A serum piRNA signature as promising non-invasive diagnostic and prognostic biomarkers for colorectal cancer

This article was published in the following Dove Press journal:
Cancer Management and Research

Ailin Qu1
Wenfei Wang2,3
Yongmei Yang4
Xin Zhang4
Yuhuan Dong4
Guixi Zheng4
Qiuyan Wu4
Mingjin Zou4
Lutao Du1
Yunshan Wang1
Chuanxin Wang1

1Department of Clinical Laboratory, The Second Hospital of Shandong University, Jinan, Shandong 250033, People’s Republic of China; 2Humanistic Medicine Research Center, Qilu Hospital, Shandong University, Jinan, Shandong 250012, People’s Republic of China; 3Humanistic Medicine Research Center, Shandong University, Jinan, Shandong 250012, People’s Republic of China; 4Department of Clinical Laboratory, Qilu Hospital, Shandong University, Jinan, Shandong 250012, People’s Republic of China

Correspondence: Chuanxin Wang
Department of Clinical Laboratory, The Second Hospital of Shandong University, 247 Beiyuan Road, Jinan, Shandong 250033, People’s Republic of China
Tel +86 531 8587 5002
Fax +86 531 8896 2544
Email cxwang@sdu.edu.cn

Purpose: Piwi-interacting RNAs (piRNAs) are a novel class of small non-coding RNAs, which are not easily degraded but detectable in human body fluids. Recent studies have shown that aberrant piRNA expression is a signature feature across multiple tumor types. However, the expressions of piRNAs in serum of tumor patients and their potential clinical values remain largely unclear.

Patients and methods: High-throughput sequencing was performed to investigate the serum piRNA profiles, followed by evaluations in serum samples of 220 colorectal cancer (CRC) patients and 220 healthy controls using reverse transcription quantitative real-time PCR (RT-qPCR). Biomarker panels including piRNA-based Panel I and carcinoembryonic antigen (CEA)-based Panel II, were developed by logistic regression model, and their diagnostic potentials were compared. Fagan’s nomogram was plotted to promote clinical application.

Results: We identified five differentially expressed serum piRNAs (piR-001311, piR-004153, piR-017723, piR-017724 and piR-020365), which, when combined in the piRNA-based Panel I, outperformed the CEA-based Panel II (P<0.001) and could detect CRC with an area under the receiver operating characteristic curve of 0.867. In addition, Kaplan–Meier analysis showed that patients with low serum piR-017724 level had worse overall survival (OS) and progression-free survival (PFS). In multivariate Cox regression analysis, serum piR-017724 was an independent prognostic factor for OS and PFS (P<0.05).

Conclusion: Our findings suggest serum piRNA expression signatures have potential for use as biomarkers for CRC detection and to predict prognosis at the time of diagnosis.

Keywords: serum piRNA, colorectal cancer, diagnosis, prognosis, nomogram

Introduction
Currently, colorectal cancer (CRC) is the third most commonly diagnosed cancer and accounts for the second leading cause of cancer death globally.1 CRC survival depends mostly on the stage of disease at diagnosis, typically ranging from a 5-year survival rate of 90% for patients diagnosed at the localized stage to 10% for cancer detected at regional or distant stage.2 Early detection is crucial for effective treatment of CRC and patient survival. Currently available modalities for CRC screening include fecal occult blood test (FOBT), stool DNA test, digital rectal examination, flexible sigmoidoscopy and colonoscopy. However, these conventional tests have several disadvantages, such as low sensitivity and specificity, high cost and discomfort to the patients.3 Therefore, it is urgently necessary to develop sensitive, specific, cost-effective and non-invasive methods to complement and improve current CRC screening strategies.
Piwi-interacting RNAs (piRNAs) are a novel class of non-coding RNAs (ncRNAs), which have been of interest primarily due to their essential role in the maintenance of germ cells. Only recently, have researchers discovered their presence in mammalian somatic cells, including human cancers. It has been demonstrated that piRNAs are expressed across various tissue types in a cancer-specific manner. Deregulated piRNAs have been detected in kidney, thyroid, uterine, bladder, breast, lung, gastric and colon cancer tissues. The aberrant piRNA expression is a signature feature with potential diagnostic or prognostic implications across multiple cancer types. Besides, recent studies have shown that some piRNAs are involved in carcinogenesis and metastasis of several types of cancers, including CRC, providing compelling evidence for the role of these piRNAs in cancer development. Notably, piRNAs have been reported to be present in human body fluids, including the serum and plasma of healthy individuals, in a markedly stable form. And most recently, Petra et al have found that circulating piR-5937 and piR-28876 are promising diagnostic biomarkers of colon cancer. In light of these studies, we hypothesized that detection of circulating piRNAs might provide a non-invasive avenue for diagnosis and prognosis of cancers. Nonetheless, until now, the expressions of circulating piRNAs in CRC patients and their potential clinical utilities remain largely unexplored.

In the present study, we systematically and comprehensively screened piRNA profiles in the serum of CRC patients and healthy controls. The diagnostic value of the markedly dysregulated piRNAs in the serum as an additional indicator for CRC was evaluated and compared with that of currently widely used biomarkers. Moreover, the correlation between the expression level of serum piRNAs identified and the survival rate of CRC patients was examined, to evaluate their potential for prognostic prediction.

**Material and Methods**

**Ethical statement**

All procedures performed in the study involving human participants were in accordance with the ethical standards of the Clinical Research Ethics Committee of Qilu Hospital, Shandong University and the Declaration of Helsinki. Also, this study was approved by Clinical Research Ethics Committee of Qilu Hospital and the experiments were undertaken with the understanding and written consent of each subject.

**Study design**

The study design consisted of three phases (Figure 1). In the screening phase, the serum samples were pooled from 10 patients with TNM I-II stage (CRC Group A), 10 patients with TNM III-IV stage (CRC Group B) and 10 healthy controls. The samples were then subjected to sequencing to identify piRNAs that were significantly differentially expressed. In the training phase, firstly tested the expression of candidate piRNAs in an independent cohort from 120 CRC patients and 120 healthy controls. The diagnostic potential of the piRNA-based panel with CEA and CA19-9 was evaluated. Moreover, the correlation between the expression level of serum piRNAs identified and the survival rate of CRC patients was examined, to evaluate their potential for prognostic prediction.

**Figure 1** The flowchart of our study design.

**Abbreviations:** CRC, colorectal cancer; piRNA, Piwi-interacting RNA; RT-qPCR, quantitative real-time PCR.
controls, and then constructed piRNA-based and carci-
noembryonicantigen (CEA)-based diagnostic panels.
A comparison between these two panels was per-
formed. In the validation phase, we evaluated the di-
agnostic performance of the piRNA-based panel in
an independent cohort of 100 CRC patients and
100 healthy controls. Table S1 lists the clinical and
demographic characteristics of CRC patients and
healthy controls.

Patients and control subjects
A total of 240 blood samples from CRC patients were
collected before surgery in the Department of General
Surgery of the Qilu Hospital, Shandong University
(Jinan, China). None of the CRC patients enrolled had
ever received any preoperative therapy (radiotherapy,
chemotherapy or chemoradiotherapy) or suffered from
other tumor diseases at the same time. Histological
specimens from all patients were reviewed to confirm
the diagnosis of CRC, and the tumors were staged and
graded according to the 2016 TNM classification and
the 1973 WHO grading system, respectively. In addi-
tion, 230 healthy controls, who visited the hospital for
a medical check-up, were enrolled in our study. The
healthy controls were screened to ensure that they
were within the normal range of all laboratory
findings and had no history of cancer. Moreover, the proportions
of age and gender in selected healthy controls were
similar to those in CRC patients. Furthermore, 20 color-
rectal adenoma (CRA) patients (12 males, 8 females; age
42–71 years, average 59 years) were included in the
study.

The demographic and clinicopathological features of
CRC patients, including sex, age, histological differentia-
tion, tumor location, tumor size, tumor stage and preopera-
tive tumor markers [serum CEA and CA19-9], were
retrospectively collected. The duration of follow-up was
calculated from the date of surgery to death or last follow-
up, and the follow-up was completed in December 2016.
The overall survival (OS) was defined as the period from
time of diagnosis until death. The progression-free survi-
vial (PFS) was calculated from the first day of diagnosis to
the date when disease progression was detected or the date
of death from any cause. The patients were excluded from
analyses of clinicopathological characteristics and prog-
nosis if they had incomplete medical records or inadequate
follow-up.

Sample collection and preparation
Briefly, 5 mL of whole blood samples were collected from
controls and patients by venipuncture. Serum was sepa-
rated within 1 hr of blood collection by sequential cen-
trifugation (2,000×g for 10 mins at 4°C, 12,000×g for
10 mins at 4°C) to completely remove cell debris. The
serum samples were then transferred to RNase/DNase-free
tubes and stored at −80°C until further analysis.

Illumina high-throughput sequencing (HTS) and data analysis
The sequencing procedure was performed as previously
described.16–18 Briefly, total RNA was isolated from serum
using the Trizol LS (Invitrogen, Waltham, MA, USA). The
quantity and integrity of RNA yield was assessed using the
Qubit2.0 (Life Technologies, Waltham, MA, USA) and
Agilent 2200 TapeStation (Agilent Technologies, Santa
Clara, CA, USA) separately. Libraries of small RNAs
were constructed with NEBNext® Multiplex Small RNA
Library Prep Set for Illumina® (Set 1) (NEB, USA) that is
suitable for methylated small RNAs (eg, piRNAs) as well
as unmethylated small RNAs, according to manufacturer’s
instructions. The libraries were sequenced by HiSeq 2500
(Illumina, USA) with single-end 50nt at Ribobio Co. Ltd
(Ribobio, Guangzhou, China).

Raw reads containing adapter or poly 'N' and reads
with low quality and lengths smaller than 17nt were fil-
tered out by FASTQC to get clean reads. Filtered reads
were mapped to reference genome by using BWA. Then,
mapping reads were matched to pirnabank (http://pirna
bank.ibab.ac.in/) to identify piRNAs by BLAST. The
expression levels of piRNAs were calculated by edger
algorithm. Differently expressed piRNAs were chosen
according to the criteria of |log2 (Fold Change) ≥1 and
P<0.05. The entire procedure was done in Ribobio.

Quantification of piRNAs by reverse
transcription quantitative real-time PCR
(Rt-qPCR) assays
Briefly, 2 μg of total RNA was reversely transcribed into cDNA
using Bulge-Loop piRNA-specific RT primer (Ribobio) and
AMV reverse transcriptase (Takara Bio Inc., Shiga, Japan).
Then, 2 μL cDNA generated from reverse transcription was
used for quantitative PCR analysis that was performed on an
ABI PRISM 7500 Sequence Detection System (Applied
Biosystems, Foster City, CA, USA) using miScript SYBR
Green PCR Kit (Qiagen, Hilden, Germany) and Bulge-Loop

---

**Cancer Management and Research (CMR)** is an international, peer-reviewed, open access journal focused on the medical and clinical aspects of cancer management. The journal covers all aspects of cancer management, from prevention and early detection to treatment and palliative care. Its mission is to provide a platform for the dissemination of high-quality research and clinical studies related to cancer management. CMR is indexed in several major databases and is committed to rapid publication of important research findings. For more information, visit [www.dovepress.com](http://www.dovepress.com).
Statistical analysis
The data distribution of each group was determined by the Kolmogorov–Smirnov test. Continuous variables were presented as the mean ± SD or median (interquartile range) when the values were normally or abnormally distributed, respectively. Statistical differences between the groups were assessed using the Mann–Whitney U-test, unpaired Student’s t-test or a Kruskal–Wallis test as appropriate. Categorical variables were presented as numbers (%) and analyzed using Pearson’s chi-squared test or Fisher’s exact test as appropriate. The Kaplan–Meier method was used to estimate survival, and the survival differences between the subgroups were examined using the log-rank test. A Cox regression model (proportional hazard model) was applied for the univariate analysis and multivariate analysis of prognostic factors. Receiver operating characteristic (ROC) curves were established by MedCalc 9.3.9.0 to discriminate the subjects with or without CRC. Logistic regression analysis was performed using Matlab software (Matlab, R2013a). Fagan’s nomogram was used to help clinicians to calculate a patient’s probability of having CRC according to diagnostic test results. The SPSS (Statistical Package for the Social Sciences) software (version 18.0, Chicago, IL, USA) and R software (version 3.2.3; http://www.Rproject.org) were used to analyze all the other data. P<0.05 was considered as statistically significant.

Evaluation of piRNAs in serum by RT-qPCR
The expressions of the 16 candidate piRNAs were evaluated by RT-qPCR in serum samples from 120 CRC patients and 120 healthy controls in the training phase. To ensure the stability of the reference genes, we firstly compared the expression levels of miR-191-5p, U6 and their combination in CRC patients and healthy controls. Data showed that there was no significant difference in the expressions of miR-191-5p, U6 and combination of miR-191-5p and U6 between CRC patients and healthy controls. Among the 16 piRNAs, the expressions of five piRNAs (piR-001311, piR-004153, piR-017723, piR-017724 and piR-020365) were significantly down-regulated in CRC patients compared with the healthy controls (Table 1, Figure S1A). Among the 16 piRNAs, the expressions of these small RNAs displayed a strong preference for uridine at the 5’ end, consistent with the findings of the bias 5’ uridine of piRNAs as previously reported by others.21

Results
Profiling of serum piRNAs in CRC patients by HTS
Analysis of the length distribution showed that all three serum samples contained a number of small RNAs with sizes that were in line with the common size of piRNAs (26–31 nucleotides). Correspondence analysis showed

primers (Ribobio), based on the previous reported method for piRNA detection.19 The reaction reagents contained 10μL 2XQuantiTect SYBR Green PCR Master Mix, 2μL forward primer of piRNA, 2μL 10XmiScript Universal Primer, 2μL template cDNA. The volume was adjusted with RNA-free H2O. The reaction parameters were incubation at 95°C for 15 mins, then 40 cycles of 95°C for 15 s, 55°C for 30 s and 70°C for 30 s. All reactions were performed in triplicate to remove any outliers. In addition, miR-191-5p and U6 were selected as the reference genes according to our previous study.20 The piRNA expression was normalized to the mean of these two reference genes. The relative expressions of piRNAs were determined using the $2^{-\Delta\Delta Ct}$ method.
Table 1 The selected serum piRNA expression levels in patients with CRC and healthy controls in training set and validation set (median [IQR])

<table>
<thead>
<tr>
<th>piRNA</th>
<th>Training set</th>
<th>Validation set</th>
</tr>
</thead>
<tbody>
<tr>
<td>piR-001311</td>
<td>Controls (n=120): 1.01 (0.72–1.68)</td>
<td>CRC (n=120): 0.55 (0.28–1.13)</td>
</tr>
<tr>
<td>piR-004153</td>
<td>1.11 (0.63–1.63)</td>
<td>0.48 (0.32–0.90)</td>
</tr>
<tr>
<td>piR-017723</td>
<td>1.16 (0.67–1.75)</td>
<td>0.57 (0.30–0.89)</td>
</tr>
<tr>
<td>piR-017724</td>
<td>0.97 (0.67–1.68)</td>
<td>0.54 (0.29–0.95)</td>
</tr>
<tr>
<td>piR-020365</td>
<td>0.92 (0.52–1.67)</td>
<td>0.35 (0.22–0.71)</td>
</tr>
</tbody>
</table>

Abbreviations: CRC, colorectal cancer; piRNA, Piwi-interacting RNA.

Figure 2 ROC curve analysis for the detection of CRC using piR-001311 (A), piR-004153 (B), piR-017723 (C), piR-017724 (D), piR-020365 (E) and the piRNA-based Panel I (F) in the training set.

Abbreviations: CRC, colorectal cancer; piRNA, Piwi-interacting RNA; ROC, receiver operating characteristic.

Confirmation of identified piRNAs stability in serum

In order to explore the feasibility of circulating piRNAs as the biomarkers for diagnosis of CRC, we investigated the stability of piR-001311, piR-004153, piR-017723, piR-017724 and piR-020365 in five serum samples by room-temperature incubation and repetitive freeze-thaw cycles. The serum samples were exposed to harsh conditions, including incubation at room temperature for 4, 8 and 24 h, incubation at −80°C for 1, 2 and 3 months, or 1, 2 and 4 repetitive freeze-thaw cycles. The Ct values of these five piRNAs showed no significant variation (Figure S2, all P>0.05), demonstrating that piRNAs were sufficiently stable in serum. Therefore, they could be used for routine processing of clinical samples.

Identification of predictive piRNA panel

A stepwise logistic regression model was constructed to estimate the probability of being diagnosed with CRC in the training set. The predicted risk of being diagnosed with CRC from the logit model based on the five-piRNA-panel (piRNA-
based Panel I) was evaluated via the following equation: Logit (P=CRC)=−1.0391+0.1621×piR-001311+0.203×piR-004153
+0.3711×piR-017723+0.0897×piR-017724+0.128×piR-
020365. We performed ROC curve analysis to assess the diag-
nostic performance of piR-based Panel I, and the results
revealed that the AUC for the five-piRNA panel was 0.867
(95% CI: 0.817–0.907) with a sensitivity of 78.3% and
a specificity of 74.2% (cutoff value = −0.1085; Figure 2 and
Table 2).

Validation of the piRNA panel
The piRNA-based Panel I obtained from the training set
was used to predict the probability of being diagnosed
with CRC in the validation set. Similar to the training
set, the AUC of piRNA-based Panel I was 0.854 (95%
CI: 0.797–0.900, Figure S3). We further pooled CRC
patients from the training set and validation set, and the
combined results showed that the piRNA-based Panel
I could detect CRC with an AUC of 0.862 (95% CI:
0.826–0.893) (Table 2). Interestingly, piRNA-based Panel
I outperformed the individual piRNAs with regard to CRC
detection in all sets (all P<0.05).

In addition, 20 CRA were enrolled in the present
study. Figure 3 shows that the expression of piR-
017723 was significantly decreased in CRA patients
compared with healthy controls.

| Table 2 Performance of piRNA-based panel I and CEA-based panel II in the detection of CRC from healthy controls |
|---------------------------------------------------------------|-----------------|-----------------|-----------------|-----------------|
| **Training set**                                             | **Cutoff value** | **Sensitivity (95% CI)** | **Specificity (95% CI)** | **AUC (95% CI)** |
| Panel I                                                      | −0.1085          | 0.783 (0.699–0.853)     | 0.742 (0.654–0.817)     | 0.867 (0.817–0.907) |
| Panel II                                                     | 0.0685           | 0.500 (0.407–0.583)     | 0.917 (0.852–0.959)     | 0.735 (0.675–0.790) |
| Panel I + Panel II                                          | 0.1110           | 0.800 (0.717–0.867)     | 0.867 (0.793–0.922)     | 0.899 (0.853–0.934) |
| **Validation set**                                          | **Cutoff value** | **Sensitivity (95% CI)** | **Specificity (95% CI)** | **AUC (95% CI)** |
| Panel I                                                      | −0.1085          | 0.780 (0.686–0.857)     | 0.760 (0.664–0.840)     | 0.854 (0.797–0.900) |
| Panel II                                                     | 0.0685           | 0.530 (0.428–0.631)     | 0.890 (0.812–0.944)     | 0.768 (0.704–0.825) |
| Panel I + Panel II                                          | 0.1110           | 0.840 (0.753–0.906)     | 0.850 (0.765–0.914)     | 0.897 (0.846–0.935) |
| **Combination set**                                         | **Cutoff value** | **Sensitivity (95% CI)** | **Specificity (95% CI)** | **AUC (95% CI)** |
| Panel I                                                      | −0.1085          | 0.782 (0.721–0.835)     | 0.750 (0.687–0.806)     | 0.862 (0.826–0.893) |
| Panel II                                                     | 0.0685           | 0.509 (0.441–0.577)     | 0.9045 (0.858–0.940)    | 0.745 (0.701–0.785) |
| Panel I + Panel II                                          | 0.1110           | 0.818 (0.761–0.867)     | 0.859 (0.806–0.902)     | 0.898 (0.866–0.925) |

Abbreviations: AUC, area under the ROC curve; CEA, carcinoembryonic antigen; CRC, colorectal cancer; piRNA, Piwi-interacting RNA; ROC, receiver operating characteristic.
other four piRNAs were not altered in CRA patients compared with healthy controls. Moreover, it seems to be impossible to distinguish CRC from CRA based on the expressions of these piRNAs. The values of piRNA-based Panel I were significantly decreased from the healthy controls through CRA patients to CRC patients ($P<0.05$).

**Comparison of diagnostic potential of the piRNA-based Panel I with CEA and CA19-9**

We further investigated the diagnostic roles of CEA and CA19-9 for CRC detection. In the training set, the AUC of the conventional markers were 0.718 and 0.571, and the ROC curves are illustrated in Figure S4. Using logistic regression model, we constructed CEA-based Panel II (including CEA and CA19-9) and plotted its ROC curve (Figure S4C). Then, we compared the AUC values of piRNA-based Panel I and CEA-based Panel II. The results indicate that Panel I showed significantly higher AUC value than Panel II ($P=0.0003$). When the piRNA-based Panel I was combined with CEA-based Panel II, significant differences were found between the combined model and the piRNA-based Panel I ($P=0.0056$). Analysis in validation set and combination set produced similar results (Figure 4). The results above confirmed the reliable detection ability of piRNA-based Panel I, which was superior to CEA-based Panel II in detecting CRC to a certain extent.

**Construction and assessment of Fagan’s nomogram to calculate the probability of CRC based on piRNA-based panel**

The results of combination set showed piRNA-based Panel I could detect CRC with a sensitivity of 0.782 (95% CI: 0.721–0.835) and a specificity of 0.750 (95% CI: 0.687–0.806). To make this information clinically useful, Fagan’s nomogram was established to help clinicians to use Panel I values to estimate the probability of a patient truly having CRC (Figure 5). We set the pre-test probability of CRC at a hypothetical value of 5%. The positive likelihood ratio (+LR) of Panel I was 3.13 and the negative likelihood ratio (-LR) was 0.29. We mark pre-test probability and LR values on the left and middle axes, respectively. Then, a line is drawn from the two marked points along to the right axis. The point at which the line intersects with the right axis is post-test probability. As shown in Figure 5, if a patient had a positive Panel I result, the post-test probability that he truly has CRC would rise to 14.1% (red line). Alternatively, if the patient had a negative result, the post-test probability would decrease to 1.5% (blue line). These data indicate that our five-piRNA based Panel I could provide useful information to clinicians.

**Correlation of piRNA expression with survival in CRC patients**

We assessed the association of five piRNAs with patients’ survival. A total of 110 patients with complete medical
records and adequate follow-up from both training and validation sets were included in the analysis. Kaplan–Meier survival analysis showed that patients with low expression of piR-017724 had significantly lower OS and PFS rate compared with those with high expression of piR-017724 ($P=0.005$ and $P=0.002$, respectively, Figure 6). Cox proportional hazards model showed that piR-017724 expression was an independent prognostic factor for OS and PFS in CRC patients ($P=0.003$ and $P=0.004$, respectively). The statistical values of piRNAs expression and the other
Clinicopathological parameters derived from the Cox proportional hazards regression model were listed in Table 3.

**Discussion**

Recently, the emergence of circulating ncRNAs as biomarkers has provided a new opportunity to facilitate non-invasive diagnosis and prognosis of cancers. In the present study, we, for the first time, demonstrated a potentially clinically applicable serum piRNA panel that may aid in the development of improved diagnosis and outcome prediction.

The Piwi-piRNA pathway was originally discovered in the Drosophila male germline. It is also active in various tumorous cell lines, implying that nonaging somatic cancer stem cells may adopt certain germline-specific characteristics (such as an unlimited proliferation capacity) via the piRNA pathway. Piwi-piRNA pathway has been implicated in tumorigenesis via DNA methylation of gene promoter regions, particularly of putative tumor-suppressor genes. Additionally, the piRNA expression can delineate clinical features of disease progression. For being mostly only 26–31 nucleotide (nt) in length, piRNAs are not easily degraded and can pass through cell membrane easily. This may allow circulating piRNA signatures to be identified in malignant diseases. A new study has revealed that circulating piRNAs are significantly deregulated between colon cancer patients and healthy controls. In the present study, we consistently found that piRNAs were differentially expressed in serum from CRC patients compared with controls, and for the first time ever, we constructed a five-piRNA-panel (including piR-001311, piR-004153, piR-017723, piR-017724 and piR-020365) for CRC diagnosis.

Conventionally, CEA and CA19-9 have been used to detect and monitor cancer relapse and prognosis of CRC patients. However, these conventional biomarkers suffer from well-documented disadvantages of limited sensitivity and specificity, especially in case of a single biomarker. In our present study, the combined use of these five deregulated piRNAs yielded better diagnostic performance than the use of individual piRNAs. Also, the five-piRNA based Panel I exhibited satisfactory diagnostic performance, showing higher AUC compared with FOBT. Furthermore, the diagnostic potential of this five-piRNA based Panel I was better than that of CEA-CA19-9 based Panel II. Since the AUC could only provide global estimation of diagnostic performance, we constructed a Fagan’s nomogram to facilitate the interpretation of test results into clinically useful information. With a hypothetical pre-test probability of 5%, the post-test probability of CRC would rise to 14.1% in condition of positive Panel I result, and decrease to 1.5% in condition of negative result.

In addition, we found the scores of the five-piRNA-panel were progressively decreased from the healthy controls through CRA patients to CRC patients. Since adenoma represents a precursor stage of CRC, we surmised that decreased panel score in adenomas might be an early indicator in the progression of cancer. Thus, the panel might help identify people with an increased probability of developing CRC in population with familial adenomatous polyposis.

Similar to miRNAs, piRNAs are easier to detect than lncRNA and stable in body fluids. Moreover, piRNAs remain largely undegraded in circulation and are capable of dynamically adapting to a wide range of laboratory conditions. Considering that stability is crucial for tumor marker, we demonstrated that serum piRNAs remained stable regardless of long-term incubation at room temperature or repetitive freeze–thawing. One possible explanation for such stability can be attributed to that piRNAs are packaged inside exosomes and secreted from cells. However, Freedman and colleagues have found that it may not be the case. An alternative explanation is that piRNAs may form RNA-protein complexes to maintain stable in the same manner as cell-free miRNAs. In this study, the piRNAs were robustly and stably expressed in human blood serum, showing lower Ct values compared with the previously detected miRNAs. Therefore, circulating piRNAs possessed the potential as a novel class of blood-based biomarkers for cancer patients.

It is crucial to select suitable reference genes in piRNA research since normalization to inappropriate endogenous control may lead to the identification of incorrect piRNAs of interest. We selected miR-191 and U6, which were used to normalize miRNA expression, as the internal reference genes because they have the advantages of stability and high abundance. In this study, we confirmed that there was no significant difference in the mean of miR-191-5p and U6 between CRC patients and healthy controls, indicating that they were also suitable reference genes for piRNA study. Based on this, differentially expressed serum piRNAs were identified in CRC patients by RT-qPCR.

Biomarkers that can identify CRC patients with high risk and poor prognosis are one of the most pressing topics...
**Table 3** Cox proportional hazards regression model analysis for OS and PFS in CRC patients

<table>
<thead>
<tr>
<th>Clinicopathological feature</th>
<th>OS</th>
<th></th>
<th>PFS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Univariate analysis</td>
<td>Multivariate analysis</td>
<td>Univariate analysis</td>
<td>Multivariate analysis</td>
</tr>
<tr>
<td></td>
<td>HR 95% CI</td>
<td>P-value</td>
<td>HR 95% CI</td>
<td>P-value</td>
</tr>
<tr>
<td>Sex</td>
<td>0.639</td>
<td>0.292–1.395</td>
<td>0.261</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.792</td>
<td>0.356–1.764</td>
<td>0.569</td>
<td></td>
</tr>
<tr>
<td>Tumor location</td>
<td>1.013</td>
<td>0.455–2.256</td>
<td>0.974</td>
<td></td>
</tr>
<tr>
<td>Differentiation</td>
<td>1.775</td>
<td>0.564–5.584</td>
<td>0.326</td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>0.899</td>
<td>0.417–1.937</td>
<td>0.785</td>
<td></td>
</tr>
<tr>
<td>Local invasion</td>
<td>2.782</td>
<td>0.657–11.776</td>
<td>0.165</td>
<td></td>
</tr>
<tr>
<td>Lymph node metastasis</td>
<td>1.934</td>
<td>1.015–3.685</td>
<td>0.045</td>
<td>1.675 0.862–3.256 0.128</td>
</tr>
<tr>
<td>TNM stage</td>
<td>2.077</td>
<td>1.127–3.830</td>
<td>0.019</td>
<td>2.477 1.405–4.366 0.002</td>
</tr>
<tr>
<td>piR-001311</td>
<td>0.626</td>
<td>0.290–1.349</td>
<td>0.232</td>
<td>2.469 1.408–4.329 0.002</td>
</tr>
<tr>
<td>piR-004153</td>
<td>1.048</td>
<td>0.493–2.230</td>
<td>0.903</td>
<td></td>
</tr>
<tr>
<td>piR-017723</td>
<td>0.769</td>
<td>0.360–1.642</td>
<td>0.497</td>
<td></td>
</tr>
<tr>
<td>piR-017724</td>
<td>0.231</td>
<td>0.093–0.572</td>
<td>0.002</td>
<td>0.250 0.100–0.624 0.003</td>
</tr>
<tr>
<td>piR-019752</td>
<td>0.611</td>
<td>0.284–1.318</td>
<td>0.209</td>
<td>0.624 0.361–1.079 0.091</td>
</tr>
</tbody>
</table>

Abbreviations: AUC, area under ROC curve; CRC, colorectal cancer; OS, overall survival; PFS, progression-free survival; ROC, receiver operating characteristic.
needs for clinicians. piRNAs have potential functions in transposon silencing as well as epigenetic and post-transcriptional regulation of gene expression. Although only a few studies suggest a role for piRNAs in carcinogenesis and progression of cancers, this newly emerging research topic will surely grow immensely in the near future. Previously, a three-piRNA-panel has been identified in gastric cancer tissue to effectively separate cancer patients by risk of recurrence, which also showed a similar association with recurrence-free survival has been observed in patients with colon cancer. In the present study, we investigated the roles of these five identified piRNAs as prognostic biomarkers. Our data demonstrated that the reduced expression of serum piR-017724 was correlated with poor OS and PFS rate. Cox regression model analysis showed that serum piR-017724 was an independent prognostic factor for CRC, suggesting its potential as biomarker for CRC prognosis. Another interesting finding was that piR-020365 for tumor detection and meanwhile, serum piR-017724 was identified as an independent prognosis predictor in patients with CRC.

Conclusion
This study revealed a novel serum five-piRNA-panel (piR-001311, piR-004153, piR-017723, piR-017724 and piR-020365) for tumor detection and meanwhile, serum piR-017724 was identified as an independent prognosis predictor in patients with CRC.

Acknowledgments
This study was funded by the National Natural Science Foundation of China (Nos 81472025, 81772271, 81601846, 81702084, 81301506, 81874040, and 81501819), Shandong Technological Development Project (2016CYS01A02 and 2016GSF201120), Taishan Scholar Program of Shandong Province, National Science of Basic Scientific Research Foundation of Shandong University (2017BTS01), Science Foundation of Qilu Hospital of Shandong University (2015QLM5S1 and 2015QLMS45), Fundamental Research Funds of Shandong University (2014QIKY03), and Key Research and Development Projects of Shandong Province (2015GSF118167).

Disclosure
The authors report no conflicts of interest in this work.

References


Supplementary materials

Figure S1 The expressions of reference genes and piRNAs between CRC patients and healthy controls in the training set. (A) Ct values of miR-191-5p, U6 and combination of miR-191-5p and U6 in CRC patients and healthy controls (all \( P > 0.05 \)). No significant difference was found within three reference genes in both groups. (B–F) The expression levels of piR-001311 (B), piR-004153 (C), piR-017723 (D), piR-017724 (E) and piR-020365 (F) were determined by RT-qPCR in 120 CRC patients and 120 healthy control.

**Abbreviations:** CRC, colorectal cancer; piRNA, Piwi-interacting RNA.
Figure S2  Stability of serum piRNA expressions. Serum levels of piR-001311 (A), piR-004153 (B), piR-017723 (C), piR-017724 (D) and piR-020365 (E) remained stable when treated with prolonged incubation at room-temperature, repetitive freeze-thaw cycles, or prolonged incubation at –80°C.

Abbreviation: piRNA, Piwi-interacting RNA.
**Figure S3** ROC curve analysis for the detection of CRC using piR-001311 (A), piR-004153 (B), piR-017723 (C), piR-017724 (D), piR-020365 (E) and the piRNA-based panel (F) in the validation set.

**Abbreviations:** AUC, area under the ROC curve; CRC, colorectal cancer; piRNA, Piwi-interacting RNA; ROC, receiver operating characteristic.

**Figure S4** ROC curve analysis for the diagnostic performance of CEA (A), CA19-9 (B) and CEA-based panel II (C) in the training set.

**Abbreviations:** AUC, area under the ROC curve; CEA, carcinoembryonic antigen; ROC, receiver operating characteristic.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Training set</th>
<th>Validation set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CRC n=120</td>
<td>Controls n=120</td>
</tr>
<tr>
<td>Average age (range)</td>
<td>61 (41-70)</td>
<td>59 (38-73)</td>
</tr>
<tr>
<td>Sex</td>
<td>male 69 (57%)</td>
<td>male 66 (55%)</td>
</tr>
<tr>
<td></td>
<td>female 51 (43%)</td>
<td>female 54 (45%)</td>
</tr>
<tr>
<td>Tumor location:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colon</td>
<td>48 (40%)</td>
<td>31 (31%)</td>
</tr>
<tr>
<td>Rectum</td>
<td>72 (60%)</td>
<td>67 (67%)</td>
</tr>
<tr>
<td>TNM</td>
<td>I 31 (26%)</td>
<td>18 (18%)</td>
</tr>
<tr>
<td></td>
<td>II 39 (33%)</td>
<td>38 (38%)</td>
</tr>
<tr>
<td></td>
<td>III 40 (33%)</td>
<td>34 (34%)</td>
</tr>
<tr>
<td></td>
<td>IV 10 (8%)</td>
<td>10 (10%)</td>
</tr>
<tr>
<td>T stage</td>
<td>T1-T2 34 (28%)</td>
<td>T1-T2 18 (18%)</td>
</tr>
<tr>
<td></td>
<td>T3-T4 86 (72%)</td>
<td>T3-T4 82 (82%)</td>
</tr>
<tr>
<td>Lymph node metastasis</td>
<td>Negative 72 (60%)</td>
<td>Negative 58 (58%)</td>
</tr>
<tr>
<td></td>
<td>Positive 48 (40%)</td>
<td>Positive 42 (42%)</td>
</tr>
</tbody>
</table>

Abbreviation: CRC, colorectal cancer.
Table S2 Selected piRNAs that were significantly downregulated in serum samples of CRC patients compared with healthy controls.
The values showing in the columns referred to RPM value.

<table>
<thead>
<tr>
<th>piRNA-ID</th>
<th>H</th>
<th>CRC Group A</th>
<th>CRC Group B</th>
<th>A/H</th>
<th>B/H</th>
</tr>
</thead>
<tbody>
<tr>
<td>piR-000765</td>
<td>8.95</td>
<td>2.23</td>
<td>1.51</td>
<td>0.25</td>
<td>0.17</td>
</tr>
<tr>
<td>piR-001311</td>
<td>1064.56</td>
<td>379.34</td>
<td>133.72</td>
<td>0.36</td>
<td>0.13</td>
</tr>
<tr>
<td>piR-001312</td>
<td>2325.22</td>
<td>841.15</td>
<td>417.92</td>
<td>0.36</td>
<td>0.18</td>
</tr>
<tr>
<td>piR-004150</td>
<td>198.35</td>
<td>63.17</td>
<td>28.46</td>
<td>0.32</td>
<td>0.14</td>
</tr>
<tr>
<td>piR-004153</td>
<td>5111.56</td>
<td>276.36</td>
<td>266.94</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>piR-009295</td>
<td>25516.63</td>
<td>7396.88</td>
<td>5951.08</td>
<td>0.29</td>
<td>0.23</td>
</tr>
<tr>
<td>piR-014620</td>
<td>1.25</td>
<td>5.30</td>
<td>5.95</td>
<td>4.26</td>
<td>4.78</td>
</tr>
<tr>
<td>piR-016677</td>
<td>4940.14</td>
<td>403.55</td>
<td>337.38</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>piR-017716</td>
<td>66490.41</td>
<td>5832.94</td>
<td>4198.87</td>
<td>0.09</td>
<td>0.06</td>
</tr>
<tr>
<td>piR-017723</td>
<td>10.99</td>
<td>24.52</td>
<td>35.22</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>piR-017724</td>
<td>166.13</td>
<td>224.52</td>
<td>324.67</td>
<td>0.15</td>
<td>0.21</td>
</tr>
<tr>
<td>piR-018569</td>
<td>1773.60</td>
<td>673.92</td>
<td>324.67</td>
<td>0.38</td>
<td>0.18</td>
</tr>
<tr>
<td>piR-020326</td>
<td>304616.11</td>
<td>102410.79</td>
<td>67569.38</td>
<td>0.34</td>
<td>0.22</td>
</tr>
<tr>
<td>piR-020365</td>
<td>1297.15</td>
<td>518.91</td>
<td>249.98</td>
<td>0.40</td>
<td>0.19</td>
</tr>
<tr>
<td>piR-020388</td>
<td>49.72</td>
<td>5.46</td>
<td>3.94</td>
<td>0.11</td>
<td>0.08</td>
</tr>
<tr>
<td>piR-020829</td>
<td>21.09</td>
<td>6.23</td>
<td>5.05</td>
<td>0.30</td>
<td>0.24</td>
</tr>
</tbody>
</table>

**Abbreviations:** A/H; CRC Group A/ healthy controls; B/H, CRC Group B/ healthy controls; CRC, colorectal cancer; H, healthy controls; RPM, the number of reads per million clean tags; piRNA, Piwi-interacting RNA.
### Table S3 Diagnostic performance of individual biomarkers in the training, validation and combination sets

<table>
<thead>
<tr>
<th></th>
<th>Cut-off</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>AUC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training set</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>piR-001311</td>
<td>0.592</td>
<td>0.575 (0.481-0.665)</td>
<td>0.842 (0.764-0.902)</td>
<td>0.726 (0.665-0.782)</td>
</tr>
<tr>
<td>piR-004153</td>
<td>0.483</td>
<td>0.508 (0.416-0.601)</td>
<td>0.892 (0.822-0.941)</td>
<td>0.728 (0.668-0.784)</td>
</tr>
<tr>
<td>piR017723</td>
<td>0.732</td>
<td>0.667 (0.575-0.750)</td>
<td>0.733 (0.645-0.810)</td>
<td>0.756 (0.697-0.809)</td>
</tr>
<tr>
<td>piR-017724</td>
<td>0.591</td>
<td>0.575 (0.481-0.665)</td>
<td>0.817 (0.736-0.881)</td>
<td>0.749 (0.690-0.803)</td>
</tr>
<tr>
<td>CEA</td>
<td>5</td>
<td>0.417 (0.327-0.510)</td>
<td>0.958 (0.905-0.986)</td>
<td>0.718 (0.656-0.774)</td>
</tr>
<tr>
<td>CA19-9</td>
<td>39</td>
<td>0.192 (0.126-0.274)</td>
<td>0.958 (0.905-0.986)</td>
<td>0.571 (0.506-0.634)</td>
</tr>
<tr>
<td><strong>Validation set</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>piR-001311</td>
<td>0.592</td>
<td>0.510 (0.408-0.611)</td>
<td>0.850 (0.765-0.914)</td>
<td>0.698 (0.629-0.761)</td>
</tr>
<tr>
<td>piR-004153</td>
<td>0.483</td>
<td>0.210 (0.135-0.303)</td>
<td>0.890 (0.812-0.944)</td>
<td>0.707 (0.639-0.769)</td>
</tr>
<tr>
<td>piR017723</td>
<td>0.732</td>
<td>0.580 (0.477-0.678)</td>
<td>0.760 (0.664-0.840)</td>
<td>0.719 (0.651-0.780)</td>
</tr>
<tr>
<td>piR-017724</td>
<td>0.591</td>
<td>0.510 (0.408-0.611)</td>
<td>0.870 (0.788-0.929)</td>
<td>0.751 (0.685-0.809)</td>
</tr>
<tr>
<td>CEA</td>
<td>5</td>
<td>0.42 (0.322-0.523)</td>
<td>0.93 (0.861-0.971)</td>
<td>0.709 (0.640-0.771)</td>
</tr>
<tr>
<td>CA19-9</td>
<td>39</td>
<td>0.19 (0.118-0.281)</td>
<td>0.97 (0.915-0.994)</td>
<td>0.577 (0.506-0.647)</td>
</tr>
<tr>
<td><strong>Combination set</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>piR-001311</td>
<td>0.592</td>
<td>0.546 (0.477-0.613)</td>
<td>0.846 (0.791-0.891)</td>
<td>0.711 (0.666-0.753)</td>
</tr>
<tr>
<td>piR-004153</td>
<td>0.483</td>
<td>0.382 (0.317-0.45)</td>
<td>0.891 (0.842-0.929)</td>
<td>0.715 (0.671-0.757)</td>
</tr>
<tr>
<td>piR017723</td>
<td>0.732</td>
<td>0.628 (0.56-0.691)</td>
<td>0.746 (0.683-0.802)</td>
<td>0.739 (0.695-0.779)</td>
</tr>
<tr>
<td>piR-017724</td>
<td>0.591</td>
<td>0.546 (0.477-0.613)</td>
<td>0.846 (0.791-0.891)</td>
<td>0.746 (0.703-0.786)</td>
</tr>
<tr>
<td>piR-020365</td>
<td>0.475</td>
<td>0.468 (0.401-0.536)</td>
<td>0.873 (0.821-0.914)</td>
<td>0.766 (0.724-0.805)</td>
</tr>
<tr>
<td>CEA</td>
<td>5</td>
<td>0.418 (0.352-0.486)</td>
<td>0.946 (0.907-0.972)</td>
<td>0.707 (0.662-0.749)</td>
</tr>
<tr>
<td>CA19-9</td>
<td>39</td>
<td>0.191 (0.141-0.249)</td>
<td>0.964 (0.930-0.984)</td>
<td>0.574 (0.526-0.620)</td>
</tr>
</tbody>
</table>

**Abbreviations:** AUC, area under the ROC curve; CEA, carcinoembryonic antigen.