

# Bioinformatics Identification of Prognostic Factors Associated with Breast Cancer

**Ying Wei**

Sichuan University <https://orcid.org/0000-0001-8178-4705>

**Shipeng Zhang**

College of Pharmacy, North Sichuan Medical College

**Li Xiao**

West China School of Basic Medical Sciences and Forensic Medicine

**Jing Zou**

West China School of Basic Medical Sciences and Forensic Medicine

**Yingqing Fu**

West China School of Basic Medical Sciences and Forensic Medicine

**Yi Ye**

West China School of Basic Medical Sciences and Forensic Medicine

**Linchuan Liao** (✉ [linchuanliao@scu.edu.cn](mailto:linchuanliao@scu.edu.cn))

West China School of Basic Medical Sciences and Forensic Medicine <https://orcid.org/0000-0003-3700-8471>

---

## Research

**Keywords:** Breast cancer, Differentially expressed genes, miRNAs, Transcription factors, Bioinformatic analysis

**Posted Date:** December 2nd, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-117477/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

**Background:** Breast cancer (BRCA) remains one of the most common forms of cancer and is the most prominent driver of cancer-related death among women. The mechanistic basis for BRCA, however, remains incompletely understood. In particular, the relationships between driver mutations and signaling pathways in BRCA are poorly characterized, making it difficult to identify reliable clinical biomarkers that can be employed in diagnostic, therapeutic, or prognostic contexts.

**Methods:** First, we downloaded publically available BRCA datasets (GSE45827, GSE42568, and GSE61304) from the Gene Expression Omnibus (GEO) database. We then compared gene expression profiles between tumor and control tissues in these datasets using Venn diagrams and the GEO2R analytical tool. We further explore the functional relevance of BRCA-associated differentially expressed genes (DEGs) via functional and pathway enrichment analyses using the DAVID tool, and we then constructed a protein-protein interaction network incorporating DEGs of interest using the Search Tool for the Retrieval of Interacting Genes (STRING) database. Modules within this PPI network were then identified using Cytoscape, leading to the identification of key candidate genes. The prognostic relevance of these candidate genes was then established through Kaplan-Meier survival analyses and further Gene Expression Profiling Interactive Analysis (GEPIA) validation. Then, key gene-target miRNA regulatory network and transcription factor-key gene regulatory relationships were established using the online miRWalk2.0, TargetScan7.2, miRDB and TRRUST tools. Moreover, four representative key molecules (AURKA, RRM2, BIRC5, and E2F1) were optionally chosen for verification by using quantitative real-time polymerase chain reaction (RT-PCR) and western blot.

**Results:** We identified 85 BRCA-related DEGs across these three datasets. The 31 upregulated DEGs were found to be enriched for pathways and functions including mitotic nuclear division, cell division, G2/M transition of mitotic cell cycle, collagen catabolic process, endodermal cell differentiation, oocyte meiosis, ECM-receptor interactions, and p53 signaling pathway. The 54 downregulated DEGs were, in contrast, enriched in pathways and functions such as lipid metabolic processes, lipid transport, regulation of cardiac muscle contraction by regulation of the release of sequestered calcium ions, positive regulation of cell proliferation, positive regulation of cell-matrix adhesion, tyrosine metabolism, cytochrome P450 drug metabolism, protein digestion and absorption, and PPAR signaling. We were further able to select 16 upregulated candidate genes of interest from our PPI network, and in subsequent Kaplan-Meier analyses we were able to determine that elevated expression of 14 of these genes was associated with a poorer BRCA patient prognosis. We then employed GEPIA to validate these 14 gene candidates, confirming them to all be expressed at elevated levels in BRCA relative to normal tissue controls. In addition, a regulatory network consisting of 9 genes, 10 miRNAs and 3 TFs was constructed, enabling the identification of potential biomarkers of BRCA, including AURKA, RRM2, BIRC5, and E2F1. RT-PCR results suggested that significantly elevated AURKA, RRM2 and BIRC5 mRNAs expressed in the breast cancer cells than in the normal cells. Western blot results shown that E2F1 protein was highly expressed in breast cancer cells compared to normal cells. In conclusion, these candidate molecules may offer insight regarding the underlying pathogenesis of BRCA and highlight a number of potential therapeutic avenues for the treatment of breast cancer patients.

## 1. Background

Breast cancer (BRCA) remains the leading cancer-related cause of death among women and accounts for almost 25% of all female cancer diagnoses, with 2.1 million new cases in 2018 alone[1]. The incidence of breast cancer has increased drastically in the past few decades and is projected over the next few years[2]. Recent research has led to the development of many novel screening and therapeutic strategies that have significantly improved the survival outcomes for many breast cancer patients. Currently, the five-year survival rate of women diagnosed with stage I BRCA is nearly 100%, but this percentage falls dramatically to 26% when women are diagnosed with stage IV disease (~ 5% of all cases)[3]. Advanced BRCA is thus associated with a very poor prognosis. Therefore, it is particularly important to identify biomarkers capable of enabling the early diagnosis of breast cancer. Additionally, significant heterogeneity between BRCA patients also means that individual responsiveness to a given therapeutic intervention can be highly variable[4, 5]. A number of different genetic and phenotypic biomarkers have been leveraged to date to guide BRCA treatment and diagnosis, including wound-response gene

signatures[6], hypoxia-response signatures[7], 21-gene recurrence scores[8], and a 70-gene signature[9]. Despite the promise of these biomarker panels, they are relatively non-specific and can necessitate costly testing that may impose a significant financial burden on patients. This, coupled with the poorly characterized mechanistic basis of BRCA heterogeneity, underscores the importance of identifying more reliable and economical therapeutic, diagnostic, and prognostic biomarkers of BRCA.

Advances in high-throughput microarray and sequencing approaches have allowed for the development of more efficient screening approaches aimed at clarifying the genetic and epigenetic changes associated with tumor development and progression. Such microarray datasets can thus be readily leveraged to identify diagnostic and prognostic biomarkers of specific cancer types[10, 11]. Microarrays can allow for the rapid detection of all genes that are differentially expressed between tumors and normal tissues[12]. In recent years, some progress has been made in screening for potential biomarkers related to the occurrence, development and prognosis of breast cancer at the whole-genome level. For instance, DNAPT3 was first found to play an oncogenic role in breast cancer initiation and development by Hou et al[13]. A recent study also conducted WGCNA and found that immune response- and ubiquitin-mediated proteolysis can be used as prognostic and predictive biomarkers in BRCA [14]. Furthermore, studies have identified numerous candidate biomarkers associated with the prognosis and metastasis of breast cancer[15, 16]. Simultaneously, several studies have demonstrated that some microRNAs (miRNAs) and transcription factors (TFs) play critical regulatory roles in BRCA[17–20]. While many BRCA-related differentially expressed genes, microRNAs, and TFs have been studied in recent years, the detailed pathogenesis of BRCA remains unclear due to limited understanding of mRNA-miRNA-TF relationships in this oncogenic context.

In the present study, we downloaded four publicly available microarray datasets (GSE45827, GSE42568, and GSE61304 from the Gene Expression Omnibus (GEO) database. Using Venn diagrams, the GEO2R analysis tool, and the DAVID platform we were able to identify BRCA-related DEGs and to conduct gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of these genes. We further generated a DEG protein-protein interaction (PPI) network and leveraged this network to identify core BRCA genes that were found to be significantly associated with BRCA patient prognosis in survival analyses conducted using the Kaplan-Meier plotter database. Subsequently, a prognosis-related gene-miRNA targets interaction network analysis was performed, and TF-mediated regulation of prognosis-related genes was analyzed. Furthermore, experimental verification was conducted. This work will enable the identification of core genes, miRNAs, and TFs associated with BRCA, thereby facilitating the construction of a prognosis-related gene-target miRNA-TF regulatory network that can clarify the mechanisms of breast cancer pathogenesis and highlight novel candidate molecular targets.

## 2. Materials And Methods

### 2.1 Microarray datasets and DEG identification

We downloaded the gene expression data from the GSE45827, GSE42568, and GSE61304 microarray datasets. All three datasets compared gene expression profiles in BRCA tumors and normal breast tissue samples from human patients, and all three microarrays were conducted using GPL570 Platforms([HG-U133\_Plus\_2] Affymetrix Human Genome U133 Plus 2.0 Array). In total, these datasets incorporated 142 BRCA tissues and 11 control tissues, 104 BRCA tissues and 17 control tissues, and 58 BRCA tissues and 4 control tissues, respectively. Subsequently, we used the GEO2R online tool[21] to identify DEGs by comparing BRCA and normal control tissue gene expression levels, using  $|\log FC| > 2$  and adjusted P-value  $< 0.05$  as criteria for differential expression. We then loaded the resultant DEG lists into an online Venn diagram software in order to identify those DEGs that were common to all three of these DEG datasets.

### 2.2 Functional enrichment analyses

In order to explore the biological significance of identified DEGs, we conducted both GO and KEGG enrichment analyses on DEG gene lists. GO analyses offer direct insight into the annotated characteristics and functions of genes contained within genomic or transcriptomic datasets, offering a high-level view of processes or properties that may be impacted in a given

biological context[22]. KEGG analyses, in contrast, provide insight into the signaling and biological pathways that are impacted in a given context, as in response to a particular treatment or in cells impacted by a particular disease state [23, 24]. We utilized the DAVID tool (<https://david.ncifcrf.gov/>) to conduct GO and KEGG analyses[25], with  $P < 0.05$  as the significance threshold.

## 2.3 PPI network construction and analysis

We constructed our PPI network using the online STRING tool (<https://string-db.org/>)[26] in an effort to further understand the mechanistic basis for BRCA development and/or progression. We employed the Cytoscape tool[27] to visualize this network and to examine the relationships between the DEGs found therein, with network edges representing confidence and with medium confidence cut-off (0.400) being used for network construction. Modules within this network were identified using the Molecular Complex Detection (MCODE) application within Cytoscape using the following criteria: degree cutoff = 2, max. Depth = 100, k-core = 2, and node score cutoff = 0.2.

## 2.4 DEG validation and prognostic analysis

We employed the Kaplan–Meier Plotter (Online: <http://kmplot.com/>) database in order to explore the prognostic relevance of key hub genes identified in our PPI network module analysis, as this database compiles cancer patient survival data from sources including GEO, TCGA, and EGA[28].

For this analysis we conducted overall survival analyses for genes of interest, with  $P < 0.05$  as the significance threshold. We additionally utilized GEPIA (<http://gepia.cancer-pku.cn/>) to explore RNA-seq datasets derived from thousands of TCGA and GTEx samples[29] in an effort to validate the relevance of specific DEGs to BRCA progression.

## 2.5 Prediction of target miRNAs associated with prognosis-related genes and construction of gene-miRNA target interaction network

miRNA targets of core genes were predicted using miRWalk 2.0 (<http://www.umm.uniheidelberg.de/apps/zmf/mirwalk/>) [30], TargetScan7.2 ([http://www.targetscan.org/vert\\_72/](http://www.targetscan.org/vert_72/)) [31] and miRDB (<http://www.mirdb.org/>) [32]. To ensure the accuracy of these results, the miRNA targets were selected based on the intersection of results from these three databases. A regulatory network incorporating these core genes and putative targeting miRNAs was then constructed using the Cytoscape software.

## 2.6 Construction of a prognosis-related gene-target miRNA-transcription factor (TF) regulatory network

miRNAs and TFs function as trans-regulatory factors that can control broad gene regulatory networks. To explore TFs that can regulate prognosis-related genes in BRCA, we screened the TRRUST v2 database (<https://www.grnpedia.org/trrust/>), which is a curated database of human and mouse transcriptional regulatory network that provides insight into the regulation of TF-target relationships [33]. For these analyses, the species was set to 'human', with  $p < 0.05$  as a threshold for significance. A mRNA-miRNA-TF regulatory network was then constructed to assess the functional roles of identified genes.

## 2.7 Quantitative real-time PCR and Western blot analysis

Human breast cancer cell line MDA-MB-231 (CL-0150) and MCF-10A (CL-0525) cell line were purchased from the Procell Life Science & Technology Co., Ltd. (Wuhan, China). Total RNA was isolated from cultured cells using by Trizol reagent (Aidlab, China), subsequently, 5  $\mu\text{g}$  RNA was converted to cDNA by using the HiScript Reverse Transcriptase (VAZYME, China) following the manufacturer's protocol. Real-time PCR was performed using SYBR Green Master Mix (VAZYME, China) and analyzed with QuantStudio 6 Real-time PCR System (Applied Biosystems, USA). GAPDH was applied as an endogenous control. Relative gene expression level was calculated by the  $2^{-\Delta\Delta\text{CT}}$  method between case and controls. Primer sequences used in the study are shown in **Supplementary Table 1**.

The MDA-MB-231 and MCF-10A cells were lysed with PMSF lysis (Aladdin, China) buffer. The concentrations of protein were determined by the BCA protein assay (Beyotime, China) according to the manufacturer's protocol. Equal amounts of protein

were loaded and analyzed by 10% SDS- polyacrylamide gel electrophoresis (Biosharp, China) and transferred to a PVDF membrane (Millipore, USA). The size-cut membrane was blocked in TBST with 5% non-fat milk for 2 h at room temperature and then incubated overnight at 4 °C with the primary antibody. After that, the membrane was incubated with HRP conjugated secondary antibody at room temperature for 2 h. Finally, ECL reagent was performed to determine the antigen-antibody complexes. An rabbit anti-human GAPDH polyclonal antibody(Goodhere, China), mouse anti-human E2F1 monoclonal antibody(Proteintech Group, Inc, China) and HRP conjugated goat anti-rabbit/mouse antibody(Boster, China) were used.

## 3. Results

### 3.1 BRCA-related DEG identification

By pooling data from three GEO microarray datasets, we were collectively able to analyze 304 BRCA tumor tissue samples and 32 normal breast tissue control samples. With the aid of the GEO2R tool, we were able to extract 2384, 1176, and 260 BRCA-related DEGs from the GSE45827, GSE42568, and GSE61304 datasets, respectively. Using a Venn diagram approach, we then identified 85 DEGs that were common to all three datasets. Of these DEGs, 54 were downregulated ( $\log_{2}FC < 0$ ) and 31 were upregulated( $\log_{2}FC > 0$ ) in BRCA tumor tissues (**Supplementary Table 2 & Fig. 1**).

### 3.2 Functional enrichment analyses of identified DEGs

In order to understand the functional roles of these 85 BRCA-associated DEGs, we next conducted GO and KEGG analyses thereof. GO analyses examined the biological processes (BPs), cellular components (CCs), and molecular functions (MFs) with which these DEGs were affiliated. The GO BP analysis revealed the upregulated DEGs to be primarily enriched for processes including mitotic nuclear division, cell division, G2/M transition of mitotic cell cycle, collagen catabolic process, and endodermal cell differentiation, whereas downregulated DEGs were enriched lipid metabolic processes, lipid transport, regulation of cardiac muscle contraction by regulation of the release of sequestered calcium ions, positive regulation of cell proliferation, and positive regulation of cell-matrix adhesion. The GO CC analysis revealed upregulated DEGs to be primarily enriched in the proteinaceous extracellular matrix, midbody, condensed chromosome kinetochore, centriole, and microtubule cytoskeleton, whereas downregulated DEGs were enriched in the extracellular space, extracellular region, lipid particles, proteinaceous extracellular matrix, and the apical plasma membrane. With respect to the GO MF analysis, upregulated DEGs were enriched for functions including collagen binding, heparin binding, ATP binding, and metalloendopeptidase activity, with downregulated DEGs being enriched lipid binding, heparin binding, lipoprotein particle binding, transporter activity, zinc-dependent alcohol dehydrogenase activity (Table 1).

We further conducted a KEGG pathway analysis of these DEGs. This approach revealed the upregulated DEGs to primarily be associated with the p53 signaling pathway, ECM-receptor interactions, and oocyte meiosis, whereas downregulated DEGs were primarily associated with the PPAR signaling pathway, tyrosine metabolism, cytochrome P450 drug metabolism, and protein digestion and absorption. KEGG pathway enrichment analysis results for upregulated and downregulated DEGs are shown in **Fig. 2**.

Table 1  
GO enrichment analysis of BRCA-related DEGs (The terms with the top 5 P-values are shown)

Expression	Category	Term	Count	P-value	Fold enrichment	FDR
Up-regulated	GOTERM_BP_DIRECT	GO:0007067 ~ mitotic nuclear division	11	9.55E-12	24.02601457	1.31E-08
	GOTERM_BP_DIRECT	GO:0051301 ~ cell division	9	1.29E-07	13.92884793	1.77E-04
	GOTERM_BP_DIRECT	GO:0000086 ~ G2/M transition of mitotic cell cycle	6	4.06E-06	23.72309866	0.005568755
	GOTERM_BP_DIRECT	GO:0030574 ~ collagen catabolic process	5	4.88E-06	42.31854839	0.006689601
	GOTERM_BP_DIRECT	GO:0035987 ~ endodermal cell differentiation	4	1.46E-05	80.24850657	0.020038146
	GOTERM_CC_DIRECT	GO:0005578 ~ proteinaceous extracellular matrix	8	2.10E-07	17.5483871	2.23E-04
	GOTERM_CC_DIRECT	GO:0030496 ~ midbody	6	2.03E-06	27.34283571	0.002166272
	GOTERM_CC_DIRECT	GO:0000777 ~ condensed chromosome kinetochore	4	3.89E-04	27.02855024	0.413450196
	GOTERM_CC_DIRECT	GO:0005814 ~ centriole	4	8.34E-04	20.80959178	0.885431205
	GOTERM_CC_DIRECT	GO:0015630 ~ microtubule cytoskeleton	4	0.001454401	17.16411585	1.539162614
	GOTERM_MF_DIRECT	GO:0005518 ~ collagen binding	4	1.62E-04	36.30322581	0.177777537
	GOTERM_MF_DIRECT	GO:0008201 ~ heparin binding	4	0.002811565	13.61370968	3.047438273
	GOTERM_MF_DIRECT	GO:0005524 ~ ATP binding	8	0.013957193	2.913971302	14.3155518
	GOTERM_MF_DIRECT	GO:0004222 ~ metalloendopeptidase activity	3	0.017098388	14.45703683	17.26865537
	Down-regulated	GOTERM_BP_DIRECT	GO:0006629 ~ lipid metabolic process	6	5.42E-05	14.26072187
GOTERM_BP_DIRECT		GO:0006869 ~ lipid transport	4	0.00103261	19.63976608	1.52497516

Abbreviations: BP, biological process; CC, cellular component; Count, the number of enriched genes in each term; FDR, false discovery rate; MF, molecular function; BRCA, breast cancer.

Expression	Category	Term	Count	P-value	Fold enrichment	FDR
	GOTERM_BP_DIRECT	GO:0010881 ~ regulation of cardiac muscle contraction by regulation of the release of sequestered calcium ion	3	0.001115403	58.91929825	1.646301724
	GOTERM_BP_DIRECT	GO:0008284 ~ positive regulation of cell proliferation	7	0.001276576	5.605340963	1.882087273
	GOTERM_BP_DIRECT	GO:0001954 ~ positive regulation of cell-matrix adhesion	3	0.001499269	50.88484848	2.207004361
	GOTERM_CC_DIRECT	GO:0005615 ~ extracellular space	14	4.96E-05	3.713932195	0.055590194
	GOTERM_CC_DIRECT	GO:0005576 ~ extracellular region	14	3.01E-04	3.107246377	0.33641381
	GOTERM_CC_DIRECT	GO:0005811 ~ lipid particle	4	7.87E-04	21.65656566	0.878449732
	GOTERM_CC_DIRECT	GO:0005578 ~ proteinaceous extracellular matrix	6	8.18E-04	8	0.912134466
	GOTERM_CC_DIRECT	GO:0016324 ~ apical plasma membrane	6	0.001180873	7.367697595	1.314786411
	GOTERM_MF_DIRECT	GO:0008289 ~ lipid binding	4	0.005831972	10.64711447	6.708154954
	GOTERM_MF_DIRECT	GO:0008201 ~ heparin binding	4	0.00684227	10.04821429	7.827449644
	GOTERM_MF_DIRECT	GO:0071813 ~ lipoprotein particle binding	2	0.012086409	160.7714286	13.44231459
	GOTERM_MF_DIRECT	GO:0005215 ~ transporter activity	4	0.012887447	7.958981612	14.27185096
	GOTERM_MF_DIRECT	GO:0004024 ~ alcohol dehydrogenase activity, zinc-dependent	2	0.014486531	133.9761905	15.90609305
Abbreviations: BP, biological process; CC, cellular component; Count, the number of enriched genes in each term; FDR, false discovery rate; MF, molecular function; BRCA, breast cancer.						

### 3.3 PPI network construction and analysis

We next generated a PPI network in an effort to better understand the relationships between the 85 BRCA-related DEGs identified above. To that end, we leveraged the STRING database to construct a network that successfully incorporated 63 of these 85 DEGs (28 and 35 up- and down-regulated, respectively). The final network contained 63 nodes and 215 edges (Fig. 3A). Using the MCODE plugin, we then identified 16 nodes corresponding to hub genes within this network using filtering degree  $\geq 14$  as a criterion for hub identification (Fig. 3B).

### 3.4 Assessment of the prognostic relevance of core genes

Next, we tested the prognostic relevance of these 16 core genes using the Kaplan-Meier Plotter database. This analysis revealed that the upregulation of 14 of these genes was significantly associated with poorer patient survival ( $P < 0.05$ , **Supplementary Table 3 & Fig. 4**). We then further validated the expression of these 14 prognostic hub genes using the GEPIA tool, confirming that all 14 of these genes were expressed at higher levels in BRCA tumor tissue samples relative to normal breast tissue ( $P < 0.01$ , **Fig. 5**). Importantly, patients with higher expression of these genes had poorer outcomes than did patients with lower expression thereof. The gene symbols, full names, and functions of these prognosis-related genes are shown in Table 2.

Table 2  
Functions of key prognosis-related genes

Gene symbol	Full name	Function
TOP2A	Topoisomerase (DNA) II $\alpha$	TOP2A can be therapeutic targets for anticancer, and mutations in this gene is associated with the development of drug resistance[34].
ASPM	Abnormal spindle microcephaly associated	ASPM plays a role in the regulation of mitotic spindle, with a priority role in regulating neurogenesis. Overexpression of ASPM links with tumor progression and poor prognosis in various tumors, including breast cancer[35].Furthermore, it has been reported that ASPM promotes tumor development by regulating Wnt- $\beta$ -catenin signaling[36, 37].
NEK2	Never in Mitosis (NIMA) Related Kinase 2	NEK2 plays a critical role in multiple aspects of mitotic processes, such as centrosome duplication and separation, microtubule stabilization, kinetochore attachment, and spindle assembly checkpoint[38].NKK2 is frequently overexpressed in a variety of human malignancies, has been implicated in tumorigenesis, tumor progression and drug resistance[39, 40].
RRM2	ribonucleotide reductase M2	RRM2, a rate-limiting enzyme for DNA synthesis and repair, overexpression exhibited enhanced cellular invasiveness, serving as a potential prognostic marker and a possible therapeutic target for cancer treatment.
UBE2T	Ubiquitin-conjugating enzyme E2T	UBE2T plays a key role in cellular processes, for example, signal transduction, cell cycle control and tumorigenesis[41].
CEP55	Centrosomal protein 55	CEP55 localizes to the centrosome throughout the cell cycle, and to regulate both the mitotic spindle and microtubule organization[42].
ANLN	Anillin, actin binding protein	ANLN plays an important role in cytokinesis, highly express in numerous cancer types[43, 44].
DLGAP5	discs large associated protein 5	DLGAP5 is a regulator of cell cycle involved in carcinogenesis. Elevated DLGAP5 expression correlates with poor prognosis in breast cancer patients[45].
CCNB2	Cyclin B2	Cyclin B2 may regulate the cell cycle with transforming growth factor beta-mediated. Aberrant expression of CCNB2 is correlated with tumor invasion, metastasis, poor prognosis of various human cancers[46].
BIRC5	Survivin, Baculoviral IAP repeat containing 5	BIRC5 may inhibit apoptosis and is usually overexpressed in most tumors[47].
TPX2	Targeting protein for Xenopus kinesin-like protein 2	TPX2 is a microtubule-associated protein required for mitosis and spindle assembly.
AURKA	Aurora kinase A	AURKA, a serine/threonine kinase, has important functions in regulating cell cycle and mitosis.
KIF2C	Kinesin family member 2C	KIF2C, a crucial mitotic regulator, is highly expressed in many cancer types[48].
UBE2C	Ubiquitin-conjugating enzyme E2C	UBE2C is overexpressed in many human tumor types[49]

### 3.5 Integrated mRNA-miRNA interaction network analysis

Subsequently, Candidate target miRNAs associated with fourteen prognosis-related genes were predicted using miRWalk, TargetScan, and miRDB. Overlapping miRNAs predicted by all three databases were then selected, with a specific focus on

miRNAs capable of binding to 3'UTR regulatory regions. An interaction network incorporating these miRNAs and core genes was then constructed using Cytoscape (Fig. 6). The network contained 217 nodes and 214 edges. By comparing the targets of core genes, RRM2, AURKA, NEK2, and BIRC5 were found to be potential targets of 53 miRNAs, 41 miRNAs, 30 miRNAs and 28 miRNAs, respectively. Those miRNAs that simultaneously regulated a high number of gene cross-links ( $\geq 2$ ) were selected (Supplementary Table 4).

### 3.6 Construction of a mRNA-target miRNA-TF regulatory network

The selected miRNAs and corresponding genes may play a crucial role in the pathogenesis of BRCA. Hence, RRM2, CEP55, ANLN, UBE2C, NEK2, TOP2A, AURKA, KIF2C, and BIRC5 were considered to be key genes of interest in this context. In order to further understand the function of these genes, a miRNA-target gene-TF regulatory network for these key genes was constructed containing 9 genes, 10 miRNAs, and 3 TFs (Fig. 7).

### 3.7 Verification of candidate molecules expression by RT-PCR and Western Blot

For the purpose of enhancing the credibility of results, the mRNA levels of the three hub genes (AURKA, RRM2 and BIRC5) were measured by RT-PCR and E2F1 protein expression was determined with western blot in MDA-MB-231 and MCF-10A cells. As illustrated in Fig. 8A, all three genes were significantly upregulated in breast cancer cells ( $P < 0.01$ ), as predicted by the bioinformatics analysis. E2F1 protein was highly expressed in breast cancer cells as shown in Fig. 8B.

## 4. Discussion

Despite many recent therapeutic innovations, BRCA remains the leading cancer-related cause of death among women. BRCA remains a highlight heterogeneous disease, with four major BRCA subtypes having been defined based upon tumor expression of estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2, and the proliferative marker Ki67 (MKI67) [50–52]. This classification scheme is typically used to predict BRCA patient prognosis and to define treatment strategies together with information pertaining to tumor histological grade, type, and TNM stage. Given the high degree of BRCA tumor heterogeneity, however, it is difficult to predict tumor response to treatment or associated patient outcomes. Additionally, the cooperative effects of TFs and miRNAs in gene regulation remain largely unclear. Recent molecular studies have highlighted a number of genes and signaling pathways that govern BRCA development and progression[53–56], but further work is needed to fully elucidate the mechanistic basis for this complex disease. This study represents the first attempt to construct a trans-regulatory mRNA-miRNA-TF network in order to identify genes, miRNAs, and TFs related to BRCA, thereby clarifying the pathogenesis of BRCA at the molecular level and improving the detection, treatment, and prognostic assessment of this disease.

In the present study, we conducted an integrated bioinformatics analysis of three microarray datasets, leading us to identify 85 BRCA-related DEGs (54 and 31 down- and up-regulated, respectively). We found that our upregulated DEGs were mainly enriched in the cell cycle, cell differentiation, oocyte meiosis, and the p53 signaling pathway, which plays a significant role in the occurrence and progression of tumors. Mitotic catastrophe/error can also contribute to tumorigenesis [57, 58]. As such, mitosis has become an attractive anti-tumor target, with paclitaxel being one of the most successful drugs targeting this pathway that is commonly used to treat BRCA[59, 60]. Previous studies have shown that the p53 signaling pathway controls cell cycle regulation, mitosis, reproduction, and the inhibition of neoplastic transformation and tumor progression[61, 62]. The down-regulated genes were enriched in the lipid metabolic process, lipid transport, regulation of cardiac muscle contraction by regulation of the release of sequestered calcium ions, positive regulation of cell proliferation, positive regulation of cell-matrix adhesion, tyrosine metabolism, cytochrome P450 drug metabolism, protein digestion and absorption, and PPAR signaling pathway. Previous studies have reported that PPAR, as a ligand-activated transcription factor, regulates lipid metabolism and inflammation and contributes to the survival of breast cancer cells[63, 64]. Six DEGs were enriched in these two signaling pathways, including CDK1, CCNB2, RRM2, and AURKA, which have been proven to be involved in the mitotic process[65]. All in

all, our results are consistent with the above theories. BRCA is mostly correlated with cell cycle and cell differentiation, mainly for mitotic cell cycle and enriching p53 signaling pathway.

We then constructed PPI network from identified DEGs and identified 16 DEGs as key genes (degree  $\geq 14$ ), of which 14 were found to be significantly associated with poorer BRCA patient survival in the Kaplan-Meier Plotter database. We further confirmed that all 14 of these were significantly upregulated in BRCA tissue samples relative to normal tissue controls in a GEPIA analysis ( $P < 0.05$ ). Mitotic error is not the only cause of BRCA, but it is an important pathogenic factor and can explain one possible mechanism whereby BRCA can develop. Most of the core genes screened out in this study were related to mitosis, including ASPM[66], NEK2[67], RRM2, CEP55[42], CCNB2, BIRC5[47], KIF2C, AURKA, GLGAP5, and TPX2[68]. Of these genes, several were previously correlated with the occurrence, development, and prognosis of BRCA. Accumulating evidence indicates that overexpression of AURKA causes drug resistance, promotes progression, and predicts poor prognosis for breast cancer. Suppression AURKA can accelerate PI3K-pathway inhibitor efficacy in breast cancer resulting in apoptosis and tumor regression[69]. Zheng et al. have also shown that AURKA translocates to the nucleus and causes carcinogenic activities in malignant tumor cells by enhancing breast cancer stem cell phenotypes[70]. Furthermore, UBE2C has been shown to regulate the activity of AURKA by controlling the activity of APC/C[49]. Overexpression of UBE2C in breast microcalcification lesions suggests that UBE2C is also related to the occurrence and development of breast cancer[71]. One study found that AURKA, TPX2, and DLGAP5 were highly upregulated in non-small cell lung cancer and correlated with one another. TPX2 and DLGAP5 are also phosphorylated by AURKA during the cell cycle[68]. Chen et al. found TPX2 silencing negatively regulates PI3K / Akt, activates the p53 signaling pathway, inhibits breast cancer cell proliferation, and promotes cell apoptosis, indicating that TPX2 may be a potential target for the treatment of breast cancer[72]. Another study has also found that TPX2 may be a novel prognostic marker of breast cancer[73].

In one prior analysis, RRM2 was found to be highly expressed in diverse cancers, and inhibiting RRM2 overexpression in retroperitoneal liposarcoma can impede tumor progression via downregulating the Akt/mTOR/4EBP1 pathway[74]. Meanwhile, Koppenhafer et al. found that CDK2-mediated RRM2 downregulation promotes DNA damage and cell apoptosis in Ewing sarcoma via inhibiting the ATR-CHK1 pathway[75]. Increasing evidence suggests that miRNAs and lncRNAs related to RRM2 can modulate tumor progression by regulating the expression of RRM2[76]. For example, in BRCA tissues, lncRNA DSCAM-AS1 directly targets miR-204-5p, stimulates proliferation and invasion, and impedes BRCA cell apoptosis by inhibiting miR-204-5p and facilitating RRM2 expression[77]. Consistent with these reports, we found that RMM2 upregulation was associated with a poorer BRCA patient prognosis. According to a previous study, CEP55 is a cell fate determinant in the context of perturbed mitosis in breast cancer. Loss of CEP55 renders breast cancer cells sensitive to anti-mitotic drugs by prematurely activating CDK1 / cyclin B and CDK1 caspase-dependent mitotic cell death. Furthermore, blocking MEK1/2-PLK1 signaling may reduce the outgrowth of MYC-CEP55-dependent basal-like, triple-negative breast cancers[78]. Our results were consistent with these findings, indicating that patients with higher levels of these genes expression may have poorer prognosis.

miRNAs are endogenous non-coding RNA molecules, that can modulate gene expression at the post-transcriptional level by reducing or inhibiting the translation of target genes[79]. This study screened out mRNA-miRNA target pairs using online prediction tools. RRM2, AURKA, NEK2, and BIRC5 were the top four potential miRNA targets in this study ( $\geq 28$ ). Recent studies have found that some miRNAs are involved in the development of cancer, including BRCA[17–19]. TFs also serve as trans-regulatory factors, much as do miRNAs. To gain a further understanding of key gene functions in this oncogenic context, an mRNA-target miRNA-TF regulatory network was therefore constructed. E2F1 was identified as a co-regulator of KIF2C, TOP2A, RRM2, AURKA, and BIRC5. MED1 was found to regulate AURKA and UBE2C, while SP1 was predicted to regulate KIF2C and BIRC5. Several studies have shown that E2F1 plays a critical role in the regulation of both cell proliferation and apoptosis[80, 81]. AURKA overexpression has also been observed in many cancers. He et al. [82] found that the transcriptional activity of E2F1 was increased by AURKA overexpression. We also found that AURKA was associated with miRNA cluster expression by regulating E2F1. AURKA may thus be a putative therapeutic target in BRCA. In addition, a previous study has reported that silencing RRM2 to inhibit the up-regulation of RRM2 can enhance DNA damage, which could in turn improve the efficacy of topoisomerase I inhibitors[83]. ASA has also been found to promote TRAIL-induced apoptosis

by down-regulating BIRC5 gene expression, which was mediated by inhibiting E2F-1 binding to the BIRC5 promoter region[84]. The experiments in vitro further conducted to validate. These results suggest that AURKA, RRM2, BIRC5, and related regulatory factors including E2F1 and miRNAs may be key molecules involved in controlling the onset and progression of BRCA.

## 5. Conclusions

In conclusion, we were able to leverage three microarray datasets in an integrated bioinformatics analysis in order to identify 85 BRCA-related DEGs, including 16 key candidate genes that were detected through a PPI network analysis. Additionally, a regulatory network consisting of 9 genes, 10 miRNAs and 3 TFs was constructed, revealing the co-regulatory associations among these key genes, corresponding targeted miRNAs, and TFs in BRCA. These key molecules may be regarded as potential biomarkers of BRCA, with AURKA, RRM2, BIRC5, and E2F1 being particularly relevant targets of interest. Nevertheless, further prospective studies will be required in order to validate our findings. Even so, the results of this analysis may offer valuable insight into the molecular basis for BRCA carcinogenesis and progression, while also identifying reliable candidate biomarkers that can guide the precision diagnosis and individualized treatment of BRCA.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Availability of data and materials

The data used to support the findings of this study are included within the article. The data and materials in this study are available from the corresponding author on reasonable request.

### Conflict of interest

All the authors declared that there are no conflicts of interest relevant to this study.

### Funding

Not applicable.

### Authors' contributions

Ying Wei contributed to conceive and design the experiments, authored or reviewed drafts of the paper. Shipeng Zhang performed the experiments, analyzed the data. Li Xiao conducted the literature search and contributed reagents/materials/analysis tools. Ying Wei gave the final approval of the version to be submitted. Yingqiang Fu prepared figures and/or tables. Jing Zhou wrote drafts of the article. Yi Ye and Linchuan Liao revised and reviewed the article. All authors read and approved the final draft.

### Acknowledgments

The authors would like to express their sincere gratitude to all those who have given their assistance in the preparation of this paper.

## References

1. 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: A Cancer Journal for Clinicians* 2018, **68**(6):394-424.
2. Cheng Y, Yan Y, Gong J, Yang N, Nie S: **Trends in incidence and mortality of female breast cancer during transition in Central China.** *Cancer Manag Res* 2018, **10**:6247-6255.
3. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, Jemal A, Kramer JL, Siegel RL: **Cancer treatment and survivorship statistics, 2019.** *CA Cancer J Clin* 2019, **69**(5):363-385.
4. Katanoda K, Matsuda T: **Five-year relative survival rate of breast cancer in the USA, Europe and Japan.** *Jpn J Clin Oncol* 2014, **44**(6):611.
5. Paterson R, Phillips KA: **Genetic testing in women with breast cancer: implications for treatment.** *Expert Rev Anticancer Ther* 2017, **17**(11):991-1002.
6. Chang HY, Nuyten DS, Sneddon JB, Hastie T, Tibshirani R, Sorlie T, Dai H, He YD, van't Veer LJ, Bartelink H *et al*: **Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival.** *Proc Natl Acad Sci U S A* 2005, **102**(10):3738-3743.
7. Nuyten DS, Hastie T, Chi JT, Chang HY, van de Vijver MJ: **Combining biological gene expression signatures in predicting outcome in breast cancer: An alternative to supervised classification.** *Eur J Cancer* 2008, **44**(15):2319-2329.
8. Puhalla SL, Davidson NE: **Breast cancer: The 21-gene recurrence score - biology remains at the forefront.** *Nat Rev Clin Oncol* 2016, **13**(8):470-472.
9. van't Veer LJ, Dai HY, van de Vijver MJ, He YD, Augustinus A M Hart MM, Hans L Peterse, Karin van der Kooy, Matthew J Marton, Anke T Witteveen, George J Schreiber, Ron M Kerkhoven, Chris Roberts, Peter S Linsley, René Bernards, Stephen H Friend: **Gene expression profiling predicts clinical outcome of breast cancer.** *Nature* 2002, **415**(6871):530-536.
10. Kulasingam V, Diamandis EP: **Strategies for discovering novel cancer biomarkers through utilization of emerging technologies.** *Nat Clin Pract Oncol* 2008, **5**(10):588-599.
11. Liu J, Meng H, Li S, Shen Y, Wang H, Shan W, Qiu J, Zhang J, Cheng W: **Identification of Potential Biomarkers in Association With Progression and Prognosis in Epithelial Ovarian Cancer by Integrated Bioinformatics Analysis.** *Front Genet* 2019, **10**:1031.
12. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW: **Cancer genome landscapes.** *Science* 2013, **339**(6127):1546-1558.
13. Hou L, Chen M, Wang M, Cui X, Gao Y, Xing T, Li J, Deng S, Hu J, Yang H *et al*: **Systematic analyses of key genes and pathways in the development of invasive breast cancer.** *Gene* 2016, **593**(1):1-12.
14. Zhao Q, Song W, He DY, Li Y: **Identification of key gene modules and pathways of human breast cancer by co-expression analysis.** *Breast Cancer* 2018, **25**(2):213-223.
15. Cai Y, Mei J, Xiao Z, Xu B, Jiang X, Zhang Y, Zhu Y: **Identification of five hub genes as monitoring biomarkers for breast cancer metastasis in silico.** *Hereditas* 2019, **156**:20.
16. Xu YH, Deng JL, Wang LP, Zhang HB, Tang L, Huang Y, Tang J, Wang SM, Wang G: **Identification of Candidate Genes Associated with Breast Cancer Prognosis.** *DNA Cell Biol* 2020, **39**(7):1205-1227.
17. Wu ZS, Wang CQ, Xiang R, Liu X, Ye S, Yang XQ, Zhang GH, Xu XC, Zhu T, Wu Q: **Loss of miR-133a expression associated with poor survival of breast cancer and restoration of miR-133a expression inhibited breast cancer cell growth and invasion.** *BMC Cancer* 2012, **12**:51.
18. Zhang C, Liu K, Li T, Fang J, Ding Y, Sun L, Tu T, Jiang X, Du S, Hu J *et al*: **miR-21: A gene of dual regulation in breast cancer.** *Int J Oncol* 2016, **48**(1):161-172.
19. Zheng S, Li M, Miao K, Xu H: **SNHG1 contributes to proliferation and invasion by regulating miR-382 in breast cancer.** *Cancer Manag Res* 2019, **11**:5589-5598.

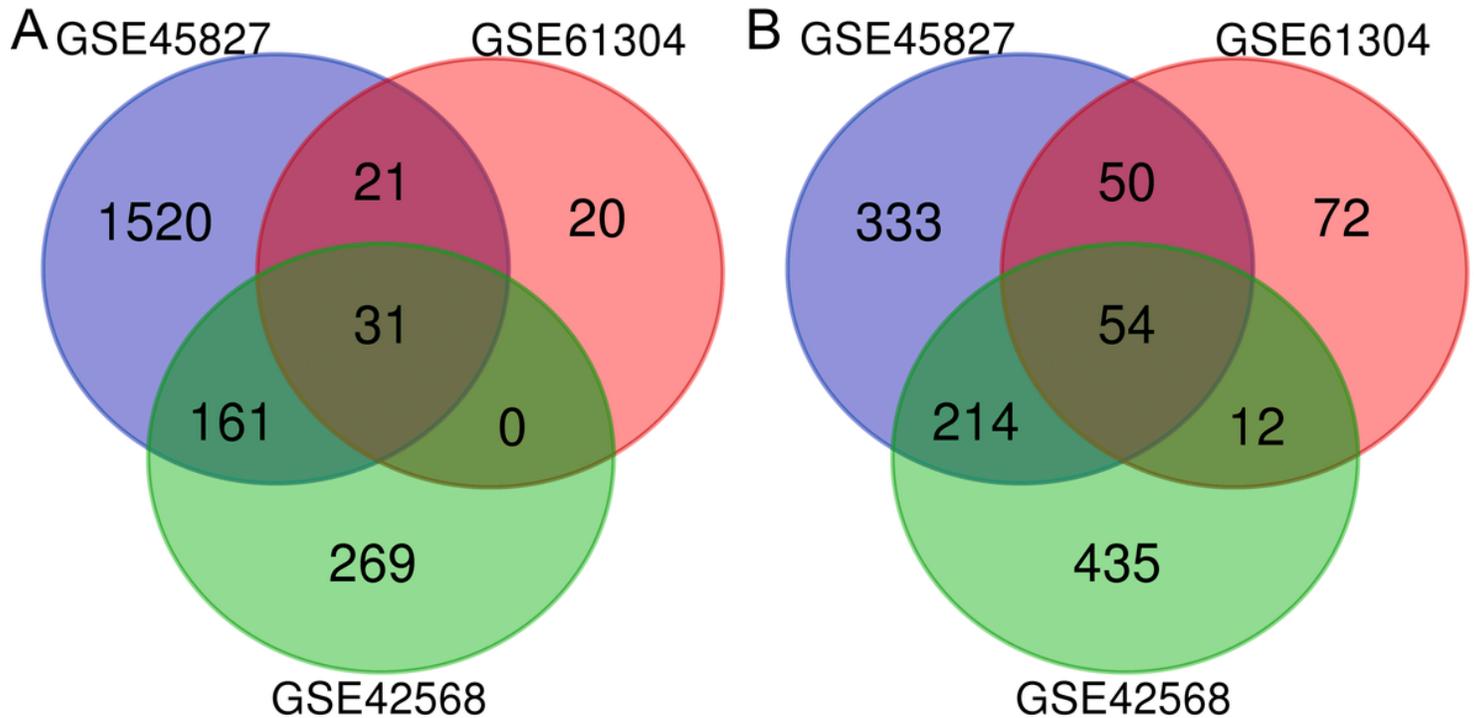
20. Yu Z, He Q, Xu G: **Screening of Prognostic Factors in Early-Onset Breast Cancer.** *Technol Cancer Res Treat* 2020, **19**:1533033819893670.
21. Davis S, Meltzer PS: **GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor.** *Bioinformatics* 2007, **23**(14):1846-1847.
22. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al.* **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25**(1):25-29.
23. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M: **KEGG: Kyoto Encyclopedia of Genes and Genomes.** *Nucleic Acids Research*, 1999, **27**(1):29-34.
24. Du J, Yuan Z, Ma Z, Song J, Xie X, Chen Y: **KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model.** *Mol BioSyst* 2014, **10**(9):2441-2447.
25. Huang da W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protoc* 2009, **4**(1):44-57.
26. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP *et al.* **STRING v10: protein-protein interaction networks, integrated over the tree of life.** *Nucleic Acids Res* 2015, **43**(Database issue):D447-452.
27. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Res* 2003, **13**(11):2498-2504.
28. Szász MA, Lániczky A, Nagy A, Förster S, Green JE, Boussioutas A, Busuttill R, Szabó A, Hark K, Gyórfy B: **Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients.** *Oncotarget* 2016, **7**(31):49322-49333.
29. 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.
30. Dweep H, Sticht C, Pandey P, Gretz N: **miRWalk – Database: Prediction of possible miRNA binding sites by “walking” the genes of three genomes.** *Journal of Biomedical Informatics* 2011, **44**(5):839-847.
31. Zhang G, Xu S, Zhang Z, Zhang Y, Wu Y, An J, Lin J, Yuan Z, Shen L, Si T: **Identification of Key Genes and the Pathophysiology Associated With Major Depressive Disorder Patients Based on Integrated Bioinformatics Analysis.** *Front Psychiatry* 2020, **11**:192.
32. Chen Y, Wang X: **miRDB: an online database for prediction of functional microRNA targets.** *Nucleic Acids Res* 2020, **48**(D1):D127-D131.
33. Han H, Cho JW, Lee S, Yun A, Kim H, Bae D, Yang S, Kim CY, Lee M, Kim E *et al.* **TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions.** *Nucleic Acids Res* 2018, **46**(D1):D380-D386.
34. Delgado JL, Hsieh CM, Chan NL, Hiasa H: **Topoisomerases as anticancer targets.** *Biochem J* 2018, **475**(2):373-398.
35. Tang J, Lu M, Cui Q, Zhang D, Kong D, Liao X, Ren J, Gong Y, Wu G: **Overexpression of ASPM, CDC20, and TTK Confer a Poorer Prognosis in Breast Cancer Identified by Gene Co-expression Network Analysis.** *Front Oncol* 2019, **9**:310.
36. Chen X, Huang L, Yang Y, Chen S, Sun J, Ma C, Xie J, Song Y, Yang J: **ASPM promotes glioblastoma growth by regulating G1 restriction point progression and Wnt-beta-catenin signaling.** *Aging (Albany NY)* 2020, **12**(1):224-241.
37. Pai VC, Hsu CC, Chan TS, Liao WY, Chuu CP, Chen WY, Li CR, Lin CY, Huang SP, Chen LT *et al.* **ASPM promotes prostate cancer stemness and progression by augmenting Wnt-Dvl-3-beta-catenin signaling.** *Oncogene* 2019, **38**(8):1340-1353.
38. Fang Y, Zhang X: **Targeting NEK2 as a promising therapeutic approach for cancer treatment.** *Cell Cycle* 2016, **15**(7):895-907.
39. Li Y, Chen L, Feng L, Zhu M, Shen Q, Fang Y, Liu X, Zhang X: **NEK2 promotes proliferation, migration and tumor growth of gastric cancer cells via regulating KDM5B/H3K4me3.** *Am J Cancer Res* 2019, **9**(11):2364-2378.
40. Cappello P, Blaser H, Gorrini C, Lin DC, Elia AJ, Wakeham A, Haider S, Boutros PC, Mason JM, Miller NA *et al.* **Role of Nek2 on centrosome duplication and aneuploidy in breast cancer cells.** *Oncogene* 2014, **33**(18):2375-2384.

41. Lim KH, Song MH, Baek KH: **Decision for cell fate: deubiquitinating enzymes in cell cycle checkpoint.** *Cell Mol Life Sci* 2016, **73**(7):1439-1455.
42. Jeffery J, Sinha D, Srihari S, Kalimutho M, Khanna KK: **Beyond cytokinesis: the emerging roles of CEP55 in tumorigenesis.** *Oncogene* 2016, **35**(6):683-690.
43. Wang A, Dai H, Gong Y, Zhang C, Shu J, Luo Y, Jiang Y, Liu W, Bie P: **ANLN-induced EZH2 upregulation promotes pancreatic cancer progression by mediating miR-218-5p/LASP1 signaling axis.** *J Exp Clin Cancer Res* 2019, **38**(1):347.
44. Wang Z, Chen J, Zhong MZ, Huang J, Hu YP, Feng DY, Zhou ZJ, Luo X, Liu ZQ, Jiang WZ *et al.*: **Overexpression of ANLN contributed to poor prognosis of anthracycline-based chemotherapy in breast cancer patients.** *Cancer Chemother Pharmacol* 2017, **79**(3):535-543.
45. Xu T, Dong M, Li H, Zhang R, Li X: **Elevated mRNA expression levels of DLGAP5 are associated with poor prognosis in breast cancer.** *Oncol Lett* 2020, **19**(6):4053-4065.
46. Nam HJ, van Deursen JM: **Cyclin B2 and p53 control proper timing of centrosome separation.** *Nat Cell Biol* 2014, **16**(6):538-549.
47. 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.
48. Zhang GP, Shen SL, Yu Y, Yue X, Hu WJ, Li SQ: **Kinesin family member 2C aggravates the progression of hepatocellular carcinoma and interacts with competing endogenous RNA.** *Journal of Cellular Biochemistry* 2020.
49. Wang R, Song Y, Liu X, Wang Q, Wang Y, Li L, Kang C, Zhang Q: **UBE2C induces EMT through Wnt/betacatenin and PI3K/Akt signaling pathways by regulating phosphorylation levels of Aurora-A.** *Int J Oncol* 2017, **50**(4):1116-1126.
50. Sørlie T, Wang Y, Xiao C, Johnsen H, Naume B, Samaha RR, Børresen-Dale A-L: **Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms.** *BMC Genomics* 2006, **7**(1).
51. Sorlie T, Tibshirani R, Parker J, Hastie T, S. MJ, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S *et al.*: **Repeated observation of breast tumor subtypes in independent gene expression data sets.** *Natl Acad Sci U S A* 2003, **100**(14):8418-8423.
52. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ, Panel m: **Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011.** *Ann Oncol* 2011, **22**(8):1736-1747.
53. Xue D, Cheng P, Han M, Liu X, Xue L, Ye C, Wang K, Huang J: **An integrated bioinformatical analysis to evaluate the role of KIF4A as a prognostic biomarker for breast cancer.** *Onco Targets Ther* 2018, **11**:4755-4768.
54. Li G, Guo X, Tang L, Chen M, Luo X, Peng L, Xu X, Wang S, Xiao Z, Yi W *et al.*: **Analysis of BRCA1/2 mutation spectrum and prevalence in unselected Chinese breast cancer patients by next-generation sequencing.** *J Cancer Res Clin Oncol* 2017, **143**(10):2011-2024.
55. Cybulski C, Carrot-Zhang J, Kluźniak W, Rivera B, Kashyap A, Wokołorczyk D, Giroux S, Nadaf J, Hamel N, Zhang S *et al.*: **Germline RECQL mutations are associated with breast cancer susceptibility.** *Nature Genetics* 2015, **47**(6):643-646.
56. Buys SS, Sandbach JF, Gammon A, Patel G, Kidd J, Brown KL, Sharma L, Saam J, Lancaster J, Daly MB: **A study of over 35,000 women with breast cancer tested with a 25-gene panel of hereditary cancer genes.** *Cancer* 2017, **123**(10):1721-1730.
57. Levine MS, Holland AJ: **The impact of mitotic errors on cell proliferation and tumorigenesis.** *Genes Dev* 2018, **32**(9-10):620-638.
58. Mc Gee MM: **Targeting the Mitotic Catastrophe Signaling Pathway in Cancer.** *Mediators Inflamm* 2015, **2015**:146282.
59. Olziersky AM, Labidi-Galy SI: **Clinical Development of Anti-mitotic Drugs in Cancer.** *Adv Exp Med Biol* 2017, **1002**:125-152.
60. Sinha D, Duijf PHG, Khanna KK: **Mitotic slippage: an old tale with a new twist.** *Cell Cycle* 2019, **18**(1):7-15.

61. Khan H, Reale M, Ullah H, Sureda A, Tejada S, Wang Y, Zhang ZJ, Xiao J: **Anti-cancer effects of polyphenols via targeting p53 signaling pathway: updates and future directions.** *Biotechnol Adv* 2020, **38**:107385.
62. Toma-Jonik A, Vydra N, Janus P, Widłak W: **Interplay between HSF1 and p53 signaling pathways in cancer initiation and progression: non-oncogene and oncogene addiction.** *Cellular Oncology* 2019, **42**(5):579-589.
63. Wang X, Wang G, Shi Y, Sun L, Gorczynski R, Li YJ, Xu Z, Spaner DE: **PPAR-delta promotes survival of breast cancer cells in harsh metabolic conditions.** *Oncogenesis* 2016, **5**(6):e232.
64. Fan P, Abderrahman B, Chai TS, Yerrum S, Jordan VC: **Targeting Peroxisome Proliferator-Activated Receptor gamma to Increase Estrogen-Induced Apoptosis in Estrogen-Deprived Breast Cancer Cells.** *Mol Cancer Ther* 2018, **17**(12):2732-2745.
65. Penna LS, Henriques JAP, Bonatto D: **Anti-mitotic agents: Are they emerging molecules for cancer treatment?** *Pharmacol Ther* 2017, **173**:67-82.
66. Higgins J, Midgley C, Bergh AM, Bell SM, Askham JM, Roberts E, Binns RK, Sharif SM, Bennett C, Glover DM *et al*: **Human ASPM participates in spindle organisation, spindle orientation and cytokinesis.** *BMC Cell Biol* 2010, **11**:85.
67. Ha Kim Y, Yeol Choi J, Jeong Y, Wolgemuth DJ, Rhee K: **Nek2 localizes to multiple sites in mitotic cells, suggesting its involvement in multiple cellular functions during the cell cycle.** *Biochem Biophys Res Commun* 2002, **290**(2):730-736.
68. Schneider MA, Christopoulos P, Muley T, Warth A, Klingmueller U, Thomas M, Herth FJ, Dienemann H, Mueller NS, Theis F *et al*: **AURKA, DLGAP5, TPX2, KIF11 and CKAP5: Five specific mitosis-associated genes correlate with poor prognosis for non-small cell lung cancer patients.** *Int J Oncol* 2017, **50**(2):365-372.
69. Donnella HJ, Webber JT, Levin RS, Camarda R, Momcilovic O, Bayani N, Shah KN, Korkola JE, Shokat KM, Goga A *et al*: **Kinome rewiring reveals AURKA limits PI3K-pathway inhibitor efficacy in breast cancer.** *Nat Chem Biol* 2018, **14**(8):768-777.
70. 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.
71. Chou CP, Huang NC, Jhuang SJ, Pan HB, Peng NJ, Cheng JT, Chen CF, Chen JJ, Chang TH: **Ubiquitin-conjugating enzyme UBE2C is highly expressed in breast microcalcification lesions.** *PLoS One* 2014, **9**(4):e93934.
72. Chen M, Zhang H, Zhang G, Zhong A, Ma Q, Kai J, Tong Y, Xie S, Wang Y, Zheng H *et al*: **Targeting TPX2 suppresses proliferation and promotes apoptosis via repression of the PI3k/AKT/P21 signaling pathway and activation of p53 pathway in breast cancer.** *Biochem Biophys Res Commun* 2018, **507**(1-4):74-82.
73. Jiang Y, Liu Y, Tan X, Yu S, Luo J: **TPX2 as a Novel Prognostic Indicator and Promising Therapeutic Target in Triple-negative Breast Cancer.** *Clin Breast Cancer* 2019, **19**(6):450-455.
74. Zhang S, Yan L, Cui C, Wang Z, Wu J, Lv A, Zhao M, Dong B, Zhang W, Guan X *et al*: **Downregulation of RRM2 Attenuates Retroperitoneal Liposarcoma Progression via the Akt/mTOR/4EBP1 Pathway: Clinical, Biological, and Therapeutic Significance.** *Onco Targets Ther* 2020, **13**:6523-6537.
75. Koppenhafer SL, Goss KL, Terry WW, Gordon DJ: **Inhibition of the ATR-CHK1 Pathway in Ewing Sarcoma Cells Causes DNA Damage and Apoptosis via the CDK2-Mediated Degradation of RRM2.** *Mol Cancer Res* 2020, **18**(1):91-104.
76. Yang Y, Li S, Cao J, Li Y, Hu H, Wu Z: **RRM2 Regulated By LINC00667/miR-143-3p Signal Is Responsible For Non-Small Cell Lung Cancer Cell Progression.** *Onco Targets Ther* 2019, **12**:9927-9939.
77. Liang WH, Li N, Yuan ZQ, Qian XL, Wang ZH: **DSCAM-AS1 promotes tumor growth of breast cancer by reducing miR-204-5p and up-regulating RRM2.** *Mol Carcinog* 2019, **58**(4):461-473.
78. Kalimutho M, Sinha D, Jeffery J, Nones K, Srihari S, Fernando WC, Duijf PH, Vennin C, Raninga P, Nanayakkara D *et al*: **CEP55 is a determinant of cell fate during perturbed mitosis in breast cancer.** *EMBO Mol Med* 2018, **10**(9).
79. Yi XH, Zhang B, Fu YR, Yi ZJ: **STAT1 and its related molecules as potential biomarkers in Mycobacterium tuberculosis infection.** *J Cell Mol Med* 2020, **24**(5):2866-2878.
80. Pan H, Yin C, Dyson NJ, Harlow E, Yamasaki L, Van Dyke T: **Key roles for E2F1 in signaling p53-dependent apoptosis and in cell division within developing tumors.** *Mol Cell* 1998, **2**(3):283-292.

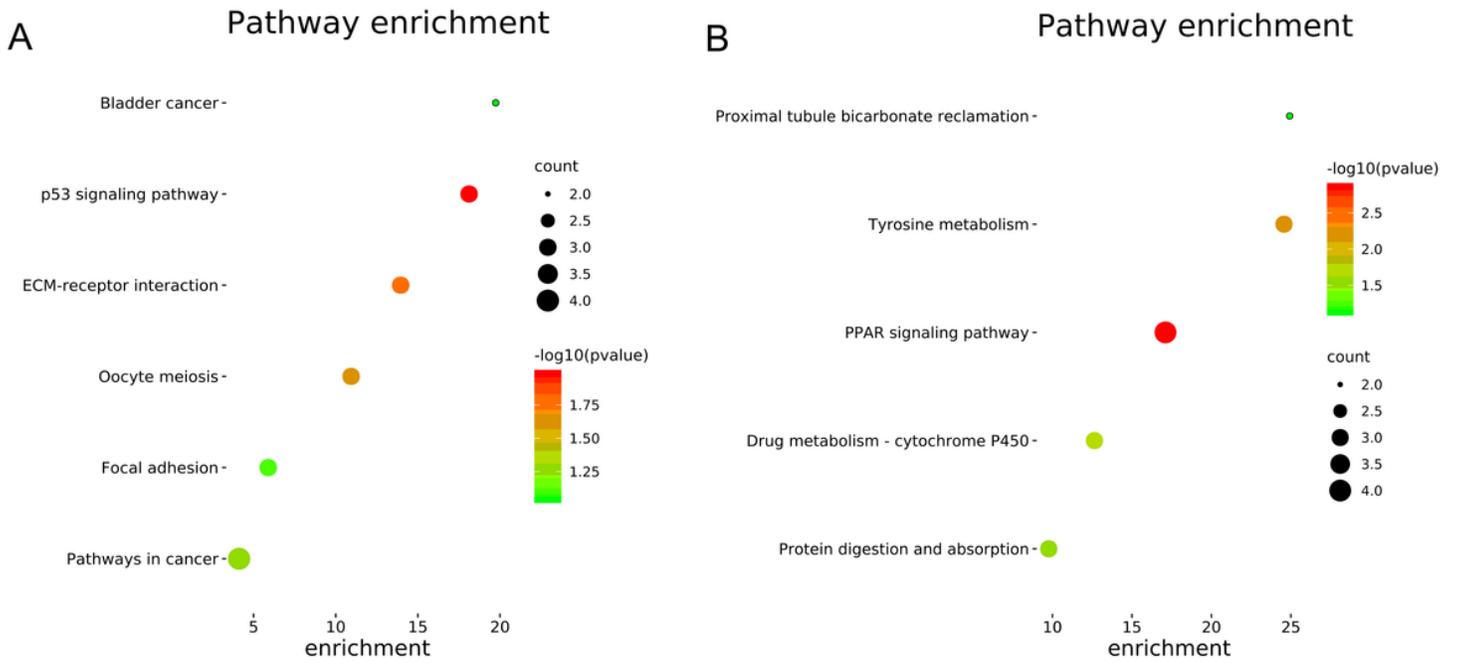
81. Ertosun MG, Hapil FZ, Osman Nidai O: **E2F1 transcription factor and its impact on growth factor and cytokine signaling.** *Cytokine Growth Factor Rev* 2016, **31**:17-25.
82. He S, Yang S, Deng G, Liu M, Zhu H, Zhang W, Yan S, Quan L, Bai J, Xu N: **Aurora kinase A induces miR-17-92 cluster through regulation of E2F1 transcription factor.** *Cell Mol Life Sci* 2010, **67**(12):2069-2076.
83. Zhang YW, Jones TL, Martin SE, Caplen NJ, Pommier Y: **Implication of checkpoint kinase-dependent up-regulation of ribonucleotide reductase R2 in DNA damage response.** *J Biol Chem* 2009, **284**(27):18085-18095.
84. Yoo J, Lee YJ: **Aspirin enhances tumor necrosis factor-related apoptosis-inducing ligand-mediated apoptosis in hormone-refractory prostate cancer cells through survivin down-regulation.** *Mol Pharmacol* 2007, **72**(6):1586-1592.

## Figures



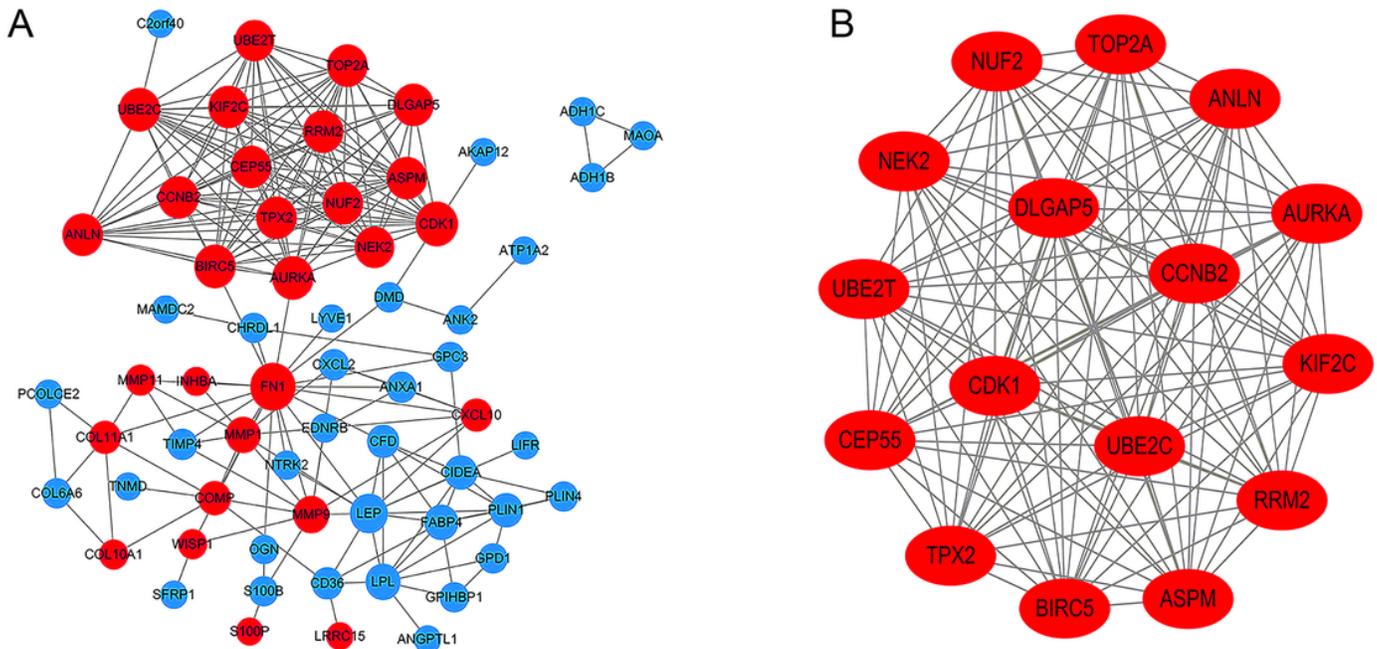
**Figure 1**

Identification of shared BRCA-associated DEGs. A total of 85 shared DEGs in the GSE45827, GSE42568, and GSE61304 datasets were identified using a Venn diagram (<http://bioinformatics.psb.ugent.be/webtools/Venn/>). Different colors correspond to different microarray datasets. A 31 shared upregulated DEGs ( $\log_{2}FC > 0$ ) and B 54 shared downregulated DEGs ( $\log_{2}FC < 0$ ) were identified via this approach.



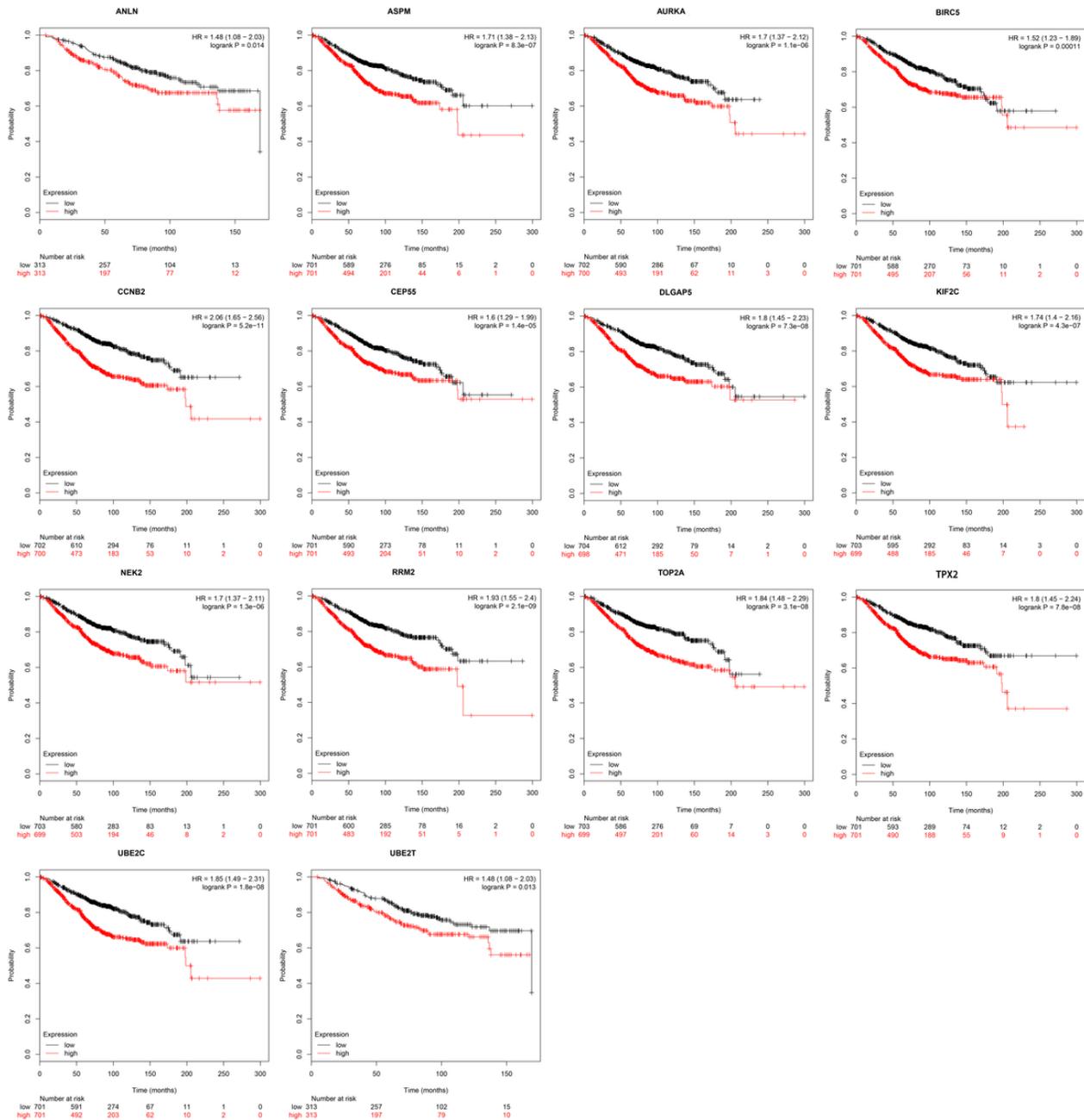
**Figure 2**

Pathway enrichment analysis of BRCA-related DEGs. A upregulated DEGs B downregulated DEGs. Bubble color represent p-value ( $-\log_{10} \text{p-value}$ ) while the size represent the number of DEGs in the corresponding pathway.



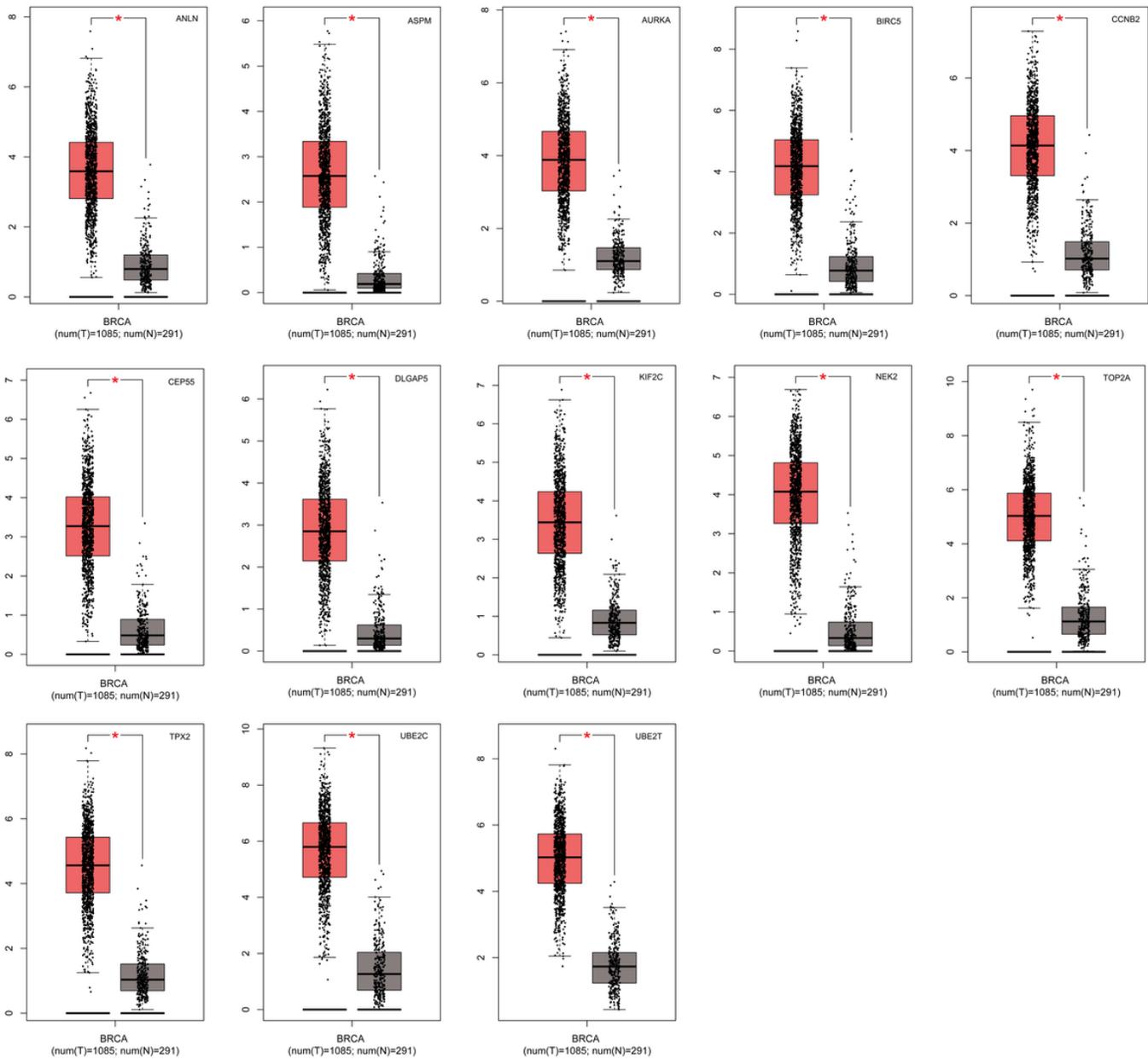
**Figure 3**

PPI network analysis of BRCA-related DEGs. A We used the STRING database and Cytoscape tools to construct a PPI network from 85 BRCA-related DEGs. B Cytoscape was used to conduct a module analysis, with up- and down-regulated genes being shown in red and blue, respectively, and with lines corresponding to gene-gene interactions.



**Figure 4**

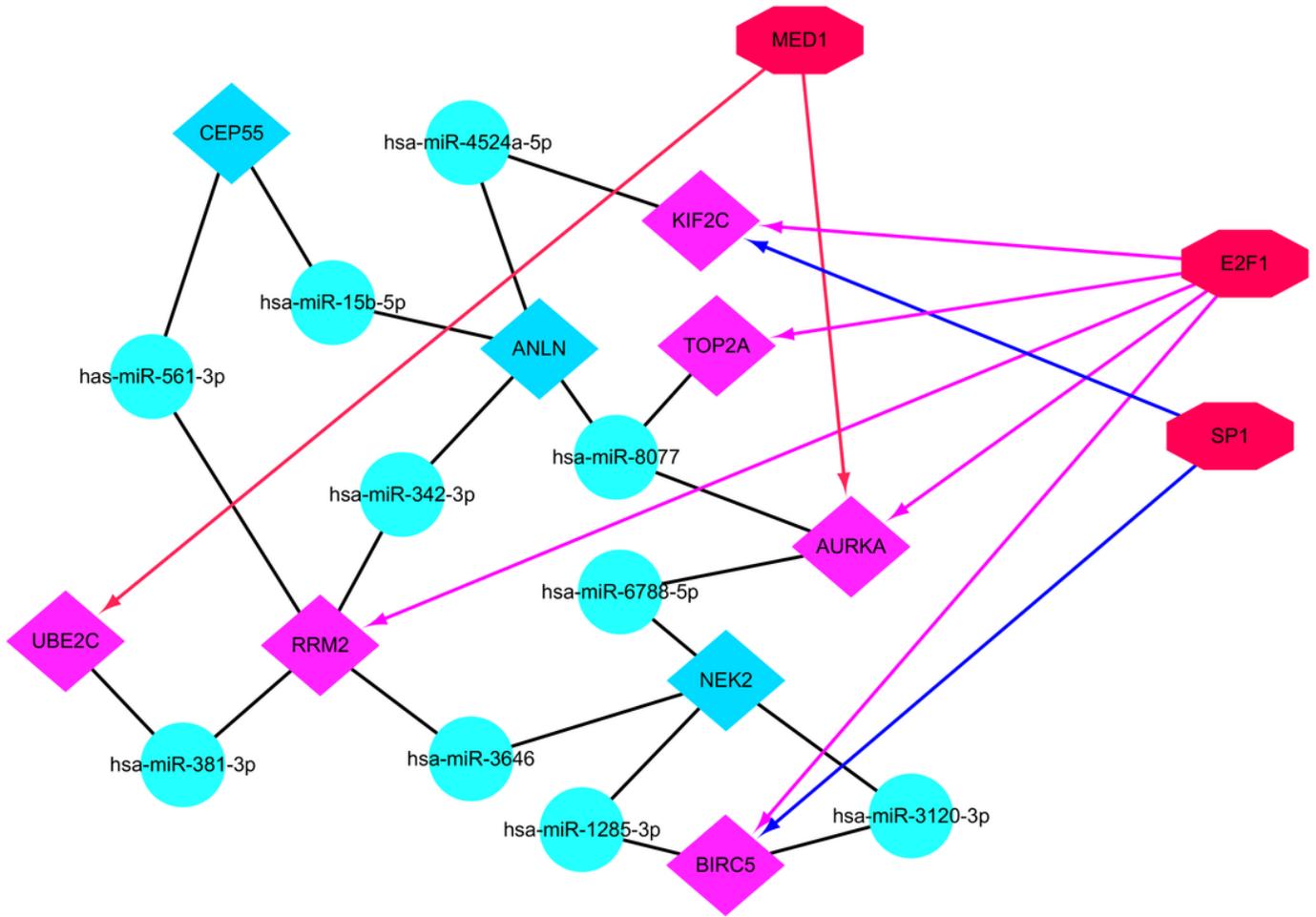
Prognostic analysis of 16 BRCA-related hub genes. The Kaplan-Meier Plotter tool was used to assess the prognostic relevance of the 16 identified core genes, with Logrank P < 0.05 as the significance threshold.



**Figure 5**

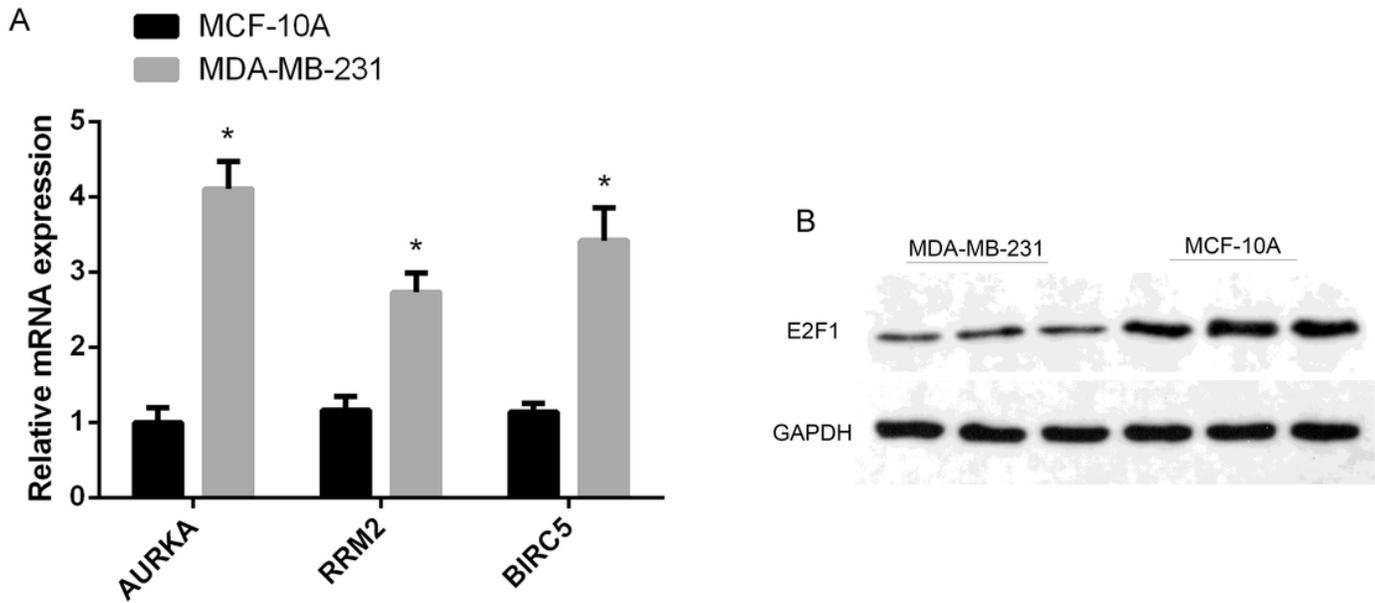
A GEPIA approach was used to confirm the differential expression of the 14 identified prognostic core genes in BRCA tumors relative to control tissues. The tumor and control tissues are indicated in red and gray, respectively, with dots corresponding to expression in individual samples. \*P < 0.01.





**Figure 7**

A mRNA-target miRNA-transcription factor interaction network. Red octagons represent transcription factors. Green circles represent miRNAs. Purple diamonds represent genes trans-regulated by miRNA and TFs. Blue diamonds represent genes trans-regulated by miRNA



**Figure 8**

The results of experimental verification. A The relative expression of AURKA, RRM2 and BIRC5 in MDA-MB-231 and MCF-10A cells by RT-PCR. B The expression of E2F1 protein in MDA-MB-231 and MCF-10A cells by Western blot. Data were calculated by paired Student's t-test. \*P < 0.01

## Supplementary Files

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

- [SupplementaryTable1.docx](#)
- [SupplementaryTable2.docx](#)
- [SupplementaryTable3.docx](#)
- [SupplementaryTable4.docx](#)