An 8 miRNA-Based Risk Score System for Predicting the Prognosis of Patients With Papillary Thyroid Cancer

Wanwan Yi, MM[1#], Jin Liu, BM[1#], Shuping Qu, MM[2#], Hengwei Fan, MM[2*] ⓘ, and Zhongwei Lv, MD[1*] ⓘ

Abstract

Background: Dysregulation of microRNAs (miRNAs) in papillary thyroid cancer (PTC) might influence prognosis of PTC. This study is aimed to develop a risk score system for predicting prognosis of PTC. **Methods:** The miRNA and gene expression profiles of PTC were obtained from The Cancer Genome Atlas database. PTC samples were randomly separated into training set (n = 248) and validation set (n = 248). The differentially expressed miRNAs (DE-miRNAs) in the training set were screened using limma package. The independent prognosis-associated DE-miRNAs were identified for building a risk score system. Risk score of PTC samples in the training set was calculated and samples were divided into high risk group and low risk group. Kaplan-Meier curves and receiver operating characteristic (ROC) curve were used to assess the accuracy of the risk score system in the training set, validation set and entire set. Finally, a miRNA-gene regulatory network was visualized by Cytoscape software, followed by enrichment analysis. **Results:** Totally, 162 DE-miRNAs between tumor and control groups in the training set were identified. An 8 independent prognosis-associated DE-miRNAs, (including miR-1179, miR-133b, miR-3194, miR-3912, miR-548j, miR-6720, miR-6734, and miR-6843) based risk score system was developed. The area under ROC curve in the training set, validation set and entire set was all above 0.93. A miRNA-gene regulatory network involving the 8 DE-miRNAs were built and functional enrichment analysis suggested the genes in the network were significantly enriched into 13 pathways, including calcium signaling pathway and hedgehog signaling pathway. **Conclusion:** The risk score system developed this study might be used for predicting the prognosis of PTC. Besides, the 8 miRNAs might affect the prognosis of PTC via hedgehog signaling pathway and calcium signaling pathway.

Keywords

papillary thyroid cancer, differential expression analysis, risk score system, stratification analysis, enrichment analysis

Received: July 23, 2019; Revised: September 2, 2020; Accepted: September 22, 2020.

Introduction

Thyroid cancer (TC) is a popular endocrine tumor, which is mainly divided into papillary thyroid cancer (PTC), medullary thyroid cancer (MTC), follicular thyroid cancer (FTC), and anaplastic thyroid cancer (ATC).[1] TC is mainly caused by enlarged thyroid, family history, and radiation exposure.[2] It usually has the symptoms of neck swelling and neck lump,[3] and is more common in young females.[4] Globally, TC affects the life of 3.2 million people and is responsible for 31,900 deaths in 2015.[5] Especially, PTC is the most predominant form, accounting for 75% to 85% of TCs.[6] Therefore, the mechanisms of PTC should be further researched.

MicroRNAs (miRNAs) are a kind of short non-coding RNAs, which can regulate the expression levels of genes

[1] Department of Nuclear Medicine, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai, China
[2] Department of Hepatic Surgery, The Eastern Hepatobiliary Surgery Hospital, Navy Medical University (Second Military Medical University), Shanghai, China
[#] Wanwan Yi, Jin Liu and Shuping Qu are co-first authors.
[*] Hengwei Fan and Zhongwei Lv are co-corresponding authors.

Corresponding Authors:
Hengwei Fan, Department of Hepatic Surgery, The Eastern Hepatobiliary Surgery Hospital, Navy Medical University (Second Military Medical University), Shanghai 200438, China.
Email: fanhengwei2006@163.com

Zhongwei Lv, Department of Nuclear Medicine, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Yanchang Middle Road 301, Shanghai 20072, China.
Email: Lvzwjs2020@163.com

Table 1. The Clinical Information of the Training Set, the Validation Set and the Entire Set.

Clinical characteristics	Training set (N = 248)	Validation set (N = 248)	Entire set (N = 496)
Age (years, mean $\pm$ SD)	47.19 $\pm$ 15.87	47.58 $\pm$ 15.80	47.20 $\pm$ 15.99
Gender (Male/Female)	74/174	57/191	131/365
Pathologic_M (M0/M1/-)	140/6/102	137/3/108	277/9/210
Pathologic_N (N0/N1/-)	101/124/23	127/94/27	228/218/50
Pathologic_T (T1/T2/T3/T4/-)	71/76/89/10/2	71/90/74/13	142/166/163/23/2
Pathologic_stage (I/II/III/IV/-)	138/18/63/28/1	140/33/48/26/1	278/51/111/54/2
Histological_type (Classical/Follicular/Tall Cell)	189/45/14	168/57/23	357/102/37
Radiation therapy (Yes/No/-)	156/81/11	143/97/8	299/178/19
Tumor recurrence (Yes/No/-)	17/223/8	17/226/5	34/449/13
Death (Dead/Alive)	7/241	9/239	16/480
Overall survival time (months, mean $\pm$ SD)	40.57 $\pm$ 33.55	40.69 $\pm$ 32.98	42.20 $\pm$ 34.58

Note: SD, standard deviation.

involved in multiple biological processes.[7,8] Several miRNAs have been reported to play important role in PTC and have potential to be biomarkers or therapy targets for PTC. For example, miR-335 expression is reduced in PTC samples and could repress cell growth and metastasis in PTC by mediating zinc finger E-box binding homeobox 2 (*ZEB2*), therefore, miR-335 can serve as a tumor suppressor in PTC.[9] Through regulating fibronectin 1 (*FN1*), miR-139 acts as a tumor suppressor and represses the tumor formation in PTC.[10] MiR-451 is a valuable tissue marker for PTC, which may also be a noninvasive blood marker for diagnosing and evaluating metastasis of PTC.[11] MiR-144-3p facilitates the cell growth and metastasis in PTC through mediating paired box gene 8 (*PAX8*), which may be useful prognostic marker and promising therapeutic target for PTC.[12,13] However, the accuracy of single miRNA for predicting prognosis of PTC is usually not ideal.

In 2019, Pak et al built a gene expression-based risk score system for predicting the event-free survival and the presence of metastasis in PTC patients.[14] Nevertheless, the miRNA expression-based risk score system for predicting the overall survival of PTC patients has not been established. Combined with the miRNA expression data of PTC samples in The Cancer Genome Atlas (TCGA) database, the current study screened the independent prognosis-associated miRNAs in PTC for constructing the risk score system. Combined with the mRNA expression profile of PTC samples, the functions of the independent prognosis-associated miRNAs were analyzed. Our findings might help to reveal the action mechanisms of key miRNAs in PTC and predict the prognosis of patients with PTC.

Materials and Methods

Data Source

The miRNA expression data (including 568 samples) and mRNA expression data (including 573 samples) of PTC (platform: Illumina HiSeq 2000 RNA Sequencing; downloaded in June 15, 2019) were obtained from TCGA database (https://cancergenome.nih.gov/). After matching with the clinical information of the samples, a total of 553 samples (including 57 normal controls and 496 PTC samples) with both miRNA

data, mRNA expression data and survival prognosis information were selected. The age of the PTC patients were 47.20 $\pm$ 15.99 years old. A total of 299 patients received radiation therapy, 178 patients didn't receive radiation therapy and the radiation therapy of 19 patients was not available. The overall survival time was 42.20 $\pm$ 34.58 years old. The 496 PTC samples (the entire set) were randomly and equally divided into 2 groups (training set: 248 samples; and validation set: 248 samples). The clinical information of the entire set, the training set, and the validation set were arranged and listed in Table 1.

Differential Expression Analysis

The samples were classified into tumor and control groups according to their source information. Using the limma package (version 3.34.7, https://bioconductor.org/packages/release/bioc/html/limma.html)[15] in R, the differentially expressed miRNAs (DE-miRNAs) between tumor and control groups were screened. The resulting P value was adjusted by Benjamini & Hochberg method to reduce false discovery rate (FDR). DE-miRNAs were selected based on FDR and $|\log_2$ fold change (FC)|. FDR < 0.05 was set and $|\log_2$ (FC)| was adjusted to obtain enough number of DE-miRNAs for further analysis. At last, FDR < 0.05 and $|\log_2$ (FC)| > 0.5 were used as the thresholds for screening DE-miRNAs. Combined with the expression values of the DE-miRNAs in the training set, hierarchical clustering was conducted using the pheatmap package (version 1.0.8, https://cran.r-project.org/web/packages/pheatmap/index.html)[16] in R.

Establishment of Risk Score System

Combined with the prognosis information of the 248 PTC samples in the training set, the DE-miRNAs significantly related to overall survival were selected using the univariate Cox regression analysis in the survival package (version 2.41 -1, http://bioconductor.org/packages/survivalr/)[17] in R. The log-rank p-value < 0.05 was taken as the threshold.

Using the multivariate Cox regression analysis in the survival package,[17] the DE-miRNAs significantly correlated with

independent prognosis were further identified. Based on the independent prognosis-associated DE-miRNAs, the risk score system was established using following formula:

$$\text{MiR Score} = \sum \beta_{miRNA} \times Exp_{miRNA}$$

Where β_{miRNA} stands for the independent prognostic coefficient of independent prognosis-associated DE-miRNA, and Exp_{miRNA} represents the expression level of independent prognosis-associated DE-miRNA.

The risk score of each sample in the training set was calculated. Then, the samples in the training set were divided into high risk group and low risk group on the basis of the median value of the risk score. Kaplan-Meier (KM) curve in the survival package[17] was draw for samples in high risk group and low risk group. The significance of the overall survival difference between high risk group and low risk group was calculated with log-rank test.

Validation of risk score system in validation set and entire set. Based on the risk score system established in the training set, the PTC samples in the validation set and the entire set were divided into high risk group and low risk group. KM method and log-rank test were used to test the efficiency of the risk score system in the the validation set and the entire set.

Screening of Independent Prognostic Factors

Using the univariate Cox regression analysis in the survival package,[17] the clinical factors significantly correlated with overall survival were screened from the training set, the validation set and the entire set, respectively. To identify the independent clinical prognostic factors, the screened clinical factors were further subjected to multivariate Cox regression analysis.[18] The threshold was set as log-rank p-value < 0.05.

Subsequently, stratification analysis was conducted for the independent clinical prognostic factors to reveal their correlations with risk grouping. Risk stratification analysis of the training set, the validation set and the entire set was respectively conducted to study the prognosis conditions of the independent clinical prognostic factors in different risk groups.

Regulatory Network and Functional Enrichment Analysis

The differentially expressed genes (DEGs) between high and low risk groups of the entire set were analyzed using the limma package.[15] The $|\log_2 FC| > 0.263$ and FDR < 0.05 were set as the thresholds for screening DEGs. Moreover, hierarchical clustering for the DEGs was performed with the pheatmap package.[16]

The starBase database (version 2.0, http://starbase.sysu.edu.cn/)[19] included the target gene prediction information in RNA22, PITA, targetScan, picTar, and miRanda databases. Using starBase database, the targets of the independent prognosis-associated DE-miRNAs were searched. The miRNA-gene regulatory relationships predicted by 2 or more databases of RNA22, PITA, targetScan, picTar, and miRanda were selected. Then, the DEGs were mapped to the target genes of the independent prognosis-associated DE-miRNAs. The overlapped miRNA-DEG relationship were filtered to build miRNA-gene regulatory network using Cytoscape software (version 3.6.1, https://cytoscape.org/).[20] Further, Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analyses were carried out for the genes involved in the regulatory network. Significant pathways and GO terms were selected with p-value < 0.05 as the threshold.

Results

Differential Expression Analysis of Samples in Training Set

After tumor (248 samples in the training set) and control (57 normal control samples) groups were divided, 162 DE-miRNAs between the 2 groups were screened, including 59 up-regulated miRNAs and 103 down-regulated miRNAs. The scatter diagram and clustering heatmap based on the expression levels of the DE-miRNAs are presented in Figure 1A and B, respectively. From Figure 1B, the expression patterns of DE-miRNAs were obvious different between PTC samples and control samples.

Construction of Risk Score System

From the 248 PTC samples in the training set, 20 DE-miRNAs significantly correlated with overall survival were identified. Based on multivariate Cox regression analysis, 8 independent prognosis-associated DE-miRNAs, including miR-1179, miR-133b, miR-3194, miR-3912, miR-548j, miR-6720, miR-6734, and miR-6843, were further screened (Table 2). Afterward, the risk score system based on the independent prognosis-associated DE-miRNAs was constructed. The calculation formula was:

$$\begin{aligned}
\text{MiR Score} = & (-0.326) \times Exp_{hsa-miR-1179} \\
& + (0.139) \times Exp_{hsa-miR-133b} + (1.050) \times \\
& Exp_{hsa-miR-3194} + (0.588) \times Exp_{hsa-miR-3912} \\
& + (1.589) \times Exp_{hsa-miR-548j} + (0.731) \times \\
& Exp_{hsa-miR-6720} + (-2.738) \times Exp_{hsa-miR-6734} \\
& + (-0.565) \times Exp_{hsa-miR-6843}
\end{aligned}$$

The risk score of each sample in the training set were calculated and the PTC samples were divided into high risk group and low risk group. The difference of overall survival in high risk group and low risk group was calculated. As shown in Figure 2A, the overall survival rate of samples in the low risk group was significant longer than that in the high risk group ($P < 0.05$). The area under the ROC curve (AUC) of 3-year prediction and 5-year prediction was both 0.998, indicating higher accuracy and credibility of the risk score model.
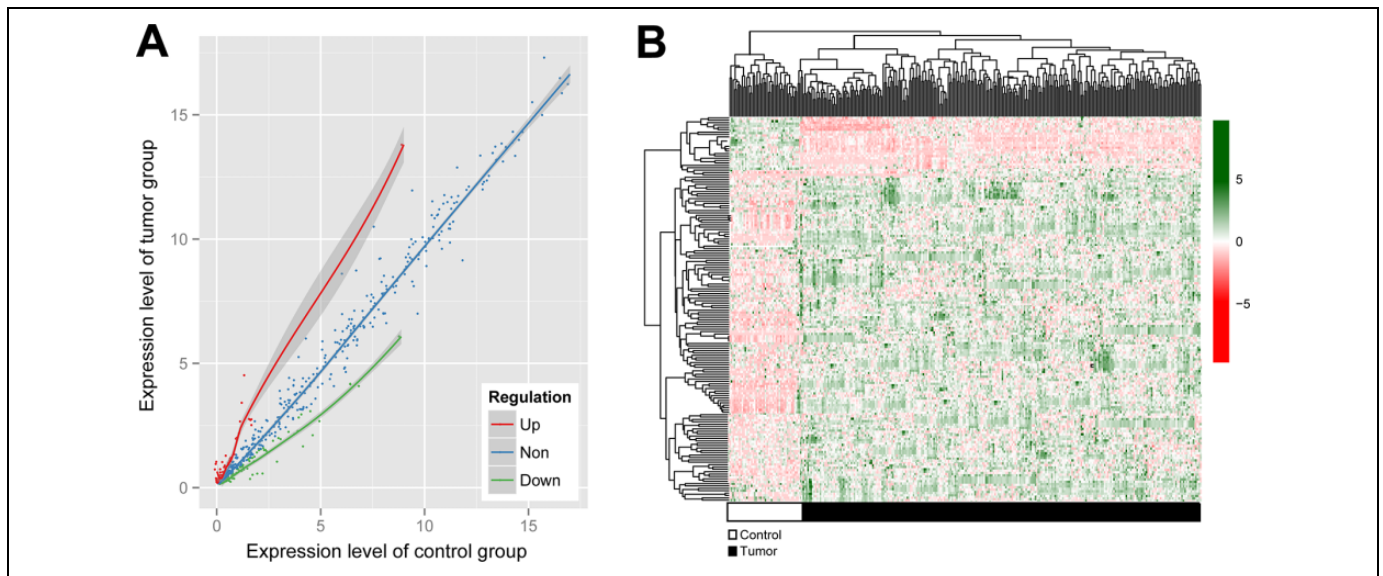
Figure 1. The scatter diagram and bidirectional hierarchical clustering heatmap based on the expression levels of the differentially expressed miRNAs (DE-miRNAs). (A) The scatter diagram (the horizontal and vertical axises represent the expression levels of control group and tumor group, respectively; red, green, and blue dots separately represent up-regulated miRNAs, down-regulated miRNAs and non-DE-miRNAs); (B) The bidirectional hierarchical clustering heatmap (in sample strip, black and white separately represent tumor samples and control samples).

Table 2. The Differentially Expressed miRNAs (DE-miRNAs) Significantly Correlated With Independent Prognosis.

ID	Coef	P-value	Hazard ratio	95% CI
hsa-miR-1179	-0.326	3.740E-02	0.722	0.504-0.933
hsa-miR-133b	0.139	4.980E-02	1.149	1.074-1.355
hsa-miR-3194	1.050	4.055E-02	2.859	1.878-4.307
hsa-miR-3912	0.588	3.550E-02	1.800	1.151-3.408
hsa-miR-548j	1.589	2.580E-02	2.898	1.989-4.267
hsa-miR-6720	0.731	4.295E-02	2.076	1.102-4.780
hsa-miR-6734	-2.738	2.880E-02	0.065	0.004-0.903
hsa-miR-6843	-0.565	2.935E-02	0.568	0.316-0.821

Note: CI, confidence interval.

Validation of the risk score system in the validation set and entire set. The credibility of the risk score system was further validated in the validation set and the entire set, respectively. The overall survival rate of samples in the low risk group was significantly longer than that in the high risk group in both validation set (Figure 2B) and entire set (Figure 2C) ($P < 0.05$). The 3-year prediction AUC and 5-year prediction AUC in both validation set and entire set were larger than 0.93.

Screening of independent prognostic factors. The association between clinical information and overall survival was analyzed using univariate Cox regression analysis and multivariate Cox regression analysis in the training set, validation set and entire set, respectively. Age and MiR Score status were identified as independent clinical prognostic factors in the training set, the validation set and the entire set, respectively (Table 3). In the entire set, the training set, and the validation set, the PTC patients with lower age had better survival prognosis (Figure 3). These results were consistent with the actual situation. Furthermore, stratification analysis was conducted for age to reveal its correlation with risk grouping in the training set, the validation set and the entire set. KM curves suggested that age was significantly related to overall survival in the high risk group (Figure 4).

Regulatory Network and Functional Enrichment Analysis

Total 469 DEGs, including 427 up-regulated genes and 42 down-regulated genes were identified between the high risk group and low risk group of the entire set (Figure 5A). The expression patterns of DEGs changed with MiR Score is presented in Figure 6B.

The target genes of the 8 independent prognosis-associated DE-miRNAs were predicted in starBase version 2.0. The obtained target genes were compared with the 469 DEGs, resulting in 176 overlapped DEGs. A miRNA-DEG regulatory network was constructed, involving 8 DE-miRNAs, 176 DEGs and their 515 interactions (Figure 6).

Functional enrichment analysis revealed that the genes in the regulatory network were significantly enriched into 17 significant GO_ Biology Process (BP) terms, such as "regulation of system process" (p-value = 1.36E-04), "regulation of cell proliferation" (p-value = 9.06E-04) and "transmission of nerve impulse" (p-value = 1.50E-03) and 13 KEGG pathways, such as "T cell receptor signaling pathway" (p-value = 1.15E-02), "hedgehog signaling pathway" (p-value = 1.26E-02) and "calcium signaling pathway" (p-value = 1.28E-02) (Table 4).
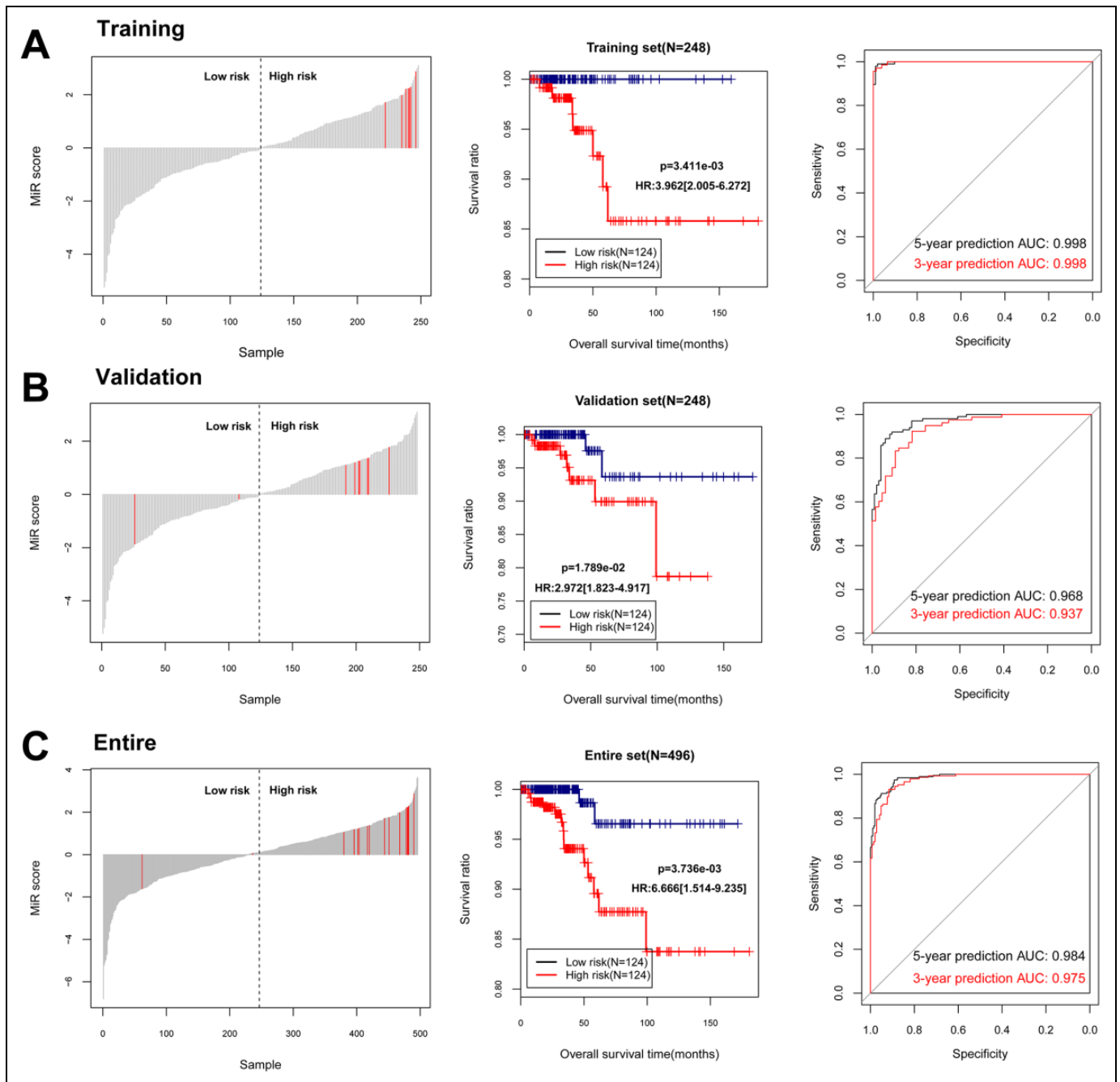
Figure 2. The distribution diagram of risk scores (RSs), the Kaplan-Meier (KM) curves and receiver operating characteristic (ROC) curves for the training set (A), validation set (B) and entire set (C). In RS distribution diagrams, the horizontal and vertical axises separately represent sample numbers and RSs. In KM curves, blue and red curves separately indicate low risk and high risk. In ROC curves, red and black curves separately represent 3-year prediction and 5-year prediction. HR: hazard ratio; AUC, area under the receiver operating characteristic curve.

Discussion

In the present study, 162 DE-miRNAs between tumor and control groups were screened. After 20 DE-miRNAs that significantly correlated with overall survival were identified, 8 independent prognosis-associated DE-miRNAs, including miR-1179, miR-133b, miR-3194, miR-3912, miR-548j, miR-

6720, miR-6734, and miR-6843 were further selected for constructing the risk score system. KM curves and ROC curves suggested that the risk score system could predict the prognosis of PTC with high accuracy and credibility. Based on Cox regression analysis, age and MiR Score status were screened as the independent clinical prognostic factors There were 469 DEGs (427 up-regulated and 42 down-regulated) between the

Table 3. The Independent Clinical Prognostic Factors Selected From the Training Set, the Validation Set and the Entire Set.

Variables	Univariate analysis			Multivariate analysis		
	HR	95% CI	P-value	HR	95% CI	P-value
Training set (N = 248)						
Age (years,mean ± SD)	1.187	1.087-1.297	1.24E-08	1.199	1.093-1.317	1.33E-04
Gender (Male/Female)	1.224	0.233-6.414	8.11E-01	-	-	-
Pathologic_M (M0/M1/-)	3.378	0.372-10.67	2.51E-01	-	-	-
Pathologic_N (N0/N1/-)	0.986	0.164-5.934	9.88E-01	-	-	-
Pathologic_T (T1/T2/T3/T4/-)	1.607	0.674-3.833	2.78E-01	-	-	-
Pathologic_stage (I/II/III/IV/-)	2.015	1.014-4.005	3.06E-02	1.066	0.712-1.401	2.80E-01
Histological_type (Classical/Follicular/Tall Cell)	0.511	0.0758-3.443	4.79E-01	-	-	-
Radiation therapy (Yes/No/-)	0.462	0.093-2.292	3.33E-01	-	-	-
Tumor recurrence (Yes/No/-)	2.123	0.255-17.68	4.76E-01	-	-	-
MiR Score status (High/Low)	3.962	2.005-6.272	3.41E-03	3.166	1.364-7.539	2.96E-02
Validation set (N = 248)						
Age (years,mean ± SD)	1.079	1.045-1.139	3.17E-06	1.070	1.012-1.132	1.76E-02
Gender (Male/Female)	2.943	0.788-3.110	9.24E-02	-	-	-
Pathologic_M (M0/M1/-)	7.434	0.811-8.170	7.60E-02	-	-	-
Pathologic_N (N0/N1/-)	1.966	0.468-4.249	3.47E-01	-	-	-
Pathologic_T (T1/T2/T3/T4/-)	3.916	1.609-9.531	1.20E-03	2.54	0.599-10.760	2.06E-01
Pathologic_stage (I/II/III/IV/-)	3.061	1.583-5.919	3.68E-05	1.251	0.330-4.735	7.42E-01
Histological_type (Classical/Follicular/Tall Cell)	0.688	0.115-1.981	1.23E-01	-	-	-
Radiation therapy (Yes/No/-)	0.401	0.193-1.269	1.27E-01	-	-	-
Tumor recurrence (Yes/No/-)	1.306	0.765-1.815	3.83E-01	-	-	-
MiR Score status (High/Low)	2.972	1.823-4.917	1.79E-02	2.581	1.446-4.921	2.90E-02
Entrie set (N = 496)						
Age (years,mean ± SD)	1.105	1.068-1.144	1.13E-11	1.099	1.054-1.147	9.75E-06
Gender (Male/Female)	1.848	0.680-5.021	2.22E-01	-	-	-
Pathologic_M (M0/M1/-)	4.176	0.892-9.561	6.96E-02	-	-	-
Pathologic_N (N0/N1/-)	1.208	0.418-3.492	7.27E-01	-	-	-
Pathologic_T (T1/T2/T3/T4/-)	2.424	1.345-4.369	2.30E-03	2.138	0.820-5.570	1.21E-01
Pathologic_stage (I/II/III/IV/-)	2.277	1.468-3.533	3.97E-05	1.709	0.726-3.824	4.76E-01
Histological_type (Classical/Follicular/Tall Cell)	0.433	0.116-1.164	1.92E-01	-	-	-
Radiation therapy (Yes/No/-)	0.411	0.135-3.058	9.38E-01	-	-	-
Tumor recurrence (Yes/No/-)	1.731	0.964-5.539	7.61E-01	-	-	-
MiR Score status (High/Low)	6.666	1.514-9.235	3.74E-03	2.073	1.540-7.956	2.88E-02

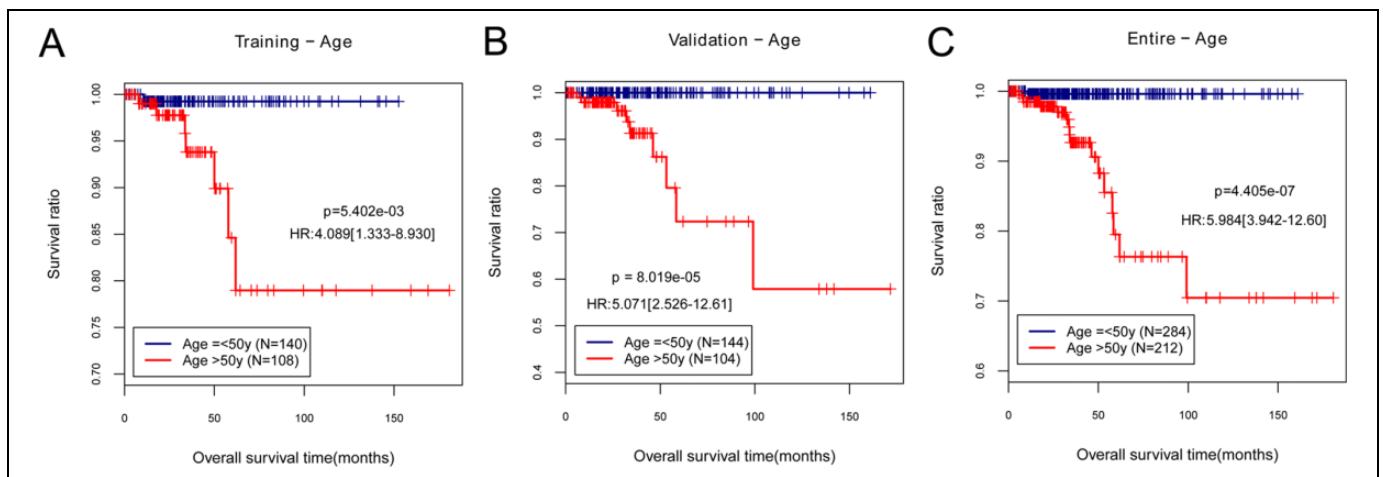Note: HR, Hazard Ratio; CI, confidence interval.

Figure 3. The Kaplan-Meier (KM) curves of different age groups in the training set (A), the validation set (B) and entire set (C). Blue and red curves indicate age ≤ 50 years and age ≥ 50 years, respectively.
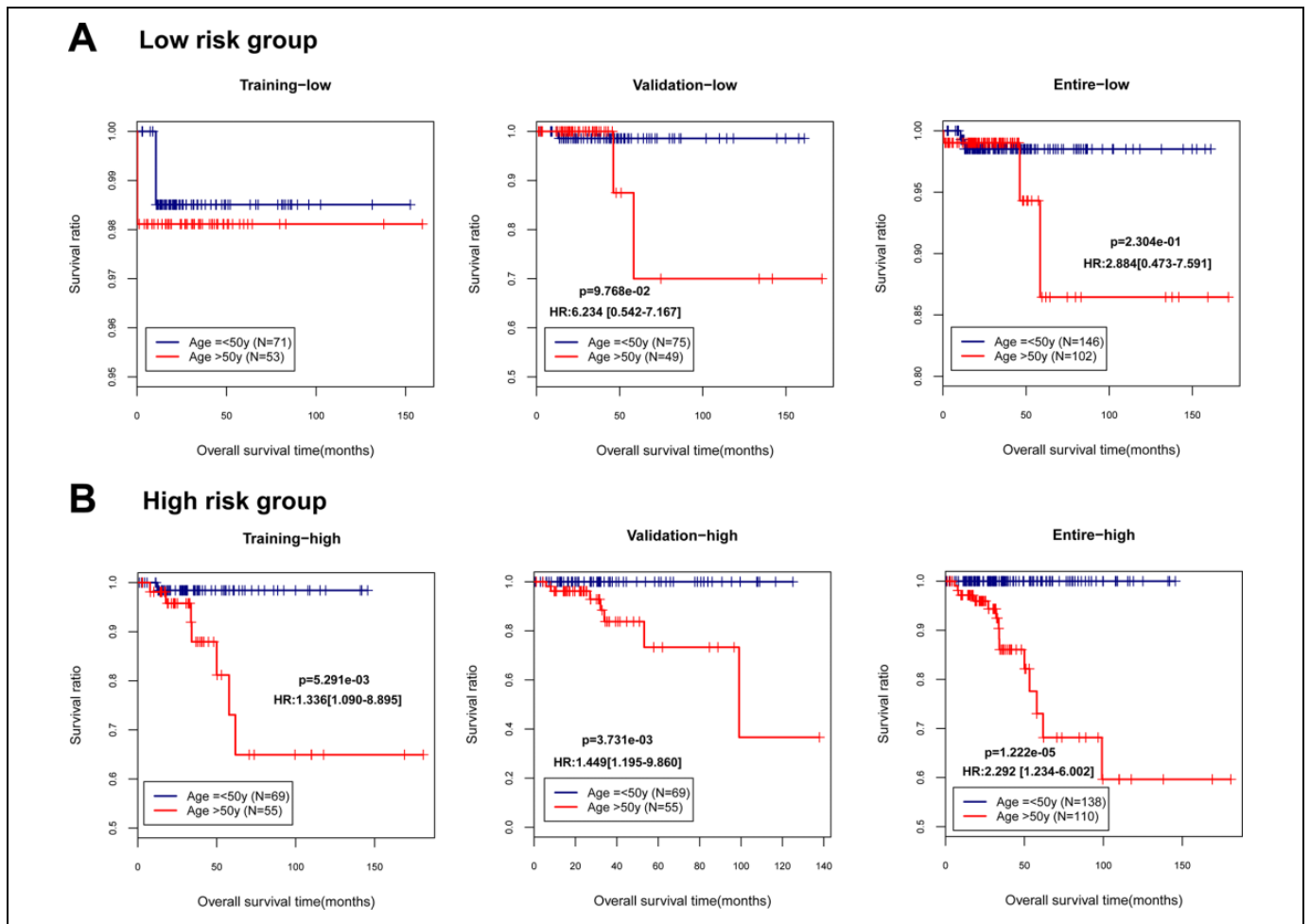
Figure 4. The Kaplan-Meier (KM) curves showing the correlations between age in different risk groups and overall survival. (A) The KM curves for the low risk group (left: the training set; middle: the validation set; and right: the entire set). (B) The KM curves for the high risk group (left: the training set; middle: the validation set; and right: the entire set). Blue and red curves represent the samples no more than 50 years and the samples older than 50 years, respectively. HR, hazard ratio.

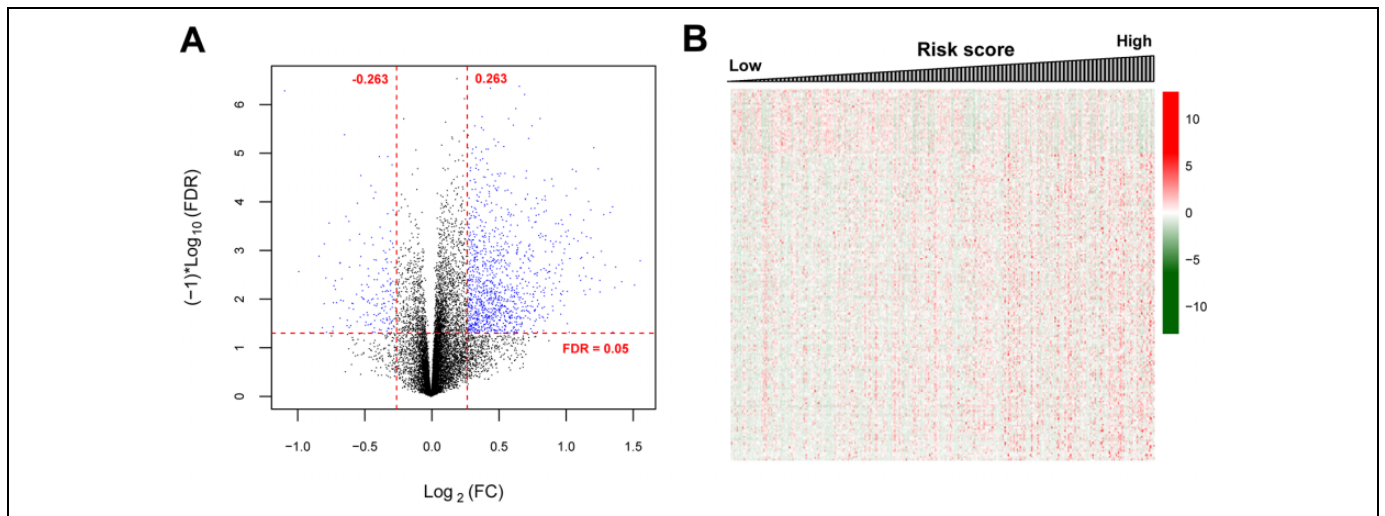**Figure 5.** The volcano plot and heatmap for the differentially expressed genes (DEGs). (A) The volcano plot (the horizontal dashed line indicates false discovery rate (FDR) < 0.05, and the vertical dashed lines represent $|\log_2$ fold change (FC)$| > 0.263$; the blue dots stand for the DEGs); (B) The heatmap based on the expression levels of the DEGs (left-to-right order indicates the changes of risk scores from low to high).

Figure 6. The miRNA-gene regulatory network. Circles and squares represent mRNAs and miRNAs, respectively. FC, fold change.

Table 4. The Enrichment Results for the Genes Involved in the Regulatory Network.

Category	Term	Count	P-value
Biology Process	GO:0044057~regulation of system process	12	1.36E-04
	GO:0042127~regulation of cell proliferation	18	9.06E-04
	GO:0019226~transmission of nerve impulse	11	1.50E-03
	GO:0008284~positive regulation of cell proliferation	12	1.57E-03
	GO:0006811~ion transport	17	1.88E-03
	GO:0007267~cell-cell signaling	14	3.63E-03
	GO:0048878~chemical homeostasis	12	7.81E-03
	GO:0030001~metal ion transport	11	1.10E-02
	GO:0007155~cell adhesion	14	1.26E-02
	GO:0022610~biological adhesion	14	1.28E-02
	GO:0006812~cation transport	12	1.34E-02
	GO:0055085~transmembrane transport	12	1.62E-02
	GO:0042592~homeostatic process	14	2.14E-02
	GO:0006955~immune response	13	2.58E-02
	GO:0007610~behavior	10	2.98E-02
	GO:0009611~response to wounding	10	5.69E-02
	GO:0010033~response to organic substance	12	6.98E-02
Pathway	hsa04660:T cell receptor signaling pathway	4	1.15E-02
	hsa04340:Hedgehog signaling pathway	3	1.26E-02
	hsa04020:Calcium signaling pathway	5	1.28E-02
	hsa00982:Drug metabolism	3	1.49E-02
	hsa04080:Neuroactive ligand-receptor interaction	6	1.49E-02
	hsa04514:Cell adhesion molecules (CAMs)	4	1.77E-02
	hsa00760:Nicotinate and nicotinamide metabolism	2	2.34E-02
	hsa00534:Heparan sulfate biosynthesis	2	2.51E-02
	hsa04060:Cytokine-cytokine receptor interaction	5	3.26E-02
	hsa04510:Focal adhesion	4	3.82E-02
	hsa00350:Tyrosine metabolism	2	3.87E-02
	hsa05320:Autoimmune thyroid disease	2	4.33E-02
	hsa00830:Retinol metabolism	2	4.52E-02

high risk group and low risk group of the entire set. A miRNA-DEG regulatory network was constructed and functional enrichment analysis suggested the genes involved in the regulatory network were significantly enriched in 17 GO_BP terms and 13 KEGG pathways.

The 8 DE-miRNAs, including miR-1179, miR-133b, miR-3194, miR-3912, miR-548j, miR-6720, miR-6734 and miR-6843 were identified as independent prognosis factors with PTC. Some of these miRNAs were reported to play important roles in PTC previously. MiR-1179 was one of the most down-regulated molecules in samples with a papillary pattern of growth.[21] The downregulation of miR-1179 in PTC could lead to the deregulation of cell cycle progression or invasiveness.[21] MiRNA-1179 was also identified in PTC samples using first next-generation sequencing study.[22] Eleven dysregulated miRNAs (including down-regulated miR 1179) are identified in PTC tissues in comparison to normal thyroid tissue, which are related to different histological types and recurrence risks of PTC.[23]

Though other 7 miRNAs were found differentially expressed in PTC and were found to be independently associated with prognosis of PTC, few studies have conducted to investigate their role in PTC. However, their roles in other cancers were reported. MiR-3194-5p may be candidate targets for blocking the metastasis of the oropharyngeal squamous cell carcinoma development.[24] MiR-3912 and miR-548j expression in embryonal tumors and the controls are differential, which may possess diagnostic value for Atypical Teratoid/Rabdoid Tumors (AT/RTs).[25] MiR-548-3p suppresses the cell proliferation in breast cancer via mediating enoyl-CoA hydratase, short chain 1 (*ECHS1*), which may be a candidate target for the treatment of the tumor.[26] The miR-548e-5p-ubiquitin conjugating enzyme E2 C (UBE2C)-ZEB1/2 axis can promote cell invasion in lung cancer, and may be valuable for diagnosing and treating lung cancer.[27] MiR-133b can significantly suppress cell proliferation, metastasis, and apoptosis in colorectal cancer (CRC) through the TAp63/miR-133b/ras homolog family member A (RhoA) loop.[28] MiR-133b dysregulation may be correlated with *TAp63* down-regulation in CRC, and miR-133b and *TAp63* may be effective prognostic markers for the disease.[29] MiR-9-5p, miR-135a-5p, miR-129-5p, and miR-6720-3p may mediate pathways in cancer, and miR-6720 expression is elevated in hepatocellular carcinoma patients treated by alternariol.[30,31] MiR-6734 suppresses colon cancer cell growth via increasing *p21* levels, therefore, miR-6734 exerts a key regulatory effect on tumor cell proliferation and survival.[32] Thus, these 7 miRNAs might play important roles in PTC and are warrant to be investigated in PTC.

Previous studies have developed other miRNA-based signatures using TCGA data for predicting prognosis of PTC. For example, a 2-miRNA signature (hsa-miR-181a-2-3p and hsa-miR-138-1-3p) was identified by Liu et al and the AUC of this 2-miRNA signature in predicting 5-year survival was 0.784.[33] Xiong et al developed a 4 miRNA signature for predicting survival of patients with PTC with an AUC of 0.886 for 5-year disease survival.[34] The AUC of the risk score system developed in this study for predicting 3-year and 5-year survival was all above 0.93 in the training set, validation set and entire set, which is higher than the miRNA signatures developed in previous study. Besides, the risk score system in this study only included 8 miRNAs. The cost of detecting the expression of these 8 miRNAs by quantitative Real-time PCR is about a few dollars to tens of dollars in different countries. Therefore, this risk score system might be a cost-effective method for predicting the prognosis of PTC patients.

The genes in the miRNA-DEG regulatory network were significantly enriched in 17 GO_ BP terms, such as "regulation of system process" (p-value = 1.36E-04), "regulation of cell proliferation" and 13 KEGG pathways, such as "T cell receptor signaling pathway," "hedgehog signaling pathway" and "calcium signaling pathway." These function terms are the canonical BPs or pathways that were dysregulated in cancer. Through activating *c-Met* and *AKT*, the sonic hedgehog pathway promotes cell motility and invasion in ATC.[35,36] In male rats, calcium signaling is implicated in the mode of action for TC induced by acrylamide exposure.[37] Hedgehog signaling pathway and calcium signaling pathway were enriched for the genes implicated in the regulatory network, indicating that the activation or suppression of the pathways might function in the prognosis of PTC.

Though the risk score system developed in this study is of high sensitively and specificity for predicting the prognosis of PTC patients in TCGA databases, there are some inevitable limitations in this study. First, the external validation of the risk score system is not conducted because there is no sufficient dataset of PTC with complete clinical information. Second, the validation in clinical practice is also warranted in the future study. We hypothesized this study, if confirmed, could significantly improve the prognosis predicting accuracy of PTC patients and gene therapy targeting these 8 miRNAs might improve the prognosis of these patients.

Conclusion

In conclusion, 162 DE-miRNAs between tumor and control groups were screened. Besides, an 8 miRNAs, including miR-1179, miR-133b, miR-3194, miR-3912, miR-548j, miR-6720, miR-6734 and miR-6843 risk score system was developed and this risk score is effective for predicting the prognosis of PTC. Furthermore, a miRNA-DEG regulatory network was constructed based on the 8 independent prognosis-associated DE-miRNAs and their target DEGs. Functional enrichment analysis suggested that these DEGs were significantly enriched in 13 pathways, including hedgehog signaling pathway and calcium signaling pathway. However, more experimental studies should be conducted in the follow-up study to confirm our results.

ORCID iDs

Hengwei Fan 🔟 https://orcid.org/0000-0002-4761-8217
Zhongwei Lv 🔟 https://orcid.org/0000-0003-2808-3357

References

1. Tahara M. Management of recurrent or metastatic thyroid cancer. *Esmo Open*. 2018;3(suppl 1):e000359.
2. Verburg FA, Luster M. Thyroid cancer: balancing benefit and risk in TSH management of DTC. *Nat Rev Endocrinol*. 2018;14(3): 136-137.
3. Nickel B, Brito JP, Moynihan R, Barratt A, Jordan S, Mccaffery K. Patients' experiences of diagnosis and management of papillary thyroid microcarcinoma: a qualitative study. *BMC Cancer*. 2018; 18(1):242.
4. GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. 2016;388(10053)**:**1545-1602.
5. Mcguire S. World Cancer Report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015. *Adv Nutr*. 2016;7(2):418.
6. Verburg FA, Santen HMV, Luster M. Pediatric papillary thyroid cancer: current management challenges. *Onco Targets Ther*. 2017;10**:**165-175.
7. Wang Y, Wang L, Chen C, Chu X. New insights into the regulatory role of microRNA in tumor angiogenesis and clinical implications. *Mol Cancer*. 2018;17(1):22.
8. Wilk G, Braun R. Single nucleotide polymorphisms that modulate microRNA regulation of gene expression in tumors. *PLoS Genet*. 2018;14(12):e1007837.
9. Kan Q, Su Y, Yang H. MicroRNA-335 is downregulated in papillary thyroid cancer and suppresses cancer cell growth, migration and invasion by directly targeting ZEB2. *Oncology Lett*. 2017; 14(16)**:**7622-7628.
10. Ying Y, Zhuang J, Wang G, He S, Jing N, Wei X. MicroRNA-139 targets fibronectin 1 to inhibit papillary thyroid carcinoma progression. *Oncol Lett*. 2017;14(6):7799-7806.
11. Zhang M, Wu W, Gao M, Fei Z. MicroRNA-451 as a prognostic marker for diagnosis and lymph node metastasis of papillary thyroid carcinoma. *Cancer Biomark*. 2017;19(14)**:**437-445.
12. Liu C, Su C, Chen Y, Li G. MiR-144-3p promotes the tumor growth and metastasis of papillary thyroid carcinoma by targeting paired box gene 8. *Cancer Cell Int*. 2018;18:54.
13. Guan H, Liang W, Xie Z, et al. Down-regulation of miR-144 promotes thyroid cancer cell invasion by targeting ZEB1 and ZEB2. *Endocrine*. 2015;48(2):566-574.
14. Pak K, Kim YH, Suh S, Goh TS, Oh SC. Development of a risk scoring system for patients with papillary thyroid cancer. *J Cell Mol Med*. 2019;23(4):3010-3015.
15. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* 2015;43(7):e47.
16. Wang L, Cao C, Ma Q, et al. RNA-seq analyses of multiple meristems of soybean: novel and alternative transcripts, evolutionary and functional implications. *BMC Plant Biol*. 2014;14:169.
17. Moreno-Betancur M, Carlin JB, Brilleman SL, Tanamas SK, Peeters A, Wolfe R. Survival analysis with time-dependent covariates subject to missing data or measurement error: Multiple Imputation for Joint Modeling (MIJM). *Biostatistics*. 2018; 19(4): 479-496.
18. Lin DY. Cox regression analysis of multivariate failure time data: the marginal approach. *Stat Med*. 2010; 13(21):2233-2247.
19. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res*. 2014;42(Database issue):D92-97.
20. Kohl M, Wiese S, Warscheid B. Cytoscape: software for visualization and analysis of biological networks. *Methods Mol Biol*. 2011;696:291-303.
21. Mancikova V, Castelblanco E, Pineiro-Yanez E, et al. MicroRNA deep-sequencing reveals master regulators of follicular and papillary thyroid tumors. *Mod pathol*. 2015;28(6):748-757.
22. Swierniak M, Wojcicka A, Czetwertynska M, et al. In-depth characterization of the microRNA transcriptome in normal thyroid and papillary thyroid carcinoma. *J Clin Endocrinol Metab*. 2013;98(8):E1401-1409.
23. Rosignolo F, Memeo L, Monzani F, et al. MicroRNA-based molecular classification of papillary thyroid carcinoma. *Int J Oncol*. 2017;50(5):1767-1777.
24. Hufbauer M, Maltseva M, Meinrath J, et al. HPV16 increases the number of migratory cancer stem cells and modulates their miRNA expression profile in oropharyngeal cancer. *Int J Cancer*. 2018;143(6):1426-1439.
25. Braoudaki M, Lambrou GI, Giannikou K, et al. MicroRNA expression signatures predict patient progression and disease outcome in pediatric embryonal central nervous system neoplasms. *J Hematol Oncol*. 2014;7:96.
26. Shi Y, Qiu M, Wu Y, Hai L. MiR-548-3p functions as an anti-oncogenic regulator in breast cancer. *Biomed Pharmacother*. 2015;75:111-116.
27. Jin D, Guo J, Wu Y, et al. UBE2C, directly targeted by miR-548e-5p, increases the cellular growth and invasive abilities of cancer cells interacting with the EMT marker protein zinc finger E-box binding homeobox 1/2 in NSCLC. *Theranostics*. 2019;9(7): 2036-2055.
28. Dai J, Wu H, Zhang Y, et al. Negative feedback between TAp63 and Mir-133b mediates colorectal cancer suppression. *Oncotarget*. 2016;7(52):87147-87160.

29. Chen Y, Zhang Y, He J, Fu Y, Lin C, Li X. MicroRNA-133b is regulated by TAp63 while no gene mutation is present in colorectal cancer. *Oncol Rep* 2017;37(3):1646-1652.

30. Vejdovszky K, Sack M, Jarolim K, Aichinger G, Somoza MM, Marko D. In vitro combinatory effects of the Alternaria mycotoxins alternariol and altertoxin II and potentially involved miRNAs. *Toxicol Lett*. 2017;267:45-52.

31. Zhang K, Chen J, Zhang D, et al. microRNA expression profiles of scar and normal tissue from patients with posterior urethral stricture caused by pelvic fracture urethral distraction defects. *Int J Mol Med*. 2018;41(5):2733-2743.

32. Kang MR, Park KH, Yang JO, et al. miR-6734 up-regulates p21 gene expression and induces cell cycle arrest and apoptosis in colon cancer cells. *PLoS One*. 2016;11(8):e0160961.

33. Liu T, You X, Sui J, et al. Prognostic value of a two-microRNA signature for papillary thyroid cancer and a bioinformatic analysis of their possible functions. *J Cell Bio*. 2019;120(5): 7185-7198.

34. Xiong C, Cai G, Zhang J, Yunxia L. MicroRNA signature predicts survival in papillary thyroid carcinoma. *J Cell Biochem*. 2019; 120(10):17050-17058.

35. Williamson AJ, Doscas ME, Ye J, et al. The sonic hedgehog signaling pathway stimulates anaplastic thyroid cancer cell motility and invasiveness by activating Akt and c-Met. *Oncotarget*. 2016; 7(9):10472-10485.

36. Xu X, Lu Y, Li Y, Prinz RA. Sonic hedgehog signaling in thyroid cancer. *Front Endocrinol (Lausanne)* 2017;8:284.

37. Chepelev NL, Gagne R, Maynor T, et al. Transcriptional profiling of male F344 rats suggests the involvement of calcium signaling in the mode of action of acrylamide-induced thyroid cancer. *Food Chem Toxicol*. 2017;107(Pt A): 186-200.