*Vietnam Journal of Biotechnology* **21**(1): 21-34, 2023

# IDENTIFICATION OF HUB GENES AND DRUG-GENE INTERACTIONS FOR TARGETED BREAST CANCER TREATMENT BY INTEGRATED BIOINFORMATICS ANALYSIS

**Dao Manh Cuong[1], Bui Van Ngoc[1,2,✉]**

[1]*Institute of Biotechnology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Road, Cau Giay District, Hanoi, Vietnam*
[2]*Graduate University of Science and Technology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Road, Cau Giay District, Hanoi, Vietnam*

✉To whom correspondence should be addressed. E-mail: bui@ibt.ac.vn

Received: 11.8.2022
Accepted: 27.02.2023

SUMMARY

Breast cancer (BC) is one of the most common cancer types in women. In addition to conventional methods for BC diagnosis, applying methods for a fast and accurate prognosis at the early stage of cancer is very meaningful for the treatment of the disease. To date, the most advanced methods are molecular diagnostics and bioinformatics. In this study, bioinformatics is applied to genetic testing for BC diagnosis; namely the R programming language combined with the bioinformatics toolkit was used to analyze gene expression levels between normal and tumor tissues in three gene expression profiles (GSE29431, GSE42568, GSE21422). The bioinformatics approaches included identification of differentially expressed genes (DEGs) and hub genes, Gen Ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) analyses, the construction of a protein-protein interaction (PPI) network, and module analysis. Following the completion of the hub gene selection process, coexpression and survival analysis were carried out. Finally, the GEPIA2 and DGIdb databases were utilized to verify the expression levels of hub genes and select the candidate drugs for BC, respectively. A total of 1369 DEGs was identified, including 400 upregulated DEGs and 969 downregulated DEGs. Thereafter, 10 hub genes (*CDK1*, *CCNA2*, *CCNB1*, *CCNB2*, *TOP2A*, *KIF11*, *RRM2*, *BUB1B*, *CDC20*, and *NCAPG*) were identified as potential biomarkers for BC diagnosis, prognosis, and therapy. Six screened small molecules, dexrazoxane, teniposide, amsacrine, etoposide, mitoxantrone and daunorubicin, were determined to be the new targeted drugs for BC treatment.

**Keywords:** Biomarker, breast cancer, drug-gene interaction, hub gene, protein-protein interaction

## INTRODUCTION

Breast cancer is the most commonly diagnosed cancer in women worldwide. According to GLOBOCAN 2020, about 2 million people worldwide were diagnosed with BC and 600,000 people died from this disease. Traditionally, some methods, including mammography, ultrasound, and some high-end molecular bioimaging, are used to diagnose breast cancer. Unfortunately, these techniques often detect BC at an advanced stage and lack sufficient specificity (Zubair *et al*., 2020; Esserman *et al*., 2007). Besides, the relative survival rate for females diagnosed with early-stage BC (stage 1) at diagnosis remained at 100% at 1, 3, and 5 years periods, despite the fact that this rate for patients with metastatic BC (stage 4) reduced to 69% at 1 year, 47% at 3 years and 32% at 5 years from diagnosis (Li *et al*., 2020).

Furthermore, chemotherapy resistance remains a significant challenge and is responsible for the majority of treatment failures, resulting in lower overall survival of BC patients (Prihantono, Faruk, 2021). Therefore, more efforts need to be invested in identifying and understanding novel biomarkers and specific targets of BC, which is considered as the key to developing more effective diagnostic and therapeutic strategies.

In recent years, gene profiles and gene chips have been employed to screen differentially expressed genes (DEGs) (Yang *et al*., 2020). Nevertheless, current research on biomarkers may be insufficient, and the DEG results could be inconsistent owing to data collection, the complex tumor heterogeneity, and the complicated molecular regulatory mechanism of BC. Therefore, reanalyzing and combining gene expression datasets can provide new insights into the current study on BC. Furthermore, a considerable number of bioinformatics studies on cancer have been proven to be effective and reliable (Feng *et al*., 2019; Ren *et al*., 2020), which means integrated bioinformatics analysis could assist with exploring the biomarkers and the mechanisms underlying the tumorigenesis and progression of cancer.

In this study, three gene expression profiles (GSE29431, GSE42568, and GSE21422) were downloaded from the Gene Expression Omnibus (GEO) database. Integrated bioinformatics analyses included identification of DEGs, Gene Ontology (GO) term analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis, protein-protein interaction (PPI) construction, discovery of hub genes, survival analysis, validation of hub genes, and detection of hub gene-drug interaction.

## MATERIALS AND METHODS

### Acquisition of microarray data

The NCBI-GEO database is a free and public database containing gene profiles. Three microarray datasets (GSE29431, GSE42568, and GSE21422) were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/gds/). These gene expression profiles consisted of tissue samples obtained from human BC tissues and normal tissues and were based on GPL570 (Affymetrix Human Genome U133 Plus 2.0 Array).

### Analysis of differentially expressed genes (DEGs)

The GEO2R tool was used to screen DEGs between breast cancer and normal breast tissues (Barrett *et al*., 2013). It was used to compare and identify DEGs presented in two sample groups in the GEO series. The cut-off criteria of |log fold change (FC)| ≥ 1 and adjusted *P-value* < 0.05 were considered statistically significant (Xu *et al*., 2016). The DEGs in the three datasets were screened using the VennDiagram package in R, with logFC ≥ 1 considered as upregulated genes and logFC ≤ -1 considered as downregulated genes.

### Enrichment analysis via GO and KEGG pathway

The DEGs data was imported into the DAVID online functional annotation bioinformatics microarray analysis (https://david.ncifcrf.gov/). Functional enrichment of DEGs was carried out, using GO analysis to examine biological process (BP), cellular component (CC), molecular function (MF) and KEGG pathway analysis. The top 10 items according to gene counts in the BP, CC, and MF categories and KEGG pathways were shown using the ggplot2 R package through the statistical software R. An adjusted *P-value* < 0.05 was used as the inclusion criterion.

### Construction of PPI network and module analysis

The PPI network of the identified DEGs was constructed by an online tool, the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING; https://string-db.org/), with a minimum required interaction score > 0.7. The active interaction sources included text mining, experiments, databases, coexpression, neighborhood, gene fusion, and corecurrence.

The degree of all nodes was calculated by the Cytoscape software (version 3.9.1) plugin cytoHubba. In this study, genes corresponding to the top 10 highest degree values were regarded as hub genes. KEGG pathway analysis, PPI network construction, and coexpression analysis of the hub genes were performed by DAVID and STRING.

**Survival analysis and validation of the hub genes**

The Kaplan-Meier plotter could assess the prognostic effect of genes on survival in many types of cancer (http://kmplot.com/analysis/). Patients with BC were split into two groups, namely, a high-expression group and a low-expression group, according to the expression of a particular gene. Then, the overall survival (OS) was analyzed for the above two groups for each hub gene in the early stages (stages 1 and 2). These analyses were shown in the form of survival curves according to the hazard ratio (HR), 95% confidence interval (95% CI) and log-rank *P-value*. Additionally, the expression levels of the hub genes between breast cancer and normal samples were verified by Gene Expression Profiling Interactive Analysis (GEPIA2, http://gepia2.cancer-pku.cn/#index). Next, UALCAN, an established resource for analyzing transcriptome data of cancers based on The Cancer Genome Atlas (TCGA) (https://doi.org/10.1016/j.neo.2017.05.002), was used to explore the variations of expression level between BC and normal samples and among BC samples at different stages. Furthermore, the cBio Cancer Genomics Portal (https://www.cbioportal.org/; version 4.1.13) online tool was used to present the genetic alteration information of the hub genes.

**Drug-hub gene interaction**

Selecting drugs based on the hub genes that served as promising targets was conducted using the Drug-Gene Interaction Database (DGIdb; https://www.dgidb.org/; v4.2.0 – shal afd9f30b). The interaction network between the potential drugs and the hub genes was constructed by Cytoscape software. In this study, the final drug list included only drugs that were approved by the Food and Drug Administration (FDA). The online tool STITCH (http://stitch.embl.de/) was applied to construct the interaction network between the potential drugs and the hub genes.

## RESULTS AND DISCUSSION

**Identification of DEGs in breast tumors**

We obtained high-throughput gene expression profiles of breast cancer and normal breast tissues from the GEO database. Three independent datasets (GSE29431, GSE42568, and GSE21422), based on the GPL570 platforms, were selected. Dataset GSE29431 consisted of 54 primary breast carcinomas and 12 normal breast tissues. Dataset GSE42568 was composed of 104 breast cancer tissues (11 tumors were grade 1, 40 were grade 2, and 53 were grade 3), and 17 normal breast tissues. Dataset GSE21422 included 14 breast cancer tissues (6 DCIS (ductal carcinoma *in situ*) and three IDC (invasive ductal carcinoma) grade 3, three DCIS, and two IDC grade 2), and five normal breast tissues. The specific details of the above datasets are presented in Table 1.

**Table 1.** Detailed information on the GEO microarray profiles of breast cancer patients.

| No. of GEO profile | Type | Source | Case | control | Platform | Annotation platform |
|---|---|---|---|---|---|---|
| GSE29431 | mRNA | breast tumor | 54 | 12 | GPL570 | Affymetrix Human Genome U133 Plus 2.0 Array |
| GSE42568 | mRNA | breast tumor | 104 | 17 | GPL570 | Affymetrix Human Genome U133 Plus 2.0 Array |
| GSE21422 | mRNA | breast tumor | 14 | 5 | GPL570 | Affymetrix Human Genome U133 Plus 2.0 Array |

A total of 2705 (899 upregulated and 1806 downregulated), 3986 (1872 upregulated and 2114 downregulated), and 2914 (1306 upregulated and 1608 downregulated) DEGs was identified from the GSE29431, GSE42568 and GSE21422 datasets, respectively. Of those, 1369 genes were presented in all three datasets, which included 400 upregulated genes and 969 downregulated genes in BC tissues (Fig. 1).


**Figure 1.** Selection of 1369 common DEGs from the three datasets (GSE29431, GSE42568, and GSE21422). (A) Venn diagram illustrating 400 upregulated DEGs (logFC ≥ 1); (B) Venn diagram illustrating 969 downregulated DEGs (logFC ≤ -1). The number denotes the number of genes shared between/among subtypes or the number of unique genes.

## Functional enrichment analysis of DEGs

DAVID, a tool for the analysis of genes and proteins, was used for gene ontology funtional annotation and biological pathway enrichment analysis of DEGs. The GO function annotation results of upregulated genes showed that these DEGs were mainly involved in the following BP: biological regulation, regulation of biological process, regulation of cellular process, response to stimulus, cellular component organization or biogenesis, cellular component organization, localization, multicellular organismal process, cellular response to stimulus, developmental process; the primary CC: intracellular anatomical structure, organelle, intracellular organelle, cytoplasm, membrane-bounded organelle, intracellular membrane-bounded organelle, membrane, intracellular non-membrane-bounded organelle, non-membrane-bounded organelle, intracellular organelle lumen, membrane-enclosed lumen, organelle lumen; the key MF: binding, protein binding, enzyme binding, identical protein binding, small molecule binding, cytoskeletal protein binding, kinase binding, protein kinase binding, protein domain specific binding, cell adhesion molecule binding (Fig. 2A); while the downregulated genes were mainly related to cellular process, biological regulation, regulation of biological process, regulation of cellular process, response to stimulus, cellular response to stimulus, multicellular organismal process, developmental process, positive regulation of biological process, cell communication (BP); cytoplasm, membrane, cell periphery, plasma membrane, cytosol, intrinsic component of membrane, integral component of membrane, endomembrane system, extracellular region, vesicle (CC); vesicle, protein binding, ion binding, catalytic activity, cation binding, metal ion binding, identical protein binding, small molecule binding, enzyme binding, molecular function regulator activity (MF) (Fig. 2B).

The KEGG pathway analysis revealed that the upregulated genes were enriched in cell cycle, human papillomavirus infection, proteoglycans in cancer, tight junction, oocyte meiosis, and ECM-receptor interaction (Fig.

3A), while the downregulated genes were enriched in metabolic pathways, PPAR signaling pathway, cAMP signaling pathway, cGMP-PKG signaling pathway, focal adhesion, carbon metabolism, AMPK signaling pathway, pyruvate metabolism, regulation of lipolysis in adipocytes, fatty acid degradation, insulin resistance, and glucagon signaling pathway (Fig. 3B).

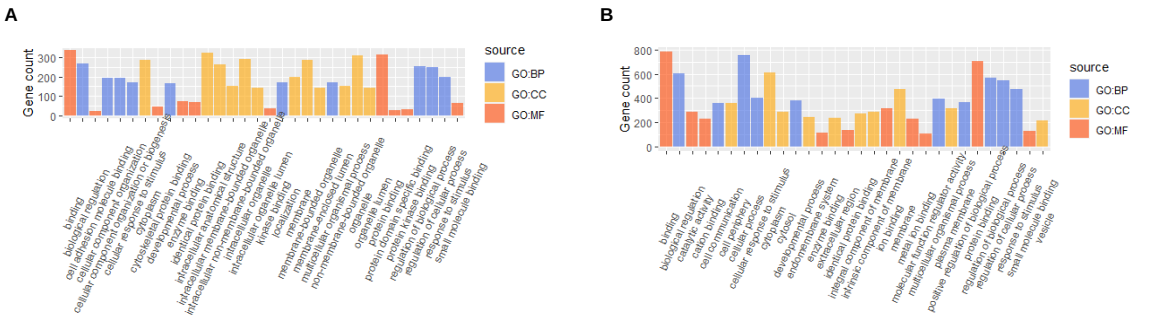


**Figure 2.** GO analyses of DEGs including biological process, cellular component, and molecular function. (A) GO analysis with upregulated DEGs; (B) GO analysis with downregulated DEGs. Top 10 GO terms according to gene count of each source were shown, with adjusted *P-value* < 0.05.

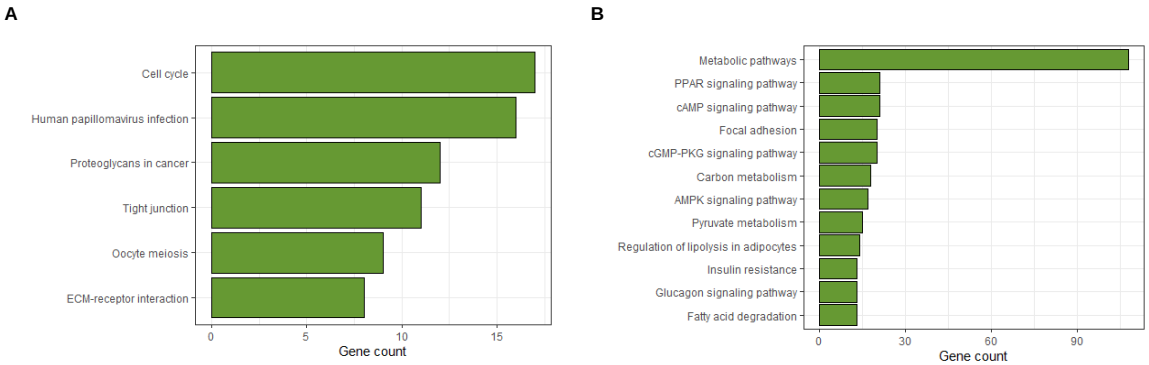


**Figure 3.** KEGG analyses of DEGs. (A) KEGG analysis with upregulated DEGs; (B) KEGG analysis with downregulated DEGs. Top 10 enriched KEGG pathways according to gene count were shown, with adjusted *P-value* < 0.05.

## PPI network construction and module analysis

The PPI network of the DEGs based on the information obtained from the STRING database was constructed. When 1369 DEGs were submitted to the STRING database, we obtained a PPI network that included 1266 nodes, 3629 edges and its PPI enrichment *P-value* was lower than 1.0E-16 (Fig. 4A). Ten genes (*CDK1*, *CCNA2*, *CCNB1*, *CCNB2*, *TOP2A*, *KIF11*, *RRM2*, *BUB1B*, *CDC20*, *NCAPG*) with the highest degree scores were identified as the hub genes for BC by applying the cytoHubba plugin and all of which were upregulated DEGs. The PPI network of the 10 screened hub genes including 10 nodes and 45 edges was constructed by cytoHubba software (Fig. 4B). Additionally, KEGG pathway analysis of ten hub genes performed by DAVID indicated that these genes mainly involved in cell cycle, p53 signaling pathway, progesteron-mediated oocyte

maturation, oocyte meiosis, cellular senescence. Furthermore, the coexpression analysis result of hub genes showed that they could actively interact with each other (Fig. 4C). Likewise, these above genes were also selected as hub genes in other researches (Deng *et al.*, 2019; Moradpoor *et al*., 2021; Weng *et al*., 2021; Zeng *et al*., 2021).

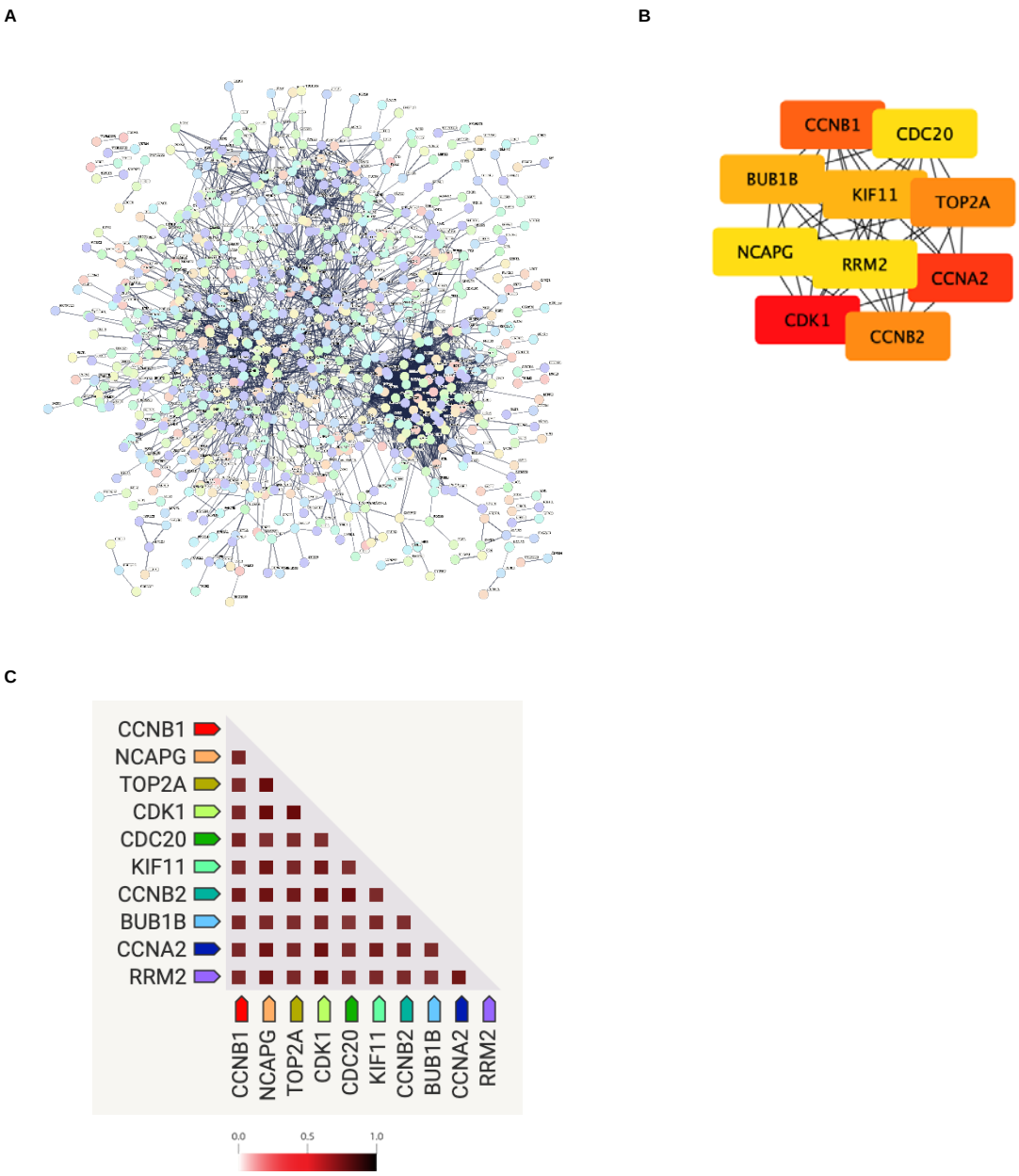


**Figure 4.** PPI network and coexpression of the hub genes in BC. (A) The PPI network of 1369 DEGs using the STRING online database. (B) The PPI network of hub genes using the cytoHubba software. (C) The coexpression analysis of the hub genes using the STRING online database. In the triangle-matrix, the intencity of color indicates the level of confidence that two proteins are functionally associated.

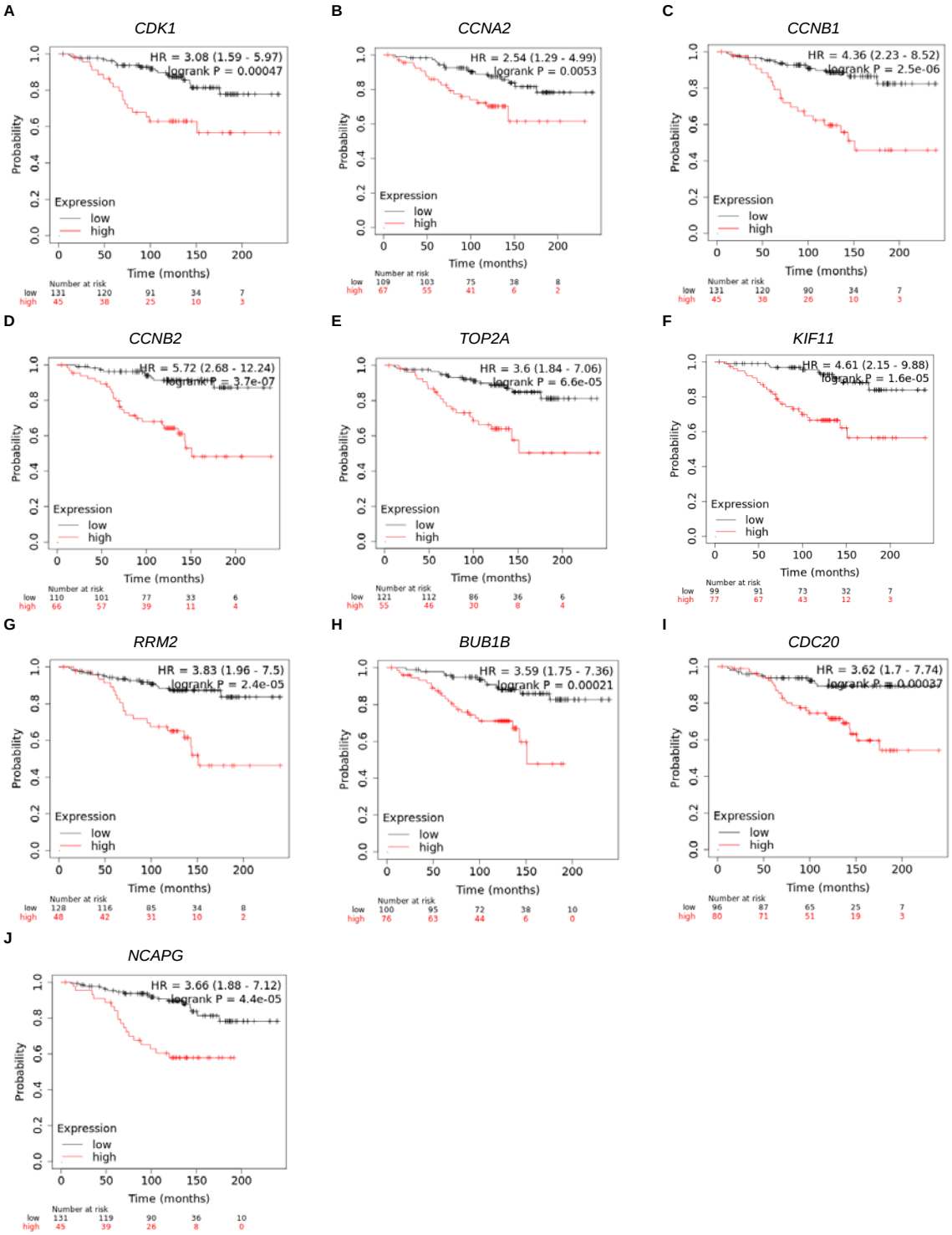


**Figure 5.** Overall survival analyses of hub genes in patients with stage 1 or 2 BC. (A-J) Survival curves were constructed by Kaplan-Meier plotter online database based on the low and high expression of hub genes in BC. Log-rank $P < 0.05$ was considered statistically significant.

**Survival analysis, genetic information and hub gene expression**

The prognostic information of 10 hub genes for BC at stages 1 and 2 was analyzed by the Kaplan-Meier plotter database. The survival curve of these genes was presented in Fig. 5A-J. All 10 hub genes were significantly associated with OS of BC patients at early stages (stages 1 and 2) (Fig. 5). Early stage BC patients with higher expression levels of *CDK1* [HR = 3.08 (1.59-5.97), *P* = 0.00047], *CCNA2* [HR = 2.54 (1.29-4.99), *P* = 0.0053], *CCNB1* [HR = 4.36 (2.23-8.52), *P* = 2.5E-06], *CCNB2* [HR = 5.72 (2.68-12.24), *P* = 3.7E-07], *TOP2A* [HR = 3.6 (1.84-7.06), *P* = 6.6E-05], *KIF11* [HR = 4.61 (2.15-9.88), *P* = 1.6E-05], *RRM2* [HR = 3.83 (1.96-7.5), *P* = 2.4E-05], *BUB1B* [HR = 3.59 (1.75-7.36), *P* = 0.00021], *CDC20* [HR = 3.62 (1.7-7.74), *P* = 0.00037], *NCAPG* [HR = 3.66 (1.88-7.12), *P* = 4.4E-05] were significantly related to poorer OS (Fig. 5A - J). Elevated *CDK1*, *CCNA2* and *CCNB1* expression levels have been strongly associated with poor overall survival, post-progression survival and recurrence-free probability rates in patients with BC (Xing *et al*., 2021; Qiu *et al*., 2021). High expression of *CDC20*, *CCNA2* and *RRM2* has been related to the low survival of HER2+ patients (Weng *et al*., 2021). High expression of *CDK1*, *TOP2A*, *RRM2* and *CCNB2* has reduced the patient's survival in BC (Jayanthi *et al*., 2020). High *TOP2A* expression has shown a worse prognosis on stage I-II luminal breast cancer (An *et al*. 2018), ER+ breast cancer (Rody *et al*., 2009), HER2+ and HER2- breast cancer (Zaczek *et al*., 2012) patients. The low survival rate of BC patients has been strongly associated with overexpression of *CDK1* and *CDC20* (Wang *et al*., 2021), *CCNB1* (Fang *et al*., 2022), *KIF11* (Wang *et al*., 2020), *BUB1B* (Koyuncu *et al*., 2021) and *NCAPG* (Dong *et al*., 2021). Therefore, upregulated expression of 10 hub genes could be prognostic biomarkers of BC and these hub genes could act as prospective targets for chemotherapy.

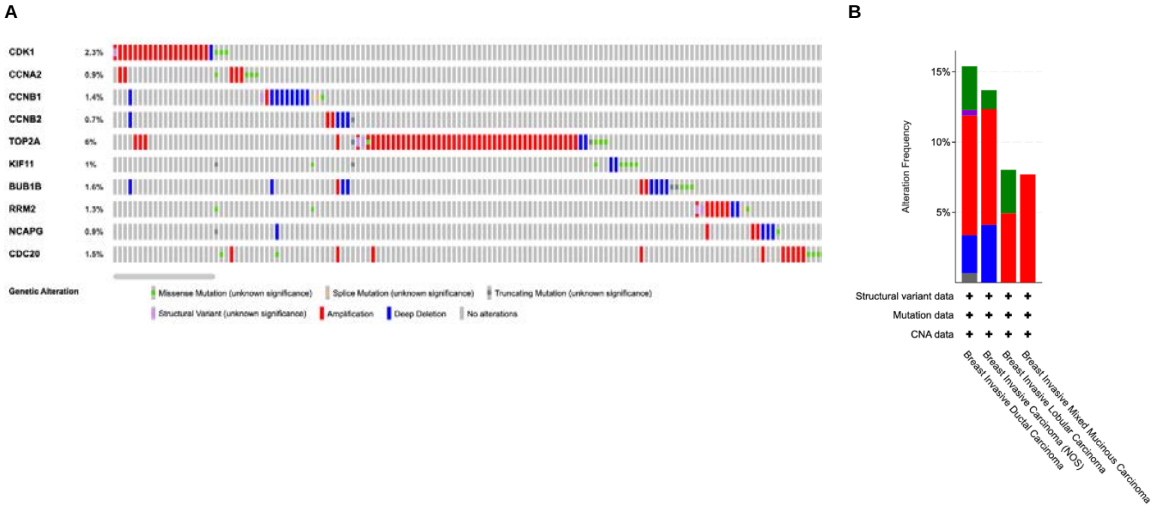


**Figure 6.** Information on the genetic alterations of the hub genes. (A) The genetic alterations related to the hub genes were shown through a visual summary across a set of breast invasive carcinoma samples (data from TCGA, PanCancer Atlas). (B) An overview of the alterations of the hub genes in the genomics datasets of breast invasive carcinoma in the TCGA database.

Subsequently, cBioPortal was used to determine the genetic alteration information of 10 hub genes (Fig. 6). As illustrated in Fig. 6A, the hub genes were altered in 140 (14%) of queried patients or samples. *TOP2A* and *CDK1* were altered most often (6% and 2.3%, respectively). These

alterations included amplification, deep deletion, truncating mutation, missense mutation, splice mutation, structural variant (Fig. 6A). Among different types of alteration, amplification, missense mutation and deep deletion accounted for the highest percentage (Fig. 6B).

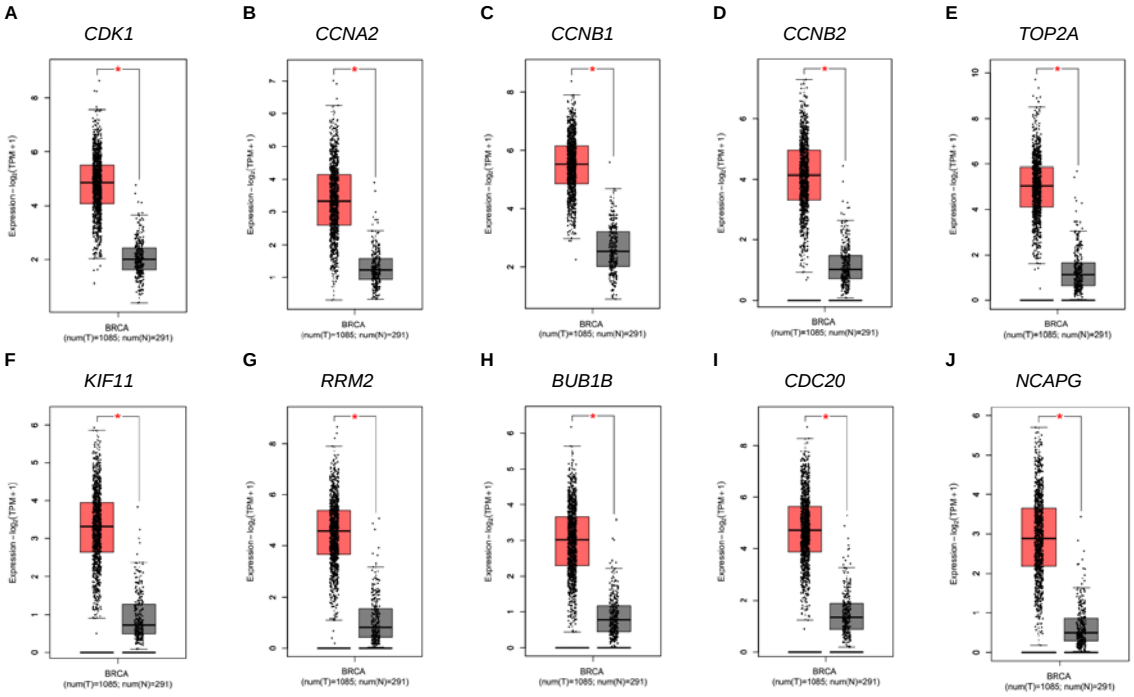


**Figure 7.** The expression level of hub genes in breast cancer tissues and normal tissues from patients (A-J). To further verify the expression level of the hub genes between BC tissues and normal tissues, the hub genes were analyzed by the GEPIA2 online database. $P < 0.01$ was considered statistically significant. *, $P < 0.01$; tumor color, red; normal color, black.

In addition, the GEPIA2 databases were used to verify the expression level of hub genes between BC and normal tissues. As shown in Fig. 7A-J, the expression levels of 10 hub genes were all statistically significant ($P < 0.01$) in BC and normal tissues on the basis of gene expression profiles from the cancer genome atlas (TCGA) and the genotype-tissue expression (GTEx) project. The findings were consistent with the obtained GEO datasets. Therefore, both the GEPIA2 and GEO databases indicated that the mRNA expression levels of the 10 hub genes were upregulated in tumor tissues. *CDK1*, *CCNA2* and *CCNB1* expression levels are higher in BC compared with control tissue samples (Xing *et al*., 2021). *CDK1*, *CCNA2* and *CCNB1* are significantly upregulated in cancer tissues (Weng *et al*., 2021). There has been a significant increase in the expression of *CDK1*, *TOP2A*, *RRM2* and *CCNB2* of BC tissues compared to normal breast tissues (Jayanthi *et al*., 2020). In BC tissues, significantly elevated expression levels compared to normal breast tissues have also been observed in *CDK1* and *CDC20* (Wang *et al*., 2021), *KIF11* (Wang *et al*., 2020), *BUB1B* (Koyuncu *et al*., 2021) and *NCAPG* (Dong *et al*., 2021).

Additionally, the UALCAN databases were used to identify the expression levels of hub genes in different stages of BC compared with normal tissues. The expression level of 10 hub genes in BC patients at different stages was shown in Fig. 8A-J. According to these results, it was easy to see that the expression of 10 hub genes was higher than those in normal tissues (*P*

< 0.05) in the early stages (stages 1 and 2). Therefore, these hub genes could become biomarkers for BC diagnosis in early stages, and hence, it could increase the possibility of treatment for patients. Moreover, there were significant variations in the expression levels of 10 hub genes in BC patients in different stages. The overall trends indicated that the expression of hub genes increased from stage 1 to stage 2 and remained at stage 3 and then decreased gradually with the continuous progression of BC (Fig. 8). Although there had been variation in expression levels of hub genes in different stages. However, hub gene expression levels in any grade of BC were all higher than those of normal tissues. Therefore, these hub genes have remained to be potential drug targets in BC treatment.

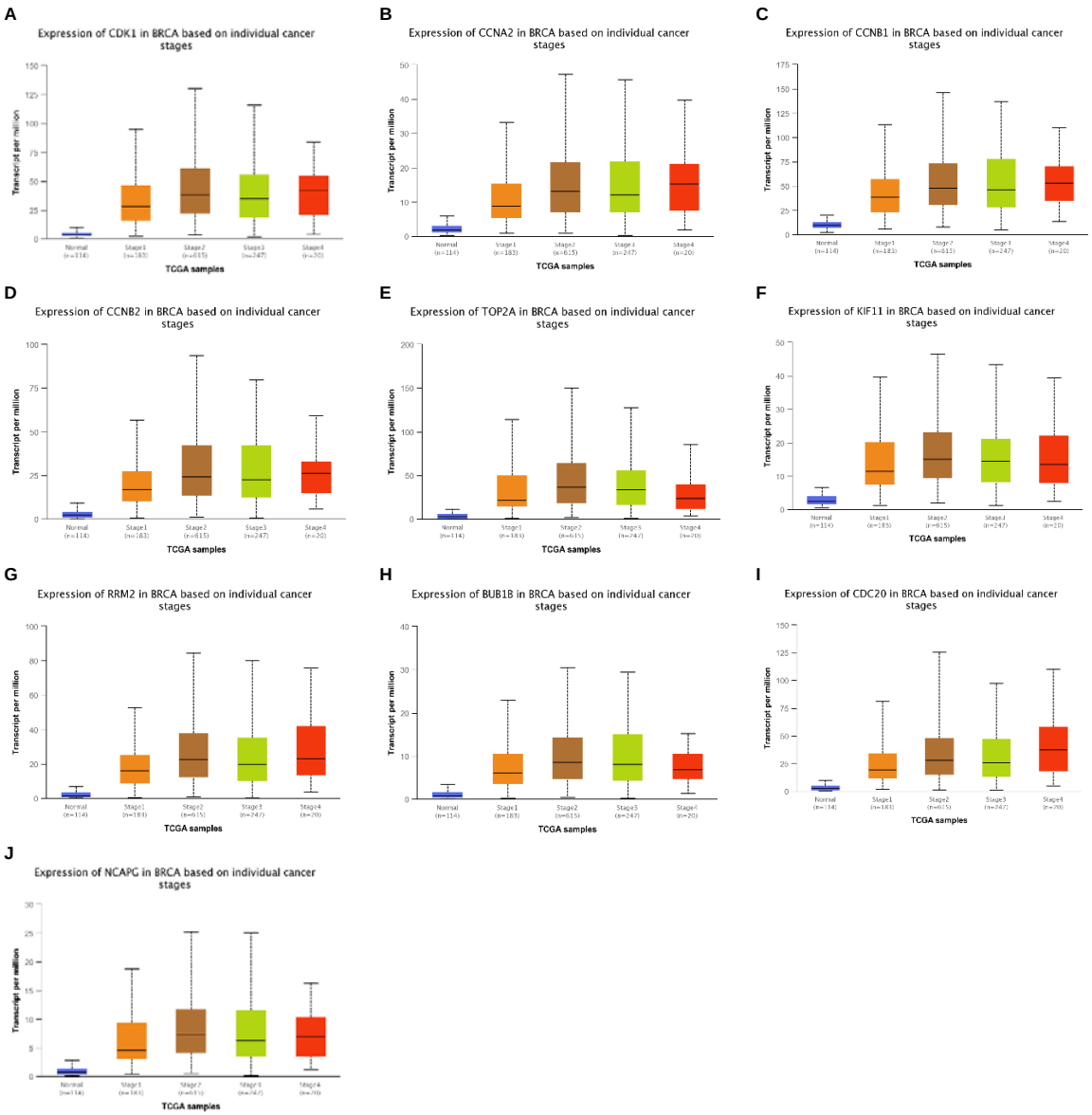


**Figure 8.** The expression of hub genes in different cancer stages compared within normal samples. Breast invasive carcinoma samples from TCGA database were taken into account. *P* < 0.05 was considered statistically significant.

**Table 2.** Candidate drugs targeting hub genes.

| No. | Gene | Drug | Interaction types | Sources | PIMDs |
|---|---|---|---|---|---|
| 1 | *TOP2A* | DEXRAZOXANE | inhibitor | NCI | 11179439, 17652819, 10194547, 11046078, 12911317, 17115008, 11752352, 11984069, 11332155 |
| 2 | *TOP2A* | TENIPOSIDE | inhibitor | TdgClinicalTrial, ChemblInteraction, NCI, TEND | 8702194, 16271071, 17361331, 17514873, 11752352, 16480143, 9426516 |
| 3 | *TOP2A* | AMSACRINE | inhibitor | DTC, NCI | 1322791, 8823806, 10691026, 8519659, 19155103, 22537681, 17911018, 8632768, 19725581, 11006484, 11716434, 11752352, 25626146, 11473732, 1311390 |
| 4 | *TOP2A* | ETOPOSIDE | inhibitor | DTC, TdgClinicalTrial, ChemblInteractions, NCI, TEND | 25466187, 20006518, 18258442, 8823806, 22867019, 25240702, 26291037, 25003995, 26216018, 26292628, 23360284, 16271071, 23920485, 21435753, 22867079, 16759114, 11678653, 19386396, 24931277, 23566520, 17361331, 25922181, 25941559, 24507920, 24775914, 9485461, 23353750, 25815139, 16309315, 24012683, 19691293, 25800514, 21644529, 22620261, 25945730, 24334150, 17514873, 8870683, 23711769, 11752352, 20863598, 24095018, 26264845, 25799376, 22364746, 16377807, 9494516, 23968711, 18816045, 24326278, 19783445, 9426516 |
| 5 | *TOP2A* | MITOXANTRONE | inhibitor | TdgClinicalTrial, NCI, TEND | 10451375, 11004693, 18687447, 11752352, 9631585, 9494516, 11278845, 9426516 |
| 6 | *TOP2A* | DOXORUBICIN | inhibitor | DTC, TdgClinicalTrial, ClearityFoundationClinicalTrial, TEND | 21388138, 17016621, 17578914, 17010609, 17351394, 26211460, 11752352, 20170164, 17089011, 22276998 |
| 7 | *TOP2A* | HYDROQUINONE | - | NCI | 15833037 |
| 8 | *TOP2A* | DAUNORUBICIN | inhibitor | DTC, TdgClinicalTrial, NCI | 22260166, 1963303, 6380596, 9494516 |
| 9 | *TOP2A* | VINCRISTINE | - | NCI | 9494516 |
| 10 | *RRM2* | GEMCITABINE | inhibitor | ClearityFoundationClinicalTrial, TTD | - |

## Drug-gene interaction

The drug-gene interactions were explored using DGIdb. Consequently, a total of 10 potential drugs for treating BC patients was identified (Table 2). In this study, according to the statistically significant results of the survival analysis, *TOP2A* and *RRM2* were selected as the potential targets of 10 drugs (Table 2). The 10 promising candidate drugs were all approved by

the FDA. Most potential drugs might interact with the *TOP2A* (9/10), either as an inhibitor or in some unknown manners (Table 2). Only a few of the *TOP2A*- and *RRM2*-targeting drugs have been used for BC treatment, such as doxorubicin and gemcitabine. The rest dexrazoxane, teniposide, amsacrine, etoposide, mitoxantrone and daunorubicin drugs have been approved by FDA and have had inhibiting effect on hub genes, therefore they could become potential new drugs in BC treatment.

The drug-genes network was drawn by Cytoscape software (Fig. 9A). Using the STITCH tool, we constructed downstream interaction networks of *TOP2A* and *RRM2* to investigate the additional effects caused by drugs of these genes. As shown in Fig. 9, *TOP2A* inhibition might have downstream effects on *TOP2B*, *SUMO1*, *SUMO2*, *PCNA*, *UBC* and *TOP1* genes (Fig. 9B); *RRM2* inhibition might have downstream effects on *RRM1*, *TYMS*, *CDK1*, *POLA1*, *PCNA*, *TFDP1* and *TFDP2* genes (Fig. 9C).

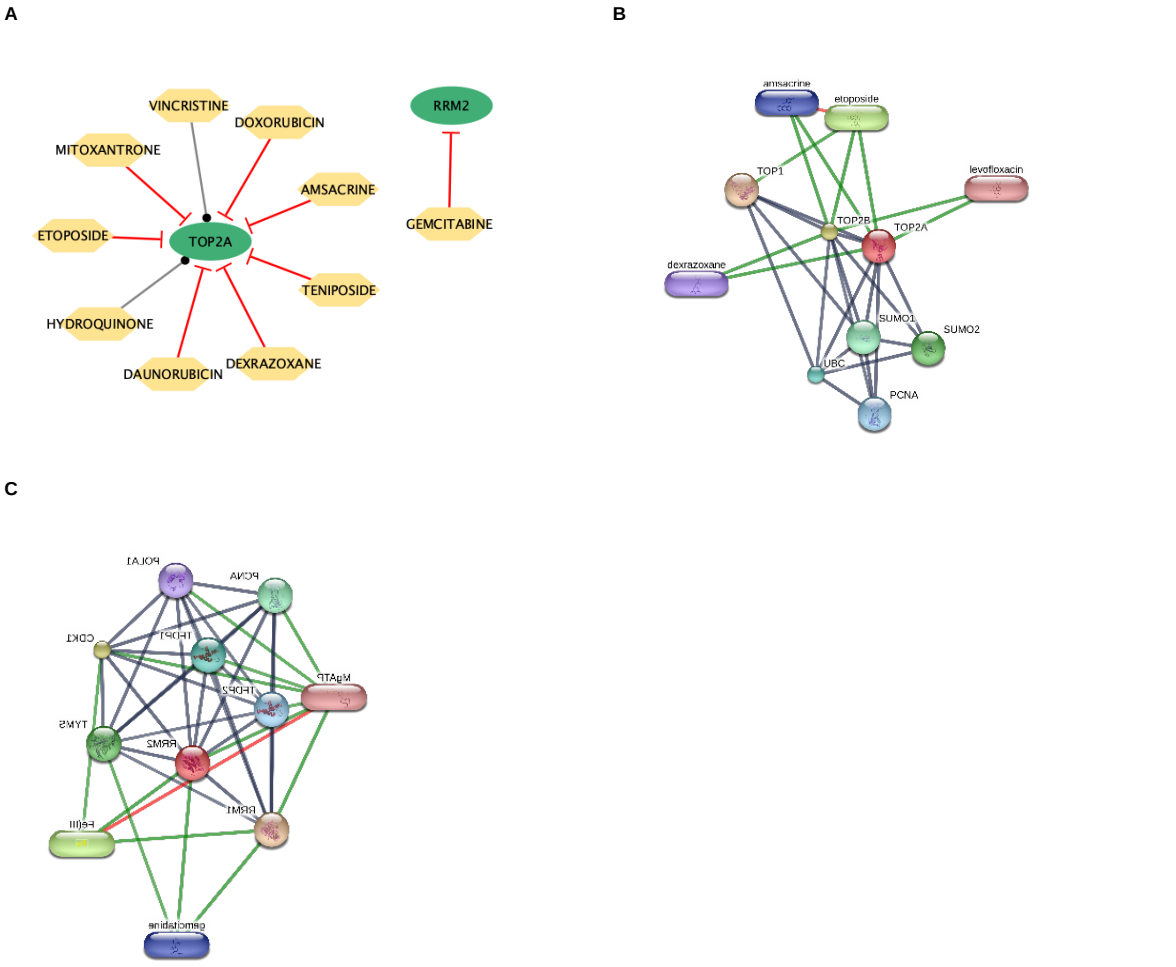


**Figure 9.** The interaction between drugs and related genes. (A) The association between the hub genes and related drugs was identified by DGIbd. Green ellipses indicate the hub genes, yellow hexagons show related drugs, T-red arrows indicate inhibitor interaction and elipse black arrows indicate unknown interaction. (B) Subnetwork of *TOP2A* performed by the STITCH online database contained 7 nodes, 18 edges and its *P-value* of PPI enrichment analysis was 0.00105. (C) Subnetwork of *RRM2* performed by the STITCH online database contained 8 nodes, 27 edges and its *P-value* of PPI enrichment analysis was 1.3E-07. For (B and C), protein-protein interactions are shown in grey, chemical-protein interactions in green and interactions between chemicals in red.

## CONCLUSION

In conclusion, a total of 1369 DEGs, including 400 upregulated DEGs and 969 downregulated DEGs in BC, have been screened through integrated bioinformatics analysis. Ten hub genes, namely, *CDK1*, *CCNA2*, *CCNB1*, *CCNB2*, *TOP2A*, *KIF11*, *RRM2*, *BUB1B*, *CDC20*, and *NCAPG*, could play crucial roles in the tumorigenesis and prognosis of BC. Additionally, the potential targeted drugs related to these genes are selected, of which dexrazoxane, teniposide, amsacrine, etoposide, mitoxantrone and daunorubicin could be new chemicals for BC treatment.

## REFERENCES

An X, Xu F, Luo R, Zheng Q, Lu J, Yang Y, Qin T, Yuan Z, Shi Y, Jiang W, Wang S (2018) The prognostic significance of topoisomerase II alpha protein in early stage luminal breast cancer. *BMC Cancer* 18: 331.

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A (2013) NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* 41: D991-995.

Deng J, Xu Y, Wang G (2019) Identification of Potential Crucial Genes and Key Pathways in Breast Cancer Using Bioinformatic Analysis. *Front Genet* 10: 695.

Dong M, Xu T, Cui X, Li H, Li X, Xia W (2021) NCAPG upregulation mediated by four microRNAs combined with activation of the p53 signaling pathway is a predictor of poor prognosis in patients with breast cancer. *Oncol Lett* 21: 323.

Esserman LJ, Shieh Y, Park JW, Ozanne EM (2007) A role for biomarkers in the screening and diagnosis of breast cancer in younger women. *Expert Rev Mol Diagn* 7: 533-544.

Fang L, Liu Q, Cui H, Zheng Y, Wu C (2022) Bioinformatics Analysis Highlight Differentially Expressed CCNB1 and PLK1 Genes as Potential Anti-Breast Cancer Drug Targets and Prognostic Markers. *Genes (Basel)* 13: 654.

Feng H, Gu Z, Li Q, Liu Q, Yang X, Zhang J (2019) Identification of significant genes with poor prognosis in ovarian cancer via bioinformatical analysis. *J Ovarian Res* 12: 35.

Jayanthi VSPKSA, Das AB, Saxena U (2020) Grade-specific diagnostic and prognostic biomarkers in breast cancer. *Genomics* 112: 388-396.

Koyuncu D, Sharma U, Goka ET, Lippman ME (2021) Spindle assembly checkpoint gene BUB1B is essential in breast cancer cell survival. *Breast Cancer Res Treat* 185: 331-341.

Li J, Guan X, Fan Z, Ching L, Li Y, Wang X, Cao W, Liu D (2020) Non-Invasive Biomarkers for Early Detection of Breast Cancer. *Cancers (Basel)* 12: 2767.

Moradpoor R, Zali H, Gharebaghian A, Akbari ME, Ajdari S, Salimi M (2021) Identification of CCNB2 as A Potential Non-Invasive Breast Cancer Biomarker in Peripheral Blood Mononuclear Cells Using The Systems Biology Approach. *Cell J* 23: 406-413.

Prihantono, Faruk M (2021) Breast cancer resistance to chemotherapy: When should we suspect it and how can we prevent it?. *Ann Med Surg (Lond)* 70: 102793.

Qiu P, Guo Q, Yao Q, Chen J, Lin J (2021) Hsa-mir-3163 and CCNB1 may be potential biomarkers and therapeutic targets for androgen receptor positive triple-negative breast cancer. *PloS One* 16: e0254283.

Ren Y, Deng R, Zhang Q, Li J, Han B, Ye P (2020) Bioinformatics analysis of key genes in triple negative breast cancer and validation of oncogene PLK1. *Ann Transl Med* 8: 1637.

Rody A, Karn T, Ruckhäberle E, Müller V, Gehrmann M, Solbach C, Ahr A, Gätje R, Holtrich U, Kaufmann M (2009) Gene expression of topoisomerase II alpha (TOP2A) by microarray analysis is highly prognostic in estrogen receptor (ER) positive breast cancer. *Breast Cancer Res Treat* 113: 457-466.

Wang B, Yu J, Sun Z, Luh F, Lin D, Shen Y, Wang T, Zhang Q, Liu X (2020) Kinesin family member 11 is a potential therapeutic target and is suppressed by microRNA-30a in breast cancer. *Mol Carcinog* 59: 908-922.

Wang N, Zhang H, Li D, Jiang C, Zhao H, Teng Y (2021) Identification of novel biomarkers in breast cancer via integrated bioinformatics analysis and experimental validation. *Bioengineered* 12: 12431-12446.

Weng Y, Liang W, Ji Y, Li Z, Jia R, Liang Y, Ning P, Xu Y (2021) Key Genes and Prognostic Analysis in HER2+ Breast Cancer. *Technol Cancer Res Treat* 20: 1-14.

Xing Z, Wang X, Liu J, Zhang M, Feng K, Wang X (2021) Expression and prognostic value of CDK1, CCNA2, and CCNB1 gene clusters in human breast cancer. *J Int Med Res* 49: 1-13.

Xu Z, Zhou Y, Cao Y, Dinh TLA, Wan J, Zhao M (2016) Identification of candidate biomarkers and analysis of prognostic values in ovarian cancer by integrated bioinformatics analysis. *Med Oncol* 33: 130.

Yang D, He Y, Wu B, Deng Y, Wang N, Li M, Liu Y (2020) Integrated bioinformatics analysis for the screening of hub genes and therapeutic drugs in ovarian cancer. *J Ovarian Res* 13: 10.

Zaczek AJ, Markiewicz A, Seroczynska B, Skokowski J, Jaskiewicz J, Pienkowski T, Olszewski WP, Szade J, Rhone P, Welnicka-Jaskiewicz M, Jassem J (2012) Prognostic significance of TOP2A gene dosage in HER-2-negative breast cancer. *Oncologist* 17: 1246-1255.

Zeng X, Shi G, He Q, Zhu P (2021) Screening and predicted value of potential biomarkers for breast cancer using bioinformatics analysis. *Sci Rep* 11: 20799.

Zubair M, Wang S, Ali N (2020) Advanced Approaches to Breast Cancer Classification and Diagnosis. *Front Pharmacol* 11: 632079.