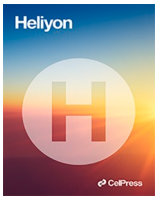
Heliyon 10 (2024) e32909

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

ScHGSC-IGDC: Identifying genes with differential correlations of high-grade serous ovarian cancer based on single-cell RNA sequencing analysis

Yuanqi Li [a,b,c,1], Qi Wang [a,b,c,1], Xiao Zheng [a,b,c], Bin Xu [a,b,c], Wenwei Hu [b,c,d], Jinping Zhang [e], Xiangyin Kong [f], Yi Zhou [a,b,c,**], Tao Huang [g,***], You Zhou [a,b,c,*]

[a] *Tumor Biological Diagnosis and Treatment Center, The Third Affiliated Hospital of Soochow University, Changzhou 213003, China*
[b] *Jiangsu Engineering Research Center for Tumor Immunotherapy, Changzhou, 213003, China*
[c] *Institute of Cell Therapy, Soochow University, Changzhou, 213003, China*
[d] *Department of Oncology, The Third Affiliated Hospital of Soochow University, Changzhou, 213003, China*
[e] *Institutes of Biology and Medical Sciences, Soochow University, Suzhou, 215123, China*
[f] *CAS Key Laboratory of Tissue Microenvironment and Tumor, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, Shanghai, 200031, China*
[g] *Bio-Med Big Data Center, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, Shanghai, 200031, China*

ARTICLE INFO

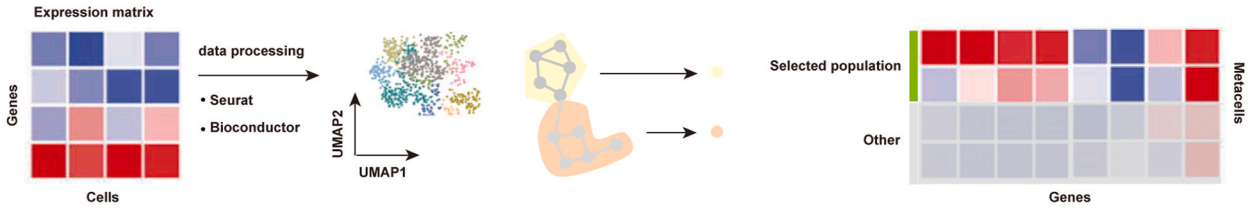
Keywords:
Single-cell sequencing
WGCNA
Differential correlation
HGSC

ABSTRACT

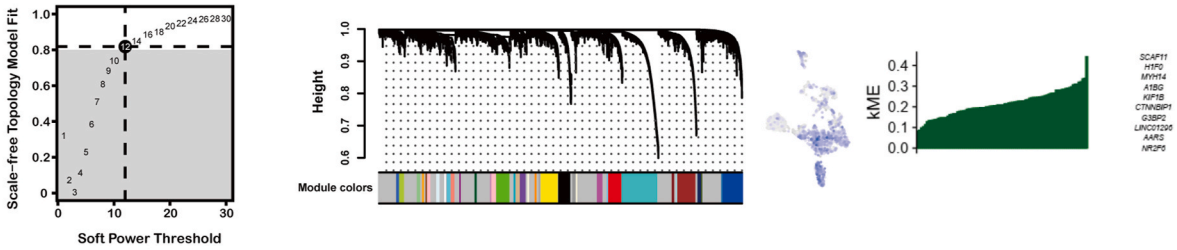
Due to the high heterogeneity of ovarian cancer (OC), it occupies the main cause of cancer-related death among women. As the most aggressive and frequent subtype of OC, high-grade serous cancer (HGSC) represents around 70 % of all patients. With the booming progress of single-cell RNA sequencing (scRNA-seq), unique and subtle changes among different cell states have been identified including novel risk genes and pathways. Here, our present study aims to identify differentially correlated core genes between normal and tumor status through HGSC scRNA-seq data analysis. R package high-dimension Weighted Gene Co-expression Network Analysis (hdWGCNA) was implemented for building gene interaction networks based on HGSC scRNA-seq data. DiffCorr was integrated for identifying differentially correlated genes between tumor and their adjacent normal counterparts. Software Cytoscape was implemented for constructing and visualizing biological networks. Real-time qPCR (RT-qPCR) was utilized to confirm expression pattern of new genes. We introduced ScHGSC-IGDC (Identifying Genes with Differential Correlations of HGSC based on scRNA-seq analysis), an *in silico* framework for identifying core genes in the development of HGSC. We detected thirty-four modules in the network. Scores of new genes with opposite correlations with others such as NDUFS5, TMSB4X, SERPINE2 and ITPR2 were identified. Further survival and literature validation emphasized their great values in the HGSC management. Meanwhile, RT-qPCR verified expression pattern of NDUFS5, TMSB4X, SERPINE2 and ITPR2 in human OC cell lines and tissues. Our research offered novel perspectives on the gene modulatory mechanisms from single cell resolution, guiding network based algorithms in cancer etiology field.

* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail addresses: czzhouyi2600@hotmail.com (Y. Zhou), tohuangtao@126.com (T. Huang), zhouyounew@163.com (Y. Zhou).
[1] These authors have contributed equally to this work.

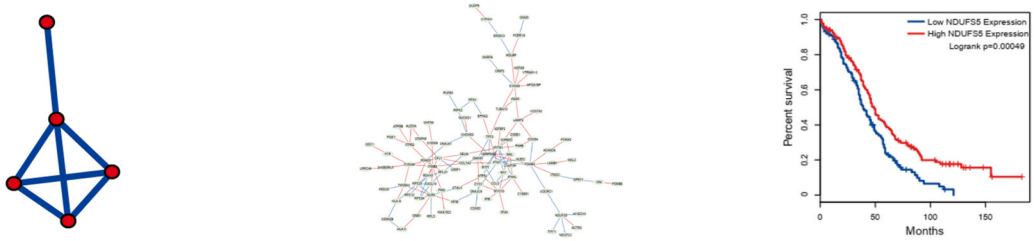
https://doi.org/10.1016/j.heliyon.2024.e32909
Received 27 November 2023; Received in revised form 29 May 2024; Accepted 11 June 2024
Available online 13 June 2024



1 Primary data processing	2 Metacells	3 Setup expression matrix
• Normalization • Dimensionality reduction • Clustering and annotation	• Merge similar cells via KNN bagging • Decrease data sparsity	• Get expression matrix • Select genes
Seurat, SetupForWGCNA	*MetacellsByGroups*	*SetDatExpr*



4 Soft power threshold	5 Construct co-expression network	6 Compute module eigengenes and connectivity
• Determine best value of β based on scale-free topology fit	• Compute topological overlap matrix (TOM) • Hierarchical clustering	• Get Module Eigengenes (MEs) • Compute eigengene-based connectivity (kME)
TestSoftPowers	*GetTOM, ConstructNetwork*	*ModuleEigengenes, ModuleConnectivity*



7 Module networks	8 Pair-wise differential correlation	9 Functional analysis of core genes
• Extract conditional modules • Draw module members for each condition	• Identify significant differences between two correlations based on its stringency test and conservative estimates	• Biological analysis of core genes that build most links with others • The Kaplan-Meier plot of core gens
Get.eigen.module, Get.eigen.molecul.graph	*Comp.2.cc.fdr*	*Survival*

Fig. 1. The schematic diagram of ScHGSC-IGDC.

1. Introduction

Due to the high heterogeneity of ovarian cancer (OC), it stands as the primary contributor to cancer-related mortality among female reproductive malignancies [1]. As the most aggressive and frequent subtype of OC, high-grade serous cancer (HGSC) accounts for about two thirds of all cases [2]. Usually diagnosed at an advanced stage, HGSC presents with massive ascites and metastases. Although great advances have been made in surgery and chemotherapy, HGSC still portends poor survival and prognosis due to relapse and treatment resistance [3]. Therefore, elucidating the pathologic mechanisms and identifying new risk genes of HGSC are of great value.

The principal challenging to study HGSC is its intra-tumor heterogeneity (ITH) distinguishing from other OC types, which is characterized by ubiquitous TP53 driver mutations, homologous recombination DNA repair deficiency and extensive somatic copy number alterations [2]. To explore the heterogeneity of HGSC and the dynamic changes between normal and tumor tissues, previous

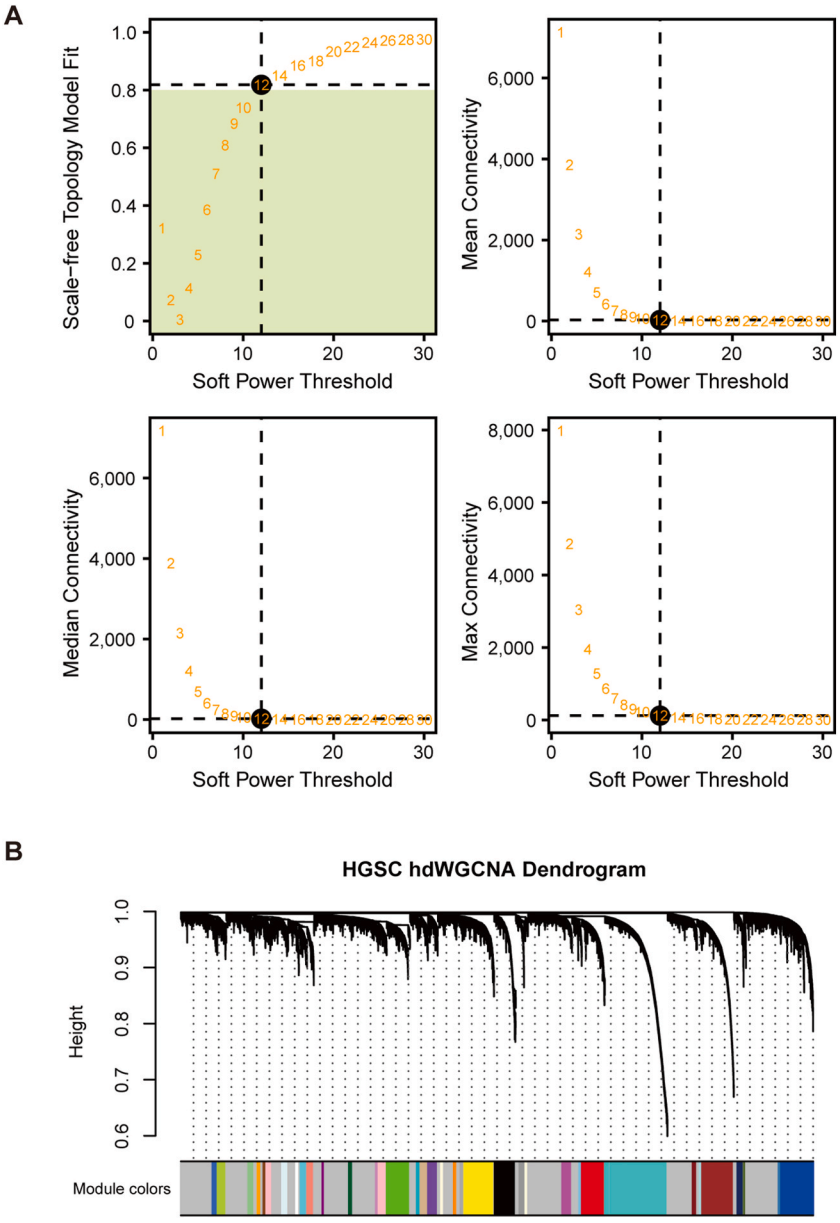


Fig. 2. Cluster dendrogram of HGSC and normal cells. (A) Relationship between soft power threshold and network properties. Upper panel: Relationship between soft power threshold and scale-free topology model fit (left). Relationship between soft power threshold and mean connectivity (right). Lower panel: Relationship between soft power threshold and median (left) and max (right) connectivity, respectively. (B) All the genes were clustered in thirty-four modules. One color represented one module. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

studies mostly based on bulk RNA sequencing have incorporated gene expression as well as gene correlation data followed by construction of gene interaction networks [4,5]. However, these studies only permit resolution of tissue-level and are unable to decipher ITH. Furthermore, our observations may be mostly redundant gene correlations as they existed in both normal and tumor status, increasing difficulties in identifying true causative genes and pathways [6,7].

Arising as a powerful tool for dissecting the ITH, single-cell RNA sequencing (scRNA-seq) offers quantification of transcripts as well as diversified information of individual cells. Accumulating evidence has used scRNA-seq to explore the unique and subtle changes among different cell states [8–11]. For example, the changes between normal and tumor status, the changes among primary tumors and metastases. Meanwhile, subtle changes among different cell states can benefit from identification of differential correlations between genes in network structure. Accumulating evidence has indicated the altered regulatory network was associated with multiple biological processes such as cellular differentiation and cancer initiation [12,13]. The important regulators with highly variable correlated expression patterns may control cell fate decisions like cancer cell formation, which were considered as empirical indicators of the upcoming transitions [14]. And a change in correlation from normal tissues to tumor tissues represented the alteration of regulatory relationship comprised of regulators and their target genes. As little is known about genes with differential correlations from scRNA-seq data, we introduced a unique *in silico* framework named ScHGSC-IGDC to distinguish core genes during the development of HGSC (Fig. 1). First, we extracted scRNA-seq data of HGSC from Gene Expression Omnibus (GSE189955). Then, high-dimension Weighted Gene Co-expression Network Analysis (hdWGCNA) as well as DiffCorr were implemented to identify differentially correlated genes. Lastly, scores of new genes such as NDUFS5, TMSB4X, SERPINE2 and ITPR2 were recognized, whose expression pattern was verified through real-time qPCR (RT-qPCR). Further survival and literature validation exhibited crucial impacts of these genes on HGSC progression.

2. Results

2.1. Co-expression network construction

On the single-cell dataset, dimension reduction clustering was performed (Supplementary Fig. S1). And 13,560 genes expressed in at least 5 % cells were selected for further analysis. The weighted gene co-expression network was built by 13,560 noncoding and coding genes using hdWGCNA method. We set soft-thresholding power twelve in order to match the scale-free topology of network

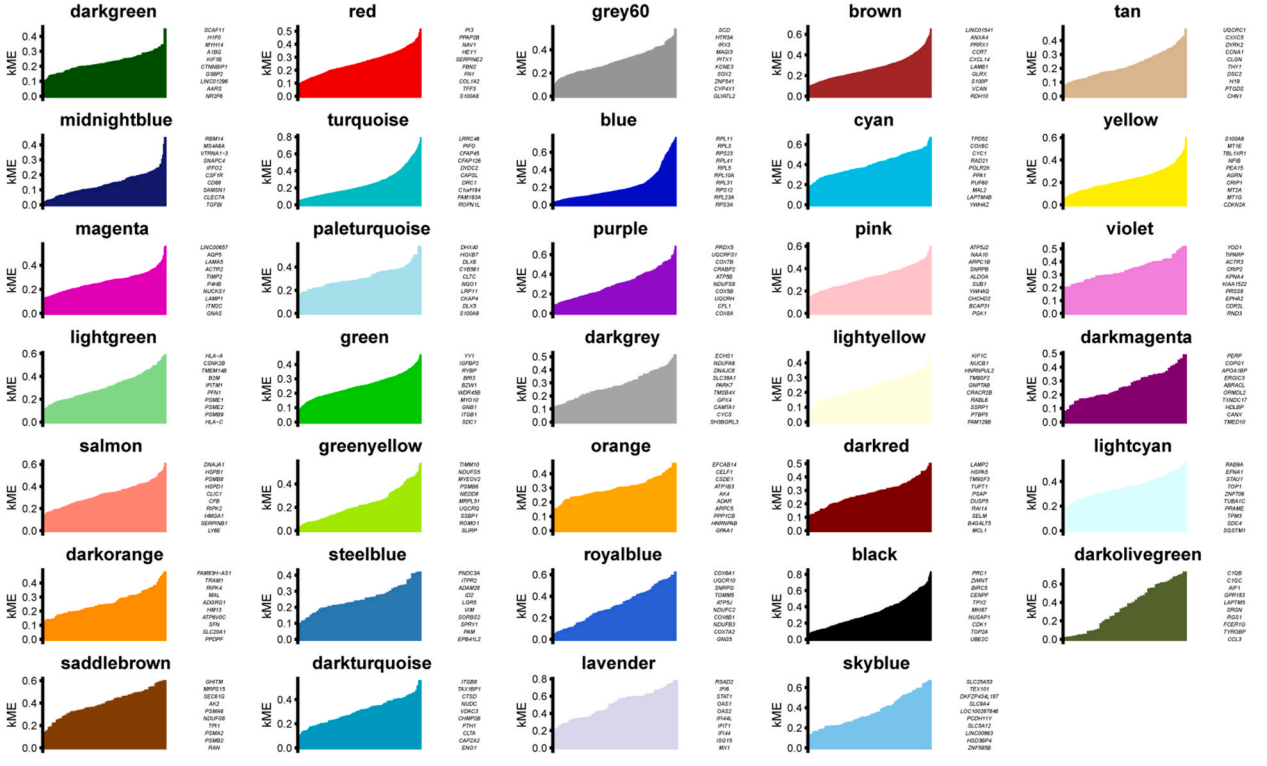
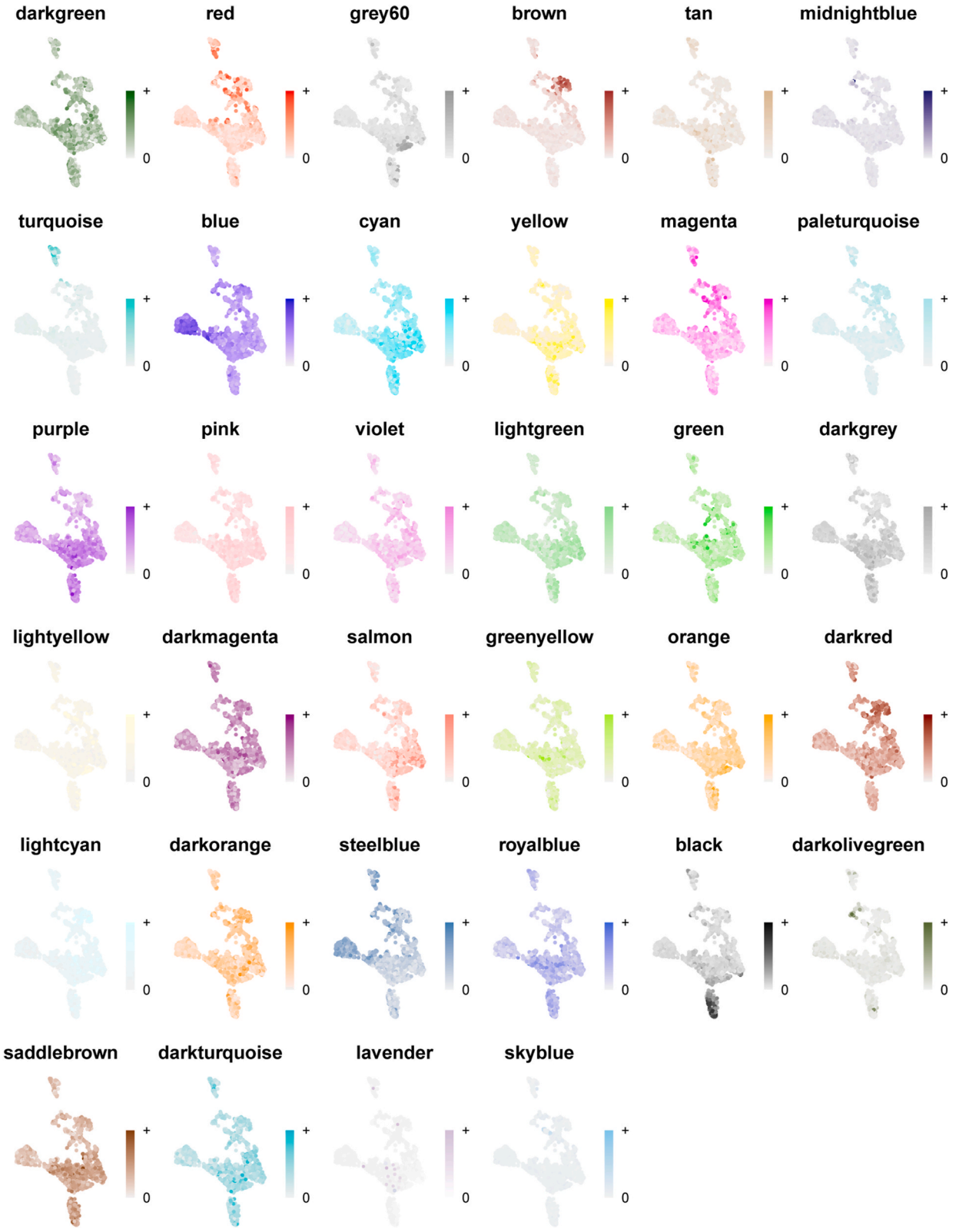


Fig. 3. Module genes ranked by eigengene-based connectivity. Displayed were ranked ten core genes of thirty-four modules ("darkgreen", "red", "grey60", "brown", "tan", "midnightblue", "turquoise", "blue", "cyan", "yellow", "magenta", "paleturquoise", "pruple", "pink", "violet", "lightgreen", "green", "darkgrey", "lightyellow", "darkmagenta", "salmon", "greenyellow", "orange", "darkred", "lightcyan", "darkorange", "steelblue", "royalblue", "black", "darkolivegreen", "saddlebrown", "darkturquoise", "lavender", "skyblue"). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)



(caption on next page)

Fig. 4. Core genes signature. The function *ModuleFeaturePlot* was used to construct FeaturePlots for each co-expression module colored by each module's uniquely assigned color. The features were core genes in each module and evaluated by gene expression levels. Displayed were core genes signature of thirty-four modules ("darkgreen", "red", "grey60", "brown", "tan", "midnightblue", "turquoise", "blue", "cyan", "yellow", "magenta", "paleturquoise", "pruple", "pink", "violet", "lightgreen", "green", "darkgrey", "lightyellow", "darkmagenta", "salmon", "greenyellow", "orange", "darkred", "lightcyan", "darkorange", "steelblue", "royalblue", "black", "darkolivegreen", "saddlebrown", "darkturquoise", "lavender", "skyblue"). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

(Fig. 2A). We detected thirty-four modules in the network displayed by cluster dendrogram (Fig. 2B). Supplementary Table S1 listed each member in each module. In addition to grey module comprised of numerous unclassified members, darkolivegreen module consisted of 50 genes was minimum, while turquoise module consisted of 1342 genes was maximum.

We next computed harmonized module eigengenes and module connectivity. Module eigengenes reflected the expression profiles of all the modules. We focused on those core genes with high connectivity in each module. As displayed in Fig. 3, ten core genes belonging to each module were ranked by eigengene-based connectivity. And their detailed information was listed in Supplementary Table S2. We constructed FeaturePlots for each co-expression module, colored according to their uniquely assigned colors (Supplementary Fig. S2). The features were core genes in each module and evaluated by module eigengene expression levels. Core gene signature score was also detected using *ModuleFeaturePlot* function of hdWGCNA package (Fig. 4). *ModuleCorrelogram* function was used for visualization of the correlation among the modules based on their core gene scores (Supplementary Fig. S3).

In order to validate the robustness of our identification method, we conducted Gene Ontology (GO) enrichment analysis using 340 core genes in all the thirty-four modules in our dataset as well as another two randomly selected subsets of cells. One included 1492 tumor cells and 400 normal cells and the other included 1119 tumor cells and 300 normal cells. We conducted hdWGCNA analysis and got core genes, respectively. The GO analysis showed that among the top ten enriched GO terms, seven of them were shared by the three datasets (Fig. 5A–C). These core genes were mainly enriched in oxidative phosphorylation, aerobic respiration and electron transport chain, endowing the significance of these biological activities to HGSC development.

We also quantified module-trait associations by selecting tumor cells. As a result, two modules steelblue and blue exhibited the most significant correlations with HGSC stage (Supplementary Figs. S4A–D). The corresponding correlation coefficients for the two modules, steelblue and blue, were 0.60 ($P = 2\times10^{-15}$) and −0.77 ($P = 3\times10^{-29}$). Obviously, Gene Significance (GS) and Module Membership (MM) analysis demonstrated genes strongly remarkably correlated to HGSC were likewise the most vital components of modules correlated to HGSC (Supplementary Fig. 4E). GO analysis of the above two modules showed that blue module genes were markedly accumulated in actin filament organization, positive regulation of proteolysis and nucleocytoplasmic transport, while no significant GO terms were found in genes of the steelblue module, endowing their potential association with HGSC stage (Supplementary Fig. 4F).

2.2. *Differential correlation identification*

Then, we chose the above 340 core genes in all the thirty-four modules in order to estimate differential correlations. We clustered these genes according to their expression patterns in normal or tumor and calculated differential correlations of genes. Afterthat, module network was constructed and visualized based on hierarchical cluster analysis (Fig. 6A and B). Finally, correlation analyses offered differential correlations between two genes among all the modules (Supplementary Table S3).

In order to provide specific biological insight for the module network, we investigated the enriched pathways of module network in tumor. The results showed that genes of module 1 were significantly accumulated in metabolic process like glycolysis, genes of module

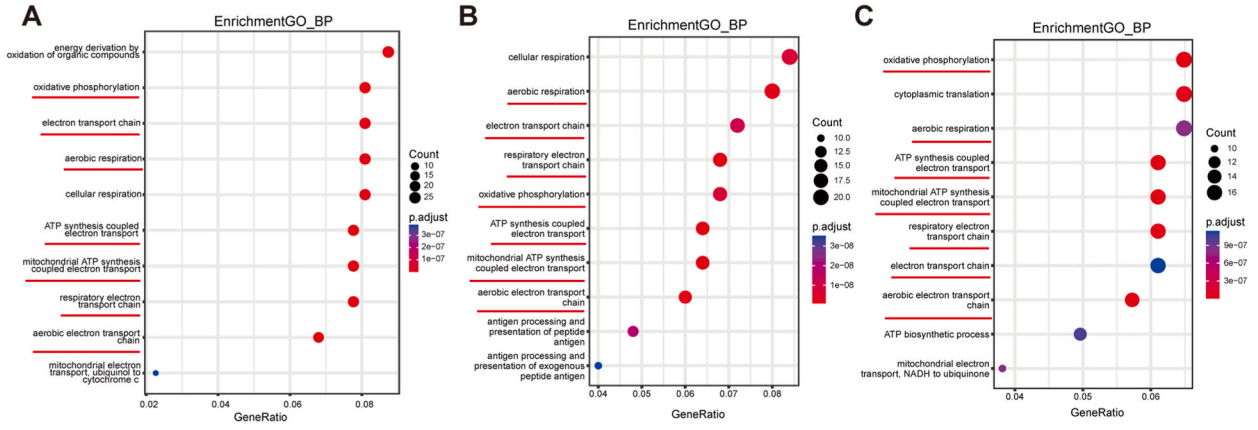


Fig. 5. Functional enrichment of core genes. GO analysis showed the top 10 enriched biological processes in our dataset (A) and two randomly selected subsets of cells (B and C). The shared processes were underlined in red. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

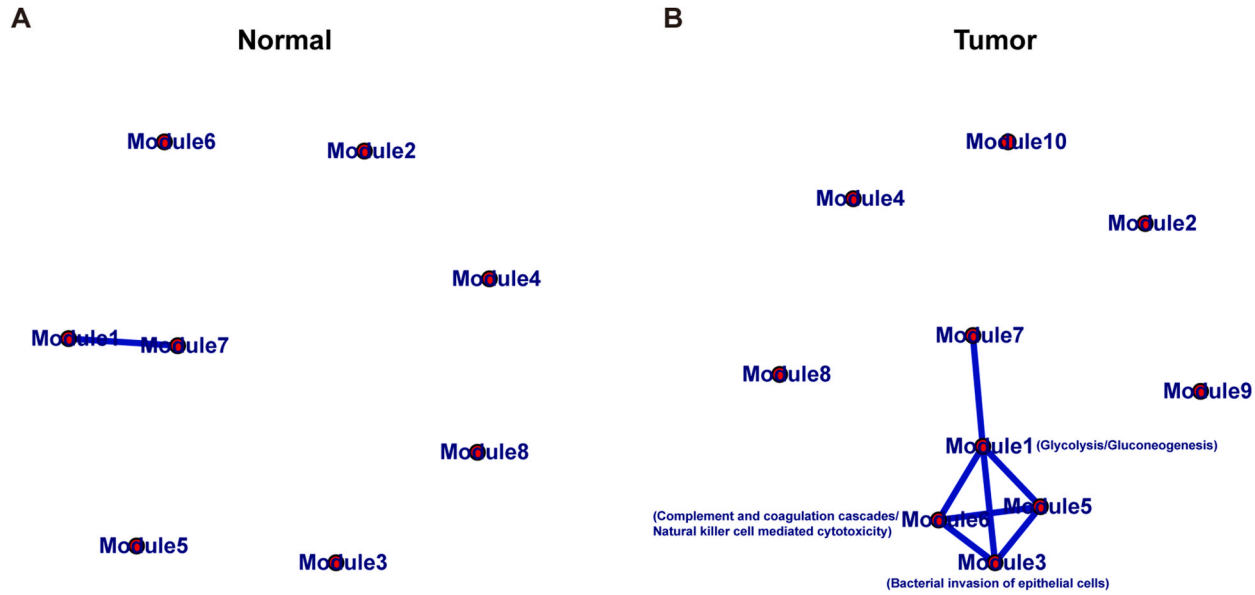


Fig. 6. Module networks. Representation of normal (A) and tumor (B) module networks from GSE189955. Each edge represented module correlation. Each node represented a module. The enriched pathways of module 1, 3 and 6 in tumor were labeled.

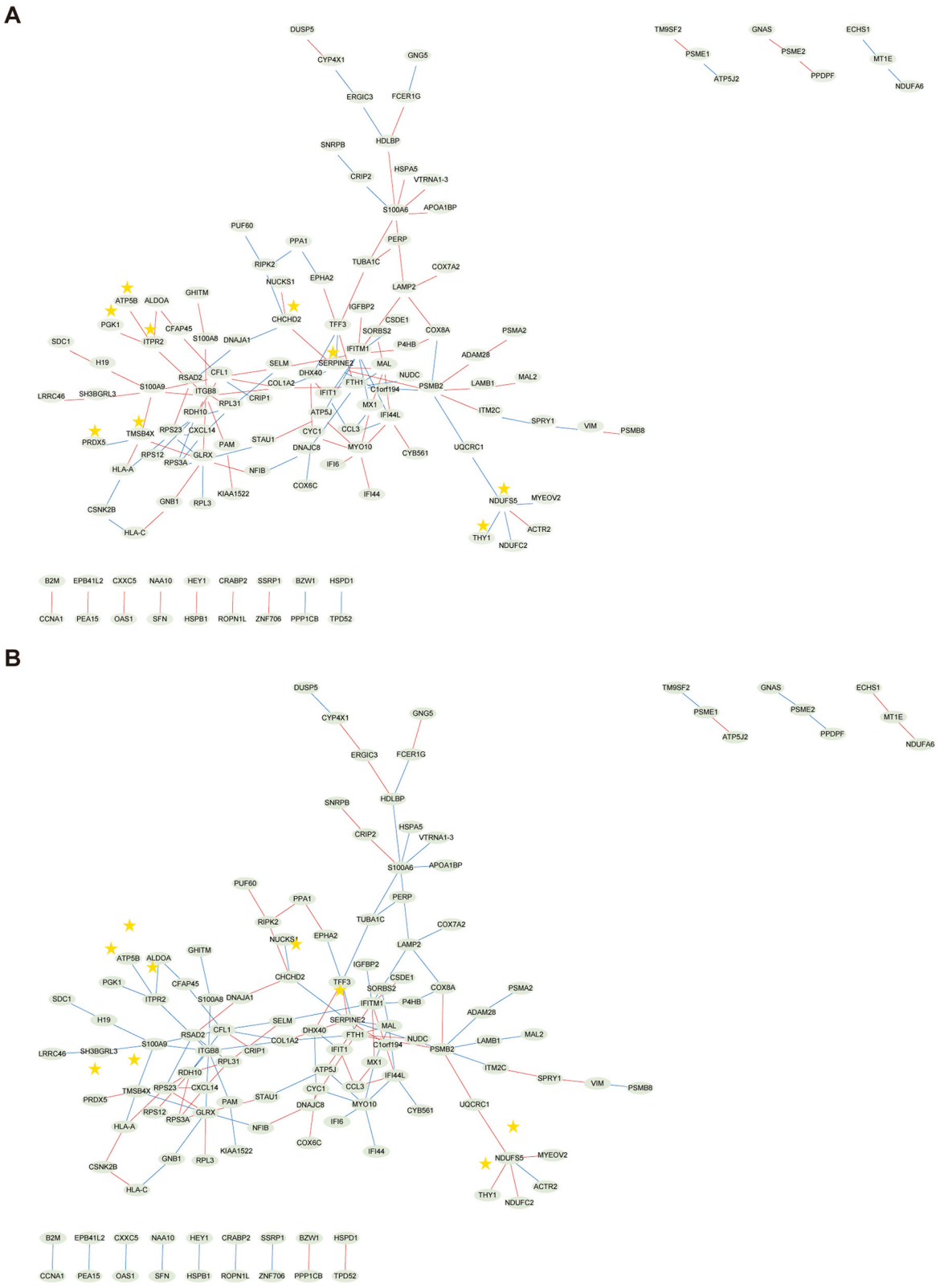
3 in bacterial invasion of epithelial cells, as well as genes of module 6 in immune responses such as natural killer cell mediated cytotoxicity and complement and coagulation cascades (Fig. 6B). Recent studies have indicated the role of bacterial and its metabolites in formation and metastasis of ovarian cancer [15,16]. Moreover, in response to bacterial infection, immune cells reacted rapidly with a series of inflammatory burst, which explained the relationship between module 3 and module 6 [17]. Meanwhile, during the inflammatory processes, immune cells underwent a switch to aerobic glycolysis for the production of Warburg effect, implying the relationship between module 1 and module 6 [18].

Also, DiffCorr algorithm discovered oppositely correlated genes. That means, two genes are negatively correlated in HGSC cells while positively correlated in normal cells. The transformed gene pairs were noteworthy, underlining key functions during HGSC progression. 143 oppositely correlated gene pairs were obtained in all (Supplementary Table S4). And their interacted networks were presented in Fig. 7A and B. Fifteen most differentially correlated gene pairs between HGSC and normal cells were shown in Table 1. We also performed functional enrichment of these oppositely correlated genes. The results indicated that they were almost enriched in oxidative phosphorylation which was hyperactive in ovarian cancer cells and strongly linked to the proliferation as well as survival of cancer stem cells [19]. And oxidative phosphorylation associated genes included COX6C, NDUFC2, CYC1 and UQCRC1. Thus, oxidative phosphorylation could play key roles in carcinogenesis and be rendered as a therapeutic target for HGSC course.

2.3. Functional analysis of differentially correlated core genes

Eventually, we centered on four novel genes which have not been previously reported in HGSC and acted as dominant regulators for building most links with others. As one of the members of NADH ubiquinone oxidoreductase subunit family, NDUFS5 mainly participates in maintaining cell proliferation and differentiation and regulates signal transducers and activators of transcription 3 (STAT3) [20,21]. Prior research has recognized its function as a tumor suppressor in glioma, consistent with our survival analysis that lower expression levels of NDUFS5 predicting poor outcome of HGSC patients (Fig. 8A, Supplementary Fig. S5A) and lower expression levels observed in HGSC patients and cell lines compared with normal (Fig. 9A, Supplementary Fig. S6A). Previous study has implicated the prognostic value of STAT3 in HGSC patients under platinum-based chemotherapy [22]. Meanwhile, dysregulation of STAT3 has been indicated to be correlated to proliferation and metastasis of OC [23]. Given the great importance of STAT3 in the progression of HGSC, it can be inferred that NDUFS5 modulates biological processes of HGSC cells by interacting with STAT3 signalling pathway and may be effective therapeutic target of HGSC prevention or treatment. Furthermore, the altered association between NDUFS5 and THY1 should not be ignored. THY1 is a glycosylphosphatidylinositol anchored protein and recently has been authenticated as an OC stem cell marker [24]. Increased expression of THY1 has been reported to predict poor prognosis and be related to self-renewal in epithelial ovarian cancer (EOC), which conferred THY1 as an oncogene [25]. Thus, NDUFS5 and THY1 were endowed with opposite functions and they were negatively correlated in normal cells. However, this correlation changed to be positive in tumor cells, indicating substantial changes of regulatory networks upon cancer occurrence.

As a widely distributed and multifunctional peptide, thymosin-β4 (Tβ4), the gene product of TMSB4X, plays pivotal roles in various physiological and pathological processes, especially angiogenesis and inflammation [26]. Prior studies have indicated the differential expression pattern of Tβ4 in a broad variety of cancers [27]. A recent scRNA-seq analysis has demonstrated the elevated levels of TMSB4X in papillary thyroid cancer cells compared with non-malignant cells [28]. Moreover, the aberrant expression of Tβ4 has been



(caption on next page)

Fig. 7. Differential co-expressed gene network of normal (A) and tumor (B) modules from GSE189955. One node represented one gene. One edge represented correlation between two genes, among which blue meant negative correlation and red positive correlation. The stronger correlation coefficient was represented by the thicker edge. NDUFS5, THY1, TMSB4X, PRDX5, SERPINE2, CHCHD2, ITPR2, ATP5B, and PGK1 gene pairs were marked by asterisks. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Fifteen most differentially correlated gene pairs transformed oppositely from modules between HGSC and normal cells.

Molecule X	Molecule Y	r1 (Normal)	r2 (Tumor)	lfdr[a] (difference)
SERPINE2	TFF3	−0.747055516	0.920468266	0
IFI44L	IFITM1	−0.7139838	0.713499528	0
IFITM1	MX1	−0.572945762	0.767236876	0
CSNK2B	HLA-C	−0.563572757	0.795383642	0
COL1A2	SERPINE2	−0.539857251	0.784304513	0
ALDOA	ITPR2	0.525296438	−0.77355076	0
GLRX	TMSB4X	0.553493633	−0.935714716	0
FTH1	TFF3	0.62875303	−0.731098244	0
CFL1	SELM	0.660930843	−0.670434993	0
GLRX	RPS23	−0.672680312	0.648752588	1.68E-13
NDUFS5	UQCRC1	−0.710706839	0.587593374	4.66E-13
COX6C	CYC1	−0.511134907	0.749010427	1.44E-12
GHITM	S100A8	0.530322574	−0.725233437	3.62E-12
ITPR2	PGK1	0.644676077	−0.630798971	3.62E-12
FTH1	MAL	−0.542917233	0.704141786	9.41E-12

[a] *Local FDR.*

correlated with advanced disease in thyroid cancer [29]. Another research on murine mammary gland cells has underlined the importance of TGFβ/Tβ4 signaling pathway in epithelial-mesenchymal transition, metastasis and cancer progression [30]. However, our survival analysis showed low expression of TMSB4X implicated high risk (Fig. 8B, Supplementary Fig. S5B), implying its distinct molecular mechanisms in HGSC. And downregulation of TMSB4X in HGSC patients and cell line was also observed in our experiments (Fig. 9B, Supplementary Fig. S6B). Notably, PRDX5, one isoform of antioxidant enzymes Peroxiredoxins, was negatively correlated with TMSB4X in normal cells whose association altered in tumor cells. PRDX5 has been observed to be frequently involved in regulating reductant-oxidant-sensitive cellular processes [31]. Prior studies have shown elevated expression of PRDX5 indicated poorer overall survival, conferring PRDX5 a negative survival predictor in OC [32]. In other words, our study expanded the function of TMSB4X as well as its interacted networks in cancer biology which needed further investigations.

As an extracellular serine proteinase inhibitor, Serpin peptidase inhibitor, clade E, member 2 (SERPINE2) participates in the progression of several malignancies [33]. Recent research has underscored the oncogenic significance of SERPINE2 across a wide range of cancer types [33,34]. For instance, SERPINE2 can act as a poor biomarker and prompted cancer metastasis. SERPINE2 can also modulate DNA damage response and radio-resistance. Conversely, tumor suppressive effects of SERPINE2 have been found in glioma and prostate cancer [35,36], suggesting its dual function in carcinogenesis. Here, the Kaplan-Meier analysis demonstrated preferable survival in the lower expression group (Fig. 8C, Supplementary Fig. S5C), emphasizing cancer accelerating influences of SERPINE2 on HGSC. Moreover, RT-qPCR confirmed significant elevated levels of SERPINE2 in HGSC cell lines and patients (Fig. 9C). Regarding the interaction networks of SERPINE2, it was positively correlated with mitochondrial nuclear retrograde regulator 1 (MNRR1, also named CHCHD2) in normal cells which reversed in tumor cells. CHCHD2 has been considered as a driver of OC progression whose deletion led to tumor growth delay and improved survival in OC mouse models [37]. Therefore, we elucidated the central position of SERPINE2 in the regulatory network of HGSC, clarifying its molecular mechanisms partly.

The endoplasmic reticulum (ER) inositol 1,4,5-trisphosphate receptor, type 2 (ITPR2) contributes to calcium ion transmembrane transportation as well as exerts vital roles in regulating cell cycle and proliferation [38]. Multiple evidence has indicated that ITPR2 controlled cellular senescence and aging through collaboration with mitochondria and ER [38–40]. Given the importance of ITPR2 in regulating different pathophysiological processes, its aberration is associated with the cancer occurrence. ITPR2 has been recognized as a novel biomarker as well as clinical target for leukemia [41]. Genome-wide association studies (GWAS) has shown great associations of germline genetic variations in ITPR2 with clear cell renal cell cancer among Chinese people [42]. Nevertheless, scant literature is found with regard to the function of ITPR2 on OC or HGSC. Compared with high group, the low ITPR2 expression group had significantly longer survival time (Fig. 8D, Supplementary Fig. S5D), consistent with the carcinogenic property of ITPR2 in previously reported cancers. And ITPR2 showed significant higher expression levels in HGSC compared with normal cell lines and tissues (Fig. 9D, Supplementary Fig. S6C). Meanwhile, several genes interacting with ITPR2 were also identified from our regulatory network including phosphoglycerate kinase 1 (PGK1) and mitochondrial β-F1-ATPase (ATP5B). PGK1 is a key factor in coordinating cellular energy production and upregulated in multiple cancers [43]. Recent research has also recognized PKG1 as a target for anti-glycolytic therapy of OC [44,45]. Similarly, as a metabolic marker, high expression of ATP5B has been revealed to predict poor overall survival of HGSC [46]. All as oncogenes, ITPR2 was positively correlated with PGK1 and ATP5B in normal cells whose aberrant association occurred in tumor cells, contributing to understanding the etiology of HGSC.

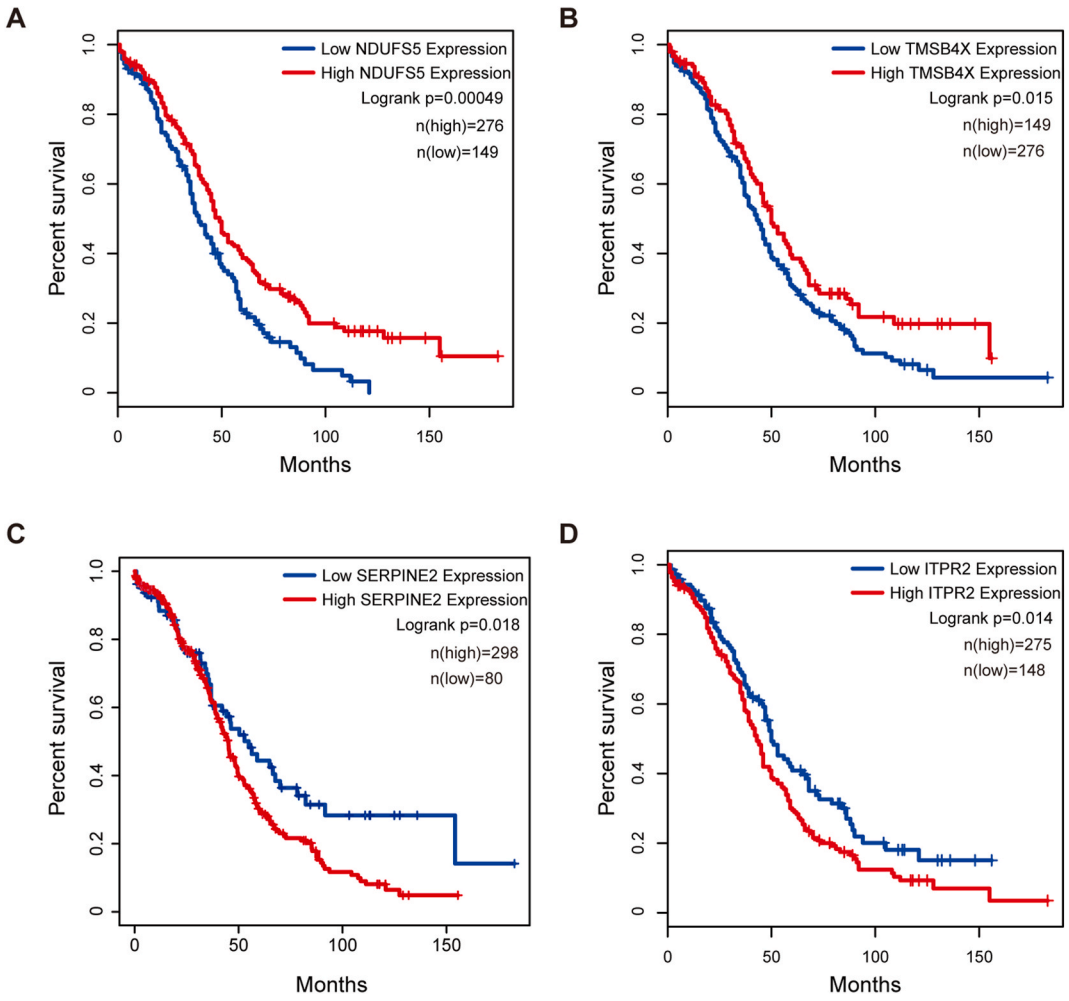


Fig. 8. The Kaplan-Meier plot of NDUFS5, TMSB4X, SERPINE2 and ITPR2 from TCGA datasets. The low expression of NDUFS5 (A) and TMSB4X (B) indicated high risk, while the high expression of SERPINE2 (C) and ITPR2 (D) indicated high risk. Different group cutoffs were used to obtain the optimal statistical result.

3. Discussion

Most HGSC patients suffered from cancer relapse and treatment resistance due to ITH, which could not be addressed by bulk RNA sequencing and gene correlation analysis. To clarify the heterogeneity of HGSC and altered regulatory network during carcinogenesis, we used scRNA-seq data and investigated differentially correlated core genes between normal and tumor status for the first time. Scores of novel genes including NDUFS5, TMSB4X, SERPINE2 and ITPR2 were identified. And their differentially correlated genes THY1, PRDX5, CHCHD2 and PGK1 were firstly proposed for their causative roles during HGSC progression. Further RT-qPCR confirmed the expression patterns of NDUFS5, TMSB4X, SERPINE2 and ITPR2 and more in-depth experimental research was required to validate their biological function in the pathogenesis of HGSC. Additionally, our computational strategy can be expanded to other cell types with more subtle differences such as different developmental stages as well as different cancer types.

In addition to alterations among genome and transcriptome, tumor development also involved vast epigenetic changes which needed to be comprehensively analyzed, so our study could benefit from integrating multi-omics sequencing data including epigenetic modifications and metabolites data in the future.

4. Materials and methods

4.1. HGSC scRNA-seq dataset

The scRNA-seq dataset of HGSC (GSE189955) was downloaded from Gene Expression Omnibus including 10 HGSC samples, 4 non-serous OC samples and 6 normal fallopian tubes [47]. Following standardization, 1940 primary tumor cells (refer to as tumor) and 520

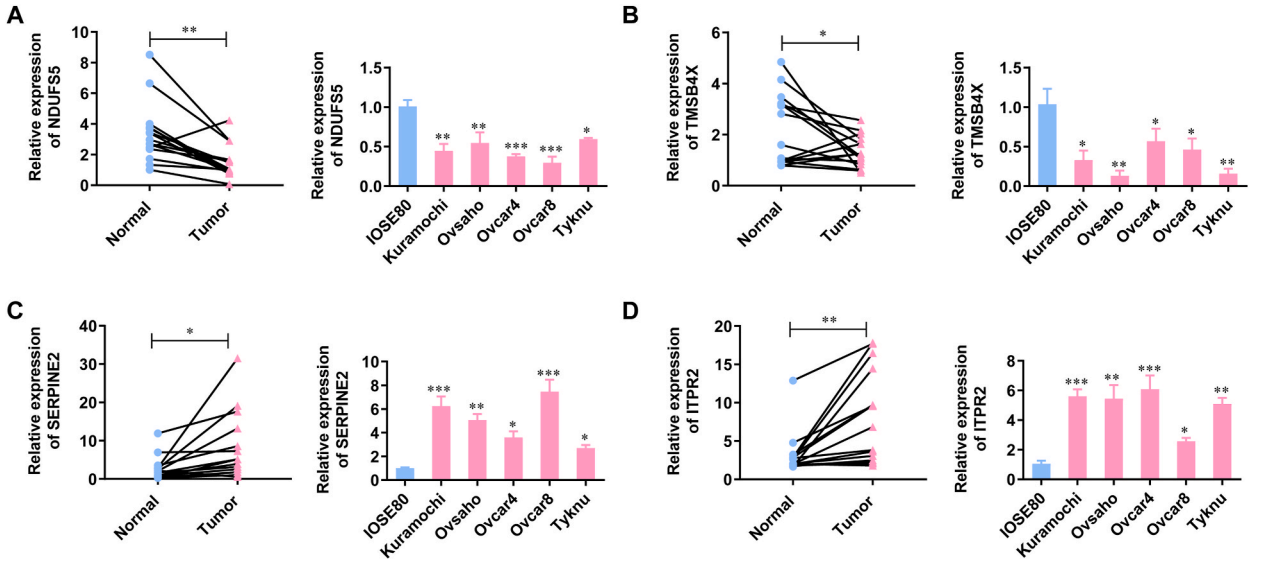


Fig. 9. The expression levels of NDUFS5 (A), TMSB4X (B), SERPINE2 (C) and ITPR2 (D) in HGSC cell lines and tissues tested by RT-qPCR. OC cell lines: Kuramochi, Ovsaho, Ovcar4, Ovcar8 and Tyknu. Human ovarian epithelial cell line: IOSE80. Data were means ± S.E.M. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. Experiments were duplicated at least three times.

epithelial cells from fallopian tubes (refer to as normal) were selected for further gene expression analysis.

4.2. Analysis of single-cell co-expression network

R package hdWGCNA was performed to screen core gene and detect co-expression of gene pair, which is optimized from WGCNA and applied for scRNA-seq data. First, we need to set up Seurat object from single-cell RNA count matrix including performing *RunPCA* for dimensionality reduction and *RunHarmony* for batch correction with default parameters. Genes expressed in at least 5 % of all cells were selected (fraction = 0.05). Next, we constructed metacells based on *k*-nearest neighbor (KNN) algorithm to reduce expression matrix sparsity with the function *MetacellsByGroups* (k = 25). Thirdly, to construct gene co-expression network, we run WGCNA after selecting the soft-thresholding power (soft_power = 12). Briefly, the network is undirected and weighted in which nodes are genes and edges are correlation coefficients between genes. Through elevating the absolute correlation values to a soft-thresholding power, the network construction underlines strong correlations at the expense of weak correlations. The selection criterion of the soft-thresholding power is making the connections conform to the distribution of scale-free network. Sequentially, we performed the following functions with default parameters: ConstructNetwork, ModuleEigengenes, and ModuleConnectivity. The overall premise and procedure of WGCNA was detailedly described in Ref. [48]. Lastly, genes and network modules were identified.

4.3. GO and KEGG analysis

R package clusterProfiler was utilized for functional analysis of core genes among all the modules. And a hypergeometric distribution test was employed to examine enriched terms. False discovery rate (FDR) method was employed to adjust *P* values with a cutoff less than 0.05 [49].

4.4. Evaluation of differential correlations

We implemented R package DiffCorr for identifying and visualizing differential correlations and details were described in Ref. [50]. This package was mainly comprised of three steps: calculation of differential correlations, identification of eigen-molecules and scaling and clustering. Firstly, we transformed correlation coefficients for each condition, r_A and r_B into Z_A and Z_B, respectively. And the Fisher's transformation of r_A/r_B was: $Z = \frac{1}{2}\log\frac{1+r}{1-r}$ (1). Differences of two correlations could be calculated as: $Z = \frac{Z_A - Z_B}{\sqrt{\frac{1}{n_A-3}+\frac{1}{n_B-3}}}$ (2), where n_A and n_B represented sample size for each gene pair under each condition. (1-correlation coefficient) was referred to distance measure according to the *cutree* function. And then local false discovery rate (lfdr) was utilized to control true estimates as well as identify remarkable differential correlations. Secondly, we calculated eigen-molecule or "eigengene" according to first principal component of a module derived from hierarchical cluster analysis. Besides pair-wise differential correlations among genes, module correlations were also evaluated by eigen-molecule modules. For module network visualization, two functions *get.eigen.molecule* and *get.eigen.molecule.graph* were implemented. Thirdly, we integrated diverse pretreatment methods with downstream correlation analysis under function *comp.2.cc.fdr*. Eventually, we got pair-wise differential correlations between genes as well as differential correlations between

modules.

4.5. Visualization analysis

Visualization of differential co-expressed network comprised of oppositely correlated genes was realized using software Cytoscape (3.9.0) [51].

4.6. Survival analysis of differentially correlated core genes

We conducted survival analysis of core genes using GEPIA from TCGA datasets. We also employed GEO datasets (GSE26193, GSE26712, GSE30161 and GSE3149) as validation groups for the survival analysis of these genes using Kaplan-Meier Plotter. $P < 0.05$ was referred as remarkable difference [52].

4.7. Cell lines and cell culture

The OC cell lines (Kuramochi, Ovsaho, Ovcar4, Ovcar8 and Tyknu) and normal human ovarian epithelial cell line IOSE80 were purchased from Cell Bank of Chinese Academy of Sciences (Shanghai, China). Medium RPMI-1640 (Thermo, USA) adding 10 % fetal bovine serum (FBS) was used to culture cells. All cell media contained penicillin (100U/mL) and streptomycin (100U/mL) at 37 °C in a 5 % CO_2 incubator.

4.8. Tissue specimens and HGSC patients

Fifteen pairs of adjacent normal and HGSC tumor tissues were gathered from the Third Affiliated Hospital of Soochow University from September 2022 to May 2023. Tissues were stored at −80 °C. Our research was permitted by Research Ethics Committee of the Third Affiliated Hospital of Soochow University, consistent with the Declaration of Helsinki. Written informed consent was obtained from all the patients.

4.9. RT-qPCR

TRIzol reagent (Thermo, USA) was utilized to isolate total RNA. We utilized a ReverTra Ace qPCR RT Kit (Toyobo, Osaka, Japan) to synthesize the first strand cDNA. For RT-qPCR, we employed the Fast SYBR Green Master Mix (Applied Biosystems Inc., CA, USA). Our loading control was GAPDH. The expression levels of genes were evaluated by the relative quantification 2-ΔΔCT method.

4.10. Expression analysis of core genes

We employed GSE223426 as a validation dataset for mRNA expression analysis of core genes generated by GEO2R [53].

4.11. Statistical analysis

Student's *t*-test was employed to distinguish differential expressed genes. Adjusted *P* values less than 0.05, computed by Benjamini-Hochberg method, represented remarkable difference [54]. Fisher's z-test was employed to assess differentially correlated gene pairs. Less than 0.05 lfdr was referred as remarkable difference.

Ethics declarations

This study was reviewed and approved by the Third Affiliated Hospital of Soochow University, and the approval number was 2022-science-158. All participants provided informed consent of the study.

Funding

This work was supported by grants from the National Key R&D Program of China (2022YFF1203202), National Natural Science Foundation of China (31701111), Natural Science Foundation of Universities in Jiangsu Province (23KJB310024), Top Talent of Changzhou "The 14th Five-Year Plan" High-Level Health Talents Training Project (2022260), Strategic Priority Research Program of Chinese Academy of Sciences (XDB38050200, XDA26040304), Self-supporting Program of Guangzhou Laboratory (SRPG22-007), Major Science and Technology Project of Changzhou Health Commission (ZD202203), Applied Basic Research Project of Changzhou (CJ20220107), Youth Talent Science and Technology Project of Changzhou Health Commission (QN202014).

Data availability statement

ScHGSC-IGDC is freely available as open-source (MIT license) at https://github.com/zhouxiaoyou2023/ScHGSC-IGDC.git. The scRNA-seq data of HGSC (GSE189955) was downloaded from Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/query/

acc.cgi?acc=GSE189955).

CRediT authorship contribution statement

Yuanqi Li: Writing – original draft, Methodology, Formal analysis, Data curation. **Qi Wang:** Writing – review & editing, Investigation, Formal analysis. **Xiao Zheng:** Writing – review & editing, Investigation. **Bin Xu:** Writing – review & editing, Investigation. **Wenwei Hu:** Writing – review & editing, Resources. **Jinping Zhang:** Writing – review & editing, Investigation. **Xiangyin Kong:** Writing – review & editing, Resources. **Yi Zhou:** Writing – review & editing, Conceptualization. **Tao Huang:** Writing – review & editing, Resources, Investigation. **You Zhou:** Writing – original draft, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e32909.

References

[1] R.L. Siegel, et al., Cancer statistics, 2023, CA A Cancer J. Clin. 73 (1) (2023) 17–48.
[2] N. Cancer Genome Atlas Research, Integrated genomic analyses of ovarian carcinoma, Nature 474 (7353) (2011) 609–615.
[3] L. Kelland, The resurgence of platinum-based cancer chemotherapy, Nat. Rev. Cancer 7 (8) (2007) 573–584.
[4] H. Zhang, S. Wang, T. Huang, Identification of chronic hypersensitivity pneumonitis biomarkers with machine learning and differential Co-expression analysis, Curr. Gene Ther. 21 (4) (2021) 299–303.
[5] L. Sheng, et al., Identification of hub genes with differential correlations in sepsis, Front. Genet. 13 (2022) 876514.
[6] S. Pido, G. Ceddia, M. Masseroli, Computational analysis of fused co-expression networks for the identification of candidate cancer gene biomarkers, NPJ Syst Biol Appl 7 (1) (2021) 17.
[7] Y. Yang, et al., Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types, Nat. Commun. 5 (2014) 3231.
[8] B. Izar, et al., A single-cell landscape of high-grade serous ovarian cancer, Nat. Med. 26 (8) (2020) 1271–1279.
[9] Q. Hao, et al., Single-cell transcriptomes reveal heterogeneity of high-grade serous ovarian carcinoma, Clin. Transl. Med. 11 (8) (2021) e500.
[10] S.A. MacParland, et al., Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations, Nat. Commun. 9 (1) (2018) 4383.
[11] N. Aizarani, et al., A human liver cell atlas reveals heterogeneity and epithelial progenitors, Nature 572 (7768) (2019) 199–204.
[12] M. Bockmayr, et al., New network topology approaches reveal differential correlation patterns in breast cancer, BMC Syst. Biol. 7 (2013) 78.
[13] T. Ando, R. Kato, H. Honda, Differential variability and correlation of gene expression identifies key genes involved in neuronal differentiation, BMC Syst. Biol. 9 (2015) 82.
[14] M. Scheffer, et al., Anticipating critical transitions, Science 338 (6105) (2012) 344–348.
[15] A. Sipos, et al., The role of the microbiome in ovarian cancer: mechanistic insights into oncobiosis and to bacterial metabolite signaling, Mol. Med. 27 (1) (2021) 33.
[16] L. Chen, et al., Altering the microbiome inhibits tumorigenesis in a mouse model of oviductal high-grade serous carcinoma, Cancer Res. 81 (12) (2021) 3309–3318.
[17] D. Zhang, et al., Metabolic regulation of gene expression by histone lactylation, Nature 574 (7779) (2019) 575–580.
[18] J. Pouyssegur, et al., 'Warburg effect' controls tumor growth, bacterial, viral infections and immunity - genetic deconstruction and therapeutic perspectives, Semin. Cancer Biol. 86 (Pt 2) (2022) 334–346.
[19] Y.J. Wu, et al., Targeting oxidative phosphorylation as an approach for the treatment of ovarian cancer, Front. Oncol. 12 (2022) 971479.
[20] S.L. Salscheider, et al., AIFM1 is a component of the mitochondrial disulfide relay that drives complex I assembly through efficient import of NDUFS5, EMBO J. 41 (17) (2022) e110784.
[21] A. Murari, et al., Phospholipids can regulate complex I assembly independent of their role in maintaining mitochondrial membrane integrity, Cell Rep. 42 (8) (2023) 112846.
[22] J.S. Raju, et al., Prognostic value of TNFR2 and STAT3 among high-grade serous ovarian cancer survivors according to platinum sensitivity, Diagnostics 11 (3) (2021) 526.
[23] C. Chen, et al., ERBB3-induced furin promotes the progression and metastasis of ovarian cancer via the IGF1R/STAT3 signaling axis, Oncogene 39 (14) (2020) 2921–2933.
[24] E. Koren, et al., Thy1 marks a distinct population of slow-cycling stem cells in the mouse epidermis, Nat. Commun. 13 (1) (2022) 4628.
[25] E.V. Connor, et al., Thy-1 predicts poor prognosis and is associated with self-renewal in ovarian cancer, J. Ovarian Res. 12 (1) (2019) 112.
[26] Y. Xing, et al., Progress on the function and application of thymosin beta4, Front. Endocrinol. 12 (2021) 767785.
[27] Z. Yang, et al., TMSB4X: a novel prognostic marker for non-small cell lung cancer, Heliyon 9 (11) (2023) e21505.
[28] W. Pu, et al., Single-cell transcriptomic analysis of the tumor ecosystems underlying initiation and progression of papillary thyroid carcinoma, Nat. Commun. 12 (1) (2021) 6058.
[29] C.Y. Kuo, et al., Aberrant expression of thymosin beta-4 correlates with advanced disease and BRAF V600E mutation in thyroid cancer, J. Histochem. Cytochem. 70 (10) (2022) 707–716.
[30] T. Morita, K. Hayashi, Tumor progression is mediated by thymosin-β4 through a TGFβ/MRTF signaling Axis, Mol. Cancer Res. 16 (5) (2018) 880–893.
[31] T. Ismail, et al., Interplay between mitochondrial Peroxiredoxins and ROS in cancer development and progression, Int. J. Mol. Sci. 20 (18) (2019) 4407.
[32] C.Y. Wei, et al., Identification of hypoxia signature to assess the tumor immune microenvironment and predict prognosis in patients with ovarian cancer, Internet J. Endocrinol. 2021 (2021) 4156187.
[33] T. Sasahira, et al., SERPINE2 is an oral cancer-promoting factor that induces angiogenesis and lymphangiogenesis, Int. J. Clin. Oncol. (2021) 1–9.
[34] J. Zhang, et al., SERPINE2/PN-1 regulates the DNA damage response and radioresistance by activating ATM in lung cancer, Cancer Lett. 524 (2022) 268–283.

[35] V. Pagliara, et al., Protease Nexin-1 affects the migration and invasion of C6 glioma cells through the regulation of urokinase Plasminogen Activator and Matrix Metalloproteinase-9/2, Biochim. Biophys. Acta 1843 (11) (2014) 2631–2644.
[36] C.M. McKee, et al., Protease nexin 1 inhibits hedgehog signaling in prostate adenocarcinoma, J. Clin. Invest. 122 (11) (2012) 4025–4036.
[37] H. Chehade, et al., MNRR1 is a driver of ovarian cancer progression, Transl Oncol 29 (2023) 101623.
[38] D.V. Ziegler, et al., Calcium channel ITPR2 and mitochondria-ER contacts promote cellular senescence and aging, Nat. Commun. 12 (1) (2021) 720.
[39] N. Martin, D. Bernard, Calcium signaling and cellular senescence, Cell Calcium 70 (2018) 16–23.
[40] C. Wiel, et al., Endoplasmic reticulum calcium release through ITPR2 channels leads to mitochondrial calcium accumulation and senescence, Nat. Commun. 5 (2014) 3792.
[41] J. Shi, L. Fu, W. Wang, High expression of inositol 1,4,5-trisphosphate receptor, type 2 (ITPR2) as a novel biomarker for worse prognosis in cytogenetically normal acute myeloid leukemia, Oncotarget 6 (7) (2015) 5299–5309.
[42] N. Zhang, et al., Germline genetic variations in PDZD2 and ITPR2 genes are associated with clear cell renal cell carcinoma in Chinese population, Oncotarget 8 (15) (2017) 24196–24201.
[43] Z.Z. Chen, et al., mcPGK1-dependent mitochondrial import of PGK1 promotes metabolic reprogramming and self-renewal of liver TICs, Nat. Commun. 14 (1) (2023) 1121.
[44] R. Gou, et al., PGK1 is a key target for anti-glycolytic therapy of ovarian cancer: based on the comprehensive analysis of glycolysis-related genes, Front. Oncol. 11 (2021) 682461.
[45] Y. Wu, et al., Pim1 promotes cell proliferation and regulates glycolysis via interaction with MYC in ovarian cancer, OncoTargets Ther. 11 (2018) 6647–6656.
[46] E. Hjerpe, et al., Metabolic markers GAPDH, PKM2, ATP5B and BEC-index in advanced serous ovarian cancer, BMC Clin. Pathol. 13 (1) (2013) 30.
[47] Y. Wang, et al., Single-cell dissection of the multiomic landscape of high-grade serous ovarian cancer, Cancer Res. 82 (21) (2022) 3903–3916.
[48] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinf. 9 (2008) 559.
[49] G. Yu, et al., clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS 16 (5) (2012) 284–287.
[50] A. Fukushima, DiffCorr: an R package to analyze and visualize differential correlations in biological networks, Gene 518 (1) (2013) 209–214.
[51] N.T. Doncheva, et al., Cytoscape StringApp: network analysis and visualization of proteomics data, J. Proteome Res. 18 (2) (2019) 623–632.
[52] F. Vandin, et al., Accurate computation of survival statistics in genome-wide studies, PLoS Comput. Biol. 11 (5) (2015) e1004071.
[53] T. Barrett, et al., NCBI GEO: archive for functional genomics data sets-update, Nucleic Acids Res. 41 (D1) (2013) D991–D995.
[54] Y. Hochberg, Y. Benjamini, More powerful procedures for multiple significance testing, Stat. Med. 9 (7) (1990) 811–818.