



# CRISPR-cas9 Screening Identified Lethal Genes Enriched in Cell Cycle Pathway and of Prognosis Significance in Breast Cancer

Xi Sun<sup>†</sup>, Zheng Wang<sup>†</sup>, Xiaosong Chen<sup>\*</sup> and Kunwei Shen<sup>\*</sup>

Department of General Surgery, Comprehensive Breast Health Center, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

## OPEN ACCESS

### Edited by:

Mingyan Zhu,  
Affiliated Hospital of Nantong  
University, China

### Reviewed by:

Shang Xie,  
Peking University, China  
Xinxiang Li,  
Fudan University, China

### \*Correspondence:

Kunwei Shen  
kwshen@medmail.com.cn  
Xiaosong Chen  
chenxiaosong0156@hotmail.com

<sup>†</sup>These authors have contributed  
equally to this work and share first  
authorship

### Specialty section:

This article was submitted to  
Molecular and Cellular Oncology,  
a section of the journal  
Frontiers in Cell and Developmental  
Biology

**Received:** 28 December 2020

**Accepted:** 15 February 2021

**Published:** 19 March 2021

### Citation:

Sun X, Wang Z, Chen X and  
Shen K (2021) CRISPR-cas9  
Screening Identified Lethal Genes  
Enriched in Cell Cycle Pathway  
and of Prognosis Significance  
in Breast Cancer.  
*Front. Cell Dev. Biol.* 9:646774.  
doi: 10.3389/fcell.2021.646774

**Background:** Lethal genes have not been systematically analyzed in breast cancer which may have significant prognostic value. The current study aims to investigate vital genes related to cell viability by analyzing the CRISPR-cas9 screening data, which may provide novel therapeutic target for patients.

**Methods:** Genes differentially expressed between tumor and normal tissue from the Cancer Genome Atlas (TCGA) and genes related to cell viability by CRISPR-cas9 screening from Depmap (Cancer Dependency Map) were overlapped. Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) analysis was conducted to identify which pathways of overlapped genes were enriched. GSE21653 set was randomized into training and internal validation dataset at a ratio of 3:1, and external validation was performed in GSE20685 set. The least absolute shrinkage and selection operator (LASSO) regression was used to construct a signature to predict recurrence-free survival (RFS) of breast cancer patients. Univariate and multivariate Cox regression were used to evaluate the prognostic value of this signature. Differentially expressed genes (DEGs) between high-risk and low-risk patients were then analyzed to identify the main pathways regulated by this signature. Weighted correlation network analysis (WGCNA) was conducted to recognize modules correlated with high risk. Enrichment analysis was then used to identify pathways regulated by genes shared in the overlapped genes, DEGs, and WGCNA.

**Results:** A total of 86 oncogenes were upregulated in TCGA database and overlapped with lethal genes in Depmap database, which were enriched in cell cycle pathway. A total of 51 genes were included in the gene signature based on LASSO regression, and the median risk score of 2.36 was used as cut-off to separate low-risk patients from high-risk patients. High-risk patients showed worse RFS compared with low-risk patients in internal training, internal validation, and external validation dataset. Time-dependent receiver operating characteristic curves of 3 and 5 years indicated that risk score was superior to tumor stage, age, and PAM50 in both entire and external validation datasets. Cell cycle was the main different pathway between the high-risk and low-risk groups.

Meanwhile, cell cycle was also the main pathway enriched in the 25 genes which were shared among 86 genes, DEGs, and WGCNA.

**Conclusion:** Cell cycle pathway, identified by CRISPR-cas9 screening, was a key pathway regulating cell viability, which has significant prognostic values and can serve as a new target for breast cancer patient treatment.

**Keywords:** CRISPR-cas9 screening, breast cancer, cell cycle, signature, cell viability

## INTRODUCTION

Breast cancer, with the highest incidence rate among female cancer, is the second leading cause of cancer-related death and poses a great threat to women's health (Siegel et al., 2020). It accounts for 30% of all new cases and is responsible for 15% of cancer deaths in women. Despite that the prognosis of patients has been improving over the past few years, the complex biological behaviors of breast cancer still hamper the progress in clinical treatment. Thus, understanding the specific vulnerability of breast carcinoma is of great importance.

Currently, CRISPR-cas9 screening is emerging as a powerful tool for precise medicine (Doudna and Charpentier, 2014; Kurata et al., 2018). Combining cas9 with pooled guide RNA libraries facilitates screening of genes that contribute to specific biologic phenotypes and diseases in a high-throughput way (Joung et al., 2017). This "phenotype-to-genotype" approach includes modifying expression of genes, selecting cells with a phenotype of interest, and sequencing the perturbation of interest, which allows for discovering genes related to cell viability (Schuster et al., 2019). Meanwhile, large-scale loss-of-function screening for cancer dependences have been performed in a variety of well-characterized cancer cell lines to assess the effect of single-gene knockout on cell viability (Meyers et al., 2017; Tsherniak et al., 2017). These data were deposited in the Cancer Dependence Map (DepMap) website.

Aberrant cell cycle is a hallmark of cancer (Hanahan and Weinberg, 2011). The evolution of cell cycle is conservative. Checkpoints have evolved to ensure that cell cycle progress is under sequential activation (Kastan and Bartek, 2004; Strzyz, 2016). Under the stimulation of mitogenic signal, cyclin-dependent kinases (CDKs) associate with cyclins and phosphorylate intracellular proteins that orchestrate cell cycle progress in a well-organized way (Malumbres and Barbacid, 2009; Malumbres, 2014). In cancer cells, aberrant signals are developed and promote the activation of CDK-cyclin complex. Deregulation of the cell cycle engine eventually leads to uncontrolled cell proliferation and genomic instability in cancer. Thus, the therapeutic potential of targeting the cell cycle has been increasingly concerned (Lim and Kaldis, 2013). In breast cancer, the application of CDK4/6 inhibitor transformed treatment landscape in estrogen (ER)-positive human epidermal growth factor receptor-2 (HER2) negative breast cancer. The improvement in prognosis indicated that targeting cell cycle is an essential way to cancer treatment (Ingham and Schwartz, 2017; Slamon et al., 2018; Tripathy et al., 2018).

Biological process involving cell viability is complex. However, cell vulnerability of breast cancer has not been systematically researched. Meanwhile, pathways and the prognostic significance of these genes have never been detailed. In the current study, we aimed to identify genes differentially expressed in tumor tissues and contributed to cell viability. Using these genes, a prediction model with prognostic significance was constructed and validated. The pathways and biological processes regulated by these genes were also evaluated.

## MATERIALS AND METHODS

### Identification of Viability Vulnerable

Dependence scores of breast cancer cell lines were downloaded from the Depmap dataset<sup>1</sup>, and this is the result from a series of loss-of-function genomic screening in different cell lines. Dependence score was calculated by CERES algorithm to identify genes essential to proliferation and survival (Meyers et al., 2017). A negative score of a gene indicates that knocking out of the gene inhibits the survival of a cell line, whereas a positive score indicates that knocking out of the gene promotes the survival and proliferation. Cut-offs of 0.5 and -0.5 were to define growth-suppressing genes and growth-promoting genes.

Read counts of breast cancer were downloaded from TCGA datasets<sup>2</sup>. Differentially expressed genes between tumor and normal patients were calculated based on the negative binomial distribution using DESeq2 package. Adjusted *p* value <0.05 and absolute fold change greater than 2 were used as cut-off to select differentially expressed genes. Growth-suppressing genes from Depmap were overlapped with downregulated genes from the TCGA dataset, and growth-promoting genes from Depmap were overlapped with upregulated genes from the TCGA dataset to select genes for further analysis.

### Data Processing

Raw data of GSE20685 and GSE21653 were downloaded from the GEO database<sup>3</sup>. These two datasets were both from [HG-U133\_Plus\_2] Affymetrix Human Genome U133 Plus 2.0 platform. The raw data of GSE20685 and GSE21653 were normalized by gcrma algorithm simultaneously. The probe ID was converted into gene symbol using the annotation platform.

<sup>1</sup><https://depmap.org/portal/download/>

<sup>2</sup><https://portal.gdc.cancer.gov/>

<sup>3</sup><http://www.ncbi.nlm.nih.gov/geo/>

When one probe was matched to the same gene, average gene expression of this gene was calculated.

### LASSO Regression Analysis

Patients from GSE21653 set were randomized into internal training dataset and internal validation dataset at a ratio of 3:1. The least absolute shrinkage and selection operator (LASSO)

model was used to remove genes of high correlation and a risk model was constructed (Tibshirani, 1997; Zhou et al., 2019). A risk score formula was established by integrating gene expression value weighted by their LASSO Cox coefficients. R package “glmnet” in R 3.5.2 was used to perform LASSO analysis (Ternès et al., 2016). Univariate and multivariate Cox regression analysis was used to assess the prognostic value of risk score in

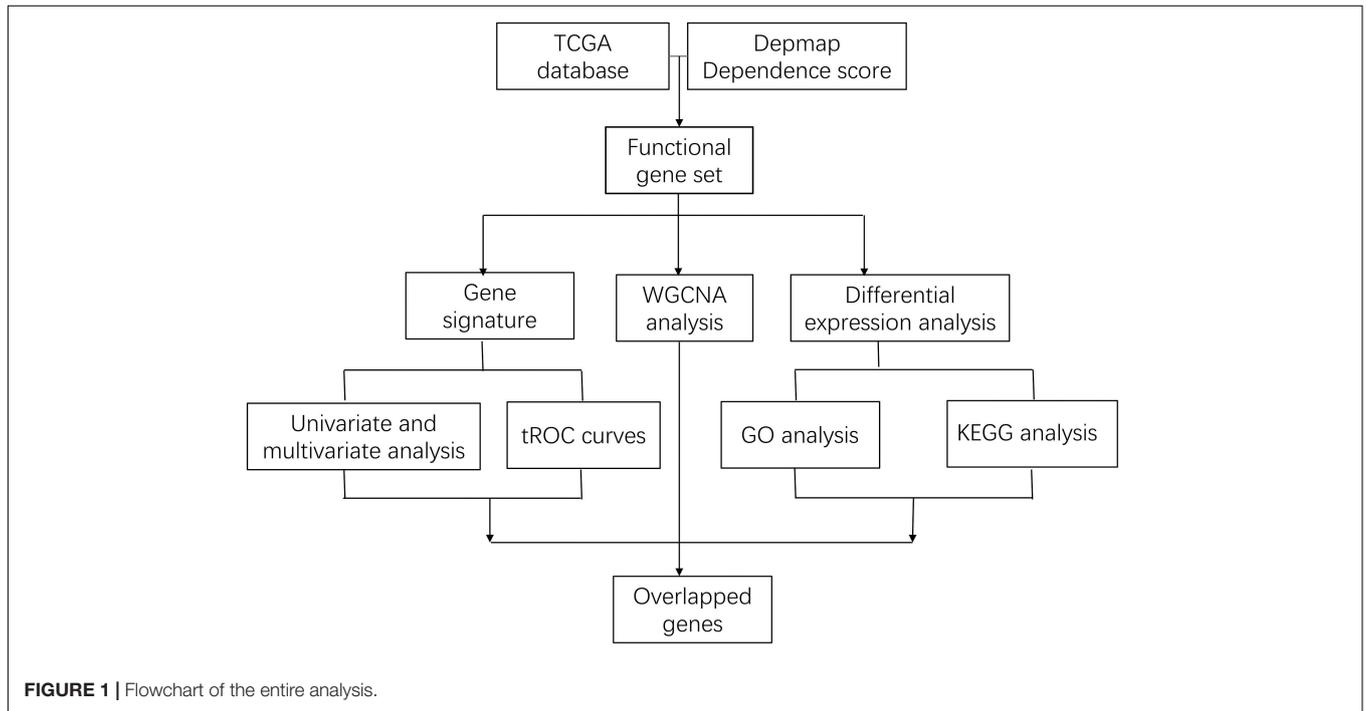


FIGURE 1 | Flowchart of the entire analysis.

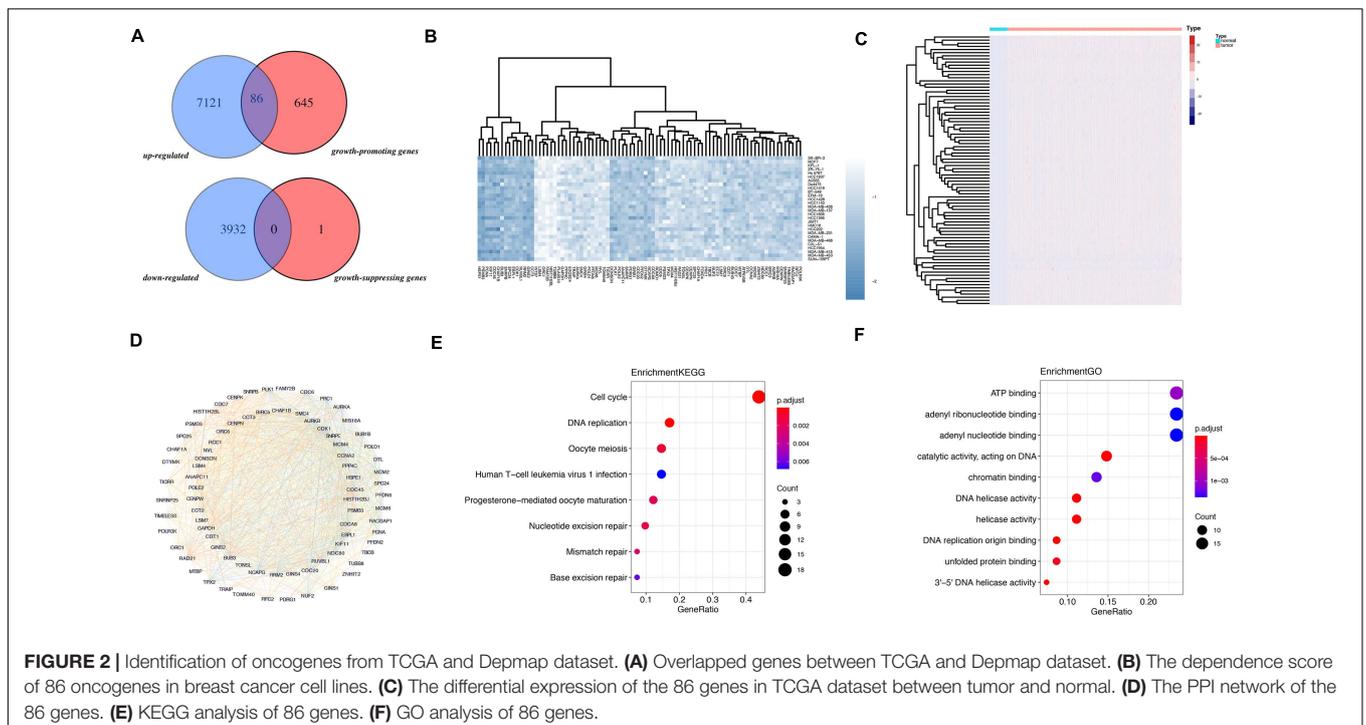


FIGURE 2 | Identification of oncogenes from TCGA and Depmap dataset. (A) Overlapped genes between TCGA and Depmap dataset. (B) The dependence score of 86 oncogenes in breast cancer cell lines. (C) The dependence score of 86 genes in TCGA dataset between tumor and normal. (D) The PPI network of the 86 genes. (E) KEGG analysis of 86 genes. (F) GO analysis of 86 genes.

entire dataset and external validation dataset. Time-dependent receiver operating characteristic (tROC) curves were used to compare the prediction accuracy of risk score with traditional clinicopathological parameters. “survivalROC” package was used to plot tROC curve and calculate Area under curve (AUC).

## Differentially Expressed Genes

In dataset GSE21653, limma package was used to calculate differentially expressed genes between high-risk group and low-risk group patients.  $P$  value  $<0.05$  and absolute fold change greater than 1.5 were defined as DEGs.

## Enrichment Analysis

Gene Ontology (GO) was used to annotate biological processes, molecular functions, and cellular components of genes. Kyoto Encyclopedia of Genes and Genomes (KEGG) was used to annotate the gene pathways (Gene Ontology Consortium, 2015; Kanehisa et al., 2017). GO and KEGG analysis was performed using clusterProfiler package (Yu et al., 2012).  $P$  value  $<0.05$  was considered as significant pathways enriched.

## PPI Network Construction and Hub Gene Identification

STRING<sup>4</sup> website was used to discover known and predicted protein-protein interactions, as well as to construct a PPI

<sup>4</sup><http://string-db.org>

network. The Cytoscape software was then employed to visualize the interactive relationship of the overlapped genes.

## Weighted Correlation Network Analysis

Weighted Correlation Network Analysis (WGCNA) was performed to find modules of highly correlated genes using WGCNA package (Langfelder and Horvath, 2008). A One-step network construction was used to construct network and modules were identified. Eigengenes were correlated with external traits to identify modules that are significantly associated with the measured clinical traits. A scatterplot of Gene Significance (GS) versus Module Membership (MM) in different modules was plotted to show the correlation of GS and MM.

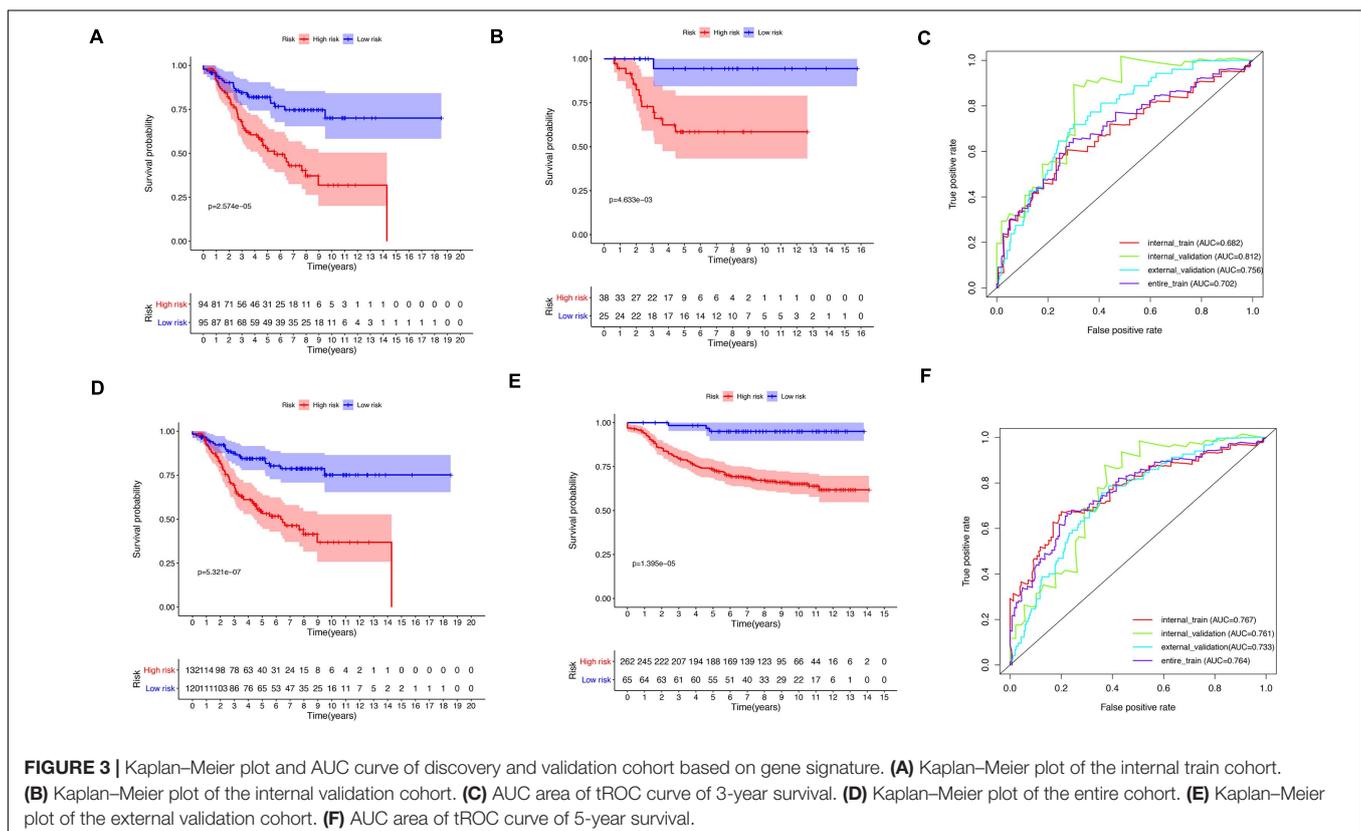
## Statistical Analysis

Survival analysis was evaluated using Kaplan–Meier analysis with the log-rank test.  $P$  value  $<0.05$  was defined as statistically significant. The time-dependent AUC value was calculated by the survivalROC package.

## RESULTS

### Identification of Functional Genomic Genes in Breast Cancer

The flowchart of analysis is shown in **Figure 1**. A total of 28 cell lines of breast cancer have dependence scores on the

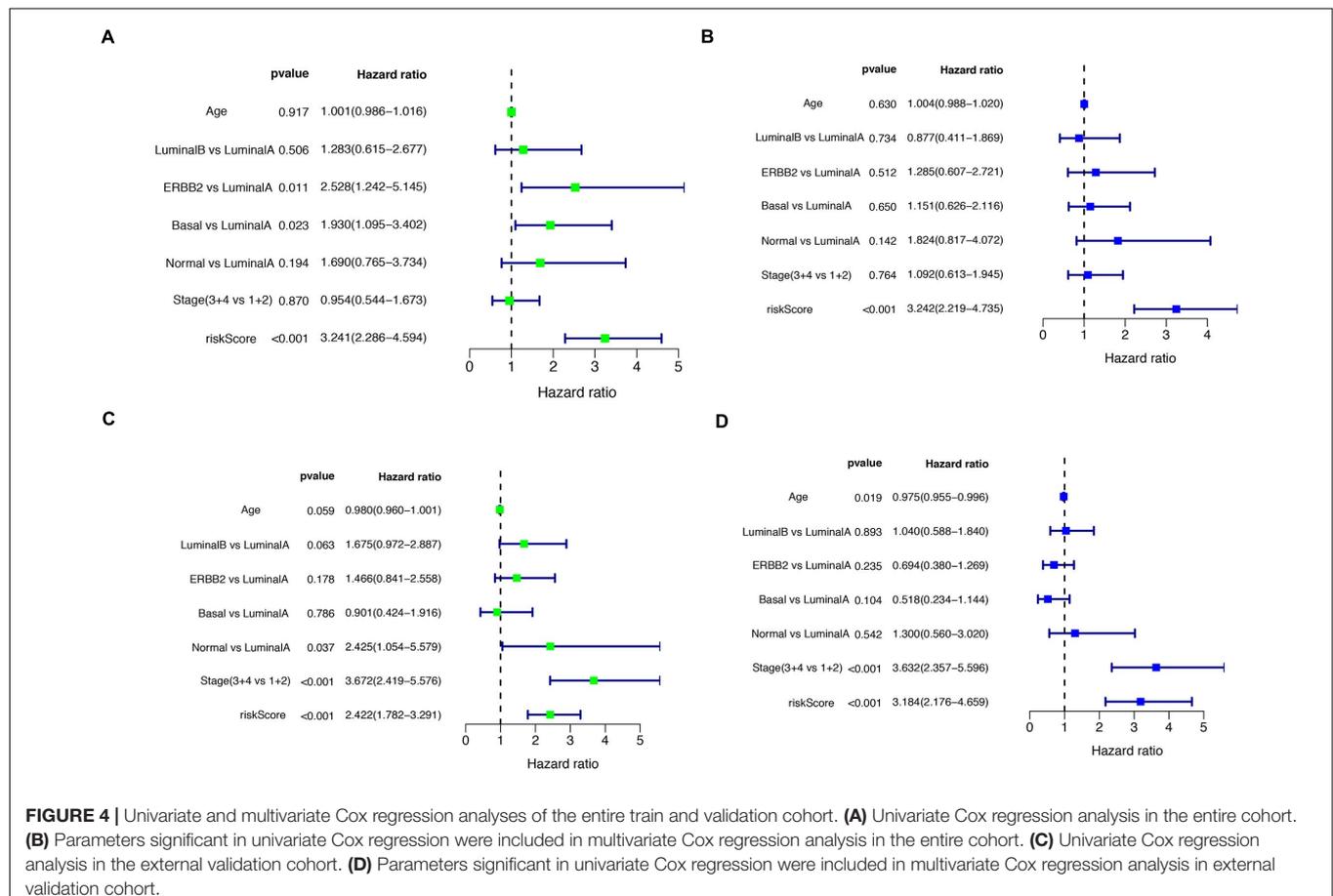


Depmap website. Genes with dependence score less than  $-0.5$  in all breast cancer cell lines were overlapped with upregulated genes in TCGA and 86 genes were discovered. Meanwhile, genes with dependence score greater than  $0.5$  in all breast cancer cell lines were overlapped with downregulated genes in TCGA and none of the genes were overlapped (Figure 2A). These 86 genes were defined as oncogenes. The dependence scores of oncogenes are shown in Figure 2B and expression of these genes in TCGA are shown in Figure 2C. PPI network revealed there contains 1117 interactions among these proteins (Figure 2D). Overall, the mutation rate of these genes was low, and INTS7 had the highest mutation rate of 0.8% (Supplementary Figure 1). KEGG analysis demonstrated that these oncogenes were enriched in pathways including cell cycle, DNA replication, oocyte meiosis, and nucleotide excision repair gene (Figure 2E). In GO analysis, the top three pathways enriched were ATP binding, adenyl ribonucleotide binding, and adenyl nucleotide binding (Figure 2F).

## Gene Signature Construction and Validation

A total of 51 genes were screened out by Lasso regression model, and a risk formula was constructed. The coefficients are listed in Supplementary Table 1. The median risk score

2.36 was used as cut-off value to divide patients into high-risk and low-risk group in internal training dataset and the same cut-off was also used in validation dataset. Patients from the high-risk group had significantly shorter median RFS of 5.52 years compared with patients from the low-risk group in internal training dataset (Figure 3A,  $p < 0.001$ , HR: 2.88, 95% CI: 1.72–4.82). This was validated in internal and external validation dataset. As expected, patients with high risk had poorer RFS compared with patients in low-risk group in internal validation dataset (Figure 3B,  $p < 0.001$ , HR: 10.54, 95% CI: 1.38–80.79), entire train dataset (Figure 3D,  $p < 0.001$ , HR: 3.28, 95% CI: 2.00–5.36), and external validation dataset (Figure 3E,  $p < 0.001$ , HR: 8.39, 95% CI: 2.65–26.48). Cut-off calculated by ROC curve was used and 2.68 was used as a cut-off to separate patients into different risks. Similarly, both in the entire dataset and validation dataset, there was statistical significance between high-risk and low-risk groups (Supplementary Figure 2). For 3-year tROC curves, AUC area was 0.682, 0.812, 0.702, and 0.756 for internal train, internal validation, entire train, and external validation dataset (Figure 3C). For 5-year tROC curves, AUC area was 0.767, 0.761, 0.764, and 0.733 for internal train, internal validation, entire train, and external validation dataset, respectively (Figure 3F). The distribution of risk scores of different risk groups are shown in risk plots (Supplementary Figure 3). After adjustment of clinicopathological variables, Cox

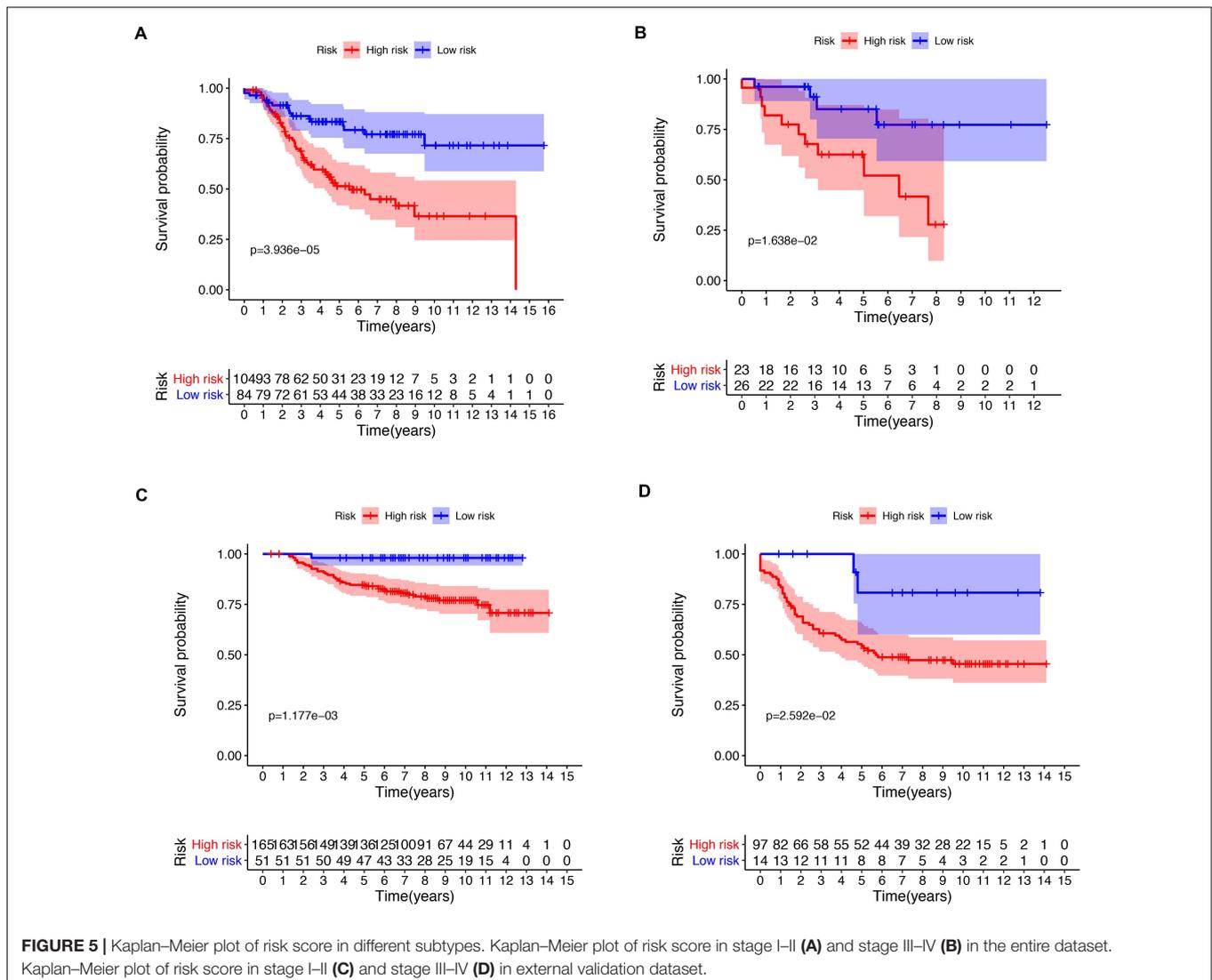


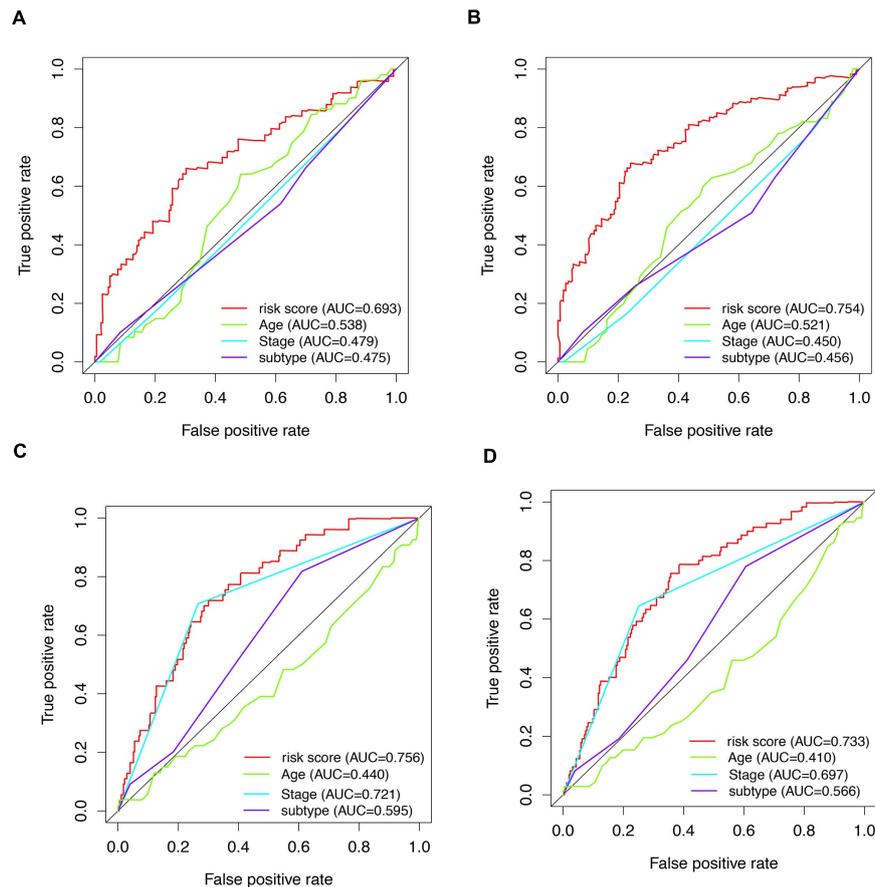
regression demonstrated that the risk score was an independent prognostic signature in the entire train set (Figures 4A,B,  $p < 0.001$ , HR: 3.24, 95% CI: 2.22–4.74) and external validation set (Figures 4C,D,  $p < 0.001$ , HR: 3.18, 95% CI: 2.18–4.66). Stratified analysis suggested that the signature was clinically significant in stage I + II, stage III + IV cases (Figures 5A–D). In both luminal and HER2 amplification groups, RFS of the high-risk group was also worse than that of the low-risk group in validation dataset (Supplementary Figure 4, luminal group:  $p < 0.0002$ , HER2 amplification group:  $p = 0.02$ ). A total of 45 patients were included in the basal group and only one patient was included in the low-risk group, so basal cohort with more patients were needed to validate the performance of this signature ( $p = 0.6$ ). Stratified analysis was done in dataset GSE20685 between patients with chemotherapy and without chemotherapy. In patients receiving chemotherapy, RFS of high-risk group was worse than that of low-risk group (Supplementary Figure 5,  $p < 0.0003$ ). In patients without therapy, the RFS of high-risk group was worse than that of low-risk group, but there was

no statistical significance ( $p = 0.1$ ). Meanwhile, our signature outperformed age, stage, and PAM50 subtype in both datasets (Figure 6). For 3-year tROC curves, AUC area was 0.693, 0.538, 0.479, and 0.475 for risk score, age, stage, and PAM50 subtype in entire dataset (Figure 6A). For 5-year tROC curves, AUC area was 0.754, 0.521, 0.450, and 0.456 for risk score, age, stage, and PAM50 subtype in entire dataset (Figure 6B). For 3-year tROC curves, AUC area was 0.756, 0.440, 0.721, and 0.596 for risk score, age, stage, and PAM50 subtype in external validation dataset (Figure 6C). For 5-year tROC curves, AUC area was 0.733, 0.410, 0.697, and 0.566 for risk score, age, stage, and PAM50 subtype in external validation dataset (Figure 6D).

### DEGs Between High-Risk and Low-Risk Groups

Differentially expressed genes were calculated between high-risk and low-risk groups in GSE21653 dataset. A total of 398 were identified, among which 285 genes were upregulated and 113





**FIGURE 6 |** AUCs of tROC curve for clinicopathologic parameters and risk score. **(A)** tROC curves of 3-year survival for clinicopathologic parameters and risk score in the entire train dataset. **(B)** tROC curves of 5-year survival for clinicopathologic parameters and risk score in the entire train dataset. **(C)** tROC curves of 3-year survival for clinicopathologic parameters and risk score in the validation dataset. **(D)** tROC curves of 5-year survival for clinicopathologic parameters and risk score in the validation dataset.

genes were downregulated (**Figures 7A,B**). Downregulated genes were only enriched in PPAR signaling pathway (**Figure 7C**), whereas upregulated genes were involved in cell cycle, oocyte meiosis, chemokine signaling pathway in KEGG analysis, and in identical protein binding, tubulin binding, and microtubule binding in GO analysis (**Figures 7D,E**). Interestingly, both in our signature and DEGs, cell cycle and oocyte meiosis were both enriched, indicating cell cycle was an important way regulating cell viability.

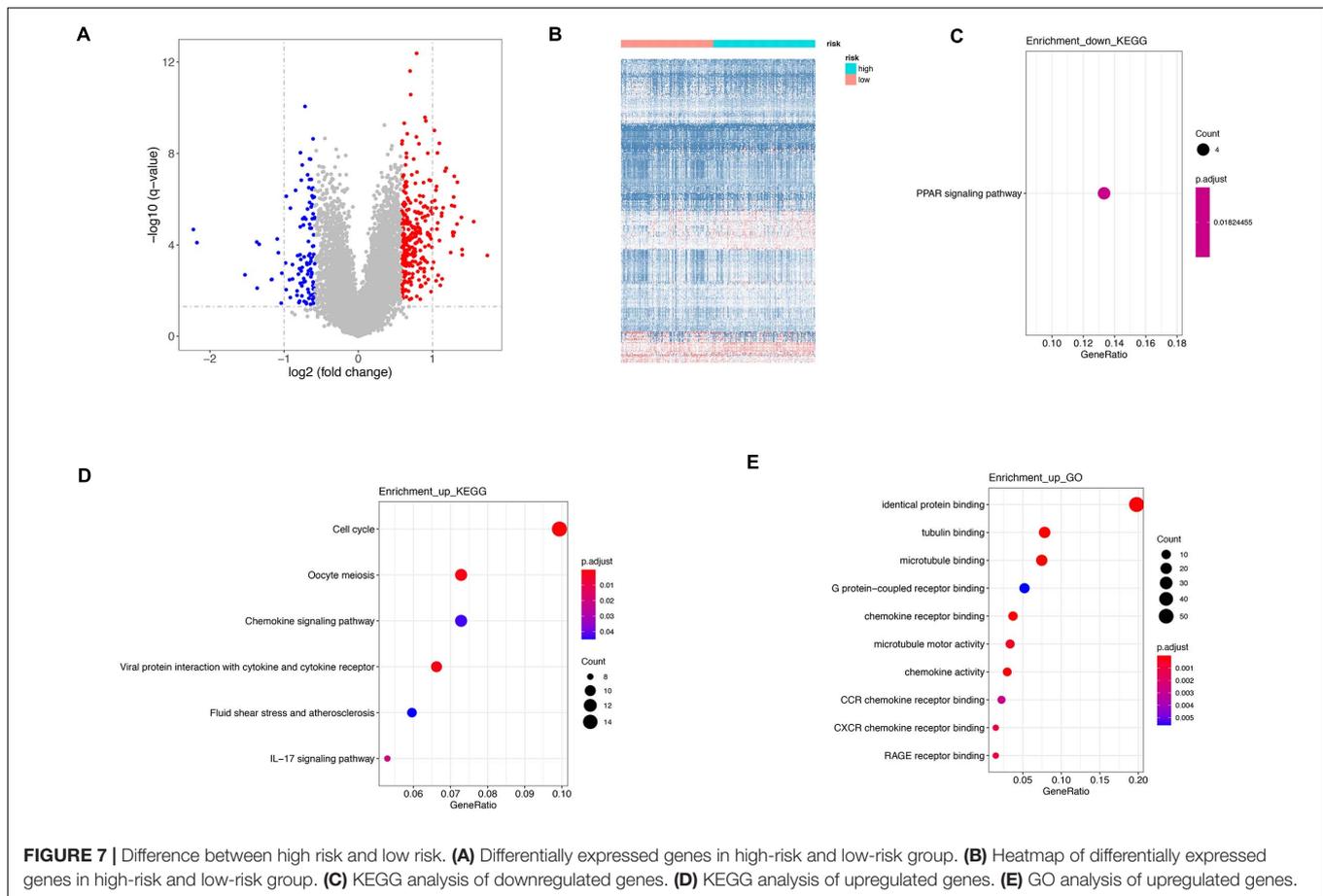
## WGCNA

WGCNA was performed in GSE21653 dataset to identify hub genes. Soft threshold power of 5 was selected to ensure a scale-free network (**Figure 8A**). A total of 37 clusters were clustered using cut height 0.25 (**Figure 8B**). Several modules were correlated with high-risk group (**Figure 8C**), including light yellow module (correlation parameter = 0.68,  $p < 0.001$ ), red module (correlation parameter = 0.5,  $p < 0.001$ ), and brown module (correlation parameter = 0.56,  $p < 0.001$ , **Figures 8D–F**). Upregulated DEGs from GSE21653 dataset, genes from three modules, and 86 oncogenes were overlapped and 25 genes were shared in these three groups (**Figure 8G**).

Enrichment analysis showed that cell cycle, progesterone-mediated oocyte maturation, and oocyte meiosis pathways were the top three pathways (**Figure 8H**). In GO analysis, the top three items enriched in these 25 genes were enzyme binding, ATP binding, and adenylyl ribonucleotide binding (**Figure 8I**). Of the 25 overlapped genes, 9 genes (CDC6, MCM2, MCM4, CDC20, CDK1, BUB1B, PLK1, CCNA2, GINS1) were enriched in cell cycle pathway. Similarly, upregulated DEGs from GSE21653 dataset, genes from three modules, and 51 oncogenes were overlapped and 25 genes were shared in these three groups (**Supplementary Figure 6**). A total of 13 genes were overlapped and cell cycle was still the main pathway regulated by these 13 genes. Six genes (CDC6, MCM4, CDK1, BUB1B, PLK1, CCNA2) were enriched in cell cycle pathway.

## DISCUSSION

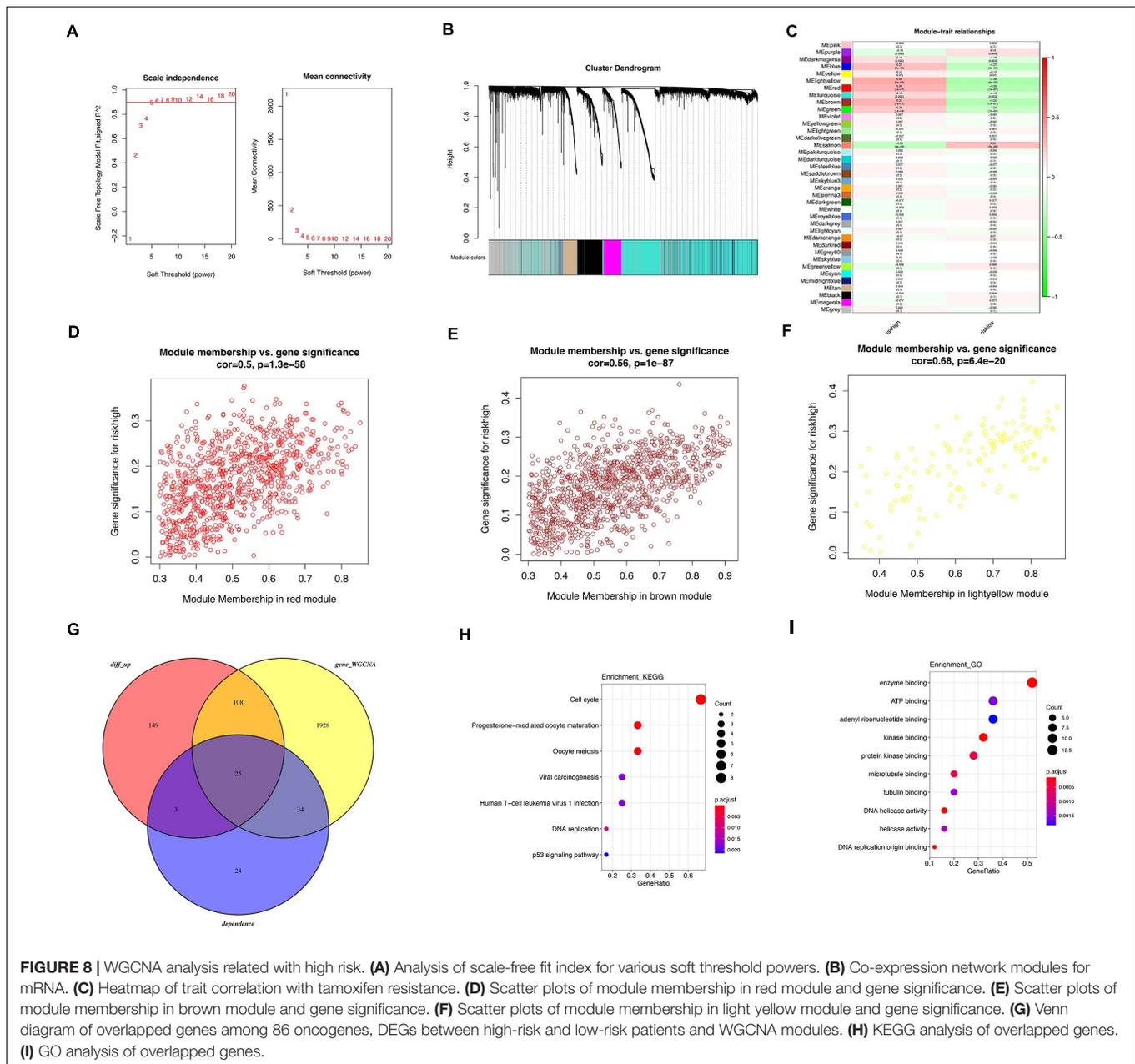
Breast cancer poses a great threat to women's health. Reliable targeted therapy provides prospects for improving the survival of breast cancer. The CRISPR-cas9 screening serves as a cornerstone



and is an outstanding way to systematically identify synthetic-lethal genes (Dhanjal et al., 2017). In the current study, we integrated CRISPR-cas9 screening results of breast cancer from DepMap with TCGA dataset, and identified 86 oncogenes that were overexpressed and have a lethal effect on breast cancer. Enrichment analysis revealed that these genes were enriched in cell cycle. Moreover, we established a gene signature screened from 86 genes, and this signature could divide patients into high risk and low risk. Interestingly, cell cycle pathway also ranked the first when we analyzed DEGs between high-risk and low-risk group and genes. Overlap among the 86 oncogenes, DEGs, and WGCNA analysis highlight the cell cycle pathways as well, indicating its importance to breast cancer viability.

In our study, 9 cell cycle related genes mainly involved in G1/S and G2/M phases, and researches on mechanism of these genes have been conducted. MCM2 and MCM4 genes are members of minichromosome maintenance (MCM) protein complex. As a crucial element of the pre-replication complex (pre-RC), they regulate the helicase activity and the formation of the replication forks. Meanwhile, they play an important role in the initiation of DNA replication and unwinding of the DNA strands in the G1 to S transition phase (Forsburg, 2004). High expression of MCM2 was associated with poor survival in breast cancer patients (Samad et al., 2020). In our analysis, we identified that MCM2

and MCM4 are vital to the survival of breast cancer. CDC6 plays an important role in DNA replication and cells could not initiate DNA replication without CDC6 (Borlado and Méndez, 2008). CDC6 helps MCM proteins load onto origins of replication and promotes them to associate with the chromatin. GINS1 also had a vital effect on G1/S transition by tightly interacting with DNA polymerase  $\epsilon$  (Pol  $\epsilon$ ) and participating in DNA replication (Takayama et al., 2003). The rest of the genes mainly had critical function in G2 to M transition. CDK1 is one of the most important proteins that regulate cell cycle progression via associating with cyclin B1 and promote G2 to M transition. High expression of CDK1 was reported in breast cancer compared with normal tissue (Barrett et al., 2002). CDC20 is co-activator of anaphase promoting complex (APC) which is a complex E3 ubiquitin ligase (Foe and Toczyński, 2011). APC<sup>Cdc20</sup> destructs critical cell cycle regulators such as cyclin B, allowing cells to progress from the metaphase to anaphase transition (Kim and Yu, 2011). CDC20 was reported higher expressed in breast cancer and functions as an oncogene, which was also validated in our analysis (Yuan et al., 2006). BUB1B, which blocks the activation of APC<sup>Cdc20</sup>, is the central component of the mitotic checkpoint for spindle assembly, and it was proved overexpressed in breast cancer and acts as a oncogene (Koyuncu et al., 2020). PLK1, as a member of polo-like kinase family, is involved in mitotic



entry, centrosome maturation, spindle assembly, and cytokinesis process (Golsteyn et al., 1995; García et al., 2020). Results from CRISPR-cas9 screening suggested that knocking out these genes leads to cell death in breast cancer of all subtypes, and inhibitors targeting these genes may be potential therapeutic strategies for breast cancer. Furthermore, upregulation of these genes in breast cancer indicates that they may be good candidates for drug development.

Previous studies constructed a prognostic signature mainly based on the gene expression, while we integrated functional genomic screening with gene expression in the current study. Our gene signature could divide patients into high risk and low risk regardless of the tumor stage. For patients with high risk,

appropriate target agent may be prescribed to improve patients' outcome. The prediction accuracy of our signature was better than classic clinicopathologic parameters such as tumor stage or PAM50 subtype. This signature was the only independent prognostic parameter in multivariate analysis, although subtypes and stage were widely used in clinical practice.

In conclusion, our research systematically studied genes vulnerable to cell viability and cell cycle is a vital pathway to this process. Functional genomic screening was integrated into our gene signature to predict the prognosis of breast cancer, which outperformed classical clinicopathological parameters in prediction accuracy. These cell cycle-related genes may serve as targets for breast cancer therapy.

## 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/s.

## AUTHOR CONTRIBUTIONS

XS participated in all experimental work and drafted the manuscript. XC and KS designed the article. ZW collected the data. All authors contributed to the article and approved the submitted version.

## ACKNOWLEDGMENTS

We would like to acknowledge the GEO database developed by the National Institutes of Health (NIH). We appreciate the financial support by the National Natural Science Foundation of China (grant nos. 81772797 and 82002773); Shanghai Municipal Education Commission—Gaofeng Clinical Medicine Grant Support (20172007); and Ruijin Hospital, Shanghai Jiao Tong University School of Medicine—“Guangci Excellent Youth Training Program” (GCQN-2017-A18). All these financial sponsors had no role in the study design, collection, analysis, or interpretation of data.

## REFERENCES

- Barrett, K. L., Demiranda, D., and Katula, K. S. (2002). Cyclin b1 promoter activity and functional cdk1 complex formation in G1 phase of human breast cancer cells. *Cell Biol. Int.* 26, 19–28. doi: 10.1006/cbir.2001.0817
- Borlado, L. R., and Méndez, J. (2008). CDC6: from DNA replication to cell cycle checkpoints and oncogenesis. *Carcinogenesis* 29, 237–243. doi: 10.1093/carcin/bgm268
- Dhanjal, J. K., Radhakrishnan, N., and Sundar, D. (2017). Identifying synthetic lethal targets using CRISPR/Cas9 system. *Methods* 131, 66–73. doi: 10.1016/j.ymeth.2017.07.007
- Doudna, J. A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346:1258096. doi: 10.1126/science.1258096
- Foe, I., and Toczyski, D. (2011). Structural biology: a new look for the APC. *Nature* 470, 182–183. doi: 10.1038/470182a
- Forsburg, S. L. (2004). Eukaryotic MCM proteins: beyond replication initiation. *Microbiol. Mol. Biol. Rev.* 68, 109–131. doi: 10.1128/mmbr.68.1.109-131.2004
- García, I. A., Garro, C., Fernandez, E., and Soria, G. (2020). Therapeutic opportunities for PLK1 inhibitors: spotlight on BRCA1-deficiency and triple negative breast cancers. *Mutat. Res.* 821:111693. doi: 10.1016/j.mrfmmm.2020.111693
- Gene Ontology Consortium (2015). Gene ontology consortium: going forward. *Nucleic Acids Res.* 43, D1049–D1056. doi: 10.1093/nar/gku1179
- Golsteyn, R. M., Mundt, K. E., Fry, A. M., and Nigg, E. A. (1995). Cell cycle regulation of the activity and subcellular localization of Plk1, a human protein kinase implicated in mitotic spindle function. *J. Cell Biol.* 129, 1617–1628. doi: 10.1083/jcb.129.6.1617
- Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646–674. doi: 10.1016/j.cell.2011.02.013
- Ingham, M., and Schwartz, G. K. (2017). Cell-Cycle therapeutics come of age. *J. Clin. Oncol.* 35, 2949–2959. doi: 10.1200/jco.2016.69.0032

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2021.646774/full#supplementary-material>

**Supplementary Figure 1** | Mutation rate of 86 genes in TCGA dataset.

**Supplementary Figure 2** | Kaplan–Meier plot of discovery and validation cohort based on gene signature using cut-off from ROC curve. **(A)** Kaplan–Meier plot of the internal train cohort. **(B)** Kaplan–Meier plot of the internal validation cohort. **(C)** Kaplan–Meier plot of the entire cohort. **(D)** Kaplan–Meier plot of the external validation cohort.

**Supplementary Figure 3** | Risk plot of the discovery and validation cohort. **(A)** Risk score distribution of patients in the prognostic model in the discovery cohort. **(B)** Relationship between the survival time and risk score rank in the discovery cohort. **(C)** Risk score distribution of patients in the prognostic model in the validation dataset. **(D)** Relationship between the survival time and risk score rank in the validation cohort.

**Supplementary Figure 4** | Kaplan–Meier plot of risk score in different subtypes. Kaplan–Meier plot of risk score in luminal **(A)** and HER2 **(B)** subtypes in GSE20685 dataset.

**Supplementary Figure 5** | Kaplan–Meier plot of risk score in different chemotherapy groups. Kaplan–Meier plot of risk score in patients with chemotherapy **(A)** and without chemotherapy **(B)** in GSE20685 dataset.

**Supplementary Figure 6** | Overlapped genes between three groups. **(A)** Venn diagram of overlapped genes among 51 oncogenes, DEGs between high-risk and low-risk patients, and WGCNA modules. **(B)** GO analysis of overlapped genes. **(C)** KEGG analysis of overlapped genes.

- Joung, J., Konermann, S., Gootenberg, J. S., Abudayyeh, O. O., Platt, R. J., Brigham, M. D., et al. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat. Protoc.* 12, 828–863. doi: 10.1038/nprot.2017.016
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361. doi: 10.1093/nar/gkw1092
- Kastan, M. B., and Bartek, J. (2004). Cell-cycle checkpoints and cancer. *Nature* 432, 316–323. doi: 10.1038/nature03097
- Kim, S., and Yu, H. (2011). Mutual regulation between the spindle checkpoint and APC/C. *Semin. Cell Dev. Biol.* 22, 551–558. doi: 10.1016/j.semcdb.2011.03.008
- Koyuncu, D., Sharma, U., Goka, E. T., and Lippman, M. E. (2020). 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
- Kurata, M., Yamamoto, K., Moriarity, B. S., Kitagawa, M., and Largaespa, D. A. (2018). CRISPR/Cas9 library screening for drug target discovery. *J. Hum. Genet.* 63, 179–186. doi: 10.1038/s10038-017-0376-9
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559
- Lim, S., and Kaldis, P. (2013). Cdks, cyclins and CKIs: roles beyond cell cycle regulation. *Development* 140, 3079–3093. doi: 10.1242/dev.091744
- Malumbres, M. (2014). Cyclin-dependent kinases. *Genome Biol.* 15:122. doi: 10.1186/gb4184
- Malumbres, M., and Barbacid, M. (2009). Cell cycle, CDKs and cancer: a changing paradigm. *Nat. Rev. Cancer* 9, 153–166. doi: 10.1038/nrc2602
- Meyers, R. M., Bryan, J. G., McFarland, J. M., Weir, B. A., Sizemore, A. E., Xu, H., et al. (2017). Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* 49, 1779–1784. doi: 10.1038/ng.3984
- Samad, A., Haque, F., Nain, Z., Alam, R., Al Noman, M. A., Rahman Molla, M. H., et al. (2020). Computational assessment of MCM2 transcriptional

- expression and identification of the prognostic biomarker for human breast cancer. *Heliyon* 6:e05087. doi: 10.1016/j.heliyon.2020.e05087
- Schuster, A., Erasmus, H., Fritah, S., Nazarov, P. V., van Dyck, E., Niclou, S. P., et al. (2019). RNAi/CRISPR screens: from a pool to a valid hit. *Trends Biotechnol.* 37, 38–55. doi: 10.1016/j.tibtech.2018.08.002
- Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer statistics, 2020. *CA Cancer J. Clin.* 70, 7–30. doi: 10.3322/caac.21590
- Slamon, D. J., Neven, P., Chia, S., Fasching, P. A., De Laurentiis, M., Im, S. A., et al. (2018). Phase III randomized study of Ribociclib and Fulvestrant in hormone receptor-positive, human epidermal growth factor receptor 2-negative advanced breast cancer: MONALEESA-3. *J. Clin. Oncol.* 36, 2465–2472. doi: 10.1200/jco.2018.78.9909
- Strzyz, P. (2016). Cell signalling: signalling to cell cycle arrest. *Nat. Rev. Mol. Cell Biol.* 17:536. doi: 10.1038/nrm.2016.108
- Takayama, Y., Kamimura, Y., Okawa, M., Muramatsu, S., Sugino, A., and Araki, H. (2003). GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. *Genes Dev.* 17, 1153–1165. doi: 10.1101/gad.1065903
- Ternès, N., Rotolo, F., and Michiels, S. (2016). Empirical extensions of the lasso penalty to reduce the false discovery rate in high-dimensional Cox regression models. *Stat. Med.* 35, 2561–2573. doi: 10.1002/sim.6927
- Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Stat. Med.* 16, 385–395.
- Tripathy, D., Im, S. A., Colleoni, M., Franke, F., Bardia, A., Harbeck, N., et al. (2018). Ribociclib plus endocrine therapy for premenopausal women with hormone-receptor-positive, advanced breast cancer (MONALEESA-7): a randomised phase 3 trial. *Lancet Oncol.* 19, 904–915. doi: 10.1016/s1470-2045(18)30292-4
- Tsherniak, A., Vazquez, F., Montgomery, P. G., Weir, B. A., Kryukov, G., Cowley, G. S., et al. (2017). Defining a cancer dependency map. *Cell* 170, 564–576.e516. doi: 10.1016/j.cell.2017.06.010
- Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118
- Yuan, B., Xu, Y., Woo, J. H., Wang, Y., Bae, Y. K., Yoon, D. S., et al. (2006). Increased expression of mitotic checkpoint genes in breast cancer cells with chromosomal instability. *Clin. Cancer Res.* 12, 405–410. doi: 10.1158/1078-0432.Ccr-05-0903
- Zhou, Z. R., Wang, W. W., Li, Y., Jin, K. R., Wang, X. Y., Wang, Z. W., et al. (2019). In-depth mining of clinical data: the construction of clinical prediction model with R. *Ann. Transl. Med.* 7:796. doi: 10.21037/atm.2019.08.63
- 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.
- Copyright © 2021 Sun, Wang, Chen and Shen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). 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.