Gynecologic Oncology 155 (2019) 499–507

Contents lists available at ScienceDirect

Gynecologic Oncology

journal homepage: www.elsevier.com/locate/ygyno

Identification of a RNA-seq-based signature to improve prognostics for uterine sarcoma

Jian-Guo Zhou [a, *, 1], He-Tong Zhao [b, 1], Su-Han Jin [c, 1], Xu Tian [d], Hu Ma [a, **]

[a] Department of Oncology, Affiliated Hospital of Zunyi Medical University, Zunyi, China
[b] Department of Traditional Chinese Medicine, Changhai Hospital, Second Military Medical University, Shanghai, China
[c] Department of Orthodontics, Affiliated Stomatological Hospital of Zunyi Medical University, Zunyi, China
[d] Chongqing University Cancer Hospital; Chongqing Cancer Hospital and Institute, Chongqing, China

ARTICLE INFO

Article history:
Received 17 May 2019
Received in revised form
27 August 2019
Accepted 31 August 2019
Available online 26 October 2019

Keywords:
Gene
RNA-Seq
Risk score
Uterine sarcoma

ABSTRACT

Objective: Uterine sarcoma (US) is a highly malignant cancer with poor prognosis and high mortality. This study focused on the identification of a RNA-Seq expression signature for prognosis prediction in uterine sarcoma.

Methods: We obtained RNA-Seq expression profiles from The Cancer Genome Atlas database, and differentially expressed genes were identified between US tissues and normal tissues. Univariate Cox proportional hazards regression analysis and LASSO Cox model were performed to identify and construct the prognostic gene signature. Time-dependent receiver operating characteristic, Kaplan-Meier curve and multivariate Cox regression analysis were used to assess the prognostic capacity of the six-gene signature. The nomogram was developed including prognostic signature and independent clinical factors to predict the overall survival (OS) of US patients. The functional enrichment and somatic mutation analysis were also analyzed by bioinformatics to understand the molecular mechanisms.

Results: This study identified a prognostic signature based on 6 genes: FGF23, TLX2, TIFAB, RNF223, HIST1H3A and AADACL4. In the training group, the median OS in the high- and low-risk groups was 19.6 vs 88.1 months (HR, 0.1412, 95% CI: 0.03295 - 0.6054; $P = 0.002$), respectively. In the testing group, the median OS in the high- and low-risk groups were 30 vs NR (not reach) months (HR, <0.0001, 95% CI: 0 - inf; $P = 0.03$). In all of patients, the low-risk group showed significant better survival compared with the high-risk group in OS, PFI, DSS and DFI. The nomogram based on the gene signature and radiation therapy was developed and successfully predicted the OS of US patients. The patients in the high-risk group displayed distinct mutation signatures comparing to patients in the low-risk group. Functional enrichment analysis indicated that the signature can play a vital role in cancer-related biological processes.

Conclusion: Our study established a novel 6-gene signature and nomogram which could improve prognosis prediction in patients with US.

1. Introduction

Uterine sarcoma (US) is a rare, aggressive and lethal gynecological cancer in women of all ages that accounts for 3-7% of all uterine cancers, presenting a higher incidence in women between 50 and 70 years old [1–3]. Uterine sarcomas include leiomyosarcoma, carcinosarcoma, and endometrial stromal sarcoma according to the common histological types [4]. The prognosis of uterine sarcomas remains poor. The 5-year survival rate varies from 50–55% of early US and 8–12% for advanced US [5]. Regardless of which therapy can be used for uterine sarcoma, the response to treatment is unsatisfactory [6]. Further insights are desperately needed to predict the outcome of uterine sarcomas to improve the 5-year survival rate.

As of yet, uterine sarcomas have numerous characteristics that are validated as prognostic elements, including tumor staging, age and necrosis [7–9]. However, these clinical characteristics could not effectively distinguish between patients with high or low survival rates and do not predict which patients are likely to benefit

* Corresponding author.
** Corresponding author.
E-mail addresses: jianguo.zhou@fau.de (J.-G. Zhou), mahuab@163.com (H. Ma).
[1] These authors have contributed equally to this work.

https://doi.org/10.1016/j.ygyno.2019.08.033
0090-8258/

from therapy. Dysregulated RNA plays a leading role in the occurrence and progression of malignant tumors, suggesting prognostic model-based RNA-Seq expression profiles can be used as prognostic signatures in US. Thus, there is a need to construct a prognostic signature that can predict overall survival (OS) in US. In the age of genomics, several researchers conducted studies to develop a gene expression signature associated with the general survival of patients with various cancers [10–14]. However, there are presently no RNA-seq-based prognostic signatures for uterine sarcoma.

In this study, we developed an individualized RNA-seq-based prognostic signature that uses the least absolute shrinkage and selection operation (LASSO) Cox regression model to predict the survival rate of US patients in the training group and then validated the prediction accuracy of this model in the testing group.

2. Materials and methods

2.1. Clinical data sources

We downloaded the RNA sequencing (RNA-seq) expression profiles and clinical information from The Cancer Genome Atlas (TCGA) data portal (http://cancergenome.nih.gov/) and the Genotype-Tissue Expression project (GTEx) (http://commonfund.nih.gov/GTEx/), which contained data from 57 US tissues (2 without follow-up data) and 78 normal tissues from US patients (Supplementary Table S1). This study met the publication guidelines provided by TCGA (http://cancergenome.nih.gov/publications/publicationguidelines).

2.2. Searching strategy

PubMed (http://www.ncbi.nlm.nih.gov/pubmed) was searched to find articles related to prognostic signatures in US using "prognostic [Title/Abstract] AND signature [Title/Abstract] AND uterine [Title/Abstract] AND sarcoma", which suggested there is no prognostic signature that had been developed in US yet.

2.3. Differentially expressed gene analysis

To select differentially expressed genes (DEGs) associated with the development of US, genes between tumor tissues and normal tissues were analyzed by using the R software edgeR package [15] (http://bioconductor.org/packages/edgeR/) according to the parameters | Log_2-fold change | <1 and *P*-value $\leq$ 0.05.

2.4. Developing and validating the RNA-seq-based prognostic model

Considering the generally unfavorable outcomes of US, we chose OS as the primary endpoint, progression-free interval (PFI), disease-specific survival (DSS) and disease-free interval (DFI) as the secondary endpoints [16]. Then we acquired all clinical data from the TCGA database. Next, the relationship between genes and patient OS was studied by univariate Cox proportional hazards regression analysis. Prognosis-associated genes were selected if | Log_2-fold change | <1 and *P*-value < 0.05. DEGs associated to a great extent with the OS of US patients were included in a LASSO Cox proportional hazards regression model to minimize overfitting in the training group with 50,000 bootstrapping repetitions conducted by the R package. Prognosis-associated DEGs with a nonzero coefficient were selected to build a prognostic signature. The prognostic value of the prognostic signature was further validated in the testing set. The patients in the training and testing groups were divided into high-risk and low-risk groups according to the median risk score calculated based on the signature. Hazard ratios (HRs) and 95% confidence intervals (95% CIs) were calculated. Kaplan-Meier analysis with the log-rank test for the difference was performed by the R package survival.

2.5. Nomogram

To evaluate the prognostic value of the risk score calculated by the signature and other clinical variables, univariate and multivariate Cox proportional hazards analyses for OS and PFI were performed, including the genetic risk score and the conventional clinicopathologic variables. The nomogram consisted of independent prognostic factors that were constructed by rms R package. The time-dependent receiver operator characteristic (ROC) area under the curve (AUC) values was drawn by timeROC package (version 0.3) [17]. The patients with US were divided into high- and low-risk groups as 1:1 according to the risk score calculated by the nomogram.

2.6. Functional enrichment and somatic mutation analysis

Functional enrichment analysis (gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis) of the prognostic signature was performed using the bioconductor package clusterprofiler [18]. Somatic mutation analysis was performed using the maftools packages for visualization and summarization of Mutation Annotation Format (MAF) files [19]. The Fisher's exact test was used to identify the differentially muted genes. The plotmafSummary function was used to plot the summary of the MAF file, which displays a number of variant types and classifications. The oncoplot function was used to plot OncoPlot of the top ten mutated genes.

2.7. Statistical analysis

DEGs were defined as an absolute fold change larger than 1 with an adjusted *P*-value less than 0.05. The Cox proportional hazards model was performed to determine which gene was an independent prognostic factor associated with survival. The overall survival curves were estimated by the Kaplan-Meier method and compared statistically by the log-rank test. The interactions between RNA-seq expression and clinicopathologic features were tested by Student's t-test for two groups or by one-way ANOVA test for more than two groups, and clinicopathologic features were compared by x2 tests or Wilcoxon tests. The area under the curve (AUC) of the receiver operating characteristic (ROC) curve was used to identify the optimum values for the signature to predict survival [20]. All analyses were performed using R (version 3.3.1) and Bioconductor [21].

3. Results

3.1. Baseline characteristics

Two of the 57 US samples did not have complete survival data, and we removed these cases. In all, 55 US samples from gene expression RNAseq - HTSeq - Counts and 78 normal samples from the GTEx dataset were included for the identification of DEGs. Fifty-five patients diagnosed with US were randomized into training (n = 39) and testing (n = 16) groups. The median (interquartile range) follow-up was 22.23 (95% CI: 18.33-27.57) months. Overall, 34 (61.82%) patients died from disease. The workflow of our study is presented in Supplemental Fig. S1. Table 1 shows the baseline characteristics of the cohort.

3.2. Identification of DEGs

We initially compared the transcriptomes of 55 US samples with 78 controls. A total of 10,466 DEGs from gene expression RNAseq - HTSeq - Counts, including 7,632 upregulated genes and 2,834

Table 1
Baseline characteristics of US patients.

Features	Training	Testing	*P-value*
N	39	16	
Age			1
≤ 65	11 (28.21%)	5 (31.25%)	
>65	28 (71.79%)	11 (68.75%)	
Stage			0.769
I + II	17 (43.59%)	8 (50.00%)	
III + IV	22 (56.41%)	8 (50.00%)	
Histological Type			0.079
Heterologous	15 (38.46%)	5 (31.25%)	
Homologous	6 (15.38%)	7 (43.75%)	
NOS	18 (46.15%)	4 (25.00%)	
Hormone therapy			1
No	16 (80.00%)	11 (78.57%)	
Yes	4 (20.00%)	3 (21.43%)	
Hypertension			0.214
No	13 (37.14%)	9 (60.00%)	
Yes	22 (62.86%)	6 (40.00%)	
Icd 10			0.543
C54	26 (66.67%)	9 (56.25%)	
C55	13 (33.33%)	7 (43.75%)	
Oct embedded			0.733
FALSE	10 (27.78%)	5 (35.71%)	
TRUE	26 (72.22%)	9 (64.29%)	
Peritoneal wash			0.432
Negative	20 (71.43%)	5 (55.56%)	
Positive	8 (28.57%)	4 (44.44%)	
Postoperative rx tx			0.533
[Discrepancy]	1 (2.70%)	0 (0.00%)	
No	13 (35.14%)	3 (18.75%)	
Yes	23 (62.16%)	13 (81.25%)	
Primary therapy outcome success			0.174
CR	18 (54.55%)	12 (85.71%)	
Partial Remission/Response	2 (6.06%)	1 (7.14%)	
PD	12 (36.36%)	1 (7.14%)	
Stable Disease	1 (3.03%)	0 (0.00%)	
Rdiation Therapy			0.685
[Discrepancy]	1 (2.70%)	0 (0.00%)	
No	18 (48.65%)	10 (62.50%)	
Yes	18 (48.65%)	6 (37.50%)	
Residual tumor.x			0.224
non-R0	19 (54.29%)	5 (33.33%)	
R0	16 (45.71%)	10 (66.67%)	
Surgical approach			0.754
Minimally Invasive	22 (62.86%)	8 (57.14%)	
Open	13 (37.14%)	6 (42.86%)	
System version			0.093
1988	7 (17.95%)	0 (0.00%)	
2009	32 (82.05%)	16 (100.00%)	
Tissue prospective collection indicator			1
No	23 (58.97%)	10 (62.50%)	
Yes	16 (41.03%)	6 (37.50%)	
Vital status			0.36
Deceased	26 (66.67%)	8 (50.00%)	
Living	13 (33.33%)	8 (50.00%)	
Survival status			0.36
Dead	13 (33.33%)	8 (50.00%)	
Alive	26 (66.67%)	8 (50.00%)	

Abberivations: NOS, not otherwise specified; CR, complete response; PD, progressive disease.

downregulated genes, were identified as being statistically significant between US tissues and normal tissues (| Log_2-fold change | < 1, *P*-value ≤ 0.05) (Fig. 1A; Supplementary Table S2). Then, 4674 genes were normalized using FPKM (Fragments Per Kilobase of transcript per Million mapped reads). Next, 714 protein coding genes were annotated using the gencode. v22. annotation.gtf file (http://www.gencodegenes.org/). Relative expression values for the top 50 DEGs between US tissues and normal tissues were illustrated in Fig. 1B and C, indicating substantial individual variation of the transcripts. We then developed univariate Cox proportional hazards regression analysis to select 7 prognosis-associated genes using data from gene expression RNAseq - HTSeq - FPKM, which were statistically significantly correlated with OS (*P*-value < 0.01). The patients with US were randomly divided into the training set (n = 35) and validating set (n = 16) as 7:3. LASSO Cox proportional hazards regression methods were used in the training set, and 6 survival-related genes were identified with the highest frequency of being selected by the method among 500,000 bootstrap replicates as follows: FGF23, TLX2, TIFAB, RNF223, HIST1H3A and AADACL4, which all reached statistical significance (P < 0.05).

3.3. Prognostic signature model development and validation based on RNA-seq

After LASSO regularization and gene selection, a prognostic signature model was established by LASSO proportional hazards Cox regression analysis. Our prognostic signature model obtained from the training cohort found that risk score= (-0.028376892 × expression score of FGF23) + (0.137019893 × expression score of TLX2) + (-0.044655391 × expression score of TIFAB) + (-0.08820314 × expression score of RNF223) + (-0.020606389 × expression score of HIST1H3A) + (0.049295477 × expression score of AADACL4). In this model, TLX2 and AADACL4 were positive coefficients, and others were negative coefficients. The results showed that patients with high expression of TLX2 and AADACL4 suffered from high survival. Nevertheless, patients with high expression of other genes had low survival.

Then, the risk score for every patient based on our prognostic signature model was calculated. The patients in the training and testing groups were separated into high-risk and low-risk subgroups by the median cut-off value of the training group to evaluate the prognostic value of the risk score by performing survival analysis.

The Kapan-Meier survival curve resulting in the training and testing groups suggested the same results: US patients with low-risk scores had much better OS than US patients with high-risk scores (*P* < 0.0001). The median OS for all patients in the two groups were 27 (95% CI: 18.36 − 35.6) months. In the training group, the median OS for the high- and low-risk groups were 19.6 (95% CI: 13.7 − 26.6) vs 88.1 (95% CI: 52.3 − NR (not reach)) months (HR, 0.1412, 95% CI: 0.03295 - 0.6054; *P* = 0.002), respectively (Fig. 2A). In the testing group, the median OS in the high- and low-risk groups were 30 (95% CI: 12.4 - NR) vs NR (95% CI: NR - NR) months (HR, <0.0001, 95% CI: 0 - inf; *P* = 0.03), respectively (Fig. 2C).

The 6-gene prognostic signature's sensitivity and specificity for predicting survival were plotted in a time-dependent ROC curve to assess the risk score's ability to predict survival condition. In the training group, the AUC values for 12-, 36- and 60-month OS prediction were 0.881, 0.864 and 0.829, respectively (Fig. 2B). In the testing group, the AUC values for 12-, 36- and 60-month OS prediction were 0.971, 0.849 and 0.854, respectively (Fig. 2D). These results indicated that the 6-gene prognostic signature can predict the prognosis of US patients.

3.4. Risk score distribution in the training and testing groups

The distribution of patient risk scores in the training group, the survival status and RNA-seq expression of 39 US patients were shown in Fig. 3A, B, 3C and ranked according to the 6-gene prognostic signature. Fig. 3A showed patients sorted based on risk score, with yellow indicating patients with a risk score below the median and blue indicating those with a risk score above the median. Fig. 3B showed that patients who had high-risk scores were inclined to express hazardous genes, while patients with low-risk scores were inclined to express protective genes. The death rate of patients with high-risk scores was higher than in those with low-

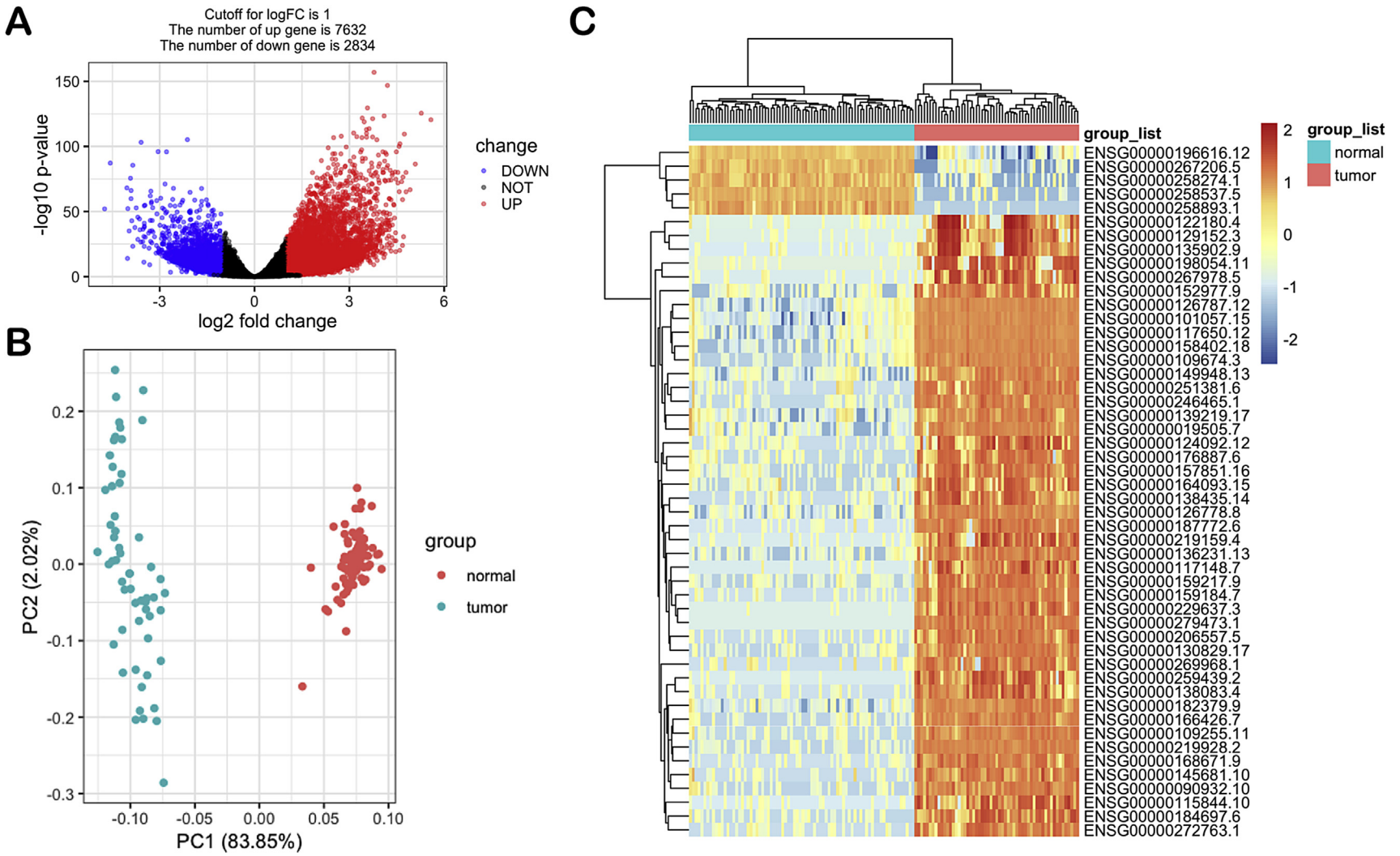

Fig. 1. Identification of the prognostic gene signatures in the dataset. (A) Volcano plot of the survival-associated genes in univariate Cox regression analysis. (B) Principal component analysis (PCA) of the top 50 DEGs between US tissues and normal tissues. (C) Heatmap of the top 50 DEGs between US tissues and normal tissues.

risk scores (Fig. 3C). Additionally, we observed similar results in the testing group (Fig. 3D, E and 3F).

3.5. Survival analysis between high- and low-risk groups according to the 6-gene signature

The 6-gene signature could assign US patients into high-group (n = 42) and low-risk group (n = 13) (Supplementary Table S3). We found that the patients in the low-risk group showed significant better survival compared with the patients in the high-risk group in OS (HR, 0.0929, 95% CI: 0.022 - 0.393; $P < 0.0001$; Supplementary Fig. S2A), PFI (HR, 0.196, 95% CI: 0.0682 - 0.563; $P < 0.0001$; Supplementary Fig. S2B), DSS (HR, 2.37e-09, 95% CI: 0-inf; $P < 0.0001$; Supplementary Fig. S2C) and DFI (HR, 0.0129, 95% CI: 0.015 - 0.99; $P = 0.02$; Supplementary Fig. S2D) (Supplementary Table S4).

3.6. The 6-gene signature was an independent prognostic factor in US

The relationships between OS and clinicopathological factors were analyzed by using Cox proportional hazard regression model (Supplementary Table S5). Univariate analysis showed that stage, radiation therapy, postoperative rx tx and 6-gene signature were significantly associated with OS ($P < 0.05$). While multivariate analysis showed that only radiation therapy and 6-gene signature remained significantly associated with OS ($P < 0.05$). The Cox proportional hazard regression model was also used to analyze the correlations between PFI and clinical factors (Supplementary Table S6). Univariate analysis showed that stage, ICD 10, radiation therapy, postoperative rx tx and 6-gene signature were significantly associated with PFI ($P < 0.05$). While multivariate analysis showed that age, histological type (heterologous vs NOS); radiation therapy, surgical approach, and 6-gene signature showed significantly associated with OS ($P < 0.05$). The results indicated that 6-gene signature was a good prognostic predictor for US patients.

3.7. Nomogram based on 6-gene signature for prognostic prediction of US patients

The nomogram including radiation therapy and 6-gene signature was constructed to predict the OS of US patients (Fig. 4A). The time-dependent AUC for the nomogram was all more than 0.5 of 0.5-5 years (Fig. 4B). The result meant that the nomogram had a good ability to discriminate patients of poor prognosis from patients of favor prognosis. The median OS for the high- and low-risk

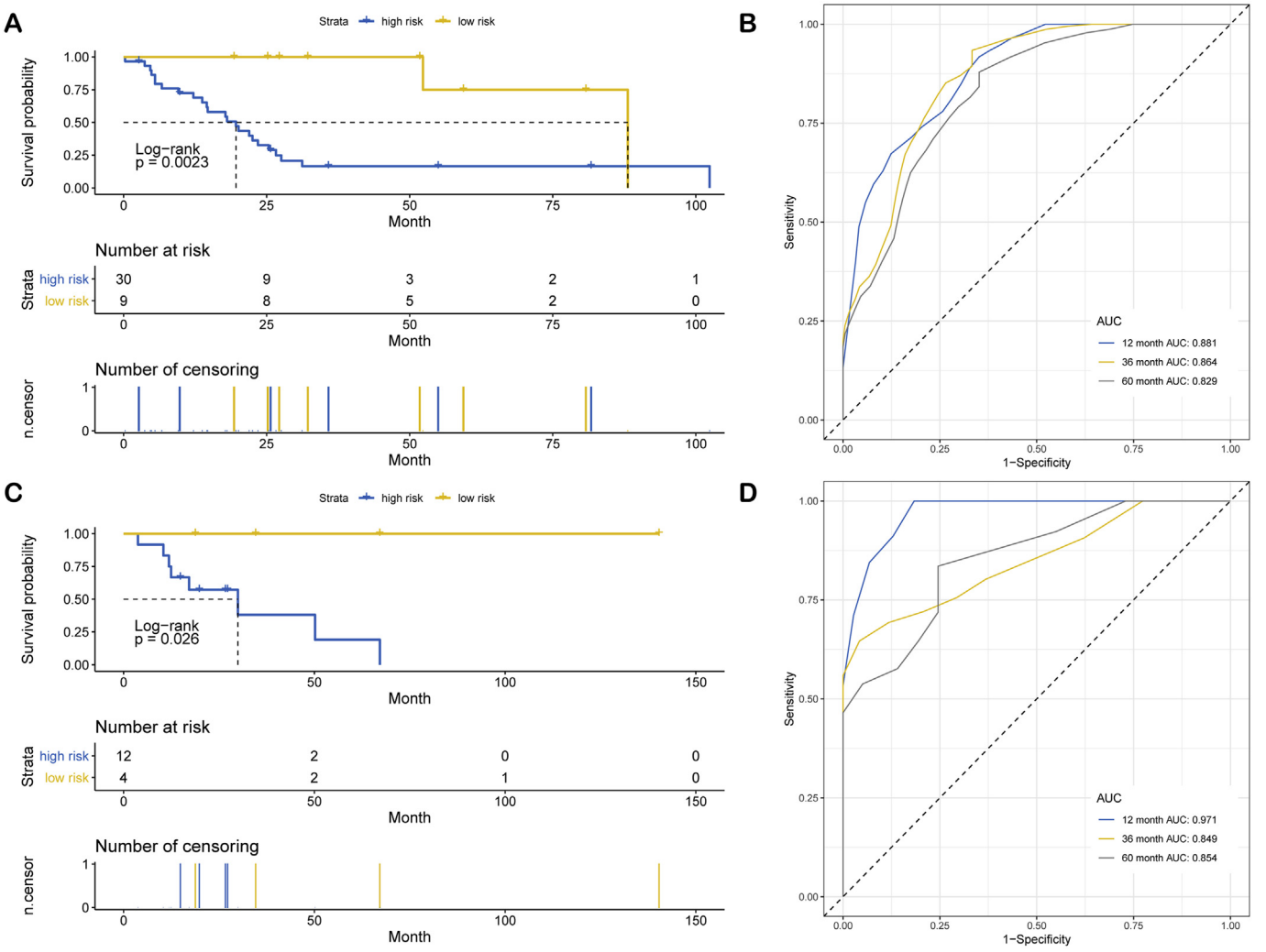

Fig. 2. Kaplan-Meier estimates for the overall survival and Time-dependent ROC curves in the training and testing sets. (A) The overall survival in training cohort stratified by the 6-gene prognostic signature into high- and low-risk with the *P*-value. (B) Time-dependent ROC curves in the training cohort. (C) The overall survival Testing cohort stratified by the 6-gene prognostic signature into high- and low-risk with the *P*-value. (D) Time-dependent ROC curves in the testing cohort. We used AUC values at 12, 36, and 60 months to assess prognostic accuracy and calculated *P*-values using the log-rank test. ROC: receiver operator characteristic. AUC: area under the ROC curve.

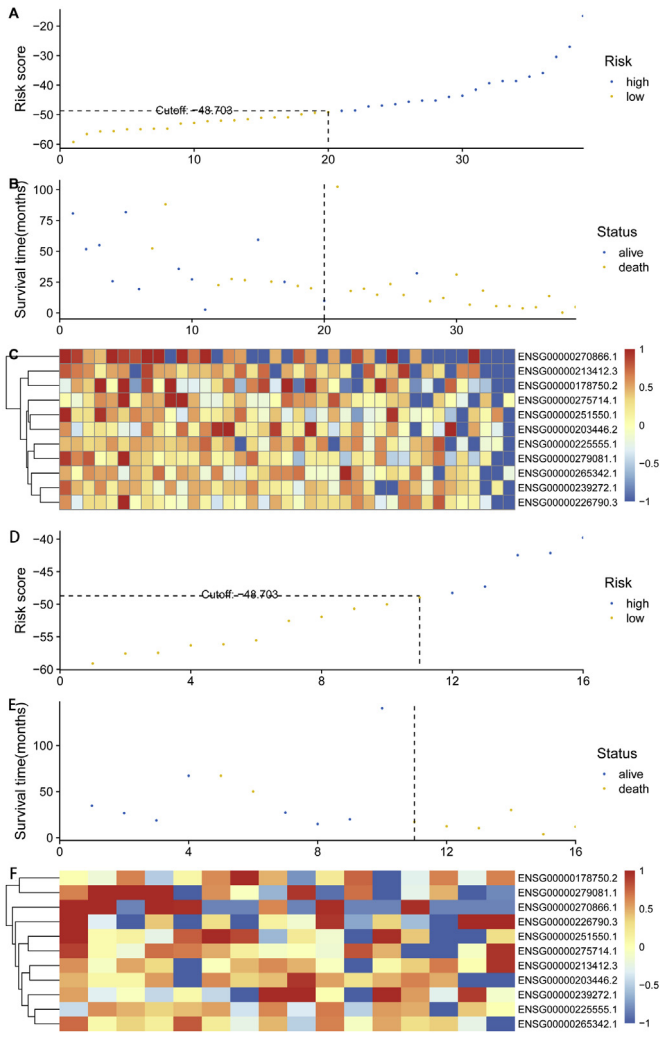

Fig. 3. Characteristics of the 6-gene prognostic signature in the training and testing groups. (A) The risk score of each US patient in the training group. (B) Patients' overall survival and survival status in the training group. (C) Heat map of gene expression profiles of US patients in the training group. (D) The risk score of each US patient in the testing group. (E) Patients' overall survival and survival status in the testing group. (F) Heat map of gene expression profiles of US patients in the testing group. The black dotted line represents the median gene signature cutoff dividing patients into low-risk and high-risk groups.

groups were 17.2 (95% CI: 13.7 − 26.6) vs 67.2 (95% CI: 30.0 - NA) months (HR, 0.0854, 95% CI, 0.0115 - 0.633; P = 0.002), respectively (Fig. 4C). It was shown that patients in the low-risk group had better survival when compared with those in the high-risk group.

3.8. Functional enrichment analysis

To further describe the function of genes in the prognostic signature, functional enrichment analysis is conducted for the sake of biological processes and pathways. First, in the TCGA dataset (Pearson correlation coefficient>0.5), the gene expression prognostic values of Pearson's correlation coefficients were calculated. Additionally, the KEGG signaling pathway analysis suggested that inflammatory bowel disease was the most significant pathway. And the genes also participated in following pathways: Th1 and Th2 cell differentiation and toll-like receptor signaling pathway (Fig. 5A; Supplementary Table S7). The GO biological process enrichment analysis was composed of three parts: biological process (BP), cellular component (CC), and molecular function (MF). The BP enrichment analysis showed that the genes were mainly enriched in signal release and axon development (Fig. 5B; Supplementary Table S8). CC enrichment analysis showed that the genes were mainly enriched in presynapse and cell-cell junction (Fig. 5C; Supplementary Table S9). MF enrichment analysis showed that the genes were mainly enriched in transcription factor activity, RNA polymerase II proximal promoter sequence−specific DNA binding and protein serine/threonine kinase activity (Fig. 5D; Supplementary Table S10). Above all, genes in the model can modulate the expression level of their target genes, which plays a role in inflammation-driven tumors [22] and tumorigenesis [23].

3.9. Somatic mutation analysis

In our study, the top ten mutated genes in high- and low-risk groups were in the Supplementary Figs. S3A and S3B. The high-risk group had somatic mutations in the following order: TP53 > FBXW7 > PIK3CA > PPP2R1 A > TTN and in the low-risk group: TP53 > FBXW7 > PPP2R1 A > CHD4 > DOPEY1. Furthermore, we observed that the patients in the high-risk group displayed distinct mutation signatures comparing to patients in the low-risk group (Supplementary Fig. S3C and Table S11).

4. Discussion

The prognosis of US remains poor, and there are few effective and responsible prognostic biomarkers or models for improving the clinical outcomes of US patients.

In the current study, our team first developed and validated a 6-gene-based prognostic signature for prognosis prediction of uterine sarcoma.

It is difficult to discover links between genes and disease prognosis. Methods of regularized regression, such as LASSO, are popular for this task [24]. In this study, LASSO was applied for gene selection to predict US survival. Precision LASSO is a type of lasso variable that promotes the selection of thin variable covariance and inverse covariance matrix of explanatory variables. By applying multiple biostatistics methods, such as LASSO proportional hazards Cox regression, on the OS of US cases in TCGA, 6 genes, FGF23, TLX2, TIFAB, RNF223, HIST1H3A and AADACL4, were identified as independent prognosis predictors in US. However, no study had been reported about the biological mechanism of these genes except FGF23 and TLX2. Fibroblast growth factor 23 (FGF23) was shown to be relevant in initiation and progression of various cancers [25,26]. T cell leukemia homeobox 2 (TLX2) was reported as a prognostic factor of gastrointestinal stromal tumors [27].

The risk score for every patient in the training group and testing group was then calculated, and the cohort was divided into high-risk and low-risk groups. We used Kaplan-Meier survival curves and log-rank tests to compare the survival differences between these groups. As we showed in the results, it is observed that obvious division exists in the survival curves of the high- and low-risk groups classified by the risk score in both the training and testing group. Hence, the 6-gene risk score signature was of prognostic significance.

Finally, we found that radiation therapy showed well in predicting the survival of the US patients. Thus, a nomogram including radiation therapy and 6-gene signature was established for real-life clinical practice. The time-dependent AUC suggested that the prognostic value of the nomogram was evident, and confirmed the clinical usability of the nomogram. These results were consistent with previous studies that radiation therapy could improve the locoregional control and survival in US patients [28,29]. However, several researchers confirmed that radiation therapy in patients with US does not improve the survival [30,31]. The role of radiation therapy for patients with US still need further investigate [32].

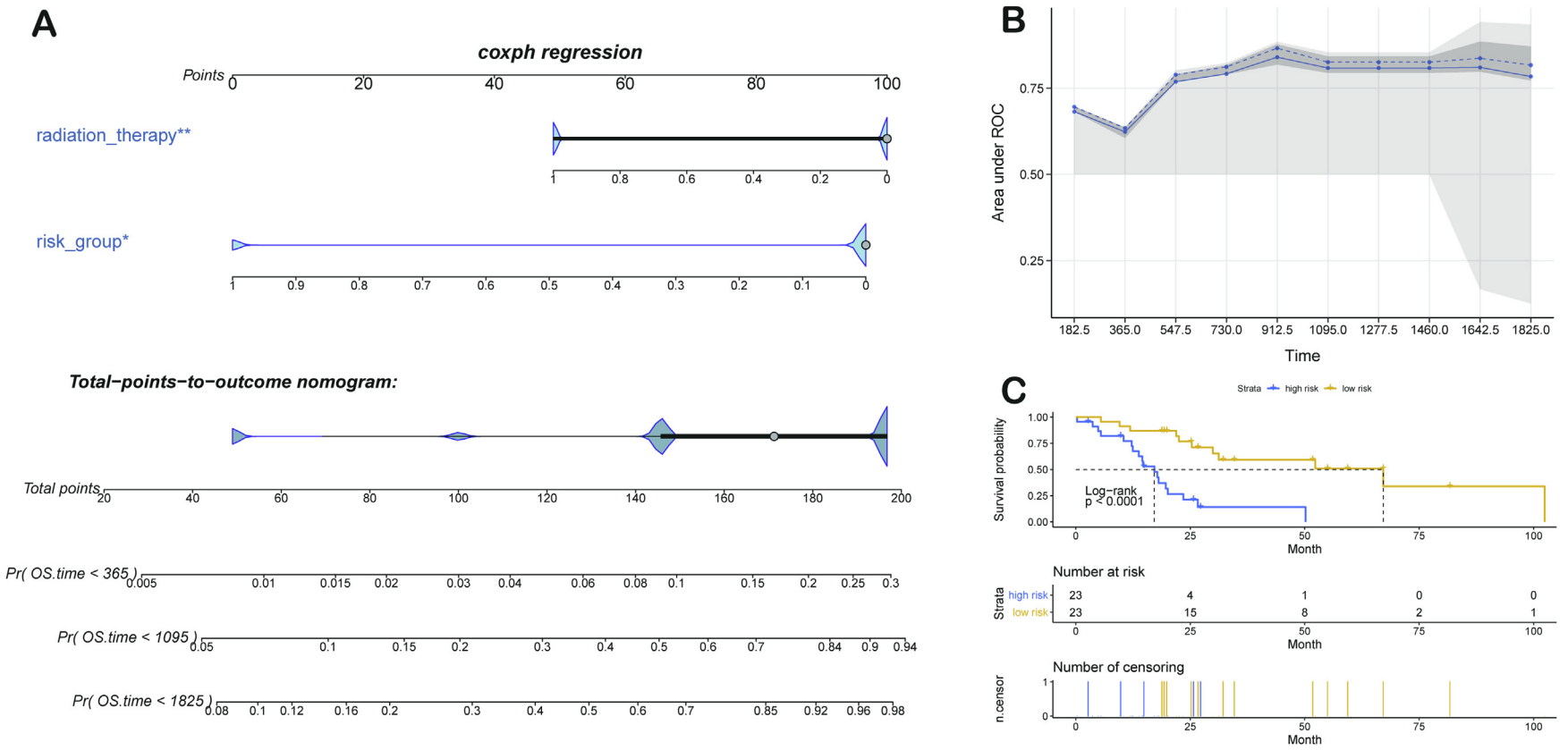

Fig. 4. Nomogram to predict overall survival in US patients. (A) Nomogram based 6-gene signature, radiation therapy for 1-,3- and 5-year overall survival prediction. (B) Time-dependent AUC for the nomogram. (C) Kaplan-Meier analysis of overall survival in high- and low-risk group patients according to the nomogram.

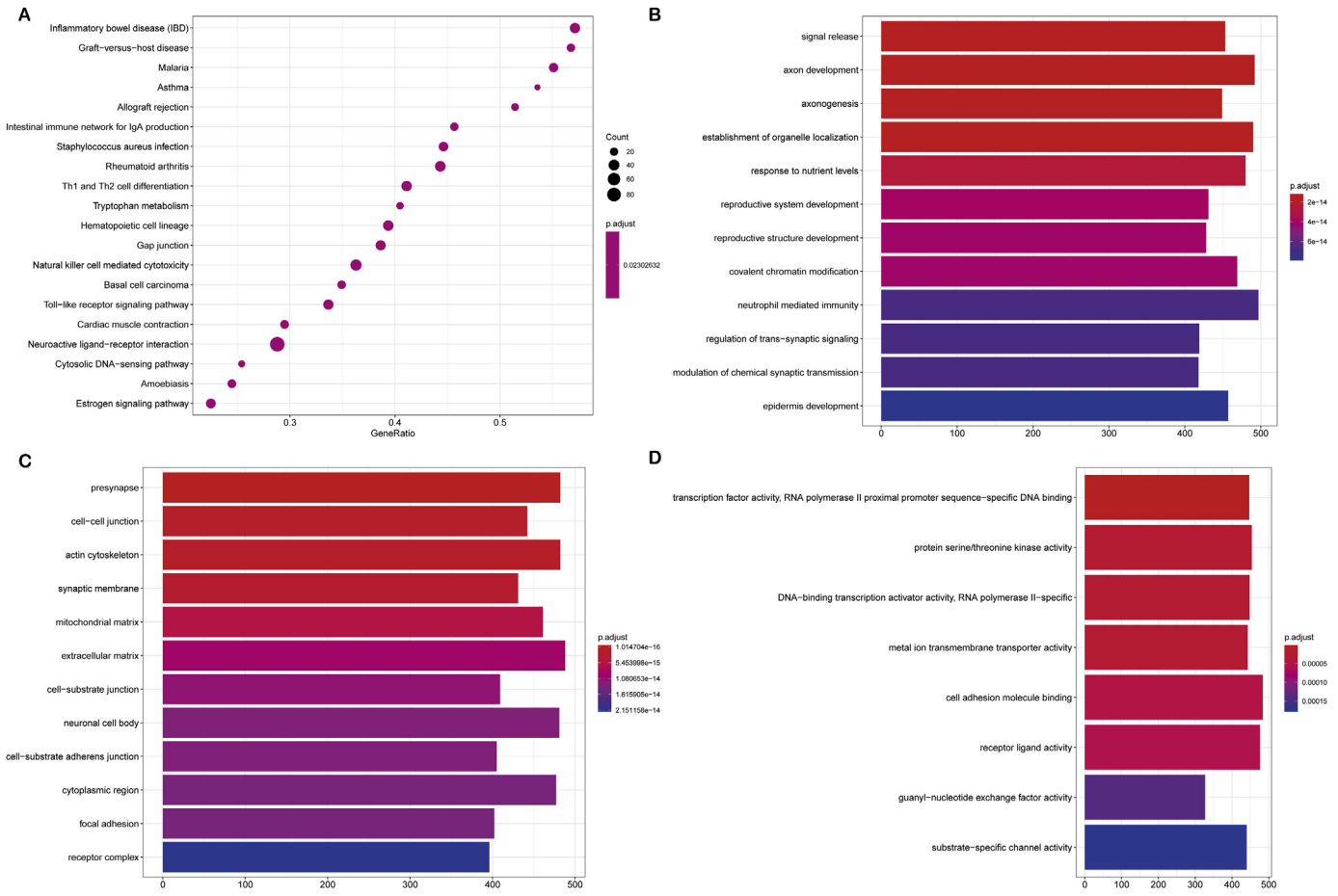

Fig. 5. Functional enrichment of the genes with (A) KEGG pathway analysis, (B) GO biological process enrichment. (C) GO cellular component enrichment. (D) GO molecular function enrichment. GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Further gene enrichment analysis indicated that the genes in the signature can play a role in cancer-related biological processes, such as transcriptional misregulation in cancer, cytokine-cytokine receptor interaction, signaling pathways regulating pluripotency of stem cells, and the chemokine signaling pathway. Due to high-throughput technology development, the gene expression profile from microarray and RNA sequencing has been a powerful tool to identify cancer-related molecular markers [33]. Compared with microarrays, the important advantage of RNA-seq is the ability to measure almost all types of RNA, such as noncoding RNAs, in one experiment. Our model is based on RNA-seq, which is repeatable and stable. To find out somatic mutations in specific genes between high- and low-groups, maftools package was used to detect and further examine the mutations. The results showed that the different mutated genes could contribute to the different sore of genes in the US patients. It was suggested that TP53, TTN, PPP2R1 A and CHD4 were important driver genes that contribute to cancer development [34–37]. What's more, PIK3CA mutation was significantly associated with poor prognostic in patients undergoing lung stereotactic body radiation therapy [36,38].

There are also limitations to our research. Although the number of samples in our study is currently the largest for US, there are only 55 cases involved, which is fewer than other cancer data sets. In addition, the testing group is not independent or external due to the low incidence rate of US. With a small sample size of US patients from TCGA, caution must be applied because the results can be limited to the 6-gene prognostic signature. Additionally, little is known about the functions and mechanisms of these genes in US. Above all, this research has raised many questions, and further studies of our prognostic signature model with larger cohorts of US are required in the future.

5. Conclusion

We constructed and validated a 6-gene-based prognostic signature for uterine sarcoma. This signature could have many potential prognostic and therapeutic implications for US patient management.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No.81360351, 81660512), the Special Fund for Traditional Chinese Medicine and Ethnic Medicine supported by the Administration of Traditional Chinese Medicine of Guizhou Province (Grant No. QZYY2017-113), Projects ZY-201751044 and ZYKY-20173830 supported by the Zunyi Medical University Training Program of Innovation and Entrepreneurship for Undergraduates, the Open Project Program of the Special Key Laboratory of Oral Diseases Research, Higher Education Institution in Guizhou Province (Grant No. K-029), and the Research Programs of Science and Technology Commission Foundation of Zunyi City (Grant No. HZ2019-11, HZ2019-07). The study funders had no role in the design of the study, the collection, analysis, or interpretation of the data, the writing of the manuscript, or the decision to submit the manuscript for publication.

Abbreviations

US	uterine sarcoma
DEG	differentially expressed gene
NR	not reach
ROC	receiver operating characteristic
AUC	area under the ROC curve
TCGA	The Cancer Genome Atlas
GO	gene ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
GTEx	Genotype-Tissue Expression project
HR	Hazard ratio
CI	confidence interval
OS	overall survival
PFI	progression-free interval
DSS	disease-specific survival
DFI	disease-free interval
LASSO	the least absolute shrinkage and selection operation
RNA-seq	RNA sequencing
MAF	Mutation Annotation Format
BP	biological process
CC	cellular component
MF	molecular function

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygyno.2019.08.033.

Author contribution

Dr. Zhou designed the study and analyzed the data. Dr. Zhao contributed significantly to the methods, results and discussion sections. Dr. Jin provided significant contribution to the abstract, introduction and results section. Dr. Tian helped extensive edits. Dr. Ma served as senior author and contributed significantly to the entire manuscript. All authors revised and reviewed this work, and all authors gave their final approval of the submitted manuscript.

References

[1] R.J. Kurman, M.L. Carcangiu, C.S. Herrington, R.H. Young, WHO Classification of Tumours of Female Reproductive Organs, 2014.

[2] V.M. Abeler, O. Rā,Yne, S. Thoresen, H.E. Danielsen, J.M. Nesland, G.B. Kristensen, Uterine sarcomas in Norway. A histopathological and prognostic survey of a total population from 1970 to 2000 including 419 patients, Histopathology 54 (2010) 355–364.

[3] J.L. Zhongqiu Lin, Rongchun Lin, Interpretation of FIGO 2015 report on women's cancer series IV: interpretation of guidelines for diagnosis and treatment of uterine sarcoma, Chin. J. Pract. Gynecol. Obstet. 31 (2015) 1082–1087.

[4] F. Amant, C. An, M. Debiec-Rychter, D. Timmerman, I. Vergote, Clinical management of uterine sarcomas, Lancet Oncol. 10 (2009) 1188–1198.

[5] A. Gadducci, S. Cosio, A. Romanini, A.R. Genazzani, The management of patients with uterine sarcoma: a debated clinical challenge, Crit. Rev. Oncol.-Hematol. 65 (2008) 129–142.

[6] C.D.M. Fletcher, The evolving classification of soft tissue tumours: an update based on the new WHO classification, Histopathology 64 (2013) 2–11.

[7] K.,. Bodner, B.,. Bodner-Adler, A.,. Obermair, G.,. Windbichler, E.,. Petru, S.,. Mayerhofer, et al., Prognostic parameters in endometrial stromal sarcoma: a clinicopathologic study in 31 patients, Gynecol. Oncol. 81 (2001) 160–165.

[8] W. Feng, A. Malpica, S.J. Robboy, E. Gudlaugsson, K. Hua, X. Zhou, et al., Prognostic value of the diagnostic criteria distinguishing endometrial stromal sarcoma, low grade from undifferentiated endometrial sarcoma, 2 entities within the invasive endometrial stromal neoplasia family, Int. J. Gynecol. Pathol. 32 (2013) 299–306.

[9] Kelly K. Hunt, Kristelle Lusby, Yiqun Zhang, et al., Uterine leiomyosarcoma management, outcome, and associated Molecular;Biomarkers: a single institution's experience, Ann. Surg. Oncol. 20 (2013) 2364–2372.

10 M.S. Leapman, H.G. Nguyen, J.E. Cowan, X. Lingru, S. Bradley, S. Jeffry, et al., Comparing Prognostic Utility of a Single-Marker Immunohistochemistry Approach with Commercial Gene Expression Profiling Following Radical Prostatectomy, Eur. Urol. 74 (5) (November 2018) 668–675, https://doi.org/10.1016/j.eururo.2018.08.020.

[11] J. Li, J. Wang, Y. Chen, L. Yang, S. Chen, A prognostic 4-gene expression signature for squamous cell lung carcinoma, J. Cell. Physiol. (2017) 232.

[12] D.Y. Zhang, N. Goossens, J. Guo, M.C. Tsai, H.I. Chou, C. Altunkaynak, et al., A hepatic stellate cell gene expression signature associated with outcomes in hepatitis C cirrhosis and hepatocellular carcinoma after curative resection, Gut 65 (2016) 1754.

[13] R. Wang, X.H. Ye, X.L. Zhao, J.L. Liu, C.Y. Zhang, Development of a five-gene signature as a novel prognostic marker in ovarian cancer, Neoplasma 66 (3) (December 2018), https://doi.org/10.4149/neo_2018_180705N447.

[14] J. Su, L.F. Miao, X.H. Ye, M.S. Cui, X.F. He, Development of prognostic signature and nomogram for patients with breast cancer, Medicine 98 (2019), e14617.

[15] D.J. McCarthy, G.K. Smyth, M.D. Robinson, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (2009) 139–140.

[16] J. Liu, T. Lichtenberg, K.A. Hoadley, L.M. Poisson, A.J. Lazar, A.D. Cherniack, et al., An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics, Cell 173 (2018), 400-16.e11.

[17] P. Blanche, J.F. Dartigues, H. Jacqmin-Gadda, Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks, Stat. Med. 32 (2013) 5381–5397.

[18] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS A J. Integr. Biol. 16 (2012) 284–287.

[19] A. Mayakonda, D.C. Lin, Y. Assenov, C. Plass, H.P. Koeffler, Maftools: efficient and comprehensive analysis of somatic variants in cancer, Genome Res. 28 (2018) 1747–1756.

[20] P.J. Heagerty, T. Lumley, M.S. Pepe, Time-dependent ROC curves for censored survival data and a diagnostic marker, Biometrics 56 (2000) 337–344.

[21] R.C. Gentleman, V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, S. Dudoit, et al., Bioconductor: open software development for computational biology and bioinformatics, Genome Biol. 5 (2004) R80.

[22] E.M. Borroni, B. Savino, R. Bonecchi, M. Locati, Chemokines sound the alarmin: the role of atypical chemokine in inflammation and cancer, Semin. Immunol. 38 (2018) 63–71.

[23] T.I. Lee, R.A. Young, Transcriptional regulation and its misregulation in disease, Cell 152 (2013) 1237–1251.

[24] H. Wang, B.J. Lengerich, B. Aragam, E.P. Xing, Precision Lasso: accounting for correlations and linear dependencies in high-dimensional genomic data, Bioinformatics 35 (2019) 1181–1187.

[25] S. Feng, J. Wang, Y. Zhang, C.J. Creighton, M. Ittmann, FGF23 promotes prostate cancer progression, Oncotarget 6 (2015) 17291–17301.

[26] A. Mansinho, A.R. Ferreira, S. Casimiro, I. Alho, I. Vendrell, A.L. Costa, et al., Levels of circulating fibroblast growth factor 23 (FGF23) and prognosis in cancer patients with bone metastases, Int. J. Mol. Sci. 20 (2019).

[27] J.T. Kaifi, M. Wagner, P.G. Schurr, R. Wachowiak, U. Reichelt, E.F. Yekebas, et al., Allelic loss of Hox11L1 gene locus predicts outcome of gastrointestinal stromal tumors, Oncol. Rep. 16 (2006) 915–919.

[28] J.R. Gunther, E.N. Christensen, P.K. Allen, L.M. Ramondetta, A. Jhingran, N.D. Fleming, et al., Role of radiation therapy in the multidisciplinary management of uterine carcinosarcoma, Int. J. Gynecol. Cancer : Off. J. Int. Gynecol. Cancer Soc. 28 (2018) 114–121.

[29] H.L. Hou, M.B. Meng, X.L. Chen, L.J. Zhao, L. Zhu, B.L. Zhang, et al., The prognosis factor of adjuvant radiation therapy after surgery in uterine sarcomas, OncoTargets Ther. 8 (2015) 2339–2344.

[30] M. Hosh, S. Antar, A. Nazzal, M. Warda, A. Gibreel, B. Refky, Uterine sarcoma: analysis of 13,089 cases based on surveillance, epidemiology, and end results database, Int. J. Gynecol. Cancer : Off. J. Int. Gynecol. Cancer Soc. 26 (2016) 1098–1104.

[31] K. Galaal, E. van der Heijden, K. Godfrey, R. Naik, A. Kucukmetin, A. Bryant, et al., Adjuvant radiotherapy and/or chemotherapy after surgery for uterine carcinosarcoma, Cochrane Database Syst. Rev. (2013), Cd006812.

[32] J.Y. Chern, L.R. Boyd, S.V. Blank, Uterine sarcomas: the latest approaches for these rare but potentially deadly tumors, Oncology (Williston Park, NY) 31 (2017) 229–236.

[33] C. Zhan, L. Yan, L. Wang, W. Jiang, Y. Zhang, J. Xi, et al., Landscape of expression profiles in esophageal carcinoma by the Cancer Genome Atlas data, Dis. Esophagus : Off. J. Int. Soc. Dis. Esophagus 29 (2016) 920–928.

[34] J. Liu, A. Near, J.A. Chiarappa, K. Wada, J. Tse, C. Burudpakdee, et al., Clinical outcomes associated with pathogenic genomic instability mutations in prostate cancer: a retrospective analysis of US pharmacy and medical claims data, J. Med. Econ. (2019) 1.

[35] S. Gohler, M.I. Da Silva Filho, R. Johansson, K. Enquist-Olsson, R. Henriksson, K. Hemminki, et al., Functional germline variants in driver genes of breast cancer, Cancer Causes Control : CCC (Cancer Causes Control) 28 (2017) 259–271.

[36] M. Rahman, K. Nakayama, M.T. Rahman, N. Nakayama, H. Katagiri, A. Katagiri, et al., PPP2R1A mutation is a rare event in ovarian carcinoma across histological subtypes, Anticancer Res. 33 (2013) 113–118.

[37] Y. Li, Q. Liu, D.J. McGrail, H. Dai, K. Li, S.Y. Lin, CHD4 mutations promote endometrial cancer stemness by activating TGF-beta signaling, Am J. Cancer Res. 8 (2018) 903–914.

[38] N.A. Lockney, T.J. Yang, D. Barron, E. Gelb, D.Y. Gelblum, E. Yorke, et al., PIK3CA mutation is associated with increased local failure in lung stereotactic body radiation therapy (SBRT), Clin. Transl. Radiat. Oncol. 7 (2017) 91–93.