Optimal subset of feature miRNAs consisting of 7 miRNAs that can serve as a novel diagnostic and prognostic predictor for the progression of cervical cancer

Shanshan Qin¹, Yingchun Gao¹, Yijun Yang¹, Juanpeng Yu¹, Fei Meng¹, Ting Zhang¹ and Can Shi¹

¹Department of Obstetrics and Gynecology, Huai'an First People's Hospital, Nanjing Medical University, Huai'an, Jiangsu 223300, P.R. China

Correspondence to: Can Shi, email: shican2015@126.com

Keywords: cervical cancer; biomarker; miRNA; prognosis; progression

Received: September 07, 2017

Accepted: November 13, 2017

Published: January 02, 2018

Copyright: Qin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Objective: Cervical cancer is the second most commonly diagnosed cancer in women. Novel prognostic biomarkers are needed to predict the progression of cervical cancer.

Results: In total, 44 significantly differentially expressed miRNAs were identified. An optimal subset of 7 feature miRNAs was identified, including hsa-miR-144, hsa-miR-147b, hsa-miR-218-2, hsa-miR-425, hsa-miR-451, hsa-miR-483, and hsa-miR-486. The feature miRNAs were used to construct an SVM classifier and showed a good performance in predicting pathologic stages of samples. SVM classification was found to be an independent prognostic factor. Functional enrichment analysis indicated that these feature miRNAs are involved in tumorigenesis.

Materials and Methods: Cervical cancer expression data were obtained from the Cancer Genome Atlas database. MicroRNAs (miRNAs) significantly differentially expressed between early- and advanced-stage samples were identified by expression analysis. An optimal subset of feature miRNAs for pathologic stage prediction was delineated using the random forest algorithm and was used for the construction of a cervical cancer-specific support vector machine (SVM) classifier. The roles of feature miRNAs in cervical cancer were analyzed by functional annotation.

Conclusions: The subset of feature miRNAs could potentially serve as a novel diagnostic and prognostic predictor for cervical cancer.

INTRODUCTION

As one of the most frequently diagnosed cancers in women worldwide, it is estimated that cervical cancer accounts for more than 500,000 new cases and 250,000 deaths each year [1]. HPV (human papillomavirus) infection is recognized as the most significant risk factor that presents in most cervical cancers [2, 3]. Integration of HPV into the cellular genome causes genome instability, transcriptional variations, and epigenetic alterations [4, 5]. However, HPV alone is not sufficient to induce malignant transformation [6]. Therefore, additional cancer-causing genetic variations may underlie the development and progression of cervical cancer.

MicroRNAs (miRNAs) are small, noncoding RNA molecules of approximately 22 nucleotides in length, functioning through targeting mRNAs to regulate gene expression at the post-transcriptional level [7, 8]. MiRNAs are involved in diverse biological processes, including cell cycle, differentiation, and metabolism [9]. Increasing evidence highlights the involvement of altered miRNA expression in cervical cancer [10]. Examples include hsa-miR-21 [11], hsa-miR-196a [12], and hsa-miR-497 [13]. Hsa-miR-21 is an oncogene overexpressed in cervical cancer, the inhibition of which upregulates the tumor suppressor, programmed cell death 4, and suppresses cell proliferation [11]. Upregulation of hsa-miR-196a has also been detected in cervical cancer tissues, in which it

promotes cancer cell proliferation [12]. Unlike hsa-miR-21 and hsa-miR-196a, hsa-miR-497 is a tumor suppressor for cervical cancer and suppresses cancer cell migration and invasion by targeting insulin-like growth factor 1 receptor [13].

Potential prognostic miRNA signatures of cervical cancer have been identified. A miRNA signature for clinical response consisting of hsa-miR-200a and hsa-miR-9 has been previously identified by the expression analysis of candidate miRNAs [14]. According to the expression levels of hsa-miR-200a and hsa-miR-9, cervical cancer samples could be divided into low- and high-risk groups. Functional analysis indicated that both hsa-miR-200a and hsa-miR-9 are likely to play important roles in cervical cancer metastasis. However, only a limited number of miRNAs were analyzed in this study [14], and miRNA expression differences between pathologic stages have not yet been examined.

To identify novel outcome predictors of cervical cancer, we comprehensively analyzed the expression levels of miRNAs using data from The Cancer Genome Atlas (TCGA). miRNAs significantly differentially expressed between early (I and II) and advanced (III and IV) pathologic stages were identified, followed by the identification of an optimal subset of feature miRNAs. The subset of feature miRNAs showed good performance in progression prediction and may serve as a promising prognostic predictor of cervical cancer in clinical practice.

RESULTS

Differential miRNA expression between early- and advanced-stage groups

Among the 143 samples in the training dataset, 90 were early-stage and 53 were advanced-stage samples.

miRNAs with low expression (median expression value < 1.0) were removed, and the remaining 318 miRNAs were used for further analysis. In total, 51 miRNAs were identified to be significantly differentially expressed between early- and advanced-stage groups by *t*-test, whereas 49 miRNAs were identified by Wilcoxon rank-sum test (Supplementary Figure 1). In total, 44 miRNAs were identified to be significantly differentially expressed by both *t*-test and Wilcoxon rank-sum test.

Two-way hierarchical clustering analysis was carried out based on the expression levels of the 44 miRNAs and separated the training dataset into 2 distinct clusters, designated as cluster 1 and cluster 2 (Figure 1A). Cluster 1 mainly consisted of early-stage samples, including 86 early- and 15 advanced-stage samples. Cluster 2 mainly consisted of advanced-stage samples, including 38 advanced- and 4 early-stage samples. The overall accuracy of hierarchical clustering in classifying samples was 86.71% (124 out of 143 samples). Moreover, the 2 clusters correlated significantly with cancer progression status ($\chi^2 = 69.5245$, $p < 2.2e-16$). Kaplan–Meier analysis showed that samples in cluster 1 were related with a significantly better prognosis than those in cluster 2 (Figure 1B; logRank $p = 1.179e-05$), which was consistent with the fact that cluster 1 mainly consisted of early-stage samples, whereas cluster 2 mainly consisted of advanced-stage samples.

Cervical cancer-related optimal subset of feature miRNAs

The random forest algorithm was used to identify an optimal subset of feature miRNAs from the 44 significantly differentially expressed miRNAs, using the training dataset. The results showed that the OOB error reached a minimum (0.181) when 7 miRNAs were used for fitting (Figure 2A). These 7 miRNAs, including miRNAs

Figure 1: Hierarchical clustering and survival analyses of samples in the training dataset. (A) Two-way hierarchical clustering of samples in the training dataset based on the expression levels of the 44 significantly differentially expressed miRNAs yielded 2 clusters, labeled as cluster 1 and cluster 2. Early- and advanced-stage samples are indicated in red and black in the horizontal bar. (B) Kaplan–Meier survival curves of the 2 clusters in (A). Survival curves for cluster 1 and 2 are indicated in green and red, respectively.

were hsa-miR-144, hsa-miR-147b, hsa-miR-218-2, hsa-miR-425, hsa-miR-451, hsa-miR-483, and hsa-miR-486, were selected as optimal miRNAs and are summarized in Table 1. According to the expression levels of these 7 miRNAs in the training dataset, hsa-miR-147b and hsa-miR-218-2 showed a significantly higher expression in the early than in the advanced stage, whereas the remaining 5 miRNAs showed a significantly lower expression level in the early stage (Figure 2B).

Hierarchical clustering analysis based on the expression levels of the feature miRNAs showed that samples in the training dataset could be separated into 2 clusters, (Figure 2C). Similar to the hierarchical clustering results based on the expression levels of all 44 significantly differentially expressed miRNAs, cluster 1 mainly consisted of early-stage samples (74 early- vs. 10 advanced-stage samples) whereas cluster 2 mainly consisted of advanced-stage samples (16 early- vs. 43 advanced-stage samples). The overall classification accuracy was 81.82% (117 out of 143 samples). Additionally, cluster 1 was related with significantly better prognosis than cluster 2 (Figure 2D; logRank $p = 0.01175$).

A cervical cancer-specific SVM classifier for pathologic stage prediction

A cervical cancer-specific SVM classifier was constructed based on the expression levels of feature miRNAs in the optimal subset. Samples in the training dataset were classified as early-stage-like or advanced-stage-like using the SVM classifier. The results showed that the SVM classifier could classify samples in the training dataset with an overall accuracy of 85.31% (122 out of 143 samples; sensitivity = 79.81%, specificity = 94.44%, positive prediction value (PPV) = 88.095%, negative prediction value (NPV) = 84.16%, and area under the receiver operating characteristic (ROC) curve (AUC) = 0.897) (Figure 2E). Kaplan–Meier survival analysis showed that the predicted early-stage-like group had a significantly better prognosis than the advanced-stage-like group (Figure 2F; logRank $p = 0.004715$).

Validation of the performance of the feature miRNAs in pathologic stage prediction

The validation dataset was used to validate the performance of the 7 feature miRNAs in predicting pathologic stage. The validation samples were first classified into cluster 1 and cluster 2 by hierarchical clustering (Figure 3A). Similar to the results for the training dataset, cluster 1 mainly consisted of early-stage samples (78 early- vs. 15 advanced-stage samples) and cluster 2 mainly consisted of advanced-stage samples (37 early- vs. 12 advanced-stage samples). The overall accuracy was 80.99% (115 out of 142 samples). Prognosis

in the 2 clusters was analyzed using Kaplan–Meier survival curve analysis. In consistence with the findings based on the training dataset (Figure 2D), classification in cluster 1 indicated significantly better prognosis (Figure 3B).

Then, we classified the validation samples into early-stage-like and advanced-stage-like groups using the SVM classifier. Consistent with the training dataset findings, the validation samples were classified with an accuracy of 80.98% (115 out of 142 samples; sensitivity = 73.46%, specificity = 91.11%, PPV = 80.49%, NPV = 81.19%, and AUC = 0.857) (Figure 3C). Kaplan–Meier survival curve analysis showed that the early-stage-like samples corresponded with significantly better prognosis than the advanced-stage-like group (Figure 3D; logRank $p = 0.006246$).

Independence of SVM-predicted group in progression prediction

Correlations between clinical variables and prognosis were analyzed via Cox regression. Both univariate and multivariate Cox regression showed that SVM-predicted group, pathologic T, and new tumor were independent prognostic factors, as they correlated significantly with prognosis ($p < 0.05$) in both the training and the validation dataset (Table 2). Kaplan–Meier survival curve analysis for both the training and the validation dataset showed that samples under pathologic T0 and T1 categories were related to a significantly better prognosis than those under pathologic T2 and T3 categories, and samples without new tumor correlated with a better prognosis than those with new tumor (Supplementary Figure 2).

Stratified analysis was performed to evaluate the independence of SVM-predicted group as a prognostic factor. Samples in both the training and the validation dataset were stratified according to age, pathologic M, pathologic N, pathologic T, smoking, new tumor, radiation therapy, and targeted molecular therapy. Univariate Cox regression showed that SVM-predicted group correlated significantly with prognosis for elder patients (> 45 years of age) in both the training and the validation dataset (Table 3). Similar results were found for patients without new tumor and radiation therapy (Table 3). Additionally, SVM-predicted group correlated significantly with pathologic N1 patients in the training dataset and pathologic N0 patients in the validation dataset (Table 3). Kaplan–Meier survival curves for samples stratified by age, pathologic N, new tumor, and radiation therapy are shown in Figure 4.

Functional annotation of the 7 feature miRNAs

mRNA expression profiles of early- and advanced-stage samples were compared, and 535 significantly

Figure 2: Optimal subset of feature miRNAs and its performance in sample classification. (A) OOB error curve for feature miRNA selection using the training dataset. The OOB error is plotted against the number of feature miRNAs used. The red vertical dashed line indicates the minimum OOB error (0.181), where the number of selected feature miRNAs is 7. (B) Expression levels of the feature miRNAs of the optimal subset in early- (red) and advanced- (black)-stage groups. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$. (C) Two-way hierarchical clustering of samples in the training dataset based on the expression levels of the 7 feature miRNAs yielded 2 clusters, which are labeled as cluster 1 and cluster 2. Early- and advanced-stage samples are indicated in red and black in the horizontal bar, respectively. (D) Kaplan-Meier survival curves of the 2 clusters in (C). Survival curves for clusters 1 and 2 are indicated in green and red, respectively. (E) ROC curves of the training dataset generated using the SVM classifier. The AUC was calculated to be 0.897. (F) Kaplan-Meier survival curves for early-stage-like (red curve) and advanced-stage-like (black curve) groups as predicted by the SVM classifier.

differentially expressed mRNAs were identified. Target mRNAs of the feature miRNAs were predicted using miRanda [15]. As a result, we found that 63, 86, 59, 14, 33, and 3 significantly differentially expressed mRNAs were targeted by hsa-miR-218, hsa-miR-144, hsa-miR-425, hsa-miR-451, hsa-miR-483, and hsa-miR-486, respectively, while none were targeted by hsa-miR-147b. We constructed a cervical cancer-related miRNA–mRNA regulation network, consisting of feature miRNAs and significantly differentially expressed mRNAs targeted by these feature miRNAs (Figure 5A). The network consisted of 207 nodes, containing 6 feature miRNAs and 201 significantly differentially expressed target mRNAs.

To interpret the potential roles of the feature miRNAs in cervical cancer, functional enrichment analysis was conducted using DAVID [16]. The results showed that mRNAs in the network were significantly enriched for cancer-related GO biological processes and KEGG pathways (Figure 5B and 5C). The GO terms included

cell-cell signaling, epithelium development, ion transport, and adhesion (Figure 5B). The KEGG terms included pathways in cancer, calcium signaling pathway, basal cell carcinoma, Hedgehog signaling pathway, Wnt signaling pathway, heparan sulfate biosynthesis, and TGF-beta signaling pathway (Figure 5C).

DISCUSSION

Although efforts have been made to identify cervical cancer-related miRNAs [10–14], few studies have focused on miRNAs related to the progression of cervical cancer, which is essential for the development of novel diagnostic biomarkers for this disease. In the present study, we comprehensively analyzed cervical cancer-related miRNA expression profiles and identified 44 miRNAs significantly differentially expressed between early- and advanced-stage samples, using a training dataset. From these miRNAs, an optimal subset of feature miRNAs was

Figure 3: Validation of the optimal subset of feature miRNAs using the validation dataset. (A) Two-way hierarchical clustering of samples in the validation dataset based on the expression levels of the 7 feature miRNAs yielded 2 clusters, which are labeled as cluster 1 and cluster 2. Early- and advanced-stage samples are indicated in red and black in the horizontal bar, respectively. (B) Kaplan–Meier survival curves for clusters 1 (green curve) and 2 (red curve) in (A). (C) ROC curves of the validation dataset generated using the SVM classifier. The AUC was calculated to be 0.897. (D) Kaplan–Meier survival curves of early-stage-like (red curve) and advanced-stage-like (black curve) groups classified by the SVM classifier.

selected, including hsa-miR-144, hsa-miR-147b, hsa-miR-218-2, hsa-miR-425, hsa-miR-451, hsa-miR-483, and hsa-miR-486. Both hierarchical clustering and SVM prediction results demonstrated the high performance of these feature miRNAs in the prediction of cervical cancer progression. The high performance of this subset of feature miRNAs was validated using a validation dataset. Similar to the findings for the training dataset, both hierarchical clustering and SVM prediction could classify the samples in the validation dataset with high accuracy. Additionally, Kaplan–Meier survival analysis showed that the SVM-

predicted early-stage-like group showed significant better prognosis than the advanced-stage-like group for both the training and the validation dataset. We also examined the independence of SVM prediction as a prognostic factor. Both univariate and multivariate Cox regression analyses confirmed the independence of SVM-predicted group in outcome prediction for the training and validation datasets. The prognostic power of SVM prediction was additionally evaluated using stratified analysis. Elder patients (> 45 years of age) in both datasets could be classified into early-stage-like and advanced-stage-like groups with

Figure 4: Stratified Kaplan–Meier survival curve analysis. Samples were stratified by age (A and B), pathologic N (C and D), new tumor (E and F), and radiation therapy (G and H) for both the training and the validation dataset. Survival curves for early-stage-like and advanced-stage-like groups in each stratum are shown in green and red, respectively.

significant differences in prognosis. Similar results were acquired for patients with no new tumor and patients that had received no radiation therapy in both datasets. Taken together, the results demonstrated that SVM prediction was a prognostic factor independent of other clinical factors, including age, new tumor, and radiation therapy.

MiRNAs function by regulating the expression of target mRNAs [8, 9]. To interpret the functions of the feature miRNAs in cervical cancer, potential target mRNAs were identified and subjected to GO and KEGG functional enrichment analyses. Cancer-related

GO biological processes and KEGG pathways were overrepresented among these mRNAs, which supported the importance of the feature miRNAs in cervical cancer. Two KEGG terms, Hedgehog signaling pathway and Wnt signaling pathway, deserve special attention. Both pathways are essential for cancer cell proliferation, migration, and invasion in various cancers, and thus, represent promising targets for cancer therapy [17, 18].

The high predictive performance of the feature miRNAs may be explained by their close relation to cervical cancer and other cancers. hsa-miR-218

Figure 5: Functional enrichment analysis of mRNAs targeted by the optimal subset of feature miRNAs. (A) Regulatory network consisting of feature miRNAs and their targeting mRNAs. Up- and downregulated feature miRNAs in advanced-stage samples are shown as orange triangles and diamonds, respectively. Up- and downregulated mRNAs are shown as purple triangles and green diamonds, respectively. Potential regulatory relationships between feature miRNAs and targeting mRNAs are indicated as gray lines. (B and C) GO (B) and KEGG (C) term enrichment analysis of mRNAs in (A). Gene numbers (vertical axis) are plotted against GO and KEGG terms (horizontal axis). $-\log(p\text{-values})$ are indicated as red dots and lines.

Table 1: Progression related feature miRNAs of cervical cancer

miRNA	Wilcox test			<i>t</i> test		
	logFC ^a	FDR ^b	Wilcox <i>p</i>	logFC	FDR	<i>t</i> test <i>p</i>
hsa-miR-144	0.701563604	4.16E-06	1.78E-07	0.699709	4.46E-06	1.95E-07
hsa-miR-147b	-0.654720572	0.005331	0.000228	-0.65658	0.009557	0.000419
hsa-miR-218-2	-0.40945555	0.000328	1.40E-05	-0.41131	0.000538	2.36E-05
hsa-miR-425	0.268232408	6.62E-06	2.83E-07	0.266381	1.48E-05	6.49E-07
hsa-miR-451	0.590364706	6.95E-07	2.97E-08	0.588512	7.74E-07	3.39E-08
hsa-miR-483	0.740086852	0.001946	8.33E-05	0.738239	0.00312	0.000137
hsa-miR-486	0.592724644	3.35E-06	1.43E-07	0.590872	8.53E-06	3.74E-07

^a Fold change;

^b False discovery rate.

reportedly is tumor suppressor that is downregulated in cervical cancer [19] and functions through suppressing migration, invasion, and metastasis by targeting survivin [20]. Additionally, lower-than-normal expression of hsa-miR-218 is related to infection with high-risk, rather than low-risk, HPV [21]. Consistent with the tumor-suppressive role of hsa-miR-218, we found that hsa-miR-218-2 expression is reduced in advanced-stage patients. hsa-miR-486, hsa-miR-425, and hsa-miR-144 are also reportedly engaged in cervical cancer [22–24]. A variant of hsa-miR-486, hsa-miR-486-3p, acts as a tumor suppressor through inhibiting cell growth and metastasis by targeting ECM1 [22]. hsa-miR-425-5p, a variant of hsa-miR-425, is significantly upregulated in cervical cancer and may serve as a prognostic indicator [23]. hsa-miR-144 is significantly downregulated in metastatic cervical cancer [24], though its role remains elusive. The remaining 3 feature miRNAs, while not directly associated with cervical cancer, have been associated with other cancer types. Dysregulation of hsa-miR-483 has been found in various cancers, including adrenocortical cancer [25], prostate cancer [26], and lung adenocarcinoma [27]. hsa-miR-451 has been reported to be a tumor suppressor in different cancers [28, 29] and a protective effect of hsa-miR-147b has been found in ovarian cancer [30]. Considering the involvement of these feature miRNAs in diverse cancer types, they may play important roles in the development and progression of cervical cancer and may serve as potential biomarkers of this disease. However, further experimental and functional studies are needed to reveal the specific roles of these feature miRNAs in cervical cancer.

In summary, we identified 7 feature miRNAs of cervical cancer by a comprehensive miRNA expression analysis. The feature miRNAs were significantly associated with the progression of cervical cancer and were used for the construction of a cervical cancer-specific SVM classifier. The SVM is a promising predictor of progression and outcome. Meanwhile, the

feature miRNAs may be novel therapeutic targets in future clinical practice.

MATERIALS AND METHODS

Data source

mRNA and miRNA expression profiles (Illumina HiSeq 2000 RNA sequencing data) of cervical cancer were downloaded from TCGA (<https://gdc-portal.nci.nih.gov/>) database. Samples for which both mRNA and miRNA data as well as information on pathologic stage and survival were available, were selected for our study. In total, data of 285 samples were collected, which were randomly divided into training (143 samples) and validation (142 samples) datasets. Clinical characteristics of these samples are summarized in Table 4.

Screening of significantly differentially expressed miRNAs

Samples in the training dataset were divided into early (I and II) and advanced (III and IV) pathologic stage groups. miRNA expression profiles were compared between the two groups using the *t* test (<http://127.0.0.1:26738/library/stats/html/t.test.html>) and Wilcoxon rank-sum test (<http://127.0.0.1:26738/library/stats/html/wilcox.test.html>) under R3.1.0. FDR (false discovery rate) < 0.05 and |logFC (fold change)| > 0.263 were set as thresholds for the selection of significantly differentially expressed miRNAs for both methods. Significantly differentially expressed miRNAs derived from both the *t* test and the Wilcoxon rank-sum test were used for further analysis.

Two-way hierarchical clustering analysis based on centered Pearson correlation [31] was performed using the pheatmap package (<https://cran.r-project.org/package=pheatmap>) in R. Correlation between clusters and pathologic stages were analyzed by the Chi-square

Table 2: Cox regression analysis of the correlations between clinical variables and prognosis for both the training and the validation dataset

Variables	Univariate analysis			Multivariate analysis		
	HR ^a	95% CI ^b	P value	HR	95% CI	P value
Training dataset (N = 143)						
SVM predicted group (Early/Advanced stage)	2.601	1.308~5.172	0.00472	2.07	0.341~2.587	0.0272
Age (≤ 45/> 45)	1.239	0.633~2.426	0.531	-	-	-
Pathologic M (M0/M1)	-	-	-	-	-	-
Pathologic N (N0/N1)	4.861	1.583~14.93	0.00227	3.037	0.902~10.22	0.0729
Pathologic T (T1/T2/T3/T4)	2.076	1.27~3.395	0.00234	2.916	1.209~7.032	0.0172
Smoking (Yes/No)	1.461	0.726~2.944	0.286	-	-	-
New tumor (Yes/No)	6.036	3.083~11.82	2.60E-09	6.101	1.964~14.24	0.00053
Radiation therapy (Yes/No)	0.9697	0.4204~2.237	0.943	-	-	-
Targeted molecular therapy (Yes/No)	0.891	0.4168~1.904	0.765	-	-	-
Validation dataset (N = 142)						
SVM predicted group (Early/Advanced stage)	2.473	1.265~4.836	0.00625	1.233	0.422~3.601	0.00702
Age (≤ 45/> 45)	1.248	0.628~2.48	0.526	-	-	-
Pathologic M (M0/M1)	-	-	-	-	-	-
Pathologic N (N0/N1)	1.81	0.732~4.474	0.193	-	-	-
Pathologic T (T1/T2/T3/T4)	1.71	1.174~2.491	0.00384	1.739	1.119~2.701	0.01388
Smoking (Yes/No)	1.39	0.683~2.827	0.362	-	-	-
New tumor (Yes/No)	3.94	1.908~8.136	6.32E-05	5.229	2.117~12.919	0.000337
Radiation therapy (Yes/No)	1.155	0.517~2.578	0.725	-	-	-
Targeted molecular therapy (Yes/No)	1.074	0.472~2.446	0.865	-	-	-

^aHazard ratio;

^bConfidence interval.

test using the `chisq.test` function (<http://127.0.0.1:21869/library/stats/html/chisq.test.html>) in R. Kaplan–Meier survival analysis was conducted for different clusters, using the survival package (version 2.40-1, <https://cran.r-project.org/package=survival>) under R.

Selection and validation of an optimal subset of feature miRNAs

Feature miRNAs were selected from differentially expressed miRNAs using the bootstrap algorithm [32]

Table 3: Stratified prognosis analysis of patients in both the training and the validation dataset

Variables	Training dataset			Variables	Validation dataset		
	HR ^a	95% CI ^b	P value		HR	95% CI	P value
Age				Age			
≤ 45 (N = 73)	0.846	0.187~3.825	0.828	≤ 45 (N = 64)	1.647	0.504~5.388	0.404
>45 (N = 70)	4.379	1.751~10.95	0.000623	> 45 (N = 78)	2.865	1.229~6.684	0.0109
Pathologic N				Pathologic N			
N0 (N = 62)	1.34	0.813~2.506	0.566	N0 (N = 67)	8.396	1.983~35.55	0.00386
N1 (N = 24)	1.321	0.571~2.285	0.004678	N1 (N = 28)	0.565	0.0677~4.715	0.593
Pathologic T				Pathologic T			
T1+T2 (N = 97)	3.153	0.976~10.18	0.04289	T1+T2 (N = 105)	2.135	0.673~6.769	0.187
T3+T4 (N = 12)	1.062	0.178~6.326	0.948	T3+T4 (N = 16)	1.359	0.320~5.77	0.676
Smoking				Smoking			
Yes (N = 56)	1.697	0.635~4.537	0.286	Yes (N = 56)	1.873	0.719~4.871	0.191
No (N = 69)	6.558	2.119~20.3	0.000186	No (N = 71)	2.138	0.741~6.174	0.1503
New tumor				New tumor			
Yes (N = 24)	4.868	1.759~13.47	0.000852	Yes (N = 19)	0.3594	0.0961~1.345	0.1134
No (N = 101)	2.769	1.012~7.576	0.0389	No (N = 102)	5.213	2.006~13.55	0.000159
Radiation therapy				Radiation therapy			
Yes (N = 88)	2.407	1.104~5.246	0.0227	Yes (N = 83)	2.12	0.948~4.74	0.0613
No (N = 30)	1.843	1.173~6.056	0.0113	No (N = 33)	1.884	1.422~4.372	0.005335
Targeted molecular therapy				Targeted molecular therapy			
Yes (N = 69)	2.405	0.995~5.812	0.0512	Yes (N = 59)	1.643	1.389~2.292	0.006816
No (N = 33)	3.257	0.622~7.05	0.139	No (N = 30)	2.33	0.829~6.541	0.0988

^a Hazard ratio;^b Confidence interval.

of the random forest package (<https://cran.r-project.org/package=randomForest>) [33] in R. The optimal subset of feature miRNAs was the one yielding the minimum out-of-bag (OOB) error.

Based on the expression values of the optimal subset of miRNAs, two-way hierarchical clustering analysis was conducted for both the training and the validation dataset, followed by Kaplan–Meier survival analysis of different clusters.

A cervical cancer-specific support vector machine (SVM) classifier was constructed based on the expression values of the optimal subset of miRNAs, using the SVM function (core function: Sigmoid kernel, cross: 10-fold cross validation) in the e1701 package in R [34]. The SVM classifier was used to predict the pathologic stages of samples. Based on the predictions, samples in both the training and the validation dataset were classified into either an early-stage-like or an advanced-stage-like group. The prognosis of different groups was analyzed using Kaplan–Meier survival curve analysis.

Independence analysis of SVM-predicted group as a prognostic factor

Clinical information on age, pathologic M, pathologic N, pathologic T, smoking, new tumor, radiation therapy, and targeted molecular therapy was extracted from both the training and the validation dataset. Correlations between clinical variables and prognoses were analyzed by univariate and multivariate Cox regressions, using the survival package under R3.1.0. Clinical variables with $p < 0.05$ were considered to be significant and independent prognostic factors.

Samples in both the training and the validation dataset were further stratified according to different clinical variables. The relation between SVM-predicted group and prognosis of cervical cancer was analyzed using univariate Cox regression and Kaplan–Meier survival curve analysis for each stratum, with $p < 0.05$ considered significant.

Table 4: Summary of clinical characteristics of the training, validation and entire dataset

Clinical characteristics	Training dataset (N = 143)	Validation dataset (N = 142)	Entire dataset (N = 285)
Age (years, mean $\pm$ sd ^a)	46.65 $\pm$ 13.56	48.92 $\pm$ 13.53	47.78 $\pm$ 13.57
Pathologic_M (M0/M1/-)	55/4/84	50/7/85	105/11/169
Pathologic_N (N0/N1/-)	62/24/57	67/28/47	129/52/104
Pathologic_T (T1/T2/T3/T4/-)	69/28/9/3/34	65/40/9/7/21	134/68/18/10/55
Pathologic_stage (I/II/III/IV)	83/7/43/10	72/18/40/12	155/25/83/22
Smoking (Yes/No/-)	69/56/18	56/71/15	125/127/33
New tumor (Yes/No/-)	24/101/18	19/102/21	43/203/39
Radiation therapy (Yes/No/-)	88/30/25	83/33/26	171/63/51
Targeted molecular therapy (Yes/No/-)	69/33/41	59/30/53	128/63/94
Dead (Death/Alive)	35/108	35/107	70/215
Overall survival months (mean $\pm$ sd)	32.07 $\pm$ 39.29	32.25 $\pm$ 38.57	32.16 $\pm$ 38.86

Functional analysis of the optimal subset of feature miRNAs

mRNAs significantly differentially expressed between early and advanced stages were screened as described above for miRNAs. mRNAs targeted by the optimal subset of feature miRNAs were predicted using miRanda (<http://www.microrna.org/microrna/home.do>) [15], which were further intersected with the significantly differentially expressed mRNAs. Thus, the selected mRNAs were used for the construction of a regulatory network of the optimal subset of feature miRNAs. mRNAs in the regulatory network were functionally annotated using DAVID (<https://david.ncifcrf.gov/>) [16] and significantly enriched Gene Ontology (GO) biological processes, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were retrieved, with $p < 0.05$ set as a threshold for significantly enriched terms.

Author contributions

Conception and design of the research: Shanshan Qin and Yingchun Gao; Acquisition of data: Yijun Yang and Juanpeng Yu; Analysis and interpretation of data: Fei Meng and Ting Zhang; Drafting the manuscript: Shanshan Qin; Revision of manuscript for important intellectual content: Can Shi.

ACKNOWLEDGMENTS

None.

CONFLICTS OF INTEREST

The authors have declared that no competing interests exist.

FUNDING

None.

REFERENCES

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer*. 2015; 136:E359–86.
2. Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, Snijders PJ, Peto J, Meijer C, Munoz N. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *The Journal of pathology*. 1999; 189:12–19.
3. Doorbar J. Molecular biology of human papillomavirus infection and cervical cancer. *Clinical science*. 2006; 110:525–541.
4. Jimenez-Wences H, Peralta-Zaragoza O, Fernandez-Tilapa G. Human papilloma virus, DNA methylation and microRNA expression in cervical cancer (Review). *Oncology reports*. 2014; 31:2467–2476.
5. Akagi K, Li J, Broutian TR, Padilla-Nash H, Xiao W, Jiang B, Rocco JW, Teknos TN, Kumar B, Wangsa D. Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome research*. 2014; 24:185–199.
6. Schiffman M, Wentzensen N, Wacholder S, Kinney W, Gage JC, Castle PE. Human papillomavirus testing in the prevention of cervical cancer. *Journal of the National Cancer Institute*. 2011; 103:368–383.
7. Liz J, Esteller M. lncRNAs and microRNAs with a role in cancer development. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*. 2016; 1859:169–176.

8. Morris KV, Mattick JS. The rise of regulatory RNA. *Nature reviews Genetics*. 2014; 15:423.
9. Garzon R, Calin GA, Croce CM. MicroRNAs in Cancer. *Annual review of medicine*. 2009; 60:167–179.
10. Reshmi G, Pillai MR. Beyond HPV: Oncomirs as new players in cervical cancer. *FEBS Letters*. 2008; 582:4113–4116.
11. Yao Q, Xu H, Zhang QQ, Zhou H, Qu LH. MicroRNA-21 promotes cell proliferation and down-regulates the expression of programmed cell death 4 (PDCD4) in HeLa cervical carcinoma cells. *Biochemical and biophysical research communications*. 2009; 388:539–542.
12. Hou T, Ou J, Zhao X, Huang X, Huang Y, Zhang Y. MicroRNA-196a promotes cervical cancer proliferation through the regulation of FOXO1 and p27Kip1. *British journal of cancer*. 2014; 110:1260–1268.
13. Luo M, Shen D, Zhou X, Chen X, Wang W. MicroRNA-497 is a potential prognostic marker in human cervical cancer and functions as a tumor suppressor by targeting the insulin-like growth factor 1 receptor. *Surgery*. 2013; 153:836–847.
14. Hu X, Schwarz JK, Lewis JS Jr, Huettner PC, Rader JS, Deasy JO, Grigsby PW, Wang X. A microRNA expression signature for cervical cancer prognosis. *Cancer research*. 2010; 70:1441–1448.
15. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. Human microRNA targets. *PLoS biology*. 2004; 2:e363.
16. Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*. 2009; 4:44–57.
17. Takebe N, Miele L, Harris PJ, Jeong W, Bando H, Kahn M, Yang SX, Ivy SP. Targeting Notch, Hedgehog, and Wnt pathways in cancer stem cells: clinical update. *Nature reviews Clinical oncology*. 2015; 12:445–464.
18. Nwabo Kamdje AH, Takam Kamga P, Tagne Simo R, Vecchio L, Seke Etet PF, Muller JM, Bassi G, Lukong E, Kumar Goel R, Mbo Amvene J, Krampera M. Developmental pathways associated with cancer metastasis: Notch, Wnt, and Hedgehog. *Cancer biology & medicine*. 2017; 14:109–120.
19. Yu J, Wang Y, Dong R, Huang X, Ding S, Qiu H. Circulating microRNA-218 was reduced in cervical cancer and correlated with tumor invasion. *Journal of cancer research and clinical oncology*. 2012; 138:671–674.
20. Kogo R, How C, Chaudary N, Bruce J, Shi W, Hill RP, Zahedi P, Yip KW, Liu FF. The microRNA-218–Survivin axis regulates migration, invasion, and lymph node metastasis in cervical cancer. *Oncotarget*. 2015; 6:1090–1100. <http://doi.org/10.18632/oncotarget.2836>.
21. Li Y, Liu J, Yuan C, Cui B, Zou X, Qiao Y. High-risk human papillomavirus reduces the expression of microRNA-218 in women with cervical intraepithelial neoplasia. *The Journal of international medical research*. 2010; 38:1730–1736.
22. Ye H, Yu X, Xia J, Tang X, Tang L, Chen F. MiR-486-3p targeting ECM1 represses cell proliferation and metastasis in cervical cancer. *Biomedicine & pharmacotherapy*. 2016; 80:109–114.
23. Sun L, Jiang R, Li J, Wang B, Ma C, Lv Y, Mu N. MicoRNA-425-5p is a potential prognostic biomarker for cervical cancer. *Annals of clinical biochemistry*. 2017; 54:127–133.
24. Ding H, Wu YL, Wang YX, Zhu FF. Characterization of the microRNA expression profile of cervical squamous cell carcinoma metastases. *Asian Pac J Cancer Prev*. 2014; 15:1675–1679.
25. Patterson EE, Holloway AK, Weng J, Fojo T, Kebebew E. MicroRNA profiling of adrenocortical tumors reveals miR-483 as a marker of malignancy. *Cancer*. 2011; 117:1630–1639.
26. Yang ZG, Ma XD, He ZH, Guo YX. miR-483-5p promotes prostate cancer cell proliferation and invasion by targeting RBM5. *Int Braz J Urol*. 2017; 43:1060–1067.
27. Song Q, Xu Y, Yang C, Chen Z, Jia C, Chen J, Zhang Y, Lai P, Fan X, Zhou X, Lin J, Li M, Ma W, et al. miR-483-5p promotes invasion and metastasis of lung adenocarcinoma by targeting RhoGDI1 and ALCAM. *Cancer research*. 2014; 74:3031–3042.
28. Pan X, Wang R, Wang ZX. The potential role of miR-451 in cancer diagnosis, prognosis, and therapy. *Molecular cancer therapeutics*. 2013; 12:1153–1162.
29. Liu D, Liu C, Wang X, Ingvarsson S, Chen H. MicroRNA-451 suppresses tumor cell growth by down-regulating IL6R gene expression. *Cancer epidemiology*. 2014; 38:85–92.
30. Kleemann M, Bereuther J, Fischer S, Marquart K, Hanle S, Unger K, Jendrosseck V, Riedel CU, Handrick R, Otte K. Investigation on tissue specific effects of pro-apoptotic micro RNAs revealed miR-147b as a potential biomarker in ovarian cancer prognosis. *Oncotarget*. 2017; 8:18773–18791. <http://doi.org/10.18632/oncotarget.13095>.
31. Li J, Xue W, Lv J, Han P, Liu Y, Cui B. Identification of potential long non-coding RNA biomarkers associated with the progression of colon cancer. *Oncotarget*. 2017; 8:75834–75843. <http://doi.org/10.18632/oncotarget.17924>.
32. Zapf A, Brunner E, Konietzschke F. A wild bootstrap approach for the selection of biomarkers in early diagnostic trials. *BMC medical research methodology*. 2015; 15:43.
33. Tološi L, Lengauer T. Classification with correlated features: unreliability of feature ranking and solutions. *Bioinformatics*. 2011; 27:1986–1994.
34. Wang Q, Liu X. Screening of feature genes in distinguishing different types of breast cancer using support vector machine. *OncoTargets and therapy*. 2015; 8:2311.