Research Paper

Identification of differentially expressed genes-related prognostic risk model for survival prediction in breast carcinoma patients

Jinyu Li[1,*], Gena Huang[1,*], Caixia Ren[2], Ning Wang[3], Silei Sui[4], Zuowei Zhao[1,5,&], Man Li[1]

[1]Department of Breast Oncology, The Second Hospital of Dalian Medical University, Dalian, Liaoning 116023, China
[2]Department of Respiratory Medicine, The Second Hospital of Dalian Medical University, Dalian, Liaoning 116023, China
[3]Institute for Genome Engineered Animal Models of Human Diseases, Dalian Medical University, Dalian, Liaoning 116044, China
[4]Institute of Cancer Stem Cell, Dalian Medical University, Dalian, Liaoning 116044, China
[5]Department of Breast Surgery, The Second Hospital of Dalian Medical University, Dalian, Liaoning 116023, China
[*]Equal contribution

Correspondence to: Zuowei Zhao, Man Li; **email:** dmuzhaozuowei@163.com, https://orcid.org/0000-0002-2569-3822; liman19890930@sina.com, https://orcid.org/0000-0002-6159-7440

Keywords: breast cancer, differentially expressed genes, prognostic risk model, prognostic outcome, Cox regression
Received: November 19, 2020 **Accepted:** May 31, 2021 **Published:** June 26, 2021

ABSTRACT

Since the imbalance of gene expression has been demonstrated to tightly related to breast cancer (BRCA) genesis and growth, common genes expressed of BRCA were screened to explore the essence in-between. In current work, most common differentially expressed genes (DEGs) in various subtypes of BRCA were identified. Functional enrichment analysis illustrated the driving factor of deactivation of the cell cycle and the oocyte meiosis, which critically triggers the development of BRCA. Herein, we constructed a 12-gene prognostic risk model relative to differential gene expression. Subsequently, the K-M curves, analysis on time-ROC curve and Cox regression were performed to assess this risk model by determining the respective prognostic value, and the prediction performance were ascertained for both training and validation cohorts. In addition, multivariate Cox regression was analysed to reveal the independence between risk score and prognostic stage, and the accuracy and sensitivity of prognosis are particularly improved after clinical indicators are included into the analysis. In summary, this study offers novel insights into the imbalance of gene expression within BRCA, and highlights 12 selected genes associated with patient prognosis. The risk model can help individualize treatment for patients at different risks, and propose precise strategies and treatments for BRCA therapy.

INTRODUCTION

As the most common malignant tumor in women, breast cancer has exhibited the fifth highest death rate after stomach cancer worldwide [1]. With the progressive increase of both incidence and mortality, it is crucial to evaluate the prognosis in clinical environment and to propose an appropriate therapeutic regimen for malignant tumor patients. In clinical practice, TNM staging is generally used, but it may lead to an entirely distinct prognosis in the same situation [2]. Therefore, a more precise and valuable method is highly desirable to predict outcomes.

The past decade has emerged several novel technologies to explore cancer's molecular characteristics, especially with the rapid advancement of high-throughput next-generation sequencing (NGS), bioinformatics analyses, machine learning, and gene microarray technique. These techniques extremely contributed to early diagnosis of tumors, prognosis prediction, and individualized treatment. In addition, the bioinformatics-based biomarker discovery renders a deeper understanding on disease-related regulatory pathways and molecular mechanisms. To identify high-risk patients, gene risk models are established via bioinformatics analysis using clinical information and gene expression data. More studies have been addressed on establishing gene risk models, resulting from the measuring capability of mRNA expression by NGS and microarray. Particularly, several risk models have played an excellent role in predicting the prognosis outcomes, including autophagy-related gene models, immune cell infiltration-related gene models, nomograms, and so on [3–6]. However, these models elucidated the prognosis of BRCA from the perspective of a single functional genome, which limited their prediction results.

The occurrence and development of tumors are highly relative to the accumulative changes in tumor suppressor genes and oncogenes [7]. The differentially expressed genes (DEGs) play varying roles during different periods of the occurrence and distinct developmental stages of cancer [8]. Genes' abnormal expression has been previously reported to accelerate the progression of malignancies, and DEGs have been targeted as a novel treatment approach in several antitumor clinical types of research [9, 10].

In this study, high-throughput mRNA expression profiles from distinct regions and races have been investigated, focusing on the differences between breast cancer and adjacent mammary tissue hence to identify potential genetic biomarkers. The selected DEGs profiles were incorporated with TCGA-BRCA clinical data to explore DEGs' role in prognosis. Moreover, the independent prediction of BRCA patients' outcomes was achieved using a risk model based on 12 DEGs-related signatures. As such, the risk prediction model is demonstrated as a reliable prognostic marker in BRCA patients. On the other hand, our functional biomarker-based study also provides a novel alternative to predict prognosis in BRCA patients.

RESULTS

Identification of DEGs in BRCA

To identify DEGs in this study, breast carcinoma samples (479 cases) and normal breast tissues (206 cases) were randomly collected from different regions and races. Using the limma package, we identified 3065 DEGs in GSE29431, 1293 DEGs in GSE32641, 1315 DEGs in GSE61304, 2252 DEGs in GSE70947 and 722 DEGs in GSE86374. DEGs in two representative samples from each of the five expression profiling databases are shown in the volcano plots (Figure 1A–1E). In Figure 1F and 1G, 61 upregulated genes and 90 downregulated genes are in common ($|\log2\text{Fold Change}| > 1$, adj.$p < 0.05$), which are displayed in the heatmap in five databases (Figure 1H).

Functional enrichment of DEGs and PPI network construction

The biological roles of 151 DEGs were further investigated using GO and KEGG pathway analysis. There are three functional categories in GO analysis: 1) in terms of biological process (BP), upregulated DEGs were merely enriched in nuclear division and organelle fission, while downregulated ones were enriched in peptide hormone and smooth muscle cell proliferation; 2) in terms of cellular component (CC), upregulated DEGs were involved in spindle, condensed chromosome and spindle pole, while downregulated ones were involved in collagen-containing extracellular matrix, lipid droplet and basement membrane; 3) in terms of molecular function (MF), microtubule motoring and binding activities were remarkably relative to upregulated DEGs, while downregulated ones were merely involved in integrin and growth factor and glycosaminoglycan binding (Figure 2A–2B). Additionally, KEGG pathway analysis demonstrated the participated roles of most upregulated genes in cell cycle and oocyte meiosis. As a contrast, downregulated genes were merely involved in PPAR signaling pathway and tyrosine metabolism (Figure 2C). The heatmap indicated the relationship between DEGs and enriched KEGG pathways (Figure 2D). Lastly, there are 151 nodes and 1169 edges in the PPI network of DEGs (51 upregulated genes and 90 downregulated genes, p = 1.0e-16), and the highest degree is determined at 48 (Supplementary Figure 1).

Creation and validation of OS-related prognostic risk signature by cox regression

TCGA-BRCA and GEO databases were employed for model training (n = 1076) and validation (n = 408), respectively. As indicated in Supplementary Table 1, there are 31 genes correlated to OS in the TCGA-BRCA cohort upon the univariate Cox regression analysis. Furthermore, the significant OS-related DEGs were identified using multivariate cox regression, in which 12 genes could serve as potential prognostic predictors in BRCA patients (Figure 3A). In Figure 3B and 3C, the correlation analysis suggested that 12 selected genes were cross-interacted.

Besides, genetic alteration of 12 risk-related genes was proceeded to identify their performance in BRCA patients (Supplementary Figure 2). PPI results suggested that most of the 12 selected genes were significantly intercorrelated (*p* < 1.0e-16, Supplementary Figure 3). Subsequently, the survival condition was evaluated in Kaplan Meier-Plotter datasets, indicating that all selected genes were included in the prognostic risk model with a promising survival predicting performance (Supplementary Figure 4).

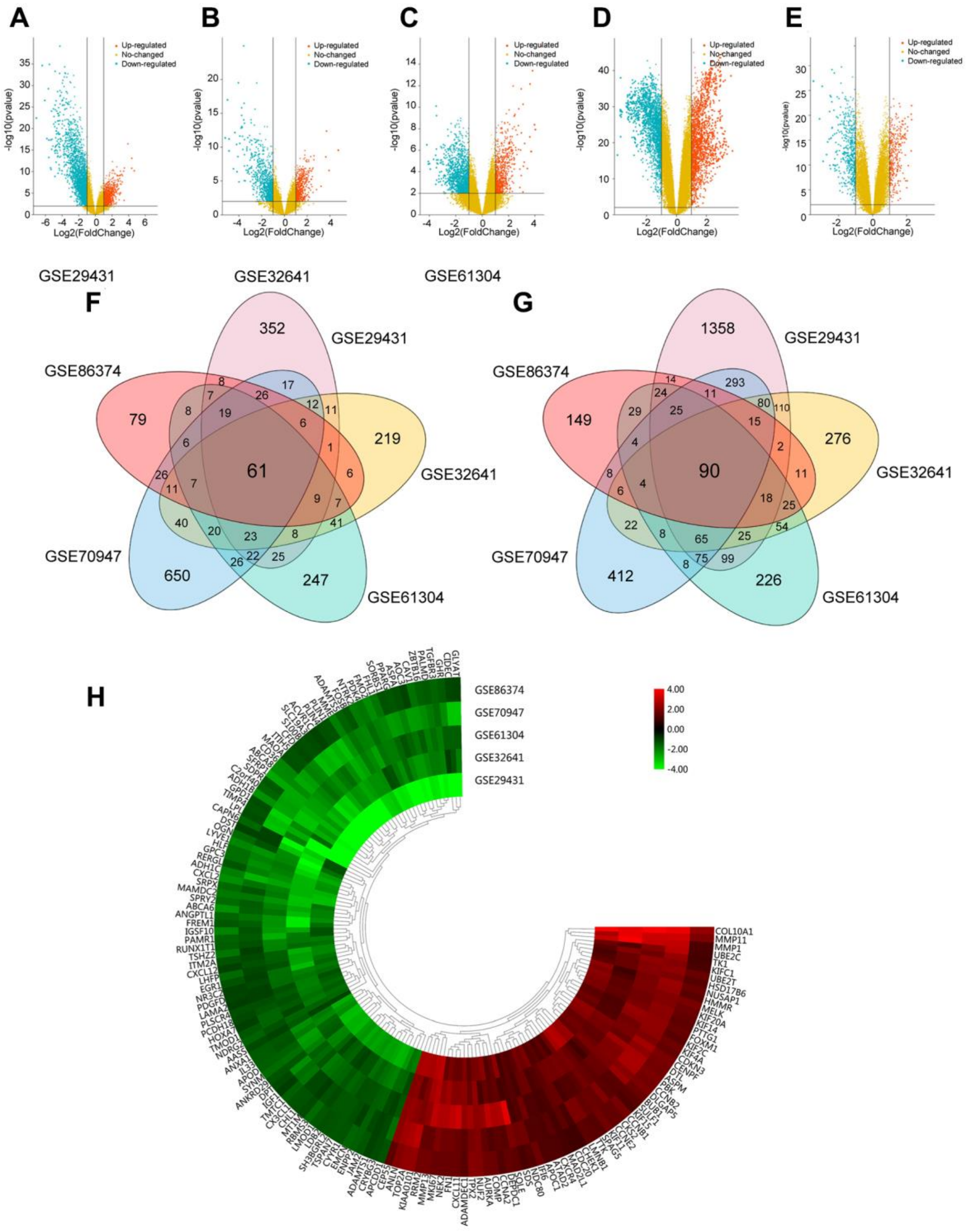


Figure 1. Foundation of DEGs in five GEO databases. (**A**–**E**) The display of DEGs in each database by volcano plots. Datasets from GSE29431 (**A**), GSE32641 (**B**), GSE61304 (**C**), GSE70947 (**D**) and GSE86374 (**E**). Orange: Up-regulated genes (logFC ≥ 1.0, adj. *P* < 0.05); Blue: Down-regulated genes (logFC ≤ -1.0, adj. *P* < 0.05); Yellow: Genes with no significance. (**F**–**G**) A Total of 61 significantly upregulated genes (**F**) and 90 significantly downregulated genes (**G**) were screened from the five GEO databases. (**H**) Hierarchical clustering heatmap showed expression of 151 DEGs in five GEO databases. Red: higher expression genes, green: lower expression genes.

Lastly, the formula was established for the prognostic risk model by referring to multiple Cox regression (Risk score = 0.30912 × ANLN expression (Exp-ANLN) + 0.26714 × Exp-CCNE2 + 0.47389 × Exp-KIF4A – 0.42876 × Exp-MELK – 0.55747 × Exp-NDC80 – 0.44425 × Exp-NEK2 – 0.23508 × Exp-TOP2A + 0.34102 × Exp-TTK + 0.46718 × Exp-UBE2T – 0.25889 × Exp-FREM1 + 0.29101 × Exp-IGF1 – 0.20025 × Exp-SORBS1). The respective risk score was determined in the training cohort, patients were then divided into high-risked group or low-risked group (Figure 4A), in accordance to the calculated median risk scores. Using Kaplan-Meier survival analysis, high-risked patients exhibit lower OS rates than that of low-risked patients in the training cohorts ($p < 0.0001$, Figure 4B).

The predictive risk model on prognosis was further investigated, and the established and acquired formula was applied for other 408 BRCA patients in separate cohorts. The validation cohorts consisted of GSE20685 ($n = 327$) and GSE48390 ($n = 81$) databases, including mRNA expression, survival status and survival time. Similarly, patients were group-categorized according to

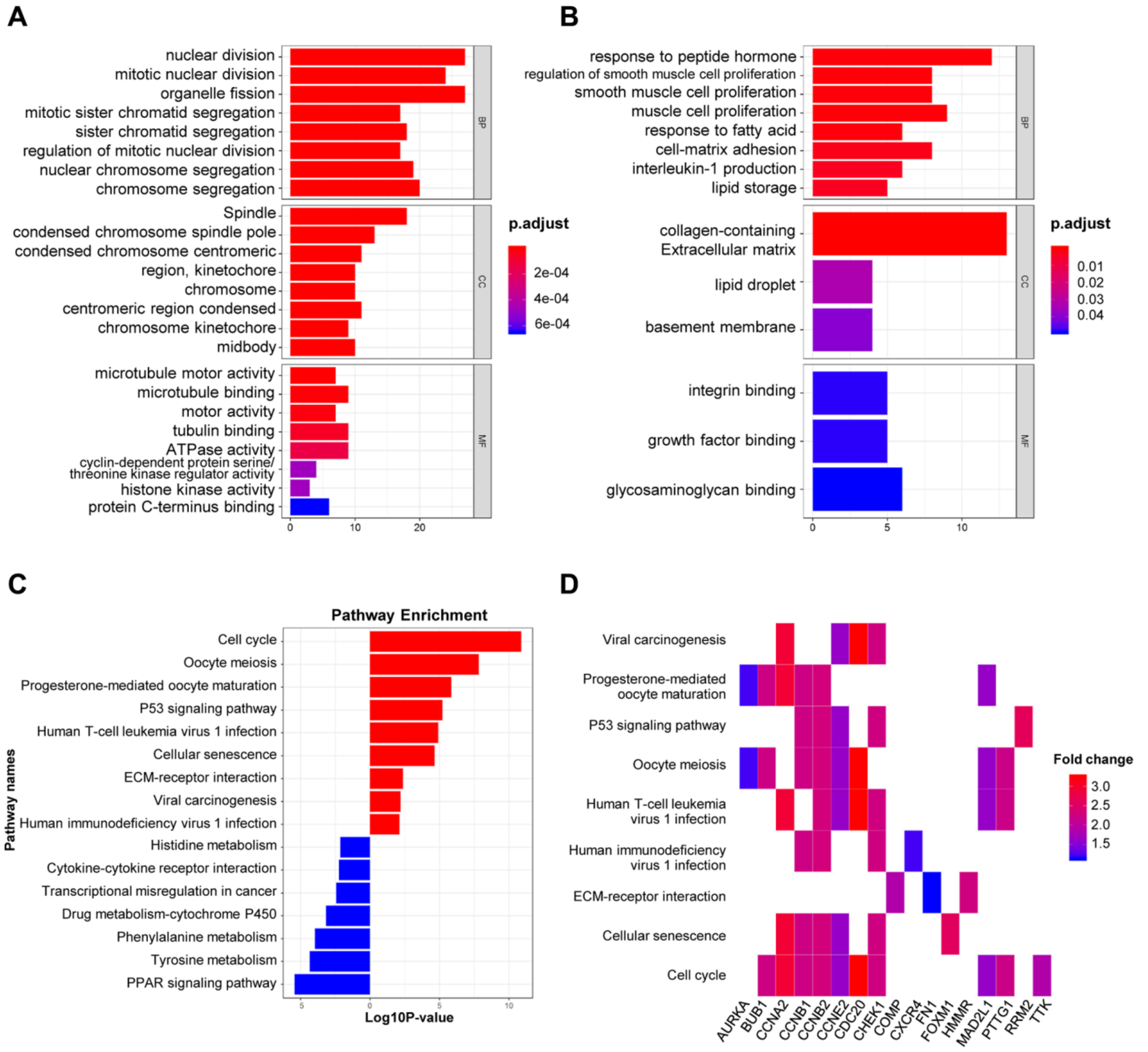


Figure 2. GO and KEGG analysis of DEGs in BRCA. (**A**–**B**) The biological processes, cellular components, and molecular functions of 61 up-regulated DEGs (**A**) and 90 down-regulated DEGs (**B**) were displayed by GO analysis. (**C**) The signaling pathways of 151 DEGs were displayed by KEGG analysis. (**D**) Heatmap of the significant enrichment results in the KEGG pathway.

their calculated risk scores. For patients with higher risk score in the validation cohorts (Figure 4C–4D), the worse OS rates were observed using Kaplan-Meier survival analysis. Moreover, receiver operating characteristic (ROC) curves were plotted to investigate the time-dependent dependability of the risk model. As shown in Figure 4E, the area under curve (AUC) in 5-year and 10-year survival was 0.74 and 0.72, respectively, in TCGA-BRCA training cohort, demonstrating the survival predicting capability of as-constructed risk model. On the other hand, ROC curves were utilized in validation cohorts. For example, the

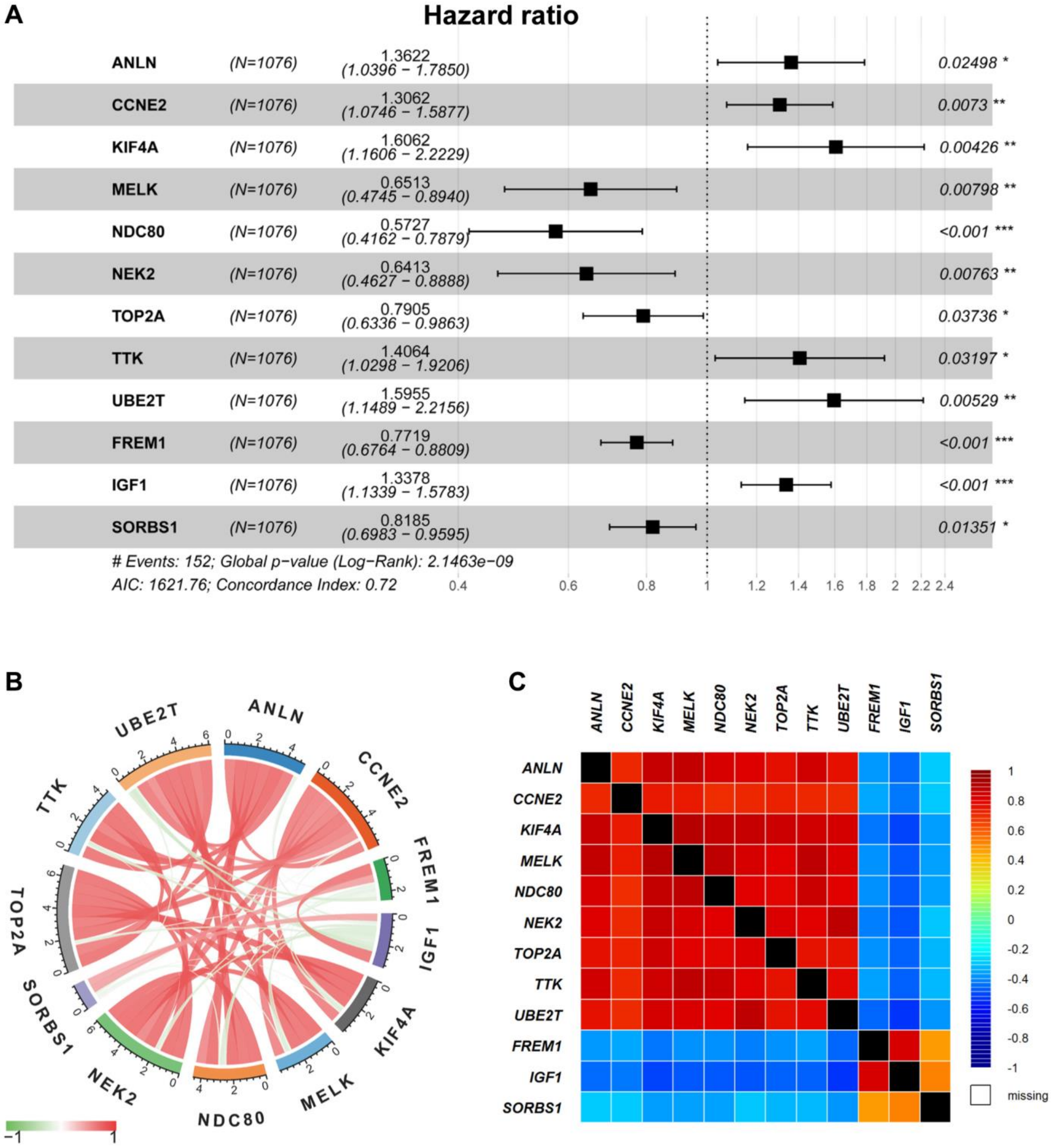


Figure 3. Multivariable Cox regression and correlation analysis. (**A**) The multivariable Cox regression analysis of 12 selected DEGs were displayed by forest plot. (**B**–**C**) The correlation analysis of 12 selected DEGs was displayed by Pearson's correlation (**B**) and the bc-GenExMiner software (**C**), respectively.

representative AUC of 5-year and 10-year survival rates were 0.73 and 0.67 in the GSE20685 cohort (Figure 4F), and that in the GSE48390 cohort were 0.81 and 0.72, respectively (Figure 4G). These results suggested that our developed risk model was a dependable prognostic indicator with improved performance. In a nutshell, the combination of the 12-DEGs-related risk signatures in the validation cohorts demonstrated its significant predictive value for prognosis.

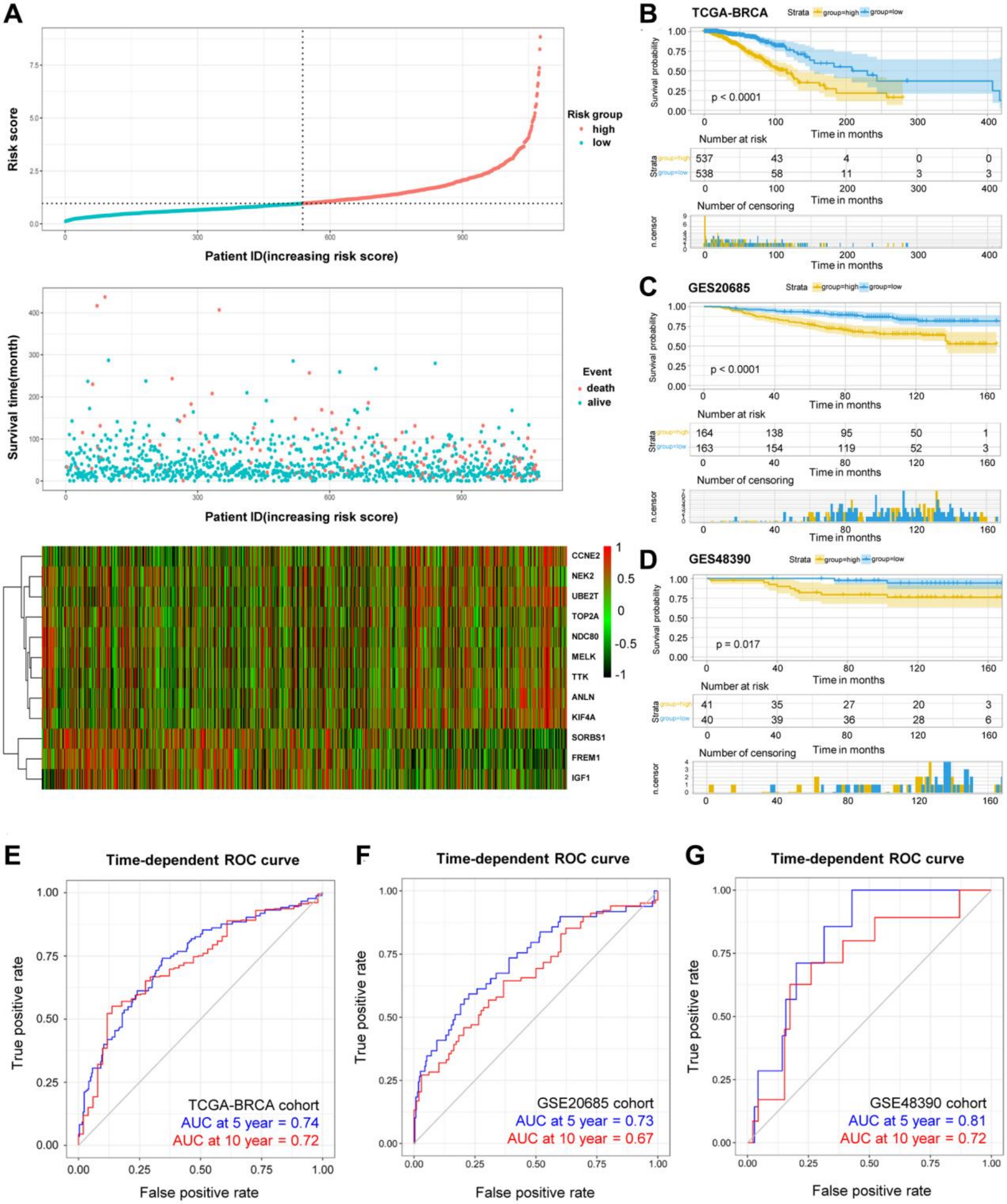


Figure 4. OS-related prognostic risk model of BRCA patients. (**A**) The display of prognostic risk model with risk score, patient survival time and status in TCGA-BRCA database. (**B**–**D**) The K-M survival curves of the high- and low-risk group of TCGA-BRCA cohort (**B**) and validation GSE20685 cohort (**C**) and GSE48390 cohort (**D**). (**E**–**G**) The prediction of 5- and 10-year survival in TCGA-BRCA cohort (**E**) and GSE20685 cohort (**F**) and GSE48390 cohort (**G**) by Time-ROC.

The correlation analysis between DEGs-related risk model and clinicopathological variables

Univariate and multivariate Cox regression were subsequently analysed to identify the roles of DEGs-related risk model in predicting prognosis. Univariate Cox regression revealed that there are 6 risk factors for survival prediction, including advanced age, pathological stage, PAM50 molecular subtypes, tumor size, lymph node metastasis, and risk score evaluated by the DEGs-related model (Figure 5A). In addition, multivariate Cox regression analysis demonstrated the high consistence between the OS of BRCA patients and the abovementioned six factors, especially the risk score (Figure 5B). In summary, these Cox regression results demonstrated the functions and contributions of risk score in predicting prognosis without restrictions from tumor clinicopathologic features.

The probability of 5-year and 10-year OS was predicted by generating a nomogram. As shown in Figure 5C, calibration curves demonstrated the favorable consistency in actual and predicted survival performance (Figure 5D–5E), especially for 5-year survival. In addition, after combination of age, pathological stage, PAM50 molecular subtypes, tumor size, lymph node metastasis and our risk score, the predictive accuracy was significantly improved in TCGA-BRCA (Table 1). As evidenced in the time-ROC results in Figure 5F, the comprehensive analysis offered more accurate predictions on the prognosis in BRCA patients (5-year AUC = 0.83, 10-year AUC = 0.79).

To verify the effect of DEGs-related risk signature for survival on the malignancy of BRCA, our risk model was correlated with clinicopathological variables. In Figure 6, the risk score was significantly increased, when the HER2 subtype and luminal B subtype, the later clinical stage, the larger tumor size, and lymph node metastasis were taken into consideration, confirming the excellent consistency between risk score and prognostic outcomes.

Exploration of the mechanism in predicted differential risk patients by GSEA

The functional differences among differential risk patients by GSEA were explored by comparing patients in low-risk and high-risk groups. For instance, high-risked patients were positively correlated with cell cycle (NES = 1.85, p = 0.012), TCA cycle (NES = 1.82, p = 0.016), and oxidative phosphorylation (NES = 2.05, p = 0.026). Meanwhile, patients in low-risk groups were negatively correlated with basal cell carcinoma (NES = –1.80, p = 0.004), Hedgehog signaling pathway (NES = –1.69, p = 0.010), and JAK-STAT signaling pathway (NES = –1.65, p = 0.035) (Figure 7). Therefore, the presence of an intensively regulatory role was observed for the development and progression in high-risk BRCA patients, exhibiting significant changes in pathways.

Prediction of targeted treatment in BRCA patients by our risk score

As presented in Figure 8, our DEGs-related risk score was closely associated with CDK4 expression (cor = 0.12, p = 9.9e-05), as well as the expressions of ERBB2 (cor = 0.14, p < 2.2E-16), EGFR (cor = –0.14, p <4.6e-06), and KIT (cor = –0.19, p < 4e-10) by Pearson’s correlation analysis. In conclusion, patients with higher DEGs-related risk score exhibited favorable therapeutic response to CDK4- and ERBB2- targeted treatments. Otherwise, the lower risk score, patients exhibited better response to EGFR and KIT targeted treatments.

Verification of the 12-prognostic mRNA expressions between BRCA specimens and adjacent breast tissues by qRT-PCR

To avoid false-positive results from public database, 12-mRNA expressions were further verified based on qRT-PCR results from 20 frozen tissues from BRCA patients (Figure 9). The experimental results indicated that the mRNA expressions of ANLN, KIF4A, MELK, NDC80, NEK2, TOP2A, TTK and UBE2T were upregulated in BRCA tissue in comparison to the adjacent tissues, while that of FREM1 and SORBS1 were downregulated in BRCA tissues compared to the adjacent tissues. These validation results were generally consistent with TCGA-BRCA database (Supplementary Figure 2), suggesting that these prognostic genes played critical roles in the initiative and developing stages of breast cancer.

DISCUSSION

The aberrant gene expression is a major threat during the progressively developing stages of BRCA, recent intensive studies have indicated that some genes could be potentially targeted for diagnosis, treatment, and prognosis in BRCA. Thus, it is highly desirable to discover effective gene signatures to identify patients' condition, not only to find applicable prognostic targets but also to provide precise therapy for patients at high risk for disease recurrence. Nowadays, full-scale genetic data from BRCA samples can be obtained using DNA microarray and next-generation sequencing, providing comprehensive assessment during diseases progression.

Model in the present study was constructed using five GEO databases and TCGA-BRCA database. Among the total 151 DEGs, 31 DEGs were significantly correlated to the prognosis in BRCA patients, and 12 DEGs (ANLN, CCNE2, KIF4A, MELK, NDC80, NEK2, TOP2A, TTK, UBE2T, FREM1, IGF1 and SORBS1) were recognized and included in the risk model for overall survival prediction. For instance, ANLN was identified as the target for BRCA patients, which correlated with poor survival [11, 12]. CCNE2

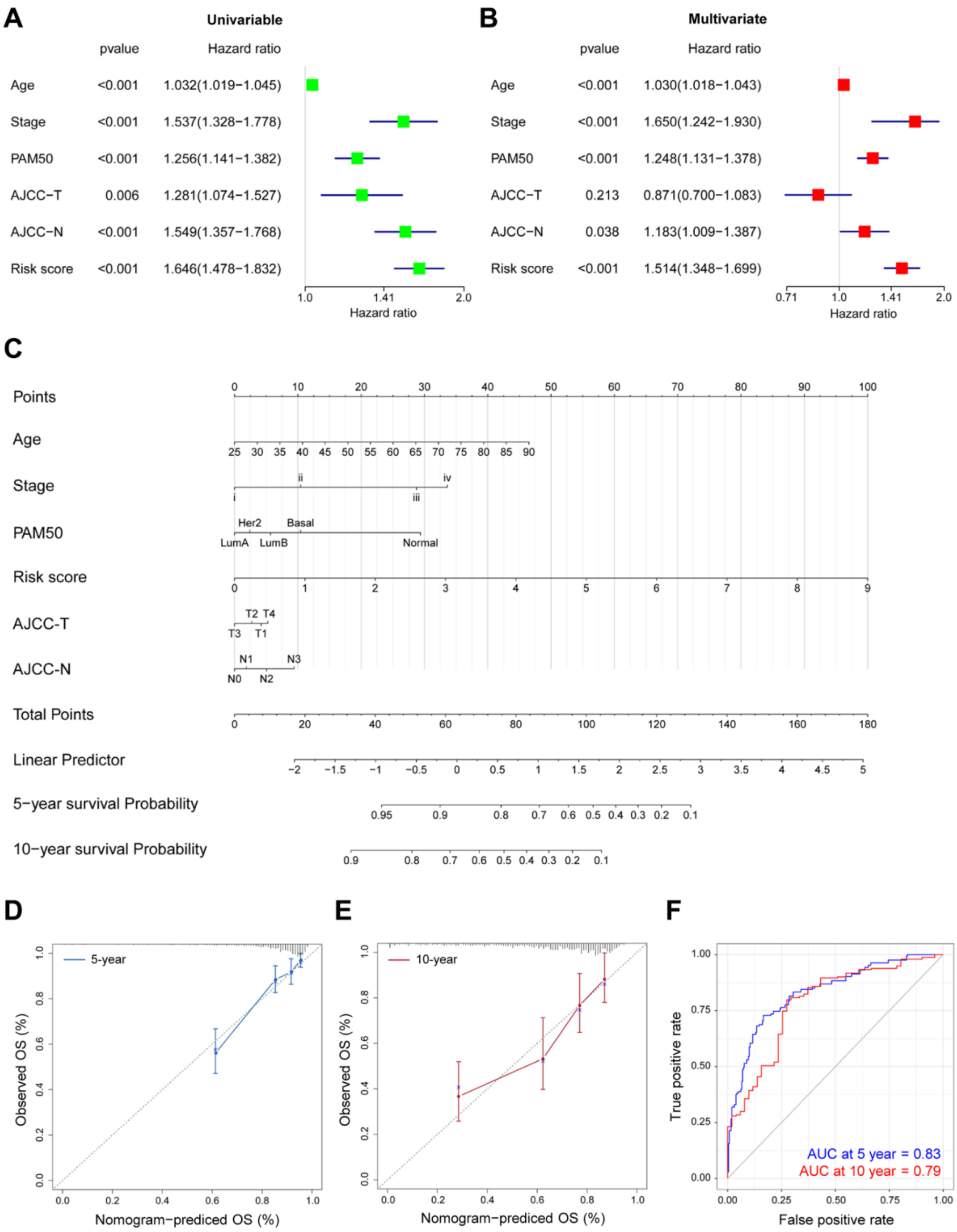


Figure 5. Prognostic nomogram for predicting survival in TCGA-BRCA patients. (**A**–**B**) The correlations of the OS risk score and clinical variables by Univariate (**A**) and Multivariate (**B**) Cox regression. (**C**) Prognostic nomogram with certain characteristics in TCGA-BRCA patients. (**D**–**E**) The prediction of 5- (**D**) and 10-year (**E**) survival by calibration curves. x-axis, predicted OS; y-axis, observed OS; the solid line, predicted nomogram; the vertical bars, 95% confidence interval. (**F**) Time-ROC curves for the combination of age, stage, PAM50, T, N and risk score.

Table 1. Comparing of the predictive efficiency of the prognostic risk models in the entire training cohorts (*n* = 1076).

Factor	Overall survival		
	C-index	95% CI	AIC
Age	0.633	0.576–0.690	1640.11
Stage	0.693	0.643–0.734	1628.64
PAM50	0.619	0.567–0.672	1622.68
T	0.608	0.552–0.663	1655.79
N	0.657	0.603–0.711	1634.36
Risk score	0.721	0.670–0.772	1621.76
Age + stage + PAM50 + T + N	0.782	0.737–0.826	1555.89
Risk score + age + stage + PAM50 + T + N	0.811	0.768–0.854	1520.41

Abbreviations: C-index: Harrell's concordance index; CI: confidence interval; AIC: Akaike information criterion.

promoted G1-S transition in HER2+ BRCA, and hence resulted in trastuzumab resistance [13]. In certain stages of mitosis, KIF4A served as a biomarker for predicting clinical prognosis [14, 15]. MELK promoted the occurrence and progression of colorectal adenocarcinomas [16, 17]. NDC80 was also recognized as a potential target in BRCA [18]. The expression of NEK2 exhibited its importance during the mitotic cell cycle of BRCA [19, 20]. As previously reported, the higher the expression of

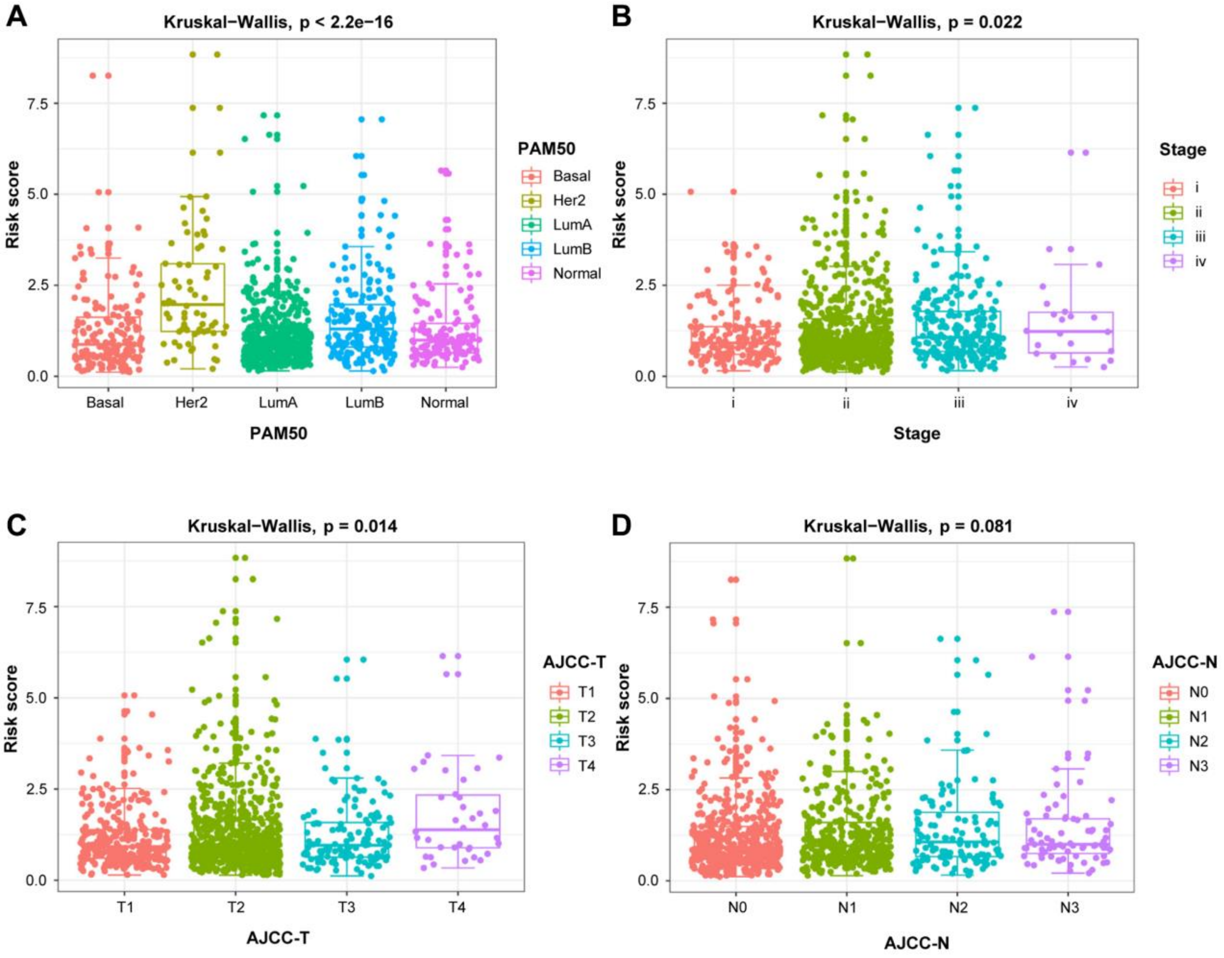


Figure 6. The relationship between the risk score and clinicopathological variables. Clinicopathological significance of the prognostic index of BRCA. (**A**) PAM50. (**B**) Stage. (**C**) T stage. (**D**) N stage.

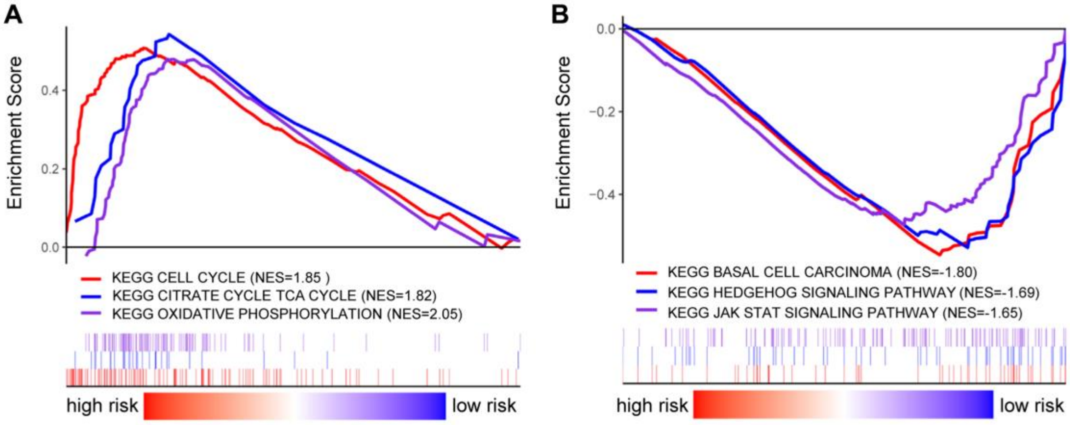


Figure 7. GSEA analysis in BRCA patients with high- and low-risk score. (**A**–**B**) GSEA displayed the KEGG enrichment pathways in BRCA patients with high- (**A**) and low-risk score (**B**).

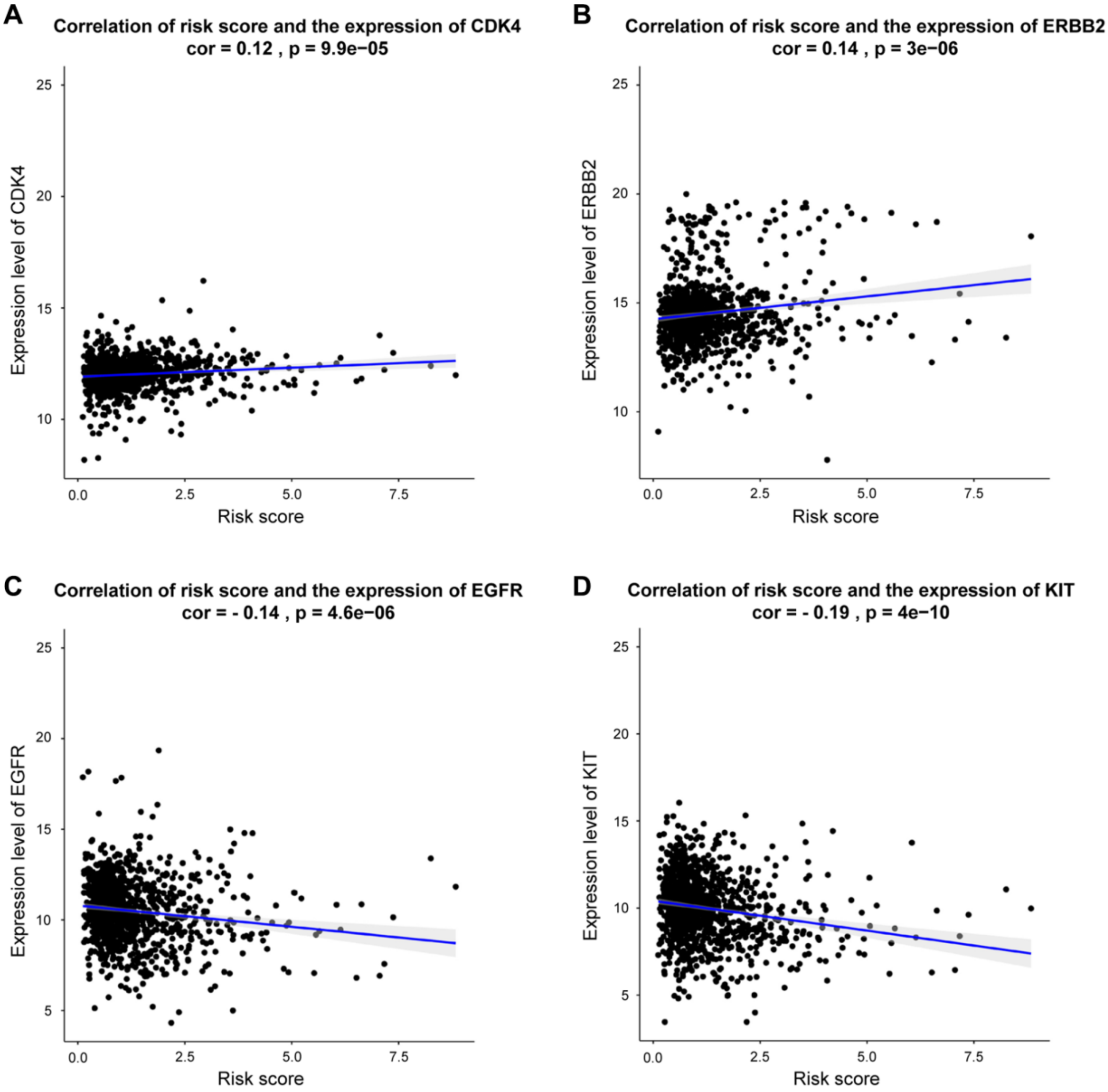


Figure 8. Correlation between risk score and genes expression for targeted treatment in BRCA. (**A**–**D**) Correlation analysis shows the results of CDK4, ERBB2, EGFR, KIT, respectively.

TOP2A, the better prognosis of BRCA [21]. Also, TTK was another promising therapeutic target due to it was overexpressed in BRCA [22]. UBE2T activated PI3K/Akt signaling pathway and stimulated tumor progression [23]. A recent study proposed that FREM1 isoform was an effective diagnostic and therapeutic marker for BRCA [24]. In certain cancer cells, binding of IGF-1 to IGF-1R could activate some signaling pathway and then promote oncogenic effect [25, 26]. SORBS1 was an adaptor protein involved in cell adhesion, growth factor signaling, and cancer metastasis [27, 28]. Next, our risk model was ascertained to independently predict prognosis in BRCA patients. As the risk score increased, the prognosis outcomes got worsen. The resulting DEGs-related risk model exhibited favorable predictive outcomes in both training (AUC at 5 year = 74%) and validation (AUC at 5 year = 73%) cohorts.

Furthermore, the 12 DEGs-related prognostic model was validated as an effective and excellent indicator of patients' tumor status and prognostic outcomes. Using our risk model, patients with particular clinicopathological features can be stratified into subgroups with varying clinical outcomes. In combination with these results, a nomogram was established by incorporating clinical features and risk score for DEGs signature, which presented excellent performance in survival prediction for BRCA patients. GSEA revealed that inhibited cell cycle is associated with better outcomes, suggesting the critical role of CDK 4/6 inhibitors for the prolonged survival time among BRCA patients. Moreover, the DEGs-related risk score significantly correlated with four targeted therapy genes, providing potential guidance for personalized treatments.

Several prognostic risk models have been previously developed, for example, Zhao et al. recently reported BRCA patients from TCGA cohorts and discussed the immune-related genes and the immune microenvironment of BRCA [5]. Lin et al. identified an autophagy-related genes prognostic model [6]. In our study, a DEGs-related signature was constructed through the different regions and races databases to predict BRCA patients' prognosis, as well as to evaluate the essence in BRCA in a comprehensive manner. The development and progression of BRCA were elaborated from multiple perspectives of differentially expressed genes, and few DEGs have been identified and verified to have potential application in clinics [29, 30]. However, even though external verification was performed in this study, validation of other cohorts is still necessary to verify the prognostic risk model's performance and efficiency. Lastly, the inevitable and inherent bias within the retrospective method should also be addressed.

In conclusion, a DEGs-related risk model was successfully constructed for predicting prognosis in BRCA, which is beneficial for patients and clinical researchers. Our systematic and comprehensive studies suggest that 12-DEGs signature might offer a more accurate evaluation system for BRCA patients' prognosis and provide more personalized therapies. In the future, more extended researches should be carried out to explore the possible mechanisms for the prediction of genes function, as well as the constitutions of the prognostic signature.

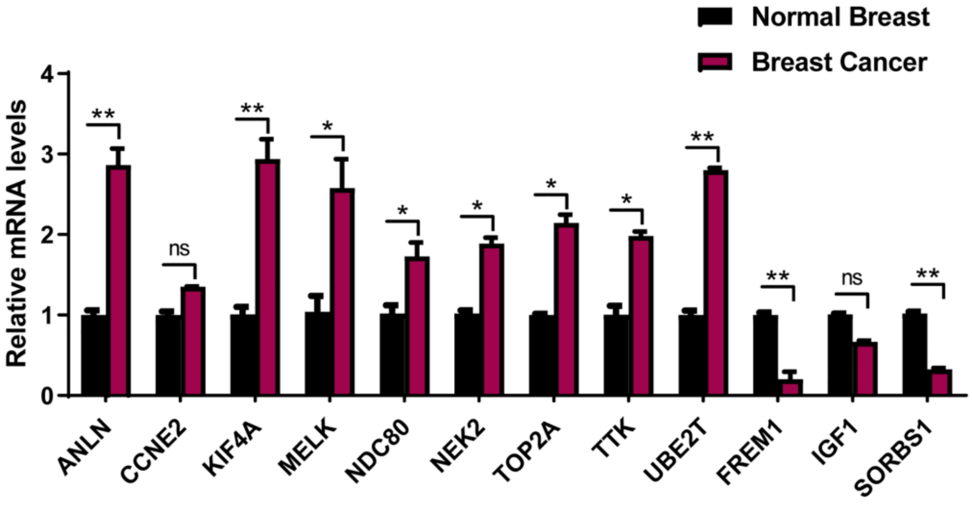


Figure 9. The expression of the 12 prognostic genes in clinical breast cancer samples. Clinical samples of breast carcinoma tissues and normal breast tissues were collected and followed by RNA extraction for qRT-PCR measurement of relative gene expression of ANLN, CCNE2, KIF4A, MELK, NDC80, NEK2, TOP2A, TTK, UBE2T FREM1, IGF1 and SORBS1. Data are means ± SEM. ns, denotes not significant. $^{*}P < 0.05$. $^{**}P < 0.001$.

MATERIALS AND METHODS

Data sources

Firstly, a flow chat was used to illustrate the entire process of our studies (Supplementary Figure 5). A total of 5 GEO datasets were used to identify DEGs in this study. GSE29431, GSE32641, GSE61304, GSE70947, and GSE86374 were downloaded from NCBI-GEO, a free and public database of transcriptional expression. GSE29431 data was obtained with the GPL570 platforms and collected from 54 breast tumors and 12 non-tumor breast tissue samples. GPL887 platforms were utilized to obtain GSE32641 data from 95 breast tumors and 7 normal breast samples. GSE61304 data was obtained with the GPL570 platforms and collected from 58 breast tumors and 4 normal breast tissues. GSE70947 data was obtained with the GPL13607 platforms and collected from 148 breast tumors and 148 normal breast samples. GSE86374 data was obtained with the GPL6422 platforms and collected from 124 breast tumors and 35 normal breast tissues. RNA-seq and survival information of TCGA-BRCA cohorts were retrieved from UCSC Xena [31]. GSE20685 and GSE 48390 data, including 408 BRCA samples from GEO datasets, were used for external validation. Details of the GEO datasets was shown in the Supplementary Table 2.

Identification of DEGs

Limma, a package that allows users to compare multiple databases in the GEO series under the R environment [32]. DEGs between BC and non-tumor breast tissue samples are identified. Removing the invalid genes, absolute log2 fold change > 1 and adjusted $p < 0.05$ were confirmed as threshold criteria for the genes, to further identify significant DEGs. The volcano plots, Venn diagrams, and heatmap were made by TBtools, and the overlapping DEGs were used to delve deeper [33].

Functional enrichment analysis of DEGs

Significant DEGs in BRCA were analysed by the "clusterProfiler" package [34], including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis [35, 36]. The biological processes, cell components and molecular functions were evaluated. Significant signaling pathways were statistically identified by KEGG analysis, where $p < 0.05$ and adjusted $p < 0.05$ were applied.

Construction of the PPI network

In this study, STRING online database (version 11.0) was utilized to construct the protein-protein interaction (PPI) network [37]. Cytoscape software (version 3.7.2) was employed to paint the integrated regulatory networks and analyzed the interaction network of different genes [38]. Using the MCODE plug-in component of the Cytoscape software, sorted and filtered critical modules of the whole network [39].

Construction and validation of OS related prognostic risk model

Univariate Cox regression analysis was firstly carried out to obtain the prognostic DEGs in BRCA, the risk model was subsequently developed using multivariate Cox regression analysis. Final optimization was incorporated into the prognostic risk model via stepwise regression. The resulting risk score was obtained using this formula:

$$\text{Risk score} = \sum_{i=1}^{n} Coef_i \times x_i$$

where n denotes the number of genes, Coef is the risk coefficient, and x indicates the expression level of individual DEGs in this risk model. In the light of the calculation formula of risk score, the median risk score was critical to divide patients into low-risked or high-risked category. Kaplan-Meier survival analysis was performed to differentiate survival rate between the high-risked and low-risked patients via log-rank test. Followed by the time-ROC analysis, the accurateness of risk model forecast was investigated [40], and verified in the independent BRCA cohorts (GES20685). In the training and validation cohorts, the identical calculation formula was used. Dual Cox regression analysis and clinic correlation analysis were combined to evaluate the effect of risk signature in predicting prognoses of the patients with BRCA.

Breast cancer gene-expression miner

The expression of 12 selected DEGs in different subtypes of BRCA were analyzed using the bcGenExMiner (version 4.5), by correlating that with the co-expressed genes [41]. The correlation of 12 selected genes was generated using the correlation module.

Nomogram construction and validation

The "rms" package was used to establish a nomogram in R, which can predict patients' prognosis by combining the risk score and multiple clinico-pathological factors [42]. As for the assessment of predictive accuracy of the model, Calibration curves were established, the concordance index (C-index) and Akaike information criterion (AIC) were further conducted to evaluate the influence of prognosis factors.

Gene set enrichment analysis (GSEA)

GSEA4.0.3 (https://www.gsea-msigdb.org/gsea/index.jsp) was used to detect which pathways genes are primarily enriched, that is, gene set enrichment analysis differences satisfying the nominal $p < 0.05$ and the FDR < 0.25 were considered statistically significant [43].

Samples and clinicopathological data

A total of 20 surgically resected breast cancer specimens and adjacent breast tissue were collected from the Second Hospital of Dalian Medical University between January 2010 and January 2018. There are 5 patients involved in each subtypes (luminal A, luminal B, HER2+, and TNBC), identified via pathological examination. No chemotherapeutic or radiotherapeutic treatments have been applied on patients before the surgery. All procedures in this research protocol were approved by the ethics committee in the Second Hospital of Dalian Medical University.

Isolation of RNA and quantitative reverse transcriptase PCR quantification

Extraction of total RNA and synthesis of complementary DNA was performed by referring to the manufacturers' instructions. TransStart Tip Green qPCR SuperMix (Transgen Biotech) was utilized for real-time qRT-PCR with specific primers against ANLN, CCNE2, KIF4A, MELK, NDC80, NEK2, TOP2A, TTK, UBE2T FREM1, IGF1, SORBS1, and glyceraldehyde-3-phosphate dehydrogenase (GAPDH), using the ABI 7900HT FAST Real-Time PCR System (Applied Biosystems, USA). GAPDH was selected for normalization. The primer sequences were shown in Supplementary Table 3.

Statistical analysis

The Wilcox test was proceeded to identify genes' expression levels of BRCA tissues against that of normal tissues. DEGs in the prognostic risk signature were screened via Cox regression analyses. The log-rank test was performed to correlate OS related Kaplan–Meier survival curve. Time-dependent ROC curve was analysed using the "timeROC" package. Step-comparison for internal-group and external-group was conducted using Mann–Whitney–Wilcoxon test and Kruskal-Wallis test, respectively. The c-index was applied to represent the prognostic nomogram for 5, and 10 years. Statistical significance was considered when two-sided p-values were smaller than 0.05, using R software (version 3.6.2).

AUTHOR CONTRIBUTIONS

J Li, G Huang, Z Zhao and M Li designed the study. J Li, G Huang and C Ren collected and analyzed the databases. J Li, N Wang, and S Sui organized and analyzed the statistics. J Li and G Huang wrote and revised the manuscript. All authors read and approved the manuscript and the submitted version.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

FUNDING

This study was funded by the Natural Science Foundation of Liaoning Province NO.2019-MS-098 to G. Huang, and National Nature Science Foundation of China (NSFC) No.81872156 to M. Li.

Editorial note

&This corresponding author has a verified history of publications using a personal email address for correspondence.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021; 71:209–49. https://doi.org/10.3322/caac.21660 PMID:33538338

2. Park YH, Lee SJ, Cho EY, Choi Y, Lee JE, Nam SJ, Yang JH, Shin JH, Ko EY, Han BK, Ahn JS, Im YH. Clinical relevance of TNM staging system according to breast cancer subtypes. Ann Oncol. 2011; 22:1554–60. https://doi.org/10.1093/annonc/mdq617 PMID:21242587

3. Liu Y, Wu L, Ao H, Zhao M, Leng X, Liu M, Ma J, Zhu J. Prognostic implications of autophagy-associated gene signatures in non-small cell lung cancer. Aging (Albany NY). 2019; 11:11440–62. https://doi.org/10.18632/aging.102544 PMID:31811814

4. Wang S, Zhang Q, Yu C, Cao Y, Zuo Y, Yang L. Immune cell infiltration-based signature for prognosis and immunogenomic analysis in breast cancer. Brief Bioinform. 2021; 22:2020–31. https://doi.org/10.1093/bib/bbaa026 PMID:32141494

5. Zhao Y, Pu C, Liu Z. Exploration the Significance of a Novel Immune-Related Gene Signature in Prognosis and Immune Microenvironment of Breast Cancer. Front Oncol. 2020; 10:1211. https://doi.org/10.3389/fonc.2020.01211

PMID:32850356

6. Lin QG, Liu W, Mo YZ, Han J, Guo ZX, Zheng W, Wang JW, Zou XB, Li AH, Han F. Development of prognostic index based on autophagy-related genes analysis in breast cancer. Aging (Albany NY). 2020; 12:1366–76. https://doi.org/10.18632/aging.102687 PMID:31967976

7. Lahouel K, Younes L, Danilova L, Giardiello FM, Hruban RH, Groopman J, Kinzler KW, Vogelstein B, Geman D, Tomasetti C. Revisiting the tumorigenesis timeline with a data-driven generative model. Proc Natl Acad Sci U S A. 2020; 117:857–64. https://doi.org/10.1073/pnas.1914589117 PMID:31882448

8. Cook DJ, Kallus J, Jörnsten R, Nielsen J. Molecular natural history of breast cancer: Leveraging transcriptomics to predict breast cancer progression and aggressiveness. Cancer Med. 2020; 9:3551–62. https://doi.org/10.1002/cam4.2996 PMID:32207233

9. Huynh MM, Pambid MR, Jayanthan A, Dorr A, Los G, Dunn SE. The dawn of targeted therapies for triple negative breast cancer (TNBC): a snapshot of investigational drugs in phase I and II trials. Expert Opin Investig Drugs. 2020; 29:1199–208. https://doi.org/10.1080/13543784.2020.1818067 PMID:32869671

10. Robert M, Frenel JS, Bourbouloux E, Rigaud DB, Patsouris A, Augereau P, Gourmelon C, Campone M. An Update on the Clinical Use of CDK4/6 Inhibitors in Breast Cancer. Drugs. 2018; 78:1353–62. https://doi.org/10.1007/s40265-018-0972-9 PMID:30143968

11. Dai X, Mei Y, Chen X, Cai D. ANLN and KDR Are Jointly Prognostic of Breast Cancer Survival and Can Be Modulated for Triple Negative Breast Cancer Control. Front Genet. 2019; 10:790. https://doi.org/10.3389/fgene.2019.00790 PMID:31636652

12. Magnusson K, Gremel G, Rydén L, Pontén V, Uhlén M, Dimberg A, Jirström K, Pontén F. ANLN is a prognostic biomarker independent of Ki-67 and essential for cell cycle progression in primary breast cancer. BMC Cancer. 2016; 16:904. https://doi.org/10.1186/s12885-016-2923-8 PMID:27863473

13. Lee C, Fernandez KJ, Alexandrou S, Sergio CM, Deng N, Rogers S, Burgess A, Caldon CE. Cyclin E2 Promotes Whole Genome Doubling in Breast Cancer. Cancers (Basel). 2020; 12:2268. https://doi.org/10.3390/cancers12082268 PMID:32823571

14. Li TF, Zeng HJ, Shan Z, Ye RY, Cheang TY, Zhang YJ, Lu SH, Zhang Q, Shao N, Lin Y. Overexpression of kinesin superfamily members as prognostic biomarkers of breast cancer. Cancer Cell Int. 2020; 20:123. https://doi.org/10.1186/s12935-020-01191-1 PMID:32322170

15. Zhu C, Jiang W. Cell cycle-dependent translocation of PRC1 on the spindle by Kif4 is essential for midzone formation and cytokinesis. Proc Natl Acad Sci U S A. 2005; 102:343–48. https://doi.org/10.1073/pnas.0408438102 PMID:15625105

16. Gong X, Chen Z, Han Q, Chen C, Jing L, Liu Y, Zhao L, Yao X, Sun X. Sanguinarine triggers intrinsic apoptosis to suppress colorectal cancer growth through disassociation between STRAP and MELK. BMC Cancer. 2018; 18:578. https://doi.org/10.1186/s12885-018-4463-x PMID:29783958

17. Ding X, Duan H, Luo H. Identification of Core Gene Expression Signature and Key Pathways in Colorectal Cancer. Front Genet. 2020; 11:45. https://doi.org/10.3389/fgene.2020.00045 PMID:32153633

18. Bièche I, Vacher S, Lallemand F, Tozlu-Kara S, Bennani H, Beuzelin M, Driouch K, Rouleau E, Lerebours F, Ripoche H, Cizeron-Clairac G, Spyratos F, Lidereau R. Expression analysis of mitotic spindle checkpoint genes in breast carcinoma: role of NDC80/HEC1 in early breast tumorigenicity, and a two-gene signature for aneuploidy. Mol Cancer. 2011; 10:23. https://doi.org/10.1186/1476-4598-10-23 PMID:21352579

19. Deb B, Sengupta P, Sambath J, Kumar P. Bioinformatics Analysis of Global Proteomic and Phosphoproteomic Data Sets Revealed Activation of NEK2 and AURKA in Cancers. Biomolecules. 2020; 10:237. https://doi.org/10.3390/biom10020237 PMID:32033228

20. Rellos P, Ivins FJ, Baxter JE, Pike A, Nott TJ, Parkinson DM, Das S, Howell S, Fedorov O, Shen QY, Fry AM, Knapp S, Smerdon SJ. Structure and regulation of the human Nek2 centrosomal kinase. J Biol Chem. 2007; 282:6833–42. https://doi.org/10.1074/jbc.M609721200 PMID:17197699

21. Fasching PA, Weihbrecht S, Haeberle L, Gasparyan A, Villalobos IE, Ma Y, Ekici AB, Wachter DL, Hartmann A, Beckmann MW, Slamon DJ, Press MF. HER2 and TOP2A amplification in a hospital-based cohort of breast cancer patients: associations with patient and tumor characteristics. Breast Cancer Res Treat. 2014;

145:193–203.
https://doi.org/10.1007/s10549-014-2922-x
PMID:24682655

22. King JL, Zhang B, Li Y, Li KP, Ni JJ, Saavedra HI, Dong JT. TTK promotes mesenchymal signaling via multiple mechanisms in triple negative breast cancer. Oncogenesis. 2018; 7:69.
https://doi.org/10.1038/s41389-018-0077-z
PMID:30206215

23. Wang Y, Leng H, Chen H, Wang L, Jiang N, Huo X, Yu B. Knockdown of UBE2T Inhibits Osteosarcoma Cell Proliferation, Migration, and Invasion by Suppressing the PI3K/Akt Signaling Pathway. Oncol Res. 2016; 24:361–69.
https://doi.org/10.3727/096504016X14685034103310
PMID:27712593

24. Kashem MA, Li H, Toledo NP, Omange RW, Liang B, Liu LR, Li L, Yang X, Yuan XY, Kindrachuk J, Plummer FA, Luo M. Toll-like Interleukin 1 Receptor Regulator Is an Important Modulator of Inflammation Responsive Genes. Front Immunol. 2019; 10:272.
https://doi.org/10.3389/fimmu.2019.00272
PMID:30873160

25. Zhu Y, Wang T, Wu J, Huang O, Zhu L, He J, Li Y, Chen W, Chen X, Shen K. Associations Between Circulating Insulin-Like Growth Factor 1 and Mortality in Women With Invasive Breast Cancer. Front Oncol. 2020; 10:1384.
https://doi.org/10.3389/fonc.2020.01384
PMID:32974138

26. Christopoulos PF, Msaouel P, Koutsilieris M. The role of the insulin-like growth factor-1 system in breast cancer. Mol Cancer. 2015; 14:43.
https://doi.org/10.1186/s12943-015-0291-7
PMID:25743390

27. Brandsma CA, Guryev V, Timens W, Ciconelle A, Postma DS, Bischoff R, Johansson M, Ovchinnikova ES, Malm J, Marko-Varga G, Fehniger TE, van den Berge M, Horvatovich P. Integrated proteogenomic approach identifying a protein signature of COPD and a new splice variant of SORBS1. Thorax. 2020; 75:180–83.
https://doi.org/10.1136/thoraxjnl-2019-213200
PMID:31937552

28. Lin WH, Chiu KC, Chang HM, Lee KC, Tai TY, Chuang LM. Molecular scanning of the human sorbin and SH3-domain-containing-1 (SORBS1) gene: positive association of the T228A polymorphism with obesity and type 2 diabetes. Hum Mol Genet. 2001; 10:1753–60.
https://doi.org/10.1093/hmg/10.17.1753
PMID:11532984

29. Zhang B, Chen MY, Shen YJ, Zhuo XB, Gao P, Zhou FS, Liang B, Zu J, Zhang Q, Suleman S, Xu YH, Xu MG, Xu JK, et al. A Large-Scale, Exome-Wide Association Study of Han Chinese Women Identifies Three Novel Loci Predisposing to Breast Cancer. Cancer Res. 2018; 78:3087–97.
https://doi.org/10.1158/0008-5472.CAN-17-1721
PMID:29572226

30. Bownes RJ, Turnbull AK, Martinez-Perez C, Cameron DA, Sims AH, Oikonomidou O. On-treatment biomarkers can improve prediction of response to neoadjuvant chemotherapy in breast cancer. Breast Cancer Res. 2019; 21:73.
https://doi.org/10.1186/s13058-019-1159-3
PMID:31200764

31. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, Banerjee A, Luo Y, Rogers D, Brooks AN, Zhu J, Haussler D. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol. 2020; 38:675–78.
https://doi.org/10.1038/s41587-020-0546-8
PMID:32444850

32. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43:e47.
https://doi.org/10.1093/nar/gkv007
PMID:25605792

33. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol Plant. 2020; 13:1194–202.
https://doi.org/10.1016/j.molp.2020.06.009
PMID:32585190

34. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012; 16:284–87.
https://doi.org/10.1089/omi.2011.0118
PMID:22455463

35. Lu Y, Rosenfeld R, Simon I, Nau GJ, Bar-Joseph Z. A probabilistic generative model for GO enrichment analysis. Nucleic Acids Res. 2008; 36:e109.
https://doi.org/10.1093/nar/gkn434
PMID:18676451

36. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017; 45:D353–61.
https://doi.org/10.1093/nar/gkw1092
PMID:27899662

37. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-

wide experimental datasets. Nucleic Acids Res. 2019; 47:D607–13.
https://doi.org/10.1093/nar/gky1131
PMID:30476243

38. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13:2498–504.
https://doi.org/10.1101/gr.1239303
PMID:14597658

39. Bandettini WP, Kellman P, Mancini C, Booker OJ, Vasu S, Leung SW, Wilson JR, Shanbhag SM, Chen MY, Arai AE. MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study. J Cardiovasc Magn Reson. 2012; 14:83.
https://doi.org/10.1186/1532-429X-14-83
PMID:23199362

40. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics. 2000; 56:337–44.
https://doi.org/10.1111/j.0006-341x.2000.00337.x
PMID:10877287

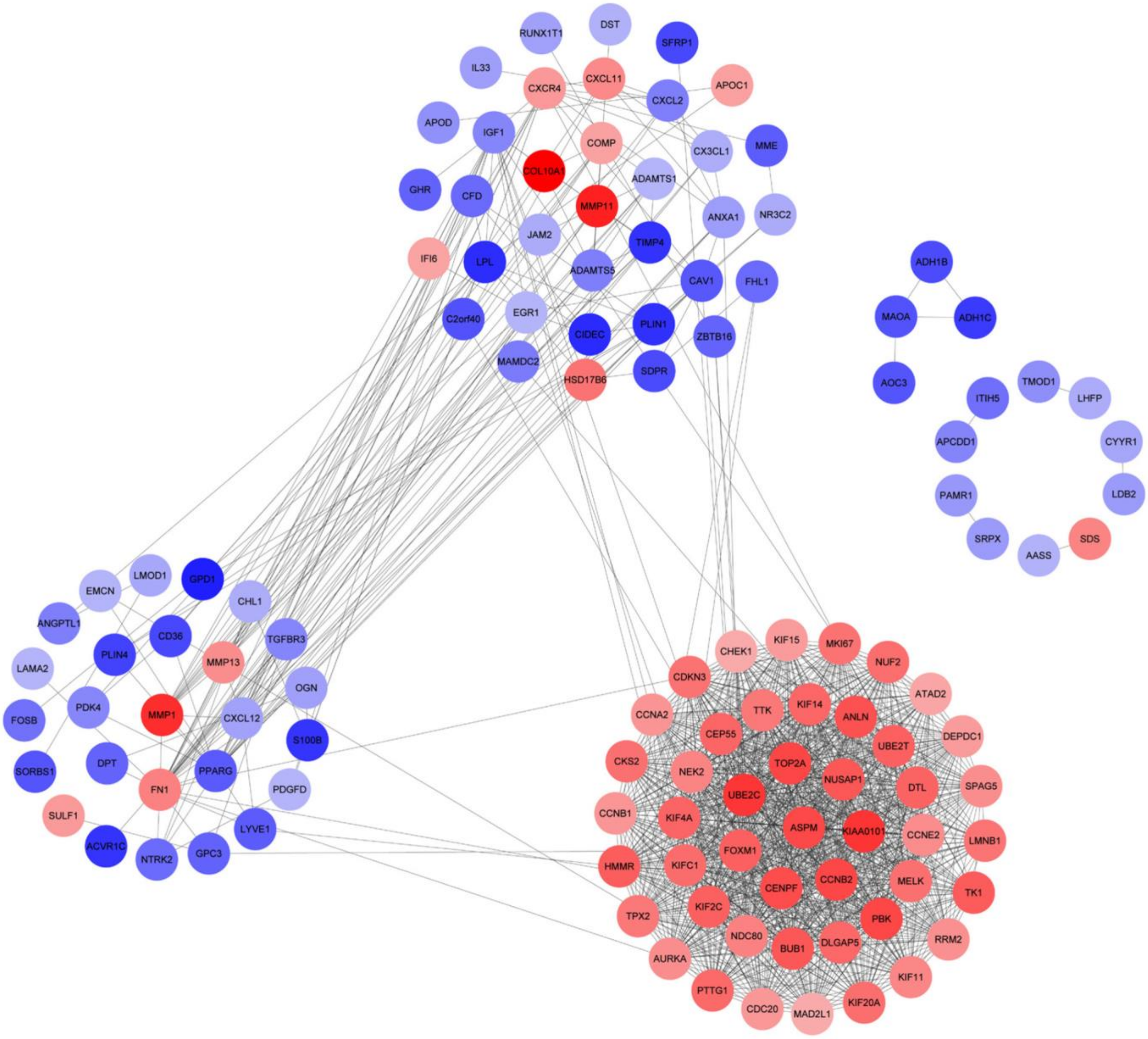
41. Jézéquel P, Frénel JS, Campion L, Guérin-Charbonnel C, Gouraud W, Ricolleau G, Campone M. bc-GenExMiner 3.0: new mining module computes breast cancer gene expression correlation analyses. Database (Oxford). 2013; 2013:bas060.
https://doi.org/10.1093/database/bas060
PMID:23325629

42. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. Lancet Oncol. 2015; 16:e173–80.
https://doi.org/10.1016/S1470-2045(14)71116-7
PMID:25846097

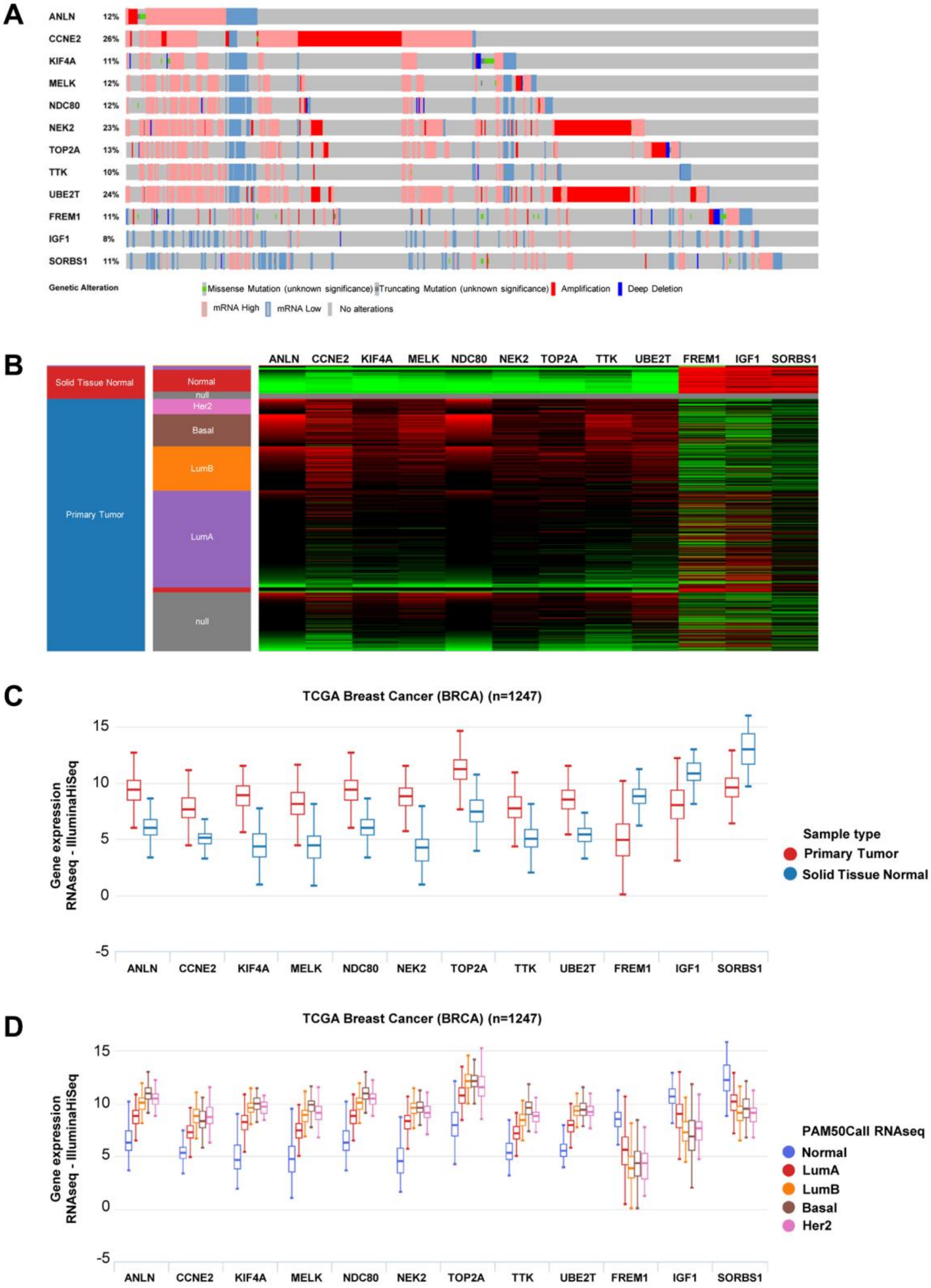
43. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102:15545–50.
https://doi.org/10.1073/pnas.0506580102
PMID:16199517

SUPPLEMENTARY MATERIALS

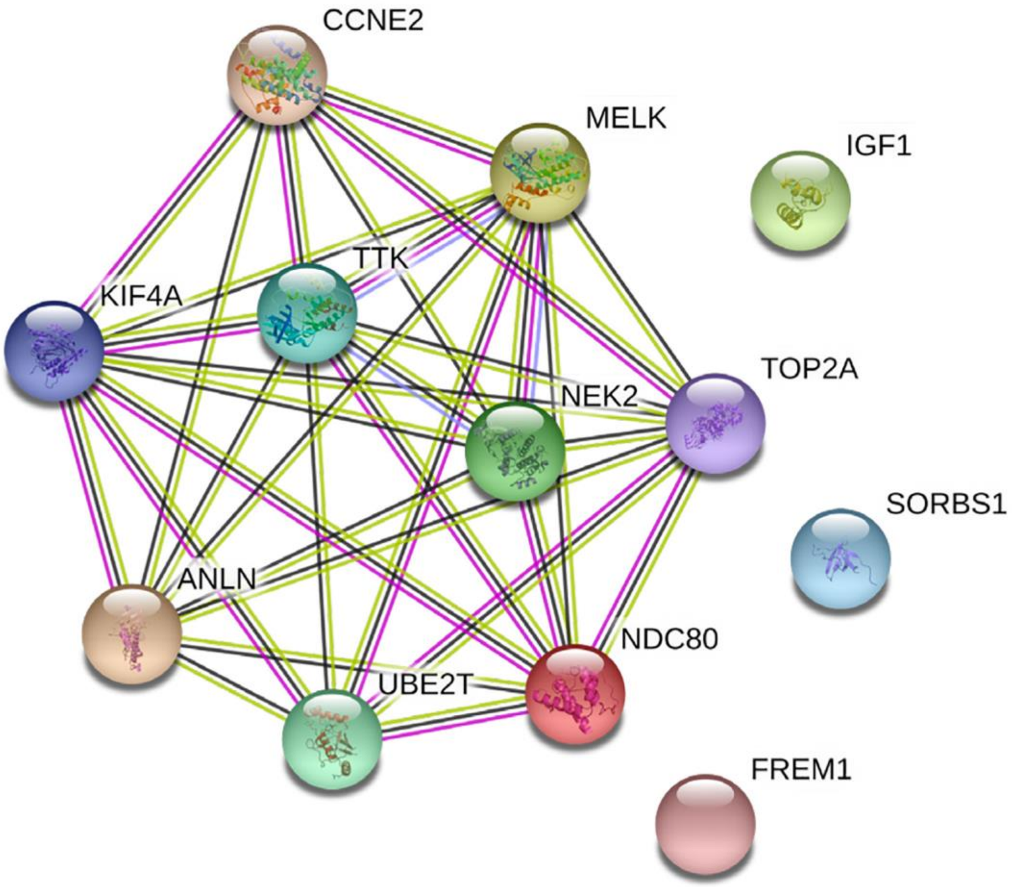
Supplementary Figures



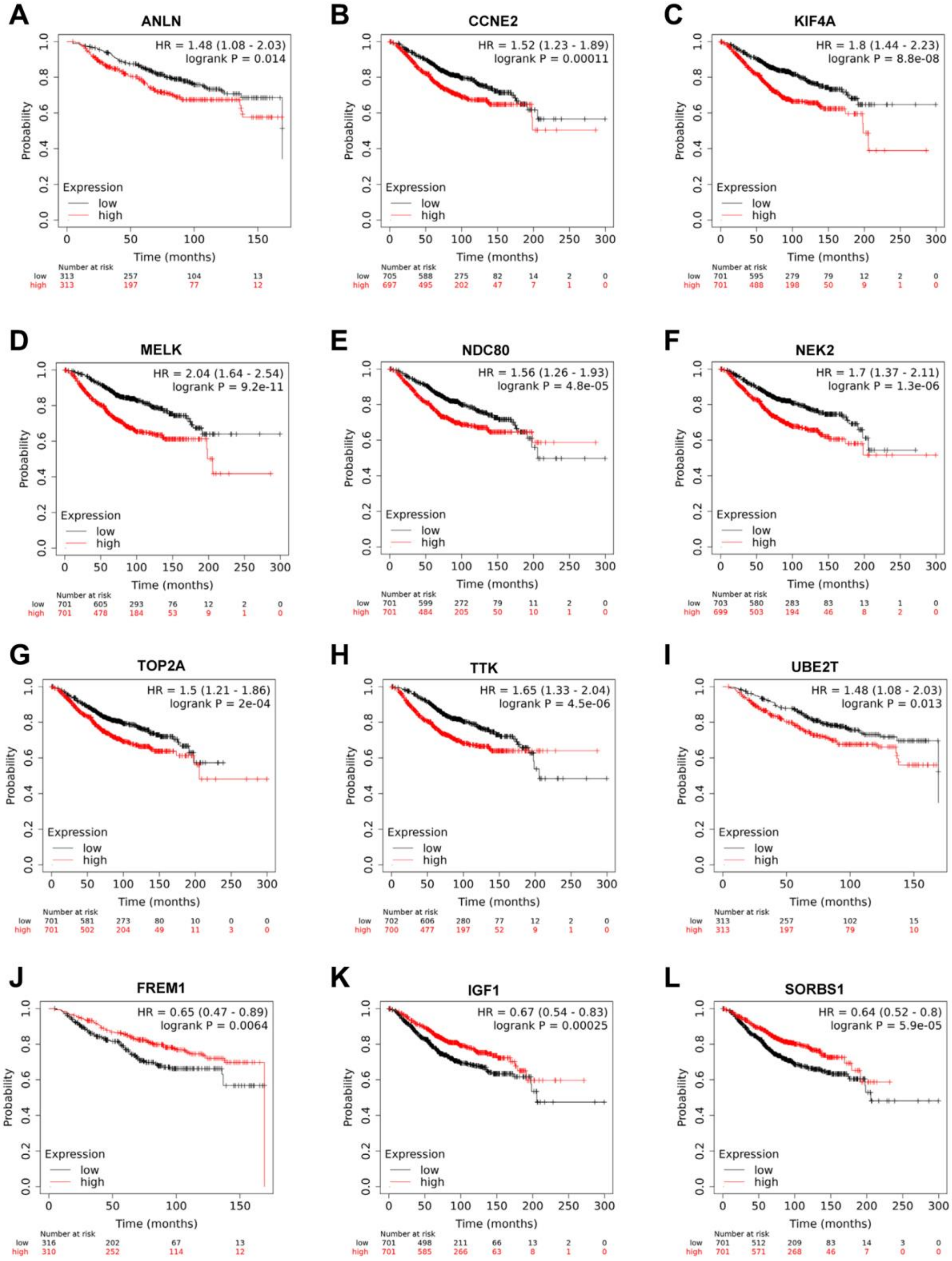
Supplementary Figure 1. PPI network of 151 DEGs was constructed in Cytoscape. Red node indicated the upregulated genes and blue node indicated the downregulated genes. The interaction relationship between nodes was connected by lines.



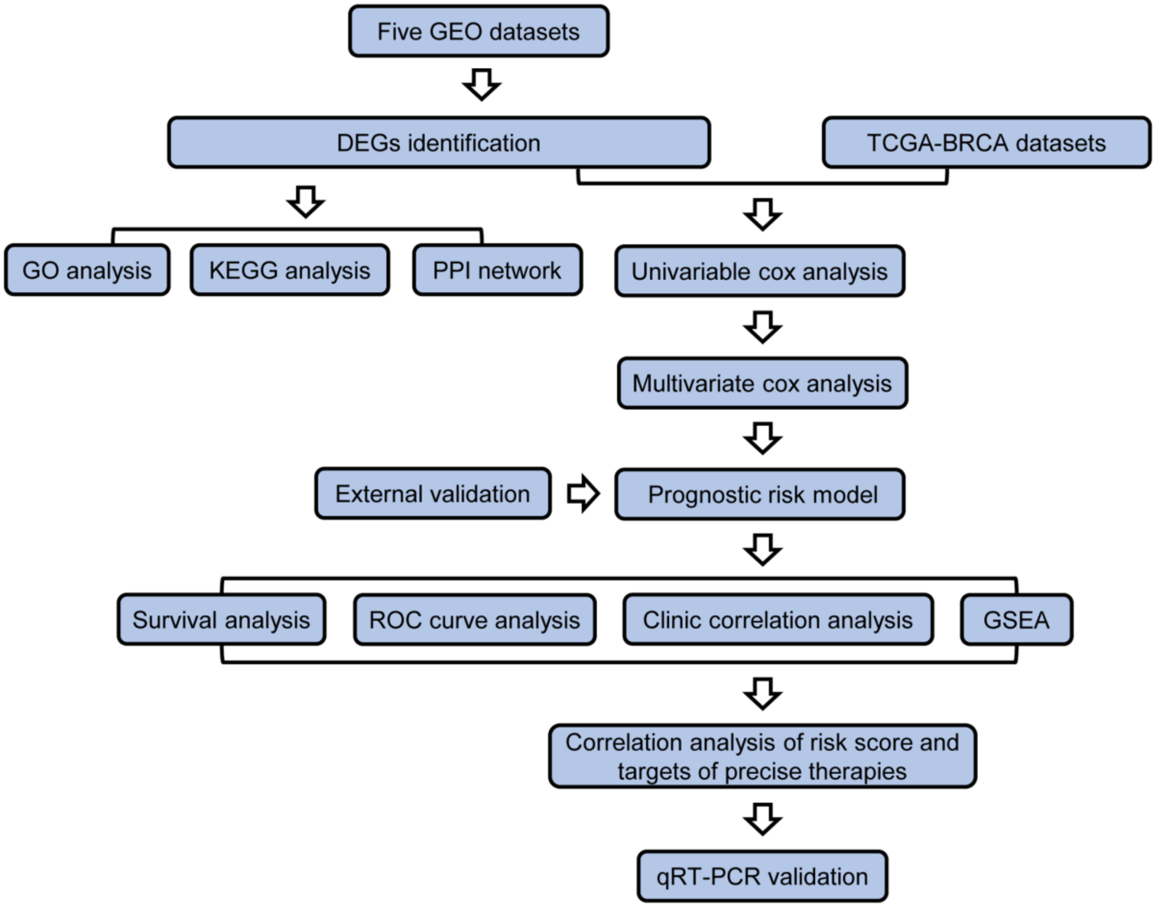
Supplementary Figure 2. Synthetical analyses of 12 selected genes in prognostic risk model. (**A**) OncoPrint demonstrating the copy number variations and mRNA expression alterations of 12 selected genes in prognostic risk model. (**B**) Heatmap of 12 selected genes expression across sample type and PAM50 subtypes in the TCGA-BRCA dataset obtained from the UCSC-Xena online tools. (**C**) Differential expression of the 12 selected genes between normal and BRCA tissues. Red and blue indicate malignant tissues and normal tissues, respectively. (**D**) Differential expression of the 12 selected genes among molecular subtypes of TCGA-BRCA.



Supplementary Figure 3. PPI analysis of 12 selected DEGs in prognostic risk model by the STRING online dataset.



Supplementary Figure 4. Kaplan-Meier survival curves of OS based on 12 genes expression in BRCA patients. (**A**–**I**). Expression level of ANLN, CCNE2, KIF4A, MELK, KNTC2, NEK2, TOP2A, TTK and UBE2T was markedly correlated with poor OS of BRCA patients. (**J**–**L**). Expression level of FREM1, IGF1, SORBS1 was markedly correlated with improved OS of BRCA patients.



Supplementary Figure 5. The flowchart of data analysis procedures.

Supplementary Tables

Supplementary Table 1. Univariate Cox regression analysis of DEGs.

Gene	Coef	*p*	HR	95%Cl	95%Cl
ANLN	0.422	0.013	1.525	1.093	2.129
CCNE2	0.597	0.001	1.817	1.293	2.554
KIF4A	0.461	0.006	1.585	1.139	2.206
MELK	0.411	0.015	1.508	1.082	2.103
NDC80	0.343	0.042	1.408	1.013	1.959
NEK2	0.425	0.011	1.53	1.101	2.126
TOP2A	0.341	0.043	1.406	1.01	1.958
TTK	0.447	0.008	1.564	1.124	2.176
UBE2T	0.365	0.029	1.441	1.037	2.002
SORBS1	-0.479	0.005	0.619	0.442	0.868
FREM1	-0.357	0.033	0.7	0.504	0.971
IGF1	-0.35	0.04	0.705	0.505	0.984
APOD	-0.347	0.038	0.707	0.509	0.981
LYVE1	0.456	0.008	1.577	1.124	2.212
BUB1	0.375	0.027	1.454	1.043	2.029
CCNA2	0.412	0.016	1.51	1.08	2.11
CCNB2	0.45	0.008	1.569	1.122	2.193
CDC20	0.353	0.036	1.423	1.023	1.978
CEP55	0.367	0.03	1.443	1.037	2.01
FOXM1	0.403	0.017	1.496	1.075	2.084
HSD17B6	0.349	0.036	1.417	1.023	1.964
KIAA0101	0.47	0.008	1.6	1.132	2.262
KIF2C	0.561	0.001	1.752	1.256	2.444
LMNB1	0.419	0.013	1.52	1.092	2.116
MAD2L1	0.374	0.027	1.454	1.043	2.028
MMP13	0.412	0.018	1.509	1.074	2.122
NUF2	0.394	0.019	1.484	1.067	2.063
SQLE	0.403	0.018	1.496	1.071	2.09
TK1	0.341	0.043	1.406	1.01	1.958
TPX2	0.4	0.018	1.492	1.071	2.078
UBE2C	0.469	0.006	1.598	1.146	2.227

Abbreviations: Coef: coefficient; HR: hazard ratio; CI: confidence interval.

Supplementary Table 2. Characteristics of GEO datasets included in the study.

Classification	Series accession ID	Country (region)	Number of samples		Platform ID
			Tumor	Normal	
Identification of DEGs	GSE29431	Spain	54	12	GPL570
	GSE32641	Taiwan	95	7	GPL887
	GSE61304	Singapore	58	4	GPL570
	GSE70947	USA	148	148	GPL13607
	GSE86374	Mexico	124	35	GPL6422
External verification cohorts	GSE20685	Taiwan	327	0	GPL570
	GSE48390	Taiwan	81	0	GPL570

Supplementary Table 3. Primer sequences for qRT-PCR.

Gene Name	Forward Primer	Reverse Primer
ANLN	5′- TGCCAGGCGAGAGAATCTTC -3′	5′- CGCTTAGCATGAGTCATAGACCT -3′
CCNE2	5′- TCAAGACGAAGTAGCCGTTTAC -3′	5′- TGACATCCTGGGTAGTTTTCCTC -3′
KIF4A	5′- TACTGCGGTGGAGCAAGAAG -3′	5′- CATCTGCGCTTGACGGAGAG -3′
MELK	5′- TATTCACCTCGATGATGATTGCG -3′	5′- AGAAAGCCTTAAACGAACTGGTT -3′
NDC80	5′- TCAAGGACCCGAGACCACTTA -3′	5′- GGGAGCTTGTAGAGATTTCATGG -3′
NEK2	5′- TGCTTCGTGAACTGAAACATCC -3′	5′- CCAGAGTCAACTGAGTCATCACT -3′
TOP2A	5′- ACCATTGCAGCCTGTAAATGA -3′	5′- GGGCGGAGCAAAATATGTTCC -3′
TTK	5′- GTGGAGCAGTACCACTAGAAATG -3′	5′- CCCAAGTGAACCGGAAAATGA -3′
UBE2T	5′- ATCCCTCAACATCGCAACTGT -3′	5′- CAGCCTCTGGTAGATTATCAAGC -3′
FREM1	5′- GCCTGTGGTAACCAGGAACAA -3′	5′- CGCAGGTGTATCAGGGTCG -3′
IGF1	5′- GCTCTTCAGTTCGTGTGTGGA -3′	5′- GCCTCCTTAGATCACAGCTCC -3′
SORBS1	5′- ATTCCCAAGCCTTTCCATCAG -3′	5′- TTTTGCTGTTCTCGATTGTGTTG -3′
GAPDH	5′- GGAGCGAGATCCCTCCAAAAT -3′	5′- GGCTGTTGTCATACTTCTCATGG -3′

Abbreviations: qRT-PCR: quantitative real-time polymerase chain reaction.