

Identification of grade-specific diagnostic and prognostic biomarkers of key candidate genes and pathways in breast cancer by integrated bioinformatics analysis

HanYu Zhang

Huashan Hospital Fudan University <https://orcid.org/0000-0001-7324-0943>

XiuLing Zhou

Huashan Hospital Fudan University

XiaoMing Zhang

Huashan Hospital Fudan University

ZhongYu Yu

Huashan Hospital Fudan University

JianTao Wang

Huashan Hospital Fudan University

XinYu Guo

Huashan Hospital Fudan University

Yue Hu

Huashan Hospital Fudan University

JingChun GUO

Fudan University School of Basic Medical Sciences

YingHong Tao

Huashan Hospital Fudan University

HouGuang Zhou (✉ zhg7376@163.com)

<https://orcid.org/0000-0002-5550-1166>

Research article

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

Posted Date: February 28th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-15432/v1>

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

Abstract

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

Methods: Samples were obtained from NCBI-GEO website. Then, GEO2R tools and Venn diagram software were used to identify the differentially expressed genes (DEGs). Next, analyze Kyoto Encyclopedia of Gene and Genome (KEGG) pathway and gene ontology (GO) were analyzed by Database for Annotation, Visualization and Integrated Discovery (DAVID). The protein-protein interaction (PPI) network was drew by Search Tool for the Retrieval of Interacting Genes (STRING). Afterwards, we selected the core genes from PPI network by Molecular Complex Detection (MCODE) plug-in. And we performed Kaplan-Meier analysis to assess the overall survival of the core genes. Last Gene Expression Profiling Interactive Analysis (GEPIA) was used to discover highly expressed genes in BC.

Results: DEGs contained 23 up-regulated and 32 down-regulated genes. GO described molecular function (MF), cellular component (CC), and biological process (BP). KEGG pathway showed DEGs were mainly involved in ECM-receptor interaction, p53 signaling pathway, PPAR signaling pathway, signaling pathways regulating pluripotency of stem cells, cGMP-PKG signaling pathway and Tyrosine metabolism. Finally, we screened 15 core genes, 14 of which had adverse prognosis and high expression in BC.

Conclusions: In the current study, 14 core genes of BC were identified based on bioinformatics method, which could useful to provide essential information for early diagnosis and treatment of BC.

Introduction

At present, breast cancer (BC) is the most common cancer diagnosis among female tumors¹. Although the treatment of BC has made great progress in recent years, it is still the main cause of female death. In 2016, the mortality of BC in the European Union was expected to decline by 8%. However, BC remains the most common cause of cancer death in less developed countries². The molecular mechanism of tumorigenesis in BC remains unclear. Identifying new genes related to tumorigenesis and prognosis of patients and elucidating the molecular mechanisms of these carcinogenic processes³ are essential for early diagnosis, prevention. Meanwhile, genomic analysis of BC will result in new therapeutic roadmap⁴ that contribute to personalized therapy.

Over the past ten years, 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^{5, 6}. 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 reduce cancer morbidity and mortality⁷.

In this study, three datasets were acquired from Gene Expression Omnibus (GEO) database. Then, we got the commonly differentially expressed genes (DEGs) by comparison between the gene expression data of BC

patients and non-BC patients in the three datasets. Afterwards, these DEGs were underwent Gene Ontology (GO) and Kyoto Encyclopedia of Gene and Genomes (KEGG) pathway analysis. In addition, we constructed protein-protein interaction (PPI) network and applied Cytotype MCODE (Molecular Complex Detection) for analysis of the DEGs to identify some core genes. And the Kaplan-Meier plotter online database were utilized for survival analysis of these core genes. Subsequently, we validated the DEGs expression of BC tissues through Gene Expression Profiling Interactive Analysis (GEPIA). Through the above bioinformatics research, this study provided some effective biomarkers and signal pathways for BC.

Materials And Methods

Microarray data acquisition

Three gene expression profiles datasets (GSE61304, GSE29044 and GSE42568) were obtained from the GEO (<http://www.ncbi.nlm.nih.gov/geo/>) database which is a free public repository for storing high-throughput microarray and sequence functional genomic datasets⁸. All three microarray datasets were based on Platforms of GPL570 ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array). GSE61304 included 4 normal breast tissues and 58 BC tissues. GSE42568 contained 17 normal breast tissues and 104 BC tissues. GSE29044 consisted of 36 normal breast tissues and 73 BC tissues.

Data processing of DEGs

We identified DEGs between BC and normal breast samples via GEO2R online tools according to the criteria $|\log_{2}FC| > 2$ and $P < 0.05$. Then, Venn diagram was plotted by online software to ascertain common DEGs among the three datasets. The DEGs with $\log_{2}FC > 2$ was regarded as up-regulated genes and those with $\log_{2}FC < -2$ was down-regulated genes.

GO and KEGG pathway enrichment analysis of DEGs

GO is a comprehensive analytical approach which can be used to calculate the functions of genes and gene products. It describes three aspects: molecular function (MF), cellular component (CC), and biological process (BP)⁹. KEGG is a collection of databases that systematically analyze gene functions and link genome information with high-level functional information¹⁰. Here we mapped the GO enrichment analysis and KEGG pathway of significant DEGs via Database for Annotation, Visualization and Integrated Discovery (DAVID, [https://david.ncifcrf.gov/version 6.8](https://david.ncifcrf.gov/version%206.8)) database. P-value < 0.05 was considered statistically significant.

PPI network establishment and modular selection

Search Tool for the Retrieval of Interacting Genes (STRING, <https://string-db.org/>) online biological database provided us with critical evaluation and collection of PPI information¹¹. Then, we utilized the STRING app in Cytoscape software to build the protein interaction network. The maximum number of interactors = 0 and confidence score ≥ 0.4 was as the threshold value. Furthermore, we analyzed modules of PPI network via MCODE plugin of Cytoscape with the included criteria as follow: degree cutoff = 2, max. Depth = 100, k-core = 2, and node score cutoff = 0.2.

Survival analysis and RNA sequencing data

We evaluated the influence of key candidate genes on survival and prognosis with 95% confidence intervals and log rank P value were computed by Kaplan-Meier plotter (<http://kmplot.com/analysis>). Afterwards, GEPIA (<http://gepia.cancer-pku.cn/>), an interactive web application for gene expression analysis¹², was applied to analyzing RNA sequencing data for the sake of validated these DEGs.

Results

Identification of DEGs in BC

In the current study, 235 BC samples and 57 non-BC samples were obtained from three microarray databases (GSE29044, GSE42568 and GSE61304). Then, we identified 504,1770 and 749 DEGs with the criterion as $|\log FC| > 2$ and $P < 0.05$ via GEO2R online tools. Subsequently, 23 up-regulated DEGs ($\log FC > 2$) and 32 down-regulated DEGs ($\log FC < -2$) intersected in three datasets were detected via VENN diagram software (Table 1 & Fig. 1).

Table 1

23 up-regulated and 32 down-regulated differentially expressed genes (DEGs) in breast cancer were accessed from three profile databases.

DEGs	Genes Name
Up-regulated	TPX2 ASPM INHBA CDCA3 ANLN UBE2C CCNB2 CDK1 TOP2A GJB2 COMP DLGAP5 HMMR KIF2C SQLE COL10A1 LRRC15 MELK COL11A1 KIAA0101 NEK2 UBE2T MMP11
Down-regulated	CHRDL1 BTNL9 ATP1A2 SDPR LIFR MAOA LEPROT///LEPR MEOX2 CD36 FGF2 GPD1 MT1M FHL1 TFPI ABCA9 CSN1S1 IL33 SEMA3G ADH1B LPL ZBTB16 CCL15-CCL14///CCL14 FABP4 EDNRB ACACB ACVR1C CD300LG ADRB2 C2orf40 PDK4 COL6A6 TMEM100

GO analysis of DEGs

GO enrichment analysis were demonstrated, in order to reflect additional biological information of 55 genes, via DAVID software. The results illustrated that for BP, up-regulated DEGs were primary enriched in mitotic nuclear division, G2/M transition of mitotic cell cycle, cell division, cell proliferation, collagen catabolic process and hematopoietic progenitor cell differentiation; down-regulated DEGs were primary enriched in positive regulation of inflammatory response, cell surface receptor signaling pathway, positive regulation of cholesterol storage, cGMP-mediated signaling, cellular response to lipopolysaccharide and positive regulation of chemokine secretion. For CC, up-regulated DEGs were particularly enriched in proteinaceous extracellular matrix, midbody, microtubule cytoskeleton, nucleoplasm, microtubule and nucleus; down-regulated DEGs were particularly enriched in extracellular space, extracellular region, membrane raft, plasma membrane, apical plasma membrane and caveola. For MF, up-regulated DEGs were mainly enriched in ATP binding, chromatin binding, ubiquitin conjugating enzyme activity, cyclin-dependent protein serine/threonine kinase activity, collagen binding and protein kinase activity; down-regulated were mainly enriched in transporter activity, growth factor binding and RNA polymerase II distal enhancer sequence-specific DNA binding (Table 2).

Table 2

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

Expression	Category	Term	Count	%	P-Value	FDR
Up-regulated	GOTERM_BP_DIRECT	GO:0007067 ~ mitotic nuclear division	8	0.197433	1.99E-08	2.58E-05
	GOTERM_BP_DIRECT	GO:0000086 ~ G2/M transition of mitotic cell cycle	6	0.148075	7.91E-07	0.001027
	GOTERM_BP_DIRECT	GO:0051301 ~ cell division	7	0.172754	4.42E-06	0.005739
	GOTERM_BP_DIRECT	GO:0008283 ~ cell proliferation	5	0.123396	0.00119	1.533912
	GOTERM_BP_DIRECT	GO:0030574 ~ collagen catabolic process	3	0.074038	0.003145	4.00682
	GOTERM_BP_DIRECT	GO:0002244 ~ hematopoietic progenitor cell differentiation	3	0.074038	0.003341	4.251492
	GOTERM_CC_DIRECT	GO:0005578 ~ proteinaceous extracellular matrix	4	0.098717	0.003938	3.929136
	GOTERM_CC_DIRECT	GO:0030496 ~ midbody	3	0.074038	0.010469	10.13987
	GOTERM_CC_DIRECT	GO:0015630 ~ microtubule cytoskeleton	3	0.074038	0.011745	11.30978
	GOTERM_CC_DIRECT	GO:0005654 ~ nucleoplasm	8	0.197433	0.040131	34.03841
	GOTERM_CC_DIRECT	GO:0005874 ~ microtubule	3	0.074038	0.053585	42.85074
	GOTERM_CC_DIRECT	GO:0005634 ~ nucleus	11	0.271471	0.086514	60.11868
	GOTERM_MF_DIRECT	GO:0005524 ~ ATP binding	8	0.197433	0.002183	2.242714
	GOTERM_MF_DIRECT	GO:0003682 ~ chromatin binding	4	0.098717	0.013696	13.33602
	GOTERM_MF_DIRECT	GO:0061631 ~ ubiquitin conjugating enzyme activity	2	0.049358	0.037143	32.48714

Expression	Category	Term	Count	%	P-Value	FDR
	GOTERM_MF_DIRECT	GO:0004693 ~ cyclin-dependent protein serine/threonine kinase activity	2	0.049358	0.043412	36.91288
	GOTERM_MF_DIRECT	GO:0005518 ~ collagen binding	2	0.049358	0.075389	55.67167
	GOTERM_MF_DIRECT	GO:0004672 ~ protein kinase activity	3	0.074038	0.078808	57.34339
Down-regulated	GOTERM_BP_DIRECT	GO:0050729 ~ positive regulation of inflammatory response	3	0.077539	0.006099	8.161591
	GOTERM_BP_DIRECT	GO:0007166 ~ cell surface receptor signaling pathway	4	0.103386	0.00941	12.32909
	GOTERM_BP_DIRECT	GO:0010886 ~ positive regulation of cholesterol storage	2	0.051693	0.011203	14.51243
	GOTERM_BP_DIRECT	GO:0019934 ~ cGMP-mediated signaling	2	0.051693	0.012794	16.4065
	GOTERM_BP_DIRECT	GO:0071222 ~ cellular response to lipopolysaccharide	3	0.077539	0.014118	17.95323
	GOTERM_BP_DIRECT	GO:0090197 ~ positive regulation of chemokine secretion	2	0.051693	0.014382	18.25872
	GOTERM_CC_DIRECT	GO:0005615 ~ extracellular space	8	0.206772	0.004368	4.307352
	GOTERM_CC_DIRECT	GO:0005576 ~ extracellular region	7	0.180925	0.03827	32.45913
	GOTERM_CC_DIRECT	GO:0045121 ~ membrane raft	3	0.077539	0.042275	35.23489
	GOTERM_CC_DIRECT	GO:0005886 ~ plasma membrane	12	0.310158	0.045492	37.38963
	GOTERM_CC_DIRECT	GO:0016324 ~ apical plasma membrane	3	0.077539	0.077815	55.72269

Expression	Category	Term	Count	%	P-Value	FDR
	GOTERM_CC_DIRECT	GO:0005901 ~ caveola	2	0.051693	0.098504	64.75675
	GOTERM_MF_DIRECT	GO:0005215 ~ transporter activity	3	0.077539	0.035688	32.87746
	GOTERM_MF_DIRECT	GO:0019838 ~ growth factor binding	2	0.051693	0.039255	35.55148
	GOTERM_MF_DIRECT	GO:0000980 ~ RNA polymerase II distal enhancer sequence-specific DNA binding	2	0.051693	0.092005	65.31201

KEGG pathway analysis

Results of KEGG pathway analysis by DAVID gene annotation tool were verified that the up-regulated DEGs were significant involved in ECM-receptor interaction and p53 signaling pathway; down-regulated DEGs were significant involved in PPAR signaling pathway, signaling pathways regulating pluripotency of stem cells, cGMP-PKG signaling pathway and Tyrosine metabolism (Table 3).

Table 3

Kyoto Encyclopedia of Gene and Genome (KEGG) pathways analysis of up-regulated and down-regulated genes in breast cancer.

Expression	Pathway ID	Name	Count	%	P-Value	FDR	Genes
Up-regulated	hsa04512	ECM-receptor interaction	3	0.074038	0.005373	4.230023	COMP, COL11A1, HMMR
	hsa04115	p53 signaling pathway	2	0.049358	0.084367	50.6899	CDK1, CCNB2
Down-regulated	hsa03320	PPAR signaling pathway	3	0.077539	0.014362	14.15501	LPL, CD36, FABP4
	hsa04550	Signaling pathways regulating pluripotency of stem cells	3	0.077539	0.056106	45.62159	LIFR, FGF2, ACVR1C
	hsa04022	cGMP-PKG signaling pathway	3	0.077539	0.06946	53.21163	EDNRB, ADRB2, ATP1A2
	hsa00350	Tyrosine metabolism	2	0.051693	0.092485	64.08027	MAOA, ADH1B

Construction of PPI network and analysis of module analysis

The PPI network of DEGs which contained 38 nodes and 130 edges was constructed by STRING online database and Cytoscape software. A total of 38 DEGs were identified, of which 21 were up regulated genes and 17 were down-regulated genes (Fig. 2a). Afterwards, we utilized Cytotype MCODE to identify significant modules. The results indicated that 15 core DEGs, which all belong up-regulated genes were recognized among the 38 DEGs (Fig. 2b). These core DEGs include KIAA0101, anillin actin binding protein (ANLN), cyclin dependent kinase 1 (CDK1), ubiquitin conjugating enzyme E2 C (UBE2C), cyclin B2 (CCNB2), kinesin family member 2C (KIF2C), DNA topoisomerase II alpha (TOP2A), TPX2 microtubule nucleation factor (TPX2), hyaluronan mediated motility receptor (HMMR), DLG associated protein 5 (DLGAP5), abnormal spindle microtubule assembly (ASPM), NIMA related kinase 2 (NEK2), maternal embryonic leucine zipper kinase (MELK), ubiquitin conjugating enzyme E2 T (UBE2T) and cell division cycle associated 3 (CDCA3).

Analysis of core genes by the Kaplan Meier plotter and GEPIA

To estimate the prognostic value of DEGs, we used Kaplan-Meier plotter to assess the correlation between the expression of 15 core genes and survival rates of BC patients. It was showed that 14 genes had a significantly worse survival ($P < 0.05$, Fig. 3). Subsequently, we used GEPIA to re-analyze the expression level of 14 genes between cancerous and normal tissues. The results reported that compared with normal BC patients, all 14 genes were highly expressed in BC samples (Table 4 & Fig. 4).

Table 4
Prognostic information of core genes via GEPIA.

Category	Genes
Genes with high expressed in BC ($P < 0.05$)	ANLN ASPM CCNB2 CDCA3 DLGAP5 HMMR KIAA0101 KIF2C MELK NEK2 TOP2A TPX2 UBE2C UBE2T

Discussion

Prognostic biomarkers in BC were explored and validated comprehensively via bioinformatical methods in the current research. First, a total of 235 BC samples and 57 non-BC samples were retrieved from GSE39044, GSE42568 and GSE61304 datasets. 23 up-regulated DEGs and 32 down-regulated DEGs were screened via GEO2R and Venn online tools. Afterwards, GO enrichment analysis via DAVID software concluded that: 1) for BP, up-regulated DEGs were mainly enriched in mitotic nuclear division, G2/M transition of mitotic cell cycle, cell division, cell proliferation, collagen catabolic process and hematopoietic progenitor cell differentiation; down-regulated DEGs were mainly enriched in positive regulation of inflammatory response, cell surface receptor signaling pathway, positive regulation of cholesterol storage, cGMP-mediated signaling, cellular response to lipopolysaccharide and positive regulation of chemokine secretion. 2) for CC, up-regulated DEGs were enriched in proteinaceous extracellular matrix, midbody, microtubule cytoskeleton, nucleoplasm, microtubule and nucleus; down-regulated DEGs were enriched in extracellular space, extracellular region, membrane raft, plasma membrane, apical plasma membrane and caveola. 3) for MF, up-regulated DEGs were enriched in ATP binding, chromatin binding, ubiquitin conjugating enzyme activity, cyclin-dependent protein serine/threonine

kinase activity, collagen binding and protein kinase activity; down-regulated were enriched in transporter activity, growth factor binding and RNA polymerase II distal enhancer sequence-specific DNA binding. KEGG pathway analysis concluded that the up-regulated DEGs were involved in ECM-receptor interaction and p53 signaling pathway; down-regulated DEGs were involved in PPAR signaling pathway, signaling pathways regulating pluripotency of stem cells, cGMP-PKG signaling pathway and Tyrosine metabolism. Then, we plotted the DEGs PPI network with the STRING online database in Cytoscape software include 38 nodes and 130 edges. In addition, via Cytotype MCODE, we detected 15 core DEGs from PPI network, which were up-regulated genes all. Moreover, 14 of 15 DEGs that had a significantly worse survival were identified though Kaplan Meier plotter analysis, contain KIAA0101, ANLN, ASPM, CCNB2, CDCA3, DLGAP5, HMMR, KIF2C, MELK, NEK2, TOP2A, TPX2, UBE2C and UBE2T. Finally, we validated 14 DEGs by GEPIA analysis, the results indicated that the expression of 14 genes in BC samples was higher than that in normal samples.

ANLN is an evolutionarily actin-binding protein, which is firstly found in *Drosophila* that play a pivotal role in cytokinesis¹³. Research indicated that knockdown of ANLN could cause G2/M phase cell cycle arrest and thus inhibit cell growth¹⁴. In BC patients, high expression of nuclear ANLN in cancer cell is associated with poor prognosis¹⁵. Meanwhile, ANLN participated in the cellular growth and migration in BC¹⁶. It has been also testified that patients with lower level of ANLN expression might have both better prognosis and response to anthracycline-based chemotherapy¹⁷.

KIAA0101 is a 15-kDa protein containing a conserved proliferating cell nuclear antigen (PCNA)-binding motif, which was identified by a yeast two hybrid screen for proteins binding to PCNA as 15PAF (PCNA-associated factor)^{18,19}. It was verified that KIAA0101 has been associated with various tumor. In hepatocellular carcinoma and gastric cancer, KIAA001 overexpression is interrelated with cancer recurrence^{19,20}. Meanwhile, In esophageal cancer, KIAA0101 expression reduced the sensitivity of cancer cells to cisplatin chemotherapy²¹. In addition, reports on BC indicated that KIAA0101 gene was positively correlated with the CCNE2, CDK6, and CDKN1A expression level and thus regulated the BC cells proliferation and delayed cell cycle G1/S progression²².

ASPM is a microtubules-associated protein that is vital for control spindle microtubules organization, spindle function, and cytokinesis from insects to mammals^{23,24}. Increasing evidence indicated that ASPM was upregulated in a variety of tumors, such as ovarian cancer, prostate cancer, glioma, and hepatocellular carcinoma²⁵. Furthermore, ASPM was reported to conduce to predicting the effect of endocrine therapy in BC²⁶, but the mechanism is still unclear and demands further exploration.

CCNB2 is a member of the cyclin family, specifically the B-type cyclins. It regulates the activities of cyclin-dependent kinases (CDKs) and spatial functions of different cyclins and temporal in specific phases of the cell cycle. CCNB2 generally triggers the progression of G2/M transition through activating CDK1 kinase²⁷. Previous studies show that CCNB2 was demonstrated to serve as an independent biomarker for invasive BC, and elevated CCNB2 was negatively correlated with the survival rate of patients²⁸.

HMMR is a hyper-expressed oncogene, the overexpression of which is transforming and is also essential for H-ras transformation²⁹. HMMR is a hyaluronan receptor, and hyaluronan is a major constituent of the

microenvironment of numerous malignant tumors³⁰. In BC, recent evidences suggest that high levels of HMMR is a helpful prognostic indicator for breast carcinoma progression³¹. In addition, HMMR locus is related to high risk of BC in human³².

CDCA3 functions as part of a SKP1-Culin-RING-F-box ubiquitin ligase, which is reported to modulate cell cycle progression³³. Previous data analysis indicated that CDCA3 showed on average > 100 – fold overexpression in BC compared with normal tissue³⁴. Meantime, CDCA3 has been demonstrated to be expressed in liver cancer, oral squamous cell carcinoma tissue, non-small cell lung cancer and prostate cancer cells³⁵. However, the role of CDCA3 in malignant tumors remains unclear.

DLGAP5 is an significant component of the mitotic apparatus³⁶. Actually, DLGAP5 was regarded as an oncogenic target of Aurora A, a main mitotic serine/threonine kinase, and DLGAP5 phosphorylated by Aurora A is precondition in regulating spindle assembly and function during mitosis³⁷. Furthermore, the current research provides new evidence that DLGAP5 is required for the G2/M transition in tumorigenesis³⁸. And in BC, DLGAP expression can inhibit cell proliferation and enhance susceptibility to epirubicin³⁹.

KIF2C is a member of the kinesin-13 subfamily of kinesin-related proteins that functions as a microtubule-dependent molecular motor⁴⁰. Its activity is regulated by aurora kinase-B, and it is important to chromosome segregation and the correction of improper kinetochore-microtubule interactions during mitosis⁴¹. Previous detailed expression-profile analysis of clinical BC indicated that downregulation of KIF2C remarkable suppresses the growth of BC cells, revealing its crucial role in the growth of BC cells⁴².

MELK belongs to the adenosine monophosphate- activated protein kinase (AMPK)-related kinase family which is a kind of serine/threonine kinase⁴³. MELK has been deemed as a cancer dependency and putative drug target in multiple tumor types, such as melanoma, colorectal cancer, and triple-negative BC. In these cancers, high expression of MELK is correlated with poor prognosis⁴⁴. Due to the possible oncogenic role of MELK, the potential of MELK inhibitors as drug targets has begun to be concerned by relevant research⁴⁵.

NEK2 is a cell cycle-regulated protein kinase which located in the centrosome, and it participates in the regulation of centrosome separation and spindle formation⁴⁶. Several researches indicated that high expression NEK2 is overexpressed in variety forms of tumor, where it function as an oncogene⁴⁷. Furthermore, cancer cells with down-regulated NEK2 represent aberrant mitosis leading to apoptosis. Inhibiting NEK2 is beneficial to disrupt the cell cycle of triple negative BC cells, enhancing the effects of chemotherapeutic interventions⁴⁸.

TOP2A is a DNA encodes topoisomerase II α , an enzyme that controls and alters the topologic states of DNA during transcription and a direct target of anthracyclines⁴⁹. In view of the role of TOP2A in the cell duplication process, it was thought that TOP2A protein was up-regulated in proliferating cancer cells⁵⁰. TOP2A is closely related to HER2. However, several studies demonstrated that TOP2A amplification and deletion is correlated with, but not limited to, BC with HER2 amplification⁵¹.

TPX2 is microtubule-associated protein which involved in the complex process of mitosis, and has been identified as a prohibition strictly regulated by cell cycle⁵². Due to the functional properties, TPX2 is considered as a good histological marker of active proliferative in tumors⁵³. Research in recent years has shown that TPX2 is closely connected with the development of BC stem cells. And the radiotherapy effect of lovastatin on BC stem cells may be related to the expression of TPX2⁵⁴.

UBE2C is a crucial component of the ubiquitin-conjugating enzyme complex, which is involved in the ubiquitin-proteasome system⁵⁵. Studies proved that overexpressed UBE2C was linked to poor prognosis in BC⁵⁶. Simultaneously, UBE2C biomarker expression in mammary molybdenum target biopsy specimens is benefit to evaluate the diagnosis of breast lesions or early breast carcinoma⁵⁷. Besides that, the increased expression of UBE2C can reduce the sensitivity of cancer cells to doxorubicin treatment⁵⁸.

UBE2T is the E2 enzyme of the Fanconi anemia pathway, which is indispensable to the repair of DNA interchain cross-links⁵⁹. Prior research support the argument that UBE2T is associated with adverse outcomes in breast and lung cancers⁶⁰. There are also several evidences imply that overexpression of UBE2T may contribute to breast carcinogenesis⁶¹. However, UBE2T as a therapeutic target for BC still needs further explore.

Previous literature has reported that these genes are correlated to the development and progression of tumors. They also perform certain function in BC. ANLN, KIAA0101, CCNB2, HMMR, DLGAP5, KIF2C, MELK, NEK2, TOP2A, UBE2C has been proven to be associated with BC in many researches. However, numerous mechanisms remain unclear yet. DEGs include ASPM, CDCA3, TPX2, UBE2T need further study. Hence, the data in our study could provide helpful information for further research.

Conclusions

Taken together, we selected 15 core genes from PPI network by Molecular Complex Detection (MCODE) plug-in and performed Kaplan–Meier analysis, and results showed that 14 of the 15 genes had adverse prognosis. Furthermore, 14 genes were validated highly expressed in BC tissues via Gene Expression Profiling Interactive Analysis (GEPIA), which could useful to provide information for early diagnosis and treatment of BC. Results indicated that these genes play roles in the development of BC. Anyway, these genes would help in the early diagnosis of BC, and would be promising targets for research of novel anticancer therapies.

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 Rugo HS. Achieving Improved Survival Outcomes in Advanced Breast Cancer. *The New England journal of medicine*. 2019; **381**: 371-72.
- 2 Harbeck N, Gnant M. Breast cancer. *Lancet (London, England)*. 2017; **389**: 1134-50.
- 3 Xu S, Kong D, Chen Q, Ping Y, Pang D. Oncogenic long noncoding RNA landscape in breast cancer. *Molecular cancer*. 2017; **16**: 129.
- 4 Ellis M. Genomic analysis of breast cancer heralds a changing treatment paradigm. *Journal of the National Comprehensive Cancer Network : JNCCN*. 2014; **12**: 750-2.
- 5 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**: 35.
- 6 Yan H, Zheng G, Qu J, *et al*. Identification of key candidate genes and pathways in multiple myeloma by integrated bioinformatics analysis. *Journal of cellular physiology*. 2019.
- 7 Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. *Science (New York, NY)*. 2013; **339**: 1546-58.
- 8 Barrett T, Wilhite SE, Ledoux P, *et al*. NCBI GEO: archive for functional genomics data sets—update. *Nucleic acids research*. 2013; **41**: D991-5.

- 9 Expansion of the Gene Ontology knowledgebase and resources. *Nucleic acids research*. 2017; **45**: D331-d38.
- 10 Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*. 2000; **28**: 27-30.
- 11 Szklarczyk D, Franceschini A, Wyder S, *et al*. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic acids research*. 2015; **43**: D447-52.
- 12 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 research*. 2017; **45**: W98-w102.
- 13 Zhang S, Nguyen LH, Zhou K, *et al*. Knockdown of Anillin Actin Binding Protein Blocks Cytokinesis in Hepatocytes and Reduces Liver Tumor Development in Mice Without Affecting Regeneration. *Gastroenterology*. 2018; **154**: 1421-34.
- 14 Zhou W, Wang Z, Shen N, *et al*. Knockdown of ANLN by lentivirus inhibits cell growth and migration in human breast cancer. *Molecular and cellular biochemistry*. 2015; **398**: 11-9.
- 15 Lian YF, Huang YL, Wang JL, *et al*. Anillin is required for tumor growth and regulated by miR-15a/miR-16-1 in HBV-related hepatocellular carcinoma. *Aging*. 2018; **10**: 1884-901.
- 16 Suzuki C, Daigo Y, Ishikawa N, *et al*. ANLN plays a critical role in human lung carcinogenesis through the activation of RHOA and by involvement in the phosphoinositide 3-kinase/AKT pathway. *Cancer research*. 2005; **65**: 11314-25.
- 17 Wang Z, Chen J, Zhong MZ, *et al*. Overexpression of ANLN contributed to poor prognosis of anthracycline-based chemotherapy in breast cancer patients. *Cancer chemotherapy and pharmacology*. 2017; **79**: 535-43.
- 18 Hosokawa M, Takehara A, Matsuda K, *et al*. Oncogenic role of KIAA0101 interacting with proliferating cell nuclear antigen in pancreatic cancer. *Cancer research*. 2007; **67**: 2568-76.
- 19 Kato T, Daigo Y, Aragaki M, Ishikawa K, Sato M, Kaji M. Overexpression of KIAA0101 predicts poor prognosis in primary lung cancer patients. *Lung cancer (Amsterdam, Netherlands)*. 2012; **75**: 110-8.
- 20 Zhu K, Diao D, Dang C, *et al*. Elevated KIAA0101 expression is a marker of recurrence in human gastric cancer. *Cancer science*. 2013; **104**: 353-9.
- 21 Cheng Y, Li K, Diao D, *et al*. Expression of KIAA0101 protein is associated with poor survival of esophageal cancer patients and resistance to cisplatin treatment in vitro. *Laboratory investigation; a journal of technical methods and pathology*. 2013; **93**: 1276-87.
- 22 Lv W, Su B, Li Y, Geng C, Chen N. KIAA0101 inhibition suppresses cell proliferation and cell cycle progression by promoting the interaction between p53 and Sp1 in breast cancer. *Biochemical and biophysical research communications*. 2018; **503**: 600-06.

- 23 Gai M, Bianchi FT, Vagnoni C, *et al.* ASPM and CITK regulate spindle orientation by affecting the dynamics of astral microtubules. *EMBO reports*. 2016; **17**: 1396-409.
- 24 Saunders RD, Avides MC, Howard T, Gonzalez C, Glover DM. The *Drosophila* gene abnormal spindle encodes a novel microtubule-associated protein that associates with the polar regions of the mitotic spindle. *The Journal of cell biology*. 1997; **137**: 881-90.
- 25 Tang J, Lu M, Cui Q, *et al.* Overexpression of ASPM, CDC20, and TTK Confer a Poorer Prognosis in Breast Cancer Identified by Gene Co-expression Network Analysis. *Frontiers in oncology*. 2019; **9**: 310.
- 26 Turnbull AK, Arthur LM, Renshaw L, *et al.* Accurate Prediction and Validation of Response to Endocrine Therapy in Breast Cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2015; **33**: 2270-8.
- 27 Qian X, Song X, He Y, *et al.* CCNB2 overexpression is a poor prognostic biomarker in Chinese NSCLC patients. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*. 2015; **74**: 222-7.
- 28 Shubbar E, Kovács A, Hajizadeh S, *et al.* Elevated cyclin B2 expression in invasive breast carcinoma is associated with unfavorable clinical outcome. *BMC cancer*. 2013; **13**: 1.
- 29 Hall CL, Yang B, Yang X, *et al.* Overexpression of the hyaluronan receptor RHAMM is transforming and is also required for H-ras transformation. *Cell*. 1995; **82**: 19-26.
- 30 Tilghman J, Wu H, Sang Y, *et al.* HMMR maintains the stemness and tumorigenicity of glioblastoma stem-like cells. *Cancer research*. 2014; **74**: 3168-79.
- 31 Wang C, Thor AD, Moore DH, *et al.* The overexpression of RHAMM, a hyaluronan-binding protein that regulates ras signaling, correlates with overexpression of mitogen-activated protein kinase and is a significant parameter in breast cancer progression. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 1998; **4**: 567-76.
- 32 Pujana MA, Han JD, Starita LM, *et al.* Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nature genetics*. 2007; **39**: 1338-49.
- 33 Ayad NG, Rankin S, Murakami M, Jebanathirajah J, Gygi S, Kirschner MW. Tome-1, a trigger of mitotic entry, is degraded during G1 via the APC. *Cell*. 2003; **113**: 101-13.
- 34 Itzel T, Scholz P, Maass T, *et al.* Translating bioinformatics in oncology: guilt-by-profiling analysis and identification of KIF18B and CDCA3 as novel driver genes in carcinogenesis. *Bioinformatics (Oxford, England)*. 2015; **31**: 216-24.
- 35 Qian W, Zhang Z, Peng W, *et al.* CDCA3 mediates p21-dependent proliferation by regulating E2F1 expression in colorectal cancer. *International journal of oncology*. 2018; **53**: 2021-33.
- 36 Koffa MD, Casanova CM, Santarella R, Köcher T, Wilm M, Mattaj IW. HURP is part of a Ran-dependent complex involved in spindle formation. *Current biology : CB*. 2006; **16**: 743-54.

- 37 Wong J, Lerrigo R, Jang CY, Fang G. Aurora A regulates the activity of HURP by controlling the accessibility of its microtubule-binding domain. *Molecular biology of the cell*. 2008; **19**: 2083-91.
- 38 Chen X, Thiaville MM, Chen L, *et al*. Defining NOTCH3 target genes in ovarian cancer. *Cancer research*. 2012; **72**: 2294-303.
- 39 Zhang X, Pan Y, Fu H, Zhang J. Nucleolar and Spindle Associated Protein 1 (NUSAP1) Inhibits Cell Proliferation and Enhances Susceptibility to Epirubicin In Invasive Breast Cancer Cells by Regulating Cyclin D Kinase (CDK1) and DLGAP5 Expression. *Medical science monitor : international medical journal of experimental and clinical research*. 2018; **24**: 8553-64.
- 40 Wordeman L, Mitchison TJ. Identification and partial characterization of mitotic centromere-associated kinesin, a kinesin-related protein that associates with centromeres during mitosis. *The Journal of cell biology*. 1995; **128**: 95-104.
- 41 Andrews PD, Ovechkina Y, Morrice N, *et al*. Aurora B regulates MCAK at the mitotic centromere. *Developmental cell*. 2004; **6**: 253-68.
- 42 Shimo A, Tanikawa C, Nishidate T, *et al*. Involvement of kinesin family member 2C/mitotic centromere-associated kinesin overexpression in mammary carcinogenesis. *Cancer science*. 2008; **99**: 62-70.
- 43 Heyer BS, Warsowe J, Solter D, Knowles BB, Ackerman SL. New member of the Snf1/AMPK kinase family, Melk, is expressed in the mouse egg and preimplantation embryo. *Molecular reproduction and development*. 1997; **47**: 148-56.
- 44 Giuliano CJ, Lin A, Smith JC, Palladino AC, Sheltzer JM. MELK expression correlates with tumor mitotic activity but is not required for cancer growth. *eLife*. 2018; **7**.
- 45 Huang HT, Seo HS, Zhang T, *et al*. MELK is not necessary for the proliferation of basal-like breast cancer cells. *eLife*. 2017; **6**.
- 46 Hayward DG, Clarke RB, Faragher AJ, Pillai MR, Hagan IM, Fry AM. The centrosomal kinase Nek2 displays elevated levels of protein expression in human breast cancer. *Cancer research*. 2004; **64**: 7370-6.
- 47 Kim S, Lee K, Choi JH, Ringstad N, Dynlacht BD. Nek2 activation of Kif24 ensures cilium disassembly during the cell cycle. *Nature communications*. 2015; **6**: 8087.
- 48 Cappello P, Blaser H, Gorrini C, *et al*. Role of Nek2 on centrosome duplication and aneuploidy in breast cancer cells. *Oncogene*. 2014; **33**: 2375-84.
- 49 Ejlersen B, Tuxen MK, Jakobsen EH, *et al*. Adjuvant Cyclophosphamide and Docetaxel With or Without Epirubicin for Early TOP2A-Normal Breast Cancer: DBCG 07-READ, an Open-Label, Phase III, Randomized Trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2017; **35**: 2639-46.
- 50 Zheng H, Li X, Chen C, *et al*. Quantum dot-based immunofluorescent imaging and quantitative detection of TOP2A and prognostic value in triple-negative breast cancer. *International journal of*

nanomedicine. **11**: 5519-29.

- 51 Chen JR, Chien HP, Chen KS, *et al*. Amplification of HER2 and TOP2A and deletion of TOP2A genes in a series of Taiwanese breast cancer. *Medicine*. 2017; **96**: e5582.
- 52 Yang Y, Li DP, Shen N, *et al*. TPX2 promotes migration and invasion of human breast cancer cells. *Asian Pacific journal of tropical medicine*. 2015; **8**: 1064-70.
- 53 Cocchiola R, Grillo C, Altieri F, Chichiarelli S, Turano C, Eufemi M. Upregulation of TPX2 by STAT3: identification of a novel STAT3 binding site. *PloS one*. 2014; **9**: e113096.
- 54 Huang C, Han Z, Wu D. Effects of gene on radiotherapy sensitization in breast cancer stem cells. *Oncology letters*. 2017; **14**: 1531-35.
- 55 Mo CH, Gao L, Zhu XF, *et al*. The clinicopathological significance of UBE2C in breast cancer: a study based on immunohistochemistry, microarray and RNA-sequencing data. *Cancer cell international*. 2017; **17**: 83.
- 56 Loussouarn D, Campion L, Leclair F, *et al*. Validation of UBE2C protein as a prognostic marker in node-positive breast cancer. *British journal of cancer*. 2009; **101**: 166-73.
- 57 Chou CP, Huang NC, Jhuang SJ, *et al*. Ubiquitin-conjugating enzyme UBE2C is highly expressed in breast microcalcification lesions. *PloS one*. 2014; **9**: e93934.
- 58 Qin T, Huang G, Chi L, *et al*. Exceptionally high UBE2C expression is a unique phenomenon in basal-like type breast cancer and is regulated by BRCA1. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*. 2017; **95**: 649-55.
- 59 Morreale FE, Bortoluzzi A, Chaugule VK, Arkinson C, Walden H, Ciulli A. Allosteric Targeting of the Fanconi Anemia Ubiquitin-Conjugating Enzyme Ube2T by Fragment Screening. *Journal of medicinal chemistry*. 2017; **60**: 4093-98.
- 60 Perez-Peña J, Corrales-Sánchez V, Amir E, Pandiella A, Ocana A. Ubiquitin-conjugating enzyme E2T (UBE2T) and denticleless protein homolog (DTL) are linked to poor outcome in breast and lung cancers. *Scientific reports*. 2017; **7**: 17530.
- 61 Ueki T, Park JH, Nishidate T, *et al*. Ubiquitination and downregulation of BRCA1 by ubiquitin-conjugating enzyme E2T overexpression in human breast cancer cells. *Cancer research*. 2009; **69**: 8752-60.

Figures

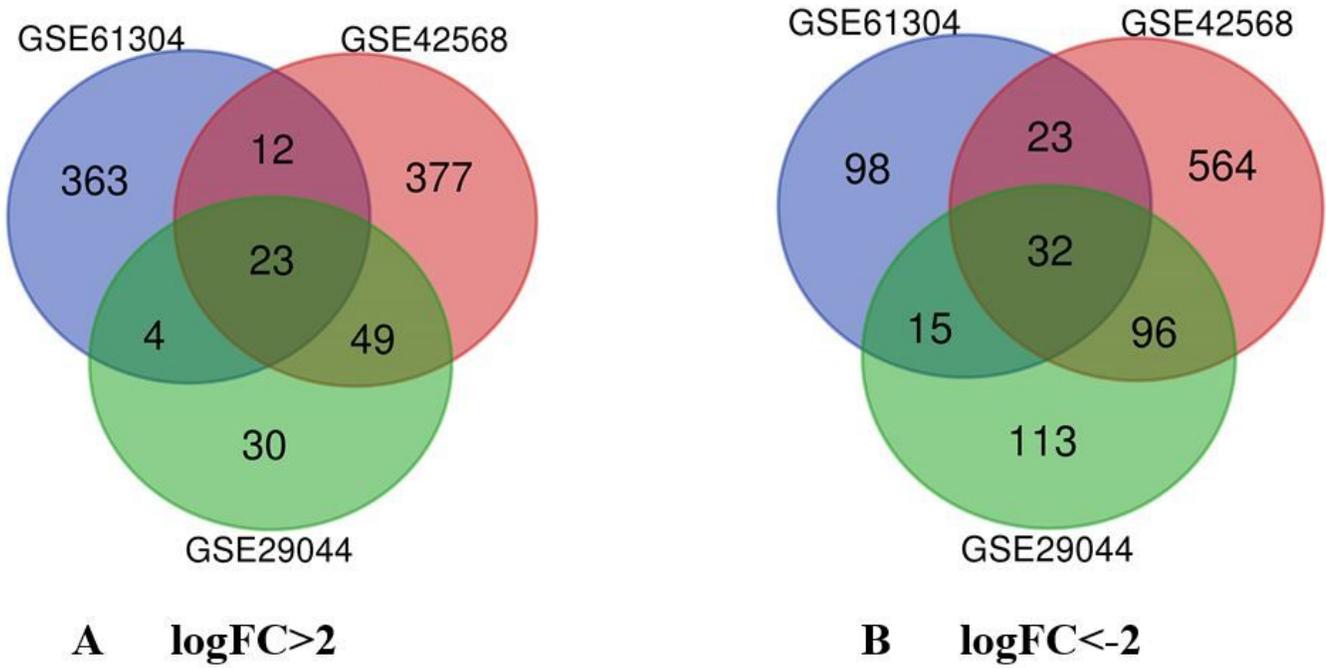


Figure 1

55 DEGs were screen out from three datasets (GSE61304, GSE29044 and GSE42568) by Venn diagrams software a. the 23 in the intersection of three circular areas were up-regulated DEGs ($\log FC > 2$). b. the 32 in the intersection of three circular areas were down-regulated DEGs ($\log FC < -2$).

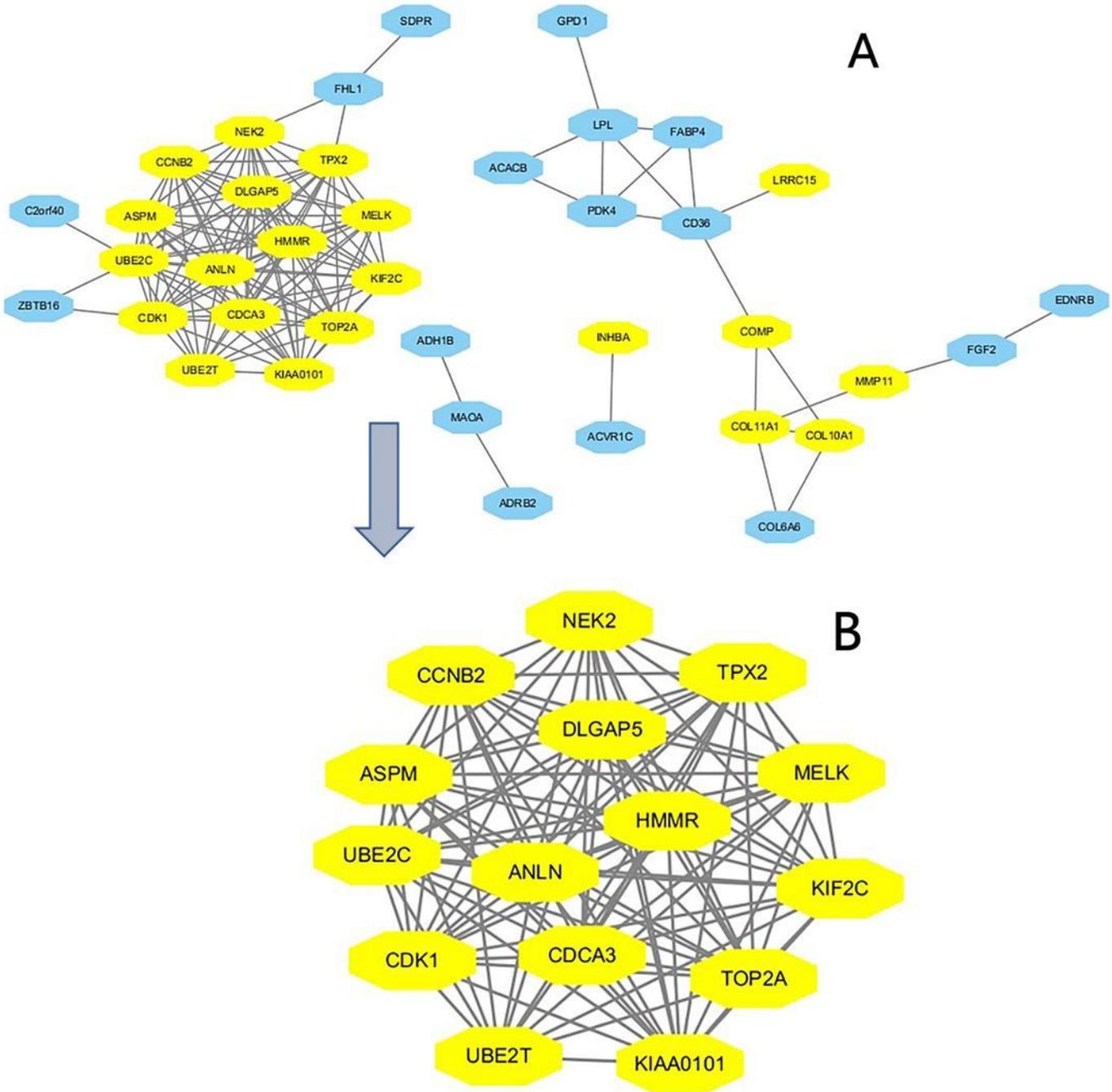


Figure 2

Construction of up-regulated and down-regulated DEGs PPI network by STRING online database and Module analysis. a The nodes represent proteins, the gray lines represent interactions of proteins, yellow octagons represent up-regulation DEGs and blue octagons represent down-regulation DEGs. b Module analysis by Cytoscape software (degree cutoff = 2, node score cutoff = 0.2, k-core = 2, and max. Depth = 100).

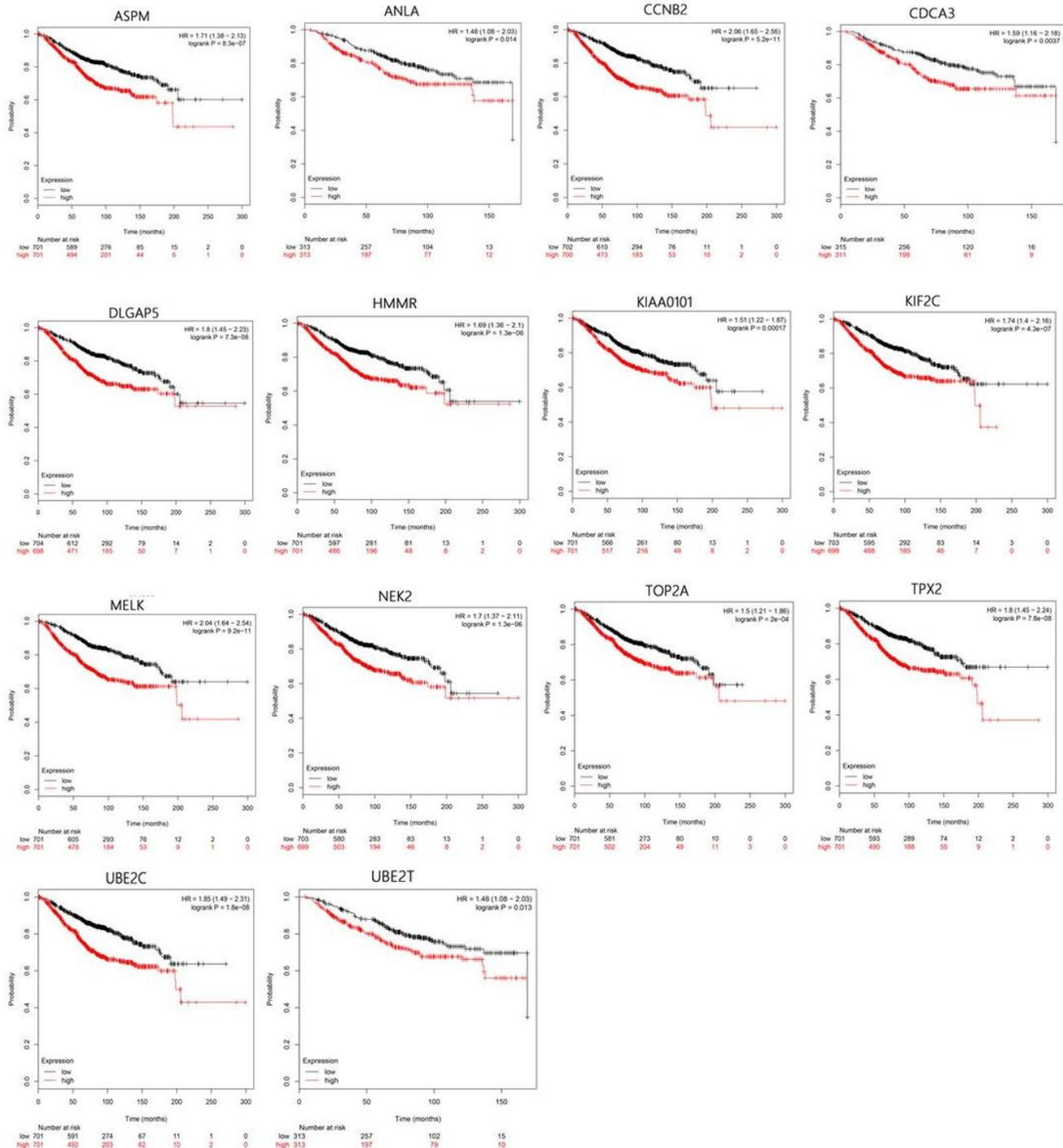


Figure 3

The prognostic of the key candidate genes. Prognostic information of the core genes was identified via Kaplan Meier plotter online tools and 14 genes had a significantly worse survival rate ($P < 0.05$).

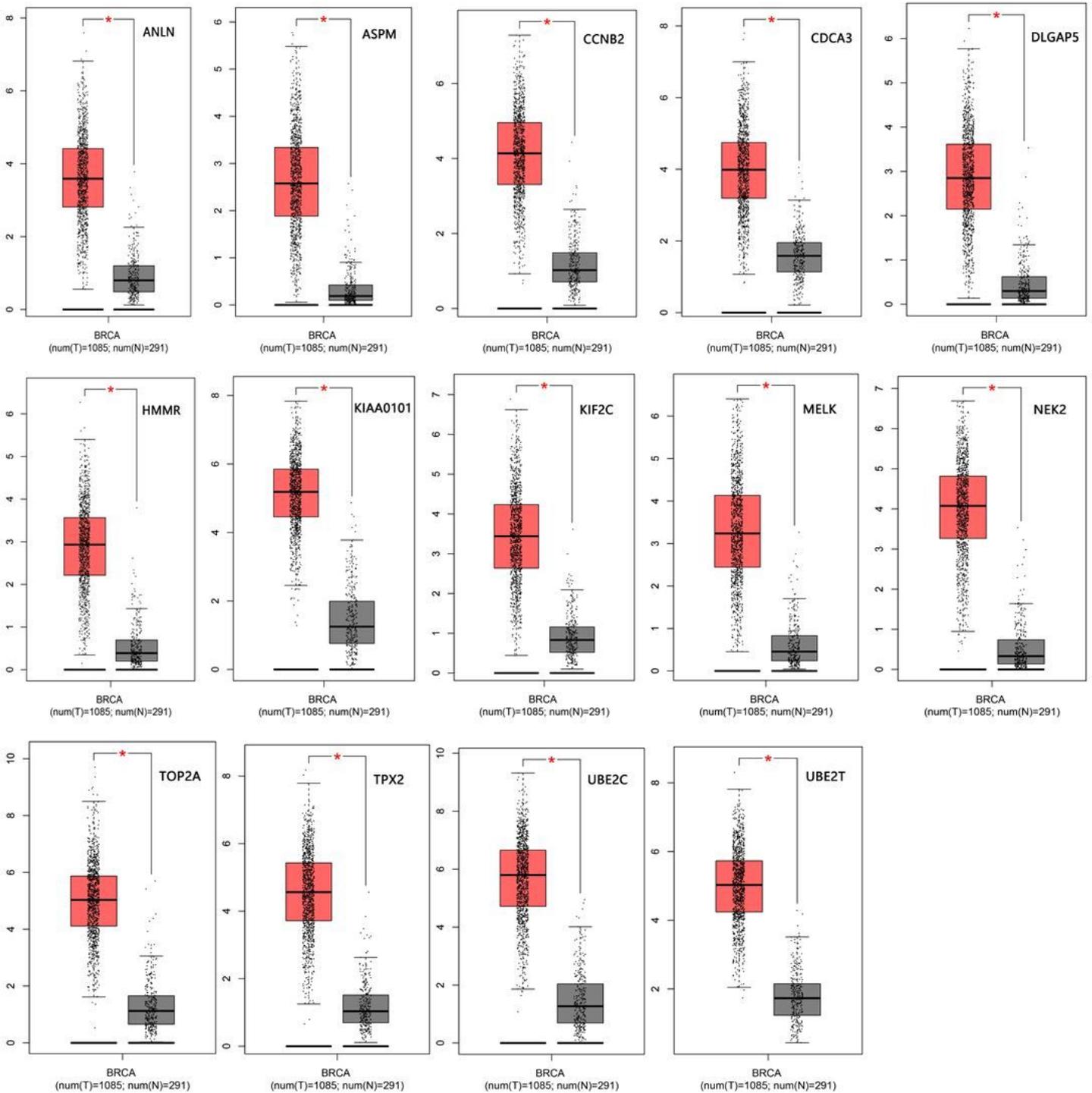


Figure 4

Compared with healthy people, 14 genes were significantly expressed in breast cancer patients. In order to further determine the gene expression level between breast cancer patients and healthy people, 14 genes related to poor prognosis were analyzed by GEPIA website. The expression level of these genes in breast cancer specimens was significantly higher than that in normal specimens (* $P < 0.05$). Red color means tumor tissues and grey color means normal tissues.