Whole transcriptome analysis reveals non-coding RNA's competing endogenous gene pairs as novel form of motifs in serous ovarian cancer

Li, Haili

2022-09

Li , H , Zheng , X , Gao , J , Leung , K S , Wong , M H , Yang , S , Liu , Y , Dong , M , Bai , H , Ye , X & Cheng , L 2022 , ' Whole transcriptome analysis reveals non-coding RNA's competing endogenous gene pairs as novel form of motifs in serous ovarian cancer ' , Computers in Biology and Medicine , vol. 148 , 105881 . https://doi.org/10.1016/j.compbiomed.2022.105881

http://hdl.handle.net/10138/347825
https://doi.org/10.1016/j.compbiomed.2022.105881

Computers in Biology and Medicine 148 (2022) 105881

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Whole transcriptome analysis reveals non-coding RNA's competing endogenous gene pairs as novel form of motifs in serous ovarian cancer

Haili Li [a,1], Xubin Zheng [a,b,1], Jing Gao [c,d], Kwong-Sak Leung [b], Man-Hon Wong [b], Shu Yang [a], Yakun Liu [e], Ming Dong [f], Huimin Bai [g], Xiufeng Ye [a,**], Lixin Cheng [a,*]

[a] Shenzhen People's Hospital, The Second Clinical Medical College of Jinan University, The First Affiliated Hospital of Southern University of Science and Technology, Shenzhen, China
[b] Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong
[c] Department of Pulmonary Medicine, University of Helsinki and Helsinki University Hospital, Finland
[d] Respiratory Medicine Unit, Department of Medicine, Karolinska Institute, Solna, Stockholm, Sweden
[e] Department of Gynecology, The Fourth Hospital of Hebei Medical University, Shijiazhuang, Hebei, China
[f] Bioland Laboratory (Guangzhou Regenerative Medicine and Health Guangdong Laboratory), Guangzhou, Guangdong, China
[g] Department of Obstetrics and Gynecology, Beijing Chaoyang Hospital, Capital Medical University, Beijing, China

ARTICLE INFO

Keywords:
Serous ovarian cancer
Transcriptome
circRNA
ceRNA
Biomarker
Gene pair

ABSTRACT

The non-coding RNA (ncRNA) regulation appears to be associated to the diagnosis and targeted therapy of complex diseases. Motifs of non-coding RNAs and genes in the competing endogenous RNA (ceRNA) network would probably contribute to the accurate prediction of serous ovarian carcinoma (SOC). We conducted a microarray study profiling the whole transcriptomes of eight human SOCs and eight controls and constructed a ceRNA network including mRNAs, long ncRNAs, and circular RNAs (circRNAs). Novel form of motifs (mRNA-ncRNA-mRNA) were identified from the ceRNA network and defined as non-coding RNA's competing endogenous gene pairs (ceGPs), using a proposed method denoised individualized pair analysis of gene expression (deiPAGE). 18 cricRNA's ceGPs (cceGPs) were identified from multiple cohorts and were fused as an indicator (SOC index) for SOC discrimination, which carried a high predictive capacity in independent cohorts. SOC index was negatively correlated with the CD8+/CD4+ ratio in tumour-infiltration, reflecting the migration and growth of tumour cells in ovarian cancer progression. Moreover, most of the RNAs in SOC index were experimentally validated involved in ovarian cancer development. Our results elucidate the discriminative capability of SOC index and suggest that the novel competing endogenous motifs play important roles in expression regulation and could be potential target for investigating ovarian cancer mechanism or its therapy.

1. Introduction

Ovarian cancer is the most lethal malignancy worldwide in gynaecology [1]. According to estimates from the American Cancer Society, 1 in 78 women will suffer from ovarian cancer. Furthermore, around 21,750 will be newly diagnosed and 13,940 will die from ovarian cancer in 2020. Epithelial ovarian carcinomas account for 90% of ovarian cancer cases and serous ovarian carcinoma (SOC) is thus far the most common subtype in epithelial ovarian carcinomas [2,3].

Non-coding RNAs (ncRNAs) are RNA molecules that are not translated into proteins, which include transfer RNAs, micro RNA (miRNAs), circular RNAs (circRNAs) and long ncRNAs (lncRNAs). It was found in the recent years that non-coding RNAs, especially miRNAs and circRNAs, were associated with complex diseases [4,5]. Many miRNAs were identified to be associated with diseases by computational methods such as learning based methods [6,7] and matrix factorization [8–11]. Therefore, ncRNAs are probably associated with SOC and may play critical role in the diagnosis and the treatment of SOC.

NcRNA could exert its influence through gene regulatory network [12]. One type of the regulatory mechanism is the competing endogenous RNAs (ceRNAs). CeRNA represents a regulation mode in which ceRNAs interact with other RNAs by competing for the shared target

* Corresponding author.
** Corresponding author.
E-mail addresses: szrosexiu@126.com (X. Ye), easonlcheng@gmail.com (L. Cheng).
[1] These authors contributed equally to this study.

https://doi.org/10.1016/j.compbiomed.2022.105881
Received 6 June 2022; Received in revised form 30 June 2022; Accepted 16 July 2022
Available online 20 July 2022

microRNAs [13,14]. Chiu et al. validated the ceRNA regulatory network in prostate and breast adenocarcinomas [15]. Wang et al. found that lncRNA small nucleolar RNA host gene 16 (SNHG16), acting as a ceRNA, played important roles in the immune processes and was upregulated in myasthenia gravis patients [16]. Liang et al. constructed a ceRNA regulatory network for mesenchymal ovarian cancer and identified the downregulation of lncRNA pro-transition associated R (PTAR) potentially inhibiting cancer metastasis by sponging miR-101 [17]. PTAF is a pivotal regulator of the epithelial-to-mesenchymal transition promoting the invasion–metastasis cascade of ovarian cancer. Furthermore, they demonstrated that the overexpression of PTAF can upregulate snail family zinc finger 2 (SNAI2) by directly sponging miR-25, leading to the promotion of ovarian cancer epithelial-to-mesenchymal transition and invasion [17].

Investigation about ceRNAs in SOC concentrated on the regulatory relation between individual ncRNAs and mRNAs, i.e., ncRNA – mRNA [10,18]. However, ncRNA may compete endogenous with several mRNAs at the same time. Moreover, current studies about ceRNA network in ovarian cancer did not cover circular RNA (circRNA) [19–23], given that no sufficient data are publicly available for ovarian cancer. As a layer of the gene regulatory network, circRNAs feature a variety of biological processes, including tumour cell proliferation, migration, and invasion [24]. Identifying motifs from ceRNA and investigating the correlation between circRNAs and other types of RNAs will shed light on the underlying molecular mechanisms of ovarian cancer.

In this study, we hypothesized the triangular relationship of ncRNA's competing endogenous gene pair (ceGP), where the genes and ncRNA composed a mRNA –ncRNA – mRNA motif in the ceRNA network and the two genes reversed in their expression between SOC and normal controls (Fig. 1A). First, we generated expression profiles from eight SOCs and eight normal ovary samples to characterise the full RNA expression pattern of human SOC using Agilent microarrays (Fig. 1B). To comprehensively understand the alteration of RNAs, we assessed the differentially expressed mRNAs, long non-coding RNAs (lncRNAs), and circular RNAs (circRNAs), and constructed a ceRNA network based on them. Next, we proposed the denoised individualized pair analysis of gene expression (deiPAGE) to select ncRNA's ceGPs (nceGPs) and developed a diagnostic indicator (SOC index) for SOC discrimination based on three cohorts including the self-profiled one. Validation on two independent cohorts and one blood cohort were performed. After that, we investigated the correlation between the SOC index and tumour infiltrating cells including the two typical T cells, $CD4^+$ and $CD8^+$. Functional enrichment and literature study were also carried out for the gene pairs in SOC index. Finally, a cricRNA's ceGP (cceGP) *BBS4*-circHUNK-*PRC1* was illustrated as an example for the nceGP relationship in SOC.

2. Results

2.1. Overall design of the study

In this study, we first generated expression profiles from eight SOCs and eight normal ovary samples using microarrays (Fig. 1B). Then, we

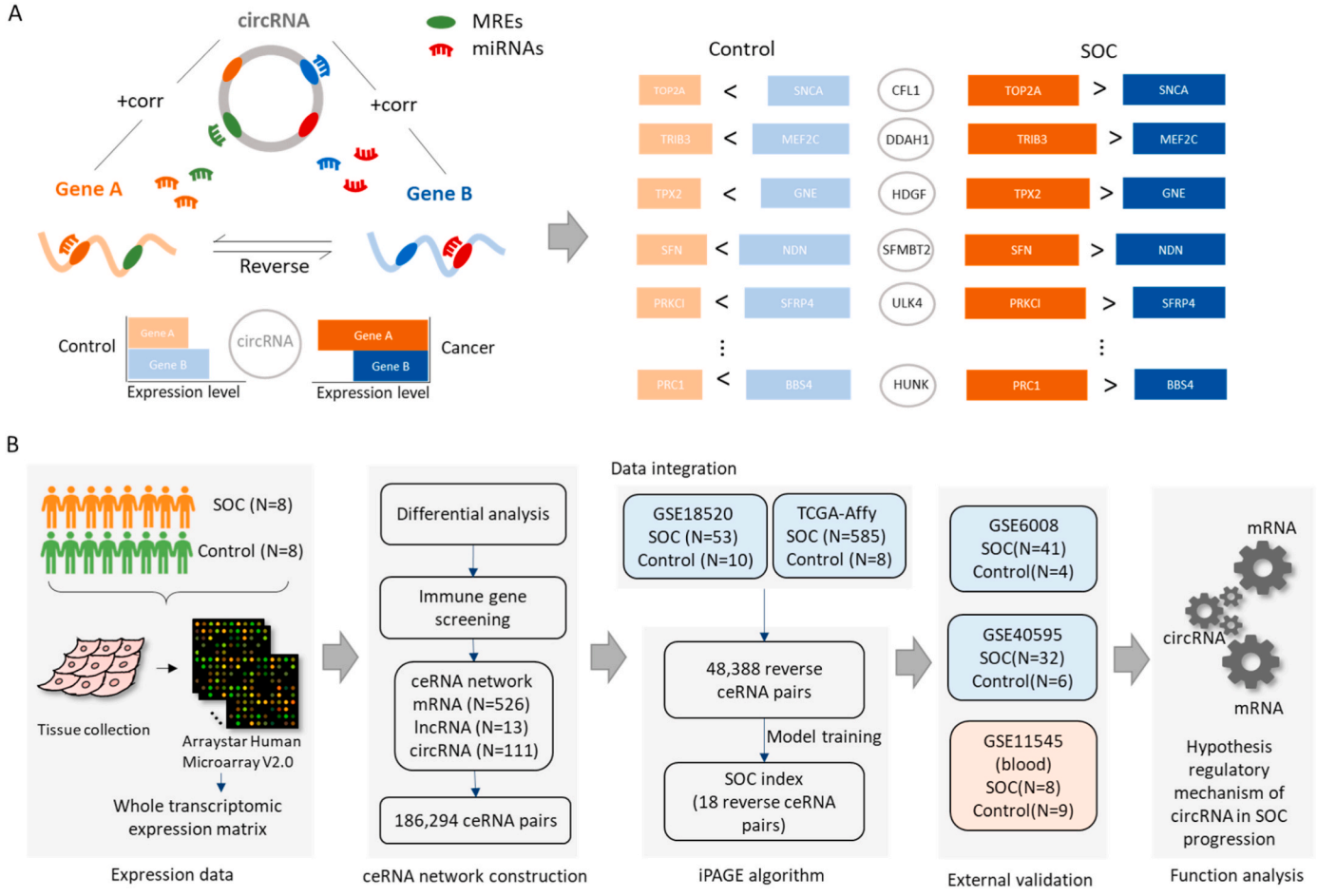


Fig. 1. Overview of this study. A) Illustration of circular RNA's competitive endogenous gene pair (cceGP), where the two genes reversed in their expression between serous ovarian carcinomas (SOCs) and normal controls. Multiple pairs were identified to discriminate SOC from normal controls. B) The workflow to study cceGPs as signature in SOC.

extracted the differentially expressed mRNAs, long non-coding RNAs (lncRNAs), and circular RNAs (circRNAs) and constructed a ceRNA network. Next, we proposed deiPAGE algorithm and identified ncRNA's ceGPs (nceGPs). Using them, we developed a diagnostic indicator, called SOC index, from TCGA, GSE18520, and the self-profiled cohorts for SOC discrimination. We evaluated the performance of SOC index on two independent cohorts GSE6008 and GSE40595 and a blood-sample cohort GSE11545. Moreover, we proposed a hypothesis about the regulatory mechanism of cceGPs in SOC progression.

2.2. Clinical characteristics and differential analysis

Aged between 41 and 66 years, eight serous ovarian carcinoma (SOC) patients and eight patients without SOC were included in this study (Table 1). Patients with SOC had lymphatic metastasis and ascites, while patients without SOC did not. The clinical tumour marker cancer antigen 125 (CA125) in two of eight SOC patients was within normal range (<35 U/mL) and only one patient had an abnormal alpha fetoprotein (AFP) level (>7 ng/mL), which reflected the low sensitivity (75% for CA125 and 12.5% for AFP) of the existing clinical protein tumour biomarkers.

We profiled these 16 samples by microarray using the Agilent platform (Table 1), yielding comprehensive expression profiles for 17,972 mRNAs, 19,394 lncRNAs, and 5,594 circRNAs. Among these, 3,876 mRNAs, 2,567 lncRNAs, and 191 circRNAs were screened as differentially expressed genes (DEGs) with an absolute fold change >2 and a false discovery rate (FDR) adjusted *P*-value from a Student's t-test <0.01 (Fig. 2A). Using GSEA [25], functional enrichment analyses showed that these DEGs are remarkably involved in the Kyoto Encyclopedia of Genes and Genomes (KEGG) [26] pathways related to cancer immunity, including T-cell activation, cytokine production, immune regulation, and tumour-infiltrating lymphocyte differentiation and migration (Fig. 2B and Fig. S1). Therefore, we only concentrated on the immune-related genes for further analysis. According to InnateDB [27] and ImmLnc [28], 884 mRNAs and 426 lncRNAs of these DEGs are immune-related (Fig. 2C and D).

2.3. Construction of competing endogenous regulatory network

The competitive endogenous RNA (ceRNA) hypothesis revealed an intrinsic mechanism within RNAs that regulate biological processes and has been validated by many studies [29]. LncRNAs, circRNAs, and mRNAs act as miRNA sponges or ceRNAs by competing for the shared microRNAs (miRNAs). For instance, in the mRNA–miRNA–lncRNA interaction, changes in the expression of lncRNA alter the number of unbound miRNAs, thereby affecting the expression abundance of the target mRNA. To disclose the RNAs participating as ceRNAs in SOC, we screened the positively correlated circRNAs/lncRNAs and mRNAs in SOC samples (Pearson's correlation coefficient (PCC) > 0.5) and selected those that have interactive miRNA in common based on RNAInter [30]. Ultimately, 526 mRNAs, 13 lncRNAs, and 111 circRNAs were identified and composed the competing endogenous regulatory network (Fig. 2E; Additional File 2: Table S1). Since all the competing endogenous partners are immune-related RNAs, the 111 circRNAs in the ceRNA network were determined as immune-related circRNAs for SOC. The detailed description of these immune-related circRNAs is listed in Additional File 2: Table S3.

2.4. Identification of non-coding RNAs' competitive endogenous gene pair

In addition to our cohort, we collected all publicly available tissue expression cohorts of SOC from The Cancer Genome Atlas (TCGA) [31] and the Gene Expression Omnibus (GEO) [32], including GSE18520, GSE6008, and GSE40595 (Table 2; Fig. 1B). We paired two genes if they connected to the same lncRNA or circRNA in the ceRNA regulatory network (Figs. 1A and 3E) and obtained 186,294 pairs from our self-profiled cohort. Among them, 111,967 pairs existed in the other four public cohorts. We combined the two largest cohorts—TCGA and GSE18520—with our cohort to comprise the training set consisting of 646 patients and 26 normal controls (Table 2). The remaining two cohorts served as independent validation sets.

As shown in Fig. 1A, if two genes that connected to the same non-coding RNA in ceRNA network reversed in their expression between SOC and controls, they were defined as non-coding RNA's competitive endogenous gene pair (nceGP).

Table 1
Clinical characteristics of patients with ovarian cancer and control samples.

ID	Age	FIGO staging system	Tissue	Lymphatic metastasis	Ascites	AFP (ng/mL)	CA125 (U/mL)	Tumour size (left; cm)	Tumour size (right; cm)
A1	53	Cervical squamous cell carcinoma stage IB1	NOET	no	no	–	–	–	–
A2	52	Uterine sarcoma	NOET	unknown	unknown	–	–	–	–
A3	41	Endometrial carcinoma stage IA	NOET	no	no	–	522	–	–
A4	54	Cervical squamous cell carcinoma stage IB1	NOET	no	no	–	–	–	–
A5	53	Cervical squamous cell carcinoma stage IB1	NOET	no	no	–	–	–	–
A6	61	Cervical squamous cell carcinoma stage IIA2	NOET	unknown	unknown	–	–	–	–
A7	55	Cervical squamous cell carcinoma stage IB1	NOET	no	no	–	–	–	–
A8	51	Endometrial carcinoma stage IA	NOET	no	no	–	–	–	–
C1	66	Papillary serous cystadenocarcinomas stage IIIC	SOC	yes	yes	6.37	122.5	5 × 2.5 × 2.5	2.5 × 2.5 × 1.2
C2	66	Papillary serous cystadenocarcinomas stage IV	SOC	yes	yes	3.44	15.53	8 × 7 × 7	4 × 3*3
C3	50	Papillary serous cystadenocarcinomas stage IIIC	SOC	yes	yes	2.77	3239	11 × 10 × 5	5.5 × 4 × 2.5
C4	51	Papillary serous cystadenocarcinomas stage IIIC	SOC	yes	yes	3.84	18.9	9 × 7 × 4	3 × 1.5 × 1,5
C5	47	Papillary serous cystadenocarcinomas stage IIIC	SOC	yes	yes	2.38	1818	15 × 14 × 5	
C6	41	Ovarian mucinous cystadenocarcinomas stage IIIC	SOC	yes	yes	8.6	115	3 × 2*1	4 × 3 × 2
C7	55	Papillary serous cystadenocarcinomas stage IIIC	SOC	yes	yes	2.96	730.9	1.8 × 1.3 × 0.9	1.8 × 1.2 × 0,7
C8	47	Papillary serous cystadenocarcinomas stage IIIC	SOC	yes	yes	2.15	1278	15 × 2 × 10	3 × 2.5 × 2.5

Note: FIGO, International Federation of Gynaecology and Obstetrics; NOET, normal ovarian epithelial tissue; AFP, alpha fetoprotein; SOC, serous ovarian carcinomas.

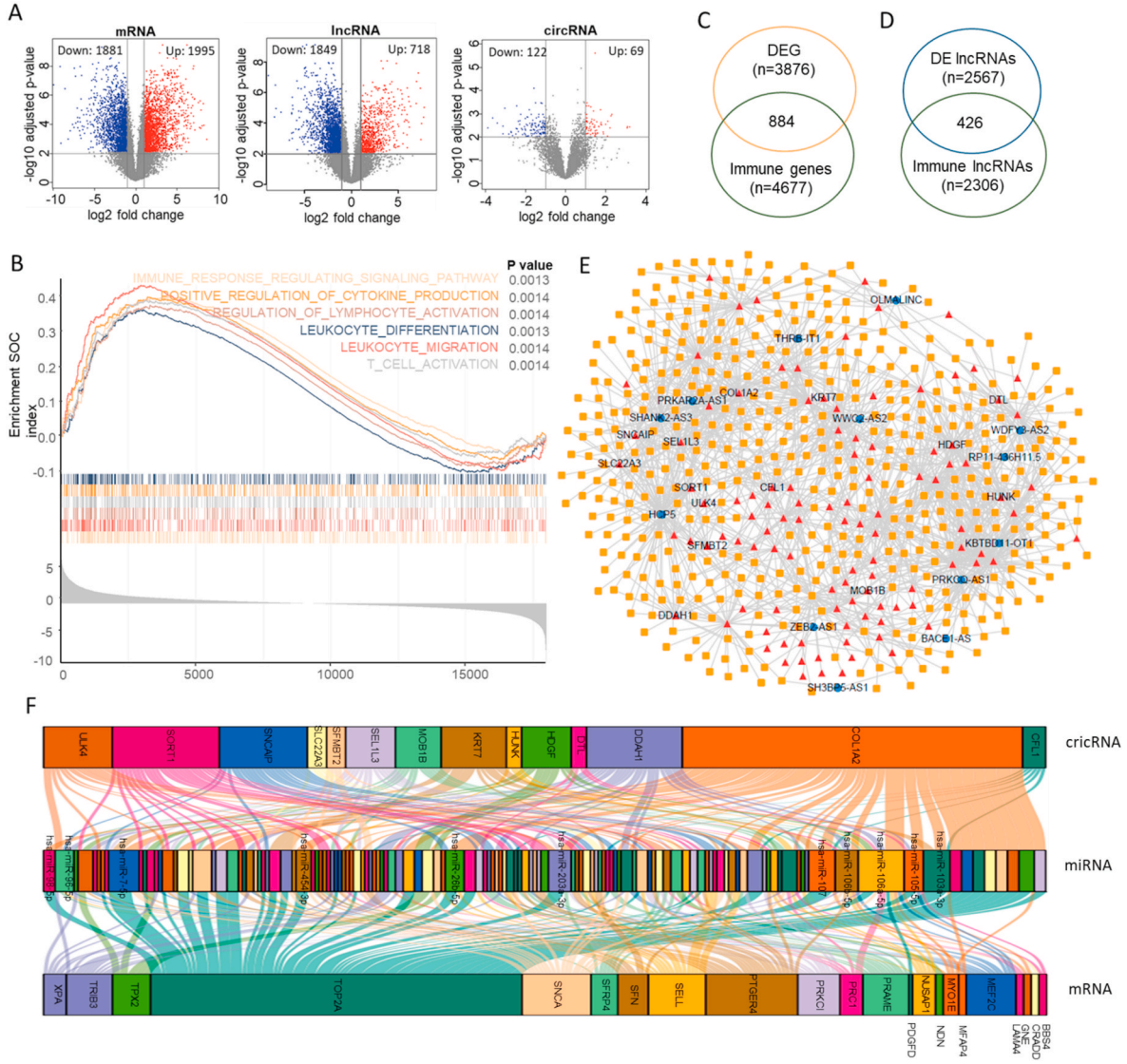


Fig. 2. Construction of the competitive endogenous RNA (ceRNA) regulatory network. A) Volcano plot of differentially expressed mRNAs, lncRNAs, and circRNAs. B) Functional analysis of differentially expressed RNAs using GSEA. A majority of the enriched biological processes are associated with immunity. C) Venn diagram of the differentially expressed genes and immune-related genes. D) Venn diagram of the differentially expressed lncRNAs and immune-related lncRNAs. E) The immune-related ceRNA regulatory network composed of mRNAs, lncRNAs, and circRNAs. F) Sankey diagram of the competitive endogenous circRNAs and mRNAs identified by deiPAGE. The miRNAs interacting with them are in the middle layer.

Table 2
Gene expression datasets used in this study.

Dataset	Cohorts	Platform	SOC	Control	Total
Training	TCGA-Affy	Affymetrix HG-U133A Array	585	8	593
	GSE18520	Affymetrix HG-U133 Plus 2.0 Array	53	10	63
	Self-profiled	Agilent- Arraystar	8	8	16
Validation 1	GSE6008	Affymetrix HG-U133A Array	41	4	45
Validation 2	GSE40595	Affymetrix HG-U133 Plus 2.0 Array	32	6	38
Blood samples Validation	GSE11545	ABI Human Genome Survey Microarray V2	8	9	17

We proposed the denoised individualized pair analysis of gene expression (deiPAGE) (Fig. 3; see the Methods section) to select nceGPs and constructed a SOC index by extracting and summarising these nceGPs. For each sample in the training cohort, we first calculated the subtraction between two genes regulated by identical ncRNAs in the ceRNA regulatory network (Fig. 3A). Then, we converted the difference between a pair of genes into a 'greater' signal (1) or 'smaller' signal (−1) with a noise interval of 0.5 to filter out some false discoveries due to technical variation (Fig. 3B). If the difference within a pair did not exceed 0.5, the pairing signal was assigned 0. The signals for all possible gene pairs were defined as the relative expression level and we derived a pairwise spectrum of 111,967 pairs. For each pair in the pairwise spectrum, we calculated the contingency table across the population (Figs. 3C), 48 and 388 DEPs were identified ($P < 0.01$, FDR corrected Fisher's exact test).

Next, we applied LASSO regression model to discriminate the SOC from the controls. After training in the training set, it selected eighteen most effective nceGPs (Table 3) to construct the SOC index as follows.

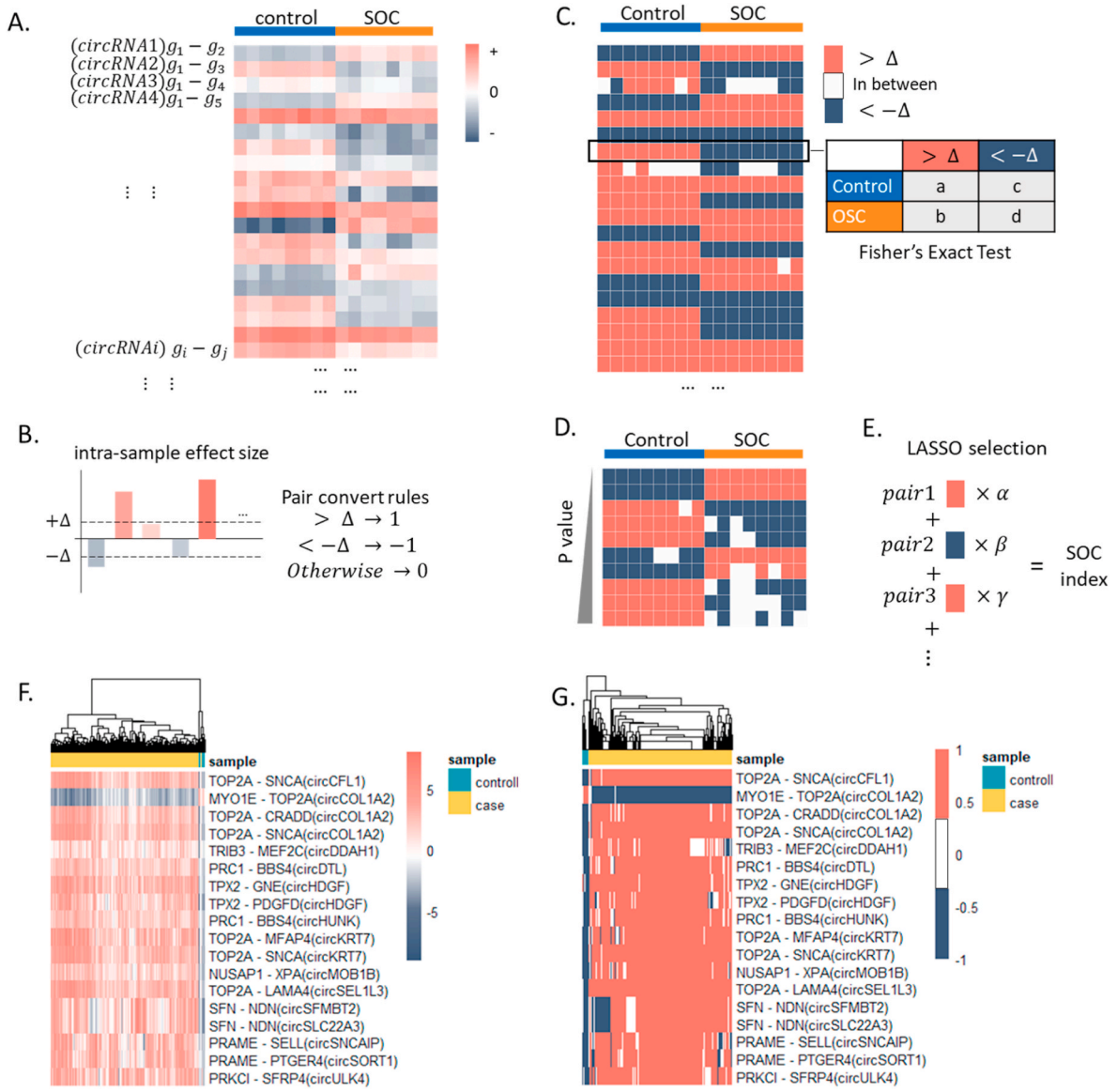


Fig. 3. The workflow of denoised individualized pair analysis of gene expression (deiPAGE) to compute the SOC index. A) Subtraction of two genes within each sample of the expression matrix. B) Gene pairs with a large effect size (Δ) were converted to intrasample relative expression indicators (1 or −1). Gene pairs with a small effect size possibly introduced from technical variation were filtered out (0). C) Based on the intra-sample rank, a Fisher's exact test was performed across the population without considering a small effect size for each gene pair. D) Gene pairs were ranked by their P-value and differentially reversed gene pairs (DRPs) were screened. E) Feature selection and construction of SOC index by machine learning method, the least absolute shrinkage and selection operator (LASSO) regression. F) Heatmap showing the subtraction result of the 18 circRNA's competitive endogenous gene pairs (cceGPs) identified by LASSO. G) Heatmap showing the subtraction result of the 18 cceGPs after conversion.

Table 3
Statistics of the 18 circular RNA's competitive endogenous gene pairs.

Non-coding RNA	Fold change	mRNA	Experimental validation	Fold change	Correlation	Reversal p value
circCFL1	1.183	TOP2A	[59,60]	5.692	0.529	1.98E-33
		SNCA	[62,68]	−2.126	0.792	
circCOL1A2	1.056	TOP2A	[58–60]	5.692	0.634	7.22E-35
		MYO1E		1.010	0.651	
circCOL1A2	1.056	TOP2A	[58–60]	5.692	0.634	1.98E-33
		CRADD		−1.057	0.738	
circCOL1A2	1.056	TOP2A	[58–60]	5.692	0.634	1.98E-33
		SNCA	[62,68]	−2.126	0.868	
circDDAH1	1.091	TRIB3	[69]	1.038	0.621	7.45E-25
		MEF2C	[70]	−2.487	0.766	
circDTL	1.082	PRC1	[38,39]	2.426	0.525	5.69E-33
		BBS4	[40]	−1.273	0.965	
cricHDGF	1.269	TPX2	[71]	4.066	0.674	1.16E-31
		GNE		−1.668	0.848	
cricHDGF	1.269	TPX2	[71]	4.066	0.674	1.66E-26
		PDGFD	[72,73]	−4.153	0.624	
cricHUNK	1.666	PRC1	[38,39]	2.426	0.741	5.69E-33
		BBS4	[40]	−1.273	0.713	
cricKRT7	1.473	TOP2A	[58–60]	5.692	0.626	6.43E-29
		MFAP4	[63]	−3.083	0.595	
cricKRT7	1.473	TOP2A	[58–60]	5.692	0.626	1.98E-33
		SNCA	[62,68]	−2.126	0.646	
cricMOB1B	−1.126	NUSAP1	[74]	1.913	0.575	2.52E-30
		XPA		−1.184	0.583	
cricSEL1L3	1.225	TOP2A	[58–60]	5.692	0.755	1.99E-35
		LAMA4	[61]	−1.415	0.835	
cricSFMBT2	−1.625	SFN	[75]	4.672	0.534	1.05E-19
		NDN	[76]	−3.375	0.637	
cricSLC22A3	−1.992	SFN	[75]	4.672	0.645	1.05E-19
		NDN	[76]	−3.375	0.522	
cricSNCAIP	−3.393	PRAME	[77]	3.491	0.522	2.31E-22
		SELL		2.024	0.521	
cricSORT1	1.349	PRAME	[77]	3.491	0.565	1.13E-24
		PTGER4		1.142	0.526	
cricULK4	1.064	PRKCI	[78,79]	1.605	0.504	2.24E-22
		SFRP4	[80,81]	−3.105	0.607	

$$
\begin{aligned}
\text{SOC index} = {} & 0.1127 * f(\text{TOP2A},\ \text{SNCA(circCFL1)}) - 0.0384 \\
& * f(\text{MYO1E, TOP2A(circCOL1A2)}) + 0.0469 \\
& * f(\text{TOP2A},\ \text{CRADD(circCOL1A2)}) + 0.0287 \\
& * f(\text{TOP2A},\ \text{SNCA(circCOL1A2)}) + 0.0034 \\
& * f(\text{TRIB3},\ \text{MEF2C(circDDAH1)}) + 0.0133 \\
& * f(\text{PRC1},\ \text{BBS4(circDTL)}) + 0.0149 \\
& * f(\text{TPX2},\ \text{GNE(cricHDGF)}) + 0.0261 \\
& * f(\text{TPX2},\ \text{PDGFD(cricHDGF)}) + 0.0012 \\
& * f(\text{PRC1},\ \text{BBS4(cricHUNK)}) + 0.0525 \\
& * f(\text{TOP2A},\ \text{MFAP4(cricKRT7)}) + 0.0029 \\
& * f(\text{TOP2A},\ \text{SNCA(cricKRT7)}) + 0.0429 \\
& * f(\text{NUSAP1},\ \text{XPA(cricMOB1B)}) + 0.0476 \\
& * f(\text{TOP2A},\ \text{LAMA4(cricSEL1L3)}) + 0.0078 \\
& * f(\text{SFN},\ \text{NDN(cricSFMBT2)}) + 3.6106 * 10^{-17} \\
& * f(\text{SFN},\ \text{NDN(cricSLC22A3)}) + 0.0022 \\
& * f(\text{PRAME},\ \text{SELL(cricSNCAIP)}) + 0.0238 \\
& * f(\text{PRAME},\ \text{PTGER4(cricSORT1)}) + 0.0027 \\
& * f(\text{PRKCI},\ \text{SFRP4(cricULK4)}) + 0.5066
\end{aligned}
$$

$$
where\ f(g_i, g_j) = \begin{cases} 1, g_i - g_j > \Delta \\ -1, g_i - g_j < -\Delta, \Delta = 0.5 \\ 0,\ otherwise \end{cases}
$$

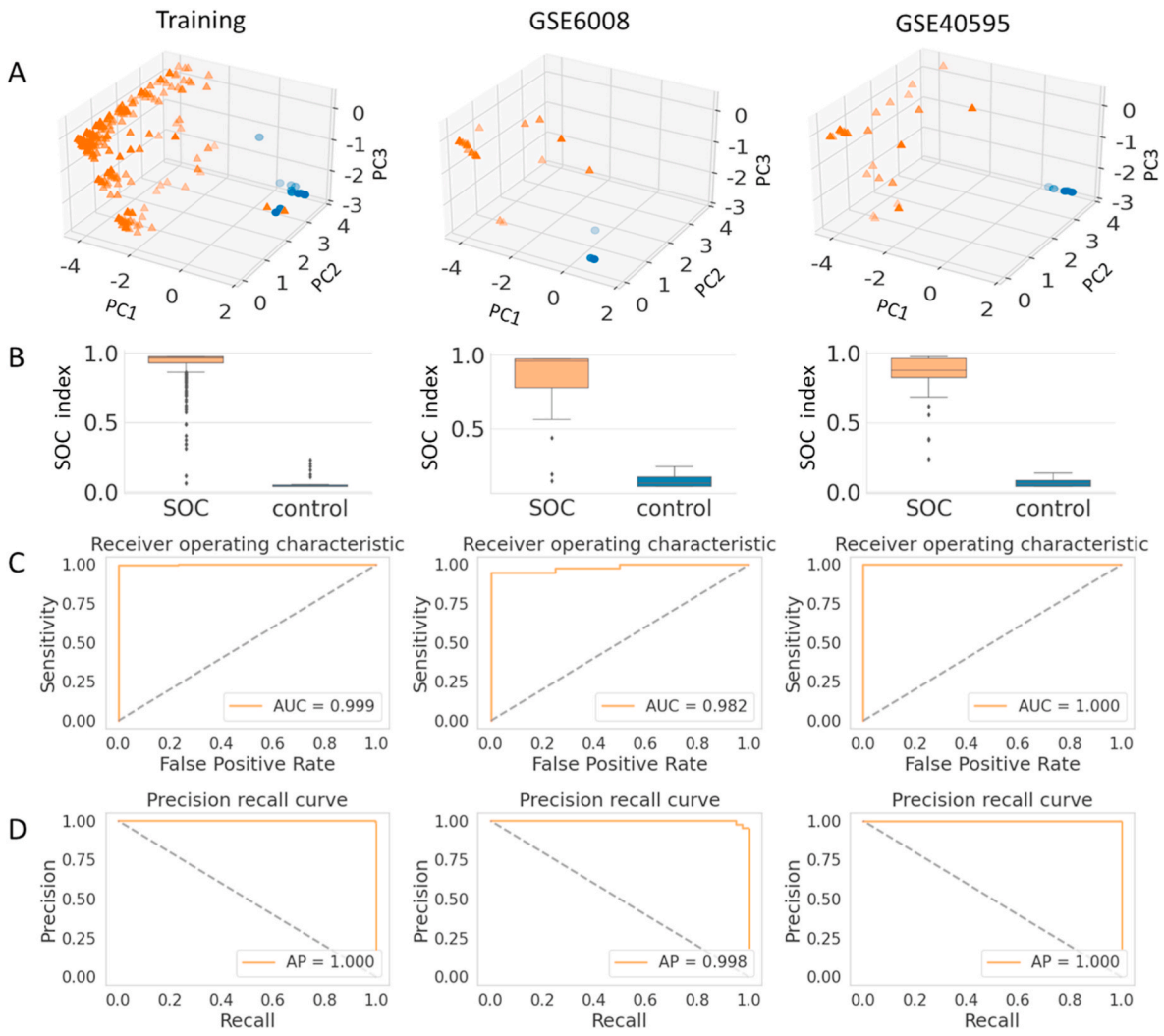
Notably, the 18 pairs selected by LASSO were all circRNA's competing endogenous gene pairs (cceGPs). A heat map displays the effect size (difference between two mRNAs in a pair) of each pair (Fig. 3F). After conversion, the denoised relative expression shows a clear distinction between SOC patients and normal controls, since all the control samples were clustered in the same group (Fig. 3G). We further retrieved the 18 cceGPs in ceRNA regulatory network and their common interactive miRNAs (Fig. 2F; Additional File 2: Table S2). Their locations in the chromosome are shown in circos plot (Additional File 1: Fig. S4).

Generally, the reversal relationships of the expression abundance of these cceGPs between SOC and control samples were consistent in the training and validation datasets (Additional File 1: Figure S5, S6 and S7). Overall, we developed a composite SOC index ranging from 0 to 1.0 on 672 samples in the training set using the LASSO regression model.

2.5. Performance of the SOC index

We performed principal components analysis (PCA) of the 18 cceGPs on the training set and two validation sets (Fig. 4A). The 3D plot of the three principal components illustrates that the 18 pairs selected by deiPAGE revealed differences between the SOCs and controls. The SOC index of SOC samples also significantly differed from controls in the three data sets (Fig. 4B). Next, we applied the receiver operating characteristic (ROC) curve and the precision-recall curve (PRC) to evaluate the model. The area under the receiver operating characteristic (AUROC) and the area under the precision-recall curve (AUPRC) of SOC index achieved 0.999 and 1.000 in the training set (Fig. 4C). The SOC index also demonstrated high sensitivity (99.7%) and specificity (92.3%) in discrimination (Fig. 4E).

To determine if the SOC index obtained from the training set is reproducible in other SOC cohorts, we applied it to two independent validation cohorts (GSE6008 and GSE40595) measured by two different platforms (Affymetrix HG-U133A Array and Affymetrix HG-U133 Plus 2.0 Array). The confusion matrix of GSE6008 (Affymetrix HG-U133A Array) indicated that the SOC index established by deiPAGE from the training set carried a sensitivity of 95.1%, a specificity of 75%, an AUROC of 0.982, and an AUPRC of 0.998 (Fig. 4C, D and 4E: GSE6008).



E

Training dataset	SOC	Normal	
Predict OSC	644	2	
Predict normal	2	24	Totals
Totals	646	26	672
Correct	644	24	668
Sensitivity (%)	99.7		
Specificity (%)		92.3	99.4

GSE6008	SOC	Normal	
Predict OSC	39	1	
Predict normal	2	3	Totals
Totals	41	4	45
Correct	39	3	42
Sensitivity (%)	95.1		
Specificity (%)		75	93.3

GSE40595	SOC	Normal	
Predict OSC	32	0	
Predict normal	0	6	Totals
Totals	32	6	38
Correct	32	6	38
Sensitivity (%)	100		
Specificity (%)		100	100

Fig. 4. Performance evaluation of the SOC index. A) The three principal components of the 18 cceGPs in the training and validation sets. B) Box plot of the SOC index in the training and validation sets. C) ROC curves of SOC index in the training and validation sets. D) Precision-recall curves of SOC index in the training and validation sets. E) Confusion matrix of SOC index in the training and validation sets.

In GSE40595, we assessed the performance of the SOC index to again distinguish SOC from normal using the Affymetrix HG-U133 Plus 2.0 Array platform (Fig. 4C, D and 4E: GSE40595). SOC index demonstrated a 100% sensitivity and 100% specificity with an AUROC of 1.000 and an AUPRC of 1.000. Current molecular biomarkers ROMA, HE4, and CA125 only achieve AUROC of 0.898, 0.857, and 0.877 [33]. In our cohort, CA125 only attains a sensitivity of 75%. Taken together, we observed highly accurate discrimination in the two independent cohorts, indicating that the SOC index carries a very high predictive value in assisting SOC detection, such as improving the accuracy of biopsy diagnosis.

Furthermore, to evaluate the non-invasive diagnosis value of the SOC index, we examined its diagnostic performance in a blood dataset GSE11545. The SOC index achieved comparable AUROC of 0.819 and AUPRC of 0.832 (Fig. 5A and B). Among the 18 cceGPs, we observed that the expression reversal of *PRKCI*-cricULK4-*SFRP4* and *TOP2A*-circKRT7-

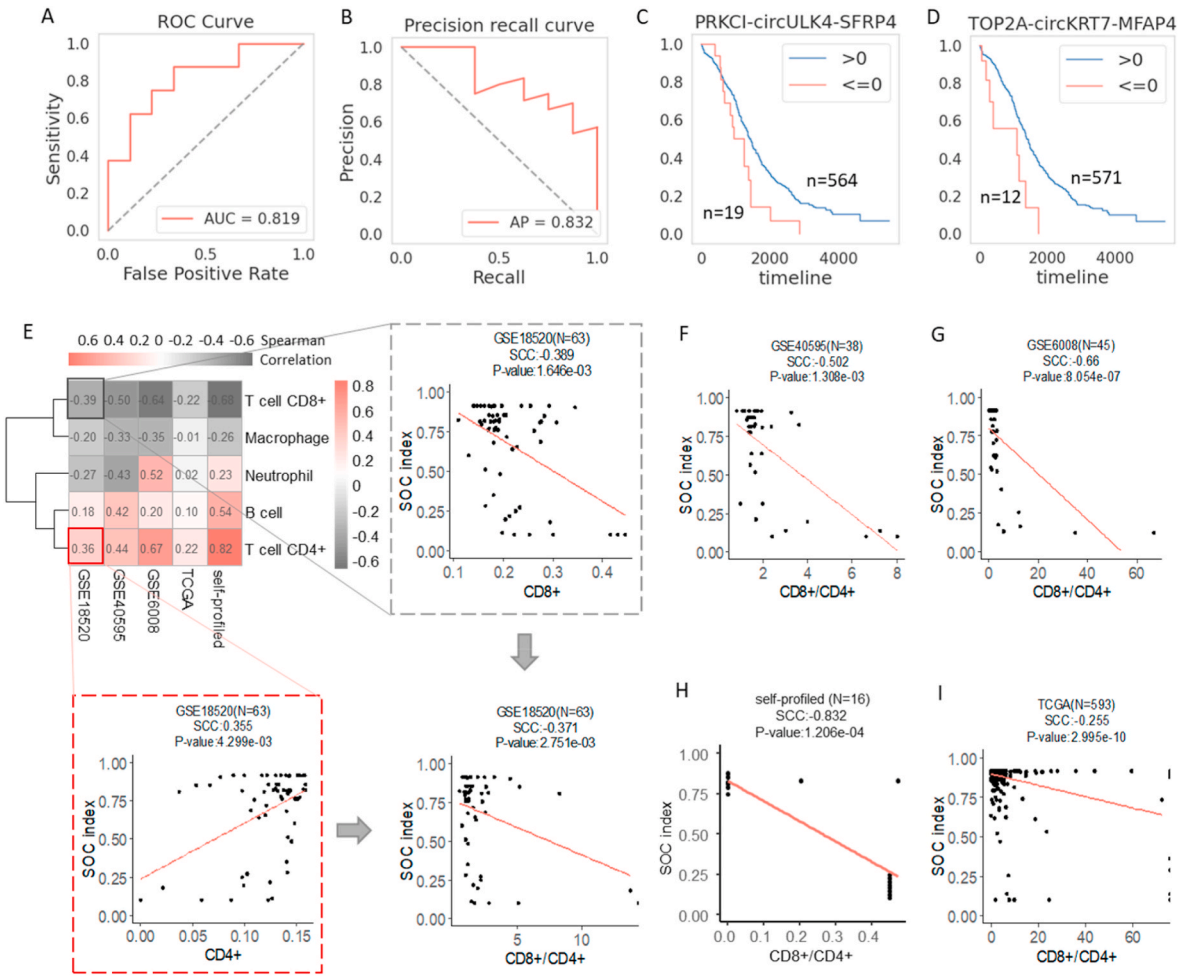


Fig. 5. The SOC index for diagnosis and prognosis, and its association with tumour progression. A) ROC curve of the SOC index in blood samples. B) Precision-recall curve of the SOC index in blood samples. C) Kaplan-Meier survival curve of PRKCI:SFRP4 pair ($P < 0.03$, log-rank test). D) Kaplan-Meier survival curve of TOP2A: MFAP4 pair ($P < 0.01$, log-rank test). E-I) Correlations between the SOC index and tumour infiltration in five cohorts (E, F, G, H, and I). The SOC index is negatively correlated with $CD8^+$ and positively correlated with $CD4^+$ in GSE18520. The SOC index is also inversely correlated with CD8+/CD4+ (E). In the other four cohorts, the SOC index and CD8+/CD4+ is consistently negatively correlated (F, G, H, and I).

MFAP4 indicated significantly better outcomes in TCGA (Fig. 5C and D).

2.6. SOC index indicates tumour infiltration

Given that the SOC index of the identified circRNA's competing endogenous gene pairs potentially discriminated between SOC and normal controls, we questioned whether the score correlated with the tumour progression. We performed a tumour-infiltrating immune cell analysis using TIMER2.0 [34], and decomposed the bulk mRNA expression into cell-type proportions for the self-profiled cohort and the other four cohorts (GSE18520, GSE40595, GSE 6008, and TCGA). We evaluated the correlation between the SOC index and the population of different cell types including $CD8^+$ T cells, $CD4^+$ T cells, B cells, macrophages, and neutrophils on five cohorts (Fig. 5E). The SOC index was consistently positively associated with $CD4^+$ T cells and negatively associated with $CD8^+$ T cells. For instance in GSE18520, the SOC index was positively correlated with $CD4^+$ T cells (Fig. 5E and Additional File 1: Fig. S8; SCC = 0.355, $P < 4.299E\text{-}03$, Spearman's correlation) and negatively correlated with $CD8^+$ T cells (Fig. 5E and Additional File 1: Fig. S8; SCC = −0.389, $P < 1.646E\text{-}03$), resulting in a significant negative correlation with the ratio of CD8+/CD4+ T cells (SCC = −0.371, $P < 2.751E\text{-}03$). The same trend was also observed in the other cohorts (Fig. 5F–I and Additional File 1: Fig. S8). The ratio of CD8+/CD4+ T cells in ovarian cancer can be used as a prognostic factor and patients with higher CD8+/CD4+ ratios tend to have improved survival [35]. The high correlation with the CD8+/CD4+ ratio indicates the potential prognostic value of the SOC index.

2.7. CircRNA's competing endogenous gene pairs

The circRNA's competing endogenous gene pair referred to the correlation between a circRNA and two genes, where the circRNA and two mRNAs are competing endogenous RNA respectively in disease, while the two genes reversed in their expression between the case and control samples. The 18 circRNA's competing endogenous gene pairs identified by deiPAGE included 14 circRNAs and 22 genes (Table 3 and Fig. 6A). Some gene pairs were regulated by multiple circRNAs while some circRNAs tartgeted more than one gene pairs. The 14 circRNAs were differentially expressed with ten up-regulated (circCFL1, circCOL1A2, circDDAH1, circDTL, circHDGF, circHUNK, circKRT7, circSEL1L3, circSORT1, and circULK4) and four down-regulated (circMOB1B, circSFMBT2, circSLC22A3, and circSNCAIP) (Fig. 6B).

The 22 genes were enriched in Reactome [36] pathways that are highly related to cancer progression, such as signal transduction, signalling by NGF, and transcriptional regulation by TP53 (Fig. 6C). Other pathways consisted of prostanoid ligand receptors, TP53 regulates transcription of caspase activators and caspases, Chk1/Chk2(Cds1) mediated inactivation of Cyclin B:Cdk1 complex, p75NTR recruits

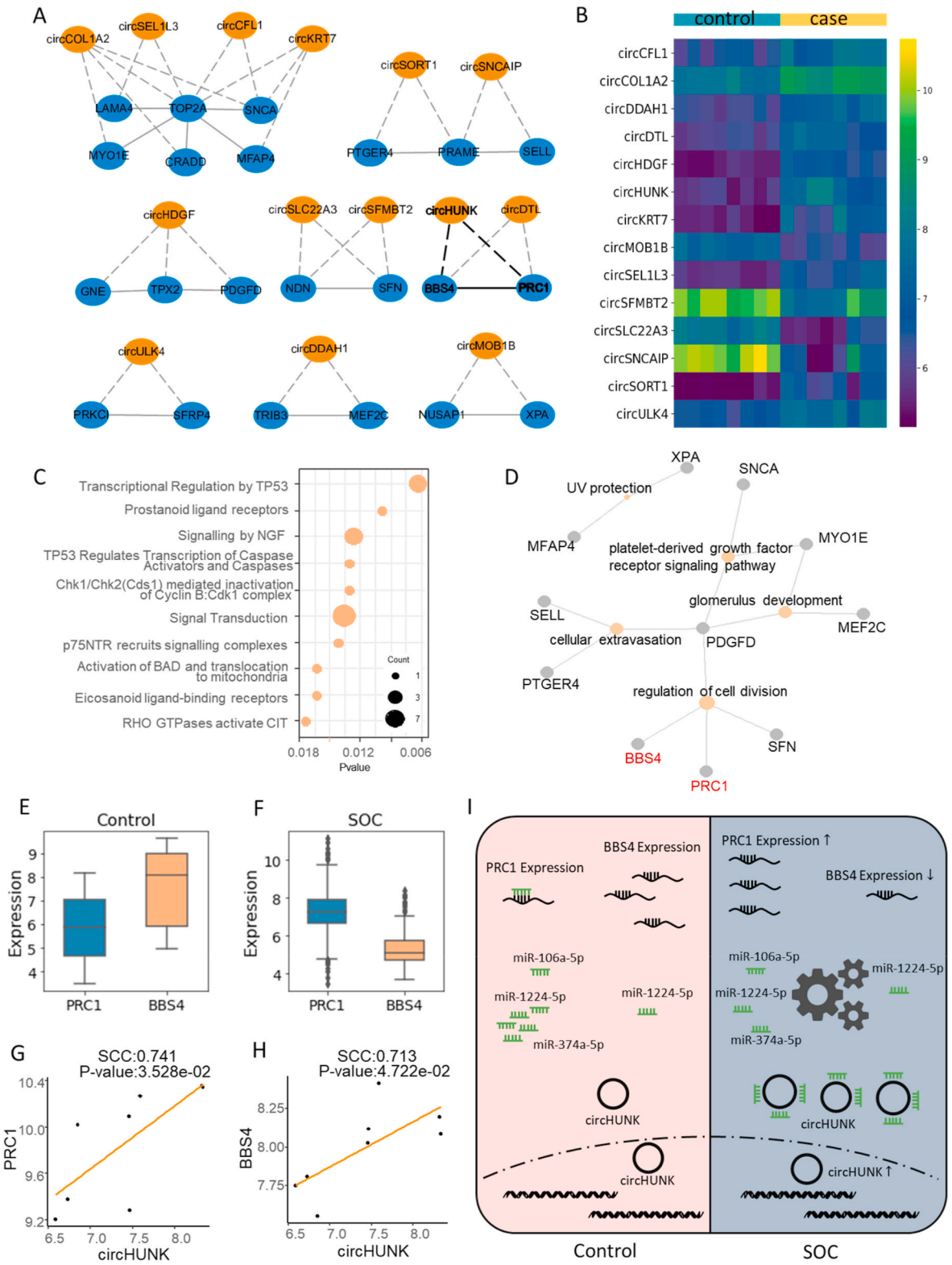


Fig. 6. Functional analysis of the cceGPs. A) The 18 cceGPs or gene pair-circRNA motifs in SOC. B) Heatmap illustrating the expression abundance of circRNAs. C) Enriched functions of the 18 cceGPs in Reactome. D) Enriched functions of the 18 cceGPs in GO. Yellow node denotes functional category while grey node represents gene. E, F) Expression abundance of *PRC1* and *BBS4* in normal controls and SOCs. G, H) Correlations between circHUNK and *PRC1* (*BBS4*) in SOCs. I) *PRC1*-circHUNK-*BBS4* as an example of cceGPs relationship in SOC progress.

signalling complexes, p75NTR recruits signalling complexes, eicosanoid ligand-binding receptors, and RHO GTPases activate CIT. We also performed functional enrichment in Gene Oncology (GO) [37] and discovered that the genes participated in the cellular growth and development such as regulation of cell division, platelet-derived growth factor receptor signalling pathway, cellular extravasation and glomerulus development (Fig. 6D). Particularly, Bardet-Biedl Syndrome 4 (*BBS4*) and Protein Regulator Of Cytokinesis 1 (*PRC1*), as competing

endogenous RNAs of circHUNK, both participated in the regulation of cell division. Other than that, most of the genes in the 18 circRNA's competing endogenous gene pairs have been reported to play key roles in carcinogenesis in previous experimental studies (Table 3).

Taking *BBS4*–circHUNK–*PRC1* as an example, we dissected the competing endogenous mechanism of the regulatory motif. The expression abundance of *PRC1* was lower than *BBS4* in normal controls, while in SOC samples it is higher than *BBS4* within most of the samples, although both of them are involved in the regulation of cell division in SOC. The boxplot showed the reverse expression pattern of *PRC1* and *BBS4* in the training set (Fig. 6E and F). The expression level of the 18 cceGPs in the training and validation sets were also provided (Additional File 1: Figure S5, S6 and S7). In the SOC samples, circHUNK was significantly positively correlated with *BBS4* and *PRC1* in expression (Fig. 6G and H). Simultaneously, circHUNK sponged the same microRNA miR-1224–5p as *BBS4* and it shared miR-106a-5p, miR-374a-5p, and miR-1224–5p with *PRC1* (Fig. 6I). It has been found that the up-regulation of *PRC1* activates the Wnt/β-catenin signalling pathway and leads to an increase in cell viability, invasion, migration and EMT of ovarian cancer cells [38]. The overexpression of *PRC1* also indicates a poor prognosis [39]. Moreover, *BBS4* has been observed down-regulated in breast cancer and indicates a shorter survival time [40]. Collectively, these findings suggest that circHUNK play potential roles in the downstream disturbance of cell division in SOV via competing endogenous RNA mechanisms (Fig. 6I). Our results provide new insight into how the ncRNA-gene pair motif works coordinately in the progression of SOC.

3. Materials and methods

3.1. Patients and samples

16 patient samples were included in this study (Table 1). Eight patients with ovarian cancer and eight patients with cervical cancer (normal ovarian tissue specimens) were recruited sterilely at the Fourth Hospital of Hebei Medical University between 2015 and 2017. This study was approved by the institutional review board of Hebei Medical University. Both cancerous and normal ovarian tissues were quickly excised and snap-frozen in a liquid nitrogen tank at −80 °C until further use.

3.2. RNA extraction

Tissues were homogenised in the TRIZOL reagent (Invitrogen, USA) using a Qiagen Tissuelyser. Total RNA was extracted in accordance with the manufacturer's protocol and then quantified using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). The RNA integrity of each sample was assessed by denaturing agarose gel electrophoresis.

3.3. Microarray experiments

To identify deregulated RNAs associated with SOC patient outcomes, we conducted a microarray study and profiled mRNAs, lncRNAs, and circRNAs, respectively. We performed Arraystar Human LncRNA Microarray V2.0 and Arraystar Human circRNA Array V2.0 analyses on all 16 samples. The expressions of lncRNAs and mRNA were quantified using the first platform, while the expressions of circRNAs were measured using the second. Total RNA from each sample was measured using NanoDrop ND-1000. Sample preparation and microarray hybridisation were performed based on the standard protocols of Arraystar (Agilent Technology, USA). Processing RNA was different between the two platforms. For the lncRNA platform, rRNA was removed from total RNA using the mRNA-ONLY™ Eukaryotic mRNA Isolation Kit (Epicentre Biotechnologies, USA). For the circRNA platform, total RNAs were digested with Rnase R (Epicentre, Inc.) to remove linear RNAs and enrich circular RNAs. Then, the two platforms followed the same steps below.

Each sample was amplified and transcribed into fluorescent cRNA utilising a random priming method (Arraystar Super RNA Labeling Kit; Arraystar). The labelled cRNAs were purified using the RNeasy Mini Kit (Qiagen, Germany). The concentration and specific activity of the labelled cRNAs (pmol Cy3/μg cRNA) were measured using NanoDrop ND-1000. Here, 1 μg of each labelled cRNA was fragmented by adding 5-μl 10 × blocking agent and 1-μl 25 × fragmentation buffer, heating the mixture to 60 °C for 30 min, and then adding 25-μl 2 × hybridisation buffer to dilute the labelled cRNA. Next, 50-μl hybridisation solution was dispensed into a gasket slide and assembled on the circRNA expression microarray slide. The slides were incubated for 17 h at 65 °C in an Agilent Hybridisation Oven. The hybridised arrays were washed, fixed, and scanned using the Agilent Scanner G2505C. We used the Agilent Feature Extraction software (version 11.0.1.1) to analyse the acquired array images.

3.4. Datasets

Other than the self-profiled cohort, we downloaded the gene expression of SOC samples from TCGA and GEO (GSE18520, GSE6008, GSE40595 and GSE11545; Table 2). We performed differential and functional analysis and constructed a competing endogenous RNA regulatory network based on the self-profiled cohort. We then combined the self-profiled cohort, GSE18520, and the TCGA cohort retrieved from UCSC Xena including 672 samples as a training set for SOC signature identification (Fig. 1B). GSE6008 and GSE40595 were applied as two independent validation sets. To verify the signature for non-invasive diagnosis, we adopted a blood cohort GSE11545 as another external validation set.

Gene expressions from the five cohorts were quantified using three platforms, i.e., the Affymetrix HG-U133A Array, the Affymetrix HG-U133 Plus 2.0 Array, and the Agilent-Arraystar. Besides, GSE11545 measured gene expression of SOC in blood using ABI Human Genome Survey Microarray V2.

3.5. Differential and functional analysis

The expression profiles of SOC patients and normal controls in the self-profiled cohort were analysed to identify differentially expressed mRNAs, lncRNAs, and circRNAs, respectively. Quantile normalization was used to normalize the profiles to render the data comparable across samples. The significance levels were estimated using the Student's t-test and, then, adjusted using the Benjamini and Hochberg (BH) multiple testing correction method. The effect size was also taken into account of using the two-fold change [41,42]. In total, we obtained 1,881 downregulated mRNAs and 1,995 upregulated mRNAs, 1,849 downregulated lncRNAs and 718 upregulated lncRNAs, as well as 122 downregulated circRNAs and 69 upregulated circRNAs.

3.6. Competing endogenous RNA network construction

The competing endogenous RNAs (ceRNAs) sponged the same micro RNAs (miRNAs) and the their expression are positively correlated. We used the Pearson's correlation test to calculate the expression correlation among mRNAs, lncRNAs, and circRNAs in SOC. Then we screened those that shared the same interactive miRNAs through RNAInter [30]. Only RNAs positively correlated (Pearson correlation coefficient (PCC) > 0.5) and interactive with the same miRNAs were considered as ceRNAs for the construction of the competing endogenous regulatory network [13,43]. The final network was illustrated using Cytoscape [44].

3.7. Denoised individualized pair analysis of gene expression

The abundance of genes may vary across different detection

platforms or preprocessing methods, but the relative ranking is stable in a pair of genes. Herein, we developed denoised individualized pair analysis of gene expression (deiPAGE). deiPAGE not only considered the relative ranking but also included the effective size between two genes. Moreover, deiPAGE applied the least absolute shrinkage and selection operator (LASSO) for feature selection.

3.7.1. Establishment of ceGPs

The gene expression profiles from different training cohorts were concatenated directly as a training set. Then, we performed pairwise subtraction for all genes to form gene pairs ($g_i - g_j$) in a single sample (Fig. 3A). However, the gene expression value may vary due to technical noise and it was denoted as $g = g' + \varepsilon$, where g' was the true value of the gene expression and $\varepsilon \in (-\infty, +\infty)$ represented the error of measurement. When subtraction was performed, it became $(g_i' + \varepsilon_i) - (g_j' + \varepsilon_j) = g_i' - g_j' + \varepsilon_i - \varepsilon_j$. The difference within a pair may not only result from the ground truth of the gene expression $g_i' - g_j'$, but was also affected by the technical variation $\varepsilon_i - \varepsilon_j$ when counting RNAs from reads. Suppose $g_i' > g_j'$, the relative expression will be opposite, i.e., $g_i < g_j$, if $-(\varepsilon_i - \varepsilon_j) > g_i' - g_j'$.

However, as ε is the noise of measurement, it is usually a small value compared to g'. Therefore, $g_i' - g_j' + \varepsilon_i - \varepsilon_j$ is in a small range because $\varepsilon_i - \varepsilon_j$ and $g_i' - g_j'$ is small, as $-(\varepsilon_i - \varepsilon_j) > g_i' - g_j'$. In such cases, to remove these opposite relative expressions caused by noise, we introduced effective size Δ, where only the difference $g_i' - g_j' + \varepsilon_i - \varepsilon_j$ exceeding Δ was regarded as effective. As a result, we applied the function $f(g_i, g_j) = \begin{cases} 1, g_i - g_j > \Delta \\ -1, g_i - g_j < -\Delta \\ 0, \ otherwise \end{cases}$ to each gene pair (Fig. 3B).

Following the intrasample analysis, we performed a population analysis to obtain the significance level of the gene pairs. For each gene pair, a contingency table was calculated without considering the samples within the difference threshold, given that those samples are largely affected by noise (Fig. 3C). We then conducted the Fisher's exact test, and 48,388 significantly different pairs (False discovery rate corrected $P < 0.01$) were preserved as non-coding RNA's competing endogenous gene pairs (nceGPs) (Fig. 3D).

3.7.2. SOC index generation

The least absolute shrinkage and selection operator (LASSO) [45–47] is a regression model with regularization in machine learning, which has been widely used in medical applications [48–50]. It can not only classify different classes, but also prune features to avoid overfitting and improve the generalizability. The linear regression was optimized using the following loss function with an L1 penalty: $\mathscr{L}(W;\alpha) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i * W) + \alpha\sum_{j=1}^{3}|w_j|$, where n is the number of samples, Y is the label for each sample, X is the vector composed by all the nceGPs and the constant term, W is the vector of weights for nceGPs, and α is the coefficient for the L1 penalty.

After 48,388 differential gene pairs were obtained from the previous step, resampling was applied to the training dataset due to the unbalanced size of the SOC samples and normal controls, which is a commonly used trick in machine learning to improve model accuracy. LASSO regression model was trained on the resampled training set and 18 circular RNA's competing endogenous gene pairs (cceGPs) were selected for the final SOC index (Fig. 3E). The code of deiPAGE algorithm is available at https://github.com/Kimxbzheng/deiPAGE.

3.8. Principle component analysis

Principal component analysis (PCA) is a machine learning method to summarize high-dimensional data into a few main components for data analysis and visualization. In this study, we performed PCA using sklearn in python to convert the 18 cceGPs into three principal components for visualization of its classification performance (Fig. 4A).

3.9. Correlation analysis with tumour infiltration

To explore the association with immune cell infiltration, Spearman's rank correlation coefficient was adopted to estimate the correlation between the SOC index and immune cells in cancer ($|R| > 0.3$ and $P < 0.01$). For each sample, we calculated the levels of immune cell infiltration using the Tumour Immune Estimation Resource (TIMER) [51].

3.10. Gene set enrichment analysis and functional enrichment analysis

We downloaded all pathways of Collection 2 (C2) and their gene sets from MSigDB (v7.4) and applied Gene Set Enrichment Analysis (GSEA) [52] with the R package clusterProfiler [53]. The differentially expressed mRNAs were subjected to GSEA [25] for functional enrichment analysis. Hypergeometric test was applied to evaluate the statistical significance of functional enrichment for overrepresented gene sets using the R package clusterProfiler [53]. The 18 cceGPs in SOC index were enriched in Gene Oncology (GO) [37] and Reactome [36].

3.11. Statistical analyses

R Project (R x64, version 3.5.2) and Python (version 3.6) were used for statistical computation in this study. Student's t-test and the Benjamini and Hochberg (BH) multiple testing correction method were used to calculate the differential expression RNAs in the self-profiled cohort. Pearson's correlation test were applied to calculate the expression correlation among RNAs. Fisher's exact test and false discovery rate (FDR) correction were employed to compute the significance of gene pair reversal. The LASSO regression was used to screen gene pairs and construct classification model for SOC. PCA was used to visualize the 18 gene pairs. Kaplan-Meier method and log-rank test were used for clinical outcome comparison. Spearman's rank correlation coefficient was adopted to estimate the correlation between the SOC index and immune cells. Hypergeometric test was applied to evaluate the statistical significance of functional enrichment for overrepresented gene sets. Statistical significance was set at $P < 0.01$.

4. Discussion

In this study, we hypothesized the triangular relation of mRNA-ncRNA-mRNA that two mRNAs and the ncRNA were competing endogenous RNAs and the two mRNAs were reversed in expression between SOCs and controls. Such mRNA-ncRNA-mRNA motifs were defined as ncRNA's competing endogenous gene pairs (nceGPs). We constructed a competing endogenous RNA (ceRNA) regulatory network on the basis of the immune-related and differentially expressed RNAs from the self-profiled cohort. We developed deiPAGE algorithm to screen nceGPs and constructed an SOC index for SOC detection, in which 18 nceGPs were included and defined as circRNA's ceGPs (cceGPs). Validation in two independent cohorts revealed that the SOC index was strongly reproducible and accurate (average AUC close to 0.99).

A preliminary study showed that the six cohorts shared few differentially expressed mRNAs (Additional File 1: Figs. S2 and S3). For the top 100 DEGs, specifically, only two genes were commonly identified across the six cohorts, indicating that the absolute gene abundance is incapable to consistently identify biomarkers across different platforms. To address this problem, we developed deiPAGE to identify the stable signals from various platforms taking advantage of the relative gene abundance. The deiPAGE we proposed in this study is a generalized algorithm for data integration and biomarker identification. Compared to iPAGE, it not only considered the relative ranking but also included

the effective size between two genes. The effective size Δ can reduce the false-positive rate that results from neglecting the variance of two genes in the relative expression.

The 18 cceGPs consist of 22 genes regulated by 14 circRNAs (Table 3). We investigated the 14 circRNAs and found that circKRT7 promotes ovarian cancer cell progression by the circKRT7-miR-29a-3p-*COL1A1* axis [54] and circSLC22A3 suppresses ovarian cancer progression by the CircSLC22A3-miR-518a-5p/Fas axis [55]. CircSFMBT2 inhibits the proliferation and metastasis of glioma cells through miR-182–5p/Mtss1 Pathway [56,57]. Furthermore, we seek for the experimental studies of the 22 genes and found that 13 genes involve in the regulation in ovarian cancer and three genes participate in the regulation of other cancers. *TOP2A* played a central role in the network as it was involved in four cceGPs (Fig. 6A). *TOP2A* (DNA Topoisomerase II Alpha) is a cancer-related gene that encodes DNA topoisomerase enzyme, which controls the structures of DNA during transcription. *TOP2A* promotes tumourigenesis of high-grade SOC by regulating the TGF-β/Smad pathway [58] and serves as the target for several anticancer agents. The *TOP2A* expression is also a marker for the response to pegylated lyposomal doxorubicin (PLD) in epithelial ovarian cancer therapy [59,60]. In this study, *TOP2A* composed reserved pairs with *LAM4*, *SNCA*, *MYO1E*, *CRADD*, and *MFAP4* in SOC, *LAM4*, *SNCA,* and *MFAP4* among which are associated with ovarian cancer [61–63]. Particularly, the cceGP *TOP2A*:*MFAP4* indicates the clinical outcomes of patients (Fig. 5D). All the genes and circRNAs in cceGPs *TOP2A*--circKRT7-*SNCA* and *TOP2A*-circKRT7- *MFAP4* have been experimental validated to be associated to the regulation of ovarian cancer. Moreover, as *TOP2A*, *SNCA*, and *LAM4* were related to the regulation of ovarian cancer, we speculated that their competing endogenous ncRNAs circCOL1A3 and circSEL1L3 participated in the regulatory mechanism in ovarian cancer.

Function analysis indicated that *BBS4* and *PRC1* in the cceGP *BBS4*-circHUNK-*PRC1* were both enriched in the regulation of cell division (Fig. 6D). *PRC1* encodes a protein that involves in mitosis and cytokinesis and the overexpression of *PRC1* can infer a poor prognosis in ovarian cancer. Wang et al. found that lncHCP5/miR-525–5p/*PRC1* crosstalk might promote malignant behaviors of ovarian cancer cells and the silencing of lncRNA HCP5 impeded growth and metastasis of tumour in mice [38]. *BBS4* was differentially expressed and downregulated in breast cancer and it was associated with poor prognosis [40]. Collectively, these findings inferred that the circHUNK in this cceGP might be associated with the regulation of gynecologic cancer and might be a potential target for anticancer drugs.

Taking *BBS4*–circHUNK–*PRC1* as an example to dissect the competing endogenous mechanism of the regulatory motif (Fig. 6I). After the reversed expression of *PRC1* and *BBS4*, the mRNA of *PRC1* turn into a high level and *BBS4* into a low level in SOC. In that case, the mRNAs of *PRC1* and *BBS4* became competing endogenous with circHUNK, as circHUNK sponged common targeted miRNAs with *PRC1* and *BBS4*, and they are positively correlated in SOC. The competing endogenous regulation of the circRNA might be induced by its competing endogenous gene expression reversal.

The proposed SOC index reflects tumour infiltration and is inversely correlated with the ratio of CD8+/CD4+ T-cells, which is lower in the SOC patients and higher among the normal controls. A higher SOC index score implies lower $CD8^+$ fraction and higher $CD4^+$ fraction (Fig. 5E–I). It suggests that although $CD4^+$ T-cells were active in SOC to help recruit and activate $CD8^+$ T-cells [64], the tumour cells somehow found a way to shut down or deactivate $CD8^+$ cytotoxic T-cells, which can eliminate tumour cells. As the ratio of CD8+/CD4+ T cells in ovarian cancer is used as a prognostic factor [35], the SOC index may be a potential prognostic indicator in cancer patients.

The deiPAGE we proposed in this study is a generalized algorithm for data integration and biomarker identification. Technically, data integration is necessary for a large-scale study to obtain accurate biomarkers [65,66]. The primary challenge lies on integrating various cohorts, due to the technical variation between platforms and the batch effect from different experiments [67]. To address this issue, the relative expression of gene pairs in each sample rather than the absolute expression value of a single gene was taken into account. Although this might lose some quantitative information from the expression data, datasets from different resources were integrated for model training, thereby substantially increasing the sample size and improving the statistical power of detecting reversal gene pairs. More importantly, the reversal gene pairs can be easily applied to independent individuals, since they do not require any extra preprocessing of population samples. Notably, deiPAGE not only considered the relative ranking but also included the effective size between two genes. Neglecting the size effect difference of two genes for relative expression ordering results in a number of false positives among significant results, given that a large expression difference of two genes contributes equally to a small one. To address this, we used a parameter, difference threshold Δ, to reduce the rate of false-positive discoveries.

Normal ovarian samples are crucial for expression profiling studies relying on comparisons with malignant ovarian tissues. However, tissues from normal donors are rare for ovarian cancer because of the invasive procedure. In total, only 33 normal ovarian samples for gene expression were publicly available among TCGA and GEO before this study. We profiled the expression of eight normal ovarian samples from cervical cancer patients not metastasised to the ovaries, offering important support and complements to ovarian research. In addition to the mRNA and lncRNA expressions, we measured genome-wide circRNAs using microarray, thereby providing an opportunity to investigate the regulatory role of non-coding RNAs in ovarian cancer. This research may also facilitate studies related to the mechanism and therapeutics involving circRNAs in ovarian cancer.

Data availability

The self-profiled data are available at https://www.jianguoyun.com/p/Dbtjn28Qm_jCRjN_ogE.

Funding

This work was supported by Guangdong Basic and Applied Basic Research Foundation (Grant no. 2022A1515012368).

Declaration of competing interest

The authors declare no competing interests.

Acknowledgements

The results are in part based upon data generated by the TCGA Research Network (https://www.cancer.gov/tcga). We are also grateful to Vanessa L Fuller, from Language Services at the University of Helsinki, Finland, for assistance in the English-language revision of this work.

ABBREVIATIONS

SOC	serous ovarian carcinomas
ceRNA	competing endogenous RNA
deiPAGE	denoised individualized pair analysis of gene expression
ceGPs	competing endogenous gene pairs
nceGPs	non-coding RNA's competing endogenous gene pairs
cceGPs	cricRNA's ceGPs
CA125	Cancer antige 125
HE4	human epididymis protein 4
FDA	Food and Drug Administration
PTAR	pro-transition associated R
SNAI2	nail family zinc finger 2

circRNA	circular RNA
ncRNA	long non-coding RNA
GSEA	Gene Set Enrichment Analysis
PCC	Pearson correlation coefficient
DEP	differentially expressed gene pairs
LASSO	least absolute shrinkage and selection operator
TIMER	Tumour Immune Estimation Resource
AFP	bnormal alpha fetoprotein
DEG	differentially expressed genes
KEGG	Kyoto Encyclopedia of Genes and Genomes
miRNA	microRNA
TCGA	The Cancer Genome Atlas
GEO	the Gene Expression Omnibus
DEP	differentially expressed gene pair
AUROC	the area under the receiver operating characteristic
IDH	isocitrate dehydrogenase

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2022.105881.

References

[1] W. Guo, L. Zhu, M. Yu, R. Zhu, Q. Chen, Q. Wang, A five-DNA methylation signature act as a novel prognostic biomarker in patients with ovarian serous cystadenocarcinoma, Clin. Epigenet. 10 (1) (2018) 142.

[2] E.J. Nam, H. Yoon, S.W. Kim, H. Kim, Y.T. Kim, J.H. Kim, J.W. Kim, S. Kim, MicroRNA expression profiles in serous ovarian carcinoma, Clin. Cancer Res. 14 (9) (2008) 2690–2695.

[3] D. Hao, J. Li, J. Wang, Y. Meng, Z. Zhao, C. Zhang, K. Miao, C. Deng, B.K. Tsang, L. Wang, et al., Non-classical estrogen signaling in ovarian cancer improves chemosensitivity and patients outcome, Theranostics 9 (13) (2019) 3952–3965.

[4] X. Chen, D. Xie, Q. Zhao, Z.H. You, MicroRNAs and complex diseases: from experimental results to computational models, Briefings Bioinf. 20 (2) (2019) 515–539.

[5] C.C. Wang, C.D. Han, Q. Zhao, X. Chen, Circular RNAs and complex diseases: from experimental results to computational models, Briefings Bioinf. 22 (6) (2021).

[6] X. Chen, C.C. Zhu, J. Yin, Ensemble of decision tree reveals potential miRNA-disease associations, PLoS Comput. Biol. 15 (7) (2019), e1007209.

[7] W. Liu, H. Lin, L. Huang, L. Peng, T. Tang, Q. Zhao, L. Yang, Identification of miRNA-disease associations via deep forest ensemble learning based on autoencoder, Briefings Bioinf. 23 (3) (2022).

[8] X. Chen, J. Yin, J. Qu, L. Huang, MDHGI: matrix decomposition and heterogeneous graph inference for miRNA-disease association prediction, PLoS Comput. Biol. 14 (8) (2018), e1006418.

[9] X. Chen, L.G. Sun, Y. Zhao, NCMCMDA: miRNA-disease association prediction through neighborhood constraint matrix completion, Briefings Bioinf. 22 (1) (2021) 485–496.

[10] H. Liu, G. Ren, H. Chen, Q. Liu, Y. Yang, Q. Zhao, Predicting lncRNA–miRNA interactions based on logistic matrix factorization with neighborhood regularized, Knowl. Base Syst. 191 (2020), 105261.

[11] X. Chen, L. Wang, J. Qu, N.N. Guan, J.Q. Li, Predicting miRNA-disease association based on inductive matrix completion, Bioinformatics 34 (24) (2018) 4256–4265.

[12] W. Liu, Y. Jiang, L. Peng, X. Sun, W. Gan, Q. Zhao, H. Tang, Inferring gene regulatory networks using the improved markov blanket discovery algorithm, Interdiscipl. Sci. Comput. Life Sci. (2021) 1–14.

[13] L. Cheng, C. Nan, L. Kang, N. Zhang, S. Liu, H. Chen, C. Hong, Y. Chen, Z. Liang, X. Liu, Whole blood transcriptomic investigation identifies long non-coding RNAs as regulators in sepsis, J. Transl. Med. 18 (1) (2020) 217.

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

[15] H.S. Chiu, M.R. Martinez, M. Bansal, A. Subramanian, T.R. Golub, X. Yang, P. Sumazin, A. Califano, High-throughput validation of ceRNA regulatory networks, BMC Genom. 18 (1) (2017) 418.

[16] J. Wang, Y. Cao, X. Lu, X. Wang, X. Kong, C. Bo, S. Li, M. Bai, Y. Jiao, H. Gao, et al., Identification of the regulatory role of lncRNA SNHG16 in myasthenia gravis by constructing a competing endogenous RNA network, Mol. Ther. Nucleic Acids 19 (2020) 1123–1133.

[17] H. Liang, T. Yu, Y. Han, H. Jiang, C. Wang, T. You, X. Zhao, H. Shan, R. Yang, L. Yang, et al., LncRNA PTAR promotes EMT and invasion-metastasis in serous ovarian cancer by competitively binding miR-101-3p to regulate ZEB1 expression, Mol. Cancer 17 (1) (2018) 119.

[18] L. Zhang, P. Yang, H. Feng, Q. Zhao, H. Liu, Using network distance analysis to predict lncRNA–miRNA interactions, Interdiscipl. Sci. Comput. Life Sci. 13 (3) (2021) 535–545.

[19] L. Cheng, K.S. Leung, Quantification of non-coding RNA target localization diversity and its application in cancers, J. Mol. Cell Biol. 10 (2) (2018) 130–138.

[20] X. Liu, Y. Xu, R. Wang, S. Liu, J. Wang, Y. Luo, K.-S. Leung, L. Cheng, A network-based algorithm for the identification of moonlighting noncoding RNAs and its application in sepsis, Briefings Bioinf. 22 (1) (2020) 581–588.

[21] L. Cheng, K.S. Leung, Identification and characterization of moonlighting long non-coding RNAs based on RNA and protein interactome, Bioinformatics 34 (20) (2018) 3519–3528.

[22] X. Wang, L. Han, L. Zhou, L. Wang, L.M. Zhang, Prediction of candidate RNA signatures for recurrent ovarian cancer prognosis by the construction of an integrated competing endogenous RNA network, Oncol. Rep. 40 (5) (2018) 2659–2673.

[23] W. Li, S. Ma, X. Bai, W. Pan, L. Ai, W. Tan, Long noncoding RNA WDFY3-AS2 suppresses tumor progression by acting as a competing endogenous RNA of microRNA-18a in ovarian cancer, J. Cell. Physiol. 235 (2) (2020) 1141–1154.

[24] Y. Geng, J. Jiang, C. Wu, Function and clinical significance of circRNAs in solid tumors, J. Hematol. Oncol. 11 (1) (2018) 98.

[25] A. Subramanian, H. Kuehn, J. Gould, P. Tamayo, J.P. Mesirov, GSEA-P: a desktop application for gene set enrichment analysis, Bioinformatics 23 (23) (2007) 3251–3253.

[26] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (1) (2000) 27–30.

[27] S. Bhattacharya, S. Andorf, L. Gomes, P. Dunn, H. Schaefer, J. Pontius, P. Berger, V. Desborough, T. Smith, J. Campbell, et al., ImmPort: disseminating data to the public for the future of immunology, Immunol. Res. 58 (2–3) (2014) 234–239.

[28] Y. Li, T. Jiang, W. Zhou, J. Li, X. Li, Q. Wang, X. Jin, J. Yin, L. Chen, Y. Zhang, et al., Pan-cancer characterization of immune-related lncRNAs identifies potential oncogenic biomarkers, Nat. Commun. 11 (1) (2020) 1000.

[29] C. Yang, D. Wu, L. Gao, X. Liu, Y. Jin, D. Wang, T. Wang, X. Li, Competing endogenous RNA networks in human cancer: hypothesis, validation, and perspectives, Oncotarget 7 (12) (2016) 13479–13490.

[30] Y. Lin, T. Liu, T. Cui, Z. Wang, Y. Zhang, P. Tan, Y. Huang, J. Yu, D. Wang, RNAInter in 2020: RNA interactome repository with increased coverage and annotation, Nucleic Acids Res. 48 (D1) (2020) D189–D197.

[31] N. Cancer Genome Atlas Research, J.N. Weinstein, E.A. Collisson, G.B. Mills, K. R. Shaw, B.A. Ozenberger, K. Ellrott, I. Shmulevich, C. Sander, J.M. Stuart, The cancer genome Atlas pan-cancer analysis project, Nat. Genet. 45 (10) (2013) 1113–1120.

[32] T. Barrett, S.E. Wilhite, P. Ledoux, C. Evangelista, I.F. Kim, M. Tomashevsky, K. A. Marshall, K.H. Phillippy, P.M. Sherman, M. Holko, et al., NCBI GEO: archive for functional genomics data sets–update, Nucleic Acids Res. 41 (Database issue) (2013) D991–D995.

[33] T. Van Gorp, I. Cadron, E. Despierre, A. Daemen, K. Leunen, F. Amant, D. Timmerman, B. De Moor, I. Vergote, HE4 and CA125 as a diagnostic test in ovarian cancer: prospective validation of the Risk of Ovarian Malignancy Algorithm, Br. J. Cancer 104 (5) (2011) 863–870.

[34] T. Li, J. Fu, Z. Zeng, D. Cohen, J. Li, Q. Chen, B. Li, X.S. Liu, TIMER2.0 for analysis of tumor-infiltrating immune cells, Nucleic Acids Res. 48 (W1) (2020) W509–W514.

[35] C.C. Preston, M.J. Maurer, A.L. Oberg, D.W. Visscher, K.R. Kalli, L.C. Hartmann, E. L. Goode, K.L. Knutson, The ratios of CD8+ T cells to CD4+CD25+ FOXP3+ and FOXP3- T cells correlate with poor clinical outcome in human serous ovarian cancer, PLoS One 8 (11) (2013), e80063.

[36] B. Jassal, L. Matthews, G. Viteri, C. Gong, P. Lorente, A. Fabregat, K. Sidiropoulos, J. Cook, M. Gillespie, R. Haw, et al., The reactome pathway knowledgebase, Nucleic Acids Res. 48 (D1) (2020) D498–D503.

[37] C. Gene Ontology, The Gene Ontology resource: enriching a GOld mine, Nucleic Acids Res. 49 (D1) (2021) D325–D334.

[38] L. Wang, M. He, L. Fu, Y. Jin, Role of lncRNAHCP5/microRNA-525-5p/PRC1 crosstalk in the malignant behaviors of ovarian cancer cells, Exp. Cell Res. 394 (1) (2020), 112129.

[39] H. Bu, Y. Li, C. Jin, H. Yu, X. Wang, J. Chen, Y. Wang, Y. Ma, Y. Zhang, B. Kong, Overexpression of PRC1 indicates a poor prognosis in ovarian cancer, Int. J. Oncol. 56 (3) (2020) 685–696.

[40] H. Zhang, B. Suo, X.P. Sun, Y.-X. Sun, H. Qiao, Bardet-biedl Syndrome 4 in early diagnosis and prognosis of breast cancer, Indian J. Pharmaceut. Sci. (2021) 145–152.

[41] L. Cheng, L.Y. Lo, N.L. Tang, D. Wang, K.S. Leung, CrossNorm: a novel normalization strategy for microarray data in cancers, Sci. Rep. 6 (2016), 18898.

[42] X. Liu, N. Li, S. Liu, J. Wang, N. Zhang, X. Zheng, K.S. Leung, L. Cheng, Normalization methods for the analysis of unbalanced transcriptome data: a review, Front. Bioeng. Biotechnol. 7 (2019) 358.

[43] C.C. Nan, N. Zhang, K.C.P. Cheung, H.D. Zhang, W. Li, C.Y. Hong, H.S. Chen, X. Y. Liu, N. Li, L. Cheng, Knockdown of lncRNA MALAT1 alleviates LPS-induced acute lung injury via inhibiting apoptosis through the miR-194-5p/FOXP2 Axis, Front. Cell Dev. Biol. 8 (2020), 586869.

[44] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (11) (2003) 2498–2504.

[45] J. Wang, X. Xiang, L. Bolund, X. Zhang, L. Cheng, Y. Luo, GNL-Scorer: a generalized model for predicting CRISPR on-target activity by machine learning and featurization, J. Mol. Cell Biol. 12 (11) (2020) 909–911.

[46] J. Wang, X. Zhang, L. Cheng, Y. Luo, An overview and metanalysis of machine and deep learning-based CRISPR gRNA design tools, RNA Biol. 17 (1) (2020) 13–22.

[47] R. Tibshirani, Regression shrinkage and selection via the lasso, J. Roy. Stat. Soc. B 58 (1) (1996) 267–288.

[48] P. Liu, X. Zheng, M.-H. Wong, K.-S. Leung, Drug2vec: a drug embedding method with drug-drug interaction as the context, in: International Conference on Engineering Applications of Neural Networks: 2020, Springer, 2020, pp. 326–337.
[49] X. Liu, X. Zheng, J. Wang, N. Zhang, K.-S. Leung, X. Ye, L. Cheng, A long non-coding RNA signature for diagnostic prediction of sepsis upon ICU admission, Clin. Transl. Med. 10 (3) (2020) e123.
[50] Y. Yang, Y. Zhang, S. Li, X. Zheng, M.H. Wong, K.S. Leung, L. Cheng, A robust and generalizable immune-related signature for sepsis diagnostics, IEEE ACM Trans. Comput. Biol. Bioinf (2021), https://doi.org/10.1109/TCBB.2021.3107874.
[51] T. Li, J. Fan, B. Wang, N. Traugh, Q. Chen, J.S. Liu, B. Li, X.S. Liu, TIMER, A web server for comprehensive analysis of tumor-infiltrating immune cells, Cancer Res. 77 (21) (2017) e108–e110.
[52] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U. S. A. 102 (43) (2005) 15545–15550.
[53] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS 16 (5) (2012) 284–287.
[54] Q. An, T. Liu, M.Y. Wang, Y.J. Yang, Z.D. Zhang, Z.J. Lin, B. Yang, circKRT7-miR-29a-3p-COL1A1 Axis promotes ovarian cancer cell progression, OncoTargets Ther. 13 (2020) 8963–8976.
[55] N. Zhang, Y. Jin, Q. Hu, S. Cheng, C. Wang, Z. Yang, Y. Wang, Circular RNA hsa_circ_0078607 suppresses ovarian cancer progression by regulating miR-518a-5p/Fas signaling pathway, J. Ovarian Res. 13 (1) (2020) 64.
[56] H. Sun, P. Xi, Z. Sun, Q. Wang, B. Zhu, J. Zhou, H. Jin, W. Zheng, W. Tang, H. Cao, Circ-SFMBT2 promotes the proliferation of gastric cancer cells through sponging miR-182-5p to enhance CREB1 expression, Cancer Manag. Res. 10 (2018) 5725.
[57] S. Zhang, W. Qin, S. Yang, N. Guan, X. Sui, W. Guo, Circular RNA SFMBT2 inhibits the proliferation and metastasis of glioma cells through mir-182-5p/Mtss1 pathway, Technol. Cancer Res. Treat. 19 (2020), 1533033820945799.
[58] Y. Gao, H. Zhao, M. Ren, Q. Chen, J. Li, Z. Li, C. Yin, W. Yue, TOP2A promotes tumorigenesis of high-grade serous ovarian cancer by regulating the TGF-beta/smad pathway, J. Cancer 11 (14) (2020) 4181–4192.
[59] E. Ghisoni, F. Maggiorotto, F. Borella, G. Mittica, S. Genta, G. Giannone, D. Katsaros, A. Sciarrillo, A. Ferrero, I. Sarotto, TOP2A as marker of response to pegylated lyposomal doxorubicin (PLD) in epithelial ovarian cancers, J. Ovarian Res. 12 (1) (2019) 1–6.
[60] J. Erriquez, P. Becco, M. Olivero, R. Ponzone, F. Maggiorotto, A. Ferrero, M. S. Scalzo, E.M. Canuto, A. Sapino, L. Verdun di Cantogno, et al., TOP2A gene copy gain predicts response of epithelial ovarian cancers to pegylated liposomal doxorubicin: TOP2A as marker of response to PLD in ovarian cancer, Gynecol. Oncol. 138 (3) (2015) 627–633.
[61] Y. Liu, Y. Xu, L. Ding, L. Yu, B. Zhang, D. Wei, LncRNA MEG3 suppressed the progression of ovarian cancer via sponging miR-30e-3p and regulating LAMA4 expression, Cancer Cell Int. 20 (2020) 181.
[62] W. Bruening, B.I. Giasson, A.J. Klein-Szanto, V.M. Lee, J.Q. Trojanowski, A. K. Godwin, Synucleins are expressed in the majority of breast and ovarian carcinomas and in preneoplastic lesions of the ovary, Cancer 88 (9) (2000) 2154–2163.
[63] H. Zhao, Q. Sun, L. Li, J. Zhou, C. Zhang, T. Hu, X. Zhou, L. Zhang, B. Wang, B. Li, et al., High expression levels of AGGF1 and MFAP4 predict primary platinum-based chemoresistance and are associated with adverse prognosis in patients with serous ovarian cancer, J. Cancer 10 (2) (2019) 397–407.
[64] W. Wang, W. Zou, J.R. Liu, Tumor-infiltrating T cells in epithelial ovarian cancer: predictors of prognosis and biological basis of immunotherapy, Gynecol. Oncol. 151 (1) (2018) 1–3.
[65] S. Liu, W. Zhao, X. Liu, L. Cheng, Metagenomic analysis of the gut microbiome in atherosclerosis patients identify cross-cohort microbial signatures and potential therapeutic target, Faseb. J. 34 (11) (2020) 14166–14181.
[66] L. Cheng, Y. Zeng, S. Hu, N. Zhang, K.C.P. Cheung, B. Li, K.S. Leung, L. Jiang, Systematic prediction of autophagy-related proteins using Arabidopsis thaliana interactome data, Plant J. 105 (3) (2021) 708–720.
[67] X. Zheng, Q. Wu, H. Wu, K.-S. Leung, M.-H. Wong, X. Liu, L. Cheng, Evaluating the consistency of gene methylation in liver cancer using bisulfite sequencing data, Front. Cell Dev. Biol. 9 (2021) 1022.
[68] K.M. Fung, L.B. Rorke, B. Giasson, V.M. Lee, J.Q. Trojanowski, Expression of alpha-, beta-, and gamma-synuclein in glial tumors and medulloblastomas, Acta Neuropathol. 106 (2) (2003) 167–175.
[69] S. Wang, C. Wang, X. Li, Y. Hu, R. Gou, Q. Guo, X. Nie, J. Liu, L. Zhu, B. Lin, Down-regulation of TRIB3 inhibits the progression of ovarian cancer via MEK/ERK signaling pathway, Cancer Cell Int. 20 (2020) 418.
[70] H. Kang, C. Kim, E. Ji, S. Ahn, M. Jung, Y. Hong, W. Kim, E.K. Lee, The MicroRNA-551a/MEF2C Axis regulates the survival and sphere formation of cancer cells in response to 5-fluorouracil, Mol. Cell. 42 (2) (2019) 175–182.
[71] S. Ma, X. Rong, F. Gao, Y. Yang, L. Wei, TPX2 promotes cell proliferation and migration via PLK1 in OC, Cancer Biomarkers 22 (3) (2018) 443–451.
[72] M. Zhang, T. Liu, B. Xia, C. Yang, S. Hou, W. Xie, G. Lou, Platelet-derived growth factor D is a prognostic biomarker and is associated with platinum resistance in epithelial ovarian cancer, Int. J. Gynecol. Cancer 28 (2) (2018).
[73] C. Yang, M. Zhang, Y. Cai, Z. Rong, C. Wang, Z. Xu, H. Xu, W. Song, Y. Hou, G. Lou, Platelet-derived growth factor-D expression mediates the effect of differentiated degree on prognosis in epithelial ovarian cancer, J. Cell. Biochem. 120 (5) (2019) 6920–6925.
[74] Y. Zhang, K. Huang, H. Cai, S. Chen, D. Sun, P. Jiang, The role of nucleolar spindle-associated protein 1 in human ovarian cancer, J. Cell. Biochem. 121 (11) (2020) 4397–4405.
[75] C.S. Bryant, S. Kumar, S. Chamala, J. Shah, J. Pal, M. Haider, S. Seward, A.M. Qazi, R. Morris, A. Semaan, et al., Sulforaphane induces cell cycle arrest by protecting RB-E2F-1 complex in epithelial ovarian cancer cells, Mol. Cancer 9 (2010) 47.
[76] H. Yang, P. Das, Y. Yu, W. Mao, Y. Wang, K. Baggerly, Y. Wang, R.T. Marquez, A. Bedi, J. Liu, NDN is an imprinted tumor suppressor gene that is downregulated in ovarian cancers through genetic and epigenetic mechanisms, Oncotarget 7 (3) (2016) 3018.
[77] W. Zhang, C.J. Barger, K.H. Eng, D. Klinkebiel, P.A. Link, A. Omilian, W. Bshara, K. Odunsi, A.R. Karpf, PRAME expression and promoter hypomethylation in epithelial ovarian cancer, Oncotarget 7 (29) (2016) 45352–45369.
[78] S. Sarkar, C.A. Bristow, P. Dey, K. Rai, R. Perets, A. Ramirez-Cardenas, S. Malasi, E. Huang-Hobbs, M. Haemmerle, S.Y. Wu, et al., PRKCI promotes immune suppression in ovarian cancer, Genes Dev. 31 (11) (2017) 1109–1121.
[79] H. Rehmani, Y. Li, T. Li, R. Padia, O. Calbay, L. Jin, H. Chen, S. Huang, Addiction to protein kinase Cι due to PRKCI gene amplification can be exploited for an aptamer-based targeted therapy in ovarian cancer, Signal Transduct. Targeted Ther. 5 (1) (2020) 1–11.
[80] A. Belur Nagaraj, M. Knarr, S. Sekhar, R.S. Connor, P. Joseph, O. Kovalenko, A. Fleming, A. Surti, E. Nurmemmedov, L. Beltrame, et al., The miR-181a-SFRP4 Axis regulates Wnt activation to drive stemness and platinum resistance in ovarian cancer, Cancer Res. 81 (8) (2021) 2044–2055.
[81] C.E. Ford, E. Jary, S.S. Ma, S. Nixdorf, V.A. Heinzelmann-Schwarz, R.L. Ward, The Wnt gatekeeper SFRP4 modulates EMT, cell migration and downstream Wnt signalling in serous ovarian cancer cells, PLoS One 8 (1) (2013), e54362.