

Identification of Molecular Markers Associated With Ovarian Cancer Prognosis Using Bioinformatics Analysis

Li Chuang

Shengjing Hospital of China Medical University <https://orcid.org/0000-0002-6346-2857>

Yuan Lyu

Shengjing Hospital of China Medical University

Caixia Liu (✉ liucx1716@163.com)

<https://orcid.org/0000-0001-7019-0266>

Primary research

Keywords: bioinformatics analysis, differentially expressed genes, Gene Expression Omnibus database, ovarian cancer, PDZ-binding kinase

Posted Date: February 21st, 2020

DOI: <https://doi.org/10.21203/rs.2.24152/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background Ovarian cancer is associated with a high mortality rate worldwide. However, the pathogenesis, clinicopathological features, and genetic mechanisms of ovarian cancer are still unclear, and it is essential to identify prognostic markers for its clinical diagnosis and treatment. Here, we utilized bioinformatic analysis to identify potential genes involved in mediating BRCA1 expression to elucidate the potential mechanisms underlying ovarian cancer.

Methods Gene expression profiling (GSE14407) was performed to identify differentially expressed genes (DEGs) and analyze the weighted gene co-expression network. We selected the key module that was significantly associated with BRCA1 expression and performed gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses for genes in the hub module. We then screened the hub genes utilizing the Search Tool for the Retrieval of Interacting Genes Database (STRING) and Molecular Complex Detection (MCODE) plug-in in Cytoscape. We validated gene expression levels through The Cancer Genome Atlas and GTEx databases for hub genes and screened genes that were related to overall survival in patients with ovarian cancer using the Kaplan-Meier plotter database.

Results In total, 3124 DEGs were detected, including 433 upregulated genes and 2691 downregulated genes. We selected the brown module, which was the most significantly associated with BRCA1 expression. GO analysis showed that the hub module genes were significantly enriched in biological processes, including the mitotic cell cycle process, chromosome segregation, and cell division. KEGG analysis showed that the hub module genes were particularly enriched in the cell cycle, p53 signaling pathway, and small cell lung cancer. We selected 30 hub genes from the protein-protein interaction network, which had 88 nodes and 721 edges. Further analyses identified PBK as a prognosis-associated hub gene. Notably, PBK expression was significantly increased in ovarian cancer tissues, as demonstrated by immunohistochemistry analysis using samples from the Human Protein Atlas database.

Conclusion PBK was found to be associated with overall survival in patients with ovarian cancer. Our results provide insights into our understanding of the molecular mechanisms and molecular diagnosis of ovarian cancer.

Background

Ovarian cancer is a common gynecological tumor that is the fifth leading cause of death among women worldwide [1]. Notably, ovarian cancer is associated with a high mortality rate because patients experience no or nonspecific symptoms during the early stages of the disease, resulting in delayed diagnosis. More than 75% of patients with ovarian cancer are diagnosed at an advanced stage, limiting treatment efficacy [2, 3]. Therefore, early screening and diagnosis of patients with ovarian cancer is essential.

At least 5–10% of ovarian cancer cases exhibit a genetic predisposition to the disease, and most are carriers of breast cancer susceptibility gene (BRCA) 1 or BRCA2 mutations [4]. The risk of developing ovarian cancer varies considerably in women with BRCA1/2 mutations, i.e., 39–58.9% for BRCA1 and 11–34.5% for BRCA2. Thus, women with BRCA1 mutations have a higher risk of developing ovarian cancer [5–7]. However, previous studies have shown that cancer is caused by combinations of several genes and pathways, and not just by a single gene.

Accordingly, in this study, we obtained mRNA expression profiles from the Gene Expression Omnibus (GEO) and identified differently expressed genes (DEGs) in ovarian cancer. Furthermore, function annotations and signal pathway enrichment were performed to identify key module genes and construct protein-protein interaction networks for elucidation of novel biomarkers in ovarian cancer.

Materials And Methods

Identification of DEGs from the GEO database

We chose the ovarian cancer dataset GSE14407 from the GEO database. GSE14407, which was based on the GPL570 platform (HG-U133_Plus_2; Affymetrix Human Genome U133 Plus 2.0 Array), contained 12 paired normal ovarian epithelia and serous papillary ovarian cancer samples. We converted probe names to gene symbols according to the Whole Human Genome Microarray 4x44K G4112F (Probe Name Version). However, many probes can correspond to the same gene symbol, and we used the p value as the criteria for downstream DEG analysis. The probes with the most significant p value were chosen as DEGs. Then, we used the limma package in R [8-10] to identify DEGs in normal ovary tissues and paired tumor tissue. Genes that met the following criteria were chosen as DEGs: $|\log_2 \text{fold change}| \geq 1.0$ and adjusted p value < 0.05 .

Weighted gene co-expression network construction

Expression data for DEGs (3125 genes) were applied to identify co-expression gene modules constructed by weighted gene co-expression network analysis (WGCNA) [11]. We set the soft threshold power as 14, which was the lowest power based on scale-free topology [12]. A topological overlap matrix (TOM) was calculated by adjacency transformation, and the value $(1 - \text{TOM})$ was designated as the distance for identification of hierarchical clustering genes and modules. The minimum module size was set to 100.

Module clinical feature associations

In order to identify modules that were significantly associated with the designated clinical trait (*BRCA1* expression level), we selected the module with most significant p value from the heat map of the module/trait relationships for downstream analysis.

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses

GO analysis was used to identify the mechanisms affected by the module genes. KEGG pathway analysis was used to identify known metabolic pathways affected by the DEGs. We used the Database for Annotation, Visualization and Integrated Discovery (DAVID) website to perform the GO function and KEGG pathway enrichment analyses of DEGs; results with p values of less than 0.05 were considered significant [13,14]. We plotted the results using a bar map for the top 10 significant biological processes, cellular components, and molecular functions using the ggplot2.R package based on the p value. The top five significant pathways were visualized using the GOplot.R package.

Establishment of a protein-protein interaction (PPI) network

We used the Search Tool for the Retrieval of Interacting Genes (STRING) to identify key proteins and construct PPI networks [15]. We uploaded the module genes to the STRING database and set the threshold for the interaction score at greater than or equal to 0.9 (the highest confidence). Additionally, we used Cytoscape software to visualize the PPI networks [16,17] and to analyze key modules constructed by the Molecular Complex Detection (MCODE) function. The k score ≥ 10 was taken as the criteria to define a module.

Validation of candidate genes

Hub genes from the MCODE analysis were chosen as potential biomarkers for deep analysis and validation. We validate the expression levels of hub genes via the GEPIA website (<http://gepia.cancer-pku.cn/>) using the following criteria as the cutoff: $p < 0.01$ and $|\log_2FC| \geq 1$ [18]. We further carried out overall survival analysis for hub genes using the Kaplan-Meier plotter database (<http://kmplot.com>).

Oncomine database extraction

The Oncomine database (<http://www.oncomine.org>) [19] is the world's largest oncogene-chip database integrated data-mining platform and contains 715 gene expression datasets and 86,733 cancer and paired normal sample sequencing data. Here, the Oncomine database was applied for analysis of DEGs and clinicopathological features for common cancer types.

The Human Protein Atlas analysis

The Human Protein Atlas database (<https://www.proteinatlas.org>) [20] provides information on the tissue and cellular distribution of 24,000 human proteins based on immunohistochemical analysis. Here, we used the Human Protein Atlas database to evaluate the proteome of patients with ovarian cancer.

Bioinformatics analysis

Ovarian cancer methylation data were downloaded from Xena (<https://xenabrowser.net/datapages/>), and the correlations between methylation levels and PDZ binding kinase (*PBK*) gene expression levels were tested by Pearson correlation analysis. Somatic cell mutations were investigated both in cases with low *PBK* expression and cases with high *PBK* expression. The ovarian cancer somatic cell mutation data

were downloaded from The Cancer Genome Atlas (TCGA). We analyzed the correlations between *PBK* expression levels and copy number variations from cBioPortal (<http://www.cbioportal.org/>). To further identify the molecular mechanism of PBK, we investigate the

Results

Identification of DEGs

In the study, we carried out DEG analysis of normal ovary tissue and paired tumor tissues from the GSE14407 dataset. We identified upregulated and downregulated genes between the two sample types and found a total of 3124 genes that met the cutoff criteria (433 upregulated genes and 2691 downregulated genes; Figure 1A). The DEG heatmap is shown in Figure 1B.

WGCNA

The expression values of 3124 DEGs from the 12 paired normal and cancer tissues were used to construct the co-expression module using the WGCNA tool. One of the most important steps was choosing the appropriate soft-thresholding power that affected the independence and average connectivity of gene co-expression modules. When the power value was selected as 14, the independence was equal to 0.8, and the average connectivity degree of modules was higher (Figure 2A). As shown in Figure 2B, a power value of 14 was selected to produce different co-expression modules.

Modules corresponding to clinical traits

We identified the co-expression modules with the most significant associations based on correlations between the module eigengene and extra traits (e.g., *BRCA1* expression). The result indicated that the brown module (183 genes) was the most positively associated with *BRCA1* expression (Figure 3A). Additionally, we found that the brown modules and *BRCA1* expression were highly related according to the results of hierarchical clustering of eigengenes and the heatmap plot of eigengene adjacencies for the expression of *BRCA1* (Figure 3B). Scatterplot analysis of gene significance versus module membership is shown for the brown modules in Figure 3C.

GO and KEGG pathway enrichment analysis

To improve our understanding of the brown module genes involved in the molecular mechanisms of ovarian cancer, we carried out GO function and KEGG pathway enrichment analysis. GO analysis revealed that the module genes were enriched in biological processes, including GO:1903047 (mitotic cell cycle process), GO:0000278 (mitotic cell cycle), GO:0022402 (cell cycle process), GO:0007059 (chromosome segregation), and GO:0051301 (cell division), as shown in Figure 4A. The module genes were enriched in cellular components, including GO:0005819 (spindle), GO:0000779 (condensed chromosome, centromeric region), GO:0000775 (chromosome, centromeric region), GO:0000777 (condensed chromosome kinetochore), and GO:0000776 (kinetochore), as shown in Figure 4B. The modules genes were also enriched in molecular functions, including GO:0008017 (microtubule binding), GO:0097367

(carbohydrate derivative binding), GO:0015631 (tubulin binding), GO:0035639 (purine ribonucleoside triphosphate binding), and GO:0032550 (purine ribonucleoside binding), as shown in Figure 4C. For pathway enrichment analysis, the module genes were enriched for genes involved in the cell cycle and p53 signaling pathway (Figure 4D).

PPI network construction

After PPI network construction, we found that the network (score > 0.900) had 88 nodes and 721 edges (Figure 5A). The most significant module consisted of 30 nodes and 398 edges (Figure 5B). This result suggested that the identified 30 hub genes could play critical roles in ovarian cancer.

Validation of hub genes

In order to validate the gene expression levels of a large number of samples to assess whether gene expression was concordant between samples from the GSE14407 and OV (TCGA/GTEX) datasets, we uploaded genes in the module to the GEPIA website. We found that the expression levels of 28 of 30 genes in the OV dataset (from TCGA/GTEX) were consistent with gene expression changes in the GSE14407 dataset. The gene expression plots of these genes are shown in Figure S1. We then further explored whether these genes were involved in overall survival; the survival prognosis forest map is shown in Figure 6, and the survival curves related to these genes are shown in Figure S2.

Analysis of the *PBK* gene in the OncoPrint and Human Protein Atlas databases

In this study, OncoPrint analysis revealed that *PBK* expression was significantly different in 66 tumor studies (upregulated in 55 studies, and downregulated in 11 studies; Figure 7A). Additionally, we further showed that *PBK* expression was upregulated in all ovarian cancer samples (Figure 7B-D). To increase the credibility of the results, we obtained immunohistochemistry data from the Human Protein Atlas database (Figure 8A-B). Despite the limitations of the use of database results, our findings showed that *PBK* expression differed between tumor and normal samples.

Molecular mechanisms of *PBK* expression in ovarian cancer

The methylation data were downloaded from Xena (<https://xenabrowser.net/datapages/>). The results revealed that the *PBK* expression level was negatively correlated with the methylation level and that hypomethylation resulted in high expression of *PBK* (Figure 9A). Somatic cell mutation data were downloaded from TCGA database, and patients were divided into high and low expression groups according to the median expression of *PBK*. In both high expression and low expression groups, *TP53*, *CSMD3*, and *TTN* mutations were significantly enriched (Figure 9B-C). We investigated the association between copy number variations and *PBK* expression levels according to the cBioportal database (<https://www.cbioportal.org/>). According to copy number variations, the samples were divided into four groups: deep deletion, shallow deletion, gain, and amplification. The results showed that deep deletion and shallow deletion of *PBK* were associated with low expression levels, whereas gain and amplification were associated with high expression levels (Figure 9D). Additionally, GO function and KEGG pathway

analyses revealed that genes associated with PBK expression were involved in biological processes, such as cell cycle process, nuclear division, DNA metabolic process, and DNA replication (Figure 9E), and molecular functions, such as ATP binding, DNA-dependent ATPase activity, and DNA binding (Figure 9F). The results of KEGG analysis indicated that the pathways were mainly involved in the p53 signaling pathway, small cell lung cancer, prostate cancer (Figure 9G).

Discussion

Despite advances in medical technology in recent years, the mortality rate of patients with ovarian cancer has not significantly changed, and ovarian cancer remains the leading cause of gynecological tumor-related deaths. Because ovarian cancer tends to metastasize and exhibits tumor heterogeneity, most patients are diagnosed only at an advanced stage. Therefore, identification of more reliable biomarkers and exploration of the underlying molecular mechanisms of ovarian cancer will improve the overall survival and prognosis of patients with ovarian cancer.

BRCA1 is a tumor-suppressor gene that contributes to DNA repair and transcriptional regulation in response to DNA damage. Carriers with mutations in the BRCA1 gene are at high risk of breast cancer, ovarian cancer, pancreatic cancer, and prostatic cancer [21]. The specific mechanism through which mutations in BRCA1 can result in increased susceptibility to ovarian cancer are still unclear, although it is expected that many genes may play important roles in this process.

In this study, we identified 3124 DEGs between normal tissues and ovarian cancer tissues and showed that the genes in modules identified by WGCNA were involved in specific biological process, cellular components, and molecular functions. Further pathway enrichment analysis showed that the module genes were enriched in cell cycle and the p53 signaling pathway. From the results of enrichment analysis, we found that these genes were involved in regulating the biological behaviors of ovarian cancer cells and affected the occurrence and development of ovarian cancer.

To further find key genes that played important roles in the occurrence and development of ovarian cancer, we construct PPI networks and screened the key networks. From this analysis, we identified PBK as contributing the overall survival in patients with ovarian cancer. PBK is a serine-threonine kinase that regulates cell cycle-related processes, including cell growth, immune responses, and DNA damage repair, and can accelerate tumorigenesis in various types of cancer, including gastric cancer, oral cancer, prostate cancer, and cervical cancer [22]. Subsequent analysis showed that PBK was strongly associated with ovarian cancer and that the expression of PBK in ovarian cancer tissues was significantly higher than that in normal tissues. Overall, these findings indicated that PBK may contribute to clinical prognosis in patients with ovarian cancer.

PBK has been studied as potential biomarker in various types of cancers, including gastric cancer, colorectal cancer, lung adenocarcinoma, hepatocellular carcinoma, and breast cancer. Overexpression of PBK protein is frequently detected in gastric cell lines and primary gastric tumor samples and is related to gastric tumor occurrence, metastasis, and invasion. Ohashi et al. found that knockdown of PBK in PBK-

overexpressing gastric cell lines inhibits cell proliferation by activating p53 in a TP53 mutation-dependent manner and inhibits cell migration/invasion by upregulating phosphatase and tensin homolog (PTEN) in a TP53 mutation-independent manner [23]. Additionally, Su et al. found that the 5-year survival rate was higher in patients with high PBK expression than in patients with low PBK expression in a study of colorectal cancer [22]. Lei et al. also reported that PBK overexpression affected the survival times of patients with lung adenocarcinoma [24], and Shih et al. found that PBK promoted cell migration by activating the phosphatidylinositol 3-kinase/PTEN/AKT signaling pathway in lung cancer [25]. Moreover, Yang et al. showed that PBK can be a potential diagnostic biomarker and therapeutic target for patients with hepatocellular carcinoma and can promote cell migration and invasion by activating the ETV4-uPAR signaling pathway [26]. Finally, Park et al. found that PBK acted as a cancer/testis antigen with oncogenic activity in breast cancer [27].

To further explore the molecular mechanisms underlying the action of PBK in ovarian cancer, we analyzed somatic cell mutation data and found that TP53, CSMD3, and TTN mutations were significantly enriched in both high and low expression groups. Additionally, the expression levels of PBK were found to be positively associated with copy number variations, suggesting that samples with copy number amplification may have high PBK expression. DNA methylation is a critical type of DNA modification occurring at the epigenetic level and is modulated by DNA methyltransferases [28]. During cancer development, abnormal methylation often occurs, resulting in abnormal expression of multiple key genes [29]. In this study, hypomethylation of PBK resulted in higher expression of PBK and could be related to poor prognosis in patients with ovarian cancer, suggesting that PBK may act as a biomarker and therapeutic target in ovarian cancer. To identify the molecular function of PBK expression in ovarian cancer, the first 500 genes significantly associated with PBK were selected from the cBioportal database for further analysis. The GO function analysis indicated that these genes were involved in cell cycle process, nuclear division, DNA metabolic process, and DNA replication. The molecular functions of these genes were mainly associated with ATP binding, DNA-dependent ATPase activity, and DNA binding. The pathways analysis were related to the p53 signaling pathway, small cell lung cancer, prostate cancer. However, further studies are necessary to fully elucidate the molecular functions of PBK in ovarian cancer.

Conclusion

In summary, we used DEG analysis and WGCNA to identify key module genes closely associated with clinical traits (e.g., BRCA1 expression). Further analysis showed that the PBK gene was associated with prognosis in patients with ovarian cancer. Bioinformatics analysis was used to elucidate the potential molecular mechanisms through which PBK contributed to ovarian cancer. In the future, more in-depth studies are needed to confirm association between PBK expression and ovarian cancer.

Abbreviations

DEGs: Differentially expressed genes; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; WGCNA: Weighted Gene Co-expression Network Analysis; PBK: PDZ-binding kinase; *BRCA1*: breast cancer susceptibility gene; TCGA: The Cancer Genome Atlas;

Declarations

Acknowledgment

Not applicable.

Authors' contributions

The idea of the paper was devised and the first draft was written by the author Li Chuang. All authors read and approved the final manuscript.

Funding

Not applicable.

Ethics approval and consent to participate

Not applicable.

Availability of data and materials

The datasets used and analyzed in the study are available from the corresponding author upon reasonable request.

Consent for publication

Not applicable.

Competing Interests

The authors have declared that no competing interest exists.

References

[1] Siegel RL, Miller KD, Jemal A. Cancer statistics. *CA Cancer J Clin.* 2017; 67(1): 7–30. doi: 10.3322/caac.21387.

[2] Emmings E, Mullany S, Chang S, Landen CN Jr, Linder S, Bazzaro M. Targeting Mitochondria for Treatment of Chemoresistant ovarian Cancer. *Int J Mol Sci.* 2019; 20(1): E229. doi: 10.3390/ijms20010229.

- [3] Elzek MA, Rodland KD. Proteomics of ovarian cancer: functional insights and clinical applications. *Cancer Metastasis Rev.* 2015; 34(1): 83-96. doi: 10.1007/s10555-014-9547-8.
- [4] Claus EB, Schildkraut JM, Thompson WD, Risch NJ. The genetic attributable risk of breast and ovarian cancer. *Cancer.* 1996; 77(11): 2318–2324. doi: 10.1002/(SICI)1097-0142(19960601)77:11<2318::AID-CNCR21>3.0.CO;2-Z
- [5] van der Kolk DM, de Bock GH, Leegte BK, Schaapveld M, Mourits MJ, de Vries J, van der Hout AH, Oosterwijk JC. Penetrance of breast cancer, ovarian cancer and contralateral breast cancer in BRCA1 and BRCA2 families: high cancer incidence at older age. *Breast Cancer Res Treat.* 2010; 124(3): 643–651. doi: 10.1007/s10549-010-0805-3.
- [6] Chen S, Parmigiani G. Meta-analysis of BRCA1 and BRCA2 penetrance. *J Clin Oncol.* 2007; 25(11): 1329–1333. doi: 10.1200/JCO.2006.09.1066
- [7] Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, Loman N, Olsson H, Johannsson O, Borg A, Pasini B, Radice P, Manoukian S, Eccles DM, Tang N, Olah E, Anton-Culver H, Warner E, Lubinski J, Gronwald J, Gorski B, Tulinius H, Thorlacius S, Eerola H, Nevanlinna H, Syrjäkoski K, Kallioniemi OP, Thompson D, Evans C, Peto J, Lalloo F, Evans DG, Easton DF. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. *Am. J. Hum. Genet.* 2003; 72(5): 1117–1130. doi: 10.1086/375033
- [8] Dessau RB, Phipps CB. "R"-project for statistical computing. *Ugeskr Laeger.* 2008; 170(5): 328-330.
- [9] Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol.* 2004; 3: Article3. doi: 10.2202/1544-6115.1027
- [10] Smyth GK, Speed T. Normalization of cDNA microarray data. *Methods.* 2003; 31(4): 265–273. doi: 10.1016/s1046-2023(03)00155-5
- [11] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008; 9: 559. doi: 10.1186/1471-2105-9-559.
- [12] Zhang B, Horvath S. A General Framework For Weighted Gene Co-Expression Network Analysis. *Stat Appl Genet Mol Biol.* 2005; 4:Article17. doi: 10.2202/1544-6115.1128
- [13] Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009; 37(1): 1-13. doi: 10.1093/nar/gkn923.
- [14] Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc.* 2009; 4(1): 44-57. doi: 10.1038/nprot.2008.211.

- [15] Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 2013; 41 (Database issue): D808–D815. doi: 10.1093/nar/gks1094.
- [16] Kohl M, Wiese S, Warscheid B. Cytoscape: software for visualization and analysis of biological networks. *Data mining in proteomics.* Humana Press. 2011; 696: 291–303.
- [17] Kohl M, Wiese S, Warscheid B. Cytoscape: software for visualization and analysis of biological networks. *Methods Mol Biol.* 2011; 696: 291-303. doi: 10.1007/978-1-60761-987-1_18.
- [18] Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 2017; 45(W1): W98–W102. doi: 10.1093/nar/gkx247.
- [19] Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincead-Beal C, Kulkarni P, Varambally S, Ghosh D, Chinnaiyan AM.. Oncomine 3.0: Genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia.* 2007; 9(2): 166–180. doi: 10.1593/neo.07112
- [20] Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Pontén F. Proteomics. Tissue-based map of the human proteome. *Science.* 2015; 347(6220):1260419. doi: 10.1126/science.1260419.
- [21] Yoshida K, Miki Y. Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. *Cancer Sci.* 2004; 95(11): 866-71. doi: 10.1111/j.1349-7006.2004.tb02195.x
- [22] Ohashi T, Komatsu S, Ichikawa D, Miyamae M, Okajima W, Imamura T, Kiuchi J, Kosuga T, Konishi H, Shiozaki A, Fujiwara H, Okamoto K, Tsuda H, Otsuji E. Overexpression of PBK/TOPK relates to tumor malignant potential and poor outcome of gastric carcinoma. *Br J Cancer.* 2017; 116(2): 218-226. doi: 10.1038/bjc.2016.394.
- [23] Su TC, Chen CY, Tsai WC, Hsu HT, Yen HH, Sung WW, Chen CJ. Cytoplasmic, nuclear, and total PBK/TOPK expression is associated with prognosis in colorectal cancer patients: A retrospective analysis based on immunohistochemistry stain of tissue microarrays. *PLoS One.* 2018; 13(10): e0204866. doi: 10.1371/journal.pone.0204866.
- [24] Lei B, Qi W, Zhao Y, Li Y, Liu S, Xu X, Zhi C, Wan L, Shen H. PBK/TOPK expression correlates with mutant p53 and affects patients' prognosis and cell proliferation and viability in lung adenocarcinoma. *Hum Pathol.* 2015; 46(2): 217-24. doi: 10.1016/j.humpath.2014.07.026.

[25] Shih MC, Chen JY, Wu YC, Jan YH, Yang BM, Lu PJ, Cheng HC, Huang MS, Yang CJ, Hsiao M, Lai JM. TOPK/PBK promotes cell migration via modulation of the PI3K/PTEN/AKT pathway and is associated with poor prognosis in lung cancer. *Oncogene*. 2012; 31(19): 2389-400. doi: 10.1038/onc.2011.419.

[26] Yang QX, Zhong S, He L, Jia XJ, Tang H, Cheng ST, Ren JH, Yu HB, Zhou L, Zhou HZ, Ren F, Hu ZW, Gong R, Huang AL, Chen J. PBK overexpression promotes metastasis of hepatocellular carcinoma via activating ETV4-uPAR signaling pathway. *Cancer Lett*. 2019; 452: 90-102. doi: 10.1016/j.canlet.2019.03.028.

[27] Park JH, Lin ML, Nishidate T, Nakamura Y, Katagiri T. PDZ-binding kinase/T-LAK cell-originated protein kinase, a putative cancer/testis antigen with an oncogenic activity in breast cancer. *Cancer Res*. 2006; 66(18): 9186-95. doi: 10.1158/0008-5472.CAN-06-1601

[28] Fernandez AF, Huidobro C, Fraga MF. De novo DNA methyltransferases: oncogenes, tumor suppressors, or both? *Trends Genet*. 2012; 28(10): 474-9. doi: 10.1016/j.tig.2012.05.006.

[29] Wang YP, Lei QY. Metabolic recoding of epigenetics in cancer. *Cancer Communications(Lond)*. 2018; 38(1): 25. doi: 10.1186/s40880-018-0302-3.

Figures

Figure 1A

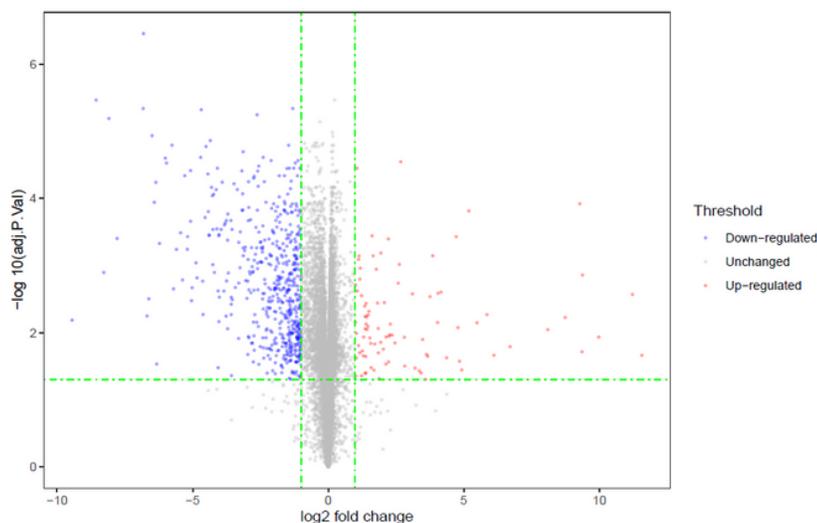


Figure 1B

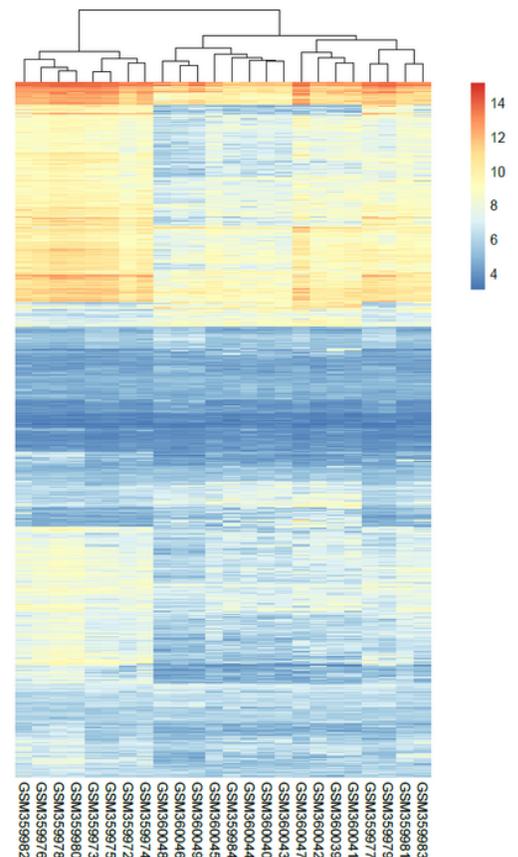


Figure 1

Differentially expressed genes (DEGs) form the dataset GSE14407. (A) DEGs volcano plot. The red indicates up-regulated genes. The blue indicated down-regulated genes. The grey indicates un-changed genes. (B) DEGs expression heatmap. The ordinate is the sample identification; is the gene name.

Figure 2A

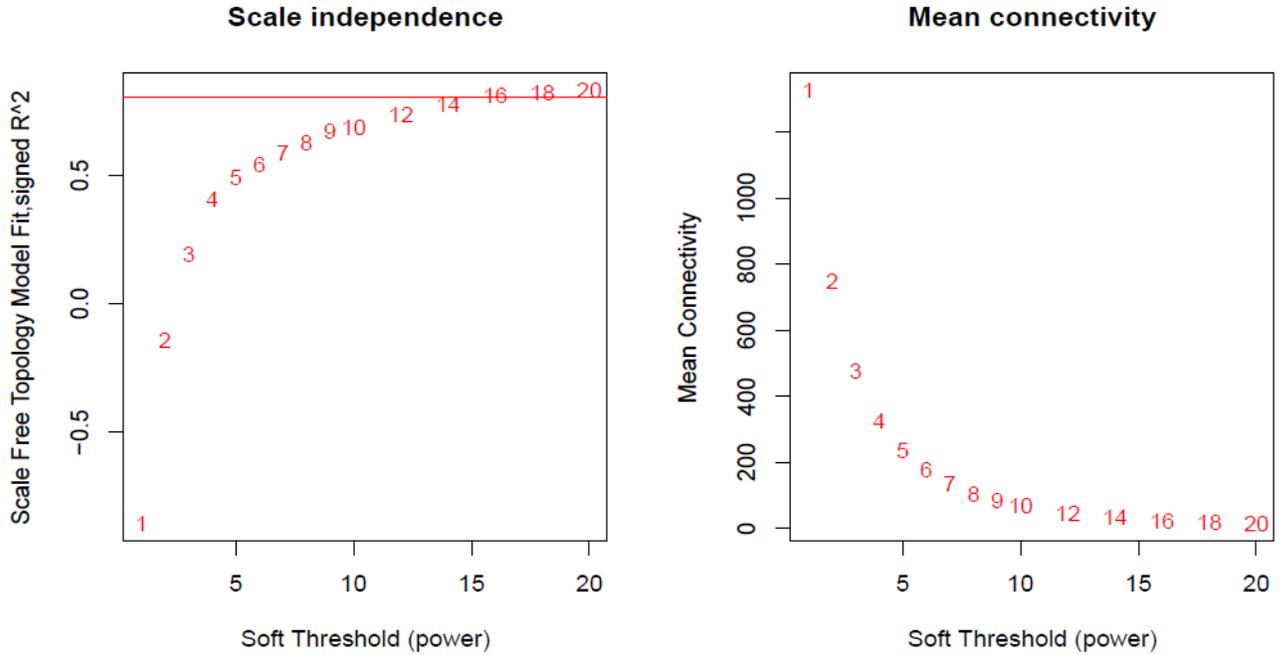


Figure 2B

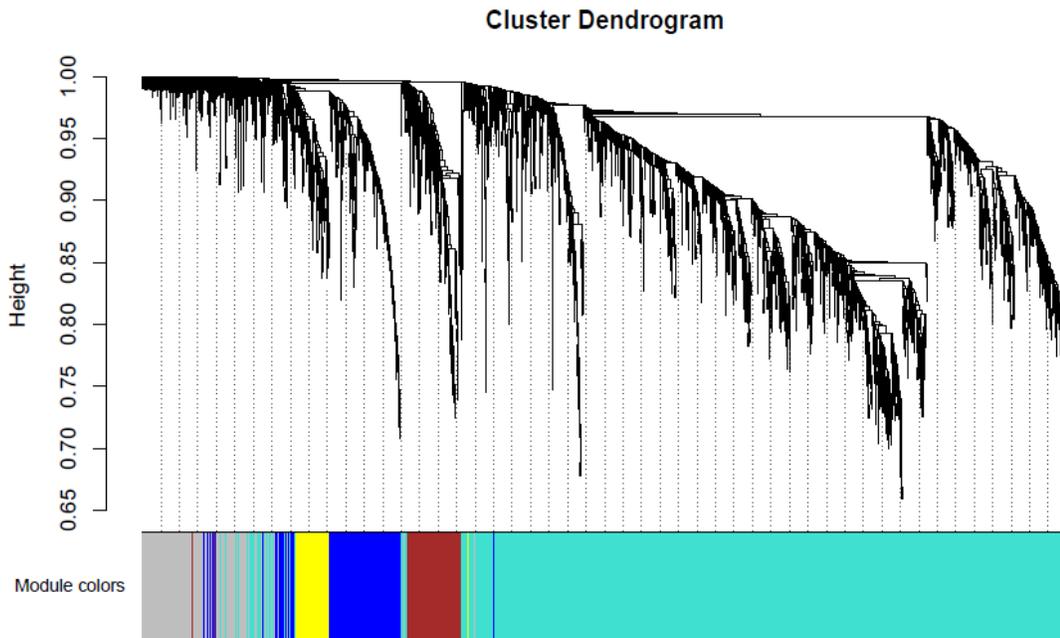


Figure 2

The clustering of samples and selection of soft-thresholding power (A) Analysis of the scale-free fit index for various soft threshold powers (B) The clustering dendrogram of genes, every color below represents one co-expression module

Figure 3A

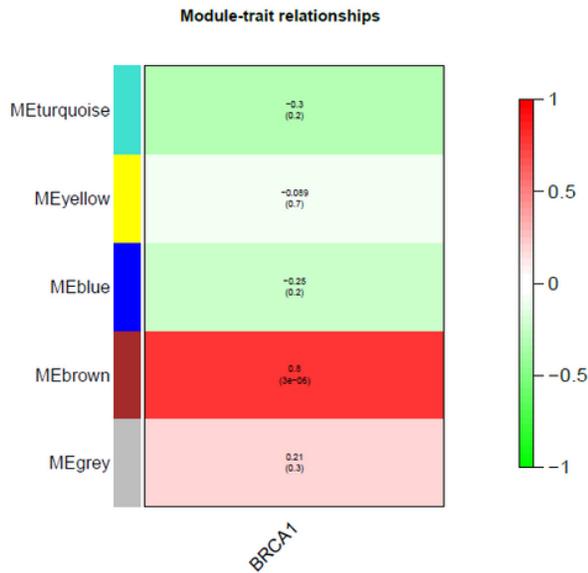


Figure 3B

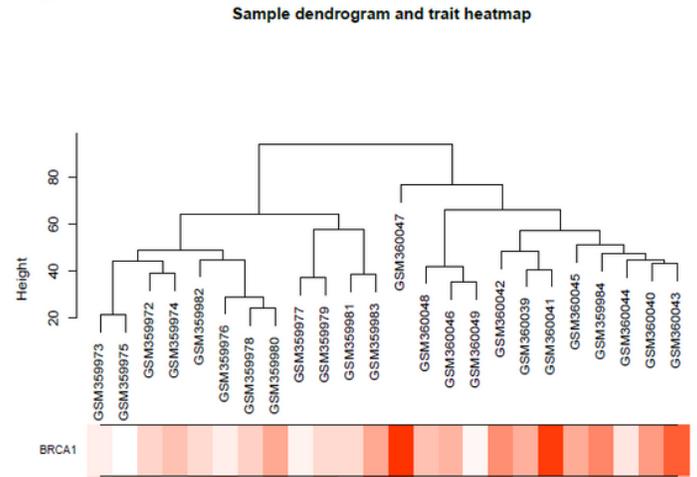


Figure 3C

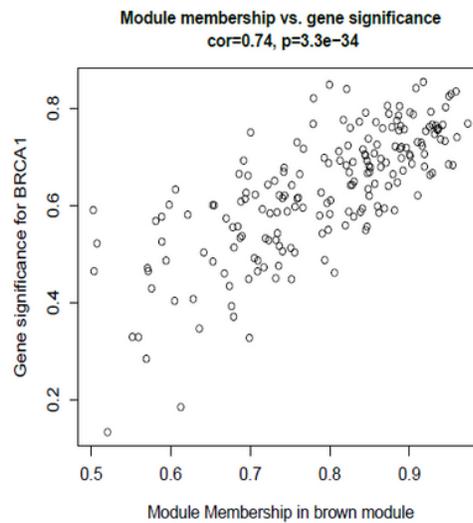


Figure 3

Heatmap of module-trait relationships. (A) The MEbrown module was the most positive module correlated with the expression of BRCA1 (B) Hierarchical clustering of module and heatmap plot of the adjacencies of modules. The red indicates high adjacency. The blue indicates low adjacency. (C) Scatterplot of Gene Significance vs. Module Membership in the brown modules

Figure 4A

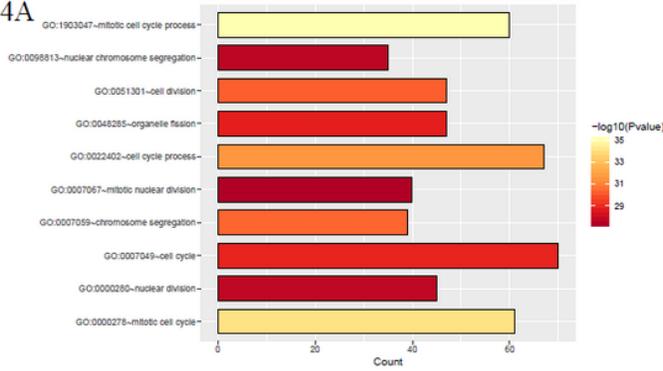


Figure 4B

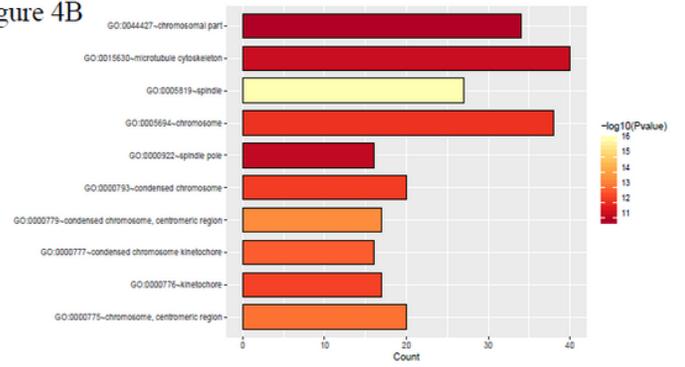


Figure 4C

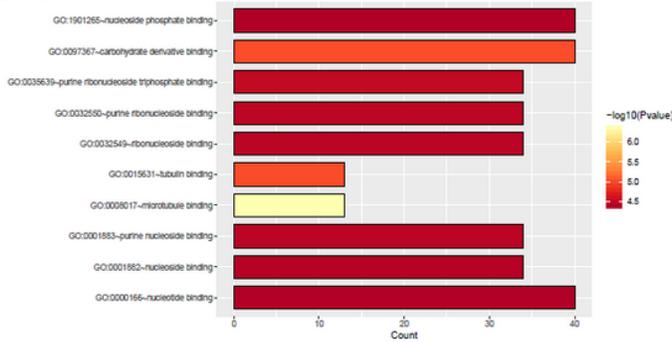


Figure 4D

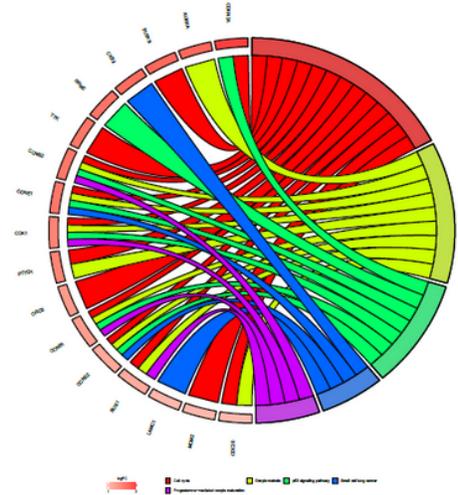


Figure 4

The plot of GO function and KEGG pathway enrichment (A) Top 10 biological processes enriched in gene ontology terms (B) Top 10 cellular component enriched in gene ontology terms (C) Top 10 molecular function enriched in gene ontology terms (d) Distribution pattern of DEGs in different pathways

Figure 5A

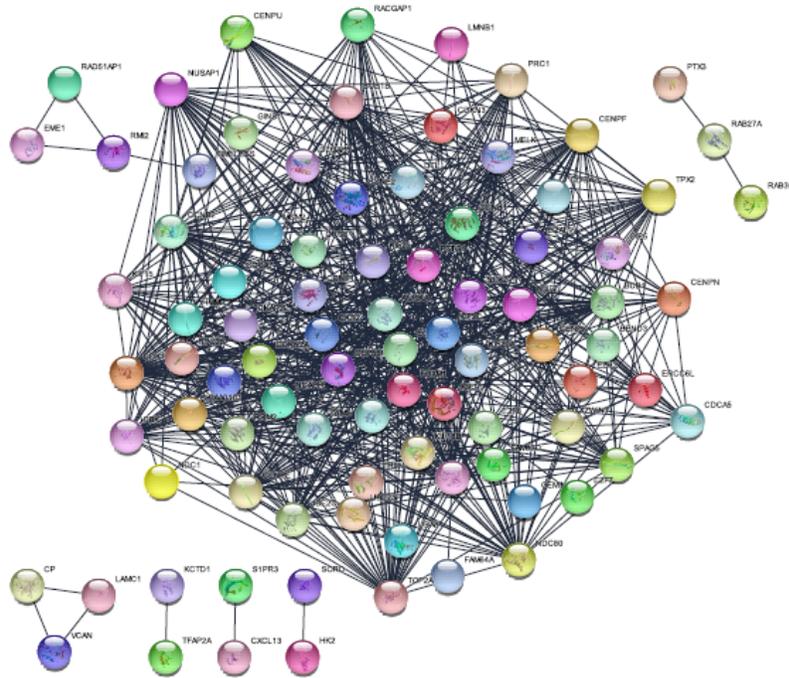


Figure 5B

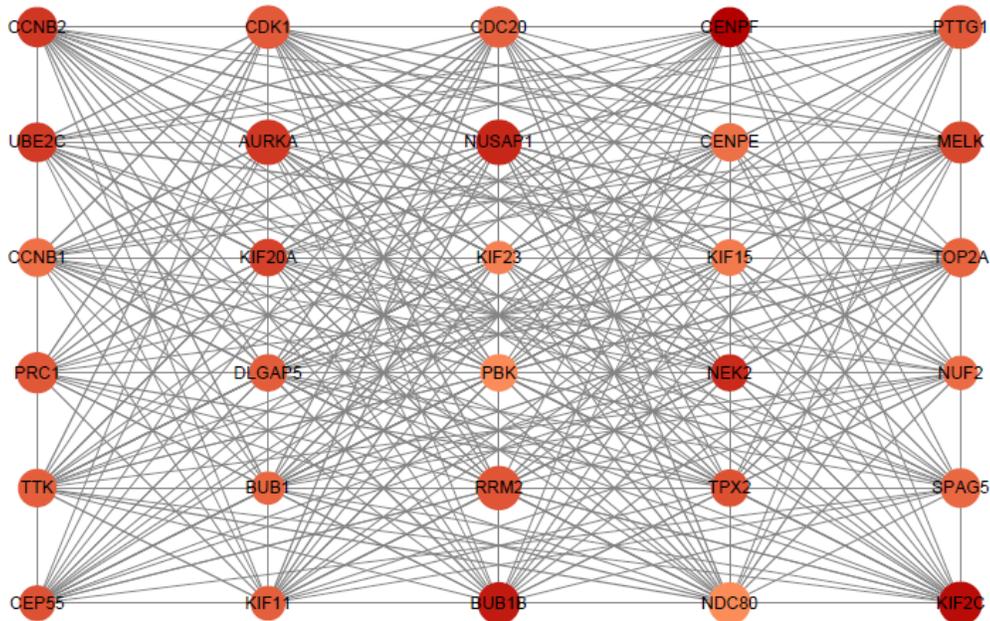


Figure 5

The protein–protein interaction network and key module selection(A) PPI network of DEGs according to the interaction score ≥ 0.9 (B) Each node indicates a protein and the edge between the nodes indicates the interaction between the two proteins. Different thicknesses of the edges indicates the strengths of the relationships. Red nodes indicate the upregulated genes; The smaller the p value is after correction, the more significant the difference.

Hazard Ratio

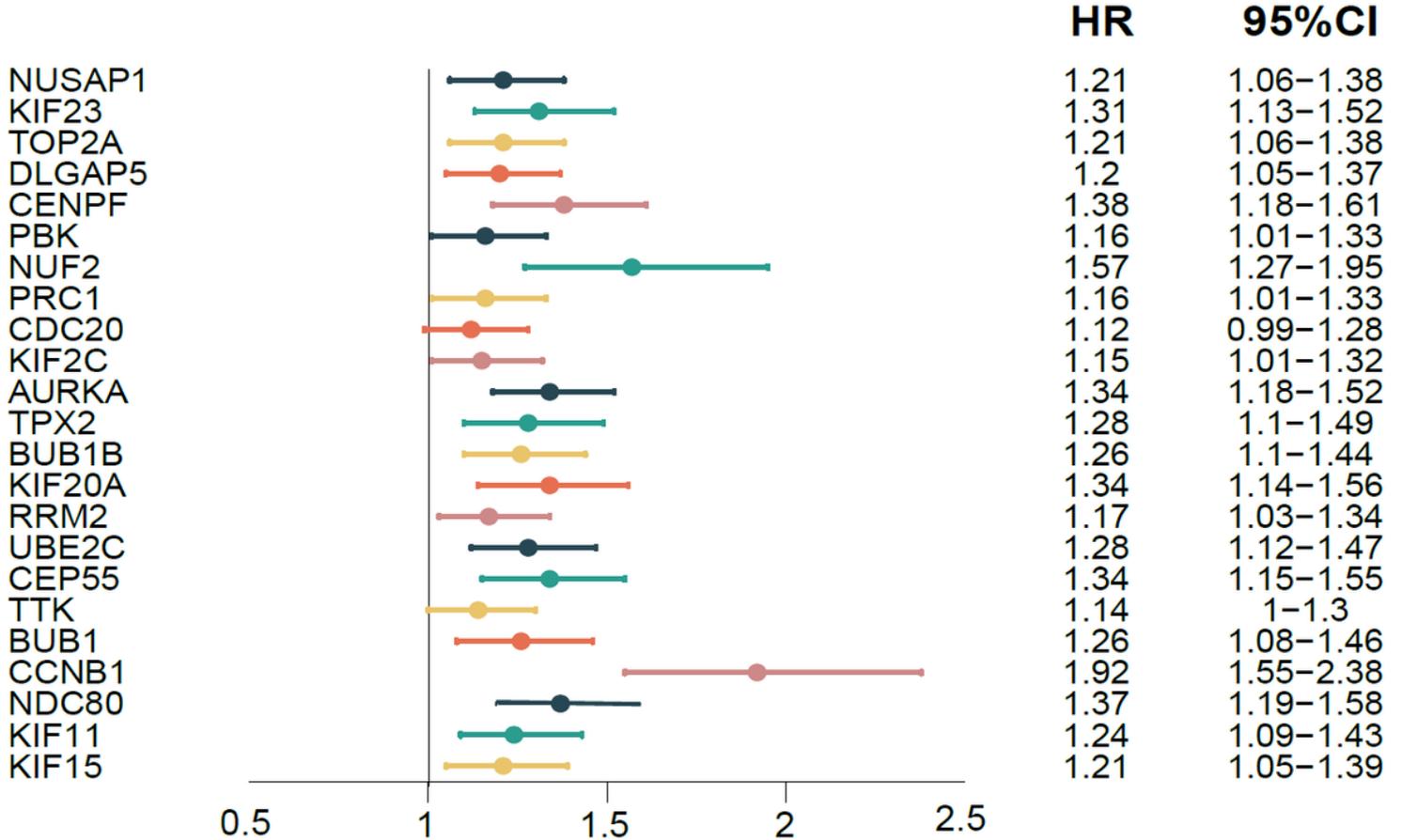


Figure 6

Survival prognosis forest map of hub genes. Each point indicates the hazard ratio of the gene. The line on both sides of the point indicates a 95% confidence interval, and the line in the middle of the abscissa represents hazard ratio = 1

Figure 7A

Analysis Type by Cancer	Cancer vs. Normal	
Bladder Cancer	2	
Brain and CNS Cancer	6	
Breast Cancer	9	1
Cervical Cancer	2	
Colorectal Cancer	2	
Esophageal Cancer	2	
Gastric Cancer	3	
Head and Neck Cancer	2	
Kidney Cancer		
Leukemia		5
Liver Cancer	3	
Lung Cancer	5	
Lymphoma	4	
Melanoma		
Myeloma		
Other Cancer	3	5
Ovarian Cancer	3	
Pancreatic Cancer	1	
Prostate Cancer	1	
Sarcoma	8	
Significant Unique Analyses	55	11
Total Unique Analyses	385	

Figure 7B

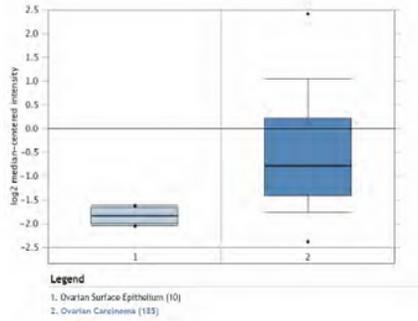


Figure 7C

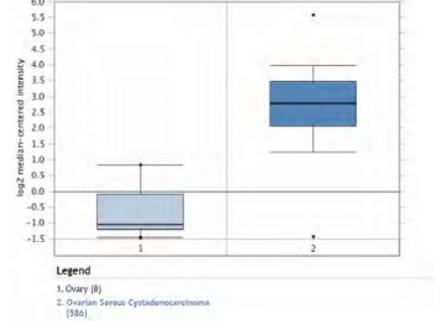


Figure 7D

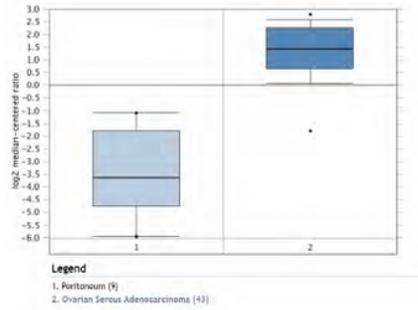


Figure 7

Expression of PBK in Oncomine database. (A) Expression of PBK in all tumor studies in the Oncomine database (B-D) Expression of PBK in chips of different ovarian cancer research in the Oncomine database.

Figure 8A

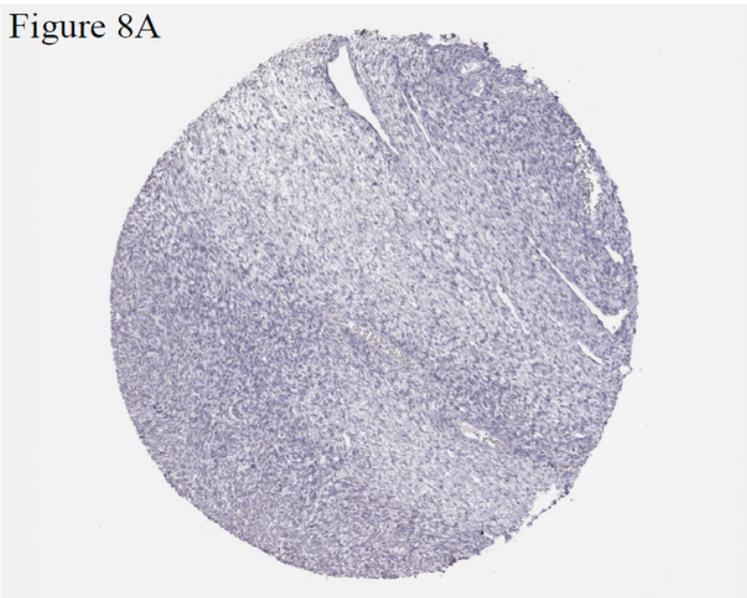


Figure 8B

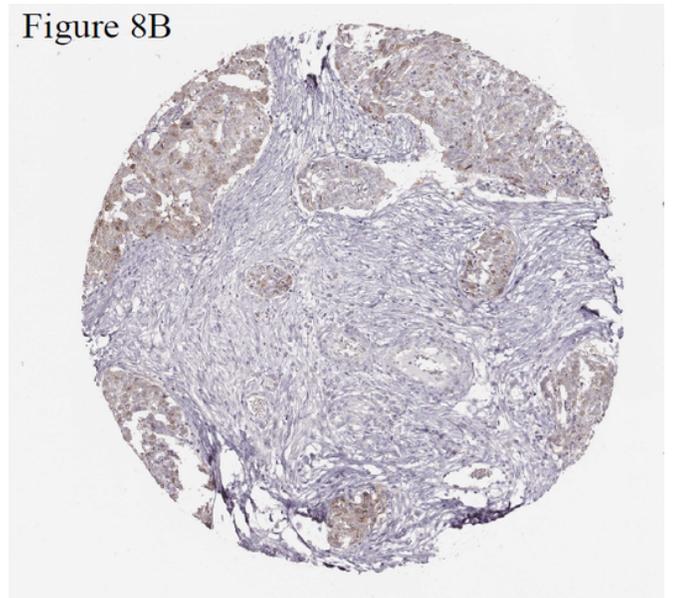


Figure 8

Immunohistochemistry of the PBK based on the Human Protein Atlas. (A) Protein levels of PBK in normal tissue (staining: not detected; intensity: negative; quantity: none). (B) Protein levels of PBK in tumor tissue (staining: medium; intensity: moderate; quantity: 75%-25%).

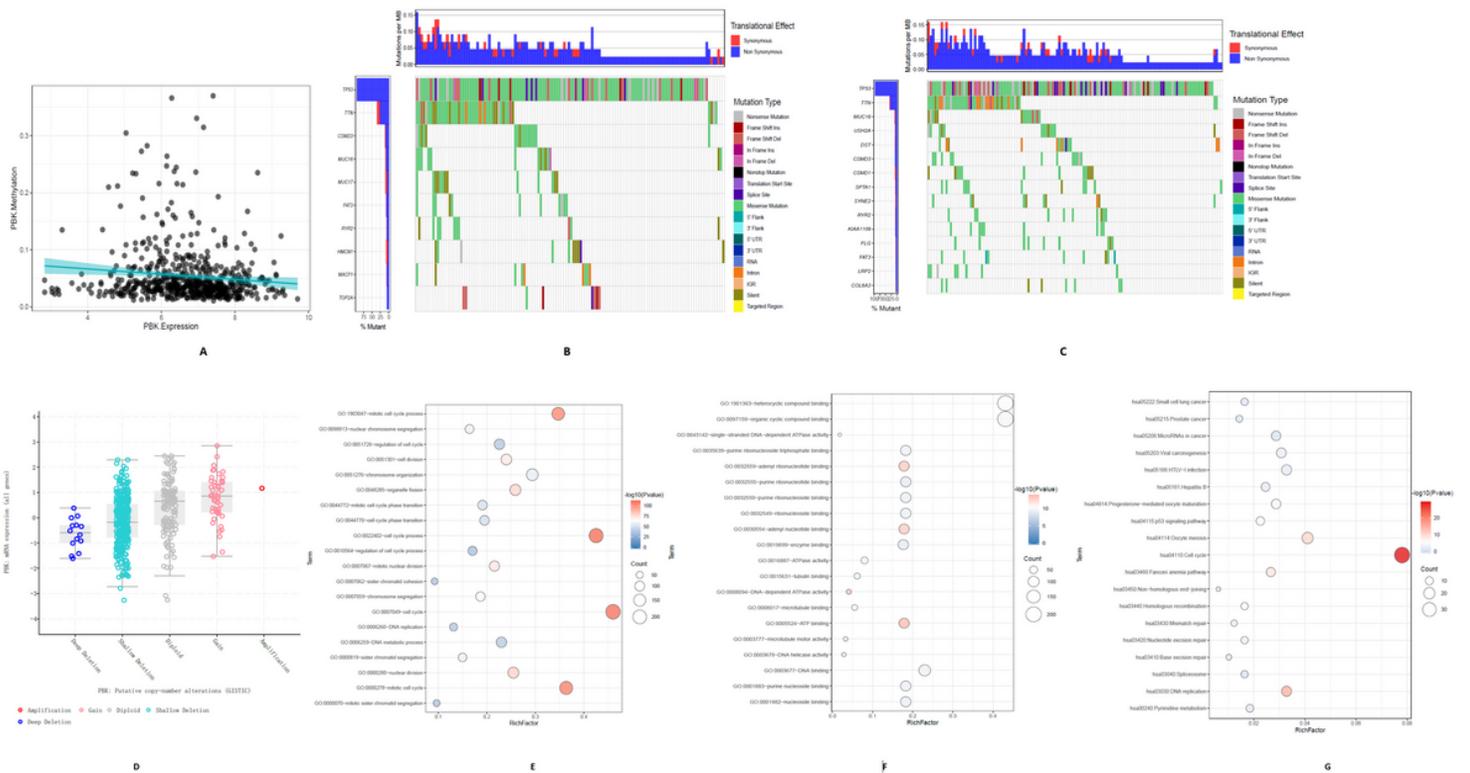


Figure 9

The molecular mechanisms of PBK in ovarian cancer. (A) Correlation analysis between PBK expression and methylation data. (B) Differential somatic mutations were detected in ovarian cancer with low PBK expression group. (C) Differential somatic mutations were detected in ovarian cancer with high PBK expression group. (D) Correlation analysis between PSAT1 expression and copy number variation. (E) Top 20 of biological process enrichment related to PBK co-expressed genes. (F) Top 20 of molecular function enrichment related to PBK co-expressed genes. (G) Top 20 of pathways enrichment related to PBK co-expressed

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigureS2.pdf](#)
- [FigureS1.pdf](#)