Construction and analysis of dysregulated lncRNA-associated ceRNA network identified novel lncRNA biomarkers for early diagnosis of human pancreatic cancer

Meng Zhou[1,*], Zhiyong Diao[2,*], Xiaolong Yue[3,*], Yang Chen[1], Hengqiang Zhao[1], Liang Cheng[1], Jie Sun[1]

[1]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, 150081, PR China

[2]Department of Plastic Surgery, the First Affiliated Hospital of Harbin Medical University, Harbin, 150001, PR China

[3]Medical Oncology Department, Affiliated Tumor Hospital, Harbin Medical University, Harbin, 150001, PR China

[*]These authors contributed equally to this work

Correspondence to: Jie Sun, **email:** suncarajie@hotmail.com
Liang Cheng, **email:** liangcheng@hrbmu.edu.cn

ABSTRACT

It is increasing evidence that ceRNA activity of long non-coding RNAs (lncRNAs) played critical roles in both normal physiology and tumorigenesis. However, functional roles and regulatory mechanisms of lncRNAs as ceRNAs in pancreatic ductal adenocarcinoma (PDAC), and their potential implications for early diagnosis remain unclear. In this study, we performed a genome-wide analysis to investigate potential lncRNA-mediated ceRNA interplay based on "ceRNA hypothesis". A dysregulated lncRNA-associated ceRNA network (DLCN) was constructed by utilizing sample-matched miRNA, lncRNA and mRNA expression profiles in PDAC and normal samples in combination with miRNA regulatory network. The results of network analysis uncovered seven novel lncRNAs as functional ceRNAs whose aberrant expression will result in the extensive variation in tumorigenic or tumor-suppressive gene expression through DLCN at the post-transcriptional level contributing to PDAC. Therefore, we developed a 7-lncRNA signature (termed LncRisk-7) based on the expression data of seven lncRNAs and SVM algorithm as a novel diagnostic tool to improve early diagnosis of PDAC. The LncRisk-7 achieved high performance in distinguishing PDAC patients from nonmalignant pancreas samples in the discovery cohort and was further confirmed in another two independent validation cohorts. Functional analysis demonstrated that seven lncRNA biomarkers act as ceRNAs involving the regulation of cell death, cell adhesion and cell cycle. This study will help to improve our understanding of the lncRNA-mediated ceRNA regulatory mechanisms in the pathogenesis of PDAC and provide novel lncRNAs as candidate diagnostic biomarkers or potential therapeutic targets.

INTRODUCTION

Pancreatic cancer (PC) remains the most common cause of cancer-related mortality worldwide [1]. Pancreatic ductal adenocarcinoma (PDAC) is the most frequent subtype of pancreatic cancer, accounting for more than 85% of pancreatic tumor cases [2]. Despite increasing efforts in PDAC, the 5-year survival rate is low at less than 5%, primarily due to late stage diagnosis at advanced stages in a large proportion of patients [1]. To decrease mortality and improve the management of PDAC, diagnostic markers is critical for early detection and risk stratification of PDAC which could aid clinicians to select early tailored treatment. However, the PDAC is a heterogeneous disease in various aspects including clinicopathological, molecular and cellular heterogeneity [3, 4]. Therefore, traditional recognizable clinical and pathological symptoms or signs have limited value in

detecting early PDAC. Molecular biomarkers have been proven to be a promising clinical tool for identifying patients' subgroup having early-stage disease.

Over the last ten years, large-scale genome and transcriptome studies have documented large numbers of non-coding RNAs (ncRNAs), including short ncRNAs and long ncRNAs [5, 6]. Long non-coding RNAs (lncRNAs), a major class of ncRNAs with larger than 200 nucleotides in length, have been reported to participate in a wide range of biological processes, including genomic imprinting, transcriptional and post-transcriptional regulation [7–9]. It has been shown that lncRNAs harboring miRNA response elements (MREs) can be regulated by miRNAs, thus acting as competing endogenous RNAs (ceRNAs) to communicate with mRNAs by competing for shared miRNAs [10, 11]. Experimental evidence revealed that dysregulated expressions of key lncRNAs of ceRNA network have greater effects on the miRNA-mediated lncRNA/mRNA ceRNA crosstalk interactions and disrupt bistable states, thus contributing to the initialization and development of cancers [11, 12]. Wang and his colleagues showed that the crosstalk between lncRNA *HULC* and *PRKACB* gene via competitive binding to *miR-372* in a ceRNA-dependent feed-forward loop resulted in highly up-regulated expression of *HULC* in liver cancer [13]. Paci *et al.* built two miRNA-mediated lncRNA-mRNA networks in normal and pathological breast tissue and investigated the potential role of lncRNA as ceRNA in breast cancer [14]. Recent two studies had led to important insights into ceRNA-mediated gene regulation in gastric cancer and ovarian cancer and identified candidate functional cancer-related lncRNAs [15, 16]. Several lncRNAs have been found to be differentially expressed in PC tissues compared to the healthy controls, such as *HOTTIP-005*, *RP11-567G11.1* and *MALAT1* [17, 18], implying their potentials as diagnostic or prognostic biomarkers in PC. A more recent study reported that lncRNA *NUTF2P3-001* function as ceRNAs to communicate with *KRAS* mRNA by competitively binding to *miR-3923*, and the overexpression of *NUTF2P3-001* will unregulated *KRAS* expression by depriving the inhibition of *miR-3923* on *KRAS* leading to the proliferation and invasion of pancreatic cancer cell [19], revealing the functional roles of lncRNA-mediated ceRNA crosstalk in PC for the first time.

In this study, as demonstrated by Chou's 5-step rule [20] to establish a really useful prediction method for a biological system, we first investigated the altered expression patterns of lncRNAs, miRNAs and mRNAs between PDAC patients and normal samples. Then a dysregulated lncRNA-associated ceRNA network (DLCN) was constructed by integrating altered lncRNAs, miRNAs, mRNAs and their co-dysregulated regulatory relationships based on "ceRNA hypothesis". We uncovered 7 novel lncRNAs as functional ceRNAs with key roles in the pathogenesis of PDAC, and developed a SVM-based 7-lncRNA signature (termed LncRisk-7) which significantly discriminate PDAC tumors from nonmalignant pancreas samples with high performance. The diagnostic values of the LncRisk-7 were further validated in another two independent testing cohorts. With further experimental validation, these novel lncRNAs acting as ceRNAs at the post-transcriptional level may become promising diagnostic biomarkers and therapeutic targets.

RESULTS

Identification of differentially expressed mRNAs, miRNAs and lncRNAs in PDAC patients

We first compared the expression profiles of mRNAs, miRNAs and lncRNAs between 25 PDAC samples and 7 nonmalignant pancreas samples from the discovery cohort, and found that 1032 mRNAs, 87 miRNAs and 78 lncRNAs are differentially expressed (Fold change ≥ 1.5 or ≤ 0.67 and FDR-adjusted $p \leq 0.1$) in PDAC samples compared with nonmalignant pancreas samples using the SAM analysis (Supplementary Table S1). Of these, 632 mRNAs, 47 miRNAs and 52 lncRNAs were over-expressed and 400 mRNAs, 40 miRNAs and 26 lncRNAs were down-expressed in PDAC patients compared with nonmalignant pancreas samples.

Construction and analysis of dysregulated lncRNA-associated ceRNA network

The preliminary dysregulated lncRNA-associated ceRNA network (DLCN) was first built by integrating expression profiles and regulatory relationships of mRNAs, miRNAs and lncRNAs of 32 samples in the discovery cohort. As described in the Materials and Methods section, we detected 290 miRNA-mediated lncRNA-mRNA competing triplets among 5 miRNAs, 7 lncRNAs and 150 mRNAs (Supplementary Table S2). These significant competing triplets were then assembled to constitute dysregulated lncRNA-associated ceRNA network for exploring the dynamic changes of ceRNA regulation in PDAC. The resulted DLCN was comprised of 467 edges among 5 miRNAs, 7 lncRNAs and 150 mRNAs (Figure 1A).

Next, we performed network analysis to study the structure and organization of the DLCN. The degree distribution of nodes in the DLCN was investigated and the power-law distribution with a slope of −1.783 and $R^2 = 0.9988$ was observed (Figure 2B), suggesting that the DLCN displayed scale-free characteristics typical of biological networks. Furthermore, the DLCN showed module characteristics with a significantly higher clustering coefficient than random networks ($p < 0.01$, Figure 2C). However, the characteristic path length of the DLCN is substantially larger than that of random networks

($p < 0.001$) (Figure 1D), implying that the DLCN had reduced global efficiency. The scale-free, module characteristics and reduced global efficiency of the DLCN suggested that dysregulated ceRNA interactions often occurred at a local scale and hubs in the modules tended to be critical in the context of the entire network.

Further comparison analysis showed that there were significant differences in the degree distribution and betweenness centrality among mRNAs, miRNAs and lncRNAs ($p = 3.356e-09$ for degree and $p = 2.65e-07$ for betweenness centrality, Kruskal-Wallis test) (Figure 2A and 2B). LncRNAs and miRNAs have significantly higher degree and betweenness centrality compared with mRNAs, suggesting that lncRNAs and miRNAs tended to be hub nodes. As showed in Figure 1A, a large proportion of mRNAs (55.3%) communicated with individual lncRNAs, and all lncRNAs acted as ceRNAs to communicate with multiple mRNAs by competing specific shared miRNAs. These results suggested that the aberrant expression of lncRNA ceRNA would result in the extensive variation in gene expression by miRNA-mediated lncRNA-mRNA ceRNA crosstalk interactions, implying that ceRNA function of lncRNAs in the DLCN is of crucial importance in the development of PDAC.

Identification of potential diagnostic lncRNA signature in PDAC from the discovery cohort

Based on the above observation, these 7 lncRNAs with ceRNA activity were considered as potential biomarkers associated with PDAC (Table 1). To test whether these 7 lncRNA biomarkers could efficiently distinguish PDAC patients from nonmalignant pancreas samples, we performed unsupervised hierarchical clustering for 32 samples in the discovery cohort according to the expression pattern of 7 lncRNA biomarkers. The results showed that all samples in the discovery cohort were grouped into two distinctive sample clusters (11 samples in Cluster 1 vs. 21 samples in Cluster 2), which were highly correlated with disease status ($p = 9.804e-05$, Fisher exact test; Figure 2C). As seen in Figure 2C, all nonmalignant pancreas samples were clustered into Cluster 1, and most of PDAC patients
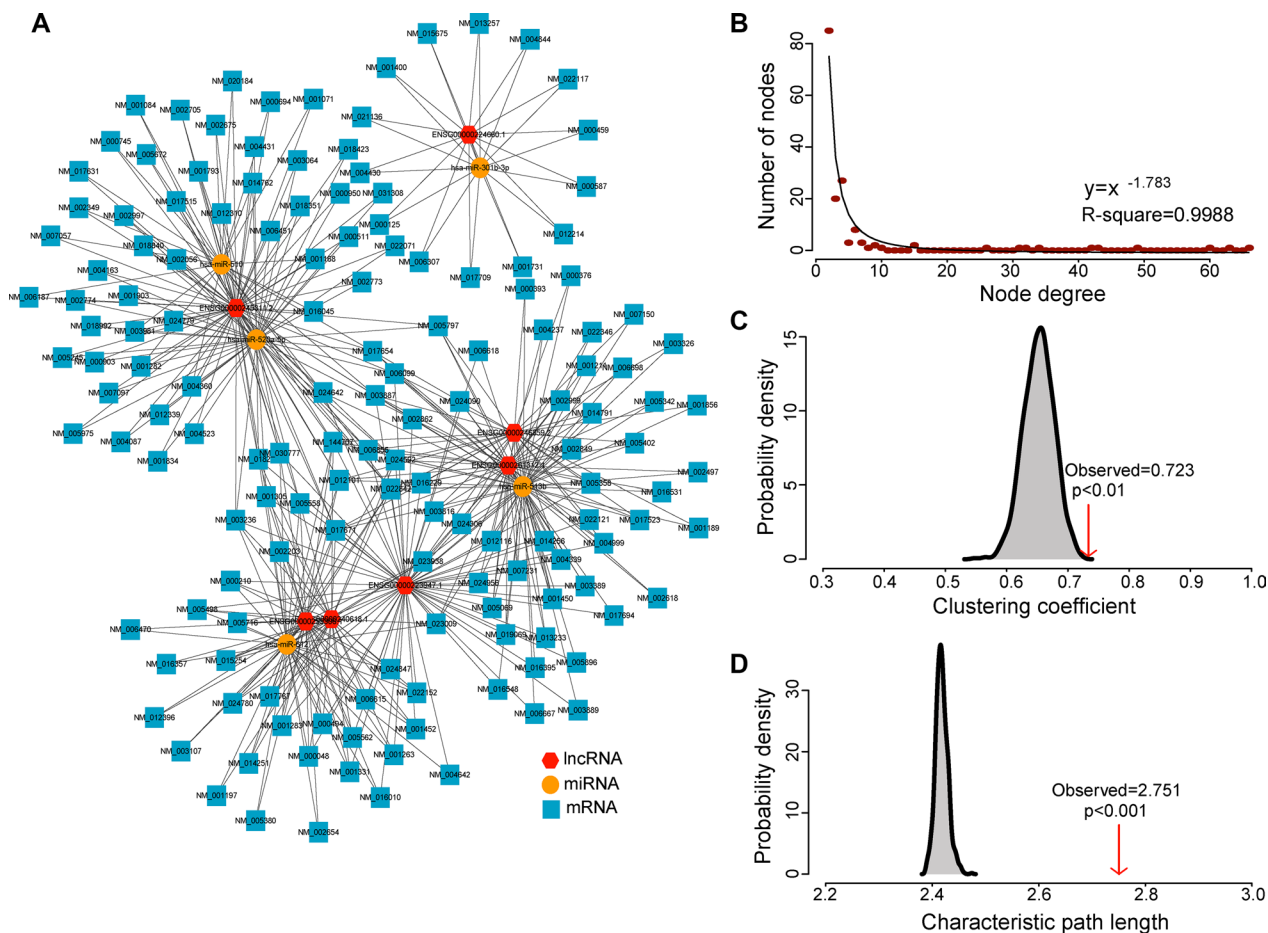
Figure 1: The layout of dysregulated lncRNA-mediated ceRNA network (DLCN) and its structural characteristics.
(**A**) Global view of the DLCN in pancreatic cancer. The DLCN was comprised of 467 edges among 5 miRNAs, 7 lncRNAs and 150 mRNAs. (**B**) Degree distribution of the DLCN. (**C**) The clustering coefficient of the DLCN is higher than randomization test. The arrow represents the clustering coefficient in the real ceRNA network. (**D**) The characteristic path length of the DLCN is higher than randomization test. The arrow represents the characteristic path length in the real ceRNA network.

Table 1: The detailed information of dysregulated 7 lncRNAs with ceRNA activity in the DLCN

Ensembl ID	Gene name	Genomic location	Fold change	FDR
ENSG00000224660	SH3BP5-AS1	Chr 3: 15,254,184–15,264,493 (+)	0.42	0.09713
ENSG00000246859	STARD4-AS1	Chr 5: 111,512,226–111,739,726 (+)	2.47	0.09713
ENSG00000245311	ARNTL2-AS1	Chr 12: 27,389,789–27,446,625 (–)	2.78	0.09713
ENSG00000261312	AC002550.5	Chr 16: 19,706,351–19,715,383 (+)	6.63	0.09713
ENSG00000240618	RP11-206L10.5	Chr 1: 759,032–764,925(–)	2.21	0.09713
ENSG00000223947	AC016738.4	Chr 2: 100,993,676–101,002,244(–)	2.67	0.09713
ENSG00000255306	RP5-901A4.1	Chr 11: 68,024,809–68,030,461(–)	3.12	0.09713
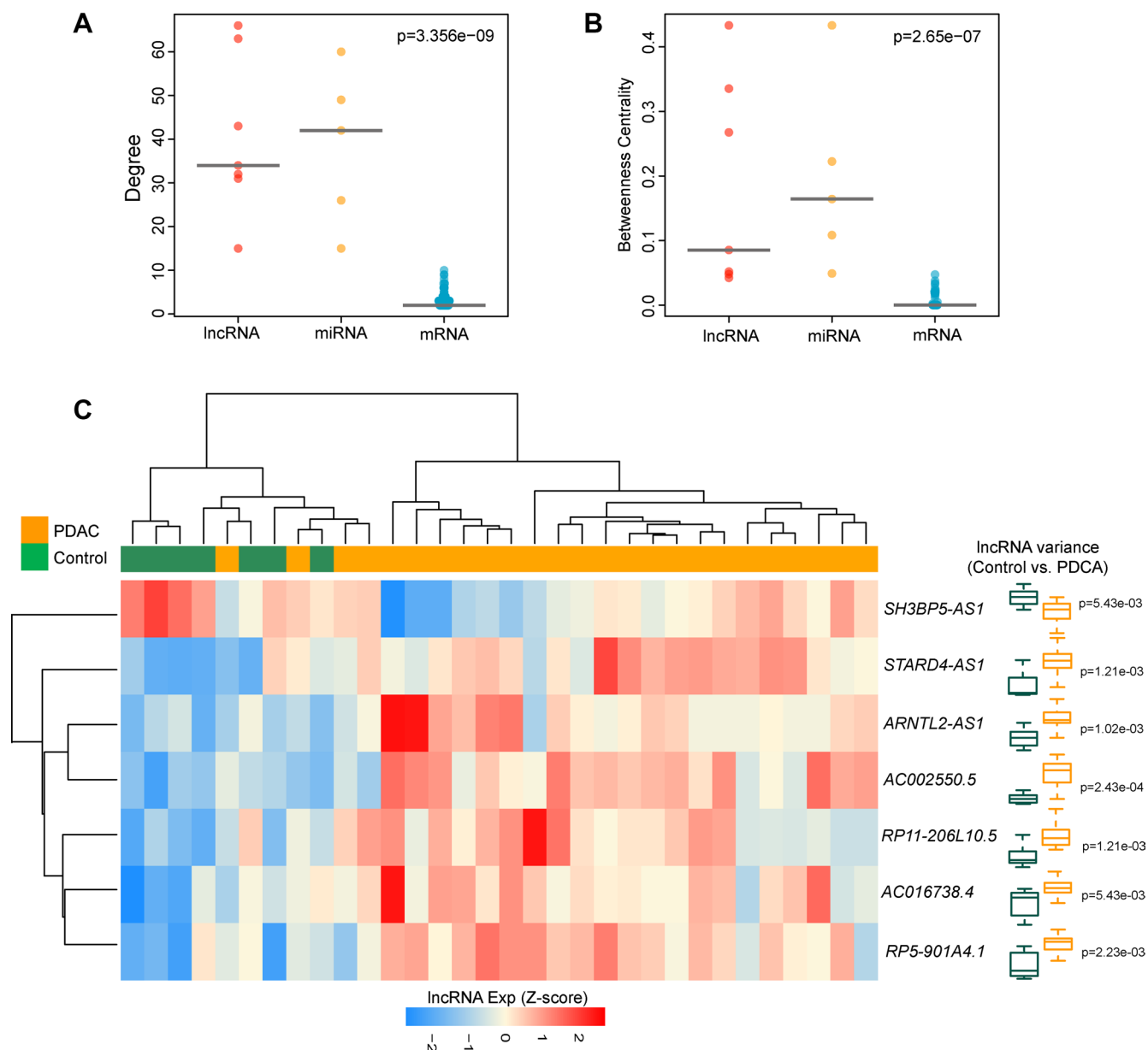
Figure 2: The lncRNA ceRNAs are more critical components compared to mRNA ceRNAs in the DLCN. (**A**) The difference of degree among lncRNAs, miRNAs and mRNAs. The lncRNA ceRNAs have significantly higher degrees than mRNA ceRNAs in the DLCN. (**B**) The difference of betweenness centrality among lncRNAs, miRNAs and mRNAs. The lncRNA ceRNAs had a higher betweenness centrality than mRNA ceRNAs in the DLCN. (**C**) The unsupervised hierarchical clustering heatmap of 32 samples based on the expression profiles of 7 lncRNA biomarkers in the discovery cohort.

(23/25, 84%) were subdivided into Cluster 2. The above results demonstrated that these 7 dysregulated lncRNAs in the DLCN might have a predictive role and could be used as biomarkers in the diagnosis of PDAC. Thus, we integrated these 7 lncRNA biomarkers to construct a 7-lncRNA signature (termed LncRisk-7) by developing an SVM classifier. The performance of the LncRisk-7 in distinguishing PDAC patients from nonmalignant pancreas samples was evaluated in the discovery cohort using the LOOCV procedure, in which 31 samples were used as training set and the remaining one was served as the test sample. Results of discovery cohort showed that the LncRisk-7 was able to correctly classify 24 out of 32 samples, achieving an overall predictive accuracy of 75% with a sensitivity of 84% and a specificity of 57.1%. The discriminatory performance of the LncRisk-7, evaluated by calculating the AUC and DOR, revealed that the AUC was 0.829 (Figure 3A and 3B) and the DOR was 3.938 (Figure 3C), which were significantly higher than those of randomization tests (Figure 3B and 3C). These results demonstrated that the LncRisk-7 had the better predictive performance for discriminating PDAC tumors from nonmalignant pancreas samples.

Further validation of the LncRisk-7 with two additional independent PDAC cohorts

To independently validate the diagnostic power of the LncRisk-7, we first test the discriminatory performance of the LncRisk-7 in an independent cohort of 52 samples from Pei's study [21]. For this, we first performed unsupervised hierarchical clustering analysis based on the expression profiles of 7 lncRNAs and found that 52 samples were grouped into two distinctive clusters based on expression patterns of seven lncRNAs, with significantly different tumor status ($p = 4.241e-05$, Chi-square test; Figure 4A). Furthermore, the LncRisk-7 efficiently distinguished PDAC patients from nonmalignant pancreas samples in the Pei cohort, achieving 76.9% prediction accuracy with 86.1% sensitivity and 43.8% specificity. The discriminatory power measured by the AUC and DOR was 0.833 and 7.97, respectively (Figure 4B).

Further validation of the discriminatory power of the LncRisk-7 was conducted using another completely independent cohort of 72 samples from Badea's study [22]. Results with unsupervised hierarchical clustering analysis were similar to those observed in the discovery cohort and Pei cohort above (Figure 5A). Most of PDAC patients (30/36, 83.3%) was clustered into one subgroup and 28 out of 36 (77.8%) normal samples was divided into another subgroup, indicating the significant association between expression patterns of 7 lncRNAs and tumor occurrence ($p = 7.144e-07$, Chi-square test; Figure 5A). Next, we performed classification of PDAC and control samples in the Badea cohort using the LncRisk-7, and found that the LncRisk-7 was able to correctly classify 26 out of 36 PDAC samples and 24 out of 36 control samples, resulting in 69.4 % accuracy, 72.2% sensitivity and 33.3% specificity. The discriminatory power measured by the AUC and DOR was 0.781 and 5.2, respectively (Figure 5B). Taken together, these results with additional independent validation cohorts demonstrated better and reproducible diagnostic performance of the LncRisk-7 in discriminating PDAC tumors from normal samples.

Functional implication of lncRNA biomarkers

To investigate the potential functional implication of the LncRisk-7, we performed functional enrichment analysis of GO and KEGG for mRNAs in the DLCN. Results of GO analysis revealed 39 enriched GO terms in the "Biological Process" (GOTERM-BP-FAT) ($p < 0.05$ and Fold Enrichment > 3.0) (Supplementary Table S3), which could be clustered into three functional sub-networks involved in cell death, cell adhesion and cell cycle (Figure 6A). KEGG analysis focusing on the biological pathways showed that these mRNAs as ceRNA interactors of lncRNA biomarkers in the LncRisk-7 were significantly enriched in several pathways involved in pathways in cancer, ECM-receptor interaction, cell adhesion molecules (CMAs) and adherens junction ($p < 0.05$ and Fold Enrichment > 3.0) (Figure 6B and Supplementary Table S3). These enriched biological processes and pathways have been reported to play important roles in PDAC pathogenesis, thus the LncRisk-7 might be involved with.

DISCUSSION

PDAC is one of the deadliest solid tumors characterized by complex molecular and cellular heterogeneity [4, 23]. During the past years, great efforts have been made to provide novel insights into the molecular mechanisms underlying PDAC, but the focus has been on protein-coding genes or miRNAs [24, 25]. A recently discovered non-coding RNA class, termed lncRNAs, has been widely reported to participate in a wide range of biological processes and their dysregulated expression is associated with many complicated human disease phenotypes including cancers [26–28]. There is increasing evidence that lncRNAs are extensively targeted by miRNAs, and function as ceRNAs. The crosstalk between ceRNAs occurred through competitively binding to shared miRNAs and formed a complex ceRNA-mediated regulatory network contributing to a novel dimension of post-transcriptional gene regulation [10, 11]. It has been shown that lncRNAs are key components of ceRNA-mediated regulatory network and their ceRNA activity has been implicated in both physiological conditions and cancer development [12, 29, 30]. Systematic analysis of ceRNA network has been performed in breast cancer [14, 31], gastric

cancer [16], glioblastoma multiforme [32] and ovarian cancer [15]. However, lncRNAs are known to have more developmental stage-, tissue-, organ- and disease-specific expression patterns than protein-coding genes, suggesting that lncRNA-associated ceRNA crosstalk will likely shift under different specific conditions and occur in a disease-specific manner. Therefore, it is critical for studying functional roles and regulatory mechanisms of lncRNA as ceRNAs in the development of PDAC, and investigating their potential implications for early diagnosis of PDAC.

In this study, we obtained genome-wide expression profiles of lncRNAs in PDAC patients and nonmalignant pancreas samples by repurposing three publicly available PDAC-related GEO cohorts and identified 78 differentially
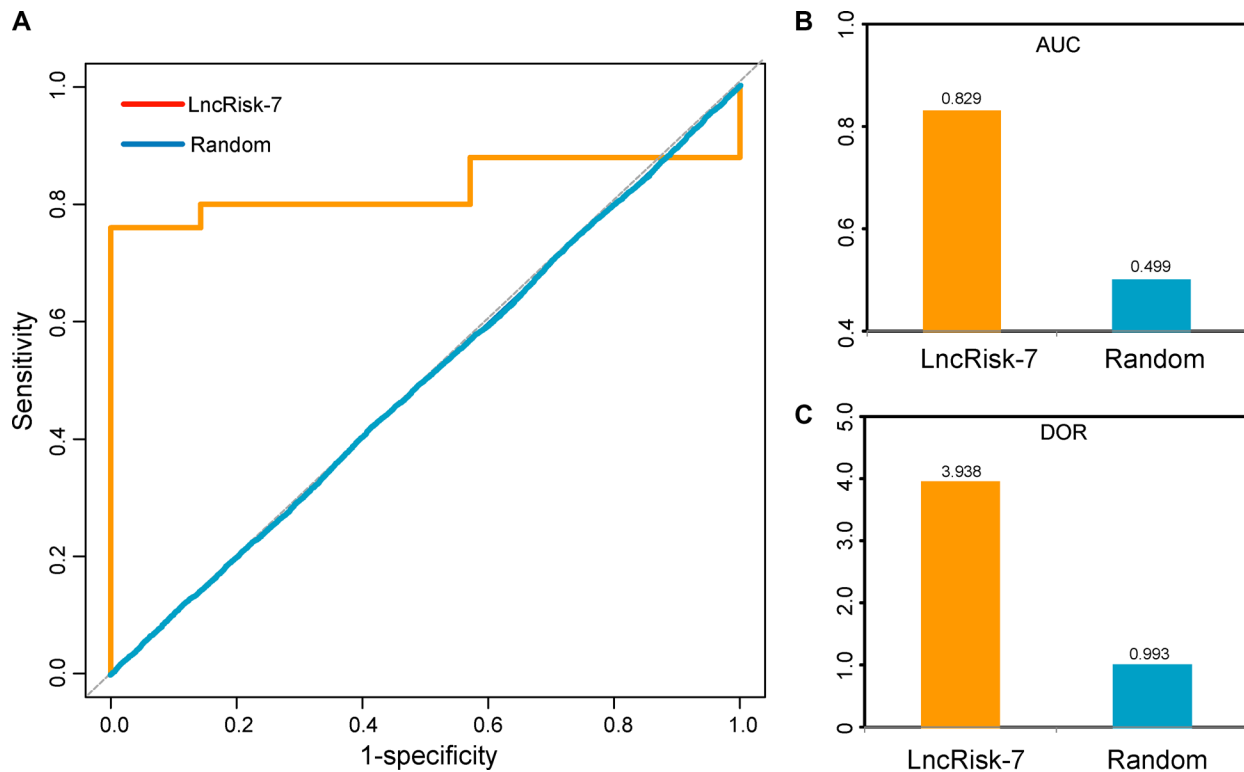

Figure 3: The performance of the LncRisk-7 in distinguishing PDAC patients from nonmalignant pancreas samples in the discovery cohort. (**A**) ROC analysis of the sensitivity and specificity of the LncRisk-7. (**B**) The AUC values of the LncRisk-7 based on leave-one-out cross validation. (**C**) The DOR values of the LncRisk-7 based on leave-one-out cross validation.
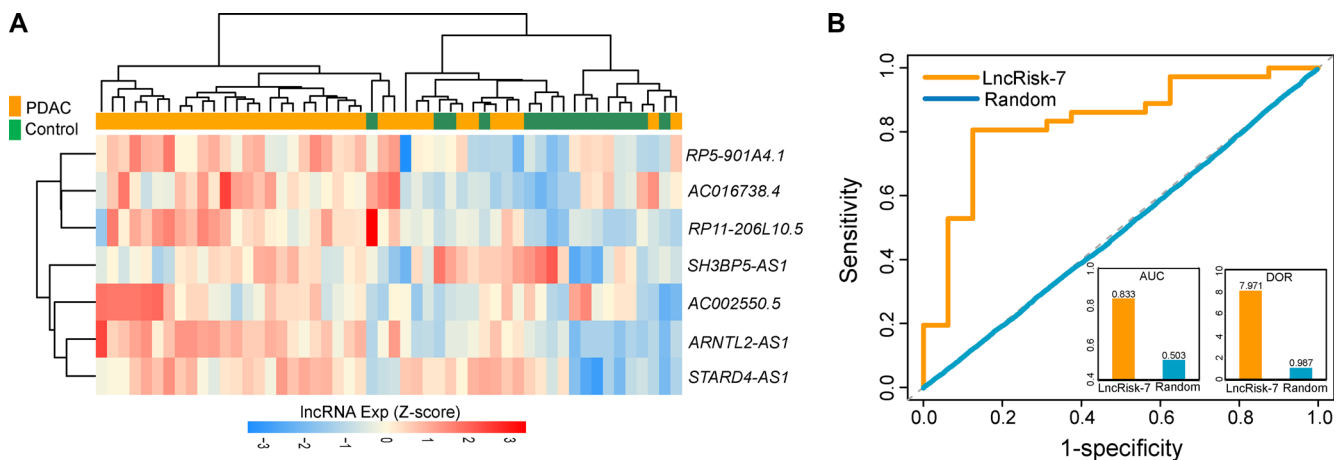

Figure 4: Independent validation of the LncRisk-7 for early diagnosis in another cohort of 52 samples from Pei's study. (**A**) Hierarchical clustering heatmap and dendrogram of 52 samples based expression patterns of 7 lncRNA biomarkers. (**B**) The discriminatory performance of the LncRisk-7, evaluated by ROC analysis and calculating the AUC and DOR.

expressed lncRNAs, implying that these lncRNAs may be associated with PDAC. In order to explore the ceRNA activity of lncRNAs in PDAC, we used an integrative computational framework to identify 290 dysregulated lncRNA-mediated ceRNA crosstalks through integrating sample-matched expression profiles of lncRNAs, miRNAs and mRNAs with miRNA-target regulatory information and constructed a lncRNA-mediated ceRNA regulatory network. Only 7 of 78 differentially expressed lncRNAs displayed ceRNA activity and communicated with 150 mRNAs by competing for 5 common miRNAs in PDAC. Network analysis showed that PDAC-specific DLCN was a scale-free and small world network as general biological networks, and the topological properties of DLCN were significantly distinguished from random networks, including high clustering coefficient and characteristic path length. We further examined the associations between mRNAs ceRNAs in the DLCN and cancers, and found that 150 mRNA ceRNAs in the DLCN were significantly enriched in the cancer class from The Genetic Association Database [33] (GAD, https://geneticassociationdb.nih.gov/) ($p = 5.0e$-03), in which 40 mRNAs are known to be related to cancers recorded in GAD (Supplementary Table S4). Previous studies have suggested that hub nodes, characterized by their high degree of connectivity to other nodes, play critical roles in a network and tend to be essential in network organization [34, 35]. In the DLCN, lncRNAs were observed to be topological key nodes whose degree and betweenness
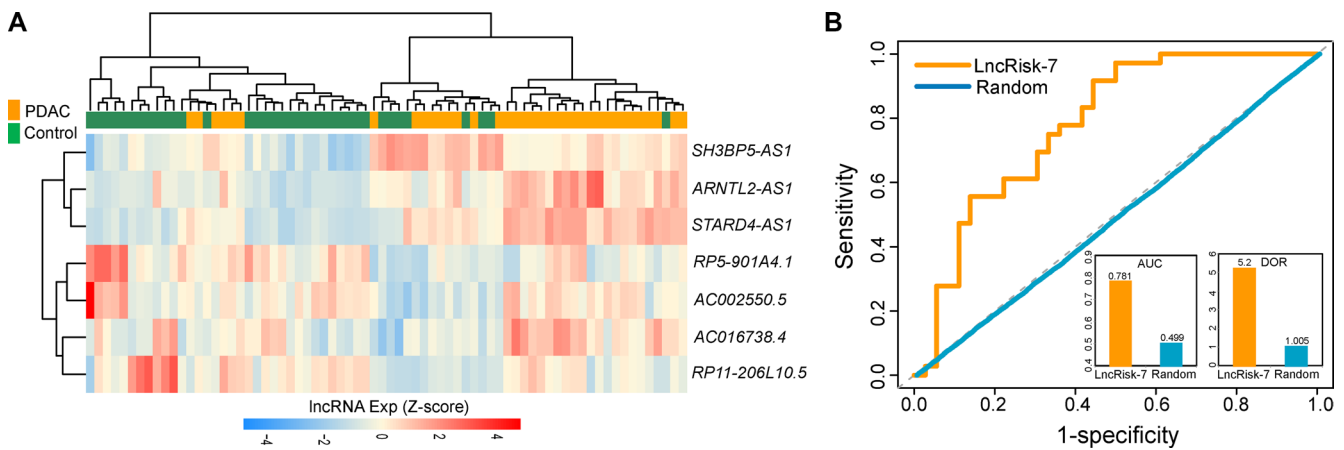
Figure 5: Further confirmation of the LncRisk-7 for early diagnosis in other independent cohort of 72 samples from Badea's study. (**A**) Hierarchical clustering heatmap and dendrogram of 72 samples based expression patterns of 7 lncRNA biomarkers. (**B**) The discriminatory performance of the LncRisk-7, evaluated by ROC analysis and calculating the AUC and DOR.
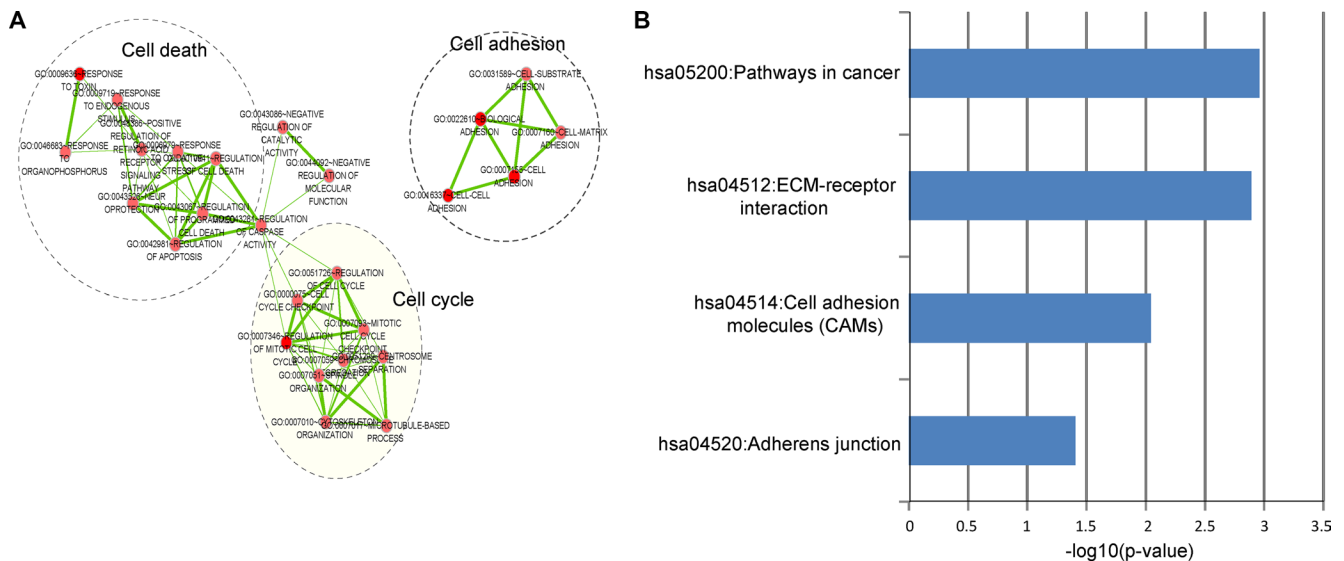
Figure 6: Functional analysis of the diagnostic lncRNAs. (**A**) The functional enrichment map of GO terms for mRNAs as ceRNA counterparts of lncRNA biomarkers. Each node represents a GO term and an edge represents the proportion of shared genes between connecting GO terms. (**B**) The enriched KEGG pathways ranked by –log10 (*p*-value).

centrality are significantly higher than mRNA ceRNAs, indicating that ceRNA activity of lncRNAs has profound implications for PDAC and the dysregulated expression of lncRNA ceRNAs will result in the extensive variation in tumorigenic or tumor-suppressive gene expression through lncRNA-mediated ceRNA regulatory network at the post-transcriptional level contributing to PDAC.

With increasing attention to the roles of lncRNAs as oncogenes and tumor suppressors in cancers, lncRNAs have exhibited superior potential as diagnostic and prognostic biomarkers than protein-coding genes owing to the more closely association between lncRNA expression and their functions [36, 37]. Recent studies have reported several lncRNA-focus signatures to improve prognosis prediction for some malignant tumors including breast cancer, lung cancer, colorectal cancer and so on [38–45]. However, the diagnostic role of lncRNAs in PDAC has not been investigated. Introducing graphic methods into biological systems can provide an intuitive vision and useful insights. It is particularly helpful to deal with complicated biological systems as demonstrated in recent some studies [46–48]. To explore whether 7 lncRNAs with ceRNA activity uncovered in the DLCN could become suitable diagnostic biomarkers for PDAC, we first performed unsupervised hierarchical clustering analysis and found that expression patterns of 7 lncRNAs could discriminate effectively between PDAC tumors and nonmalignant pancreas samples in the discovery cohort and two independent validation cohorts. In statistical prediction, the following three cross-validation methods are often used to examine a predictor for its effectiveness in practical application: independent dataset test, subsampling test, and jackknife test [49]. We adopted the leave one out cross-validation in this study as done by many investigators with SVM as the prediction engine and chose independent dataset test method to examine the accuracy of 7 lncRNAs in dignostic precdition. Therefore, we developed a 7-lncRNA signature based on the expression data of 7 lncRNAs and SVM algorithm as a novel diagnostic tool to improve early diagnosis of PDAC. The 7-lncRNA signature achieved a high performance in distinguishing PDAC patients from nonmalignant pancreas samples in the discovery cohort and was further confirmed in another two independent validation cohorts. These results demonstrated that the 7-lncRNA signature had the robust discriminatory power and may become a reliable and powerful predictor for early diagnosis of patients with PDAC.

As functional characterization of lncRNA is still in its infancy, only very few lncRNAs have been well functionally annotated until now. The ceRNA activity of these 7 lncRNAs was reported to be associated with PDAC for the first time and their function is unknown. It is increasing evidence that ceRNA crosstalk has become a powerful approach to infer functional roles of lncRNA with unknown function based on their ceRNA activity [27, 28, 50–52]. Therefore, as a preliminary exploration for the potential functional implication of the LncRisk-7,

we used 'guilt by association' principle to investigate biological processes and pathways regulated by the dysregulation of lncRNAs biomarkers and associated ceRNA crosstalk by functional enrichment analysis for mRNAs in the DLCN. We found that mRNAs as ceRNA counterparts of lncRNA biomarkers were involved in three GO biological processes (cell death, cell adhesion and cell cycle) and four KEGG pathways (pathways in cancer, ECM-receptor interaction, cell adhesion molecules (CMAs) and adherens junction). Three GO biological processes(cell death, cell adhesion and cell cycle) are well known to be associated with the hallmarks of cancer [53]. The extracellular matrix (ECM) not only provided mechanical and structural support but also played important roles in the development processes [54]. The dysregulated expression of ECM receptors contributing to the alterations of ECM formation and composition has been implicated in PDAC [55, 56]. Therefore, it is a plausible inference that the seven lncRNA biomarkers act as ceRNAs involving the regulation of cell death, cell adhesion and cell cycle. The dysregulation of these lncRNA ceRNAs and the resultant perturbation in lncRNA-mediated ceRNA network will have important effects in cancer-related biological processes contributing to tumorigenesis and progress of PDAC.

In summary, we performed a genome-wide analysis to investigate potential lncRNA-mediated ceRNA interplay by utilizing sample-matched miRNA, lncRNA and mRNA expression profiles in PDAC patients and normal samples in combination with miRNA regulatory network based on "ceRNA hypothesis". Then a PDAC-specific dysregulated lncRNA-mediated ceRNA network was constructed, which for the first time enables an overall view and analysis of lncRNA-associated ceRNA-mediated gene regulation in the development of PDAC on a system-wide level. This study will help to improve our understanding of lncRNA-mediated ceRNA regulatory mechanisms in the pathogenesis of PDAC and provide novel lncRNAs as candidate diagnostic biomarkers or potential therapeutic targets.

MATERIALS AND METHODS

Patient datasets

Three independent, nonoverlapping PDAC patient cohorts were used in this study. The initial discovery cohort of 25 PDAC samples and 7 nonmalignant pancreas samples with whole genome gene expression profiles and miRNA expression profiles were retrieved from Donahue' study [63] and were used to identify dysregulated lncRNA-mediated ceRNA interplay in PDAC. Another two PDAC patient cohorts composed of 52 samples (36 tumor samples and 16 normal samples, denoted by " Pei cohort") and 72 samples (36 tumor samples and 36 normal samples, denoted by " Badea cohort") only with

whole genome gene expression profiles were collected from Pei's study [21] and Badea's study [22], and were used as validation cohorts to test the diagnostic power of lncRNA biomarkers.

Acquisition and analysis of expression profiles

The sample-matched whole genome gene expression data and miRNA expression profiles data of 32 samples in the discovery cohort and the whole genome gene expression data of 52 samples in the Pei cohort and 72 samples in the Badea cohort were obtained from the public available GEO database (the GEO accession number is GSE32688, GSE16515 and GSE15471). Raw gene expression data profiled from Affymetrix Human Genome U133 Plus 2.0 Array (HG-U133_Plus_2.0) in the three patient cohorts were processed and normalized using the Robust Multichip Average (RMA) algorithm for background adjustment [64] and log-transformed (base 2). miRNA expression data produced by the Exiqon miRNA arrays (miRCURY LNA microRNA Array v.11.0 -hsa, mmu & rno) were adjusted and normalized by variance stabilizing transformation [63]. lncRNA expression data were obtained by repurposing the probes in the HG-U133_Plus_2.0 array to lncRNAs based on the annotation from the GENCODE project (http://www.gencodegenes.org, release 22) as previously described [36, 41, 65, 66]. Finally, expression data of 11005 protein-coding genes (PCGs), 826 miRNAs and 2330 lncRNAs were retained for further analysis.

Differential expression analysis of mRNAs, miRNAs and lncRNAs between PDAC samples and normal samples was carried out using the significance analysis of microarrays (SAM) method [67]. Unsupervised hierarchical clustering was used to investigate the expression pattern between PDAC samples and normal samples, and the Chi-square test or Fisher exact test was used to analyze the correlations between tumor status and lncRNA biomarkers.

MiRNA-mRNA and miRNA-lncRNA interaction data

The experimentally validated miRNA-mRNA interaction data was collected and integrated from TarBase (version 6.0) [68], miRTarBase (version 4.5) [69] and miRecords [70], including 37659 interactions between 402 miRNAs and 12360 PCGs. The putative interactions of miRNA-lncRNA were downloaded from lnCeDB database [71], including 1562845 interactions between 1394 miRNA and 28364 lncRNAs.

Construction and analysis of dysregulated lncRNA-associated ceRNA network

The dysregulated lncRNA-associated ceRNA network (DLCN) in PDAC was constructed based on

"ceRNA hypothesis" as follows: Firstly, expression correlation between differentially expressed mRNAs and differentially expressed lncRNAs was evaluated using Pearson correlation coefficient (PCC) from matched mRNA and lncRNA expression profiles data. Those dysregulated lncRNA-mRNA pairs with PCC > 0.5 and $p < 0.05$ were selected as co- dysregulated lncRNA-mRNA pairs. Then, the Pearson correlation coefficient between differentially expressed miRNAs and differentially expressed mRNAs, and between differentially expressed miRNAs and differentially expressed lncRNAs was computed from paired miRNA, mRNA and lncRNA expression profile data. For a given co-dysregulated lncRNA-mRNA pair, both mRNAs and lncRNAs in this pair are targeted and co-expressed negatively with a certain common miRNA, this miRNA-mRNA-lncRNA was identified as dysregulated competing triplets (DCTs). Finally, a DLCN was built for PDAC by assembling all DCTs identified above.

The network characteristics of DLCN, including degree, characteristic path length (CPL), clustering coefficient (CC) and betweenness centrality were analyzed.

Development of lncRNA-based signature in PDAC diagnosis

For classification of PDAC vs. normal samples, lncRNA biomarkers were integrated to form a lncRNA-focus signature using support vector machine (SVM) with the sigmoid kernel. An unbiased performance estimate in identifying PDAC patients was carried out using leave one out cross-validation (LOOCV). Diagnostic ability of the lncRNA-focus prediction model was evaluated by obtaining the area under a receiver operating characteristic (ROC) curve (AUC) and diagnostic odds ratio (DOR). The ROC curve was produced by plotting true positive rates (sensitivity) against false positive rates (1-specificity). The DORs were calculated as follows:

$$DOR = \frac{Sensitivity \times Specificity}{(1 - Sensitivity) \times (1 - Specificity)}.$$

The permutation p-value of AUC and DOR was obtained from 1,000 randomization tests by randomizing lncRNA expression data for testing the null hypothesis.

Functional analysis of lncRNA biomarkers

Functional enrichment analysis of Gene Ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) for mRNAs in the DLCN was performed to infer potential biological processes and pathways of lncRNA biomarkers using DAVID Bioinformatics Tool (version 6.7) [72] limited to GO terms in the "Biological Process"(GOTERM-BP-FAT) and KEGG pathway categories. The biological processes and pathways with p-value of < 0.05 and an enrichment score of > 3.0

using the whole human genome as background were considered as significant functional categories which were organized into an interaction network with similar functions using the Enrichment Map plugin in Cytoscape environment [73].

CONFLICTS OF INTEREST

The authors declare that they have no of interest.

REFERENCES

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA Cancer J Clin. 2013; 63:11–30.

2. Li D, Xie K, Wolff R, Abbruzzese JL. Pancreatic cancer. Lancet. 2004; 363:1049–1057.

3. Stotz M, Eisner F, Szkandera J, Absenger G, Kornprat P, Lackner C, Samonigg H, Gerger A, Pichler M. Clinico-pathological characteristics and clinical outcome of different histological types of pancreatic cancer in a large Middle European series. J Clin Pathol. 2013; 66:753–757.

4. Samuel N, Hudson TJ. The molecular and cellular heterogeneity of pancreatic ductal adenocarcinoma. Nat Rev Gastroenterol Hepatol. 2012; 9:77–87.

5. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, Kodzius R, Shimokawa K, Bajic VB, et al. The transcriptional landscape of the mammalian genome. Science. 2005; 309:1559–1563.

6. Consortium EP, Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007; 447:799–816.

7. Shi X, Sun M, Wu Y, Yao Y, Liu H, Wu G, Yuan D, Song Y. Post-transcriptional regulation of long noncoding RNAs in cancer. Tumour Biol. 2015; 36:503–513.

8. Kornienko AE, Guenzl PM, Barlow DP, Pauler FM. Gene regulation by the act of long non-coding RNA transcription. BMC Biol. 2013; 11:59.

9. Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. Nat Rev Genet. 2014; 15:7–21.

10. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. Nature. 2014; 505:344–352.

11. Tan JY, Sirey T, Honti F, Graham B, Piovesan A, Merkenschlager M, Webber C, Ponting CP, Marques AC. Extensive microRNA-mediated crosstalk between lncRNAs and mRNAs in mouse embryonic stem cells. Genome Res. 2015; 25:655–666.

12. Karreth FA, Pandolfi PP. ceRNA cross-talk in cancer: when ce-bling rivalries go awry. Cancer Discov. 2013; 3:1113–1121.

13. Wang J, Liu X, Wu H, Ni P, Gu Z, Qiao Y, Chen N, Sun F, Fan Q. CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. Nucleic Acids Res. 2010; 38:5366–5383.

14. Paci P, Colombo T, Farina L. Computational analysis identifies a sponge interaction network between long non-coding RNAs and messenger RNAs in human breast cancer. BMC Syst Biol. 2014; 8:83.

15. Zhou M, Wang X, Shi H, Cheng L, Wang Z, Zhao H, Yang L, Sun J. Characterization of long non-coding RNA-associated ceRNA network to reveal potential prognostic lncRNA biomarkers in human ovarian cancer. Oncotarget. 2016; 7:12598–12611. doi: 10.18632/oncotarget.7181.

16. Xia T, Liao Q, Jiang X, Shao Y, Xiao B, Xi Y, Guo J. Long noncoding RNA associated-competing endogenous RNAs in gastric cancer. Sci Rep. 2014; 4:6088.

17. Wang Y, Li Z, Zheng S, Zhou Y, Zhao L, Ye H, Zhao X, Gao W, Fu Z, Zhou Q, Liu Y, Chen R. Expression profile of long non-coding RNAs in pancreatic cancer and their clinical significance as biomarkers. Oncotarget. 2015; 6:35684–35698. doi: 10.18632/oncotarget.5533.

18. Liu JH, Chen G, Dang YW, Li CJ, Luo DZ. Expression and prognostic significance of lncRNA MALAT1 in pancreatic cancer tissues. Asian Pac J Cancer Prev. 2014; 15:2971–2977.

19. Li X, Deng SJ, Zhu S, Jin Y, Cui SP, Chen JY, Xiang C, Li QY, He C, Zhao SF, Chen HY, Niu Y, Liu Y, et al. Hypoxia-induced lncRNA-NUTF2P3–001 contributes to tumorigenesis of pancreatic cancer by derepressing the miR-3923/KRAS pathway. Oncotarget. 2016; 7:6000–6014. doi: 10.18632/oncotarget.6830.

20. Chou KC. Some remarks on protein attribute prediction and pseudo amino acid composition. Journal of theoretical biology. 2011; 273:236–247.

21. Pei H, Li L, Fridley BL, Jenkins GD, Kalari KR, Lingle W, Petersen G, Lou Z, Wang L. FKBP51 affects cancer cell response to chemotherapy by negatively regulating Akt. Cancer Cell. 2009; 16:259–266.

22. Badea L, Herlea V, Dima SO, Dumitrascu T, Popescu I. Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia. Hepatogastroenterology. 2008; 55:2016–2027.

23. Bardeesy N, DePinho RA. Pancreatic cancer biology and genetics. Nat Rev Cancer. 2002; 2:897–909.

24. Zhang L, Jamaluddin MS, Weakley SM, Yao Q, Chen C. Roles and mechanisms of microRNAs in pancreatic cancer. World J Surg. 2011; 35:1725–1731.

25. Hong SM, Park JY, Hruban RH, Goggins M. Molecular signatures of pancreatic cancer. Arch Pathol Lab Med. 2011; 135:716–727.

26. Cheetham SW, Gruhl F, Mattick JS, Dinger ME. Long noncoding RNAs and the genetics of cancer. Br J Cancer. 2013; 108:2419–2425.

27. Zhou M, Wang X, Li J, Hao D, Wang Z, Shi H, Han L, Zhou H, Sun J. Prioritizing candidate disease-related long non-coding RNAs by walking on the heterogeneous lncRNA and disease network. Mol Biosyst. 2015; 11:760–769.

28. Sun J, Shi H, Wang Z, Zhang C, Liu L, Wang L, He W, Hao D, Liu S, Zhou M. Inferring novel lncRNA-disease associations based on a random walk model of a lncRNA functional similarity network. Mol Biosyst. 2014; 10:2074–2081.

29. Wang Y, Hou J, He D, Sun M, Zhang P, Yu Y, Chen Y. The Emerging Function and Mechanism of ceRNAs in Cancer. Trends Genet. 2016; 32:211–224.

30. Sanchez-Mejias A, Tay Y. Competing endogenous RNA networks: tying the essential knots for cancer biology and therapeutics. J Hematol Oncol. 2015; 8:30.

31. Zhou X, Liu J, Wang W. Construction and investigation of breast-cancer-specific ceRNA network based on the mRNA and miRNA expression data. IET Syst Biol. 2014; 8:96–103.

32. Chiu YC, Hsiao TH, Chen Y, Chuang EY. Parameter optimization for constructing competing endogenous RNA regulatory network in glioblastoma multiforme and other cancers. BMC Genomics. 2015; 16 Suppl 4:S1.

33. Becker KG, Barnes KC, Bright TJ, Wang SA. The genetic association database. Nat Genet. 2004; 36:431–432.

34. He X, Zhang J. Why do hubs tend to be essential in protein networks? PLoS Genet. 2006; 2:e88.

35. Gursoy A, Keskin O, Nussinov R. Topological properties of protein interaction networks from a structural perspective. Biochem Soc Trans. 2008; 36:1398–1403.

36. Du Z, Fei T, Verhaak RG, Su Z, Zhang Y, Brown M, Chen Y, Liu XS. Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. Nat Struct Mol Biol. 2013; 20:908–913.

37. Hauptman N, Glavac D. Long non-coding RNA in cancer. Int J Mol Sci. 2013; 14:4655–4669.

38. Sun J, Cheng L, Shi H, Zhang Z, Zhao H, Wang Z, Zhou M. A potential panel of six-long non-coding RNA signature to improve survival prediction of diffuse large-B-cell lymphoma. Sci Rep. 2016; 6:27842.

39. Zhou M, Sun Y, Sun Y, Xu W, Zhang Z, Zhao H, Zhong Z, Sun J. Comprehensive analysis of lncRNA expression profiles reveals a novel lncRNA signature to discriminate nonequivalent outcomes in patients with ovarian cancer. Oncotarget. 2016; 7:32433–48. doi: 10.18632/oncotarget.8653.

40. Sun J, Chen X, Wang Z, Guo M, Shi H, Wang X, Cheng L, Zhou M. A potential prognostic long non-coding RNA signature to predict metastasis-free survival of breast cancer patients. Sci Rep. 2015; 5:16553.

41. Zhou M, Guo M, He D, Wang X, Cui Y, Yang H, Hao D, Sun J. A potential signature of eight long non-coding RNAs predicts survival in patients with non-small cell lung cancer. J Transl Med. 2015; 13:231.

42. Zhou M, Xu W, Yue X, Zhao H, Wang Z, Shi H, Cheng L, Sun J. Relapse-related long non-coding RNA signature to improve prognosis prediction of lung adenocarcinoma. Oncotarget. 2016; 7:29720–38. doi: 10.18632/oncotarget.8825.

43. Hu Y, Chen HY, Yu CY, Xu J, Wang JL, Qian J, Zhang X, Fang JY. A long non-coding RNA signature to improve prognosis prediction of colorectal cancer. Oncotarget. 2014; 5:2230–2242. doi: 10.18632/oncotarget.1895.

44. Zhang XQ, Sun S, Lam KF, Kiang KM, Pu JK, Ho AS, Lui WM, Fung CF, Wong TS, Leung GK. A long non-coding RNA signature in glioblastoma multiforme predicts survival. Neurobiol Dis. 2013; 58:123–131.

45. Li J, Chen Z, Tian L, Zhou C, He MY, Gao Y, Wang S, Zhou F, Shi S, Feng X, Sun N, Liu Z, Skogerboe G, et al. LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. Gut. 2014; 63:1700–1710.

46. Chou KC. Applications of graph theory to enzyme kinetics and protein folding kinetics. Steady and non-steady-state systems. Biophysical chemistry. 1990; 35:1–24.

47. Xiao X, Shao S, Ding Y, Huang Z, Chen X, Chou KC. An application of gene comparative image for predicting the effect on replication ratio by HBV virus gene missense mutation. Journal of theoretical biology. 2005; 235:555–565.

48. Chou KC. Graphic rule for drug metabolism systems. Current drug metabolism. 2010; 11:369–378.

49. Chou KC, Zhang CT. Prediction of protein structural classes. Critical reviews in biochemistry and molecular biology. 1995; 30:275–349.

50. Cheng L, Shi H, Wang Z, Hu Y, Yang H, Zhou C, Sun J, Zhou M. IntNetLncSim: an integrative network analysis method to infer human lncRNA functional similarity. Oncotarget. 2016; 7:47864–47874. doi: 10.18632/oncotarget.10012.

51. Zhou M, Han L, Zhang J, Hao D, Cai Y, Wang Z, Zhou H, Sun J. A computational frame and resource for understanding the lncRNA-environmental factor associations and prediction of environmental factors implicated in diseases. Mol Biosyst. 2014; 10:3264–3271.

52. Liu K, Yan Z, Li Y, Sun Z. Linc2GO: a human LincRNA function annotation resource based on ceRNA hypothesis. Bioinformatics. 2013; 29:2221–2222.

53. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000; 100:57–70.

54. Kim SH, Turnbull J, Guimond S. Extracellular matrix and cell signalling: the dynamic cooperation of integrin, proteoglycan and growth factor receptor. J Endocrinol. 2011; 209:139–151.

55. Gress TM, Menke A, Bachem M, Muller-Pillasch F, Ellenrieder V, Weidenbach H, Wagner M, Adler G. Role of extracellular matrix in pancreatic diseases. Digestion. 1998; 59:625–637.

56. Apte MV, Yang L, Phillips PA, Xu Z, Kaplan W, Cowley M, Pirola RC, Wilson JS. Extracellular matrix composition significantly influences pancreatic stellate cell gene expression pattern: role of transgelin in PSC function. Am J Physiol Gastrointest Liver Physiol. 2013; 305:G408–417.

57. Xiao X, Ye HX, Liu Z, Jia JH, Chou KC. iROS-gPseKNC: Predicting replication origin sites in DNA by incorporating dinucleotide position-specific propensity into general pseudo nucleotide composition. Oncotarget. 2016; 7: 34180–9. doi: 10.18632/oncotarget.9057.

58. Qiu WR, Xiao X, Xu ZC, Chou KC. iPhos-PseEn: Identifying phosphorylation sites in proteins by fusing different pseudo components into an ensemble classifier. Oncotarget. 2016; 7:51270–51283. doi: 10.18632/oncotarget.9987.

59. Qiu WR, Sun BQ, Xiao X, Xu ZC, Chou KC. iHyd-PseCp: Identify hydroxyproline and hydroxylysine in proteins by incorporating sequence-coupled effects into general PseAAC. Oncotarget. 2016; 7:44310–44321. doi: 10.18632/oncotarget.10027.

60. Jia J, Liu Z, Xiao X, Liu B, Chou KC. iCar-PseCp: identify carbonylation sites in proteins by Monto Carlo sampling and incorporating sequence coupled effects into general PseAAC. Oncotarget. 2016; 7:34558–70. doi: 10.18632/oncotarget.9148.

61. Lin H, Deng EZ, Ding H, Chen W, Chou KC. iPro54-PseKNC: a sequence-based predictor for identifying sigma-54 promoters in prokaryote with pseudo k-tuple nucleotide composition. Nucleic Acids Res. 2014; 42:12961–12972.

62. Chou KC. Impacts of bioinformatics to medicinal chemistry. Medicinal chemistry. 2015; 11:218–234.

63. Donahue TR, Tran LM, Hill R, Li Y, Kovochich A, Calvopina JH, Patel SG, Wu N, Hindoyan A, Farrell JJ, Li X, Dawson DW, Wu H. Integrative survival-based molecular profiling of human pancreatic cancer. Clin Cancer Res. 2012; 18:1352–1363.

64. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003; 4:249–264.

65. Sorensen KP, Thomassen M, Tan Q, Bak M, Cold S, Burton M, Larsen MJ, Kruse TA. Long non-coding RNA expression profiles predict metastasis in lymph node-negative breast cancer independently of traditional prognostic markers. Breast Cancer Res. 2015; 17:55.

66. Zhou M, Zhao H, Wang Z, Cheng L, Yang L, Shi H, Yang H, Sun J. Identification and validation of potential prognostic lncRNA biomarkers for predicting survival in patients with multiple myeloma. J Exp Clin Cancer Res. 2015; 34:102.

67. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A. 2001; 98:5116–5121.

68. Vergoulis T, Vlachos IS, Alexiou P, Georgakilas G, Maragkakis M, Reczko M, Gerangelos S, Koziris N, Dalamagas T, Hatzigeorgiou AG. TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. Nucleic Acids Res. 2012; 40:D222–229.

69. Hsu SD, Tseng YT, Shrestha S, Lin YL, Khaleel A, Chou CH, Chu CF, Huang HY, Lin CM, Ho SY, Jian TY, Lin FM, Chang TH, et al. miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. Nucleic Acids Res. 2014; 42:D78–85.

70. Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. miRecords: an integrated resource for microRNA-target interactions. Nucleic Acids Res. 2009; 37:D105–110.

71. Das S, Ghosal S, Sen R, Chakrabarti J. lnCeDB: database of human long noncoding RNA acting as competing endogenous RNA. PLoS One. 2014; 9:e98965.

72. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009; 37:1–13.

73. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One. 2010; 5:e13984.