Am J Transl Res 2022;14(11):7670-7688
www.ajtr.org /ISSN:1943-8141/AJTR0144974

Original Article

MACF1 mutations predict poor prognosis: a novel potential therapeutic target for breast cancer

Ye Tian[1], Kongjun Zhu[1], Yuefei Li[2], Zhen Ren[2], Juan Wang[2]

[1]Department of Thyroid and Breast Surgery, Wuhan No. 1 Hospital, Wuhan, Hubei, China; [2]Department of Blood Transfusion, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China

Received June 27, 2022; Accepted October 19, 2022; Epub November 15, 2022; Published November 30, 2022

Abstract: Objective: Microtubule actin cross-linking factor 1 (MACF1) mutations are known to play an important role in the progression of various cancers. However, its role in breast cancer remains to be determined. In this study, we investigated how MACF1 mutations may play a role in breast cancer development. Methods: The gene-expression profile data of patients with breast cancer were obtained from The Cancer Genome Atlas (TCGA)-Breast cancer cohort. We estimated the influence of MACF1 mutations on patient clinical prognosis using the Kaplan-Meier method. Further, patients with MACF1-mutant (MACF1-MT) and MACF1-wild-type (MACF1-WT) were compared to identify the differentially expressed genes (DEGs). We also performed functional enrichment analyses, constructed protein-protein interaction (PPI) and competing endogenous RNA (ceRNA) networks, and investigated the correlation between MACF1 mutations and immune-cell infiltration. To explore the prognostic value of MACF1 mutations, a nomogram was developed based on MACF1 mutations and other clinicopathological parameters. Results: Patients with MACF1-MT had a worse prognosis and higher tumor mutation burden score ($P < 0.05$) than patients with MACF1-WT. MACF1 mutations were demonstrated to upregulate the mTOR signaling pathway and alter energy metabolism and tumor immune microenvironment. Thus, MACF1 mutations might affect immunogenicity and result in a lower response to immunotherapy. By analyzing the Genomics of Drug Sensitivity in Cancer (GDSC), the sensitivity of breast cancer cells to 13 drugs was found to be significantly enhanced by MACF1 mutations. The prognostic model was verified in predicting the outcome of breast cancer patients. Conclusion: MACF1 mutations might be a potential prognostic biomarker and a therapeutic target for breast cancer.

Keywords: MACF1, breast cancer, immunotherapy, prognosis, tumor immune microenvironment

Introduction

Breast cancer is a major cause of cancer-associated deaths in women worldwide [1]. Despite significant development in the diagnosis and therapy for breast cancer patients, 20-30% of early-stage cases develop distant metastasis [2]. Patients with advanced breast cancer often have a poor prognosis [3], with low treatment responses. This outcome is attributed to tumor heterogeneity at the molecular level, including processes such as signal transduction, cell cycle regulation, and the tumor microenvironment (TME), which have molecular functions and interactions at multiple steps [4]. Although various biomarkers, such as CA153 and CEA, have been related to breast cancer [5], the identification of more specific and sensitive biomarkers are still needed to improve the current diagnosis, treatment, and prognosis of patients with breast cancer.

The cytoskeleton comprises microtubules (MTs), microfilaments (F-actin), and intermediate filaments (IFs). Previous studies have demonstrated that specific cross-linking of proteins is significantly involved in various functions of the cytoskeleton [6]. Microtubule actin cross-linking factor 1 (*MACF1*) was the first specific cross-linking protein that were identified [7]. *MACF1* is located on chromosome 1p34.3 and contains at least 110 exons. *MACF1* is widely expressed in tissues, especially the cytoplasm, and can combine with the microtubule and microfilament cytoskeleton network, which are significantly involved in cell proliferation, migra-

tion, and maintenance of tissue integrity [8]. *MACF1* also participates in cell signal transduction, embryo development, and several other diseases.

Given its role in cell signal transduction, *MACF1* may be involved in cancer development. With the advent of whole genome sequencing, *MACF1* has been shown to have a high mutation rate in various cancers, including lung, colon, and gastric cancer. In gastric cancer, *MACF1* mutations were found to be related to cancer aggressiveness by promoting cellular metastasis and were correlated with a poor prognosis [9]. Another study found that the mutation rate of *MACF1* in breast cancer is 12% [10]. *MACF1* appears to be involved in the Wnt/β-catenin signaling pathway [11, 12]. However, only a few studies have explored the relationship between *MACF1* mutations and breast cancer. Accordingly, the underlying functions and mechanisms of *MACF1* mutations in tumor progression and TME in breast cancer remain unknown.

In the present study, mutational and clinical data were obtained from TCGA-Breast Cancer cohort to investigate the association between *MACF1* mutations and prognosis in breast cancer patients. Patients enrolled in this study were divided into *MACF1*-mutant (MT) and *MACF1*-wild-type (WT) groups. Based on our findings, *MACF1*-MT patients had a higher tumor mutation burden (TMB) than *MACF1*-WT patients, which might affect immunogenicity. Furthermore, to identify the differentially-expressed genes (DEGs) in *MACF1*-MT and *MACF1*-WT patients, the biological pathways and functions relating to *MACF1* mutations were elucidated. Our results provide new insights into the mechanisms underlying breast cancer development and *MACF1* mutations might serve as a novel prognostic biomarker and a potential therapeutic target for patients with breast cancer.

Materials and methods

Data processing and analysis of mutations

Genomic somatic mutation data (MAF files) and the corresponding clinical and prognostic data of breast cancer patients were obtained from TCGA (http://cancergenome.nih.gov/) [13] and UCSC Xena (http://xena.ucsc.edu/). Data for 955 breast cancer patients with mutational and survival information were collected for further analyses and annotated using the GH-Ch38 version of the genome annotation file from the Ensembl database (ftp://ftp.ensembl.org/pub/current_gtf) [14]. Copy number variation data were also collected from TCGA database. The clinical parameters of the breast cancer patients are summarized in Supplementary Table 1. As all data were collected from TCGA for this study, institutional ethics committee approval was not required.

Somatic mutation data were visualized using the R packages, Maftools [15] and GenVisR [16], and the *MACF1* mutations were visualized using the R package G3viz [17]. GISTIC2.0 analysis by Genepattern (http://cloud.genepattern.org/) was used to assess gene copy number variations. To estimate the association between *MACF1* mutations and patient prognosis, patients were divided into *MACF1*-MT and *MACF1*-WT groups, and survival analysis was performed using the Kaplan-Meier method. TMB [18] levels and microsatellite instability (MSI) [19] were compared between the *MACF1*-MT and *MACF1*-WT groups.

In addition, the Human Protein Atlas (HPA) database [20] was used to evaluate MACF1 protein expression levels in breast cancer tissues and normal breast tissues. TCGA pan-cancer samples were analyzed with the cBioPortal database [21] for mutation status of MACF1. The gene expression levels, copy number variation (CNV), and protein expression levels of MACF1 in various types of cancer cell lines were compared using the Cancer Cell Line Encyclopedia (CCLE) database [22].

Analysis of drug sensitivity and DEGs

Whole exome sequencing (WES) data were downloaded from GDSC and differences in drug sensitivity were compared using the R package oncoPredict [23]. We used the R package GSVA [24] for gene set variation analysis of the *MACF1*-MT and *MACF1*-WT groups. Using this R package, the reference gene sets "h.all.V7.4.symbols.gmt" and "C2.cp.kegg.V7.4.symbols.gmt" were downloaded from the MSigDB database [25] to calculate the enrichment score for each sample. Dysregulated pathways in the *MACF1*-MT and *MACF1*-WT groups were analyzed. Moreover, mutation data for the *MACF1*-MT and *MACF1*-WT groups were compared to identify DEGs using the limma pack-

age in R. The thresholds for this analysis were $|\log_2(\text{fold-change (FC)})| > 1$ and *P*-value < 0.05. Genes with logFC > 1 and *P*-value < 0.05 were upregulated, while genes with logFC < -1 and *P*-value < 0.05 were downregulated. The DEGs were divided into mRNA, lncRNA, and miRNA groups. DEGs were visualized using volcano plots.

Gene function and pathway enrichment analysis

Functional and pathway enrichment analyses, including gene ontology (GO) [26], analysis of the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, and the identification of cellular components (CC), molecular functions (MF), and biological processes (BP) [27], of the DEGs were performed using the R package clusterProfiler [28]. Gene set enrichment analysis (GSEA) [29] was also performed using clusterProfiler to clarify the significant functions and pathway differences between *MACF1*-MT and *MACF1*-WT. Both "c2.cp.kegg.V7.4.entrez.gmt" and "c5.go.v7.4.entrez.gmt" were downloaded as reference gene sets from the MSigDB database to perform GSEA.

Protein-protein interaction (PPI) network construction and ceRNA network construction

The Search Tool for the Retrieval of Interacting Genes (STRING) was used to construct a PPI network for DEGs [30]. Statistically significant interactions were identified using a combined score > 400. In this study, to identify crucial subnetworks, we used MCODE, extracted hub genes in the PPI network [31], and set an MCODE score > 10 as the cut-off value. The Cytoscape software (version 3.7.0) was used to visualize the PPI networks, whereas clueGO [32] was used for functional annotation. Similar to ceRNAs, lncRNAs can sponge miRNAs and reduce miRNA inhibitory effects on target mRNAs. To analyze the relationship between DEGs, miRNAs, and lncRNAs in the post-transcriptional stage, mRNAs, and lncRNAs related to differentially-expressed miRNAs from the miRNet database [33] were obtained, and an mRNA-miRNA-lncRNA regulatory network was constructed based on the intersection of differentially expressed mRNAs and lncRNAs identified in TCGA-Breast cancer cohort. The Cytoscape software (version 3.7.0) was used to identify the mRNA-miRNA-LncRNA regulatory networks.

Immune infiltration analysis

To investigate the association between *MACF1* mutations and the TME, differences in immune cell content between *MACF1*-MT and *MACF1*-WT patients were assessed. Further, the correlations among immune cell types in patients with *MACF1* mutations were determined. Stromal and immune cell abundance was quantified using the ESTIMATE package [34] in the R software. Immune infiltration analysis for the 22 types of immune cells in the tumor samples was carried out using CIBERSORT [35]. Correlations between *MACF1*, immune cells, and human leukocyte antigen (HLA) genes were analyzed using Spearman's correlation.

Construction and evaluation of a nomogram

We collected the clinical and mutational data of 955 patients with breast cancer from TCGA Breast Cancer cohort. The 955 patients were randomly divided into training and testing (5:5) sets for constructing and validating a prognostic model. Based on the results of Cox regression analysis, we constructed a nomogram that integrated the valuable clinicopathological variables from the training set. The rms R package was used to construct the nomograms. The nomogram was assessed using receiver operating characteristic (ROC) analyses, calibration plots, and Decision Curve Analysis (DCA). In addition, validation of the model was carried out using the testing set.

Statistical analysis

All statistical analyses and plots were performed and generated using the R (v3.6.3) software. Multiple testing corrections were performed using Benjamin Hochberg (BH). FDR correction was also performed to reduce the false-positive rate. For comparisons between two groups, Student's *t*-test was employed to estimate the significance of normally distributed variables, while the Mann-Whitney U test was used to analyze non-normally distributed variables. Survival analysis was performed using the Kaplan-Meier method with the R package survival [36], and log-rank tests in separate curves. Univariate and multivariate Cox regression analyses were used to identify independent variables. P < 0.05 was considered to indicate statistical significance, and all statistical tests were bilateral.

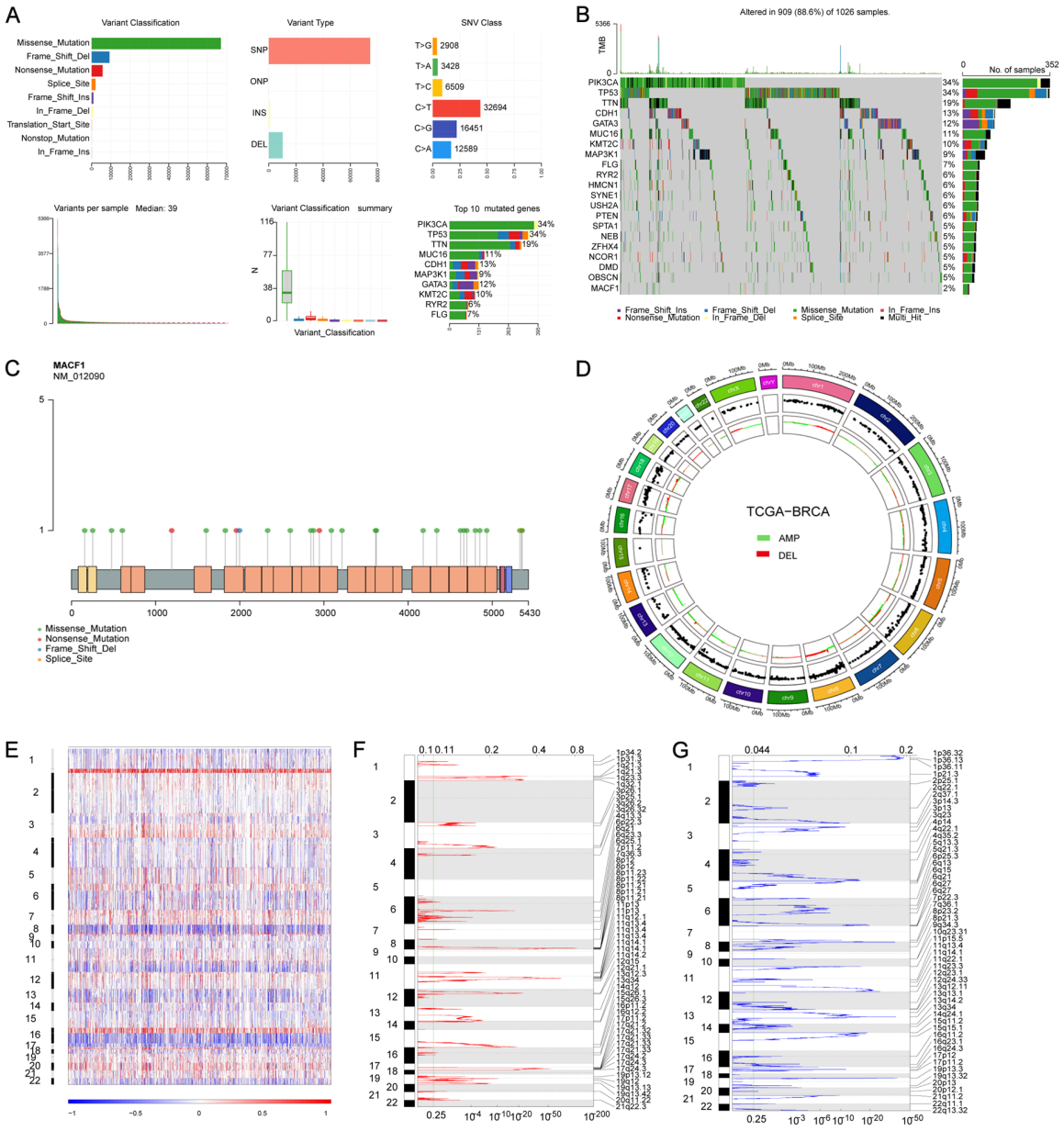


Figure 1. Analysis of somatic mutations and copy number variations in breast cancer patients. A. Statistics of mutation information in The Cancer Genome Atlas (TCGA)-Breast cancer cohort. B. Genes with the most frequent mutations were identified and presented. The panel on the left shows the genes ordered by their frequencies of mutation, while the panel on the right shows the different mutation types expressed as different colors. C. Microtubule actin cross-linking factor 1 (*MACF1*) mutations in TCGA-Breast cancer cohort. D. Copy number alteration (CNV) analysis; the outmost edge represents chromosomes, green represents amplification, and red represents deletion. E-G. Identification of genes with significant amplification or deletion.

Results

High gene mutation frequencies in breast cancer patients from TCGA-Breast cancer dataset

We analyzed the mutation types of mutational genes in breast cancer patients from the Breast Cancer dataset (**Figure 1A**). As shown in **Figure 1A**, the most common mutations were missense mutations, and the second most common mutation was deletion. Genes with the highest mutation frequencies were *PIK3CA*, *TP53*, and *TTN*. As shown in **Figure 1B**, a high frequency of *MACF1* mutations was observed.

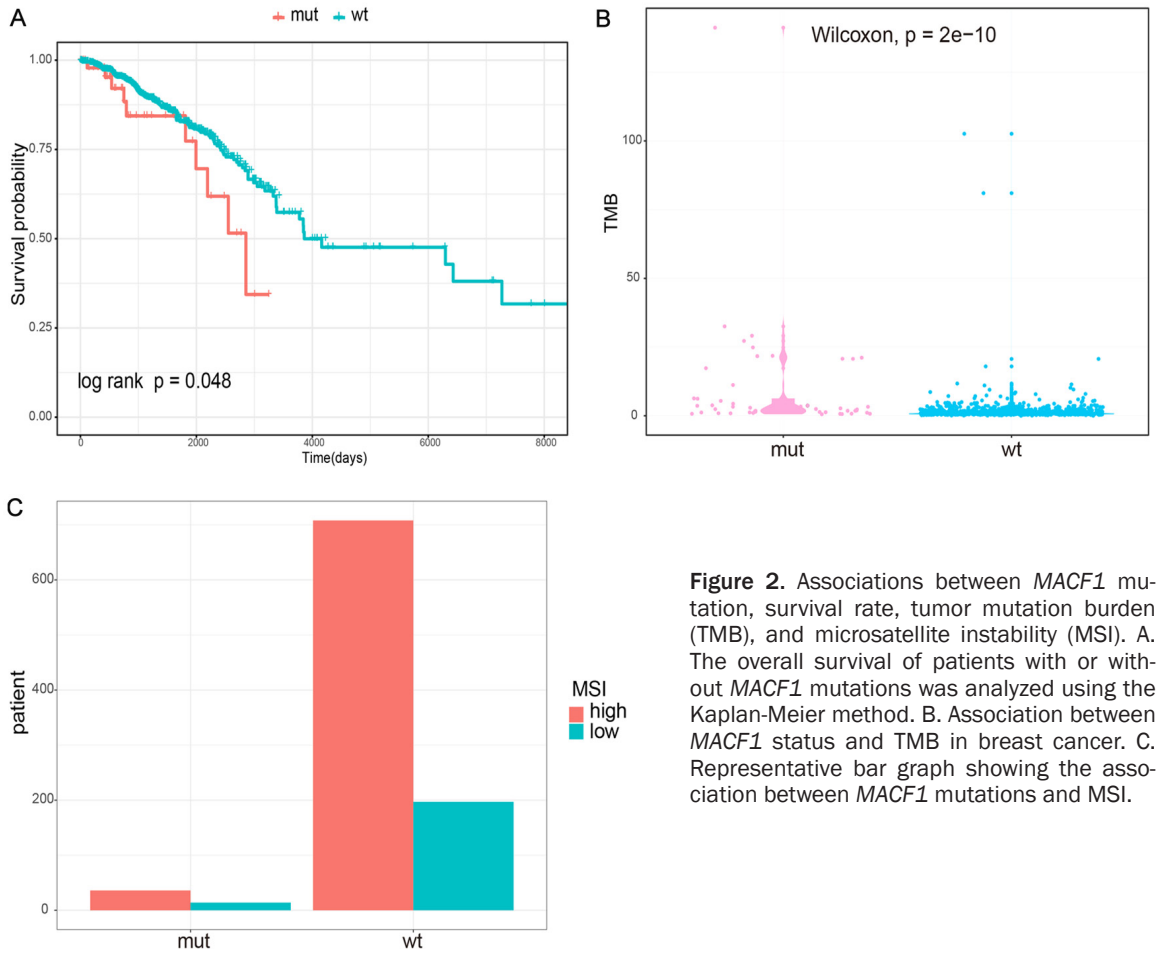


Figure 2. Associations between *MACF1* mutation, survival rate, tumor mutation burden (TMB), and microsatellite instability (MSI). A. The overall survival of patients with or without *MACF1* mutations was analyzed using the Kaplan-Meier method. B. Association between *MACF1* status and TMB in breast cancer. C. Representative bar graph showing the association between *MACF1* mutations and MSI.

Using copy number variation data from TCGA, genes that showed significant amplification or deletions were identified. *MACF1* was not significantly amplified or deleted (**Figure 1D-F**). The relationship between *MACF1* mutations and the patient prognosis was evaluated using Kaplan-Meier survival analysis. Compared to the *MACF1*-WT group, *MACF1*-MT was more strongly associated with poor prognosis (P = 0.048) (**Figure 2A**). MACF1 protein levels were significantly higher in normal breast tissues compared with breast cancer tissues according to the HPA database (Supplementary Figure 1A). According to the cBioPortal database, skin cutaneous melanoma had the highest alteration frequency of MACF1 mutations, and breast invasive carcinoma had high alteration frequency of MACF1 mutations (Supplementary Figure 1B). Based on the CCLE database, the gene expression levels and protein expression levels of MACF1 in breast cancer cell lines was lower than that in most types of cancer cell lines (Supplementary Figure 1C, 1E), and the difference of CNV of MACF1 in various types of cancer cell lines was not significant (Supplementary Figure 1D).

Drug sensitivity and analysis of TMB and MSI

Compared with patients without *MACF1* mutations, those with *MACF1* mutations exhibited a higher TMB score (P < 0.05) (**Figure 2B**). The differences in MSI levels between patients with and without *MACF1* mutations were not statistically significant (**Figure 2C**). The correlations between *MACF1* mutations and drug sensitivity in breast cancer cells were analyzed using the GDSC cohort. Patients with *MACF1* mutations were found to have lower IC_{50} values for several drugs, including GSK1059615, rapamycin, and mitomycin-C. Moreover, these patients displayed poor sensitivity to AZD8055 (**Figure 3A**). We proceeded to determine the effect of *MACF1* mutations on biological characteristics and oncogenic signaling pathways using the R

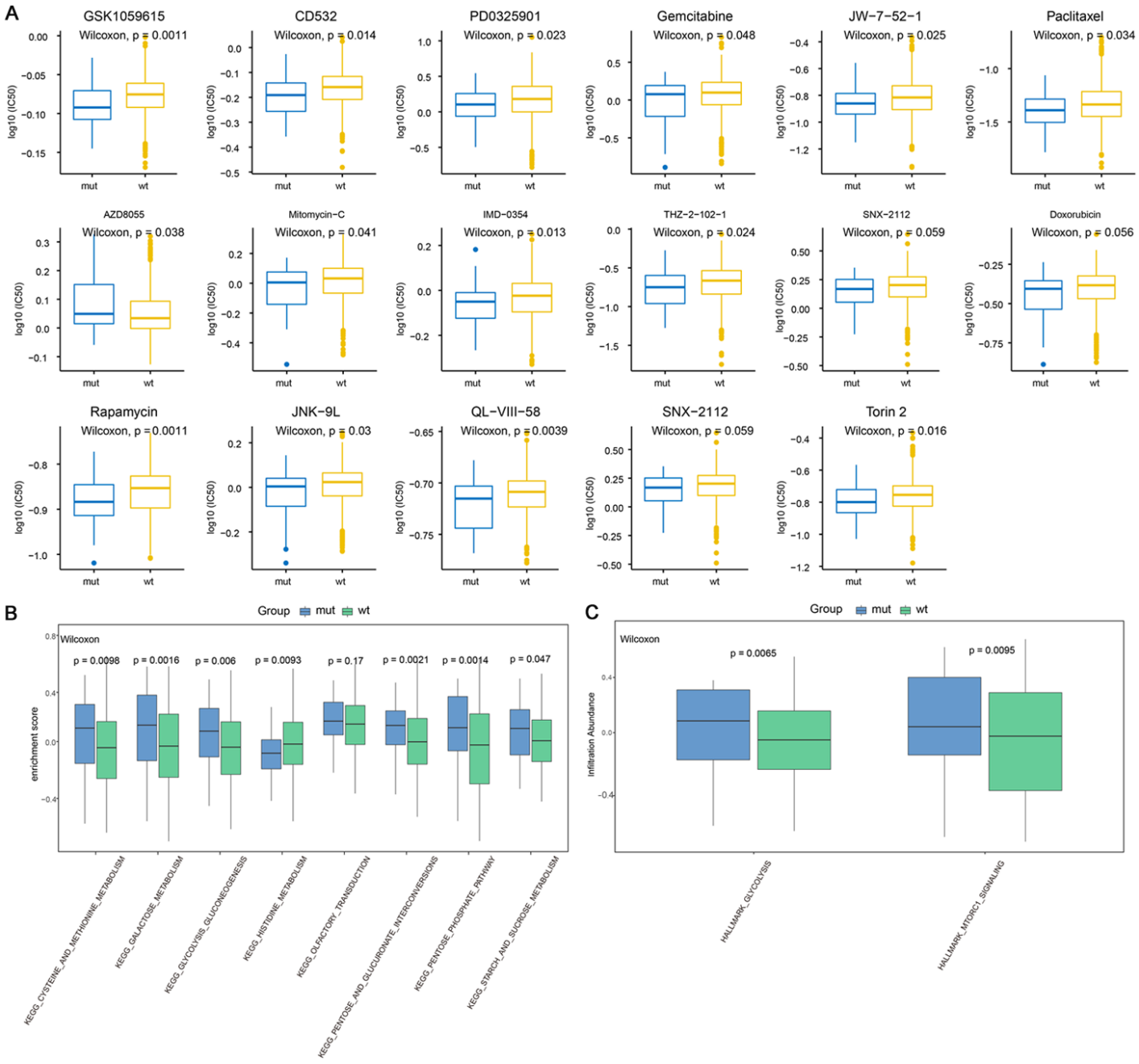

A
GSK1059615
Wilcoxon, p = 0.0011
CD532
Wilcoxon, p = 0.014
PD0325901
Wilcoxon, p = 0.023
Gemcitabine
Wilcoxon, p = 0.048
JW−7−52−1
Wilcoxon, p = 0.025
Paclitaxel
Wilcoxon, p = 0.034
AZD8055
Wilcoxon, p = 0.038
Mitomycin−C
Wilcoxon, p = 0.041
IMD−0354
Wilcoxon, p = 0.013
THZ−2−102−1
Wilcoxon, p = 0.024
SNX−2112
Wilcoxon, p = 0.059
Doxorubicin
Wilcoxon, p = 0.056
Rapamycin
Wilcoxon, p = 0.0011
JNK−9L
Wilcoxon, p = 0.03
QL−VIII−58
Wilcoxon, p = 0.0039
SNX−2112
Wilcoxon, p = 0.059
Torin 2
Wilcoxon, p = 0.016
log10 (IC50)
mut
wt
B
Group mut wt
Wilcoxon
p = 0.0098
p = 0.0016
p = 0.006
p = 0.0093
p = 0.17
p = 0.0021
p = 0.0014
p = 0.047
enrichment score
KEGG_CYSTEINE_AND_METHIONINE_METABOLISM
KEGG_GALACTOSE_METABOLISM
KEGG_GLYCOLYSIS_GLUCONEOGENESIS
KEGG_HISTIDINE_METABOLISM
KEGG_OLFACTORY_TRANSDUCTION
KEGG_PENTOSE_AND_GLUCURONATE_INTERCONVERSIONS
KEGG_PENTOSE_PHOSPHATE_PATHWAY
KEGG_STARCH_AND_SUCROSE_METABOLISM
C
Group mut wt
Wilcoxon
p = 0.0065
p = 0.0095
Infiltration Abundance
HALLMARK_GLYCOLYSIS
HALLMARK_MTORC1_SIGNALING

Figure 3. Drug sensitivity analysis and differences in biological characteristics. A. Representative box plots showing drug sensitivity of the breast cancer cells based on their *MACF1* mutations. B. Results of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment between the *MACF1*-mutant (*MACF1*-MT) and MACF1-wild-type (*MACF1*-WT) groups. C. Results of hallmark enrichment between the *MACF1*-MT and *MACF1*-WT groups.

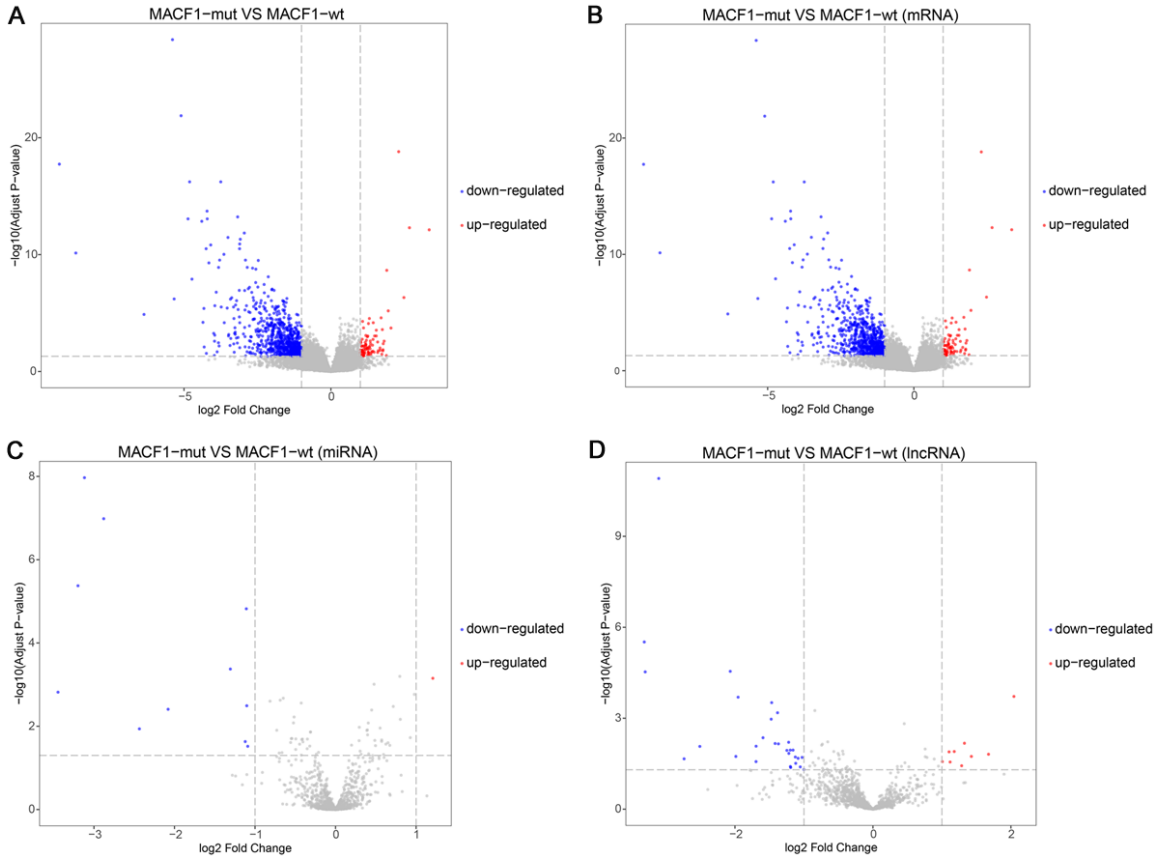


Figure 4. Identifying differentially expressed genes (DEGs) between the *MACF1*-MT and *MACF1*-WT groups. A. Representative Volcano plot showing DEGs in patients with or without *MACF1* mutations. B. Representative Volcano plot showing mRNAs with different expression between the *MACF1*-MT and *MACF1*-WT groups. C. Representative Volcano plot showing miRNAs with different expression between *MACF1*-MT and *MACF1*-WT groups. D. Representative Volcano plot showing LncRNAs with different expression between the *MACF1*-MT and *MACF1*-WT groups.

package GSVA. Based on the results for the *MACF1*-MT group, hallmark glycolysis and mTORC1 signaling were significantly upregulated (**Figure 3C**), while KEGG histidine metabolism was significantly downregulated. In parallel, KEGG cysteine and methionine metabolism, galactose metabolism, and glycolytic gluconeogenesis were significantly upregulated (**Figure 3B**).

DEGs identification and functional enrichment analysis

Based on the cut-off criteria (|logFC| > 1 and P < 0.05), the gene profile data of *MACF1*-MT patients (n = 50) or *MACF1*-WT patients (n = 905) were analyzed using the R package limma. As shown in **Figure 4A**, 949 DEGs were identified (**Figure 4A**). Further, as shown in **Figure 4B** and **4D**, 38 differentially expressed lncRNAs and 911 mRNAs were found. Among these 38 lncRNAs, 9 were upregulated and 29 were downregulated. Of the 911 mRNAs, 79 were upregulated and 832 were downregulated (**Figure 4B**, **4D**).

The miRNA profile data were analyzed using the same cutoff criteria. A total of 12 miRNAs were identified (**Figure 4C**). Among the 12 identified miRNAs, one was upregulated and 11

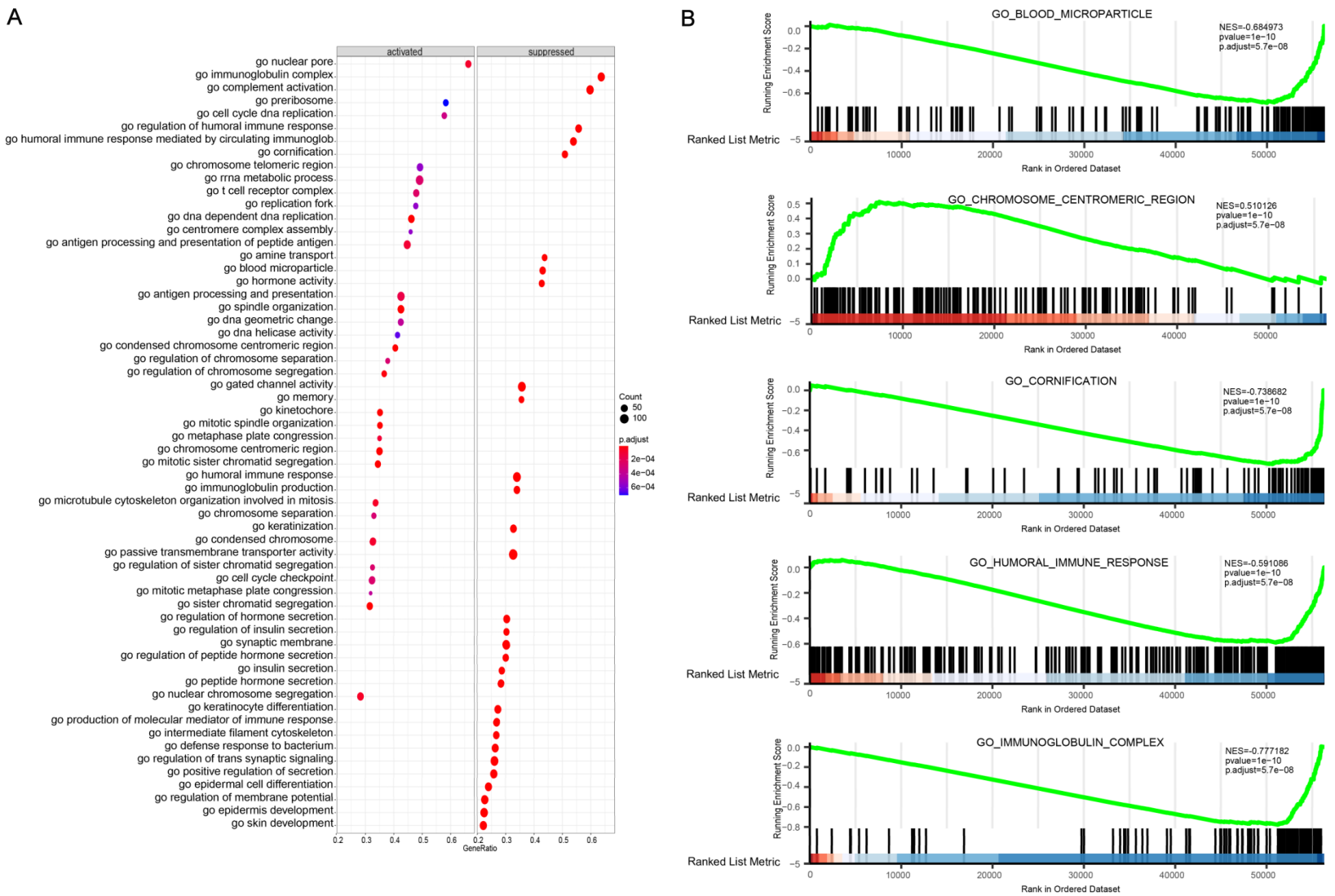
were downregulated (**Figure 4C**). To better understand the functional enrichment of *MACF1* mutations in breast cancer, GO (Supplementary Figure 2A; Supplementary Table 2) and KEGG functional enrichment analyses were performed using the R package clusterProfiler. Ten enriched GO terms were identified in the BP category. These terms were enriched in response to chemical synaptic transmission, epidermis development, the epoxygenase P450 pathway, proteolysis, cytoskeleton organization, cell-cell signaling, response to nicotine, positive regulation of synaptic transmission, glutamatergic, memory, and digestion (Supplementary Figure 2B). Categorization by CC revealed 10 enriched GO terms, which were associated with organelle membrane, extracellular space, postsynaptic membrane, intermediate filament, plasma membrane, keratin filament, extracellular region, and blood microparticles and so on (Supplementary Figure 2C). Moreover, the MF category revealed 10 significant enrichments in GO terms associated with monooxygenase activity, serine-type endopeptidase activity, structural constituent of the cytoskeleton, acting on paired donors, structural molecule activity, oxidoreductase activity, arachidonic acid epoxygenase activity, oxygen binding, heme binding and so on (Supplementary Figure 2D). Diseases, such as visual seizure, opioid addiction, epileptic seizures, nicotine dependence, withdrawal symptoms, melancholia, cocaine abuse, and palmoplantar keratosis, also correlated with the 911 DEGs (Supplementary Figure 2E). Finally, the pathways that were significantly affected by DEGs were analyzed (Supplementary Figure 2F; Supplementary Table 3). The results showed that neuroactive ligand-receptor interactions, gastric acid secretion, and nicotine addiction were affected. We showed the top two significantly enriched pathways (Supplementary Figure 2G, 2H).

MACF1 mutations related biological functions and pathways obtained by GSEA

GSEA of the differences between *MACF1*-MT and *MACF1*-WT groups was conducted to identify key signaling pathways and biological functions associated with *MACF1* mutations (**Figure 5A**, **5B**; Supplementary Table 4). Biological functions, including blood microparticles, chromosome centromeric region, cornification, humoral immune response, immunoglobulin complex, intermediate filament cytoskeleton, keratinization, keratinocyte differentiation, and mitotic sister chromatid segregation, showed significant differential enrichment in *MACF1*-MT. Further, signaling pathways, such as drug metabolism cytochrome p450, antigen processing and presentation, graft versus host disease, cell cycle, retinol metabolism, and natural killer cell-mediated cytotoxicity, were significantly enriched (**Figure 5C**, **5D**; Supplementary Table 4).

PPI and ceRNA networks

A PPI network was constructed and visualized (Supplementary Figure 3A) using a total of 507 DEGs and 2,001 PPI pairs. The top five genes that interacted with other DEGs were *ALB* (interacting with 86 DEGs), *SNAP25* (interacting with 52 DEGs), *GRIA2* (interacting with 51 DEGs), *SST* (interacting with 46 DEGs), and *NRXN1* (interacting with 45 DEGs). In the PPI network related to DEGs, enrichment of genes associated with immunity, such as antimicrobial peptide production, neutrophil chemotaxis, regulation of lymphocyte proliferation, and regulation of osteoclast differentiation, was observed (Supplementary Figure 3B). The MCODE algorithm was applied to identify densely connected network components. The two most significant MCODE components were extracted from the PPI network and comprised 12 (Supplementary Figure 3C) and 21 genes (Supplementary Figure 3E), respectively. Genes in MCODE component 1 were enriched in syntaxin-1 binding and neuron projection membrane (Supplementary Figure 3D), whereas genes in MCODE component 2 were enriched in startle response and calcium-dependent protein binding (Supplementary Figure 3F). Additionally, mRNAs, miRNAs, and lncRNAs that were differentially expressed in *MACF1*-MT were identified. Using the miRNet database, lncRNAs and mRNAs associated with miRNAs were identified, and an mRNA-miRNA-lncRNA regulatory network was constructed based on DEGs identified in TCGA-Breast cancer cohort (Supplementary Figure 3G). The network contained 47 interaction relationships, comprising 15 mRNAs, 5 miRNAs, and 2 lncRNAs. The differentially expressed mRNAs regulated by the


A
activated
suppressed
go nuclear pore
go immunoglobulin complex
go complement activation
go preribosome
go cell cycle dna replication
go regulation of humoral immune response
go humoral immune response mediated by circulating immunoglob
go cornification
go chromosome telomeric region
go rrna metabolic process
go t cell receptor complex
go replication fork
go dna dependent dna replication
go centromere complex assembly
go antigen processing and presentation of peptide antigen
go amine transport
go blood microparticle
go hormone activity
go antigen processing and presentation
go spindle organization
go dna geometric change
go dna helicase activity
go condensed chromosome centromeric region
go regulation of chromosome separation
go regulation of chromosome segregation
go gated channel activity
go memory
go kinetochore
go mitotic spindle organization
go metaphase plate congression
go chromosome centromeric region
go mitotic sister chromatid segregation
go humoral immune response
go immunoglobulin production
go microtubule cytoskeleton organization involved in mitosis
go chromosome separation
go keratinization
go condensed chromosome
go passive transmembrane transporter activity
go regulation of sister chromatid segregation
go cell cycle checkpoint
go mitotic metaphase plate congression
go sister chromatid segregation
go regulation of hormone secretion
go regulation of insulin secretion
go synaptic membrane
go regulation of peptide hormone secretion
go insulin secretion
go peptide hormone secretion
go nuclear chromosome segregation
go keratinocyte differentiation
go production of molecular mediator of immune response
go intermediate filament cytoskeleton
go defense response to bacterium
go regulation of trans synaptic signaling
go positive regulation of secretion
go epidermal cell differentiation
go regulation of membrane potential
go epidermis development
go skin development
0.2 0.3 0.4 0.5 0.6
GeneRatio
Count
50
100
p.adjust
2e−04
4e−04
6e−04
B
GO_BLOOD_MICROPARTICLE
NES=-0.684973
pvalue=1e−10
p.adjust=5.7e−08
GO_CHROMOSOME_CENTROMERIC_REGION
NES=0.510126
pvalue=1e−10
p.adjust=5.7e−08
GO_CORNIFICATION
NES=-0.738682
pvalue=1e−10
p.adjust=5.7e−08
GO_HUMORAL_IMMUNE_RESPONSE
NES=-0.591086
pvalue=1e−10
p.adjust=5.7e−08
GO_IMMUNOGLOBULIN_COMPLEX
NES=-0.777182
pvalue=1e−10
p.adjust=5.7e−08
Running Enrichment Score
Ranked List Metric
Rank in Ordered Dataset
0 10000 20000 30000 40000 50000

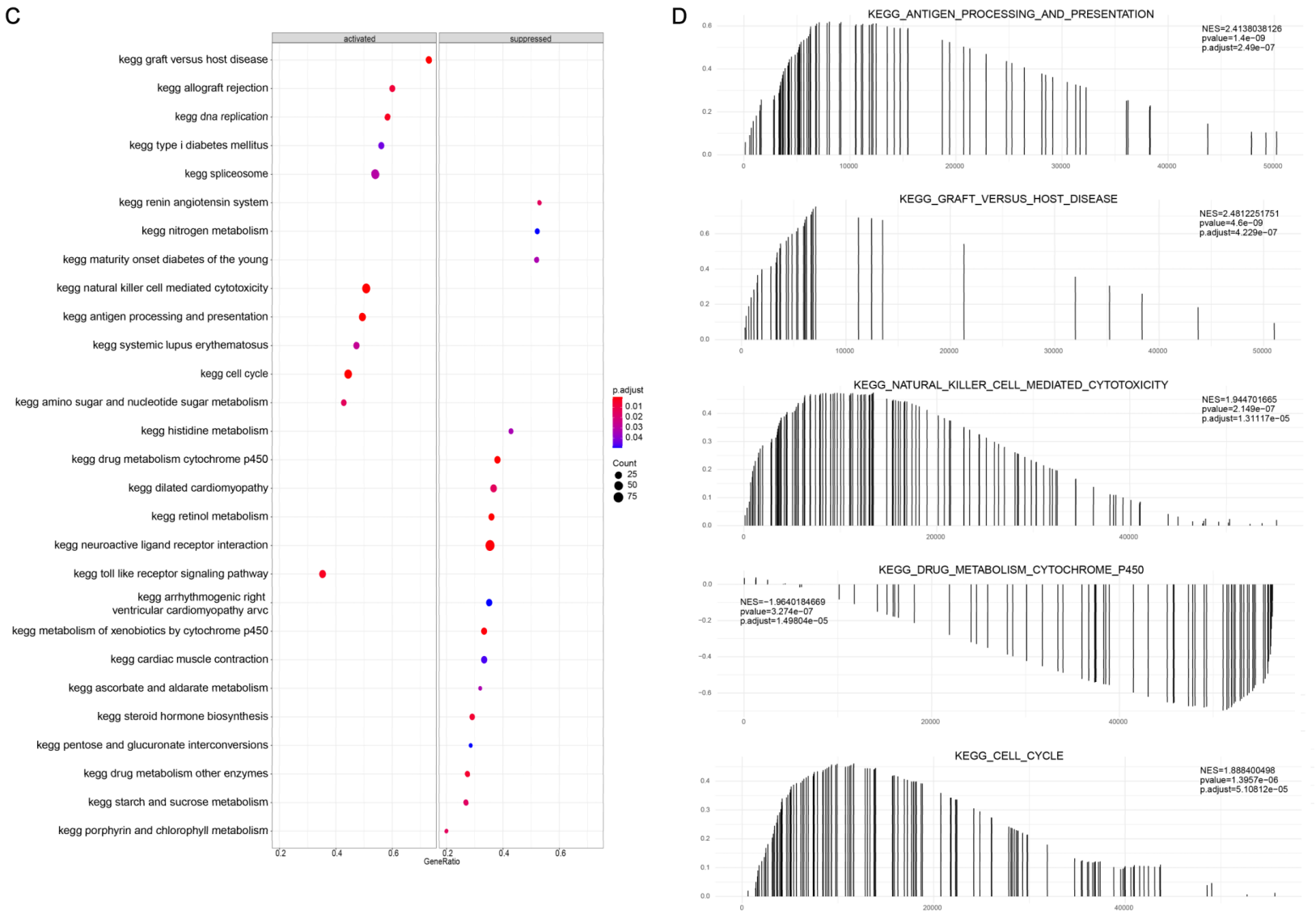


Figure 5. *MACF1* mutations related biological functions and pathways based on Gene set enrichment analysis (GSEA). A, B. GSEA of biological function enrichments of breast cancer patients with or without *MACF1* mutations. C, D. GSEA results showing enrichment of the functional pathways of breast cancer patients with or without *MACF1* mutations. The dot size depicts the number of genes contained in the enrichment functions, while the color of dots depicts adjusted *P*-values.

miRNAs were *AMER2*, *ANK1*, *NPTX1*, and *TMEM151B*. Further, the identified miRNAs were *hsa-mir-153-3p* (regulating 12 mRNAs), *hsa-mir-135a-5*p (regulating 6 mRNAs), *hsa-mir-105-5p* (regulating 4 mRNAs), *hsa-mir-216a-5p* (regulating 4 mRNAs and 1 lncRNA), and *hsa-mir-153-5p* (regulating 3 mRNAs).

Immune cell infiltration analysis

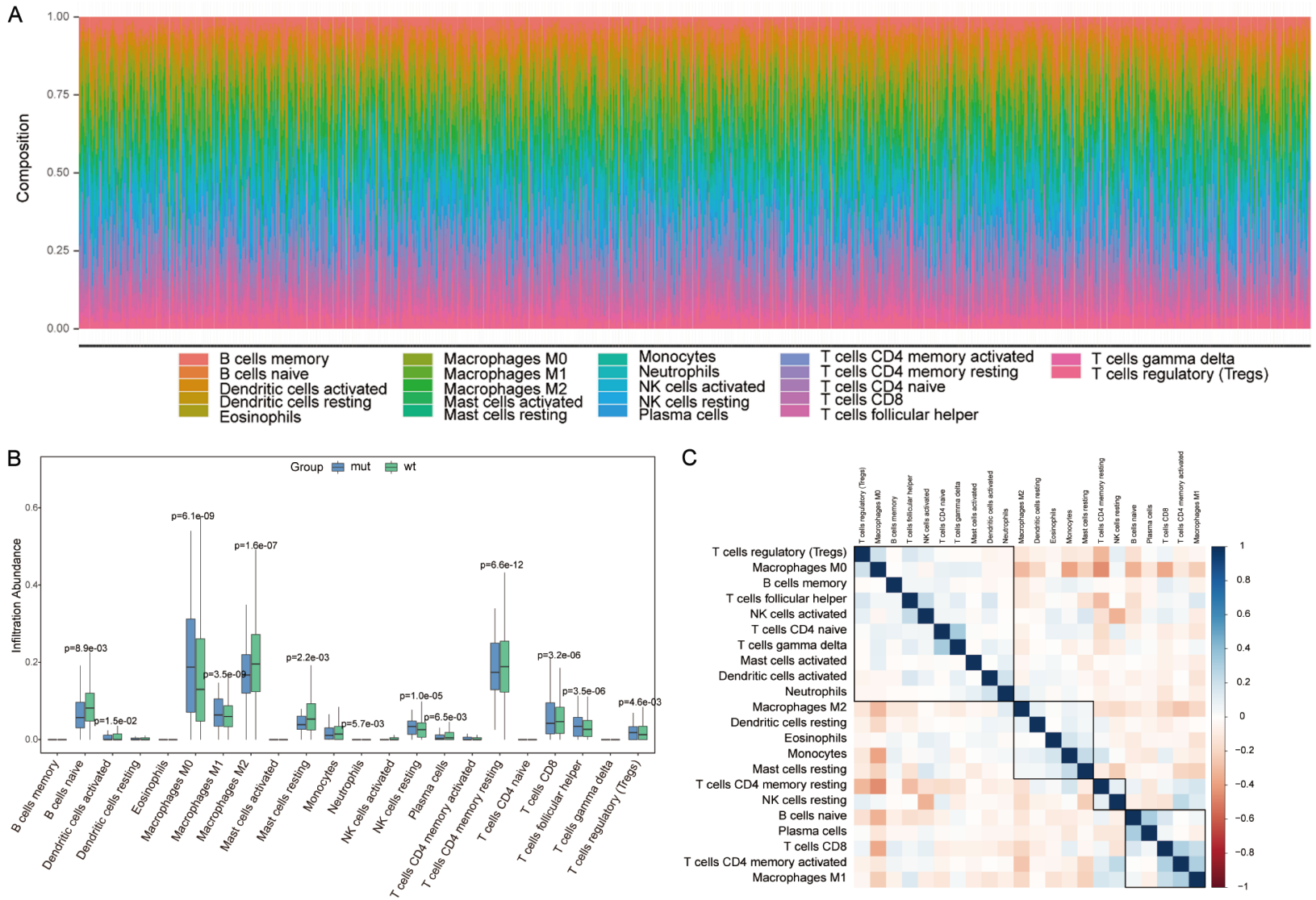
The immune cell content in patients with or without *MACF1* mutations was analyzed (**Figure 6A**). As a result, *MACF1*-MT patients were found to have a higher content of immune cells than *MACF1*-WT patients, suggesting that *MACF1* mutations may activate immune cells ($P < 0.05$, **Figure 6B**). We analyzed the correlation between immune cell types in *MACF1*-MT patients and found that most immune cells showed a negative correlation ($P < 0.05$, **Figure 6C**), indicating that different immune cells inhibit each other. Such finding suggests that the immunotherapeutic response of *MACF1*-MT patients might be poor. The correlation between 22 immunocytes and HLA genes was also determined, which revealed that activated memory $CD4^+$ and $CD8^+$ T cells, resting and activated mast cells, and M1 and M2 macrophages were significantly correlated with several HLA family genes ($P < 0.05$, **Figure 6D**). In addition, *MACF1* gene expression was positively correlated with monocytes, resting mast cells, naïve B cells, neutrophils, resting NK cells, and resting memory $CD4^+$ T cells. In parallel, *MACF1* expression was negatively correlated with activated NK cells, regulatory T cells (Tregs), follicular helper T cells, and $CD8^+$ T cells ($P < 0.05$, **Figure 6E**). *MACF1* gene expression was also negatively correlated with the immune score ($P < 0.05$, **Figure 6F**) and positively correlated with the stromal score ($P < 0.05$, **Figure 6G**).

Nomogram model construction and prediction efficacy validation

The training set included 477 patients with breast cancer, and we performed both univariate and multivariate Cox regression analyses based on clinicopathological characteristics including *MACF1* mutations (group), *MACF1* expression, age, and gender and so on. Univariate analysis revealed that *MACF1* mutations (group), ER status, N/M stage, and tumor stage were significantly correlated with patient survival (**Figure 7A**; Supplementary Table 5). In the multivariate Cox regression model, Black or African American race and N3 stage were identified as prognostic factors (**Figure 7B**; Supplementary Table 6). Therefore, a nomogram was constructed to integrate *MACF1* mutations (group) and other clinicopathological variables, such as ER status, N/M stage, and tumor stage (**Figure 7C**) [37]. Thereafter, the prediction efficacy of this model was determined. Based on the results, Calibration plots for the overall survival (OS) at 1, 3, and 5 years showed good agreement with actual observations (**Figure 7D-F**). Using time-dependent ROC analysis, the prediction efficiencies of 1-, 3-, and 5-year OS were 82.7%, 78%, and 73%, respectively (**Figure 7G**). The DCA for the prognostic model was performed, showing a good net benefit (**Figure 7H**). The prognostic model was validated by using the testing set (478 breast cancer patients). The results showed that OS was significantly shorter for patients with *MACF1* mutations compared to patients without *MACF1* mutations both for the training and the testing sets (**Figure 8A**, **8B**). Another nomogram, comprising clinicopathological variables selected from the training set, was built by the testing set (**Figure 8C**). According to the results, the prognostic model had a good predictive effect on OS at 1, 3, and 5 years in breast cancer patients in the testing set (**Figure 8D-F**). The prediction efficiencies of 1-, 3-, and 5-year OS in patients with breast cancer in the testing set were 83.7%, 78.7%, and 76.3%, respectively (**Figure 8G**). DCA demonstrated that the net benefit of the model for breast cancer patients in the testing set was significant (**Figure 8H**).

Discussion

Breast cancer has the highest incidence and mortality rates among women worldwide [38]. Despite improved screening, diagnosis, and therapy, prognosis for breast cancer patients remains poor [39]. Owing to its relatively high immunogenicity and mutational load, immunotherapy has been considered the best therapeutic choice for breast cancer [40]. Nevertheless, the efficacy of breast cancer immunotherapy has not been efficiently predicted owing to the lack of reliable biomarkers. In this regard, aberrant expression of *MACF1* was found significantly associated with the pathogenesis, progression, and metastasis of breast


A
Composition
1.00
0.75
0.50
0.25
0.00
B cells memory
B cells naive
Dendritic cells activated
Dendritic cells resting
Eosinophils
Macrophages M0
Macrophages M1
Macrophages M2
Mast cells activated
Mast cells resting
Monocytes
Neutrophils
NK cells activated
NK cells resting
Plasma cells
T cells CD4 memory activated
T cells CD4 memory resting
T cells CD4 naive
T cells CD8
T cells follicular helper
T cells gamma delta
T cells regulatory (Tregs)
B
Group mut wt
Infiltration Abundance
p=8.9e-03
p=1.5e-02
p=6.1e-09
p=3.5e-09
p=1.6e-07
p=2.2e-03
p=5.7e-03
p=1.0e-05
p=6.5e-03
p=6.6e-12
p=3.2e-06
p=3.5e-06
p=4.6e-03
C

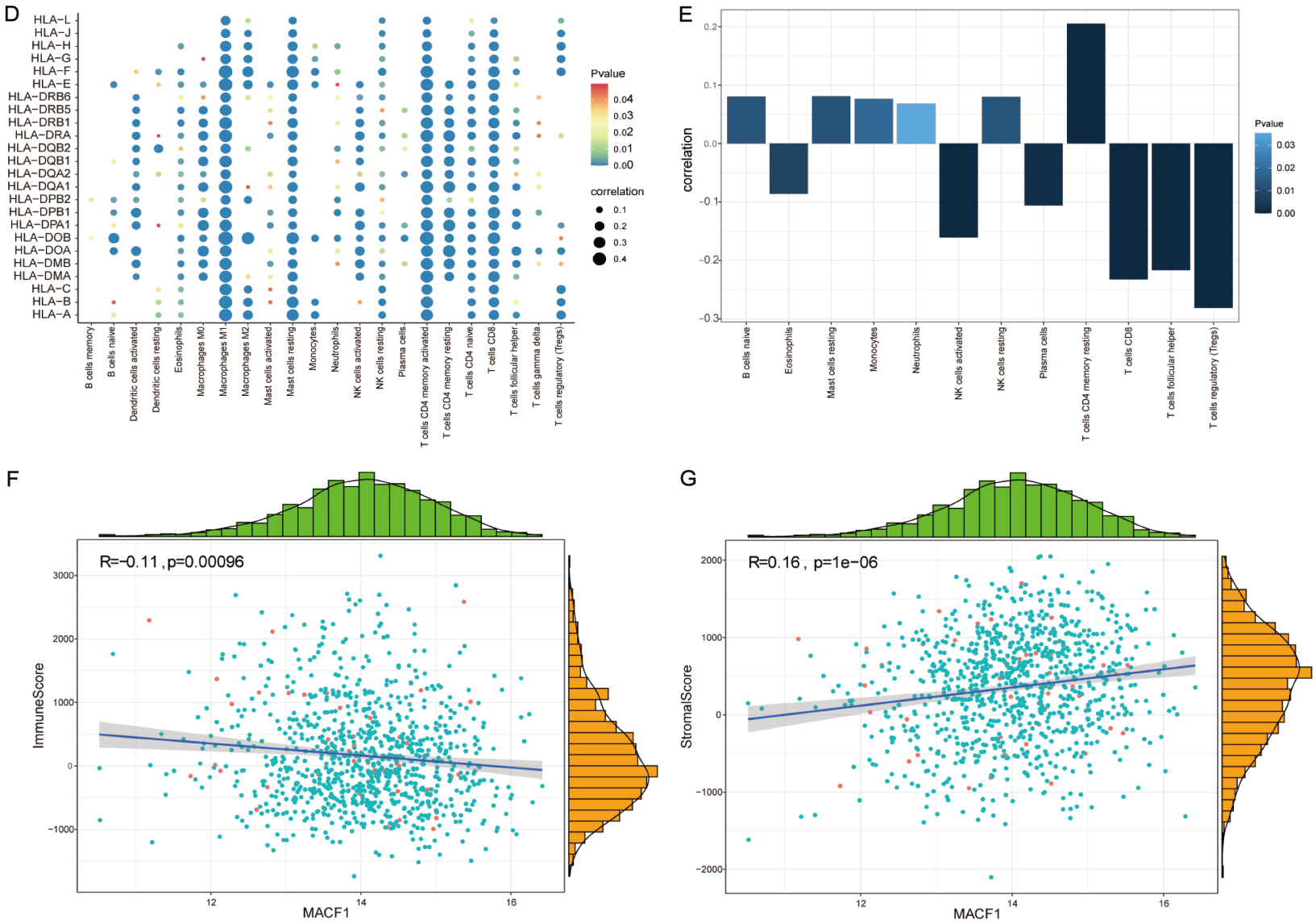


Figure 6. The association between *MACF1* mutations and immune infiltration. A. Distribution map of immune cells of breast cancer patients with or without *MACF1* mutations. The X-axis represents patient ID, and the Y-axis represents immune cells proportion. B. Difference in immune cell content between the *MACF1*-MT and *MACF1*-WT groups. Blue represents *MACF1*-MT patients and green represents *MACF1*-WT patients. C. Correlation heat map of immune cell types. D. Correlations between 22 immunocytes and human leukocyte antigen (HLA) family genes. E. Correlations between *MACF1* gene expression level and immune cell types. F. Correlation between *MACF1* gene expression level and immune score. G. Correlation between *MACF1* gene expression level and stromal score.

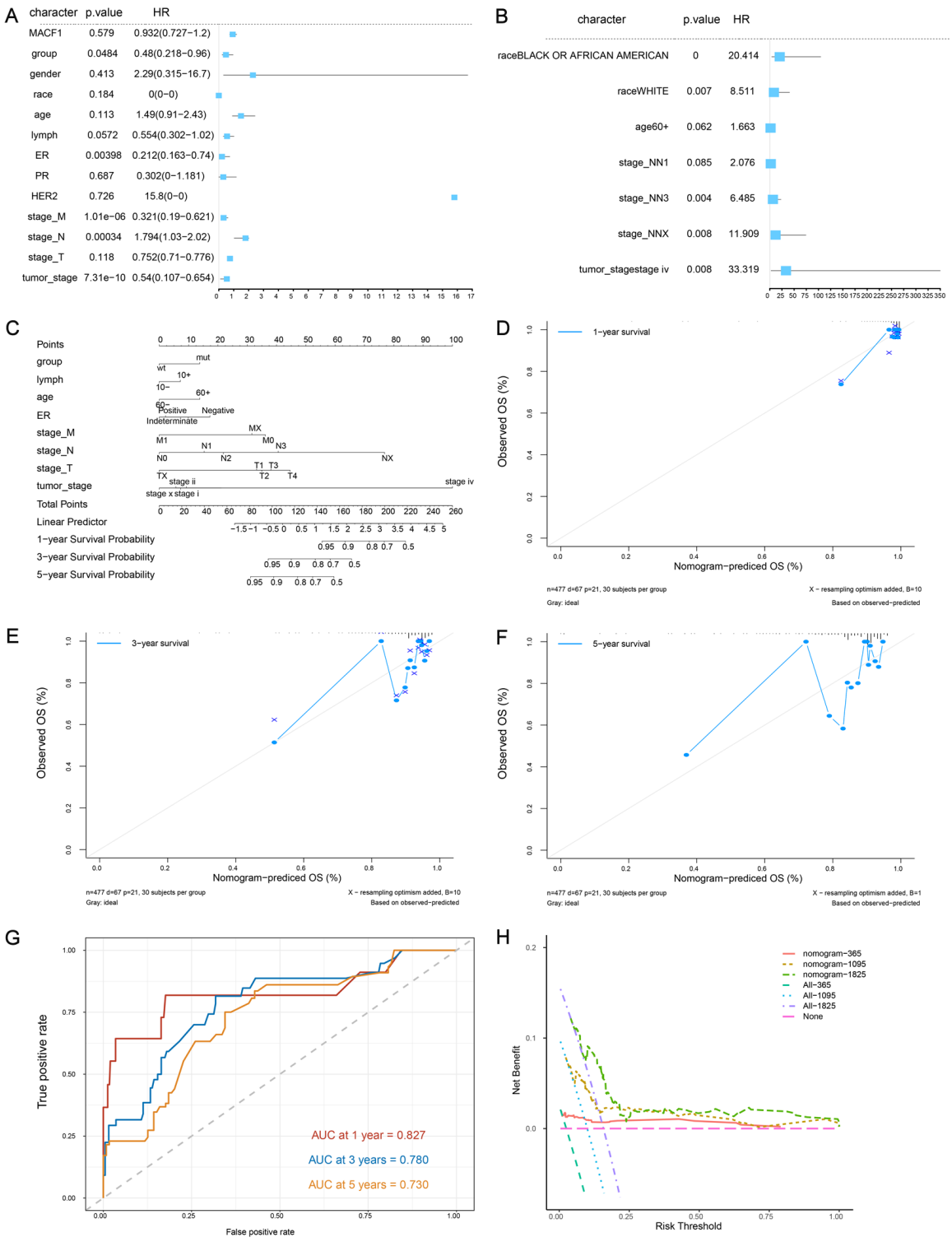


Figure 7. *MACF1* mutation-based nomogram construction and performance validation in patients with breast cancer in training set. A. Cox univariate regression. B. Cox multivariate regression. C. Nomogram for the prediction of the breast cancer patients' survival probability at 1, 3, and 5 years. D-F. Representative calibration plots showing the prediction of the overall survival (OS) at 1, 3, and 5 years using the nomogram. G. Time-dependent receiver operating characteristic (ROC) curve analysis of the nomogram. H. The Decision Curve Analysis (DCA) of the nomogram.

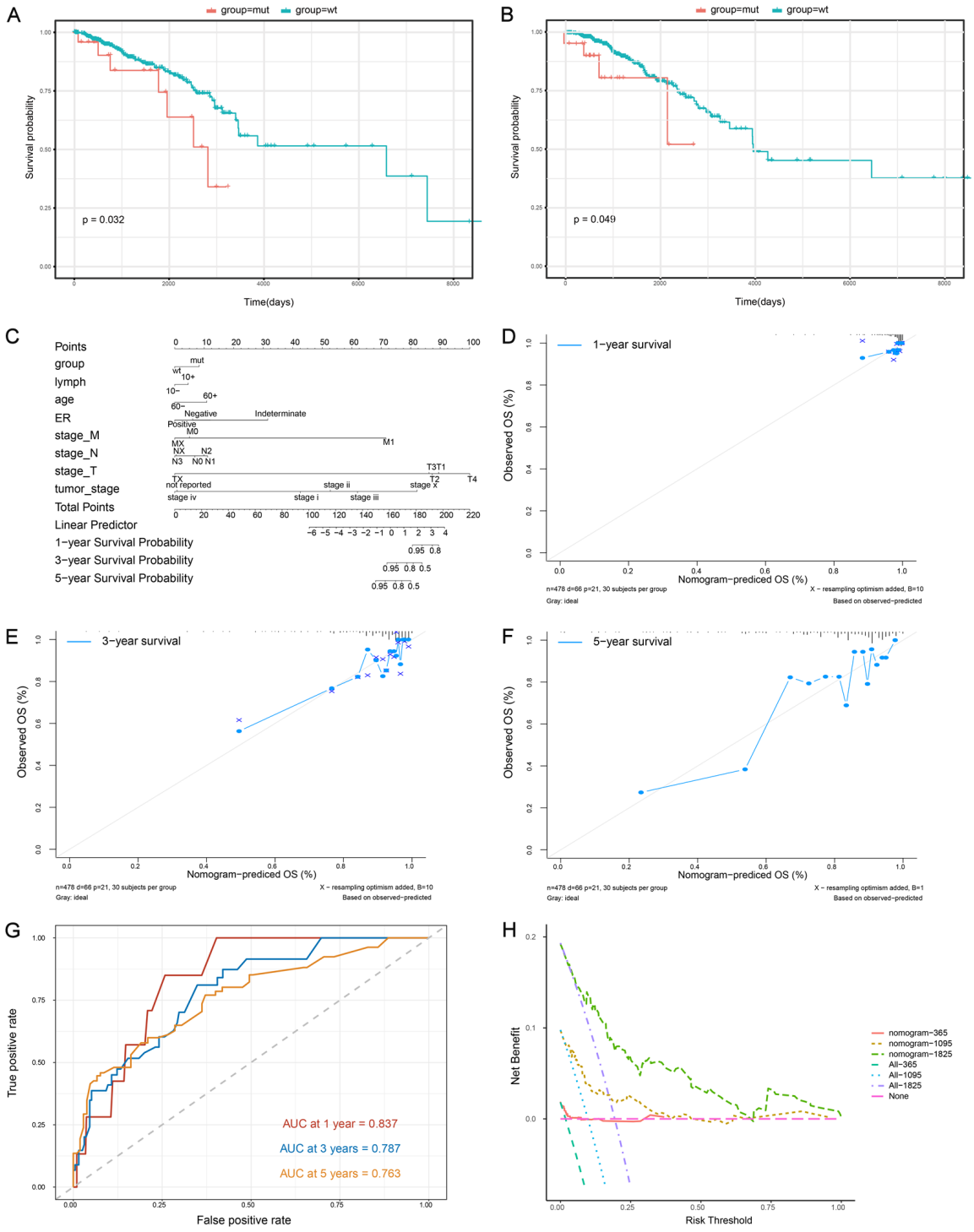


Figure 8. *MACF1* mutation-based nomogram construction and performance validation in patients with breast cancer in testing set. A. The OS of patients with or without *MACF1* mutations was analyzed in training set. B. The OS of patients with or without *MACF1* mutations was analyzed in testing set. C. Nomogram for the prediction of the breast cancer patients' survival probability at 1, 3, and 5 years. D-F. Representative calibration plots showing the prediction of the OS at 1, 3, and 5 years using the nomogram. G. Time-dependent ROC curve analysis of the nomogram. H. The DCA of the nomogram.

cancer [41]. The underlying mechanisms of *MACF1* mutations in tumor progression and TME remain largely unknown. However, a worse prognosis was found for patients with *MACF1*

mutations compared to patients without *MACF1* mutations.

In this study, the mutation frequency of *MACF1* was found to be relatively high in breast cancer patients and *MACF1* was neither amplified nor deleted. Further, *MACF1*-MT patients had higher TMB scores than *MACF1*-WT patients. As a result, we opted to compare the immune cell content between patients, which revealed that patients with *MACF1* mutations had a significantly higher immune-cell content than *MACF1*-WT patients. Such findings indicate that mutations in *MACF1* may activate immune cells. Therefore, owing to the higher mutation load, *MACF1*-MT patients may present altered immunogenicity. Most immune cell types were found to be negatively correlated with the *MACF1*-MT group, suggesting that the immunotherapeutic response in these patients might be unsatisfactory.

A total of the 22 types of immune cells were investigated. MACF1 expression was found to be positively correlated with naïve B cells, resting mast cells, monocytes, resting memory $CD4^+$ T cells, neutrophils, and resting NK cells, and negatively correlated with activated NK cells, $CD8^+$ T cells, Tregs, and follicular helper T cells. Further, activated and resting mast cells, M1 and M2 macrophages, and activated memory $CD4^+$ and $CD8^+$ T cells were found to be significantly correlated with multiple HLA family genes. Consequently, the overexpression of *MACF1* may upregulate immunosuppressive immune cells and molecules in the tumor immune microenvironment. By quantifying immune and stromal cell abundance, significant negative and positive correlations were observed between the expression of *MACF1* and immune and stromal scores, respectively. Gene overexpression is generally caused by amplification or mutation [42]. Therefore, the overexpression of *MACF1* in breast cancer may arise from mutations rather than amplification. However, further research is required to test this hypothesis.

MACF1 mutations upregulated hallmark glycolysis, hallmark mTORC1 signaling, KEGG cysteine and methionine metabolism, KEGG galactose metabolism, and KEGG glycolysis gluconeogenesis, and downregulated KEGG histidine metabolism. Previous studies have shown that cancer cells do not use mitochondrial oxidative phosphorylation to generate energy, even under aerobic conditions. Instead, these cells use aerobic glycolysis, defined as the Warburg effect, which provides growth and proliferation advantages to cancer cells [43]. Owing to the frequent activation of the PI3K/AKT/mTOR pathway in breast cancer, PI3K, mTORC1, and mTORC2 can be exploited as potential therapeutic targets [44, 45]. Thus, *MACF1* mutations may promote the growth, migration, and invasion of tumor cells, leading to poorer immunotherapeutic results. Therefore, novel clinical strategies for breast cancer therapy need to be explored.

The regulatory effects of *MACF1* mutations were investigated and identified by analyzing the drug sensitivity data. The sensitivity of breast cancer cells to 13 drugs, such as GSK-1059615 (PI3K inhibitor), rapamycin (mTOR inhibitor), and Mitomycin-C (chemotherapeutic drug), was enhanced in *MACF1*-MT patients compared to *MACF1*-WT patients. Thus, *MACF1* mutations might lead to a worse prognosis, which may be due to upregulation of the mTOR signaling pathway, and modifications in energy metabolism and within the tumor immune microenvironment. Based on these observations, we hypothesized that *MACF1*-MT patients would gain more benefits from an immunotherapeutic combination of PI3K and mTOR inhibitors.

DEGs were identified in *MACF1*-MT patients. Additionally, GO and KEGG functional enrichment analyses were performed to further reveal the functional implications of *MACF1* mutations within the 911 DEGs. GO and KEGG functional enrichment analyses revealed that these DEGs were mainly related to signal transduction and substance metabolism. In the present study, *MACF1* mutations were found to be associated with graft-versus-host disease and the processing and presentation of antigens, drug metabolism cytochrome P450, NK cell-mediated cytotoxicity, cell cycle, and retinol metabolism by GSEA. Of note, these biological pathways are primarily related to immunological functions. By constructing a PPI network, DEGs were recognized to be enriched in immune and inflammatory responses. According to previous studies, inflammatory responses can promote tumor progression and are a hallmark of numerous cancers [46-48]. Finally,

DEGs were used to construct a ceRNA network, which lays the foundation for more in-depth experimental research to identify the underlying mechanisms of breast cancer.

For more accurate prognostic prediction, we constructed a nomogram based on *MACF1* mutations and other important parameters in the training set. According to the calibration plots, we found favorable predictive values for the OS of patients at 1-, 3-, and 5-year follow-ups. Further, the time-dependent ROC curves performed well for the prediction of patient 1-, 3-, and 5-year OS rates. The DCA for the prognostic model revealed a good net benefit. Furthermore, there were similar results in the testing set. Further research is needed to validate the nomogram for use in clinical practice.

Although our research revealed the relationship between *MACF1* mutations and breast cancer, some limitations are worth noting. First, the cohort data contained a small sample of patients despite the use of TCGA-Breast cancer cohort data, and we need an external dataset to validate our results. Second, the number of *MACF1*-WT patients differed from that of the *MACF1*-MT patients. Future studies should balance sample sizes across these groups. Third, the caveats that are inherent to retrospective studies should be validated in prospective studies. Nonetheless, we propose that immunotherapy with PI3K and mTOR inhibitors might achieve more beneficial outcomes in *MACF1*-MT patients; however, further experimental and clinical verification is required. More mechanistic studies are also needed to clarify the signal transduction and molecular mechanisms of *MACF1* mutations in breast cancer.

In conclusion, *MACF1* mutations may predict poor prognosis in breast cancer patients and may regulate signaling pathways, energy metabolism, and the tumor immune microenvironment. Further, *MACF1* mutations in breast cancer were significantly associated with the sensitivity of breast cancer cells to 13 drugs, such as GSK1059615, rapamycin, and mitomycin-C. Our findings suggest that *MACF1* mutations may serve as a potential prognostic biomarker and a therapeutic target in breast cancer.

Acknowledgements

We are grateful to the TCGA Research Network, the GDSC data portal, and the authors for making their valuable research results public.

Disclosure of conflict of interest

None.

Address correspondence to: Juan Wang, Department of Blood Transfusion, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Jiefang Avenue 1095, Wuhan 430030, Hubei, China. E-mail: juanwangsxk@tjh.tjmu.edu.cn

References

[1] Huo D, Hu H, Rhie SK, Gamazon ER, Cherniack AD, Liu J, Yoshimatsu TF, Pitt JJ, Hoadley KA, Troester M, Ru Y, Lichtenberg T, Sturtz LA, Shelley CS, Benz CC, Mills GB, Laird PW, Shriver CD, Perou CM and Olopade OI. Comparison of breast cancer molecular features and survival by African and European Ancestry in the Cancer Genome Atlas. JAMA Oncol 2017; 3: 1654-1662.

[2] Chaffer CL and Weinberg RA. A perspective on cancer cell metastasis. Science 2011; 331: 1559-1564.

[3] Peart O. Metastatic breast cancer. Radiol Technol 2017; 88: 519M-539M.

[4] Early Breast Cancer Trialists' Collaborative Group (EBCTCG), Davies C, Godwin J, Gray R, Clarke M, Cutter D, Darby S, McGale P, Pan HC, Taylor C, Wang YC, Dowsett M, Ingle J and Peto R. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. Lancet 2011; 378: 771-784.

[5] Kwan TT, Bardia A, Spring LM, Giobbie-Hurder A, Kalinich M, Dubash T, Sundaresan T, Hong X, LiCausi JA, Ho U, Silva EJ, Wittner BS, Sequist LV, Kapur R, Miyamoto DT, Toner M, Haber DA and Maheswaran S. A digital RNA signature of circulating tumor cells predicting early therapeutic response in localized and metastatic breast cancer. Cancer Discov 2018; 8: 1286-1299.

[6] Fuchs E and Yang Y. Crossroads on cytoskeletal highways. Cell 1999; 98: 547-550.

[7] Hu L, Su P, Li R, Yin C, Zhang Y, Shang P, Yang T and Qian A. Isoforms, structures, and functions of versatile spectraplakin MACF1. BMB Rep 2016; 49: 37-44.

[8] Hu L, Xiao Y, Xiong Z, Zhao F, Yin C, Zhang Y, Su P, Li D, Chen Z, Ma X, Zhang G and Qian A. MACF1, versatility in tissue-specific function and in human disease. Semin Cell Dev Biol 2017; 69: 3-8.

[9] Chen C, Shi C, Huang X, Zheng J, Zhu Z, Li Q, Qiu S, Huang Z, Zhuang Z, Wu R, Liu P, Wu F, Lin S, Li B, Zhang X and Chen Q. Molecular pro-

files and metastasis markers in Chinese patients with gastric carcinoma. Sci Rep 2019; 9: 13995.

[10] Sjöblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JK, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW and Velculescu VE. The consensus coding sequences of human breast and colorectal cancers. Science 2006; 314: 268-274.

[11] Zaoui K, Benseddik K, Daou P, Salaun D and Badache A. ErbB2 receptor controls microtubule capture by recruiting ACF7 to the plasma membrane of migrating cells. Proc Natl Acad Sci U S A 2010; 107: 18517-18522.

[12] Yucel G and Oro AE. Cell migration: GSK3beta steers the cytoskeleton's tip. Cell 2011; 144: 319-321.

[13] Tomczak K, Czerwinska P and Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn) 2015; 19: A68-77.

[14] Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, Bhai J, Billis K, Boddu S, Charkhchi M, Cummins C, Da Rin Fioretto L, Davidson C, Dodiya K, El Houdaigui B, Fatima R, Gall A, Garcia Giron C, Grego T, Guijarro-Clarke C, Haggerty L, Hemrom A, Hourlier T, Izuogu OG, Juettemann T, Kaikala V, Kay M, Lavidas I, Le T, Lemos D, Gonzalez Martinez J, Marugán JC, Maurel T, McMahon AC, Mohanan S, Moore B, Muffato M, Oheh DN, Paraschas D, Parker A, Parton A, Prosovetskaia I, Sakthivel MP, Salam AIA, Schmitt BM, Schuilenburg H, Sheppard D, Steed E, Szpak M, Szuba M, Taylor K, Thormann A, Threadgold G, Walts B, Winterbottom A, Chakiachvili M, Chaubal A, De Silva N, Flint B, Frankish A, Hunt SE, Ilsley GR, Langridge N, Loveland JE, Martin FJ, Mudge JM, Morales J, Perry E, Ruffier M, Tate J, Thybert D, Trevanion SJ, Cunningham F, Yates AD, Zerbino DR and Flicek P. Ensembl 2021. Nucleic Acids Res 2021; 49: D884-D891.

[15] Mayakonda A, Lin DC, Assenov Y, Plass C and Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res 2018; 28: 1747-1756.

[16] Skidmore ZL, Wagner AH, Lesurf R, Campbell KM, Kunisaki J, Griffith OL and Griffith M. GenVisR: genomic visualizations in R. Bioinformatics 2016; 32: 3012-3014.

[17] Guo X, Zhang B, Zeng W, Zhao S and Ge D. G3viz: an R package to interactively visualize genetic mutation data using a lollipop-diagram. Bioinformatics 2020; 36: 928-929.

[18] Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A and Peters S. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. Ann Oncol 2019; 30: 44-56.

[19] Hile SE, Shabashev S and Eckert KA. Tumor-specific microsatellite instability: do distinct mechanisms underlie the MSI-L and EMAST phenotypes? Mutat Res 2013; 743-744: 67-77.

[20] Thul PJ and Lindskog C. The human protein atlas: a spatial map of the human proteome. Protein Sci 2018; 27: 233-244.

[21] Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C and Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 2013; 6: pl1.

[22] Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D, Reddy A, Liu M, Murray L, Berger MF, Monahan JE, Morais P, Meltzer J, Korejwa A, Jané-Valbuena J, Mapa FA, Thibault J, Bric-Furlong E, Raman P, Shipway A, Engels IH, Cheng J, Yu GK, Yu J, Aspesi P Jr, de Silva M, Jagtap K, Jones MD, Wang L, Hatton C, Palescandolo E, Gupta S, Mahan S, Sougnez C, Onofrio RC, Liefeld T, MacConaill L, Winckler W, Reich M, Li N, Mesirov JP, Gabriel SB, Getz G, Ardlie K, Chan V, Myer VE, Weber BL, Porter J, Warmuth M, Finan P, Harris JL, Meyerson M, Golub TR, Morrissey MP, Sellers WR, Schlegel R and Garraway LA. Addendum: the cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 2019; 565: E5-E6.

[23] Maeser D, Gruener RF and Huang RS. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. Brief Bioinform 2021; 22: bbab260.

[24] Hanzelmann S, Castelo R and Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 2013; 14: 7.

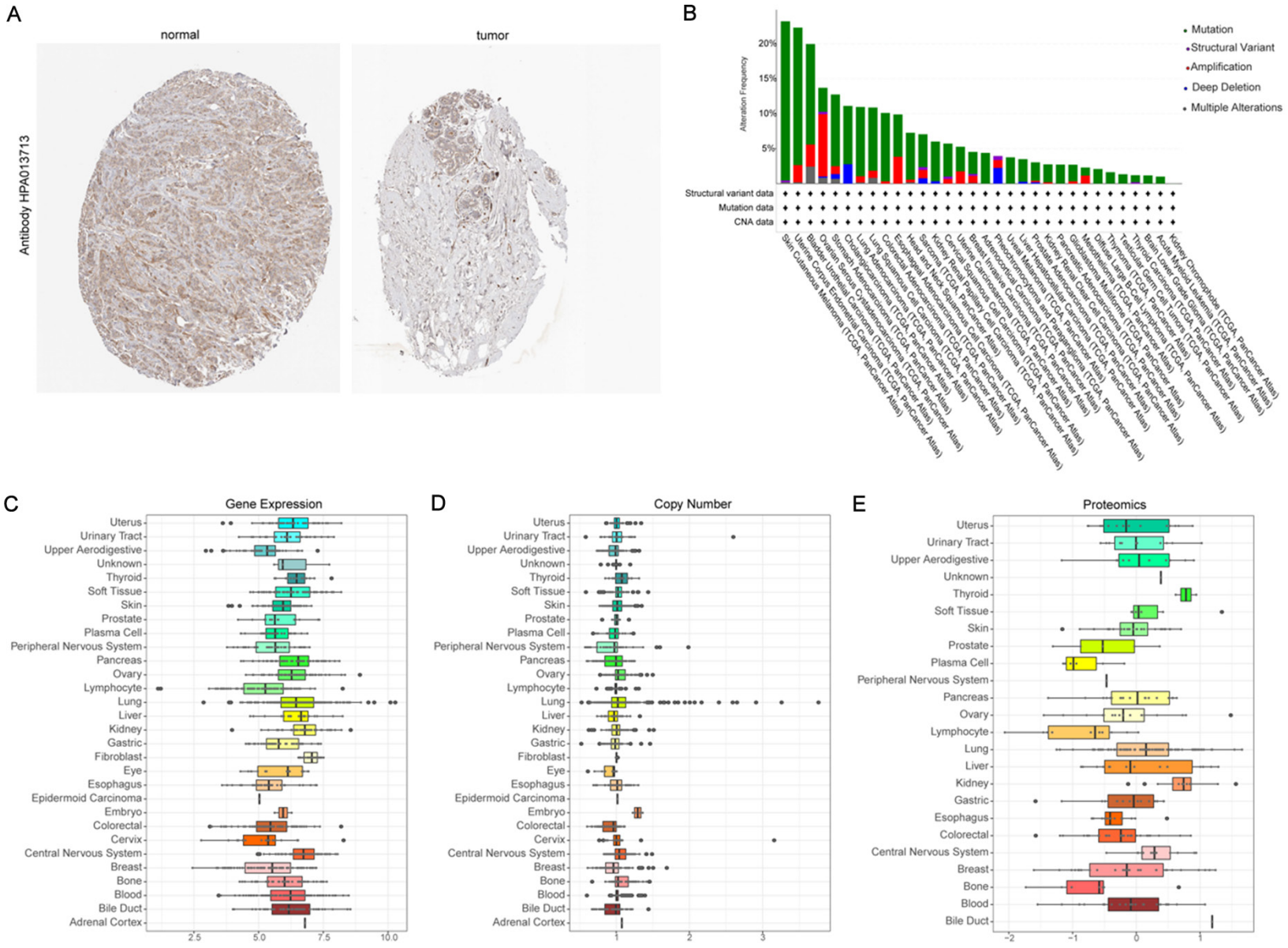
[25] Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP and Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. Cell Syst 2015; 1: 417-425.

[26] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM and Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000; 25: 25-29.

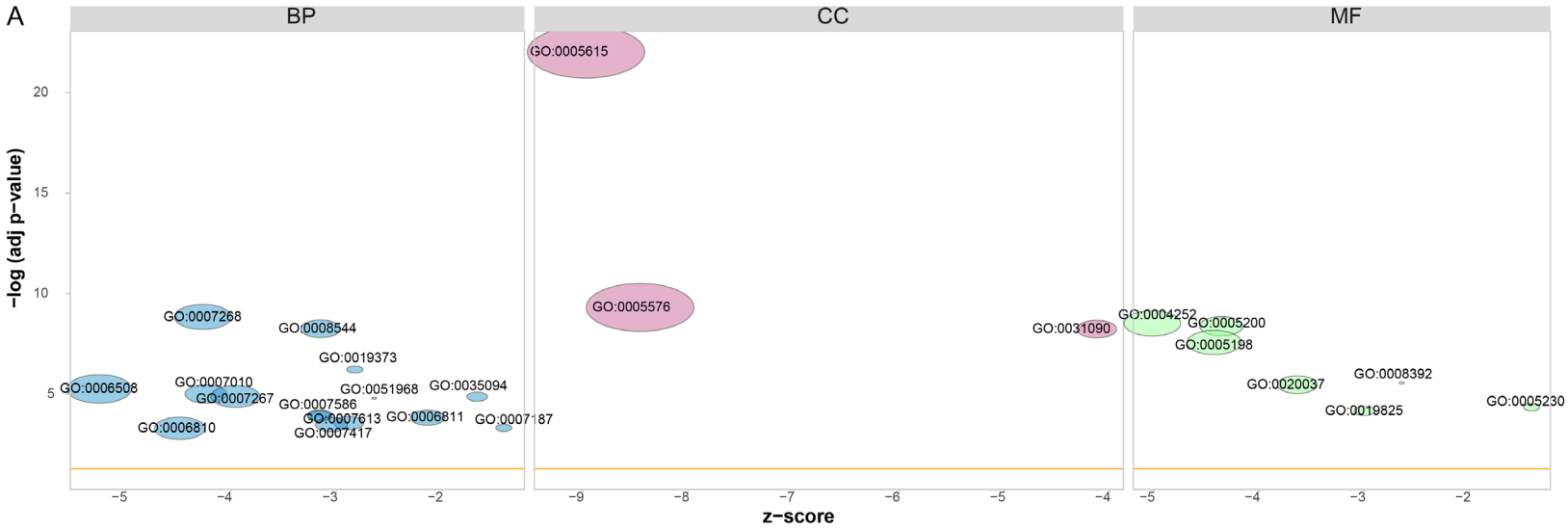
[27] Kanehisa M and Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000; 28: 27-30.
[28] Yu G, Wang LG, Han Y and He QY. ClusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012; 16: 284-287.
[29] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES and Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005; 102: 15545-15550.
[30] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ and Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019; 47: D607-D613.
[31] Bandettini WP, Kellman P, Mancini C, Booker OJ, Vasu S, Leung SW, Wilson JR, Shanbhag SM, Chen MY and Arai AE. Multicontrast delayed enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study. J Cardiovasc Magn Reson 2012; 14: 83.
[32] Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pages F, Trajanoski Z and Galon J. ClueGO: a cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 2009; 25: 1091-1093.
[33] Chang L, Zhou G, Soufan O and Xia J. MiRNet 2.0: network-based visual analytics for miRNA functional analysis and systems biology. Nucleic Acids Res 2020; 48: W244-W251.
[34] Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB and Verhaak RG. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 2013; 4: 2612.
[35] Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, Khodadoust MS, Esfahani MS, Luca BA, Steiner D, Diehn M and Alizadeh AA. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol 2019; 37: 773-782.
[36] Durisova M and Dedik L. SURVIVAL--an integrated software package for survival curve estimation and statistical comparison of survival rates of two groups of patients or experimental animals. Methods Find Exp Clin Pharmacol 1993; 15: 535-540.
[37] Gittleman H, Sloan AE and Barnholtz-Sloan JS. An independently validated survival nomogram for lower-grade glioma. Neuro Oncol 2020; 22: 665-674.
[38] Zhang Y, Dian L, Wei X, Huang J, Sun Y, Song X, Yang C, Kang M, Ou A, Chen Q and Xu R. Physicians' attitudes towards reproduction in young patients with early breast cancer in China. Breast Cancer Res Treat 2020; 184: 567-583.
[39] Wang H, Xiao Y, Wu L and Ma D. Comprehensive circular RNA profiling reveals the regulatory role of the circRNA-000911/miR-449a pathway in breast carcinogenesis. Int J Oncol 2018; 52: 743-754.
[40] Clark NM, Martinez LM, Murdock S, deLigio JT, Olex AL, Effi C, Dozmorov MG and Bos PD. Regulatory T cells support breast cancer progression by opposing ifn-gamma-dependent functional reprogramming of myeloid cells. Cell Rep 2020; 33: 108482.
[41] Quick Q, Paul M and Skalli O. Roles and potential clinical applications of intermediate filament proteins in brain tumors. Semin Pediatr Neurol 2015; 22: 40-48.
[42] Lin G, Yin G, Yan Y and Lin B. Identification of prognostic biomarkers for malignant melanoma using microarray datasets. Oncol Lett 2019; 18: 5243-5254.
[43] Hanahan D and Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011; 144: 646-674.
[44] Kim SB, Dent R, Im SA, Espie M, Blau S, Tan AR, Isakoff SJ, Oliveira M, Saura C, Wongchenko MJ, Kapp AV, Chan WY, Singel SM, Maslyar DJ and Baselga J; LOTUS investigators. Ipatasertib plus paclitaxel versus placebo plus paclitaxel as first-line therapy for metastatic triple-negative breast cancer (LOTUS): a multicentre, randomised, double-blind, placebo-controlled, phase 2 trial. Lancet Oncol 2017; 18: 1360-1372.
[45] Dolly SO, Wagner AJ, Bendell JC, Kindler HL, Krug LM, Seiwert TY, Zauderer MG, Lolkema MP, Apt D, Yeh RF, Fredrickson JO, Spoerke JM, Koeppen H, Ware JA, Lauchle JO, Burris HA 3rd and de Bono JS. Phase I study of apitolisib (GDC-0980), dual phosphatidylinositol-3-kinase and mammalian target of rapamycin kinase inhibitor, in patients with advanced solid tumors. Clin Cancer Res 2016; 22: 2874-2884.
[46] Greten FR and Grivennikov SI. Inflammation and cancer: triggers, mechanisms, and consequences. Immunity 2019; 51: 27-41.
[47] Hamarsheh S, Gross O, Brummer T and Zeiser R. Immune modulatory effects of oncogenic KRAS in cancer. Nat Commun 2020; 11: 5439.
[48] Brown JM, Recht L and Strober S. The promise of targeting macrophages in cancer therapy. Clin Cancer Res 2017; 23: 3241-3250.

Supplementary Table 1. Associations between Microtubule actin cross-linking factor 1 (*MACF1*) mutations and clinicopathologic features

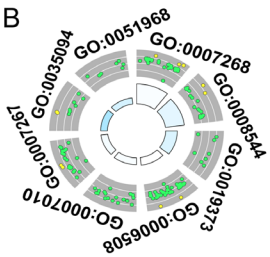
	MACF1-MT	MACF1-WT	p	test
sample	50	905		
MACF1 = low (%)	25 (50.0)	433 (47.8)	0.773	exact
group = wt (%)	0 (0.0)	905 (100.0)	< 0.001	exact
gender = MALE (%)	0 (0.0)	11 (1.2)	1	exact
race (%)			0.357	exact
AMERICAN INDIAN OR ALASKA NATIVE	0 (0.0)	1 (0.1)		
ASIAN	2 (4.0)	53 (5.9)		
BLACK OR AFRICAN AMERICAN	8 (16.0)	151 (16.7)		
WHITE	34 (68.0)	630 (69.6)		
age = 60+ (%)	22 (44.0)	410 (45.3)	0.885	exact
lymph = 10+ (%)	37 (74.0)	737 (81.4)	0.196	exact
ER (%)			0.012	exact
Indeterminate	0 (0.0)	2 (0.2)		
Negative	20 (40.0)	190 (21.0)		
Positive	27 (54.0)	673 (74.4)		
PR (%)			0.252	exact
Indeterminate	0 (0.0)	3 (0.3)		
Negative	21 (42.0)	275 (30.4)		
Positive	26 (52.0)	586 (64.8)		
HER2 (%)			0.907	exact
Equivocal	6 (12.0)	150 (16.6)		
Indeterminate	0 (0.0)	11 (1.2)		
Negative	26 (52.0)	457 (50.5)		
Positive	8 (16.0)	137 (15.1)		
stage_M (%)			0.024	exact
M0	42 (84.0)	760 (84.0)		
M1	4 (8.0)	17 (1.9)		
MX	4 (8.0)	128 (14.1)		
stage_N (%)			0.008	exact
N0	28 (56.0)	419 (46.3)		
N1	11 (22.0)	310 (34.3)		
N2	3 (6.0)	109 (12.0)		
N3	4 (8.0)	55 (6.1)		
NX	4 (8.0)	12 (1.3)		
stage_T (%)			0.242	exact
T1	9 (18.0)	245 (27.1)		
T2	33 (66.0)	524 (57.9)		
T3	4 (8.0)	102 (11.3)		
T4	4 (8.0)	31 (3.4)		
TX	0 (0.0)	3 (0.3)		
tumor_stage (%)			0.016	exact
stage I	7 (14.0)	156 (17.2)		
stage II	28 (56.0)	518 (57.2)		
stage III	8 (16.0)	199 (22.0)		
stage IV	4 (8.0)	15 (1.7)		
stage X	2 (4.0)	7 (0.8)		



Supplementary Figure 1. Expression analysis of MACF1 from databases. A. MACF1 protein expression levels in breast cancer tissues and in normal breast tissues. B. The MACF1 mutation status analysis in various types of cancer. C. The gene expression levels of MACF1 in various types of cancer cell lines. D. The copy number variation (CNV) analysis of MACF1 in various types of cancer cell lines. E. The protein expression levels of MACF1 in various types of cancer cell lines.

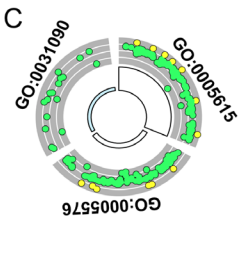


B



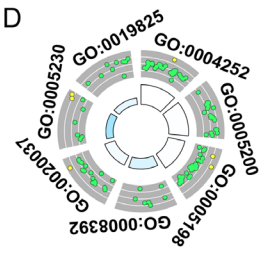
ID	Description
GO:0007268	chemical synaptic transmission
GO:0008544	epidermis development
GO:0019373	epoxygenase P450 pathway
GO:0006508	proteolysis
GO:0007010	cytoskeleton organization
GO:0007267	cell−cell signaling
GO:0035094	response to nicotine
GO:0051968	positive regulation of synaptic transmission, glutar

C

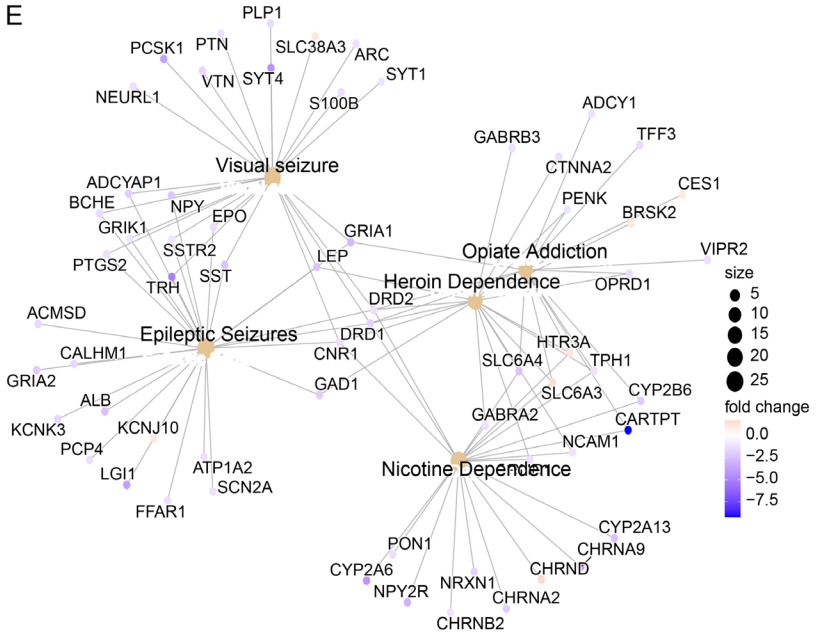


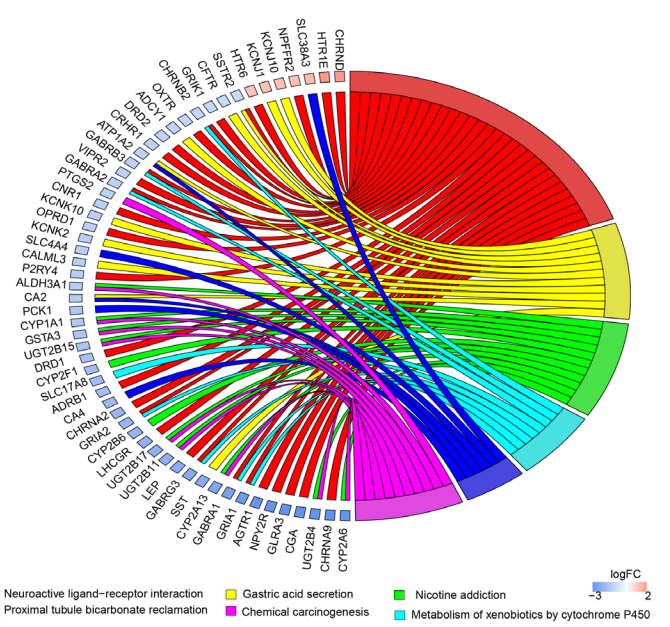
ID	Description
GO:0005615	extracellular space
GO:0005576	extracellular region
GO:0031090	organelle membrane

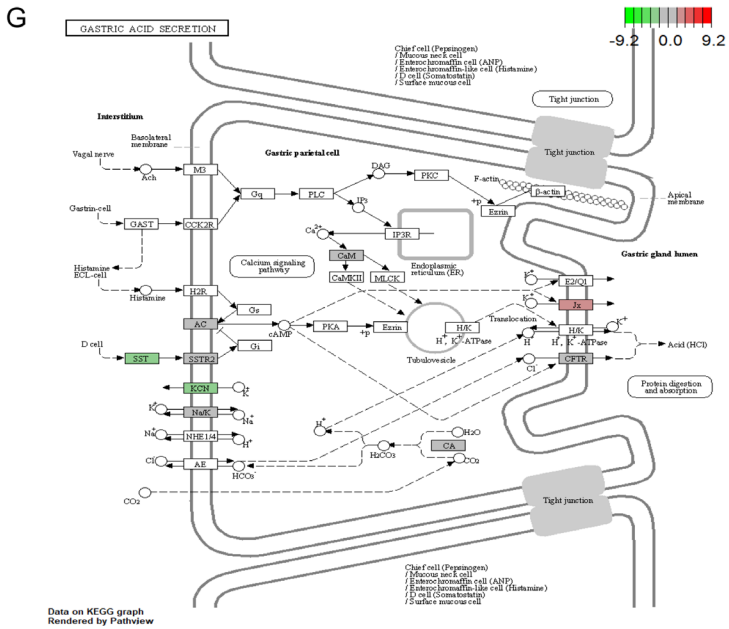
D

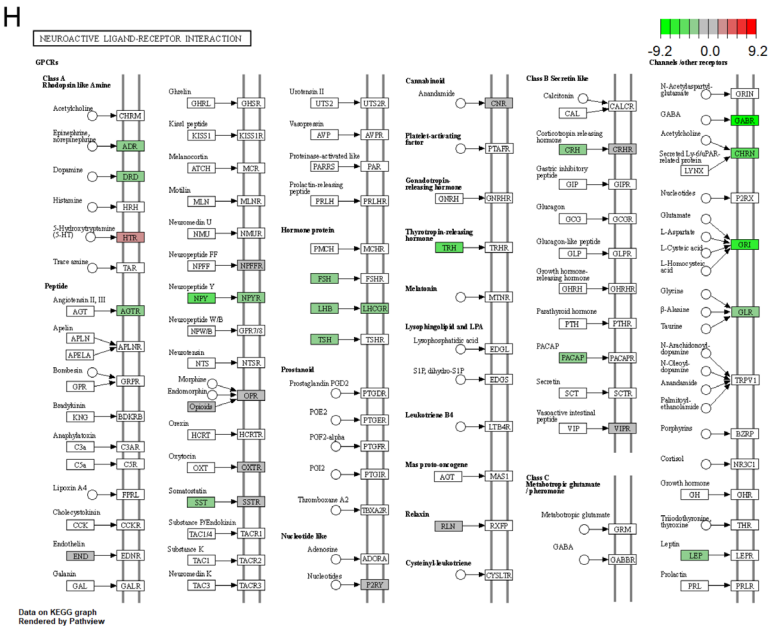


ID	Description
GO:0004252	serine−type endopeptidase activity
GO:0005200	structural constituent of cytoskeleton
GO:0005198	structural molecule activity
GO:0008392	arachidonic acid epoxygenase activity
GO:0020037	heme binding
GO:0005230	extracellular ligand−gated ion channel activity
GO:0019825	transcriptional activator activity


E
PLP1
PCSK1
PTN
SLC38A3
ARC
SYT1
VTN
SYT4
NEURL1
S100B
Visual seizure
ADCYAP1
BCHE
NPY
EPO
GRIK1
SSTR2
PTGS2
TRH
SST
GRIA1
LEP
ACMSD
Epileptic Seizures
CALHM1
GRIA2
ALB
KCNK3
KCNJ10
PCP4
LGI1
ATP1A2
SCN2A
FFAR1
DRD1
CNR1
GAD1
ADCY1
GABRB3
TFF3
CTNNA2
PENK
CES1
BRSK2
Opiate Addiction
VIPR2
Heroin Dependence
OPRD1
DRD2
HTR3A
SLC6A4
TPH1
SLC6A3
CYP2B6
GABRA2
CARTPT
NCAM1
Nicotine Dependence
CYP2A13
CHRNA9
PON1
CYP2A6
CHRND
NPY2R
NRXN1
CHRNA2
CHRNB2
size
5
10
15
20
25
fold change
0.0
−2.5
−5.0
−7.5


F
CHRND
HTR1E
SLC38A3
NPFFR2
KCNJ10
KCNJ1
HTR6
SSTR2
CTTR
GRIK1
CHRNB2
OXTR
ADCY1
DRD2
CRHR1
ATP1A2
GABRB3
VIPR2
GABRA2
PTGS2
CNR1
KCNK10
OPRD1
KCNK2
SLC4A4
CALML3
P2RY4
ALDH3A1
CA2
PCK1
CYP1A1
GSTA3
UGT2B15
DRD1
CYP2F1
SLC17A8
ADRB1
CA4
CHRNA2
GRIA2
CYP2B6
LHCGR
UGT2B17
UGT2B11
LEP
GABRG3
SST
CYP2A13
GABRA1
GRIA1
AGTR1
NPY2R
GLRA3
CGA
UGT2B4
CHRNA9
CYP2A6
KEGG
Neuroactive ligand−receptor interaction
Proximal tubule bicarbonate reclamation
Gastric acid secretion
Chemical carcinogenesis
Nicotine addiction
Metabolism of xenobiotics by cytochrome P450
logFC
−3
2


G
GASTRIC ACID SECRETION
-9.2
0.0
9.2
Interstitium
Basolateral membrane
Vagal nerve
Gastrin-cell
Histamine
ECL-cell
D cell
Gastric parietal cell
Gastric gland lumen
Apical membrane
Tight junction
Calcium signaling pathway
Endoplasmic reticulum (ER)
Translocation
Tubulovesicle
Acid (HCl)
Protein digestion and absorption
Data on KEGG graph
Rendered by Pathview


H
NEUROACTIVE LIGAND-RECEPTOR INTERACTION
-9.2
0.0
9.2
GPCRs
Class A Rhodopsin-like
Class B Secretin-like
Class C Metabotropic glutamate / pheromone
Ion channels
Other receptors
Data on KEGG graph
Rendered by Pathview

Supplementary Figure 2. Gene ontology (GO) and KEGG functional enrichment analysis of the identified DEGs. (A) GO functional enrichment analysis of differentially expressed mRNAs. (B-D) Enriched GO terms in the "biological process" (B), "cellular component" (C), and "molecular function" (D) categories. The node tones represent gene expression levels. Yellow and green tones represent upregulated and downregulated genes, respectively. (E) GO functional enrichment analysis of differentially expressed mRNAs. The node tones indicate gene expression levels, while the gene amount contained in enriched diseases is depicted by the node size. (F) KEGG pathway enrichment analysis. Node colors indicate gene expression levels, and different colored lines indicate enrichment in different pathways. (G, H) The first two KEGG pathways with significant enrichment.

Supplementary Table 2. Gene ontology (GO) enrichment in breast cancer

ONTOLOGY	ID	Description	*P* value
BP	GO:0007268	chemical synaptic transmission	1.44E-09
BP	GO:0008544	epidermis development	5.58E-09
BP	GO:0019373	epoxygenase P450 pathway	5.89E-07
BP	GO:0006508	proteolysis	5.4E-06
BP	GO:0007010	cytoskeleton organization	9.66E-06
BP	GO:0007267	cell-cell signaling	1.29E-05
BP	GO:0035094	response to nicotine	1.32E-05
BP	GO:0051968	positive regulation of synaptic transmission, glutamatergic	1.58E-05
BP	GO:0007613	memory	0.000108
BP	GO:0007586	digestion	0.000123
CC	GO:0005615	extracellular space	9.66E-23
CC	GO:0005576	extracellular region	4.83E-10
CC	GO:0031090	organelle membrane	5.77E-09
CC	GO:0045211	postsynaptic membrane	4.15E-07
CC	GO:0005887	integral component of plasma membrane	4.62E-07
CC	GO:0005882	intermediate filament	1.44E-06
CC	GO:0030054	cell junction	2.44E-05
CC	GO:0005886	plasma membrane	6.98E-05
CC	GO:0072562	blood microparticle	0.0002
CC	GO:0045095	keratin filament	0.000204
MF	GO:0004252	serine-type endopeptidase activity	3.01E-09
MF	GO:0005200	structural constituent of cytoskeleton	4.13E-09
MF	GO:0005198	structural molecule activity	2.66E-08
MF	GO:0008392	arachidonic acid epoxygenase activity	2.74E-06
MF	GO:0020037	heme binding	3.34E-06
MF	GO:0005230	extracellular ligand-gated ion channel activity	4.41E-05
MF	GO:0019825	oxygen binding	6.74E-05
MF	GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	0.000273
MF	GO:0004497	monooxygenase activity	0.000308
MF	GO:0008236	serine-type peptidase activity	0.000547

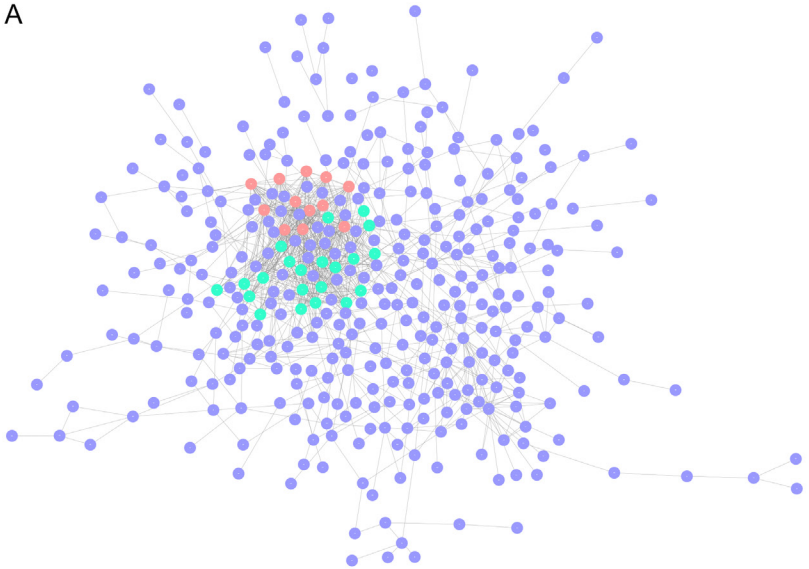
Supplementary Table 3. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment in breast cancer

ONTOLOGY	ID	Description	*P* value
KEGG_PATHWAY	hsa04080	Neuroactive ligand-receptor interaction	1.79056E-09
KEGG_PATHWAY	hsa04971	Gastric acid secretion	4.92554E-05
KEGG_PATHWAY	hsa00980	Metabolism of xenobiotics by cytochrome P450	5.55076E-05
KEGG_PATHWAY	hsa05033	Nicotine addiction	0.000143148
KEGG_PATHWAY	hsa04964	Proximal tubule bicarbonate reclamation	0.000464464
KEGG_PATHWAY	hsa05204	Chemical carcinogenesis	0.000536946
KEGG_PATHWAY	hsa04024	cAMP signaling pathway	0.000714667
KEGG_PATHWAY	hsa04723	Retrograde endocannabinoid signaling	0.000743243
KEGG_PATHWAY	hsa00830	Retinol metabolism	0.002623913
KEGG_PATHWAY	hsa04727	GABAergic synapse	0.003439928

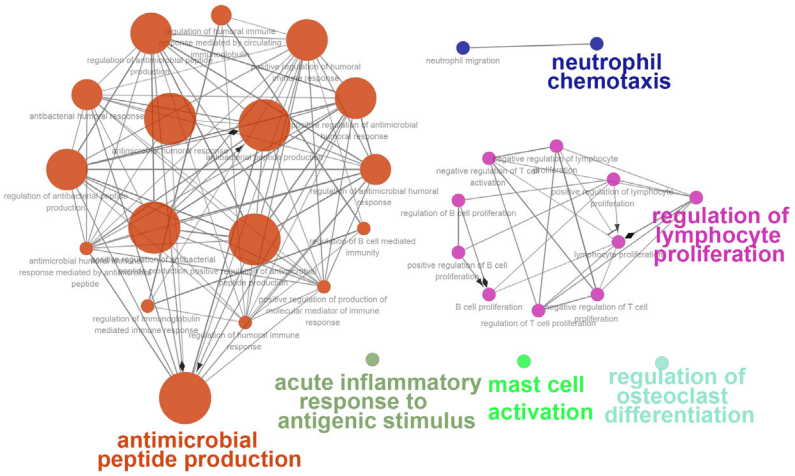
Supplementary Table 4. Gene set enrichment analysis (GSEA) of breast cancer samples

Description	ES	*P* value
go_blood_microparticle	-0.684973713	7.56556E-08
go_chromosome_centromeric_region	0.510126928	7.56556E-08
go_cornification	-0.738682009	7.56556E-08
go_humoral_immune_response	-0.591086935	7.56556E-08
go_immunoglobulin_complex	-0.77718297	7.56556E-08
go_intermediate_filament_cytoskeleton	-0.625770202	7.56556E-08
go_keratinization	-0.645482527	7.56556E-08
go_keratinocyte_differentiation	-0.593230188	7.56556E-08
go_mitotic_sister_chromatid_segregation	0.523549156	7.56556E-08
go_kinetochore	0.553538662	9.73932E-08
kegg_antigen_processing_and_presentation	0.619550907	1.15747E-07
kegg_graft_versus_host_disease	0.75359945	3.54466E-07
kegg_drug_metabolism_cytochrome_p450	-0.706974918	1.61982E-05
kegg_natural_killer_cell_mediated_cytotoxicity	0.47439649	1.61982E-05
kegg_cell_cycle	0.461369722	2.08361E-05
kegg_retinol_metabolism	-0.701874909	0.000119487
kegg_neuroactive_ligand_receptor_interaction	-0.51619541	0.0007227
kegg_metabolism_of_xenobiotics_by_cytochrome_p450	-0.655231647	0.001453848
kegg_dna_replication	0.612598844	0.002222122
kegg_allograft_rejection	0.601829389	0.004042925

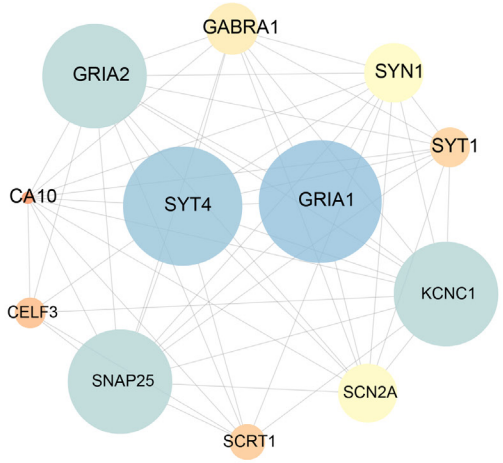
A



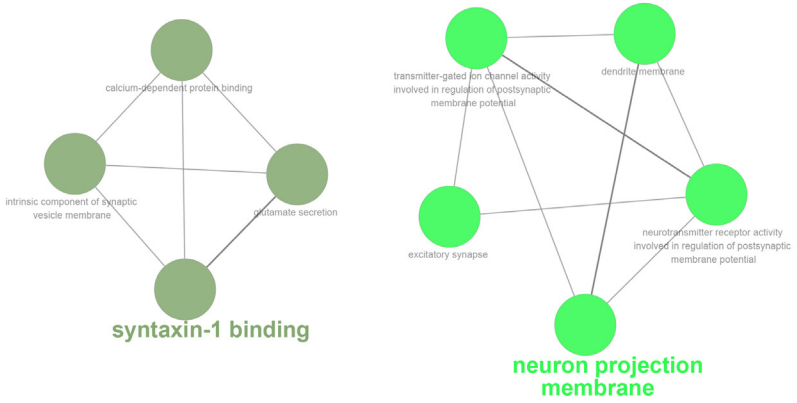
B

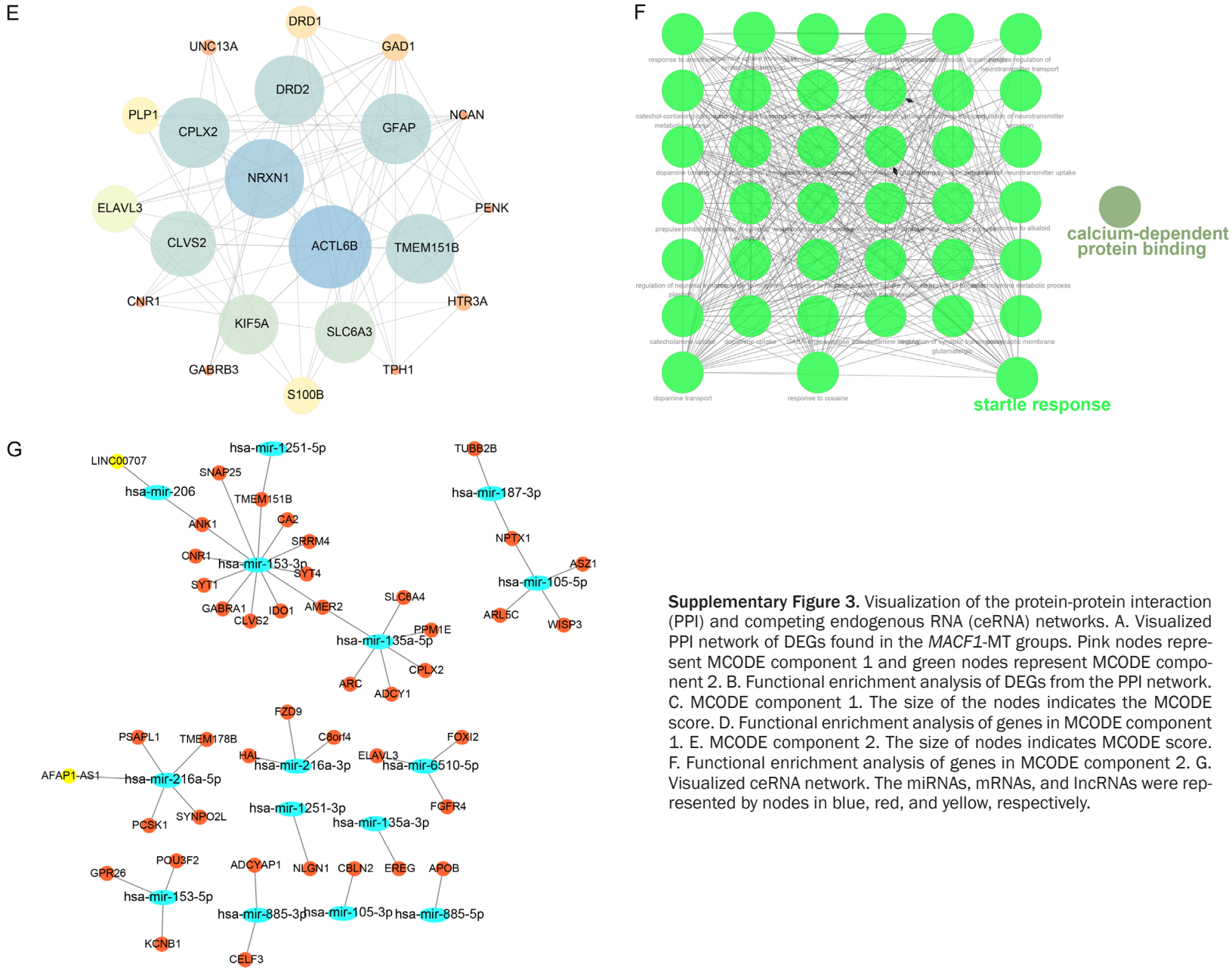

neutrophil migration
neutrophil chemotaxis
regulation of lymphocyte proliferation
antimicrobial peptide production
acute inflammatory response to antigenic stimulus
mast cell activation
regulation of osteoclast differentiation

C


GABRA1
GRIA2
SYN1
SYT1
CA10
SYT4
GRIA1
KCNC1
CELF3
SNAP25
SCN2A
SCRT1

D


calcium-dependent protein binding
intrinsic component of synaptic vesicle membrane
glutamate secretion
syntaxin-1 binding
dendrite membrane
transmitter-gated ion channel activity involved in regulation of postsynaptic membrane potential
excitatory synapse
neurotransmitter receptor activity involved in regulation of postsynaptic membrane potential
neuron projection membrane



Supplementary Figure 3. Visualization of the protein-protein interaction (PPI) and competing endogenous RNA (ceRNA) networks. A. Visualized PPI network of DEGs found in the *MACF1*-MT groups. Pink nodes represent MCODE component 1 and green nodes represent MCODE component 2. B. Functional enrichment analysis of DEGs from the PPI network. C. MCODE component 1. The size of the nodes indicates the MCODE score. D. Functional enrichment analysis of genes in MCODE component 1. E. MCODE component 2. The size of nodes indicates MCODE score. F. Functional enrichment analysis of genes in MCODE component 2. G. Visualized ceRNA network. The miRNAs, mRNAs, and lncRNAs were represented by nodes in blue, red, and yellow, respectively.

Supplementary Table 5. Cox univariate analysis

character	*p*.value	HR	HR.confint.lower	HR.confint.upper
MACF1	0.579	0.932	0.727	1.2
group	0.0484	0.48	0.218	0.96
gender	0.413	2.29	0.315	16.7
race	0.184	0	0	0
age	0.113	1.49	0.91	2.43
lymph	0.0572	0.554	0.302	1.02
ER	0.00398	0.212	0.163	0.74
PR	0.687	0.302	0	1.181
HER2	0.726	15.8	0	0
stage_M	1.01e-06	0.321	0.19	0.621
stage_N	0.00034	1.794	1.03	2.02
stage_T	0.118	0.752	0.71	0.776
tumor_stage	7.31e-10	0.54	0.107	0.654

Supplementary Table 6. Cox multivariate analysis

character	*p*.value	HR	HR.confint.lower	HR.confint.upper
groupwt	0.817061	0.8947163	0.348600459	2.2963750
raceASIAN	0.142087	6.67855892	0.529266530	84.273512
raceBLACK OR AFRICAN AMERICAN	0.000299	20.413514	3.97925576	104.72098
raceWHITE	0.007243	8.511090	1.7833290	40.61991
age60+	0.061575	1.663025	0.97564381	2.834697
lymph10+	0.327405	0.621204	0.2395732	1.610760
ERIndeterminate	0.428915	0.245458	0.007560351	7.969197
ERNegative	0.885386	0.856086	0.10350131	7.080911
ERPositive	0.555375	0.535115	0.066978373	4.275234
stage_MM1	0.282870	0.206818	0.01165388	3.670358
stage_MMX	0.408839	0.651632	0.23584544	1.800436
stage_NN1	0.085155	2.075941	0.90378423	4.768318
stage_NN2	0.180352	2.341400	0.67441314	8.128783
stage_NN3	0.00386	6.485180	1.82412461	23.05630
stage_NNX	0.007978	11.90927	1.91005113	74.2549
stage_TT2	0.302841	1.657782	0.6337812	4.33626
stage_TT3	0.401120	1.680156	0.500386	5.641487
stage_TT4	0.203205	2.2492347	0.6453616	7.83910
stage_TTX	0.430341	0.3195711	0.018760	5.443651
tumor_stagestage I	0.756241	0.7101584	0.0818133	6.1643333
tumor_stagestage II	0.346206	0.4408707	0.0802100	2.423224
tumor_stagestage III	0.465136	0.55594518	0.1150434	2.686594
tumor_stagestage IV	0.007949	33.319392	2.5020144	443.71523