Heliyon 10 (2024) e37288

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Identification of ubiquitin markers for survival and prognosis of ovarian cancer

Yiwen Feng [a,1], Liyun Shan [a,1], Yanping Gong [b], Wenzhao Hang [a], Zhenyu Sang [a], Yunyan Sun [a], Kefu Tang [c], Yulan Wang [b], Binjie Hu [b], Xiaowei Xi [a,*]

[a] Departments of Obstetrics and Gynecology, Shanghai General Hospital, Shanghai Jiaotong University School of Medicine, Shanghai 200080, PR China
[b] Departments of Oncology, Shanghai General Hospital, Shanghai Jiaotong University School of Medicine, Shanghai 200080, PR China
[c] Center for Reproductive Medicne, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai 200135, PR China

ARTICLE INFO

Keywords:
Ovarian cancer
Ubiquitin
SNP
Survival
Drug

ABSTRACT

Ovarian cancer (OC) is one of the most common malignancies and a leading cause of death among women worldwide. The ubiquitin pathway plays an important role in OC development. Using the single nucleotide polymorphism data obtained using the prevalence and dominance strategies, four ubiquitin marker genes were identified according to their expression levels: BARD1, BRCA2, FANCA, and BRCA1. Based on these four genes, a consensus clustering of OC expression data was performed. The significant differences in the survival analysis, ESTIMATE, and CIBERSORT results among the clusters indicated the pivotal role of these four genes in OC development. Of the ubiquitin-representative genes in each cluster, two ubiquitin genes, *TOP2A* and *MYLIP*, were identified in the survival risk model after univariate survival, Least Absolute Shrinkage and Selection Operator regression, and multivariate survival analyses. The reliability and robustness of both the training and validation data were confirmed by comparing the significant survival difference between high- and low-risk patients. We further explored the association between our risk model and clinical outcomes as well as predicted potentially interacting drugs. The co-expression network showed clear interactions among the four marker genes and two model genes and between high- and low-risk differentially expressed genes. Significantly enriched genes were found in pathways associated with ion channels, channel activity, and neuroactive ligand–receptor interactions. Therefore, suggesting the involvement of ubiquitin genes in the survival and development of OC through neurohumoral regulation. Our results will provide valuable reference or supplementary information for studies investigating OC diagnosis and therapies.

1. Introduction

Ovarian cancer (OC) is one of the most common malignancies affecting women worldwide [1], there reported about 61,100 people diagnosed ovarian cancer patients and about 32,600 deaths by 2022 [2].OC occurs in women of all ages, and its incidence usually increases with age [3]. Epithelial OC (EOC) is the predominant clinical subtype, accounting for approximately 80–90 % of all cases.

* Corresponding author.
E-mail address: xixiaowei2018@163.com (X. Xi).
[1] These authors contributed to this work equally.

https://doi.org/10.1016/j.heliyon.2024.e37288
Received 30 August 2023; Received in revised form 11 August 2024; Accepted 30 August 2024
Available online 7 September 2024

EOCs are highly heterogeneous malignancies and can be classified into invasive and borderline types. Invasive epithelial OC usually has a poor prognosis, with a 5-year survival rate of approximately 35 %, posing a significant threat to women's lives. Certain rare pathological types of OC predominantly affect younger women, severely affecting their quality of life and fertility. Understanding the molecular mechanisms of epithelial OC and developing targeted therapeutic drugs are crucial for clinical and social significance.

OC development is influenced by genetic, environmental, and endocrine factors, with genetic factors being particularly significant. Approximately 20 % of patients have a familial EOC risk attributed to mutations in genes carrying high to moderate risk [4,5]. In OC, deleterious mutations have been identified in several genes, including *BRCA1*, *BRCA2*, *RAD51D*, *HOXD9*, *TERT*, *OBFC1*, *BRIP1*, *FANCM*, *MLH1*, *MSH2*, *MSH6*, and *PMS2* [6–8]. However, owing to the particularity of genomic coding, only a few damaging mutations may lead to OC development. Mutations detected in *BRCA1* and *BRCA2*, the most frequently reported marker genes for OC, were found in only 40 % of patients with familial OC [9].

Aberrant ubiquitination plays an important role in cancer development by regulating metabolic reprogramming in cancer cells [10], modulating the immunological tumor microenvironment, and maintaining the stemness of cancer stem cells [11]. *BRCA1*, a ubiquitination-related tumor suppressor gene, encodes breast cancer type 1 susceptibility protein, which increases the risk of carcinogenesis [12]. In OC, the interaction between BRCA1 and BRCA1-associated ring domain protein 1 (BARD1) generates significant

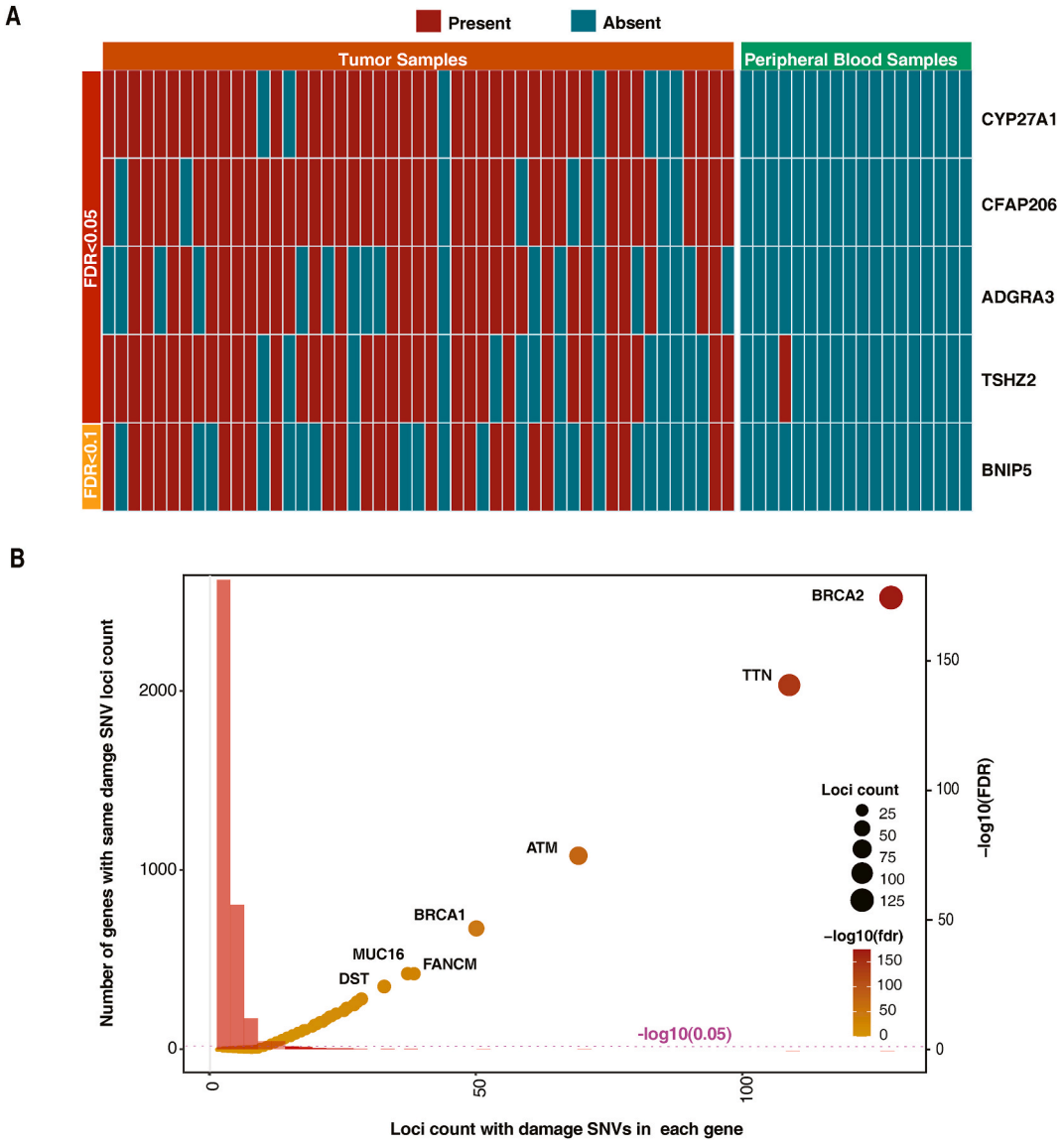


Fig. 1. Significant marker genes identified from the SNP data. A. Marker genes that were significantly enriched in tumor samples according to the count of sample with or without damage SNPs. FDR <0.1. Fisher's test. B. Marker genes with significantly higher frequency (loci) of damage mutations within a specific gene in the tumor samples compared with the expected median frequency. In the poison test, FDR value < 0.05. SNP, single nucleotide variant; FDR, false discovery rate.

ubiquitin ligase activity involved in the DNA damage response. The downregulation of ubiquitin ligase activity of BRCA1–BARD1 is significantly involved in tumor development and regression [13]. In addition to *BRCA* genes, the ubiquitin system plays critical roles in OC development through the ERK, cyclin-dependent, *ERBB2* gene, and p53-dependent pathways [14]. Mitogen-activated protein kinase phosphatases 3 (MPK-3) is a negative regulator of ERK1/2, low MPK-3 expression due to ubiquitination was significantly associated with high ERK1/2 activity in promoting proliferation of primary human ovarian cancer cells [15]. And enforced MPK-3 expression reduced ERK1/2 activity lead significantly inhibition in ovarian cancer cell proliferation and tumor development in nude mice [15]. The ubiquitin pathways could regulate the activity and location of p53. At the same time, p53 could regulate protein homeostasis by downregulating ubiquitin–proteasome system function in response to the cellular stress in ovarian cancer cells [14]. In cisplatin-mediated OC treatment, Lactacystin was report to increase cisplatin toxicity in cisplatin-resistant OC cell lines by inhibiting ubiquitination [16].

Due to the crucial role of ubiquitin in OC development and the impact of single nucleotide variants (SNPs) on the alteration of expression, we investigated the survival implications of ubiquitin genes in patients with OC. By constructing a survival risk model using SNP and expression data, we identified gene markers. This approach aims to provide valuable insights for future studies on OC diagnosis and therapies.

2. Materials and methods

2.1. Clinical samples

Blood samples and tumor samples were collected from patients diagnosed with epithelial serous OC who underwent surgical treatment at the Department of Gynecology of Shanghai First People's Hospital from April 2007 to June 2018 with the following inclusion criteria were included in the study.

1) Non-pregnant or non-lactating women aged >18 years,
2) women with pathologically diagnosed epithelial serous OC,
3) women who did not receive radiotherapy or chemotherapy.

All selected patients were from an unrelated Han population. Tumor tissue and blood matching samples were collected from 18 patients, and besides paired samples, another batch of tumor samples were collected from other 36 patients. The Ethics Committee of Shanghai First People Hospital approved this study, and informed consent was obtained from all participants.

2.2. DNA isolation for Infinium OncoArray-500K BeadChip

Total DNA was isolated, digested, and quantified using a NanoDrop microspectrophotometer and fluorescence spectrophotometer. The DNA quality was assessed through agarose gel electrophoresis.

2.3. SNP quality control (QC)and annotation

SNPs were detected using Illumina Infinium OncoArray-500K BeadChip following the recommended experimental procedure. SNP QC was conducted following OncoArray QC Guidelines [17]. The raw intensity data were processed using Genome Studio (Illumina) with default settings. The SNP calling rates were calculated using PLINK. Replacement quality control was performed based on the IMPUTE information score, and single nucleotide polymorphisms (SNPs) with a score of <0.80 or minor alleles of <0.01 were excluded. SNPs were annotated using ANNOVAR [18].

With the SNPs identified we performed the downstream analysis as showed in flowchart Fig. 1.

2.4. Analysis of damage SNPs

Exome SNPs that were annotated with startloss, stoploss, stopgain, frameshift_deletion, and nonsynonymous_SNP with a combined annotation dependent depletion (CADD) score of ≥15 [19,20] were considered as damage SNPs. Genes with a significantly higher frequency of damaged SNPs in the unique loci of tumor samples (Poisson distribution test) or compared with the control samples according to sample count (Fisher test) were identified as potentially damaged genes. Blood samples were used as the controls.

2.5. Download of transcriptome data and ubiquitin gene set

The gene expression data sets GSE18521 (platforms: GPL570, 53 case/10 control) and GSE102073 (platforms: GPL16791, 85 cases) were searched using the keyword "OC" and downloaded. For genes with multiple probe ID, the mean of these probe values was calculated as the gene expression value. All collections of ubiquitin pathway genes were searched in The Molecular Signatures Database [21] using the keyword "ubiquitin." The identified genes were used as the ubiq gene set for the downstream analysis (Table S1).

2.6. Differential expressed gene and consensus clustering analyses

R package "limma" v3.16 was used for differential expressed gene analysis, and genes with a p value of <0.05 and $|\log2FC|$ value of >0.58 were considered as significant differentially expressed genes (DEGs). DEGs were identified by intersecting ubiq genes and damaged SNP candidate genes to obtain the ubiq-damaged genes. Consensus clustering of OVC samples were performed using R packages "ConsensusClusterPlus" v3.16 with code "ConsensusClusterPlus (d,maxK = 10,reps = 1,000,pItem = 0.9,pFeature = 1,title = title, clusterAlg = "hc",distance = "pearson",seed = 1262118388.71279,plot = "png")." Survival analyses were performed using R package "survival" v3.5-5 and "survminer" v0.4.9.

2.7. Identification of representative genes for clusters

To identify the representative genes of each cluster, DEG analysis was performed between samples in each cluster and other OC samples using R package "limma" v3.16. Genes with a p value of <0.05 and log2FC value of >0.58 were selected as the representative genes of a specific cluster.

2.8. Construction and validation of the risk model and survival analysis of high-low risk samples

After identifying the ubiquitin genes among the cluster representative genes, a survival risk model was constructed using Cox survival regression with a single variable, Least Absolute Shrinkage and Selection Operator (LASSO) regression, and Cox survival regression with multiple variables. The risk model was defined as Rsikscore$\sim\beta_1$Gene 1$+\beta_2$Gene 2 …. $+\beta_N$GeneN, where β_n is the coefficient of a specific gene in the multiple variable Cox survival regression, and GeneN is the expression value of a specific gene. The patients were divided into high- and low-risk groups based on the median risk score. Survival analysis was conducted using R package "survival" v3.5-5 and "survminer" v0.4.9. The risk model was validated using dataset GSE102073.

2.9. CERBORSORT and ESTIMATE analyses

CERBORSORT and ESTIMATE analyses were conducted using R package "CIBERSORT" v3.0 [22] and "estimate" v2.0.0 [23].

2.10. Drug interaction prediction

Survival-associated ubiquitin genes were integrated into drug interaction predictions in DGIdb v4.0 [24], tested using AutoDock vina [25] and illustrated in pymol v4.6 [26]. Protein and drug molecule Protein Data Bank files were downloaded from AlphaFold (https://alphafold.ebi.ac.uk/) [27] and DrugBank (https://go.drugbank.com/) [28].

2.11. Construction of co-expression network

DEGs identified between high- and low-risk samples were utilized to construct a co-expression network based on Spearman's correlation. The significance for each correlation pair was adjusted using the "Benjamini–Hochberg" (BH) method, with correlation pairs having an adjusted p value (FDR) of <0.05 deemed significant. Network visualization was performed using Cytoscape V3.9.1 [29].

2.12. Western blot analysis

Anti-*MYLIP* antibody and anti-*TOP2A* antibody were purchased from Proteintech (USA), and total cellular protein was obtained through 10000–14000g centrifuging for 10min after RIPA lysis buffer (Beyotime, China) infiltration. After blocking with 5 % skimmed milk powder solution at 37 °C for 2h, anti-*MYLIP* and anti-*TOP2A* antibody were incubated on protein sample at 4 °C for 12 h. Secondary antibody (1:10000) was then added onto membrane for incubation for 2h at 37 °C. Detection was conducted using an enhanced chemiluminescence kit (NCM Biotech, China). The protein expression level was analyzed using ImageJ software.

2.13. Statistical analysis

R v4.3.2 was used for all statistical analyses and the generation of graphs except that in Western blot analysis. Prism 8 software (Graphpad, La Jolla, CA, USA) was used for the statistical analysis in Western blot analysis. The statistical significance of genes was obtained using Fisher test or Poison test (with Bonferroni post hoc test) and the difference of ESTIMATE or CIBERSORT analysis was analyzed using Wilcox Test. The comparison of relative protein level was performed using T test. Cluster Enrichment was performed with Fisher Test. P value < 0.05 was considered statistically significant.

3. Results

3.1. SNP calling, annotation, and identification of candidate pathogenic genes

After SNP quality control, 1,129,847 unique SNPs were identified and annotated using ANNOVAR [18]. Among the annotated SNPs, 16,141 exonic SNPs annotated as frameshift deletion, startloss, stopgain, stoploss, and nonsynonymous SNPs with a CADD score of ≥15 [19,20] as well as some SNPs with unknown annotation that were deletion or nonsynonymous SNPs with a CADD score ≥15 (Table S2) were retained due to their potential pathogenicity in OC. Among the 1,6141 SNPs, 13,388 (Table S3) were found in genes with at least two damaged SNPs in the tumor samples. Two strategies were used to identify the candidate pathological genes for these recurrent genes. Both tumor and blood samples (serving as control samples) containing damaged SNPs belonging to each specific recurrent gene were counted. The enrichment of each gene in the tumor or blood samples was performed (Fisher's test, BH-adjusted FDR <0.05), identifying five genes that were significantly enriched in the tumor samples (Fig. 1A–Table S4). However, based on the frequency of unique damage SNPs in each gene observed in tumor samples, 118 genes were found to significantly contribute to the

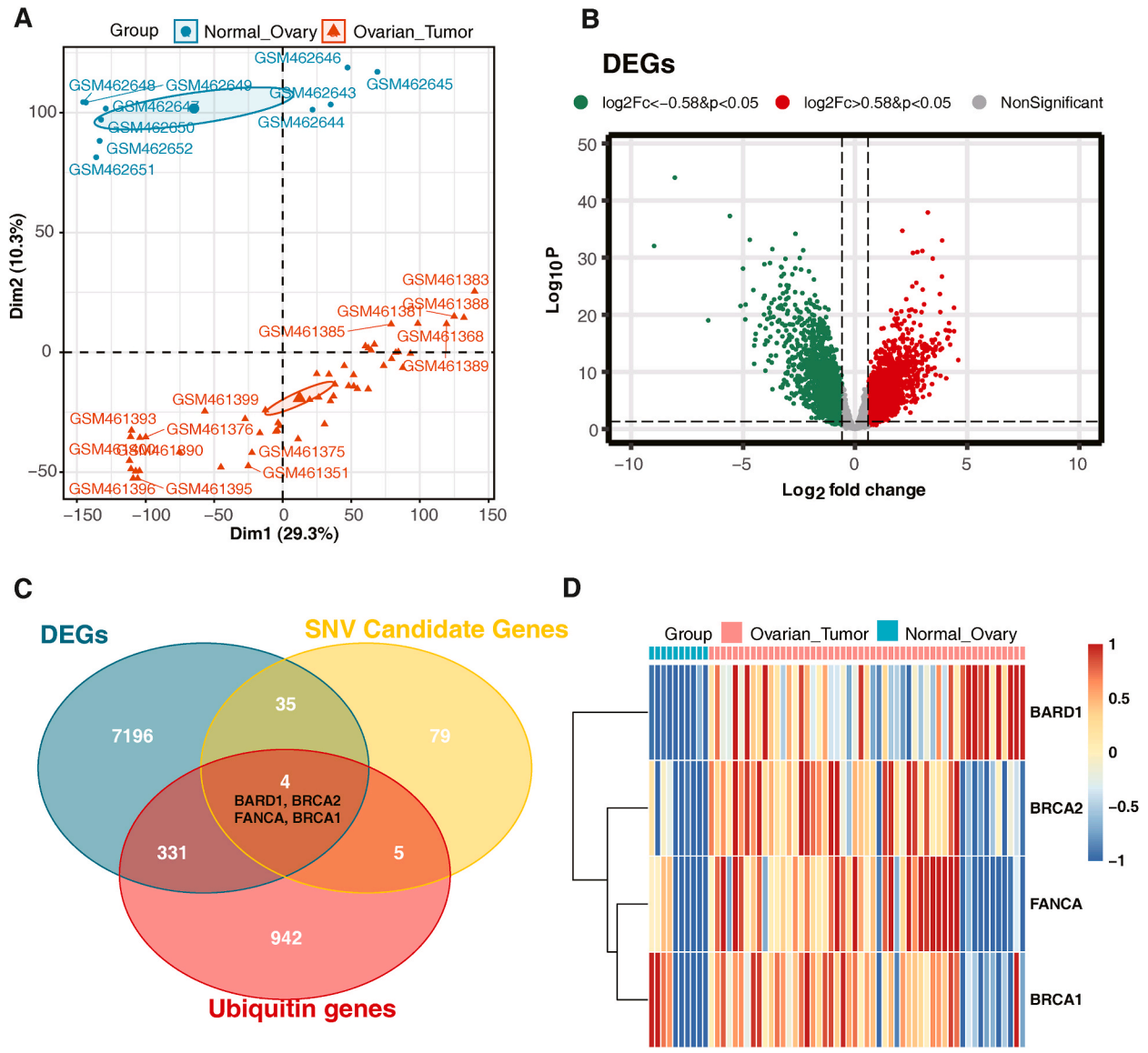


Fig. 2. Identification of four ubiquitin marker genes from the SNP and expression data. A. PCA of ovarian tumor and normal ovarian tissues according to the mRNA expression value. B. DEG analysis between ovarian tumor tissue and normal ovarian tissue. mRNA with a p value of <0.05 and a |log2FoldChange| value of >0.58 was considered significant. C. Intersection among DEGs, SNP candidate genes acquired with significant enrichment, higher frequency, and ubiquitin pathway genes. D. Expression of four ubiquitin marker genes. DEGs, differentially expressed genes; PCA, principal component analysis.

pathogenic risk (Poisson distribution test, BH-adjusted FDR <0.05, expectation = 2) (Fig. 1B–Table S5). Finally, we combined 123 candidate genes for downstream analysis.

3.2. Gene differential expression analysis and consensus clustering in ovarian cancer samples

The expression data GSE18521 of patients with OC were further downloaded, and performing principal component analysis suggested an obvious distinction between the case and control groups (Fig. 2A). Overall, 7566 DEGs were identified based on a |log2FC| value of >0.58 and p value of <0.05. Among these DEGs, 5150 were upregulated, while 2416 were downregulated in cancer samples (Fig. 2B–Table S6). After intersecting with 123 candidate SNP genes, 7566 DEGs, and 1282 ubiquitin genes, four genes (*BARD1*, *BRCA2*, *FANCA*, and *BRCA1*) were found to be significantly associated with OC in the SNPs and expression data (Fig. 2C), while four genes were found to be upregulated in OC samples (Fig. 2D). Using the four genes, OC samples were grouped into three clusters (Fig. 3A and B) with significant inter-cluster separation according to the results of non-metric multidimensional scaling (NMDS) and analysis of similarities (Fig. 3C). Significant survival differences were found among samples from clusters; samples from cluster3 exhibited worse survival, whereas those from cluster2 demonstrated the best survival (Fig. 3D), indicating the potential association of these four genes with the survival of patients with OC. In ESTIMATE analysis, consistent with the outcomes of the survival analysis, samples in cluster2 showed the highest stromal, immune, and ESTIMATE scores but the lowest tumor purity score (Fig. 3E). In the CIBERSORT analysis, samples from cluster2 showed the highest proportion of cluster of differentiation CD8 T cells, which are the most potent effectors in the anticancer immune response (Fig. 3F) [30], contributing to better survival.

3.3. Construction of survival risk model from representative genes for each cluster

When the significant survival differences among clusters were assessed, 2,962, 3,465, and 129 representative genes were identified in cluster 1, cluster2, and cluster3 and intersected with the ubiquitin gene set separately (Table S7), leaving 379 ubiquitin-representative genes for downstream analysis (Fig. 4A, Table S8). To obtain the survival risk score, five genes ($p < 0.1$ from the single-variable Cox model) were selected for LASSO regression (Fig. 4B–Table S9). After LASSO regression, *MYLIP*, *TOP2A*, and *RAG1* ($p < 0.05$) (Fig. 4C) were selected for further multivariate Cox survival analysis. However, only *MYLIP* and *TOP2A* ($p < 0.05$) were significantly associated with the survival of OC samples (Fig. 4D). Using the coefficients of *MYLIP* and *TOP2A*, a survival risk model was constructed as follows.

$$Riskscore \sim (-0.7629768)^{*}\ \mathrm{EXP}_{MYLIP} + 0.6215511^{*}\mathrm{EXP}_{TOP2A}$$

where EXP_{MYLIP} and EXP_{TOP2A} represent the expression values of *MYLIP* and *TOP2A* in the samples.

According to the patients' median risk score, significant differences were observed in the expression of model genes and survival between the high- and low-risk groups (Fig. 4E and F). Based on the risk score model, receiver operating characteristic (ROC) analysis showed robustness in predicting the 1-, 3-, and 5-year survival rates (Fig. 4G). The nomogram and calibration curves indicated the effectiveness and independence of the risk score in survival prediction in patients with OC (Fig. 4H and I).

3.4. Validation of the risk score model in different datasets

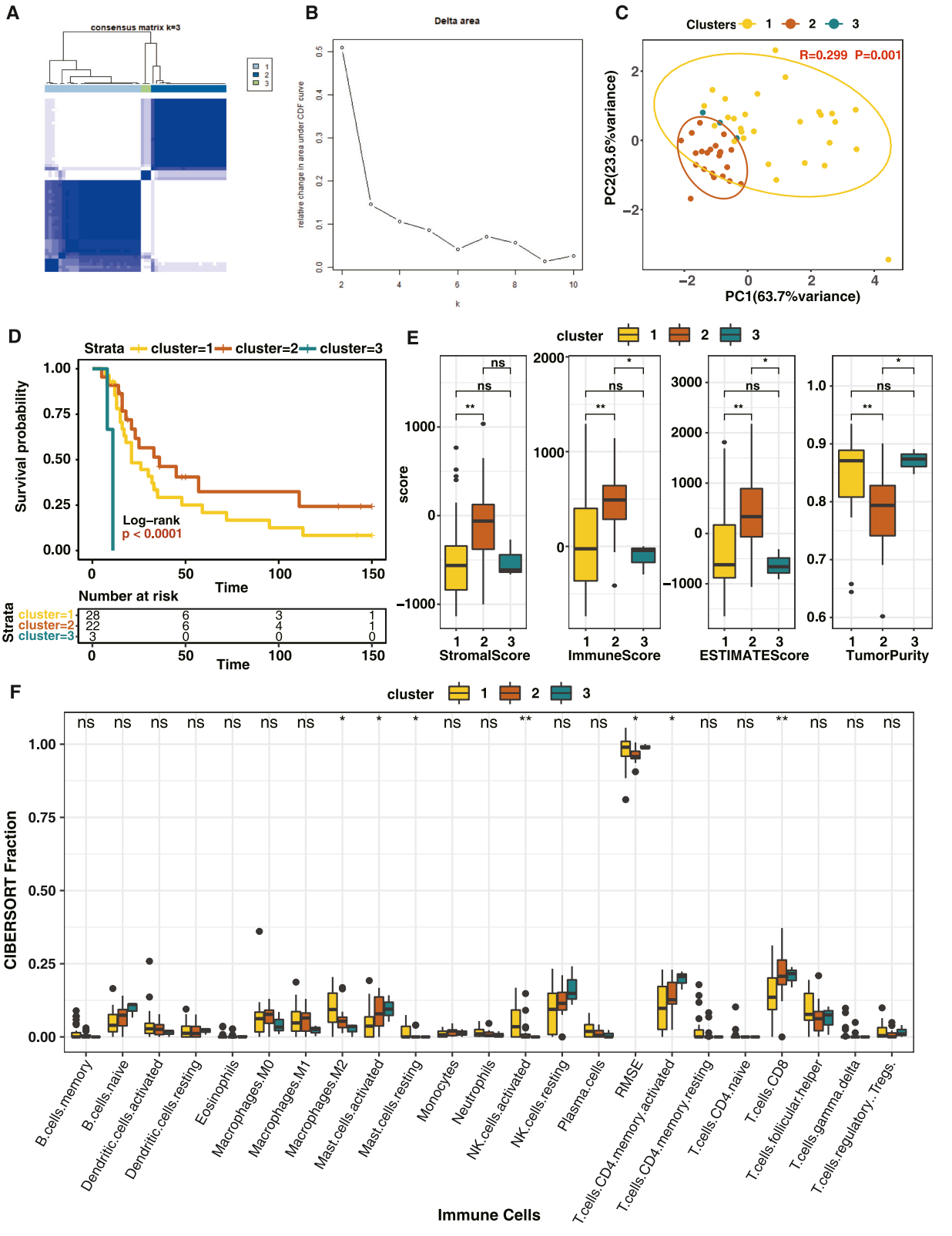
To validate the constructed risk score model, another set of expression data from patients with OC, GSE102073, was downloaded and rated. The significant differences observed in the expression of model genes and survival analysis between high- and low-risk patients validated the robustness of the survival risk score model and high 1-, 3-, and 5-year survival prediction ROC (Fig. 5A–C). Some clinical information revealed that samples with serous tubal intraepithelial carcinoma (STIC) were significantly enriched in the high-risk group (Fig. 5D) and showed higher risk scores (Fig. 5E). Patients in late pathological stages also had significantly higher risk scores (Fig. 5F).

3.5. Prediction of drug interaction with model genes

The association between risk score model genes and survival was validated. To identify the drug molecules that could potentially improve the prognosis of patients with OC, we predicted the drugs targeting the two model genes using DGIdb [24]. Eighty drugs interacted with TOP2A, most of which played an inhibitory role (Table S10). As in our risk-survival model, TOP2A showed a positive association with the risk score. Therefore, inhibitory drugs with interacting TOP2A could potentially improve the survival of patients with OC. AMRUBICIN interacted with TOP2A with a clear inhibitory role and the highest interaction score; thus, it was selected for drug docking analysis. As shown in Fig. 5G, AMRUBICIN could show a tight interaction with TOPA2 (P11388) through hydrogen bonding, with a minimum binding affinity of −9 Å (Table S11).

3.6. Construction of correlation network between SNP marker genes, model genes, and high-to low-risk DEGs

To explore the relationship between SNP marker ubiquitin genes and risk model genes, and their roles in classifying OC risk, a DEG analysis was performed between high- and low-risk samples (Table S12), constructing a co-expression network between SNP ubiquitin genes, risk model genes, and high-risk DEGs. As shown in Fig. 6, the model genes *TOP2A* and *MYLIP* clustered separately, while four



(caption on next page)

Fig. 3. Construction of consensus clustering using four ubiquitin marker genes and related analysis among clusters.
A and B. We identified a consensus expression matrix and delta areas with three clusters. C. NMDS and ANOSIM analyses of three clusters. An R value of >0 suggested a difference among clusters, whereas a p value of <0.05 suggested a significant difference. D. Survival analysis showed a significant difference in survival among clusters. E. ESTIMATE analysis. *:0.01< p < 0.05, **:0.001< p < 0.01. F. CIBERSORT analysis. *:0.01< p < 0.05, **:0.001< p < 0.01. NMDS, nonmetric multidimensional scaling; ANOSIM, analysis of similarities; ns, not significant.

SNP ubiquitin maker genes, *BRCA2*, *FANCA*, *BARD1*, and *BRCA1*, were closely connected with the two model genes, playing hub roles in mediating the risk. Both model genes showed significant differences between high- and low-risk samples. Although no significant differences were found in the four SNP ubiquitin marker genes, a correlation was still confirmed. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed on the genes in the co-expression network. The GO enrichment analysis revealed significant enrichment in several pathways, including hormone activity, ligand-gated cation channel activity, signal release, ion channel, and associated components or complexes. In KEGG analysis, pathways such as neuroactive ligand-receptor interaction, serotonergic synapse, calcium signaling pathway, and nicotine addiction were significantly enriched. Therefore, co-expressed genes associated with OC risk are likely involved in neurohumoral regulation (Figure S2, Figure S3, Table S13).

3.7. In vivo experimental validation

With collection of tumor tissues and tumor-adjacent normal tissues from 4 OC patients as well as tumor tissues from 4 OC patients in I & II stage and 4 OC patients in III & IV stage, we performed the Western-Blot analysis for identified 2 ubiquitin markers TOP2A and MYLIP. It was found that MYLIP showed no significant different difference in relative protein level between tumor tissues and normal tissues (Fig. 7A and B, Fig. 4) and TOP2A showed significantly higher protein level in tumor tissues than in normal tissues (Fig. 7A and C, Fig. 5). Furthermore, we found that in patients in earlier stage (I&II) MYLIP was given a higher relative protein level than that in late stage (III&IV) (Fig. 7D and E, Fig. 6). And TOP2A was found higher in relative protein level in late stage (III&IV) samples than in earlier stage (I&II) samples (Fig. 7D and F, Fig. 7) though without statistical significance. The Western-Blot result further validated the roles of TOP2A and MYLIP played in OC development.

4. Discussion

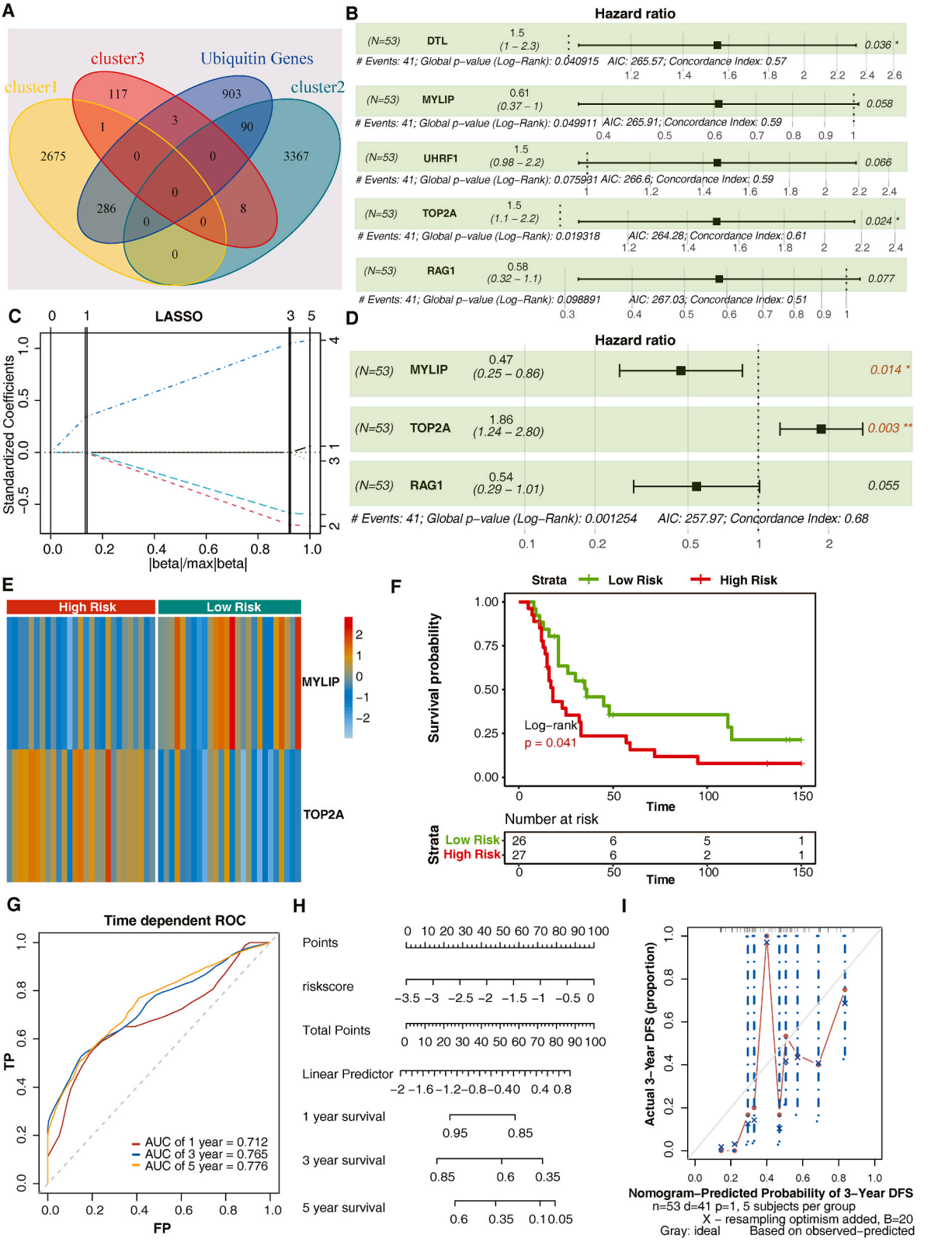
OC is one of the most common type of cancer and the leading cause of death in women worldwide [31]. Due to the significant association between ubiquitination and OC, we constructed a ubiquitin gene-associated survival risk model and detected background pathways by exploring SNP and expression data.

Using damaged SNPs scored using the CADD tool, with a focus on genes with damaged SNPs in at least two patients, four ubiquitin marker genes were identified: *BRCA1*, *BRCA2*, *BARD1*, and *FANCA*. *BRCA1* and *BRCA2* are both tumor suppressor genes, encoding proteins involved in DNA repair and transcriptional regulation for DNA damage [32]. Mutations in both genes significantly contribute to the development of breast cancer and OC [33], with cumulative risks of 30–40 % and 10–15 %, respectively [34]. *BARD1*, whose interaction with *BRCA1* through homologous domains, plays critical roles in inhibiting the progression of several cancers, such as breast cancer and OC, in *BRCA1*-dependent pathways [35]. BARD1 functions as both a tumor suppressor gene and an oncogene, with an increased risk of carcinogenesis when downregulated. Fanconi anemia (FA) is a rare genetic condition with a broad spectrum of clinical features due to DNA repair deficiencies, including progressive bone marrow failure, congenital defects, and cancer predisposition [36]. Furthermore, *BRCA1* and *BRCA2* are FA genes, and significant associations of FA genes, including *FANCA*, with breast and OCs have been reported in several studies [37,38]. Regarding the expression level, the consensus cluster and consistency between ESIMATE and CIBERSORT results suggested the pivotal role of the four genes in regulating OC development and their relationship with the survival of patients with OC.

Due to the significant difference in survival among the clusters, we identified two ubiquitin genes, *TOP2A* and *MYLIP*, from the cluster of representative ubiquitin genes and constructed a survival risk model. The significant survival differences and high area under the ROC curve for survival prediction in the trained and validation datasets suggested the reliability and robustness of the different OC data. As STIC is the most likely precursor lesion for most high-grade serous pelvic carcinomas, significant enrichment of STIC and higher pathological stage in the high-risk group indicated that, despite being constructed with only two genes, our survival risk model showed sensitivity and potential for distinguishing early-stage OC from late-stage OC along with providing early warning signs for high-grade serous OCs.

TOP2A and *MYLIP* were positively and negatively correlated with risk scores, respectively, in our model. *TOP2A* encodes a DNA topoisomerase that modulates the DNA topological structure during transcription. It is a crucial prognostic indicator of tumor advancement and recurrence, particularly associated with poor survival in various cancer types such as breast, ovarian, and lung cancers [39]. *TOP2A* was significantly over-presented in high-grade serous OC and accelerated cancer progression through the transforming growth factor-β/Smad pathway [39]. MYLIP was significantly downregulated in lung cancer, and patients with lung cancer who demonstrated a higher MYLIP expression frequently showed better prognosis. Furthermore, MYLIP showed a significant suppressive effect on lung cancer cell proliferation, migration, and invasion [40]. In breast cancer, MYLIP inhibition is significantly associated with migration and metastasis [41]. In OC, the role of MYLIP has rarely been reported; however, our risk model suggests a protective role of MYLIP in OC development.

Significantly interacting genes within the co-expression network were significantly enriched in pathways associated with ion



(caption on next page)

Fig. 4. Construction of survival risk model and evaluation in the training data set.
A. Veen plot showing the intersection among cluster representative genes and ubiquitin genes. B. Significant genes acquired from single-factor Cox survival analysis. p value < 0.1. C. LASSO regression. D. Significant genes acquired from multi-factor model. p value < 0.05. E. Expression of two model genes in high- and low-risk samples, which was leveled based on the median risk score. F. Survival analysis between high- and low-risk samples. G. ROC plot in prediction of 1-, 3-, 5-years survival training data. H. Nomogram plot for risk score model in training data. I. Calibration plot for risk score model with 3 years survival. LASSO, Least Absolute Shrinkage and Selection Operator; ROC, receiver operating characteristic.

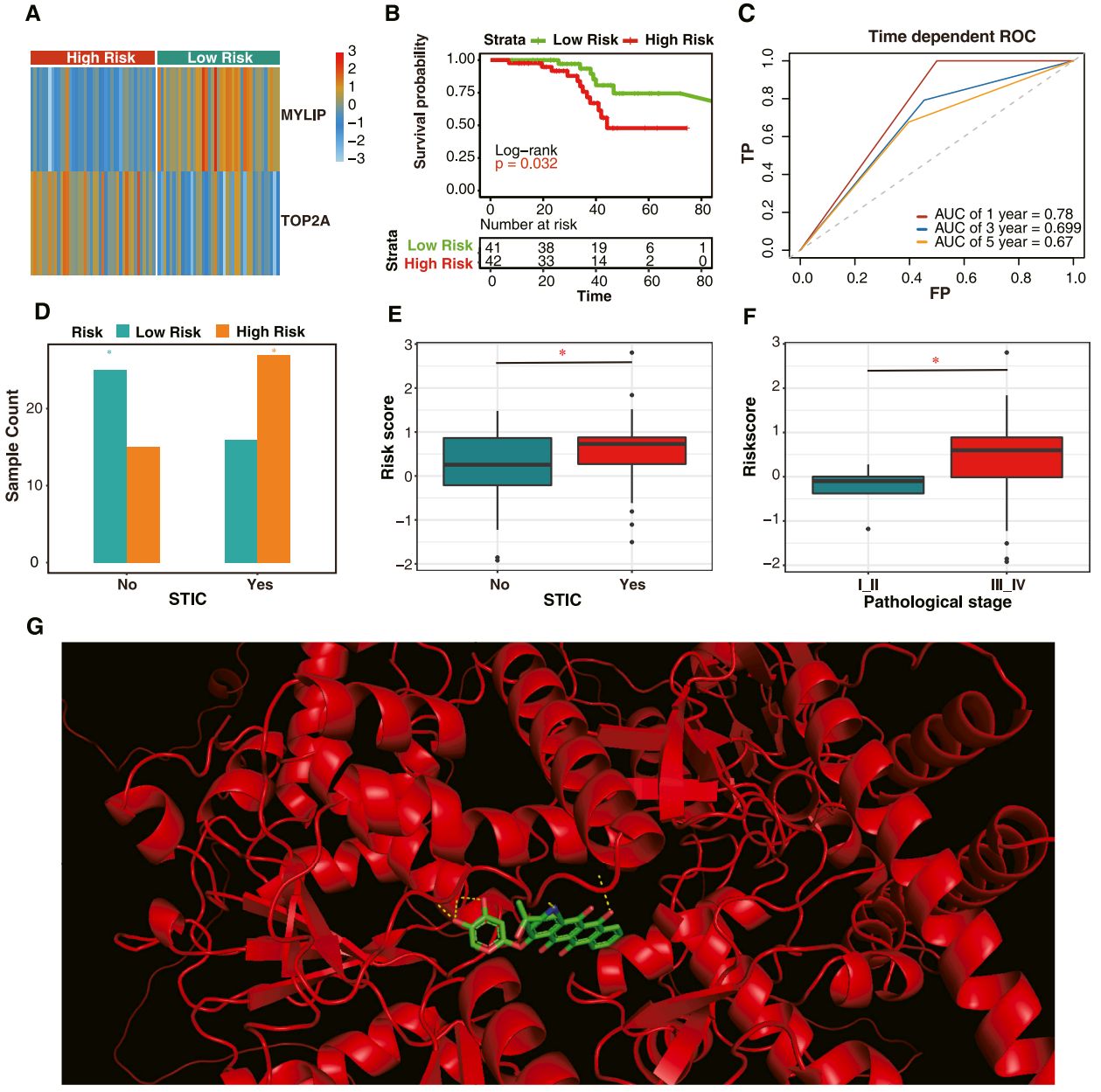


Fig. 5. Estimation of the survival risk model in the validation data set. A. Expression of model genes in high- and low-risk samples in the validation data set. B. Survival analysis between high- and low-risk samples in the validation data set. C. ROC plot showing the prediction of 1-, 3-, and 5-year survival outcomes in the validation data set. D. Significant enrichment of samples with STIC in high-risk samples. *:0.01< p < 0.05. E. Comparison of risk scores between patients with and without STIC. *:0.01<p < 0.05. F. Comparison of risk scores between patients with early-stage and late-stage ovarian cancer. *:0.01< p < 0.05. G. Interaction between AMRUBICIN and TOP2A. STIC: Serous tubal intraepithelial carcinoma lesions.

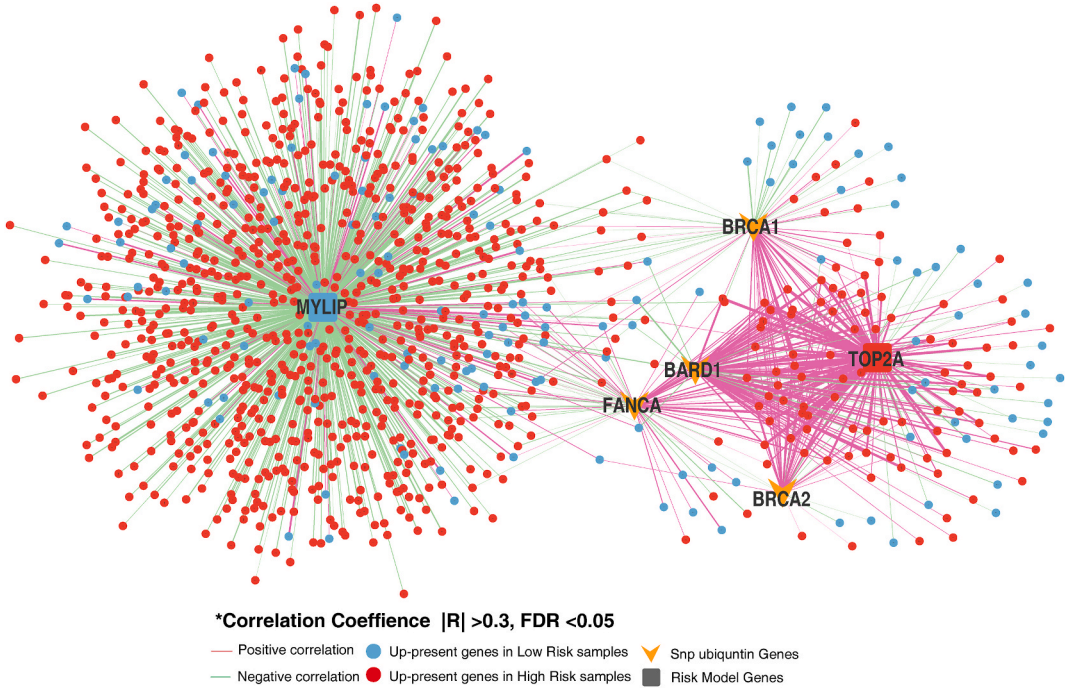


Fig. 6. Co-expression network among four ubiquitin genes, two model genes, and high- and low-risk DEGs. Co-expression network showed a coefficient of >0.3 and a BH-adjusted FDR of <0.05. BH, Benjamini–Hochberg; FDR, false discovery rate.

channels, channel activation, and KEGG pathways, including neuroactive ligand–receptor interaction and nicotine addiction. This finding suggests the deep involvement of neurohumoral regulation in the survival of patients with OC. A previous study reported that significant alterations in the expression of estrogen and progesterone receptors might play pivotal roles in OC development [42]. Neuroactive ligand–receptor interactions have also been reported to play critical roles in tumor development [43,44]. In relation to nicotine addiction, some studies reported that nicotine could suppress the proliferation and invasion of OC cells [45]. Furthermore, another study reported that smoking was significantly associated with invasive/borderline mucinous OC in another larger cohort [46]. Considering that our data were derived from epithelial samples, the nicotine addiction pathway might be associated with lower risk, suggesting a potential protective role.

We identified four ubiquitin marker genes from the SNP data, constructed a robust and reliable survival risk model with two ubiquitin genes, and explored the involvement of ubiquitin genes in OC survival risk. Despite the model incorporating only two genes, it demonstrated potential value as an indicator of early high-grade serious OC. As our results were derived from bioinformatics analysis, further validation of confidence is needed through experimental studies. Considering our risk model and the six identified genes, we hope that our results will supplement existing studies on OC and serve as a reference for the diagnosis and therapy of patients with OC.

All samples were collected in this study were with the informed consent and all methods were carried out in accordance with relevant guidelines and regulations.

Consent to publication

As no identifiable information (image, face, name etc.) of participants is provided, consent to publication is not applicable.

Data availability

The data associated with this study were presented in Supplementary Tables. And the called raw SNP data could be request to correspondence.

Funding

Medical Engineering Cross Research Fund of Shanghai Jiao Tong University (YG2019QNA69).

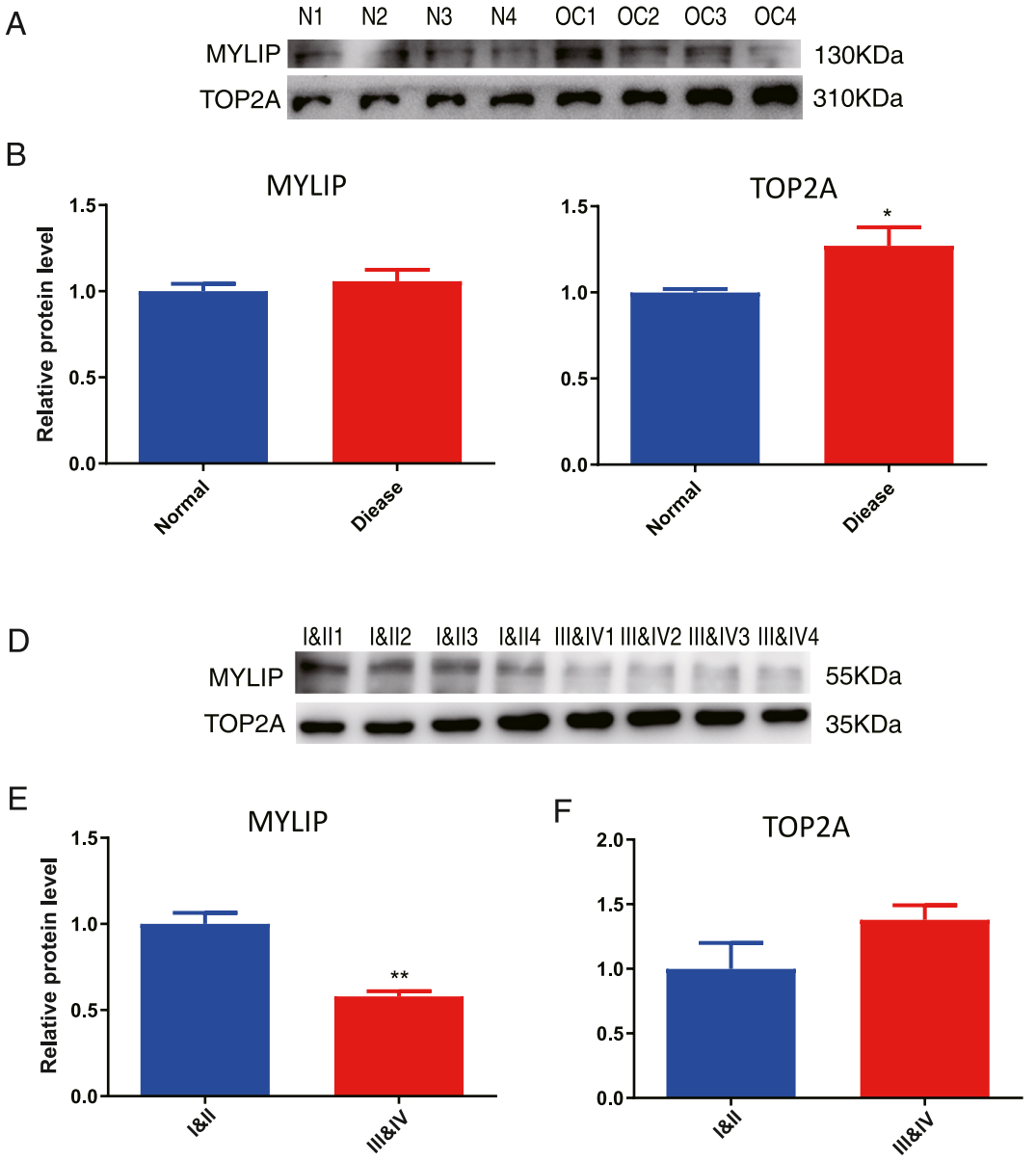


Fig. 7. In vivo validation of TOP2A and MYLIP. A. Expression patterns of MYLIP and TOP2A validated by Western-blot between tumor-adjacent normal tissue and tumor tissue. B. Grayscale value analysis of MYLIP using Image J between tumor-adjacent normal tissue and tumor tissue (T test). C. Grayscale value analysis of TOP2A using Image J between tumor-adjacent normal tissue and tumor tissue (T test). D. Expression patterns of MYLIP and TOP2A validated by Western-blot in tumor tissue from I &II stage patients and tumor tissue from III &IV stage patients. E. Grayscale value analysis of MYLIP using Image J between tumor tissue from I &II stage patients and tumor tissue from III &IV stage patients (T test). F. Grayscale value analysis of TOP2A using Image J between tumor tissue from I &II stage patients and tumor tissue from III &IV stage patients (T test). *:0.01< p < 0.05, **:0.001< p < 0.01.

Ethics declarations and ethical approval

This study was approved by the Ethics Committee of Shanghai General Hospital.

CRediT authorship contribution statement

Yiwen Feng: Writing – original draft. **Liyun Shan:** Resources. **Yanping Gong:** Funding acquisition. **Wenzhao Hang:** Resources. **Zhenyu Sang:** Resources. **Yunyan Sun:** Software. **Kefu Tang:** Investigation. **Yulan Wang:** Writing – review & editing. **Binjie Hu:** Writing – review & editing. **Xiaowei Xi:** Supervision, 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.

Acknowledgements

All authors have read the journal's authorship agreement. All authors have no question on reviewing and approving the final manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e37288.

References

[1] L.A. Torre, F. Bray, R.L. Siegel, J. Ferlay, J. Lortet-Tieulent, A. Jemal, Global cancer statistics, 2012, Ca - Cancer J. Clin. 65 (2) (2015) 87–108.
[2] R.S. Zheng, R. Chen, B.F. Han, S.M. Wang, L. Li, K.X. Sun, H.M. Zeng, W.W. Wei, J. He, [Cancer incidence and mortality in China, 2022], Zhonghua zhong liu za zhi [Chinese journal of oncology] 46 (3) (2024) 221–231.
[3] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, Ca - Cancer J. Clin. 69 (1) (2019) 7–34, 2019.
[4] E. Manrriquez, J.S. Chapman, J. Mak, A.M. Blanco, L.M. Chen, Disparities in genetics assessment for women with ovarian cancer: can we do better? Gynecol. Oncol. 149 (1) (2018) 84–88.
[5] P.D. Pharoah, B.A. Ponder, The genetics of ovarian cancer, Best Pract. Res. Clin. Obstet. Gynaecol. 16 (4) (2002) 449–468.
[6] S.P. Kar, A. Berchuck, S.A. Gayther, E.L. Goode, K.B. Moysich, C.L. Pearce, S.J. Ramus, J.M. Schildkraut, T.A. Sellers, P.D.P. Pharoah, Common genetic variation and susceptibility to ovarian cancer: current insights and future directions, Cancer Epidemiol. Biomark. Prev. : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 27 (4) (2018) 395–404.
[7] H. Song, M.S. Cicek, E. Dicks, P. Harrington, S.J. Ramus, J.M. Cunningham, B.L. Fridley, J.P. Tyrer, J. Alsop, M. Jimenez-Linan, et al., The contribution of deleterious germline mutations in BRCA1, BRCA2 and the mismatch repair genes to ovarian cancer in the population, Hum. Mol. Genet. 23 (17) (2014) 4703–4709.
[8] S.J. Ramus, H. Song, E. Dicks, J.P. Tyrer, A.N. Rosenthal, M.P. Intermaggio, L. Fraser, A. Gentry-Maharaj, J. Hayward, S. Philpott, et al., Germline mutations in the BRIP1, BARD1, PALB2, and NBN genes in women with ovarian cancer, J. Natl. Cancer Inst. (Bethesda) 107 (11) (2015).
[9] M. Earp, J.P. Tyrer, S.J. Winham, H.Y. Lin, G. Chornokur, J. Dennis, K.K.H. Aben, H. Anton-Culver, N. Antonenkova, E.V. Bandera, et al., Variants in genes encoding small GTPases and association with epithelial ovarian cancer susceptibility, PLoS One 13 (7) (2018) e0197561.
[10] T. Sun, Z. Liu, Q. Yang, The role of ubiquitination and deubiquitination in cancer metabolism, Mol. Cancer 19 (1) (2020) 146.
[11] L. Deng, T. Meng, L. Chen, W. Wei, P. Wang, The role of ubiquitination in tumorigenesis and targeted drug discovery, Signal Transduct. Targeted Ther. 5 (1) (2020) 11.
[12] H. Kim, J. Chen, X. Yu, Ubiquitin-binding protein RAP80 mediates BRCA1-dependent DNA damage response, Science (New York, NY) 316 (5828) (2007) 1202–1205.
[13] R. Hayami, K. Sato, W. Wu, T. Nishikawa, J. Hiroi, R. Ohtani-Kaneko, M. Fukuda, T. Ohta, Down-regulation of BRCA1-BARD1 ubiquitin ligase by CDK2, Cancer Res. 65 (1) (2005) 6–10.
[14] Z. Rao, Y. Ding, Ubiquitin pathway and ovarian cancer, Curr. Oncol. 19 (6) (2012) 324–328.
[15] D.W. Chan, V.W. Liu, G.S. Tsao, K.M. Yao, T. Furukawa, K.K. Chan, H.Y. Ngan, Loss of MKP3 mediated by oxidative stress enhances tumorigenicity and chemoresistance of ovarian cancer cells, Carcinogenesis 29 (9) (2008) 1742–1750.
[16] Q.Q. Li, M.K. Yunmbam, X. Zhong, J.J. Yu, E.G. Mimnaugh, L. Neckers, E. Reed, Lactacystin enhances cisplatin sensitivity in resistant human ovarian cancer cell lines via inhibition of DNA repair and ERCC-1 expression, in: Cellular and Molecular Biology, vol. 47, Online Pub, Noisy-le-Grand, France, 2001. Ol61-72.
[17] C.I. Amos, J. Dennis, Z. Wang, J. Byun, F.R. Schumacher, S.A. Gayther, G. Casey, D.J. Hunter, T.A. Sellers, S.B. Gruber, et al., The OncoArray consortium: a network for understanding the genetic architecture of common cancers, Oncology 26 (1) (2017) 126–135. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive*.
[18] H. Yang, K. Wang, Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR, Nat. Protoc. 10 (10) (2015) 1556–1566.
[19] S.E. Jamieson, M. Fakiola, D. Tang, E. Scaman, G. Syn, R.W. Francis, H.L. Coates, D. Anderson, T. Lassmann, H.J. Cordell, et al., Common and rare genetic variants that could contribute to severe otitis media in an Australian aboriginal population, Clin. Infect. Dis. 73 (10) (2021) 1860–1870.
[20] M. Kircher, D.M. Witten, P. Jain, B.J. O'Roak, G.M. Cooper, J. Shendure, A general framework for estimating the relative pathogenicity of human genetic variants, Nat. Genet. 46 (3) (2014) 310–315.
[21] A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J.P. Mesirov, P. Tamayo, The Molecular Signatures Database (MSigDB) hallmark gene set collection, Cell systems 1 (6) (2015) 417–425.
[22] B. Chen, M.S. Khodadoust, C.L. Liu, A.M. Newman, A.A. Alizadeh, Profiling tumor infiltrating immune cells with CIBERSORT, Methods Mol. Biol. 1711 (2018) 243–259.
[23] K. Yoshihara, M. Shahmoradgoli, E. Martínez, R. Vegesna, H. Kim, W. Torres-Garcia, V. Treviño, H. Shen, P.W. Laird, D.A. Levine, et al., Inferring tumour purity and stromal and immune cell admixture from expression data, Nat. Commun. 4 (1) (2013) 2612.
[24] S.L. Freshour, S. Kiwala, K.C. Cotto, A.C. Coffman, J.F. McMichael, J.J. Song, M. Griffith, L. Griffith Obi, A.H. Wagner, Integration of the drug–gene interaction database (DGIdb 4.0) with open crowdsource efforts, Nucleic Acids Res. 49 (D1) (2021) D1144–D1151.
[25] S. Forli, R. Huey, M.E. Pique, M.F. Sanner, D.S. Goodsell, A.J. Olson, Computational protein–ligand docking and virtual drug screening with the AutoDock suite, Nat. Protoc. 11 (5) (2016) 905–919.
[26] L.L.C. Schrodinger, The PyMOL Molecular Graphics System, 2015.
[27] M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, et al., AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models, Nucleic Acids Res. 50 (D1) (2022) D439–D444.
[28] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, et al., DrugBank 5.0: a major update to the DrugBank database for 2018, Nucleic Acids Res. 46 (D1) (2018) D1074–d1082.
[29] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (11) (2003) 2498–2504.
[30] H. Raskov, A. Orhan, J.P. Christensen, I. Gögenur, Cytotoxic CD8+ T cells in cancer and cancer immunotherapy, Br. J. Cancer 124 (2) (2021) 359–367.
[31] S. Sambasivan, Epithelial ovarian cancer: review article, Cancer Treatment and Research Communications 33 (2022) 100629.

[32] S.O. Katy, I.S. Kienan, The BRCA1 and BRCA2 breast and ovarian cancer susceptibility genes — implications for DNA damage response, DNA repair and cancer therapy, in: C.C. Clark (Ed.), Advances in DNA Repair, IntechOpen, Rijeka, 2015. Ch. 7.
[33] N.S. Abul-Husn, E.R. Soper, J.A. Odgis, S. Cullina, D. Bobo, A. Moscati, J.E. Rodriguez, R.J.F. Loos, J.H. Cho, G.M. Belbin, et al., Exome sequencing reveals a high prevalence of BRCA1 and BRCA2 founder variants in a diverse population-based biobank, Genome Med. 12 (1) (2019) 2.
[34] P.J. O'Donovan, D.M. Livingston, BRCA1 and BRCA2: breast/ovarian cancer susceptibility gene products and participants in DNA double-strand break repair, Carcinogenesis 31 (6) (2010) 961–967.
[35] Y.M. Hawsawi, A. Shams, A. Theyab, W.A. Abdali, N.A. Hussien, H.E. Alatwi, O.R. Alzahrani, A.A.A. Oyouni, A.O. Babalghith, M. Alreshidi, BARD1 mystery: tumor suppressors are cancer susceptibility genes, BMC Cancer 22 (1) (2022) 599.
[36] N.E. Mamrak, A. Shimamura, N.G. Howlett, Recent discoveries in the molecular pathogenesis of the inherited bone marrow failure syndrome Fanconi anemia, Blood Rev. 31 (3) (2017) 93–99.
[37] S. Seal, R. Barfoot, H. Jayatilake, P. Smith, A. Renwick, L. Bascombe, L. McGuffog, D.G. Evans, D. Eccles, D.F. Easton, et al., Evaluation of Fanconi Anemia genes in familial breast cancer predisposition, Cancer Res. 63 (24) (2003) 8596–8599.
[38] J. del Valle, P. Rofes, J.M. Moreno-Cabrera, A. López-Dóriga, S. Belhadj, G. Vargas-Parra, À. Teulé, R. Cuesta, X. Muñoz, O. Campos, et al., Exploring the role of mutations in fanconi anemia genes in hereditary cancer patients, in: Cancers, vol. 12, 2020.
[39] Y. Gao, H. Zhao, M. Ren, Q. Chen, J. Li, Z. Li, C. Yin, W. Yue, TOP2A promotes tumorigenesis of high-grade serous ovarian cancer by regulating the TGF-β/smad pathway, J. Cancer 11 (14) (2020) 4181–4192.
[40] W. Wang, F. Li, P. Gan, D. Su, G. Li, L. Dang, Y. Peng, The expression of myosin-regulated light chain interacting protein (MYLIP) in lung cancer and its inhibitory effects on lung carcinomas, Transl. Cancer Res. 10 (5) (2021) 2389–2398.
[41] L. Zhao, Y. Zhao, Y. He, Y. Mao, miR-19b promotes breast cancer metastasis through targeting MYLIP and its related cell adhesion molecules, Oncotarget 8 (38) (2017) 64330–64343.
[42] J. Escobar, A.C. Klimowicz, M. Dean, P. Chu, J.G. Nation, G.S. Nelson, P. Ghatage, S.E. Kalloger, M. Köbel, Quantification of ER/PR expression in ovarian low-grade serous carcinoma, Gynecol. Oncol. 128 (2) (2013) 371–376.
[43] X. Liu, J. Wang, G. Sun, Identification of key genes and pathways in renal cell carcinoma through expression profiling data, Kidney Blood Press. Res. 40 (3) (2015) 288–297.
[44] X. Wu, W. Zang, S. Cui, M. Wang, Bioinformatics analysis of two microarray gene-expression data sets to select lung adenocarcinoma marker genes, Eur. Rev. Med. Pharmacol. Sci. 16 (11) (2012) 1582–1587.
[45] S.J. Harmych, J. Kumar, M.E. Bouni, D.N. Chadee, Nicotine inhibits MAPK signaling and spheroid invasion in ovarian cancer cells, Exp. Cell Res. 394 (1) (2020) 112167.
[46] I. Licaj, B.K. Jacobsen, R.M. Selmer, G. Maskarinec, E. Weiderpass, I.T. Gram, Smoking and risk of ovarian cancer by histological subtypes: an analysis among 300 000 Norwegian women, Br. J. Cancer 116 (2) (2017) 270–276.