Original Article
Sequencing of high-frequency mutated genes in breast cancer (BRCA) and associated-functions analysis

Xuelian Li[1*], Mei Yang[1*], Liyuan Yang[1], Xuefei Dang[1], Xueqing Li[2], Gang Li[1]

[1]Department of Medical Oncology, Brunch of Minhang, Fudan University Shanghai Cancer Hospital, Shanghai, The People's Republic of China; [2]Department of Surgery, Fifth People's Hospital Affiliated to Fudan University, Shanghai, The People's Republic of China. *Equal contributors.

Abstract: Objective: Mutations or aberrant expression of genes in an organism tend not to be completely random and this cumulative effect predisposes to the development of malignant tumours. This study aims to reveal the possible aberrant expression of high frequency mutated genes, and then to investigate their role in development, prognosis, signalling pathway function and drug resistance in breast cancer. Methods: The mutated genes in breast cancer (BRCA) clinical samples were identified and detected by high-throughput sequencing. High-frequency mutant genes were counted. Gene expression profiles and the relationship with prognosis were analysed throughout TCGA database. qRT-PCR was used to analyse the mRNA levels of the six high-frequency mutant genes in BRCA tissues and cell lines. IHC was used to analyse the protein levels of the six high-frequency mutant genes in BRCA tissues. The linear interaction, single-cell layer clustering status and the influence in immune cell infiltration degree among these six high-frequency mutant genes were analysed by bioinformatics analysis. The STITCH and cMAP datasets were used for high-frequency mutant gene interaction networks, association signalling pathway enrichment and drug-transcriptome analyses. The effects of trastuzumab on the proliferative capacity of breast cancer cells, as well as on the expression of six high-frequency mutated genes were determined by CCK8 assay. Results: The genes that were statistically found to have high-frequency mutations in the samples recruited in the present study by high-throughput sequencing analysis included TP53, PIK3CA, NF1, TBX3, BRCA1 and BRCA2. The expression profiles of these genes and the correlation with prognosis were further demonstrated using the TCGA database: the trend in this study was similar to that of BRCA in TCGA. The mRNA and protein expression of these genes showed that the expression of TP53, NF1, TBX3, BRCA1 and BRCA2 was higher in tumor samples than that in normal samples, with an opposite trend for PIK3CA, a similar trend was observed in BRCA cell lines. The protein expressions of TP53, NF1, TBX3, BRCA1 and BRCA2 displayed the same trend by IHC. Other correlation results include 1) the single cell layer clustering of these six genes resulted in significant clustering with few overlapping regions; 2) these six genes showed different degrees of influence on BRCA immune cell infiltration; 3) these six genes showed a significant correlation between each other; 4) the network of each gene had partially overlapping molecules; and 5) the PI3K pathway was a key association pathway in BRCA. Finally, the cell proliferation ability results confirmed the optimal concentration of trastuzumab and its effect on mRNA expression of these six genes.

Keywords: Breast cancer, high-throughput sequencing, mutant genes, trastuzumab

Introduction

Breast carcinoma is one of the most frequently diagnosed cancer worldwide, seriously threatening women's life and health with millions of new cases annually [1-3]. The high incidence of BRCA may result from various physiological characteristic such as expression of hormonal factors, diabetes, or other endocrine issues [4-6]. In addition to physiological characteris-

tics, an individual's physical condition (such as the microenvironment) also has an impact on the occurrence of breast cancer [7, 8]. For example, prognosis of breast cancer was demonstrated to be affected by increased diabetes-related comorbidity or insulin disorders [9]. It is indisputable that progression of metastasis leading to a poor prognosis is one of the main reasons for the increase in mortality from breast cancer [10-14]. Based on the above con-

siderations, exploring the causes of disease, the mechanism of action and the mode of occurrence at the gene and molecule levels, as well as identifying and development new therapeutic targets are still urgent issues that we need to solve in the clinical diagnosis and treatment of breast cancer.

Over the past few years, gene sequencing has emerged as a new technology, analysing and determining the full sequence of genes from tumour tissue samples, as well as from blood or saliva samples, and these targets are able to target an individual's aberrantly expressed genes and predict the likelihood of developing a disease [15, 16]. Compared with traditional single-gene sequencing, NGS can simultaneously detect mutations, rearrangements, and amplifications of multiple target genes, which can effectively avoid sample waste and save testing time. NGS is an important part of contemporary precision medicine and an important link in the realization of precision diagnosis and treatment [17, 18].

Although there have been some reports of mutated genes in breast cancer, there are still many gaps in the causative factors affecting the occurrence and progression of breast cancer, and the functions of many of these high-frequency mutated genes have not been fully revealed, which necessitates and urgently requires further study of them. By analysing clinical sequencing data, it is possible to identify clinically relevant somatic mutations, noncoding region mutations, and mutation tags on both common and rare tumors; mutations are usually at the DNA level, but these genes that show mutations are often also abnormally expressed at the RNA level or protein level, which in turn causes tumour progression or metastasis [19-23]. In breast cancer, the genes that are prone to mutation include TP53, BRCA1, BRCA2, PIK3CA, HER-2, etc. [24-26]. TP53 is an oncogene, which can inhibit the growth and division of cancer cells, and it can also cause apoptosis of the cancer cells; numerous data shows that mutations in the TP53 gene can lead to breast cancer [27-29]. The oncogenes, BRCA1 and BRCA2, play roles in biological functions such as inhibiting cell proliferation, repairing DNA damage and participating in DNA replication, respectively. If BRCA1 or BRCA2 mutation occurs, it may also lead to breast cancer [30-33]. PIK3CA is a membrane-bound protein tyrosine kinase, which is involved in regulating biological functions such as cell growth, proliferation and migration, and plays an important role in breast cancer development [34]. In this study, high frequency mutated genes were analysed and screened from clinical blood samples by NGS sequencing, initially determining the RNA and protein expression of these genes. We investigated them to find out the signalling pathways as they may be affected by breast cancer cell lines, and further attempt to develop targeted drugs and related research.

The largest-scale genome analysis was performed to show the high association between breast cancer and abnormal regulation of most molecular pathways [35-37]. Wang et al. found significant changes in the immune response category GO:000695, cytokine pathways and primary immunodeficiency associated with metastasis in breast cancer bone metastases by layer clustering and gene enrichment analysis [38]. Also, a recent study by Chhichholiya et al. described that the ADAM8 gene, EN1 transcription factor, WNT and VEGF signalling pathways are associated with primary breast tumours; while MMP1, COX2, XCR4, PI3k/Akt, ERK and MAPK related pathways are involved in angiogenesis [39]. This suggests that the discovery and validation of new functions or mechanisms of genes and associated pathways could provide assistance in the development of new oncology therapeutic agents, and further improve the survival status of breast cancer patients and curb the risk of breast cancer progression and metastases.

In this study, we started from the high-frequency mutated genes obtained from clinical samples, and initially analyzed the expression of these high-frequency mutated genes at the tissue and cellular levels of breast cancer patients and their relationship with the signaling pathway-PI3K. Also, we tried to explore the possible targeting relationships in the complex interactions, which is expected to provide a possibility for the subsequent development of new drugs.

Materials and methods

Cell culture

Breast cancer cells MCF-7 (BNCC100137) and MDA-MB-231 (BNCC337894) were purchased

from BeNa Culture Collection (Shanghai, China). MCF-7 cells were cultured in DMEM-H complete medium (SIGMA, St. Louis, MO, USA), containing 90% DMEM-H and 10% FBS; while MDA-MB-231 cells were cultured in conditions of L-15 complete medium (SIGMA, St. Louis, MO, USA), containing 90% L-15 and 10% FBS. In this study, the normal cell line: human mammary cells MDA-kb2 (BNCC241999, BeNa Culture Collection) was the control group with L-15 medium (containing 90% L-15 and 10% FBS) being required for cell culture. The incubation temperature was 37°C and the carbon dioxide content of the air was 7.5%. Cells were routinely trypsinized, and plated at $4 \times 10^4/cm^2$ flasks.

Sample collection

Ten pairs cancer tissues were obtained from April 2021 to October 2022 at the Fudan University Shanghai Cancer Hospital in the present study and all samples were snap-frozen in liquid nitrogen immediately after collection and stored at -80°C. The study complied with the 1964 Declaration of Helsinki and was approved by the Ethics Committee. Informed consent for participation in the study was obtained from the patients (the ethical approval number: 2022 ethics investigation No. 002).

High-throughput sequencing analysis

The DNA extraction kit by magnetic bead method (IVD5435, Meiji, Guangzhou, China) was purchased from Guangzhou Meiji Biotechnology Co. Plasma was separated from 10 mL of peripheral blood and free DNA was extracted. At the same time, the corresponding sample purification and other pre-treatment were performed. Qsep100 capillary electrophoresis macromolecule analyzer was used for fragment quality control and quantification by Qubit® fluorescent dye. The Rapid plus DNA Library Kit (RK20255, Wuhan Abletech Biotechnology Co., Wuhan, China) was employed. Transformation of cDNA into the form and concentration required for libraries construction on Illumina® sequencing platform was accompanied by agarose gel nucleic acid electrophoresis. The Qubit® fluorescent dye method was employed for quantification and quality control. DNA fragments from the library were mixed at equimolar numbers and subjected to PE150 high-throughput sequencing using the S4 chip flow tank of the Illumina® Nova6000 platform. The

raw data obtained by sequencing was subjected to quality control, bioinformatics analysis, and copy number variant comparison analysis with the medication database to finally gather information including the percentage of circulating tumor DNA corresponding to the positive variant, copy number variant, and medication guidance.

Data analysis and interaction network analysis

Gene expression profiles of BRCA from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/) were downloaded and analyzed. This series contains breast cancer tissues samples (n=1,085) and normal breast tissues (n=291). The expression of six high-frequency mutated genes (TP53, PIK3CA, NF1, TBX3, BRCA1 and BRCA2) was measured and analysed in the database of breast cancer cell lines (n=62) datasets.

The tool database STITCH (http://stitch.embl.de/) was used to analyze and identify protein interaction networks of high frequency mutant genes. The Connectivity Map (CMap) analysis was conducted through the Web interface (http://www.broadinstitute.org/cmap) using version, which contains more than 7,000 expression profiles representing effects of 1,309 compounds on several cultured human cells. The CMap shows functional connections between drugs, genes and disease. We selected candidate drugs for validation *in vitro* on the basis of the connectivity score, correlation, and *P*-value. The PI3K signaling pathway was identified by Enrichr online (http://amp.pharm.mssm.edu/Enrichr/).

RNA extraction and reverse transcription

Total RNA was extracted from cell pellets using the RNA B™ extraction kit (Polyplus Transfection, Illkirch, France) according to the manufacturer's instructions. RNA was recovered in 50 µl nuclease-free water and either used immediately or stored at -80°C until further analysis. About 2 µg RNA was treated with DNase to remove genomic contamination and followed by reverse transcription. The synthesis of first strand cDNA used the First Strand cDNA Synthesis kit (Amersham Pharmacia Biotech, Uppsala, Sweden) following the manufacturer's instructions.

Table 1. Sequences of primers for qRT-PCR

Primer name	Sequence (5' to 3')
TP53 forward	CAGCCAAGTCTGTGACTTGCACGTAC
TP53 reverse	CTATGTCGAAAAGTGTTTCTGTCATC
PIK3CA forward	ACCCGCCCTATCTCAACTACC
PIK3CA reverse	AGGACACCATAATGACAGCCT
NF1 forward	TTCCCAGGGAAGGGCTCGTTTGG
NF1 reverse	CCAGAGAAATGAGTTTGTCTGGG
TBX3 forward	AAGAAGAGGTGGAGGACGAC
TBX3 reverse	TGAGGTTCGATGTCCCTACA
BRCA1 forward	ACAAATACTCATGCCAGCTCATT
BRCA1 reverse	GGCTCCTTGCTAAGCCAGG
BRCA2 forward	CCAAGTGGTCCACCCCAAC
BRCA2 reverse	GAAAATCAAGAAAAATCCTTAAAGGCT
GAPDH forward	TTACTCCTTGGAGGCCATGTGGGCC
GAPDH reverse	ACTGCCACCCAGAAGACTGTGGATGG

Quantitative real-time PCR (qRT-PCR)

The qRT-PCR tests were performed using SYBR Premix ExTaqTM II (Perfect Real Time, TAKARA, Japan) following the manufacturer's instructions. Reactions contained SYBR Green 1 qRT-PCR primer sets specific for human TP53, PIK3CA, NF1, TBX3, BRCA1 and BRCA2. The GAPDH gene served as an internal control. Breast cancer clinical tissue samples, breast cancer cell lines MCF-7 and MDA-MB-231 were performed by qRT-PCR, with at least three replicates of each sample in parallel experiments. Each gene expression level was measured using the $2^{-\Delta\Delta Ct}$ method. The results were presented as the fold change of each gene in the PCR sample relative to the parental PCR sample. Gene primer sequences were listed in **Table 1**.

CCK-8 assay

A Cell Counting Kit-8 (CCK-8; Dojindo Laboratories, Japan) was used according to the manufacturer's instructions to assess cell viability. MCF-7 and MDA-MB-231 cells treated with 10, 30 and 50 µM resveratrol were seeded at 3,000/well, and cultured in 96-well plates. The test was performed after exposure of 24 h and 48 h. The cell viabilities were measured from the absorbance (optical density) read with a microplate reader (Bio-Rad Instruments) at 450 nm and calculated using the following formula: Cell viability = (Treatment Group OD - Blank Group OD)/(Control Group OD - Blank Group OD).

Statistical analysis

All the statistical analyses were carried out by SPSS version 26.0 for Windows (SPSS Inc., Chicago, IL, USA). Results are presented as mean ± SD. Statistical significance between groups was determined as indicated in the text. The analyses were performed and the graphs were plotted in Prism V6.0 (GraphPad Software Inc., USA). A *P*-value of ≤0.05 was considered to be statistically significant.

Results

High-frequency mutant gene analysis based on high-throughput sequencing

In order to more accurately understand the possible new targets in breast cancer, the results of high-throughput sequencing analysis based on the collected clinical samples were statistically and visually analyzed in this study. **Figure 1A** shows the results of statistical analysis of mutated genes after NGS sequencing of 30 breast cancer clinical blood samples, indicating that TP53, PIK3CA and other genes were detected to be mutated in several breast cancer blood samples. **Figure 1B** shows that TP53, PIK3CA, NF1, TBX3, BRCA1 and BRCA2 had the highest mutation frequency, which may be useful for the development of new therapeutic methods and new targeted drugs for breast cancer.

It is important to note that the disease progresses of some patients has progressed rapidly and developed brain metastases. For these patients, NGS sequencing detected not only their blood samples, and their cerebrospinal fluid (CSF) samples in order to find the consistency and complementarity of mutated genes in blood and cerebrospinal fluid. Three of the 30 blood samples for which high-throughput sequencing was performed also underwent throughput sequencing of matched cerebrospinal fluid (CSF) samples. The results showed that in sample NO.2, sample NO.4, and sample NO.21, the number of mutated genes detected by CSF-high-throughput sequencing and blood-high-throughput sequencing were 13 and 5; 13 and 12; and 8 and 1, respectively (**Figure 1C**). In addition, considering the effect of high-frequency mutated genes on the prognosis of breast cancer, this study statistically analyzed the prognosis of the samples included in TCGA database, and the results are shown in **Figure**

Study of high-frequency mutated genes in breast cancer
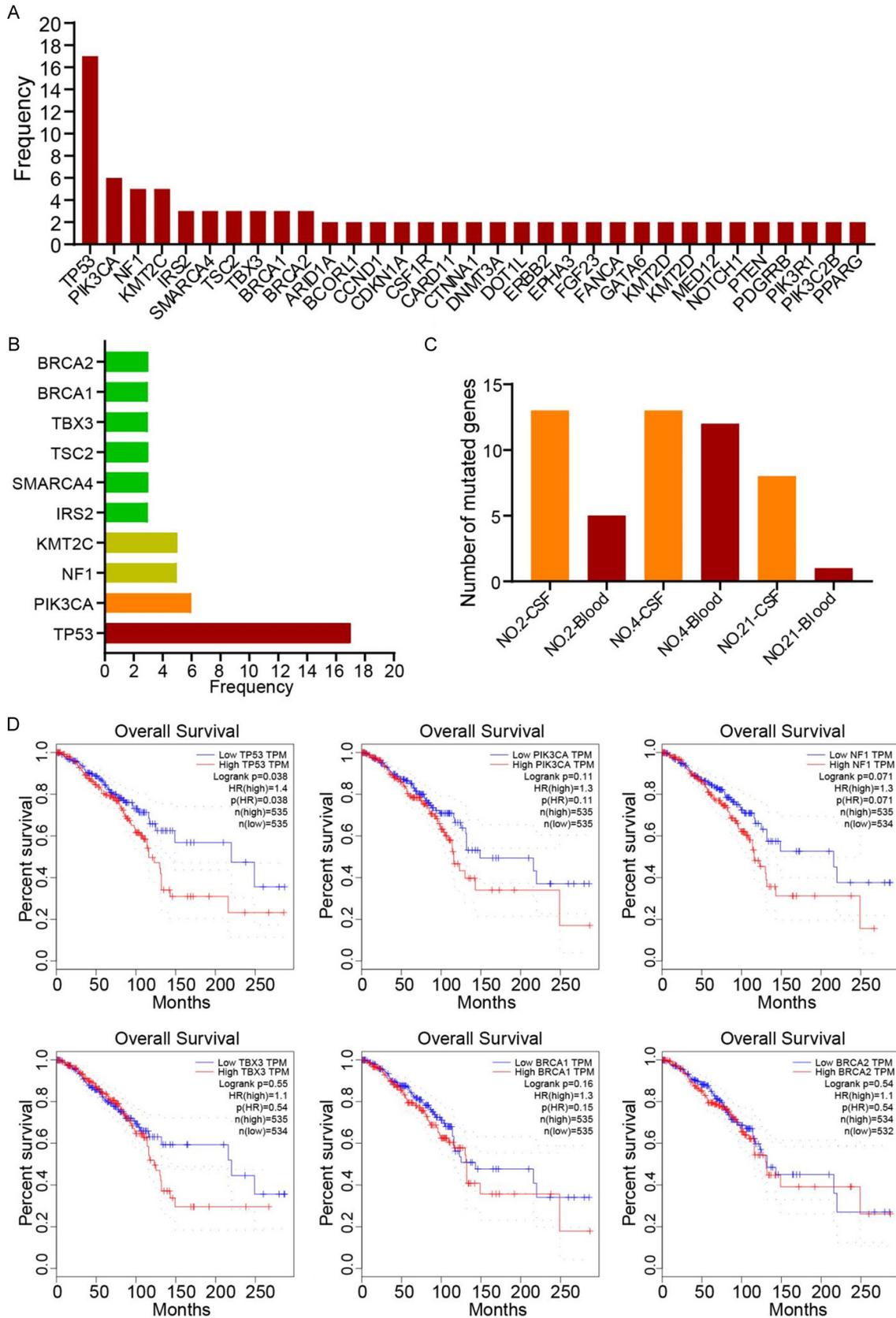
Figure 1. High-frequency mutant gene analysis based on high-throughput sequencing. A. The results of statistical analysis of mutated genes after NGS sequencing of 30 breast cancer samples; B. TP53, PIK3CA, NF1, TBX3, BRCA1

and BRCA2 the mutated genes with the highest mutation frequency; C. NO.2, NO.4 and NO.21 were detected by CSF and blood high-throughput sequencing; D. Correlation analysis of several high-frequency mutated genes in TCGA on breast cancer prognosis.

1D, which uncovered that their expression was closely related to the prognosis of breast cancer.

High-throughput result analysis of paired samples

Results from the NO.2, NO.4, and NO.21 patients, mentioned above who underwent both blood sample-high-throughput sequencing and cerebrospinal fluid sample-high-throughput sequencing were statistically analyzed, including the name of the mutated gene, exon region, the corresponding nucleic acid location and protein location, and the abundance of the mutation. This revealed that in patient NO.2, the mutated genes that were completely consistent between the blood and CSF samples were: AMER1, PIK3CA, SUZ12 and TP53; the percentage of genes that were identical was 30.8% as shown by Veen diagram (**Figure 2A**). For patient NO.4, the mutated genes that were completely consistent between the blood and CSF samples were: CCND1, CDH1, EPHB1, NTRK1, PIK3CA, PTEN, TP53 and VHL; the percentage of genes that were identical was 47.1% as shown by Veen diagram (**Figure 2B**). For patient NO.21, the mutated genes that were completely consistent between the blood and CSF samples were: KMT2D; the percentage of genes that were identical was 12.5% as shown by Veen diagram (**Figure 2C**).

Analysis of mRNA expression of high-frequency mutant genes

To inform the expression of the above mentioned high-frequency mutated genes in breast cancer, qRT-PCR was employed for the determination of mRNA expression. The expression of TP53, PIK3CA, BRCA1 and BRCA2 was determined, which are susceptible genes that have been reported to be mutated in breast cancer. The expression of two other high-frequency mutated genes, NF1 and TBX3, were also determined based on the analysis of the samples included in this study. TP53, PIK3CA, NF1, TBX3, BRCA1 and BRCA2 expression were measured in the breast cancer tissues and cell lines, respectively. The results, as shown in **Figure 3B** and **3C**, confirmed the up-regulated

expression of TP53, NF1, TBX3, BRCA1 and BRCA2 and the down-regulated expression of PIK3CA in breast cancer tissues and cell lines, respectively, compared with paracancerous tissues and normal breast cells. Breast cancer cell line (n=62) datasets were employed and analyzed, and each gene was ranked according to its level of expression from lowest to highest (Supplementary Figure 1).

The above results were consistent with the trend of TP53, PIK3CA, NF1, TBX3, BRCA1 and BRCA2 expression profiles in breast cancer samples collected on TCGA, the expression of TP53, PIK3CA, NF1, TBX3, BRCA1 and BRCA2 in the TCGA dataset is shown in **Figure 3A**. In addition, the protein expression of TP53, PIK3CA, NF1, BRCA1 and BRCA2 in breast cancer tissues were examined by IHC (**Figure 3D**); and the protein structures of TP53, PIK3CA, NF1, BRCA1 and BRCA2 were predicted, labelling their potential units of action (**Figure 3E**).

Correlation analysis among high-frequency mutated genes

In order to investigate whether there was a correlation between the high-frequency mutated genes counted in the enrolled breast cancer samples a linear correlation analysis was performed in this study between the TCGA database data and the data from the included samples (**Figure 4**). The results showed that TP53 was closely correlated with PIK3CA, NF1, BRCA1 and BRCA2 (**Figure 4A**); PIK3CA was closely correlated with TBX3 and BRCA1 (**Figure 4B**); NF1 was closely correlated with TBX3, BRCA1 and BRCA2 (**Figure 4C**); TBX3 was closely correlated with BRCA1 and BRCA2 (**Figure 4D**); and there was a significant linear relationship between BRCA1 and BRCA2 (**Figure 4E**).

In addition, the single-cell clustering results of breast cancer tissue samples corresponding to TP53, PIK3CA, NF1, TBX3, BRCA1 and BRCA2 were shown in Supplementary Figure 2, and that the cells can be effectively clustered by T-cells, Breast glandular cells, Macrophages, Breast myoepithelial cells, etc. The clustering was obvious and the main difference was
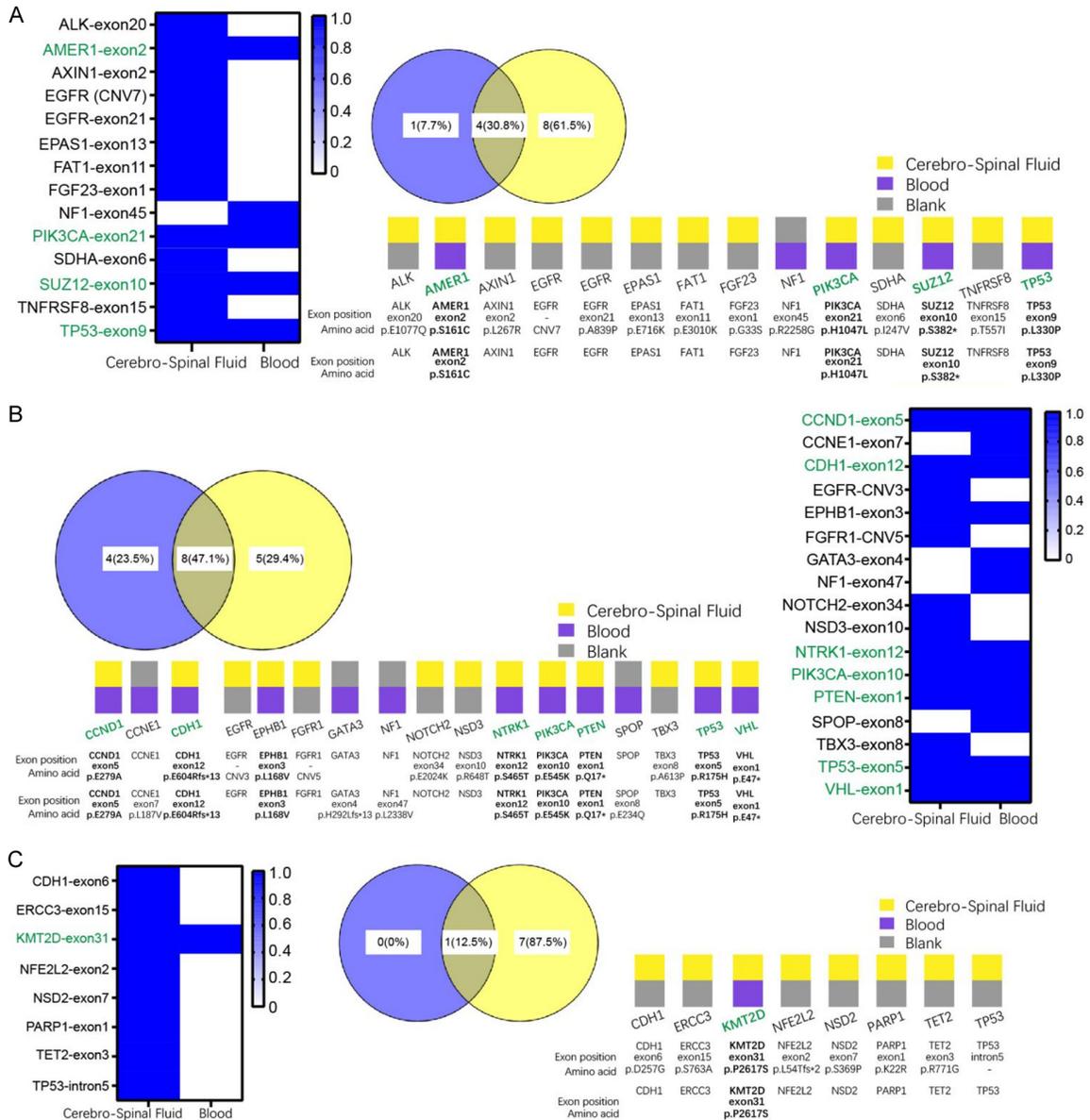
Figure 2. High-throughput results analysis of paired samples. NO.2 (A), NO.4 (B) and NO.21 (C). Statistics and analysis of high-throughput sequencing data of paired blood samples and high-throughput sequencing data of CSF samples, including the name of the mutated genes, the exonic regions of the mutated genes, the corresponding nucleic acid sites and protein sites, and the abundance of the mutations.

shown in the expression abundance of cells in each cluster. Meanwhile, TP53, PIK3CA, NF1, TBX3, BRCA1 and BRCA2 had different degrees of influence on BRCA immune cell infiltration, and the results showed that TP53 and PIK3CA showed significant differences on the degree of immune cell infiltration (Supplementary Figure 3). NF1 and TBX3 had an effect on the extent of breast cancer immune cell infiltration, but there was no significant difference; BRCA1 and BRCA2 had no effect on the extent of breast cancer immune cell infiltration.

Networks of association roles for high-frequency mutant genes

In addition to the correlation between two of the high-frequency genes, these high-frequency mutated genes were also associated with some essential signaling pathways. Therefore, STITCH tool was employed to perform the target protein/pathway interaction network analysis of the above several high-frequency mutated genes. The results are shown in **Figure 5**. The results showed that BRCA1 and TP53 were
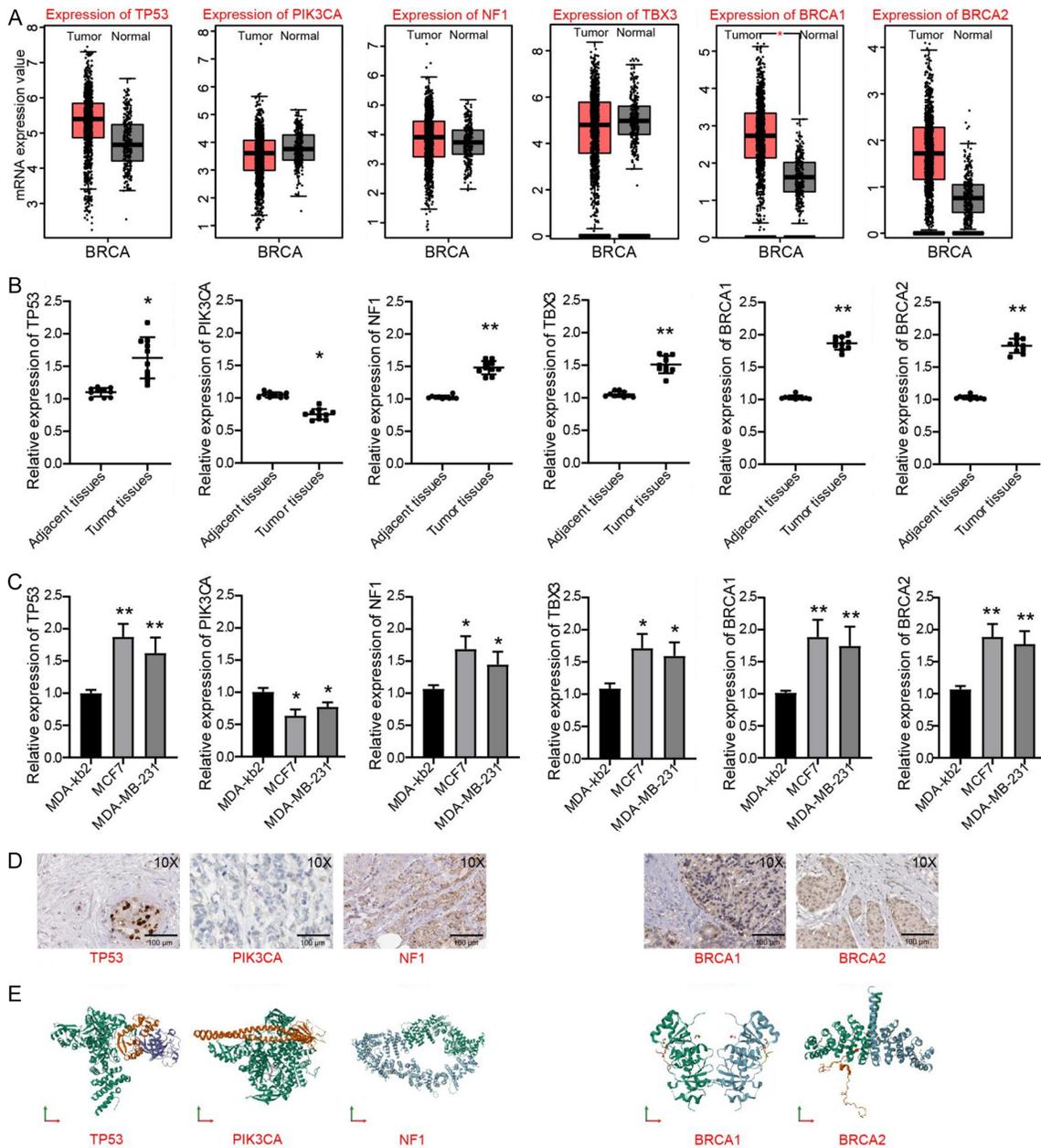
Figure 3. Analysis of mRNA expression of high-frequency mutant genes. The mRNA expression of TP53, PIK3CA, NF1, TBX3, BRCA1 and BRCA2 in the TCGA dataset (A), breast cancer tissue (B) and breast cancer cell line (C); the protein expression of TP53, PIK3CA, NF1, TBX3, BRCA1 and BRCA2 in breast cancer tissue (D); the protein structures of TP53, PIK3CA, NF1, TBX3, BRCA1 and BRCA2 (E).

associated with each other (**Figure 5A**, **5E**); BRCA1 was on the target protein/pathway interaction network of NF1 and BRCA2 (**Figure 5C**, **5F**). The target protein/pathway interaction network of TBX3 is shown in **Figure 5D**. There were many hotspots/key proteins on the target protein/pathway interaction network of PIK3CA (**Figure 5B**).

Pharmacological transcriptomics analysis

The differentially expressed genes analyzed from the database of breast cancer tissue samples and normal breast tissues were mapped to protein interaction networks for visual analysis and clustering. The largest clusters that were directly connected are plotted as shown
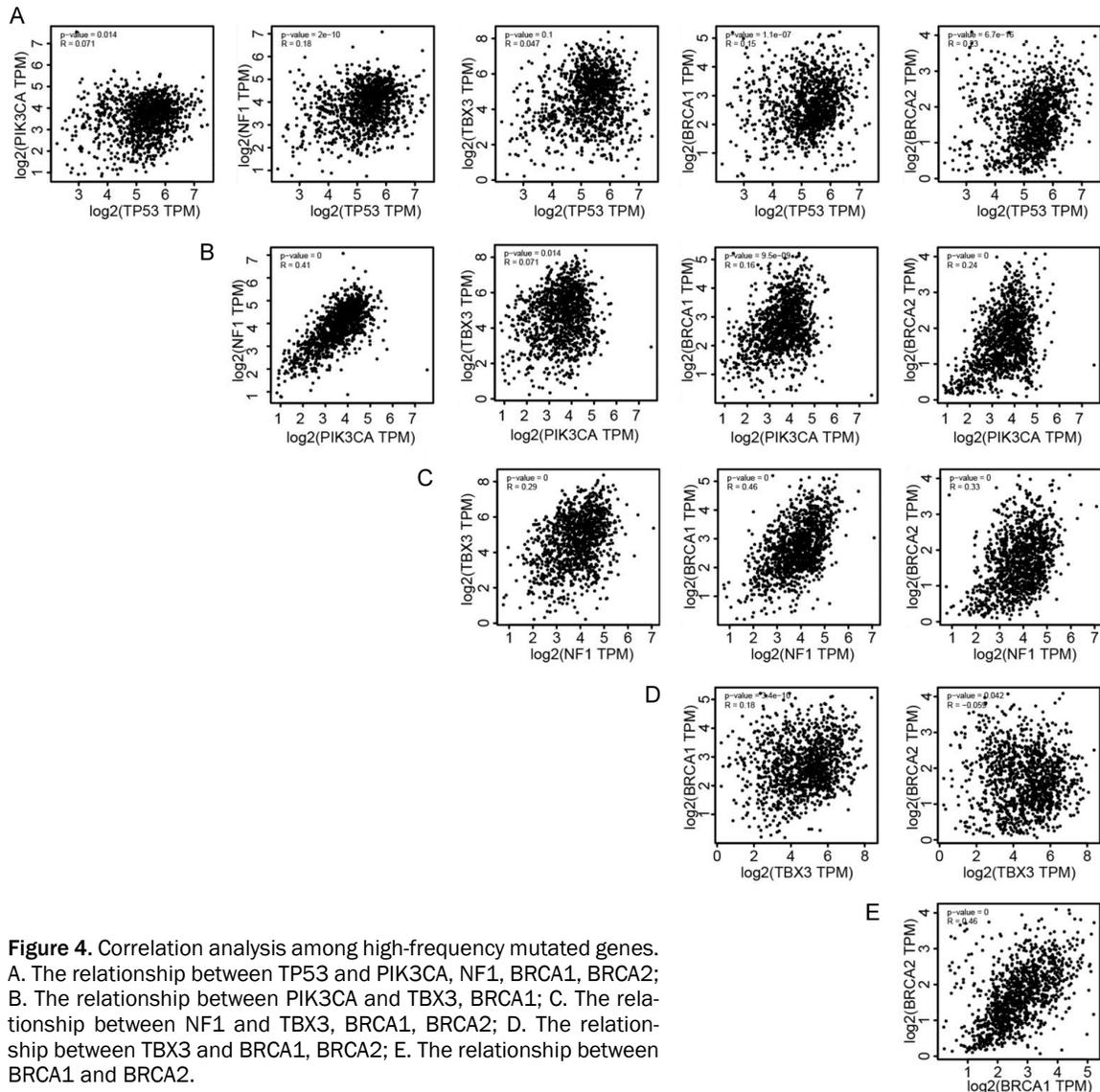
Figure 4. Correlation analysis among high-frequency mutated genes. A. The relationship between TP53 and PIK3CA, NF1, BRCA1, BRCA2; B. The relationship between PIK3CA and TBX3, BRCA1; C. The relationship between NF1 and TBX3, BRCA1, BRCA2; D. The relationship between TBX3 and BRCA1, BRCA2; E. The relationship between BRCA1 and BRCA2.

in **Figure 6A**, where each module was represented by a color. On this basis, the cMAP network drug transcriptome analysis was combined to map the signaling pathways with potential action relationships with trastuzumab, in which the PIK-ATK signaling pathway was among them (**Figure 6C**, **6D**). The flow graph of the cMAP network drug transcriptome analysis is shown in **Figure 6B**. In addition, the enrichment and specificity of cMAP network drug transcriptome analyses to PI3K inhibitors and others were shown in **Figure 6E**.

Cytotoxicity of trastuzumab on breast cancer cell

The cytotoxic effect of trastuzumab on breast cancer cells was examined through testing cell viability by CCK-8 assay. MCF-7 and MDA-MB-231 cells were cultured with different concentrations of trastuzumab (10, 30, and 50 μM) for 24 h and 48 h. The results shown in **Figure 7A** and **7B** revealed that the percentage of cytotoxicity induced by trastuzumab with a concentration of 50 μM was higher than 50% after 48 h, and the percentage of cytotoxicity induced by trastuzumab within the range of 0-30 μM is lower than 50%. In addition, the mRNA levels of the six studied high-frequency mutated genes in breast cancer cells at different trastuzumab concentrations were detected by qRT-PCR. The results showed that trastuzumab increased the mRNA expression of PIK3CA in MCF-7 and MDA-MB-231 cells compared with the control group; while trastuzum-
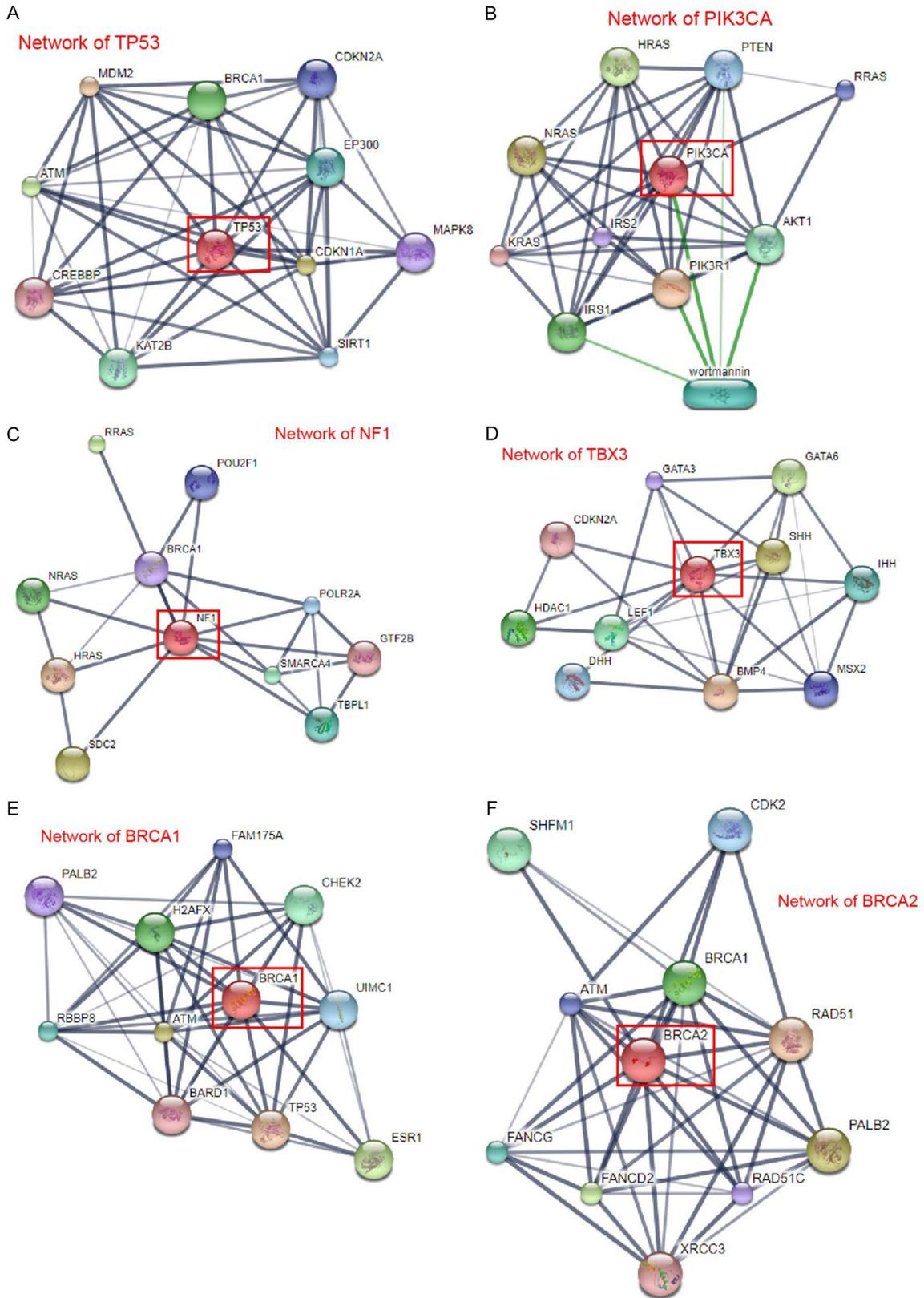
Figure 5. Networks of association roles for high-frequency mutant genes. A. The network of TP53; B. The network of PIK3CA; C. The network of NF1; D. The network of TBX3; E. The network of BRCA1; F. The network of BRCA2.
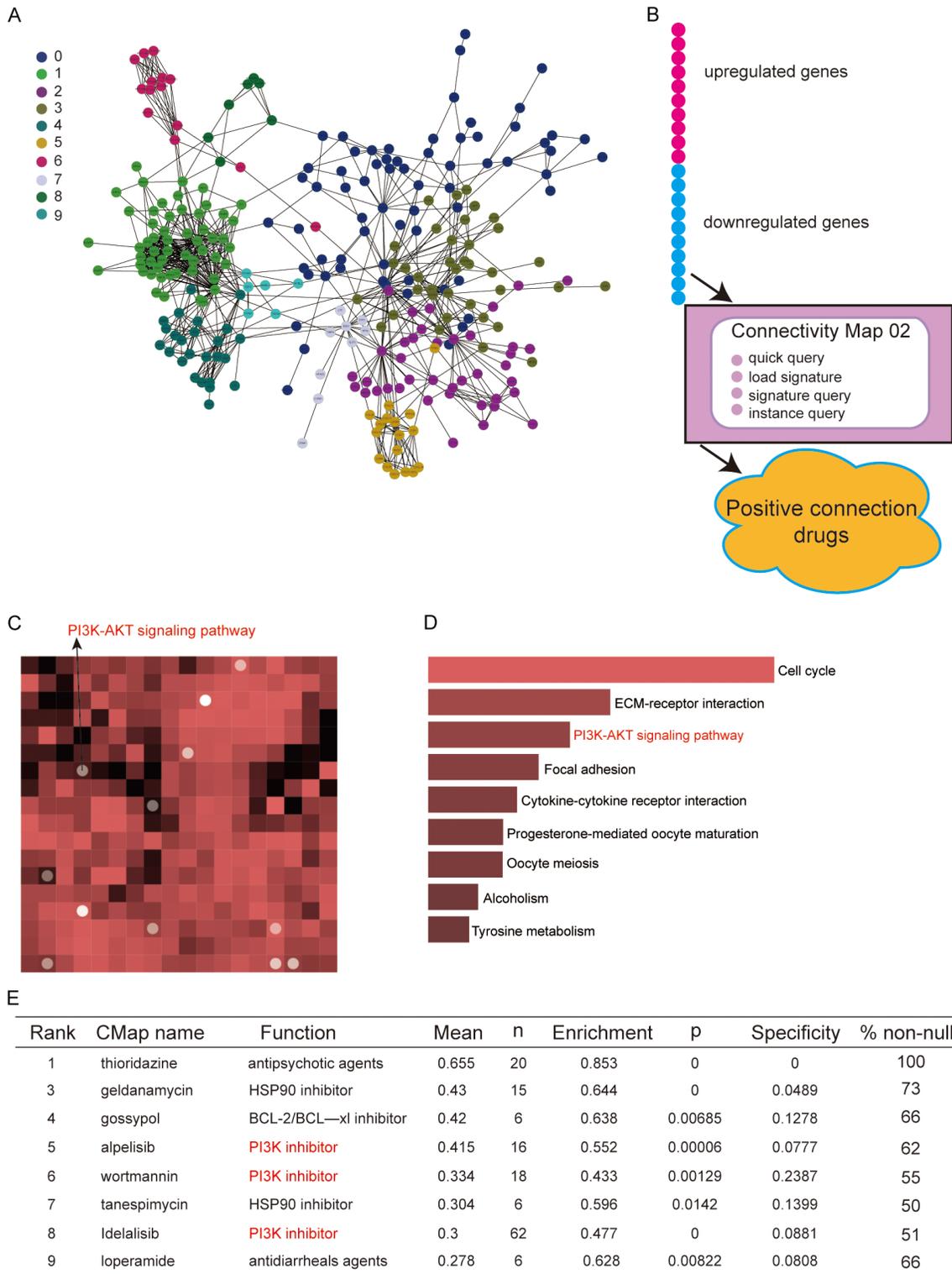
Figure 6. Networks of association roles for high-frequency mutant genes. A. The largest clusters that are directly connected in breast cancer; B. The flow chart of the cMAP network drug transcriptome analysis; C. Signaling pathways with potential action in relation to trastuzumab via cMAP network drug transcriptome analysis; D. The PIK-ATK signaling pathway was visualized by heatmap; E. The enrichment and specificity of cMAP network drug transcriptome analyses.

The table in panel E:

Rank	CMap name	Function	Mean	n	Enrichment	p	Specificity	% non-null
1	thioridazine	antipsychotic agents	0.655	20	0.853	0	0	100
3	geldanamycin	HSP90 inhibitor	0.43	15	0.644	0	0.0489	73
4	gossypol	BCL-2/BCL—xl inhibitor	0.42	6	0.638	0.00685	0.1278	66
5	alpelisib	PI3K inhibitor	0.415	16	0.552	0.00006	0.0777	62
6	wortmannin	PI3K inhibitor	0.334	18	0.433	0.00129	0.2387	55
7	tanespimycin	HSP90 inhibitor	0.304	6	0.596	0.0142	0.1399	50
8	Idelalisib	PI3K inhibitor	0.3	62	0.477	0	0.0881	51
9	loperamide	antidiarrheals agents	0.278	6	0.628	0.00822	0.0808	66

Figure 7. Cytotoxicity of trastuzumab on breast cancer cell. (A, B) The results of cytotoxicity on MCF-7 (A) and MDA-MB-231 (B) cells with different trastuzumab concentration; (C, D) The mRNA levels of the above six high-frequency mutated genes in MCF-7 (C) and MDA-MB-231 (D) cells at different trastuzumab concentrations were detected by qRT-PCR.

ab decreased the mRNA expression of TP53, NF1, TBX3, BRCA1 and BRCA2 in breast cancer cells (*P*<0.05, **Figure 7C**, **7D**).

Discussion

In this study, genes with high-frequency mutations in breast cancer were statistically analyzed by the results of high-throughput sequencing analysis of the enrolled samples. It was confirmed that the consistency of the mutated genes in the paired samples between blood sample and CSF sample was high. Analysis of expression, correlation, target proteins and signaling pathways of TP53, PIK3CA, NF1, TBX3, BRCA1, and BRCA2 in breast cancer, have multiple highly correlated axes of action that affect signaling pathways such as PIK-AKT. The work flow was displayed in **Figure 8**. They are expected to be important biomarkers for breast cancer.

Breast cancer accounts for nearly a quarter of all cancer diagnoses and is the leading cause of cancer-related deaths in women worldwide [40, 41]. Currently, research relying on high-throughput sequencing has become a hotspot of scientific study in the past decades, and related studies have contributed to the improvement of breast cancer treatment [42-45]. By analyzing the results of high-throughput sequencing and screening high-frequency mutant genes, it is possible to carry out the development of potential target-genes, new drug exploitation and other clinical applications in breast cancer [46-48]. Meanwhile, the screened high-frequency mutant genes can also be used to do experimental studies and analysis of expression and functional mechanisms, etc. [49-52]. We have analysed as comprehensively as possible the similarities and differences between mutated genes originating from the blood and
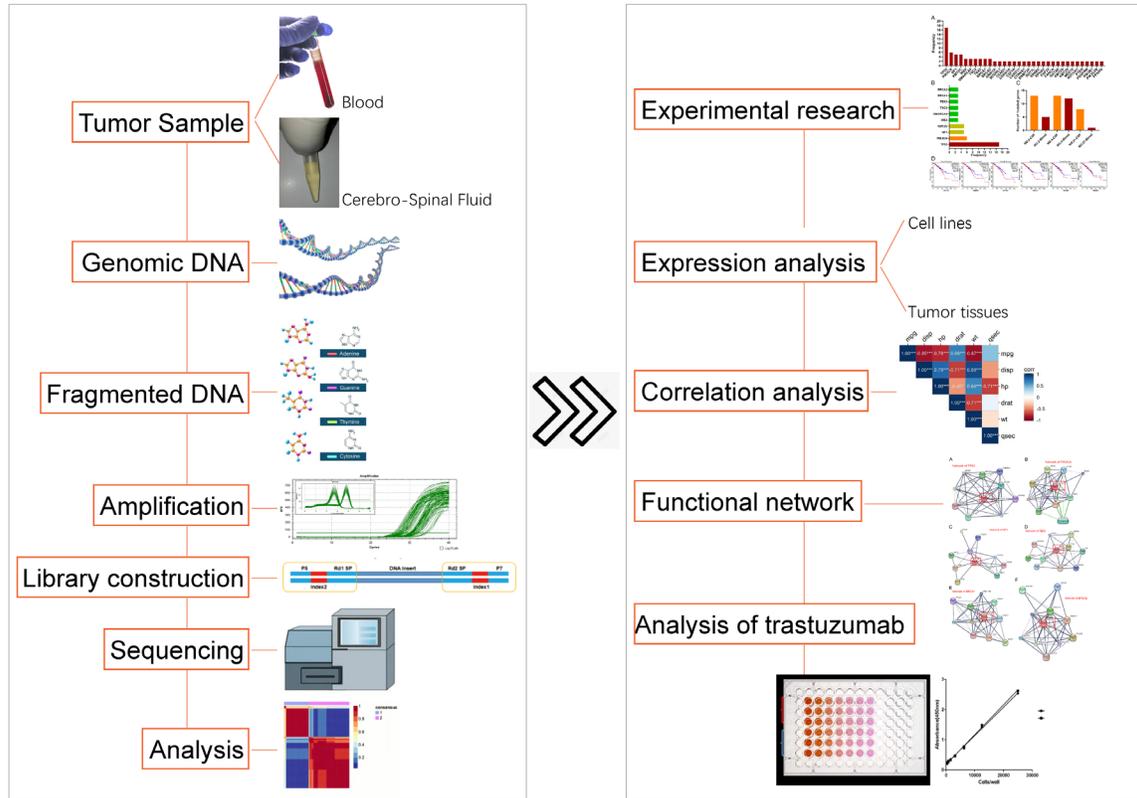
Figure 8. The technical route and analytical approach.

those originating from cerebrospinal fluid samples of brain metastases from breast cancer patients in an attempt to discover new high-frequency mutated genes and their functions.

Previous studies have shown that trastuzumab has inhibitory effects on tumors such as gastric and breast cancer [53-55]. It has been shown that they exert anticancer effects on tumor growth and metastasis of triple-negative breast cancer cells by inhibiting the expression of TGF-β1 [56]. Several studies have shown that trastuzumab, in combination with other enzymes and small molecules, has a synergistic effect on inhibiting cell death, migration, invasion, tumor growth and metastasis [57-60]. Similarly, in our study, the effect of trastuzumab on the proliferative abilities of breast cancer cells were investigated and several high-frequency mutant gene expression assays were performed by setting concentration gradients. The results showed that the effect of trastuzumab on breast cancer cells was concentration-dependent, which has important implications for further clinical studies.

In the present study, in addition to relatively well-known genes like TP53 and BRCA1/2, we also found some genes that are relatively underreported and this leaves a lot of research gaps, such as NF1 and TBX3. This finding was made possible by the samples that were enrolled in this study, which included samples of breast cancer with brain metastases in addition to samples with breast cancer. Because this study is ongoing and continuous, the samples collected were dynamically monitored and continuously followed. Like NF1 and TBX3, other targets need to be progressively recognized and researched in order to confirm their key role in oncology research [61, 62]. Their biological functions in breast cancer need to be urgently explored. In addition, the high-frequency mutated genes originating from clinical samples are the main target for our research, and our team is also still constructing animal models in the hope of mimicking the *in vivo* environments to probe the functional gene regulation patterns from our included samples. This study does have a few shortcomings; we did not include a large number of samples. In addition,

the *in vivo* and *in vitro* experiments conducted are not enough, which is limited by the project funding. In future work, we will continue to conduct in-depth research on these genes, focusing on exploring the biological functions of these genes, and providing data to support the clinical search for new drug targets. All in all, our study reveals high-frequency mutated genes in breast cancer obtained by high-throughput sequencing, and preliminarily explores their effects on breast cancer development and related signaling pathways, which is of prospective research value.

Acknowledgements

Disclosure of conflict of interest

None.

Address correspondence to: Gang Li, MBBS Minhang Branch, Fudan University Shanghai Cancer Center, No. 106 Ruili Road, Minhang District, Shanghai, The People's Republic of China. E-mail: ligang_0113@163.com

References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A and Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021; 71: 209-249.

[2] Giaquinto AN, Sung H, Miller KD, Kramer JL, Newman LA, Minihan A, Jemal A and Siegel RL. Breast cancer statistics, 2022. CA Cancer J Clin 2022; 72: 524-541.

[3] Li Q, Xia C, Li H, Yan X, Yang F, Cao M, Zhang S, Teng Y, He S, Cao M and Chen W. Disparities in 36 cancers across 185 countries: secondary analysis of global cancer statistics. Front Med 2024; 18: 911-920.

[4] Arnold M, Morgan E, Rumgay H, Mafra A, Singh D, Laversanne M, Vignat J, Gralow JR, Cardoso F, Siesling S and Soerjomataram I. Current and future burden of breast cancer: global statistics for 2020 and 2040. Breast 2022; 66: 15-23.

[5] Wang H, Zhang N, Sun Q, Zhao Z, Pang H, Huang X, Zhang R, Kang W and Shan M. Comparison of the efficacy of taxanes with carboplatin and anthracyclines with taxanes in neoadjuvant chemotherapy for stage II-III triple negative breast cancer: a retrospective analysis. J Cancer Res Clin Oncol 2024; 150: 291.

[6] Sun K, Zhang B, Lei S, Zheng R, Liang X, Li L, Feng X, Zhang S, Zeng H, Yao Y, Ma P, Wang S, Chen R, Han B, Wei W and He J. Incidence, mortality, and disability-adjusted life years of female breast cancer in China, 2022. Chin Med J (Engl) 2024; 137: 2429-2436.

[7] Youn HJ and Han W. A review of the epidemiology of breast cancer in Asia: focus on risk factors. Asian Pac J Cancer Prev 2020; 21: 867-880.

[8] Pedersen RN, Esen BÖ, Mellemkjær L, Christiansen P, Ejlertsen B, Lash TL, Nørgaard M and Cronin-Fenton D. The incidence of breast cancer recurrence 10-32 years after primary diagnosis. J Natl Cancer Inst 2022; 114: 391-399.

[9] Hosio M, Urpilainen E, Hautakoski A, Marttila M, Arffman M, Sund R, Ahtikoski A, Puistola U, Läärä E, Karihtala P and Jukkola A. Association of antidiabetic medication and statins with survival from ductal and lobular breast carcinoma in women with type 2 diabetes. Sci Rep 2021; 11: 10445.

[10] Ibragimova MK, Tsyganov MM, Kravtsova EA, Tsydenova IA and Litviakov NV. Organ-specificity of breast cancer metastasis. Int J Mol Sci 2023; 24: 15625.

[11] Kim MY. Breast cancer metastasis. Adv Exp Med Biol 2021; 1187: 183-204.

[12] Nolan E, Kang Y and Malanchi I. Mechanisms of organ-specific metastasis of breast cancer. Cold Spring Harb Perspect Med 2023; 13: a041326.

[13] Ivanova M, Porta FM, Giugliano F, Frascarelli C, Sajjadi E, Venetis K, Cursano G, Mazzarol G, Guerini-Rocco E, Curigliano G, Criscitiello C and Fusco N. Breast cancer with brain metastasis: molecular insights and clinical management. Genes (Basel) 2023; 14: 1160.

[14] McGinnis CS, Miao Z, Superville D, Yao W, Goga A, Reticker-Flynn NE, Winkler J and Satpathy AT. The temporal progression of lung immune remodeling during breast cancer metastasis. Cancer Cell 2024; 42: 1018-1031, e6.

[15] Chang L, Jiao H, Chen J, Wu G, Liu P, Li R, Guo J, Long W, Tang X, Lu B, Xu H and Wu H. Single-cell whole-genome sequencing, haplotype analysis in prenatal diagnosis of monogenic diseases. Life Sci Alliance 2023; 6: e202201761.

[16] Liu SQ, Gao ZJ, Wu J, Zheng HM, Li B, Sun S, Meng XY and Wu Q. Single-cell and spatially resolved analysis uncovers cell heterogeneity

of breast cancer. J Hematol Oncol 2022; 15: 19.

[17] Liu YY, Cheng K, Just R, Enke S and Bright JA. Sequencing-induced artefacts in NGS STR data. Forensic Sci Int Genet 2024; 72: 103086.

[18] Andre F, Filleron T, Kamal M, Mosele F, Arnedos M, Dalenc F, Sablin MP, Campone M, Bonnefoi H, Lefeuvre-Plesse C, Jacot W, Coussy F, Ferrero JM, Emile G, Mouret-Reynier MA, Thery JC, Isambert N, Mege A, Barthelemy P, You B, Hajjaji N, Lacroix L, Rouleau E, Tran-Dien A, Boyault S, Attignon V, Gestraud P, Servant N, Le Tourneau C, Cherif LL, Soubeyran I, Montemurro F, Morel A, Lusque A, Jimenez M, Jacquet A, Gonçalves A, Bachelot T and Bieche I. Genomics to select treatment for patients with metastatic breast cancer. Nature 2022; 610: 343-348.

[19] Rogozin IB, Pavlov YI, Goncearenco A, De S, Lada AG, Poliakov E, Panchenko AR and Cooper DN. Mutational signatures and mutable motifs in cancer genomes. Brief Bioinform 2018; 19: 1085-1101.

[20] Roos A and Byron SA. Genomics-enabled precision medicine for cancer. Cancer Treat Res 2019; 178: 137-169.

[21] Breast Cancer Association Consortium; Dorling L, Carvalho S, Allen J, González-Neira A, Luccarini C, Wahlström C, Pooley KA, Parsons MT, Fortuno C, Wang Q, Bolla MK, Dennis J, Keeman R, Alonso MR, Álvarez N, Herraez B, Fernandez V, Núñez-Torres R, Osorio A, Valcich J, Li M, Törngren T, Harrington PA, Baynes C, Conroy DM, Decker B, Fachal L, Mavaddat N, Ahearn T, Aittomäki K, Antonenkova NN, Arnold N, Arveux P, Ausems MGEM, Auvinen P, Becher H, Beckmann MW, Behrens S, Bermisheva M, Białkowska K, Blomqvist C, Bogdanova NV, Bogdanova-Markov N, Bojesen SE, Bonanni B, Børresen-Dale AL, Brauch H, Bremer M, Briceno I, Brüning T, Burwinkel B, Cameron DA, Camp NJ, Campbell A, Carracedo A, Castelao JE, Cessna MH, Chanock SJ, Christiansen H, Collée JM, Cordina-Duverger E, Cornelissen S, Czene K, Dörk T, Ekici AB, Engel C, Eriksson M, Fasching PA, Figueroa J, Flyger H, Försti A, Gabrielson M, Gago-Dominguez M, Georgoulias V, Gil F, Giles GG, Glendon G, Garcia EBG, Alnæs GIG, Guénel P, Hadjisavvas A, Haeberle L, Hahnen E, Hall P, Hamann U, Harkness EF, Hartikainen JM, Hartman M, He W, Heemskerk-Gerritsen BAM, Hillemanns P, Hogervorst FBL, Hollestelle A, Ho WK, Hooning MJ, Howell A, Humphreys K, Idris F, Jakubowska A, Jung A, Kapoor PM, Kerin MJ, Khusnutdinova E, Kim SW, Ko YD, Kosma VM, Kristensen VN, Kyriacou K, Lakeman IMM, Lee JW, Lee MH, Li J, Lindblom A, Lo WY, Loizidou MA, Lophat-

ananon A, Lubiński J, MacInnis RJ, Madsen MJ, Mannermaa A, Manoochehri M, Manoukian S, Margolin S, Martinez ME, Maurer T, Mavroudis D, McLean C, Meindl A, Mensenkamp AR, Michailidou K, Miller N, Mohd Taib NA, Muir K, Mulligan AM, Nevanlinna H, Newman WG, Nordestgaard BG, Ng PS, Oosterwijk JC, Park SK, Park-Simon TW, Perez JIA, Peterlongo P, Porteous DJ, Prajzendanc K, Prokofyeva D, Radice P, Rashid MU, Rhenius V, Rookus MA, Rüdiger T, Saloustros E, Sawyer EJ, Schmutzler RK, Schneeweiss A, Schürmann P, Shah M, Sohn C, Southey MC, Surowy H, Suvanto M, Thanasitthichai S, Tomlinson I, Torres D, Truong T, Tzardi M, Valova Y, van Asperen CJ, Van Dam RM, van den Ouweland AMW, van der Kolk LE, van Veen EM, Wendt C, Williams JA, Yang XR, Yoon SY, Zamora MP, Evans DG, de la Hoya M, Simard J, Antoniou AC, Borg Å, Andrulis IL, Chang-Claude J, García-Closas M, Chenevix-Trench G, Milne RL, Pharoah PDP, Schmidt MK, Spurdle AB, Vreeswijk MPG, Benitez J, Dunning AM, Kvist A, Teo SH, Devilee P and Easton DF. Breast cancer risk genes - association analysis in more than 113,000 women. N Engl J Med 2021; 384: 428-439.

[22] Martínez-Jiménez F, Movasati A, Brunner SR, Nguyen L, Priestley P, Cuppen E and Van Hoeck A. Pan-cancer whole-genome comparison of primary and metastatic solid tumours. Nature 2023; 618: 333-341.

[23] Dietlein F, Wang AB, Fagre C, Tang A, Besselink NJM, Cuppen E, Li C, Sunyaev SR, Neal JT and Van Allen EM. Genome-wide analysis of somatic noncoding mutation patterns in cancer. Science 2022; 376: eabg5601.

[24] Brown NA and Elenitoba-Johnson KSJ. Enabling precision oncology through precision diagnostics. Annu Rev Pathol 2020; 15: 97-121.

[25] Liu PF, Zhuo ZL, Xie F, Xian HP, Liu C, Wang S and Zhao XT. Multiple mutation detection for risk assessment in patients with breast cancer by using next-generation sequencing. Ann Clin Lab Sci 2021; 51: 670-677.

[26] Ma W, Wu H, Chen Y, Xu H, Jiang J, Du B, Wan M, Ma X, Chen X, Lin L, Su X, Bao X, Shen Y, Xu N, Ruan J, Jiang H and Ding Y. New techniques to identify the tissue of origin for cancer of unknown primary in the era of precision medicine: progress and challenges. Brief Bioinform 2024; 25: bbae028.

[27] Shahbandi A, Nguyen HD and Jackson JG. TP53 mutations and outcomes in breast cancer: reading beyond the headlines. Trends Cancer 2020; 6: 98-110.

[28] Ragu ME, Lim JMC, Ng PS, Yip CH, Rajadurai P, Teo SH and Pan JW. TP53 somatic mutations in Asian breast cancer are associated with

subtype-specific effects. Breast Cancer Res 2023; 25: 48.

[29] Lin CJ, Jin X, Ma D, Chen C, Ou-Yang Y, Pei YC, Zhou CZ, Qu FL, Wang YJ, Liu CL, Fan L, Hu X, Shao ZM and Jiang YZ. Genetic interactions reveal distinct biological and therapeutic implications in breast cancer. Cancer Cell 2024; 42: 701-719, e12.

[30] Narod SA and Salmena L. BRCA1 and BRCA2 mutations and breast cancer. Discov Med 2011; 12: 445-53.

[31] Liede A, Mansfield CA, Metcalfe KA, Price MA; Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer, Snyder C, Lynch HT, Friedman S, Amelio J, Posner J, Narod SA, Lindeman GJ and Evans DG. Preferences for breast cancer risk reduction among BRCA1/BRCA2 mutation carriers: a discrete-choice experiment. Breast Cancer Res Treat 2017; 165: 433-444.

[32] Martelli G, Barretta F, Vernieri C, Folli S, Pruneri G, Segattini S, Trapani A, Carolla C, Spatti G, Miceli R and Ferraris C. Prophylactic salpingo-oophorectomy and survival after BRCA1/2 breast cancer resection. JAMA Surg 2023; 158: 1275-1284.

[33] Pourmasoumi P, Moradi A and Bayat M. BRCA1/2 mutations and breast/ovarian cancer risk: a new insights review. Reprod Sci 2024; 31: 3624-3634.

[34] Martínez-Sáez O, Chic N, Pascual T, Adamo B, Vidal M, González-Farré B, Sanfeliu E, Schettini F, Conte B, Brasó-Maristany F, Rodríguez A, Martínez D, Galván P, Rodríguez AB, Martinez A, Muñoz M and Prat A. Frequency and spectrum of PIK3CA somatic mutations in breast cancer. Breast Cancer Res 2020; 22: 45.

[35] Jiang L, Yu H, Wang C, He F, Shi Z, Tu H, Ning N, Duan S and Zhao Y. The anti-cancer effects of mitochondrial-targeted triphenylphosphonium-resveratrol conjugate on breast cancer cells. Pharmaceuticals (Basel) 2022; 15: 1271.

[36] Gupta R, Ponangi R and Indresh KG. Role of glycosylation in breast cancer progression and metastasis: implications for miRNA, EMT and multidrug resistance. Glycobiology 2023; 33: 545-555.

[37] Itah Z, Chaudhry S, Raju Ponny S, Aydemir O, Lee A, Cavanagh-Kyros J, Tournier C, Muller WJ and Davis RJ. HER2-driven breast cancer suppression by the JNK signaling pathway. Proc Natl Acad Sci U S A 2023; 120: e2218373120.

[38] Wang P, Sun Z, Zhang Z and Yin Q. Immune response pathways enriched in breast cancer samples with brain metastasis. Gland Surg 2021; 10: 3334-3341.

[39] Chhichholiya Y, Ruthuparna M, Velagaleti H and Munshi A. Brain metastasis in breast can-

cer: focus on genes and signaling pathways involved, blood-brain barrier and treatment strategies. Clin Transl Oncol 2023; 25: 1218-1241.

[40] Wilkinson L and Gathani T. Understanding breast cancer as a global health concern. Br J Radiol 2022; 95: 20211033.

[41] Benitez Fuentes JD, Morgan E, de Luna Aguilar A, Mafra A, Shah R, Giusti F, Vignat J, Znaor A, Musetti C, Yip CH, Van Eycken L, Jedy-Agba E, Piñeros M and Soerjomataram I. Global stage distribution of breast cancer at diagnosis: a systematic review and meta-analysis. JAMA Oncol 2024; 10: 71-78.

[42] Mu H, Zhang W, Qiu Y, Tao T, Wu H, Chen Z and Xu G. miRNAs as potential markers for breast cancer and regulators of tumorigenesis and progression (Review). Int J Oncol 2021; 58: 16.

[43] Wu L, Yao H, Chen H, Wang A, Guo K, Gou W, Yu Y, Li X, Yao M, Yuan S, Pang F, Hu J, Chen L, Liu W, Yao J, Zhang S, Dong X, Wang W, Hu J, Ling Q, Ding S, Wei Y, Li Q, Cao W, Wang S, Di Y, Feng F, Zhao G, Zhang J, Huang L, Xu J, Yan W, Tong Z, Jiang D, Ji T, Li Q, Xu L, He H, Shang L, Liu J, Wang K, Wu D, Shen J, Liu Y, Zhang T, Liang C, Wang Y, Shang Y, Guo J, Liang G, Xu S, Liu J, Wang K and Wang M. Landscape of somatic alterations in large-scale solid tumors from an Asian population. Nat Commun 2022; 13: 4264.

[44] Neagu AN, Whitham D, Bruno P, Morrissiey H, Darie CA and Darie CC. Omics-based investigations of breast cancer. Molecules 2023; 28: 4768.

[45] Tierno D, Grassi G, Scomersi S, Bortul M, Generali D, Zanconati F and Scaggiante B. Next-generation sequencing and triple-negative breast cancer: insights and applications. Int J Mol Sci 2023; 24: 9688.

[46] Jia Q, Chu H, Jin Z, Long H and Zhu B. High-throughput single-cell sequencing in cancer research. Signal Transduct Target Ther 2022; 7: 145.

[47] Agathangelidis A, Vlachonikola E, Davi F, Langerak AW and Chatzidimitriou A. High-throughput immunogenetics for precision medicine in cancer. Semin Cancer Biol 2022; 84: 80-88.

[48] Jacobs AT, Martinez Castaneda-Cruz D, Rose MM and Connelly L. Targeted therapy for breast cancer: an overview of drug classes and outcomes. Biochem Pharmacol 2022; 204: 115209.

[49] Veiga DFT, Nesta A, Zhao Y, Deslattes Mays A, Huynh R, Rossi R, Wu TC, Palucka K, Anczukow O, Beck CR and Banchereau J. A comprehensive long-read isoform analysis platform and sequencing resource for breast cancer. Sci Adv 2022; 8: eabg6711.

[50] Hou YC, Neidich JA, Duncavage EJ, Spencer DH and Schroeder MC. Clinical whole-genome sequencing in cancer diagnosis. Hum Mutat 2022; 43: 1519-1530.

[51] Clusan L, Ferrière F, Flouriot G and Pakdel F. A basic review on estrogen receptor signaling pathways in breast cancer. Int J Mol Sci 2023; 24: 6834.

[52] Qian J, Zhao L, Xu L, Zhao J, Tang Y, Yu M, Lin J, Ding L and Cui Q. Cell death: mechanisms and potential targets in breast cancer therapy. Int J Mol Sci 2024; 25: 9703.

[53] Chen YZ, Xue JY, Chen CM, Yang BL, Xu QH, Wu F, Liu F, Ye X, Meng X, Liu GY, Shen ZZ, Shao ZM and Wu J. PPAR signaling pathway may be an important predictor of breast cancer response to neoadjuvant chemotherapy. Cancer Chemother Pharmacol 2012; 70: 637-44.

[54] von Minckwitz G, Huang CS, Mano MS, Loibl S, Mamounas EP, Untch M, Wolmark N, Rastogi P, Schneeweiss A, Redondo A, Fischer HH, Jacot W, Conlin AK, Arce-Salinas C, Wapnir IL, Jackisch C, DiGiovanna MP, Fasching PA, Crown JP, Wülfing P, Shao Z, Rota Caremoli E, Wu H, Lam LH, Tesarowski D, Smitt M, Douthwaite H, Singel SM and Geyer CE Jr; KATHERINE Investigators. Trastuzumab emtansine for residual invasive HER2-positive breast cancer. N Engl J Med 2019; 380: 617-628.

[55] Shi W, Zhang G, Ma Z, Li L, Liu M, Qin L, Yu Z, Zhao L, Liu Y, Zhang X, Qin J, Ye H, Jiang X, Zhou H, Sun H and Jiao Z. Hyperactivation of HER2-SHCBP1-PLK1 axis promotes tumor cell mitosis and impairs trastuzumab sensitivity to gastric cancer. Nat Commun 2021; 12: 2812.

[56] Rapti V, Moirogiorgou E, Koliou GA, Papadopoulou K, Binas I, Pentheroudakis G, Bafaloukos D, Bobos M, Chatzopoulos K, Chrisafi S, Christodoulou C, Nicolaou I, Sotiropoulou M, Magkou C, Koutras A, Papakostas P, Kotsakis A, Razis E, Psyrri A, Tryfonopoulos D, Pectasides D, Res E, Alexopoulos A, Kotoula V and Fountzilas G. mRNA expression of specific HER ligands and their association with clinical outcome in patients with metastatic breast cancer treated with trastuzumab. Oncol Lett 2022; 23: 23.

[57] Yeh Y, Chen C and Ko Y. Effectiveness and cost-effectiveness of trastuzumab emtansine in women with HER2-positive locally advanced or metastatic breast cancer: a systematic review and meta-analysis. J Cancer Res Ther 2022; 18: 1061-1072.

[58] Aapro M, Cardoso F, Curigliano G, Eniu A, Gligorov J, Harbeck N, Mueller A, Pagani O, Paluch-Shimon S, Senkus E, Thürlimann B and Zaman K. Current challenges and unmet needs in treating patients with human epidermal growth factor receptor 2-positive advanced breast cancer. Breast 2022; 66: 145-156.

[59] Yamaguchi H, On J, Morita T, Suzuki T, Okada Y, Ono J and Evdokiou A. Combination of near-infrared photoimmunotherapy using trastuzumab and small protein mimetic for HER2-positive breast cancer. Int J Mol Sci 2021; 22: 12213.

[60] Chen YF, Xu YY, Shao ZM and Yu KD. Resistance to antibody-drug conjugates in breast cancer: mechanisms and solutions. Cancer Commun (Lond) 2023; 43: 297-337.

[61] Pacot L, Masliah-Planchon J, Petcu A, Terris B, Gauthier Villars M, Lespinasse J, Wolkenstein P, Vincent-Salomon A, Vidaud D and Pasmant E. Breast cancer risk in NF1-deleted patients. J Med Genet 2024; 61: 428-429.

[62] Krstic M, Kolendowski B, Cecchini MJ, Postenka CO, Hassan HM, Andrews J, MacMillan CD, Williams KC, Leong HS, Brackstone M, Torchia J, Chambers AF and Tuck AB. TBX3 promotes progression of pre-invasive breast cancer cells by inducing EMT and directly up-regulating SLUG. J Pathol 2019; 248: 191-203.

Study of high-frequency mutated genes in breast cancer
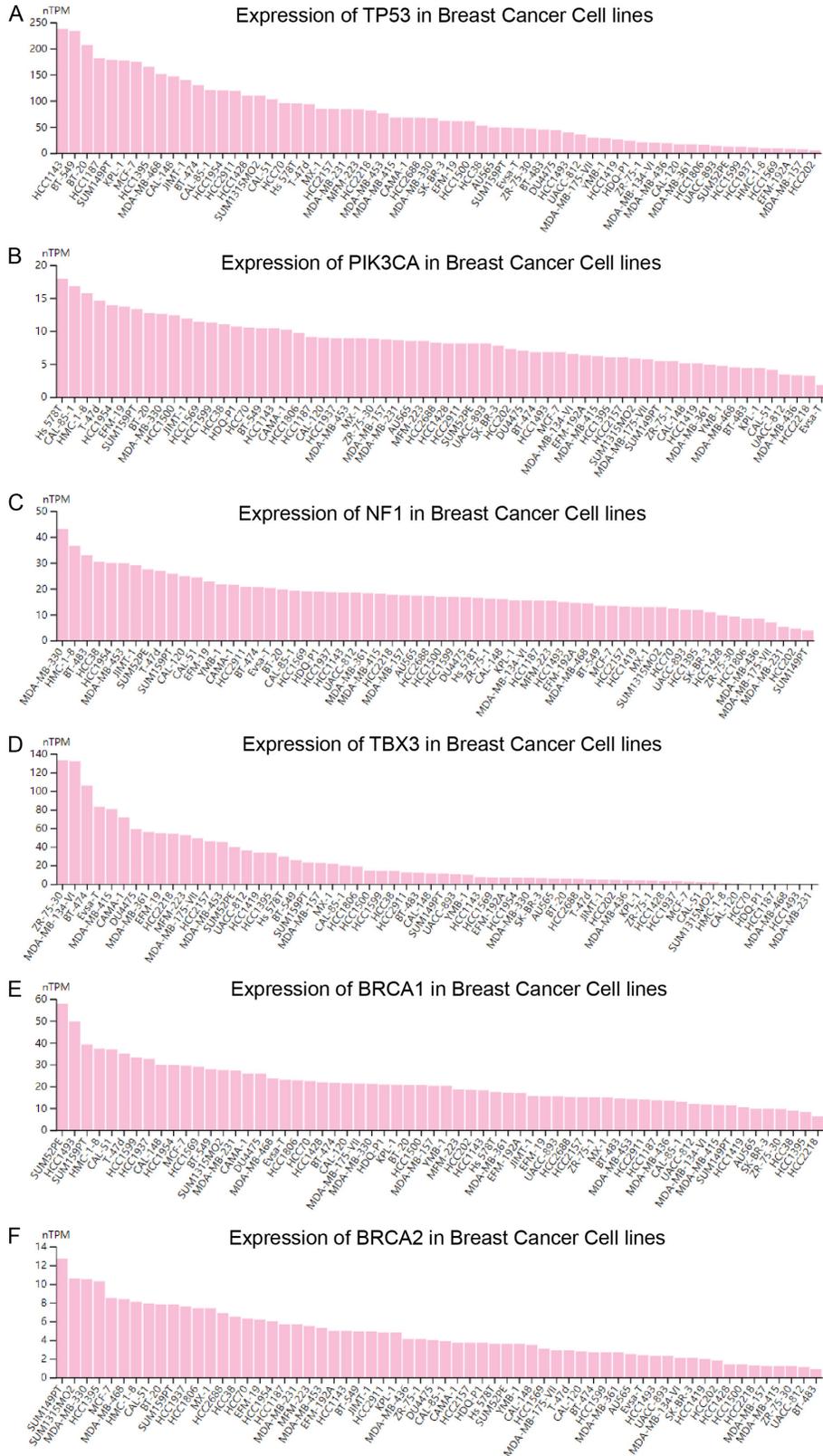
Supplementary Figure 1. Statistical collation of the BRCA cell line (n=62) dataset presenting the expression of TP53 (A), PIK3CA (B), NF1 (C), TBX3 (D), BRCA1 (E), and BRCA2 (F) at the cellular level and ranked according to the level of expression of each gene from lowest to highest.

Study of high-frequency mutated genes in breast cancer

A

Single Cell Type by TP53 Expression

Cluster	Cell type	Expression (nTPM) 0–80
c-0	T-cells	
c-1	Breast glandular cells	
c-2	T-cells	
c-3	Macrophages	
c-4	Breast glandular cells	
c-5	Fibroblasts	
c-6	Adipocytes	
c-7	Adipocytes	
c-8	Adipocytes	
c-9	Breast glandular cells	
c-10	T-cells	
c-11	Breast glandular cells	
c-12	Fibroblasts	
c-13	Fibroblasts	
c-14	Smooth muscle cells	
c-15	T-cells	
c-16	Breast glandular cells	
c-17	Breast glandular cells	
c-18	Macrophages	
c-19	Macrophages	
c-20	Breast myoepithelial cells	
c-21	Breast glandular cells	
c-22	Endothelial cells	
c-23	Adipocytes	
c-24	T-cells	

B

Single Cell Type by PIK3CA Expression

Cluster	Cell type	Expression (nTPM) 0–30
c-0	T-cells	
c-1	Breast glandular cells	
c-2	T-cells	
c-3	Macrophages	
c-4	Breast glandular cells	
c-5	Fibroblasts	
c-6	Adipocytes	
c-7	Adipocytes	
c-8	Adipocytes	
c-9	Breast glandular cells	
c-10	T-cells	
c-11	Breast glandular cells	
c-12	Fibroblasts	
c-13	Fibroblasts	
c-14	Smooth muscle cells	
c-15	T-cells	
c-16	Breast glandular cells	
c-17	Breast glandular cells	
c-18	Macrophages	
c-19	Macrophages	
c-20	Breast myoepithelial cells	
c-21	Breast glandular cells	
c-22	Endothelial cells	
c-23	Adipocytes	
c-24	T-cells	

C

Single Cell Type by NF1 Expression

Cluster	Cell type	Expression (nTPM) 0–50
c-0	T-cells	
c-1	Breast glandular cells	
c-2	T-cells	
c-3	Macrophages	
c-4	Breast glandular cells	
c-5	Fibroblasts	
c-6	Adipocytes	
c-7	Adipocytes	
c-8	Adipocytes	
c-9	Breast glandular cells	
c-10	T-cells	
c-11	Breast glandular cells	
c-12	Fibroblasts	
c-13	Fibroblasts	
c-14	Smooth muscle cells	
c-15	T-cells	
c-16	Breast glandular cells	
c-17	Breast glandular cells	
c-18	Macrophages	
c-19	Macrophages	
c-20	Breast myoepithelial cells	
c-21	Breast glandular cells	
c-22	Endothelial cells	
c-23	Adipocytes	
c-24	T-cells	

D

Single Cell Type by TBX3 Expression

Cluster	Cell type	Expression (nTPM) 0–300
c-0	T-cells	
c-1	Breast glandular cells	
c-2	T-cells	
c-3	Macrophages	
c-4	Breast glandular cells	
c-5	Fibroblasts	
c-6	Adipocytes	
c-7	Adipocytes	
c-8	Adipocytes	
c-9	Breast glandular cells	
c-10	T-cells	
c-11	Breast glandular cells	
c-12	Fibroblasts	
c-13	Fibroblasts	
c-14	Smooth muscle cells	
c-15	T-cells	
c-16	Breast glandular cells	
c-17	Breast glandular cells	
c-18	Macrophages	
c-19	Macrophages	
c-20	Breast myoepithelial cells	
c-21	Breast glandular cells	
c-22	Endothelial cells	
c-23	Adipocytes	
c-24	T-cells	

2

Supplementary Figure 2. The single-cell clustering results of breast cancer tissue samples corresponding to TP53 (A), PIK3CA (B), NF1 (C), TBX3 (D), BRCA1 (E) and BRCA2 (F) and the cells can be effectively clustered by T-cells, Breast glandular cells, Macrophages, Breast myoepithelial cells.

Study of high-frequency mutated genes in breast cancer

Supplementary Figure 3. TP53 (A), PIK3CA (B), NF1 (C) and TBX3 (D) had different degrees of influence on BRCA immune cell infiltration: TP53 and PIK3CA showed significant differences on the degree of immune cell infiltration; NF1 and TBX3 had an effect on the extent of breast cancer immune cell infiltration, but there was no significant difference.