Research Paper

Protein-coding genes, long non-coding RNAs combined with microRNAs as a novel clinical multi-dimension transcriptome signature to predict prognosis in ovarian cancer

Xu Meng[1,*], Guo Jin-Cheng[2], Zhang Jue[1], Ma Quan-Fu[1], Yan Bin[1] and Wu Xu-Feng[1]

[1]Department of Gynecology, Maternal and Child Health Hospital of Hubei Province, Wuhan, China

[2]Department of Biochemistry and Molecular Biology, Shantou University Medical College, Shantou, China

[*]First author

Correspondence to: Wu Xu-Feng, *email:* zwuxufeng@163.com

ABSTRACT

Ovarian cancer is prevalent in women which is usually diagnosed at an advanced stage with a high mortality rate. The aim of this study is to investigate protein-coding gene, long non-coding RNA, and microRNA associated with the prognosis of patients with ovarian serous carcinoma by mining data from TCGA (The Cancer Genome Atlas) public database. The clinical data of ovarian serous carcinoma patients was downloaded from TCGA database in September, 2016. The mean age and survival time of 407 patients with ovarian serous carcinoma were 59.71 ± 11.54 years and 32.98 ± 26.66 months. Cox's proportional hazards regression analysis was conducted to analyze genes that were significantly associated with the survival of ovarian serous carcinoma patients in the training group. Using the random survival forest algorithm, Kaplan–Meier and ROC analysis, we kept prognostic genes to construct the multi-dimensional transcriptome signature with max area under ROC curve (AUC) (0.69 in the training group and 0.62 in the test group). The selected signature composed by *VAT1L*, *CALR*, LINC01456, RP11-484L8.1, MIR196A1 and MIR148A, separated the training group patients into high-risk or low-risk subgroup with significantly different survival time (median survival: 35.3 months *vs.* 64.9 months, *P* < 0.001). The signature was validated in the test group showing similar prognostic values (median survival: 41.6 months in high-risk *vs.* 57.4 months in low-risk group, *P*=0.018). Chi-square test and multivariable Cox regression analysis showed that the signature was an independent prognostic factor for patients with ovarian serous carcinoma. Finally, we validated the expression of the genes experimentally.

INTRODUCTION

Ovarian cancer (OC) is a deadly female reproductive cancer, accounting for 5% of female cancer deaths [1]. The majority of women with ovarian cancer are always diagnosed in an advanced stage, which substantially increases the risk of early death [2, 3]. Despite advances in imaging diagnosis, preoperative and postoperative care and chemotherapy, there has been little improvement in overall survival [4–6]. Identification of clinical markers in OC is of significance to early diagnosis, select appropriate treatment and improve prognosis of patients with OC.

As the development of high-throughput sequencing technology, attempts have been made to identify molecular markers from sequencing data that affect clinical outcomes by integrating multiple profiles and clinical data [7–10]. A number of studies have shown that protein-coding genes (PCGs) involved in the many important

Table 1: Summary of patient demographics and clinical characteristics

Characteristic	Training set	Testing set	Total
Age			
Median	59	58	59
Range	38~85	30~87	30~87
Clinical stage			
Stage I	0	0	0
Stage II	8	13	21
Stage III	163	158	321
Stage IV	31	31	62
Vital status			
Living	91	88	179
Dead	112	116	228

biological processes and could be powerful predictors of tumor staging for patients in different cancers. A five-gene (*CKAP4*, *SLC40A1*, *OTOF*, *MAN2A2* and *ISPD*) signature is significantly related to patient survival in renal clear cell carcinoma patients from The Cancer Genome Atlas (TCGA) database [11]. A prognostic 7-Gene (*NHLRC3*, *ZDHHC21*, *PRR14L*, *CCBL1*, *PTPRB*, *PNPO*, and *PPIP5K2*) expression signature for stage III is constructed in colorectal cancer [12]. Recent years, long non-coding RNAs (lncRNAs) have become new players in tumorigenesis and tumor progression with an important clinical significance in prognosis due to their gene regulation function at the transcriptional, posttranscriptional and epigenetic levels [13, 14]. The well-known lncRNA named *HOTAIR* is significantly associated with breast cancer metastasis [15]. *GAS5* and *Yiya* are promising prognostic biomarkers of liver metastasis for early stage colorectal cancer patients [16]. Two immune-associated lncRNA biomarkers (RP11-284N8.3.1 and AC104699.1.1) could independently predict the survival of patients with different ovarian cancer stages [17]. Another lncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma [10]. Researchers have identified key lncRNAs associated with distinct stages of OC progression using a ceRNA-network driven method, and developed a ten-lncRNA signature to predict the clinical outcome of OC [18]. An eight-lncRNA signature has been found by a comprehensive analysis for lncRNA expression profile and clinical outcome of a large number of OC patients from TCGA [19]. Apart from PCGs, lncRNAs and microRNAs (miRNAs, miRs), a class of small noncoding RNAs of 18–25 nucleotides are thought to inhibit gene expression post-transcriptionally by causing mRNA degradation and/or repressing mRNA

translation [20]. MiRNAs are frequently found their dysregulated expression in multiple cancers, and may function as both oncogenes and tumor suppressors [21]. *BCL11A* overexpression modulated by microRNA-30a could predict survival and relapse in non-small cell lung cancer [22]. Several prognostic and predictive microRNA markers have been identified for many types of cancers, for instance, breast cancer, hepatocellular carcinoma, lung cancer, glioma and colorectal cancer [23–29].

In summary, protein-coding genes, long non-coding RNAs and microRNAs have the prognostic potential. Therefore, the combination of PCGs, lncRNAs and microRNAs could show the clinical outcome alteration of patients with ovarian cancer more elaborately since it revealed difference in multiple transcriptome dimension. Here, we obtained the expression level of PCGs, lncRNAs and micoRNAs from a large dataset (n=407) in TCGA database and reported the first clinical multi-dimension transcriptome molecular signature which had the ability to predict prognosis in ovarian serous carcinoma patients.

RESULTS

Patient characteristics

All 407 patients used in this study were clinically and pathologically diagnosed with ovarian serous carcinoma. According to the International Federation of Gynecology and Obstetrics (FIGO) classification, clinical stages of the tumor were classified into stages I to IV. In our study, there were 0, 21, 321 and 62 patients in stage I, II, III and IV, respectively. Patients with missing data were not included in the study. All the other statistical information was summarized in Table 1. We divided the downloaded datasets into training and test

groups randomly using the algorithm called "sample" of R program.

Identification of three PCGs, three lncRNAs and three microRNAs associated with survival from the training group

We filtered out gene expression data to generate new PCGs, lncRNAs and microRNAs expression profiles (see method). Then 15426 PCGs, 8335 lncRNAs and 340 microRNAs expression values were obtained from the TCGA and TANRIC databases.

The training group (n = 203) including a relatively large patient sample size and relatively complete clinical information, were used to explore the association of survival with PCGs, lncRNAs and microRNAs. Firstly, we conducted a univariate Cox proportional hazards regression analysis of the PCGs, lncRNAs and microRNAs expression profiling data with survival time and survival status as the dependent variable, and identified a 1061-PCGs-lncRNAs-microRNAs set composed by 730 PCGs, 313 lncRNAs and 18 microRNAs which were significantly correlated with patients' OS (P value <0.05,

Figure 2A. Supplementary Table 2). Secondly, using random forest supervised classification algorithm, two PCGs, two lncRNAs and two microRNAs mostly related to the prognostic classification were selected among the 1061-PCGs-lncRNAs-microRNAs set according to the permutation important score by random survival forests-variable hunting (RSFVH) algorithm (Supplementary Figure 1).

Acquisition of the prognostic PCGs-lncRNAs-microRNAs signature in the training dataset

The PCGs-lncRNAs-microRNAs set in the training dataset could have $2^9-1=511$ combination and corresponding risk score according to the signature based risk-score model (Supplementary Table 1). In order to select a better prognostic signature, we performed time-dependent ROC curve. All the risk scores of the patients were calculated as the methods described. Then the PCGs-lncRNAs-microRNAs combination composed by *VAT1L*, *CALR*, LINC01456, RP11-484L8.1, MIR196A1 and MIR148A with the max AUC was selected (Figure 2B, Table 2). The risk score of the combination composed by *VAT1L*, *CALR*,
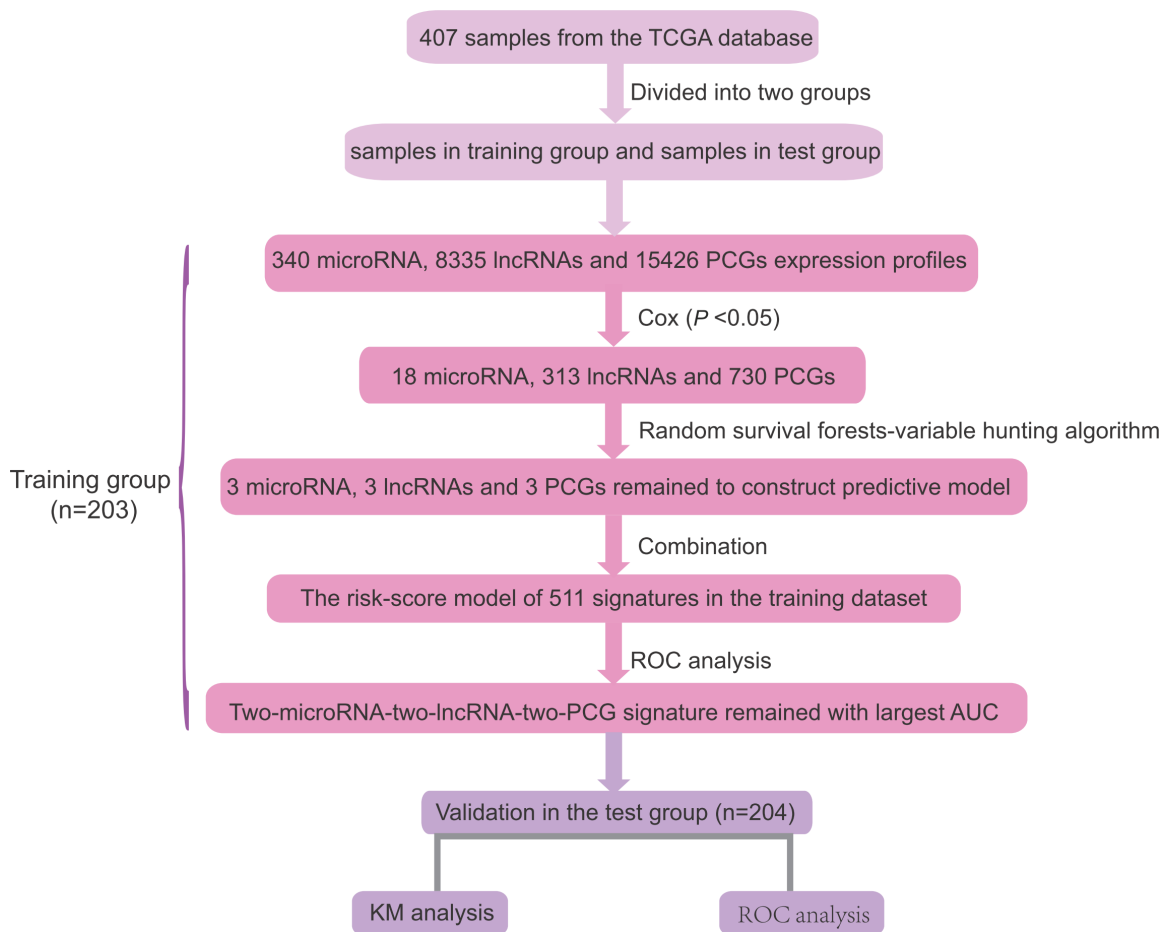

Figure 1: Schedule of the study. The order of analyses to develop the risk score model and validate the efficiency of the signature to predict prognostic outcomes.

LINC01456, RP11-484L8.1, MIR196A1 and MIR148A was got as follows: Risk score = (0.47 × expression value of *VAT1L*) + (-0.52 × expression value of *CALR*) + (-0.66 × expression value of LINC01456) + (2.56 × expression value of RP11-484L8.1) + (0.14 × expression value of MIR196A1) + (-0.19 × expression value of RP11-484L8.1) + (0.14 × expression value of MIR148A). AUC of the PCGs-lncRNAs-microRNAs signature in the prognostic model was 0.69 (Figure 2B) demonstrating its good performance for survival prediction.

Validation of the prediction performance of the PCGs-lncRNAs-microRNAs signature in the training dataset and the test dataset.

The training group patients were divided into either the high-risk group (n =102) or low-risk group (n = 101) using the median risk score as the cutoff point. Patients in the high-risk group had a significantly shorter OS than those in the low-risk group (median survival: 35.3 months *vs.* 64.9 months, log-rank test *P* < 0.001; Figure 3A, *left*). OS rates of
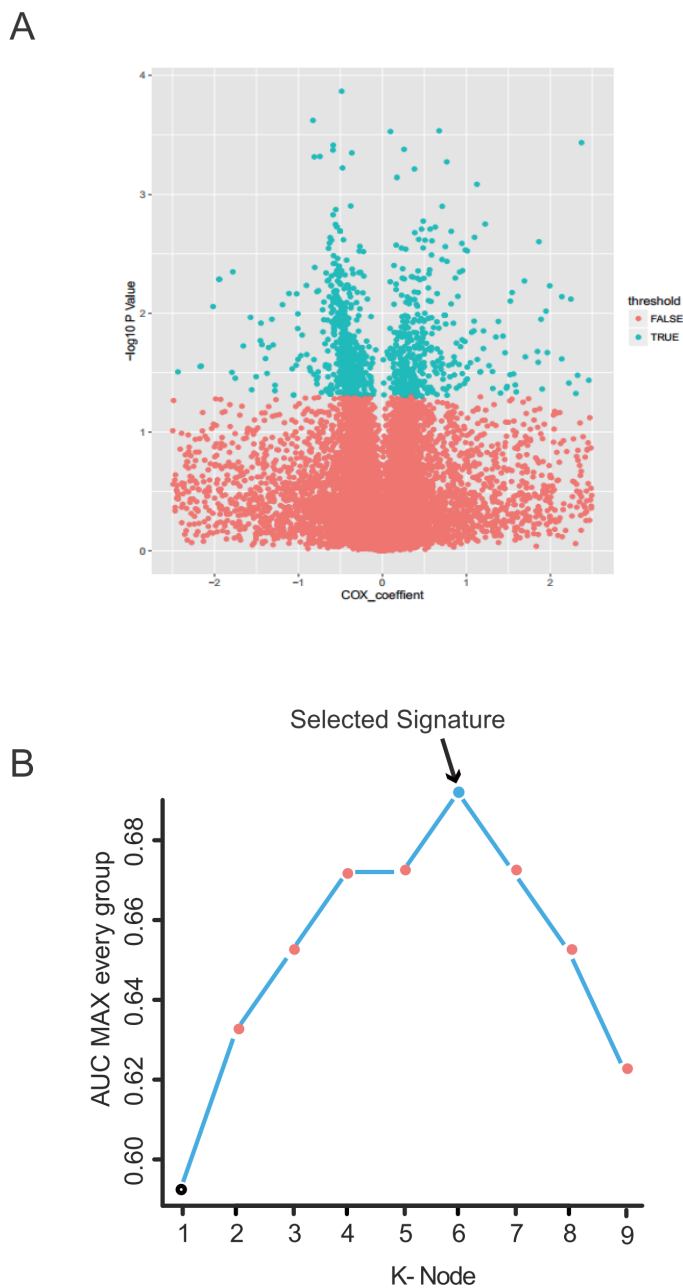
Figure 2: Identification of the PCGs-lncRNAs-microRNAs signature in the training dataset. (A) Univariate Cox proportional hazards regression analysis of the PCGs, lncRNAs and microRNAs expression profiling data in the training dataset. **(B)** The procedure for identifying the final signature. The accuracies of all 511 signatures were calculated and the nine highest accuracies for k=1, 2...... 9 were shown in the plot.

Table 2: Identities of PCGs, lncRNAs and microRNAs in the prognostic expression signature and their univariable cox association with prognosis

Ensembl ID	Gene symbol	Gene name	Coefficient[a]	*P* value [a]	Gene expression level association with poor prognosis	Chromosome location
ENSG00000171724[b]	VAT1L	Vesicle amine transport 1 like	0.47	0.01	high	chr16:77788530-77980107:[+]
ENSG00000179218[b]	CALR	Calreticulin	-0.52	0.00	low	chr19:12938578-12944489:[+]
ENSG00000225882[b]	LINC01456		-0.66	0.05	low	chrX:17970197-18104644:[-]
ENSG00000267764[b]	RP11-484L8.1		2.56	0.01	high	chr18:48826051-48834770:[-]
hsa-miR-196a-1[c]	MIR196A1		0.14	0.03	high	chr17: 48632490-48632559 [-]
hsa-miR-148a[c]	MIR148A		-0.19	0.02	low	chr7: 25949919-25949986 [-]

a: Derived from the univariable Cox regression analysis in the training set.
b: Ensembl database
c: miRBase database

patients in the high-risk group were less than 20% at 5 years, while more than 50% in the low-risk group.

To validate the prognostic power of the PCGs-lncRNAs-microRNAs signature for survival prediction, the constructed expression-defined PCGs-lncRNAs-microRNAs prognostic model was also evaluated in the test dataset. The test dataset patients were also classified into high-risk group and low-risk group with the same cutoff value in the training group. Kaplan–Meier curves for the high- and low-risk groups in the test dataset were shown in the *right* of Figure 3A (median survival: 41.6 months *vs.* 57.4 months, log-rank test *P* =0.018). The OS rate of patients in the high-risk group was about 33% at 5 years versus 54% in the low-risk group.

Survival prediction performance of the PCGs-lncRNAs-microRNAs signature is independent of clinical features

To obtain a better understanding of the clinical significance of the PCGs-lncRNAs-microRNAs signature in ovarian serous carcinoma, we correlated the signature with a series of clinicopathological parameters in the combined-two groups. As shown in Table 3, there was no association between PCGs-lncRNAs-microRNAs signature and clinicopathological variables, including age and TNM stage in the training dataset.

To assess whether the prognostic power of the PCGs-lncRNAs-microRNAs signature was independent of other clinical features, multivariable Cox regression analysis was performed using the signature-based risk score and other clinical features. The results of multivariable Cox regression analysis from two ovarian serous carcinoma datasets showed that the prognostic power of the PCGs-lncRNAs-microRNAs signature risk score for prediction of survival was indeed independent of these clinical features in the training group (High-risk group *vs.* Low-risk group, HR = 2.80, 95% CI 1.87–4.18, *P*< 0.001, n=203), and the same result was seen in the test group and entire dataset (Table 4).
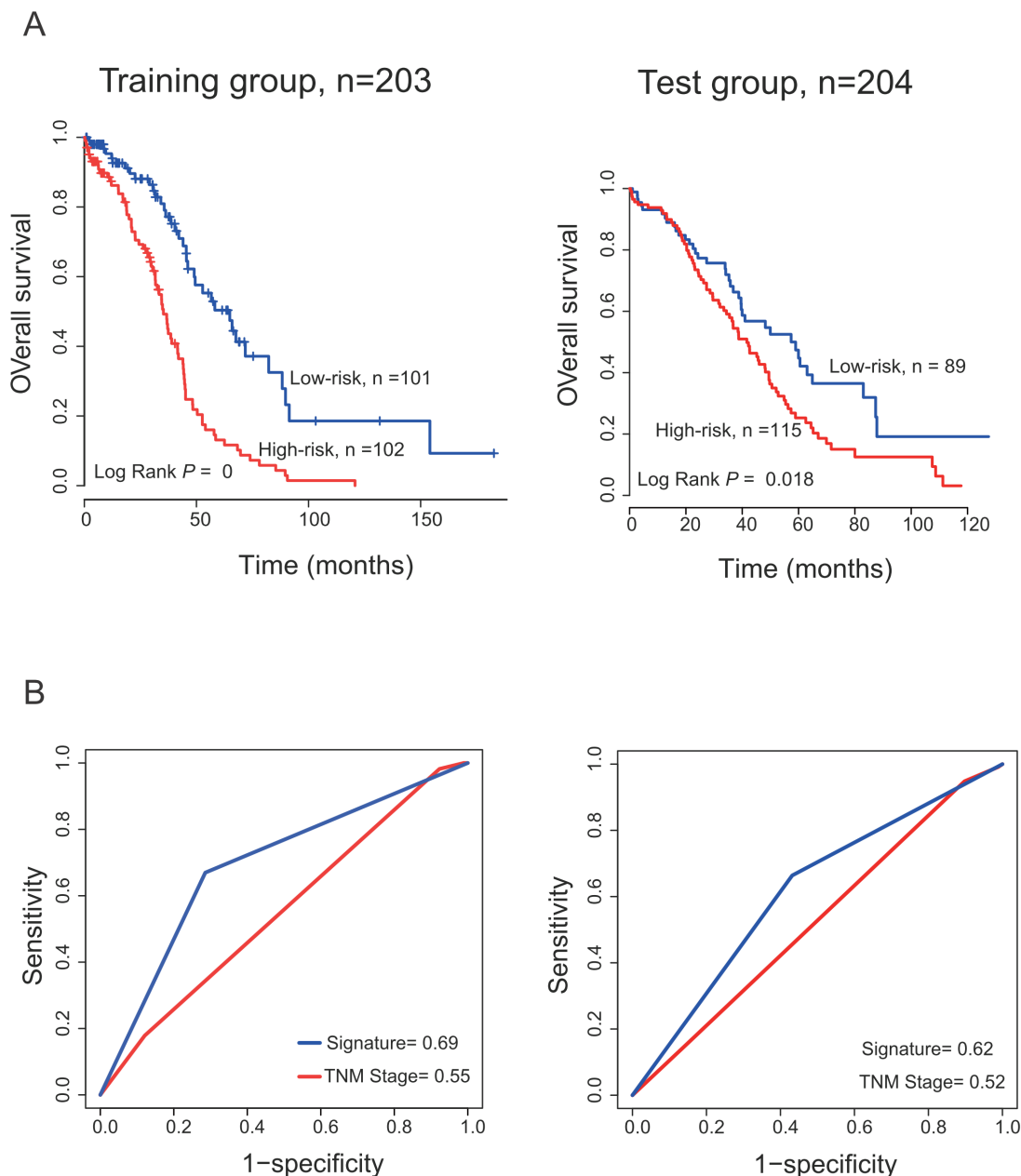
Comparison of the survival prediction power of the PCGs-lncRNAs-microRNAs signature with TNM stage

To compare the sensitivity and specificity in survival prediction between TNM stage and the PCGs-lncRNAs-microRNAs signature, we performed ROC analysis, considering that the larger area under the ROC curve (AUC) usually implied a better model for prediction

[35, 36]. In the training dataset (n=203), predictive ability of the PCGs-lncRNAs-microRNAs signature was significantly better than TNM stage (AUC $_{Signature}$=0.69 vs. AUC $_{TNM}$ =0.55), which further demonstrated that the signature in our study was a novel prognostic marker with higher accuracy and had important clinical significance (Figure 3). The similar result could be seen in the test group (AUC $_{Signature}$=0.62 vs. AUC $_{TNM}$ =0.52, n=204).

Functional characterization of the selected prognostic PCGs, lncRNAs and microRNAs

To further investigate the potential biological roles of this signature, the co-expressed relationships of the two PCGs, two lncRNAs and two microRNAs with those corresponding co-expressed protein-coding genes were computed using Pearson correlation coefficients in the training/test group dataset. The expression levels of

Figure 3: The PCGs-lncRNAs-microRNAs signature predicts overall survival of patients with OC and comparison the survival prediction power of the PCGs-lncRNAs-microRNAs signature and TNM stage. (A) Kaplan–Meier survival curves classified patients into high- and low-risk groups using the PCGs-lncRNAs-microRNAs signature in the training and test datasets. *P* Values were calculated by log-rank test. (B) ROC analysis was used to compare survival prediction power between the PCGs-lncRNAs-microRNAs signature and TNM stage.

Table 3: Association of the PCG-lncRNA-microRNA signature with clinicopathological characteristics in OV patients (n=203)

Variables	PCG-lncRNA signature		P
	Low risk *	High risk *	
Age			0.94
≤59	51	52	
>59	51	49	
pTNM stage			
unkown	1	0	0.5
II	5	3	
III	83	80	
IV	13	18	

* Low risk ≤Median of risk score, High risk >Median of risk score; The Chi-squared test; *P* value <0.05 was considered significant.

Table 4: Univariable and multivariable Cox regression analysis of the PCG-lncRNA-microRNA signature and survival of OV patients in the training, test and entire group

Variables		The training set (n=203)				The Test set (n=204)				The entire dataset (n=407)			
		HR	95% CI of HR		P	HR	95% CI of HR		P	HR	95% CI of HR		P
			lower	upper			lower	upper			lower	upper	
Univariable analysis													
Age	>59 vs.≤59	1.19	0.82	1.73	0.36	0.92	0.55	1.53	0.74	1.24	0.96	1.61	0.10
pTNM stage	IV vs.II+III	1.65	1.06	2.57	0.03	1.30	0.90	1.88	0.16	1.20	0.84	1.70	0.31
PCG-lncRNA-microRNA signature	High risk vs. low risk	2.80	1.87	4.18	0.00	1.59	1.07	2.33	0.02	2.05	1.55	2.71	0.00
Multivariable analysis													
Age	>59 vs.≤59	1.24	0.85	1.81	0.43	1.48	1.01	2.16	0.04	1.36	1.05	1.77	0.02
pTNM stage	IV vs.II+III	1.49	0.95	2.33	0.21	0.97	0.67	1.38	0.86	1.16	0.87	1.55	0.30
PCG-lncRNA-microRNA signature	High risk vs. low risk	2.68	1.79	4.01	0.00	1.75	1.19	2.57	0.00	2.16	1.64	2.84	0.00

1218/1226 protein-coding genes were highly correlated with that of at least one of the selected PCGs, lncRNAs and microRNAs (Pearson correlation coefficient >0.30, *P*<0.05). GSEA analysis for these co-expressed protein-coding genes was then performed based on the whole C2 set. Several clusters of functionally related terms were observed and implied that the two PCGs, two lncRNAs and two microRNAs might be involved in tumorigenesis through interacting with those protein-coding genes that affect important biological processes such as EMT, cell adhesion, GNF FEMALE (Both in training and test groups, *P*<0.05, Figure 4).

Validation of the PCGs, lncRNAs and microRNAs expression in an experimental cohort

As an additional confirmatory method of public data analysis, RNA was extracted from 2 fresh tissues of ovarian serous carcinoma patients who were operated in

our hospital in September 2016 (Supplementary Table 4). The results of semi-quantified PCR for *VAT1L*, *CALR*, LINC01456 and RP11-484L8.1 were shown in Figure 5A. Using $2^{-\Delta\Delta Ct}$ value of the two microRNAs in each sample to evaluate the expression for MIR196A, MIR148A (Figure 5B). All six genes in the signature which was associated with the survival of OV patients by public data analysis were positive in the ovarian cancer tissues. In the future, we will enlarge the ovarian serous carcinoma samples for the real-time PCR.

DISCUSSION

Ovarian serous carcinoma represents one of the leading causes of cancer mortality in women, exhibiting a low five-year survival rate [37]. Therefore,
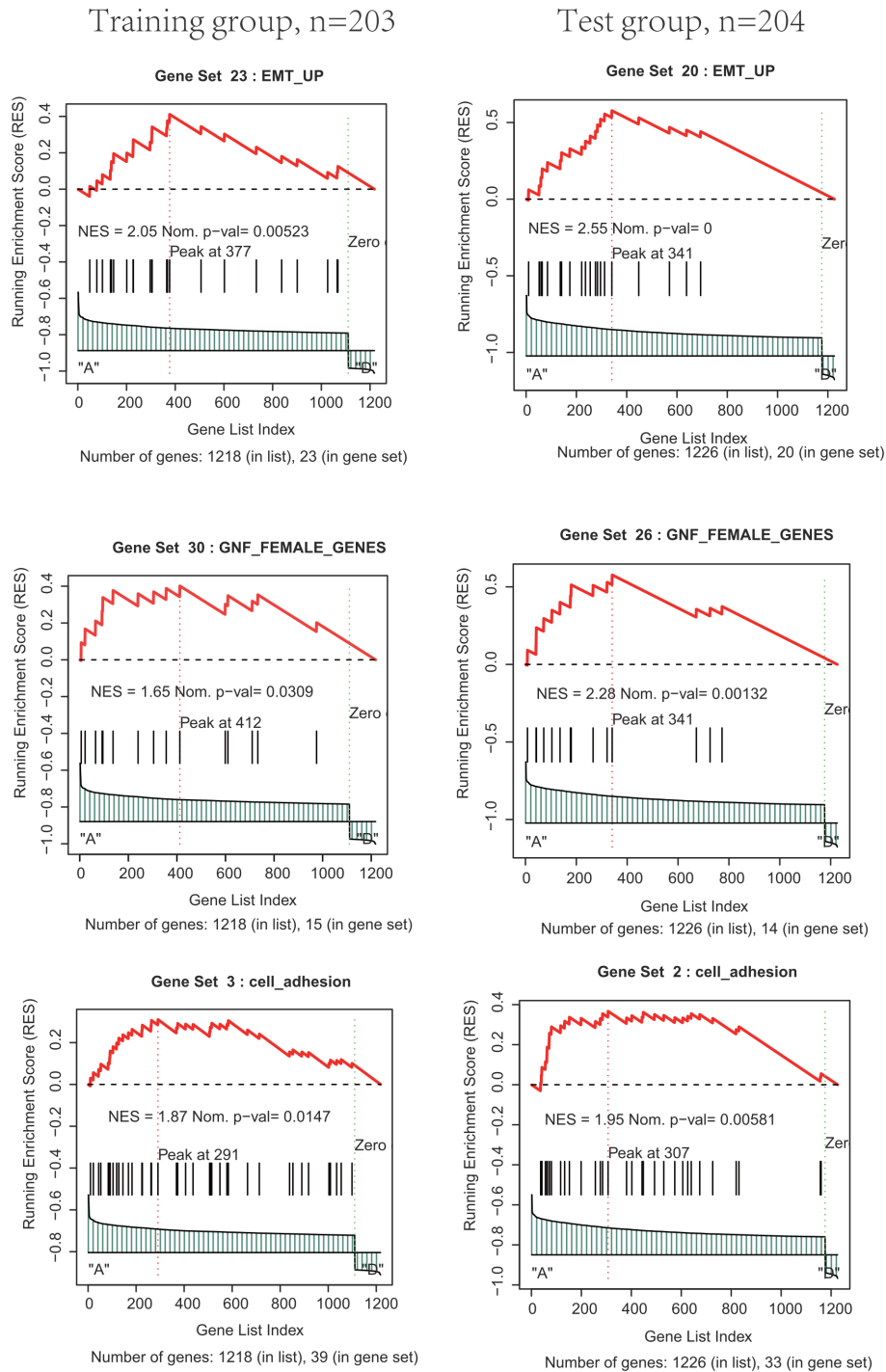
Figure 4: Functional enrichment of the co-expressed protein-coding genes with prognostic the two PCGs, two lncRNAs and two microRNAs by GSEA in training and test group.

the identification and validation of novel biomarkers account for an important part of ovarian serous carcinoma study. On the other hand, high-throughput sequencing become more and more popular in biological studies, the complexity and multiple dimensions of these datasets make the statistical analysis difficult and challenging. Accumulating evidence suggests that PCGs, lncRNAs and microRNAs are involved in oncogenic and tumor suppressive pathways and they may serve as biomarkers [14, 38–43]. The signature composed by PCGs, lncRNAs and microRNAs, which can reflect the important biological processes from multi-dimensional levels, showed a prognostic power in OC patients.

In this study, we used different statistics and machine learning methods to identify a PCGs-lncRNAs-microRNAs expression signature that was associated with survival of ovarian serous carcinoma patients. We further revealed that the signature was an independent predictor of ovarian serous carcinoma patient survival. The multivariable Cox regression analysis was to assess the independence of the selected PCGs-lncRNAs-microRNAs signature in predicting OS. With age, pTNM stage as covariates in the regression analysis, risk score of patients based on the PCGs-lncRNAs-microRNAs signature maintained an independent correlation with OS. Taken together, these results suggested that the prognostic power of the PCGs-lncRNAs-microRNAs signature for predicting OS of OC patients was independent of other clinical features.

The expression of the six genes was validated by experiment. As for the characteristics of two PCGs, two lncRNAs and two microRNAs, the overexpression of *VAT1L*, RP11-484L8.1 and MIR196A1 was associated with shorter OS (coefficient >0) while the overexpression of the remaining *CALR*, LINC01456 and MIR148A was associated with longer OS (coefficient < 0). There is few literature about the function of the two PCGs. *VAT1L* is mainly expressed in the brain and *CALR* mutation status defined subtypes of essential thrombocythemia with

substantially different clinical course and outcomes, thus it could be a potential biomarker for myeloproliterative neoplasm [44, 45]. On the another hand, miR-148a played a pivotal role in the liver by promoting the hepatospecific phenotype and suppressing the invasion of transformed cells [46], promoting cell proliferation by targeting p27 in gastric cancer cells [47], and silencing of miR-148a in cancer-associated fibroblasts resulted in WNT10B-mediated stimulation of tumor cell motility [48]. MiR-196 appears to be a vertebrate specific microRNA and it has been suggested that a rare SNP (rs11614913) that overlaps miR-196 has been found to be associated with non-small cell lung carcinoma [49, 50]. MiR-196 is correlated with metastasis and prognosis of human colorectal cancer [51], and may serve as an emerging cancer biomarker for digestive tract cancers [52]. Although the functions of these PCGs, lncRNAs, microRNAs have been inferred by bioinformatics analysis, the biological roles of the selected two PCGs and two lncRNAs in tumorigenesis are still not clear and should be investigated further.

A few of limitations in this study need to be acknowledged except limited available data about ovarian serous carcinoma. Firstly, in this study, only a fraction of human PCGs (15426 out of 30000+), lncRNAs (8335 out of 15000+) and microRNAs (340 out of 2000+) were included in the analyses. So, the prognostic PCGs, lncRNAs and microRNAs identified here might not represent all the candidates that were potentially correlated with ovarian serous carcinoma overall survival. Secondly, the specific predicted mechanisms of these PCGs, lncRNAs and microRNAs in OC need to be further study. Finally, the signature has not yet been tested prospectively in a clinical trial. Despite these drawbacks, however, the significant and consistent correlation of our PCGs-lncRNAs-microRNAs signature with overall survival in two independent datasets indicated that it was a potentially powerful prognostic marker for OC.

In conclusion, it is the first study to investigate signature composed by prognostic PCGs, lncRNAs and
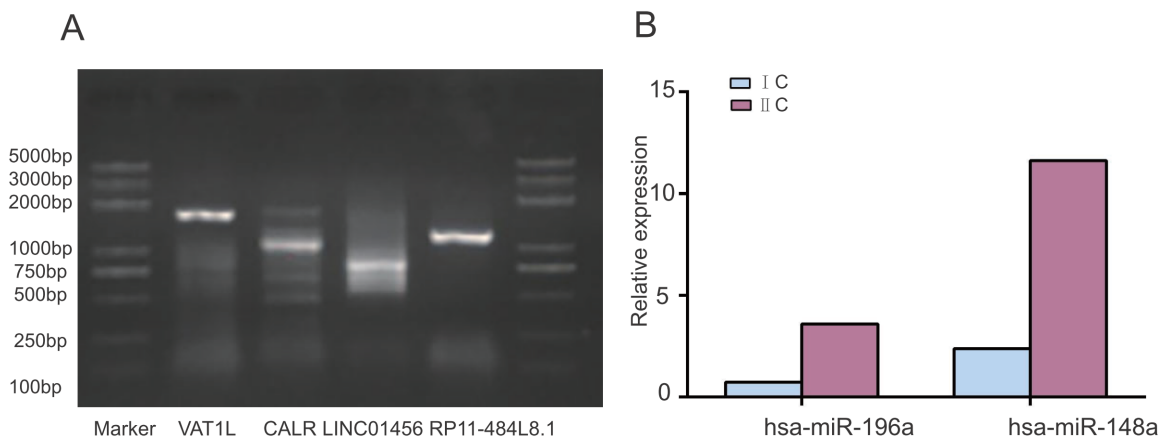
Figure 5: Validation of the PCGs, lncRNAs and microRNAs expression by experiment **(A–B).**

microRNAs in patients with ovarian serous carcinoma. The PCGs-lncRNAs-microRNAs signature can predict the survival of ovarian serous carcinoma patients with more prediction accuracy showing the signature has a bright clinical significance.

MATERIALS AND METHODS

Cancer PCGs, lncRNAs and microRNAs expression data in TCGA

We downloaded the microRNAs (Illumina HiSeq microRNA Seq) and mRNA (Illumina HiSeq RNA Seq V2) level 3 expression data of ovarian carcinoma from the TCGA database through the Data portal (https://genome-cancer.ucsc.edu/proj/site/hgHeatmap/). LncRNA expression datasets were obtained from the TANRIC database (http://ibl.mdanderson.org/tanric/_design/basic/index.html). All genes with missing expression values in >30% samples were removed, and then we imputed the remaining missing values by the k-nearest neighbor method. The gene expression values were log2 transformed for all subsequent analysis [30]. The selection process of the prognostic signature was shown in Figure 1.

Construction of a weighted overall survival (OS) predictive score algorithm

We used a univariable Cox regression analysis to evaluate the relationship between the continuous expression level of each PCG or lncRNA or microRNA and patients' OS in the training dataset. Subsequently, we developed a model for estimation of prognosis similar to what was described as follows [31, 32].

Risk score (RS) $=\sum_{i=1}^{N}$(explg *coef)

Where N was the number of prognostic PCGs, lncRNAs and microRNAs, Explg was the expression value of PCGs, lncRNAs or microRNAs, and Coef was the estimated regression coefficient of PCGs, lncRNAs or microRNAs in the univariable Cox regression analysis. This risk score model was made up of the prognostic PCGs, lncRNAs and microRNAs.

Statistical analyses

Considering that a smaller number of PCGs, lncRNAs and microRNAs in the model would make the model more practical, we performed the random survival forests-variable hunting (RSFVH) algorithm to filter genes until two PCGs, two lncRNAs and two microRNAs were screened out [10]. The time-dependent receiver operating characteristic (ROC) curve was used to compare the sensitivity and specificity of the survival prediction of the risk score of the 511 combinations or signatures composed by the selected nine genes in the training dataset. Area under the curve (AUC) value was calculated from the ROC curve [10]. Then using the median risk score in the training dataset as a cutoff value, ovarian serous carcinoma patients in each dataset were divided into high- and low-risk groups [33]. Kaplan–Meier survival analyses were performed to test the equality for survival distributions in different groups for each ovarian serous carcinoma cohort, and statistical significance was assessed using the two-sided log-rank test. Additionally, chi-square test was used to analyze the association of the survival with the clinical attributes. And multivariable Cox regression analysis was performed to test whether the risk score was independent of other clinical features within the available data. Significance was defined as $P < 0.05$. All analyses were performed using R program (http://www.r-project.org) including packages named survival ROC, survival and random Forest SRC downloaded from Bio-conductor.

Sample collection and preparation

The study was approved by the Hubei Maternal and Child Health Hospital, P.R. China. Fresh tissues of two ovarian serous carcinoma patients were collected, from whom underwent surgical resection (clinicopathological characteristics were listed in Supplementary Table 4). The utilization of the tissues had been through the complete patient informed consent. After being examined by a pathologist, tissues were immediately frozen in liquid nitrogen and stored at -80°C. All tumor samples contained more than 80% tumor tissue free of necrosis. All samples were coded to protect patient anonymity.

PCR and RT-PCR experiments for validation of the gene expression

Total RNA was extracted using TRIzol (Life Technologies, USA) according to the manufacturer's protocol. cDNA was synthesized with random hexamers, PCR kit was completed by HotStar Taq Master Mix kit (QIAGEN, Germany) and real-time PCR was performed by using a QuantiTect SYBR® Green PCR kit (QIAGEN, Germany). Briefly, reverse transcription was performed according to the following conditions: 16°C, 30 min; 42°C, 40 min; 85°C, 5 min (cDNA for the RT-PCR) and 30°C, 10 min; 42°C, 60 min; 95°C, 5 min (cDNA for the PCR). RT-PCR was performed using an PAC 3000 real-time PCR system (BIO RAD, USA) according to the following conditions: 95°C, 2 min; 94°C, 10 sec; 56°C, 10 sec; 72°C, 40 sec. PCR was completed according to the criteria: 96°C, 20 sec; 52°C, 20 sec; 72°C, 5 min. Relative quantification of microRNA expression was calculated by the $2^{-\Delta\Delta Ct}$ method. The primer sequences of *VAT1L*, *CALR*, LINC01456, RP11-484L8.1 for PCR, and MIR196A1, MIR148A for RT-PCR were shown in Supplementary Table 3, and *ACTB* (β-actin) was used as the internal

control. All were accomplished in triplicate and repeated at least three times. All methods were performed in accordance with guidelines and regulations set by the above Ethical Committee.

Function prediction of PCGs, lncRNAs, and microRNAs by gene set enrichment analysis (GSEA)[34]

The co-expressed relationships between the prognostic PCGs, lncRNAs, microRNAs and their corresponding co-expressed PCGs were computed using Pearson correlation coefficients. Gene set enrichment analysis (GSEA) was used to predict the roles of above-obtained co-expressed PCGs (Pearson coefficient > 0.3, $P < 0.05$). Functional annotation with $P < 0.05$ was considered significant.

Abbreviations

ROC: receiver operating characteristic; AUC: area under the ROC curve; CI: confidence interval; HR: hazard ratio; PCG: protein-coding gene; LncRNA: long non-coding RNA; OS: overall survival; OV: ovarian cancer; SD: standard deviation; GSEA: Gene set enrichment analysis; TCGA: The Cancer Genome Atlas.

Author contributions

The authors contributed in the following way: data collection, data analysis, interpretation, and drafting.

The authors contributed in the following way: Xu Meng, Guo Jin-Cheng,: data collection, data analysis, interpretation, and drafting, study design, study supervision, and final approval of the manuscript; Wu Xufeng, Zhang Que, Ma Quanfu, Yan Bing: technical support, experiment advice and data collection. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST

No potential conflicts of interest were disclosed.

GRANT SUPPORT

REFERENCES

1. Gansler T, Ganz PA, Grant M, Greene FL, Johnstone P, Mahoney M, Newman LA, Oh WK, Thomas CR Jr, Thun MJ, Vickers AJ, Wender RC, Brawley OW. Sixty years of CA: a cancer journal for clinicians. CA Cancer J Clin. 2010; 60:345-350.

2. Holschneider CH, Berek JS. Ovarian cancer: epidemiology, biology, and prognostic factors. Semin Surg Oncol. 2000; 19:3-10.

3. Klint A, Tryggvadottir L, Bray F, Gislum M, Hakulinen T, Storm HH, Engholm G. Trends in the survival of patients diagnosed with cancer in female genital organs in the Nordic countries 1964-2003 followed up to the end of 2006. Acta Oncol. 2010; 49:632-643.

4. Markman M, Bundy BN, Alberts DS, Fowler JM, Clark-Pearson DL, Carson LF, Wadler S, Sickel J. Phase III trial of standard-dose intravenous cisplatin plus paclitaxel versus moderately high-dose carboplatin followed by intravenous paclitaxel and intraperitoneal cisplatin in small-volume stage III ovarian carcinoma: an intergroup study of the Gynecologic Oncology Group, Southwestern Oncology Group, and Eastern Cooperative Oncology Group. J Clin Oncol. 2001; 19:1001-1007.

5. Armstrong DK, Bundy B, Wenzel L, Huang HQ, Baergen R, Lele S, Copeland LJ, Walker JL, Burger RA, Gynecologic Oncology Group. Intraperitoneal cisplatin and paclitaxel in ovarian cancer. N Engl J Med. 2006; 354:34-43.

6. Coleman MP, Forman D, Bryant H, Butler J, Rachet B, Maringe C, Nur U, Tracey E, Coory M, Hatcher J, McGahan CE, Turner D, Marrett L, et al. Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. Lancet. 2011; 377:127-138.

7. Zhou X, Huang Z, Xu L, Zhu M, Zhang L, Zhang H, Wang X, Li H, Zhu W, Shu Y, Liu P. A panel of 13-miRNA signature as a potential biomarker for predicting survival in pancreatic cancer. Oncotarget. 2016; 7:69616-69624. http://doi.org/18632/oncotarget.11903.

8. Xu M, Kuang Y, Wang M, Han X, Yang Q. A microRNA expression signature as a predictor of survival for colon adenocarcinoma. Neoplasma. 2016; 64:56-64.

9. Meng J, Li P, Zhang Q, Yang Z, Fu S. A four-long non-coding RNA signature in predicting breast cancer survival. J Exp Clin Cancer Res. 2014; 33:84.

10. 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.

11. Zhan Y, Guo W, Zhang Y, Wang Q, Xu XJ, Zhu L. A Five-Gene Signature Predicts Prognosis in Patients with Kidney Renal Clear Cell Carcinoma. Comput Math Methods Med. 2015; 2015:842784.

12. Sveen A, Agesen TH, Nesbakken A, Meling GI, Rognum TO, Liestol K, Skotheim RI, Lothe RA. ColoGuidePro: a prognostic 7-gene expression signature for stage III colorectal cancer patients. Clin Cancer Res. 2012; 18:6001-6010.

13. Maass PG, Luft FC, Bahring S. Long non-coding RNA in health and disease. J Mol Med (Berl). 2014; 92:337-346.

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

15. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010; 464:1071-1076.

16. Kong H, Wu Y, Zhu M, Zhai C, Qian J, Gao X, Wang S, Hou Y, Lu S, Zhu H. Long non-coding RNAs: novel prognostic biomarkers for liver metastases in patients with early stage colorectal cancer. Oncotarget. 2016; 7:50428-50436. http://doi.org/10.18632/oncotarget.10416.

17. Guo Q, Cheng Y, Liang T, He Y, Ren C, Sun L, Zhang G. Comprehensive analysis of lncRNA-mRNA co-expression patterns identifies immune-associated lncRNA biomarkers in ovarian cancer malignant progression. Sci Rep. 2015; 5:17683.

18. 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. http://doi.org/10.18632/oncotarget.7181.

19. 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; 7:32433-32448. http://doi.org/10.18632/oncotarget.8653.

20. Carthew RW, Sontheimer EJ. Origins and Mechanisms of miRNAs and siRNAs. Cell. 2009; 136:642-655.

21. Kent OA, Mendell JT. A small piece in the cancer puzzle: microRNAs as tumor suppressors and oncogenes. Oncogene. 2006; 25:6188-6196.

22. Jiang BY, Zhang XC, Su J, Meng W, Yang XN, Yang JJ, Zhou Q, Chen ZY, Chen ZH, Xie Z, Chen SL, Wu YL. BCL11A overexpression predicts survival and relapse in non-small cell lung cancer and is modulated by microRNA-30a and gene amplification. Mol Cancer. 2013; 12:61.

23. Cascione L, Gasparini P, Lovat F, Carasi S, Pulvirenti A, Ferro A, Alder H, He G, Vecchione A, Croce CM, Shapiro CL, Huebner K. Integrated microRNA and mRNA signatures associated with survival in triple negative breast cancer. PLoS One. 2013; 8:e55910.

24. Fan MQ, Huang CB, Gu Y, Xiao Y, Sheng JX, Zhong L. Decrease expression of microRNA-20a promotes cancer cell proliferation and predicts poor survival of hepatocellular carcinoma. J Exp Clin Cancer Res. 2013; 32:21.

25. Jiang X, Du L, Wang L, Li J, Liu Y, Zheng G, Qu A, Zhang X, Pan H, Yang Y, Wang C. Serum microRNA expression signatures identified from genome-wide microRNA profiling serve as novel noninvasive biomarkers for diagnosis and recurrence of bladder cancer. Int J Cancer. 2015; 136:854-862.

26. Hu Z, Chen X, Zhao Y, Tian T, Jin G, Shu Y, Chen Y, Xu L, Zen K, Zhang C, Shen H. Serum microRNA signatures identified in a genome-wide serum microRNA expression profiling predict survival of non-small-cell lung cancer. J Clin Oncol. 2010; 28:1721-1726.

27. Landi MT, Zhao Y, Rotunno M, Koshiol J, Liu H, Bergen AW, Rubagotti M, Goldstein AM, Linnoila I, Marincola FM, Tucker MA, Bertazzi PA, Pesatori AC, et al. MicroRNA expression differentiates histology and predicts survival of lung cancer. Clin Cancer Res. 2010; 16:430-441.

28. Qin CZ, Lv QL, Yang YT, Zhang JM, Zhang XJ, Zhou HH. Downregulation of MicroRNA-320d predicts poor overall survival and promotes the growth and invasive abilities in glioma. Chem Biol Drug Des. 2016; 89:806-814.

29. Mullany LE, Herrick JS, Wolff RK, Slattery ML. MicroRNA Seed Region Length Impact on Target Messenger RNA Expression and Survival in Colorectal Cancer. PLoS One. 2016; 11:e0154177.

30. Xu J, Li Y, Lu J, Pan T, Ding N, Wang Z, Shao T, Zhang J, Wang L, Li X. The mRNA related ceRNA-ceRNA landscape and significance across 20 major cancer types. Nucleic Acids Res. 2015; 43:8169-8182.

31. Cao HH, Zhang SY, Shen JH, Wu ZY, Wu JY, Wang SH, Li EM, Xu LY. A three-protein signature and clinical outcome in esophageal squamous cell carcinoma. Oncotarget. 2015; 6:5435-5448. http://doi.org/10.18632/oncotarget.3102

32. Guo JC, Li CQ, Wang QY, Zhao JM, Ding JY, Li EM, Xu LY. Protein-coding genes combined with long non-coding RNAs predict prognosis in esophageal squamous cell carcinoma patients as a novel clinical multi-dimensional signature. Mol Biosyst. 2016; 12:3467-3477.

33. 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.

34. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102:15545-15550.

35. Tse LA, Dai J, Chen M, Liu Y, Zhang H, Wong TW, Leung CC, Kromhout H, Meijer E, Liu S, Wang F, Yu IT, Shen H, Chen W. Prediction models and risk assessment for silicosis using a retrospective cohort study among workers exposed to silica in China. Sci Rep. 2015; 5:11059.

36. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics. 2000; 56:337-344.

37. Baldwin LA, Huang B, Miller RW, Tucker T, Goodrich ST, Podzielinski I, DeSimone CP, Ueland FR, van Nagell JR,

Seamon LG. Ten-year relative survival for epithelial ovarian cancer. Obstet Gynecol. 2012; 120:612-618.

38. He L, Hannon GJ. MicroRNAs: small RNAs with a big role in gene regulation. Nat Rev Genet. 2004; 5:522-531.

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

40. 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.

41. Wan J, Wu W, Che Y, Kang N, Zhang R. Insights into the potential use of microRNAs as a novel class of biomarkers in esophageal cancer. Dis Esophagus. 2016; 29:412-420.

42. Zhang H, Chen Z, Wang X, Huang Z, He Z, Chen Y. Long non-coding RNA: a new player in cancer. J Hematol Oncol. 2013; 6:37.

43. Geng Q, Fan T, Zhang B, Wang W, Xu Y, Hu H. Five microRNAs in plasma as novel biomarkers for screening of early-stage non-small cell lung cancer. Respir Res. 2014; 15:149.

44. Rumi E, Pietra D, Ferretti V, Klampfl T, Harutyunyan AS, Milosevic JD, Them NC, Berg T, Elena C, Casetti IC, Milanesi C, Sant'antonio E, Bellini M, et al. JAK2 or CALR mutation status defines subtypes of essential thrombocythemia with substantially different clinical course and outcomes. Blood. 2014; 123:1544-1551.

45. Yuan J, Hao HL, Li Y, Wang RC, Zhang XX. [Research Progress on CALR Mutation in the Myeloproliterative Neoplasm -Review]. Zhongguo Shi Yan Xue Ye Xue Za Zhi. 2016; 24:1252-1255. [Article in Chinese].

46. Gailhouste L, Gomez-Santos L, Hagiwara K, Hatada I, Kitagawa N, Kawaharada K, Thirion M, Kosaka N, Takahashi RU, Shibata T, Miyajima A, Ochiya T. miR-148a plays a pivotal role in the liver by promoting the hepatospecific phenotype and suppressing the invasiveness of transformed cells. Hepatology. 2013; 58:1153-1165.

47. Guo SL, Peng Z, Yang X, Fan KJ, Ye H, Li ZH, Wang Y, Xu XL, Li J, Wang YL, Teng Y, Yang X. miR-148a promoted cell proliferation by targeting p27 in gastric cancer cells. Int J Biol Sci. 2011; 7:567-574.

48. Aprelikova O, Palla J, Hibler B, Yu X, Greer YE, Yi M, Stephens R, Maxwell GL, Jazaeri A, Risinger JI, Rubin JS, Niederhuber J. Silencing of miR-148a in cancer-associated fibroblasts results in WNT10B-mediated stimulation of tumor cell motility. Oncogene. 2013; 32:3246-3253.

49. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 2011; 39:D152-157.

50. Griffiths-Jones S. miRBase: the microRNA sequence database. Methods Mol Biol. 2006; 342:129-138.

51. Shen S, Pan J, Lu X, Chi P. Role of miR-196 and its target gene HoxB8 in the development and proliferation of human colorectal cancer and the impact of neoadjuvant chemotherapy with FOLFOX4 on their expression. Oncol Lett. 2016; 12:4041-4047.

52. Lu YC, Chang JT, Chan EC, Chao YK, Yeh TS, Chen JS, Cheng AJ. miR-196, an Emerging Cancer Biomarker for Digestive Tract Cancers. J Cancer. 2016; 7:650-655.