Tsinghua Science and Technology

Volume 23 | Issue 2 Article 3

2018

Prediction of Cancer-Associated piRNA–mRNA and piRNA–lncRNA Interactions by Integrated Analysis of Expression and Sequence Data

Yajun Liu
Xi'an University of Technology, Xi'an 710048, China. the School of Computer Science and Technology, Xidian University, Xi'an 710071, China.

Junying Zhang
the School of Computer Science and Technology, Xidian University, Xi'an 710071, China.

Aimin Li
Xi'an University of Technology, Xi'an 710048, China.

Zhaowen Liu
the School of Computer Science and Technology, Xidian University, Xi'an 710071, China.

Zhongzhen He
the School of Computer Science and Technology, Xidian University, Xi'an 710071, China.

See next page for additional authors

Recommended Citation

Yajun Liu, Junying Zhang, Aimin Li et al. Prediction of Cancer-Associated piRNA–mRNA and piRNA–lncRNA Interactions by Integrated Analysis of Expression and Sequence Data. Tsinghua Science and Technology 2018, 23(2): 115-125.

Prediction of Cancer-Associated piRNA–mRNA and piRNA–lncRNA Interactions by Integrated Analysis of Expression and Sequence Data

Authors

Yajun Liu, Junying Zhang, Aimin Li, Zhaowen Liu, Zhongzhen He, Xiguo Yuan, and Shouheng Tuo

TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 01/09 pp115–125
DOI: 10.26599 / TST.2018.9010056
Volume 23, Number 2, April 2018

Prediction of Cancer-Associated piRNA–mRNA and piRNA–lncRNA Interactions by Integrated Analysis of Expression and Sequence Data

Yajun Liu, Junying Zhang*, Aimin Li, Zhaowen Liu, Zhongzhen He, Xiguo Yuan, and Shouheng Tuo

Abstract: piwi-interacting RNAs (piRNAs) are valuable biomarkers, but functional studies are still very limited. Recent research shows that piRNA-mediated cleavage acts on Transposable Elements (TEs), messenger RNAs (mRNAs), and long non-coding RNAs (lncRNAs). This study aimed to predict cancer-associated piRNA-mRNA and piRNA-lncRNA interactions as well as piRNA regulatory functions. Four cancer types (BRCA, HNSC, KIRC, and LUAD) were investigated. Interactions were identified by integrated analysis of the expression and sequence data. For the expression analysis, only piRNA–mRNA and piRNA–lncRNA pairs with expression profiles that were significantly inversely correlated were retained to reduce false-positive rates during the prediction. For the sequence analysis, miRanda was used for the target prediction. We identified 198 piRNA–mRNA and 10 piRNA–lncRNA pairs. Unlike mRNA and lncRNA expressions, the piRNA expression was relatively consistent across the cancer types. Furthermore, the identified piRNAs were consistent with previously published cancer biomarkers, such as piRNA-36741, piR-21032, and piRNA-57125. More importantly, predicted piRNA functions were determined by constructing an interaction network, and piRNA targets were placed in gene ontology categories related to the cancer hallmarks “activating invasion and metastasis” and “sustained angiogenesis”.

Key words: co-expression; piRNA-mRNA interaction; piRNA-lncRNA interaction; integrated analysis; target prediction

1 Introduction

Non-coding RNAs (ncRNAs) can be grouped into two major classes based on size. Piwi-interacting RNAs (piRNAs) comprise the largest class of small ncRNAs and are expressed in animal cells[1]. They are primarily expressed in the germline, specifically with Piwi-class proteins[2]. Similar to micro-RNAs (miRNAs), the 5’ ends of piRNAs also contain a strong uracil bias[2]. Based on their origins, piRNAs can be divided into three groups: transposon-derived piRNAs[3], mRNA-derived piRNAs, and lncRNA-derived piRNAs[4].

1.1 piRNAs in human cancers

Although piRNAs are generally thought to be expressed in the germline[5], recent research indicates that piRNAs can be found in somatic[6] and cancer cells[7]. Cancer and germ cells share several characteristics, such as rapid proliferation. Therefore, it is not surprising that germline factors would also be involved in oncogenesis[8]. With developments in cancer research, more studies have highlighted the role of piRNAs in cancer, but functional

- Yajun Liu and Aimin Li are with Xi’an University of Technology, Xi’an 710048, China. E-mail: liuyajun1007@163.com; liaiminmail@gmail.com.
- Yajun Liu, Junying Zhang, Zhaowen Liu, Zhongzhen He, Xiguo Yuan, and Shouheng Tuo are with the School of Computer Science and Technology, Xidian University, Xi’an 710071, China. E-mail: jyzhang@mail.xidian.edu.cn; liukaka928@126.com; hzz2007.good@163.com; xiguoyuan@ mail.xidian.edu.cn; tuo_sh@126.com.

* To whom correspondence should be addressed.
Manuscript received: 2016-11-07; revised: 2017-03-07; accepted: 2017-03-10

studies remain limited.

Cheng et al.[9] showed that the expression of piR-651 in gastric, colon, lung, and breast cancer tissues was higher than that in paired noncancerous tissues. Their findings suggest that piR-651 might be involved in the development of gastric and other cancers and may be a potential marker for cancer diagnosis. Cheng et al.[9] also found that the expression level of piR-823 in gastric cancer tissues was significantly lower than that in noncancerous tissues. They observed that after increasing the level of piR-823 in gastric cancer cells, cell growth was inhibited, suggesting that piR-823 is a putative biomarker for gastric cancer diagnosis.

Similarly, Cui et al.[10] observed that the levels of piR-651 and piR-823 in peripheral blood from patients with gastric cancer were significantly lower than those in healthy controls. Moreover, they demonstrated that piR-823 levels were positively associated with the tumor-node-metastasis stage and distant metastasis.

Related studies have also revealed similar correlations in recent years, but are limited to a single piRNA or a few expression profiles. Martinez et al.[7] analyzed 6260 human piRNA transcriptomes derived from non-malignant and tumor tissues from 11 organs. They discovered that a total of 522 piRNAs were expressed in all tumor tissues. This public piRNA expression data could be used to support future piRNA functional studies.

1.2 piRNA co-expression analysis

One of the main tasks in piRNA research is functional analysis. Genes typically have dependencies on each other. Martinez et al.[7] analyzed correlations in piRNA expression patterns across samples within the same tumor type and found distinct clusters in similarity matrices. Intriguingly, piRNA expression was associated with clinically relevant features of individual tumor types. We believe that piRNAs form an information network inside the cell and that appropriate statistical methods can be used for elucidating these networks.

1.3 piRNA–mRNA and piRNA–lncRNA interactions

The most well-known piRNA function is piRNA-mediated cleavage of Transposable Elements (TEs). However, recent research has shown that piRNAs are also involved in the regulation of mRNAs and lncRNAs.

Watanabe et al.[12] showed that piRNAs derived from transposons and pseudogenes mediate the degradation of many mRNAs and lncRNAs in murine late spermatocytes. Their genomic and in vivo functional analyses revealed that retrotransposon sequences in the 3' UTRs of mRNAs are targeted by piRNAs for degradation. Similarly, the degradation of spermatogenic cell-specific lncRNAs by piRNAs is mediated by retrotransposon sequences.

Gou et al.[13] reported the involvement of pachytene piRNAs in directing massive elimination of mRNAs in murine Elongating Spermatids (ES). They demonstrated that a piRNA-Induced Silencing Complex (pi-RISC) containing Murine PIWI (MIWI) and the deadenylase CAF1 is selectively assembled in ES and is responsible for inducing mRNA deadenylation and decay via a mechanism that resembles the action of miRNAs in somatic cells.

Prediction algorithms for piRNA targets in mice have begun to emerge. Yuan et al.[14] trained a support vector machine classifier using a combination of MIWI CLIP-Seq-derived features and position-derived features to predict potential mRNA targets of piRNAs. However, this software is specifically trained for mice rather than humans.

Considering the well-known roles of miRNAs in mRNA and lncRNA deadenylation in oncogenesis, we utilized this piRNA mechanism as a focus of cancer research and hypothesized that the expression of a piRNA may be inversely correlated to that of its target. As piRNA regulation results in some disease conditions, an increase in the expression levels of the target should correspond to a decrease in the expression levels of its piRNA, and vice versa. Our study aimed to predict cancer-associated piRNA–mRNA and piRNA-lncRNA interactions and to predict piRNA functions using these interactions.

2 Materials and Methods

2.1 Datasets

2.1.1 Expression data

piRNA expression data was obtained from Martinez et al.[7] The expression data contained a total of 522 expressed piRNAs in corresponding tumor tissues. Meanwhile, gene expression profiles containing coded RNAs and lncRNAs were obtained from the Synapse archive (www.synapse.org), including the following accessions: syn1461149 (BLCA), syn1461151 (BRCA), syn1461155 (COAD), syn1461156 (HNSC), syn1461159 (KIRC), syn1461166 (LUAD), syn1461168 (LUSC), syn1461171 (OV), syn1461173 (PRAD), syn1461177 (STAD), syn1461178 (THCA), and syn1461180 (UCEC). Categorical annotations of the genes were acquired using NCBI FTP[15], and 18 633 protein-coding RNAs and 795

lncRNAs were considered in the next research step.

For subsequent correlation coefficient analysis, four tumor types (BRCA, HNSC, KIRC, and LUAD) with more than 20 samples each were selected. The number of samples for each corresponding tumor type is listed in Table 1.

2.1.2 Sequence data

We obtained 75 669 human 3' UTRs from among 18 171 mRNA and 167 lncRNA sequences from the Ensembl database build 19_34b[16] using the R package "biomaRt"[17]. Human piRNA sequence data were collected by the University of California, Santa Cruz (UCSC) and accessed using the Genome Bioinformatics site (http://genome.ucsc.edu/)[18].

2.1.3 Gene-derived piRNA data

In the piRBase database[19], piRNAs mapped to RefSeq genes[20] or repeat elements annotated by RepeatMasker[21] were identified. Among these, mRNA-derived piRNAs were downloaded.

2.2 Overview of methodology

The structure of our method is presented in Fig. 1 and is divided into two main parts: expression data analysis and sequence data analysis. Expression data analysis was used for reducing false-positive rates during predictions, and sequence data analysis was employed for predicting piRNA targets and target down-regulation scores.

During expression data analysis, piRNA–target pair identification was performed based on a multi-stage filtering approach. First, only differentially expressed mRNAs and lncRNAs remained after stage 1, and piRNAs expressed in the corresponding cancers were selected in stage 2. Next, correlations between the piRNAs and candidate targets were analyzed in stage 3. Only piRNA–candidate pairs whose expression profiles were significantly inversely correlated were retained. Further details are explained in the following section.

For sequence data analysis, only the expressed piRNAs and differentially expressed candidates in the four cancer types were considered as input. miRanda[22, 23] (www.microrna.org), a well-known miRNA target software, was used for the target prediction. The intersection of miRanda results and correlation analysis results was retained. A potential set of piRNA–target pairs was obtained after filtering out mRNA-derived piRNAs and precursor mRNAs as well as piRNA–candidate pairs that only exist in one cancer type.

Table 1 Samples and tissue types studied.

Expression profile type	TCGA code (for tumor types)	Number of normal samples	Number of tumor samples
Genes (coding RNAs and lncRNAs)	BRCA	104	822
	HNSC	37	303
	KIRC	68	470
	LUAD	57	355
piRNAs	BRCA	103	1043
	HNSC	43	455
	KIRC	71	529
	LUAD	46	497

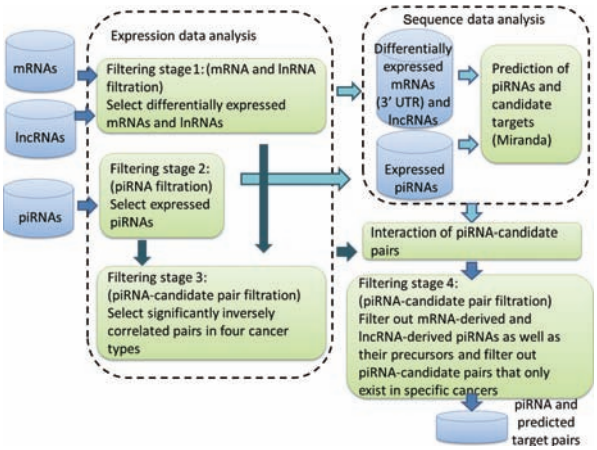


Fig. 1 Schematic of prediction pipeline.

2.2.1 Filtering stage 1 (mRNA and lncRNA filtering)

To identify cancer-related piRNA–target interactions, differentially expressed mRNAs and lncRNAs in case-control paired samples were considered as candidates, a majority of which were shown to be correlated with carcinogenesis and metastasis in most types of cancer. Therefore, the DESeq library[24] was applied to analyze the mRNA and lncRNA expression data in this study. There were three standards when considering whether a gene was significantly differentially expressed: an FDR value < 0.05, a p-value < 0.05, and a fold-change > 2. No differentially expressed mRNAs or lncRNAs were filtered out in the first stage of prediction.

2.2.2 Filtering stage 2 (piRNA filtering)

As a class of small ncRNAs, piRNA expression patterns in tumors are unique. Martinez et al.[7] noticed that only 522 piRNAs were expressed in corresponding tumor tissues. Globally, piRNA expression levels were considerably lower in non-malignant tissues compared with levels in tumor tissues, except for BRCA and KIRC tumors, which showed levels comparable to those observed in non-malignant tissues. In addition, of the 522 expressed piRNAs, 324 were significantly differentially expressed between non-malignant and tumor tissues, most of which

were overexpressed in tumors compared to levels in non-malignant tissues, with 135 piRNAs exclusively expressed in tumors[7].

In this study, all 522 expressed piRNAs in the corresponding tumor tissues were retained in the second stage of prediction. This was because piRNAs that target differentially expressed genes may not themselves be differentially expressed.

2.2.3 Filtering stage 3 (piRNA–candidate pair filtering)

Correlation analysis between piRNAs–mRNAs and piRNAs–lncRNAs was performed by sample-paired expression profiles. It is worth noting that expression levels were calculated in normal and tumor samples for each cancer. Spearmans rank correlation was used, which is ideal for detecting nonlinear relationships. The Spearmans rank correlation and the corresponding p-value were calculated using the function cor() in R language.

piRNAs are known to degrade mRNAs and regulate lncRNAs. Based on the evidence, we hypothesized that the expression of a target may be inversely correlated to that of its corresponding piRNA. Hence, piRNA-candidate pairs were retained only when their expression profiles were significantly inversely correlated both in normal and tumor samples.

In addition, piRNA co-expression analysis was used for inferring piRNA associations. Spearmans correlation-based matrices were calculated for each tumor type and piRNA co-expression networks were constructed for piRNA function research.

2.2.4 Prediction of piRNA–target pairs (miRanda)

Because of the similarities in patterns of degradation/regulation between piRNAs and miRNAs, a well-known miRNA–target prediction software program was used for piRNA–target prediction. Gou et al.[13] searched for potential targeting sites of MIWI-associated piRNAs within the 3' UTRs of MIWI-associated mRNAs using the miRanda software[23]. Based on their approach, we also used miRanda for target prediction. The 3' UTR sequences of differentially expressed mRNAs, sequences of differentially expressed lncRNAs, and sequences of expressed piRNAs were collected as input for miRanda.

2.2.5 Intersection & filtering stage 4 (piRNA–candidate pair filtering)

To improve the prediction credibility, the intersection of the miRanda and correlation analysis results was selected as input for filtering stage 4. In the last filtering stage, piRNA–candidate pairs that only existed in a specific cancer type were removed. In addition, mRNA-derived and lncRNA-derived piRNAs as well as their precursors were first filtered out. As mentioned in Section 1, some mRNAs, transposons, and lncRNAs are likely precursor molecules that are processed into small RNAs such as piRNAs. Among these, mRNA-derived piRNAs are on the same sense strand as the mRNA from which they are processed and often originate from 3' UTRs. Those piRNAs and their precursors were identified and filtered out based on the piRBase database[19]. lncRNA-derived piRNAs and their precursors were tested by UCSC[18].

DAVID 6.7 (https://david.ncifcrf.gov/home.jsp)[25, 26] was used for gene ID conversion. A classic alignment algorithm, the Needleman-Wunsch algorithm, was used for piRNA sequence alignment.

3 Results and Discussion

3.1 Results of differential expression analysis

Gene expression data for four cancer types were used for differential expression analysis of case-control paired samples. Gene probes of expression data contained 795 ncRNAs and 18 633 protein-coding RNAs.

The results are displayed in Venn diagrams (Figs. 2a and 2b), which show that a significant number of mRNAs and lncRNAs were exclusively expressed in individual tumor types. In contrast, the variation in the expression of piRNAs among cancers was small (Fig. 2c). As a class of small ncRNAs, the number of expressed piRNAs was far less than the number of differentially expressed

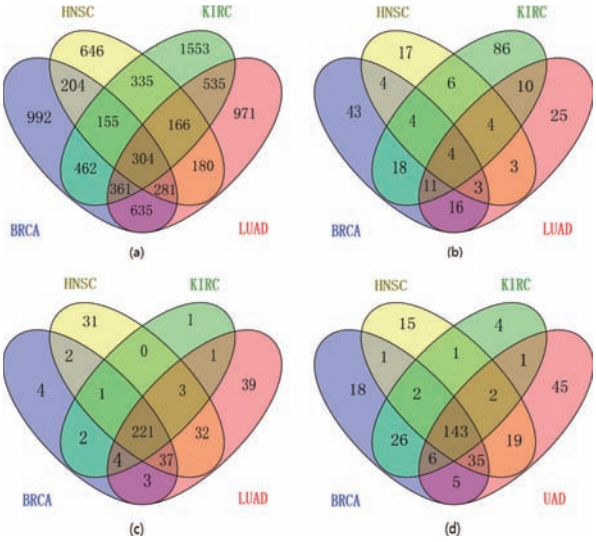


Fig. 2 Venn diagrams for four cancers. (a) Differentially expressed mRNAs, (b) differentially expressed lncRNAs, (c) expressed piRNAs, and (d) differentially expressed piRNAs.

mRNAs for each cancer type.

Furthermore, referring to Martinez et al.[7], differential expression analyses of piRNAs were performed using the Comparative Marker Selection module implemented in the gene pattern[27, 28]. The results showed that there is still a large overlap in the piRNAs that are differentially expressed in the four cancers (Fig. 2d). We believe that research on the piRNAs involved in pan-cancer represents a new area of focus in the field of functional analysis.

3.2 piRNA co-expression analysis

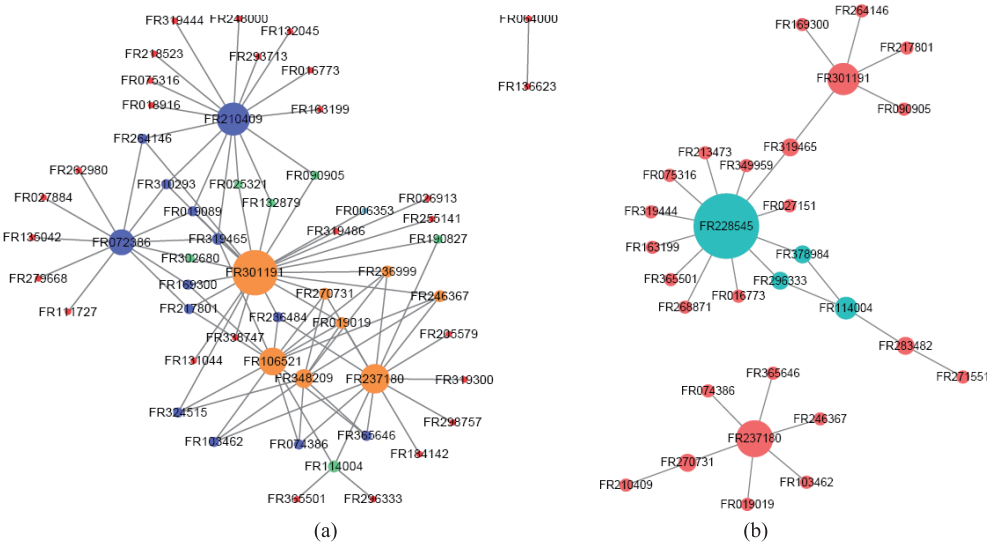
Co-expression analysis is a common tool for inferring gene associations and functions using expression data and can be applied to piRNAs. Spearman's correlation-based matrices were calculated as well as piRNA co-expression networks for the four cancer tissues and noncancerous (control) tissues. Figure 3 demonstrates the use of CytoScape[29] to visualize interactions between the BRCA tumor and adjacent noncancerous tissues.

The degree of centrality is defined as the number of links one node has to another, which determines the relative importance of a gene. In addition, k-cores were applied as a method of simplifying graph topology analyses. Core regulatory factors (piRNAs), which have the highest degrees, connect the most adjacent piRNAs and build the structure of the network.

There were more connections among piRNAs in cancer tissues than in control tissues, suggesting that most of the physiological piRNA–piRNA interactions and linkages in normal tissues had been altered in the cancer tissues. The piRNAs with high degrees of centrality and k-cores, indicating that they possessed most of the interactions with other piRNAs, were considered key piRNAs in the interaction network and are listed in Table 2. The biological relevance of these key piRNAs requires further

Table 2 Top 5 piRNAs with highest degrees of centrality and k-cores in co-expression networks of four cancers and adjacent noncancerous tissues.

Tumor type	Co-expression networks in tumor samples			Co-expression networks in normal samples		
	PiRNA name	k-core	Degree	PiRNA name	k-core	Degree
BRCA	FR301191	4	23	FR228545	2	12
	FR237180	4	14	FR114004	2	3
	FR106521	4	13	FR296333	2	2
	FR348209	4	8	FR378984	2	2
	FR270731	4	4	FR301191	1	5
HNSC	FR301191	6	56	FR301191	6	38
	FR237180	6	26	FR202919	6	28
	FR214430	6	24	FR037883	6	26
	FR090905	6	6	FR270731	6	14
	FR236484	6	6	FR190827	6	10
KIRC	FR237180	4	28	FR237180	2	22
	FR072386	4	22	FR136623	2	20
	FR136623	4	16	FR072386	2	6
	FR210409	4	10	FR163199	2	6
	FR131044	4	6	FR381169	2	6
LUAD	FR301191	10	52	FR228545	4	46
	FR321895	10	36	FR252722	4	22
	FR384068	10	34	FR136216	4	12
	FR202919	10	32	FR125672	4	10
	FR145670	10	24	FR384068	4	10



(a) (b)

Fig. 3 Co-expression network of piRNAs in normal and cancerous tissues. (a) Co-expression network in BRCA, and (b) co-expression network in adjacent noncancerous tissues.

exploration.

3.3 Correlation analysis

piRNAs exert their function through mRNA and lncRNA deadenylation. Consequently, we hypothesized that the expression of a piRNA target would be inversely correlated to that of the piRNA.

Correlation analysis was performed using piRNA, mRNA, and lncRNA expression data in individual tumor types. In BRCA, for example, 3391 differentially expressed mRNAs, 103 differentially expressed lncRNAs, and 274 expressed piRNAs were considered. Paired gene (coding RNAs and lncRNAs) and piRNA expression data were calculated both in normal and tumor tissues.

From Fig. 4, each histogram of correlation values conforms to the normal distribution. In addition, the wilcoxon rank sum test is used to test the difference between the case and control correlation, with a p-value $< 2.2-16$.

In this study, we focused on significant inverse correlations between piRNA–candidate pairs (cor < 0, $p < 0.05$) in both normal and tumor samples simultaneously. The results are provided in Table 3.

3.4 Results of miRanda

Based on expression data analysis, 8221 mRNAs, 254 lncRNAs, and 323 piRNAs were obtained. Among these mRNAs, 18 173 3' UTR sequences from 7321 mRNAs were available in the Ensembl database[16] and downloaded by the R package "biomaRt"[17]. In addition, 167 differentially expressed lncRNA sequences were obtained

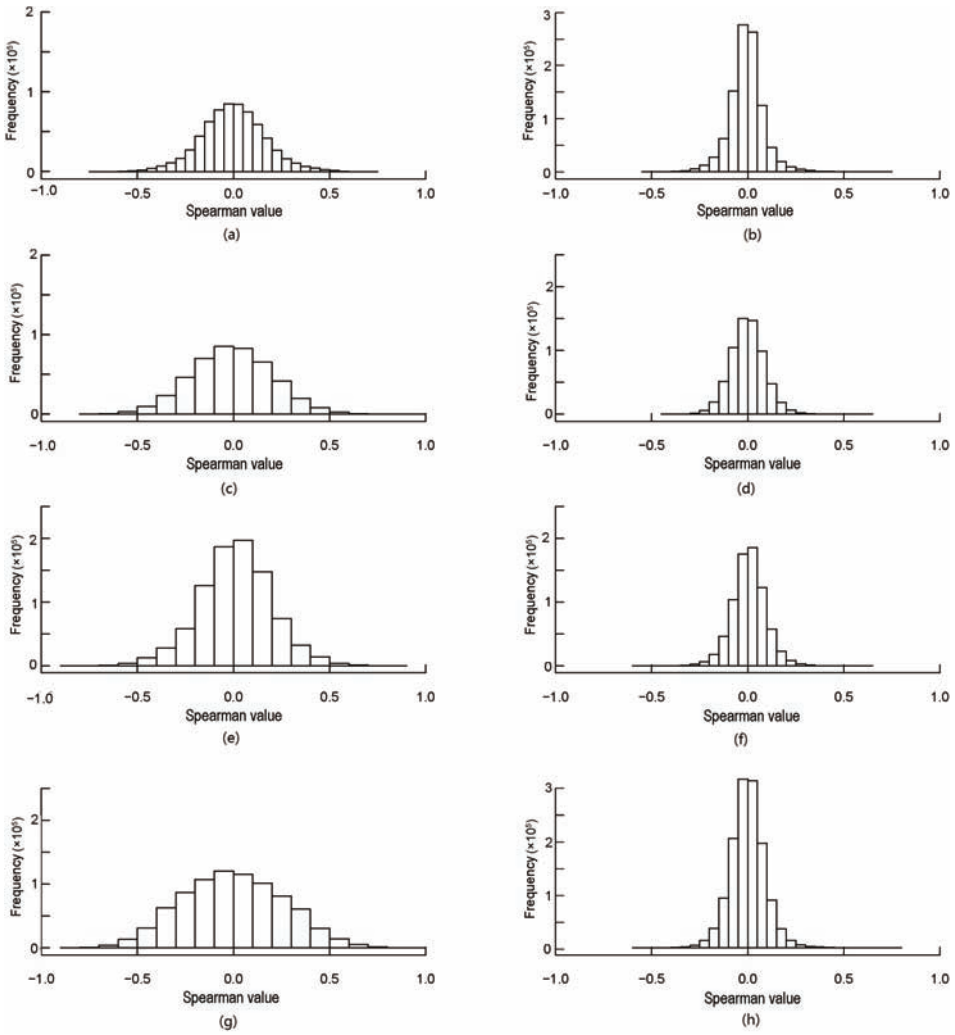


Fig. 4 Histograms of correlation values between piRNA-gene pairs in (a) BRCA control samples and (b) BRCA cases; (c) HNSC control samples and (d) HNSC cases; (e) KIRC control samples and (f) KIRC cases; (g) LUAD control samples; and (h) LUAD cases.

Table 3 Results of significant inverse correlation analysis.

TCGA code (for tumor code)	Number of significant inverse correlation piRNA-gene pairs		
	Normal samples	Tumor samples	Intersection
BRCA	90362	156130	26519
HNSC	36846	61006	5351
KIRC	76242	87480	14921
LUAD	29061	116482	4529

the same way. Furthermore, 323 piRNA sequences were collected by UCSC[18].

The above sequences were used as input for the miRanda software[23] to predict piRNA targets from the 3' UTRs of mRNAs and lncRNAs using the options "-strict -quiet". As a result, 711073 piRNA–mRNA pairs (458008 unique pairs) and 100620 piRNA–lncRNA pairs (38770 unique pairs) were obtained by miRanda.

3.5 piRNA–mRNA and piRNA–lncRNA interaction analysis

3.5.1 Analysis of predicted pair sets

After filtering of mRNA-derived piRNAs, 3687 piRNA–mRNA pairs in BRCA, 800 pairs in HNSC, 1955 pairs in KIRC, and 693 pairs in LUAD were preserved. By analyzing these sets of predicted mRNAs, 2143 of 2639 mRNAs (81.2%) were found to be expressed in a specific cancer (Fig. 5a). After lncRNA-derived piRNAs were filtered, 479 piRNA–lncRNA pairs in BRCA, 62 pairs in HNSC, 451 pairs in KIRC, and 31 pairs in LUAD were preserved. By analyzing these sets of predicted lncRNAs, 104 out of 132 mRNAs (78.8%) were found to be expressed in a specific cancer (Fig. 5b).

In contrast, a small percentage of piRNAs (24.6%) were specifically expressed, and the intersection of piRNAs among the four sets was 55 (29.4%) (Fig. 5c). These results imply that the predicted piRNAs are consistent across cancer types, and their functions may therefore be related to pan-cancer.

3.5.2 Agreement with known piRNA cancer biomarkers

To date, the exact role of piRNAs in cancer has not been extensively studied. Relative data and reports mainly concern mice and rarely involve humans. A closer look at the predicted piRNAs revealed that three piRNAs had been reported as potential cancer biomarkers in previous studies, including piR-36741 (5'-GTTTAGACGGGCTCACATCAC-3'), piR-21032 (5'-TAAAGTGCTGACAGTGCAGATAGTGGTCCTC-3'), and piR-57125 (5'-TGGTCGTGGTTG TAGTCCGTG-3')[30, 31]. Yang et al.[31] reported that these piRNAs are stably expressed in human serum or plasma samples and can serve as valuable blood-based biomarkers for disease detection and monitoring.

3.5.3 Interaction network analysis

We focused on piRNA–mRNA and piRNA–lncRNA interactions exclusively related to tumors and selected 198 piRNA–mRNA and 10 piRNA–lncRNA pairs shared by at least two cancers from which to construct a network.

Similar to the miRNA regulatory mechanism, we speculated that the lower the free energy of the piRNA–target pair, the greater the binding affinity. Hence, we used the minimum free energy of the piRNA–target binding and CytoScape[29] to visualize these interactions, as shown in Fig. 6. In this network, most mRNAs can be identified as potential targets of one or two piRNAs, and a piRNA may target one or more mRNAs. While half of the key piRNAs in the network have only one or two interactions, we believe that the biological meaning of the interactions between key piRNAs and their target mRNAs and lncRNAs is worthy of future attention.

3.5.4 Functional analysis

We investigated the potential functions of piRNAs by

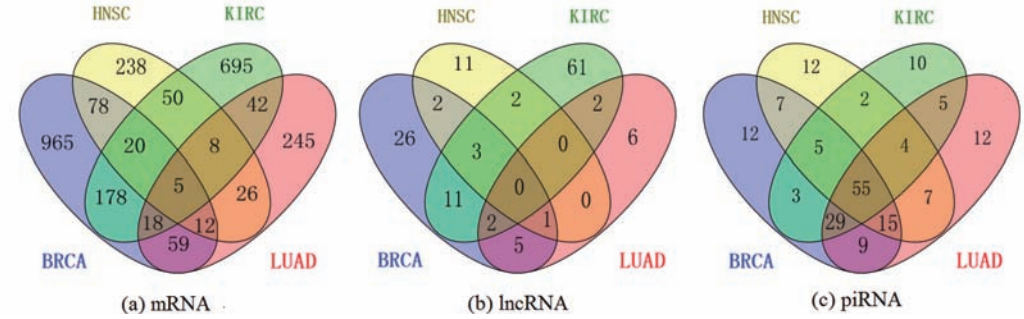


Fig. 5 Venn diagrams of (a) predicted mRNAs, (b) predicted lncRNAs, and (c) predicted piRNAs in the four cancers.

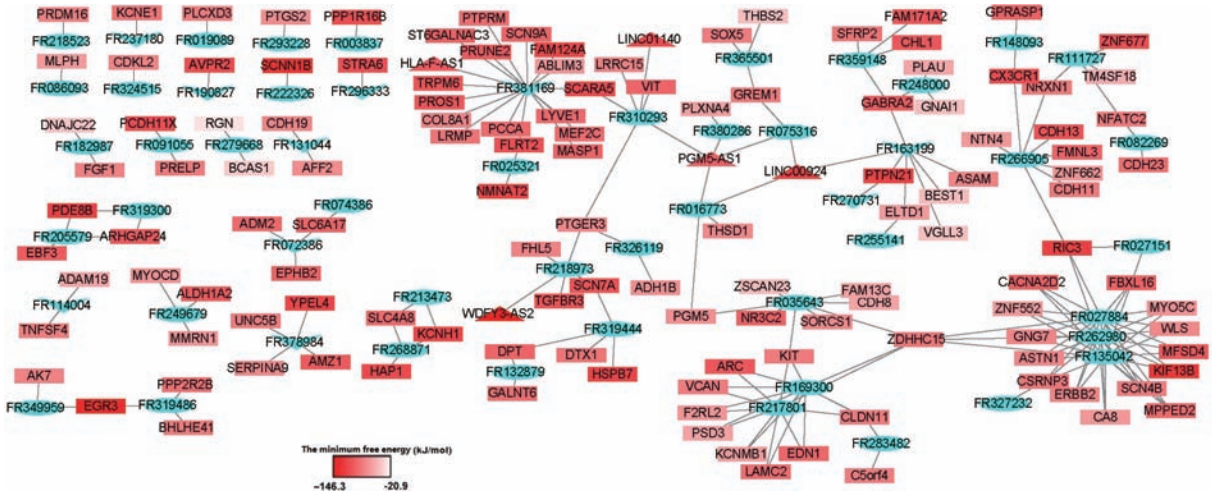


Fig. 6 piRNA–mRNA interaction network showing key piRNAs listed in Table 3 (blue chevrons), remaining piRNAs (blue ovals), mRNAs (red rectangles), and lncRNAs (red triangles). Intensity of each red node indicates the minimum free energy for all interactions with that node.

analyzing their mRNA targets using Webgestalt[32, 33], which is a suite of tools for functional enrichment analysis in various biological contexts. We analyzed all 128 non-cancer-type-specific genes, and the most enriched biological processes are shown in red as a Directed Acyclic Graph (DAG) in Fig. 7.

The most significant biological function was blood vessel development, which is closely related to the cancer hallmark sustained angiogenesis. Another significant GO term cell adhesion is highly related to the cancer hallmark activating invasion and metastasis. These indicate potential roles for piRNAs in cancer, which should be studied further.

4 Conclusion

In addition to their transposon regulatory function, piRNAs may also target and regulate the expression of mRNAs. It is also possible that piRNAs may be involved in tumorigenesis through the regulation of cancer-related genes. Furthermore, gene expression analyses conducted on cancers have previously been used to predict diagnostic markers and to identify gene expression patterns associated with prognosis but have not been used to reveal interactions between piRNAs and mRNAs/lncRNAs. This study aimed to predict cancer-associated piRNA–mRNA and piRNA–lncRNA interactions and to predict piRNA regulatory functions based on those interactions.

Based on the quality and quantity of expression data, four cancers (BRCA, HNSC, KIRC, and LUAD) were selected. Corresponding piRNA–mRNA interactions were identified by integrated analysis of piRNA and mRNA expression and sequence data. We predicted 198 piRNA–mRNA interactions related to cancer.

The results showed that in contrast with gene expression, piRNA expression is relatively consistent among the four cancers. Furthermore, the identified piRNAs are consistent with previously published cancer biomarkers, as in the case of piRNA-36741, piRNA-21032, and piRNA-57125. More importantly, potential piRNA functions were predicted by constructing an interaction network and analyzing piRNA targets. These targets were enriched in gene ontology categories related to the cancer hallmarks activating invasion and metastasis and sustained angiogenesis. We believe that piRNAs are closely related to cancer development and require further study.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Nos. 91530113, 61571341, and 61201312); the Research Fund for the Doctoral Program of Higher Education of China (No. 20130203110017); the Fundamental Research Funds for the Central Universities of China (Nos. BDY171416, JBZ170301, 20101164977, and JB140306); and the Natural Science Foundation of Shaanxi Province in China (Nos. 2015JM6275 and 207JM6024).

References

[1] A. G. Seto, R. E. Kington, and N. C. Lau, The coming of age for piwi proteins, *Molecular Cell*, vol. 26, no. 5, pp. 603–609, 2007.

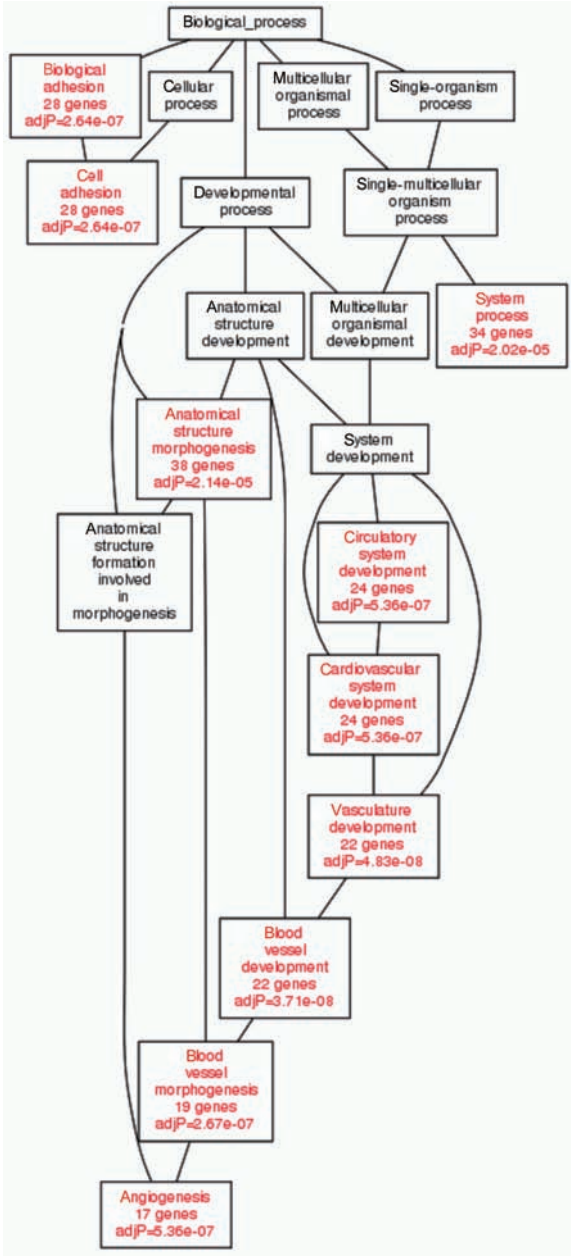


Fig. 7 A directed acyclic graph of piRNA target mRNAs based on GO analysis. GO terms in red are significant.

[2] A. Aravin, D. Gaidatzis, S. Pfeffer, M. Lagosquintana, P. Landgraf, N. Iovino, P. Morris, M. J. Brownstein, S. Kuramochimiyagawa, T. Nakano, et al., A novel class of small rnas bind to mili protein in mouse testes, *Nature*, vol. 442, no. 7099, pp. 203–207, 2006.

[3] E. M. Weick and E. A. Miska, piRNAs: From biogenesis to function, *Development*, vol. 141, no. 18, pp. 3458–3471, 2014.

[4] T. Thomson and H. Lin, The biogenesis and function of piwi proteins and piRNAs: Progress and prospect, *Annual Review of Cell & Developmental Biology*, vol. 2009, no. 1, pp. 355–376, 2009.

[5] S. Houwing, L. M. Kamminga, E. Berezikov, D. Cronembold, A. Girard, D. E. H. Van, D. V. Filippov, H. Blaser, E. Raz, and C. B. Moens, A role for piwi and piRNAs in germ cell maintenance and transposon silencing in zebrafish, *Cell*, vol. 129, no. 1, pp. 69–82, 2007.

[6] Z. Yan, H. Y. Hu, X. Jiang, V. Maierhofer, E. Neb, L. He, Y. Hu, H. Hu, N. Li, W. Chen, et al., Widespread expression of piRNA-like molecules in somatic tissues, *Nucleic Acids Research*, vol. 39, no. 15, pp. 6596–6607, 2011.

[7] V. D. Martinez, E. A. Vucic, K. L. Thu, R. Hubaux, K. S. Enfield, L. A. Pikor, D. D. Beckersantos, C. J. Brown, S. Lam, and W. L. Lam, Unique somatic and malignant expression patterns implicate piwi-interacting RNAs in cancer-type specific biology, *Scientific Reports*, vol. 5, p. 10423, 2015.

[8] R. Suzuki, S. Honda, and Y. Kirino, Piwi expression and function in cancer, *Frontiers in Genetics*, vol. 3, p. 204, 2012.

[9] J. Cheng, J. M. Guo, B. X. Xiao, Y. Miao, Z. Jiang, H. Zhou, and Q. N. Li, piRNA, the new non-coding rna, is aberrantly expressed in human cancer cells, *Clinica Chimica Acta; International Journal of Clinical Chemistry*, vol. 412, nos. 17&18, p. 1621, 2011.

[10] J. Cheng, H. Deng, B. Xiao, H. Zhou, F. Zhou, Z. Shen, and J. Guo, pir-823, a novel non-coding small RNA, demonstrates in vitro and in vivo tumor suppressive activity in human gastric cancer cells, *Cancer Letters*, vol. 315, no. 1, pp. 12–17, 2012.

[11] L. Cui, Y. Lou, X. Zhang, H. Zhou, H. Deng, H. Song, X. Yu, B. Xiao, W. Wang, and J. Guo, Detection of circulating tumor cells in peripheral blood from patients with gastric cancer using piRNAs as markers, *Clinical Biochemistry*, vol. 44, no. 13, p. 1050, 2011.

[12] T. Watanabe, E. C. Cheng, M. Zhong, and H. Lin, Retrotransposons and pseudogenes regulate mRNAs and lncRNAs via the piRNA pathway in the germline, *Genome Research*, vol. 25, no. 3, p. 368, 2014.

[13] L. T. Gou, P. Dai, J. H. Yang, E. D. Wang, and M. F. Liu, Pachytene piRNAs instruct massive mRNA elimination during latespermiogenesis, *Cell Research*, vol. 24, no. 6, pp. 680–700, 2014.

[14] J. Yuan, P. Zhang, Y. Cui, J. Wang, G. Skogerbϕ, D. W. Huang, R. Chen, and S. He, Computational identification of pirna targets on mouse mrnas, *Bioinformatics*, vol. 32, no. 8, p. 1170, 2016.

[15] NCBI FTP, ftp://ftp.ncbi.nih.gov/gene/DATA/GENE_INFO/Mammalia/Homo_sapiens.gene_info.gz, 2016.

[16] P. Flicek, B. L. Aken, B. Ballester, K. Beal, E. Bragin, S. Brent, Y. Chen, P. Clapham, G. Coates, S. Fairley, et al., Ensembl's 10th year, *Nucleic Acids Research*, vol. 38, no. 1, pp. D557–D562, 2010.

[17] S. Haider, B. Ballester, D. Smedley, J. Zhang, P. Rice, and

A. Kasprzyk, Biomart central portal—Unified access to biological data, *Nucleic Acids Research*, vol. 37, no. Web Server issue, p. W23, 2009.

[18] D. Karolchik, A. S. Hinrichs, T. S. Furey, K. M. Roskin, C. W. Sugnet, D. Haussler, and W. J. Kent, The ucsc table browser data retrieval tool, *Nucleic Acids Research*, vol. 32, no. Database issue, pp. 493–496, 2004.

[19] P. Zhang, X. Si, G. Skogerbϕ, J. Wang, D. Cui, Y. Li, X. Sun, L. Liu, B. Sun, and R. Chen, pirbase: A web resource assisting piRNA functional study, *Database*, vol. 2014, p. bau110, 2014.

[20] K. D. Pruitt, T. Tatusova, and D. R. Maglott, Ncbi reference sequence (refseq): A curated non-redundant sequence database of genomes, transcripts and proteins, *Nucleic Acids Research*, vol. 33, no. Database issue, pp. D501–D504, 2005.

[21] J. Jurka, Repbase update: A database and an electronic journal of repetitive elements. *Trends in Genetics*, vol. 16, no. 9, p. 418, 2000.

[22] A. J. Enright, B. John, U. Gaul, T. Tuschl, C. Sander, and D. S. Marks, Microrna targets in drosophila, *Genome Biology*, vol. 5, no. 1, p. R1, 2003.

[23] J. Bino, A. J. Enright, A. Alexei, T. Thomas, S. Chris, and D. S. Marks, Human microRNA targets, *Plos Biology*, vol. 2, no. 11, p. e363, 2004.

[24] S. Anders and W. Huber, Differential expression analysis for sequence count data, *Genome Biology*, vol. 11, no. 10, p. R106, 2010.

[25] D. W. Huang, B. T. Sherman, and R. A. Lempicki, Systematic and integrative analysis of large gene lists using david bioinformatics resources, *Nature Protocol*, vol. 4, no. 1, pp. 44–57, 2009.

[26] D. Huang, B. T. Sherman, and R. A. Lempicki, Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists, *Nucleic Acids Research*, vol. 37, no. 1, pp. 1–13, 2009.

[27] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, Cluster analysis and display of genome-wide expression patterns, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 25, pp. 14 863–14 868, 1998.

[28] M. Reich, T. Liefeld, J. Gould, J. Lerner, P. Tamayo, and J. P. Mesirov, Genepattern 2.0, *Nature Genetics*, vol. 38, no. 5, pp. 500–501, 2006.

[29] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, Cytoscape: A software environment for integrated models of biomolecular interaction networks, *Genome Research*, vol. 13, no. 11, pp. 2498–2504, 2003.

[30] C. H. Kwon, H. Tak, M. Rho, H. R. Chang, Y. H. Kim, K. T. Kim, C. Balch, E. K. Lee, and S. Nam, Detection of piwi and piRNAs in the mitochondria of mammalian cancer cells, *Biochemical & Biophysical Research Communications*, vol. 446, no. 1, pp. 218–223, 2014.

[31] X. Yang, Y. Cheng, Q. Lu, J. Wei, H. Yang, and M. Gu, Detection of stably expressed pirnas in human blood, *International Journal of Clinical & Experimental Medicine*, vol. 8, no. 8, pp. 13 353–13 358, 2015.

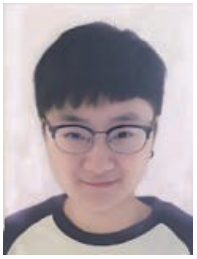
[32] B. Zhang, S. Kirov, and J. Snoddy, Webgestalt: An integrated system for exploring gene sets in various biological contexts, *Nucleic Acids Research*, vol. 33, no. Web Server issue, p. W741, 2005.

[33] J. Wang, D. Duncan, Z. Shi, and B. Zhang, Web-based gene set analysis toolkit (webgestalt): Update 2013, *Nucleic Acids Research*, vol. 41, no. Web Server issue, p. W77, 2013.

Junying Zhang received the PhD degree in signal and information processing from Xidian University, Xi'an, China, in 1998. From 2001 to 2002, she was a visiting scholar at the Department of Electrical Engineering and Computer Science, The Catholic University of America, Washington, DC, USA, and in 2007, she was a visiting professor at the Department of Electrical Engineering and Computer Science, Virginia Polytechnical University, USA. She is currently a professor in the School of Computer Science and Technology, Xidian University, Xi'an, China. Her research interests focus on intelligent information processing, including machine learning and its application to cancer related bioinformatics, causative learning, and pattern discovery.

Aimin Li received the PhD degree in computer application technology from Xidian University, Xi'an, China, in 2015. He is currently a primary lecturer in the School of Computer Science and Engineering, Xi'an University of Technology, China. His research interests focus on non-coding RNAs and epigenetics, including machine learning and its application to non-coding RNA bioinformatics, methylation, gene regulation, and pattern discovery.

Yajun Liu received the BS and MS degrees in computer applications from Beijing Institute of Technology, China, in 2006 and 2009, respectively. Since 2012 she has been working towards PhD degree in computer application technology at Xidian University. Her current research interests are non-coding RNA and gene expression data analysis.



Zhaowen Liu is a PhD student at School of Computer Science and Technology, Xidian University, China. She received the BSc degree from Hunan University, China in 2008. Her research interests include complex network, bioinformatics, and computational neuroscience.

Zongzhen He is a PhD student at the School of Computer Science and Technology, Xidian University, China. She received the MSc degree in computer application technology from Zhengzhou University, China, in 2014, and the BSc degree from Luoyang Institute of Science and Technology, Henan, China, in 2011. Her research interests include data mining, computational bioinformatics, tumor informatics, and molecular biology.

Xiguo Yuan received the BS and MS degrees in computer applications from Wuhan University of Science & Technology, China, in 2005 and 2008, respectively, the PhD degree in computer application technology from Xidian University, China, in 2011. His current research interests are bioinformatics and pattern recognition.

Shouheng Tuo received the BS degree in computer application from the Lanzhou University of Finance and Economics, China, in 2001, and received MS degree in computer application technology from the University of Electronic Science and Technology of China, Chengdu, China, in 2008. Since 2014, he has been working towards PhD degree in computer application technology at Xidian University, Xi'an, China. His current research interests are intelligent computing and bioinformatics.