A novel twelve-gene signature to evaluate neoadjuvant chemotherapy response and predict prognosis in breast cancer

Peng Tang (tp1232000@sina.com)
Southwest Hospital, Army Medical University (Third Military Medical University)
Jin Wu
Southwest Hospital, Army Medical University (Third Military Medical University)
Yuan Tian
Linyi People's Hospital
Bin Liao
Chongqing General Hospital
Minghao Wang
Southwest Hospital, Army Medical University (Third Military Medical University)
Wei Liu
Southwest Hospital, Army Medical University (Third Military Medical University)
Yuzhao Yan
Southwest Hospital, Army Medical University (Third Military Medical University)
Ying Hu
Southwest Hospital, Army Medical University (Third Military Medical University)
Hong Zheng
Xinqiao Hospital, Army Medical University (Third Military Medical University)

Research Article

Keywords: Breast cancer, Evaluations of neoadjuvant chemotherapy, Chemo-resistant, Prognosis of breast cancer

Posted Date: February 8th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1308081/v1

Abstract

Background: Accurate evaluation of the response to neoadjuvant chemotherapy (NAC) provides important information about systemic therapies for breast cancer, which implies tumor biology, prognosis, and guide further therapy. Gene profiles overcome many limitations observed with classic histopathological variables but are complicated and expensive. Therefore, it is essential to develop a more accurate, repeatable, and economical evaluation approach for neoadjuvant chemotherapy responses.

Methods: We analyzed the transcriptional profiles of chemo-resistant breast cancer cell lines and tumors of chemo-resistant breast cancer patients in the GEO25066 dataset. We preliminarily screened out common significantly differentially expressed genes and constructed a NAC response risk model using least absolute shrinkage and selection operator (LASSO) regression and univariate and multivariate analyses. The differences in prognosis, clinical features, tumor microenvironment components, and immune characteristics were compared between risk groups. The connectivity map (CMap) database was used to screen potential drugs that could reverse chemotherapy resistance in breast cancer.

Results: Thirty-six genes were commonly overexpressed or downregulated in both NAC chemo-resistant tumors and cells compared to the sensitive tumors and wild-type cells. Through LASSO regression, we obtained a risk model composed of 12 genes. The risk model divided patients into high-risk and low-risk groups. Univariate and multivariate Cox regression analyses indicated that the risk score can be used as an independent prognostic factor for evaluating NAC response in breast cancer. Tumors in the high-risk and low-risk groups showed significant differences in molecular biological characteristics, tumor-infiltrating lymphocytes, and immunosuppressive molecule expression. Our results showed that the risk score was also an independent prognostic factor for breast cancer. Finally, we screened potential drugs that could reverse chemotherapy resistance in breast cancer.

Conclusion: Our results suggest that the novel signature of 12 genes could be used to evaluate NAC response and predict prognosis in breast cancer.

Background

Breast cancer has become the cancer with the world's highest incidence rate [1]. Neoadjuvant chemotherapy (NAC) refers to systemic cytotoxic drug treatment before surgery or radiotherapy, and is considered as the standard treatment for patients with locally advanced or inoperable breast cancer [2,3]. Accurate evaluation of the response to neoadjuvant chemotherapy provides important information about tumor biology and prognosis and guides further therapy [4,5,6]. In addition to clinical and pathological evaluation criteria, gene expression signatures have been developed to evaluate the response to NAC [7,8]. Different multi-gene expression signatures, such as genomic grade index (GGI), MammaPrint, and Oncotype DX, have been shown to outperform classic histopathological variables and represents an important step toward personalizing breast cancer treatment [9,10,11]. In particular, gene profiles overcome many of the limitations observed in classic histopathologic variables [12].

GGI is a gene expression signature developed to better define histologic grade assessment and evaluate the response to chemotherapy [7]. Surname et al. evaluated the ability of GGI to predict the response to NAC was evaluated in 229 tumor samples collected before neoadjuvant chemotherapy with paclitaxel, fluorouracil, doxorubicin, and cyclophosphamide (T/FAC). Chemotherapy responsiveness increased with higher GGI values. Better prognosis and higher endocrine responsiveness were observed in the subgroup of patients with lower GGI values [13]. In the evaluation of GGI as a predictor of response, a more precise method for evaluating pathologic response called residual cancer burden (RCB) was used as a comparator [14,15]. Unlike pathological evaluation, GGI assessment tends to be reproducible. However, GGI involves too many genes (97 genes), resulting in high detection costs and difficulty in clinical promotion.

In this study, we analyzed the transcriptional patterns of breast cancer cell lines and tumors of NAC-resistant patients evaluated by GGI and screened potential genes associated with chemoresistance. Furthermore, we constructed a neoadjuvant chemotherapy response risk model and examined the evaluation accuracy of the risk score for NAC response. We conducted molecular bioinformatics analysis of the genes that constitute the chemotherapy resistance risk score and explored potential drugs to reverse breast cancer chemotherapy resistance. Finally, we examined the risk score for predicting the prognosis of breast cancer. In summary, we have reported a novel signature to evaluate neoadjuvant chemotherapy response and predict prognosis in breast cancer and screened out potential drugs to reverse chemotherapy resistance in breast cancer.

Methods

Cell Cultures and Chemo-resistant Cell Line Induction

Human breast cancer cell lines MCF-7 (luminal breast cancer), SKBR3 (HER2+ breast cancer), and MDA-MB-231 (triple negative breast cancer) were purchased from FuHeng Biology. All cell lines were authenticated using short tandem repeat analysis. MCF-7 cells were cultured in DMEM (Gibco, USA) supplemented with 10% fetal bovine serum (FBS, Gibco, USA), penicillin (100 U/mL, Gibco, USA), and streptomycin (100 μg/mL, Gibco, USA). SKBR3 cells were cultured in McCoy’s 5A medium (McCoy's 5A, Gibco, USA) supplemented with 10% FBS, penicillin (100 U/mL, Gibco, USA), and streptomycin (100 μg/mL, Gibco, USA). MDA-MB-231 cells were cultured in Leibovitz's L-15 medium with 10% FBS, penicillin (100 U/mL, Gibco, USA), and streptomycin (100 μg/mL, Gibco, USA). Cells were incubated in a humidified atmosphere containing 5% CO_2 at 37 °C and were regularly checked for mycoplasma infection.

The epirubicin (EPI, s1223, Selleck, CHN) -resistant variants of MCF-7, SKBR3, and MDA-MB-231 cells were generated by pulse selection. Cells were exposed to the respective maximal inhibitory concentration (IC_{90}) values of EPI or docetaxel (DOC, s1148, Selleck) for 4 h, on-ce a week for 10 weeks, to obtain resistant variants T47D/EPI, SKBR3/EPI, and MDA-MB-231/EPI [16,17]. The EPI-resistant breast cancer cells were washed with PBS, fully lysed with Trizol reagent (15596026, Invitrogen, USA), and stored at - 80°C until use.

Cytotoxicity Assay

Five thousand MCF-7, SKBR3, and MDA-MB-231 cells were seeded in each well of 96-well plates. After the cells adhered completely, gradient concentrations EPI and DOC were added 72 h later, cells were stained with sterile methylthiazolyldiphenyl-tetrazolium bromide (MTT, C0009, Beyotime, CHN) in culture media (1:10) for 2 h at 37 °C. The absorbance was measured at 570 nm [18, 19].

Five thousand MCF-7/EPI cells were seeded in each well of a 96-well plate. After the cells adhered completely, EPI (0.04 μM), DOC (0.01 μM), bambuterol (HY-17501A, MCE, 0.04 μM), pravastatin (HY-B0165A, MCE, 0.37 μM), isocarboxazid (HY-13929, MCE, 10 μM), Imexon (HY-15385, MCE,0.125μM), temozolomide (HY-17364, MCE, 0.12 μM), axitinib (HY-10065, MCE, 1.11 μM), semaxanib (s2845, Selleck) and crizotinib (HY-50878, MCE, 0.37 μM) were added 72 hours later and cells were stained with sterile methylthiazolyldiphenyl-tetrazolium bromide (MTT, C0009, Beyotime, CHN) in culture media (1:10) for 2 h at 37 °C. Absorbance was measured at 570 nm. The drug concentration was the same as the concentration used for the connectivity map (CMap) (https://clue.io/).

RNA Preparation and RNA-seq

Total RNA from MCF-7, SKBR3, MDA-MB-231, MCF-7/EPI, SKBR3/EPI, and MDA-MB-231/EPI was extracted using TRIzol reagent and treated with RNase-free DNase I to remove genomic DNA contamination. Thereafter, RNA was assessed using a Nano Photometer® spectrophotometer (IMPLEN, CA, USA) and a Qubit® 2.0 Fluorometer (Invitrogen, USA). The high-quality RNA samples were subsequently submitted to Sangon Biotech Co., Ltd. (Shanghai, China). Sequencing libraries were generated using the VAHTSTM mRNA-seq V2 Library Prep Kit (Illumina®, USA). Paired-end sequencing of the library was performed using NovaSeq sequencers (Illumina, USA). Gene expression values of the transcripts were computed using the String Tie software (version 1.3.3b).

Differential Gene Expression and Enrichment Analysis

Breast cancer transcriptome and clinical data GSE25066 (n = 509) were downloaded from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/). The transcriptome data included original data files (CEL files) and platform files. The differential gene expression profiles of NAC-resistant- and -sensitive patients were analyzed using the R language package (limma 3.20.9). GO and KEGG pathway analysis were applied to annotate the biological functions of DEGs by R language package (GOplot, kegg plot function R). Gene set enrichment analysis (GSEA) was performed to investigate the functions correlated with different risk groups using GSEA 4.1.0 (http://www.gsea-msigdb.org/gsea/downloads) [20].

Diagnostic Model Construction and Validation

After deleting the patient samples with missing data, there were 492 samples in the dataset. In total, 246 samples were randomly selected as the training set, and the remaining 246 samples were used as the validation set. The “survival” R package was used to perform univariate Cox regression analysis of the differentially expressed genes between clusters GGI-low and GGI-high, the “glmnet” R package was used for LASSO analysis [21], and the “survival” R package was used for multivariate Cox regression analysis to establish a risk model. The risk score was obtained using the “predict” function in R software, and the mathematical model of the risk score is as follows: Risk score = $h0(t) * \exp(\beta 1X1 + \beta 2X2 + \ldots + \beta nXn)$, where n is the representative number of modeling genes; β and X are the correlation coefficient and expression level of model gene prediction, respectively; and h0(t) is derived from the “predict” function. The ROC curve was drawn with the true positive rate (sensitivity) as the ordinate and the false positive rate (1-specificity) as the abscissa to obtain the Jordan index (sensitivity + specificity-1). The boundary value or decision threshold corresponding to the maximum youden index is the best classification boundary value; that is, the Youden index criterion [22] was also applied in order to select the optimal cutoff.

Bioinformatics Analysis

To identify the protein–protein interactions, the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) was employed [24]. The R package “survival” was used for univariate and multivariate analyses of the signature-based risk factor score, age, stage, ER, PR, HER2, and grade to NAC resistant to prognosis [23]. The survival curves of nine hub genes and seven genes used for modeling were obtained from the Kaplan–Meier plotter (http://kmplot.com/analysis) [25]. The CMap database was used to identify compounds that were negatively correlated with the input differential gene profile after acting on MCF-7 cells [26,27].

Immune Cell Infiltration Analysis

In the TIMER database, we explored the impact of mutations in the seven genes on the degree of infiltration of eight immune cell types within the tumor microenvironment (TME). Based on the GEO25066 database, we further analyzed the immune scores of 492 patients in the different risk groups. Next, we used three analysis methods (CIBERSORT, QUANTISEQ, and XCELL) to determine the immune cell infiltration status within the TME of patients in the TIMER2 database. [28].

The correlation between risk scores and various immunosuppressive molecules was graphically displayed using the online analysis tool hiplot. The R package was used to analyze the correlation between the risk score and multiple treatment targets.

Construction and Assessment of the Nomogram

We used the “rms” R package to build the nomogram and calibration chart. A calibration chart was used to evaluate the performance of the nomogram.

Statistical Analysis

GraphPad Prism 8.0 was used to draw various bar graphs. The data were reported as the mean ± SEM. In this study, analyses between the two groups were performed using Student’s t-tests or Mann–Whitney U-tests. One-way analysis of variance was used to analyze the difference between multiple groups. Categorical variables in different groups were analyzed used the chi-square test by SPSS22.0. Spearman’s test was used to analyze the correlation between the two groups. Statistical significance was described as follows: n.s., not significant; *P < 0.05; **P < 0.01; and ***P < 0.001.

Results

Screening of Hub Genes related to NAC resistance in Breast Cancer

To explore the common mechanism that promotes chemoresistance in breast cancer, we generated EPI-resistant cell lines of different subtypes of breast cancer, including MCF-7/EPI, SKBR3/EPI, and MDA-MB-231/EPI. The IC50 of EPI in resistant cells was seven times higher than that of parental wild-type (WT) cells (Fig. 1A and Table 1). Furthermore, consistent with clinical experience, tumor chemoresistance showed characteristics of multidrug resistance in our experiment. Breast cancer drug-resistant cells induced by EPI (a cytotoxic agent that inhibits RNA and DNA synthesis) are also resistant to DOC (a cytotoxic agent that stabilizes the tubulin polymer bundles and therefore interferes with microtubular assembly and cell replication). We then compared the expression profiles of EPI-resistant cells with WT cells. Three hundred and two genes were commonly upregulated or downregulated at least 2-fold in MCF-7/EPI, SKBR3/EPI, and MDA-MB-231/EPI compared to that in parental MCF-7, SKBR3, and MDA-MB-231 cells (Fig. 1B and C). In GO analysis, we analyzed the top 30 GO with the highest enrichment in cellular component (CC), biological process (BP), and molecular function (MF). Upregulated DEGs were widely distributed in the intracellular, intracellular part, intracellular organelle, organelle, and nucleus of the breast cancer resistant cells, and were enriched in “regulation of cellular metabolic process,” “negative regulation of cellular process” and “negative regulation of biological process.” The molecular functions of the upregulated DEGs were “protein and transcription regulatory region DNA” and “RNA polymerase II proximal promoter sequence-specific DNA binding.” Different to upregulated DEGs, downregulated DEGs were enriched in extracellular of the breast cancer resistant cells and were enriched in “cell adhesion” and “regulation of cell motility” (Supplementary material 1). The results showed that slowing down the cell cycle and decreasing biological processes and metabolic abnormalities are important mechanisms for the survival of chemo-resistant cells. In KEGG enrichment analysis, upregulated chemoresistance cell feature genes were enriched in “DNA replication,” “HIF-1 signaling pathway,” “p53 signaling pathway, and “Pentose phosphate pathway.” Downregulated chemoresistance cell feature genes were enriched in “Glutathione metabolism” and “Chemokine signaling pathway” (Supplementary material 1). The enrichment of these functions suggested that breast cancer chemo-resistant cells may resist chemotherapy by slowing down the cell cycle and strengthening DNA repair and synthesis. In order to more comprehensively characterize the biological characteristics of breast cancer drug resistance, we analyzed the hallmarks of breast cancer chemo-resistant cells and tumor tissues of breast cancer chemo-resistant patients by GSEA. We found that E2F and MYC targets, mTORC1 signaling, P53 pathway and KRAS signal gene sets were significantly upregulated in breast cancer resistant cells (Supplementary material 1). This suggests that the chemoresistance of breast cancer cells is closely related to the regulation of the cell cycle and apoptosis.

We also introduced clinical data and transcriptome profiles of breast cancer patients undergoing NAC using the GSE25066 database. According to the GGI evaluation method, GSE25066 samples were divided into two groups: GGI-low (NAC resistance, n=157) and GGI-high (NAC sensitivity, n=335) (Fig. 1D). Principal component analysis of gene expression showed that there was a significant difference in gene expression between the two groups (Fig. 1E). There were three hundred and forty-seven DEGs between the two groups. Thirty-six genes were commonly overexpressed or downregulated in both NAC chemo-resistant tumors and cells compared to the sensitive tumors and WT cells (Fig. 1F and G).

A Risk Model with NAC response was Constructed Based on GGI Lever

To further explore the clinical value of the differentially expressed genes, we performed the LASSO regression model based on the expression and prognosis data of 246 breast cancer patients who received NAC from the GSE25066 training set (Fig. 2A and B). We obtained two gene sets: 1se (containing 12 genes) and min (containing 18 genes). The ROC analysis showed that the 1se set (AUC=0.97) and min set (AUC=0.98) of characteristic genes both have good diagnostic values for evaluating the resistance of breast cancer to neoadjuvant chemotherapy (Fig. 2C). Considering the cost of patient detection, we chose the 1se set: *HJURP, IFI27, RAD51AP1, EZH2, DNMT3B, SLC7A5, DBF4, USP18, ELOVL5, PTGER3, KIAA1324,* and *CYBRD1* (all 12 genes mentioned below refer to these 12 genes). The results of the validation set analysis were consistent with those of the training set. (Supplementary material 2). The complete names of the 12 genes and their main functions are listed in Table 2. The mathematical model of the risk score was as follows: Risk score =h0(t)×exp (HJURP0.0887787219+IFI27×0.0047893433+
RAD51AP1×0.0554382065+EZH2×0.0351434032+DNMT3B×0.0301466701+SLC7A5×0.0511106486+DBF4×0.0100252843-USP18×0.0251921906-ELOVL5×0.0158776302-PTGER3×0.0116989697-KIAA1324×0.0002110923- CYBRD1×0.0062603332). The cutoff risk score was 0.51 (Fig. 2D). Univariate and multivariate analyses also indicated that the signature-based risk factor score was a good prognostic indicator (Fig. 2E and F).

Characteristics of Tumor Cells and Tumor Microenvironment in High-risk Patients

To explore the internal and external factors of breast cancer cell NAC resistance, we analyzed the gene expression characteristics and tumor microenvironment of NAC-resistant cancer cells. According to the GSE25066 gene expression profile, the positive genes of risk score (*HJURP, IFI27, RAD51AP1, EZH2, DNMT3B, SLC7A5, DBF4, USP18,* and *ELOVL5*) were negatively correlated with the expression of negative genes (*PTGER3, KIAA1324,* and *CYBRD1*) (Fig. 3A). Utilizing the

HitPredict database, we searched for interaction partners of their protein products, which were displayed in the protein interaction network diagram (Fig. 3B). We then performed GO and KEGG enrichment analyses of these genes. GO enrichment analysis indicated that 12 genes promote breast cancer resistance to NAC by upregulating DNA repair and metabolism-related pathways and downregulating some membrane receptor signaling mechanisms (Fig. 3C and Table 3). Furthermore, KEGG pathway enrichment analysis also revealed that the G protein-coupled receptor signaling pathway and transition metal ion transport were enriched by these genes (Table 3). The chord plot showed that 12 genes had complex interactions with the enriched GO pathway (Fig. 3D). For example, *EZH2* and *DNMT3B* are involved in transferase activity (GO0016740) and negative regulation of gene expression (GO0045892 and GO0045814). DNMT3B, DBF4, and CYBRD1 jointly regulate metal ion binding (GO0046872). The above results suggest that these 12 genes cause drug-resistant cells in dormant cells to resist the killing effect of chemotherapeutic drugs by increasing DNA repair, reducing cell macromolecule synthesis, and cell metabolism. We also analyzed the correlation between the risk score and some known molecules associated with chemotherapy resistance in breast cancer. The risk score was significantly positively correlated with genes related to cell cycle regulation, DNA repair, transport, and efflux (Fig. 3E).

Tumor-infiltrating lymphocytes are critical components of the tumor microenvironment and are important external factors of chemotherapy resistance [29,30,31]. Therefore, we analyzed the difference in the proportion of tumor-infiltrating lymphocytes between the high-risk and low-risk groups. Compared with the low-risk group, the high-risk group recruited more CD4, CD8, and NK cells (Fig. 3F and G). Interestingly, we found that there were more tumor-infiltrating cells in the high-risk group, but the survival rate was significantly lower than that in the low-risk group (Fig. 2). We further explored the correlation between the risk score and immunosuppressive molecules in the GSE25066 dataset. The risk score was positively correlated with many immunosuppressive molecules, such as *CTLA4*, *LAG3*, *ICOS*, *IDO1,* and *ADORA2A* (Fig. 3H). The above results indicate that a large number of tumor-infiltrating cells in the high-risk group are depleted, which may lead to the failure of NAC for breast cancer and the lower survival rate of patients.

Evaluation of Risk Score on Prognosis of Overall Breast Cancer

To confirm whether the risk score for evaluating NAC resistance can predict the prognosis of breast cancer, we conducted the following study. According to TCGA BRCA data, the expression of the positive genes of risk score (*HJURP, IFI27, RAD51AP1, EZH2, DNMT3B, SLC7A5, DBF4, USP18,* and *ELOVL5*) in tumors was higher than that in normal and tumor-adjacent tissues. The expression of the negative genes (*PTGER3, KIAA1324,* and *CYBRD1*) in tumors was lower than that in normal and tumor-adjacent tissues (Supplementary materiale 3). We also found that high expression of all the positive genes and low expression of all negative genes were associated with poor prognosis of breast cancer (Supplementary material 3). The above results suggest that the risk score composed of these 12 genes may also serve as a prognostic factor for breast cancer. The expression heatmap showed that *RAD51AP1, EZH2, DNMT3B, SLC7A5, DBF4,* and *USP18* were highly expressed in the high-risk group, whereas *ELOVL5, PTGER3, KIAA1324,* and *CYBRD1* were expressed at low levels in the high-risk group (Fig. 4A). Survival status and risk score showed the distribution of patients into different risk groups (Fig. 4B). Kaplan–Meier survival analysis showed that patients in the high-risk group had a worse prognosis than those in the low-risk group (Fig. 4C). To facilitate clinical application, we constructed a nomogram based on data from the GSE25066 database in order to predict patient survival probability by weighing age, ER, PR, HER2, grade, stage, lymph nodes, tumor, and the signature-based risk score (Fig. 4D and Supplementary material 3). The above results suggest that the novel signature of the 12 genes can not only evaluate NAC response but also predict prognosis in breast cancer.

Compound Screening for Reversing Breast Cancer Resistance

We further explored other potential drugs that could be used for the treatment of patients in the high-risk group. We compared the transcriptome data of samples from the high- and low-risk groups. Under the conditions of FoldChange >1, 169 genes were upregulated, and 92 genes were downregulated in the high-risk group in GSE25066. Differentially expressed genes were imported into the CMap database and the top 30 potential drugs that could be used to treat high-risk patients were screened (Fig. 5A). The top three compounds are bambuterol (bronchodilators) [32], pravastatin (lipid-lowering agents) [33], and isocarboxazid (antidepressant, a non-selective and irreversible inhibitor of monoamine oxidase) [34]. It is noteworthy that there were five anticancer drugs among the candidate compounds. They are imexon (alkylating agents) [35], temozolomide (alkylating agents) [36], axitinib (inhibition of tumor growth and phosphorylation of VEGFR-2) [37], semaxanib (VEGFR (Flk-1/KDR) inhibitor) [38] and crizotinib (ATP competitive multi-target protein kinase inhibitor that inhibits met/ALK/ROS) [39]. Furthermore, we used the previously induced chemo-resistant cell line MCF7/EPI to verify the function of the selected candidate drugs to reverse chemoresistance. Here, we selected the top three compounds (bambuterol, pravastatin, and isocarboxazid) and five antitumor drugs (imexon, temozolomide, axitinib, semaxanib, and crizotinib) as candidate compounds for verification. Pravastatin, isocarboxazid, imexon, axitinib, and crizotinib had significant cytotoxic effects on MCF7/EPI cells (Fig. 5B). We also observed the morphology of cells in each group after 72 h of drug treatment. Cells treated with pravastatin, isocarboxazid, temozolomide, crizotinib, and fiacitabine were swollen, had many protrusions, and tended to die (Fig. 5C). Finally, we detected the apoptosis level of cells in each group using a TUNEL apoptosis detection kit. This is consistent with the previous results that groups treated with pravastatin, isocarboxazid, imexon, axitinib, and crizotinib displayed numerous dead cells (Fig. 5D). This suggests that pravastatin, isocarboxazid, imexon, axitinib, and crizotinib could inhibit or kill chemo-resistant cells.

Discussion

In this study, we constructed a NAC response risk model based on GGI and obtained a novel signature of twelve different genes to evaluate NAC response and predict prognosis in breast cancer. Through bioinformatic analysis of the 12 genes, we found that breast cancer NAC-resistant cells have powerful survival strategies, such as cell cycle regulation, DNA repair, transport, and efflux. TME analysis showed that there were a large number of exhausted tumor-infiltrating lymphocytes (TILs) in the tumor tissues of patients in the high-risk group. These results indicate possible strategies to reverse breast cancer resistance. Another important result of this study is that we screened out potential drugs targeting the gene expression characteristics of patients in the high-risk group

through CMap. We further verified the cytotoxic effects of these drug candidates in the self-induced breast cancer chemoresistance cell line MCF7/EPI. According to the cytotoxicity assays, ravastatin, isocarboxazid, imexon, axitinib, and crizotinib may be potential drugs to inhibit or kill chemo-resistant cells.

Bioinformatics analysis suggested that these 12 genes promote breast cancer resistance to NAC by upregulating DNA repair, metabolism-related pathways, and downregulating some membrane receptor signaling mechanisms. Based on these characteristics, we found that the neoadjuvant chemotherapy-resistant cells in breast cancer are in a dormant state and there is increasing evidence that non-genetic processes drive drug tolerance, presenting a major hurdle to successful cancer therapy [40,41]. Drug-tolerant persister (DTP) cells are emerging as key players in the field of non-genetic heterogeneity of tumors and have been identified across a wide range of tumors in response to chemotherapy and targeted agents [42,43,44,45]. Therefore, DTPs represent a potential therapeutic opportunity prior to the development of classic irreversible genetic mutation-driven drug resistance. Here, the 12 genes may be a potential target for reversing the drug resistance of tumor cells by breaking the "cold state" of drug-resistant cells. This hypothesis can be verified by knocking out or overexpressing these 12 genes.

An increasing number of studies have confirmed that in addition to the internal factors of tumor cells, the tumor microenvironment also plays a key role in tumor drug resistance [29,30,31]. Among them, TILs are widely recognized as one of the most promising targets for reversing tumor drug resistance. Our results showed that there were a large number of TILs (such as CD8+ T cells, CD4+, and NK cells) in the high-risk group which were exhausted. This also explains why patients in the high-risk group have a poor prognosis despite abundant immune cell infiltration. Such patients may benefit from adoptive cellular immunotherapy. A clinical trial reported a patient with breast cancer who still had extensive metastases after surgery, chemotherapy, and targeted therapy. After 22 months of treatment with TILs, the tumor completely disappeared, and the patient survived [46]. Our results also showed that the tumors of patients in the high-risk group highly expressed immunosuppressive molecules such as *CTLA4, LAG3, ICOS, IDO1,* and *ADORA2A*. Thus, *CTLA4* immunosuppressants may contribute to further treatment of these patients.

The results of cytotoxicity assays showed that pravastatin, isocarboxazid, imexon, axitinib, and crizotinib, as selected candidate drugs based on CMap, have significant cytotoxic effects on MCF7/EPI. Although imexons, axitinib, and crizotinib are known antitumor drugs, their effect on chemo-resistant tumors have not yet been reported. Pavastatin and isocarboxazid A and B were originally used to treat hyperlipidemia and depression, but their significant killing effect on drug-resistant cells suggests that these two drugs may have other mechanisms of action to inhibit tumors. The antitumor activity of these drugs needs to be further confirmed by in vivo and in vitro trials.

Conclusion

Our results suggest a novel signature of the 12-gene to evaluate neoadjuvant chemotherapy response and predict prognosis in breast cancer.

Abbreviations

BP: Biological process

CC: Cellular component

DOC: Docetaxel

DTP: Drug-tolerant persister

EPI: Epirubicin

FBS: Fetal bovine serum

GEO: Gene Expression Omnibus

GGI: Genomic grade index

GSEA: Gene set enrichment analysis

LASSO: Least absolute shrinkage and selection operator

MF: Molecular function

RCB: Residual cancer burden

ROC: Receiver operating characteristic

TIL: Tumor-infiltrating lymphocytes

WT: Wild-type

Declarations

Ethics approval and consent to participate:

All procedures performed in this study were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Consent for publication:

Informed consent was obtained from all individual participants included in the study.

Ethical approval

All procedures performed in this study were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Availability of data and materials:

Funding

This work was supported by grants from the National Natural Science Foundation of China (81302315, 82102992), Natural Science Foundation of Chongqing (cstc2021jcyj-msxmX0473) and Personnel Training Program of Army Medical University (XZ-2019-505-045).

Authors' contribution

PT and HZ contributed substantially to the conception and design, and gave final approval of the version to be published. JW, LB collected and analyzed the data and performed the expression verification of clinical samples. JW and YT wrote the manuscript. WL, YZY, MHW and YH critically revised the manuscript. All authors contributed to the article and approved the submitted version.

Conflict of interest

All authors declare that they have no conflict of interest.

References

1. https://www.iarc.who.int/faq/latest-global-cancer-data-2020-qa/
2. Mittendorf EA, Jeruss JS, Tucker SL, et al. Validation of a novel staging system for disease-specific survival in patients with breast cancer treated with neoadjuvant chemotherapy. J Clin Oncol. 2011;29(15):1956-1962.
3. Radovich M, Jiang G, Hancock BA, et al. Association of Circulating Tumor DNA and Circulating Tumor Cells After Neoadjuvant Chemotherapy With Disease Recurrence in Patients With Triple-Negative Breast Cancer: Preplanned Secondary Analysis of the BRE12-158 Randomized Clinical Trial. JAMA Oncol. 2020;6(9):1410-1415.
4. Wang H, Mao X. Evaluation of the Efficacy of Neoadjuvant Chemotherapy for Breast Cancer. Drug Des Devel Ther. 2020; 14:2423-2433.
5. Kim R, Kawai A, Wakisaka M, et al. Immune factors associated with the pathological and therapeutic effects of preoperative chemotherapy in patients with breast cancer. Transl Oncol. 2021;14(1):100927.
6. Lin L, Tong X, Hu P, Invernizzi M, Lai L, Wang LV. Photoacoustic Computed Tomography of Breast Cancer in Response to Neoadjuvant Chemotherapy. Adv Sci (Weinh). 2021;8(7):2003396.
7. Metzger Filho O, Ignatiadis M, Sotiriou C. Genomic Grade Index: An important tool for assessing breast cancer tumor grade and prognosis. Crit Rev Oncol Hematol. 2011 Jan;77(1):20-9.
8. Longacre TA, Ennis M, Quenneville LA, et al. Interobserver agreement and reproducibility in classification of invasive breast carcinoma: an NCI breast cancer family registry study. Mod Pathol. 2006 Feb;19(2):195-207.
9. Sotiriou C, Wirapati P, Loi S, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst. 2006 Feb 15;98(4):262-72.
10. Bertucci F, Ueno NT, Finetti P, et al. Gene expression profiles of inflammatory breast cancer: correlation with response to neoadjuvant chemotherapy and metastasis-free survival. Ann Oncol. 2014;25(2):358-365.
11. Echeverria GV, Ge Z, Seth S, et al. Resistance to neoadjuvant chemotherapy in triple-negative breast cancer mediated by a reversible drug-tolerant state. Sci Transl Med. 2019;11(488):eaav0936.
12. Wirapati P, Sotiriou C, Kunkel S, et al. Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. Breast Cancer Res. 2008;10(4):R65.
13. Liedtke C, Hatzis C, Symmans WF, et al. Genomic grade index is associated with response to chemotherapy in patients with breast cancer. J Clin Oncol. 2009;27(19):3185-3191.

14. Symmans WF, Peintinger F, Hatzis C, et al. Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. J Clin Oncol. 2007 Oct 1;25(28):4414-22.
15. Sotiriou C, Wirapati P, Loi S, Harris A, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst. 2006 Feb 15;98(4):262-72.
16. Glynn S A, Gammell P, Heenan M et al. A new superinvasive in vitro phenotype induced by selection of uman breast carcinoma cells with the chemotherapeutic drugs paclitaxel and doxorubicin. British Journal of Cancer 2004;91(10):1800-1807.
17. Martinez V G, O Connor R, Liang Y et al. CYP1B1 expression is induced by docetaxel: effect on cell viability and drug resistance. British Journal of Cancer. 2008; 98(3):564-570.
18. Tao Qian, Chengmin Liu, Yanyan Ding, et al. PINCH-1 interacts with myoferlin to promote breast cancer progression and metastasis. Oncogene. 2020 Mar;39(10):2069-2087.
19. Qingjian Li, Tao Qin, Zhuofei Bi, et al. Rac1 activates non-oxidative pentose phosphate pathway to induce chemoresistance of breast cancer. Nat Commun. 2020;11, 1456.
20. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013 Jan 16;14:7.
21. Zhang XF, Ou-Yang L, Zhao XM, et al. Differential network analysis from cross-platform gene expression data. Sci Rep. 2016; 6:34112.
22. López-Ratón M, Rodríguez-Álvarez MX, Suárez CC, et al. OptimalCutpoints: an R package for selecting optimal cutpoints in diagnostic tests. J Stat Softw 2014; 61:1–36.
23. Son J, Lee SE, Kim EK, et al. Prediction of breast cancer molecular subtypes using radiomics signatures of synthetic mammography from digital breast tomosynthesis. Sci Rep. 2020;10(1):21566.
24. Zhang Z, Jing J, Ye Y, et al. Characterization of the dual functional effects of heat shock proteins (HSPs) in cancer hallmarks to aid development of HSP inhibitors. Genome Med. 2020;12(1):101.
25. Kim J, Piao HL, Kim BJ, et al. Long noncoding RNA MALAT1 suppresses breast cancer metastasis. Nat Genet. 2018;50(12):1705-1715.
26. Lamb J, Crawford ED, Peck D, et al. The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease. Science (2006) 313:1929–35.
27. Subramanian A, Narayan R, Corsello SM, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell. 2017 Nov 30;171(6):1437-1452.e17.
28. Wang G, Xu J, Zhao J, et al. Arf1-mediated lipid metabolism sustains cancer cells and its ablation induces anti-tumor immune responses in mice. Nat Commun. 2020;11(1):220. Published 2020 Jan 10.
29. Graeser M, Feuerhake F, Gluz O, et al. Immune cell composition and functional marker dynamics from multiplexed immunohistochemistry to predict response to neoadjuvant chemotherapy in the WSG-ADAPT-TN trial. J Immunother Cancer. 2021 May;9(5):e002198.
30. Gu SS, Wang X, Hu X, et al. Clonal tracing reveals diverse patterns of response to immune checkpoint blockade. Genome Biol. 2020;21(1):263.
31. Pietilä EA, Gonzalez-Molina J, Moyano-Galceran L, et al. Co-evolution of matrisome and adaptive adhesion dynamics drives ovarian cancer chemoresistance. Nat Commun. 2021;12(1):3904.
32. Laursen LC. Bambuterol [Bambuterol]. Ugeskr Laeger. 1992 Feb 17;154(8):505-6. Danish.
33. Al-Badr AA, Mostafa GA. Pravastatin sodium. Profiles Drug Subst Excip Relat Methodol. 2014;39: 433-513.
34. Zisook S. Isocarboxazid in the treatment of depression. Am J Psychiatry. 1983 Jun;140(6):792-4.
35. Evens AM, Prachand S, Shi B, et al. Imexon-induced apoptosis in multiple myeloma tumor cells is caspase-8 dependent. Clin Cancer Res. 2004 Feb 15;10(4):1481-91.
36. Schreck KC, Grossman SA. Role of Temozolomide in the Treatment of Cancers Involving the Central Nervous System. Oncology (Williston Park). 2018 Nov 15;32(11):555-60, 569.
37. Rini BI, Plimack ER, Stus V, et al. Pembrolizumab plus Axitinib versus Sunitinib for Advanced Renal-Cell Carcinoma. N Engl J Med. 2019 Mar 21;380(12):1116-1127.
38. Sakamoto KM. Semaxanib (SUGEN). IDrugs. 2001 Sep;4(9):1061-7.
39. Ayoub NM, Ibrahim DR, Alkhalifa AE, et al. Crizotinib induced antitumor activity and synergized with chemotherapy and hormonal drugs in breast cancer cells via downregulating MET and estrogen receptor levels. Invest New Drugs. 2021 Feb;39(1):77-88.
40. Echeverria GV, Ge Z, Seth S, et al. Resistance to neoadjuvant chemotherapy in triple-negative breast cancer mediated by a reversible drug-tolerant state. Sci Transl Med. 2019 Apr 17;11(488): eaav0936.
41. Recasens A, Munoz L. Targeting Cancer Cell Dormancy. Trends Pharmacol Sci. 2019 Feb;40(2):128-141.
42. Guler GD, Tindell CA, Pitti R, et al. Repression of Stress-Induced LINE-1 Expression Protects Cancer Cell Subpopulations from Lethal Drug Exposure. Cancer Cell. 2017 Aug 14;32(2):221-237.e13.
43. Hangauer MJ, Viswanathan VS, Ryan MJ, et al. Drug-tolerant persister cancer cells are vulnerable to GPX4 inhibition. Nature. 2017 Nov 9;551(7679):247-250.
44. Liau BB, Sievers C, Donohue LK, er al. Adaptive Chromatin Remodeling Drives Glioblastoma Stem Cell Plasticity and Drug Tolerance. Cell Stem Cell. 2017 Feb 2;20(2):233-246.e7.
45. Sharma SV, Lee DY, Li B, et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell. 2010 Apr 2;141(1):69-80.

46. Zacharakis N, Chinnasamy H, Black M, et al. Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer. Nat Med. 2018 Jun;24(6):724-730.

Tables

Table 1

EPI IC50 values in breast cancer cell lines and their chemoresistant variants (n=3)

IC50 values (ng/mL)	WT	EPI resistance	Fold
MCF-7	25.37±1.93	235.93±6.27	9.30
SKBR3	15.66±0.86	120.27±3.24	7.68
MDA-MB-231	12.53±0.42	171.26±3.73	13.67

Results are expressed as Mean ±S.D. and represent the average of three independent experiments. Fold resistance of each variant is shown in bold and represents the IC50 value of the variants divided by the IC50 value of the WT cells for each particular drug tested. *Indicates significance (P<0.05).

Table 2

The function of the twelve genes

Gene	Full name	Function summary
HJURP	Holliday junction recognition protein	a Protein Coding gene related to Cell Cycle, Mitotic and Chromatin Regulation / Acetylation
IFI27	interferon alpha inducible protein 27 like 2	a Protein Coding gene related to RNA polymerase II activating transcription factor binding and lamin binding.
RAD51AP1	RAD51 associated protein 1	a Protein Coding gene related to RNA binding and single-stranded DNA binding
EZH2	enhancer of zeste 2 polycomb repressive complex 2 subunit	The protein encoded by this gene is a member of the Polycomb-group (PcG) family
DNMT3B	DNA methyltransferase 3 beta	The protein encoded by this gene is a DNA methyltransferase which is thought to function in de novo methylation, rather than maintenance methylation.
SLC7A5	solute carrier family 7 member 5	a protein Coding gene related to t peptide antigen binding and antiporter activity
DBF4	DBF4 zinc finger	a Protein Coding gene related to nucleic acid binding and enzyme activator activity
USP18	ubiquitin specific peptidase 18	The protein encoded by this gene belongs to the ubiquitin-specific proteases (UBP) family of enzymes that cleave ubiquitin from ubiquitinated protein substrates.
ELOVL5	ELOVL Fatty Acid Elongase 5	This gene belongs to the ELO family and involved in the elongation of long-chain polyunsaturated fatty acids
PTGER3	Prostaglandin E Receptor 3	The protein encoded by this gene is a member of the G-protein coupled receptor family
KIAA1324	estrogen-induced gene 121	estrogen-induced gene
CYBRD1	Cytochrome B Reductase 1	a member of the cytochrome b(561) family that encodes an iron-regulated protein

Gene name	GO
HJURP	GO:0005515;GO:0005694;GO:0003677;GO:0005739;GO:0005829;GO:0005730;GO:0005634;GO:0007049;GO:0042393;GO:0042802;GO:000705
IFI27	GO:0016021;GO:0016020;GO:0060135;GO:0034097;GO:0032355
RAD51AP1	GO:0006281;GO:0003723;GO:0005634;GO:0003690;GO:0003697;GO:0005515;GO:0003677;GO:0006974;GO:0006310;GO:0005654;GO:001056
EZH2	GO:0018024;GO:0034968;GO:0048511;GO:0005515;GO:0000790;GO:0016569;GO:0031490;GO:0006355;GO:0006351;GO:0032259;GO:001674
DNMT3B	GO:0005737;GO:0005634;GO:0071549;GO:0071455;GO:0045892;GO:0045666;GO:0042493;GO:0042220;GO:0033189;GO:0032355;GO:003100
SLC7A5	GO:0005886;GO:0005829;GO:0016020;GO:0043231;GO:0016021;GO:0015171;GO:0003333;GO:0015807;GO:0015179;GO:1902475;GO:000573
DBF4	GO:0005515;GO:0043085;GO:0005634;GO:0046872;GO:0008270;GO:0003676;GO:0007049;GO:0006260;GO:0005654;GO:0000082;GO:000804
USP18	GO:0006508;GO:0008234;GO:0008233;GO:0005515;GO:0016787;GO:0016579;GO:0005829;GO:0005737;GO:0005634;GO:0036459;GO:000651
ELOVL5	GO:0005783;GO:0042761;GO:0034626;GO:0019367;GO:0006633;GO:0006631;GO:0102338;GO:0102337;GO:0102336;GO:0009922;GO:000662
PTGER3	GO:0007165;GO:0004871;GO:0005886;GO:0007186;GO:0004930;GO:0016021;GO:0016020;GO:0004955;GO:0004957;GO:0005887;GO:000563
KIAA1324	GO:0005886;GO:0005765;GO:0005794;GO:0016021;GO:0016020;GO:0003723;GO:0006914;GO:0005764;GO:0005768;GO:0031902;GO:007006
CYBRD1	GO:0005515;GO:0055114;GO:0016491;GO:0046872;GO:0016021;GO:0016020;GO:0005886;GO:0070062;GO:0000293;GO:0010039;GO:003152

Table 4

Bioactivity of candidate drugs

Drug	Bioactivity
bambuterol	Bambuterol hydrochloride ((±)-Bambuterol hydrochloride; KWD-2183 hydrochloride) is a long acting beta-adrenoceptor agonist (LABA) used in the treatment of asthma; it also is a prodrug of terbutaline.
pravastatin	a competitive HMG-CoA reductase inhibitor against sterol synthesis
isocarboxazid	a non-selective and irreversible inhibitor of monoamine oxidase
imexon	an iminopyrrolidone aziridine with anti-cancer activity.
temozolomide	an oral active DNA alkylating agent and a proautophagic and proapoptotic agent,which against tumor cells that are characterized by low levels of O6-alkylguanine DNA alkyltransferase (OGAT) and a functional mismatch repair system.
axitinib	a multi-targeted tyrosine kinase inhibitor for VEGFR1, VEGFR2, VEGFR3 and PDGFRβ, respectively.
semaxanib	Selective VEGFR (Flk-1/KDR) inhibitor
crizotinib	an ATP-competitive ALK, c-Met and ROS1 inhibitor, which has effective tumor growth inhibition

Figures

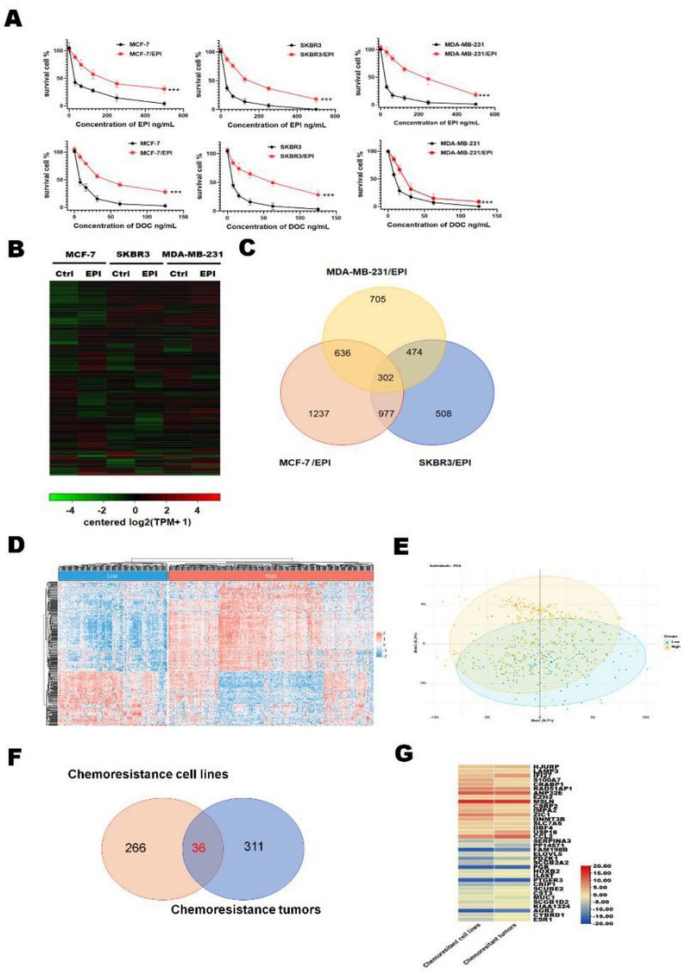

Figure 1

Visualization of genes differentially expressed in breast cancer chemoresistance cell. A, Drug toxicity of EPI and DOC to resistant cells. Data are presented as mean ± SD ($P < 0.001$). **B**, Heatmaps of DEGs in EPI-resistance cells. **C**, Over-lapping DEGs that were up/downregulated over 2-folds among EPI-resistance cells lines. **D**, Heat map of DEGs that were up/downregulated over 1-fold among breast cancer chemotherapy-resistant (n = 157) and sensitive patients (n = 335). **E**, Principal component comparison of gene expression in breast cancer chemotherapy-resistant and sensitive patients. **F**, Over-lapping DEGs among EPI-resistance cells lines and tumor of chemotherapy-resistant patients.

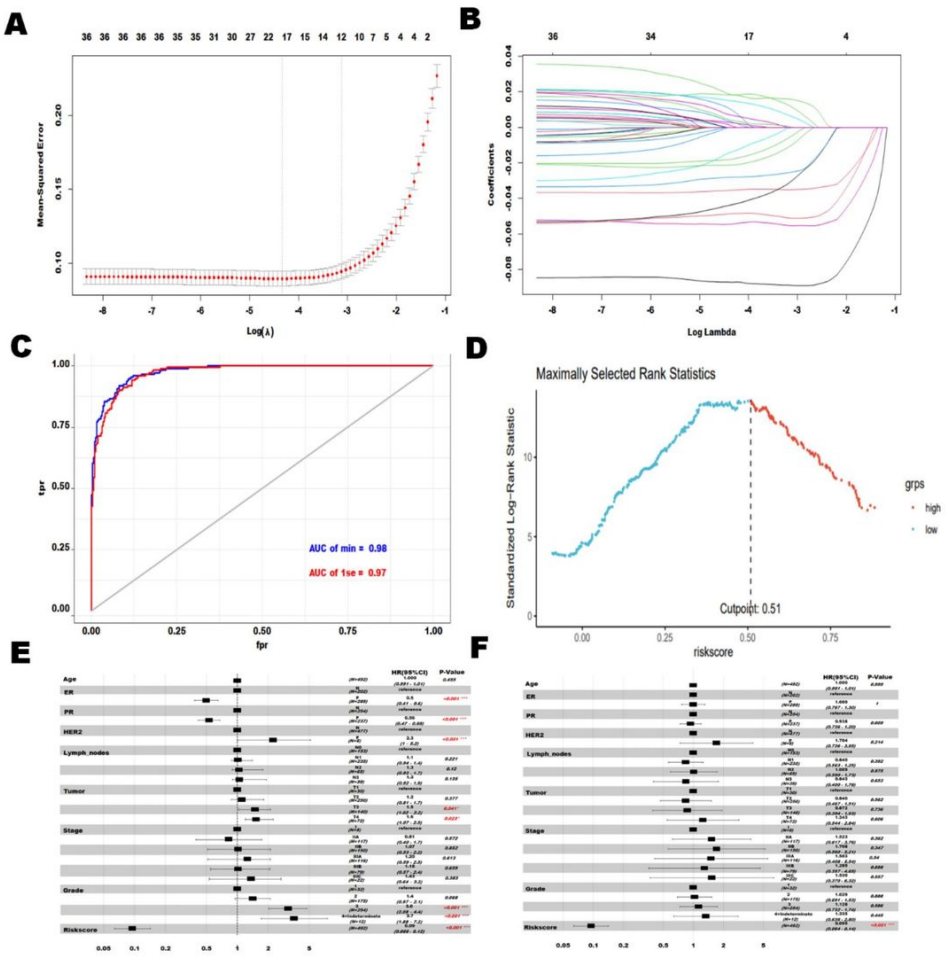

Figure 2

Risk model with neoadjuvant chemotherapy response constructed based on GGI lever between the two clusters. A, B, LASSO regression prognostic model of breast cancer patients treated with neoadjuvant chemotherapy (n=246). **C,** Receiver operating characteristic (ROC) analysis of the risk score in breast cancer patients treated with neoadjuvant chemotherapy. **D**, Cut-point of the risk score in breast cancer patients treated with neoadjuvant chemotherapy. **E**, **F**, Univariate and multivariate analysis of the twelve-gene neoadjuvant chemotherapy response risk model. All data are from the randomly selected training set in GSE25066 dataset. Significant difference between the two groups were: *P < 0.05; **P < 0.01; and ***P < 0.001.

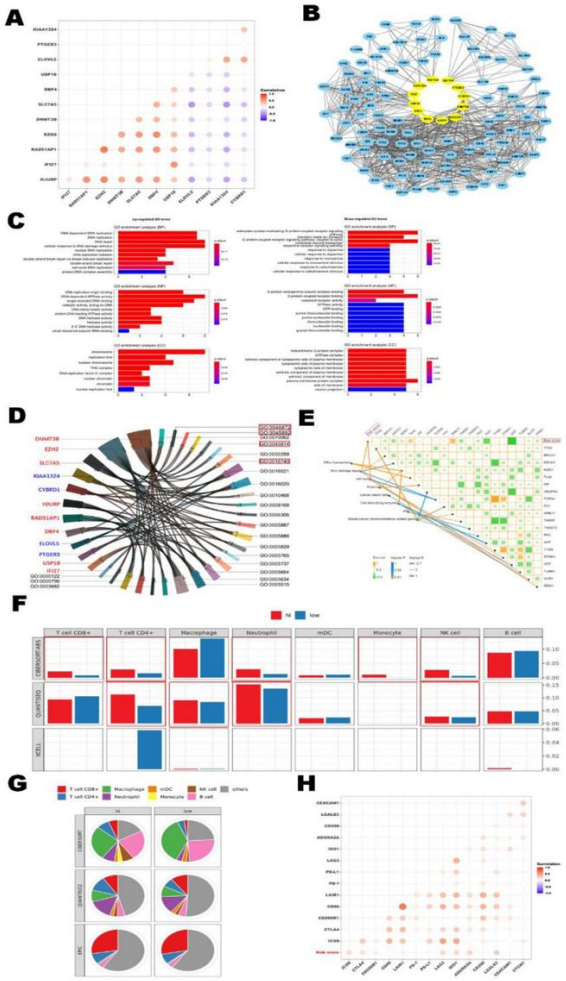

Figure 3

Characteristics of tumor cells and tumor microenvironment in high-risk patients. A, The correlation analysis of the expression of 12 genes. **B**, Protein network interaction map for 12 genes based on the HitPredict database. **C**, GO enrichment analysis of the 12 genes (P < 0.01). **D**, Chord plot of interaction between twelve genes and GO enrichment terms (P < 0.01). **E**, The correlation analysis of risk score and gene expression in breast cancer resistance-related pathways. **F**, **G**, The difference analysis of TIL levels between high and low-risk group. **H**, Heat map of the correlation between risk score and the expression of multiple immunosuppressive regulatory molecules.

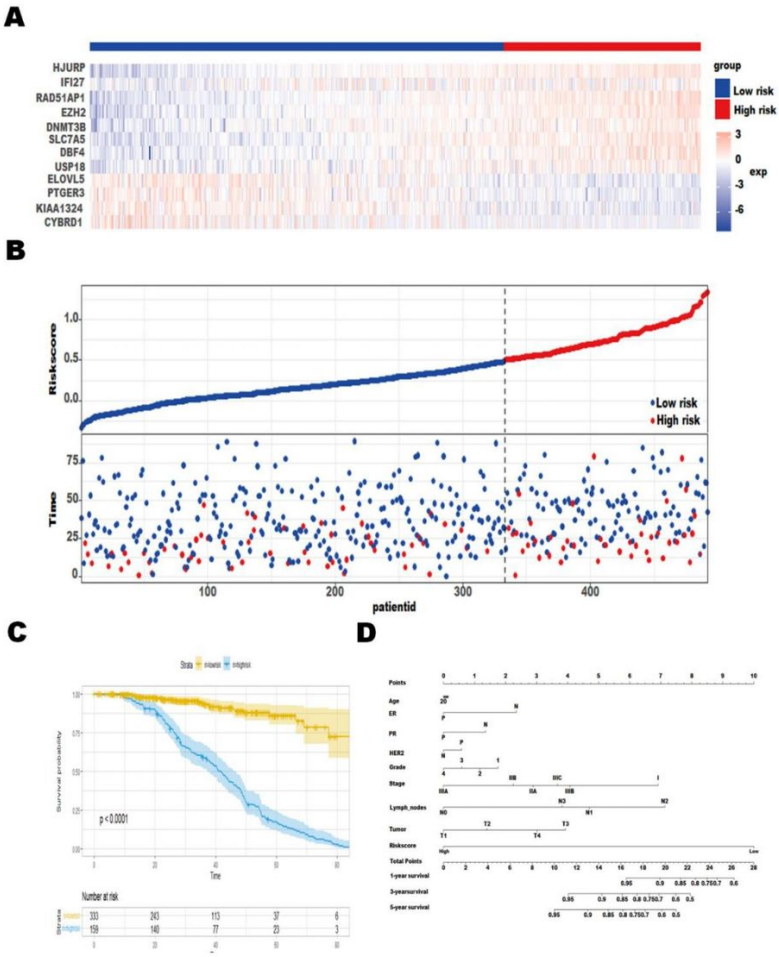

Figure 4

12-genes to predicts prognosis in breast cancer. A, The expression heatmap of 12 genes in patients in different risk groups (high-risk n=157, low-risk n=335). **B**, Survival time, survival status, and risk score of patients in different risk groups (high-risk, n=157; low-risk, n=335). **C**, Kaplan–Meier survival analysis of the different patient risk groups (high-risk n=157, low-risk n=335, P < 0.001). **D**, A nomogram for clinical diagnosis was constructed based on clinical characteristics and risk scores.

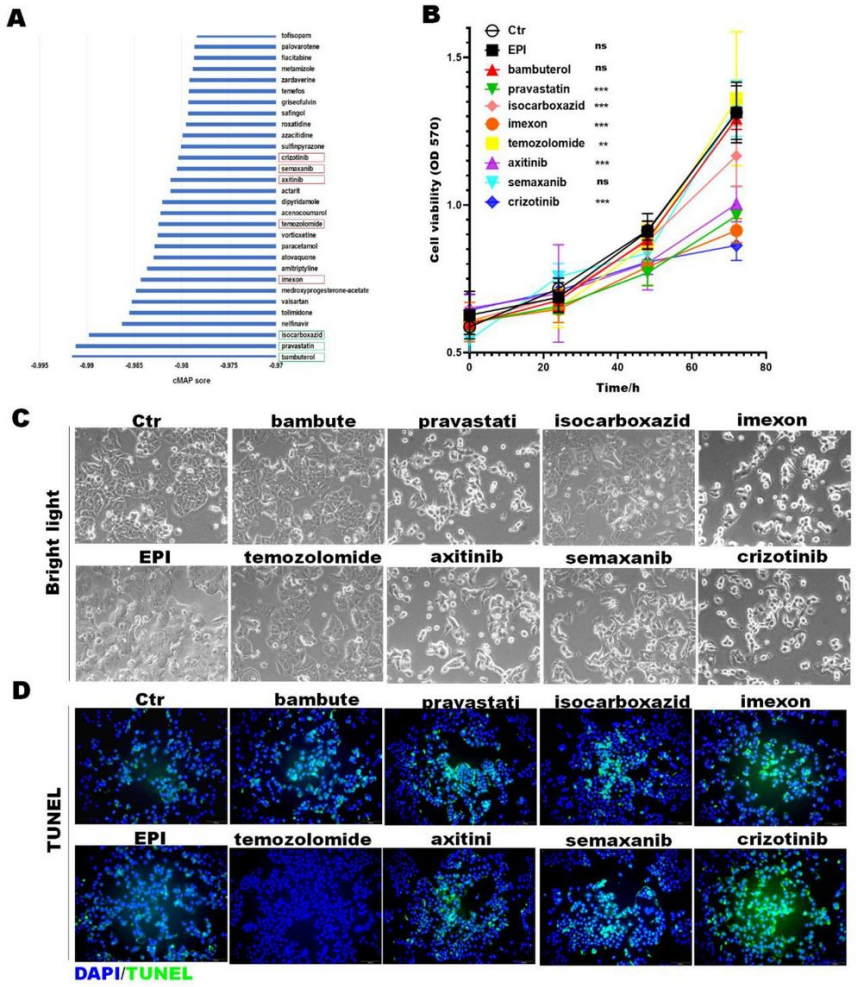

Figure 5

Screening potential drugs for the treatment of high-risk patients. A, The Connectivity Map (CMap) database was used to screen potential drugs for the treatment of high-risk patients. The three best drug candidates are in the green box, and the known anticancer drugs in the TOP30 are in the red box. **B**, Cytotoxic effects of preferred drug candidates on MCF-7/EPI after 72 h of drug treatment. Data are presented as mean ± SD. Significant difference between the two groups: *P < 0.05; **P < 0.01; and ***P < 0.001. **C**, The morphology of cells in each group after 72 h of drug treatment. **D**, The apoptosis level of cells in each group after 72 h of drug treatment.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial1.docx
- Supplementarymaterial2.docx
- Supplementarymaterial3.docx