



## OPEN ACCESS

## EDITED BY

Fu Wang,  
Xi'an Jiaotong University, China

## REVIEWED BY

ChuanGui Song,  
Fujian Medical University Union  
Hospital, China  
Xinglong Fan,  
Qilu Hospital of Shandong University,  
China

## \*CORRESPONDENCE

Xiaolei Hu,  
xlhu@smu.edu.cn

<sup>†</sup>These authors have contributed equally  
to this work

## SPECIALTY SECTION

This article was submitted to Cancer  
Genetics and Oncogenomics,  
a section of the journal  
Frontiers in Genetics

RECEIVED 08 July 2022

ACCEPTED 16 September 2022

PUBLISHED 12 October 2022

## CITATION

Wang S, Shang P, Yao G, Ye C, Chen L  
and Hu X (2022), A genomic and  
transcriptomic study toward  
breast cancer.  
*Front. Genet.* 13:989565.  
doi: 10.3389/fgene.2022.989565

## COPYRIGHT

© 2022 Wang, Shang, Yao, Ye, Chen and  
Hu. This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# A genomic and transcriptomic study toward breast cancer

Shan Wang<sup>1,2†</sup>, Pei Shang<sup>1†</sup>, Guangyu Yao<sup>1</sup>, Changsheng Ye<sup>1</sup>,  
Lujia Chen<sup>1</sup> and Xiaolei Hu<sup>1\*</sup>

<sup>1</sup>Department of Breast Surgery, Nanfang Hospital, Southern Medical University, Guangzhou, China,  
<sup>2</sup>Department of Critical Care Medicine, Nanfang Hospital, Southern Medical University, Guangzhou,  
China

**Background:** Breast carcinoma is well recognized to be having the highest global occurrence rate among all cancers, being the leading cause of cancer mortality in females. The aim of this study was to elucidate breast cancer at the genomic and transcriptomic levels in different subtypes so that we can develop more personalized treatments and precision medicine to obtain better outcomes.

**Method:** In this study, an expression profiling dataset downloaded from the Gene Expression Omnibus database, GSE45827, was re-analyzed to compare the expression profiles of breast cancer samples in the different subtypes. Using the GEO2R tool, different expression genes were identified. Using the STRING online tool, the protein–protein interaction networks were conducted. Using the Cytoscape software, we found modules, seed genes, and hub genes and performed pathway enrichment analysis. The Kaplan–Meier plotter was used to analyze the overall survival. MicroRNAs and transcription factors targeted different expression genes and were predicted by the Enrichr web server.

**Result:** The analysis of these elements implied that the carcinogenesis and development of triple-negative breast cancer were the most important and complicated in breast carcinoma, occupying the most different expression genes, modules, seed genes, hub genes, and the most complex protein–protein interaction network and signal pathway. In addition, the luminal A subtype might occur in a completely different way from the other three subtypes as the pathways enriched in the luminal A subtype did not overlap with the others. We identified 16 hub genes that were related to good prognosis in triple-negative breast cancer. Moreover, *SRSF1* was negatively correlated with overall survival in the Her2 subtype, while in the luminal A subtype, it showed the opposite relationship. Also, in the luminal B subtype, *CCNB1* and *KIF23* were associated with poor prognosis. Furthermore, new transcription factors and microRNAs were introduced to breast cancer which would shed light upon breast cancer in a new way and provide a novel therapeutic strategy.

**Conclusion:** We preliminarily delved into the potentially comprehensive molecular mechanisms of breast cancer by creating a holistic view at the genomic and transcriptomic levels in different subtypes using computational tools. We also introduced new prognosis-related genes and novel therapeutic strategies and cast new light upon breast cancer.

## KEYWORDS

breast cancer, protein–protein interaction, signal pathway, microarray, survival

## Introduction

Breast carcinoma is well recognized to be having the highest global occurrence rate among all types of cancers, being the leading cause of cancer mortality in females worldwide (Ferlay et al., 2021). In the United States, it is estimated that approximately 281550 new female cases were diagnosed in 2021, and it accounted for 15% of estimated deaths due to cancer among women (Siegel et al., 2021). It is well known that breast cancer, which harbors high biological heterogeneity both between and within tumors, is not a single disease and can be classified into four subtypes according to the molecular types, such as luminal A, luminal B, Her2-overexpressed, and triple-negative breast cancer (TNBC) (Perou et al., 2000). Luminal A and luminal B subtypes express the hormone receptors and have a better prognosis than the other two subtypes (Harbeck et al., 2019). The Her2-overexpressed subtype only has Her2 expression and lacks expression of the estrogen receptor (ER) and progesterone receptor (PR), and this subtype has achieved tremendous clinical success because of effective therapy targeting Her2 (Cancer Genome Atlas Network, 2012; Harbeck et al., 2019). TNBC is characterized by the absence of ER, PR, and Her2 expression, which possesses distinct molecular traits and unique recurrence and metastatic patterns (Sørlie et al., 2001; Nielsen et al., 2004; Harbeck et al., 2019).

Currently, the clinical approach to treating breast cancer has been mainly composed of surgery, radiotherapy, chemotherapy, endocrine treatment, and targeted therapy (Harbeck et al., 2019). Although the treatment has been relatively perfect, the reduction of decline in the death rate for breast cancer slowed in females over the past decade, which suggests that we should elucidate the pathogenesis, occurrence, and development of cancer more accurately and find new potential prognostic biomarkers so that we can ensure early diagnosis and develop more personalized treatments and precision medicine to obtain better outcomes (Ferlay et al., 2021). For this to be possible, we think that it is sensible to get a holistic view of the mechanism of breast cancer with system biology approaches. By analyzing high-throughput data extracted from omics data, these approaches present an opportunity to depict the behavior of networks and offer novel therapeutics.

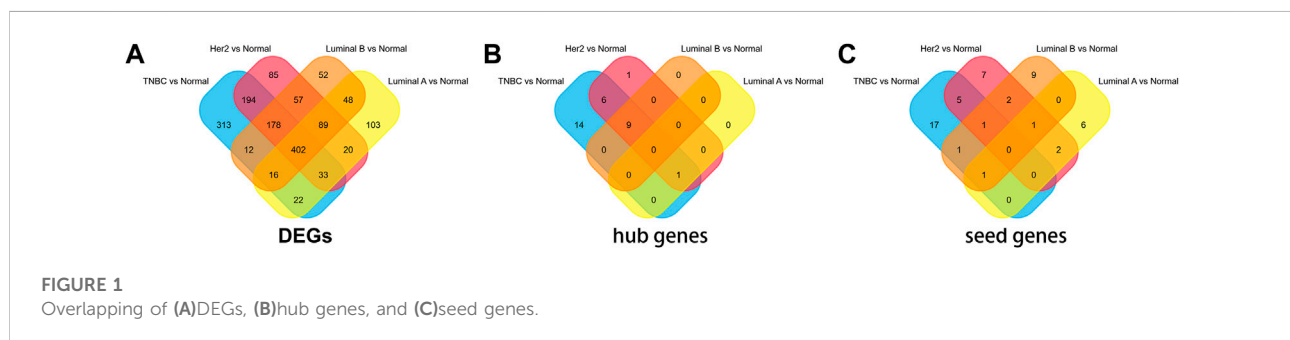
Previous studies have mostly analyzed the molecular mechanisms by comparing the difference between tumor and normal tissues of the breast or focused only on one subtype of breast cancer (Yang et al., 2019; Lin et al., 2020; Liu S. et al., 2020). Actually, during clinical treatment, different measures will be performed according to their subtype, so it is inappropriate to consider different subtypes as a whole to analyze. Also, focusing only on a single subtype cannot help us identify the difference and similarities among different subtypes. In addition, most studies are limited to exploring biomarkers and do not combine the genome with the transcriptome for further exploration.

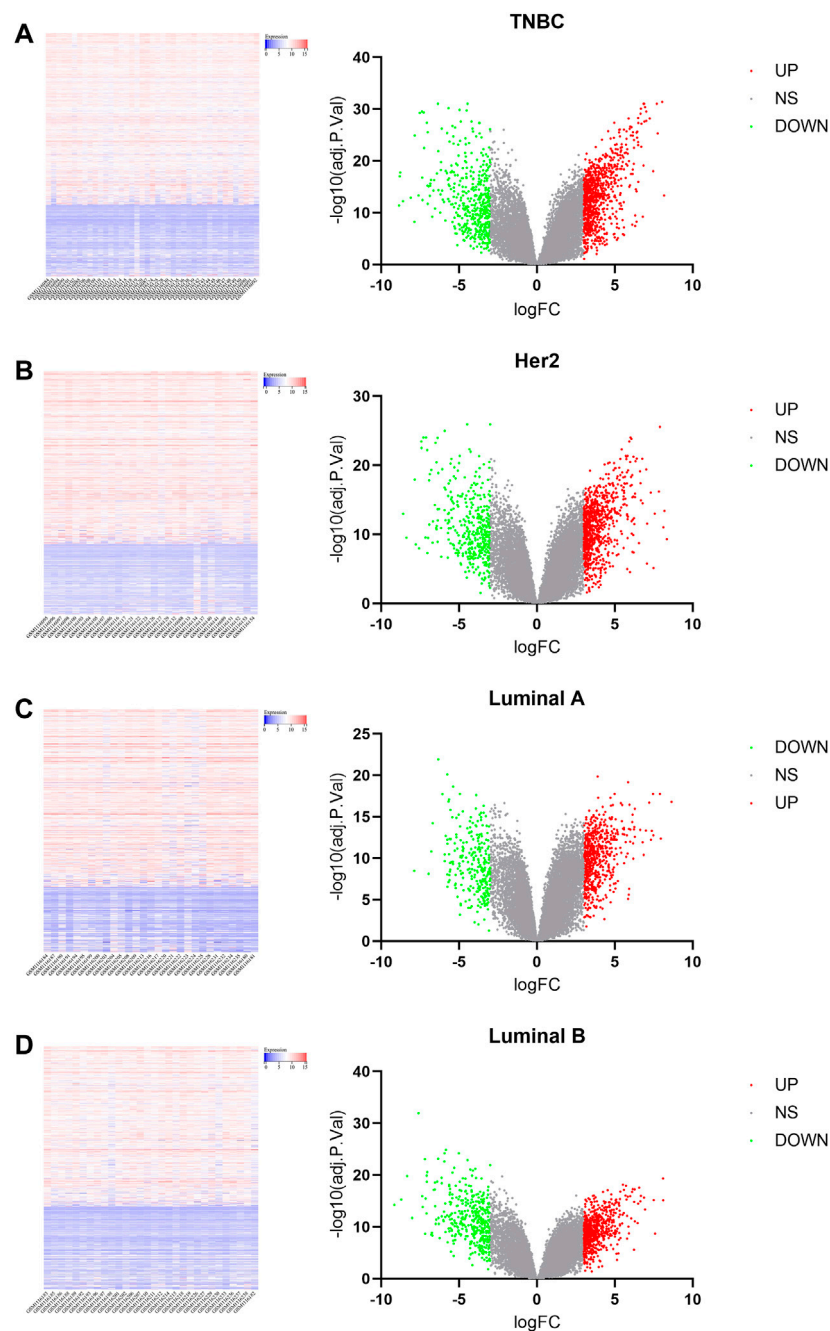
In this study, we preliminarily delved into the potentially comprehensive molecular mechanisms of breast cancer by creating a holistic view at the genomic and transcriptomic level in four different subtypes using computational tools. To the best of our knowledge, this is the first time such a systematic biological study was performed on breast cancer according to its subtypes and at the genomic and transcriptomic level. We also first explored genes such as *SRSF1*, *BUB1B*, *KIF23*, *HNRNPF*, and *ELAVL1* and obtained an exact result in our study. We re-analyzed the dataset deposited by Gruosso et al. (2016) and exhibited considerable protein–protein interaction networks. In addition, we performed network, clustering, and functional analysis so that we could have a deep understanding of the central genes of each subtype. Otherwise, pathways of different subtypes were identified with enrichment analysis, and new micro-RNAs (miRNAs) and transcription factors (TFs) were introduced to assay the regulatory mechanisms of differential expression genes (DEGs).

## Materials and methods

### Microarray data and DEG screening

A microarray dataset with accession number “GSE45827” from the GEO database was downloaded (Gruosso et al., 2016). This dataset includes 14 cell line samples, 41 TNBC cancer samples, 30 Her2 cancer samples, 29 luminal A cancer samples, 30 luminal B cancer samples, and 11 normal breast tissue samples, and we only used the cancer samples and normal breast tissue samples to analyze. GEO2R (RRID:SCR\_016569, <http://www.ncbi.nlm.nih.gov/geo/>

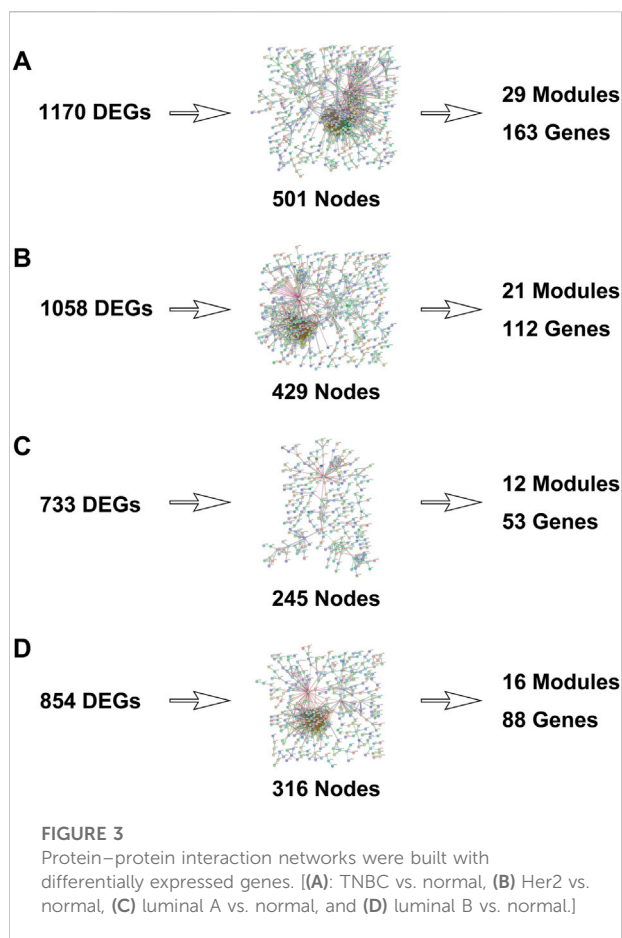


**FIGURE 2**

Heat map and volcano plot analysis of DEGs. In the volcano plot, blue dots on the left indicate the downregulated genes, gray dots in the middle indicate genes that are not differentially expressed, and red dots in the right indicate the upregulated genes. [(A): TNBC vs. normal, (B) Her2 vs. normal, (C) luminal A vs. normal, and (D) luminal B vs. normal.]

[geo2r/](#)) is an online tool that can be used to screen DEGs across different groups. Using GEO2R of GEO, groups (TNBC vs. normal, Her2 vs. normal, luminal A vs. normal, and luminal B vs. normal) were compared to identify DEGs of the four subtypes.

Benjamini–Hochberg false discovery was used for  $p$ -value adjustment. Genes were declared as DEGs when  $|\logFC| \geq 3$  and the adjusted  $p$ -value (adj.p)  $< 0.01$ . The heat map was performed by SangerBox online tool version 3.0 (<http://www.sangerbox.com/>)



tool), and the volcano plot was drawn by GraphPad Prism for Windows (v9.2.0, RRID:SCR\_002798, GraphPad Software, San Diego, California United States, [www.graphpad.com](http://www.graphpad.com)).

## PPI network construction

The PPI networks of DEGs were built with the STRING online tool (v11.0, RRID:SCR\_005223, <https://string-db.org/>) (Szkarczyk et al., 2019). DEGs were mapped to the STRING database to estimate the interactive relationships, setting the confidence cutoff to 0.95. Then, Cytoscape (v3.8.2, RRID:SCR\_003032) software was used to visualize the PPI network. For network analysis, the MCODE plugin (v2.0.0, RRID:SCR\_015828) of Cytoscape software was used to investigate modules, highly connected sub-networks, and seed genes based on default settings (Bader and Hogue, 2003; Shannon et al., 2003). CytoHubba plugin version 0.1 of Cytoscape (v3.8.2, RRID:SCR\_003032) was applied to detect hub genes (Chin et al., 2014). The criteria of hub genes were as follows: MCC cutoff = 1000, degree cutoff = 10, closeness cutoff = 50, and betweenness cutoff = 1000. In addition, Venn diagrams were drawn by FunRich software (v3.1.3, RRID:SCR\_014467).

## Pathway enrichment analysis

Genes clustered with MCODE were analyzed by the Cytoscape ClueGO plugin (v2.5.8, RRID:SCR\_005748), choosing Reactome and KEGG databases to retrieve pathways (Kanehisa and Goto, 2000; Bindea et al., 2009; Jassal et al., 2020). Bonferroni step down was used to adjust the  $p$ -value, and signal pathways with  $\text{adj.}p \leq 0.05$  were recognized.

## Survival analysis

The Kaplan–Meier plotter mRNA breast cancer database (RRID:SCR\_018753, <https://kmplot.com/analysis/>), an online database, was used to analyze the overall survival (OS) with hazard ratios (HRs), 95% confidence intervals (95% CIs), and logrank  $p$ -value. The JetSet best probe set was selected as gene probes. During the prognosis analysis, patients were split into two groups in accord with the auto-select best cutoff. The logrank  $p$ -value  $< 0.05$  was considered to show a statistical significance. The forest plot was drawn using Xiantao scholar (<https://www.xiantao.love/>), another online platform for data analysis.

## Expression analysis of prognosis-related hub genes

The differential expression of prognosis-related hub genes was analyzed in the GSE45827 dataset and validated in TCGA dataset using the ggplot2 package of R software. During the analysis, the Mann–Whitney U test, Welch’s  $t$ -test, and Student’s  $t$ -test were used, respectively, depending on the normality and homogeneity of variance. Similarly the  $p$ -value  $< 0.05$  was considered to show a statistical significance.

## Functional exploration of each prognosis-related hub gene

GeneMANIA (<http://www.genemania.org>) was used to evaluate the functions of prognosis-related hub genes according to several bioinformatics methods, such as co-expression, physical interaction, prediction, co-localization, and shared protein domains and pathways (Warde-Farley et al., 2010).

## miRNA and TF enrichment analysis

The microRNAs and TFs were predicted using the Enrichr online server (RRID:SCR\_001575) (Kuleshov et al., 2016). MiRNAs were predicted by the TargetScan microRNA 2017 database, while TFs were predicted by the ChEA2016 database.  $\text{Adj.}p \leq 0.01$  was considered to show

statistical significance. The miRNAs with higher combined scores were selected.

## Results

### By analyzing the GEO database, DEGs were identified

The microarray dataset “GSE45827” which includes primary invasive breast carcinoma (41 TNBC, 30 Her2, 29 luminal A, and 30 luminal B) and 11 normal tissues has been analyzed. Using the GEO2R tool, we found 1170, 1058, 733, and 854 DEGs which are significantly variably expressed between TNBC vs. normal, Her2 vs. normal, luminal A vs. normal, and luminal B vs. normal, respectively. Most of the DEGs overlapped between the four molecular subtypes (Figure 1A). The heat map and volcano plot are shown in Figure 2.

### Protein–protein interaction networks were constructed

The PPI networks with DEGs were conducted using the STRING online tool. The edges indicate both functional and physical protein associations. The TNBC subtype has the most nodes and edges. A total of 501, 429, 245, and 316 nodes (genes) are in PPI networks TNBC vs. normal, Her2 vs. normal, luminal A vs. normal, and luminal B vs. normal, respectively (Figures 3A–D). The topological clusters also called modules found in MCODE identified groups of genes with a similar function, and each module has the most effective genes, called seed genes. Similarly, the TNBC subtype has the most modules and seed genes. Interestingly, these sets of seed genes in different subtypes exhibited few overlaps (Figure 1C). Network topology was measured based on the graph theory concepts such as MCC, degree, closeness, and betweenness. Seed genes such as *CDC6* and *RFC3* were hub genes in the TNBC. In the Her2 subtype, *AURKB* was identified as both a seed gene and a hub gene. Only *SRSF1* which coincided with the TNBC and Her2 subtype was introduced as a hub gene in the luminal A subtype. All the hub genes in the luminal B subtype overlapped with the TNBC and Her2 subtype (Figure 1B). The hub genes are represented in Table 1.

### Pathway enrichment analysis was performed

The pathway enrichment analysis was executed based on genes identified by MCODE. We reached 51, 25, 10, and 15 pathways by performing the pathway enrichment analysis from 163, 112, 53, and 88 genes, respectively (Figure 4). The genes used to analyze pathways were those that were included in

PPI modules. The pathways involved in TNBC were mainly about DNA replication, DNA repair, and mitosis, while in the Her2 and luminal B types, they were mitosis, and most of the pathways involved in Her2 and luminal B subtypes were included in the TNBC subtype. In addition, the pathways that play a role in luminal A were totally different from the other three subtypes, especially associated with extracellular matrix organization and collagen formation. In the TNBC subtype, the top three pathways that contain the most genes are a condensation of prometaphase chromosomes, Chk1/Chk2(Cds1)-mediated inactivation of cyclin, and activation of ATR in response to replication stress. In the Her2 subtype, condensation of prometaphase chromosomes, resolution of sister chromatid cohesion, and amplification of signals from the kinetochores are the top three pathways. Syndecan interactions, MET-activated PTK2 signaling, and MET-promoted cell motility are the top three pathways in the luminal A subtype. In the luminal B subtype, the top three pathways are an amplification of the signal from the kinetochores, amplification of the signal from unattached kinetochores via a MAD2 inhibitory signal, and resolution of sister chromatid cohesion. Furthermore, there were also some pathways that are unique to specific subtypes, for example, the ERBB4 pathway in TNBC and the NOTCH4 pathway in luminal B.

### Survival analysis of hub genes in different subtypes was carried out

We then considered whether the hub genes in the different subtypes of breast cancer were associated with prognosis. The relationship between hub gene expression and survival rates was evaluated using the Kaplan Meier plotter. The prognostic analysis demonstrated that hub genes such as *CDC6*, *NDC80*, *BUB1B*, *FOXM1*, *NUF2*, *MCM4*, *CDC20*, *BUB1*, *MCM2*, *CCNB2*, *ASPM*, *PRC1*, *PLK1*, *HNRNPF*, *CCNA2*, and *KIF2C* were related to good prognosis in TNBC (Figure 5A). In addition, *SRSF1* was negatively correlated with overall survival (OS) in the Her2 subtype, while in the luminal A subtype, it showed the opposite relationship (Figures 5B,C). In the luminal B subtype, *CCNB1* and *KIF23* were associated with poor prognosis (Figure 5D).

### The expression of prognosis-related hub genes in each subtype was analyzed in the GSE45827 dataset and validated in another independent dataset

The expression of hub genes that were related to prognosis in GSE45827 was analyzed using ggplot2 of R software. All the hub genes that were analyzed in TNBC were upregulated, including *ASPM*, *BUB1*, *BUB1B*, *CCNA2*, *CCNB2*, *CDC6*, *CDC20*, *FOXM1*, *HNRNPF*, *KIF2C*, *MCM2*, *MCM4*, *NDC80*, *NUF2*, *PLK1*, and

TABLE 1 Hub genes in the PPI network.

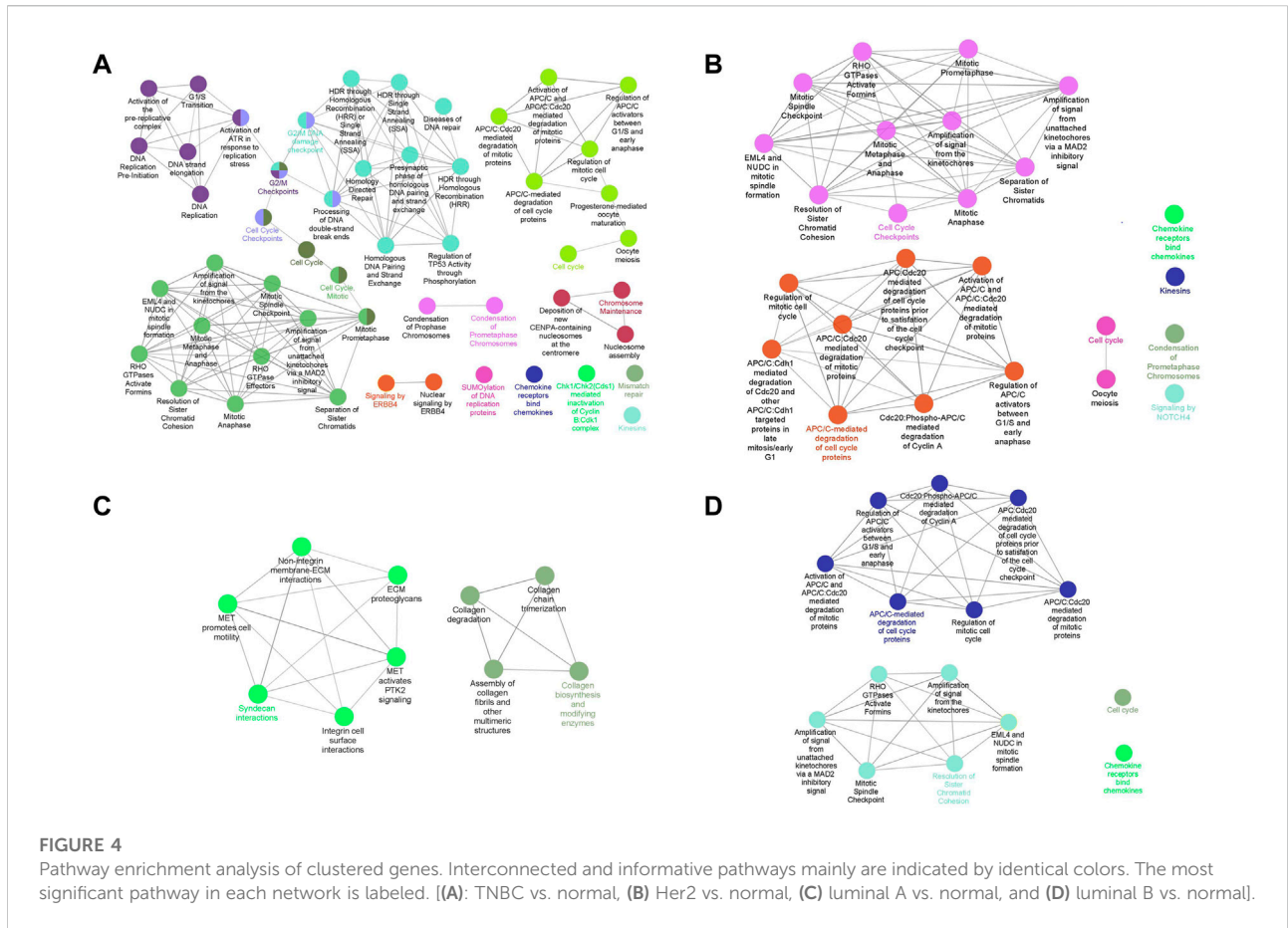
ID	MCC	Degree	Closeness	Betweenness	ID	MCC	Degree	Closeness	Betweenness
TNBC vs. normal					Her2 vs. normal				
<i>CDK1</i>	9.22E+13	89	192.19008	22078.67341	<i>NDC80</i>	9.22E+13	40	140.9357	1013.13909
<i>CDC5L</i>	8106206	77	192.6067	69392.04571	<i>KIF2C</i>	9.22E+13	34	139.8357	1240.99518
<i>CCNB1</i>	9.22E+13	64	178.27341	7504.97904	<i>KIF11</i>	9.22E+13	43	145.0857	2836.58396
<i>CDC20</i>	9.22E+13	62	165.25675	3752.95365	<i>CDK1</i>	9.22E+13	73	163.2691	20555.50241
<i>BUB1</i>	9.22E+13	57	172.85675	3211.85824	<i>CDC20</i>	9.22E+13	52	138.8952	2567.19563
<i>PLK1</i>	9.22E+13	57	171.69008	4346.90413	<i>CCNB2</i>	9.22E+13	47	136.4191	1209.68801
<i>CCNB2</i>	9.22E+13	56	160.54008	2061.86361	<i>CCNB1</i>	9.22E+13	50	149.0857	3631.78964
<i>CCNA2</i>	9.22E+13	55	161.05675	4757.15522	<i>CCNA2</i>	9.22E+13	44	134.2286	2002.89489
<i>BUB1B</i>	9.22E+13	52	168.52341	5703.00881	<i>BUB1B</i>	9.22E+13	45	145.3357	4540.65528
<i>AURKB</i>	9.22E+13	51	168.02341	2567.01982	<i>BUB1</i>	9.22E+13	49	148.6691	3185.181
<i>KIF11</i>	9.22E+13	48	167.77341	2555.27954	<i>AURKB<sup>a</sup></i>	9.22E+13	46	145.8357	3137.25928
<i>NDC80</i>	9.22E+13	47	165.22341	1237.29857	<i>KIF23</i>	5.50E+12	27	131.3333	9398.7775
<i>KIF2C</i>	9.22E+13	42	163.02341	1465.60477	<i>UBE2C</i>	1.07E+12	31	123.4036	2928.73296
<i>NUF2</i>	9.22E+13	35	155.40317	1007.52523	<i>CDC5L</i>	822688	65	163.5071	53316.02571
<i>ASPM</i>	9.22E+13	34	145.35913	1083.30296	<i>RACGAP1</i>	8.86E+07	22	141.2333	18065.07384
<i>KIF23</i>	9.22E+13	31	149.33095	11155.80535	<i>ECT2</i>	41048	11	123	4077.48928
<i>CDC6<sup>a</sup></i>	47892	27	140.84127	1232.15458	<i>SRSF1</i>	1082	12	112.3119	2071.32644
<i>MCM2</i>	8932	26	141.43889	3024.46868	Luminal A vs. normal				
<i>MCM4</i>	9062	25	141.10556	1725.51468	<i>SRSF1</i>	1034	11	51.00238	1044.8
<i>PRC1</i>	3.82E+11	24	141.62341	1039.07441	Luminal B vs. normal				
<i>RACGAP1</i>	1.25E+10	24	159.54762	17261.72834	<i>CDK1</i>	1.57E+11	57	110.2952	6869.988
<i>PCNA</i>	1845	23	139.58889	2598.0432	<i>CDC20</i>	1.57E+11	45	96.84524	1313.656
<i>CHEK1</i>	6622	20	148.45317	8464.48184	<i>BUB1</i>	1.57E+11	42	102.3786	1944.034
<i>RPA1</i>	2406	18	138.29603	2279.10647	<i>BUB1B</i>	1.27E+11	38	99.2119	1242.749
<i>RFC3<sup>a</sup></i>	1648	15	146.13651	1025.54515	<i>CCNB1</i>	7.79E+10	40	100.7119	1303.47
<i>SRSF1</i>	1826	15	131.52937	1327.95608	<i>RACGAP1</i>	987987	19	94.56667	4407.43
<i>FOXM1</i>	6673	13	132.69881	3099.92655	<i>KIF11</i>	1.56E+11	37	98.87857	1629.572
<i>ECT2</i>	367926	11	138.58095	4860.54991	<i>KIF23</i>	2.62E+09	23	90.05	3930.105
<i>HNRNPf</i>	1729	11	128.79603	2045.45414	<i>CDC5L</i>	26009	47	109.5952	21827.75
<i>ELAVL1</i>	1693	11	128.62937	1102.88098					

<sup>a</sup>Seed genes.

*PRC1* (Figure 6A–P). *SRSF1* was a hub gene that was related to prognosis in both Her2 and Luminal A subtypes, and its expression profiles in the two subtypes were similar—it had a higher expression level in the normal tissues than in the tumor tissue (Figure 6Q–6R). The two hub genes that were associated with prognosis in the luminal B subtype were also upregulated (Figure 6S–6T). Then, we validated the expression of prognosis-related hub genes in TCGA dataset. The expression of *SRSF1* in Her2 and luminal A subtypes is not consistent with the result of the GEO profile analysis, and it was highly expressed in tumor tissues rather than normal tissues (Figure 7Q–7R). The expression levels of the rest hub genes were in accordance with the GEO profile analysis (Figure 7A–P, Figure 7S,T).

## Potential functions for each prognosis-related hub gene were explored

We then investigated the functions of the prognosis-related hub genes using GeneMANIA. It showed that these genes were correlated with mitotic nuclear division (FDR=2.13e-33), chromosome segregation (FDR=1.68e-24), microtubule cytoskeleton organization involved in mitosis (FDR=8.79e-20), spindle (FDR=8.94e-18), mitotic cell cycle checkpoint (FDR=9.00e-17), negative regulation of the mitotic cell cycle (FDR=3.42e-14), and metaphase/anaphase transition of the mitotic cell cycle (FDR=3.00e-17) (Figure 8).



**FIGURE 4** Pathway enrichment analysis of clustered genes. Interconnected and informative pathways mainly are indicated by identical colors. The most significant pathway in each network is labeled. [(A): TNBC vs. normal, (B) Her2 vs. normal, (C) luminal A vs. normal, and (D) luminal B vs. normal].

### miRNAs and TFs enriched with DEGs were determined

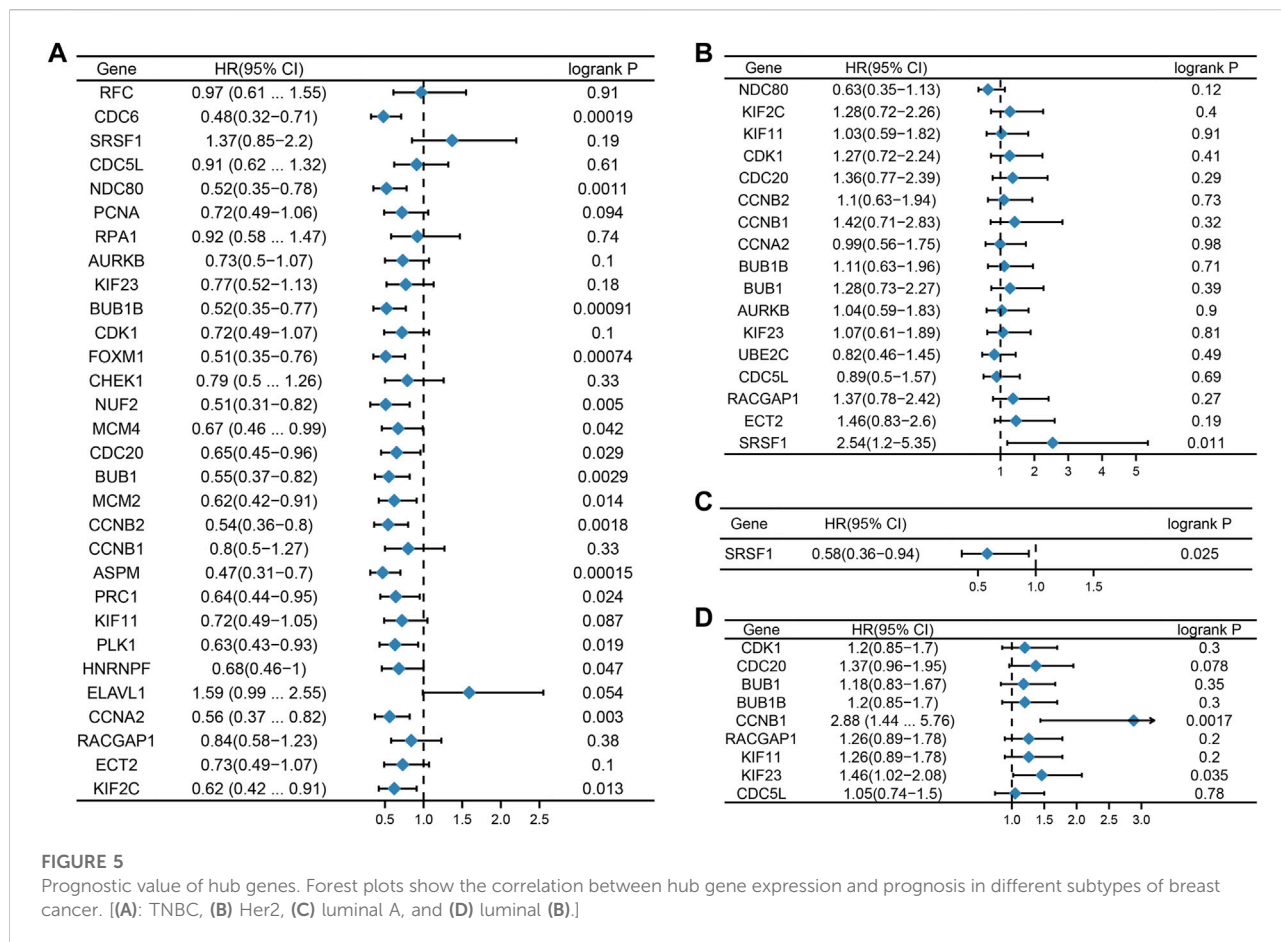
The miRNAs and TFs, as important regulators of DEGs, were predicted using the Enrichr web server. It is worth noting that the luminal A subtype had the most TFs, while it had the least hub genes among the four molecular subtypes. Seven TFs exerted their function in all four subtypes (Figure 9). TFs that were meaningful in breast cancer development are shown in Table 2. The top 10 miRNAs enriched with DEGs in each subtype are also shown (Figure 10).

### Discussion

In the present study, bioinformatic approaches were carried out to show the DEGs, modules, seed genes, PPI, and hub genes in each subtype (TNBC, Her2, luminal A, and luminal B). The topological clusters which have high-density regions in the network, also called modules, find in MCODE-identified groups of genes with a similar function. Genes in the highly interconnected subnetwork modules are expected to be involved

in the same pathways or in roles with related biological functions. Each module has a most effective gene which has high centrality, named seed genes. The nodes in the PPI network represent genes, while the edges indicate both functional and physical protein associations. Nodes with high degree, betweenness, closeness, and MCC are significant for the network and are called hub genes which can serve as targets. The analysis of these elements implied that the carcinogenesis and development of TNBC were the most important and complicated processes in breast carcinoma, occupying the most DEGs, modules, seed genes, hub genes, and the most complex PPI network.

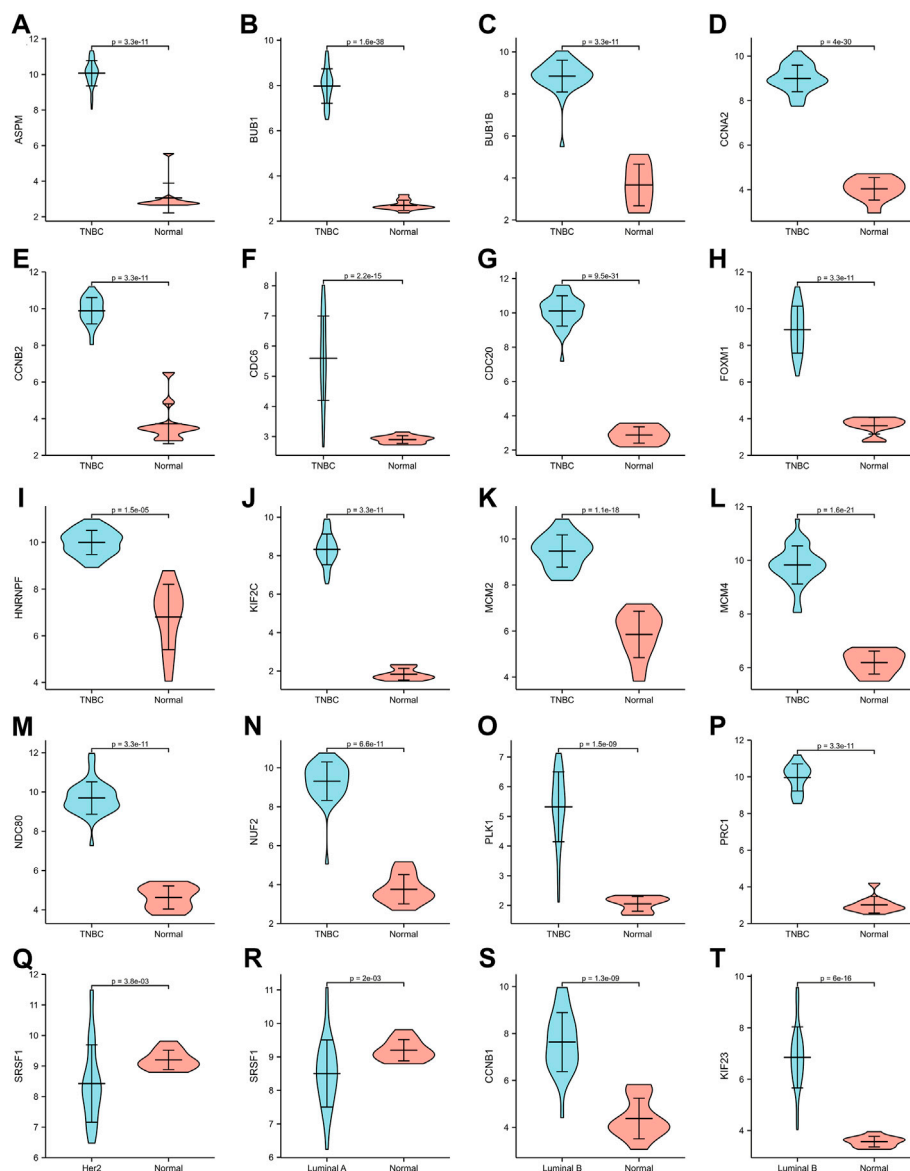
The role of some of those hub genes that we identified in our study has been verified in breast cancer, such as UBE2C whose overexpression plays a critical role in the incidence and development of breast cancer, and such a therapeutic strategy that combines palbociclib with tamoxifen might be promising in patients with HR+/HER2-breast cancer overexpressing UBE2C (Mo et al., 2017; Kim et al., 2019). Otherwise, it is reported that MCM2 and MCM4, which have a higher expression in high histological grade breast cancer, may be used as useful parameters to distinguish luminal A and luminal B subtypes instead of ki-67 and are related to poor prognosis (Issac et al.,



2019), which is partly consistent with our results in the TNBC subtype. Also, high expression of *RACGAP1* is supposed to be not only a strong poor prognostic marker in luminal-like breast cancer but might also be a predictor of response to treatment with tamoxifen and adjuvant chemotherapy (Milde-Langosch et al., 2013). In addition, the expression levels of *CDK1* and *CCNA2* have been previously revealed to be considerably higher in breast cancer tissues than those in normal tissues, and these genes lead to breast cancer development and are related to poor prognosis (Xing et al., 2021); however, compared with their study, we found that *CDK1* had an independent association with prognosis, and the discrepancy may be due to different analysis methods, as we analyzed according to the subtype while they did not. Others have shown that overexpression of *CDC20* indicates unfavorable prognosis and poor response to endocrine therapy in ER + breast cancer (Alfarsi et al., 2019; Tang et al., 2019); in contrast, we discovered that *CDC20* was related to worse prognosis only in the TNBC subtype, and we think that the different datasets that we analyzed result in the inconsistency. Other genes such as *BUB1*, *NUF2*, *CDC20*, *ASPM*, *KIF2C*, and *PRC1* have biological relevance to breast cancer progression, and *PLK1*, *NDC80*, and *CCNB2* only to TNBC progression, and these genes predict worse prognosis

(Wang et al., 2015; Tang et al., 2019; Yang et al., 2019; Lv et al., 2020; Ren et al., 2020; Chen et al., 2021; Jiang et al., 2021; Koyuncu et al., 2021); likewise, we also found that these genes correlated negatively with prognosis in the TNBC subtype. Moreover, it is also reported that high expression levels of *AURKB*, *CDC6*, and *ECT2* suggest a poor prognosis for breast cancer (Mahadevappa et al., 2017; Daulat et al., 2019; Huang et al., 2019; Xiu et al., 2019), whereas we only found that *CDC6* was negatively positively related to overall survival, and we consider that it is also the different datasets and analysis methods that cause the disparity. Furthermore, it is previously revealed that it might be an efficient therapeutic method to target *FOXMI* to impede advanced relapse and treat endocrine resistance (Roßwag et al., 2021). In addition, genes such as *SRSF1*, *BUB1B*, *KIF23*, and *HNRNPF* also play an important role in breast cancer, but so far, there is few reports on the association between these genes and the treatment or prognosis of breast cancer in the previous studies (Tyson-Capper and Gautrey, 2018; Du et al., 2021; Jian et al., 2021; Koyuncu et al., 2021); inspiringly, we performed it and obtained an exact result in our study. Also, we newly introduced applicant genes such as *RPA1* and *CDC5L* in TNBC, but their mechanism remains to be discovered with exploratory studies.





**FIGURE 6**

Expression of hub genes in GSE45827. (A–P) *ASPM*, *BUB1*, *BUB1B*, *CCNA2*, *CCNB2*, *CDC6*, *CDC20*, *FOXM1*, *HNRNP*, *KIF2C*, *MCM2*, *MCM4*, *NDC80*, *NUF2*, *PLK1*, *PRC1* expression in TNBC subtype. (Q) *SRSF1* expression in Her2 subtype. (R) *SRSF1* expression in Luminal A subtype. (S, T) *CCNB1*, *KIF23* expression in Luminal B subtype.

In terms of the pathway enrichment analysis, the nodes represent the pathway, while the edges mean that there is a functional similarity between the two pathways. As was consistent with the abovementioned analysis, the pathway enrichment analysis also showed that TNBC is the most complex subtype in breast cancer. In addition, there is an overlap of the pathways in the TNBC, Her2, and luminal B subtypes, while the pathways enriched in the luminal A subtype were unique. This suggested that the luminal A subtype occurs in a completely different way. If so, the treatment of the luminal A subtype should be different

from the other three subtypes, especially the postsurgical adjuvant therapy, or in other words, there should be a specific treatment strategy to be formulated just for the luminal A subtype compared with the others. Of course, pathways about DNA replication and DNA repair are only included in TNBC, which means studies focused on these pathways can help shed light upon TNBC and develop treatments that are only indicated for TNBC. Also, for the same reason, patients of TNBC, Her2, and luminal B subtypes may benefit from studies on mitosis in the future. Otherwise, in addition to pathways related to DNA replication, DNA repair,

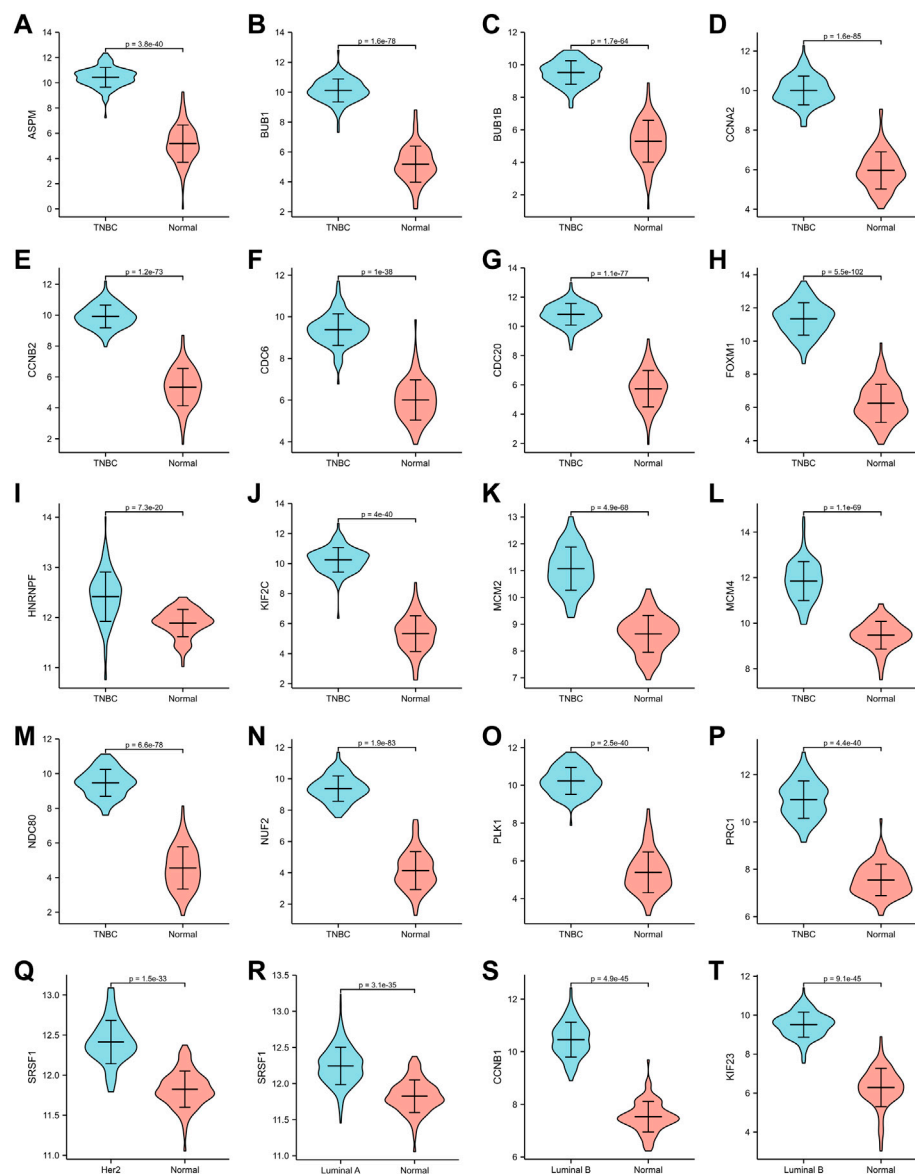
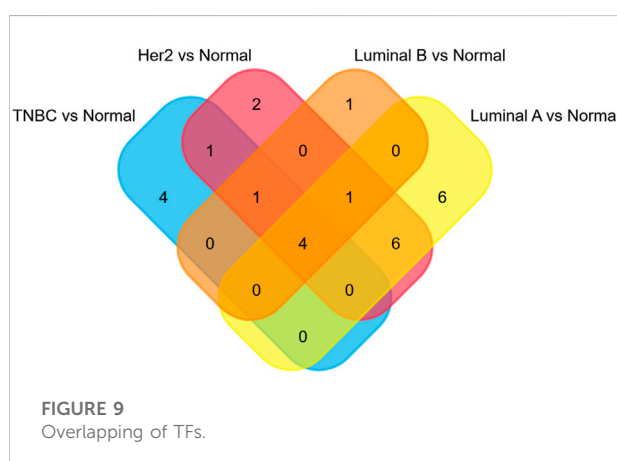
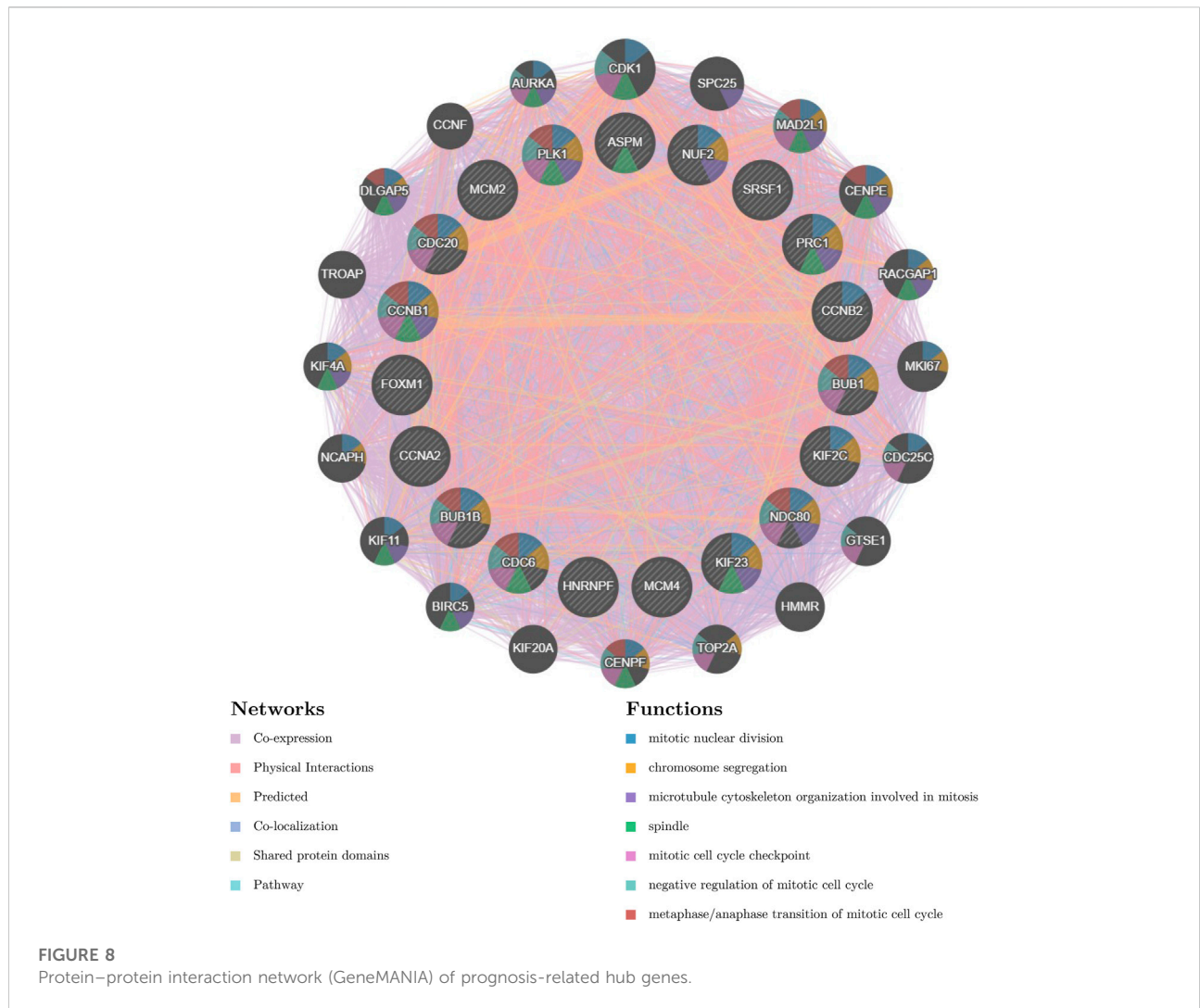


FIGURE 7

Validation of hub genes in TCGA dataset. (A–P) *ASPM*, *BUB1*, *BUB1B*, *CCNA2*, *CCNB2*, *CDC6*, *CDC20*, *FOXM1*, *HNRNP*, *KIF2C*, *MCM2*, *MCM4*, *NDC80*, *NUF2*, *PLK1*, *PRC1* expression in TNBC subtype. (Q) *SRSF1* expression in Her2 subtype. (R) *SRSF1* expression in Luminal A subtype. (S,T) *CCNB1*, *KIF 23* expression in Luminal B subtype.

and mitosis, the ERBB4 signal was only enriched in TNBC, while the NOTCH4 signal was only in the Her2 subtype, and chemokine and kinesins in both of them, compared with the luminal B subtype. ERBB4, a member of the human epidermal growth factor receptor family, has been previously reported to be a valuable prognostic marker when united with the pathologic stage in TNBC and may be helpful in predicting the therapeutic efficacy for TNBC (Kim et al., 2016). However, the biological function of ERBB4 and its potential as a cancer drug target have not been explicitly described, and we also

hope that patients suffering from TNBC with ERBB4 overexpression would benefit from further clinical trials on receptor tyrosine kinase (RTKs). NOTCH4 has been identified to hinder differentiation, functional development, and branching morphogenesis of the mammary epithelium (Uyttendaele et al., 1998). In breast cancer, NOTCH4 is predominantly expressed in the Her2 subtype, and the expression is also discovered to be associated with bad prognostic factors (Wang et al., 2018). As the NOTCH4 signal pathway is found to be enriched in the



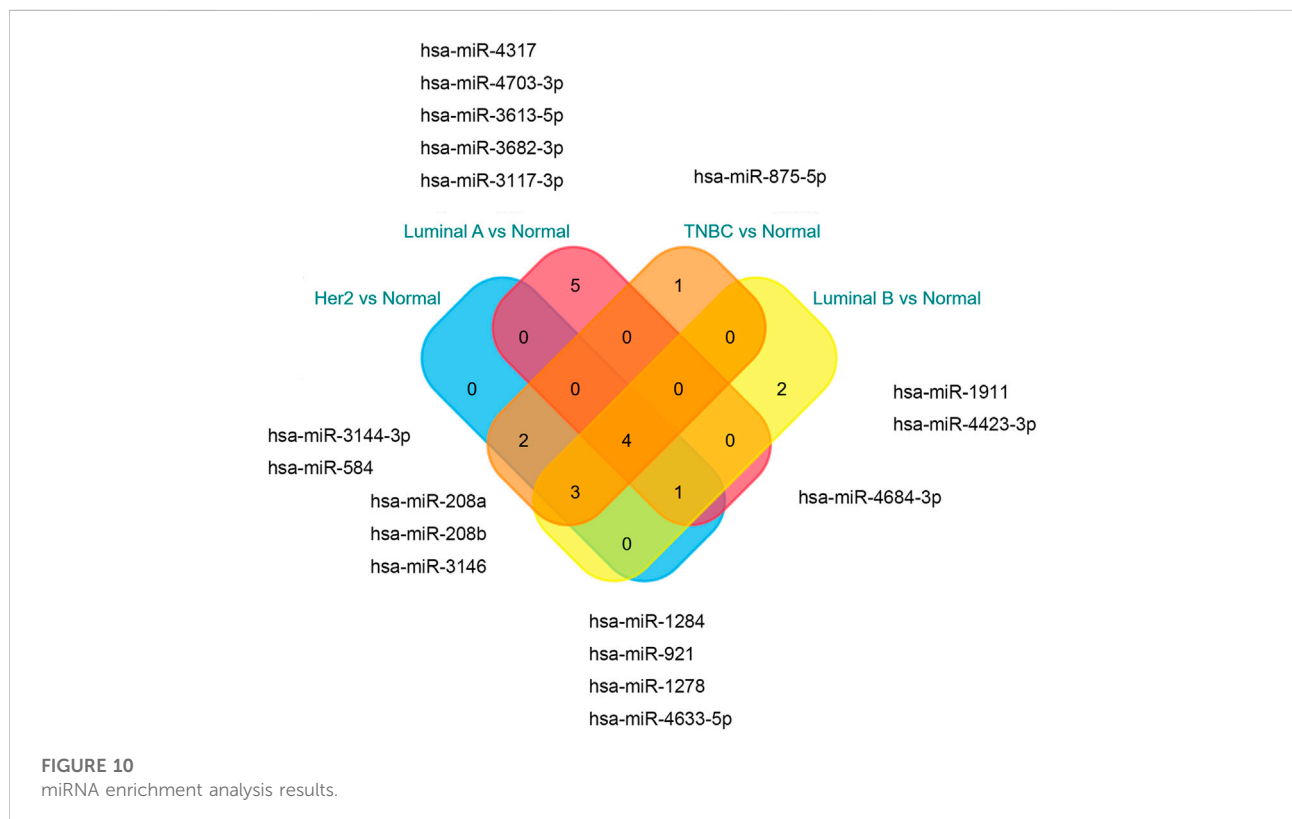
Her2 subtype in our study, using NOTCH4 antagonists to suppress NOTCH4 signaling may be a novel and individually distinct strategy to treat Her2 subtype breast cancer.

According to the result of TF analysis, although most of the TFs have been reported to be associated with breast cancer, there were still TFs such as EOMES, POU3F2, NR3C1, and RUNX1 that were less reported in breast cancer. Studies focused on these TFs may shed light upon breast cancer in a new way and provide a novel therapeutic strategy. For microRNA analysis, the miR-875 family has been reported to serve as a marker for detection and prognosis in breast cancer (Liu et al., 2021). Others, such as miR-1284, miR-3613, and miR-208a families play a role in the progression of breast cancer (Zhang et al., 2019; Zou et al., 2019; Liu Y. et al., 2020). It is valuable to investigate other miRNAs in experimental studies.

The present research also includes some limitations. First, our study only looks at data in one dataset, and the GPL platform used in the dataset is now not universally applicable. Second, most of the clinicopathologic features are not included in the dataset, and we cannot rule out that these factors might have influenced our results. Third, we just

TABLE 2 Transcription factor enrichment analysis.

TNBC vs. normal	Her2 vs. normal	Luminal A vs. normal	Luminal B vs. normal
E2F4	CLOCK	CLOCK	CLOCK
AR	FOXM1	FOXM1	FOXM1
CLOCK	RELA	ZNF217	RELA
RELA	POU3F2	SOX2	BACH1
E2F7	BACH1	WT1	SOX2
BACH1	ZNF217	RELA	EOMES
CHD1	AR	CJUN	AR
KDM6A	SOX2	AHR	
E2F1	FOXA1	ESR1	
FOXM1	PIAS1	AR	
	E2F7	ESR2	
	AHR	ARNT	
	CJUN	NR3C1	
	WT1	PIAS1	
	SMAD4	RUNX1	
		PPAR	
		SMAD4	



predicted the TFs and miRNAs through DEGs and did not calculate the relationship between them or carry out experiments to validate it, but this is just what we are doing now.

## Conclusion

In this study, we preliminarily delved into the potentially comprehensive molecular mechanisms of breast cancer by

creating a holistic view at the genomic and transcriptomic levels in different subtypes using computational tools. Our study introduced a network of genes, pathways, prognosis-related genes, TFs, and miRNAs which are possibly associated with a different subtype of breast cancer, and they can be good candidates for further analysis and provide novel approaches to treat breast cancer.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

## Author contributions

(I) Conception and design: XH and SW; (II) administrative support: SW and PS; (III) provision of study materials or patients: SW and GY; (IV) collection and assembly of data: SW and CY; (V) data analysis and interpretation: SW and LC; (VI) manuscript writing: all authors; (VII) final approval of manuscript: all authors.

## Funding

This work was supported by the Nature Science Foundation of Guangdong Province, China [Grant No. 2019A1515011331],

## References

- Alfarsi, L. H., Ansari, R. E., Craze, M. L., Toss, M. S., Masisi, B., Ellis, I. O., et al. (2019). CDC20 expression in oestrogen receptor positive breast cancer predicts poor prognosis and lack of response to endocrine therapy. *Breast Cancer Res. Treat.* 178, 535–544. doi:10.1007/s10549-019-05420-8
- Bader, G. D., and Hogue, C. W. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinforma.* 4, 2. doi:10.1186/1471-2105-4-2
- Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25, 1091–1093. doi:10.1093/bioinformatics/btp101
- Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61–70. doi:10.1038/nature11412
- Chen, G., Yu, M., Cao, J., Zhao, H., Dai, Y., Cong, Y., et al. (2021). Identification of candidate biomarkers correlated with poor prognosis of breast cancer based on bioinformatics analysis. *Bioengineered* 12, 5149–5161. doi:10.1080/21655979.2021.1960775
- Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., and Lin, C. Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* 8, S11. doi:10.1186/1752-0509-8-S4-S11
- Daulat, A. M., Finetti, P., Revinski, D., Silveira Wagner, M., Camoin, L., Audebert, S., et al. (2019). ECT2 associated to PRICKLE1 are poor-prognosis markers in triple-negative breast cancer. *Br. J. Cancer* 120, 931–940. doi:10.1038/s41416-019-0448-z
- Du, J. X., Luo, Y. H., Zhang, S. J., Wang, B., Chen, C., Zhu, G. Q., et al. (2021). Splicing factor SRSF1 promotes breast cancer progression via oncogenic splice switching of PTPMT1. *J. Exp. Clin. Cancer Res.* 40, 171. doi:10.1186/s13046-021-01978-8
- Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., et al. (2021). Cancer statistics for the year 2020: An overview. *Int. J. Cancer*, 778–789. Epub ahead of print. PMID: 33818764. doi:10.1002/ijc.33588
- Guosso, T., Mieulet, V., Cardon, M., Bourachot, B., Kieffer, Y., Devun, F., et al. (2016). Chronic oxidative stress promotes H2AX protein degradation and enhances chemosensitivity in breast cancer patients. *EMBO Mol. Med.* 8, 527–549. doi:10.15252/emmm.201505891
- Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P., et al. (2019). Breast cancer. *Nat. Rev. Dis. Prim.* 5, 66. doi:10.1038/s41572-019-0111-2
- Huang, D., Huang, Y., Huang, Z., Weng, J., Zhang, S., and Gu, W. (2019). Relation of AURKB over-expression to low survival rate in BCRA and reversine-modulated aurora B kinase in breast cancer cell lines. *Cancer Cell Int.* 19, 166. doi:10.1186/s12935-019-0885-z
- Issac, M. S. M., Yousef, E., Tahir, M. R., and Gaboury, L. A. (2019). MCM2, MCM4, and MCM6 in breast cancer: Clinical utility in diagnosis and prognosis. *Neoplasia* 21, 1015–1035. doi:10.1016/j.neo.2019.07.011
- Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., et al. (2020). The reactome pathway knowledgebase. *Nucleic Acids Res.* 48, D498–D503. doi:10.1093/nar/gkz1031
- Jian, W., Deng, X. C., Munankarmy, A., Borkhoo, O., Ji, C. L., Wang, X. H., et al. (2021). KIF23 promotes triple negative breast cancer through activating epithelial-mesenchymal transition. *Gland. Surg.* 10, 1941–1950. doi:10.21037/gs-21-19
- Jiang, C. F., Xie, Y. X., Qian, Y. C., Wang, M., Liu, L. Z., Shu, Y. Q., et al. (2021). TBX15/miR-152/KIF2C pathway regulates breast cancer doxorubicin resistance via promoting PKM2 ubiquitination. *Cancer Cell Int.* 21, 542. doi:10.1186/s12935-021-02235-w
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi:10.1093/nar/28.1.27

and the funding provides the article processing fee to Shan Wang and Xiaolei Hu. There are no other conflicts of interest to declare for all other authors.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.989565/full#supplementary-material>

- Kim, J. Y., Jung, H. H., Do, I. G., Bae, S., Lee, S. K., Kim, S. W., et al. (2016). Prognostic value of ERBB4 expression in patients with triple negative breast cancer. *BMC Cancer* 16, 138. doi:10.1186/s12885-016-2195-3
- Kim, Y. J., Lee, G., Han, J., Song, K., Choi, J. S., Choi, Y. L., et al. (2019). UBE2C overexpression aggravates patient outcome by promoting estrogen-dependent/independent cell proliferation in early hormone receptor-positive and HER2-negative breast cancer. *Front. Oncol.* 9, 1574. doi:10.3389/fonc.2019.01574
- Koyuncu, D., Sharma, U., Goka, E. T., and Lippman, M. E. (2021). Spindle assembly checkpoint gene BUB1B is essential in breast cancer cell survival. *Breast Cancer Res. Treat.* 185, 331–341. doi:10.1007/s10549-020-05962-2
- Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44, W90–W97. doi:10.1093/nar/gkw377
- Lin, Y., Fu, F., Lv, J., Wang, M., Li, Y., Zhang, J., et al. (2020). Identification of potential key genes for HER-2 positive breast cancer based on bioinformatics analysis. *Med. Baltim.* 99, e18445. doi:10.1097/MD.00000000000018445
- Liu, X., Liu, M., Ma, H., Wang, J., and Zheng, Y. (2021). miR-875 serves as A candidate biomarker for detection and prognosis and is correlated with PHH3 index levels in breast cancer patients. *Clin. Breast Cancer* 22, e199–e205. doi:10.1016/j.clbc.2021.06.008
- Liu, S., Liu, X., Wu, J., Zhou, W., Ni, M., Meng, Z., et al. (2020). Identification of candidate biomarkers correlated with the pathogenesis and prognosis of breast cancer via integrated bioinformatics analysis. *Med. Baltim.* 99, e23153. doi:10.1097/MD.00000000000023153
- Liu, Y., Yang, Y., Du, J., Lin, D., and Li, F. (2020). MiR-3613-3p from carcinoma-associated fibroblasts exosomes promoted breast cancer cell proliferation and metastasis by regulating SOCS2 expression. *IUBMB Life* 72, 1705–1714. doi:10.1002/iub.2292
- Lv, S., Xu, W., Zhang, Y., Zhang, J., and Dong, X. (2020). NUF2 as an anticancer therapeutic target and prognostic factor in breast cancer. *Int. J. Oncol.* 57, 1358–1367. doi:10.3892/ijo.2020.5141
- Mahadevappa, R., Neves, H., Yuen, S. M., Bai, Y., McCrudden, C. M., Yuen, H. F., et al. (2017). The prognostic significance of Cdc6 and Cdt1 in breast cancer. *Sci. Rep.* 7, 985. doi:10.1038/s41598-017-00998-9
- Milde-Langosch, K., Karn, T., Muller, V., Witzel, I., Rody, A., Schmidt, M., et al. (2013). Validity of the proliferation markers Ki67, TOP2A, and RacGAP1 in molecular subgroups of breast cancer. *Breast Cancer Res. Treat.* 137, 57–67. doi:10.1007/s10549-012-2296-x
- Mo, C. H., Gao, L., Zhu, X. F., Wei, K. L., Zeng, J. J., Chen, G., et al. (2017). The clinicopathological significance of UBE2C in breast cancer: a study based on immunohistochemistry, microarray and RNA-sequencing data. *Cancer Cell Int.* 17, 83. doi:10.1186/s12935-017-0455-1
- Nielsen, T. O., Hsu, F. D., Jensen, K., Cheang, M., Karaca, G., Hu, Z., et al. (2004). Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin. Cancer Res.* 10, 5367–5374. doi:10.1158/1078-0432.CCR-04-0220
- Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., et al. (2000). Molecular portraits of human breast tumours. *Nature* 406, 747–752. doi:10.1038/35021093
- Ren, Y., Deng, R., Zhang, Q., Li, J., Han, B., and Ye, P. (2020). Bioinformatics analysis of key genes in triple negative breast cancer and validation of oncogene PLK1. *Ann. Transl. Med.* 8, 1637. doi:10.21037/atm-20-6873
- Roßwag, S., Cotarello, C. L., Pantel, K., Riethdorf, S., Sleeman, J. P., Schmidt, M., et al. (2021). Functional characterization of circulating tumor cells (CTCs) from metastatic ER+/HER2- breast cancer reveals dependence on HER2 and FOXM1 for endocrine therapy resistance and tumor cell survival: Implications for treatment of ER+/HER2- breast cancer. *Cancers (Basel)* 13, 1810. doi:10.3390/cancers13081810
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi:10.1101/gr.1239303
- Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer statistics, 2021. *Ca. Cancer J. Clin.* 71, 7–33. doi:10.3322/caac.21654
- Sørli, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U. S. A.* 98, 10869–10874. doi:10.1073/pnas.191367098
- Szklarczyk, D., Gable, A. L., Lyon, D., Jung, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D13. doi:10.1093/nar/gky1131
- Tang, J., Lu, M., Cui, Q., Zhang, D., Kong, D., Liao, X., et al. (2019). Overexpression of ASPM, CDC20, and TTK confer a poorer prognosis in breast cancer identified by gene Co-expression network analysis. *Front. Oncol.* 9, 310. doi:10.3389/fonc.2019.00310
- Tyson-Capper, A., and Gautrey, H. (2018). Regulation of Mcl-1 alternative splicing by hnRNP F, H1 and K in breast cancer cells. *RNA Biol.* 15, 1448–1457. doi:10.1080/15476286.2018.1551692
- Uyttendaele, H., Soriano, J. V., Montesano, R., and Kitajewski, J. (1998). Notch4 and Wnt-1 proteins function to regulate branching morphogenesis of mammary epithelial cells in an opposing fashion. *Dev. Biol.* 196, 204–217. doi:10.1006/dbio.1998.8863
- Wang, Z., Katsaros, D., Shen, Y., Fu, Y., Canuto, E. M., Benedetto, C., et al. (2015). Biological and clinical significance of MAD2L1 and BUB1, genes frequently appearing in expression signatures for breast cancer prognosis. *PLoS One* 10, e0136246. doi:10.1371/journal.pone.0136246
- Wang, J. W., Wei, X. L., Dou, X. W., Huang, W. H., Du, C. W., and Zhang, G. J. (2018). The association between Notch4 expression, and clinicopathological characteristics and clinical outcomes in patients with breast cancer. *Oncol. Lett.* 15, 8749–8755. doi:10.3892/ol.2018.8442
- Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., et al. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 38, W214–W220. doi:10.1093/nar/gkq537
- Xing, Z., Wang, X., Liu, J., Zhang, M., and Feng, K. (2021). Expression and prognostic value of CDK1, CCNA2, and CCNB1 gene clusters in human breast cancer. *J. Int. Med. Res.* 49, 300060520980647. doi:10.1177/0300060520980647
- Xiu, Y., Liu, W., Wang, T., and Ha, M. (2019). Overexpression of ECT2 is a strong poor prognostic factor in ER(+) breast cancer. *Mol. Clin. Oncol.* 10, 497–505. doi:10.3892/mco.2019.1832
- Yang, K., Gao, J., and Luo, M. (2019). Identification of key pathways and hub genes in basal-like breast cancer using bioinformatics analysis. *Oncol. Targets. Ther.* 12, 1319–1331. doi:10.2147/OTT.S158619
- Zhang, P., Yang, F., Luo, Q., Yan, D., and Sun, S. (2019). miR-1284 inhibits the growth and invasion of breast cancer cells by targeting ZIC2. *Oncol. Res.* 27, 253–260. doi:10.3727/096504018X15242763477504
- Zou, Y., Zheng, S., Xiao, W., Xie, X., Yang, A., Gao, G., et al. (2019). circRAD18 sponges miR-208a/316a to promote triple-negative breast cancer progression through regulating IGF1 and FGF2 expression. *Carcinogenesis* 40, 1469–1479. doi:10.1093/carcin/bgz071