

PRIMARY RESEARCH

Open Access



# Identification of four hub genes as promising biomarkers to evaluate the prognosis of ovarian cancer in silico

Jingxuan Chen<sup>1,2†</sup>, Yun Cai<sup>3†</sup>, Rui Xu<sup>1,2†</sup>, Jiadong Pan<sup>4</sup>, Jie Zhou<sup>5\*</sup> and Jie Mei<sup>2,4\*</sup> 

## Abstract

**Background:** Ovarian cancer (OvCa) is one of the most fatal cancers among females in the world. With growing numbers of individuals diagnosed with OvCa ending in deaths, it is urgent to further explore the potential mechanisms of OvCa oncogenesis and development and related biomarkers.

**Methods:** The gene expression profiles of GSE49997 were downloaded from the Gene Expression Omnibus (GEO) database. Weighted gene co-expression network analysis (WGCNA) was applied to explore the most potent gene modules associated with the overall survival (OS) and progression-free survival (PFS) events of OvCa patients, and the prognostic values of these genes were exhibited and validated based on data from training and validation sets. Next, protein–protein interaction (PPI) networks were built by GeneMANIA. Besides, enrichment analysis was conducted using DAVID website.

**Results:** According to the WGCNA analysis, a total of eight modules were identified and four hub genes ( $MM > 0.90$ ) in the blue module were reserved for next analysis. Kaplan–Meier analysis exhibited that these four hub genes were significantly associated with worse OS and PFS in the patient cohort from GSE49997. Moreover, we validated the short-term (4-years) and long-term prognostic values based on the GSE9891 data, respectively. Last, PPI networks analysis, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis revealed several potential mechanisms of four hub genes and their co-operators participating in OvCa progression.

**Conclusion:** Four hub genes (*COL6A3*, *CRISPLD2*, *FBN1* and *SERPINF1*) were identified to be associated with the prognosis in OvCa, which might be used as monitoring biomarkers to evaluate survival time of OvCa patients.

**Keywords:** Ovarian cancer, WGCNA, Bioinformatic analysis, Prognosis

## Background

Ovarian cancer (OvCa) is a common cancer which has the highest morbidity and quietly poor prognosis among gynecological malignancies worldwide. In United States, OvCa causes approximately 14 thousands death patients in 2018 [1]. With the continuous improvement of comprehensive therapy, patients with early stage OvCa seem to have satisfactory prognosis that the 5-year survival rate reaching 93%. Nevertheless, since the majority of patients, precisely more than 80%, would be hard to be diagnosed until the tumor at FIGO stage III or stage IV,

\*Correspondence: zhoujie1982\_1@163.com; meijie1996@njmu.edu.cn

†Jingxuan Chen, Yun Cai and Rui Xu contributed equally to this manuscript

<sup>2</sup> Cytoskeleton Research Group & First Clinical Medicine College, Nanjing Medical University, No. 101 Longmian Road, Nanjing 211166, China

<sup>5</sup> Department of Gynecology and Obstetrics, Affiliated Wuxi Maternal and Child Health Hospital of Nanjing Medical University, No.48, Huaihua Road, Wuxi 214023, China

Full list of author information is available at the end of the article



© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

leading to a considerable number of mortality [2]. Hence, increasing number of researchers focus on this horrible disease and attempt to explore novel procedures for early diagnosis and treatment. However, early diagnostic strategies and reliable models to guide therapy and evaluate prognosis have been lacking up to now.

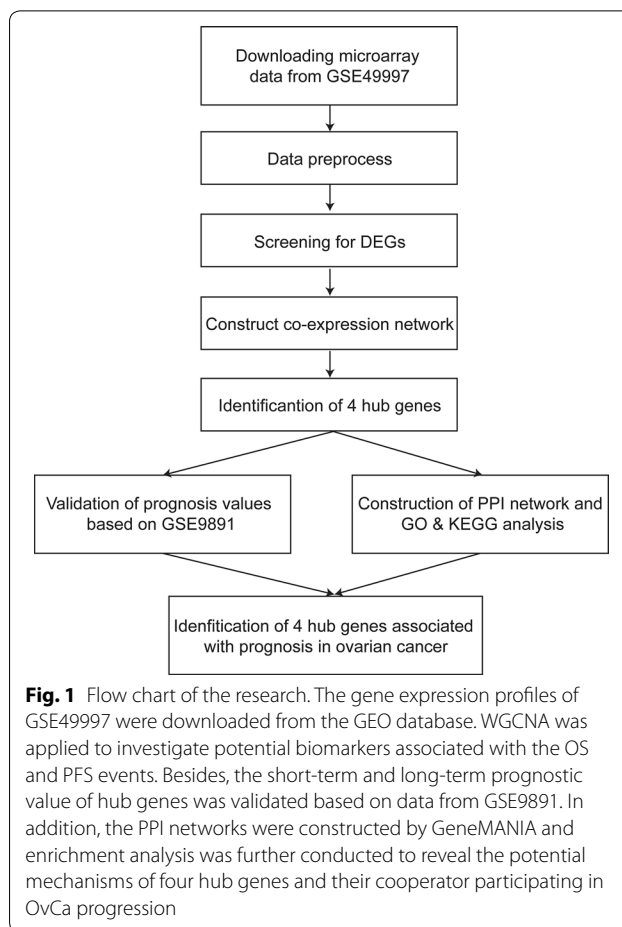
In the past decades, vigorously developing computer technology has largely promoted the flourish of big data applications. As an emerging biomedical auxiliary research technology, bioinformatics analysis has been widely applied to several aspects of clinical or basic medical research. Weighted gene co-expression network analysis (WGCNA) is a systematic biological method which describes the pattern of gene association between different samples [3]. Researchers utilize WGCNA to identify gene sets of interest with information of thousands of significantly altered genes or all genes, and then perform significant association analyses with phenotypes. At present, several laboratories have applied this technology into their researches [4–6]. Meanwhile, in the field of cancer research, investigators tend to take advantage of WGCNA for systematic analysis of phenotypes, especially for developing novel prognostic models [7, 8].

GSE49997, a microarray containing 204 OvCa samples, was contributed by Dietmar et al. in 2014 [9]. Dietmar et al. validated the prognostic impacts of a molecular subtype in OvCa on overall survival (OS) and progression-free survival (PFS) based on this microarray [10]. Given the integrated follow-up information and gene expression data in this microarray, we re-assessed the above-mentioned data and finally appraised four potential biomarkers predicting prognosis of OvCa patients through WGCNA analysis. We further conducted verification analysis on prognostic values of four genes through another microarray, GSE9891. In conclusion, we found out four genes (*COL6A3*, *CRISPLD2*, *FBN1* and *SERPINF1*) associated poor prognosis, suggesting these genes function as potential biomarkers to evaluate the prognosis of OvCa patients.

## Materials and methods

### Acquisition of microarray data and pre-process

The workflow of our research was summarized in Fig. 1. The array profiles of GSE49997 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE49997>) [9] were downloaded from the Gene Expression Omnibus (GEO) database. Besides, four hub gene expression data and survival data corresponding to GSE9891 were downloaded from the Kaplan–Meier Plotter (<http://kmplot.com/analysis/>) [11]. The basic clinic-pathological features of two microarrays were described in Table 1 according to original publications of GSE49997 and GSE9891 [9, 10]. Before further analysis, array profiles of GSE49997 were



pre-processed by background correction, quantile normalization and probe summarization. After matching the gene expression data and survival information, 194 samples from GSE49997 were retained in the current research. For further WGCNA analysis, the top 25% different expression genes (DEGs) from GSE49997 dataset according to analysis of variance (3,837 genes) were retained.

### Co-expression network construction and identification of hub genes

After pre-processing the GSE49997 microarray data, the expression profile of these 3,837 genes was sent to construct a gene co-expression network using the WGCNA package in R language [12]. The idea of a soft threshold is to continually elementize the elements in the Adjacency Matrix through a weight function and the choice of the soft threshold  $\beta$  is bound to affect the result of module identification. To create a network with a nearly scale-free topology, we installed the soft threshold power of  $\beta=3$  (scale free  $R^2=0.868$ ). Adjacency matrices were calculated and transformed into the topological overlap matrix (TOM). The dynamic tree cut algorithm was

**Table 1 The basic clinic-pathological features of OvCa patients in two datasets**

Characteristics	n	Cases (%)	
		GSE49997	GSE9891
Histologic subtype			
Serous	435	171 (88.1%)	264 (92.6%)
Non-serous	44	23 (11.9%)	21 (7.4%)
FIGO stage			
Stage 1	24	0 (0%)	24 (8.4%)
Stage 2	27	9 (4.6%)	18 (6.3%)
Stage 3	371	154 (79.4%)	217 (76.1%)
Stage 4	53	31 (16.0%)	22 (7.7%)
Unknown	4	0 (0%)	4 (1.4%)
Grade			
Grade 1	30	11 (5.7%)	19 (6.7%)
Grade 2	136	39 (20.1%)	97 (34.0%)
Grade 3	307	143 (73.7%)	164 (57.5%)
Unknown	6	1 (0.5%)	5 (1.8%)
Residual tumor			
No	221	137 (70.6%)	84 (29.5%)
Yes	221	57 (29.4%)	164 (57.5%)
Unknown	37	0 (0%)	37 (13.0%)

applied to detect gene modules. Gene significance (GS) was defined as the correlation coefficient between gene expression and module traits. The module eigengene was calculated as a summary profile for each module. Module significance was defined as the correlation coefficient between a module's eigengene and traits. Module membership (MM) was defined by the correlation coefficient of the module eigengene and gene expression profile. Genes with MM values above 0.90 were considered to be the modules' representative genes with potential critical functions.

#### Survival analyses and further authentication of key genes

The prognostic impacts of four genes for OS and PFS were evaluated by Kaplan–Meier analysis. To further verify the significant prognostic values of four hub genes, we used the array profiles and clinical data from GSE9891 to conduct survival analysis. Given that the follow-up information from training set was collected within about 4 year, we firstly set 48 months as the end point of follow-up to evaluate the short-term prognostic values of four genes. Besides, the long-term prognostic values were also validated by taking full advantage of the survival data from the validation set. For Kaplan–Meier analysis, all cases were ranked based on hub genes expression levels and further divided into two groups according to the median expression of these genes.

#### Construction of protein–protein interaction network

Protein–protein interaction (PPI) network was been constructed by GeneMANIA (<https://genemania.org/>) [13], an online server that explore interconnections between proteins in term of physical interaction, co-expression, predicted, co-localization, common pathway, genetic interaction and shared protein domains. In this research, GeneMANIA was used for PPI analysis of four hub genes at the gene level.

#### Gene Ontology and Kyoto Encyclopedia of Genes and Genomes analysis

The Database for Annotation, Visualization and Integrated Discovery (DAVID, <https://david.ncifcrf.gov/>) [14] was applied to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses of four hub genes and their most relevant cooperators. The human genome (*Homo sapiens*) was selected as the background variables. Enrichment terms were considered statistically significant when the FDR were less than 0.05 and the top 10 terms of each analysis were retained to plot bubble chart.

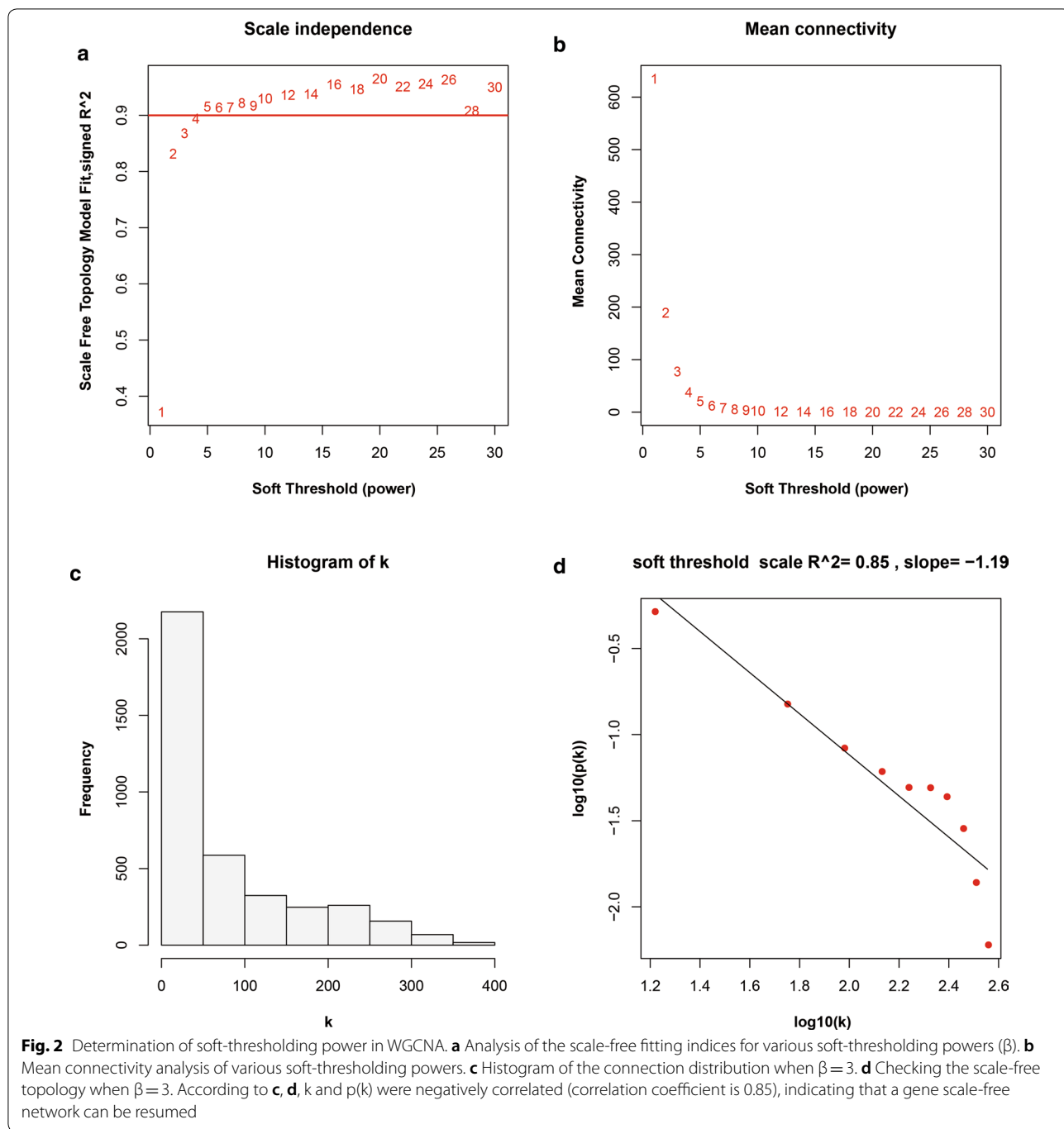
#### Statistical analysis

All statistical analyses were performed using SPSS 25.0 software and R 3.5.1 software. Kaplan–Meier survival plots were generated with survival curves compared by log-rank test, but Tarone–Ware test was used when obvious crossover between the groups was observed in survival plots. For all analyses, differences were considered statistically significant if the P values were less than 0.05.

## Results

#### Weighted co-expression network construction and key hub genes identification

We applied the R package for WGCNA in the construction of a co-expression network and then 3837 DEGs with similar expression features were submitted to modules through cluster analysis. In our research, we selected the power of  $\beta=3$  (scale free  $R^2=0.868$ ) as the soft threshold to ensure a scale-free network (Fig. 2a–d). Then, we extracted eight modules for next analysis (Fig. 3a). Afterwards, we took advantage of a heatmap and meta-modules aiming to visualize the gene network (Fig. 3b, c). The blue module, most remarkably correlated with both OS ( $R^2=0.25$ ;  $P=5e-4$ ) and PFS ( $R^2=0.14$ ;  $P=0.05$ ) events, was shown to be of remarkable value in the evaluation of OvCa prognosis. Subsequently, we selected 4 genes (*COL6A3*, *CRISPLD2*, *FBNI* and *SERPINF1*) in the blue module with MM values above 0.90 for further analysis,

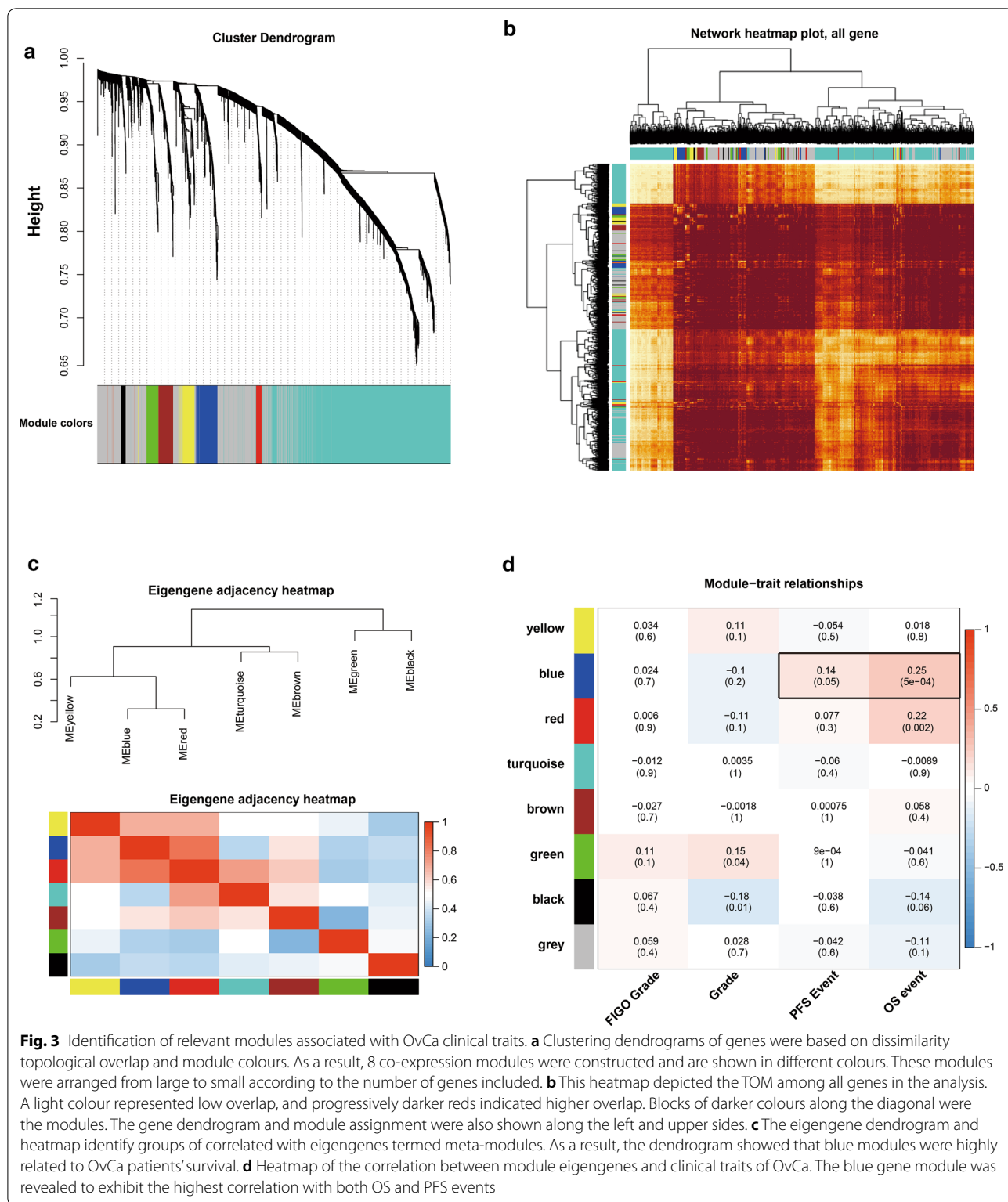


which were regarded as typical genes to exhibit crucial functions (Table 2).

**Prognostic value of key genes in predict OS and RFS in ovarian**

To assess the prognostic values of four hub genes, we next performed Kaplan–Meier analysis. As shown in Fig. 4, high mRNA expression levels of COL6A3 ( $P = 0.007$ ),

CRISPLD2 ( $P < 0.001$ ), FBN1 ( $P = 0.012$ ) and SERPINF1 ( $P = 0.021$ ) were significantly associated with worse OS in OvCa patients. Meanwhile, high mRNA expression levels of COL6A3 ( $P = 0.031$ ), CRISPLD2 ( $P = 0.011$ ), FBN1 ( $P = 0.021$ ) and SERPINF1 ( $P = 0.034$ ) were notably associated with poor PFS as well. Overall, these findings revealed the promising prognostic values of four hub genes in OvCa.



**Validation of hub genes based on GSE9891 microarray data**

To validate the prognostic values of these four hub genes

in OvCa, we used microarray data from GSE9891 to execute survival analysis for the four hub genes. Given that the follow-up information from training dataset

**Table 2 Four hub genes identified by WGCNA analysis in OvCa**

Gene ID	Ensembl ID	Gene description
COL6A3	ENSG00000163359.15	Collagen, type VI, alpha 3
CRISPLD2	ENSG00000103196.11	Cysteine-rich secretory protein LCCL domain containing 2
FBN1	ENSG00000166147.13	Fibrillin 1
SERPINF1	ENSG00000132386.10	Serpin peptidase inhibitor, clade F, member 1

(GSE49997) was collected within 4 year, we firstly set 48 months as the end point of follow-up to assess the short-term prognostic values of four genes. At first, to guarantee the homogeneity of two datasets, we compared the difference between OS and PFS events, and the result showed that no statistical significance was found (Table 3). As shown in Fig. 5, high mRNA expression levels of COL6A3 ( $P=0.002$ ), CRISPLD2 ( $P<0.001$ ), FBN1 ( $P=0.005$ ) and SERPINF1 ( $P=0.002$ ) were remarkably associated with worse OS in OvCa patients. Meanwhile, high mRNA expression levels of COL6A3 ( $P=0.002$ ), CRISPLD2 ( $P<0.001$ ), FBN1 ( $P=0.005$ ) and SERPINF1 ( $P<0.001$ ) were also notably associated with unfavorable PFS.

Furthermore, the long-term prognostic values were also validated by taking full advantage of the survival data from the validation dataset. As shown in Fig. 6, high mRNA expression levels of COL6A3 ( $P=0.001$ ),

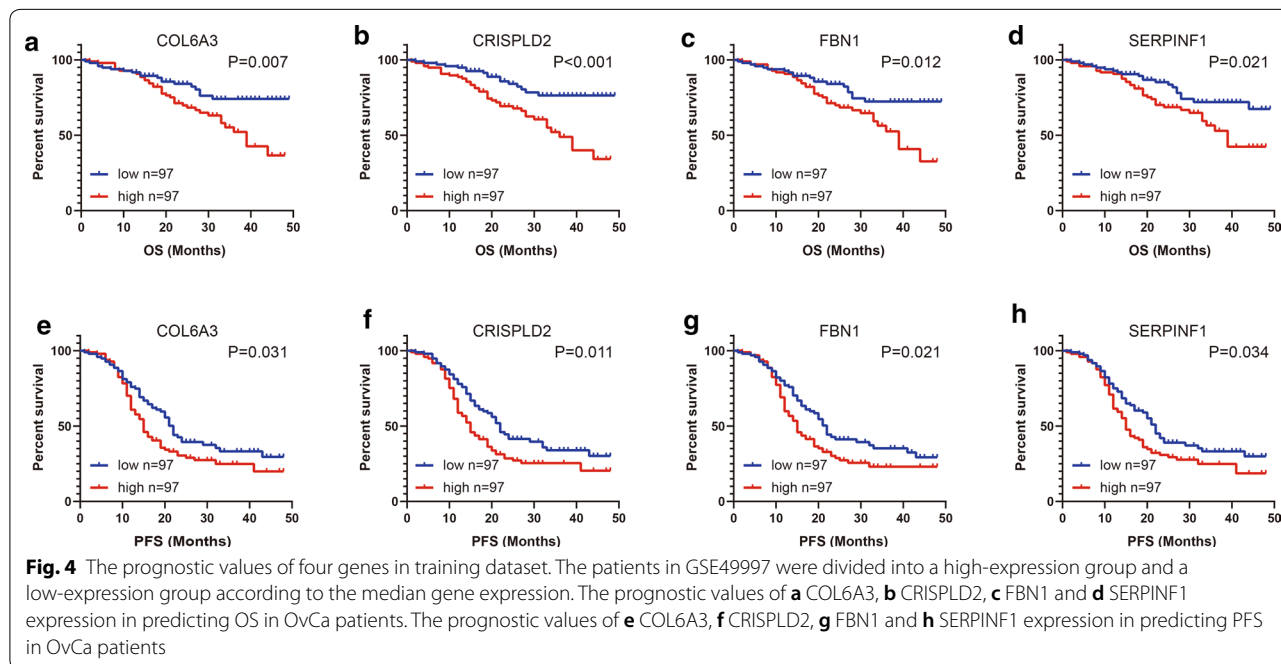
**Table 3 Comparison of patients' survival events in two datasets**

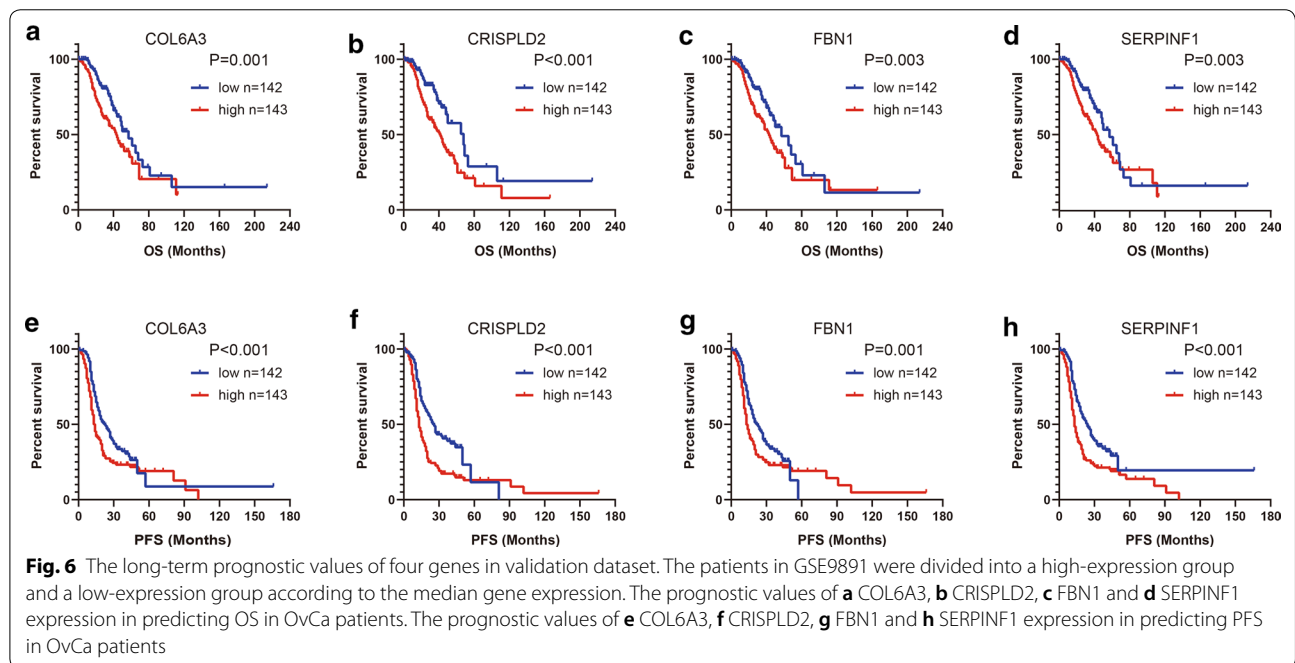
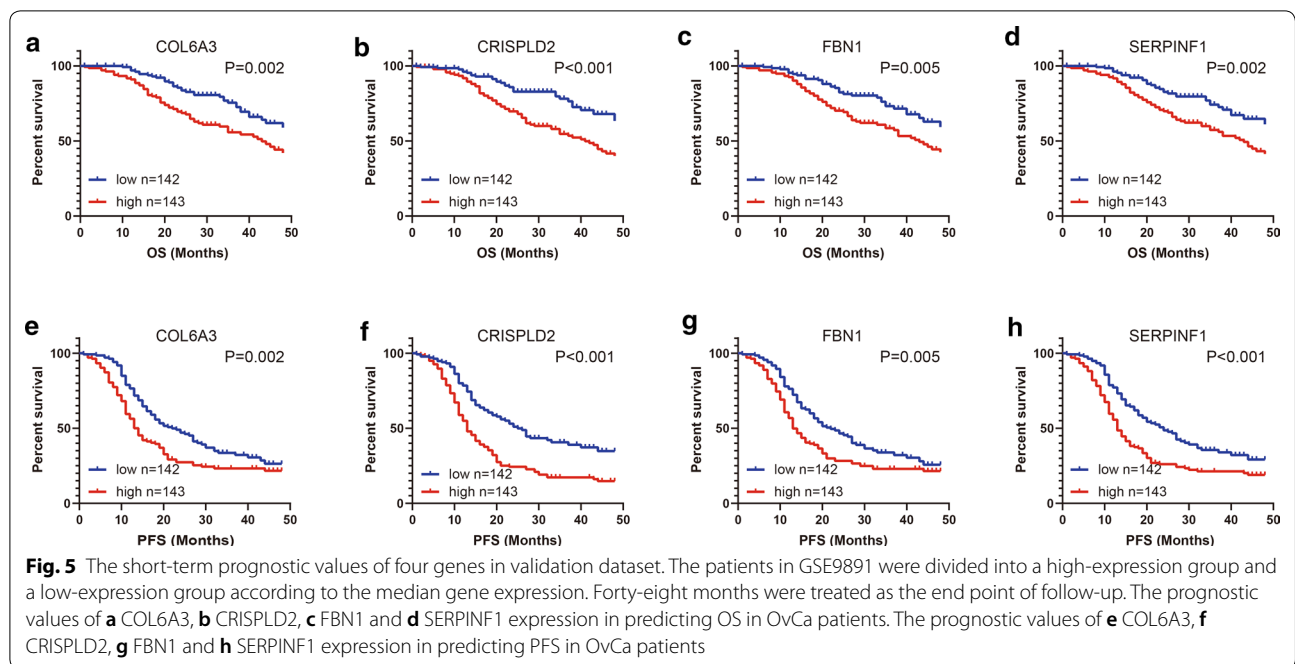
Survival events	n	Cases (%)		$\chi^2$	P value
		GSE49997	GSE9891		
OS event				0.83	> 0.05
Alive	312	137 (70.6%)	190 (66.7%)		
Dead	167	57 (29.4%)	95 (33.3%)		
PFS event				0.15	> 0.05
Non-progressive	162	70 (36.1%)	98 (34.4%)		
Progressive	317	124 (63.9%)	187 (65.6%)		

CRISPLD2 ( $P<0.001$ ), FBN1 ( $P=0.003$ ) and SERPINF1 ( $P=0.003$ ) were remarkably associated with worse OS. Besides, high mRNA expression levels of COL6A3 ( $P<0.001$ ), CRISPLD2 ( $P<0.001$ ), FBN1 ( $P=0.001$ ) and SERPINF1 ( $P<0.001$ ) were significantly associated with unfavorable PFS in OvCa patients as well. Overall, these findings validated the prognostic values of these four hub genes in OvCa.

**Construction of PPI networks and enrichment analysis**

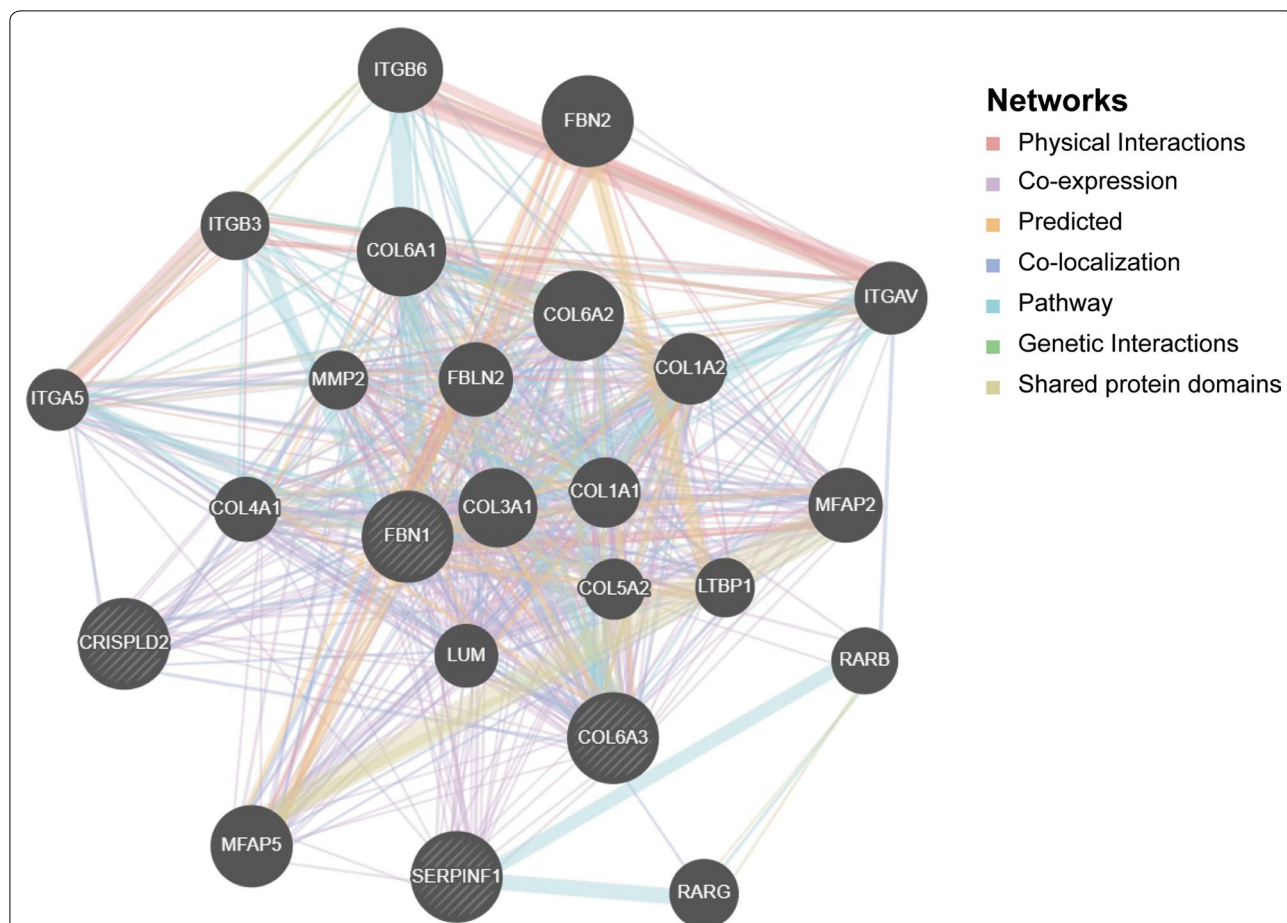
To explore potential mechanisms that these four hub genes participate in the carcinogenesis of OvCa, we applied GeneMANIA to set up a PPI network for these genes and the results revealed a series of cooperators (Fig. 7). Next, GO and KEGG analyses based on DAVID website were performed to identify the





functional enrichment of four hub genes and the inter-acted genes. GO analysis included three main functions of selected genes, including biological process (BP), cellular components (CC), and molecular functions (MF). The relative findings were represented in Fig. 8.

Besides, we supplied all GO and KEGG terms with statistical difference as Additional file 1: Table S1. These results revealed several potential mechanisms of four hub genes and their co-operators participating in OvCa progression, which provide novel insights for further study.



**Fig. 7** The PPI network of four genes constructed by GeneMANIA. PPI network for four hub genes was constructed in GeneMANIA website. The interconnections between proteins were explored in term of physical interaction, co-expression, predicted, co-localization, common pathway, genetic interaction and shared protein domains

**Discussion**

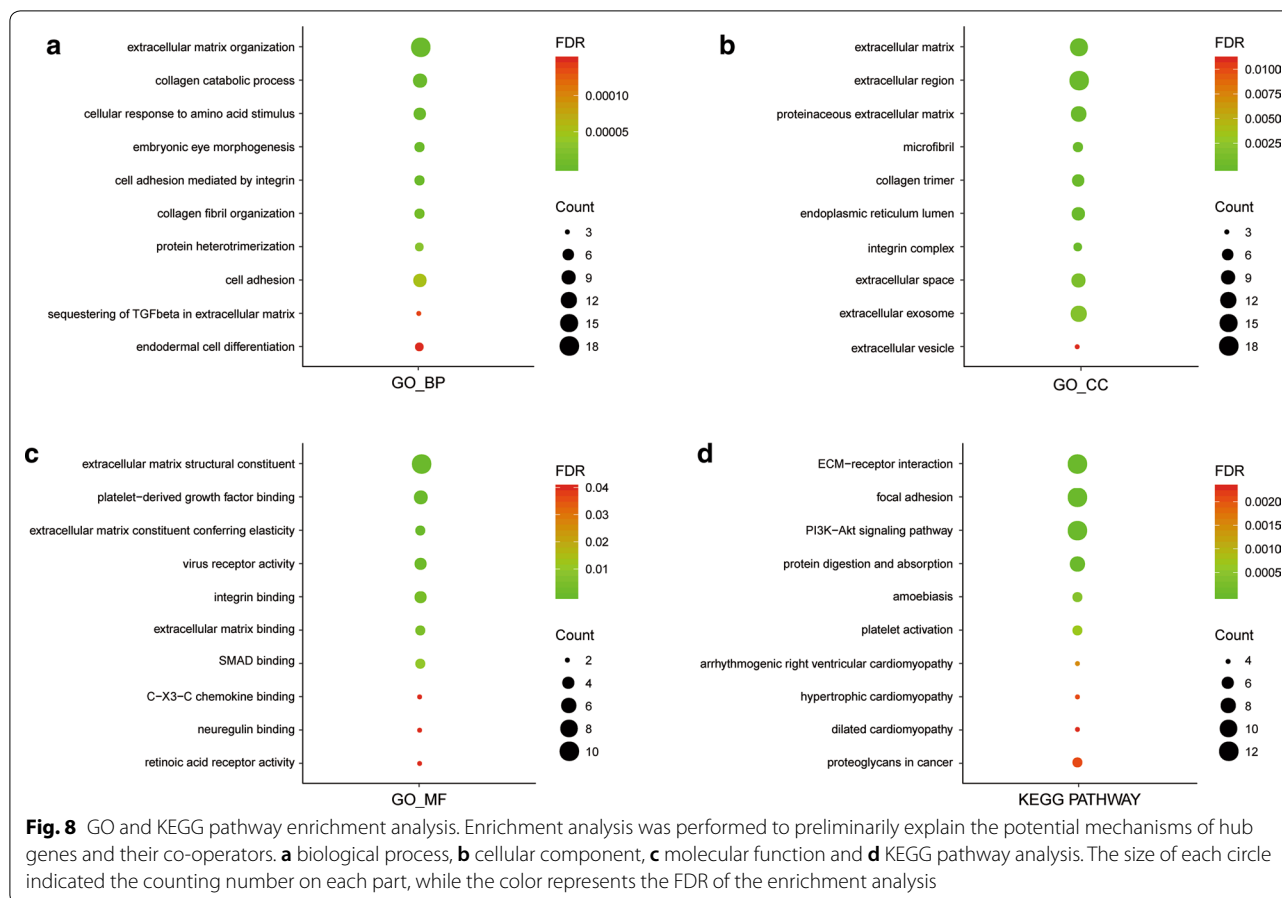
Currently, human genome research has moved towards a central phase. The successful expiry of the Human Genome Project marks a turning point of genomic research, meaning that the human genome research approaching a brand new stage of information extraction and data analysis [15, 16]. Bioinformatics is a new interdisciplinary subject developed in this context that combines the theoretical methods of biology, mathematics, physics, information science, and computer science. The core of bioinformatics is genomic informatics, including the acquisition, processing, storage, distribution and interpretation of genomic information. The key to genomic informatics is to “read” the nucleotide sequence of the genome, precisely, the exact location of all genes on the chromosome and the function of each DNA fragment [17]. Increasing evidence indicates that with the development of bioinformatics, we afford to use data analysis techniques to determine the role of mutations in

tumor formation. Identification of molecular biomarkers and expression profiles is frequently adopted in tumor systematization, diagnosis, and prediction of prognosis. The identification of oncogenesis-related genes, proteins, and cellular pathways facilitates the researches for more effective therapeutic drugs [18].

Co-expression analysis is an efficient strategy for gene/disease prediction analysis in large-scale datasets. In this research, we applied WGCNA to construct a gene co-expression network, to evaluate the relationships between genes and modules and to study the relationships between modules and clinical traits. In the analysis of the top 25% most variant genes, the blue module was identified to present the closest correlation with OS and PFS events, and 4 genes (*COL6A3*, *CRISPLD2*, *FBN1* and *SERPINF1*) with high connectivity were filtered from this module.

*COL6A3* encodes proteins to be the component of a beaded filament collagen, which can be found in majority





of connective tissues. Alpha-3 chain presents larger than other two kinds of chains composing collagen VI, for increasing number of a shared subdomain. As a primary cell-adhesive protein, collagen VI family frames a microfibrillar network supporting the function of skeletal muscle, skin, and cartilage, widely found in extracellular matrices [19, 20]. Mutations in type VI collagen genes easily lead to a range of muscle disorders, from normal defect like Bethlem Myopathy to the severer like Ullrich Scleroatonic Muscular Dystrophy. Recently, researchers have found conclusive evidences suggesting the association between COL6A3 and plurality of cancers [21–23]. COL6A3 is the most upregulated extracellular matrix (ECM) gene in cisplatin-resistant OvCa cells, and cultivation of cisplatin-sensitive cells in the presence of type VI collagen protein promotes resistance in vitro [22]. Besides, COL6A3 is also revealed to be associated with oxaliplatin resistance in OvCa [24].

CRISPLD2 was first identified in rat by Kaplan et al. [25, 26]. There is a potential mechanism that mediating CRISP and LCCL domain containing protein 2 to function in branching morphogenesis, with its nature of a glucocorticoid-inducible gene and part of a cytoskeletal

network [27]. Broadly expressing in gall bladder and placenta, LCCL domain variants affect diversity of adult human height [28]. CRISPLD2 has been proved to play significant role in either maintaining cell structure, participating in immune response, inhibiting inflammatory or involving in cell motility [29–32]. Actually, this gene is also implicated in cancers, lung defects, and epithelium growth [32, 33]. However, the role of CRISPLD2 in OvCa has not been observed yet.

As an extracellular matrix glycoprotein, FBN1 has been reported to promote the structure formation of calcium-binding microfibrils. Marfan syndrome seems to show great associations with mutations in *FBN1*, for the reason that the proteins encoded by this gene belongs to a member of the fibrillin family and can be hydrolyzed into FBN1 and hormone Asprosin protein [34, 35]. The former is a type of extracellular matrix component that serves as force-bearing structure support for connective tissue, whether elastic or nonelastic. The latter originates from white adipose tissue and serves as a regulator of glucose homeostasis [36]. Several lines of evidence indicate that FBN1 down-regulates the growth and sprouting of tumor endothelial cells via promoter histone modifications [37].

In OvCa, FBN1 functions as the downstream of Aurora-A and BRCA2, and promoted tumor metastasis through mediating the p53 and SLUG-associated signals [38].

SERPINF1 has nothing to do with serine protease inhibitory activity though this gene encodes a member of serpin family protein. Alternatively, the secreted protein is a type of strong angiogenesis blocker [39]. Moreover, researchers investigated patients with retinoblastoma and found that pathological cells in their body rely on SERPINF1, with its contribution to neuron differentiation. Mutations of this gene are also detected in osteogenesis imperfecta, type VI [39, 40]. Individuals suffering from type II diabetes have increasing circulating level of PEDE, confirmed a strong association with this gene at genome-wide significance [41, 42]. In cancerous diseases, SERPINF1 functions as a tumor suppressor in cervical cancer, which is downregulated by TXNDC5, resulting in stimulating cell migration, vasculogenic mimicry and angiogenesis [43]. Nwani et al. uncovered that SERPINF1 could maintain tumor suppressive functions in fibroblasts to prevent cancer-associated fibroblasts conversion [44]. However, in our research, SERPINF1 expression was associated with unfavorable prognosis, suggesting an opposite role in OvCa.

In our research, we further validated the prognostic values of 4 genes (*COL6A3*, *CRISPLD2*, *FBN1* and *SERPINF1*) in OvCa patients based on another microarray and the findings confirmed the reliable values for prognostic evaluation. Furthermore, we employed GeneMANIA to construct a PPI network and applied GO and KEGG analyses to identify the functional enrichment of four hub genes and the interacted genes and the results revealed several potential mechanisms of four hub genes and their co-operators in participating in OvCa progression. As results shown, four hub genes and their co-operators mainly participated in ECM regulation. The ECM mediates tissue development and homeostasis, while dys-regulation of ECM dynamics have been thought of as crucial drivers for both tumorigenesis and progression of cancers [45]. Taken together, our findings provided novel insights for hub genes mediating ECM-associated OvCa oncogenesis and development.

## Conclusion

Overall, our results furnished valuable directions for biomarker research on OvCa prognostic evaluation. We obtained crucial prognosis-associated genes based on WGCNA analysis and confirmed the prognostic values using data from GSE9891. Finally, four hub genes (*COL6A3*, *CRISPLD2*, *FBN1* and *SERPINF1*) were identified for further research, which might be employed as prospective biomarkers to assess OS and PFS in OvCa patients.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12935-020-01361-1>.

**Additional file 1: Table S1.** GO and KEGG pathway enrichment analysis (all terms with statistical difference).

### Abbreviations

OvCa: Ovarian cancer; WGCNA: Weighted gene co-expression network analysis; OS: Overall survival; PFS: Progression-free survival; GEO: The Gene Expression Omnibus; DEGs: Different expression genes; TOM: Topological overlap matrix; GS: Gene significance; MM: Module membership; PPI: Protein-protein interaction; DAVID: The Database for Annotation, Visualization and Integrated Discovery; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; COL6A3: Collagen, type VI, alpha 3; CRISPLD2: Cysteine-rich secretory protein LCCL domain containing 2; FBN1: Fibrillin 1; SERPINF1: Serpin peptidase inhibitor, clade F, member 1; BP: Biological process; CC: Cellular components; MF: Molecular functions; ECM: Extracellular matrix.

### Acknowledgements

Not applicable.

### Authors' contributions

JZ and JM conceived the study and participated in the study design, performance, coordination and manuscript writing. JC, YC, RX and JP carried out the assays and analysis. JZ and JM revised the manuscript. All authors reviewed the manuscript. All authors read and approved the final manuscript.

### Funding

Not applicable.

### Availability of data and materials

All data are included in the article.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup> School of Basic Medical Sciences, Nanjing Medical University, Nanjing 211166, China. <sup>2</sup> Cytoskeleton Research Group & First Clinical Medicine College, Nanjing Medical University, No. 101 Longmian Road, Nanjing 211166, China. <sup>3</sup> Department of Bioinformatics, Nanjing Medical University, Nanjing 211166, China. <sup>4</sup> First Clinical Medicine College, Nanjing Medical University, Nanjing 211166, China. <sup>5</sup> Department of Gynecology and Obstetrics, Affiliated Wuxi Maternal and Child Health Hospital of Nanjing Medical University, No.48, Huaishu Road, Wuxi 214023, China.

Received: 28 October 2019 Accepted: 17 June 2020

Published online: 24 June 2020

### References

1. Torre LA, Trabert B, DeSantis CE, Miller KD, Samimi G, Runowicz CD, et al. Ovarian cancer statistics, 2018. *CA Cancer J Clin*. 2018;68(4):284–96.
2. Cress RD, Chen YS, Morris CR, Petersen M, Leiserowitz GS. Characteristics of long-term survivors of epithelial ovarian cancer. *Obstet Gynecol*. 2015;126(3):491–7.
3. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005;4:Article17.
4. Langfelder P, Horvath S. Fast R functions for robust correlations and hierarchical clustering. *J Stat Softw*. 2012;46(11):i11.

5. Luo Y, Coskun V, Liang A, Yu J, Cheng L, Ge W, et al. Single-cell transcriptome analyses reveal signals to activate dormant neural stem cells. *Cell*. 2015;161(5):1175–86.
6. Cai Y, Mei J, Xiao Z, Xu B, Jiang X, Zhang Y, et al. Identification of five hub genes as monitoring biomarkers for breast cancer metastasis in silico. *Hereditas*. 2019;156:20.
7. Chen X, Hu L, Wang Y, Sun W, Yang C. Single cell gene co-expression network reveals FECH/CROT signature as a prognostic marker. *Cells*. 2019;8(7):698.
8. Chen X, Yang C, Xie S, Cheung E. Long non-coding RNA GASS and ZFAS1 are prognostic markers involved in translation targeted by miR-940 in prostate cancer. *Oncotarget*. 2018;9(1):1048–62.
9. Pils D, Hager G, Tong D, Aust S, Heinze G, Kohl M, et al. Validating the impact of a molecular subtype in ovarian cancer on outcomes: a study of the OVCAD Consortium. *Cancer Sci*. 2012;103(7):1334–41.
10. Tothill RW, Tinker AV, George J, Brown R, Fox SB, Lade S, et al. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin Cancer Res*. 2008;14(16):5198–208.
11. Lanczky A, Nagy A, Bottai G, Munkacsy G, Szabo A, Santarpia L, et al. miRpower: a web-tool to validate survival-associated miRNAs utilizing expression data from 2178 breast cancer patients. *Breast Cancer Res Treat*. 2016;160(3):439–46.
12. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform*. 2008;9:559.
13. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res*. 2010;38(Web Server issue):W214–20.
14. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol*. 2003;4(5):P3.
15. Yalcin D, Hakguder ZM, Otu HH. Bioinformatics approaches to single-cell analysis in developmental biology. *Mol Hum Reprod*. 2016;22(3):182–92.
16. Giancarlo R, Scaturro D, Utro F. Textual data compression in computational biology: a synopsis. *Bioinformatics*. 2009;25(13):1575–86.
17. Angermueller C, Parnamaa T, Parts L, Stegle O. Deep learning for computational biology. *Mol Syst Biol*. 2016;12(7):878.
18. Azuaje F. Computational models for predicting drug responses in cancer research. *Brief Bioinform*. 2017;18(5):820–9.
19. Keene DR, Engvall E, Glanville RW. Ultrastructure of type VI collagen in human skin and cartilage suggests an anchoring function for this filamentous network. *J Cell Biol*. 1988;107(5):1995–2006.
20. Bonaldo P, Braghetta P, Zanetti M, Piccolo S, Volpin D, Bressan GM. Collagen VI deficiency induces early onset myopathy in the mouse: an animal model for Bethlem myopathy. *Hum Mol Genet*. 1998;7(13):2135–40.
21. Gadd S, Huff V, Walz AL, Ooms A, Armstrong AE, Gerhard DS, et al. A Children's Oncology Group and TARGET initiative exploring the genetic landscape of Wilms tumor. *Nat Genet*. 2017;49(10):1487–94.
22. Sherman-Baust CA, Weeraratna AT, Rangel LB, Pizer ES, Cho KR, Schwartz DR, et al. Remodeling of the extracellular matrix through overexpression of collagen VI contributes to cisplatin resistance in ovarian cancer cells. *Cancer Cell*. 2003;3(4):377–86.
23. Yu J, Wu WK, Li X, He J, Li XX, Ng SS, et al. Novel recurrently mutated genes and a prognostic mutation signature in colorectal cancer. *Gut*. 2015;64(4):636–45.
24. Varma RR, Hector SM, Clark K, Greco WR, Hawthorn L, Pendyala L. Gene expression profiling of a clonal isolate of oxaliplatin-resistant ovarian carcinoma cell line A2780/C10. *Oncol Rep*. 2005;14(4):925–32.
25. Oyewumi L, Kaplan F, Gagnon S, Sweezey NB. Antisense oligodeoxynucleotides decrease LGL1 mRNA and protein levels and inhibit branching morphogenesis in fetal rat lung. *Am J Respir Cell Mol Biol*. 2003;28(2):232–40.
26. Oyewumi L, Kaplan F, Sweezey NB. Lgl1, a mesenchymal modulator of early lung branching morphogenesis, is a secreted glycoprotein imported by late gestation lung epithelial cells. *Biochem J*. 2003;376(Pt 1):61–9.
27. Kaplan F, Ledoux P, Kassamali FQ, Gagnon S, Post M, Koehler D, et al. A novel developmentally regulated gene in lung mesenchyme: homology to a tumor-derived trypsin inhibitor. *Am J Physiol*. 1999;276(6):L1027–36.
28. Gudbjartsson DF, Walters GB, Thorleifsson G, Stefansson H, Halldorsson BV, Zusmanovich P, et al. Many sequence variants affecting diversity of adult human height. *Nat Genet*. 2008;40(5):609–15.
29. Liepinsh E, Trexler M, Kaikkonen A, Weigelt J, Banyai L, Patthy L, et al. NMR structure of the LCCL domain and implications for DFNA9 deafness disorder. *EMBO J*. 2001;20(19):5347–53.
30. Nagai H, Sugito N, Matsubara H, Tatematsu Y, Hida T, Sekido Y, et al. CLCP1 interacts with semaphorin 4B and regulates motility of lung cancer cells. *Oncogene*. 2007;26(27):4025–31.
31. Trexler M, Banyai L, Patthy L. The LCCL module. *Eur J Biochem*. 2000;267(18):5751–7.
32. Zhang H, Kho AT, Wu Q, Halayko AJ, Limbert Rempel K, Chase RP, et al. CRISPLD2 (LGL1) inhibits proinflammatory mediators in human fetal, adult, and COPD lung fibroblasts and epithelial cells. *Physiol Rep*. 2016;4(17):e12942.
33. Gibbs GM, Roelants K, O'Bryan MK. The CAP superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins—roles in reproduction, cancer, and immune defense. *Endocr Rev*. 2008;29(7):865–97.
34. Duerrschmid C, He Y, Wang C, Li C, Bournat JC, Romere C, et al. Asprosin is a centrally acting orexigenic hormone. *Nat Med*. 2017;23(12):1444–53.
35. Granata A, Serrano F, Bernard WG, McNamara M, Low L, Sastry P, et al. An iPSC-derived vascular model of Marfan syndrome identifies key mediators of smooth muscle cell death. *Nat Genet*. 2017;49(1):97–109.
36. Romere C, Duerrschmid C, Bournat J, Constable P, Jain M, Xia F, et al. Asprosin, a fasting-induced glucogenic protein hormone. *Cell*. 2016;165(3):566–79.
37. Hellebrekers DM, Melotte V, Vire E, Langenkamp E, Molema G, Fuks F, et al. Identification of epigenetically silenced genes in tumor endothelial cells. *Cancer Res*. 2007;67(9):4138–48.
38. Wang Z, Liu Y, Lu L, Yang L, Yin S, Wang Y, et al. Fibrillin-1, induced by Aurora-A but inhibited by BRCA2, promotes ovarian cancer metastasis. *Oncotarget*. 2015;6(9):6670–83.
39. Becker J, Semler O, Gilissen C, Li Y, Bolz HJ, Giunta C, et al. Exome sequencing identifies truncating mutations in human SERPINF1 in autosomal-recessive osteogenesis imperfecta. *Am J Hum Genet*. 2011;88(3):362–71.
40. Ziff JL, Crompton M, Powell HR, Lavy JA, Aldren CP, Steel KP, et al. Mutations and altered expression of SERPINF1 in patients with familial otosclerosis. *Hum Mol Genet*. 2016;25(12):2393–403.
41. Cheung CYY, Lee CH, Tang CS, Xu A, Au KW, Fong CHY, et al. Genetic regulation of pigment epithelium-derived factor (PEDF): an exome-chip association analysis in chinese subjects with type 2 diabetes. *Diabetes*. 2019;68(1):198–206.
42. Crowe S, Wu LE, Economou C, Turpin SM, Matzaris M, Hoehn KL, et al. Pigment epithelium-derived factor contributes to insulin resistance in obesity. *Cell Metab*. 2009;10(1):40–7.
43. Xu B, Li J, Liu X, Li C, Chang X. TXNDC5 is a cervical tumor susceptibility gene that stimulates cell migration, vasculogenic mimicry and angiogenesis by down-regulating SERPINF1 and TRAF1 expression. *Oncotarget*. 2017;8(53):91009–24.
44. Nwani NG, Deguiz ML, Jimenez B, Vinokour E, Dubrovskiy O, Ugolkov A, et al. Melanoma cells block PEDF production in fibroblasts to induce the tumor-promoting phenotype of cancer-associated fibroblasts. *Cancer Res*. 2016;76(8):2265–76.
45. Walker C, Mojares E, Del Rio Hernandez A. Role of extracellular matrix in development and cancer progression. *Int J Mol Sci*. 2018;19(10):3028.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.