Grade-Specific Diagnosis, Prognostic and Therapeutic Biomarkers in Breast Cancer Identified by Integrated Bioinformatics Analysis

zhang hanyu
Huashan Hospital Fudan University https://orcid.org/0000-0001-7324-0943
Zhou XiuLing
Huashan Hospital Fudan University
Zhang XiaoMing
Huashan Hospital Fudan University
Yu ZhongYu
Huashan Hospital Fudan University
Wang JianTao
Huashan Hospital Fudan University
Guo XinYu
Shang Hai lujiazui community hospital
Hu Yue
Huashan Hospital Fudan University
Guo JingChun
Shanghai Medical University: Fudan University
Tao YingHong
Huashan Hospital Fudan University
Zhou HouGuang (zhouhouguang@huashan.org.cn)
Fudan University https://orcid.org/0000-0003-3892-7437

Primary research

Keywords: Bioinformatical analysis, breast cancer, differentially expressed genes, core genes, early diagnosis, treatment information

Posted Date: January 14th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-143765/v1

Abstract

Background

Breast cancer is the most common malignant tumor in women. Due to the mechanism of breast cancer has not yet been completely clear, we aim to identify the key pathway and genes in breast cancer based on bioinformatics method.

Methods

The samples were obtained from NCBI-GEO website. Then, GEO2R tools and Venn diagram software were used to identify the differentially expressed genes (DEGs). Gene enrichment analysis was performed by ClusterProfiler package (version 3.16.1) of R. The protein-protein interaction (PPI) network was plotted by Search Tool for the Retrieval of Interacting Genes (STRING) website and candidate genes were selected from PPI network by Molecular Complex Detection (MCODE) plug-in. Afterwards, we assess the overall survival and expression level of the core genes by Kaplan-Meier plotter and Gene Expression Profiling Interactive Analysis 2 (GEPIA2). Finally, we selected the core genes through the analysis of different grades cancer and analyzed the gene-drug interaction by Comparative Toxicogenomics Database (CTD).

Results

DEGs contained 23 up-regulated and 32 down-regulated genes. GO described biological process, cellular component, and molecular function. KEGG analysis showed up-regulated DEGs were mainly involved in breast cancer-related pathway like P53 signaling pathway, platinum drug resistance and bladder cancer. 15 candidate genes were screened through PPI network, 14 of which had adverse prognosis and high expression in breast cancer. At last, we selected 5 core genes from different grades analysis and found the interaction of these genes and anthracyclines and cyclophosphamide.

Conclusions

In the current study, 5 core genes of breast cancer were identified based on bioinformatics method, which could useful to provide essential information for diagnosis, prognosis and treatment of breast cancer.

Background

Breast cancer is the most common malignant tumor diagnosis among female patients in American [1]. Although the treatment of breast cancer has made great progress in recent years, it remains the main cause of female death. In 2018, the number of new cases and deaths of breast cancer were more than 18 million and 9.5 million in the whole world [2, 3]. In addition, breast cancer is still the most common cause of cancer death in less developed countries [4]. However, the molecular mechanism of tumorigenesis in breast cancer remains unclear. As the caner is not a single disease, the progression of cancer is co-regulated by multiple genes and gene products [5]. Therefore, identifying new genes related to tumorigenesis and prognosis of patients and elucidating the molecular mechanisms of these carcinogenic processes [6] are essential for early diagnosis and prevention. Meanwhile, genomic analysis of breast cancer will help to new therapeutic roadmap [7] that enable personalization in treatment.

Over the past decades, gene chip technology has been widely used. This technology can recognize different expressed genes and store related-data in public databases. Summarize and reanalysis of these genomic data can advantageous to identify biomarker sand gene expression profiles associated with certain diseases [8, 9]. Furthermore, the analysis and sequencing of genomic data to identify cancer-related driving genes and signaling pathways is also one of the most urgent needs in basic cancer research. And our knowledge of the cancer genomes can guide us to exploit more effective ways to prevent the cancer incidence and mortal [10].

In this study, we purpose to screen the differentially expression genes (DEGs) from gene expression datasets and perform the enrichment, overall survival and drug-related analysis to obtain the core genes and signaling pathways which are associated with the oncogenesis, prognosis and treatment of breast cancer. In addition, we analyzed the expression of genes with different grades to get more information about the development of breast cancer.

Materials And Methods

Microarray data acquisition

Four gene expression profile datasets (GSE20713, GSE29044, GSE42568 and GSE61304) were obtained from the NCBI-GEO (http://www.ncbi.nlm.nih.gov/geo/) database [11]. All of the four microarray datasets were based on Platforms of GPL570 ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array). Meanwhile, each of these datasets contained the gene expression profiles of normal tissues and three grades of BC tissues.

Data processing of DEGs

The gene expression values between breast cancer samples and normal samples were analyzed by GEO2R online tool. Differentially expressed genes (DEGs) were screened out by the criterion with |logFC| > 2 and $P < 0.05$. Then, the common DEGs among the four datasets were plotted in Venn diagrams web page tool. The up-regulated and down-regulated genes were identified by logFC > 2 and logFC<-2, respectively.

Analysis of function profiles for DEGs

ClusterProfiler package (version 3.16.1) of R was used to analyze Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Gene and Genomes (KEGG) pathway enrichment [12]. GO enrichment of up-regulated and down-regulated DEGs were characterized in three aspects: molecular function (MF), cellular component (CC), and biological process (BP). *P*-value cutoff = 0.05 was determined as a criterion for statistics significance in GO function and KEGG pathway.

PPI network establishment

Protein-protein interaction (PPI) network information of DEGs was analyzed by Search Tool for the Retrieval of Interacting Genes (STRING, https://string-db.org/) online biological database [13]. Afterwards, Cytoscape software was used to construct the protein interaction networks. The maximum number of interactors = 0 and confidence score ≥ 0.4 was as the threshold value. The densely connected regions in PPI network were analyzed via MCODE plugin of Cytoscape with the criterion as follow: degree cutoff = 2, max. Depth = 100, k-core = 2, and node score cutoff = 0.2.

Survival analysis

The candidate key genes were selected from densely connected regions of PPI networks. Then the prognostic influence of these genes in breast cancer were evaluated by Kaplan-Meier plotter (http://kmplot.com/analysis) web-tool [14]. Gene Expression Profiling Interactive Analysis 2 (GEPIA2) (http://gepia2.cancer-pku.cn/) was applied to comparing the tissue-wise expression of key genes with poor-prognosis between breast cancer and normal samples [15].

Analysis of grade-specific gene expression

DEGs were re-selected from the comparison of three grades (grade1, grade 2 and grade 3) breast cancer samples with normal samples in the four datasets and drew the PPI networks, respectively. Gene set enrichment analysis (GSEA) with Wilcoxon rank sum test and Kolmogorov-Smirnov (KS) test was implemented in comparison of grade 3 breast cancer samples against normal samples of the four datasets via Genetrail2 Web Service (https://genetrail2.bioinf.uni-sb.de/) [16]. KEGG pathway enrichment was re-analyzed by clusterProfiler package of R for the core genes which selected by GSEA and PPI network.

Chemical-Gene interaction analysis

The interaction network of core genes and chemotherapeutics drugs was integrated by using the Comparative Toxicogenomics Database (CTD) (http://ctdbase.org/) [17]. Cytoscape software was used to draw the reaction structure plot of chemical-gene interaction.

Results

Identification of DEGs

In all of the four datasets (GSE20713, GSE29044, GSE42568 and GSE61304), 323 breast cancer tissue samples and 59 normal tissue samples were compared to obtain the DEGs. Then, the DEGs with the criterion as |logFC| > 2 and $P < 0.05$ were performed for analysis by Venn diagram, which showed that 23 up-regulated (logFC > 2) and 32 down-regulated common DEGs (logFC<-2) in the intersection regions were screened out (Table 1 & Fig. 1).

Table 1
23 up-regulated and 30 down-regulated differentially expressed genes(DEGs)in breast cancer were accessed from four profile databases.

DEGs	Genes Name
Up-regulated	TPX2 ASPM ANLN BIRC5 UBE2C CCNB2 CDK1 CEP55 RRM2 TOP2A GJB2 AURKA DLGAP5 KIF2C SQLE COL10A1 COL11A1 NUF2 MMP1 NEK2 UBE2T MMP9 MMP11
Down-regulated	CHRDL1 OGN SCARA5 BTNL9 IGSF10 NTRK2 SDPR CXCL2 TNXB AKAP12 MEOX2 FGF2 APOD PIR-FIGF SPRY2 TSHZ2 ABCA9 ABCA8 MIR548F5 LRRN4CL ADH1B ANK2 DMD ADH1C FOSB MAMDC2 C2orf40 SFRP1 CFD PLIN1

Enrichment analysis of DEGs

The result of GO enrichment analysis implicated that DEGs are involved in many biological functions. The up-regulated DEGs were mainly enriched in nuclear division, organelle fission and mitotic nuclear division while the down-regulated DEGs were mainly enriched in regulation of cellular response to growth factor stimulus, release of sequestered calcium ion into cytosol and negative regulation of sequestering of calcium ion (Table 2 & Fig. 2). KEGG pathway enrichment analysis indicated that up-regulated DEGs were mainly involved in P53

signaling pathway, progesterone-mediated oocyte maturation, oocyte meiosis, platinum drug resistance and bladder cancer (Fig. 3).

Table 2
Gene ontology(GO)analysis of up-regulated and down- regulated genes in breast cancer.

Expression	Ontology	ID	Description	Gene Ratio	p-value	q-value
Up-regulated	BP	GO:0000280	nuclear division	11/23	4.94E-13	2.25E-10
		GO:0048285	organelle fission	11/23	1.44E-12	3.28E-10
		GO:0140014	mitotic nuclear division	9/23	1.36E-11	2.06E-09
		GO:0007051	spindle organization	6/23	4.63E-08	5.28E-06
		GO:0007052	Mitotic spindle organization	5/23	1.66E-07	1.39E-05
		GO:0051225	spindle assembly	5/23	1.83E-07	1.39E-05
		GO:1902850	microtubule cytoskeleton organization involved in mitosis	5/23	4.79E-07	3.12E-05
		GO:1901990	regulation of mitotic cell cycle phase transition	7/23	7.23E-07	4.12E-05
		GO:1901987	regulation of cell cycle phase transition	7/23	1.22E-06	6.17E-05
		GO:0007088	regulation of mitotic nuclear division	5/23	1.46E-06	6.64E-05
	CC	GO:0030496	midbody	7/23	7.72E-10	2.11E-08
		GO:0005819	spindle	7/23	9.47E-08	1.30E-06
		GO:0000793	condensed chromosome	6/23	1.68E-07	1.49E-06
		GO:0000779	condensed chromosome, centromeric region	5/23	2.18E-07	1.49E-06
		GO:0005874	microtubule	7/23	3.24E-07	1.77E-06
	MF	GO:0035173	histone kinase activity	2/23	0.000217175	0.009512334
		GO:0004222	metalloendopeptidase activity	3/23	0.000311611	0.009512334
		GO:0061631	ubiquitin conjugating enzyme activity	2/23	0.001104122	0.01322531
		GO:0061650	ubiquitin-like protein conjugating enzyme activity	2/23	0.001223123	0.01322531

Expression	Ontology	ID	Description	Gene Ratio	p-value	q-value
		GO:0030020	extracellular matrix structural constituent conferring tensile strength	2/23	0.001284832	0.01322531
Down-regulated	BP	GO:0006069	ethanol oxidation	2/26	0.000122031	0.024026518
		GO:0006067	ethanol metabolic process	2/26	0.000385298	0.024026518
		GO:0010881	regulation of cardiac muscle contraction by regulation of the release of sequestered calcium ion	2/26	0.000505092	0.024026518
		GO:0051209	release of sequestered calcium ion into cytosol	3/26	0.000664923	0.024026518
		GO:0090287	regulation of cellular response to growth factor stimulus	4/26	0.000667909	0.024026518
		GO:0051283	negative regulation of sequestering of calcium ion	3/26	0.000696604	0.024026518
		GO:0051282	regulation of sequestering of calcium ion	3/26	0.000729234	0.024026518
		GO:0010882	regulation of cardiac muscle contraction by calcium ion signaling	2/26	0.000739824	0.024026518
		GO:0051208	sequestering of calcium ion	3/26	0.000779983	0.024026518
		GO:0001508	action potential	3/26	0.000815036	0.024026518
	CC	GO:0043034	costamere	2/28	0.000327618	0.019657093
	MF	GO:0004024	alcohol dehydrogenase activity, zinc-dependent	2/23	3.38E-05	0.001242336
		GO:0004022	alcohol dehydrogenase (NAD+) activity	2/23	5.79E-05	0.001242336
		GO:0018455	alcohol dehydrogenase [NAD(P)+] activity	2/23	7.23E-05	0.001242336

Expression	Ontology	ID	Description	Gene Ratio	p-value	q-value
		GO:0004745	retinol dehydrogenase activity	2/23	0.000302687	0.003903069
		GO:0005319	lipid transporter activity	3/23	0.000630027	0.006499226

Table 3
KEGG pathway enrichment analysis of 11 hub-genes.

ID	Description	p-value	q-value	gene
hsa04914	Progesterone-mediated oocyte maturation	3.59E-05	0.00034	AURKA/CCNB2/CDK1
hsa04114	Oocyte meiosis	7.69E-05	0.000364	AURKA/CCNB2/CDK1
hsa01524	Platinum drug resistance	0.001180874	0.002797	BIRC5/TOP2A
hsa04115	p53 signaling pathway	0.001180874	0.002797	CCNB2/CDK1
hsa04110	Cell cycle	0.003369166	0.006384	CCNB2/CDK1
hsa04218	Cellular senescence	0.005284801	0.008344	CCNB2/CDK1
hsa05170	Human immunodeficiency virus 1 infection	0.009595818	0.012987	CCNB2/CDK1

Table 4
Chemical–gene interaction of 5 core genes and chemotherapeutic drugs of breast cancer.

Chemical Name	Gene Symbol	Interaction
Daunorubicin	AURKA	Daunorubicin results in decreased expression of AURKA protein
Doxorubicin	AURKA	Doxorubicin results in decreased expression of AURKA mRNA
Doxorubicin	AURKA	bisphenol A inhibits the reaction [Doxorubicin results in increased expression of AURKA mRNA]
Daunorubicin	BIRC5	Daunorubicin results in increased expression of BIRC5 mRNA
Daunorubicin	BIRC5	[bromotetrandrine co-treated with Daunorubicin] results in decreased expression of BIRC5 mRNA
Doxorubicin	BIRC5	BIRC5 protein results in decreased susceptibility to Doxorubicin
Doxorubicin	BIRC5	Doxorubicin results in decreased expression of BIRC5 mRNA
Doxorubicin	BIRC5	Doxorubicin promotes the reaction [DNMT1 protein binds to BIRC5 promoter]
Doxorubicin	BIRC5	Doxorubicin promotes the reaction [EHMT2 protein binds to BIRC5 promoter]
Doxorubicin	BIRC5	Doxorubicin promotes the reaction [HDAC1 protein binds to BIRC5 promoter]
Doxorubicin	BIRC5	Doxorubicin promotes the reaction [SP1 protein binds to BIRC5 promoter]
Doxorubicin	BIRC5	Doxorubicin promotes the reaction [TP53 protein binds to BIRC5 promoter]
Doxorubicin	BIRC5	Doxorubicin results in decreased expression of BIRC5 promoter
Doxorubicin	BIRC5	[Folic Acid co-treated with Doxorubicin] results in decreased expression of BIRC5 mRNA
Doxorubicin	BIRC5	alvocidib inhibits the reaction [Doxorubicin results in increased expression of and results in increased phosphorylation of BIRC5 protein]
Doxorubicin	BIRC5	BIRC5 protein inhibits the reaction [FOXO3 protein results in increased susceptibility to Doxorubicin]
Doxorubicin	BIRC5	[FOXO3 protein results in decreased expression of BIRC5 protein] which results in increased susceptibility to Doxorubicin
Epirubicin	BIRC5	BIRC5 protein results in decreased susceptibility to Epirubicin
mithramycin A	BIRC5	mithramycin A results in decreased expression of BIRC5 mRNA
Doxorubicin	CCNB2	Doxorubicin results in decreased expression of CCNB2 mRNA
Doxorubicin	CDK1	Doxorubicin results in decreased expression of CDK1 mRNA
Doxorubicin	CDK1	Doxorubicin results in increased phosphorylation of and results in decreased activity of CDK1 protein
Doxorubicin	CDK1	Doxorubicin promotes the reaction [RBL2 protein binds to CDK1 promoter]

Chemical Name	Gene Symbol	Interaction
Doxorubicin	CDK1	CDC25C protein inhibits the reaction [Doxorubicin results in increased phosphorylation of CDK1 protein]
Doxorubicin	CDK1	Go 6976 inhibits the reaction [Doxorubicin results in increased phosphorylation of and results in decreased activity of CDK1 protein]
Doxorubicin	CDK1	[GDC-0941 co-treated with Doxorubicin] results in decreased phosphorylation of CDK1 protein
Liposomal doxorubicin	CDK1	liposomal doxorubicin results in decreased expression of CDK1 protein
pirarubicin	CDK1	pirarubicin affects the phosphorylation of and results in decreased activity of CDK1 protein
Daunorubicin	TOP2A	Daunorubicin results in decreased activity of TOP2A protein
Doxorubicin	TOP2A	Doxorubicin results in decreased activity of TOP2A protein
Doxorubicin	TOP2A	TOP2A mRNA results in decreased susceptibility to Doxorubicin
Doxorubicin	TOP2A	[DPP4 protein results in increased expression of TOP2A protein] which results in increased susceptibility to Doxorubicin
Doxorubicin	TOP2A	[EGF protein results in decreased expression of TOP2A protein] which results in decreased susceptibility to Doxorubicin
Doxorubicin	TOP2A	[Estradiol results in increased expression of TOP2A mRNA] which results in increased susceptibility to Doxorubicin
Doxorubicin	TOP2A	Lovastatin inhibits the reaction [Doxorubicin results in increased expression of TOP2A mRNA]
Doxorubicin	TOP2A	[TOP2A protein co-treated with ERBB2 protein] results in increased susceptibility to Doxorubicin
Doxorubicin	TOP2A	palbociclib inhibits the reaction [Doxorubicin results in increased expression of TOP2A protein]
Doxorubicin	TOP2A	TOP2A protein mutant form results in decreased susceptibility to Doxorubicin
Cyclophosphamide	TOP2A	Cyclophosphamide results in increased expression of TOP2A mRNA
CMF regimen	TOP2A	TOP2A gene mutant form results in decreased susceptibility to CMF regimen
CEF regimen	TOP2A	TOP2A gene mutant form results in increased susceptibility to CEF regimen
FEC protocol	TOP2A	ERBB2 protein promotes the reaction [TOP2A protein results in increased susceptibility to FEC protocol]
Ifosfamide	TOP2A	TOP2A results in decreased susceptibility to [Carboplatin co-treated with Etoposide co-treated with Ifosfamide]
M-VAC protocol	TOP2A	TOP2A protein results in increased susceptibility to M-VAC protocol

Construction of PPI network

The analysis of PPI network construction conferred that two densely connected regions were contained in the gene interaction networks of common DEGs (Fig. 4A). Additionally, of the two densely regions, the more centralized interaction region included 16 core genes (ANLN, ASPM, AURKA, BIRC5, CCNB2, CDK1, CEP55, DLGAP5, KIF2C, NEK2, NUF2, RRM2, TOP2A, TPX2, UBE2C and UBE2T)(Fig. 4B).

Survival analysis of core genes

Hub-genes survival analysis showed 15 of these core genes were highly related to the poor prognosis of breast cancer (log-rank $P < 0.05$) (Fig. 5). Re-analysis of poor prognosis related genes by using GEPIA2 showed that expression level of all 15 genes were significantly ($P < 0.05$) increased compared with normal samples (Fig. 6).

Grade-specific genes analysis

In the comparison of three grades breast cancer samples with normal samples, we identified that the number of common DEGs was increased gradually in grade 1, 2 and 3. The grade 3 gene interaction network construction was obviously more closely linked than grade1 and grade 2 (Fig. 7). In addition, the result showed that the 15 core genes with poor prognosis were all present in the densely region of grade 3 PPI network construction. After the Gene set enrichment analysis of all the four datasets, 11 genes (ASPM, AURKA, BIRC5, CCNB2, CDK1, DLGAP5, KIF2C, NEK2, TOP2A, TPX2 and UBE2C) were performed for the re-analysis of KEGG pathway enrichment. Results of the analysis showed that there were 5 genes (AURKA, BIRC5, CCNB2, CDK1 and TOP2A) remarkable involved in these pathway (Fig. 8 & Table. 3).

Chemical-Gene interaction

Anthracyclines and cyclophosphamide have been identified as the response chemotherapeutics drugs for breast cancer [18]. Therefore, we analyzed the interaction between these drugs and the 5 core genes (Fig. 9 & Table. 4). Results demonstrated that anthracyclines and cyclophosphamide were strongly associated with the 5 hub-genes.

Discussion

To identify the valid biomarkers of diagnosis, prognosis and treatment in breast cancer, we have analyzed four datasets via bioinformatical methods in the present work. A total of 53 common DEGs were found by comparing breast cancer samples with normal samples in the four datasets. Then, GO enrichment and KEGG pathway analysis were performed on the common DEGs to investigate the different biological information related to these genes. It was observed that up-regulated genes were mainly involved in nuclear division, organelle fission and mitotic nuclear division related biological processes, and down-regulated genes were mainly involved in regulation of cellular response to growth factor stimulus related biological process. Meanwhile, the result proved that the up-regulated DEGs were involved in the breast cancer-related pathways such as P53 signaling pathway [19], platinum drug resistance [20] and bladder cancer [21]. Next, we selected 16 genes from the densely regions of PPI networks, and found that 15 of them were associated with poor prognosis of breast cancer, which the expression level in breast cancer was significantly higher than that in normal tissue samples. For further study, we re-constructed the PPI networks with the common DEGs which were screened from the individual comparisons of three grades breast cancer with normal samples. The number

of common DEGs of grade 3 was more increasingly than other grades, and the poor prognosis genes in the above research were reappeared in densely regions of grade PPI networks. It was conferred that the genes in grade 3 lead the dominant role in breast cancer tumorigenesis and progression. Therefore, we chose these 15 poor prognosis genes for the subsequent study. After gene set enrichment analysis and KEGG analysis, 5 genes (AURKA, BIRC5, CCNB2, CDK1 and TOP2A) were identified to be the potential targets for diagnosis and prognosis of breast cancer. Results showed that the cancer-related signaling pathways regulated by these 5 genes were also involved in grade 3 breast cancer. Finally, we explored the interaction of genes and chemotherapeutics dugs and found these 5 genes were strongly correlated with anthracyclines and cyclophosphamide which has been proven to be the effective treatment drugs of breast cancer [22, 23].

Because of the high expression in many cancers, increasingly researches has been focused on AURKA as a target for cancer therapy [24]. Clinical trials of AURKA inhibitors are currently undergoing and have shown the effective anti-tumor activity in multiple cancers [25], but the studies on breast cancer have progressed slowly. AURKA has been well proved involved in the cell cycle and cell division [26], it functions as a key role in mitotic entry, centrosome functions and bipolar spindle assembly especially regulates the G2/M transition [27]. Evidences indicated that AURKA is associated with higher tumor stage and cancer cell migration [28]. Additionally, the breast cancer-related studies have reported that AURKA induces different carcinogenic properties in malignant cells by regulating stem cell function [29]. In a word, the research prospects of AURKA inhibitors are promising in the breast cancer therapy.

Early studies observed that BIRC5 is involved in mitosis by forming the CPC (chromosomal passenger complex) and regulating microtubule dynamics and nucleation [30, 31]. Several researches demonstrated that BIRC5 has been considered as a prognostic biomarker for many cancers in that its closely correlation with chemotherapy resistance, neoplasm metastasis and tumor recurrence [32–34]. To date, the role of BIRC5 as a poor prognostic factor for breast cancer has been confirmed in the past studies [35, 36]. Additionally, in the field of tumor targeted therapy, BIRC5 has also been received much attention [37]. Research data suggested that the BIRC5 inhibitor could increase the antineoplastic activity of paclitaxel in breast cancer treatment [38]. Taken together, BIRC5 has a bright perspective as biomarker for prognostic and targeted therapy of breast cancer in the future.

In the regulation of cell cycle phase, the complex formed by CCNB2 binds to CDK1 acts a key role in G2/M transition [39, 40].Besides, cell cycle disorder in the process of tumorigenesis is associated with dysfunction of CCNB2 and CDK1 [41]. Previous reports demonstrated that CCNB2 not only has negative impact on the survival of breast cancer patients [42], but also could be correlated with the tumor recurrence [43]. Also, CDK1 has been proved to be a predictor of clinical outcome for breast cancer and related with the unfavorable clinicopathologic features [44]. For decades, numerous clinical studies have been performed on CDK1 inhibitors, but the results are disappointed [45]. The reason could be the CDK1 targeted drugs cause the cell toxicity, thus restricting the therapeutic window [46]. However, the role of CDK1 in the breast cancer treatment is still worth exploring. Although the current results of CCNB2 and CDK1 in breast cancer therapy research are dissatisfied, they could have well prospects as factors for diagnosis, prognostic and treatment in future work.

It was reported that TOP2A is typically highly expressed in rapidly proliferating cells and functioned in DNA replication and cell division [47]. Recent study showed that TOP2A was closely related with tumor size differentiation and lymph node metastasis in triple-negative breast cancer patients [48]. While in estrogen receptor positive breast cancer patients, TOP2A could be a evaluation factor of prognostic and recurrence [49].

Previously, numerous researches have been conducted on the correlation between TOP2A and anthracycline-based chemotherapy treatment [50, 51], but the predication effectiveness is uncertain and follow-up studies are required [52]. As the close location with the HER2 gene on chromosome [53], TOP2A should be well studied for prognosis an therapy in breast cancer.

Since the safety and efficacious of anthracyclines and cyclophosphamide has been shown in chemotherapy for breast cancer [54, 55], we performed the gene-drug interaction analysis for further study. Indeed, our work demonstrated that these 5 genes were interacted with the drugs through different ways. These data provide evidence for future research on the influence of these 5 genes on breast cancer treatment.

In the present study, detailed observation showed that the common genes of grade 3 are the most numerous and closely linked in breast cancer. As the heterogeneity of cancer, we conferred that genes in grade 3 were dominated the progression of breast cancer. Thus, the 5 core genes and the cancer-related pathways might be the key factors in tumorigenesis and development of breast cancer. In brief, we hope these data could be helpful for future study on the pathogenesis, prognosis and treatment of breast cancer.

Conclusions

In conclusion, we analyzed the information of four datasets by integrating bioinformatic methods. Results demonstrated that genes in grade 3 acted the leading role in the progression of breast cancer. After expression, survival, gene functional and gene-drug interaction analysis, 5 core genes (AURKA, BIRC5, CCNB2, CDK1 and TOP2A) are identified as the candidate biomarkers for diagnosis, prognosis and targeted therapy of breast cancer. Additionally, the related signaling pathway are taken effect on the regulation of breast cancer.

Abbreviations

DEGs Differentially expression genes

GO Gene Ontology

KEGG Kyoto Encyclopedia of Gene and Genomes

MF Molecular function

CC Cellular component

BP Biological process

PPI Protein-protein interaction

STRING Search Tool for the Retrieval of Interacting Genes

GEPIA2 Gene Expression Profiling Interactive Analysis 2

GSEA Gene set enrichment analysis

CTD Comparative Toxicogenomics Database

ANLN anillin actin binding protein

ASPM assembly factor for spindle microtubules

AURKA aurora kinase A

BIRC5 baculoviral IAP repeat containing 5

CCNB2 cyclin B2

CDK1 cyclin dependent kinase 1

CEP55 centrosomal protein 55

DLGAP5 DLG associated protein 5

KIF2C kinesin family member 2C

NEK2 NIMA related kinase 2

NUF2 NUF2 component of NDC80 kinetochore complex

RRM2 ribonucleotide reductase regulatory subunit M2

TOP2A DNA topoisomerase II alpha

TPX2 TPX2 microtubule nucleation factor

UBE2C ubiquitin conjugating enzyme E2 C

UBE2T ubiquitin conjugating enzyme E2 T

CPC chromosomal passenger complex

HER2 human epidermal growth factor receptor 2

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

There are no human subjects in this article and consent for publication is not applicable.

Availability of data and materials

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

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by grants from the National Natural Scientific Foundation, China (8117032, 81571361, 81871098), Key projects of Shanghai Health and Health Commission on Integration of traditional Chinese and Western Medicine (ZY (2018-2020) -FWTX - 3007), and the Shanghai Municipal Outstanding Talent Development Foundation (2012052).

Authors' contributions

*These authors contributed equally to this work; #Corresponding authors.

Acknowledgements

We would like to thank all those who participated in this study.

References

1. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2020. *CA Cancer J Clin* 2020, 70(1).
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018, 68(6):394-424.
3. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, Znaor A, Bray F: Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer* 2019, 144(8):1941-1953.
4. Harbeck N, Gnant M: Breast cancer. *Lancet (London, England)* 2017, 389(10074):1134-1150.
5. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA *et al*: Molecular portraits of human breast tumours. *Nature* 2000, 406(6797):747-752.
6. Xu S, Kong D, Chen Q, Ping Y, Pang D: Oncogenic long noncoding RNA landscape in breast cancer. *Molecular cancer* 2017, 16(1):129.
7. Ellis M: Genomic analysis of breast cancer heralds a changing treatment paradigm. *Journal of the National Comprehensive Cancer Network : JNCCN* 2014, 12(5 Suppl):750-752.
8. Feng H, Gu ZY, Li Q, Liu QH, Yang XY, Zhang JJ: Identification of significant genes with poor prognosis in ovarian cancer via bioinformatical analysis. *Journal of ovarian research* 2019, 12(1):35.
9. Yan H, Zheng G, Qu J, Liu Y, Huang X, Zhang E, Cai Z: Identification of key candidate genes and pathways in multiple myeloma by integrated bioinformatics analysis. *Journal of cellular physiology* 2019.
10. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW: Cancer genome landscapes. *Science (New York, NY)* 2013, 339(6127):1546-1558.
11. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M *et al*: NCBI GEO: archive for functional genomics data sets–update. *Nucleic acids research*

2013, 41(Database issue):D991-995.

12. Yu G, Wang L-G, Han Y, He Q-Y: clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012, 16(5):284-287.
13. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P *et al*: STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 2019, 47(D1):D607-D613.
14. Lánczky A, Nagy Á, Bottai G, Munkácsy G, Szabó A, Santarpia L, Győrffy B: miRpower: a web-tool to validate survival-associated miRNAs utilizing expression data from 2178 breast cancer patients. *Breast Cancer Res Treat* 2016, 160(3):439-446.
15. Tang Z, Kang B, Li C, Chen T, Zhang Z: GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res* 2019, 47(W1):W556-W560.
16. Stöckel D, Kehl T, Trampert P, Schneider L, Backes C, Ludwig N, Gerasch A, Kaufmann M, Gessler M, Graf N *et al*: Multi-omics enrichment analysis using the GeneTrail2 web service. *Bioinformatics* 2016, 32(10):1502-1508.
17. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, Wiegers J, Wiegers TC, Mattingly CJ: Comparative Toxicogenomics Database (CTD): update 2021. *Nucleic Acids Res* 2020.
18. Harbeck N, Gnant M: Breast cancer. *Lancet* 2017, 389(10074):1134-1150.
19. Gasco M, Shami S, Crook T: The p53 pathway in breast cancer. *Breast Cancer Res* 2002, 4(2):70-76.
20. Eckstein N: Platinum resistance in breast and ovarian cancer cell lines. *J Exp Clin Cancer Res* 2011, 30:91.
21. Damrauer JS, Hoadley KA, Chism DD, Fan C, Tiganelli CJ, Wobker SE, Yeh JJ, Milowsky MI, Iyer G, Parker JS *et al*: Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. *Proc Natl Acad Sci U S A* 2014, 111(8):3110-3115.
22. Palmieri C, Krell J, James CR, Harper-Wynne C, Misra V, Cleator S, Miles D: Rechallenging with anthracyclines and taxanes in metastatic breast cancer. *Nat Rev Clin Oncol* 2010, 7(10):561-574.
23. Jones S, Holmes FA, O'Shaughnessy J, Blum JL, Vukelja SJ, McIntyre KJ, Pippen JE, Bordelon JH, Kirby RL, Sandbach J *et al*: Docetaxel With Cyclophosphamide Is Associated With an Overall Survival Benefit Compared With Doxorubicin and Cyclophosphamide: 7-Year Follow-Up of US Oncology Research Trial 9735. *J Clin Oncol* 2009, 27(8):1177-1183.
24. Wu C, Lyu J, Yang EJ, Liu Y, Zhang B, Shim JS: Targeting AURKA-CDC25C axis to induce synthetic lethality in ARID1A-deficient colorectal cancer cells. *Nat Commun* 2018, 9(1):3212.
25. Falchook GS, Bastida CC, Kurzrock R: Aurora Kinase Inhibitors in Oncology Clinical Trials: Current State of the Progress. *Semin Oncol* 2015, 42(6):832-848.
26. Mahankali M, Henkels KM, Speranza F, Gomez-Cambronero J: A non-mitotic role for Aurora kinase A as a direct activator of cell migration upon interaction with PLD, FAK and Src. *J Cell Sci* 2015, 128(3):516-526.
27. Do TV, Xiao F, Bickel LE, Klein-Szanto AJ, Pathak HB, Hua X, Howe C, O'Brien SW, Maglaty M, Ecsedy JA *et al*: Aurora kinase A mediates epithelial ovarian cancer cell migration and adhesion. *Oncogene* 2014, 33(5):539-549.
28. Guan Z, Wang X-r, Zhu X-f, Huang X-f, Xu J, Wang L-h, Wan X-b, Long Z-j, Liu J-n, Feng G-k *et al*: Aurora-A, a negative prognostic marker, increases migration and decreases radiosensitivity in cancer cells. *Cancer Res* 2007, 67(21):10436-10444.

29. Zheng F, Yue C, Li G, He B, Cheng W, Wang X, Yan M, Long Z, Qiu W, Yuan Z *et al*: Nuclear AURKA acquires kinase-independent transactivating function to enhance breast cancer stem cell phenotype. *Nat Commun* 2016, 7:10180.
30. Ruchaud S, Carmena M, Earnshaw WC: The chromosomal passenger complex: one for all and all for one. *Cell* 2007, 131(2):230-231.
31. Rosa J, Canovas P, Islam A, Altieri DC, Doxsey SJ: Survivin modulates microtubule dynamics and nucleation throughout the cell cycle. *Mol Biol Cell* 2006, 17(3):1483-1493.
32. Shariat SF, Karakiewicz PI, Godoy G, Karam JA, Ashfaq R, Fradet Y, Isbarn H, Montorsi F, Jeldres C, Bastian PJ *et al*: Survivin as a prognostic marker for urothelial carcinoma of the bladder: a multicenter external validation study. *Clin Cancer Res* 2009, 15(22):7012-7019.
33. Chuwa AH, Sone K, Oda K, Ikeda Y, Fukuda T, Wada-Hiraike O, Inaba K, Makii C, Takeuchi M, Oki S *et al*: Significance of survivin as a prognostic factor and a therapeutic target in endometrial cancer. *Gynecol Oncol* 2016, 141(3):564-569.
34. Rödel F, Sprenger T, Kaina B, Liersch T, Rödel C, Fulda S, Hehlgans S: Survivin as a prognostic/predictive marker and molecular target in cancer therapy. *Curr Med Chem* 2012, 19(22):3679-3688.
35. Hamy AS, Bieche I, Lehmann-Che J, Scott V, Bertheau P, Guinebretière JM, Matthieu MC, Sigal-Zafrani B, Tembo O, Marty M *et al*: BIRC5 (survivin): a pejorative prognostic marker in stage II/III breast cancer with no response to neoadjuvant chemotherapy. *Breast Cancer Res Treat* 2016, 159(3):499-511.
36. Zhang M, Zhang X, Zhao S, Wang Y, Di W, Zhao G, Yang M, Zhang Q: Prognostic value of survivin and EGFR protein expression in triple-negative breast cancer (TNBC) patients. *Target Oncol* 2014, 9(4):349-357.
37. Li F, Aljahdali I, Ling X: Cancer therapeutics using survivin BIRC5 as a target: what can we do after over two decades of study?*J Exp Clin Cancer Res* 2019, 38(1):368.
38. Lyu H, Wang S, Huang J, Wang B, He Z, Liu B: Survivin-targeting miR-542-3p overcomes HER3 signaling-induced chemoresistance and enhances the antitumor activity of paclitaxel against HER2-overexpressing breast cancer. *Cancer Lett* 2018, 420.
39. Petri ET, Errico A, Escobedo L, Hunt T, Basavappa R: The crystal structure of human cyclin B. *Cell Cycle* 2007, 6(11):1342-1349.
40. Santamaría D, Barrière C, Cerqueira A, Hunt S, Tardy C, Newton K, Cáceres JF, Dubus P, Malumbres M, Barbacid M: Cdk1 is sufficient to drive the mammalian cell cycle. *Nature* 2007, 448(7155):811-815.
41. Vermeulen K, Van Bockstaele DR, Berneman ZN: The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. *Cell Prolif* 2003, 36(3):131-149.
42. Shubbar E, Kovács A, Hajizadeh S, Parris TZ, Nemes S, Gunnarsdóttir K, Einbeigi Z, Karlsson P, Helou K: Elevated cyclin B2 expression in invasive breast carcinoma is associated with unfavorable clinical outcome. *BMC Cancer* 2013, 13:1.
43. Zhu J, Muskhelishvili L, Tong W, Borlak J, Chen M: Cancer genomics predicts disease relapse and therapeutic response to neoadjuvant chemotherapy of hormone sensitive breast cancers. *Sci Rep* 2020, 10(1):8188.
44. Kim SJ, Nakayama S, Miyoshi Y, Taguchi T, Tamaki Y, Matsushima T, Torikoshi Y, Tanaka S, Yoshida T, Ishihara H *et al*: Determination of the specific activity of CDK1 and CDK2 as a novel prognostic indicator for early breast cancer. *Ann Oncol* 2008, 19(1):68-72.

45. Asghar U, Witkiewicz AK, Turner NC, Knudsen ES: The history and future of targeting cyclin-dependent kinases in cancer therapy. *Nat Rev Drug Discov* 2015, 14(2):130-146.
46. Kang J, Sergio CM, Sutherland RL, Musgrove EA: Targeting cyclin-dependent kinase 1 (CDK1) but not CDK4/6 or CDK2 is selectively lethal to MYC-dependent human breast cancer cells. *BMC Cancer* 2014, 14:32.
47. Liu T, Zhang H, Yi S, Gu L, Zhou M: Mutual regulation of MDM4 and TOP2A in cancer cell proliferation. *Mol Oncol* 2019, 13(5):1047-1058.
48. Zheng H, Li X, Chen C, Chen J, Sun J, Sun S, Jin L, Li J, Sun S, Wu X: Quantum dot-based immunofluorescent imaging and quantitative detection of TOP2A and prognostic value in triple-negative breast cancer. *Int J Nanomedicine* 2016, 11:5519-5529.
49. Sparano JA, Goldstein LJ, Davidson NE, Sledge GW, Gray R: TOP2A RNA expression and recurrence in estrogen receptor-positive breast cancer. *Breast Cancer Res Treat* 2012, 134(2):751-757.
50. Romero A, Caldés T, Díaz-Rubio E, Martín M: Topoisomerase 2 alpha: a real predictor of anthracycline efficacy? *Clin Transl Oncol* 2012, 14(3):163-168.
51. Wang J, Xu B, Yuan P, Zhang P, Li Q, Ma F, Fan Y: TOP2A amplification in breast cancer is a predictive marker of anthracycline-based neoadjuvant chemotherapy efficacy. *Breast Cancer Res Treat* 2012, 135(2):531-537.
52. Pritchard KI: Are HER2 and TOP2A useful as prognostic or predictive biomarkers for anthracycline-based adjuvant chemotherapy for breast cancer? *J Clin Oncol* 2009, 27(24):3875-3876.
53. Watters AD, Going JJ, Cooke TG, Bartlett JMS: Chromosome 17 aneusomy is associated with poor prognostic factors in invasive breast carcinoma. *Breast Cancer Res Treat* 2003, 77(2):109-114.
54. Paradiso A, Schittulli F, Cellamare G, Mangia A, Marzullo F, Lorusso V, De Lena M: Randomized clinical trial of adjuvant fluorouracil, epirubicin, and cyclophosphamide chemotherapy for patients with fast-proliferating, node-negative breast cancer. *J Clin Oncol* 2001, 19(19):3929-3937.
55. Masuda N, Lee S-J, Ohtani S, Im Y-H, Lee E-S, Yokota I, Kuroi K, Im S-A, Park B-W, Kim S-B *et al*: Adjuvant Capecitabine for Breast Cancer after Preoperative Chemotherapy. *N Engl J Med* 2017, 376(22):2147-2159.

Figures

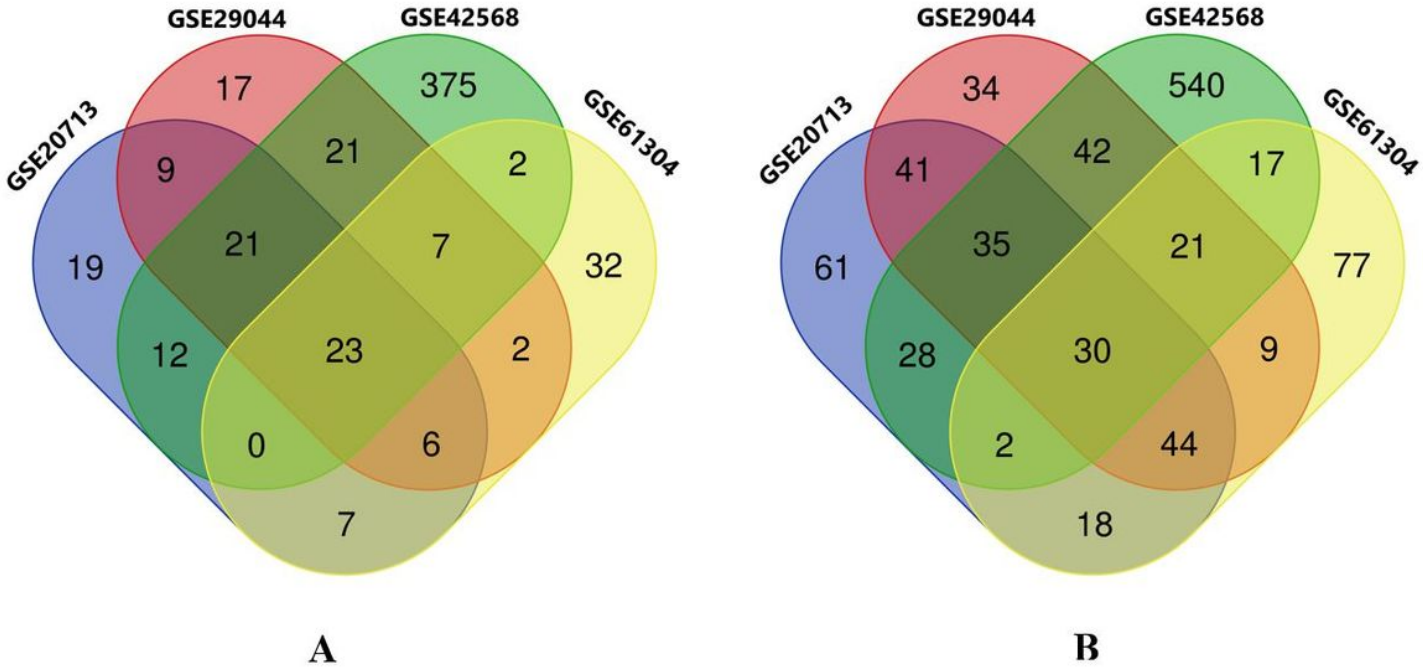


Figure 1

Common DEGs were screened out from four datasets (GSE20713, GSE29044, GSE42568, GSE61304 and GSE20713). A) The intersection of four rounded rectangle areas represents the number of up-regulated common DEGs (log FC>2). B) The intersection of four rounded rectangle areas represents the number of down-regulated common DEGs (log FC<-2).

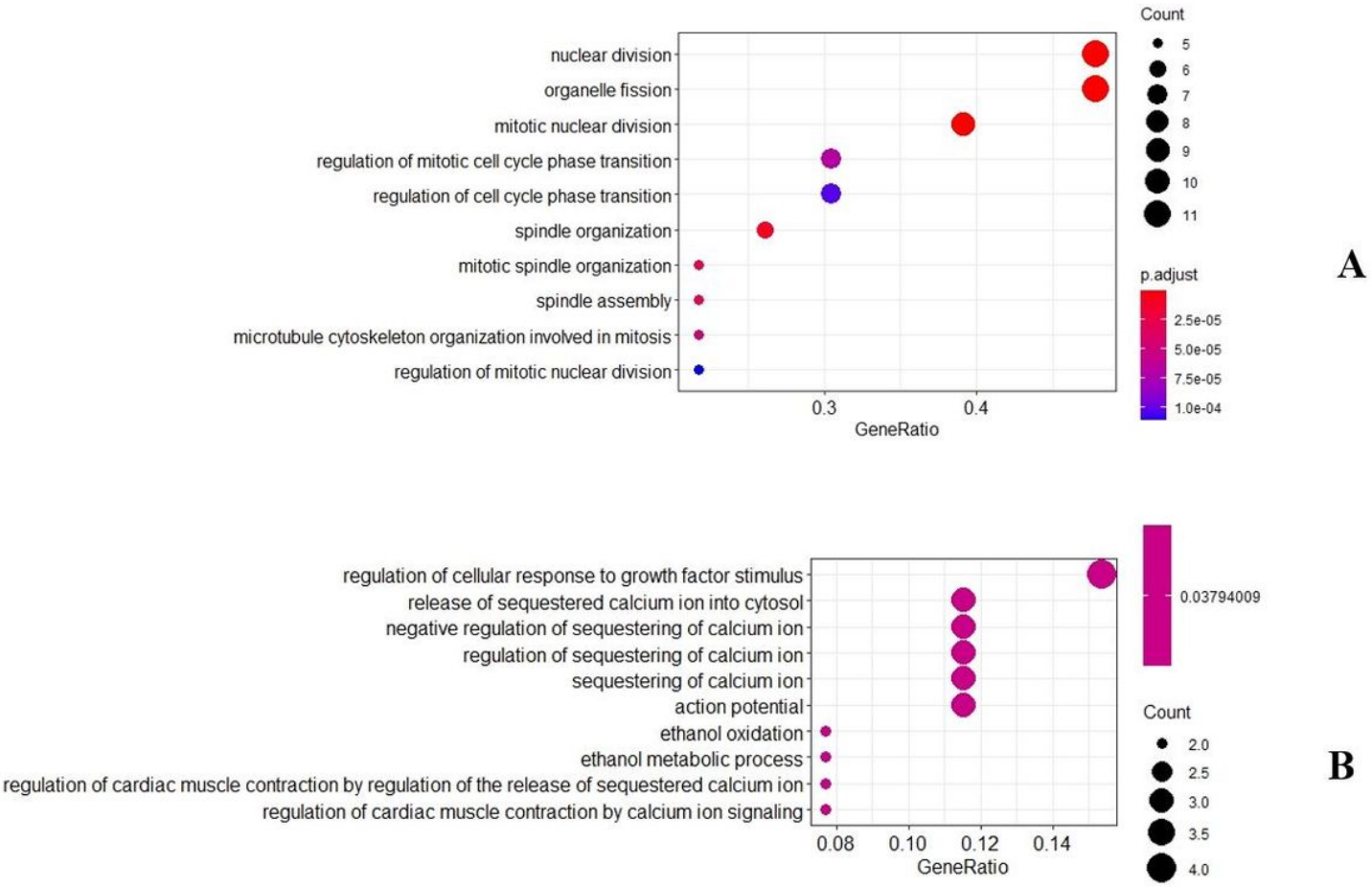
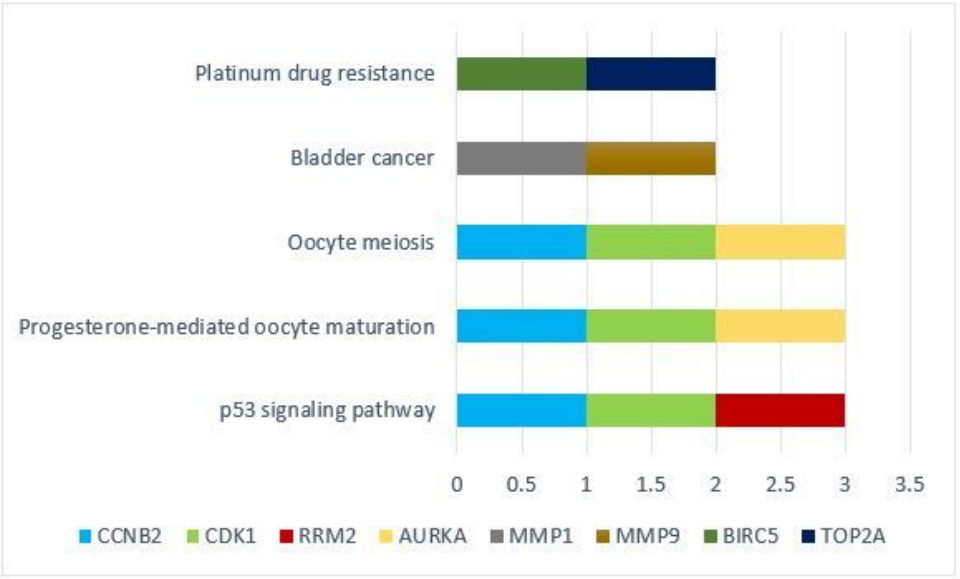


Figure 2

Bubble chart of GO enrichment analysis. A) and B) chart show the 10 terms of most enriched biological functions in up-regulated and down-regulated genes, respectively. Bubble size and color represent the number of genes and adjust p-value respectively.



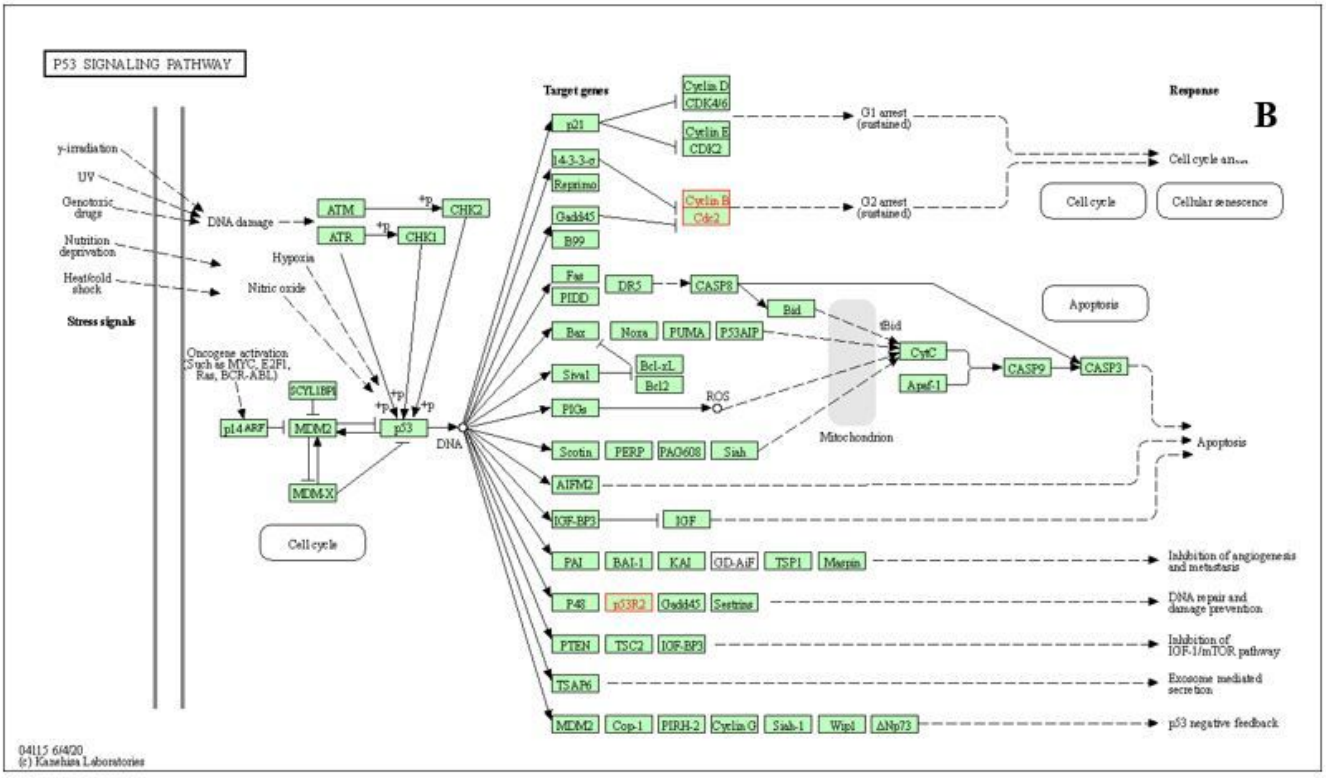


Figure 3

KEGG pathway enrichment analysis diagram of common DEGs. A) The result of up-regulated DEGs pathway enrichment analysis. B) P53 signaling pathway. Three genes (CyclinB means CCNB2, Cdc2 means CDK1, p53R2 means RRM2) are significantly enriched in this signaling pathway.

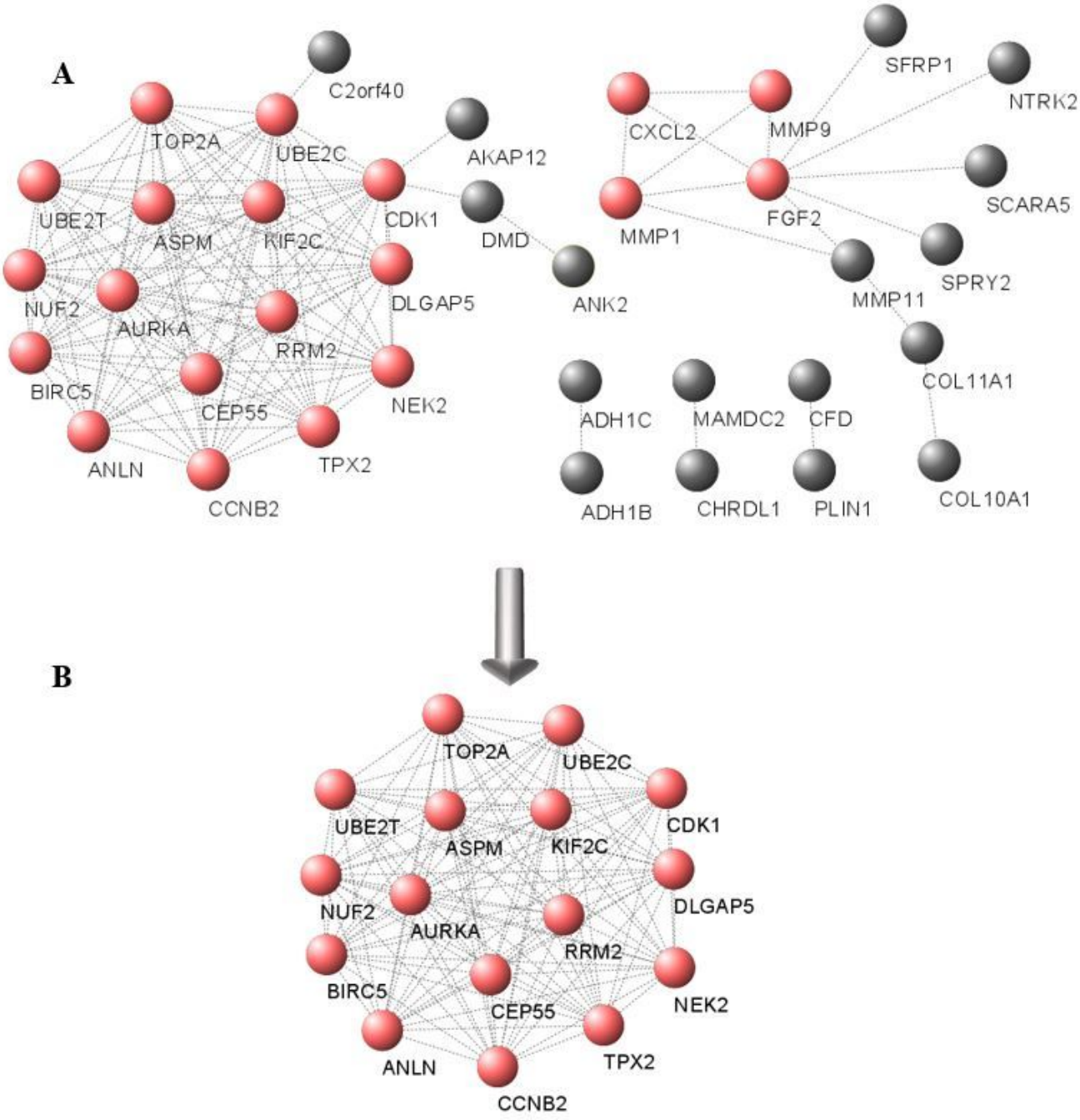


Figure 4

PPI network construction diagram of common DEGs. A) The nodes and lines represent proteins and interactions of proteins individually. Red and gray spheres represent densely connected and non-densely connected regions in PPI network individually. B) Proteins of the main densely connected region in PPI network. (degree cutoff = 2, node score cutoff = 0.2, k-core = 2, and max. Depth = 100).

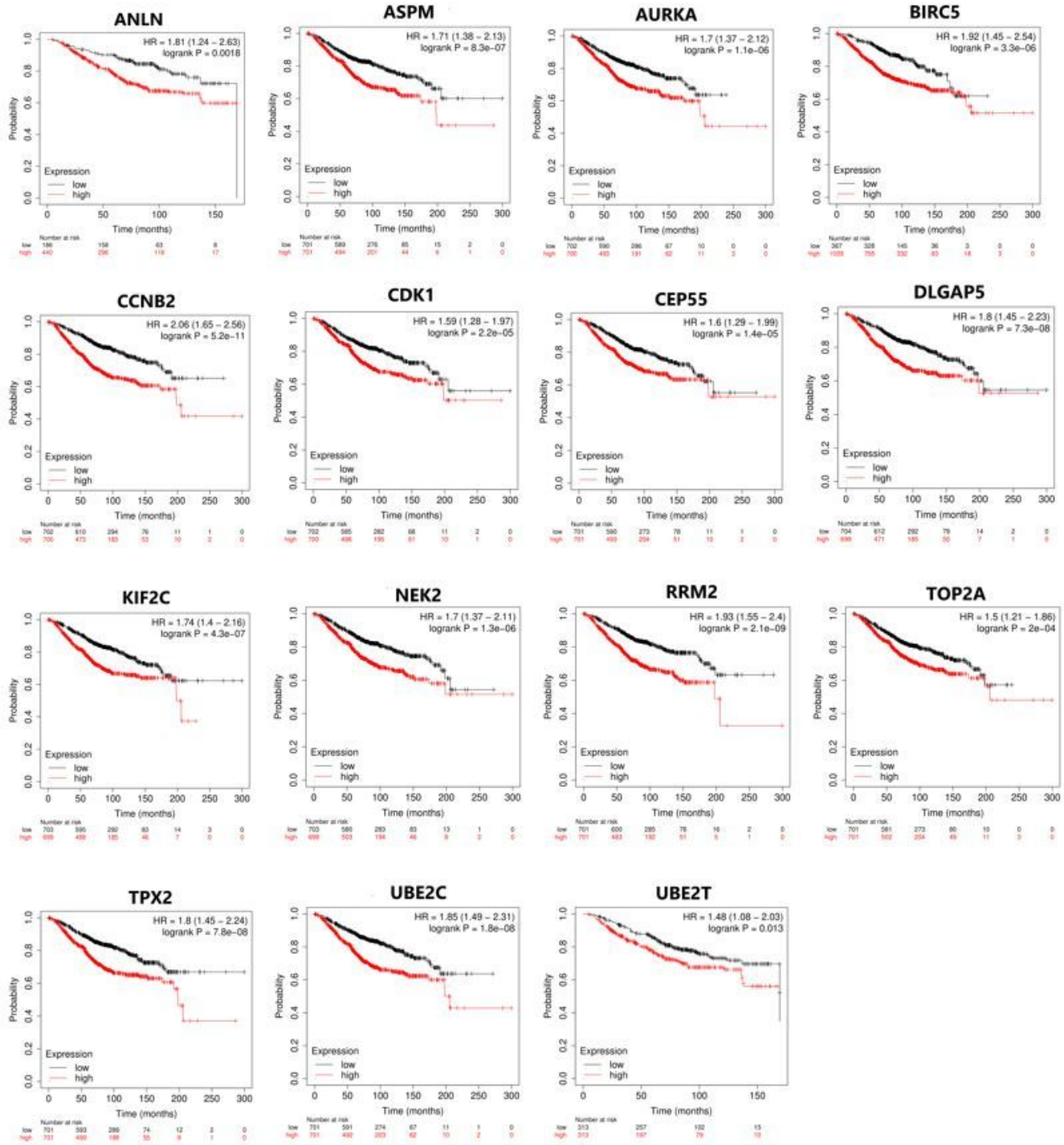


Figure 5

The Kaplan-Meier survival plot of the 15 core genes. In each plot, HR>1 and HR<1 represent gene related to poor and good prognosis. Log-rank P represents the significant difference in OS time between the high and low expression of each gene. The difference is statistically significant with log-rank P<0.05. (OS: overall survival).

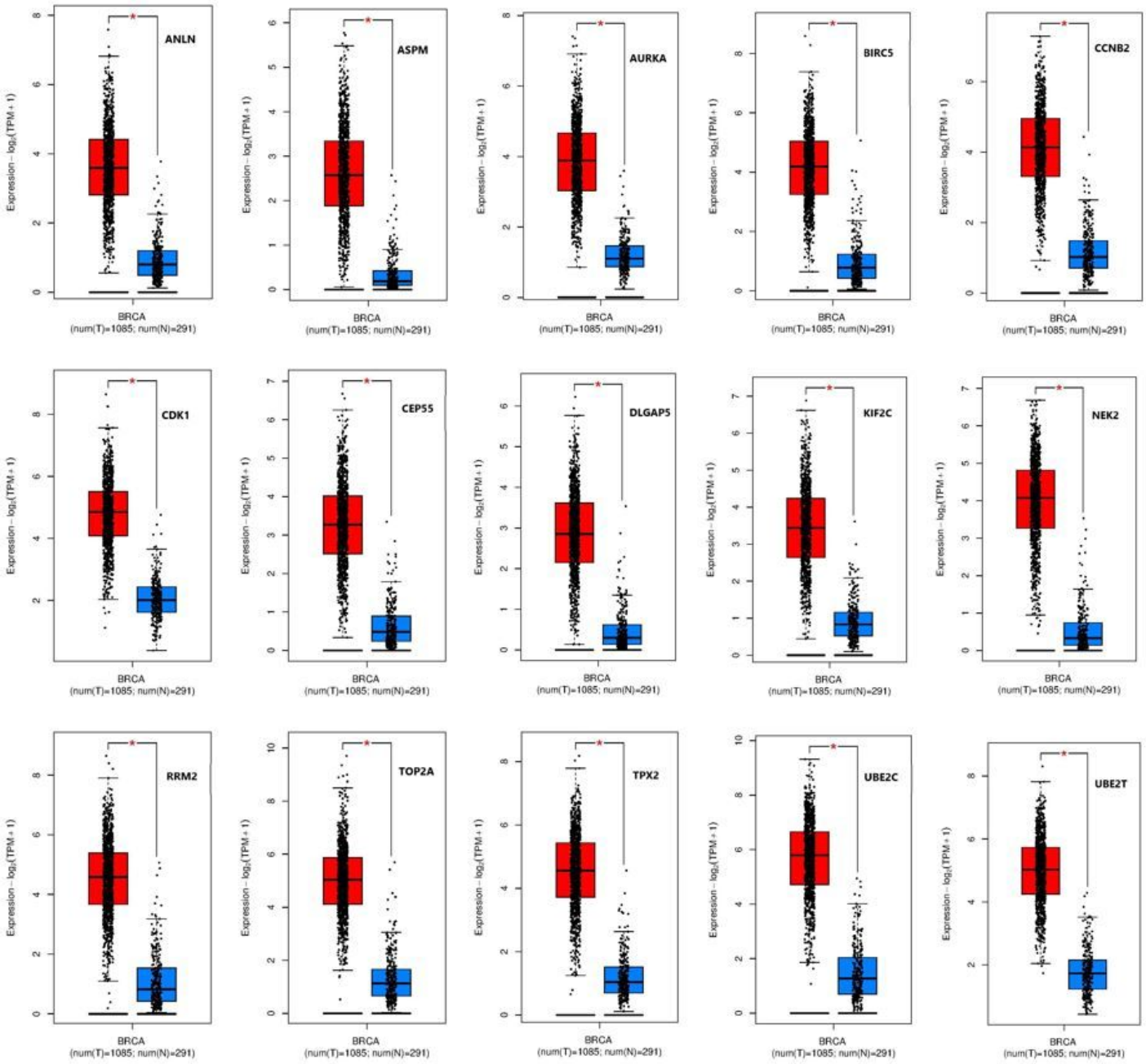


Figure 6

Comparison of 15 hub-genes expression level between breast cancer and normal tissues. Red and blue color represent tumor and normal tissue samples, respectively. Plots with asterisk represent expression level difference is statistically significant (P<0.05).

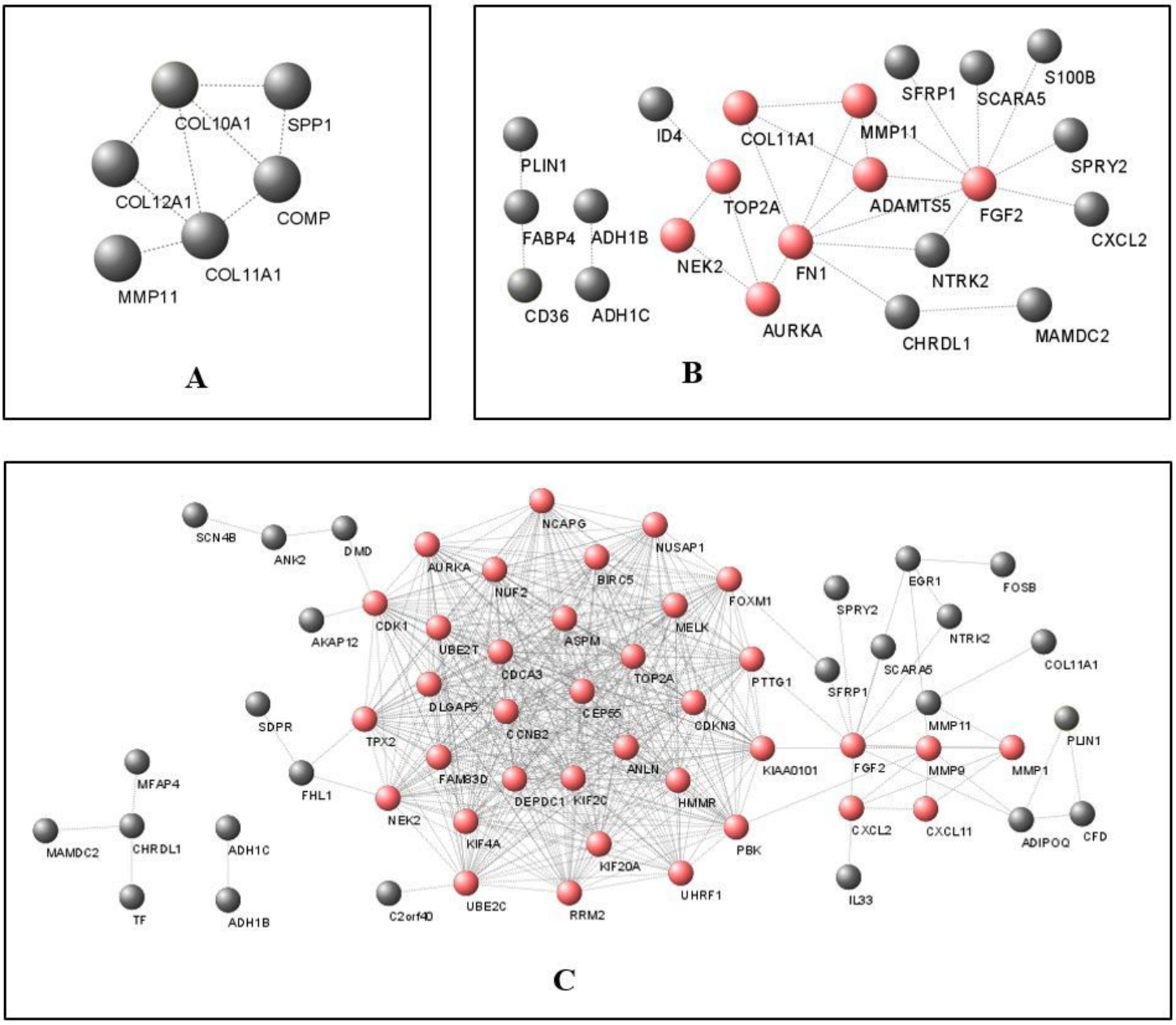


Figure 7

Grade-specific PPI network construction plot. A) Grade 1, B) Grade 2 and C) Grade 3 PPI network of DEGs in breast cancer. The nodes and lines represent proteins and interactions of proteins, respectively. Red and gray spheres represent densely connected and non-densely connected regions in PPI network , respectively.

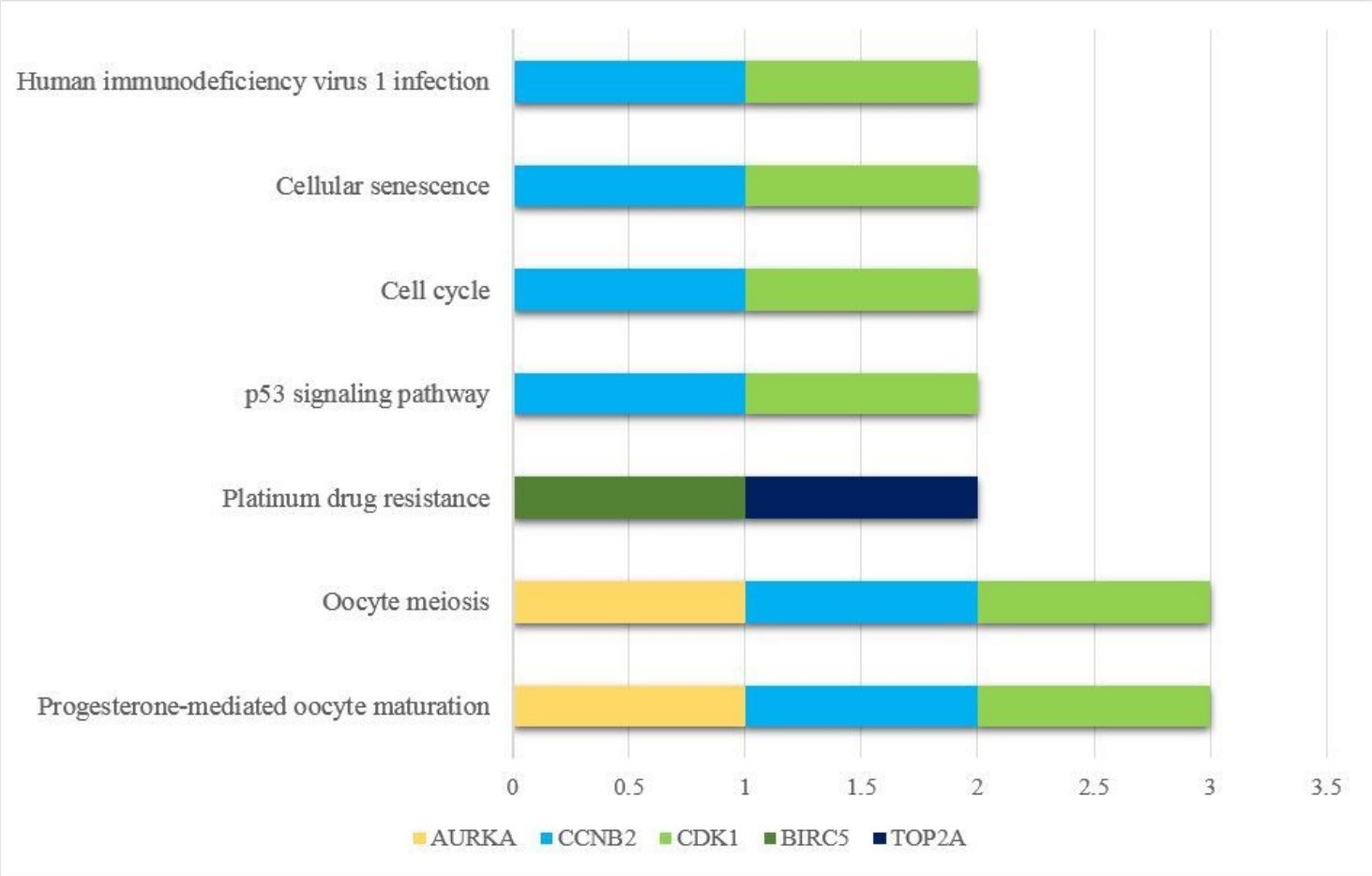


Figure 8

KEGG pathway enrichment histogram of 11 hub-genes. Five genes (AURKA, CCNB2, CDK1, BIRC5, TOP2A) are significantly enriched in seven pathways.

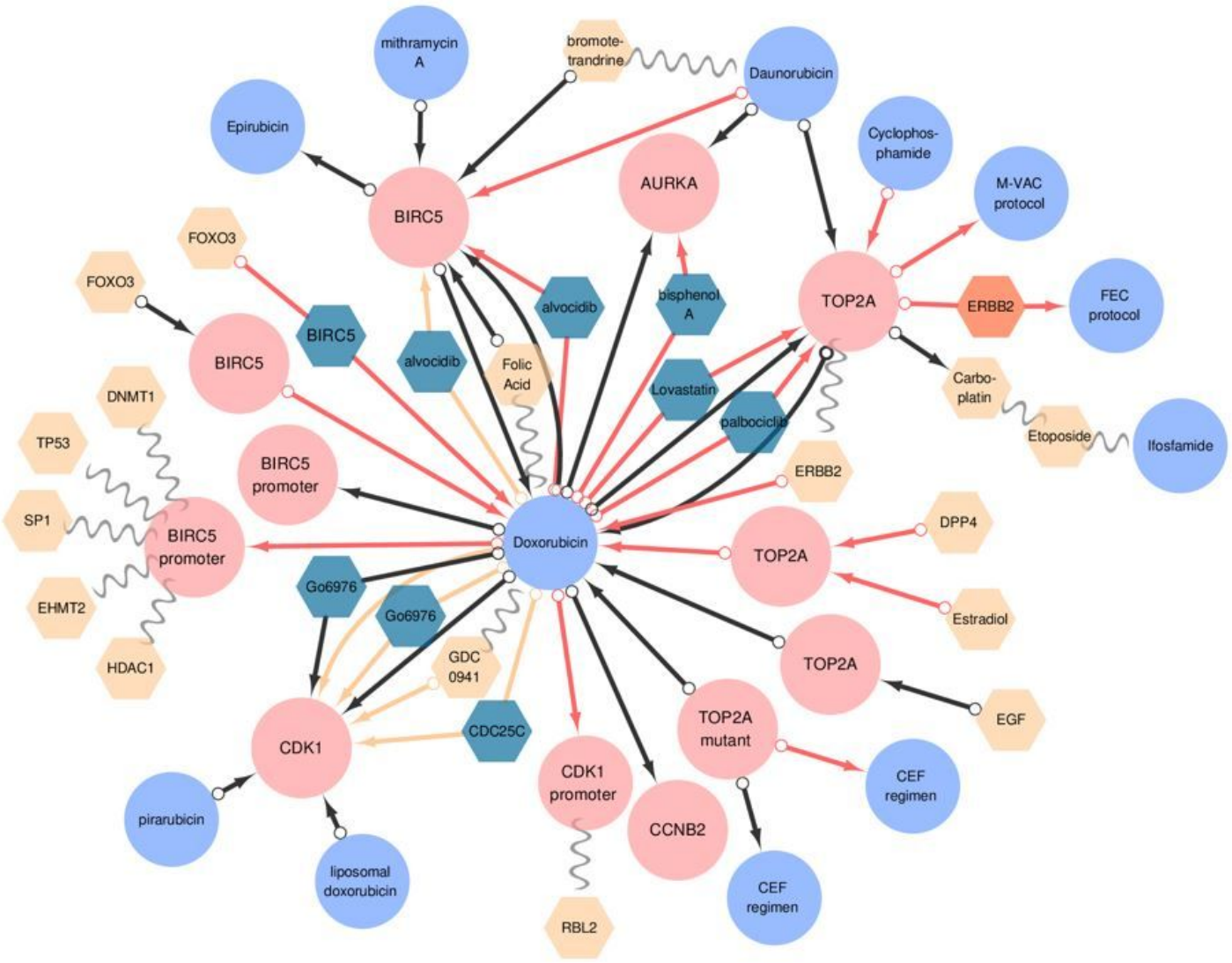


Figure 9

Chemical–gene interaction network of 5 hub-genes and chemotherapeutic drugs. Each line represents one interaction. Open circles and arrows represent the initiation and direction of interaction, respectively. Solid circles represent crucial genes and chemotherapeutic drugs. Hexagons represent other related factors of the interaction. Red lines represent promoting the reaction, increasing the expression of genes or increasing susceptibility of drugs. Black lines represent inhibiting the reaction, decreasing the expression of genes or decreasing susceptibility of drugs. Yellow lines represent affecting the phosphorylation or acetylation of genes. Wave lines represent co-treating with drugs or binding with genes. Red and blue circles represent crucial genes and main drugs for BC individually. Red and blue hexagons represent promoting and inhibiting the reaction individually.