

# A novel microRNA signature predicts survival in stomach adenocarcinoma

Bowen Ding<sup>1,\*</sup>, Xujie Gao<sup>1,\*</sup>, Hui Li<sup>1</sup>, Liren Liu<sup>1</sup>, Xishan Hao<sup>1</sup>

<sup>1</sup>Department of Gastrointestinal Cancer Biology, National Clinical Research Center of Cancer, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, China

\*These authors contributed equally to this work

**Correspondence to:** Liren Liu, **email:** liuliren@tjmuch.com  
Xishan Hao, **email:** haoxishan@tjmuch.com

**Keywords:** miRNA signature, prognostic biomarkers, overall survival, stomach adenocarcinoma

**Received:** November 10, 2016

**Accepted:** February 22, 2017

**Published:** March 07, 2017

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

## ABSTRACT

**Recent microRNA (miRNA) expression profiling studies suggest the clinical use of miRNAs as potential prognostic biomarkers in various malignancies. In this study, aiming to identify microRNAs with prognostic value for overall survival (OS) in stomach adenocarcinoma (STAD) patients, we analyzed the miRNA expression profiles and the associated clinical characteristics in 380 STAD samples from The Cancer Genome Atlas (TCGA) dataset. An eight-miRNA signature for predicting OS in STAD patients was identified and self-validated by survival analysis and semi-supervised principal components method. We developed a linear prognostic model composed of these miRNAs to divide patients into high- and low-risk groups according to the calculated prognostic scores. Kaplan-Meier analysis demonstrated that patients in the high-risk group had worse OS compared with patients in the low-risk group. Notably, this miRNA prognostic model showed prognostic significance to the STAD patients in early stages and the chemo-resistant patients, who would potentially benefit from additional medical interventions. Finally, this eight-miRNA signature is an independent prognostic biomarker and demonstrates a good predictive performance for 5-year survival. Thus, this signature may serve as a novel biomarker for predicting survival as well as chemotherapy response in STAD patients.**

## INTRODUCTION

Gastric cancer (GC) represents a major health problem worldwide, ranking the fifth most common malignancy and the third leading cause of global cancer mortality [1]. Early diagnosis is critical to the prognosis of stomach adenocarcinoma (STAD) patients, which constitute the majority (> 90%) of GC patients. Unfortunately, most STAD patients are diagnosed at advanced stage disease which is manifested by extensive tumor invasion and/or distant metastasis, and results in a poor 5-year survival rate with the median survival of less than 1 year [2–4]. Over the past few decades, despite significant progress in diagnostics and therapeutics, the overall outcomes of GC patients has only modestly improved [5]. Thus, it is essential to discover specific prognostic factors that may help guide the clinical therapeutic implications and improve the overall survival (OS) for these patients.

In recent years, various miRNAs and their targets have been found to be dysregulated in cancer. Due to the stability and expression specificity of miRNAs in different tissues, there is increasing evidence to suggest that they can serve as novel prognostic biomarkers with potential clinical significance for various cancer types. Thus far, dysregulation of miRNAs in GC has been reported to be associated with histology, Epstein-Barr virus infection, chemotherapy, and metastasis [6–11]. A number of miRNAs were further proposed to be potential prognostic biomarkers for outcomes of GC patients, such as miR-20b, miR-150, miR-214, miR-375, Let-7g, miR-125-5p, miR-146a, miR-218, miR-433, miR-451 and miR-200b/c [6, 12–14] [15–19]. Moreover, a seven-miRNA signature (miR-10b, miR-21, miR-223, miR-338, let-7a, miR-30a-5p, miR-126) has recently been identified as an independent predictor for relapse-free survival of GC patients [16]. However, the miRNAs characterized from these studies showed a wide range of diversity and

inconsistency, which may be primarily due to the fact that these results were mostly derived from the analysis based on microarray data. As we know, besides the limitations in the dynamic range, sensitivity, and specificity of the microarray data themselves, different technological detection platforms, small sample size, various methods employed for data processing and analysis may result in significant variations.

With the advent of next-generation sequencing and bioinformatics as well as the launches of large-scale cancer genome projects, such as The Cancer Genome Atlas (TCGA), comprehensive and multi-dimensional maps of the key genomic and epigenomic changes in cancer can be readily achieved and accessed. Bioinformatic analysis of TCGA datasets has been shown to be an outstanding tool in identifying genetic and epigenetic changes related to clinical outcomes, thus opening a new avenue for the discovery of novel prognostic markers for various malignancies [20, 21]. In this study, aiming to identify a panel of specific miRNAs associated with OS in STAD patients, we applied the bioinformatics analysis based on the genome-wide miRNA-Seq profiles derived from TCGA dataset, which represents the largest STAD cohort available up to date. To minimize the variations among different individuals, we first characterized the differentially expressed miRNAs by comparing miRNA profiles in paired STAD and normal tissues. We then developed a linear prognostic model composed of eight miRNAs to divide patients into high- and low-risk groups according to their calculated prognostic scores. The following Kaplan-Meier analysis demonstrated that the eight-miRNA signature correlated with a good predictive performance for 5-year survival and chemotherapy response in STAD patients, suggesting that this signature may serve as a novel biomarker to complement the traditional histopathological prognostic factors and help guide individual therapy for these patients.

## RESULTS

### Differentially expressed miRNAs in STAD versus paired adjacent normal tissue

After filtering out unqualified cases, miRNA expression and clinical data of 380 STAD patients were retained for survival analysis. It contained 252 male and 128 female, among all the participants a total of 41 patients with adjacent non-tumor tissues. The patients were randomly divided into the training set ( $n = 190$ ) and testing set ( $n = 190$ ). No significant difference in clinical covariates was observed between the two sets (Table 1).

Analysis of miRNA expression profiles in 41 pairs of STAD and normal tissues identified 138 differentially expressed miRNAs ( $\log_{2}FC > 1$  or  $\log_{2}FC < -1$ ,  $P < 0.05$  after FDR adjustment). Among these, 77 miRNAs (55.8%) were up-regulated, including miR-1269 and miR-196a-1,

which exhibited  $> 5$ -fold increased expression. Conversely, 61 miRNAs (44.2%) were down-regulated, including miR-490 and miR-1-2, which showed 4.8-fold and 3.5-fold reduced levels in STAD, respectively (Supplementary Table 1).

### Establishment of the miRNA prognostic model

Using univariate Cox regression, we characterized the common miRNAs that were associated with OS within each of the following independent subclass: differentiation, pathologic N stage, pathologic T stage, and pathologic M stage. Within each subset of clinical characteristics, the patient subclasses represented non-overlapping sets, respectively. MiRNAs were selected if they were associated significance with OS in at least two independent subclasses. The respective HRs for the association of miRNA with OS in each subclass were shown in Table 2. Seventeen miRNAs were identified in this analysis.

Eight of 17 miRNAs were selected using the supervised principal component method in the training set. Among these eight miRNAs, five were associated with high risk (miR-145, miR-184, miR-20b, miR-9-1 and miR-9-2,  $HR > 1$ ) and three were shown to be protective (miR-1537, miR-549 and miR-802,  $HR < 1$ ). We then developed a miRNA prognostic model for predicting 5-year survival in the training set, by which the samples were classified into high-risk or low-risk groups using the optimum cutoff point of miRNA scores according to ROC curve. Figure 1 showed the distribution of patient prognostic scores and miRNA expression of all 380 STAD patients, ranked by the prognostic score values for the eight-miRNA signature. Patients with high prognostic scores tended to express high-risk miRNAs, whereas patients with low prognostic scores tended to express protective miRNAs (Figure 1A and 1B).

### Validation of the eight-miRNA signature in STAD patients

Using the optimum cutoff value obtained from the training set, patients were assigned into high-risk and low-risk groups. The ability of prognostic prediction of the eight-miRNA signature was examined in the testing set and the entire STAD cohort, respectively. Kaplan-Meier analysis revealed that patients in the high-risk group had poor OS, compared with patients in the low-risk group in both testing set ( $p < 0.001$ , using the log-rank test, Figure 2A) and the entire STAD cohort ( $p < 0.001$ , using the log-rank test, Figure 3A). Time-dependent ROC curves were also used to assess the prognostic power of the eight-miRNA signature. The AUC of the eight-miRNA signature prognostic model for the testing set and the entire STAD cohort was 0.586 (Figure 2B) 0.607 (Figure 3B), respectively.

Since patients with early stage disease may benefit significantly from a prognostic biomarker signature,

**Table 1: Clinicopathological characteristics of the study cohort**

Variable	Total (n = 380)	Training Set (n = 190)	Testing Set (n = 190)	P
Age(years)				
< 60	120 (31.6%)	67 (35.3%)	53 (27.9%)	0.122 *
≥ 60	260 (68.4%)	123 (64.7%)	137 (72.1%)	
Sex				0.664
Male	252 (66.3%)	124 (65.3%)	128 (67.4%)	
Female	128 (33.7%)	66 (34.7%)	62 (32.6%)	
Vital status				1.000
Alive	230 (60.5%)	115 (60.5%)	115 (60.5%)	
Dead	150 (39.5%)	75 (39.5%)	75 (39.5%)	
Stage				0.708
I	47 (12.4%)	24 (12.6%)	23 (12.1%)	
II	132 (34.7%)	67 (35.3%)	65 (34.2%)	
III	182 (47.9%)	92 (48.4%)	90 (47.4%)	
IV	19 (5.0%)	7 (3.7%)	12 (6.3%)	
T stage				0.987
T1	20 (5.3%)	10 (5.3%)	10 (5.3%)	
T2	71 (18.7%)	36 (18.9%)	35 (18.4%)	
T3	126 (33.1%)	61 (32.1%)	65 (34.2%)	
T4	163 (42.9%)	83 (43.7%)	80 (42.1%)	
N stage				0.750
N0	118 (31.0%)	61 (32.1%)	57 (30.0%)	
N1	106 (27.9%)	51 (26.8%)	55 (28.9%)	
N2	74 (19.5%)	40 (21.1%)	34 (17.9%)	
N3	82 (21.6%)	38 (20.0%)	44 (23.2%)	
M stage				0.282
M0	357 (93.9%)	181 (95.3%)	176 (92.6%)	
M1	23 (6.1%)	9 (4.7%)	14 (7.4%)	
Adjuvant treatment				0.993
None	166 (58.9%)	83 (43.7%)	83 (43.7%)	
Chemotherapy	153 (40.3%)	79 (41.5%)	74 (38.9%)	
Radiotherapy	3(0.8%)	2(1.1%)	1 (0.6%)	
Chemoradiotherapy	58 (15.3%)	26 (13.7%)	32 (16.8%)	

we evaluated the prognostic power of the eight-miRNA signature in stage I and II patients ( $n = 179$ ). Kaplan-Meier analysis revealed that patients in the high-risk group associated with worse OS ( $p < 0.001$ , using the log-rank test). Thus, this eight-miRNA signature could distinctly predict the survival for STAD patients in an early stage, suggesting that it could have potential clinical value for help in guiding additional medical interventions to this patient subgroup. (Figure 4)

It is often challenging to distinguish chemoresistant STAD patients from those that have good responses to chemotherapy. In order to evaluate the prognostic value of the eight-miRNA signature to chemotherapy response in STAD patients, we applied the eight-miRNA signature to the 112 patients, whose post-operative chemotherapy responses were recorded in the

TCGA STAD cohort. Using our eight-miRNA prognostic model, the 112 patients were divided into low-risk and high-risk subgroups according to their responses to chemotherapy. As shown in Table 3, among 68 low-risk patients, most patients had either a complete or partial response after chemotherapy, while only 7 patients in this subgroup showed disease progression. In contrast, among 44 high-risk patients, approximately one-third of patients showed progression after chemotherapy. Thus, this eight-miRNAs signature showed a good prognostic power for predicting chemotherapy response in STAD patients ( $P$  value  $< 0.05$ ).

We also examined the association of eight-miRNA signature with clinical characteristics in STAD patients. No significant differences were observed when patients were stratified by gender and age (Supplementary Table 2).

**Table 2: MiRNAs associated with prognosis in different clinical subclasses**

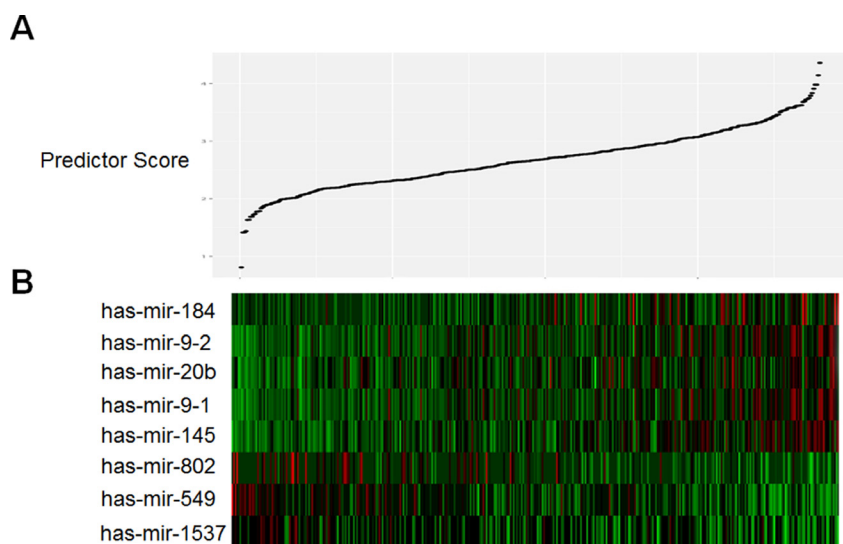
miRNA	T1-2 HR (95% CI)	T3-4 HR (95%CI)	N0 HR (95% CI)	N1-3 HR (95% CI)	M0 HR (95% CI)	M1 HR (95% CI)	Grade1-2 HR (95% CI)	Grade3 HR (95% CI)
miR-100	1.31 (1.03–1.66)	-	-	1.20 (1.03–1.40)	1.21 (1.05–1.40)	-	-	-
miR-125a	1.58 (1.10–2.28)	1.29 (1.00–1.67)	-	1.40 (1.10–1.78)	1.32 (1.06–1.63)	-	1.72 (1.23–2.39)	-
miR-125b-1	1.30 (1.02–1.65)	-	-	-	1.17 (1.01–1.34)	-	-	-
<b>miR-145</b>	-	-	-	1.16 (1.01–1.34)	1.16 (1.02–1.31)	-	1.25 (1.01–1.56)	-
<b>miR-1537</b>	-	-	-	0.79 (0.68–0.93)	0.84 (0.72–0.98)	-	-	-
<b>miR-184</b>	-	-	-	-	-	1.23 (1.02–1.46)	1.18 (1.04–1.34)	-
<b>miR-20b</b>	-	1.14 (1.01–1.30)	1.30 (1.07–1.57)	-	-	-	1.19 (1.01–1.41)	-
miR-28	-	1.45 (1.06–1.98)	-	-	-	-	1.68 (1.04–2.71)	-
miR-30a	-	1.22 (1.03–1.51)	-	-	1.19 (1.02–1.40)	-	-	-
miR-328	-	1.25 (1.03–1.51)	-	1.25 (1.05–1.49)	1.19 (1.02–1.39)	-	1.33 (1.05–1.69)	-
miR-365-1	-	-	1.72 (1.08–2.74)	-	-	-	1.54 (1.04–2.27)	-
miR-383	-	-	1.33 (1.07–1.66)	-	-	-	1.24 (1.01–1.52)	-
<b>miR-549</b>	0.79 (0.63–0.98)	-	0.73 (0.55–0.98)	-	0.87 (0.76–0.99)	-	-	-
<b>miR-802</b>	-	-	-	-	0.88 (0.77–0.99)	-	-	0.85 (0.73–0.99)
<b>miR-9-1</b>	-	1.14 (1.02–1.26)	1.22 (1.00–1.48)	-	1.10 (1.00–1.20)	--	-	1.15 (1.02–1.31)
<b>miR-9-2</b>	-	1.14 (1.03–1.27)	1.21 (1.00–1.47)	-	1.10 (1.00–1.20)	--	-	1.16 (1.02–1.31)
miR-99a	1.23 (1.00–1.50)	-	-	-	1.13 (1.01–1.26)	-	-	-

**The eight-miRNA signature is an independent prognostic factor**

We further conducted a multivariate analysis to evaluate the independent prognostic value of the eight-miRNA signature. Age, gender, Grade, T stage, N stage, M stage and the miRNA signature were used as covariates. The Cox multivariate regression analysis revealed that the miRNA signature is an independent prognostic factor associated with OS (Table 4; HR = 2.003,  $p < 0.001$ ).

**In silico analysis of target genes and pathways**

The list of predicted target genes of these eight miRNAs was downloaded from miRecords. A total of 3510 target genes which predicted by more than 5 data sets were identified to be potentially regulated by the eight miRNAs. We then performed a functional enrichment analysis to elucidate the biological function of these target genes by Kyoto Encyclopedia of Genes and Genomes (KEGG) categories and Gene Ontology (GO) categories.

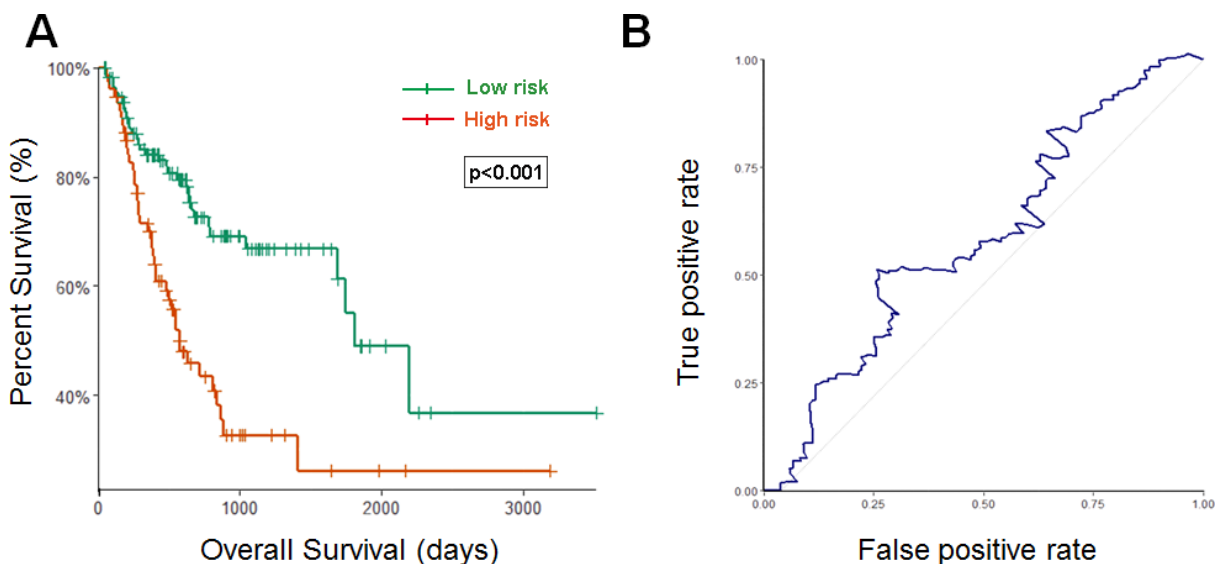


**Figure 1: Heatmap and predictor-score of eight-MicroRNA signature of STAD cohort. (A) MicroRNA predictor-score distribution. (B) Heatmap of eight miRNAs expression profiles of STAD patients.**

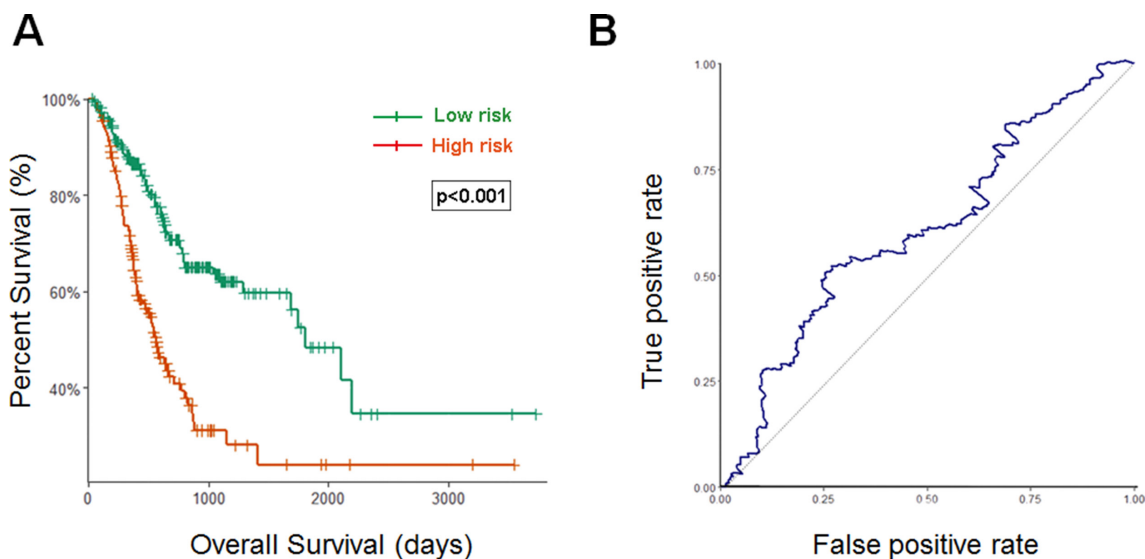
As shown in Supplementary Table 3, the analysis revealed enrichment of 45 KEGG categories and 714 GO categories ( $P$ -values  $< 0.05$  after FDR adjustment), demonstrating the predicted miRNA targets are involved in many important pathways associated with cancer development, including adherent junction, Wnt, TGF-beta, VEGFR and MAPK signaling pathways (Table 5). These results highlighted critical roles of these eight miRNAs in STAD onset and progression, and the underlying mechanisms warrant further investigation.

## DISCUSSION

MiRNAs are 22–26 nucleotides small RNAs that play vital roles in modulating gene expression at the post-transcriptional level by binding to the 3' or 5' untranslated region of targeted mRNAs. Here we found 138 miRNAs were differentially expressed between STAD and adjacent normal tissues, among which 17 miRNAs levels were associated with at least two of following histopathological factors: T stage, lymph



**Figure 2: Kaplan–Meier and ROC curves for the eight-miRNA signature in STAD testing set.** (A) The Kaplan–Meier curves for testing set ( $n = 190$ ) divided by the optimum cutoff point. Patients with high scores had the poor outcome in terms of OS (Median OS: 1811days vs. 570 days,  $p < 0.001$ ). (B) The ROC curve for predicting 60-month survival for testing set.



**Figure 3: Kaplan–Meier and ROC curves for the eight-miRNA signature in STAD cohort.** (A) The Kaplan–Meier curves for entire STAD cohort divided by the optimum cutoff point. Patients with high scores had the poor outcome in terms of OS (Median OS: 1811 days vs. 562 days,  $p < 0.001$ ). (B) The ROC curve for predicting 60-month survival for STAD cohort.



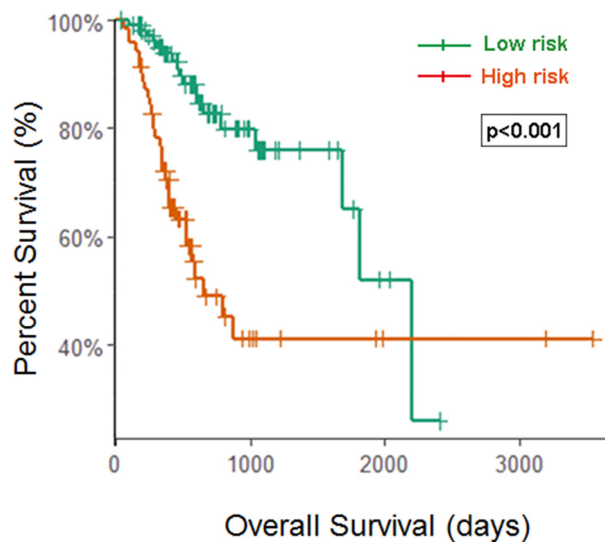
node metastasis, distant metastasis and tumor grade. Using survival analysis and semi-supervised principal components method, an eight-miRNA signature (miR-145, miR-184, miR-20b, miR-9-1, miR-9-2, miR-1537, miR-549 and miR-802) for predicting OS of STAD patients was identified and self-validated in a large cohort. Importantly, this eight-miRNA signature was further confirmed to be an independent prognostic factor. Furthermore, with respect to the association between their expression levels and patient survival, the eight miRNAs in the signature were divided into two groups: five risky miRNAs that were negatively associated with the survival and three protective miRNAs which were positively associated with the survival.

Among the five risky miRNAs (miR-145, miR-184, miR-20b, miR-9-1 and miR-9-2), overexpression of miR-145 has been reported to be a key factor for the prediction of poor overall survival in bladder cancer [22]. However, a recent study based on qRT-PCR showed that downregulation of miR-145 correlates with a poor survival in a clinical cohort composed of 145 GC patients from the north of China, which is inconsistent with our result that miR-145 was identified as one of the five risky miRNAs leading to poor 5-year survival. The discrepancy among these studies could be because there are distinct expression patterns of miR-145 in either different cancer types or the same cancer type with subtle variations, such as race. Upregulation of miR-184 enhanced the malignant phenotype of glioma cancer cells by reducing FIH-1 protein expression and facilitated the proliferation and invasion in pancreatic ductal adenocarcinoma [23, 24]. Overexpression of miR-20b was reported to be associated with poor outcomes in GC patients [6, 25]. Also, high level of miR-20b facilitated brain metastasis of breast cancer and promoted hepatocellular carcinoma invasion

and progression [26, 27]. The role of miR-9 has been investigated in many types of malignancy, but the results were inconsistent and inconclusive. Recently, a systematic meta-analysis, representing a quantified synthesis of all published studies of miR-9, found that high expression of miR-9 was significantly associated with poor survival in patients with malignancies, including colorectal cancer, non-small lung cancer, breast cancer, hepatocellular carcinoma, laryngeal and esophageal squamous cell carcinomas, glioma, ovarian, osteosarcoma, adrenocortical cancer, bladder cancer, and leukemia [28].

Among the protective miRNAs (miR-1537, miR-549 and miR-802), deletion of miR-1537 has been reported in aggressive neuroblastoma [29]. During neoadjuvant therapy in the patients with advanced adenocarcinomas of the gastroesophageal junction, there was a significant increase of serum miR-549, suggesting a suppressive role it played in tumor progression [30]. Downregulation of miR-802 was significantly associated with overall survival and cancer-specific survival in rectal cancer patients [31]. Taken together, these biological and clinical studies of the miRNAs have provided some insights into their potential prognostic value, although future work is needed to validate their roles in clinical applications.

It is of note that this eight-miRNA prognostic model could predict the high-risk patients in early stages (stage I-II), which accounts for approximately 10% of all STAD recurrences [32]. These patients may benefit significantly from a prognostic biomarker to guide their further medical interventions, such as the shorter interval between the exams for monitoring the recurrence or alternative treatment approaches. Since chemotherapy resistance is one of the major factors leading to a poor OS for STAD patients, we applied our miRNA signature for the prediction of chemotherapy response. Indeed, the



**Figure 4: Kaplan–Meier curves for the eight-miRNA signature in early stage patients.** Patients with high scores had poor outcome in terms of OS (Median OS: 2197 days vs. 652 day,  $p < 0.001$ ).

**Table 3: Association of eight-miRNA signature with chemotherapy response**

miRNA signature (#)	CR&PR	SD	Progress	p-value
Low-risk (68)	55	6	7	0.017
High-risk (44)	27	3	14	

CR: complete response, PR: partial response, SD: stable disease.

**Table 4: Multivariate analysis of overall survival of patients**

Characteristic	HR (95% CI)	P value
Gender (male vs. female)	0.781 (0.551–1.107)	0.165
Age (< 60 vs. ≥ 60 years)	1.719 (1.186–2.490)	<b>0.004</b>
Grade (Grade 1–2 vs Grade 3)	1.315 (0.928–1.863)	0.123
T stage (T1–2 vs T3–4)	1.417 (0.896–2.240)	0.136
N stage (N0 vs N1–3)	1.786 (1.184–2.695)	<b>0.006</b>
M stage (M0 vs M1)	2.346 (1.328–4.146)	<b>0.003</b>
miRNA signature	2.003 (1.458–2.752)	<b>&lt; 0.001</b>

**Table 5: Results of pathway analysis of the target genes**

Pathways	Target genes
Wnt signaling pathway	CAMK2D, CCND2, CTNBP1, LRP6, PPP3CA, PRKX, ROCK1, SENP2, SMAD3, VANGL1
MAKP signaling pathway	CRK, CRKL, ELK4, FLNB, MAP4K2, PAK1, PLA2G4A, PPP3CA, PRKX, RAPGEF2, RASA1, TAOK1, TGFB2
Adherens junction	ACTB, ACTG1, IGF1R, SMAD3, TGFB2, TJP1, YES1
TGF-beta signaling pathway	ACVR1B, ACVR2A, ROCK1, SMAD3, SMAD5, TGFB2, ZFYVE9
VEGF receptor signaling pathway	ARNT, NEDD4, LECT1, BMPR2, HIF1A, FLT1, VEGFA, HHEX, GRB10, FOXC1

chemoresponsive or chemoresistant patients were well distinguished using this eight-miRNA signature, which made it possible in clinical practice to stratify patients into two groups: one that could benefit from standard therapy and another that should be placed on alternative therapeutic protocols or novel clinical trials.

Bioinformatic analysis of potential and validated targets of miRNA and the subsequent analysis of KEGG pathways clustered by target genes are promising strategies for gaining insights into plausible biomarkers and key events involved in cancer development and progression. Enriched KEGG pathway analysis of the eight-miRNA model indicated that both predicted and validated target genes of this miRNA signature were clustered in cancer-associated KEGG pathways. Furthermore, our *in silico* analysis proved that the eight-miRNA signature is biologically meaningful. Specifically, Ago2, the co-target of miRNA-145, miR-184 and miR-802, played an important role in GC differentiation, lymph node invasion and clinical stage [33]. Myc, a multiplayer in carcinogenesis, progression and metabolism, was also the predicted target of miR-145 and miR-184 [34]. Moreover, many target genes related to VEGF receptor signaling pathway have been identified, such as VEGFR-3. Indeed, high expression level of VEGFR-3 has been found to be associated poor survival of gastric adenocarcinoma [35]. Certainly, further

molecular biological experiments need to be performed to validate these predictions.

In summary, using integrated bioinformatic analysis of the largest available cohort of STAD patients and the corresponding genome-wide miRNA sequencing results, we identified a specific eight-miRNA signature that could serve as an independent prognostic factor for the prediction of OS in STAD patients. However, before this finding can be applied to clinical practice, further validation in independent large cohorts is required in the future.

## MATERIALS AND METHODS

### TCGA dataset

Level 3 of 1049 miRNAs expression profiles in stomach adenocarcinoma (STAD) patients and their clinical information dataset were downloaded from The Cancer Genome Atlas (TCGA) data portal (March 2016). To exclude unrelated causes of death, only patients with follow-up longer than 1 month were included in the subsequent analysis.

### Identification of dysregulated miRNAs in STAD

The raw counts of miRNA expression data of 41STAD with their paired normal tissue were obtained

from the TCGA dataset (Illumina HiSeq Systems). MiRNA-expression data was normalized by the R/Bioconductor package edgeRv [36]. The expression differences were characterized by logFC (log2 fold change). MiRNAs with logFC < -1 or logFC > 1 (FDR-adjusted  $p < 0.05$ ) were considered as differentially expressed miRNAs and were selected for further analysis.

### Identification of miRNAs with prognostic value in STAD

The semi-supervised method which combines both the gene expression data and the clinical data was used to identify candidate miRNAs with prognostic value [37, 38]. Univariate Cox regression analyses were conducted to identify common miRNAs related to OS within each of the subgroups stratified by the TMN stage. Within each group of clinical characteristics, the patient subclasses represented non-overlapping sets. Common miRNAs associated with OS in at least two independent subgroups were selected for the subsequent studies, using an HR > 1 or HR < 1 with  $p < 0.05$  as the cutoff.

### Definition of prognostic risk model and ROC curve analysis

An importance score was calculated by the supervised principal components method and was assigned to each miRNA [37]. Ten-fold cross-validation was used to estimate the optimal feature threshold in supervised principal components and to select significant miRNAs. The TCGA dataset was randomly divided into the training set and the testing set. The linear miRNA signature prognostic model was developed based on the supervised principal component method. The miRNAs expression level was as the log2 reads per million of total aligned miRNA reads. The prognostic score was calculated as follows: Prognostic-score =  $(0.1482 \times \text{miR-145}) + (-0.0987 \times \text{miR-1537}) + (0.1126 \times \text{miR184}) + (0.0964 \times \text{miR20b}) + (-0.1662 \times \text{miR-549}) + (-0.1374 \times \text{miR-802}) + (0.0915 \times \text{miR-9-1}) + (-0.0148 \times \text{miR-9-2})$ . Then, the prognostic scores were computed for the 380 patients using our miRNA prognostic model. The best cutoff values of the prognostic score were decided in the ROC curve analysis for predicting 5-year survival of the training set. The OS curves were evaluated using the Kaplan–Meier and log-rank method. Time-dependent ROC curves were used to evaluate the predicted power of the miRNAs signature model. All analyses were performed using the R/Bioconductor (version 3.3.1).

### Bioinformatic analysis of miRNA-target genes and pathways

The list of predicted target genes of the candidate miRNAs was obtained from miRecords v4.0 ([www.mirecords.biolead.org](http://www.mirecords.biolead.org)).

database, which offers a comprehensive data of possible miRNA-targets of 11 different data sets. The pathway enrichment analysis was conducted with the GeneTrail gene set enrichment tool. The results were considered significant when the  $p$ -value was less than 0.05 after FDR corrected [39, 40].

### ACKNOWLEDGMENTS AND FUNDING

This work was supported by the National Natural Science Foundation of China (No.81572416), the National Key Technologies R&D Program of China (No.2016YFC1303200) and the Tianjin Natural Science Funds (No.15JCYBJC54500). Our gratitude also goes to The Cancer Genome Atlas project for generating and providing high-quality biological and clinical data on stomach adenocarcinoma.

### CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

### REFERENCES

1. Colquhoun A, Arnold M, Ferlay J, Goodman KJ, Forman D, Soerjomataram I. Global patterns of cardia and non-cardia gastric cancer incidence in 2012. *Gut*. 2015; 64:1881–8. doi: 10.1136/gutjnl-2014-308915.
2. Correa P. Is gastric cancer preventable? *Gut*. 2004; 53:1217–9. doi: 10.1136/gut.2004.039834.
3. Cervantes A, Roda D, Tarazona N, Rosello S, Perez-Fidalgo JA. Current questions for the treatment of advanced gastric cancer. *Cancer Treat Rev*. 2013; 39:60–7. doi: 10.1016/j.ctrv.2012.09.007.
4. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA Cancer J Clin*. 2014; 64:9–29. doi: 10.3322/caac.21208.
5. Lazar DC, Taban S, Cornianu M, Faur A, Goldis A. New advances in targeted gastric cancer treatment. *World J Gastroenterol*. 2016; 22:6776–99. doi: 10.3748/wjg.v22.i30.6776.
6. Mathers JC, Strathdee G, Relton CL. Induction of epigenetic alterations by dietary and other environmental factors. *Adv Genet*. 2010; 71:3–39. doi: 10.1016/B978-0-12-380864-6.00001-8.
7. Mitra SA, Mitra AP, Triche TJ. A central role for long non-coding RNA in cancer. *Front Genet*. 2012; 3:17. doi: 10.3389/fgene.2012.00017.
8. Shrestha S, Hsu SD, Huang WY, Huang HY, Chen W, Weng SL, Huang HD. A systematic review of microRNA expression profiling studies in human gastric cancer. *Cancer Med*. 2014; 3:878–88. doi: 10.1002/cam4.246.
9. Katada T, Ishiguro H, Kuwabara Y, Kimura M, Mitui A, Mori Y, Ogawa R, Harata K, Fujii Y. microRNA expression profile in undifferentiated gastric cancer. *Int J Oncol*. 2009; 34:537–42.



10. Treece AL, Duncan DL, Tang W, Elmore S, Morgan DR, Dominguez RL, Speck O, Meyers MO, Gulley ML. Gastric adenocarcinoma microRNA profiles in fixed tissue and in plasma reveal cancer-associated and Epstein-Barr virus-related expression patterns. *Lab Invest.* 2016; 96:661–71. doi: 10.1038/labinvest.2016.33.
11. Kim CH, Kim HK, Rettig RL, Kim J, Lee ET, Aprelikova O, Choi IJ, Munroe DJ, Green JE. miRNA signature associated with outcome of gastric cancer patients following chemotherapy. *BMC Med Genomics.* 2011; 4:79. doi: 10.1186/1755-8794-4-79.
12. Huang KH, Lan YT, Fang WL, Chen JH, Lo SS, Li AF, Chiou SH, Wu CW, Shyr YM. The correlation between miRNA and lymph node metastasis in gastric cancer. *Biomed Res Int.* 2015; 2015:543163. doi: 10.1155/2015/543163.
13. Tchernitsa O, Kasajima A, Schafer R, Kuban RJ, Ungethum U, Gyorffy B, Neumann U, Simon E, Weichert W, Ebert MP, Rocken C. Systematic evaluation of the miRNA-ome and its downstream effects on mRNA expression identifies gastric cancer progression. *J Pathol.* 2010; 222:310–9. doi: 10.1002/path.2759.
14. Liu K, Qian T, Tang L, Wang J, Yang H, Ren J. Decreased expression of microRNA let-7i and its association with chemotherapeutic response in human gastric cancer. *World J Surg Oncol.* 2012; 10:225. doi: 10.1186/1477-7819-10-225.
15. Ueda T, Volinia S, Okumura H, Shimizu M, Taccioli C, Rossi S, Alder H, Liu CG, Oue N, Yasui W, Yoshida K, Sasaki H, Nomura S, et al. Relation between microRNA expression and progression and prognosis of gastric cancer: a microRNA expression analysis. *Lancet Oncol.* 2010; 11:136–46. doi: 10.1016/S1470-2045(09)70343-2.
16. Xiong X, Ren HZ, Li MH, Mei JH, Wen JF, Zheng CL. Down-regulated miRNA-214 induces a cell cycle G1 arrest in gastric cancer cells by up-regulating the PTEN protein. *Pathol Oncol Res.* 2011; 17:931–7. doi: 10.1007/s12253-011-9406-7.
17. Zhang X, Yan Z, Zhang J, Gong L, Li W, Cui J, Liu Y, Gao Z, Li J, Shen L, Lu Y. Combination of hsa-miR-375 and hsa-miR-142-5p as a predictor for recurrence risk in gastric cancer patients following surgical resection. *Ann Oncol.* 2011; 22:2257–66. doi: 10.1093/annonc/mdq758.
18. Zhang Y, Hu Y, Fang JY, Xu J. Gain-of-function miRNA signature by mutant p53 associates with poor cancer outcome. *Oncotarget.* 2016; 7: 11056-66. doi: 10.18632/oncotarget.7090.
19. Li X, Zhang Y, Zhang Y, Ding J, Wu K, Fan D. Survival prediction of gastric cancer by a seven-microRNA signature. *Gut.* 2010; 59:579–85. doi: 10.1136/gut.2008.175497.
20. Tsai MM, Wang CS, Tsai CY, Huang HW, Chi HC, Lin YH, Lu PH, Lin KH. Potential Diagnostic, Prognostic and Therapeutic Targets of MicroRNAs in Human Gastric Cancer. *Int J Mol Sci.* 2016; 17. doi: 10.3390/ijms17060945.
21. Wang QX, Zhu YQ, Zhang H, Xiao J. Altered MiRNA expression in gastric cancer: a systematic review and meta-analysis. *Cell Physiol Biochem.* 2015; 35:933–44. doi: 10.1159/000369750.
22. Song F, Yang D, Liu B, Guo Y, Zheng H, Li L, Wang T, Yu J, Zhao Y, Niu R, Liang H, Winkler H, Zhang W, et al. Integrated microRNA network analyses identify a poor-prognosis subtype of gastric cancer characterized by the miR-200 family. *Clin Cancer Res.* 2014; 20:878–89. doi: 10.1158/1078-0432.CCR-13-1844.
23. Fleming NH, Zhong J, da Silva IP, Vega-Saenz de Miera E, Brady B, Han SW, Hanniford D, Wang J, Shapiro RL, Hernando E, Osman I. Serum-based miRNAs in the prediction and detection of recurrence in melanoma patients. *Cancer.* 2015; 121:51–9. doi: 10.1002/cncr.28981.
24. Barbano R, Palumbo O, Pasculli B, Galasso M, Volinia S, D'Angelo V, Icolaro N, Coco M, Dimitri L, Graziano P, Copetti M, Valori VM, Maiello E, et al. A miRNA signature for defining aggressive phenotype and prognosis in gliomas. *PLoS One.* 2014; 9:e108950. doi: 10.1371/journal.pone.0108950.
25. Clark NR, Szymkiewicz M, Wang Z, Monteiro CD, Jones MR, Ma'ayan A. Principal Angle Enrichment Analysis (PAEA): Dimensionally Reduced Multivariate Gene Set Enrichment Analysis Tool. *Proceedings (IEEE Int Conf Bioinformatics Biomed).* 2015; 2015:256–62. doi: 10.1109/BIBM.2015.7359689.
26. Gao C, Weisman D, Lan J, Gou N, Gu AZ. Toxicity mechanisms identification via gene set enrichment analysis of time-series toxicogenomics data: impact of time and concentration. *Environ Sci Technol.* 2015; 49:4618–26. doi: 10.1021/es505199f.
27. Charmpi K, Ycart B. Weighted Kolmogorov Smirnov testing: an alternative for Gene Set Enrichment Analysis. *Stat Appl Genet Mol Biol.* 2015; 14:279–93. doi: 10.1515/sagmb-2014-0077.
28. Xue TM, Tao LD, Zhang M, Xu GC, Zhang J, Zhang PJ. miR-20b overexpression is predictive of poor prognosis in gastric cancer. *Onco Targets Ther.* 2015; 8:1871–6. doi: 10.2147/OTT.S85236.
29. Xue TM, Tao LD, Zhang M, Zhang J, Liu X, Chen GF, Zhu YJ, Zhang PJ. Clinicopathological Significance of MicroRNA-20b Expression in Hepatocellular Carcinoma and Regulation of HIF-1 $\alpha$  and VEGF Effect on Cell Biological Behaviour. *Dis Markers.* 2015; 2015:325176. doi: 10.1155/2015/325176.
30. Ahmad A, Ginnebaugh KR, Sethi S, Chen W, Ali R, Mittal S, Sarkar FH. miR-20b is up-regulated in brain metastases from primary breast cancers. *Oncotarget.* 2015; 6:12188–95. doi: 10.18632/oncotarget.3664.
31. Lee SM, Wu B, Kersey JH. Likelihood-Based Approach to Gene Set Enrichment Analysis with a Finite Mixture Model. *Stat Biosci.* 2014; 6:38–54. doi: 10.1007/s12561-012-9076-3.
32. Hung JH, Yang TH, Hu Z, Weng Z, DeLisi C. Gene set enrichment analysis: performance evaluation and usage guidelines. *Brief Bioinform.* 2012; 13:281–91. doi: 10.1093/bib/bbr049.

33. He K, Wang Q, Yang Y, Wang M, Pan Y. A Comparative Study of Mouse Hepatic and Intestinal Gene Expression Profiles under PPARalpha Knockout by Gene Set Enrichment Analysis. *PPAR Res.* 2011; 2011:629728. doi: 10.1155/2011/629728.
34. Ning QY, Wu JZ, Zang N, Liang J, Hu YL, Mo ZN. Key pathways involved in prostate cancer based on gene set enrichment analysis and meta analysis. *Genet Mol Res.* 2011; 10:3856–87. doi: 10.4238/2011.December.14.10.
35. Gunji Y, Suzuki T, Hori S, Hayashi H, Matsubara H, Shimada H, Ochiai T. Prognostic significance of the number of metastatic lymph nodes in early gastric cancer. *Dig Surg.* 2003; 20:148–53. doi: 69392.
36. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010; 26: 139–40. doi: 10.1093/bioinformatics/btp616.
37. Bair E, Tibshirani R. Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biol.* 2004; 2:E108. doi: 10.1371/journal.pbio.0020108.
38. Shi H, Chen J, Li Y, Li G, Zhong R, Du D, Meng R, Kong W, Lu M. Identification of a six microRNA signature as a novel potential prognostic biomarker in patients with head and neck squamous cell carcinoma. *Oncotarget.* 2016; 7:21579–90. doi: 10.18632/oncotarget.7781.
39. Li X, Shi Y, Yin Z, Xue X, Zhou B. An eight-miRNA signature as a potential biomarker for predicting survival in lung adenocarcinoma. *J Transl Med.* 2014; 12:159. doi: 10.1186/1479-5876-12-159.
40. Backes C, Keller A, Kuentzer J, Kneissl B, Comtesse N, Elnakady YA, Muller R, Meese E, Lenhof HP. GeneTrail—advanced gene set enrichment analysis. *Nucleic Acids Res.* 2007; 35: W186-92. doi: 10.1093/nar/gkm323.