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

Jin Wu
Department of Breast and Thyroid Surgery, Southwest Hospital, Army Medical University

Yuan Tian
Department of General surgery, Linyi People's Hospital

Bin Liao
Department of Neurosurgery, Chongqing General Hospital

Hong Zheng
Department of Thoracic Surgery, Xinqiao Hospital, Army Medical University

Wei Liu
Department of Breast and Thyroid Surgery, Southwest Hospital, Army Medical University

Yuzhao Yan
Department of Breast and Thyroid Surgery, Southwest Hospital, Army Medical University

Ying Hu
Department of Breast and Thyroid Surgery, Southwest Hospital, Army Medical University

Minghao Wang
Department of Breast and Thyroid Surgery, Southwest Hospital, Army Medical University

Peng Tang (✉ tp1232000@sina.com)
Department of Breast and Thyroid Surgery, Southwest Hospital, Army Medical University

Research Article

Keywords: Breast cancer, Neoadjuvant chemotherapy evaluation, Chemo-resistant, Breast cancer prognosis

Posted Date: August 2nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1308081/v2

License: ☑️ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background

Accurate evaluation of breast cancer response to neoadjuvant chemotherapy (NAC) provides important information about systemic therapies, which elucidates tumor biology and prognosis and guides further therapies. Gene profiles overcome the limitations of some indicators of the classical pathological evaluation criteria and the subjectivity of observation, but are complicated and expensive. Therefore, it is essential to develop a more accurate, repeatable, and economical evaluation approach for NAC responses.

Methods

We analyzed the transcriptional profiles of epirubicin-resistant breast cancer cell lines and the tumors derived from NAC-resistant patients from the GEO25066 dataset. We initially screened common significantly differentially expressed genes and constructed a NAC response risk model using LASSO regression and univariate and multivariate analyses. Bioinformatics Analysis were performed to find the differences in bioinformatic features of tumor cells, immune characteristics, and prognosis between the high- and low-risk groups. We screened candidate drugs that could reverse chemotherapy resistance in breast cancer by the Connectivity Map database. Cytotoxicity assay and TUNEL assay were performed to verify the killing effect of drug candidates on drug-resistant breast cancer cells.

Results

Thirty-six genes were commonly up/down-regulated in both the chemo-resistant tumors and cell lines, compared to the sensitive tumors and wild-type cell lines. We obtained a risk model composed of 12 genes through LASSO regression. The risk model divided patients into high- and low-risk groups. Univariate and multivariate Cox regression analyses suggested that the risk score is an independent prognostic factor for the evaluation of NAC response to breast cancer. Tumors in these risk groups exhibited significant differences in molecular biological characteristics, tumor-infiltrating lymphocytes, and immunosuppressive molecule expression. Our results suggested that the risk score was also a good prognostic factor for breast cancer. Finally, our results suggest that pravastatin, isocarboxazid, imexon, axitinib, and crizotinib could inhibit or kill chemoresistant cells.

Conclusion

A novel 12 gene-signature could be used to evaluate NAC response and predict prognosis in breast cancer.
Background

Currently, breast cancer has the highest incidence rate among all cancers worldwide [1]. Neoadjuvant chemotherapy (NAC) refers to systemic cytotoxic drug treatment before surgery or radiotherapy and is considered the standard treatment regimen for patients with locally advanced or inoperable breast cancer [2, 3]. Accurate evaluation of tumor response to NAC provides important information about tumor biology and prognosis and guides further therapies [4–6]. In addition to clinical and pathological evaluation criteria, gene expression signatures have been developed to evaluate 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 represent an important step towards personalized breast cancer treatment [9–11]. In particular, gene profiles overcome the drawbacks of the relatively limited detection indicators of the classical pathological evaluation criteria and the subjectivity of observation [12].

GGI is a gene expression signature developed to ameliorate histologic grade assessment and to evaluate the response to chemotherapy [7]. Using residual cancer burden index, which is a more accurate pathological evaluation method used as a control, researchers studied 229 postoperative tumor samples from patients who had received NAC (paclitaxel, fluorouracil, doxorubicin, and cyclophosphamide). The results showed that the values of GGI increased with chemotherapy responsiveness [9, 13, 14]. It is noteworthy that, unlike pathological evaluation, GGI assessment is more reproducible. However, it involves a large number of genes (97 genes), resulting in high detection costs and difficulty in clinical application.

In this study, we analyzed the transcriptional patterns of breast cancer cell lines and tumors of NAC-resistant patients evaluated by GGI and screened candidate genes associated with chemoresistance. Furthermore, we constructed a NAC 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.

Methods

Cell Cultures and Chemo-resistant Cell Line Induction

Human breast cancer cell lines MCF-7 (luminal subtype), SKBR3 (HER2+ subtype), and MDA-MB-231 (triple-negative subtype) were purchased from Procell Life Science&Technology Co., Ltd. All cell lines were authenticated by profiling of short tandem repeat analysis. Cell lines were frequently tested for mycoplasma to ensure contamination-free conditions. MCF-7 cells were cultured in DMEM (Gibco, USA) 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 (Gibco, USA) with 10% FBS,
penicillin, and streptomycin. MDA-MB-231 cells were cultured in Leibovitz's L-15 medium with 10% FBS, penicillin, and streptomycin. Cells were incubated in 5% CO$_2$ at 37.5 °C.

Epirubicin (EPI; s1223, Selleck, CHN)-resistant variants of the human breast cancer cell lines were established by pulse selection. Cells were exposed to the respective maximal inhibitory concentration (IC90) values of EPI for 4 h, once a week for 10 weeks, to obtain resistant variants T47D/EPI, SKBR3/EPI, and MDA-MB-231/EPI [15, 16]. The EPI-resistant breast cancer cells were washed with PBS, fully lysed with TRIzol reagent (15596026, Invitrogen, USA) and preserved at -80 °C until use.

**Cytotoxicity Assay**

MCF-7, SKBR3, and MDA-MB-231 cells were seeded in 96-well plates at 5,000 cells/well. After the cells adhered completely, gradient concentrations of EPI and docetaxel (DOC) were added. After 72 h, cells were stained with sterile methylthiazolyldiphenyl-tetrazolium bromide (MTT; C0009, Beyotime, CHN) in culture media (1:10) for 2 h at 37.5 °C. The absorbance of MTT was detected at 570 nm [17, 18].

MCF-7/EPI cells were seeded in a 96-well plate at 5,000 cells/well. 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 for 72 h, and cells were stained with sterile MTT in culture media (1:10) for 2 h at 37.5 °C. The absorbance of MTT was detected at 570 nm. The drug concentration was the same as that used for the connectivity map (CMap) (https://clue.io/).

**TUNEL Assay**

TUNEL assay was performed using TUNEL Apoptosis Assay Kit-FITC (C1088, Beyotime) according to the manufacturer's instructions. Briefly, the fixed cells with 4% paraformaldehyde and incubated with 0.3% Triton X-100 for 30 min at room temperature. Cells were rinsed three times with 1× PBS and then incubated in TdT reaction mixture for 1 h at 37°C in dark. After three washes with 1× PBS, Cells were imaged under fluorescence microscope.

**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. Genomic DNA contamination of samples was eliminated by RNase-free DNase I. Thereafter, RNA was assessed using a Nano Photometer® spectrophotometer (IMPLEN, CA, USA) and a Qubit® 2.0 Fluorometer (Invitrogen, USA). The 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 differential gene expression profiles of NAC-resistant and -sensitive patients were analyzed using the R language package (limma 3.20.9). Gene ontology (GO) and the Kyoto encyclopedia of genes and genomes (KEGG) pathway analysis were applied to annotate the biological functions of differentially expressed genes (DEGs) by the R language package (GO plot, KEGG plot function R). The hallmarks of breast cancer chemo-resistant cells were investigated using gene set enrichment analysis (GSEA) (http://www.gsea-msigdb.org/gsea/downloads) [19].

Diagnostic Model Construction and Validation

After excluding the patient samples with missing data, a total of 492 samples remained in the dataset and were randomly divided into the training (n=246) and validation sets (n=246). Using the “glmnet” R package, we performed the least absolute shrinkage and selection operator (LASSO) regression analysis [20]. The Youden index criterion was defined as the boundary-value or decision threshold corresponding to the maximum Youden index, which is the best classification boundary value [21]. This was also applied to select the optimal cutoff. We performed Cox regression analysis using the “survival” R package. The “predict” R package was used to obtain the risk score.

Bioinformatics Analysis

To identify the protein-protein interactions between the positive genes of risk score, we employed the search tool for the retrieval of interacting genes/proteins (STRING) [22]. The chord diagram was drawn using the Power BI software. The R package “survival” was used for univariate and multivariate analyses of the age, stage, ER, PR, HER2, the signature-based risk factor score, and grade to assess the correlation of NAC resistance with prognosis [23]. The survival curves of 12 genes of risk score were obtained from the Kaplan–Meier plot [24]. The CMap database was used to identify compounds that were negatively correlated with the input differential gene profile after testing on MCF-7 cells [25, 26].

Immune Cell Infiltration Analysis

We used CIBERSORT, QUANTISEQ, and XCELL analyses to assess immune cell infiltration within the tumor microenvironment using the GEO25066 database in the different risk groups [27]. The online analysis tool hiplot (https://hiplot.com.cn/) was used to analyze the correlation between immunosuppressive molecules and risk scores.

Construction and Assessment of the Nomogram

The nomogram was established by the R package “rms”. We evaluated the performance of the nomogram by generating a calibration chart.
Statistical Analysis

Using GraphPad Prism 8.0, data of three independent experiments were presented as mean ± SD for statistical analysis. Student's t-tests or Mann–Whitney U-tests were performed for comparison between two groups. The chi-square test was used to analyze the categorical variables between two groups. The correlation between the two groups was analyzed using Spearman's test. P < 0.05 was considered significantly different.

Results

Screening of Hub Genes related to NAC Resistance in Breast Cancer

To explore the chemoresistance promoting mechanism in breast cancer, we generated EPI-resistant cell lines of different breast cancer subtypes, 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. Drug-resistant breast cancer cells induced by EPI were also resistant to DOC. We then compared the expression profiles of EPI-resistant cells with those of WT cells. Three hundred and two genes were commonly upregulated or downregulated at least two-fold in MCF-7/EPI, SKBR3/EPI, and MDA-MB-231/EPI compared with those in parental MCF-7, SKBR3, and MDA-MB-231 cells (Fig. 1B and C). In addition, 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 parts, intracellular organelle, and nucleus of the breast cancer resistant cells, and were enriched in “negative regulation of biological process”, “negative regulation of cellular process”, and “regulation of cellular metabolic process”. The molecular functions of the upregulated DEGs were the “protein and transcription regulatory region DNA” and “RNA polymerase II proximal promoter sequence-specific DNA binding.” In contrast to upregulated DEGs, downregulated DEGs were enriched in the extracellular space of the breast cancer resistant cells in “cell adhesion” and “regulation of cell motility” (Additional file 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 the KEGG enrichment analysis, upregulated chemoresistance cell feature genes were enriched in the “HIF-1 signaling pathway”, “Pentose phosphate pathway”, “p53 signaling pathway”, and “DNA replication”. Downregulated chemoresistance cell feature genes were enriched in “Glutathione metabolism” and “Chemokine signaling pathway” (Additional file 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. To characterize more comprehensively 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. E2F and MYC targets, mTORC1 signaling, P53 pathway, and KRAS signal gene sets were significantly upregulated in resistant breast cancer cells (Additional file 1), suggesting that 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). The results of the principal component analysis indicated that the gene expression differences between the GGI-high and -low groups were significant (Fig. 1E). There were 347 DEGs between the two groups. Thirty-six genes were commonly up/down-regulated in both the NAC chemo-resistant cells and tumors compared in contrast to the WT cells and sensitive tumors (Fig. 1F and G).

**Risk Model with NAC Response was Constructed Based on GGI Level**

In order to construct a more simplified diagnostic model of NAC response, we established a 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 and min containing 12 and 18 genes, respectively. The ROC analysis presented that the 1se (AUC=0.97) and min (AUC=0.98) of characteristic genes both have good diagnostic values for evaluating the resistance of breast cancer to NAC (Fig. 2C). Considering the cost of detection for the patients, we selected the 1se set: HJURP, IFI27, RAD51AP1, EZH2, DNMT3B, SLC7A5, DBF4, USP18, ELOVL5, PTGER3, KIAA1324, and CYBRD1. The validation set indicated that the results were similar to those of the training set (Additional file 2). The complete names and main function of the 12 genes are listed in Table 2. The cutoff risk score was 0.51 (Fig. 2D). The results of multivariate and univariate analyses suggested that the risk factor score was an indicator for NAC response (Fig. 2E and F).

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

To explore the internal and external factors of NAC resistance in breast cancer cells, we analyzed the gene expression characteristics and tumor microenvironment of NAC-resistant cancer cells. According to the GSE25066 gene expression profile, a negative correlation was found between the expression of positive genes (HJURP, IFI27, RAD51AP1, EZH2, DNMT3B, SLC7A5, DBF4, USP18, and ELOVL5) and the expression of negative genes (PTGER3, KIAA1324, and CYBRD1) in the risk score model (Fig. 3A). Utilizing the Hit Predict database, we searched for proteins that interact with those encoded by the above twelve genes. By analyzing their protein interaction network, it was observed that the proteins encoded by these 12 genes and their related proteins have many-to-many complex interactions (Fig. 3B). The results of GO enrichment analyses suggested that 12 genes promote breast cancer resistance to NAC by upregulating DNA repair and metabolism-related pathways and downregulating membrane receptor signaling mechanisms (Fig. 3C and Table 3). KEGG pathway enrichment analysis showed that “transition metal ion transport” and the “G protein-coupled receptor signaling pathway” were enriched (Table 3). The chord plot showed that 12 genes had complex interactions with the enriched GO pathway (Fig. 3D). For example, EZH2 and DNMT3B were 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 promote drug resistance by increasing DNA repair, reducing cell macromolecule synthesis, and cell metabolism among dormant cells.
The risk score was positively related to some known molecules associated with chemotherapy resistance in breast cancer. Those molecules were involved in the pathways of cell cycle regulation, DNA repair, transport, and efflux (Fig. 3E).

Tumor-infiltrating lymphocytes (TILs) are critical components of the tumor microenvironment and are important external factors of chemotherapy resistance [28-30]. Therefore, we assessed the level of immune cell infiltration within the tumor microenvironment of patients in the different risk groups of the GEO25066 database. The high-risk group recruited more CD4, CD8, and NK cells than did the low-risk group (Fig. 3F and G). Interestingly, patients from the high-risk group showed a significantly lower survival rate even though they had more tumor-infiltrating cells, than those from the low-risk group (Fig. 2). We further explored the correlation between the risk score and immunosuppressive molecules in the GSE25066 dataset. Many immunosuppressive molecules, such as CTLA4, LAG3, ICOS, IDO1, and ADORA2A, were positively correlated with the risk score (Fig. 3H). All results indicate that a large amount of tumor-infiltrating cells were depleted in the high-risk group, leading to the failure of NAC for breast cancer and the lower survival rate of patients.

**Evaluation of the Risk Score on Prognosis of Overall Breast Cancer**

To confirm whether the risk score for evaluating NAC resistance was a good prognostic predictor of clinical outcomes, we conducted the following study. According to The Cancer Genome Atlas 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 (Additional file 3). We also found a association of high expression of all positive genes and low expression of all negative genes with poor prognosis of breast cancer (Additional file 3). The above results illustrate that the risk score composed of these 12 genes may also serve as a prognostic factor for breast cancer. The expression heatmap showed that high expression of RAD51AP1, EZH2, DNMT3B, SLC7A5, DBF4, and USP18 and low expression of ELOVL5, PTGER3, KIAA1324, and CYBRD1 was detected in patients of the high-risk group (Fig. 4A). The patients of the different risk groups showed a difference in risk score and patient survival (Fig. 4B). The result of Kaplan–Meier survival analysis suggested that high risk means poor prognosis (Fig. 4C). Based on the GSE25066 database, we also constructed a nomogram to facilitate clinical application. Using the nomogram, the patient survival probability could be predicted by the weighing age, ER, PR, HER2, grade, stage, lymph nodes, tumor, and signature-based risk score. (Fig. 4D and Additional file 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**

To explore the use of candidate drugs to overcome NAC resistance in breast cancer, we analyzed the DEGs of two risk groups. We have screened 169 upregulated DEGs and 92 downregulated DEGs in the high-risk group of GSE25066 (| Fold Change | >1). To screen out the top 30 candidate drugs to overcome
NAC resistance in breast cancer, we imported DEGs into the CMap database (Additional file 4 and Fig. 5A). The top three compounds were bambuterol (bronchodilator) [31], pravastatin (lipid-lowering agent) [32], and isocarboxazid (antidepressant, a non-selective and irreversible inhibitor of monoamine oxidase) [33]. It is noteworthy that there were five anticancer drugs among the candidate compounds, namely, imexon (alkylating agent) [34], temozolomide (alkylating agent) [35], axitinib (inhibitor of tumor growth and phosphorylation of VEGFR-2) [36], semaxanib (VEGFR (Flk-1/KDR) inhibitor) [37] and crizotinib (ATP competitive protein kinase inhibitor met/ALK/ROS) [38]. Furthermore, we used the previously established chemoresistant cell line MCF7/EPI to verify the function of the selected candidate drugs to reverse chemoresistance. Here, we selected the top three compounds and the five antitumor drugs as candidates 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 flacitabine 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. The obtained data were consistent with the previous results showing that groups treated with pravastatin, isocarboxazid, imexon, axitinib, and crizotinib displayed numerous dead cells (Fig. 5D). Our results suggest that pravastatin, isocarboxazid, imexon, axitinib, and crizotinib could inhibit or kill chemoresistant cells.

**Discussion**

In this study, we constructed a NAC response risk model based on GGI and obtained a novel signature of 12 genes to evaluate NAC response and predict prognosis in breast cancer. Through bioinformatic analysis of the 12 genes, we found that NAC-resistant breast cancer cells have powerful survival strategies, such as cell cycle regulation, DNA repair, transport, and efflux. TME analysis showed that there were a large amount 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. The cytotoxic effects of these drug candidates were further verified in the self-induced chemo-resistant breast cancer cell line MCF7/EPI. According to the cytotoxicity assays, pravastatin, 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- and metabolism-related pathways and downregulating membrane receptor signaling mechanisms. Based on these characteristics, we found that the NAC-resistant cells in breast cancer were in a dormant state. Accumulating evidence reveals that non-genetic processes drive drug tolerance, regarded as a novel mechanism of failure in cancer therapy [11, 39]. The most recent studies found that as key players in the field of non-genetic heterogeneity of tumors, drug-tolerant persister (DTP) cells were confirmed to be associated with resistance to chemotherapy and targeted agents in a wide range of tumors [40–43]. Therefore, DTPs is a therapeutic opportunity before the patients develop irreversible genetic mutation-driven drug resistance. Here, the 12 genes serve as a potential target for
reversing 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.

The tumor microenvironment plays a crucial role in tumor drug resistance [28–30]. Among them, TILs are widely recognized as one of the most promising targets for reversing tumour 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 a poor prognosis is observed in the high-risk group although the patients have 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 [44]. Our results also showed that immunosuppressive molecules, such as CTLA4, LAG3, ICOS, IDO1, and ADORA2A, were highly expressed in the high-risk group. Thus, CTLA4 immunosuppressants may contribute to further treatment of these patients.

The results of the cytotoxicity assays showed that the CMap-based selected candidate drugs, pravastatin, isocarboxazid, imexon, axtinib, and crizotinib, have significant cytotoxic effects on MCF7/EPI. Although imexons, axtinib, and crizotinib are known antitumor drugs, their effect on chemoresistant tumors has not yet been reported. Pravastatin and isocarboxazid A and B were originally used to treat hyperlipidemia and depression, but their significant toxic 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 requires further validation.

**Conclusion**

Our results suggest that a novel signature of 12 genes can be used to evaluate NAC response and predict prognosis in breast cancer.

**List Of Abbreviations**

BP  
Biological process  
CC  
Cellular component  
CMap  
Connectivity map  
DEG  
Differentially expressed gene  
DOC  
Docetaxel  
DTP
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 to participate was obtained from all study participants.

Consent for publication:

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

Availability of data and materials:

The datasets generated and/or analyzed during the current study are available in the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/).
Competing interests

The authors declare that they have no competing interests

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 the Personnel Training Program of Army Medical University (XZ-2019-505-045).

Authors’ contributions

PT and MHW conceived the idea and research method of the study. JW and YT inducted the drug-resistant cell lines and conducted the cytotoxicity assays. JW and BL performed bioinformatics analysis. WL, YZY, and YH collected the samples. JW and HZ were responsible for the writing of the manuscript.

Acknowledgements

We would like to thank Editage (www.editage.cn) for English language editing.

References


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 |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MCF-7</td>
<td>25.37±1.93</td>
<td>235.93±6.27</td>
<td>9.30</td>
</tr>
<tr>
<td>SKBR3</td>
<td>15.66±0.86</td>
<td>120.27±3.24</td>
<td>7.68</td>
</tr>
<tr>
<td>MDA-MB-231</td>
<td>12.53±0.42</td>
<td>171.26±3.73</td>
<td>13.67</td>
</tr>
</tbody>
</table>

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

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

Figures

Figure 1
Visualization of genes differentially expressed in chemo-resistant breast cancer cells. **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-resistant cells. **C,** Overlapping DEGs that were up/downregulated over 2-fold among EPI-resistant cells lines. **D,** Heat map of DEGs that were up/downregulated over 1-fold among chemotherapy-resistant (n = 157) and -sensitive patients with breast cancer (n = 335). **E,** Principal component comparison of gene expression in chemotherapy-resistant and -sensitive patients with breast cancer. **F,** Overlapping DEGs among EPI-resistant cells lines and tumors of chemotherapy-resistant patients.
**Figure 2**

**Risk model with neoadjuvant chemotherapy response based on the GGI level between the two clusters.**

A, B, LASSO regression prognostic model of patients with breast cancer treated with NAC (n=246). C, ROC analysis of the risk score in patients with breast cancer treated with neoadjuvant chemotherapy. D, Cutpoint of the risk score in patients with breast cancer treated with NAC. E, F, Univariate and multivariate analysis of the twelve-gene NAC response risk model. *P < 0.05; **P < 0.01; and ***P < 0.001.
Figure 3

**Characteristics of tumour cells and tumour microenvironment in high-risk patients.**

**A,** Correlation analysis of the expression of 12 genes. **B,** Protein network interaction map for the 12 genes. **C,** GO enrichment analysis of the 12 genes (P < 0.01). **D,** Chord plot of interaction between the 12 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 the high- and low-risk groups. **h,** Heat map of the correlation between risk score and the expression of multiple immunosuppressive regulatory molecules.
Figure 4

Twelve genes to predict breast cancer prognosis. **A,** The expression heatmap of 12 genes in patients from 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 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. **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. Data are presented as mean ± SD. *P < 0.05; **P < 0.01; and ***P < 0.001.

**Supplementary Files**

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

- Table3.docx
- Supplementarymaterial.docx