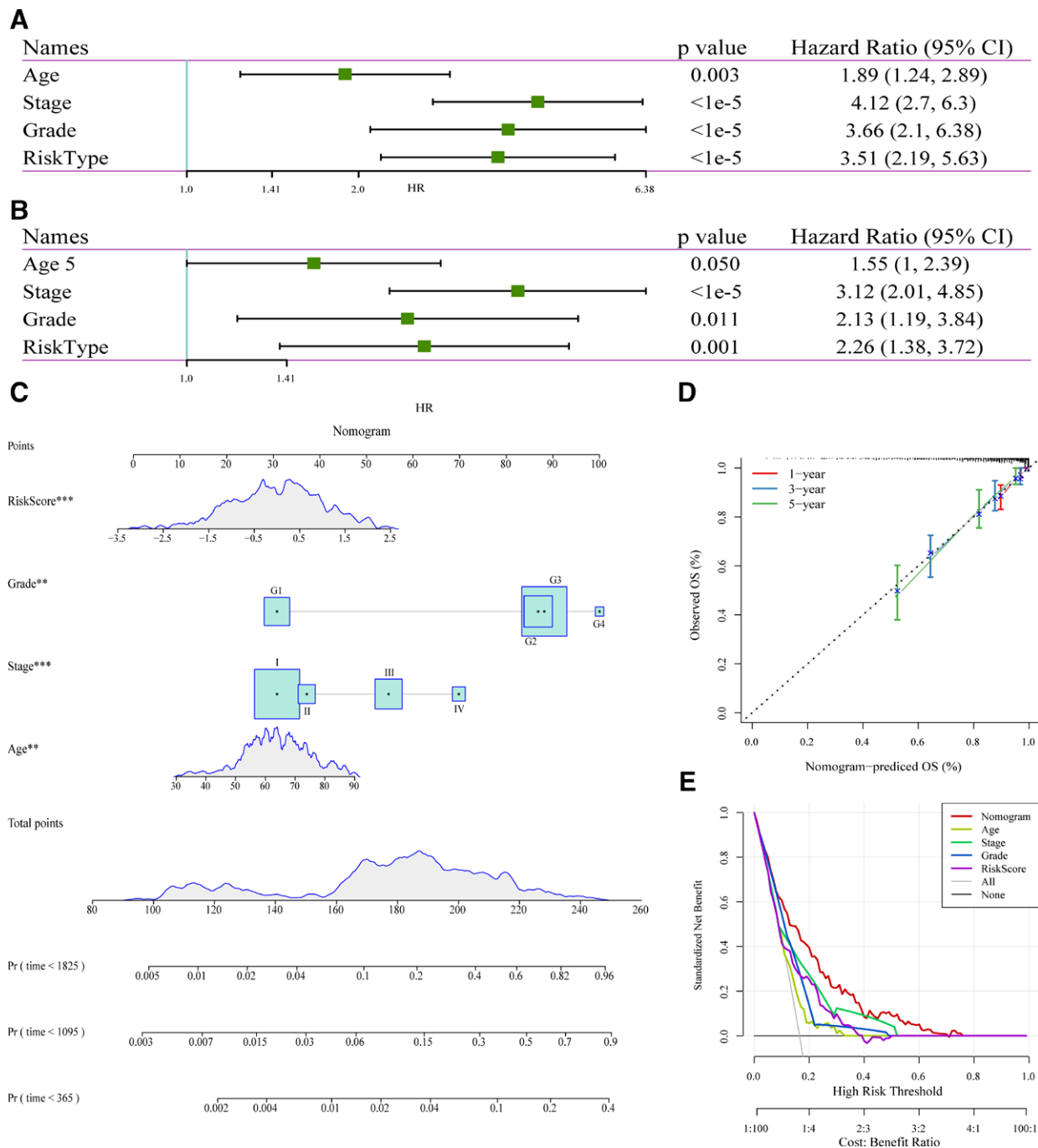
The target genes of 4 miRNA were predicted by miRWalk database. These target genes were analyzed by KEGG pathway analysis and GO functional enrichment analysis. GO functional annotation showed that the target genes of 4 miRNAs were significantly enriched in 75 biological processes, 20 cellular components and 82 molecular functions. Biological processes included, for example, positive regulation of dendrite morphogenesis, positive regulation of smooth, cellular response to decreased oxygen levels (Fig. 8A). Cellular components included ruffle, spindle and membrane raft, for example (Fig. 8B). The mainly enriched molecular function were death receptor binding, molecular adaptor activity and mRNA binding and so on (Fig. 8C). KEGG pathway enrichment analysis showed that the target genes of 4 miRNAs were significantly enriched in NF-kappa B signaling pathway, miRNAs in cancer and pathways in cancer pathways (Fig. 8D).

## 4. Discussion

Accurate treatment management of UCEC patients depends on comprehensive clinical and pathological evaluation.<sup>[24]</sup> Data from emerging studies clearly reveal the clinical significance of miRNAs in the diagnosis and prognosis of UCEC.<sup>[25,26]</sup> Previously, Wang et al detected 6 miRNAs related to the occurrence and development of UCEC, and constructed a 6-miRNA signature to predict the prognosis of UCEC.<sup>[27]</sup> In addition, Schmidt established a prognostic ratio model consisting of 4 miRNAs, which can predict the time of biochemical recurrence independent of routine clinicopathological variables.<sup>[28]</sup> However, due to insufficient sample size, lack of external verification and some other reasons, the accuracy of these prognostic signature is limited to different degrees. Therefore, a richer sample size is required to construct a prognostic signature based on miRNAs and external verification to improve the current clinicopathological prognostic model of patients.

In this study, we obtained 569 UCEC samples from TCGA and screened 388 differential miRNAs from the samples. Five hundred sixty-nine samples were divided into training set and test set. Then univariate Cox risk regression identified 70 miRNAs significantly related to the prognosis of UCEC in the training set. Six prognostic miRNAs were further screened by LASSO cox analysis, an effective method to establish prognostic risk characteristics, but there is often the problem of data over-fitting.<sup>[29]</sup> Here, to avoid the impact of over-fitting, we carried out stepwise regression according to AIC, and finally extracted 4 miRNAs, namely, miR-31-5p and miR-34a-5p, miR-26a-13p and miR-4772-3p, with the least number of predictors of prognosis.

Among the 4 miRNAs, miR-31-5p plays an important role in many types of cancer. It is reported that miR-31 can

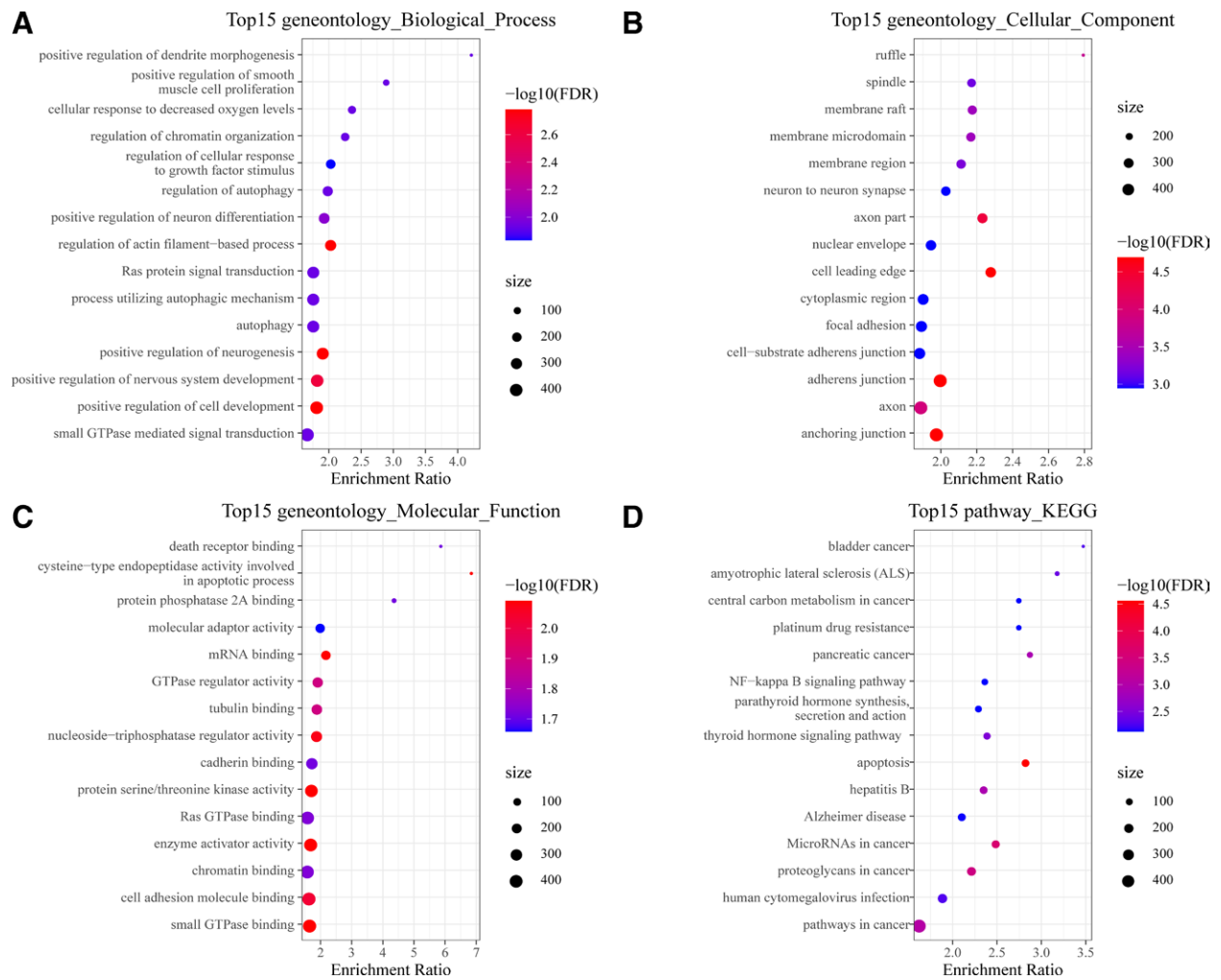


**Figure 7.** The 4-miRNA signature was an independent prognostic factor for UCEC. (A) Univariate COX regression analysis of patients with UCEC. (B) Multivariate COX regression analysis of patients with UCEC. (C) Nomogram based on clinical characteristics of UCEC patients and risk score. (D) Calibration chart of nomogram. (E) DCA based on age, clinical stag, Grade, pathological grade, risk score and nomogram. DCA = decision curve analysis, miRNAs = microRNAs, UCEC = uterine corpus endometrial carcinoma.

independently predict lymph node metastasis and survival status of patients with lung cancer. MiR-31-5p promotes metastasis of lung cancer, but it has anti-metastasis effect on breast cancer, and it is also a prognostic marker of prostate cancer.<sup>[30]</sup> Through LinkedOmics database analysis, some studies found that colon adenocarcinoma patients with a high expression of miR-31-5p tend to develop a poor prognosis.<sup>[31]</sup> MiR-31-5p also has anti-proliferative effect on ovarian cancer, osteosarcoma and prostate cancer.<sup>[32]</sup> The results of a meta-analysis suggested that the high-expressed miR-31 is related to a poor OS and cancer-specific survival of cancer patients.<sup>[33]</sup> In

addition, the regulatory role of miR-34a5p in cancer has also been reported. In colorectal cancer, miR-34a-5p prevents tumor progression by inhibiting the growth and metastasis of cancer cells.<sup>[34]</sup> On the other hand, the study of miR-34a-5p in oral carcinoma found that miR-34a-5p increases the invasiveness of oral cancer cells by regulating the cascade of AXL/AKT/GSK-3  $\beta$ / $\beta$ -catenin/Snail signal transduction.<sup>[35]</sup> MiR-4772-3p in serum exocrine can help predict tumor recurrence of stage II and III colon cancer.<sup>[36]</sup> The evidence points to an important role of these miRNAs in cancer, but their prognostic potential in UCEC is rarely reported.





**Figure 8.** Functional enrichment analysis of target genes of 4 miRNAs. (A) Biological process annotation map of miRNA target gene. (B) Cellular component annotation map of miRNA target genes. (C) Molecular function annotated map of miRNA target genes. (D) Enrichment analysis map of KEGG pathway of miRNA target genes. KEGG = Kyoto Encyclopedia of Genes and Genomes, miRNAs = microRNAs.

Here, the panel consisting of these 4 miRNAs made up a 4-miRNA prognosis signature. The risk score of each patient was calculated according to the expression level of the 4 miRNAs and regression coefficient. After z-score, the patients were divided into low-risk group and high-risk group, with 0 as the cutoff point. Survival analysis showed that OS, PFS, DFS and DSS in high-risk patients were shorter than those in low-risk patients. The prediction accuracy, reliability and clinical practicability of the prognosis signature was verified by verification set and data set. Most of the endometrial tumors, such as TP53, PTEN, PIK3CA, CTNNB1, tend to have different frequencies of mutations.<sup>[37]</sup> In this study, we found that there was a high mutation of TP53 in the high-risk group of UCEC patients and a high mutation of PTEN in the low-risk group. More importantly, the results of univariate and multivariate Cox regression analysis also revealed the independence of the 4-miRNA prognosis signature in evaluating the prognosis of UCEC.

The emergence of nomograms greatly facilitates the research of tumor prognosis.<sup>[21,38]</sup> This study established a nomogram based on age, clinical stag, grade, pathological grade and risk score of UCEC patients. According to calibration chart and DCA, nomogram, we can effectively predict the prognosis of UCEC patients. At the same time, target genes regulated by 4 miRNAs were explored, and the potential biological functions and regulatory pathways of the target genes were analyzed. The functions and pathways of enrichment of the miRNA target

genes play an indispensable role in the malignant progression of tumors.

We developed a UCEC prognosis signature based on 4 miRNAs and validated it in 2 different queues. The 4-miRNA prognosis signature can be used as an independent prognostic factor to effectively predict the prognosis of patients with UCEC. These findings provide new markers for predicting the prognosis of patients with UCEC. However, there are also some limitations in this study, for instance, the current data source came from public databases and there were relatively few clinical variables. The training set and the test set originated from the same database, and the effectiveness of prognosis signature has not been confirmed in other databases or tested by biological experiments. These limitations will be improved in our future study.

**Author contributions**

JH was responsible for study conceptualization and writing. FD participated in data collection, data analysis and data interpretation. NW was responsible for reviewing the manuscript. All authors have read and approved the final manuscript.

**Conceptualization:** Jiazhen Huang.

**Data curation:** Furong Du.

**Formal analysis:** Furong Du.

**Methodology:** Furong Du.

**Software:** Furong Du.

**Supervision:** Ning Wang.

**Writing – original draft:** Jiazhen Huang.

**Writing – review & editing:** Ning Wang.

## References

- [1] Van Nyen T, Moiola CP, Colas E, et al. Modeling endometrial cancer: past, present, and future. *Int J Mol Sci.* 2018;19:2348.
- [2] Hussein YR, Soslow RA. Molecular insights into the classification of high-grade endometrial carcinoma. *Pathology.* 2018;50:151–61.
- [3] Morice P, Leary A, Creutzberg C, et al. Endometrial cancer. *Lancet.* 2016;387:1094–108.
- [4] Arora V, Quinn MA. Endometrial cancer. *Best Pract Res Clin Obstet Gynaecol.* 2012;26:311–24.
- [5] Berger AA, Dao F, Levine DA. Angiogenesis in endometrial carcinoma: Therapies and biomarkers, current options, and future perspectives. *Gynecol Oncol.* 2021;160:844–50.
- [6] Wu H, Zhang J. Decreased expression of TFAP2B in endometrial cancer predicts poor prognosis: a study based on TCGA data. *Gynecol Oncol.* 2018;149:592–7.
- [7] Liu D, Gunther K, Enriquez LA, et al. ROR1 is upregulated in endometrial cancer and represents a novel therapeutic target. *Sci Rep.* 2020;10:13906.
- [8] Li C, Long Q, Zhang D, et al. Identification of a four-gene panel predicting overall survival for lung adenocarcinoma. *BMC Cancer.* 2020;20:1198.
- [9] Torres A, Torres K, Pesci A, et al. Diagnostic and prognostic significance of miRNA signatures in tissues and plasma of endometrioid endometrial carcinoma patients. *Int J Cancer.* 2013;132:1633–45.
- [10] Zhou C, Li C, Yan F, et al. Identification of an immune gene signature for predicting the prognosis of patients with uterine corpus endometrial carcinoma. *Cancer Cell Int.* 2020;20:541.
- [11] Long J, Zhang L, Wan X, et al. A four-gene-based prognostic model predicts overall survival in patients with hepatocellular carcinoma. *J Cell Mol Med.* 2018;22:5928–38.
- [12] Li CW, Shi X, Ma B, et al. A 4 gene-based immune signature predicts dedifferentiation and immune exhaustion in thyroid cancer. *J Clin Endocrinol Metab.* 2021;106:e3208–20.
- [13] A J, Zhang B, Zhang Z, et al. Novel gene signatures predictive of patient recurrence-free survival and castration resistance in prostate cancer. *Cancers (Basel).* 2021;13:917.
- [14] Srivastava SK, Ahmad A, Zubair H, et al. MicroRNAs in gynecological cancers: small molecules with big implications. *Cancer Lett.* 2017;407:123–38.
- [15] Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43:e47.
- [16] Zhang Z. Semi-parametric regression model for survival data: graphical visualization with R. *Ann Transl Med.* 2016;4:461.
- [17] Engebretsen S, Bohlin J. Statistical predictions with glmnet. *Clin Epigenetics.* 2019;11:123.
- [18] Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 2010;33:1–22.
- [19] Zhang Z. Variable selection with stepwise and best subset approaches. *Ann Transl Med.* 2016;4:136.
- [20] Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med.* 2013;32:5381–97.
- [21] Iasonos A, Schrag D, Raj GV, et al. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol.* 2008;26:1364–70.
- [22] Dweep H, Sticht C, Pandey P, et al. miRWalk--database: prediction of possible miRNA binding sites by “walking” the genes of three genomes. *J Biomed Inform.* 2011;44:839–47.
- [23] Wang J, Vasaikar S, Shi Z, et al. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* 2017;45:W130–7.
- [24] Yokohata K, Manabe H, Nishikata F, et al. Successfully treated primary germ cell tumor of the mediastinum: report of 2 cases. *Kyobu Geka.* 1988;41:399–402.
- [25] Rizner TL. Discovery of biomarkers for endometrial cancer: current status and prospects. *Expert Rev Mol Diagn.* 2016;16:1315–36.
- [26] Banno K, Yanokura M, Kisu I, et al. MicroRNAs in endometrial cancer. *Int J Clin Oncol.* 2013;18:186–92.
- [27] Wang Y, Xu M, Yang Q. A six-microRNA signature predicts survival of patients with uterine corpus endometrial carcinoma. *Curr Probl Cancer.* 2019;43:167–76.
- [28] Schmidt L, Fredsoe J, Kristensen H, et al. Training and validation of a novel 4-miRNA ratio model (MiCaP) for prediction of postoperative outcome in prostate cancer patients. *Ann Oncol.* 2018;29:2003–9.
- [29] Tu Z, Shu L, Li J, et al. A novel signature constructed by RNA-binding protein coding genes to improve overall survival prediction of glioma patients. *Front Cell Dev Biol.* 2021;8:588368.
- [30] Meng W, Ye Z, Cui R, et al. MicroRNA-31 predicts the presence of lymph node metastases and survival in patients with lung adenocarcinoma. *Clin Cancer Res.* 2013;19:5423–33.
- [31] Mi B, Li Q, Li T, et al. High miR-31-5p expression promotes colon adenocarcinoma progression by targeting TNS1. *Aging (Albany NY).* 2020;12:7480–90.
- [32] Creighton CJ, Fountain MD, Yu Z, et al. Molecular profiling uncovers a p53-associated role for microRNA-31 in inhibiting the proliferation of serous ovarian carcinomas and other cancers. *Cancer Res.* 2010;70:1906–15.
- [33] Wang S, Hu J, Zhang D, et al. Prognostic role of microRNA-31 in various cancers: a meta-analysis. *Tumour Biol.* 2014;35:11639–45.
- [34] Gao J, Li N, Dong Y, et al. miR-34a-5p suppresses colorectal cancer metastasis and predicts recurrence in patients with stage II/III colorectal cancer. *Oncogene.* 2015;34:4142–52.
- [35] Li YY, Tao YW, Gao S, et al. Cancer-associated fibroblasts contribute to oral cancer cells proliferation and metastasis via exosome-mediated paracrine miR-34a-5p. *EBioMedicine.* 2018;36:209–20.
- [36] Liu C, Eng C, Shen J, et al. Serum exosomal miR-4772-3p is a predictor of tumor recurrence in stage II and III colon cancer. *Oncotarget.* 2016;7:76250–60.
- [37] Cancer Genome Atlas Research Network, Kandoth C, Schultz N, et al. Integrated genomic characterization of endometrial carcinoma. *Nature.* 2013;497:67–73.
- [38] Balachandran VP, Gonen M, Smith JJ, et al. Nomograms in oncology: more than meets the eye. *Lancet Oncol.* 2015;16:e173–80.