BIOCELL
2022 46(5): 1215-1243

Tech Science Press

Prognostic tumor microenvironment gene and the relationship with immune infiltration characteristics in metastatic breast cancer

Lu YANG[1,#]; Yun LIU[1,#]; Boke ZHANG[2]; Mengsi YU[3]; Fen HUANG[1]; Yang WEN[1]; Jiangzheng ZENG[1]; Yanda LU[1,*]; Changcheng YANG[1,*]

[1] Department of Medical Oncology, The First Affiliated Hospital of Hainan Medical University, Haikou, 570102, China
[2] Clinical Laboratory Center, The First Affiliated Hospital of Anhui University of Chinese Medicine, Hefei, 230031, China
[3] Department of Clinical Laboratory, The First Affiliated Hospital of Xinjiang Medical University, Urumqi, 830054, China

Key words: Tumor microenvironment, Immune infiltration, Prognostic biomarker, Metastatic breast cancer

Abstract: The aim of this study was to reveal genes associated with breast cancer metastasis, to investigate their intrinsic relationship with immune cell infiltration in the tumor microenvironment, and to screen for prognostic biomarkers. Gene expression data of breast cancer patients and their metastases were downloaded from the GEO, TCGA database. R language package was used to screen for differentially expressed genes, enrichment analysis of genes, PPI network construction, and also to elucidate key genes for diagnostic and prognostic survival. Spearman's r correlation was used to analyze the correlation between key genes and infiltrating immune cells. We screened 25 hub genes, FN1, CLEC5A, ATP8B4, TLR7, LY86, PTGER3 and other genes were differentially expressed in cancer and paraneoplastic tissues. However, patients with higher expression of CD1C, IL-18 breast cancer had a better prognosis in the 10 years survival period, while patients with high expression of FN1, EIF4EBP1 tumors had a worse prognosis. In addition, TP53 and HIF1 genes are closely related to the signaling pathway of breast cancer metastasis. In this study, gene expression of ATP8B4 and CD1C were correlated with cancer tissue infiltration of $CD8^+$ T lymphocytes, while GSE43816, GSE62327 and TCGA databases showed that $CD8^+$ T lymphocytes were closely associated with breast cancer progression. Functional enrichment analysis of genes based on expression differences yielded key genes of prognostic value in the breast cancer microenvironment.

Introduction

Breast cancer is the most common malignancy in women worldwide and the most common cancer surpassed lung cancer (Sung *et al.*, 2021). Currently, with the increasing life expectancy of breast cancer and the rapid development of tumor diagnostic technology, breast cancer can be detected early, which lead to its increasing incidence (Sung *et al.*, 2021). In addition, great progress has been made in the comprehensive treatment of breast cancer, such as targeted anti-HER2 therapy and endocrine therapy, but the prognosis of patients is still not very optimistic (Qiu *et al.*, 2021; Yuan *et al.*, 2020). Since metastasis of breast cancer seriously affects the prognosis of patients, we still do not fully understand the underlying molecular mechanisms of its development. Therefore, it is of great importance to explore the underlying molecular mechanisms of breast cancer development by identifying new diagnostic and prognostic biomarkers and potential therapeutic targets.

Currently, numerous studies have confirmed that tumorigenesis, progression and prognosis are closely related to tumor tissue gene expression (Li *et al.*, 2021b; Lin *et al.*, 2017; Zhang *et al.*, 2019). The genetic expression features of the tumor microenvironment (TME) play an important impact in predicting the clinical outcome of tumor treatment (Llanos *et al.*, 2021; Yang *et al.*, 2021). The TME is mainly composed of cells and extracellular matrix. The cells in TME include lymphocytes, tumor-associated macrophages, tumor-associated fibroblasts, mesenchymal cells, endothelial cells, and other cells in addition to tumor cells. Immune cells and stromal cells are the two main non-tumor cells that are important for tumor diagnosis and assessment of tumor prognosis. Moreover, studies had found that altered gene expression in tumor tissue might lead to changes in non-tumor cell infiltration. For example,

*Address correspondence to: Changcheng Yang, yanggreatwall@hainmc.edu.cn; Yanda Lu, yandalu986@163.com
#These authors contributed equally to this work
Received: 07 July 2021; Accepted: 15 September 2021

Doi: 10.32604/biocell.2022.018221

SFRP4 positively correlates with infiltration of FOXP3$^+$ Treg cells in hepatocellular carcinoma, and downregulation of SFRP4 in tumor cells affects T cell recruitment (Yang *et al.*, 2019b). It has also been found that high expression of PRC1 was associated with low macrophage infiltration in HCC tissues, which in turn lead to poor prognosis (Zhou *et al.*, 2020a). Studies have reported genes closely related to breast cancer such as BRCA1, BRCA2, ATM, CHEK2, etc. (Dorling *et al.*, 2021). These genes can be used not only for cancer diagnosis, but also seriously affect the development, metastasis, and treatment outcome of breast cancer. Studies have confirmed that BRCA gene mutations may not directly drive immune infiltration, but the number of infiltrating lymphocytes and monocytes is significantly higher in breast cancer tissues with decreased BRCA expression (Kubli *et al.*, 2019; Solinas *et al.*, 2019). However, understanding of the key genes and their associated immune cell infiltration significantly associated with breast cancer prognosis is still very limited.

In this study, we first obtain differentially expressed gene signatures from the Gene Expression Omnibus (GSE) datasets and The Cancer Genome Atlas (TCGA) to evaluate genes associated with breast cancer metastasis. Then, functional analyses such as diagnosis and prediction of prognosis of disease were performed for important genes. In addition, we evaluate the relationship between relevant genes and infiltrating immune cells to provide an immunological basis for the tumor microenvironment and to reveal potential immunotherapeutic approaches.

Materials and Methods

Breast cancer datasets acquisition and processing

The workflow of analysis in our study is illustrated in Fig. S1. Gene expression datasets and corresponding clinical information of breast cancer patients were collected from the dataset of GEO (Barrett *et al.*, 2013) and the TCGA dataset included 1124 patients, and those microarray data without clinical information were discarded. In our study, 2 GEO datasets contained GSE62327 (Triulzi *et al.*, 2015) (7 cases in the pre-radiation group and 7 cases in the post-radiation group) and GSE43816 (Gruosso *et al.*, 2016) (14 cases in the disease group and 10 cases in the control group) were downloaded (https://www.ncbi.nlm.nih.gov/geo/) (Tables S1–S3). The primary data for the GSE62327 generated by Affymetrix Human Gene 1.1 ST Array [transcript (gene) version]). And the primary data for the GSE43816 generated by Illumina HumanHT-12 WG-DASL V4.0 R2 expression beadchip. The Cancer Genome Atlas (TCGA) data referenced in this study were available in a public repository from TCGA website (https://portal.gdc.cancer.gov/).

Differential gene expression analysis

The differential expression mRNAs were analyzed according to breast cancer and normal groups using the limma package (Ritchie *et al.*, 2015). Log fold change (logFC) > 1.0 and adjusted *P*-value < 0.05 were set as the threshold for differential genes. The results of differential analysis of mRNA will be presented using volcano plot (ggplot2) and heat map (pheatmap) (https://CRAN.R-project.org/package=pheatmap). Where the volcano plot is plotted with ggplot2 and the heatmap is plotted with the pheatmap package.

Enrichment analysis

The metascape (http://metascape.org) (Zhou *et al.*, 2019) database was used to analyze the gene ontology (GO) enrichment analysis of the list of differentially expressed mRNA target genes after intersection, mainly including biological processes (BP), cellular component (CC), molecular function (MF); the parameters were set $P < 0.01$, minimum count >3, enrichment factor >1.5. The GSEA method (Subramanian *et al.*, 2005) enriched differential mRNAs for expression-related pathways, and the groups were aligned 10,000 times for each analysis. The KEGG pathways dataset from curated gene sets was selected as the reference set. The threshold for statistically significant GSEA analysis was set at a corrected *P*-value < 0.05 and FDR < 0.25. The results of the enrichment analysis will be characterized by the corrected *P*-value as well as the NES. GSEA enrichment analysis and visualization were performed using GSEA native software.

PPI network construction

The STRING database (http://string-db.org, version 11) (Szklarczyk *et al.*, 2019) was used to construct the protein–protein interaction network. All the differential mRNA intersections after the associated differential molecules were put into the database for analysis, and the interaction threshold was set to 0.4. Cytoscape software (Shannon *et al.*, 2003) was used to visualize the molecular interaction network. a plug-in for Cytoscape software: CytoHubba (Chin *et al.*, 2014) was used to analyze the hub gene in the network.

Immune infiltration analysis

We extracted marker gene of 24 immune cells and analyzed the infiltration of 24 immune cells within the tumor using the EPIC method (Racle *et al.*, 2017), and analyzed the degree of correlation of the HUB gene and the expression matrix with these 24 cells using the Spearman correlation method.

Statistical analysis

Most of the statistical analyses in this study were done by the bioinformatics tools mentioned above. Other statistical analyses such as ROC curve analysis, survival curve analysis, etc. were performed by the R language package with version 3.4.1. The different expression levels of the genes were analyzed by two-tailed Student's *t*-test and the GO terms and KEGG signaling pathway were done by Fisher's test. Spearman's r correlation analysis was used to analyze the correlation of non-normally distributed data. *P*-values less than 0.05 were considered statistically significant.

Results

Screening of differentially gene expression changes in breast cancer

According to appropriate criteria in GEO database, two genome-wide gene expression datasets (GSE43816 and GSE62327) on the chemotherapy of breast cancer were

selected and analyzed using the R project limma package with a threshold of adjust *P*-value < 0.01 and $|\log_2 FC| > 3.0$. In GSE43816 datasets, a total of 5190 mRNAs were screened out, of which 3166 up-regulated mRNA and 2024 down-regulated mRNA were identified (Figs. 1A and 1B). In GSE62327 datasets, there were 2938 up-regulated mRNA and 703 down-regulated mRNA in the comparison of chemotherapy *versus* non-chemotherapy of breast cancer (Figs. 1C and 1D). To clarify the genetic changes between metastatic and non-metastatic breast cancer, we compared TCGA data to make new discoveries. As shown in the volcano plot, the data from TCGA were analyzed and 7566 differentially expressed mRNAs (including 3759 were upregulated and 3807 were downregulated) were identified (Figs. 1E and 1F). When we intersected the GEO chemotherapy dataset and the dataset of breast cancer survival status, there were 81 differentially expressed genes that were upregulated (Fig. 2A). When we take the dataset

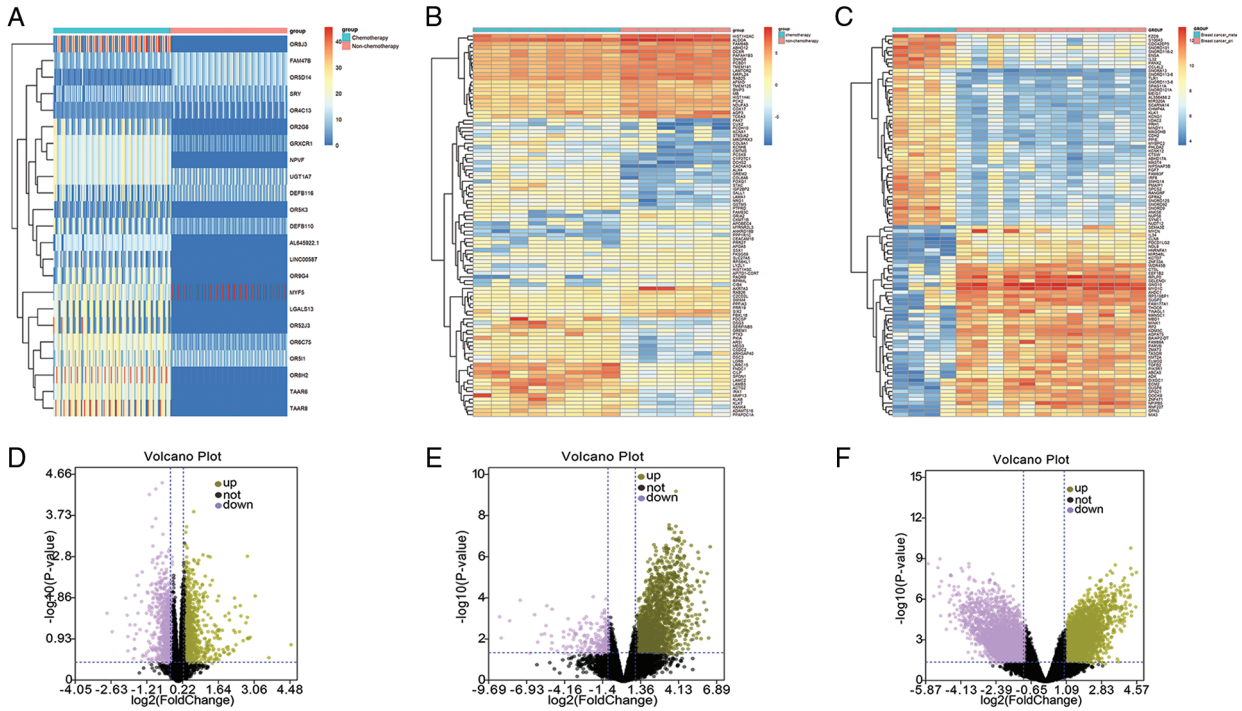

FIGURE 1. Heatmap and volcano plot of differentially expressed mRNA of breast cancer in TCGA and GEO datasets. Heatmap and volcano plot of differentially expressed mRNA in breast cancer chemotherapy *vs*. non-chemotherapy data in GSE43816 dataset (A–B). Heatmap and volcano plot of differentially expressed genes in breast cancer chemotherapy *vs*. non-chemotherapy microarray data in GSE62327 dataset (C–D). Heatmap and volcano plot of differential mRNA expression in TCGA breast cancer metastasis or not (E–F). The x-axis represents the logarithm in base 2 and the y-axis represents the negative logarithm in base 10 of the false discovery rate (FDR). Each dot in the heatmap represents a gene. Brown dots represent down-regulated mRNAs, red dots represent up-regulated mRNAs, and black dots represent genes with no differences.

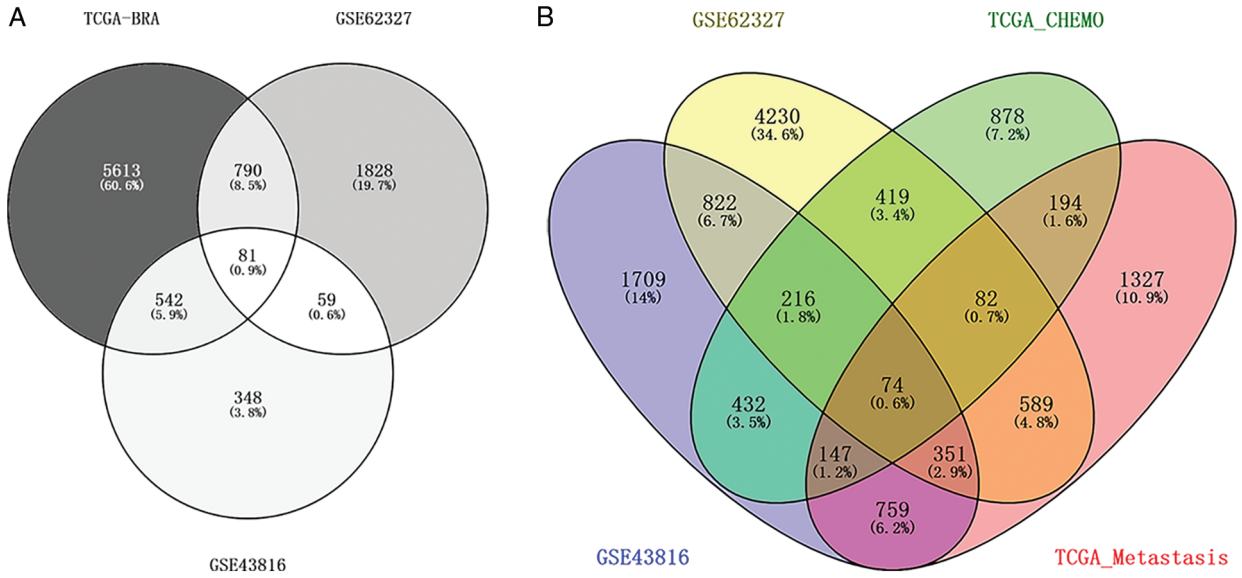

FIGURE 2. Intersection of differential genes with metastasis-associated genes. Intersection of differentially upregulated genes in three datasets (GSE43816, GSE62327 and TCGA) (A). Intersection of co-expressed genes with metastasis-associated genes (B).

of TCGA chemotherapy and metastasis and the GSE dataset to intersect, there were totally 74 upregulated mRNAs that were commonly found in four datasets (Fig. 2B).

Functional of pathway enrichment analysis of differential genes

In order to explore the biological functions of different genes, we analyzed and visualized the enrichment of 79 differential genes in the GO pathway (Figs. 3A–3C, Table S4); the enrichment results were mainly presented in the three directions of BP, MF, and CC of the GO pathway; the enrichment results revealed that the differential genes were mainly involved in biological processes such as reproductive structure development, epithelial to mesenchymal transition, cellular aldehyde metabolic process and extracellular matrix disassembly. We also found that it is closely related to the hypodifferentiation process: epithelial cell differentiation, which occurs mainly in the cellular matrix: extracellular matrix and is closely related to molecular functions such as ion binding. In summary, differential genes are mainly involved in the metastatic process and can promote epithelial to mesenchymal transition of cells. The results of the

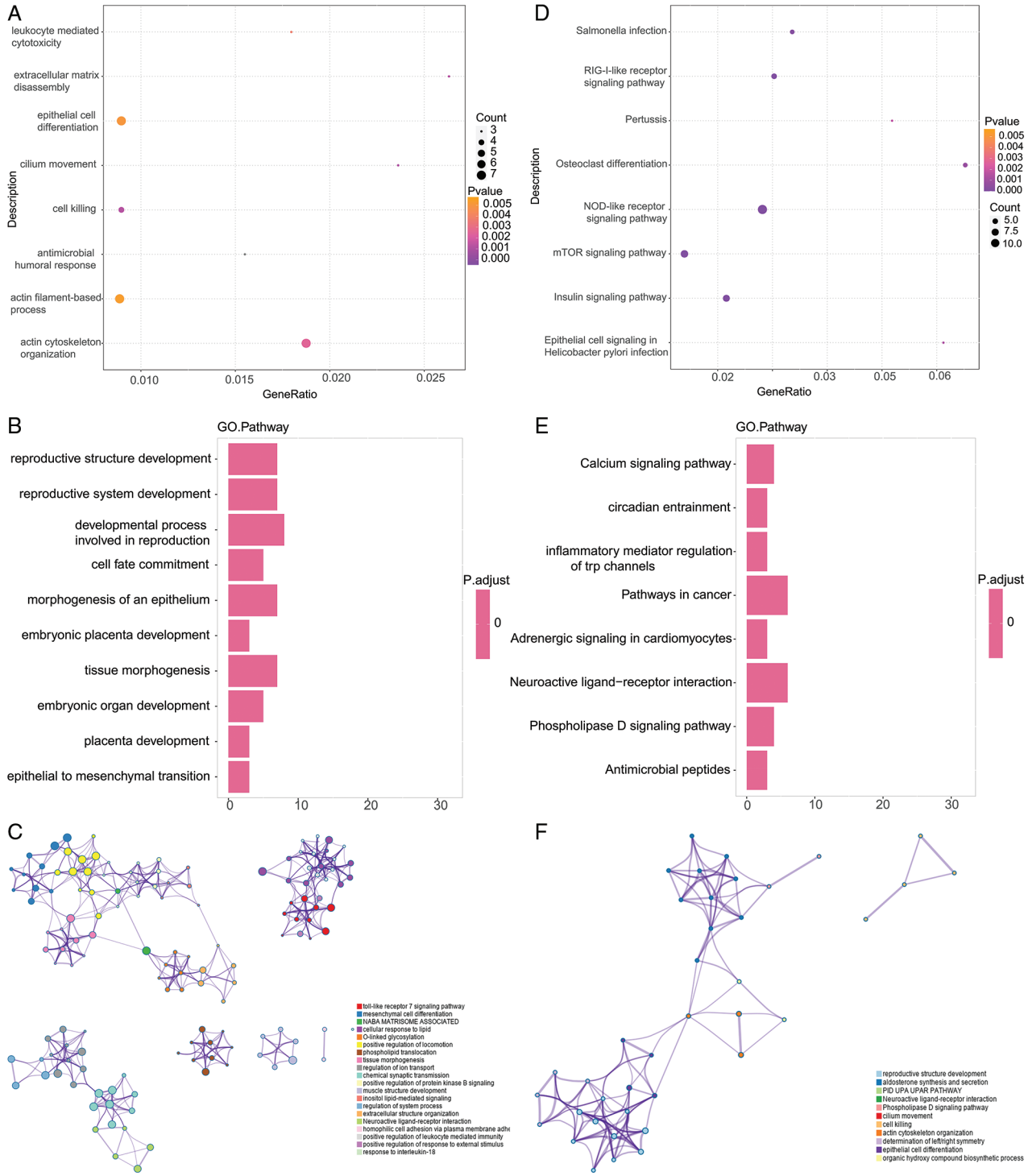

FIGURE 3. KEGG/GO pathway enrichment analysis of differential genes. GO pathway enrichment analysis of biological process, cellular component, and molecular function of differential genes (A–C). KEGG signaling pathway enrichment analysis of differential genes and the interrelationship between signaling pathways (D–F).

KEGG pathway enrichment analysis of differential genes show that differential genes are mainly associated with calcium signaling pathway, circadian entrainment, inflammatory mediator regulation of trp channels, pathways in cancer, neuroactive ligand-receptor interaction, and other biological processes (Figs. 3D–3F, Table S5).

Enrichment analysis of grouped GSEA based on tumor metastasis Based on the metastasis or not and grouping of TCGA samples, we obtained positive correlation pathway enrichment (Figs. 4A–4C, Table S6) and found that the grouping of high metastasis was more likely to be involved in biological processes of cell metastasis-related and chemotherapy-tolerant signaling pathways such as cell development, cell differentiation, regulation of cilium movement, cilium movement, response to iron ion, response to X-ray, myelin assembly. Similarly, we also obtained the negative pathway enrichment pathway (Figs. 4D–4F, Table S6) and found that the low metastasis subgroup was more likely to participate in natural killer cell mediated immunity, leukotriene biosynthetic process, positive regulation of BMP signaling pathway, response to oxygen radical leukotriene metabolic process, monocyte differentiation and other biological processes.

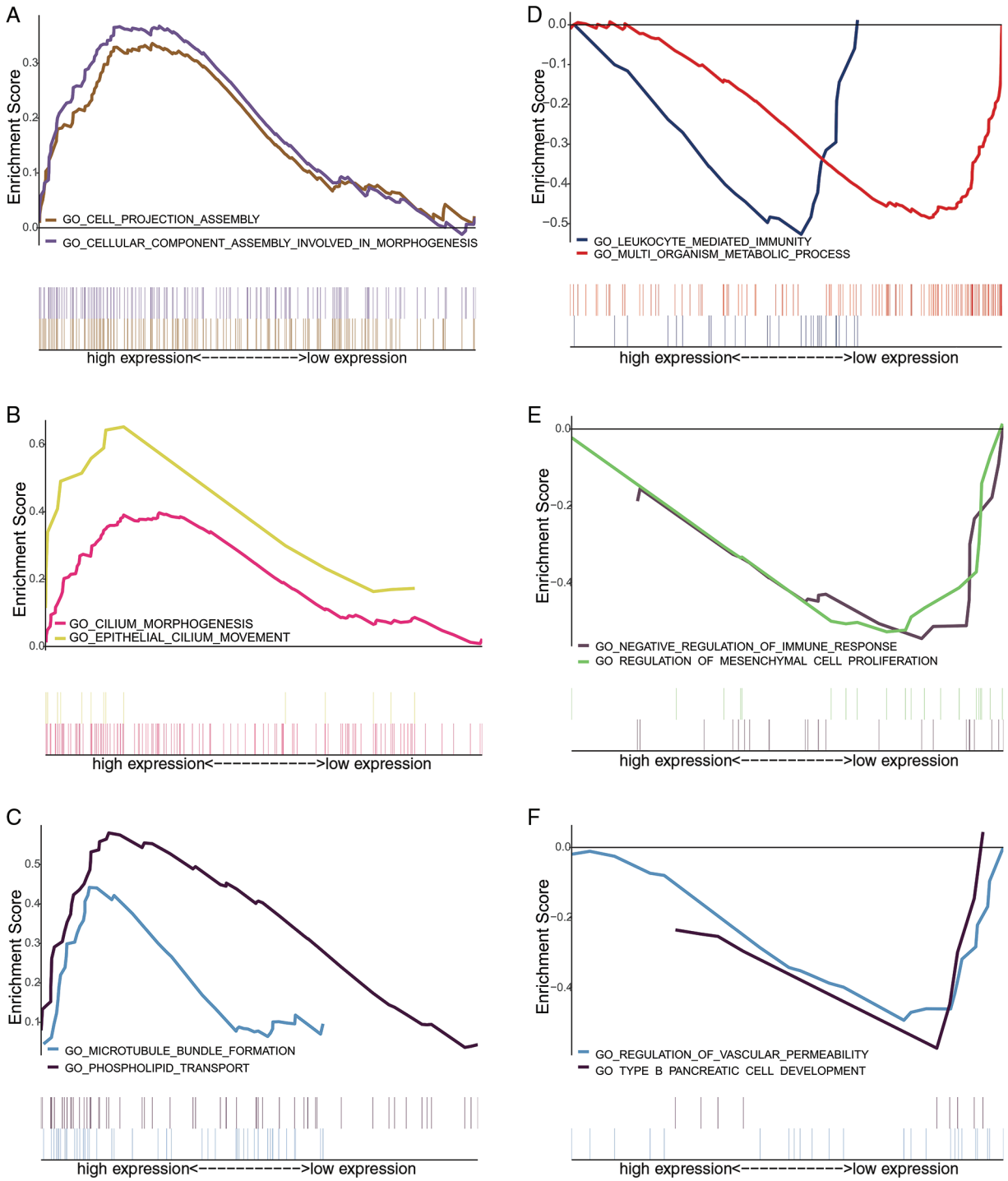

FIGURE 4. Grouped GSEA enrichment analysis based on metastasis expression. Positive correlation pathway enrichment (A–C) and negative correlation pathway enrichment pathway (D–F) were obtained based on metastasis or not and grouping of TCGA samples.

PPI network interactions and top HUB genes

In order to understand the interaction between different genes in the three database intersections, we obtained the interaction network map of tumor metastasis-related genes by PPI networks (Fig. 5A). We then performed the identification of the core molecules of the network graph (Fig. 5B) and finally obtained the top 25 HUB genes (Figs. 5B–5D, Table S7). These 25 HUB genes are IL-18, CD1C, PTGER3, TP53, TSC2, STK11, NPY, BCL1, TLR7, CD80, PRTOR, C3AR1, MAPK, FN1, LY86, CCL2, BAX, BID, TNFSF10, FADD, CLEC5A, PARP1, ATP8B4, TMEM30A, EIF4EBP1. Then we selected these hub different genes for following analysis.

Analysis of expression levels based on key genes and survival analysis in TCGA breast cancers

To further confirm the expression of hub genes in breast cancer, we examined the expression of 25 hub genes in TCGA database and subjected these genes to correlation analysis of TCGA expression in breast cancer, we found that in the screened differential genes TLR7, LY86, CLEC5A, CXCR4, PARP1, FN1 were expressed at lower levels in breast cancer cancer than in the paracancer ($P < 0.05$, Figs. 6I, 6K, 6O, 6Q, 6R, and 6S). On the other hand, PTGER3, ATP8B4 was expressed at higher levels in cancer than in the paracancer ($P < 0.05$, Figs. 6N and 6X), and then we subjected the related genes to survival analysis to determine the prognosis. Overall survival (OS) was performed to obtain risk coefficients and determine the prognosis for the relevant target genes. During the 10 years survival period, we found that lower expression of FN1, EIF4EBP1, TP53 had lower risk coefficients and were positively associated with good prognosis ($P < 0.05$, Figs. 7B, 7C and 7Q). However, high expression of CD1C, PTGER3, IL18 gene had lower risk factors and longer overall survival ($P < 0.05$, Figs. 7A, 7I, and 7L).

Genes with differential expression levels after ROC diagnostic prediction analysis

In order to further clarify the clinical diagnostic value of genes with significant differential expression, we subjected these genes to ROC predictive diagnostic analysis and found that FN1, LY86, IL-18, CD1C, C3AR1, TLR7, and TMEM30A genes were genetic indicators with significant predictive diagnoses, and their diagnostic ROC curves had areas under

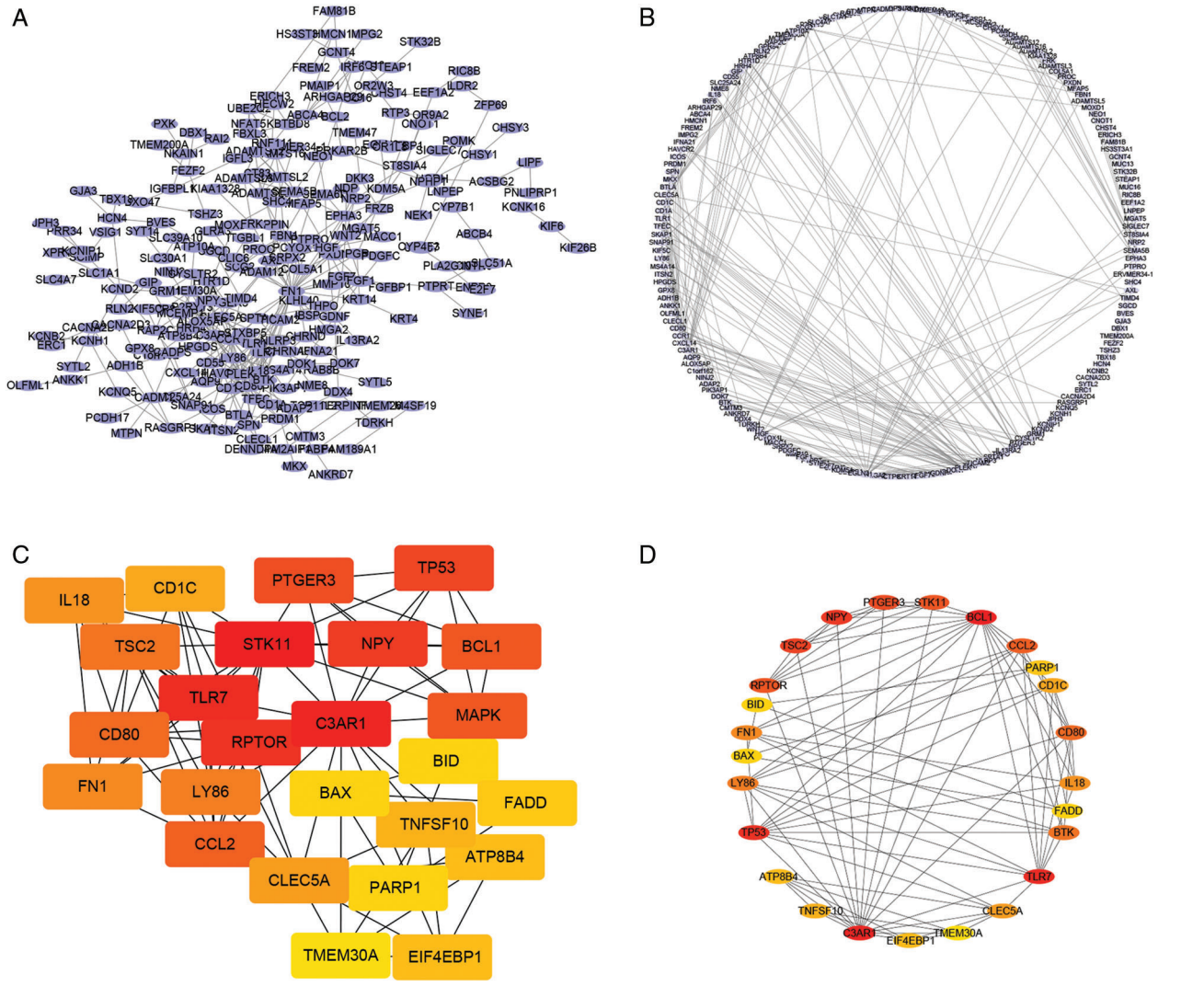

FIGURE 5. PPI network interactions of key invasion-associated genes and top HUB genes. The interplay network map was obtained by subjecting the co-expression metastasis-related genes after intersection to PPI network interplay analysis (A). The identification of the core molecules of the network map was performed and finally the top 25 HUB genes were obtained (B–D).

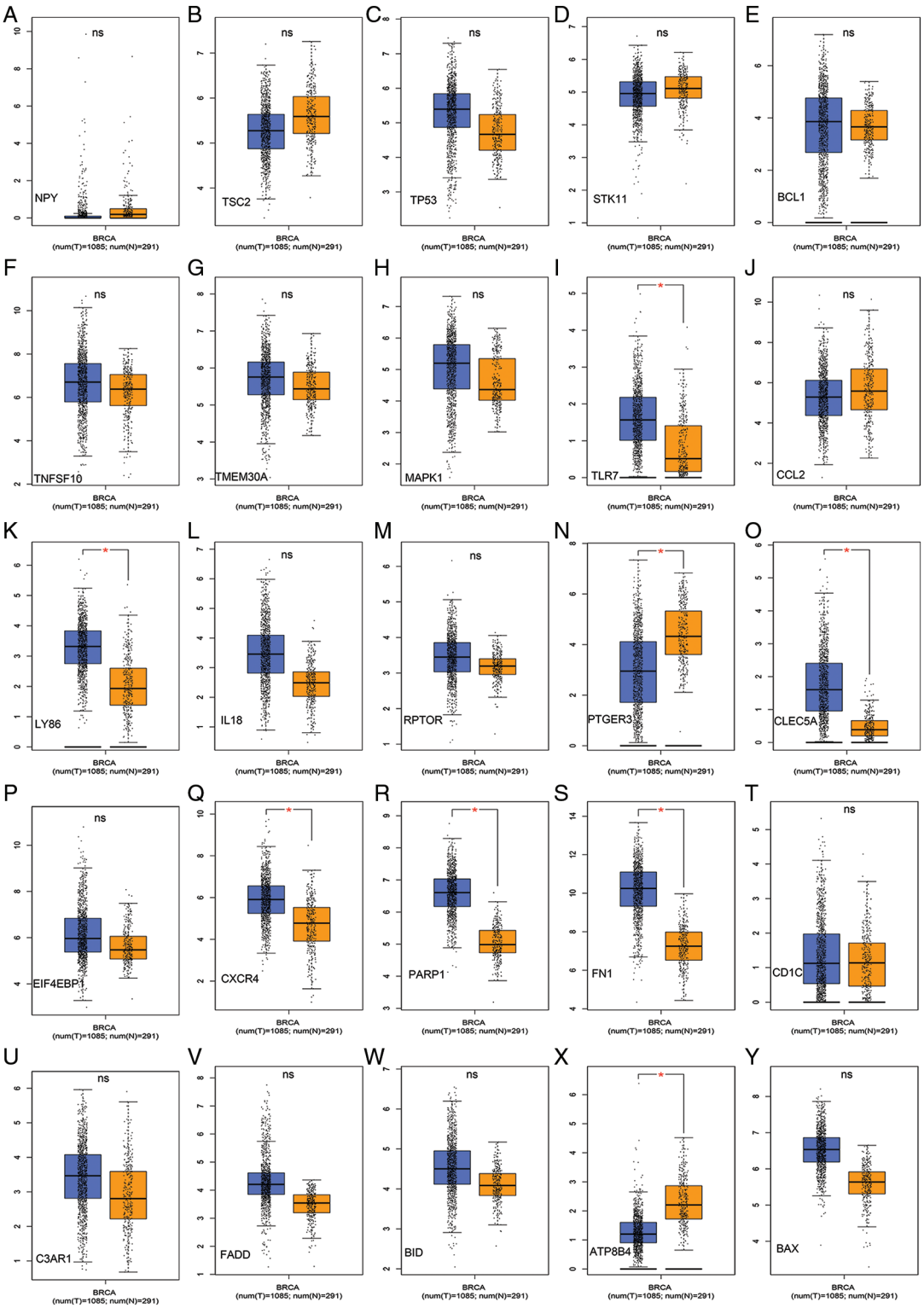

FIGURE 6. Analysis of expression levels based on key genes in TCGA breast cancers. About 25 genes obtained were analyzed for expression levels in TCGA, and the expression levels of TLR7, LY86, CLEC5A, CXCR4, PARP1, FN1, PTGER3, ATP8B4 which had significant differences between cancer and paracancer, were identified (A–Y).

the curves (AUC) of 0.968, 0.966, 0.897, 0.874, 0.846, 0.812, and 0.781, respectively (Figs. 8A and 8B).

Screening of key transcription factors for key genes and analysis of interplay networks and functions

In order to further screen the target transcription factors of key genes associated with metastasis, in this study, we took two key genes, TP53 and HIF1A, as examples, and obtained the transcription factors and their interplay networks that had binding to multiple differential genes through network interactions (Fig. 9A). Based on the network interactions between TP53 and HIF1A and the functional analysis of the pathway (Figs. 9B and 9C), we found that TP53 and HIF1A are mainly involved in cell-cell adhesion, focal adhesion, cell

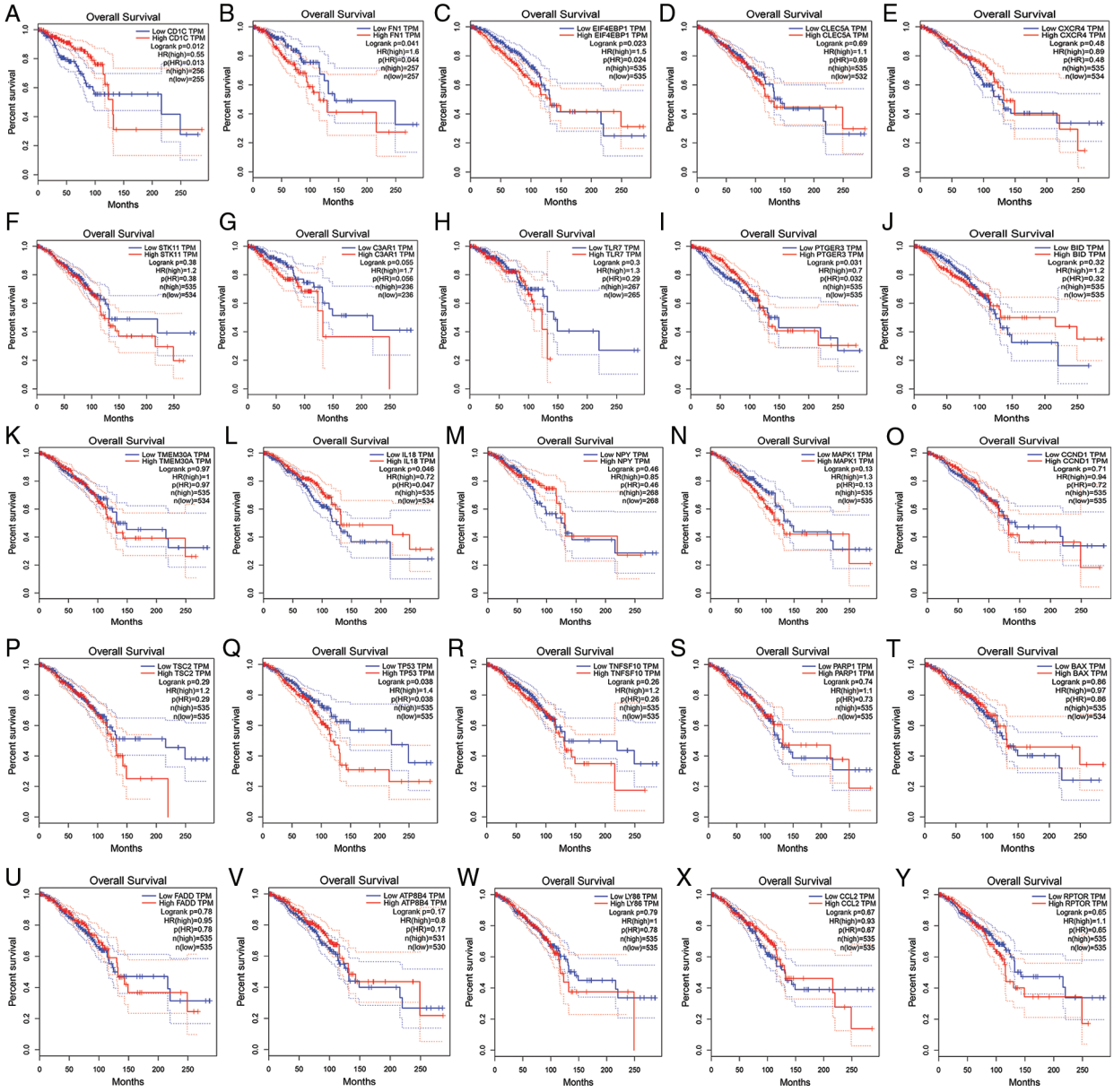

FIGURE 7. Screening of meaningful key genes in TCGA based on OS. Target gene identification by OS prognostic analysis was performed to obtain OS levels for risk factors and relevant prognosis. In the 10 years survival period, breast cancers with high expression of CD1C, PTGER3, IL18 gene were significantly better than low expression ($P < 0.05$) (A, I, L). However, Low expression of FN1, EIF4EBP1, TP53 in breast cancer showed higher OS in the 10 years survival period ($P < 0.05$) (B, C, Q). Although high expression of other hub genes possessed higher OS, it was not statistically significant ($P > 0.05$) (D–H, J, K, R–Y).

adhesion molecules cams and cell cycle. In addition, the GSEA pathway enrichment analysis of the target gene TP53 and HIF1A interaction network revealed that these two classical genes were mainly enriched to cell cycle, cell adhesion molecules after low expression (Figs. 9D and 9E).

Risk prediction of key metastasis-associated genes

To further identify risk factors associated with disease progression, we performed univariate and multivariate COX regression analysis and created forest plots for prediction and analysis based on metastasis-related genes and clinicopathological data, and found that Event (non-metastatic breast cancer had a low risk factor = 0.096, $P < 0.001$) (Fig. 10A), CD1C expression for predicting breast cancer recurrence and metastasis and OS (risk factor = 1.08719, $P = 0.01392$) (Fig. 10A), BCL1 expression for breast cancer recurrence and metastasis and OS (risk factor 1.129, $P = 0.018$) (Fig. 10A). However, although the risk coefficients of genes such as TLR7, CCR1, HRH4, HTR1D, P2RY13, FN1 were high, these genes were not statistically significant for predicting breast cancer ($P > 0.05$, Figs. 10B–10E).

Analysis of immune infiltration assessment of key genes

Finally, we performed immune infiltration assessment analysis of the screened related genes in the breast cancer TCGA database to assess their correlation analysis with $CD8^+$ T cells. In the immune infiltration assessment analysis of this study, we found that ATP8B4 showed a negative correlation with immune $CD8^+$ T-cell infiltration with an R-value of −0.134 ($P < 0.01$, Fig. 11A). In addition, FN1 also showed a significant negative correlation

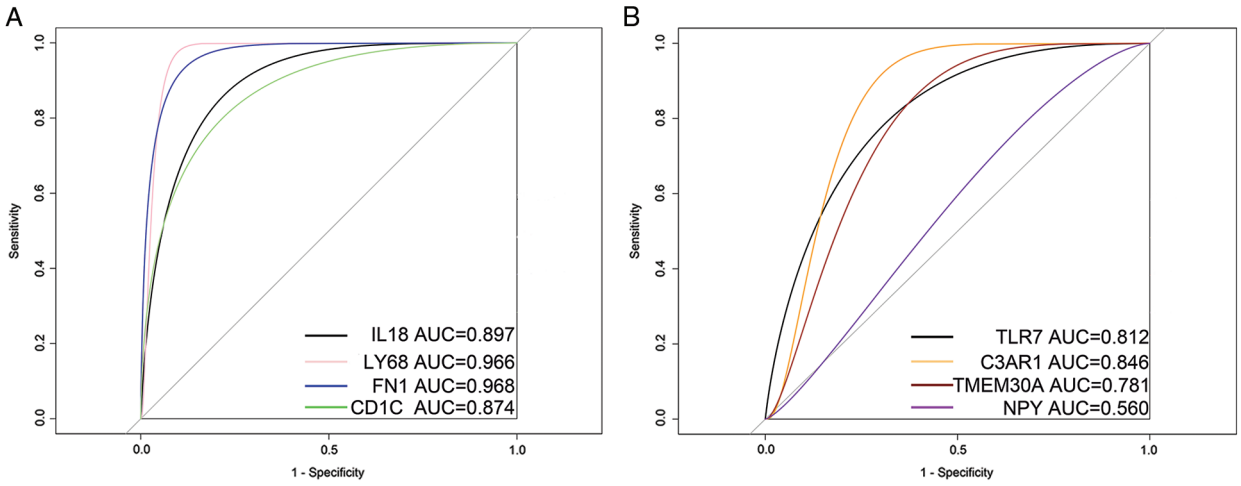

FIGURE 8. Predictive analysis of ROC diagnosis based on genes with differential expression. Genes with significant differential expression were subjected to ROC predictive diagnostic analysis, and CD1C, TLR7, FN1, IL18, C3AR1, LY86 were found to be genetic indicators with significant predictive diagnosis (A–B).

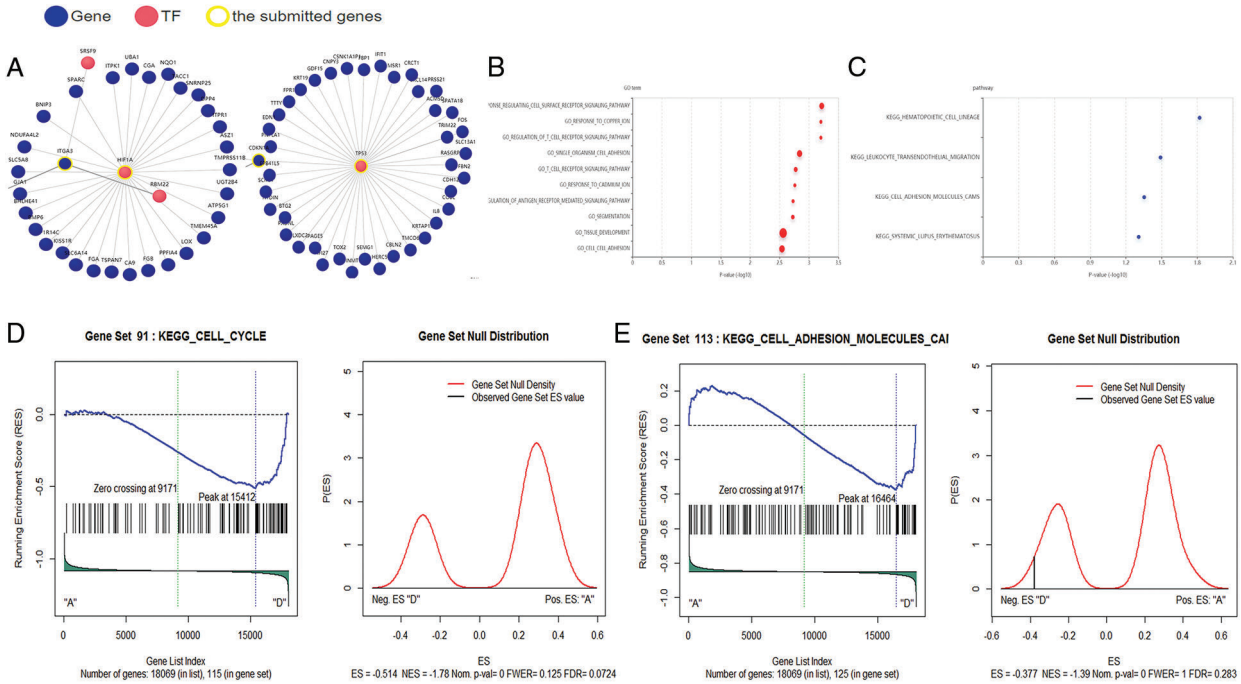

FIGURE 9. Screening of targeting key transcription factors based on key gene prediction and their interaction network and functional analysis. Key transcription factor prediction based on HIF1A and TP53 yielded multiple transcription factors and their interplay networks (A). GO signaling pathway and functional analysis of HIF1A (B–C). GSEA pathway enrichment analysis of the target gene interaction network of TP53 (D–E).

with the proportion of immune $CD8^+$ T-cell infiltration, with an R-value = −0.368 ($P < 0.01$, Fig. 11E). However, as shown in the figure, TLR7, LY86, IL18, and CD1C showed positive correlation with the infiltration proportion of $CD8^+$ T cells with R-values of 0.347, 0.257, 0.451, and 0.253, respectively, and their *P*-values were all less than 0.01 (Figs. 11B–11D and 11F).

Evaluation of immune cell infiltration from TCGA, GEO sample data

To further clarify the immune status in breast cancer tissues, samples from TCGA and GEO were evaluated for immune cell infiltration and mesenchymal cell infiltration, resulting in the infiltration of immune cells $CD4^+$, $CD8^+$ T cells and other mesenchymal cells in the relevant datasets, as well as the proportion of each immune cell. As shown by the immune infiltration levels of metastatic and non-metastatic samples of breast cancer in the TCGA dataset, cancer-associated fibroblasts, vascular endothelial cells, and $CD8^+$ T cells are strongly associated with breast cancer metastasis (Figs. 12A–12C); In addition, the levels of $CD8^+$ T cells, B cells, and NK cells were higher in the group of chemotherapy without metastasis than in the group of chemotherapy with metastasis and the group of non-chemotherapy with metastasis, while the levels of $CD8^+$ T cells, NK cells in the levels in the group of non-chemotherapy without metastasis were higher than those in the group of non-chemotherapy with metastasis (Fig. S2A). The immune infiltration levels of the GSE43816 dataset suggest that vascular endothelial cells, $CD4^+$ T cells, $CD8^+$ T cells, cancer-associated

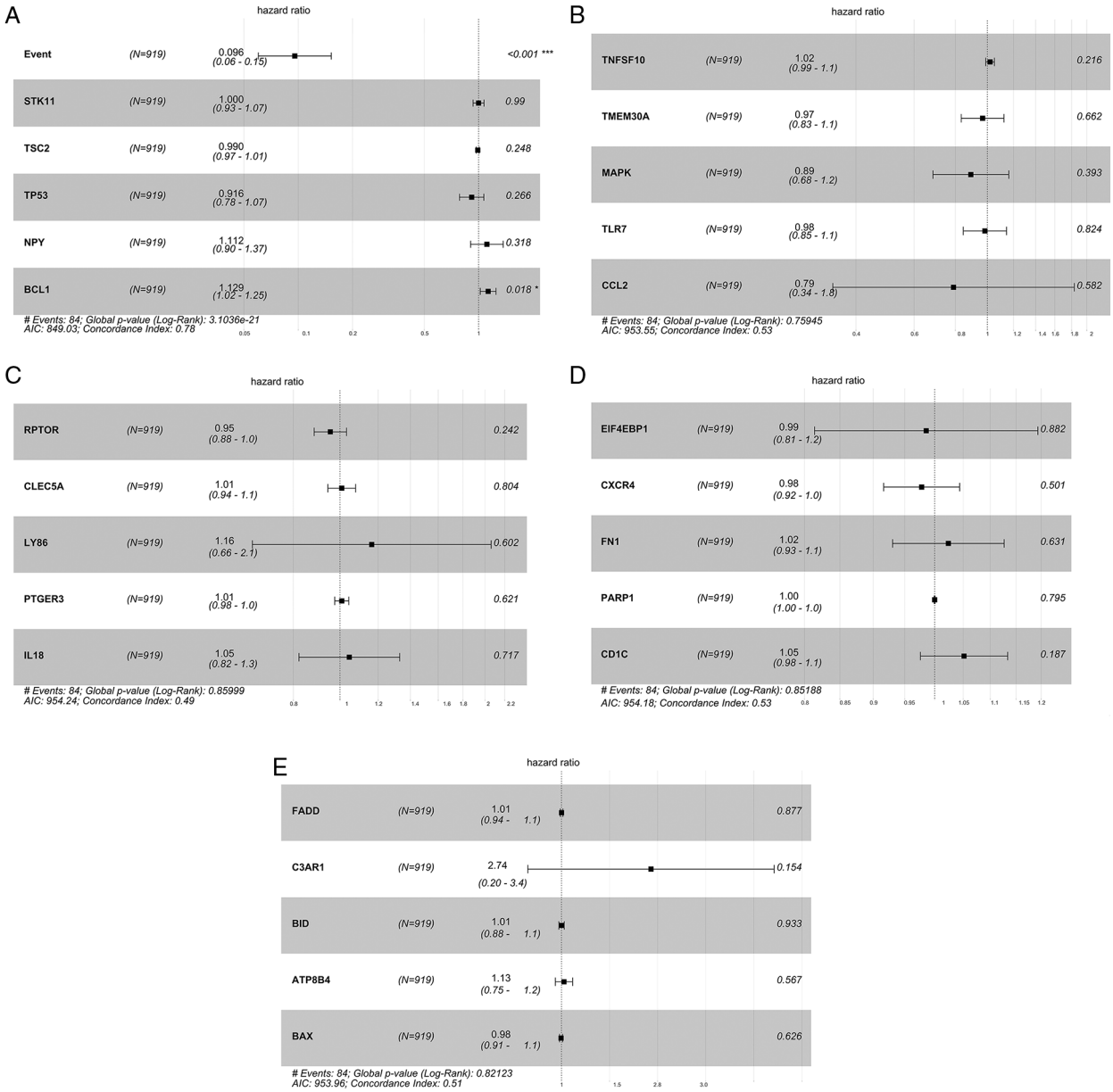

FIGURE 10. Risk prediction of key metastasis-associated genes and forest plotting, Univariate and multifactorial COX regression analysis based on metastasis-associated genes and clinicopathological data were performed to further identify factors and associated indicators with significant risk for disease progression by drawing forest plots (A–E).

fibroblasts are closely related to breast cancer progression (Figs. 12D–12F). Moreover, the expression levels of $CD8^+$ T cells, B cells, $CD4^+$ T cells, and NK cells were higher in the chemotherapy group than in the non-chemotherapy group (Fig. S2B). The GSE62327 dataset on the level of immune infiltration in breast cancer tissues suggested that cancer-associated fibroblasts, vascular endothelial cells, and $CD8^+$ T cells were strongly associated with breast cancer progression (Figs. 12G–12I). Furthermore, the expression levels of $CD8^+$ T cells, B cells, and NK cells were higher in the chemotherapy-sensitive group of breast cancer than in the chemoresistant group (Fig. S2C).

Discussion

With the rapid development of genetic testing technologies (e.g., microarray technology, sequencing technology, etc.) and the establishment of genetic databases, genes associated with cancer are increasingly analyzed and validated as new targets (Wang *et al.*, 2016). However, most of the current studies do not correlate oncogenes with immune cells infiltrated in TME, which may affect the immunotherapeutic response to cancer. In this study, we sought to explore the gene expression associated with chemotherapy and metastasis of breast cancer, to obtain important genes with prognostic and diagnostic value, and to analyze their associated signaling pathways to understand the progression of breast cancer patients. By comparing the transcriptional expression profiles of breast cancer patients and metastasis groups, a total of 81 genes involved in chemotherapy-related breast cancer gene expression and 79 genes associated with breast cancer metastasis were identified. These genes were mainly involved in reproductive structure development, epithelial to mesenchymal transition, cellular aldehyde metabolic process, extracellular

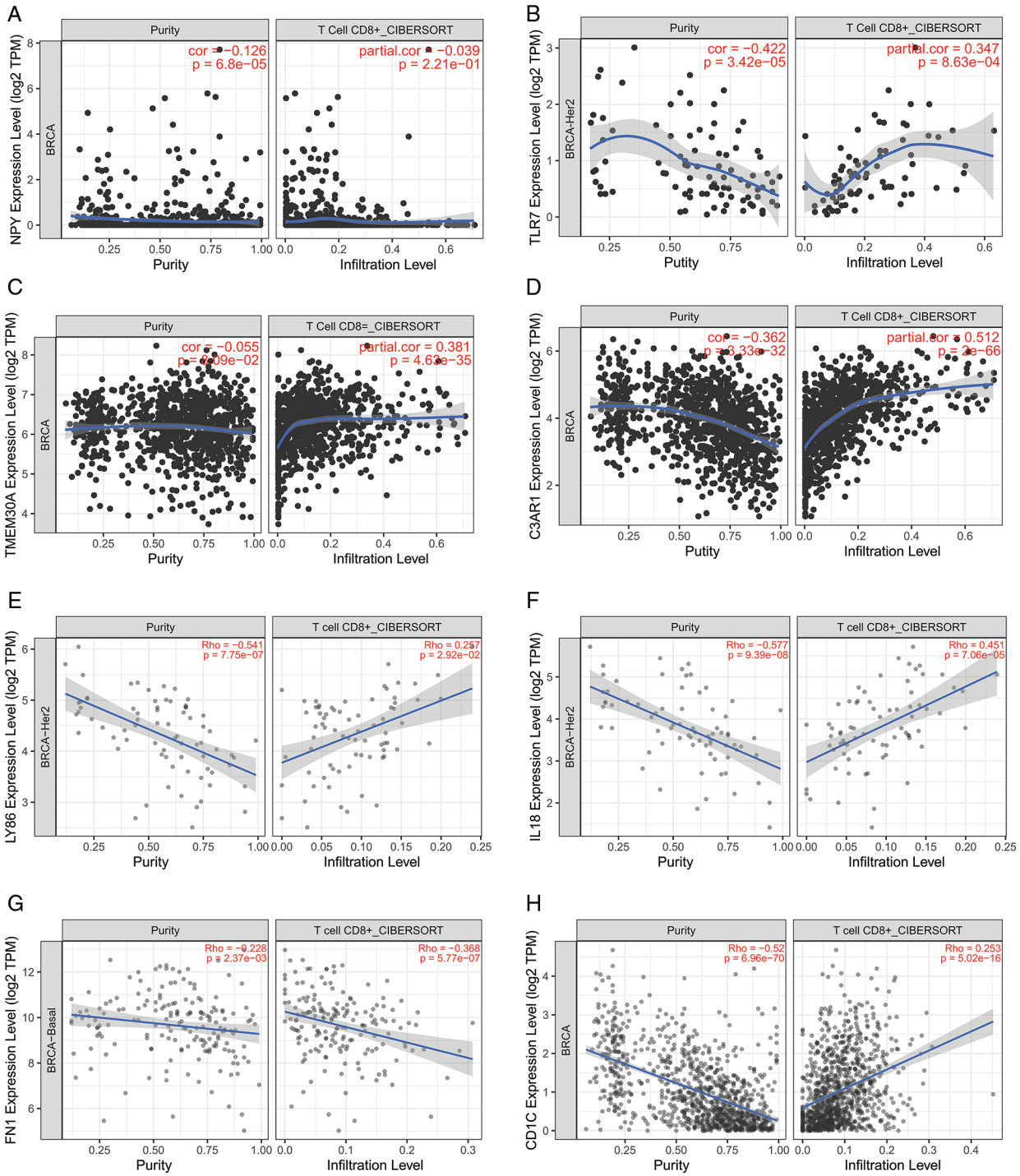

FIGURE 11. Immune infiltration assessment analysis based on key gene prediction. The screened target genes were analyzed in the breast cancer TCGA database for immune infiltration assessment, and the correlation analysis of the screened genes (ATP8B4, TLR7, LY86, IL18, FN1, CD1C) with $CD8^+$ T cells was evaluated (A–H).

matrix disassembly, epithelial cell differentiation and other biological processes. Moreover, the main genes associated with metastasis were CXCR4, ERBB2, EIF4EBP1, TSC2, MAPK1, CCL2, RPTOR, BCL2, STK11, TP53, PARP1, FADD, TNFSF10, BID, BAX and so on. However, FN1, CLEC5A, ATP8B4, TLR7, LY86 PTGER3 expression significantly predicted the overall survival of breast cancer patients. Subsequently, the expression of genes including CD1C, TLR7, FN1, IL18, C3AR1, LY86 were included in the ROC curve analysis of breast cancer diagnosis and the associated risk prediction analysis. Importantly, ATP8B4 and FN1 expression were significantly associated with $CD8^+$ T immune cell infiltration by Spearman's correlation analysis.

Chemotherapy is still the main strategy for the treatment of cancer. However, studies have confirmed that chemotherapy may promote cancer metastasis and tumor metastasis is an important factor contributing to poor tumor prognosis (Karagiannis *et al.*, 2017; Keklikoglou *et al.*, 2019). In this study, by analyzing GSE43816, GSE62327, and TCGA datasets, we identified some differential genes associated with breast cancer metastasis, and the enrichment results showed that these genes are closely related to cell stromal remodeling,

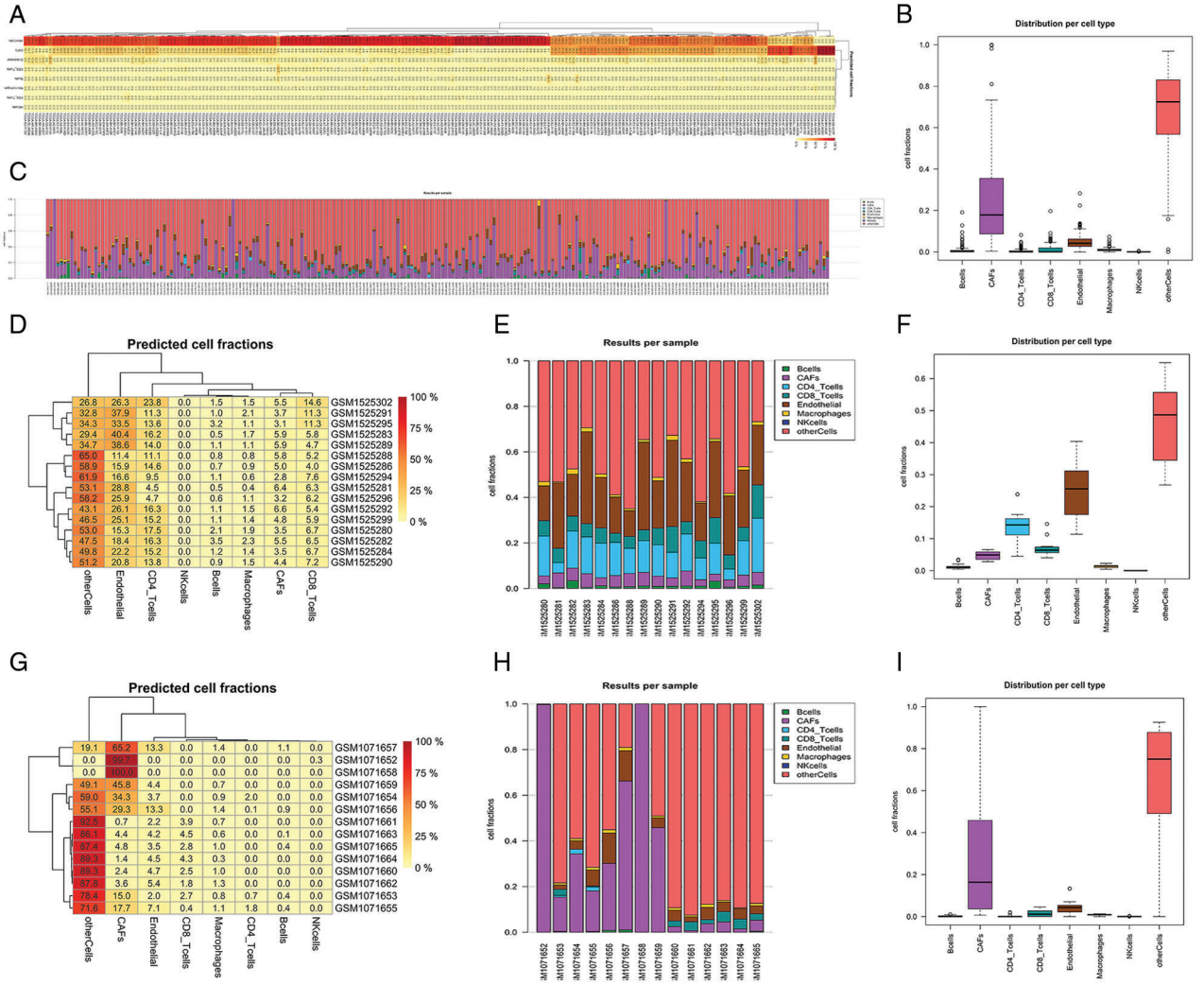

FIGURE 12. Evaluation of immune cell infiltration based on TCGA, GEO sample data. Samples in TCGA, GEO were evaluated for immune cell infiltration and mesenchymal cell infiltration to obtain the immune cell ratio for each sample in TCGA dataset. The level of immune infiltration in metastatic and non-metastatic samples of breast cancer in TCGA dataset (A–C), in GSE43816 dataset (D–F) and in GSE62327 dataset (G–I), respectively.

epithelial mesenchymal transition, intracellular metabolism, and hypofractionation process of cells. Several studies have confirmed that EMT not only confers the ability of tumor cells to invade and metastasize, but also is a major factor affecting the efficacy of chemotherapy (Li and Kang, 2016). Furthermore, epithelial mesenchymal transformation affect changes in internal cellular metabolism, involving cytoskeletal alterations and matrix remodeling, and further affects cell differentiation (Ijaz *et al.*, 2014; Tsubakihara and Moustakas, 2018). Currently, the main signaling pathways associated with EMT are TGF-β signaling pathway, Wnt signaling pathway and Notch signaling pathway (Kyuno *et al.*, 2021). In the present study, the main signaling pathways involved in the differential genes were Calcium signaling pathway, NOD-like receptor signaling pathway and mTOR signaling pathway, which were also found to affect EMT directly or indirectly in previous studies (Figiel *et al.*, 2019; Hu *et al.*, 2018; Luo *et al.*, 2021). Our study also found that high metastasis of breast cancer is closely related to biological processes such as cell development, cell differentiation, and regulation of ciliary motility, while in low metastasis natural killer cell-mediated immunity, leukotriene biosynthesis process, positive regulation of BMP signaling pathway, leukotriene metabolic process in response to oxygen free radicals, and monocyte differentiation. Thus, by searching for key genes associated with altered metastatic ability after chemotherapy resistance in breast cancer, these genes can be studied not only as targets for future new drugs, but also as relevant biomarkers to reflect the effect of cancer chemotherapy. It is important to note that additional prospective studies are needed as well to further validate these findings due to sample size limitations in the dataset.

Next, we performed overall survival analysis and ROC curve analysis on some of the screened HUB genes. Firstly, we found that the expression levels of FN1, CLEC5A, ATP8B4, TLR7, and LY86 genes were higher in breast cancer tissues than in paracancerous tissues, while the expression levels of PTGER3 genes were lower in breast cancer tissues than in paracancerous tissues. Consistent with previous studies, in addition to ATP8B4 and LY86, FN1, CLEC5A and TLR7 were found to be highly expressed in breast cancer cells or tissues by clinical trial or bioinformatics studies, but PTGER3 expression was reduced in breast cancer, and the expression levels of these genes affects the prognosis of breast cancer (Bao *et al.*, 2019;

Semmlinger *et al*., 2018; Shi *et al*., 2020; Zhang *et al*., 2021; Zhang *et al*., 2020). Moreover, these genes play an important role in the progression of several other cancers, for example, FN1 and CLEC5A have been reported to promote proliferation, invasion and metastasis of gastric cancer and glioblastoma (Fan *et al*., 2019; Han *et al*., 2020; Wang *et al*., 2020; Yang *et al*., 2019a). In line with other findings (Inoue *et al*., 2019; Milioli *et al*., 2017; Saatci *et al*., 2020), we also found a significant positive correlation between low expression of IL18 genes and good prognosis of breast cancer through survival curves, while high expression of FN1 and C3AR1 genes implied poor disease prognosis. Finally, we verified that some genes (e.g., FN1, LY86, IL-18, CD1C, C3AR1, TLR7, and TMEM30A) have good diagnostic efficacy in predicting the development of breast cancer by ROC curves, and all of these genes may be used as potential biomarkers.

To further confirm the relationship between key genes and breast cancer metastasis, in the present study TP53 and HIF1A genes were used as examples, and we found that TP53 and HIF1A are mainly involved in the biology of cell-cell adhesion, cell adhesion molecules cams and cell cycle processes. Many studies have confirmed that TP53 and HIF1A genes are closely related to the metastasis of tumors (Li *et al*., 2021a; Tang *et al*., 2021; Tiwari *et al*., 2020). In our study, CD1C and ATP8B4 were closely associated with recurrence, metastasis, and prognosis of breast cancer. Curiously, we found that the ATP8B4 gene was negatively correlated with $CD8^+$ T lymphocyte infiltration in breast cancer tissues, whereas CD1C was positively correlated with $CD8^+$ T lymphocyte infiltration in cancer tissues. In addition, we also found that genes such as TLR7, LY86, and IL18 were positively correlated with the proportion of $CD8^+$ T lymphocytes infiltrating in cancerous tissues. Recent studies have shown that IL-18 and its receptor can activate specific high expression of $CD8^+$ T cells, suggesting a correlation between IL-18 not only with $CD8^+$ T cells, but also its pathway to tumor-killing effector cell function (Zhou *et al*., 2020b). And the results by all three databases, GSE43816, GSE62327, and TCGA, showed a close relationship between $CD8^+$ T cells and the progression of breast cancer. However, whether CD1C and ATP8B4 are abundantly expressed in many malignancies, are associated with migration of regulatory T cells, and their role in tumor-associated immunosuppression still need further experimental study.

Conclusions

In our study, we extracted genes of prognostic value from the GEO and TCGA databases based on functional enrichment analysis of breast cancer metastasis. These genes can be used to predict the occurrence and prognosis of breast cancer patients and show their potential to be biomarkers for breast cancer diagnosis as well as prognosis. In addition, further study of the relevant genes will help to understand the immune status of the tumor microenvironment and explore potential therapeutic targets. However, there are still some limitations of this study. First, in order to provide a comprehensive picture of the factors and effects affecting the microenvironmental phenotype of breast cancer, additional clinical features of breast cancer patients should be included in subgroup analysis. Second, the small sample size in the dataset may lead to biased results or high false positive rates, and the sample size will be appropriately increased in future studies on the immune microenvironment of breast cancer. Third, the data obtained such as altered genes and associated enriched signaling pathways were retrieved from public databases, and experiments are needed to validate these findings. Fourth, up- and down-regulation of genes may not be the real cause of the altered tumor microenvironment (e.g., lymphocyte infiltration) and breast cancer metastasis, so further mechanistic studies of these genes are needed.

Acknowledgement: We thank the TCGA and GEO databases for providing the platform and contributors for uploading their meaningful datasets.

Availability of Data and Materials: The datasets analyzed in the current study are available in the GEO (https://www.ncbi.nlm.nih.gov/geo/) and TCGA database (https://portal.gdc.cancer. gov/).

Ethics Approval: TCGA and GEO both are public databases. The patients involved in the database have obtained ethical approval in the original studies. Users can download relevant data for research and publish relevant articles. Our study is based on these open source data, so there are no ethical issues.

Authors' Contribution: The authors confirm contribution to the paper as follows: study conception and design: Changcheng Yang and Yanda Lu; data collection: Lu Yang, Fen Huang, Yang Wen and Boke Zhang; analysis and interpretation of results: Lu Yang, Yun Liu, Boke Zhang, Jiangzheng Zeng and Mengsi Yu; draft manuscript preparation: Lu Yang, Boke Zhang and Changcheng Yang. All authors reviewed the results and approved the final version of the manuscript.

Funding Statement: This work was supported by Hainan Provincial Natural Science Foundation of China (No. 820RC765).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

Bao Y, Wang L, Shi L, Yun F, Liu X et al. (2019). Transcriptome profiling revealed multiple genes and ECM-receptor interaction pathways that may be associated with breast cancer. *Cellular & Molecular Biology Letters* **24**: 38. DOI 10.1186/s11658-019-0162-0.

Barrett T, Wilhite S, Ledoux P, Evangelista C, Kim I et al. (2013). NCBI GEO: Archive for functional genomics data sets–update. *Nucleic Acids Research* **41**: D991–D995. DOI 10.1093/nar/gks1193.

Chin C, Chen S, Wu H, Ho C, Ko M, Lin C (2014). cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology* **8**: S11. DOI 10.1186/1752-0509-8-S4-S11.

Dorling L, Carvalho S, Allen J, González-Neira A, Luccarini C et al. (2021). Breast cancer risk genes—Association analysis in

more than 113,000 women. *The New England Journal of Medicine* **384**: 428–439. DOI 10.1056/NEJMoa1913948.

Fan H, Ni Q, Fan Y, Ma Z, Li Y (2019). C-type lectin domain family 5, member A (CLEC5A, MDL-1) promotes brain glioblastoma tumorigenesis by regulating PI3K/Akt signalling. *Cell Proliferation* **52**: e12584. DOI 10.1111/cpr.12584.

Figiel S, Bery F, Chantôme A, Fontaine D, Pasqualin C et al. (2019). A novel calcium-mediated EMT pathway controlled by lipids: An opportunity for prostate cancer adjuvant therapy. *Cancers* **11**: 1814. DOI 10.3390/cancers11111814.

Gruosso T, Mieulet V, Cardon M, Bourachot B, Kieffer Y et al. (2016). Chronic oxidative stress promotes H2AX protein degradation and enhances chemosensitivity in breast cancer patients. *EMBO Molecular Medicine* **8**: 527–549. DOI 10.15252/emmm.201505891.

Han C, Jin L, Ma X, Hao Q, Lin H, Zhang Z (2020). Identification of the hub genes RUNX2 and FN1 in gastric cancer. *Open Medicine* **15**: 403–412. DOI 10.1515/med-2020-0405.

Hu B, Ding G, Fu P, Zhu X, Ji Y et al. (2018). NOD-like receptor X1 functions as a tumor suppressor by inhibiting epithelial-mesenchymal transition and inducing aging in hepatocellular carcinoma cells. *Journal of Hematology & Oncology* **11**: 28. DOI 10.1186/s13045-018-0573-9.

Ijaz T, Pazdrak K, Kalita M, Konig R, Choudhary S et al. (2014). Systems biology approaches to understanding Epithelial Mesenchymal Transition (EMT) in mucosal remodeling and signaling in asthma. *The World Allergy Organization Journal* **7**: 13. DOI 10.1186/1939-4551-7-13.

Inoue N, Li W, Fujimoto Y, Matsushita Y, Katagiri T et al. (2019). High serum levels of interleukin-18 are associated with worse outcomes in patients with breast cancer. *Anticancer Research* **39**: 5009–5018. DOI 10.21873/anticanres.13691.

Karagiannis G, Pastoriza J, Wang Y, Harney A, Entenberg D et al. (2017). Neoadjuvant chemotherapy induces breast cancer metastasis through a TMEM-mediated mechanism. *Science Translational Medicine* **9**: 131672. DOI 10.1126/scitranslmed.aan0026.

Keklikoglou I, Cianciaruso C, Güç E, Squadrito M, Spring L et al. (2019). Chemotherapy elicits pro-metastatic extracellular vesicles in breast cancer models. *Nature Cell Biology* **21**: 190–202. DOI 10.1038/s41556-018-0256-3.

Kubli S, Bassi C, Roux C, Wakeham A, Göbl C et al. (2019). AhR controls redox homeostasis and shapes the tumor microenvironment in BRCA1-associated breast cancer. *Proceedings of the National Academy of Sciences of the United States of America* **116**: 3604–3613. DOI 10.1073/pnas.1815126116.

Kyuno D, Takasawa A, Kikuchi S, Takemasa I, Osanai M, Kojima T (2021). Role of tight junctions in the epithelial-to-mesenchymal transition of cancer cells. *Biochimica et Biophysica Acta. Biomembranes* **1863**: 183503. DOI 10.1016/j.bbamem.2020.183503.

Li H, Peng C, Zhu C, Nie S, Qian X et al. (2021a). Hypoxia promotes the metastasis of pancreatic cancer through regulating NOX4/KDM5A-mediated histone methylation modification changes in a HIF1A-independent manner. *Clinical Epigenetics* **13**: 18. DOI 10.1186/s13148-021-01016-6.

Li W, Kang Y (2016). Probing the fifty shades of EMT in metastasis. *Trends in Cancer* **2**: 65–67. DOI 10.1016/j.trecan.2016.01.001.

Li W, Qin Y, Zhou R, Liu Y, Zhang G (2021b). High expression of SIX1 is an independent predictor of poor prognosis in endometrial cancer. *American Journal of Translational Research* **13**: 2840–2848.

Lin H, Sun S, Lu X, Chen P, Chen C et al. (2017). CCR10 activation stimulates the invasion and migration of breast cancer cells through the ERK1/2/MMP-7 signaling pathway. *International Immunopharmacology* **51**: 124–130. DOI 10.1016/j.intimp.2017.07.018.

Llanos A, Yao S, Singh A, Aremu J, Khiabanian H et al. (2021). Gene expression of adipokines and adipokine receptors in the tumor microenvironment: Associations of lower expression with more aggressive breast tumor features. *Breast Cancer Research and Treatment* **185**: 785–798. DOI 10.1007/s10549-020-05972-0.

Luo X, Cao M, Gao F, He X (2021). YTHDF1 promotes hepatocellular carcinoma progression via activating PI3K/AKT/mTOR signaling pathway and inducing epithelial-mesenchymal transition. *Experimental Hematology & Oncology* **10**: 35. DOI 10.1186/s40164-021-00227-0.

Milioli H, Tishchenko I, Riveros C, Berretta R, Moscato P (2017). Basal-like breast cancer: Molecular profiles, clinical features and survival outcomes. *BMC Medical Genomics* **10**: 19. DOI 10.1186/s12920-017-0250-9.

Qiu Y, Yang L, Liu H, Luo X (2021). Cancer stem cell-targeted therapeutic approaches for overcoming trastuzumabresistance in HER2-positive breast cancer. *Stem Cells* **39**: 1125–1136.

Racle J, De Jonge K, Baumgaertner P, Speiser D, Gfeller D (2017). Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *eLife* **6**: e26476. DOI 10.7554/eLife.26476.

Ritchie M, Phipson B, Wu D, Hu Y, Law C et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**: e47. DOI 10.1093/nar/gkv007.

Saatci O, Kaymak A, Raza U, Ersan P, Akbulut O et al. (2020). Targeting lysyl oxidase (LOX) overcomes chemotherapy resistance in triple negative breast cancer. *Nature Communications* **11**: 2416. DOI 10.1038/s41467-020-16199-4.

Semmlinger A, Von Schoenfeldt V, Wolf V, Meuter A, Kolben T et al. (2018). EP3 (prostaglandin E2 receptor 3) expression is a prognostic factor for progression-free and overall survival in sporadic breast cancer. *BMC Cancer* **18**: 431. DOI 10.1186/s12885-018-4286-9.

Shannon P, Markiel A, Ozier O, Baliga N, Wang J et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**: 2498–2504. DOI 10.1101/gr.1239303.

Shi S, Xu C, Fang X, Zhang Y, Li H et al. (2020). Expression profile of Toll-like receptors in human breast cancer. *Molecular Medicine Reports* **21**: 786–794.

Solinas C, Marcoux D, Garaud S, Vitória J, van den Eynden G et al. (2019). BRCA gene mutations do not shape the extent and organization of tumor infiltrating lymphocytes in triple negative breast cancer. *Cancer Letters* **450**: 88–97. DOI 10.1016/j.canlet.2019.02.027.

Subramanian A, Tamayo P, Mootha V, Mukherjee S, Ebert B et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 15545–15550. DOI 10.1073/pnas.0506580102.

Sung H, Ferlay J, Siegel R, Laversanne M, Soerjomataram I et al. (2021). Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **71**: 209–249.

Szklarczyk D, Gable A, Lyon D, Junge A, Wyder S et al. (2019). STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research* **47**: D607–D613. DOI 10.1093/nar/gky1131.

Tang Q, Efe G, Chiarella A, Leung J, Chen M et al. (2021). Mutant p53 regulates Survivin to foster lung metastasis. *Genes & Development* **35**: 528–541. DOI 10.1101/gad.340505.120.

Tiwari A, Tashiro K, Dixit A, Soni A, Vogel K et al. (2020). Loss of HIF1A from pancreatic cancer cells increases expression of PPP1R1B and degradation of p53 to promote invasion and metastasis. *Gastroenterology* **159**: 1882–1897. DOI 10.1053/j.gastro.2020.07.046.

Triulzi T, De Cecco L, Sandri M, Prat A, Giussani M et al. (2015). Whole-transcriptome analysis links trastuzumab sensitivity of breast tumors to both HER2 dependence and immune cell infiltration. *Oncotarget* **6**: 28173–28182. DOI 10.18632/oncotarget.4405.

Tsubakihara Y, Moustakas A (2018). Epithelial-mesenchymal transition and metastasis under the control of transforming growth factor β. *International Journal of Molecular Sciences* **19**: 3672. DOI 10.3390/ijms19113672.

Wang H, Yin Y, Wang P, Xiong C, Huang L et al. (2016). Current situation and future usage of anticancer drug databases. *Apoptosis: An International Journal on Programmed Cell Death* **21**: 778–794. DOI 10.1007/s10495-016-1250-5.

Wang Q, Shi M, Sun S, Zhou Q, Ding L et al. (2020). CLEC5A promotes the proliferation of gastric cancer cells by activating the PI3K/AKT/mTOR pathway. *Biochemical and Biophysical Research Communications* **524**: 656–662. DOI 10.1016/j.bbrc.2019.10.122.

Yang B, Xie J, Li Z, Su D, Lin L et al. (2021). Seven-gene signature on tumor microenvironment for predicting the prognosis of patients with pancreatic cancer. *Gland Surgery* **10**: 1397–1409. DOI 10.21037/gs-21-28.

Yang C, Wang Y, Sims M, Cai C, Pfeffer L (2019a). MicroRNA-1 suppresses glioblastoma in preclinical models by targeting fibronectin. *Cancer Letters* **465**: 59–67. DOI 10.1016/j.canlet.2019.08.021.

Yang M, Tao L, Yang J, Jiang Y, Fu X et al. (2019b). SFRP4 is a prognostic marker and correlated with Treg cell infiltration in pancreatic ductal adenocarcinoma. *American Journal of Cancer Research* **9**: 363–377.

Yuan C, Xie Z, Bian J, Huo J, Daily K (2020). Outcomes of primary endocrine therapy in elderly women with stage I-III breast cancer: A SEER database analysis. *Breast Cancer Research and Treatment* **180**: 819–827. DOI 10.1007/s10549-020-05591-9.

Zhang D, An X, Yu H, Li Z (2021). The regulatory effect of 6-TG on lncRNA-miRNA-mRNA ceRNA network in triple-negative breast cancer cell line. *Bioscience Reports* **41**: BSR20203890. DOI 10.1042/BSR20203890.

Zhang G, Zhong L, Luo H, Wang S (2019). MicroRNA-155-3p promotes breast cancer progression through down-regulating CADM1. *OncoTargets and Therapy* **12**: 7993–8002. DOI 10.2147/OTT.

Zhang L, Liu Q, Mu Q, Zhou D, Li H et al. (2020). MiR-429 suppresses proliferation and invasion of breast cancer via inhibiting the Wnt/β-catenin signaling pathway. *Thoracic Cancer* **11**: 3126–3138. DOI 10.1111/1759-7714.13620.

Zhou P, Zheng G, Li Y, Wu D, Chen Y (2020a). Construction of a circRNA-miRNA-mRNA network related to macrophage infiltration in hepatocellular carcinoma. *Frontiers in Genetics* **11**: 1026. DOI 10.3389/fgene.2020.01026.

Zhou T, Damsky W, Weizman OE, Mcgeary MK, Hartmann KP et al. (2020b). IL-18BP is a secreted immune checkpoint and barrier to IL-18 immunotherapy. *Nature* **583**: 609–614. DOI 10.1038/s41586-020-2422-6.

Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi A et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications* **10**: 1523. DOI 10.1038/s41467-019-09234-6.

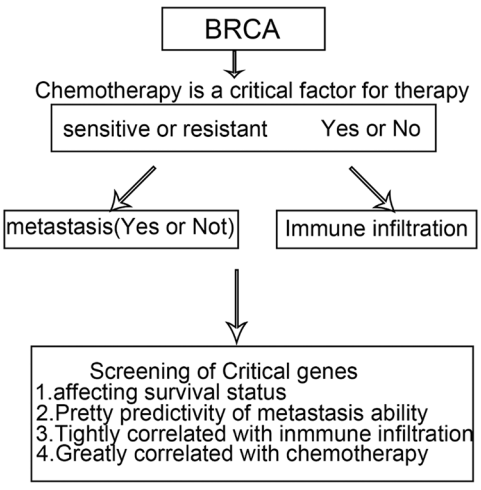

FIGURE S1. The workflow of analysis in our study.

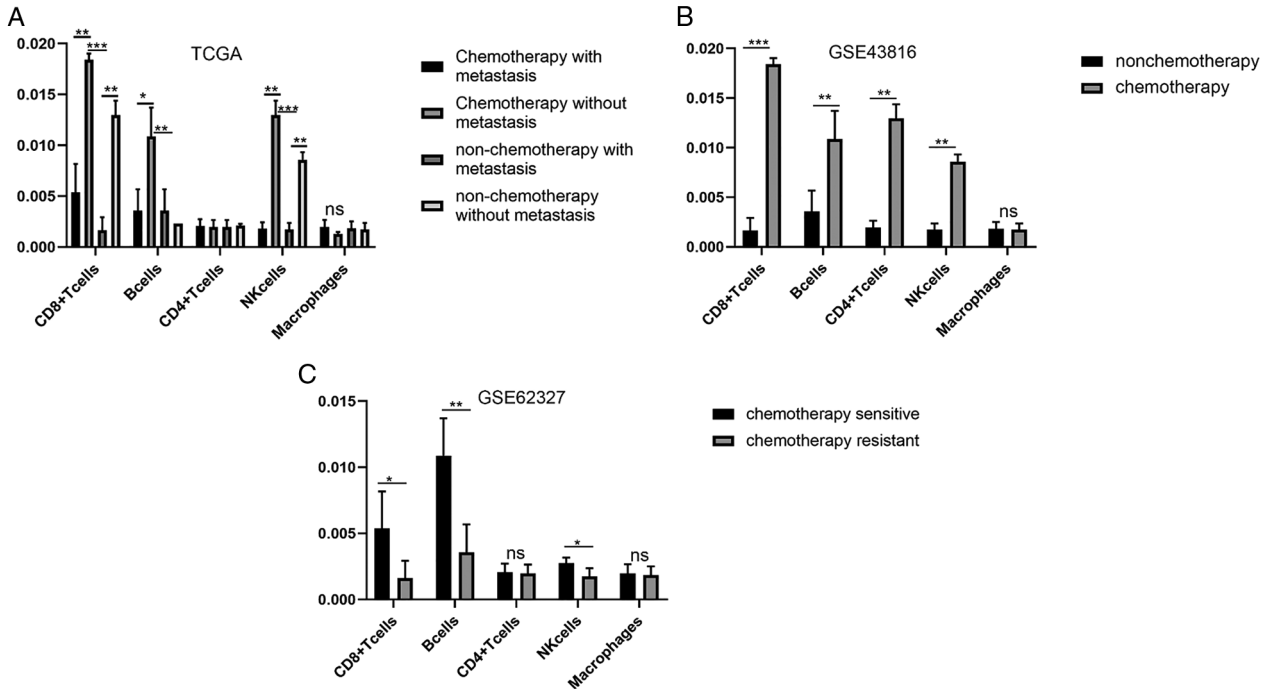

FIGURE S2. Expression of immune cell infiltration based on TCGA, GEO sample data. The level of immune cell in metastatic and non-metastatic samples of breast cancer in TCGA dataset (A). The level of immune cell in chemotherapy and non-chemotherapy samples of breast cancer in GSE43816 dataset (B). The level of immune cell in chemotherapy sensitive and chemotherapy resistant samples of breast cancer in GSE62327 dataset (C).

SUPPLEMENTARY TABLE 1

Clinical parameters of patients in TCGA

Samples	Chemotherapy	Sex	OS	Histological_type	Tumor_tissue_site	Event	T	N	M	Stage	Cancer_status	Age
TCGA-XX-A89A	Non-chemotherapy	FEMALE	488	Infiltrating lobular carcinoma	Breast	Alive	T3	N0	MX	Stage IIB	TUMOR FREE	68
TCGA-UU-A93S	Non-chemotherapy	FEMALE	116	Infiltrating ductal carcinoma	Breast	Dead	T4d	N3b	M1	Stage IV	WITH TUMOR	63
TCGA-PL-A8LX	Non-chemotherapy	FEMALE	5	Infiltrating ductal carcinoma	Breast	Alive	T4	N1a	M1	Stage IV	WITH TUMOR	35
TCGA-OL-A97C	Non-chemotherapy	FEMALE	271	Other, specify	Breast	Alive	T3	N0	M0	Stage IIB	TUMOR FREE	67
TCGA-OL-A6VR	Non-chemotherapy	FEMALE	1220	Infiltrating ductal carcinoma	Breast	Alive	T1b	N0	MX	Stage IA	TUMOR FREE	48
TCGA-OL-A66 K	Non-chemotherapy	FEMALE	1275	Infiltrating Lobular Carcinoma	Breast	Dead	T2	N0	MX	Stage IIA	TUMOR FREE	72
TCGA-OL-A5DA	Non-chemotherapy	FEMALE	1783	Infiltrating lobular carcinoma	Breast	Alive	T2	N0	MX	Stage IIA	TUMOR FREE	61
TCGA-OL-A5D8	Non-chemotherapy	FEMALE	973	Infiltrating Ductal Carcinoma	Breast	Alive	T3	N0	MX	Stage IIB	TUMOR FREE	40
TCGA-OL-A5D6	Non-chemotherapy	FEMALE	1104	Infiltrating ductal carcinoma	Breast	Dead	T2	N0	MX	Stage IIA	TUMOR FREE	71
TCGA-LL-A6FQ	Non-chemotherapy	FEMALE	80	Infiltrating ductal carcinoma	Breast	Alive	T2	N2a	MX	Stage IIIA	Unknown	77
TCGA-GI-A2C9	Chemotherapy	FEMALE	3342	Infiltrating ductal carcinoma	Breast	Alive	T3	N0	MX	Stage IIB	TUMOR FREE	58
TCGA-EW-A6SB	Non-chemotherapy	FEMALE	760	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage II	TUMOR FREE	62
TCGA-EW-A1PC	Chemotherapy	FEMALE	187	Infiltrating ductal carcinoma	Breast	Alive	T3	N0 (i-)	MX	Stage IIB	TUMOR FREE	66
TCGA-EW-A1PB	Non-chemotherapy	FEMALE	608	Infiltrating Ductal Carcinoma	Breast	Alive	T3	N1a	MX	Stage IIIA	TUMOR FREE	70
TCGA-EW-A1P0	Chemotherapy	FEMALE	1251	Other, specify	Breast	Alive	T2	N1b	MX	Stage IIB	WITH TUMOR	55
TCGA-EW-A1J6	Chemotherapy	FEMALE	875	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0 (i-)	M0	Stage I	TUMOR FREE	70
TCGA-EW-A1IY	Chemotherapy	FEMALE	258	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0	MX	Stage I	TUMOR FREE	38
TCGA-EW-A1IX	Chemotherapy	FEMALE	1208	Other, specify	Breast	Alive	T1b	N1a	MX	Stage IIA	TUMOR FREE	48
TCGA-EW-A1IW	Chemotherapy	FEMALE	371	Infiltrating lobular carcinoma	Breast	Alive	T2	N1a	MX	Stage IIB	TUMOR FREE	80
TCGA-E9-A5FL	Non-chemotherapy	FEMALE	24	Metaplastic carcinoma	Breast	Alive	T3	N0	M0	Stage IIB	TUMOR FREE	65
TCGA-E9-A24A	Chemotherapy	FEMALE	747	Infiltrating ductal carcinoma	Breast	Alive	T1c	N1	M0	Stage IIA	TUMOR FREE	69
TCGA-E9-A247	Chemotherapy	FEMALE	1186	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0 (i-)	M0	Stage IA	TUMOR FREE	59
TCGA-E9-A245	Chemotherapy	FEMALE	26	Infiltrating ductal carcinoma	Breast	Alive	T2	N1	M0	Stage IIB	TUMOR FREE	47
TCGA-E9-A22B	Chemotherapy	FEMALE	1167	Mixed histology (please specify)	Breast	Alive	T1c	N0	M0	Stage IA	TUMOR FREE	71
TCGA-E9-A22A	Chemotherapy	FEMALE	1189	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	74
TCGA-E9-A1RD	Chemotherapy	FEMALE	34	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	67
TCGA-E9-A1R6	Chemotherapy	FEMALE	339	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	63
TCGA-E9-A1R5	Chemotherapy	FEMALE	92	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0	M0	Stage IA	TUMOR FREE	63
TCGA-E9-A1QZ	Non-chemotherapy	FEMALE	755	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	61
TCGA-E2-A1LK	Non-chemotherapy	FEMALE	266	Infiltrating ductal carcinoma	Breast	Dead	T4b	N3a	M0	Stage IIIC	WITH TUMOR	84
TCGA-E2-A1LA	Chemotherapy	FEMALE	748	Infiltrating ductal carcinoma	Breast	Alive	T1c	N1a	M0	Stage IIA	TUMOR FREE	59

(Continued)

Supplementary Table 1 (continued).

Samples	Chemotherapy	Sex	OS	Histological_type	Tumor_tissue_site	Event	T	N	M	Stage	Cancer_status	Age
TCGA-E2-A1IU	Chemotherapy	FEMALE	337	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0 (mol+)	M0	Stage IA	TUMOR FREE	60
TCGA-E2-A1IN	Chemotherapy	FEMALE	675	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0	M0	Stage I	TUMOR FREE	60
TCGA-E2-A1IL	Chemotherapy	FEMALE	118	Infiltrating lobular carcinoma	Breast	Alive	T1c	N1a	M0	Stage IIA	Not Available	78
TCGA-E2-A1IK	Chemotherapy	FEMALE	1800	Infiltrating ductal carcinoma	Breast	Alive	T1c	N1mi	M0	Stage IIA	TUMOR FREE	71
TCGA-E2-A1IJ	Chemotherapy	FEMALE	865	Infiltrating lobular carcinoma	Breast	Alive	T1c	N0	M0	Stage I	TUMOR FREE	57
TCGA-E2-A1IH	Chemotherapy	FEMALE	1026	Infiltrating lobular carcinoma	Breast	Alive	T1c	N0	M0	Stage I	TUMOR FREE	80
TCGA-E2-A1IF	Chemotherapy	FEMALE	1138	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0	M0	Stage I	TUMOR FREE	74
TCGA-E2-A1BD	Chemotherapy	FEMALE	1133	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	53
TCGA-E2-A1BC	Chemotherapy	FEMALE	501	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0	M0	Stage IA	TUMOR FREE	63
TCGA-E2-A1B4	Chemotherapy	FEMALE	1004	Infiltrating ductal carcinoma	Breast	Dead	T1c	N2a	M0	Stage IIIA	TUMOR FREE	74
TCGA-E2-A15 T	Chemotherapy	FEMALE	1563	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	65
TCGA-E2-A15 M	Chemotherapy	FEMALE	336	Infiltrating lobular carcinoma	Breast	Dead	T2	N0	M0	Stage IIA	TUMOR FREE	66
TCGA-E2-A15L	Chemotherapy	FEMALE	626	Infiltrating lobular carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	65
TCGA-E2-A15I	Chemotherapy	FEMALE	1692	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	44
TCGA-E2-A15H	Chemotherapy	FEMALE	393	Infiltrating ductal carcinoma	Breast	Alive	T1c	N1mi	M0	Stage IIA	TUMOR FREE	38
TCGA-E2-A15G	Chemotherapy	FEMALE	554	Mixed histology (please specify)	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	76
TCGA-E2-A15F	Chemotherapy	FEMALE	658	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0	M0	Stage I	TUMOR FREE	64
TCGA-E2-A15D	Chemotherapy	FEMALE	526	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	47
TCGA-E2-A15C	Chemotherapy	FEMALE	694	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0	M0	Stage I	TUMOR FREE	61
TCGA-E2-A156	Chemotherapy	FEMALE	726	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0	M0	Stage I	TUMOR FREE	61
TCGA-E2-A154	Chemotherapy	FEMALE	591	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0	M0	Stage I	TUMOR FREE	68
TCGA-E2-A14Q	Chemotherapy	FEMALE	1163	Infiltrating ductal carcinoma	Breast	Alive	T2	N1mi	M0	Stage IIB	TUMOR FREE	50
TCGA-E2-A14P	Chemotherapy	FEMALE	1246	Infiltrating ductal carcinoma	Breast	Alive	T2	N3	M0	Stage IIIC	TUMOR FREE	79
TCGA-E2-A14O	Chemotherapy	FEMALE	1359	Infiltrating ductal carcinoma	Breast	Alive	T3	N1a	M0	Stage IIIA	TUMOR FREE	76
TCGA-E2-A105	Chemotherapy	FEMALE	1308	Infiltrating ductal carcinoma	Breast	Alive	T2	N0 (i-)	M0	Stage IIA	TUMOR FREE	79
TCGA-D8-A3Z5	Non-chemotherapy	FEMALE	1015	Infiltrating lobular carcinoma	Breast	Alive	T2	N3a	M0	Stage IIIC	TUMOR FREE	54
TCGA-D8-A27 V	Chemotherapy	FEMALE	381	Infiltrating lobular carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	62
TCGA-D8-A27R	Chemotherapy	FEMALE	307	Infiltrating ductal carcinoma	Breast	Alive	T2	N3a	M0	Stage IIIC	TUMOR FREE	41
TCGA-D8-A27P	Non-chemotherapy	FEMALE	49	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0	M0	Stage IA	TUMOR FREE	64
TCGA-D8-A27 N	Chemotherapy	FEMALE	519	Infiltrating ductal carcinoma	Breast	Alive	T2	N2a	M0	Stage IIIA	TUMOR FREE	36
TCGA-D8-A27 M	Chemotherapy	FEMALE	410	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0	M0	Stage IA	TUMOR FREE	59
TCGA-D8-A27L	Chemotherapy	FEMALE	499	Infiltrating ductal carcinoma	Breast	Alive	T1c	N2a	M0	Stage IIIA	TUMOR FREE	49
TCGA-D8-A27I	Chemotherapy	FEMALE	439	Infiltrating lobular carcinoma	Breast	Alive	T1c	N2a	M0	Stage IIIA	TUMOR FREE	58

(Continued)

Supplementary Table 1 (continued).

Samples	Chemotherapy	Sex	OS	Histological_type	Tumor_tissue_site	Event	T	N	M	Stage	Cancer_status	Age
TCGA-D8-A27H	Chemotherapy	FEMALE	397	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	72
TCGA-D8-A27F	Chemotherapy	FEMALE	488	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	40
TCGA-D8-A27E	Chemotherapy	FEMALE	530	Other, specify	Breast	Alive	T1c	N0	M0	Stage IA	TUMOR FREE	66
TCGA-D8-A1Y2	Chemotherapy	FEMALE	433	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	MX	Stage IIA	TUMOR FREE	71
TCGA-D8-A1Y1	Chemotherapy	FEMALE	302	Infiltrating ductal carcinoma	Breast	Dead	T3	N1a	MX	Stage IIIA	TUMOR FREE	80
TCGA-D8-A1XZ	Chemotherapy	FEMALE	466	Infiltrating ductal carcinoma	Breast	Alive	T1c	N2a	MX	Stage IIIA	TUMOR FREE	81
TCGA-D8-A1XY	Chemotherapy	FEMALE	503	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	MX	Stage IIA	TUMOR FREE	74
TCGA-D8-A1XW	Chemotherapy	FEMALE	1309	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	53
TCGA-D8-A1XT	Chemotherapy	FEMALE	506	Infiltrating ductal carcinoma	Breast	Alive	T1c	N1a	M0	Stage IIA	TUMOR FREE	61
TCGA-D8-A1XS	Chemotherapy	MALE	496	Other, specify	Breast	Alive	T2	N3a	M0	Stage IIIC	TUMOR FREE	48
TCGA-D8-A1XR	Chemotherapy	FEMALE	482	Infiltrating ductal carcinoma	Breast	Alive	T2	N1a	M0	Stage IIB	TUMOR FREE	56
TCGA-D8-A1XQ	Non-chemotherapy	FEMALE	499	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	69
TCGA-D8-A1XM	Non-chemotherapy	FEMALE	538	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0	M0	Stage IA	TUMOR FREE	57
TCGA-D8-A1XJ	Chemotherapy	FEMALE	664	Other, specify	Breast	Alive	T3	N1a	MX	Stage IIIA	TUMOR FREE	76
TCGA-D8-A1XF	Chemotherapy	FEMALE	463	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	45
TCGA-D8-A1XD	Non-chemotherapy	FEMALE	522	Infiltrating ductal carcinoma	Breast	Alive	T1c	N2a	M0	Stage IIIA	TUMOR FREE	36
TCGA-D8-A1XB	Non-chemotherapy	FEMALE	552	Infiltrating ductal carcinoma	Breast	Alive	T2	N1a	M0	Stage IIB	TUMOR FREE	62
TCGA-D8-A1X9	Chemotherapy	FEMALE	727	Infiltrating ductal carcinoma	Breast	Alive	T2	N1a	M0	Stage IIB	TUMOR FREE	66
TCGA-D8-A1X8	Chemotherapy	FEMALE	783	Infiltrating lobular carcinoma	Breast	Alive	T1c	N2a	M0	Stage IIIA	TUMOR FREE	62
TCGA-D8-A1X7	Chemotherapy	FEMALE	509	Other, specify	Breast	Alive	T2	N0	MX	Stage IIA	Not Available	40
TCGA-D8-A1X6	Chemotherapy	FEMALE	541	Infiltrating ductal carcinoma	Breast	Alive	T3	N2a	MX	Stage IIIA	TUMOR FREE	80
TCGA-D8-A1JU	Chemotherapy	FEMALE	447	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0	M0	Stage IA	TUMOR FREE	51
TCGA-D8-A1JT	Chemotherapy	FEMALE	405	Other, specify	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	70
TCGA-D8-A1JS	Chemotherapy	FEMALE	371	Other, specify	Breast	Alive	T1c	N0	M0	Stage IA	TUMOR FREE	77
TCGA-D8-A1JP	Chemotherapy	FEMALE	639	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0	M0	Stage IA	Not Available	73
TCGA-D8-A1JN	Chemotherapy	FEMALE	620	Infiltrating lobular carcinoma	Breast	Alive	T3	N3a	MX	Stage IIIC	TUMOR FREE	80
TCGA-D8-A1JM	Chemotherapy	FEMALE	590	Infiltrating ductal carcinoma	Breast	Alive	T2	N1a	M0	Stage IIB	TUMOR FREE	59
TCGA-D8-A1JH	Chemotherapy	FEMALE	426	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0	M0	Stage IA	TUMOR FREE	56
TCGA-D8-A1J9	Non-chemotherapy	FEMALE	532	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0	M0	Stage IA	TUMOR FREE	48
TCGA-D8-A1J8	Chemotherapy	FEMALE	431	Infiltrating ductal carcinoma	Breast	Alive	T2	N1a	M0	Stage IIB	TUMOR FREE	77
TCGA-D8-A145	Chemotherapy	FEMALE	410	Infiltrating ductal carcinoma	Breast	Alive	T1c	N1a	M0	Stage IIA	TUMOR FREE	80
TCGA-D8-A140	Chemotherapy	FEMALE	403	Infiltrating ductal carcinoma	Breast	Alive	T2	N1a	M0	Stage IIB	Not Available	62
TCGA-C8-A8HQ	Non-chemotherapy	FEMALE	380	Mucinous carcinoma	Breast	Alive	T2	N1	M0	Stage IIB	TUMOR FREE	53

(Continued)

Supplementary Table 1 (continued).

Samples	Chemotherapy	Sex	OS	Histological_type	Tumor_tissue_site	Event	T	N	M	Stage	Cancer_status	Age
TCGA-C8-A3M7	Non-chemotherapy	FEMALE	1034	Infiltrating lobular carcinoma	Breast	Dead	T4b	N0	M0	Stage IIIB	TUMOR FREE	60
TCGA-C8-A26Y	Non-chemotherapy	FEMALE	0	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	90
TCGA-C8-A1HN	Non-chemotherapy	FEMALE	394	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	56
TCGA-C8-A1HG	Non-chemotherapy	FEMALE	345	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	50
TCGA-C8-A1HF	Non-chemotherapy	FEMALE	332	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	48
TCGA-C8-A133	Non-chemotherapy	FEMALE	0	Other, specify	Breast	Alive	T3	N1	M0	Stage IIIA	TUMOR FREE	65
TCGA-C8-A12X	Non-chemotherapy	FEMALE	385	Mucinous carcinoma	Breast	Alive	T2	N1	M0	Stage IIB	TUMOR FREE	62
TCGA-C8-A12 V	Non-chemotherapy	FEMALE	385	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	55
TCGA-BH-A5J0	Non-chemotherapy	FEMALE	715	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0	M0	Stage IA	TUMOR FREE	63
TCGA-BH-A42 V	Non-chemotherapy	FEMALE	635	Infiltrating ductal carcinoma	Breast	Alive	T1c	N1mi	M0	Stage IB	TUMOR FREE	41
TCGA-BH-A42U	Non-chemotherapy	FEMALE	3364	Infiltrating lobular carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	Unknown	80
TCGA-BH-A208	Non-chemotherapy	FEMALE	1759	Infiltrating ductal carcinoma	Breast	Dead	T2	N1b	M0	Stage IIB	TUMOR FREE	48
TCGA-BH-A204	Non-chemotherapy	FEMALE	2534	Infiltrating ductal carcinoma	Breast	Dead	T2	N1b	M0	Stage IIB	Not Available	80
TCGA-BH-A1FR	Non-chemotherapy	FEMALE	1642	Other, specify	Breast	Dead	T4b	N1a	M0	Stage IIIB	WITH TUMOR	73
TCGA-BH-A1FH	Non-chemotherapy	FEMALE	1034	Infiltrating ductal carcinoma	Breast	Dead	T2	N1b	M1	Stage IV	WITH TUMOR	47
TCGA-BH-A1FC	Non-chemotherapy	FEMALE	3472	Medullary carcinoma	Breast	Dead	T1c	N1b	M0	Stage IIA	TUMOR FREE	78
TCGA-BH-A1F5	Non-chemotherapy	FEMALE	2712	Infiltrating ductal carcinoma	Breast	Dead	T1c	N1a	M0	Stage IIA	TUMOR FREE	62
TCGA-BH-A1F2	Non-chemotherapy	FEMALE	959	Infiltrating ductal carcinoma	Breast	Dead	T4b	N1b	M0	Stage IIIB	TUMOR FREE	53
TCGA-BH-A1EX	Non-chemotherapy	FEMALE	1508	Infiltrating ductal carcinoma	Breast	Dead	T2	N1b	M0	Stage IIB	WITH TUMOR	67
TCGA-BH-A1EV	Non-chemotherapy	FEMALE	365	Infiltrating ductal carcinoma	Breast	Dead	T3	N1b	M0	Stage IIIA	WITH TUMOR	45
TCGA-BH-A18 V	Non-chemotherapy	FEMALE	1556	Infiltrating ductal carcinoma	Breast	Dead	T2	N1b	M0	Stage IIB	Not Available	48
TCGA-BH-A18 T	Non-chemotherapy	FEMALE	224	Infiltrating ductal carcinoma	Breast	Dead	T2	N0	M0	Stage IIA	WITH TUMOR	70
TCGA-BH-A18R	Non-chemotherapy	FEMALE	1142	Infiltrating ductal carcinoma	Breast	Dead	T2b	N1	M0	Stage IIA	TUMOR FREE	50
TCGA-BH-A18Q	Non-chemotherapy	FEMALE	1692	Infiltrating ductal carcinoma	Breast	Dead	T2	N1b	M0	Stage IIB	Not Available	56
TCGA-BH-A18P	Non-chemotherapy	FEMALE	921	Infiltrating ductal carcinoma	Breast	Dead	T1	N0	M0	Stage I	WITH TUMOR	60
TCGA-BH-A18L	Non-chemotherapy	FEMALE	811	Infiltrating ductal carcinoma	Breast	Dead	T3	N1mi	M0	Stage IIIA	Not Available	50
TCGA-BH-A18 K	Non-chemotherapy	FEMALE	2763	Infiltrating ductal carcinoma	Breast	Dead	T1	N0	M0	Stage I	WITH TUMOR	46
TCGA-BH-A0WA	Non-chemotherapy	FEMALE	701	Not available	Breast	Alive	T1c	N0 (i-)	M0	Stage I	TUMOR FREE	82
TCGA-BH-A0HQ	Chemotherapy	FEMALE	1121	Infiltrating ductal carcinoma	Breast	Alive	T2	N0 (i+)	M0	Stage IIA	TUMOR FREE	56
TCGA-BH-A0HN	Chemotherapy	FEMALE	516	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0 (i-)	M0	Stage IA	TUMOR FREE	67
TCGA-BH-A0HK	Chemotherapy	FEMALE	178	Infiltrating ductal carcinoma	Breast	Alive	T2	N1	M0	Stage IIB	TUMOR FREE	81
TCGA-BH-A0HI	Chemotherapy	FEMALE	620	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0 (i-)	M0	Stage IA	TUMOR FREE	78
TCGA-BH-A0HF	Chemotherapy	FEMALE	727	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0 (i-)	M0	Stage IA	TUMOR FREE	77

(Continued)

Supplementary Table 1 (continued).

Samples	Chemotherapy	Sex	OS	Histological_type	Tumor_tissue_site	Event	T	N	M	Stage	Cancer_status	Age
TCGA-BH-A0H9	Chemotherapy	FEMALE	1247	Infiltrating ductal carcinoma	Breast	Alive	T2	N0 (i-)	M0	Stage IIA	TUMOR FREE	69
TCGA-BH-A0H6	Non-chemotherapy	FEMALE	747	Infiltrating ductal carcinoma	Breast	Alive	T1b	NX	M0	Stage I	TUMOR FREE	82
TCGA-BH-A0H3	Non-chemotherapy	FEMALE	1928	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0	M0	Stage I	TUMOR FREE	46
TCGA-BH-A0H0	Chemotherapy	FEMALE	461	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0 (i-)	M0	Stage IA	TUMOR FREE	69
TCGA-BH-A0GZ	Chemotherapy	FEMALE	328	Infiltrating ductal carcinoma	Breast	Alive	T1c	N1a	M0	Stage IIA	TUMOR FREE	62
TCGA-BH-A0EI	Non-chemotherapy	FEMALE	1926	Infiltrating ductal carcinoma	Breast	Alive	T1c	N1a	M0	Stage IIA	TUMOR FREE	51
TCGA-BH-A0EB	Chemotherapy	FEMALE	745	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0 (i-)	M0	Stage IA	TUMOR FREE	69
TCGA-BH-A0E7	Chemotherapy	FEMALE	1363	Infiltrating ductal carcinoma	Breast	Alive	T2	N1a	M0	Stage IIB	TUMOR FREE	79
TCGA-BH-A0DP	Chemotherapy	FEMALE	476	Infiltrating lobular carcinoma	Breast	Alive	T3	N0 (i-)	M0	Stage IIB	TUMOR FREE	60
TCGA-BH-A0DO	Non-chemotherapy	FEMALE	1644	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0	M0	Stage I	TUMOR FREE	78
TCGA-BH-A0DK	Chemotherapy	FEMALE	423	Infiltrating ductal carcinoma	Breast	Alive	T2	N0 (i-)	M0	Stage IIA	TUMOR FREE	49
TCGA-BH-A0BR	Non-chemotherapy	FEMALE	2330	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0	M0	Stage I	TUMOR FREE	59
TCGA-BH-A0BQ	Non-chemotherapy	FEMALE	2255	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0	M0	Stage I	TUMOR FREE	39
TCGA-BH-A0BP	Non-chemotherapy	FEMALE	2296	Infiltrating ductal carcinoma	Breast	Dead	T1c	N0	M0	Stage I	TUMOR FREE	76
TCGA-BH-A0BO	Non-chemotherapy	FEMALE	2197	Infiltrating ductal carcinoma	Breast	Alive	T1b	N0	M0	Stage I	TUMOR FREE	54
TCGA-BH-A0B8	Chemotherapy	FEMALE	1569	Infiltrating ductal carcinoma	Breast	Alive	T1b	N0 (i-)	M0	Stage I	TUMOR FREE	64
TCGA-BH-A0AU	Non-chemotherapy	FEMALE	1914	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	45
TCGA-B6-A400	Non-chemotherapy	FEMALE	215	Infiltrating ductal carcinoma	Breast	Alive	T2	N2a	M0	Stage IIIA	TUMOR FREE	43
TCGA-B6-A2IU	Non-chemotherapy	FEMALE	5176	Infiltrating lobular carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	Not Available	62
TCGA-B6-A1KC	Non-chemotherapy	FEMALE	1326	Infiltrating ductal carcinoma	Breast	Alive	T2	N1	M0	Stage IIB	TUMOR FREE	67
TCGA-B6-A0X7	Non-chemotherapy	FEMALE	1781	Mixed histology (please specify)	Breast	Dead	T1c	NX	MX	Stage X	Not Available	62
TCGA-B6-A0X4	Non-chemotherapy	FEMALE	860	Infiltrating ductal carcinoma	Breast	Dead	T2	N1b	M0	Stage IIB	WITH TUMOR	62
TCGA-B6-A0WY	Non-chemotherapy	FEMALE	3461	Infiltrating ductal carcinoma	Breast	Dead	T3	N1b	M0	Stage IIIA	WITH TUMOR	40
TCGA-B6-A0WV	Non-chemotherapy	FEMALE	2417	Infiltrating ductal carcinoma	Breast	Dead	T2	N1b	M0	Stage IIB	TUMOR FREE	67
TCGA-B6-A0RV	Non-chemotherapy	FEMALE	5156	Mixed histology (please specify)	Breast	Alive	T3	N2	M0	Stage IIIA	TUMOR FREE	42
TCGA-B6-A0RT	Non-chemotherapy	FEMALE	2721	Infiltrating ductal carcinoma	Breast	Alive	T3	N1	M0	Stage IIIA	TUMOR FREE	39
TCGA-B6-A0RO	Non-chemotherapy	FEMALE	4929	Infiltrating ductal carcinoma	Breast	Alive	T4	N1a	M0	Stage IIIB	TUMOR FREE	71
TCGA-B6-A0RL	Non-chemotherapy	FEMALE	2469	Infiltrating ductal carcinoma	Breast	Dead	T2	N0 (i-)	M0	Stage IIA	TUMOR FREE	60
TCGA-B6-A0RH	Non-chemotherapy	FEMALE	6456	Infiltrating ductal carcinoma	Breast	Dead	T2	N0 (i-)	M0	Stage IIA	TUMOR FREE	51
TCGA-B6-A0IQ	Non-chemotherapy	FEMALE	4285	Infiltrating ductal carcinoma	Breast	Alive	T3	N1b	M0	Stage IIIA	TUMOR FREE	40
TCGA-B6-A0IE	Non-chemotherapy	FEMALE	1993	Mixed histology (please specify)	Breast	Dead	T3	N1b	M0	Stage IIIA	WITH TUMOR	38
TCGA-B6-A0IC	Non-chemotherapy	FEMALE	0	Other, specify	Breast	Dead	T2	NX	MX	Stage X	Not Available	90
TCGA-B6-A0IB	Non-chemotherapy	FEMALE	3941	Infiltrating ductal carcinoma	Breast	Dead	T3	N3	M1	Stage IV	WITH TUMOR	64

(Continued)

Supplementary Table 1 (continued).

Samples	Chemotherapy	Sex	OS	Histological_type	Tumor_tissue_site	Event	T	N	M	Stage	Cancer_status	Age
TCGA-B6-A0I9	Non-chemotherapy	FEMALE	362	Infiltrating ductal carcinoma	Breast	Dead	T3	NX	M1	Stage IV	WITH TUMOR	62
TCGA-B6-A0I6	Non-chemotherapy	FEMALE	991	Infiltrating ductal carcinoma	Breast	Dead	T1c	N1	M0	Stage IIA	WITH TUMOR	49
TCGA-B6-A0I5	Non-chemotherapy	FEMALE	8556	Infiltrating ductal carcinoma	Breast	Alive	T2	N1b	M0	Stage IIB	TUMOR FREE	49
TCGA-B6-A0I2	Non-chemotherapy	FEMALE	4361	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0 (i-)	M0	Stage IA	TUMOR FREE	45
TCGA-AR-A1AP	Chemotherapy	FEMALE	2856	Infiltrating ductal carcinoma	Breast	Alive	T1	N0	M0	Stage I	TUMOR FREE	80
TCGA-AR-A1AK	Chemotherapy	FEMALE	3159	Infiltrating lobular carcinoma	Breast	Alive	T1	N0	M0	Stage I	TUMOR FREE	70
TCGA-AR-A0U0	Non-chemotherapy	FEMALE	1988	Infiltrating ductal carcinoma	Breast	Alive	T2	N1	M0	Stage IIB	Not available	73
TCGA-AR-A0TR	Chemotherapy	FEMALE	160	Infiltrating ductal carcinoma	Breast	Dead	T2	N1	M0	Stage IIB	Not available	68
TCGA-AQ-A54 N	Non-chemotherapy	FEMALE	78	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	51
TCGA-AO-A1KS	Chemotherapy	FEMALE	350	Infiltrating lobular carcinoma	Breast	Alive	T2	N0 (i-)	M0	Stage IIA	TUMOR FREE	69
TCGA-AO-A1KQ	Chemotherapy	MALE	1882	Infiltrating ductal carcinoma	Breast	Alive	T4	N1	M0	Stage IIIB	TUMOR FREE	84
TCGA-AO-A12H	Chemotherapy	FEMALE	1234	Other, specify	Breast	Alive	T2	N0 (i-)	M0	Stage IIA	TUMOR FREE	69
TCGA-AO-A12G	Chemotherapy	FEMALE	1639	Other, specify	Breast	Alive	T2	N0 (i+)	M0	Stage IIA	TUMOR FREE	75
TCGA-AN-A0XT	Non-chemotherapy	FEMALE	10	Infiltrating ductal carcinoma	Breast	Alive	T2	N2	M0	Stage IIIA	Not available	54
TCGA-AN-A0XR	Non-chemotherapy	FEMALE	10	Infiltrating ductal carcinoma	Breast	Alive	T2	N2	M0	Stage IIIA	Not available	55
TCGA-AN-A0XP	Non-chemotherapy	FEMALE	9	Infiltrating ductal carcinoma	Breast	Alive	T2	N2	M0	Stage IIIA	Not available	69
TCGA-AN-A0XN	Non-chemotherapy	FEMALE	10	Infiltrating ductal carcinoma	Breast	Alive	T2	N2	M0	Stage IIIA	TUMOR FREE	68
TCGA-AN-A0FZ	Non-chemotherapy	FEMALE	10	Infiltrating ductal carcinoma	Breast	Alive	T2	N2	M0	Stage IIIA	TUMOR FREE	45
TCGA-AN-A0FX	Non-chemotherapy	FEMALE	10	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	52
TCGA-AN-A0FL	Non-chemotherapy	FEMALE	231	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	62
TCGA-AN-A0AS	Non-chemotherapy	FEMALE	10	Infiltrating ductal carcinoma	Breast	Alive	T2	N2	M0	Stage IIIA	TUMOR FREE	70
TCGA-AN-A04C	Non-chemotherapy	FEMALE	54	Infiltrating ductal carcinoma	Breast	Alive	T2	N1	M0	Stage IIB	TUMOR FREE	51
TCGA-AC-A8OR	Non-chemotherapy	FEMALE	40	Mucinous carcinoma	Breast	Alive	T1c	N0 (i-)	MX	Stage IA	TUMOR FREE	75
TCGA-AC-A7VB	Non-chemotherapy	FEMALE	250	Infiltrating ductal carcinoma	Breast	Alive	T1c	N1a	MX	Stage IIA	TUMOR FREE	51
TCGA-AC-A6IX	Non-chemotherapy	FEMALE	373	Infiltrating lobular carcinoma	Breast	Alive	T2	N3a	MX	Stage IIIC	TUMOR FREE	49
TCGA-AC-A62Y	Non-chemotherapy	FEMALE	530	Infiltrating lobular carcinoma	Breast	Alive	T2	N1	MX	Stage IIB	TUMOR FREE	79
TCGA-AC-A62X	Non-chemotherapy	FEMALE	417	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	MX	Stage IIA	TUMOR FREE	72
TCGA-AC-A5EH	Non-chemotherapy	FEMALE	511	Infiltrating ductal carcinoma	Breast	Alive	T2	N1mi	MX	Stage IIB	TUMOR FREE	76
TCGA-AC-A3W6	Non-chemotherapy	FEMALE	0	Infiltrating lobular carcinoma	Breast	Alive	T3	N1	MX	Stage IIIA	TUMOR FREE	90
TCGA-AC-A3QQ	Non-chemotherapy	FEMALE	734	Infiltrating lobular carcinoma	Breast	Alive	T1c	N0 (i-)	MX	Stage IA	TUMOR FREE	54
TCGA-AC-A2FM	Non-chemotherapy	FEMALE	792	Infiltrating lobular carcinoma	Breast	Dead	T2	N1mi	M0	Stage IIB	WITH TUMOR	87
TCGA-A8-A0A7	Non-chemotherapy	FEMALE	30	Infiltrating lobular carcinoma	Breast	Alive	T2	N1a	M0	Stage IIB	TUMOR FREE	57
TCGA-A8-A09W	Non-chemotherapy	FEMALE	30	Infiltrating lobular carcinoma	Breast	Alive	T2	N3	M0	Stage IIIB	Not available	70

(Continued)

Supplementary Table 1 (continued).

Samples	Chemotherapy	Sex	OS	Histological_type	Tumor_tissue_site	Event	T	N	M	Stage	Cancer_status	Age
TCGA-A8-A096	Non-chemotherapy	FEMALE	0	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	73
TCGA-A8-A090	Non-chemotherapy	FEMALE	0	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	74
TCGA-A8-A08 T	Non-chemotherapy	FEMALE	3409	Infiltrating ductal carcinoma	Breast	Dead	T2	N1a	M1	Stage IV	WITH TUMOR	64
TCGA-A8-A08L	Non-chemotherapy	FEMALE	30	Infiltrating ductal carcinoma	Breast	Dead	T3	N2a	M0	Stage IIIA	TUMOR FREE	89
TCGA-A8-A08H	Non-chemotherapy	FEMALE	0	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	66
TCGA-A8-A06U	Non-chemotherapy	FEMALE	883	Infiltrating ductal carcinoma	Breast	Dead	T2	N1a	M0	Stage IIB	TUMOR FREE	80
TCGA-A7-A26 J	Chemotherapy	FEMALE	627	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	WITH TUMOR	49
TCGA-A7-A0DB	Chemotherapy	FEMALE	1007	Infiltrating ductal carcinoma	Breast	Alive	T2	N0 (i-)	M0	Stage IIA	TUMOR FREE	56
TCGA-A7-A0CH	Chemotherapy	FEMALE	1079	Infiltrating ductal carcinoma	Breast	Alive	T2	N0 (i-)	M0	Stage IIA	TUMOR FREE	79
TCGA-A7-A0CD	Chemotherapy	FEMALE	1165	Infiltrating ductal carcinoma	Breast	Alive	T1	N0	M0	Stage I	TUMOR FREE	66
TCGA-A2-A4S3	Non-chemotherapy	FEMALE	666	Infiltrating ductal carcinoma	Breast	Alive	T2	N1a	M0	Stage IIB	TUMOR FREE	59
TCGA-A2-A4S1	Non-chemotherapy	FEMALE	820	Metaplastic carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	TUMOR FREE	66
TCGA-A2-A4S0	Non-chemotherapy	FEMALE	706	Mucinous carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	Unknown	77
TCGA-A2-A259	Chemotherapy	FEMALE	1596	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0 (i-)	M0	Stage I	TUMOR FREE	70
TCGA-A2-A1FZ	Chemotherapy	FEMALE	683	Infiltrating ductal carcinoma	Breast	Alive	T2	N0 (i-)	M0	Stage IIA	TUMOR FREE	63
TCGA-A2-A0YT	Chemotherapy	FEMALE	723	Infiltrating ductal carcinoma	Breast	Dead	T4b	N2a	M0	Stage IIIB	TUMOR FREE	56
TCGA-A2-A0YI	Chemotherapy	FEMALE	1505	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0 (i+)	M0	Stage I	TUMOR FREE	62
TCGA-A2-A0YF	Chemotherapy	FEMALE	1535	Infiltrating ductal carcinoma	Breast	Alive	T1b	N0 (i+)	M0	Stage I	TUMOR FREE	67
TCGA-A2-A0YC	Chemotherapy	FEMALE	990	Infiltrating ductal carcinoma	Breast	Alive	T2	N1mi	M0	Stage IIB	TUMOR FREE	59
TCGA-A2-A0T4	Chemotherapy	FEMALE	624	Infiltrating lobular carcinoma	Breast	Alive	T2	N0 (i-)	M0	Stage IIA	TUMOR FREE	62
TCGA-A2-A0T2	Chemotherapy	FEMALE	255	Infiltrating ductal carcinoma	Breast	Dead	T3	N3	M1	Stage IV	WITH TUMOR	66
TCGA-A2-A0SU	Chemotherapy	FEMALE	1662	Infiltrating ductal carcinoma	Breast	Alive	T2	N0 (i+)	M0	Stage IIA	TUMOR FREE	66
TCGA-A2-A0EV	Non-chemotherapy	FEMALE	968	Infiltrating ductal carcinoma	Breast	Alive	T1c	N0 (i-)	M0	Stage IA	TUMOR FREE	80
TCGA-A2-A0EM	Chemotherapy	FEMALE	3094	Infiltrating ductal carcinoma	Breast	Alive	T1	N0 (i-)	M0	Stage IA	TUMOR FREE	73
TCGA-A2-A0CW	Chemotherapy	FEMALE	3283	Infiltrating ductal carcinoma	Breast	Alive	T2	N1a	M0	Stage IIB	TUMOR FREE	67
TCGA-A2-A0CU	Chemotherapy	FEMALE	158	Infiltrating ductal carcinoma	Breast	Dead	T2	N0 (i+)	M0	Stage IIA	TUMOR FREE	73
TCGA-A1-A0SP	Chemotherapy	FEMALE	584	Infiltrating ductal carcinoma	Breast	Alive	T2	N0 (i-)	M0	Stage IIA	TUMOR FREE	40
TCGA-A1-A0SO	Chemotherapy	FEMALE	852	Infiltrating ductal carcinoma	Breast	Alive	T2	N1	M0	Stage IIB	TUMOR FREE	67
TCGA-A1-A0SM	Non-chemotherapy	MALE	242	Infiltrating ductal carcinoma	Breast	Alive	T2	N0 (i-)	M0	Stage IIA	TUMOR FREE	77
TCGA-A1-A0SH	Chemotherapy	FEMALE	1437	Infiltrating ductal carcinoma	Breast	Alive	T2	N0 (i-)	M0	Stage IIA	TUMOR FREE	39
TCGA-A1-A0SG	Chemotherapy	FEMALE	434	Other, specify	Breast	Alive	T2	N1a	M0	Stage IIB	Not Available	61
TCGA-A1-A0SD	Non-chemotherapy	FEMALE	437	Infiltrating ductal carcinoma	Breast	Alive	T2	N0	M0	Stage IIA	Not Available	59

SUPPLEMENTARY TABLE 2

Clinical parameters of patients in GSE62327

Sample_title	ID	Tissue	Species	Tumor size	Lymphatic status	Grade	Response to chemotherapy	Tissue source	Dealing	Paltform
CQ18	GSM1525279	Breast cancer tumor	Homo sapiens	Tumor size: T2	Lymph node status: NA	grade: III	Response: PR	Tissue: breast cancer tumor	Total RNA	GPL14951
CP76	GSM1525280	Breast cancer tumor	Homo sapiens	Tumor size: T2	Lymph node status: neg	grade: III	Response: PR	Tissue: breast cancer tumor	Total RNA	GPL14951
CP59	GSM1525281	Breast cancer tumor	Homo sapiens	Tumor size: T2	Lymph node status: neg	grade: III	Response: CR	Tissue: breast cancer tumor	Total RNA	GPL14951
CQ00	GSM1525282	Breast cancer tumor	Homo sapiens	Tumor size: T2	Lymph node status: pos	grade: III	Response: CR	Tissue: breast cancer tumor	Total RNA	GPL14951
CP87	GSM1525283	Breast cancer tumor	Homo sapiens	Tumor size: T2	Lymph node status: pos	grade: III	Response: PR	Tissue: breast cancer tumor	Total RNA	GPL14951
CP67	GSM1525284	Breast cancer tumor	Homo sapiens	Tumor size: T3	Lymph node status: pos	grade: III	Response: PR	Tissue: breast cancer tumor	Total RNA	GPL14951
CQ06	GSM1525285	Breast cancer tumor	Homo sapiens	Tumor size: T2	Lymph node status: neg	grade: III	Response: CR	Tissue: breast cancer tumor	Total RNA	GPL14951
CQ09	GSM1525286	Breast cancer tumor	Homo sapiens	Tumor size: T4	Lymph node status: NA	grade: III	Response: PR	Tissue: breast cancer tumor	Total RNA	GPL14951
CQ03	GSM1525287	Breast cancer tumor	Homo sapiens	Tumor size: T2	Lymph node status: neg	grade: III	Response: PR	Tissue: breast cancer tumor	Total RNA	GPL14951
CQ12	GSM1525288	Breast cancer tumor	Homo sapiens	Tumor size: T3	Lymph node status: NA	grade: II	Response: CR	Tissue: breast cancer tumor	Total RNA	GPL14951
CQ23	GSM1525289	Breast cancer tumor	Homo sapiens	Tumor size: T2	Lymph node status: neg	grade: I	Response: PR	Tissue: breast cancer tumor	Total RNA	GPL14951
CP97	GSM1525290	Breast cancer tumor	Homo sapiens	Tumor size: T2	Lymph node status: NA	grade: III	Response: CR	Tissue: breast cancer tumor	Total RNA	GPL14951
CQ20	GSM1525291	Breast cancer tumor	Homo sapiens	Tumor size: T2	Lymph node status: neg	grade: II	Response: PR	Tissue: breast cancer tumor	Total RNA	GPL14951
CQ28	GSM1525292	Breast cancer tumor	Homo sapiens	Tumor size: T1	Lymph node status: neg	grade: II	Response: PR	Tissue: breast cancer tumor	Total RNA	GPL14951
CP61	GSM1525293	Breast cancer tumor	Homo sapiens	Tumor size: T3	Lymph node status: pos	grade: III	Response: PR	Tissue: breast cancer tumor	Total RNA	GPL14951
CP64	GSM1525294	Breast cancer tumor	Homo sapiens	Tumor size: T3	Lymph node status: neg	grade: III	Response: PR	Tissue: breast cancer tumor	Total RNA	GPL14951
CP70	GSM1525295	Breast cancer tumor	Homo sapiens	Tumor size: T3	Lymph node status: pos	grade: III	Response: PR	Tissue: breast cancer tumor	Total RNA	GPL14951

(Continued)

Supplementary Table 2 (continued).

Sample_title	ID	Tissue	Species	Tumor size	Lymphatic status	Grade	Response to chemotherapy	Tissue source	Dealing	Paltform
CP73	GSM1525296	Breast cancer tumor	Homo sapiens	Tumor size: T3	Lymph node status: neg	grade: III	Response: PR	Tissue: breast cancer tumor	Total RNA	GPL14951
CP89	GSM1525297	Breast cancer tumor	Homo sapiens	Tumor size: T3	Lymph node status: neg	grade: III	Response: PR	Tissue: breast cancer tumor	Total RNA	GPL14951
CP82	GSM1525298	Breast cancer tumor	Homo sapiens	Tumor size: T2	Lymph node status: pos	grade: III	Response: PR	Tissue: breast cancer tumor	Total RNA	GPL14951
CP79	GSM1525299	Breast cancer tumor	Homo sapiens	Tumor size: T2	Lymph node status: pos	grade: II	Response: PR	Tissue: breast cancer tumor	Total RNA	GPL14951
CP95	GSM1525300	Breast cancer tumor	Homo sapiens	Tumor size: T1	Lymph node status: neg	grade: III	Response: PR	Tissue: breast cancer tumor	Total RNA	GPL14951
CQ15	GSM1525301	Breast cancer tumor	Homo sapiens	Tumor size: T2	Lymph node status: neg	grade: III	Response: PR	Tissue: breast cancer tumor	Total RNA	GPL14951
CQ25	GSM1525302	Breast cancer tumor	Homo sapiens	Tumor size: T2	Lymph node status: pos	grade: III	Response: CR	Tissue: breast cancer tumor	Total RNA	GPL14951

SUPPLEMENTARY TABLE 3

: Clinical parameters of patients in GSE43816

ID	Group	Species	Tissue	Time point
GSM1071652	1_before_R	Homo sapiens	Tissue: breast cancer tissue	Before neoadjuvant chemotherapy
GSM1071653	1_after_R	Homo sapiens	Tissue: breast cancer tissue	After neoadjuvant chemotherapy
GSM1071654	2_before_R	Homo sapiens	Tissue: breast cancer tissue	Before neoadjuvant chemotherapy
GSM1071655	2_after_R	Homo sapiens	Tissue: breast cancer tissue	After neoadjuvant chemotherapy
GSM1071656	3_before_R	Homo sapiens	Tissue: breast cancer tissue	Before neoadjuvant chemotherapy
GSM1071657	3_after_R	Homo sapiens	Tissue: breast cancer tissue	After neoadjuvant chemotherapy
GSM1071658	4_before_R	Homo sapiens	Tissue: breast cancer tissue	Before neoadjuvant chemotherapy
GSM1071659	4_after_R	Homo sapiens	Tissue: breast cancer tissue	After neoadjuvant chemotherapy
GSM1071660	5_before_R	Homo sapiens	Tissue: breast cancer tissue	Before neoadjuvant chemotherapy
GSM1071661	5_after_R	Homo sapiens	Tissue: breast cancer tissue	After neoadjuvant chemotherapy
GSM1071662	6_before_R	Homo sapiens	Tissue: breast cancer tissue	Before neoadjuvant chemotherapy
GSM1071663	6_after_R	Homo sapiens	Tissue: breast cancer tissue	After neoadjuvant chemotherapy
GSM1071664	7_before_R	Homo sapiens	Tissue: breast cancer tissue	Before neoadjuvant chemotherapy
GSM1071665	7_after_R	Homo sapiens	Tissue: breast cancer tissue	After neoadjuvant chemotherapy

SUPPLEMENTARY TABLE 4

ID	Description	*P* value	P. adjust	Count	Total	Generatio
GO:0048608	Reproductive structure development	9.70031E-05	0	7	406	0.017241379
GO:0061458	Reproductive system development	0.000101536	0	7	409	0.01217656
GO:0003006	Developmental process involved in reproduction	0.00032934	0	8	657	0.020080321
GO:0045165	Cell fate commitment	0.000511153	0	5	249	0.013059701
GO:0002009	Morphogenesis of an epithelium	0.000523908	0	7	536	0.0375
GO:0001892	Embryonic placenta development	0.001241146	0	3	80	0.010971787
GO:0048729	Tissue morphogenesis	0.001443812	0	7	638	0.012077295
GO:0048568	Embryonic organ development	0.004726569	0	5	414	0.022058824
GO:0001890	Placenta development	0.005576953	0	3	136	0.019480519
GO:0001837	Epithelial to mesenchymal transition	0.00785017	0	3	154	0.039215686
GO:0006081	Cellular aldehyde metabolic process	0.001111289	0	3	77	0.03
GO:0022617	Extracellular matrix disassembly	0.001332807	0	3	82	0.026315789
GO:0003341	Cilium movement	0.001153569	0	3	78	0.023622047
GO:0001906	Cell killing	0.001342636	0	4	181	0.008962868
GO:0001909	Leukocyte mediated cytotoxicity	0.003407998	0	3	114	0.017964072
GO:0019730	Antimicrobial humoral response	0.006405946	0	3	143	0.015503876
GO:0030036	Actin cytoskeleton organization	0.002360082	0	7	696	0.01875
GO:0030029	Actin filament-based process	0.004698779	0	7	789	0.00887199
GO:0030855	Epithelial cell differentiation	0.004447186	0	7	781	0.008962868

SUPPLEMENTARY TABLE 5

ID	Description	*P* value	P. adjust	Count	Total	Generatio
hsa04925	Aldosterone synthesis and secretion	0.000154216	0	4	102	0.02919708
hsa04270	Vascular smooth muscle contraction	0.000475528	0	4	137	0.023391813
hsa04927	Cortisol synthesis and secretion	0.000952013	0	3	73	0.038961039
hsa04934	Cushing syndrome	0.00108879	0	4	171	0.019417476
hsa04020	Calcium signaling pathway	0.002153188	0	4	206	0.028846154
hsa04713	Circadian entrainment	0.002352657	0	3	100	0.010416667
hsa04750	Inflammatory mediator regulation of trp channels	0.002629719	0	3	104	0.071428571
hsa05200	Pathways in cancer	0.004100232	0	6	576	0.024390244
hsa04261	Adrenergic signaling in cardiomyocytes	0.006910896	0	3	147	0.038461538
hsa04080	Neuroactive ligand-receptor interaction	0.000452559	0	6	373	0.020979021
hsa04072	Phospholipase D signaling pathway	0.000932561	0	4	164	0.009493671

SUPPLEMENTARY TABLE 6

Name	Size	ES	NES	NOM p-val
GO_SERTOLI_CELL_DEVELOPMENT	12	0.74155694	1.8930616	0.004106776
GO_AXONEMAL_DYNEIN_COMPLEX_ASSEMBLY	15	0.78123885	1.8458536	0.004008016
GO_SERTOLI_CELL_DIFFERENTIATION	15	0.65457696	1.8127494	0.001992032
GO_REGULATION_OF_CILIUM_MOVEMENT	8	0.7736576	1.7660143	0.006085193
GO_INNER_DYNEIN_ARM_ASSEMBLY	9	0.8271263	1.764317	0.005928854
GO_REGULATION_OF_SISTER_CHROMATID_COHESION	11	0.6970716	1.7283411	0.016032064
GO_INTRA_GOLGI_VESICLE_MEDIATED_TRANSPORT	37	0.54170877	1.7083043	0.012931035
GO_OUTER_DYNEIN_ARM_ASSEMBLY	9	0.7990409	1.7036562	0.008281574
GO_POSITIVE_REGULATION_OF_PHOSPHOPROTEIN_PHOSPHATASE_ACTIVITY	12	0.619845	1.6891166	0.014084507
GO_CILIUM_MOVEMENT	26	0.62749183	1.6815434	0.022222223
GO_MITOTIC_SISTER_CHROMATID_COHESION	8	0.76784676	1.6774932	0.002057613
GO_NEGATIVE_REGULATION_OF_PROTEIN_LOCALIZATION_TO_CELL_PERIPHERY	12	0.6362113	1.6602308	0.042510122
GO_SIGNAL_PEPTIDE_PROCESSING	18	0.5500574	1.6463372	0.014373717
GO_AXONEME_ASSEMBLY	26	0.6190203	1.6379871	0.044806518
GO_RESPONSE_TO_IRON_ION	26	0.4724917	1.6342541	0.014675053
GO_RESPONSE_TO_X_RAY	20	0.5482543	1.6266178	0.03187251
GO_ER_TO_GOLGI_VESICLE_MEDIATED_TSRANSPORT	103	0.38668948	1.6009423	0.025423728
GO_ZINC_II_ION_TRANSPORT	17	0.51907265	1.5884295	0.031380754
GO_UTP_METABOLIC_PROCESS	7	0.68488634	1.5806099	0.030425964
GO_ANDROGEN_RECEPTOR_SIGNALING_PATHWAY	26	0.48187757	1.5614797	0.04517454
GO_UBIQUINONE_METABOLIC_PROCESS	6	0.7250298	1.5484565	0.038709678
GO_FUCOSE_METABOLIC_PROCESS	14	0.5319754	1.5310597	0.04809619
GO_MOTILE_CILIUM_ASSEMBLY	8	0.67667264	1.5244037	0.04733728
GO_INTRACELLULAR_STEROID_HORMONE_RECEPTOR_SIGNALING_PATHWAY	49	0.4016375	1.5136701	0.03411514

SUPPLEMENTARY TABLE 7

Gene	Ensembl ID	Description	Summary
IL18	ENSG00000150782.11	Interleukin 18	The protein encoded by this gene is a proinflammatory cytokine that augments natural killer cell activity in spleen cells, and stimulates interferon gamma production in T-helper type I cells.
LY86	ENSG00000112799.8	Lymphocyte antigen 86	————————————————————————————————————
FN1	ENSG00000115414.18	Fibronectin 1	This gene encodes fibronectin, a glycoprotein present in a soluble dimeric form in plasma, and in a dimeric or multimeric form at the cell surface and in extracellular matrix.
CD1C	ENSG00000158481.12	CD1c molecule	This gene encodes a member of the CD1 family of transmembrane glycoproteins, which are structurally related to the major histocompatibility complex (MHC) proteins and form heterodimers with beta-2-microglobulin.
TLR7	ENSG00000196664.4	Toll-like receptor 7	The protein encoded by this gene is a member of the Toll-like receptor (TLR) family which plays a fundamental role in pathogen recognition and activation of innate immunity.
C3AR1	ENSG00000171860.4	Complement component 3a receptor 1	————————————————————————————————————
TMEM30A	ENSG00000112697.15	Transmembrane protein 30A	————————————————————————————————————
MAPK1	ENSG00000100030.14	Mitogen-activated protein kinase 1	Extracellular signal-regulated kinases (ERKs), act as an integration point for multiple biochemical signals, and are involved in a wide variety of cellular processes such as proliferation, differentiation, transcription regulation and development.
TP53	ENSG00000141510.15	Tumor protein p53	Umor suppressor protein containing transcriptional activation, DNA binding, and oligomerization domains. The encoded protein responds to diverse cellular stresses to regulate expression of target genes, thereby inducing cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism.
CXCR4	ENSG00000121966.6	Chemokine (C-X-C motif) receptor 4	This gene encodes a CXC chemokine receptor specific for stromal cell-derived factor-1. The protein has 7 transmembrane regions and is located on the cell surface.
BID	ENSG00000015475.18	BH3 interacting domain death agonistkinase 2	This gene encodes a death agonist that heterodimerizes with either agonist BAX or antagonist BCL2.
FADD	ENSG00000168040.4	Fas (TNFRSF6)-associated via death domain	The protein encoded by this gene is an adaptor molecule that interacts with various cell surface receptors and mediates cell apoptotic signals.
EIF4EBP1	ENSG00000187840.4	Eukaryotic translation initiation factor 4E binding protein 1	This gene encodes one member of a family of translation repressor proteins. The protein directly interacts with eukaryotic translation initiation factor 4E (eIF4E), which is a limiting component of the multisubunit complex that recruits 40S ribosomal subunits to the 5' end of mRNAs.
BCL2	ENSG00000171791.11	B-cell CLL/lymphoma 2	This gene encodes an integral outer mitochondrial membrane protein that blocks the apoptotic death of some cells such as lymphocytes.
TNFSF10	ENSG00000121858.10	Tumor necrosis factor (ligand) superfamily, member 10	This protein preferentially induces apoptosis in transformed and tumor cells, but does not appear to kill normal cells although it is expressed at a significant level in most normal tissues.
NPY	ENSG00000122585.7	Neuropeptide Y	This gene encodes a neuropeptide that is widely expressed in the central nervous system and influences many physiological processes, including cortical excitability, stress response, food intake, circadian rhythms, and cardiovascular function.
TCS2	ENSG00000186184.15	Polymerase (RNA) I polypeptide D, 16 kDa	The protein encoded by this gene is a component of the RNA polymerase I and RNA polymerase III complexes, which function in the synthesis of ribosomal RNA precursors and small RNAs

(Continued)

Supplementary Table 7 (continued).

Gene	Ensembl ID	Description	Summary
CCL2	ENSG00000108691.9	Chemokine (C-C motif) ligand 2	This gene is one of several cytokine genes clustered on the q-arm of chromosome 17. Chemokines are a superfamily of secreted proteins involved in immunoregulatory and inflammatory processes.
RPTOR	ENSG00000141564.13	Regulatory associated protein of MTOR, complex 1	This gene encodes a component of a signaling pathway that regulates cell growth in response to nutrient and insulin levels.
PTGER3	ENSG00000050628.20	Prostaglandin E receptor 3 (subtype EP3)	The protein encoded by this gene is a member of the G-protein coupled receptor family. This protein is one of four receptors identified for prostaglandin E2 (PGE2).
CLEC5A	ENSG00000258227.6	C-type lectin domain family 5, member A	This gene encodes a member of the C-type lectin/C-type lectin-like domain (CTL/CTLD) superfamily.
PARP1	ENSG00000143799.12	poly (ADP-ribose) polymerase 1	This gene encodes a chromatin-associated enzyme, poly(ADP-ribosyl)transferase, which modifies various nuclear proteins by poly(ADP-ribosyl)ation.
ATP8B4	ENSG00000104043.14	ATPase, class I, type 8B, member 4	This gene encodes a member of the cation transport ATPase (P-type) family and type IV subfamily.
BAX	ENSG00000087088.19	BCL2-associated X protein	The protein encoded by this gene belongs to the BCL2 protein family.
STK11	ENSG00000118046.14	Serine/threonine kinase 11	This gene, which encodes a member of the serine/threonine kinase family, regulates cell polarity and functions as a tumor suppressor.