A panel of 7 prognosis-related long non-coding RNAs to improve platinum-based chemoresistance prediction in ovarian cancer

JING SONG[1], WANFENG ZHANG[1], SEN WANG[1], KUN LIU[1], FANGZHOU SONG[2] and LONGKE RAN[1,2]

[1]Department of Bioinformatics, The Basic Medical School of Chongqing Medical University;
[2]Molecular and Tumor Research Center, Chongqing Medical University, Chongqing 400016, P.R. China

Received January 25, 2018; Accepted May 2, 2018

DOI: 10.3892/ijo.2018.4403

Abstract. In order to study the role of long non-coding RNAs (lncRNAs) in predicting platinum-based chemoresistance in patients with high-grade serous ovarian carcinoma (HGS-OvCa), a=7-lncRNA signature was developed by analyzing 561 microarrays and 136 specimens from RNA-sequencing (RNA-seq) obtained from online databases [odds ratio (OR), 2.859; P<0.0001]. The upregulated lncRNAs (RP11-126K1.6, ZBED3-AS1, RP11-439E19.10 and RP11-348N5.7) and downregulated lncRNAs [RNF144A-AS1, growth arrest specific 5 (GAS5) and F11-AS1] exhibited high sensitivity and specificity in predicting chemoresistance in the Gene Expression Omnibus and the Cancer Genome Atlas (area under curve >0.8). The lncRNA signature was independent of clinical characteristics and 4 HGS-OvCa molecular subtypes. This signature was negatively associated with disease-free survival (n=47; log-rank, P<0.01). Furthermore, the expression of the 7 lncRNAs was consistent with microarray (GSE63885, GSE51373, GSE15372 and GSE9891) and RNA-seq data. In *in vitro* experiments, ZBED3-AS1, F11-AS1 and GAS5 were differentially expressed in cell lines that are known to be resistant and non-resistant to platinum-based drugs, which was consistent with the results in the present study. This lncRNA signature may be used as a prognostic marker for predicting resistance to platinum-based chemotherapeutics in HGS-OvCa. These findings may contribute to individualized therapies in patients with HGS-OvCa in the future.

Introduction

In the past 30 years, ovarian cancer has exhibited the highest mortality rate of all types of lethal gynecological cancer worldwide, with a 5-year survival rate of <30% (1,2). The characteristics of ovarian carcinoma encompass the development of chemoresistance (3), recurrence (4) and poor prognosis (5), which pose huge challenges for drug development and therapeutics (6). Once chemoresistance occurs, patients may not only experience adverse effects, but they may also eventually relapse (7). Although ~75% of patients initially respond to the platinum-based chemotherapy, the majority relapse with resistance, which leads to treatment failure and causes >90% of mortalities (8). To date, patients with ovarian cancer have been defined as treatment-resistant if the duration of disease-free survival (DFS) or the platinum-free interval was <6 months, while patients with longer durations have been defined as treatment-sensitive (3,9). However, this approach cannot provide guidance for individualized treatment of patients with drug resistance.

Long non-coding (lnc)RNA is a type of RNA that has no protein coding potential (10). However, lncRNAs have been reported to be critical in multiple stages of tumorigenesis and progression, including apoptosis, DNA damage response and metastasis (11,12).

lncRNAs may also be potential markers for the prediction of platinum-based chemoresistance (11,12). LncRNA ABHD11 antisense RNA 1 (tail to tail) was demonstrated to promote the proliferation, invasion and migration of ovarian cancer cells (13). LncRNA nuclear enriched abundant transcript 1 was reported to interact with BAI1-associated protein 1, which may contribute to the sensitivity against gemcitabine in cholangiocarcinoma (14). The role of lncRNAs in drug resistance of high-grade serous ovarian cancer (HGS-OvCa) remains to be fully investigated. Therefore, examining the possibilities of individualized treatment in patients with chemoresistance to platinum-based therapeutics and ovarian cancer from the perspective of lncRNAs may be useful to improve our understanding of drug resistance (15,16). The present study investigated the role of lncRNAs in platinum-based chemoresistance and survival in patients with HGS-OvCa in order to contribute to current research on individualized therapies.

Materials and methods

Training and validation datasets. Raw human microarray expression profiles were retrieved and downloaded from Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/). First of all, the key words 'high-grade serous ovarian

Correspondence to: Professor Longke Ran, Department of Bioinformatics, The Basic Medical School of Chongqing Medical University, 1 Yixueyuan Road, Yuzhong, Chongqing 400016, P.R. China
E-mail: longkeran@aliyun.com

Key words: high-grade serous ovarian cancer, long non-coding RNA, drug resistance, platinum-based chemotherapy, disease-free survival

cancer' and 'chemotherapeutic resistance' were used to search for samples. Secondly, as more lncRNAs could be detected using the Affymetrix Human Genome U133 Plus2.0 microarray (Affynetrix; Thermo Fisher Scientific, Inc., Waltham, MA, USA), the samples not from that platform were filtered. Finally, the samples with data associated with chemotherapeutic response were selected. In total, 328 microarray profiles with corresponding clinical information [i.e. age, tumor grade, International Federation of Gynecology and Obstetrics (FIGO) stage (17), residual tumor size and platinum sensitivity] were downloaded. The accession numbers of GEO for the datasets were as follows: GSE63885 (18) (n=70); GSE51373 (19) (n=28) and GSE15372 (20) (n=10). Another dataset GSE9891 (21) (n=220) was used as a GEO validation dataset. The expression levels were normalized using the robust multi-array average (RMA) algorithm (22). Batch effects were removed using the Combat algorithm (23).

The expression profiles and clinical information of 233 patients with HGS-OvCa were downloaded from The Atlas of Noncoding RNAs in Cancer (24) and cBioportal (25) databases, respectively, as The Cancer Genome Atlas (TCGA) validation dataset. The expression levels were processed using the same methods as aforementioned. All datasets were analyzed to determine the differences in expression (Table I).

Integration and probe annotation of lncRNA datasets. LNCipeida (26) is a publicly available database for annotated human lncRNA transcript sequences and structures. The public repository produced by Du *et al* (27) provides a resource of clinically relevant lncRNAs for the development of lncRNA biomarkers and the identification of lncRNA therapeutic targets. The present study integrated the records of lncRNAs from GENCODE (28), LNCipedia, NONCODE (29), TCGA and the public repository by Du *et al* (27). In total, 24,372 lncRNAs were included in the present study. To avoid inaccurate annotations, the probes were cross-referenced by gene name and Ensembl transcript ID (release 89). For multiple probes of a single lncRNA, the median expression of the unique transcript ID was preserved. Finally, the expression profiles of 6,955 lncRNAs were obtained.

Clustering and differential expression analysis. The platinum-based sensitivity of patients with HGS-OvCa varies greatly, and the DFS ranged from 180 days to several thousand days. In this context, unsupervised hierarchical clustering based on Pearson distance for platinum-sensitive samples were performed using the ConcensusClusterPlus package in R (30).

The differential expression analysis was performed using R/Limma package (31). The common differentially expressed lncRNAs between each sensitive subgroup and resistant group were screened [differential expressed (DE) set A]. The differentially expressed lncRNAs between the complete sensitive group and the resistant group were obtained (DE set B); DE sets A and B were then merged.

Univariate and multivariate logistic regression analysis and survival analysis. To investigate the platinum-based chemoresistance-associated lncRNAs and the clinical characteristics of patients with HGS-OvCa, an independent univariate logistic regression was performed on 6,955 lncRNAs and clinical characteristics. The lncRNAs and clinical characteristics with $P<0.05$ were considered statistically significant and associated with chemoresistance. The survival package in R (32) was used to investigate DFS-associated lncRNAs. The Kaplan-Meier survival curves of highly expressed and weakly expressed lncRNAs were constructed. The median expression level was used as a cut-off for determining high and low expression.

The lncRNAs and clinical characteristics that were statistically significant in the univariate logistic regression were used as candidate parameters for constructing the multivariate logistic regression model. A best multivariate logistic regression model was generated after fitting the model thousands of times. The expression levels of each lncRNA and its coefficient in the best multivariate logistic model were used to calculate the platinum-based chemoresistance risk score of patients with HGSC-OvCa. The risk score was computed by the following formula:

$$\text{Risk score} = \sum_{i=1}^{n} (\text{coef}_i * \text{expr}_i)$$

where n is the number of prognostic lncRNAs in the model, 'expr_i' is the expression level of lncRNA_i and coef_i is the estimated regression coefficient of lncRNA_i in the multivariate logistic regression model. Patients who have higher risk scores are expected to have higher probability of platinum-based chemoresistance. The median risk score was used as a cut-off to divide the patients in the GEO validation dataset into low-risk and high-risk groups. Furthermore, the multivariate logistic regression analysis was performed to test whether the risk score was independent of clinical covariates.

Subtype classification of patients with HGS-OvCa. The 100-gene set containing four gene signatures corresponding to four HGS-OvCa subtypes (differentiated, immunoreactive, proliferative and mesenchymal) (33) were used to train the support vector machine for subtype classification. The univariate and multivariate logistic regression were performed to investigate whether the HGS-OvCa subtypes were associated with platinum-based chemoresistance, and whether these subtypes associated with our signature, respectively.

Classifier performance evaluation. To assess the sensitivity and specificity of the risk scores in prediction of chemoresistance in patients with HGS-OvCa, ROC (34) and ROCR package (35) in R software were used to construct the receiver operating characteristic curve and perform the area under curve (AUC) analysis, respectively. Furthermore, the predictive performance of the risk score and three signatures previously developed by Zhou *et al* (36), Liu *et al* (37) and TCGA mRNA signatures (38) was compared.

Verification of the expression level of 7 lncRNAs. Microarrays and RNA-seq data were used to validate the expression of lncRNAs. The RNA-seq data of 136 patients with HGS-OvCa and platinum-based chemosensitivity was derived from the Gaucher Disease Advisory Committee (39). The reads per kilobase per million mapped reads were obtained by chromosome location (GRCH37, hg19) of lncRNAs and normalized using Z-scores.

First, the gene expression in the Cancer Cell Line Encyclopedia (CCLE) (40) database had been pre-processed using the RMA algorithm. The chemotherapeutic status

Table I. Patient characteristics in GEO and TCGA datasets.

Characteristics	GEO training dataset	GEO validation dataset	TCGA validation dataset	P-value
Sample size, n	108	328	233	
Mean age (SD), years	59.3 (8.3)	60.4 (9.8)	59.8 (11.4)	0.714
Tumor grade, n (%)				<0.001
G2	9 (12.9)	105 (32.0)	32 (13.7)	
G3	46 (65.7)	170 (51.8)	197 (84.5)	
G4	15 (21.4)	15 (4.6)	0 (0)	
Unknown	38 (35.2)	38 (11.6)	4 (1.7)	
FIGO stage, n (%)				0.073
II	8 (7.4)	34 (10.4)	11(4.7)	
III	78 (72.2)	255 (77.7)	190 (81.5)	
IV	12 (11.1)	29 (8.8)	32 (13.7)	
Unknown	10 (9.3)	10 (3.0)	0 (0)	
Residual tumor size, n (%)				<0.001
R0 (1-10 mm)	14 (13.0)	14 (4.3)	116 (49.8)	
R1 (10-20 mm)	35 (32.4)	35 (10.7)	15 (6.4)	
R2 (>20 mm)	21 (19.4)	21 (6.4)	41 (17.6)	
Unknown	38 (35.2)	258 (78.7)	61 (26.2)	
Platinum sensitivity, n (%)				0.021
Sensitive	59 (54.6)	217 (66.2)	163 (70.0)	
Resistance	49 (45.4)	111 (33.8)	70 (30.0)	
Molecular subtypes, n (%)				0.114
Differentiated	28 (25.9)	77 (23.5)	66 (28.3)	
Immunoreactive	26 (24.1)	94 (28.6)	44 (18.9)	
Mesenchymal	18 (16.7)	59 (18.0)	56 (24.0)	
Proliferative	36 (33.3)	98 (29.9)	67 (28.8)	

P-values for the difference among the GEO and TCGA cohorts were calculated using analysis of variance (for age) and the χ^2 test (for tumor grade, FIGO stage, residual tumor size, platinum sensitivity, and molecular subtypes). GEO, Gene Expression Omnibus; TCGA, The Cancer Genome Atlas; SD, standard deviation; FIGO, International Federation of Gynecology and Obstetrics.

of 52 ovarian cancer cell lines was investigated. The data regarding the cell lines of clear cell, granulosa cell, undifferentiated cell, mixed cell, endometrioid cell, Brenner cell, mucinous cell and cross-contaminated cell were filtered. The cell lines that were treated with cisplatin or carboplatin underwent >3 cycles of treatment. The cell lines exhibited resistance to platinum and the degree of resistance varied. Secondly, gene expression levels in 46 ovarian cancer cell lines with or without cisplatin treatment were obtained from GSE47856 (41). To investigate the changes in the expression of lncRNAs that was induced by platinum, cell lines from the CCLE were analyzed using the Integrative Genomics Viewer (IGV) software (version 2.3) (42). In addition, cell lines from GSE47856 with ≥3 replicates were analyzed. For multi-probe lncRNA, the median expression was calculated.

Functional prediction analysis. To investigate the biological processes and cellular components implicated in the drug resistance mechanisms, the enriched Gene Ontology (GO) (43) terms of 7 lncRNAs were identified using the NONCODE database and were visualized using Cytoscape (version 3.2.1) (44). The potential functions of the lncRNAs were further analyzed using non-coding RNA in the Drug Resistance (ncDR) database (45); P<0.05 was used to identify resistant compounds of lncRNAs in cancer.

Statistical analysis. The empirical Bayes method and a two-tailed Student's t-test were used in differential expression analysis. The log-rank test was used for survival analysis, with Kaplan-Meier curves. Clinical characteristics and molecular subtypes of three datasets were tested using the χ^2 tests or two-way analysis of variance, where appropriate. In multivariate logistic regression, the χ^2 test was used to add or filter features and avoid overdispersion. The features were also filtered using the methods of Watt *et al* (46). P<0.05 was considered to indicate a statistically significant difference. All statistical analyses were conducted using the R software (version 3.2.1).

Results

Cluster analysis of chemosensitive patients reveals three subgroups. The results of unsupervised hierarchical clustering

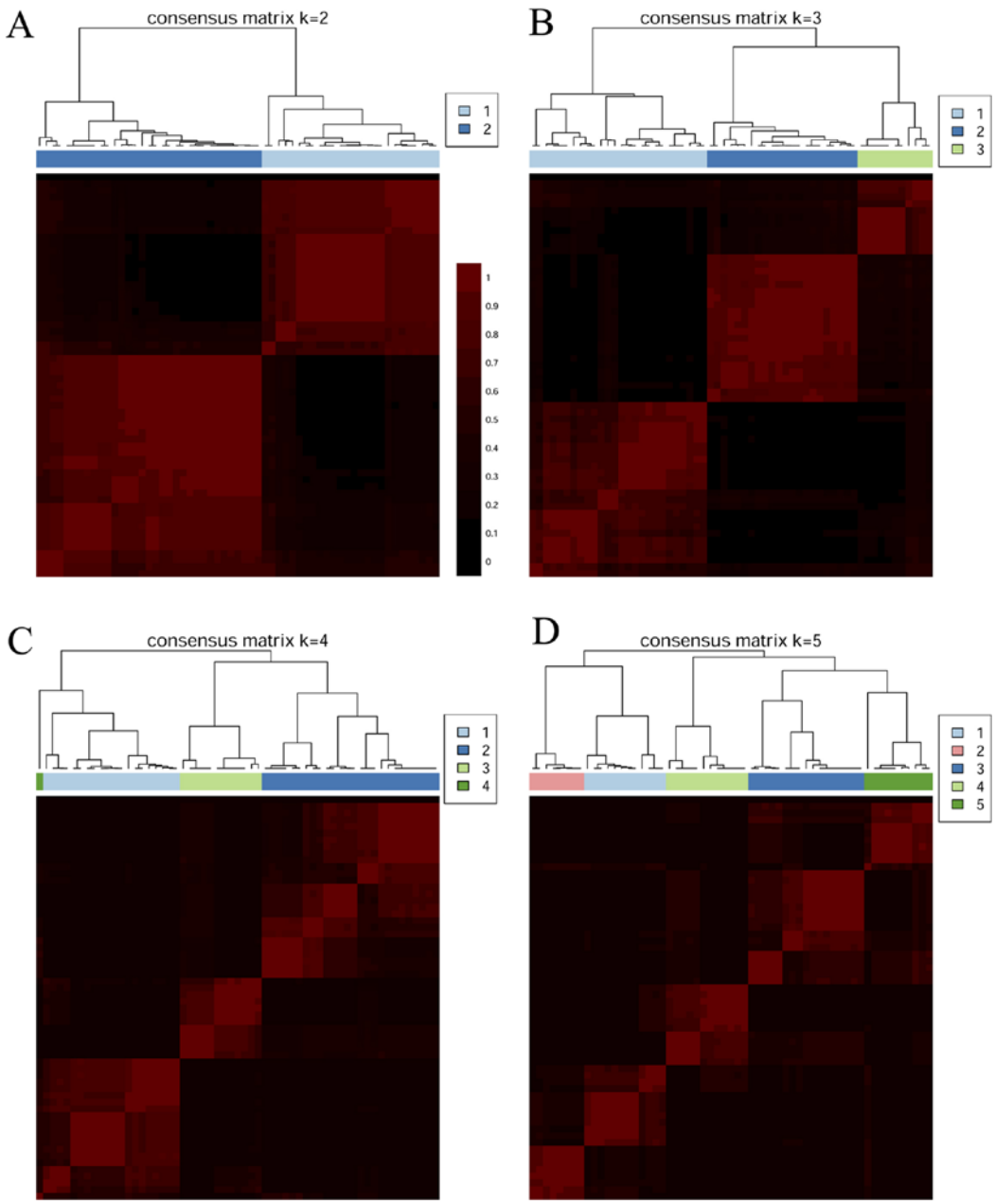

Figure 1. Unsupervised hierarchical clustering of the platinum-based chemotherapeutic sensitive specimens in the GEO training dataset. (A) A total of two dominant clusters are present and identified as cluster 1 and cluster 2 by gene expression patterns. (B) A solid third cluster (identified as cluster 3) emerged when k=3. (C and D) Consensus matrices for the GEO training array do not show more than three definitive clusters even though additional clustering (k=4 and 5) was performed. Red color indicates the similarity between samples, and the samples that were clustered together are displayed. GEO, Gene Expression Omnibus.

were obtained when clustering center k was set from 2 to 5 (Fig. 1). The specimens were perfectly divided into two clusters (Fig. 1A); a third cluster was observed when k=3 (Fig. 1B). However, fourth stable clusters were not found when the number of cluster centers k continued to increase (Fig. 1C and D). These results indicate that there may be three stable subgroups with different expression patterns in chemosensitive patients.

Identification and assessment of chemoresistance-associated lncRNA signature. In the present study, 386 differential expressed lncRNAs were obtained subsequent to merging 262 lncRNAs in DE set A and 365 lncRNAs in DE set B (Fig. 2A). In univariate regression, 337 lncRNAs (288 plus 43 plus 7 minus 1) were associated with platinum-based chemoresistance after merging the lncRNA transcripts corresponding to the same lncRNA gene, 50 lncRNAs (43 plus 7) (Fig. 2B) were associated with DFS (log-rank P<0.05). Next, a 7-lncRNA signature was identified (Table II and Fig. 2C), which was associated with platinum-based chemoresistance in the univariate and multivariate logistic regression model. The risk scores based on the 7-lncRNA signature for the prediction of platinum-based chemoresistance in patients with HGS-OvCa were also calculated. According to the risk scores, patients in the GEO training dataset were divided into low-score and high-score groups using the median risk score as a cut-off in the multivariate logistic regression model (Fig. 2D). The high-score group exhibited a higher probability of drug resistance [univariate model: Odds ratio (OR), 2.675; 95% confidence

interval (CI), 1.841-3.897; P=6.36x10^{-7}; multivariate model: OR, 2.859; 95% CI, 1.801-5.257; P=9.36x10^{-5}). The results of Kaplan-Meier curves also indicated that the high risk score was significantly associated with the poor survival of patients with HGS-OvCa and drug resistance (n=47; log-rank, P<0.05; Fig. 2E).

Risk score can be used as an independent predictor for drug resistance. In the multivariate logistic regression model of the training datasets, platinum-based chemosensitivity was a dependent variable. Platinum-based chemosensitivity has multiple covariates, including tumor stage, grade, tumor size, molecular subtypes and risk score. However, the results indicated that the risk score was an independent variable when adjusting using the covariates as aforementioned (OR, 2.859; 95% CI, 1.801-5.257; P=9.36x10^{-5}; Table III).

The results of the univariate and multivariate logistic regression analysis for the GEO validation set (n=220) and the TCGA validation set (n=233) confirmed that the risk score of the training set was associated with platinum-based chemoresistance of patients with HGS-OvCa (univariate model of the GEO dataset: OR, 4.190; 95% CI, 2.974-4.685; P<4.68x10^{-14}; multivariate model of the GEO dataset: OR, 4.210; 95% CI, 2.967-6.366; P<1.15x10^{-13}; univariate model of TCGA dataset: OR, 2.315; 95% CI, 1.852-2.956; P=2.33x10^{-5}; and multivariate model of TCGA dataset: OR, 2.514; 95% CI, 1.817-3.354; P=2.65x10^{-5}; Table III). Furthermore, clinical characteristics and HGS-OvCa subtypes (differentiated, immunoreactive, proliferative and mesenchymal) were not associated with platinum-based resistance in two validation datasets. Collectively, these results suggest that risk score has the potential to be an independent predictor for chemoresistance of HGS-OvCa cancer.

Predictive lncRNA signature performance evaluation. The risk score of the 7-lncRNA signature indicated a high sensitivity and specificity in predicting platinum-based chemoresistance in the GEO training (AUC=0.848), GEO validation (AUC=0.901) and TCGA validation (AUC=0.818) datasets (Fig. 3A). Compared with the lncRNA signature of the study by Liu *et al* (37), the lncRNA signature developed in the study by Zhou *et al* (36) and the TCGA mRNA signature, the 7-lncRNA signature exhibited an improved predictive performance (Fig. 3B-D).

Verification of the expression level of 7 lncRNAs in microarray and RNA-seq data. A total of 7 lncRNAs were observed in the GEO validation dataset, while F11-AS1 and RP11-348N5.7 were not observed in the TCGA validation dataset and RNA-seq dataset, respectively. However, all of the 7 lncRNAs were covered by validation datasets. Notably, the expression levels of 7 lncRNAs were concordant in at least two validation datasets (P<0.05; Fig. 3E).

Predicted function of the 7 lncRNAs may be associated with drug resistance mechanisms. With several cycles of cisplatin or carboplatin treatment, ZBED1-AS1 was upregulated, while F11-AS1 was downregulated by platinum in OV56 and OVKATE cell lines (P<0.05; Fig. 3F). Growth arrest specific 5 (GAS5) was downregulated by cisplatin in A2008,

Table II. Univarate and multivariate logistic regression model of 7 long non-coding RNAs in the Gene Expression Omnibus training dataset.

			Univariate regression		Multivariate regression	
Transcript ID	Gene name	Chromosome in GRCh38 (strand)	OR (95% CI)	P-value	OR (95% CI)	P-value
ENST00000426122	RNF144A-AS1	chr2:6913627-6918667 (-)	0.26 (0.080-0.797)	0.022	0.088 (0.017- 0.388)	0.026
ENST00000434796	GAS5	chr1:173863980-173864901 (-)	0.476 (0.221-0.914)	0.041	0.1898 (0.062- 0.521)	0.021
ENST00000455503	RP11-126K1.6	chr1:151346967-151348027 (+)	2.297 (1.177-4.740)	0.018	3.1549 (1.224- 9.360)	0.025
ENST00000508110	F11-AS1	chr4:186423728-186500997 (-)	0.039 (0.003-0.472)	0.014	0.009101 (0.0003- 0.208)	0.005
ENST00000515419	ZBED3-AS1	chr5:77086798-77147745 (+)	3.842 (1.173-14.310)	0.034	5.624 (1.097- 30.824)	0.003
ENST00000567832	RP11-439E19.10	chr1:246772301-246775772 (+)	43.995 (5.072-49.635)	0.001	36.666 (2.372- 826.504)	0.014
ENST00000603915	RP11-348N5.7	chr10:110910596-110912244 (+)	4.658 (1.068-22.672)	0.04	11.194 (1.470-107.695)	0.002

OR, odds ratio; 95% CI, 95% confidence interval of OR.

Table III. Univariate and multivariate logistic regression models in GEO and TCGA datasets.

	Training set (GSE63885, GSE15372, GSE51373; n=108)				GEO validation set (GSE9891; n=220)				TCGA validation set (n=233)			
	Univariate regression		Multivariable regression		Univariate regression		Multivariable regression		Univariate regression		Multivariable regression	
Variable	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)
lncRNA signature (high/low)	$6.36x10^{-7a}$	2.675 (1.841-3.897)	$9.36x10^{-5a}$	2.859 (1.801-5.257)	$4.68x10^{-14a}$	4.190 (2.974-4.685)	$1.15x10^{-13a}$	4.21 (2.967-6.366)	$2.33x10^{-5a}$	2.315 (1.852-2.956)	$2.65x10^{-5a}$	2.514 (1.817-3.354)
Age (mean, 62 years)	0.359	0.978 (0.930-1.026)	0.99	1.001 (0.929-1.081)	0.962	0.999 (0.964-1.035)	0.961	0.999 (0.947-1.053)	0.28	1.014 (0.989-1.039)	0.321	1.014 (0.986-1.045)
FIGO stage (ref=II)												
III	0.992	7.572 (3.361-21.245)	0.993	3.438 (1.625-5.218)	0.967	1.032 (0.245-5.164)	0.695	1.481 (0.219-1.204)	$4.95x10^{-3a}$	7.077 (1.964-33.252)	0.19	6.254 (1.443-34.206)
IV	0.992	5.111 (1.252-20.579)	0.993	1.615 (0.354-3.548)	0.794	1.250 (0.239-7.412)	0.262	3.869 (0.381-4.602)	0.023[a]	5.867 (1.381-31.421)	0.26	6.38 (1.213-41.476)
Tumor grade (ref=G2)												
G3	0.207	0.341 (0.047-1.589)	0.486	0.429 (0.313-4.219)	0.489	2.192 (0.311-4.656)	0.228	11.131 (0.469-65.332)	0.264	0.259 (0.012-2.274)	0.367	0.357 (0.013-3.258)
G4	0.084	0.192 (0.023-1.111)	0.114	0.105 (0.048-1.509)	0.147	6.011 (0.677-9.967)	0.056	60.499 (1.707-73.757)	0.968	0.954 (0.047-7.642)	0.726	1.555 (0.068-14.942)
Residual tumor size (ref=R0, 1-10 mm)												
R1 (11-20 mm)	0.207	0.424 (0.101-1.533)	0.155	0.225 (0.035-1.543)	0.201	2.361 (0.652-9.946)	0.101	4.071 (0.819-25.049)	0.982	1.016 (0.249-4.152)	0.558	1.598 (0.331-7.899)
R2 (>20 mm)	0.169	0.364 (0.079-1.472)	0.242	0.282 (0.029-2.182)	0.192	2.751 (0.679-12.741)	0.085	4.932 (0.869-35.164)	0.445	1.136 (0.359-3.559)	0.942	1.049 (0.282-3.835)
Histology subtypes (ref, differentiated)												
Immunoreactive	0.221	0.407 (0.089-1.642)	0.404	2.81 (0.249-3.531)	0.433	1.389 (0.614-3.201)	0.837	0.885 (0.274-2.860)	0.409	1.469 (0.601-3.791)	0.377	1.549 (0.597-4.221)
Proliferative	0.04[a]	0.221 (0.047-0.889)	0.855	1.26 (0.106-1.644)	0.02[a]	2.428 (1.134-5.385)	0.388	1.651 (0.533-5.269)	0.792	0.905 (0.428-1.904)	0.774	0.885 (0.382-2.041)
Mesenchymal	0.162	0.333 (0.066-1.505)	0.63	1.793 (0.171-2.157)	0.241	1.681 (0.707-4.058)	0.585	1.397 (0.422-4.725)	0.584	0.802 (0.362-1.781)	0.964	0.98 (0.414-2.336)

[a]P<0.05. OR, odds ratio; 95 CI%, 95% confidence interval; GEO, Gene Expression Omnibus; TCGA, The Cancer Genome Atlas; lncRNA, long non-coding RNA.

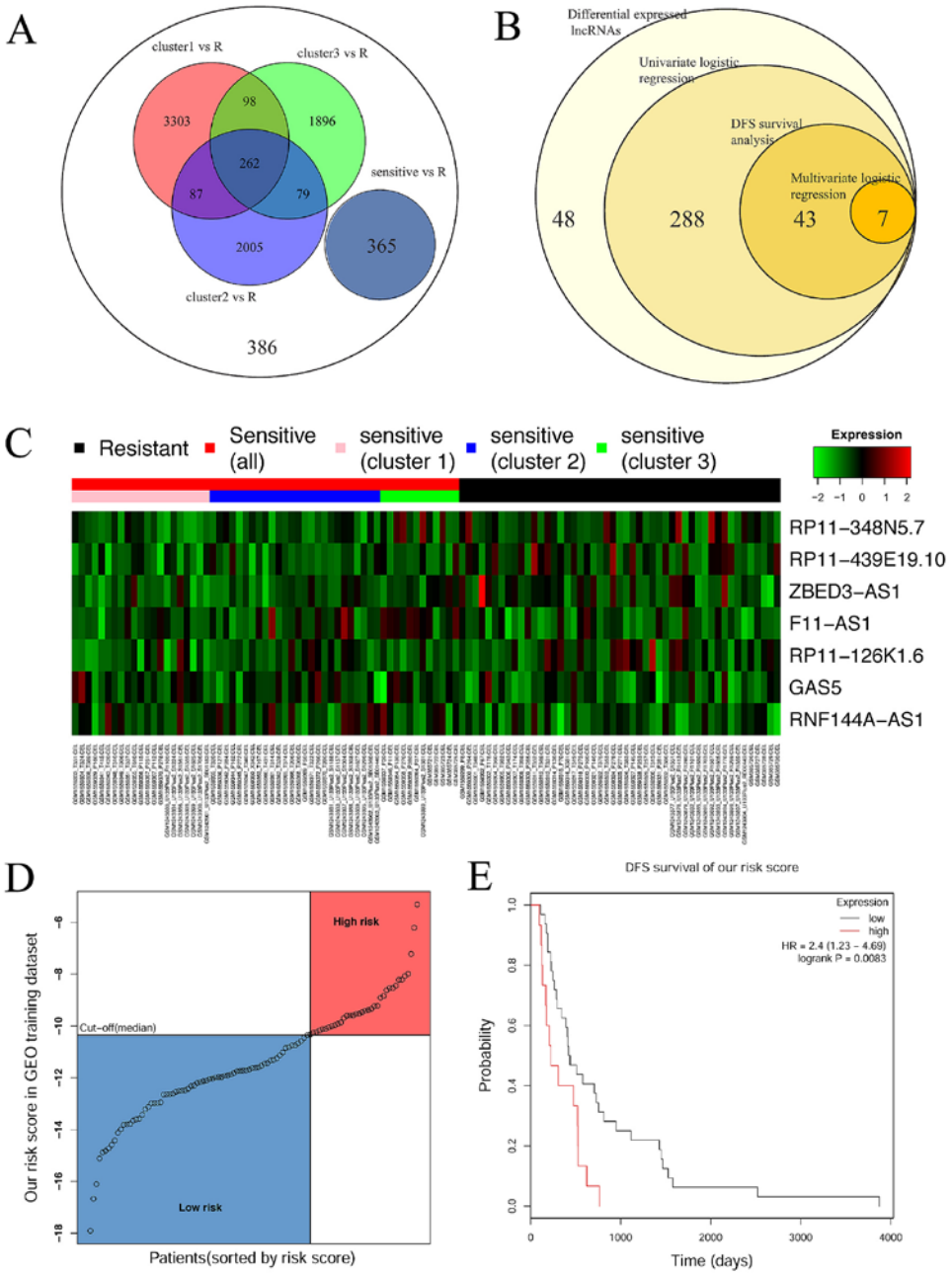

Figure 2. Heatmap of 7 lncRNAs and the distribution and DFS survival of risk score. (A) Differential expression analysis of the clusters. (B) The core steps involved in the identification of the 7 lncRNAs. (C) A heatmap based on the 7 lncRNAs (rows) of patients with ovarian cancer (columns) in the GEO training dataset. Red and blue indicate high and low expression levels, respectively. (D) The range of risk scores in the GEO training dataset. The median risk score was used as a cut-off value. Red and blue indicate high and low risk scores, respectively. (E) The Kaplan-Meier curve of DFS between low- and high-risk patients. Patients with high-risk scores had lower survival rates. Risk scores of each patient in the GEO entire dataset were sorted by risk score. DFS, disease-free survival; GEO, Gene Expression Omnibus. HR, hazard ratio; R, resistance; lncRNA, long non-coding RNA.

DOV13B and HeyA8 cell lines (P<0.05; Fig. 3G). This indicates that these lncRNAs may be involved in drug resistance mechanisms. Therefore, the functional prediction analysis of 7 lncRNAs based on NONCODE database was further performed. The results showed that a total of 42 GO terms were involved in the biological functions of 7 lncRNAs, which were distributed in 'biological processes' (91.3%) and 'cellular components' (8.7%). Furthermore, 16 of the 42 GO terms was associated with >1 lncRNA (Fig. 4). Based on ncDR, the aberrant regulation of lncRNAs was predicted to be resistant to dozens of compounds across various cancer types. For example, the upregulation of ZBED3-AS1 is associated with resistance to cisplatin, dasatinib and bicalutamide in low-grade glioma. The downregulation of RNF144A-AS1 is associated with resistance to dacinostat in sarcoma, and the downregulation of GAS5 is associated with resistance to cisplatin and cytarabine in pancreatic adenocarcinoma.

Discussion

The compilation of multiple datasets offers a unique opportunity to discover transcriptional variation in HGS-OvCa. This type of combinatorial computational analysis must take the context of numerous confounding factors, including

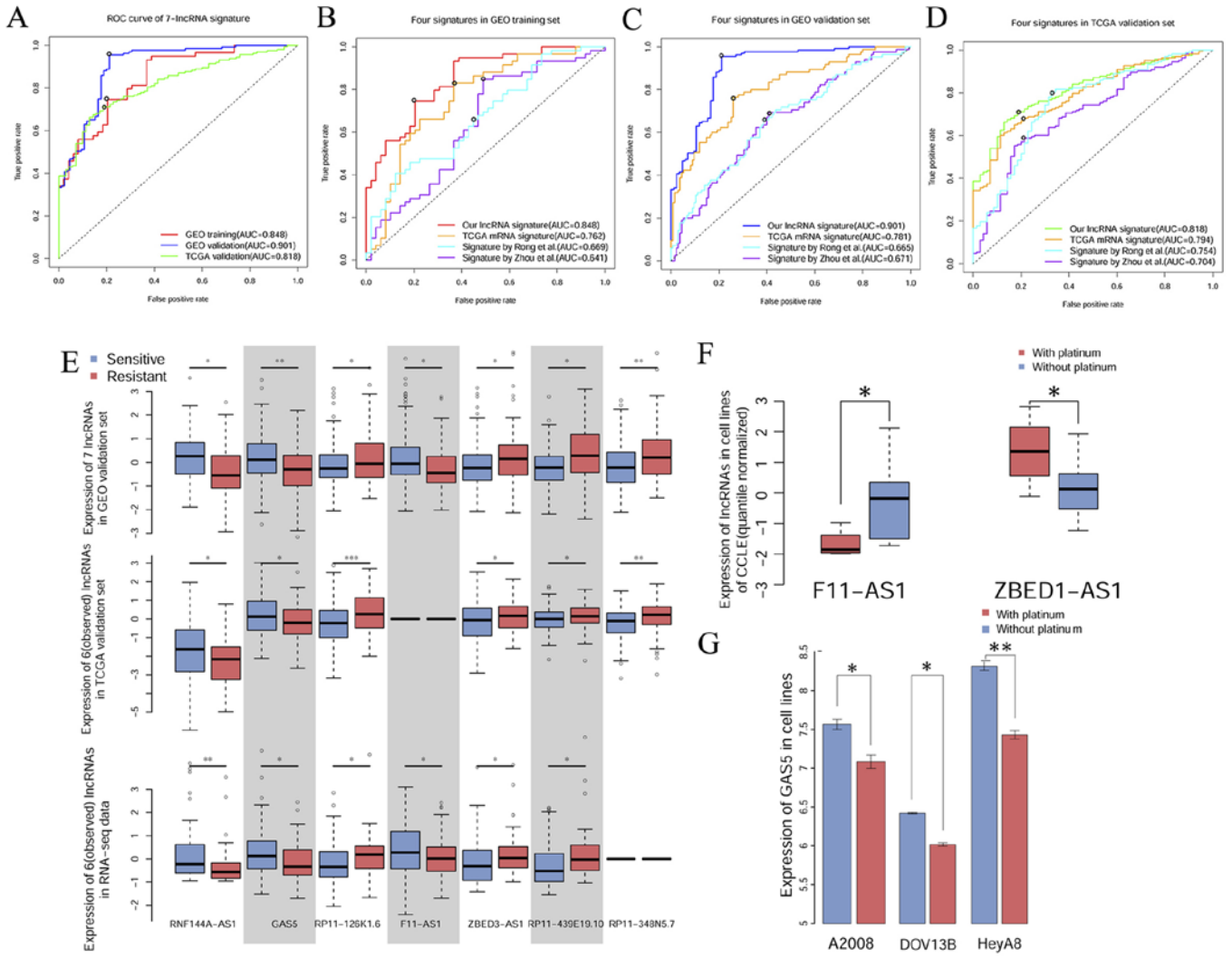

Figure 3. ROC curves of 4 signatures in 3 datasets and validation of the expression of the 7 identified lncRNAs. (A) ROC curves of the seven-lncRNA signature in three datasets. (B) ROC curves of the 4 signatures in the GEO training dataset. (C) ROC curves of the 4 signatures in the GEO validation dataset. (D) ROC curves of the 4 signatures in the TCGA validation dataset. (E) The expression levels of 7 lncRNAs in 3 different datasets, including 2 types of data sources (microarray or RNA-seq). A total of 6/7 lncRNAs were observed in the TCGA validation dataset and RNA-seq data, respectively. Blue and red boxes indicate platinum-based chemotherapeutic sensitive and resistant high-grade serous ovarian carcinoma patients, respectively. (F) The expression levels of lncRNA ZBED1-AS1 and F11-AS1 in 23 types of cell lines from CCLE. (G) The expression levels of lncRNA GAS5 in 3 types of cell lines. The best cutoff of each ROC curve was calculated and indicated by a small circle. True-positive and false-positive rates indicate sensitivity and one minus specificity, respectively. For visualization and comparison, expression levels of 3 datasets and cell lines were previously normalized using Z-score. *P<0.05, **P<0.01 and ***P<0.001. GAS5, growth arrest specific 5; lncRNA, long non-coding RNA; RNA-seq, RNA-sequencing; ROC, receiver operating characteristic curve; GEO, Gene Expression Omnibus; TCGA, The Cancer Genome Atlas; AUC, area under the curve.

distinct platforms, different protocols of sample collection and processing, and different normalization algorithms of expression level, which may not associate with tumor biology or consider cancer behavior (45). Therefore, in the present study, the independent analysis of multiple datasets was performed, and batch effect was removed. Furthermore, the verification of the expression levels of the identified lncRNAs was conducted using multiple datasets and cell lines. The results from multiple datasets were concordant and verified that the 4 upregulated lncRNAs (RP11-126K1.6, ZBED3-AS1, RP11-439E19.10 and RP11-348N5.7) and 3 downregulated lncRNAs (RNF144A-AS1, GAS5 and F11-AS1) were differentially expressed in response to platinum. In addition, not only the 7-lncRNA signature, but also the individual lncRNAs themselves were associated with the prognosis of patients with HGS-OvCa.

The molecular mechanisms of drug resistance in ovarian cancer are mainly involved in apoptosis evasion, escape from immune surveillance and natural killer cell-mediated oncolysis, promotion of DNA repair, and sustained angiogenesis, proliferation, invasion and metastasis (6,47,48). To date, the majority of the lncRNAs reported have not been functionally characterized in drug-resistant ovarian cancer.

GAS5 has been reported as a biomarker of drug resistance in non-small cell lung cancer (49) and breast cancer (50). The downregulation of GAS5 promotes the proliferation, migration and invasion of ovarian cancer cells, and inhibits apoptosis, which leads to a poor prognosis (51,52). From the results of functional prediction in the present study, the downregulation of RNF144A-AS1 and GAS5, and the upregulation of RP11-126K1.6, RP11-439E19.10 and RP11-348N5.7 in patients with ovarian cancer may contribute to platinum resistance

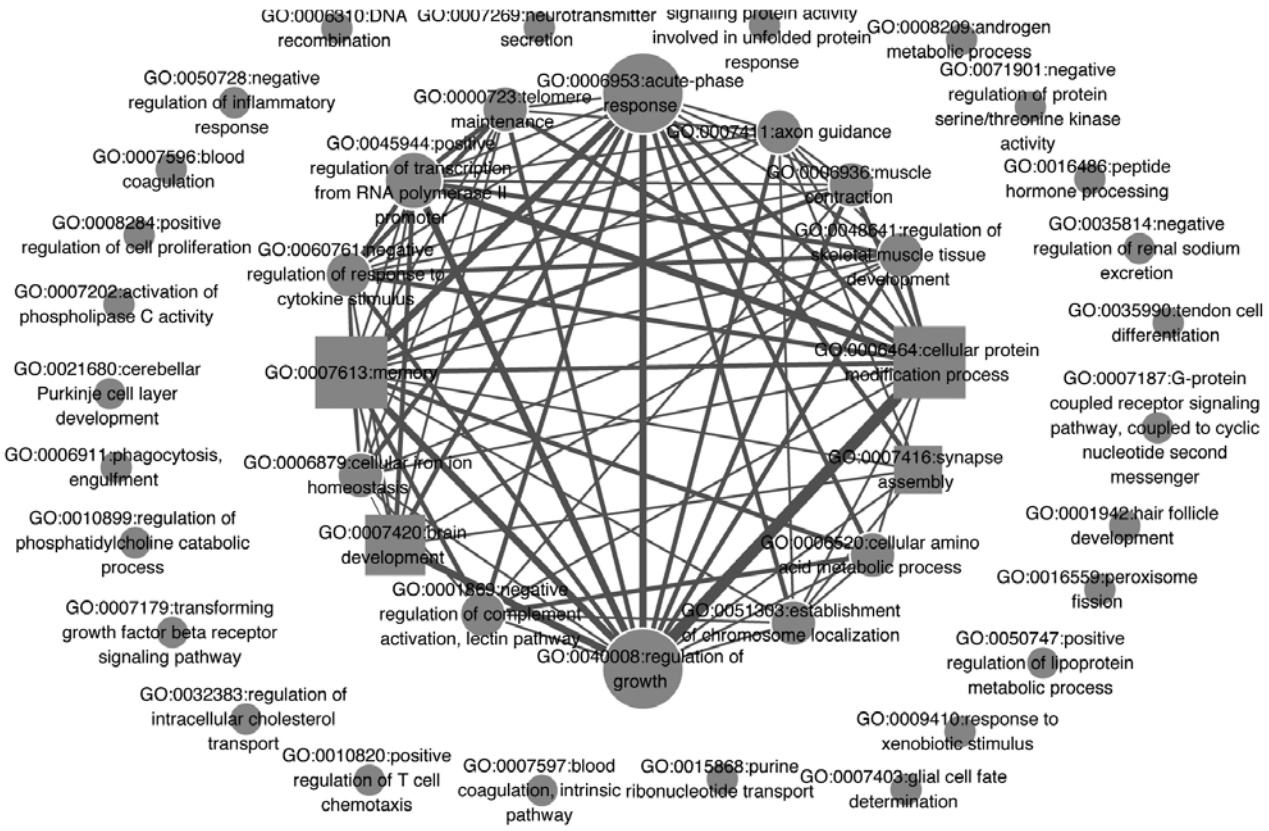

Figure 4. Interaction network of the predicted function of 7 lncRNAs based on the NONCODE database. Red nodes represent GO biological process terms and yellow nodes represent GO cellular component terms. Node size is proportional to the total number of lncRNAs in that term. Thickness of green lines is proportional to the shared lncRNAs of two connected terms. GO, Gene Ontology; lncRNA, long non-coding RNA.

by affecting the regulation of tumor growth. RNF144A-AS1, F11-AS1, ZBED3-AS1, RP11-439E19.10 and RP11-348N5.7 were predicted to participate in the acute-phase response. The acute-phase response serves as a core of the innate immune response, which can be triggered by inflammation (53). It has been confirmed that inflammation and immunogenic-tumor microenvironment interactions will increase the risk of ovarian tumor initiation and progression by overexpression of several proinflammatory cytokines (54,55). For example, the complex formed by the binding of the proinflammatory cytokine interleukin-6 (IL-6) to its receptor IL-6R activates Janus kinase and various downstream effectors, including signal transducer and activator of transcription 3 and mitogen-activated protein kinase, which are critical for cell proliferation, apoptosis evasion, survival and drug resistance in ovarian cancer (6,56,57). Taken together, the present results indicated that the 5 lncRNAs (RNF144A-AS1, F11-AS1, ZBED3-AS1, RP11-439E19.10 and RP11-348N5.7) may interact with proinflammatory cytokines to promote ovarian tumor initiation and progression. Additionally, the ncDR database also indicated that the identified lncRNAs were involved in resistance to platinum-based chemotherapeutics across various cancer types.

The present study has a number of limitations. Once multiple-transcript probes have been identified as candidate features of the multivariate logistic regression model, multicollinearity could be a concern for the accuracy of predictors. Multicollinearity does not reduce the predictive reliability of the entire signature. In the present study, there was no collinearity between other lncRNAs and the 7 candidate lncRNAs. In addition, lncRNA records vary in well-known biological databases. Although in the present study the lncRNAs records from multiple databases were re-integrated for lncRNA identification, the omission of identification of lncRNAs may still occur.

In the present study, a 7-lncRNA signature for predicting platinum-based chemoresistance in patients with HGS-OvCa was developed. The signature is independent of clinical characteristics and molecular subtypes of HGS-OvCa, and can be used as a novel prognostic marker, with the potential to be a therapeutic molecular target in the future. Further clinical studies are required for validating the lncRNAs identified in the present study and for investigating the underlying mechanisms of drug resistance in HGS-OvCa.

Acknowledgements

Not applicable.

Funding

The present study was supported by a grant from the Chongqing Science and Technology Foundation of Yuzhong District (no. 20150107).

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Authors' contributions

JS designed the study, analyzed and interpreted the data, and was a major contributor in writing the manuscript. WZ made critical advices to the conception and design of the study. SW, KL and FS contributed to data collection. LR designed the study, supervised the project and revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no conflict of interests.

References

1. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T and Thun MJ: Cancer statistics, 2008. CA Cancer J Clin 58: 71-96, 2008.
2. Bast RC Jr, Hennessy B and Mills GB: The biology of ovarian cancer: New opportunities for translation. Nat Rev Cancer 9: 415-428, 2009.
3. Alvero AB, Chen R, Fu HH, Montagna M, Schwartz PE, Rutherford T, Silasi DA, Steffensen KD, Waldstrom M, Visintin I, *et al*: Molecular phenotyping of human ovarian cancer stem cells unravels the mechanisms for repair and chemoresistance. Cell Cycle 8: 158-166, 2009.
4. Fotopoulou C, Richter R, Braicu IE, Schmidt SC, Neuhaus P, Lichtenegger W and Sehouli J: Clinical outcome of tertiary surgical cytoreduction in patients with recurrent epithelial ovarian cancer. Ann Surg Oncol 18: 49-57, 2011.
5. Husseinzadeh N: Status of tumor markers in epithelial ovarian cancer has there been any progress? A review. Gynecol Oncol 120: 152-157, 2011.
6. Hiss D: Optimizing molecular-targeted therapies in ovarian cancer: The renewed surge of interest in ovarian cancer biomarkers and cell signaling pathways. J Oncol 2012: 737981, 2012.
7. Kaye SB, Colombo N, Monk BJ, Tjulandin S, Kong B, Roy M, Chan S, Filipczyk-Cisarz E, Hagberg H, Vergote I, *et al*: Trabectedin plus pegylated liposomal doxorubicin in relapsed ovarian cancer delays third-line chemotherapy and prolongs the platinum-free interval. Ann Oncol 22: 49-58, 2011.
8. Siegel RL, Miller KD and Jemal A: Cancer statistics, 2018. CA Cancer J Clin 68: 7-30, 2018.
9. Zhang S, Balch C, Chan MW, Lai HC, Matei D, Schilder JM, Yan PS, Huang TH and Nephew KP: Identification and characterization of ovarian cancer-initiating cells from primary human tumors. Cancer Res 68: 4311-4320, 2008.
10. Lipovich L, Johnson R and Lin CY: MacroRNA underdogs in a microRNA world: Evolutionary, regulatory, and biomedical significance of mammalian long non-protein-coding RNA. Biochim Biophys Acta 1799: 597-615, 2010.
11. Wang H, Huo X, Yang XR, He J, Cheng L, Wang N, Deng X, Jin H, Wang N, Wang C, *et al*: STAT3-mediated upregulation of lncRNA HOXD-AS1 as a ceRNA facilitates liver cancer metastasis by regulating SOX4. Mol Cancer 16: 136, 2017.
12. Cui Z, Ren S, Lu J, Wang F, Xu W, Sun Y, Wei M, Chen J, Gao X, Xu C, *et al*: The prostate cancer-up-regulated long noncoding RNA PlncRNA-1 modulates apoptosis and proliferation through reciprocal regulation of androgen receptor. Urol Oncol 31: 1117-1123, 2013.
13. Wu DD, Chen X, Sun KX, Wang LL, Chen S and Zhao Y: Role of the lncRNA ABHD11-AS1 in the tumorigenesis and progression of epithelial ovarian cancer through targeted regulation of RhoC. Mol Cancer 16: 138, 2017.
14. Parasramka M, Yan IK, Wang X, Nguyen P, Matsuda A, Maji S, Foye C, Asmann Y and Patel T: BAP1 dependent expression of long non-coding RNA NEAT-1 contributes to sensitivity to gemcitabine in cholangiocarcinoma. Mol Cancer 16: 22, 2017.
15. Yap TA, Carden CP and Kaye SB: Beyond chemotherapy: Targeted therapies in ovarian cancer. Nat Rev Cancer 9: 167-181, 2009.
16. Rowan K: Intraperitoneal therapy for ovarian cancer: Why has it not become standard? J Natl Cancer Inst 101: 775-777, 2009.
17. Benedet JL, Bender H, Jones H III, Ngan HY and Pecorelli S; FIGO Committee on Gynecologic Oncology: FIGO staging classifications and clinical practice guidelines in the management of gynecologic cancers. Int J Gynaecol Obstet 70: 209-262, 2000.
18. Lisowska KM, Olbryt M, Dudaladava V, Pamuła-Piłat J, Kujawa K, Grzybowska E, Jarząb M, Student S, Rzepecka IK, Jarząb B, *et al*: Gene expression analysis in ovarian cancer - faults and hints from DNA microarray study. Front Oncol 4: 6, 2014.
19. Koti M, Gooding RJ, Nuin P, Haslehurst A, Crane C, Weberpals J, Childs T, Bryson P, Dharsee M, Evans K, *et al*: Identification of the IGF1/PI3K/NF κB/ERK gene signalling networks associated with chemotherapy resistance and treatment response in high-grade serous epithelial ovarian cancer. BMC Cancer 13: 549, 2013.
20. Li M, Balch C, Montgomery JS, Jeong M, Chung JH, Yan P, Huang TH, Kim S and Nephew KP: Integrated analysis of DNA methylation and gene expression reveals specific signaling pathways associated with platinum resistance in ovarian cancer. BMC Med Genomics 2: 34, 2009.
21. Tothill RW, Tinker AV, George J, Brown R, Fox SB, Lade S, Johnson DS, Trivett MK, Etemadmoghadam D, Locandro B, *et al*: Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. Clin Cancer Res 14: 5198-5208, 2008.
22. Sugio A, Yang B, Zhu T and White FF: Two type III effector genes of *Xanthomonas oryzae* pv. oryzae control the induction of the host genes OsTFIIAgamma1 and OsTFX1 during bacterial blight of rice. Proc Natl Acad Sci USA 104: 10720-10725, 2007.
23. Miller JA, Cai C, Langfelder P, Geschwind DH, Kurian SM, Salomon DR and Horvath S: Strategies for aggregating gene expression data: The collapseRows R function. BMC Bioinformatics 12: 322, 2011.
24. Li J, Han L, Roebuck P, Diao L, Liu L, Yuan Y, Weinstein JN and Liang H: TANRIC: An interactive open platform to explore the function of lncRNAs in cancer. Cancer Res 75: 3728-3737, 2015.
25. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, *et al*: Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 6: pl1, 2013.
26. Volders PJ, Helsens K, Wang X, Menten B, Martens L, Gevaert K, Vandesompele J and Mestdagh P: LNCipedia: A database for annotated human lncRNA transcript sequences and structures. Nucleic Acids Res 41: D246-D251, 2013.
27. Du Z, Fei T, Verhaak RG, Su Z, Zhang Y, Brown M, Chen Y and Liu XS: Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. Nat Struct Mol Biol 20: 908-913, 2013.
28. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, *et al*: GENCODE: The reference human genome annotation for The ENCODE Project. Genome Res 22: 1760-1774, 2012.
29. Liu C, Bai B, Skogerbø G, Cai L, Deng W, Zhang Y, Bu D, Zhao Y and Chen R: NONCODE: An integrated knowledge database of non-coding RNAs. Nucleic Acids Res 33: D112-D115, 2005.
30. Wilkerson MD and Hayes DN: ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. Bioinformatics 26: 1572-1573, 2010.
31. Smyth GK: Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3: 1-25, 2004.
32. Hakulinen T and Abeywickrama KH: A computer program package for relative survival analysis. Comput Programs Biomed 19: 197-207, 1985.
33. Verhaak RG, Tamayo P, Yang JY, Hubbard D, Zhang H, Creighton CJ, Fereday S, Lawrence M, Carter SL, Mermel CH, *et al*; Cancer Genome Atlas Research Network: Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. J Clin Invest 123: 517-525, 2013.
34. Duda RO, Hart PE and Stork DG: Pattern Classification. China Machine Press, 2012.

35. Sing T, Beerenwinkel N and Lengauer T: Learning mixtures of localized rules by maximizing the area under the ROC curve. In: ROCAI-2004. José Hernández-Orallo (ed). 1st International Workshop on ROC Analysis in Artificial Intelligence, Valencia, pp89-96, 2004.
36. Zhou M, Sun Y, Sun Y, Xu W, Zhang Z, Zhao H, Zhong Z and Sun J: Comprehensive analysis of lncRNA expression profiles reveals a novel lncRNA signature to discriminate nonequivalent outcomes in patients with ovarian cancer. Oncotarget 7: 32433-32448, 2016.
37. Liu R, Zeng Y, Zhou CF, Wang Y, Li X, Liu ZQ, Chen XP, Zhang W and Zhou HH: Long noncoding RNA expression signature to predict platinum-based chemotherapeutic sensitivity of ovarian cancer patients. Sci Rep 7: 18, 2017.
38. Network TCGA; Cancer Genome Atlas Research Network: Integrated genomic analyses of ovarian carcinoma. Nature 474: 609-615, 2011.
39. Tomczak K, Czerwińska P and Wiznerowicz M: The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. Contemp Oncol (Pozn) 19: A68-A77, 2015.
40. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D, *et al*: The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 483: 603-607, 2012.
41. Miow QH, Tan TZ, Ye J, Lau JA, Yokomizo T, Thiery JP and Mori S: Epithelial-mesenchymal status renders differential responses to cisplatin in ovarian cancer. Oncogene 34: 1899-1907, 2015.
42. Thorvaldsdóttir H, Robinson JT and Mesirov JP: Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. Brief Bioinform 14: 178-192, 2013.
43. Consortium GO: The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res 32: D258-D261, 2004.
44. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T: Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498-2504, 2003.
45. Brannon AR, Haake SM, Hacker KE, Pruthi RS, Wallen EM, Nielsen ME and Rathmell WK: Meta-analysis of clear cell renal cell carcinoma gene expression defines a variant subgroup and identifies gender influences on tumor biology. Eur Urol 61: 258-268, 2012.
46. Watt JP, O'Brien KL, Benin AL, McCoy SI, Donaldson CM, Reid R, Schuchat A, Zell ER, Hochman M, Santosham M, *et al*: Risk factors for invasive pneumococcal disease among Navajo adults. Am J Epidemiol 166: 1080-1087, 2007.
47. Dhillon KK, Swisher EM and Taniguchi T: Secondary mutations of BRCA1/2 and drug resistance. Cancer Sci 102: 663-669, 2011.
48. Sakai W, Swisher EM, Karlan BY, Agarwal MK, Higgins J, Friedman C, Villegas E, Jacquemont C, Farrugia DJ, Couch FJ, *et al*: Secondary mutations as a mechanism of cisplatin resistance in BRCA2-mutated cancers. Nature 451: 1116-1120, 2008.
49. Zhang N, Yang GQ, Shao XM and Wei L: GAS5 modulated autophagy is a mechanism modulating cisplatin sensitivity in NSCLC cells. Eur Rev Med Pharmacol Sci 20: 2271-2277, 2016.
50. Li J, Huang H, Li Y, Li L, Hou W and You Z: Decreased expression of long non-coding RNA GAS5 promotes cell proliferation, migration and invasion, and indicates a poor prognosis in ovarian cancer. Oncol Rep 36: 3241-3250, 2016.
51. Gao J, Liu M, Zou Y, Mao M, Shen T, Zhang C, Song S, Sun M, Zhang S, Wang B, *et al*: Long non-coding RNA growth arrest-specific transcript 5 is involved in ovarian cancer cell apoptosis through the mitochondria-mediated apoptosis pathway. Oncol Rep 34: 3212-3221, 2015.
52. Nagini S: Breast Cancer: Current molecular therapeutic targets and new players. Anticancer Agents Med Chem 17: 152-163, 2017.
53. Cray C, Zaias J and Altman NH: Acute phase response in animals: A review. Comp Med 59: 517-526, 2009.
54. Clendenen TV, Lundin E, Zeleniuch-Jacquotte A, Koenig KL, Berrino F, Lukanova A, Lokshin AE, Idahl A, Ohlson N, Hallmans G, *et al*: Circulating inflammation markers and risk of epithelial ovarian cancer. Cancer Epidemiol Biomarkers Prev 20: 799-810, 2011.
55. Kulbe H, Thompson R, Wilson JL, Robinson S, Hagemann T, Fatah R, Gould D, Ayhan A and Balkwill F: The inflammatory cytokine tumor necrosis factor-alpha generates an autocrine tumor-promoting network in epithelial ovarian cancer cells. Cancer Res 67: 585-592, 2007.
56. Bollrath J and Greten FR: IKK/NF-kappaB and STAT3 pathways: Central signalling hubs in inflammation-mediated tumour promotion and metastasis. EMBO Rep 10: 1314-1319, 2009.
57. Duan Z, Foster R, Bell DA, Mahoney J, Wolak K, Vaidya A, Hampel C, Lee H and Seiden MV: Signal transducers and activators of transcription 3 pathway activation in drug-resistant ovarian cancer. Clin Cancer Res 12: 5055-5063, 2006.