European Review for Medical and Pharmacological Sciences 2019; 23: 8257-8264

A comprehensive understanding of ovarian carcinoma survival prognosis by novel biomarkers

Y. WANG, L. LEI, Y.-G. CHI, L.-B. LIU, B.-P. YANG

Department of Gynecology, Chongqing Maternal and Child Health Hospital, Chongqing, P.R. China

Abstract. – OBJECTIVE: Ovarian cancer is one of the most common causes of cancer-related deaths in women. Many studies show that dysregulated gene expression plays a key role in tumorigenesis and development. Therefore, a comprehensive understanding of ovarian serous cystadenocarcinoma survival prognosis is needed.

PATIENTS AND METHODS: A large number of high-dimensional RNA-sequencing files and clinical datasets collected from the Genomic Data Commons Data Portal were utilized to identify novel potential biomarkers for determining the prognosis of patients with ovarian serous cystadenocarcinoma (OVSC). We adopted a new strategy to identify these biomarkers by integrating co-expression network analysis and the Kaplan-Meier estimation with a non-parametric bootstrapping procedure.

RESULTS: Functional enrichment analysis of gene modules of interest revealed several dysregulated genes in OVSC, suggesting a close relationship between hormones and angiogenesis. In combination with this comprehensive approach, 14 genes, including ABCA10, DCX, LRRC30, ALX4, DKK4, SGCZ, ANKS4B, FHL5, SPRR2F, CHRNG, GABRR1, STMN2, CRHBP, and GSTM5, were shown to serve as candidate biomarkers for predicting the prognosis of patients with OVSC.

CONCLUSIONS: The current study identified several valuable prognostic biomarkers and several potential therapeutic targets for treating OVSC.

Key Words:
Ovarian serous cystadenocarcinoma, Survival prognosis, Co-expression network, Survival analysis.

Introduction

Ovarian serous cystadenocarcinoma (OVSC) is one of the four most fatal cancers in women, and the number of newly diagnosed cases exceeds 200,000 each year[1]. The 5-year survival rate of most patients with OVSC is less than 50%. The prognosis is poor, and patients with metastasis due to delayed detection and chemotherapy tolerance have even worse prognosis[2]. Therefore, identifying clinical biomarkers to determine the prognosis of OVSC is critically important, and may lead to new therapeutic targets that ultimately reduce the risk of death in patients with OVSC. It is generally believed that the dysregulation of the gene expression acts as an important role in tumorigenesis and development. More and more studies showed that the diagnostic/prognostic markers and even new therapeutic targets were revealed by analyzing gene expression patterns in various cancers types[3,4]. In recent years, with the rapid spread of high-throughput technology, high-dimensional data on gene expression has provided comprehensive information to better understand the survival and prognosis of OVSC patients; however, the findings of these studies still have great deficiencies in the survival prognosis of OVSC patients so far[5,6]. In most cases, many rare genes with no functional relationships were selected and considered as potential biomarkers for OVSC. Given the complexity of the interactions between genes, a single genetic abnormality can propagate along with a complex intracellular network, leading to a series of common changes in cell functional properties, either directly or indirectly[7,8]. Therefore, a dynamic, modular network-based analysis for survival-related biomarkers can provide accurate information and an underlying understanding of OVSC patients' outcomes and clinical treatment options. More importantly, in order to better understand the survival prognosis and best treatment of OVSC, researchers do not only need to obtain credible biomarkers but also trace the key role of their biological functions. In this study, we

Corresponding Author: Boping Yang, MD; e-mail: gxingzhe116@hotmail.com

combined the weighted genes co-expression network analysis and Kaplan-Meier estimation with a non-parametric bootstrapping procedure using survival information, protein-protein interactions (PPI) in STRING[9], and gene expression data from OVSC. We took an integrated approach to capture a complete understanding of the prognostic survival biomarkers in OVSC patients.

Patients and Methods

Gene Expression Profiles and Clinical Datasets

According to the dissemination and application policy requirements of the Genomic Data Commons Data Portal (https://gdc-portal.nci.nih.gov/), RNA sequencing data profiles and clinical information of 305 OVSC patients including 2,744 samples (Table I) were collected. Most of the samples (89.14%) were from white patients, and 85.9% of patients received chemotherapy. The median age of selected patients was 55.7 years, the median survival time was 973 days, and the median survival time was 1102 days. This study has been approved by the Ethics Committee of the Institute of Biomedicine Research of the Chongqing Maternal and Child Health Hospital (Chongqing, China).

Genes Co-Expression Network Construction

Generally, weighted genes co-expression network analysis (WGCNA) was an important strategy to find main gene modules that represent critical characteristics for gene expression

Table I. Patients and samples distribution characteristics.

Groups	Patient	Counts	Sample	Counts
Total		305		2744
Race				
NA	16	(5.25%)	70	(2.55%)
American Indian or Alaska Native	4	(1.31%)	-11	(0.40%
Asian	26	(8.52%)	61	(2.22%)
Black or African American	39	(12.79%)	153	(5.58%)
Native Hawaiian or Pacific Islander	1	(0.33%)	3	(0.11%)
White	219	(71.80%)	2446	(89.14%)
Therapy Type				
NA	3	(0.98%)	12	(0.44%)
Chemotherapy	262	(85.90%)	2586	(94.24%)
Hormone therapy	16	(5.25%)	56	(2.04%)
Immunotherapy	13	(4.26%)	18	(0.66%)
Targeted molecular therapy	11	(3.61%)	72	(2.62%)
Clinical Stage				
NA	4	(1.31%)	17	(0.62%)
Stage IA	1	(0.33%)	3	(0.11%)
Stage IB	2	(0.66%)	6	(0.22%)
Stage IC	11	(3.61%)	30	(1.09%)
Stage IIA	5	(1.64%)	15	(0.55%)
Stage IIB	8	(2.62%)	23	(0.84%)
Stage IIC	15	(4.92%)	76	(2.77%)
Stage IIIA	17	(5.57%)	42	(1.53%)
Stage IIIB	21	(6.89%)	128	(4.66%)
Stage IIIC	182	(59.67%)	1987	(72.41%)
Stage IV	39	(12.79%)	417	(15.20%)
Vital				
Alive	143	(46.89%)	1085	(39.54%)
Dead	162	(53.11%)	1659	(60.46%)
Residual Nodule				
NA	25	(8.20%)	260	(9.48%)
No macroscopic nodule	50	(16.39%)	524	(19.10%)
1-10 mm	135	(44.26%)	1287	(46.90%)
11-20 mm	35	(11.48%)	167	(6.09%)
> 20 mm	60	(19.67%)	506	(18.44%)

NA: not available.

profiles. In the WGCNA network, each node was weighted according to the correlation matrices[10]. By using a dynamic cutting tree algorithm, gene modules were detected using power = 10 and deep split = 2 parameters. In brief, the Pearson correlation coefficients comparing all pairs of genes in all samples were converted into an adjacency matrix, resulting in a weighted network. A dynamic tree-cutting algorithm is then used to calculate the Topological Overlap Measurement (TOM), a biologically meaningful nodal similarity measure that hierarchically clusters genes, as distance measures using a height cutoff 0.95.

Functional Annotation of Gene Modules

The functional enrichment of genes in each gene module was performed using the ClusterProfiler[11] software package based on the gene ontology database. In functional annotation results, the overexpression of terms is defined as Fisher's exact *p*-value of the correction using the Benjamini method to adjust multiple tests. Due to different biological process shared common genes, as well as a gene shared more than one biological process, identifying that the hub genes are usually used as the critical step to verify potential biomarkers for diagnosis/prognosis of many diseases. In this study, the hub genes were used to assess the hazard ratios of OVSC patients based on their survival times.

Survival Analysis

To determine the prognostic capacity of potential biomarkers, the survival curves were estimated by the Kaplan-Meier method and were compared with a log-rank test of the hub genes. Here, we split our cohort using the expression of each gene whose aberrant expression (decreased/increased) was known to be related to poor prognosis in a mixed population of OVSC. To determine and ensure reliable prognostic association as a measure of significance, a non-parametric bootstrapping procedure was performed for each point-of-separation through distribution of expected hazard ratios (HRs), exactly as we would for a biomarker under investigation, against which we are able to compare our observed HRs as part of our analysis. Tukey's HSD (honestly significant difference) test is used in conjunction with an ANOVA to find means that are significantly different from each other. $p<0.05$ was recognized as a statistically significant difference.

Results

Overall Survival of OVSC Patients' Facet by Different Characteristics

At first, we performed a survival curve analysis among different races, clinical stages, therapy types, and residual nodules based on Kaplan-Meier estimates **(Supplementary Figure 1)**. The results showed that OVSC patients with no macroscopic nodules and clinical stage I or II had a higher probability of survival, while the other groups were not lucky. Although immunotherapeutic or targeted molecular therapy theoretically have better treatment effects, no significance was observed in the results of our analysis due to the abnormal distribution of races or therapy types.

Differentially Expressed Genes (DEGs)

Each probe of the OVSC microarray was separated into high and low groups according to its own median expression level. Differential expression of these two group samples was performed with the estimation of the *t*-test. The cut-off values of $p < 0.05$ and log2FoldChange ($|\log2FC| > 2$ were set as the threshold for significant DEGs selection. Finally, a total of 1,366 DEGs were exploited for gene modules detection and functional analysis.

Co-Expression Modules and Hub Genes Detection

An overview of the system-genetics of the DEGs was illustrated in Figure 1A. Six gene modules were generated based on the criteria of $|\log2FC| > 2$ and $p < 0.05$ in these 1,366 DEGs. Each leaf (short vertical line) corresponds to a gene clustered together to form a gene module. After excluding non-significantly expressed genes, we used these 1,366 significant genes and applied the WGCNA package to compile the systematic network (Figure 1B). In addition, by integrating these six gene modules, 36 hub genes **(Supplementary Table I)** were detected in the system network with the following parameters: $|\log2FC| > 2$, $p < 0.05$, and the degree of the connectivity is greater than 10.

Functional Analysis of Gene Modules

Functional enrichment of the significant gene modules was performed to determine whether the outcomes were significantly associated with the biological processes influencing the hazards of OVSC patients' survival prognosis. Significant enrichments were observed in brown and purple gene modules,

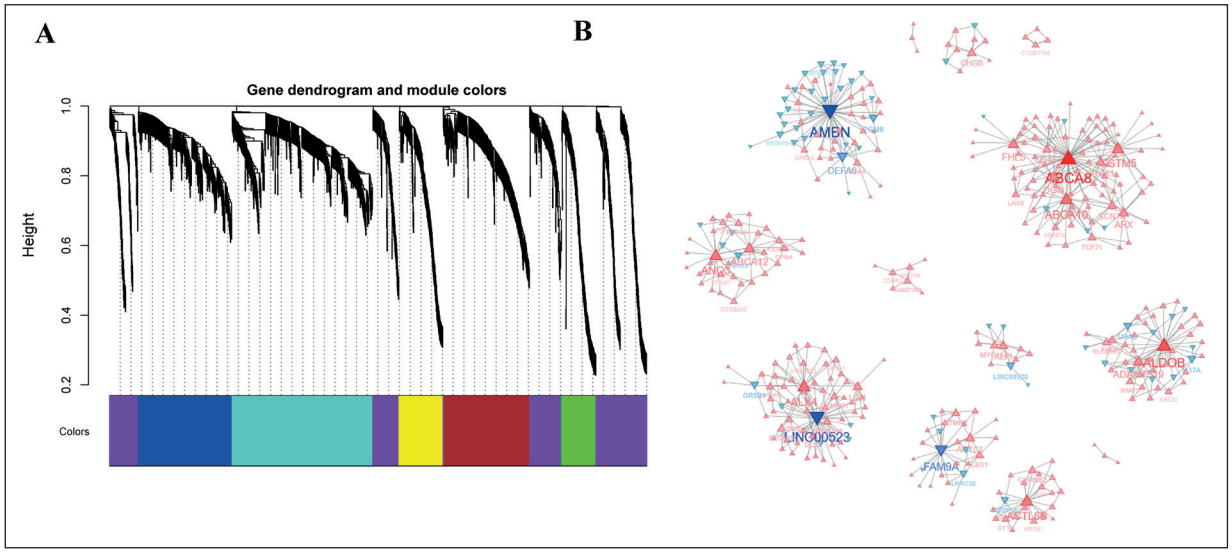

Figure 1. Gene co-expression modules and hub genes detection. **A,** Gene dendrograms through dynamic tree-cutting method. Six gene modules were generated with the following parameters: min cluster size = 50, cut height = 0.95, deep split = 2. **B,** Weighted co-expression network interconnection of these six gene modules. Each node represents a gene whose size and shape indicating its own weighted power and expressed level. Red and blue symbols indicating high-expressed and low-expressed genes, respectively).

such as hormone activity and receptor-binding biological processes (Table II). These gene terms include a few hub genes (ABCA8, ABCA10, TCF21, CRHBP, FHL5, ACTC1, ADAMTS19, GSTM5, ALX4, SPRR2F, etc.), which imply a close relationship between hormones and angiogenesis.

The enrichment results of the Turquoise module mainly focused on immune functions, including T cell activation, leukocyte differentiation, and cytokine receptor activity. The development of the genitourinary and renal system was mainly associated with the yellow module consisting of 96 genes (Table II). It is worth noting that the Wnt signaling pathway, which has critical functions in carcinogenesis and embryonic development, is significantly enriched in the biological processes and molecular functions in the green module. The main hub genes listed in Table II correspond to the enrichment results for each gene module.

Potential Prognostic Biomarkers

Prognostic biomarkers may not be hub genes; therefore, we will focus on the 36 hub genes that have patients' survival time. Due to their varied expressing levels and vital status, prognostic biomarkers differ in their ability to predict one gene from another. Here, we defined the prognostic biomarkers based on the significance values within each sample group and the Kaplan-Meier estimates. Finally, we obtained 14 significant genes, including ABCA10, DCX, LRRC30, ALX4, DKK4, SGCZ, ANKS4B, FHL5, SPRR2F, CHRNG, GABRR1, STMN2, CRHBP, and GSTM5 (Figure 2) that can be used as prognostic biomarkers in evaluating the prognosis of OVSC. Interestingly, we observed that all these candidates were significantly expressed between the dead and living patients in the two-tailed student *t*-test **(Supplementary Figure 2)**.

In addition, in order to determine and ensure the prognostic significance of these biomarkers, a non-parametric bootstrapping procedure was performed based on the distribution of expected HRs at each time point. As with the distribution of biomarkers being studied, we can use the observed HR distribution, as part of the assessment. As results showed in Figure 3, all these genes exhibited different line trends along with the time elapsed, which were used to evaluate HR with probability. Furthermore, we divided the 14 genes into high-expressed and low-expressed groups and compared them separately. Then, we found that the bootstrapping procedure showed discreteness at both the head and tail of the lifespan, which were not considered in the assessment. According to the stimulating results, the genes, including ALX4, DCX, GABRR1, and STMN2 were positively correlated with HR, whereas ABCA10, CHRNG, CRHBP, GSTM5, and SPRR2F were negatively correlated (Figure 3).

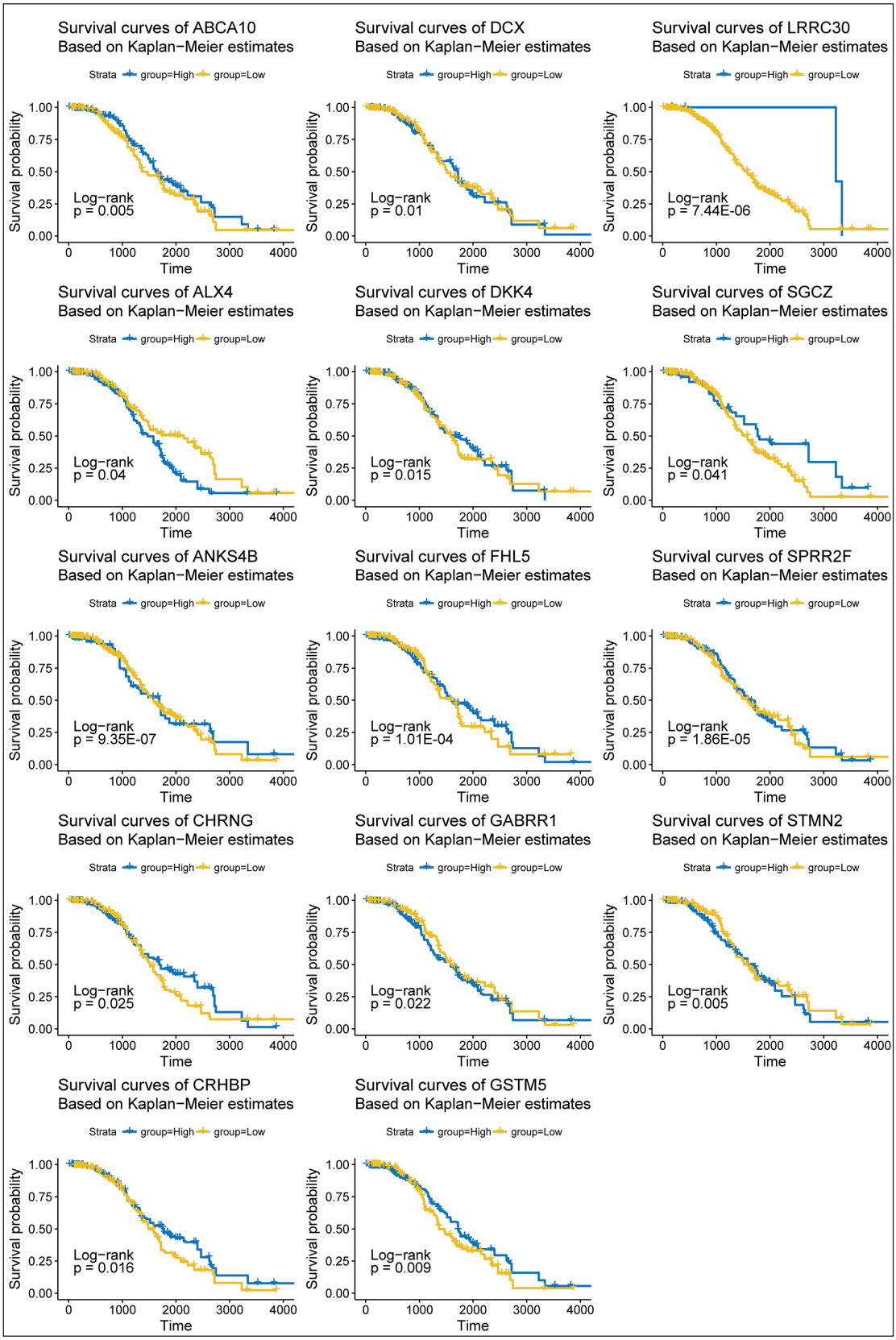

Figure 2. Kaplan-Meier estimated survival curves of potential biomarkers. Blue line indicates high-expression group and brown line indicates low-expression group.

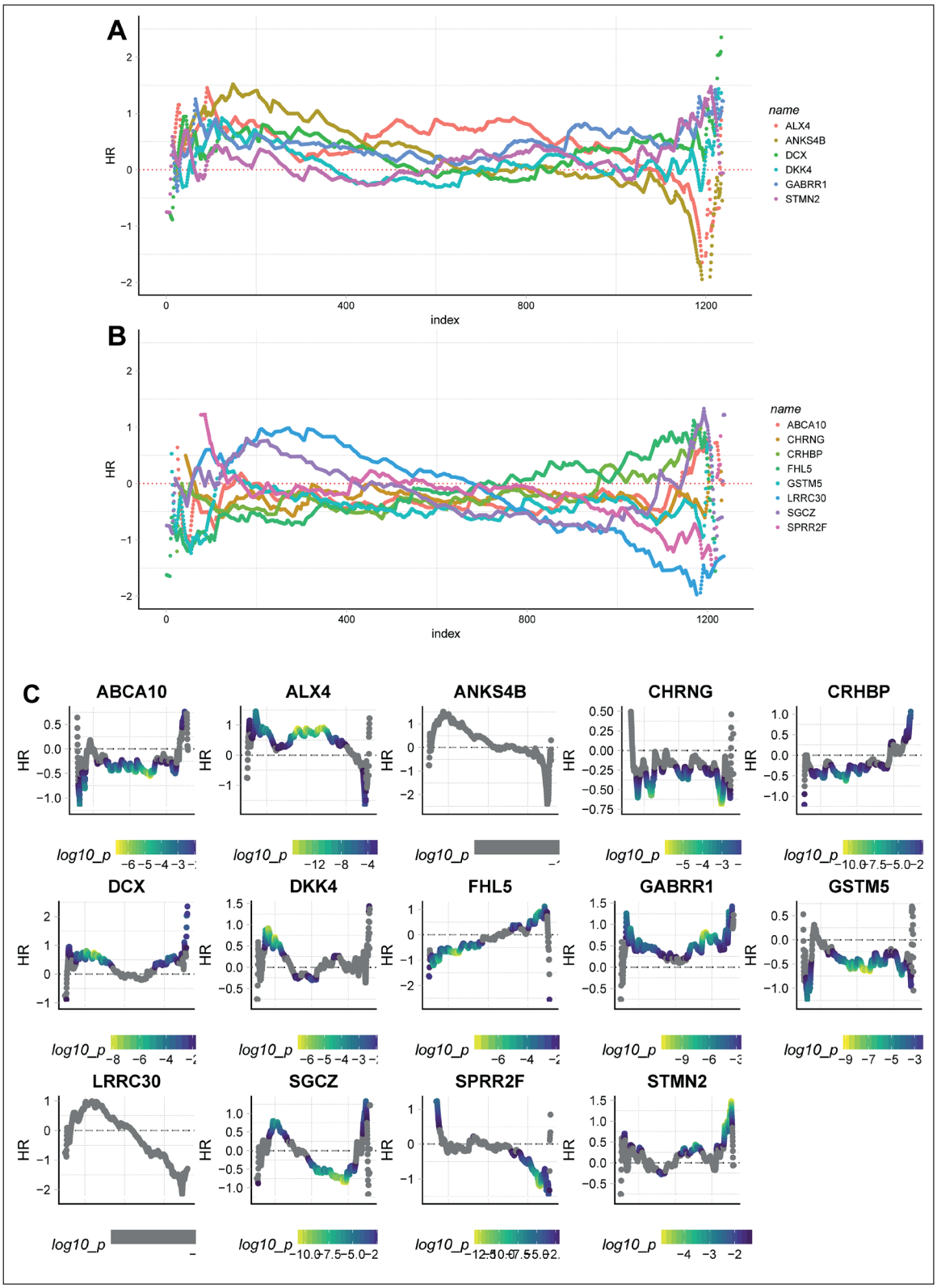

Figure 3. Distribution of expected HRs for potential prognostic biomarkers. **A, B,** Compared HRs distribution lines of highly and lowly expressed genes, respectively. **C,** HRs distribution line of each gene corresponding log10 *p*-values.

Discussion

In this study, we performed a systemic gene expression of the weighted network by integrating RNA-sequencing profiles of OVSC through WGCNA. We have developed a PPI network containing 1,336 DEGs that were significantly associated with OVSC. Functional enrichment analysis further revealed that the gene modules were enriched in genes involved in angiogenesis, T cell activation, and hormone receptor binding. It is biologically considered that the 36 important hub genes (ABCA10, TCF21, CRHBP, etc.) of this integrated network are related to the survival prognosis of OVSC. To further determine and ensure the capacity of the potential prognostic biomarkers, Kaplan-Meier estimated the survival curves of these hub genes, which were used to detect significantly expressed genes between dead and living OSVC patients. Finally, 14 genes included ABCA10, DCX, LRRC30, ALX4, DKK4, SGCZ, ANKS4B, FHL5, SPRR2F, CHRNG, GABRR1, STMN2, CRHBP, and GSTM5 were selected as potential biomarkers for prognosis of OVSC survival.

As a complex malignancy, ovarian cancer is characterized by a variety of cellular dysregulation that interacts in a complex network environment[12]. Furthermore, as the basic components of the network, the intersections of genes and their dynamic connections are essential to the coordination of biological processes[12]. Therefore, it is reasonable to perform a dynamic network-based modular analysis of biomarkers discovery. Unlike traditional network-based analyses, which usually ignore the existence of a patients' survival hazards or correlation between gene expressions[13,14], we have facilitated the dynamic responded-intersections based on survival and weighted co-expression PPI network between high-expressed and low-expressed DEGs, and have identified 14 potential prognostic biomarkers for OVSC. We have further evaluated their predictive ability through a non-parametric bootstrapping procedure[15]. Homeoprotein, Aristaless-Like Homeobox4 (ALX4), alias of Homeobox Protein Aristaless-Like 4, one of the potential prognostic biomarkers, was reported related to epithelial to mesenchymal transition (EMT) and intrusion of ovarian cancer[16]. The exogenous expression of ALX4 promoted EMT and invasion, while its depletion suppressed invasion and induced reversion of epithelium transformation. Homeoproteins are a family of transcription factors with conserved homeobox domains[17]. Researchers[18,19] have discovered that the deregulated expression of homeoproteins is related to the development of various tumors. Multiple homeoproteins are overexpressed in several cancers, and their expression induces invasion, angiogenesis, and tumor progression[20,21]. In addition, another highly expressed homeoprotein distal-less homeobox 2 (DLX2) was established in the analysis results of the hub genes; however, no significant survival prognosis was found, although in another microarray analysis it was reported that DLX2 is related to the metastasis of ovarian cancer[22]. Wnt signaling pathway, a noteworthy advance in this research in the Green module, in company with high-expressed hub genes of Dickkopf-4 (DKK4) and Small Proline-Rich Protein 4 (SPRR4), have been reported to be involved in proliferation, migration, and invasion of ovarian cancer[23,24]. Additionally, DKK4 may help to predict the progression and prognosis of epithelial ovarian cancer[23] by promoting the phosphorylation of c-JUN and JNK. DKK4 is a member of the DKK family and first binds to lipoprotein receptor-related protein 5/6 (LRP5/6), acting as an antagonist of WNT proteins, inducing the binding complex endocytosis and inhibiting WNT/β-catenin activation[25-27]. This, directly and indirectly, scientific evidence indicates the prognostic value of DKK4 in OVSC patients.

Conclusions

In summary, in order to determine the potential prognostic biomarkers of OVSC, we primarily adopt an integrated approach that has not yet been reported. Despite the limitation of this study, the normal samples were excluded due to their insufficient proportion. The conclusions of this study are still developing several novel potential prognostic biomarkers that will benefit the survival prognosis of OVSC. In this survival prognosis associated with OVSC events, a number of valuable prognostic biomarkers and therapeutic targets were provided for validations in the future and ultimately contributing to the underlying mechanisms of OVSC.

Conflict of Interest

The Authors declare that they have no conflict of interests.

Sources of Funding
This research was supported by the National Natural Science Foundation of China (No. 81671441).

References

1) Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. CA Cancer J Clin 2017; 67: 7-30.
2) Slomski A. Screening women for ovarian cancer still does more harm than good. JAMA 2012; 307: 2474-2475.
3) Simon R. Diagnostic and prognostic prediction using gene expression profiles in high-dimensional microarray data. Br J Cancer 2003; 89: 1599-1604.
4) Francis P, Namløs HM, Müller C, Edén P, Fernebro J, Berner JM, Bjerkehagen B, Akerman M, Bendahl PO, Isinger A, Rydholm A, Myklebost O, Nilbert M. Diagnostic and prognostic gene expression signatures in 177 soft tissue sarcomas: hypoxia-induced transcription profile signifies metastatic potential. BMC Genomics 2007; 8: 73.
5) Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. Nature 2011; 474: 609-615.
6) Bast RC Jr, Hennessy B, Mills GB. The biology of ovarian cancer: new opportunities for translation. Nat Rev Cancer 2009; 9: 415-428.
7) Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet 2011; 12: 56-68.
8) Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011; 144: 646-674.
9) Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 2013; 41: D808-D815.
10) Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008; 9: 559.
11) Yu G, Wang LG, Han Y, He QY. ClusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012; 16: 284-287.
12) Duan C, Cao Z, Tang F, Jian Z, Liang C, Liu H, Xiao Y, Liu L, Ma R. MiRNA-mRNA crosstalk in myocardial ischemia induced by calcified aortic valve stenosis. Aging (Albany NY) 2019; 11: 448-466.
13) Li J, Roebuck P, Grünewald S, Liang H. SurvNet: a web server for identifying network-based biomarkers that most correlate with patient survival data. Nucleic Acids Res 2012; 40(Web Server issue): W123-W126.
14) Lee YC, Huang CC, Lin DY, Chang WC, Lee KH. Overexpression of centromere protein K (CENPK) in ovarian cancer is correlated with poor patient survival and associated with predictive and prognostic relevance. PeerJ 2015; 3: e1386.
15) Ge J, Wu XM, Yang XT, Gao JM, Wang F, Ye KF. Role of long non-coding RNA SNHG1 in occurrence and progression of ovarian carcinoma. Eur Rev Med Pharmacol Sci 2018; 22: 329-335.
16) Yuan H, Kajiyama H, Ito S, Chen D, Shibata K, Hamaguchi M, Kikkawa F, Senga T. HOXB13 and ALX4 induce SLUG expression for the promotion of EMT and cell invasion in ovarian cancer cells. Oncotarget 2015; 6: 13359-13370.
17) Zhang SR, Li DB, Xue JW. The expression of TP53 pathway-related proteins in ovarian carcinoma transplanted subcutaneously in nude mice. Eur Rev Med Pharmacol Sci 2018; 22: 1258-1264.
18) Rawat VP, Humphries RK, Buske C. Beyond Hox: the role of ParaHox genes in normal and malignant hematopoiesis. Blood 2012; 120: 519-527.
19) Cillo C, Cantile M, Faiella A, Boncinelli E. Homeobox genes in normal and malignant cells. J Cell Physiol 2001; 188: 161-169.
20) Fung FK, Chan DW, Liu VW, Leung TH, Cheung AN, Ngan HY. Increased expression of PITX2 transcription factor contributes to ovarian cancer progression. PLoS One 2012; 7: e37076.
21) Lee WK, Chakraborty PK, Thévenod F. Pituitary homeobox 2 (PITX2) protects renal cancer cell lines against doxorubicin toxicity by transcriptional activation of the multidrug transporter ABCB1. Int J Cancer 2013; 133: 556-567.
22) Lancaster JM, Dressman HK, Clarke JP, Sayer RA, Martino MA, Cragun JM, Henriott AH, Gray J, Sutphen R, Elahi A, Whitaker RS, West M, Marks JR, Nevins JR, Berchuck A. Identification of genes associated with ovarian cancer metastasis using microarray expression analysis. Int J Gynecol Cancer 2006; 16: 1733-1745.
23) Wang S, Wei H, Zhang S. Dickkopf-4 is frequently overexpressed in epithelial ovarian carcinoma and promotes tumor invasion. BMC Cancer 2017; 17: 455.
24) Liu H, Shi H, Fan Q, Sun X. Cyclin Y regulates the proliferation, migration, and invasion of ovarian cancer cells via Wnt signaling pathway. Tumour Biol 2016; 37: 10161-10175.
25) Niehrs C. Function and biological roles of the Dickkopf family of Wnt modulators. Oncogene 2006; 25: 7469-7481.
26) Hiramitsu S, Terauchi M, Kubota T. The effects of dickkopf-4 on the proliferation, differentiation, and apoptosis of osteoblasts. Endocrinology 2013; 154: 4618-4626.
27) Sima J, Piao Y, Chen Y, Schlessinger D. Molecular dynamics of Dkk4 modulates Wnt action and regulates Meibomian gland development. Development 2016; 143: 4723-4735.