BMC Genomic Data

RESEARCH

Open Access

ERBB3-related gene PBX1 is associated with prognosis in patients with HER2-positive breast cancer

Shufen Mo[1], Haiming Zhong[1], Weiping Dai[2], Yuanyuan Li[1], Bin Qi[2], Taidong Li[3,4*] and Yongguang Cai[1,4*]

Abstract

Background HER2-positive breast cancer (BC) is a subtype of breast cancer. Increased ERBB3 expression has been implicated as a potential cause of resistance to other HER-targeted therapies. Our study aimed to screen and validate prognostic markers associated with ERBB3 expression by bioinformatics and affecting the prognosis of HER2 staging.

Methods Analyzing differences in ERBB3-related groups. ERBB3 expression-related differentially expressed genes (DEGs) were identified and intersected with survival status-related DEGs to obtain intersected genes. Three algorithms, LASSO, RandomForest and XGBoost were combined to identify the signature genes. we construct risk models and generate ROC curves for prediction. Furthermore, we delve into the immunological traits, correlations, and expression patterns of signature genes by conducting a comprehensive analysis that encompasses immune infiltration analysis, correlation analysis, and differential expression analysis.

Results Significant variability in ERBB3 expression and prognosis in high and low ERBB3 expression groups. Twenty-five candidate DEGs were identified by intersecting ERBB3-related DEGs with survival-related DEGs. Utilizing three distinct machine learning algorithms, we identified three signature genes-PBX1, IGHM, and CXCL13-that exhibited significant diagnostic value within the diagnostic model. In addition, the risk model had better prognostic and predictive effects, and the immune infiltration analysis showed that IGHM, CXCL13 might affect the proliferation of BC cells through immune cells. Functional studies demonstrated that interference with PBX1 inhibited the proliferation, migration, and epithelial-mesenchymal transition process of HER2-positive BC cells.

Conclusion PBX1, IGHM and CXCL13 are associated with the expression level of the ERBB3 and are prognostic markers for HER2-positive in BC, which may play an important role in the development and progression of BC.

Keywords HER2-positive breast cancer, ERBB3, Machine learning, Prognosis, Immune infiltration analysis

*Correspondence:
Taidong Li
nkyyltd@163.com
Yongguang Cai
Caiyongguang12@126.com
[1]Medical Oncology, Central Hospital of Guangdong Provincial Nongken, Zhanjiang, Guangdong, China
[2]Pathology, Central Hospital of Guangdong Provincial Nongken, Zhanjiang, Guangdong, China
[3]Thoracic Surgery, Central Hospital of Guangdong Provincial Nongken, Zhanjiang, Guangdong, China
[4]Central Hospital of Guangdong Provincial Nongken, No.2 Renmin Avenue Middle, Xiashan District, Zhanjiang, Guangdong 524002, China

Introduction

Human epidermal growth factor receptor 2 (HER2)-positive breast cancer (BC) represents a distinct pathological subtype of BC that is characterized by a high degree of aggressiveness and often results in an unfavorable prognosis [1, 2]. The overexpression of the HER2 protein is closely linked to the acceleration of tumor proliferation and invasion [3]. The advent of anti-HER2 targeted drugs has markedly enhanced the prognosis of patients with this particular subtype [2, 4]. Nevertheless, a subset of metastatic HER2-positive BC patients demonstrate inadequate responses to targeted pharmaceutical agents [5]. Expanding the number of early cured patients and preventing recurrence are therapeutic goals for HER2-positive BC. Consequently, the development of new therapeutic modalities is imperative, given the dependence of these tumors on HER2 signaling.

The HER receptors are comprised of three distinct domains: an extracellular ligand-binding domain, a transmembrane domain, and an intracellular tyrosine kinase domain. Upon ligand binding to the HER proteins, either homodimerization or heterodimerization of these receptors is triggered, resulting in the activation of downstream signaling pathways. These pathways stimulate cell division and growth while inhibiting cell apoptosis [6–8]. ERBB3, also designated HER3, is the preferred dimerization partner of HER2, forming ERBB3-HER2 heterodimers primarily through the corresponding dimerization arms of the extracellular regions of the two proteins [9–11]. A substantial contributing factor contributing to the failure of HER2-targeted therapies has been identified as the activation of ERBB3 and its subsequent downstream PI3K/AKT signaling pathway [12]. Moreover, elevated ERBB3 expression has been postulated as a potential mechanism underlying resistance to other HER-targeted therapies [13]. The overexpression of HER2 in mouse mammary tissue has been demonstrated to induce tumor formation. However, the formation of tumors is prevented when the ERBB3 gene is knocked out in the tissues [14]. It is therefore imperative to conduct a comprehensive investigation into the bioinformatics of HER2 and ERBB3 expression in HER2-positive BC patients in relation to tumor immunity, with the objective of identifying reliable biomarkers capable of forecasting or monitoring the therapeutic outcomes of HER2-positive BC. This approach is crucial for elucidating the mechanisms underlying tumor immunity and advancing treatment effectiveness.

In the present study, we utilized data from the TCGA and GEO databases were employed. By intersecting the differentially expressed genes (DEGs) identified in the high and low ERBB3 expression groups with from the survival status group, we were able to identify genes that are both associated with ERBB3 expression levels and have the potential to impact prognosis. By employing three distinct machine learning algorithms, namely Least Absolute Shrinkage and Selection Operator (LASSO), Random Forest (RF), and eXtreme Gradient Boosting (XGBoost), we could identify PBX1, IGHM, and CXCL13 as signature genes. The risk models for all three genes demonstrated superior prognostic and predictive outcomes. The immune mechanisms in patients with HER2-positive BC were investigated through the analysis of immune infiltration and correlation analysis between characterized genes and immune cell infiltration. In conclusion, the results demonstrated that PBX1, IGHM and CXCL13 are associated with the expression level of the ERBB3 and are prognostic markers for HER2-positive in BC, which may play an important role in the development and progression of BC.

Table 1 Clinical information of breast patiens

Clinicopathologic variable	Category	TCGA	GSE20711
Sample type	HER2-positive	82	22
Age (years)	< 50	23	/
	≥ 50	59	/
Tumor stage	I	7	0
	II	50	0
	III	20	22
	IV	3	0
	NA	2	0
States of survival	Alive	65	12
	Dead	17	10

Materials and methods

Data sources and processing

Bulk RNA-seq data and relevant clinical information for TCGA-BRCA from the Cancer Genome Atlas (TCGA) database, while the GSE20711 dataset was sourced from the Gene Expression Omnibus (GEO) database. The TCGA-BRCA cohort was employed as the training set, while the GSE20711 cohort served as the validation set. We conducted our study with 82 samples in the TCGA-BRCA training set and 22 samples in the GSE20711 validation set, all of which were HER2-positive BC cases with a clinical survival time exceeding 0. As illustrated in Table 1. Furthermore, the GSE20711 validation set contains 2 Normal breast tissue samples. The differences between groups with high and low ERBB3 expression, and between alive and dead groups were analyzed using the "limma" R package [15]. Genes with a P-value < 0.05 and |log2-fold change (logFC)| >1 were designated as DEGs. The R package "ggplot2" was used to create expression volcano plots of the DEGs [16].

Functional enrichment analysis

Utilizing the "clusterProfiler" R package, we conducted Gene Ontology (GO) and Kyoto Encyclopedia of Genes

and Genomes (KEGG) enrichment analyses. The GO system is comprised of three main categories: cellular components (CC), molecular functions (MF), and biological pathways (BP), which are pivotal in analyzing gene functions and biological processes. On the other hand, KEGG analysis provides information on functional annotation of genes, metabolic pathways and related pathways [17]. Additionally, Gene Set Enrichment Analysis (GSEA) input files comprise expression profiling data that has been normalized by the TCGA HER2-positive BC patient training set and biomarker sample labels. In order to identify relevant pathways and potential molecular mechanisms, the threshold for selecting enriched pathways was set at a false discovery rate (FDR) of $q < 0.01$ to assess relevant pathways and potential molecular mechanisms [18].

Machine learning algorithms to acquire signature genes

Three machine learning algorithms-LASSO, RF, and XGBoost-were employed for the purpose of identifying signature genes. The LASSO logistic regression method is supported by the "glmnet" R software package, which is designed to construct the optimal classification model by determining the classification error minimizing λ, screening the feature variables [19]. The RF algorithm ranks the DEGs and the genes with higher scores are used as signature genes [20]. XGBoost calculates the global importance score of each gene's contribution to the objective function [21]. Finally, the genes obtained from the intersection of the three machine learning algorithms are considered as signature genes.

Construction and validation of risk models

The predict function of the "survuval" R package was used to predict the risk scores of BC patients, and all BC patients were categorized into high- and low-risk groups according based on the median value of the Riskscore. Subsequently, Kaplan-Meier survival curves were plotted for the high- and low-risk groups in the training and validation sets. Furthermore, risk curves for the two groups as well as ROC curves at 1-, 3-, and 5-years were generated using the "timeROC" R software package to observe the differential expression of relevant genes between the high- and the low-risk model group.

Immune infiltration analysis

To evaluate the proportional distribution of immune cells within the BC risk model, the CIBERSORT algorithm was employed to estimate the extent of immune cell infiltration in BC patients. The algorithm in question determines the abundance of immune cells across 22 distinct cell types and functional states [22]. The differences in immune cell infiltration between the high- and low-risk BC groups was conducted using the Wilcoxon rank-sum test. Signature genes linked to immune cells were identified, and the correlation between the expression of these signature genes and immune cell levels was assessed using Spearman correlation analysis.

Cells, and shRNA

Human normal mammary epithelial cells MCF-10 A, HER2-positive breast cancer cell lines HCC1954 (HR-negative) and BT474 (HR-positive) cells were purchased from Procell Life Sciences Co. (Wuhan, China). These cells were maintained in culture following the manufacturer's established protocols. The shRNA targeting PBX1 was synthesized by Sangyo Bio. Subsequently, the cells were assigned to three distinct groups: Control, sh-NC (non-targeting shRNA control), and sh-PBX1. Utilizing Lipofectamine 3000 (Invitrogen) as the transfection agent, the shRNA was introduced into HCC1954 cells. Following a 24 h incubation period post-transfection, the cells were harvested for further analysis.

Quantitative real-time PCR (qRT-PCR)

Total RNA was extracted using RNAiso Plus reagent (Takara), following the manufacturer's instructions. The isolated RNA was then reverse transcribed into cDNA using the PrimeScript RT Master Mix Kit (Takara). qRT-PCR was performed with the GoTaq qPCR Master Mix (Vazyme), adhering strictly to the manufacturer's protocols. The specific primers used for PBX1 were: forward primer 5'-ATGAATCTCCTGCGAGAGCAA-3' and reverse primer 5'-CATCCAGAAATCGGGAACGC-3'. For IGHM, the primers were: forward primer 5'-CCCA CGACCTACAAGGTGAC-3' and reverse primer 5'-ATT CTGCTGGAAGGTCAGGC-3'. Lastly, for CXCL13, the primers were: forward primer 5'-GCTTGAGGTGTAGA TGTGTCC-3' and reverse primer 5'-CCCACGGGGCA AGATTTGAA-3'.

Western blot

Total sample protein was extracted according to the manufacturer's instructions (Proteintech). Protein concentration was measured using the BCA assay (Proteintech). Protein samples were separated by molecular weight size by polyacrylamide electrophoresis and transferred to a polyvinylidene difluoride membrane (Millipore). The samples were then incubated with anti-PBX1 (Abcam, 1:10000), anti-IGHM (CUSABIO, 1:5000), anti-CXCL13 (Abcam, 1:3000), anti-E-Cadherin (Cell Signaling, 1:1000), anti-N-Cadherin (Cell Signaling, 1:1000), anti-Vimentin (Cell Signaling, 1:1000), anti-MMP2 (Abcam, 1:5000) and anti-MMP9 (Abcam, 1:10000) were incubated overnight at 4 °C, followed by incubation with horseradish peroxidase-coupled secondary antibodies for 2 h at room temperature. The target proteins were

specifically detected by treatment with ECL reagent (MultiSciences).

CCK8

Cells from each group were harvested, and a cell proliferation assay was conducted using the CCK-8 kit (Beyotime), adhering to the manufacturer's instructions. Initially, CCK-8 solution, which contains WST-8, was dispensed into 96-well plates containing the cells. These plates were then incubated in a cell culture incubator for one hour. Subsequently, the absorbance was quantified at 450 nm using a microplate reader (Bio-Rad Laboratories, Inc.) equipped with an enzyme marker.

Scratch wound healing assay

After stabilizing the HCC1954 cells transfected with plasmids, they were seeded into a 6-well plate. Upon reaching confluence, a standard 10 μL pipette tip was used to create a straight and consistent scratch across the center of each well. Images of the various cell lines were captured at predetermined time points. The widths of the scratches were then measured and compared to their baseline values across three replicate experiments. For statistical analysis, a two-tailed unpaired t-test was employed.

Transwell cell migration and invasion assay

HCC1954 cells were inoculated with serum-free medium into the upper chamber of Transwell inserts (Corning Incorporated). Matrigel was not added for migration assays and was added for invasion assays (BD Biosciences). The lower chamber was injected with medium containing 10% FBS as chemokine. After 24 h of incubation, cells on the membrane surface that migrated or invaded the lower chamber were fixed with methanol, stained with 0.1% crystal violet solution, and counted in 10 randomly selected fields of view. The whole experiment was repeated three times.

Statistical analyses

All statistical analyses in this study were performed using R software (v4.4.0) and GraphPad Prism (v8.0), with the Wilcoxon rank-sum test used to assess the significance of differences in immune infiltration levels between the high- and low-risk groups. $p < 0.05$ was considered indicative of statistical significance.

Results

Identification of ERBB3 and survival-related DEGs and functional enrichment analysis

The flowchart of this study is presented in supplementary Fig. 1. Firstly, we halved into high and low expression groups based on the expression level of ERBB3. In the HER2-positive BC cohort, there was a significant difference in ERBB3 expression and prognosis, ERBB3 plays an important role in the survival and prognosis of patient (Fig. 1A-C). Therefore, we analyzed gene expression differences in ERBB3 expression-related groups. A total of 129 DEGs were identified, comprising 17 genes with upregulated expression and 112 genes with downregulated expression. The results of the differential expression analysis are presented in the form of volcano plots (Fig. 1D). The GO analysis revealed that DEGs were primarily concentrated in biological processes related to immunoglobulin, chemokine binding, and antigen binding (Fig. 1E). KEGG results showed that DEGs were mainly enriched in signaling pathways such as NF-κB, chemokines, and IL-17 (Fig. 1F).

Subsequently, patients were divided into two groups, Alive and Dead, based on their survival status. A total of 80 DEGs were identified. Of the identified genes, 40 genes were found to have increased expression levels, and the other 40 genes displayed reduced expression. Results of differential expression analysis are presented using volcano plots (Fig. 1G). The GO analysis revealed that the DEGs were primarily enriched in biological processes pertaining to immunoglobulin production, antigen binding, and sex differentiation (Fig. 1H). KEGG results showed that DEGs were mainly enriched in signaling pathways such as prolactin, AMPK, and thyroid hormone (Fig. 1I). These results suggest that ERBB3 and survival-related DEGs are mainly associated with biological processes related to immunity and metabolism.

Selection of the signature genes

A comparable similar pathway metabolism is observed for ERBB3-related DEGs and survival-related DEGs. Consequently, we intersected the gene sets from the ERBB3-related DEGs with survival-related DEGs, resulting in 25 DEGs (Fig. 2A, Supplementary Fig. 2).

Three machine learning algorithms, LASSO regression analysis, RF and XGBoost, were employed to identify signature genes [23]. The results demonstrate that six genes (CXCL13, IGHM, IGHV1-69D, IGHV3-53, PBX1, and SRARP) performed best when they were included in the LASSO model (Fig. 2B, C). In the RF algorithm screening, the top 20 genes in terms of relative importance (Fig. 2D). 13 genes in XGBoost had an impact on prognostic relevance (Fig. 2E). Ultimately, three genes PBX1, IGHM, and CXCL13 were identified as signature genes through the intersection of three machine learning algorithms (Fig. 2F).

Expression level and ROC validation

In the training set, a comprehensive analysis was conducted of the expression levels of three specific genes-PBX1, IGHM, and CXCL13-comparing groups with high and low ERBB3 expression as well as considering their

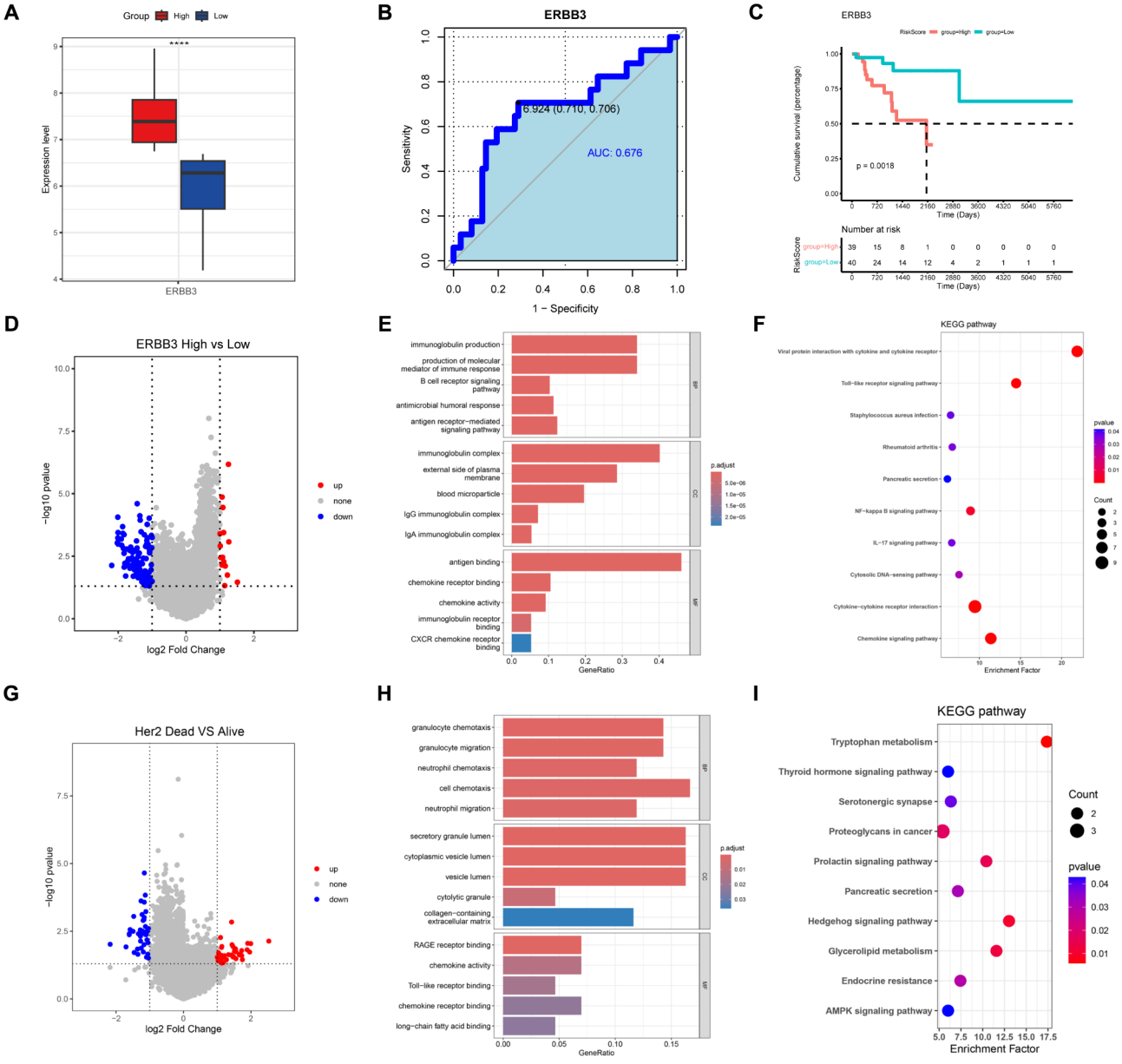

Fig. 1 Identification and enrichment analysis of DEGs grouped by high and low ERBB3 expression and survival state. **A** Box plots of ERBB3 expression levels in the high and low ERBB3 expression groups in the TCGA-HER2 positive BC patient cohort. **B** ROC analysis of ERBB3 gene. **C** Kaplan-Meier analysis of ERBB3. **D** Volcano plot of ERBB3-related DEGs(group: High and Low). **E-F** GO analysis (**E**) and KEGG enrichment analysis (**F**) based on ERBB3-related DEGs. **G** Volcano map of survival-related DEGs (group: Dead and Alive). **H-I** GO analysis (**H**) and KEGG enrichment analysis (**I**) based on survival-related DEGs. In the volcano plots, each dot represents a gene (red dots indicate up-regulated genes with logFC > 1 and p_value < 0.05, and blue dots indicate down-regulated genes with logFC < -1 and p_value < 0.05). GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; FC, fold difference; DEGs, differentially expressed genes; CC, cellular components; MF, molecular functions; BP, biological pathways

survival status. The results demonstrated that the expression of PBX1 gene was higher in the high ERBB3 expression group, and the expression of IGHM and CXCL13 genes was higher in the ERBB3 low expression group (Fig. 3A). Moreover, the Dead group exhibited a higher expression of the PBX1 gene, whereas the alive group displayed a higher expression of IGHM and CXCL13 genes ($p < 0.05$) (Fig. 3B).

Following this, we established a correlation analysis between the three signature genes and patient prognosis. The results revealed promising predictive accuracy, as evidenced by the area under the curve (AUC) values of 0.741 for PBX1, 0.717 for CXCL13, and 0.716 for IGHM in their respective ROC curves (Fig. 3C-E). Moreover, a notable correlation was observed between the genes PBX1, IGHM, and CXCL13 and the prognosis of patients (Fig. 3F-H).

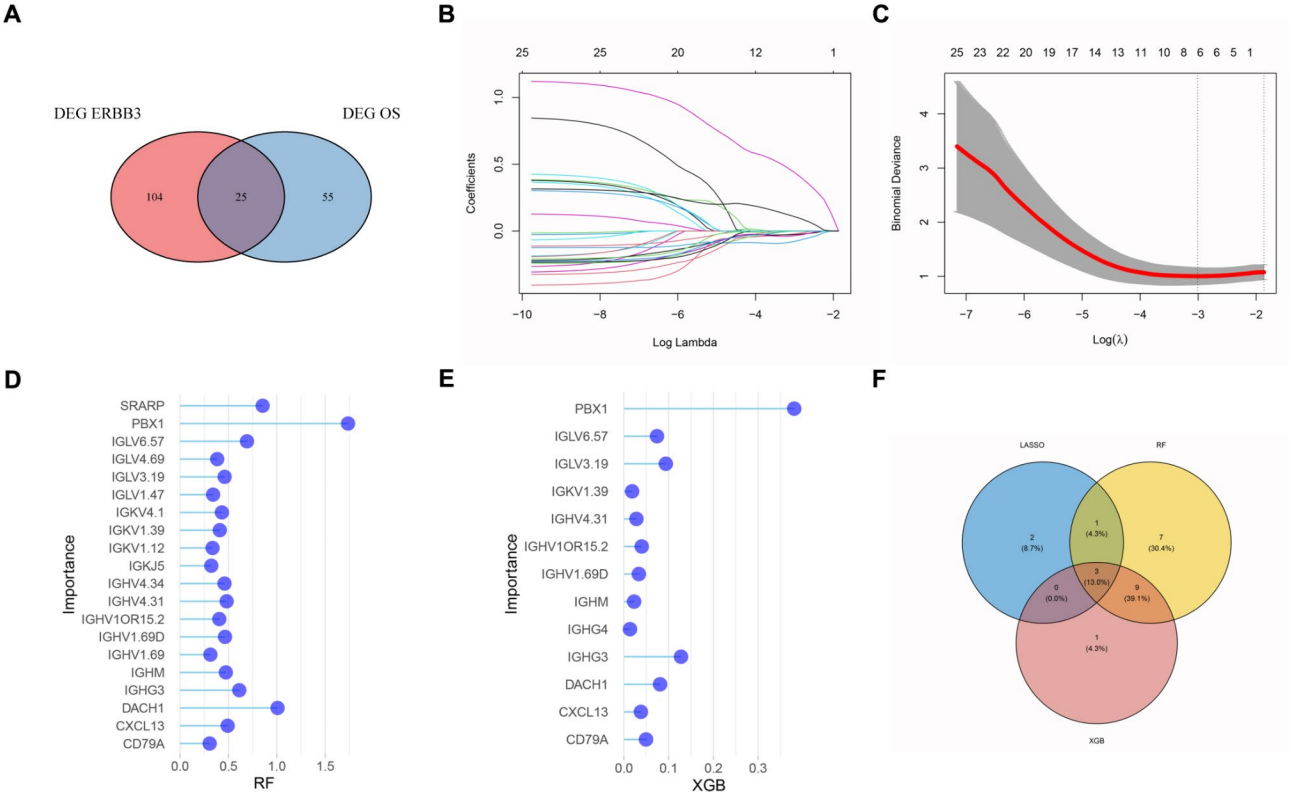

Fig. 2 Signature gene selection. **A** Venn diagram between ERBB3-related DEGs and survival-related DEGs. **B-C** Lasso regression plot (**B**) and validation plot (**C**). **D** Top 20 relatively important genes based on RF algorithm. **E** The top 13 relatively important genes based on the XGBoost algorithm. **F** Venn diagrams of the characterized genes identified by the three machine learning algorithms. LASSO, least absolute shrinkage and selection operator; RF, Random Forest; XGBoost, eXtreme Gradient Boosting

Construction and validation of risk models

A risk model was constructed using the biomarkers PBX1, IGHM, and CXCL13, which had previously been identified. The risk scores for BC patients were predicted, and subsequently, all BC patients were stratified into high- and low-risk groups using the median risk score as the threshold (Fig. 4A-C). The prognostic value was assessed by constructing Kaplan-Meier survival curves for the high- and low-risk groups in both the training and validation datasets. The findings showed that the high-risk cohort had a notably diminished overall survival (OS) in comparison to the low-risk cohort (Fig. 4C). The risk model was employed to generate ROC curves for the 1-, 3-, and 5-year training sets. Notably, the AUC values for these respective training sets exceeded 0.75, indicating that the risk model is highly efficacious (Fig. 4D). Furthermore, we investigated the differential expression patterns of PBX1, CXCL13, and IGHM genes between the high- and low-risk models. Of note, PBX1 displayed elevated expression in the high-risk group, whereas CXCL13 and IGHM exhibited increased expression in the low-risk group (Fig. 4E). The model was additionally subjected to functional validation using the independent validation set GSE20711, and the results confirmed that the model possesses a certain level of accuracy (Fig. 4F-J). In conclusion, the model had a better prognosis as well as predictive results.

Enrichment analysis of risk models

To gain further insight into the relevant signaling pathways and potential biological mechanisms associated with the high- and low-risk groups, we employed GSEA in the TCGA training set to identify pathways that were significantly enriched. GO analysis showed that the high-risk group was predominantly enriched in processes related to energy transport, beta-catenin binding, and regulation of chromosome organization (Fig. 5A). The results of the KEGG analysis revealed significant enrichment in pathways linked to the hedgehog signaling pathway, mRNA surveillance pathway, and nucleocytoplasmic transport (Fig. 5B). In addition, REACTOME analysis demonstrated that it was predominantly enriched in pathways associated with nucleoplasmic transport, ECM proteoglycans, and non-integrin membrane-ECM interactions (Fig. 5C).

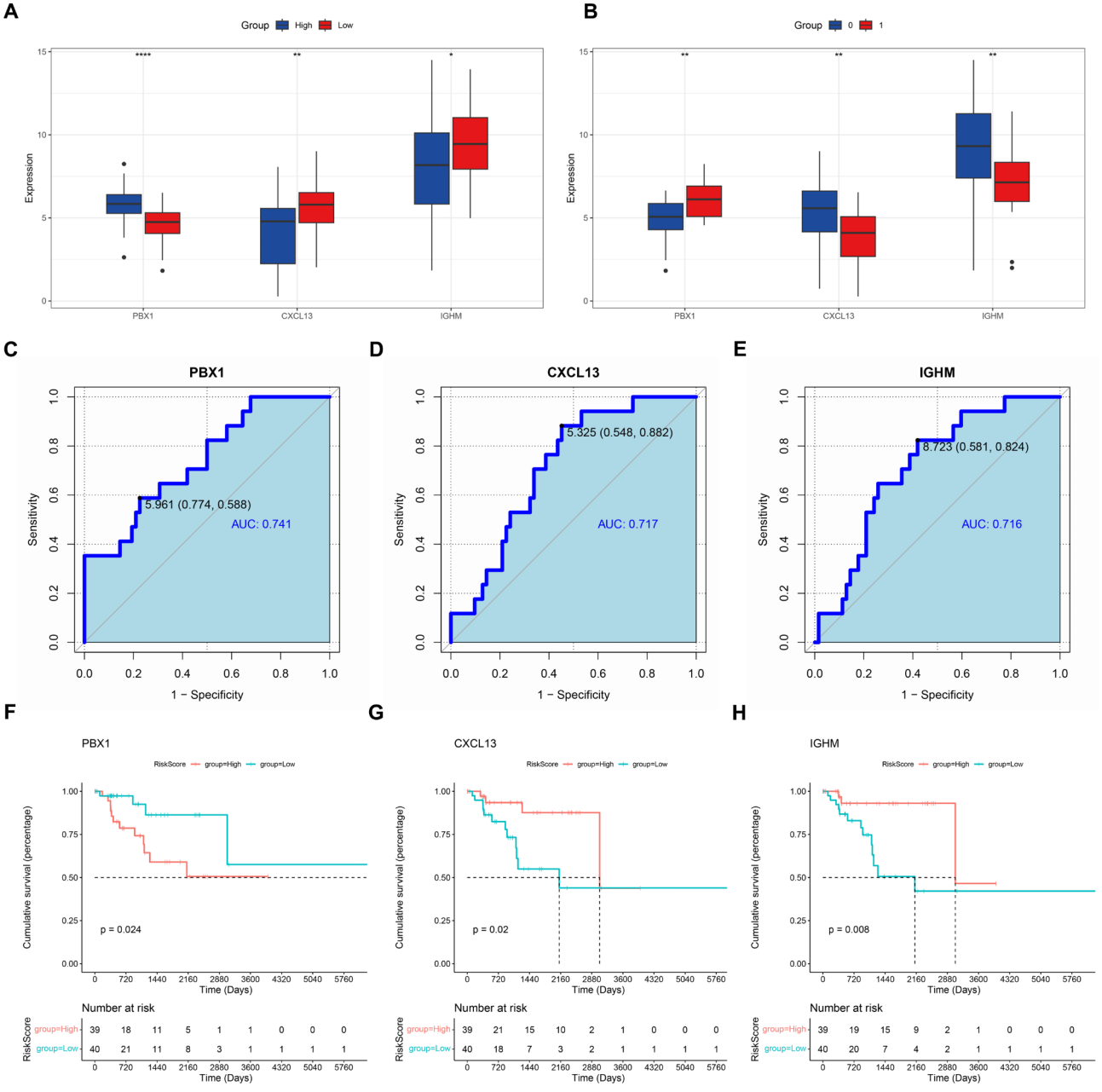

Fig. 3 Expression of genes signature and association with prognosis. **A** Box plots of expression levels of PBX1, CXCL13 and IGHM in the high and low ERBB3 expression groups. **B** Box plots of expression levels of PBX1, CXCL13 and IGHM in the survival status groups.**C-E** ROC analysis of PBX1 (**C**), CXCL13 (**D**) and IGHM (**E**) genes. **F-H** Kaplan-Meier analysis of PBX1 (**F**), CXCL13 (**G**) and IGHM (**H**) genes. 0 represents Alive and 1 represents Dead. BC, breast cancer; AUC, area under the curve; ROC, receiver operating characteristic. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and "ns" represents non-significant

Immune infiltration analysis

The degree of immune cell infiltration in BC patients was assessed using the CIBERSORT algorithm [24], and the proportional abundance of immune cells between samples from high- and low-risk BC groups was compared (Fig. 5D). A comparison of the immune cell distribution in BC patients from the high- and low-risk groups revealed that the majority of immune cell types did not exhibit significant differences. However, notable variations were observed in the levels of activated dendritic cells, plasma cells, T cells follicular helper, T cells CD8, and macrophages M0 (Fig. 5E). Furthermore, following the exclusion of items that were not statistically significant, we found strong correlations between the expression of signature genes and immune cells. These included a negative correlation between PBX1 and B cell memory, as well as T cells follicular helper. Additionally, IGHM and CXCL13 exhibits a negative correlation with

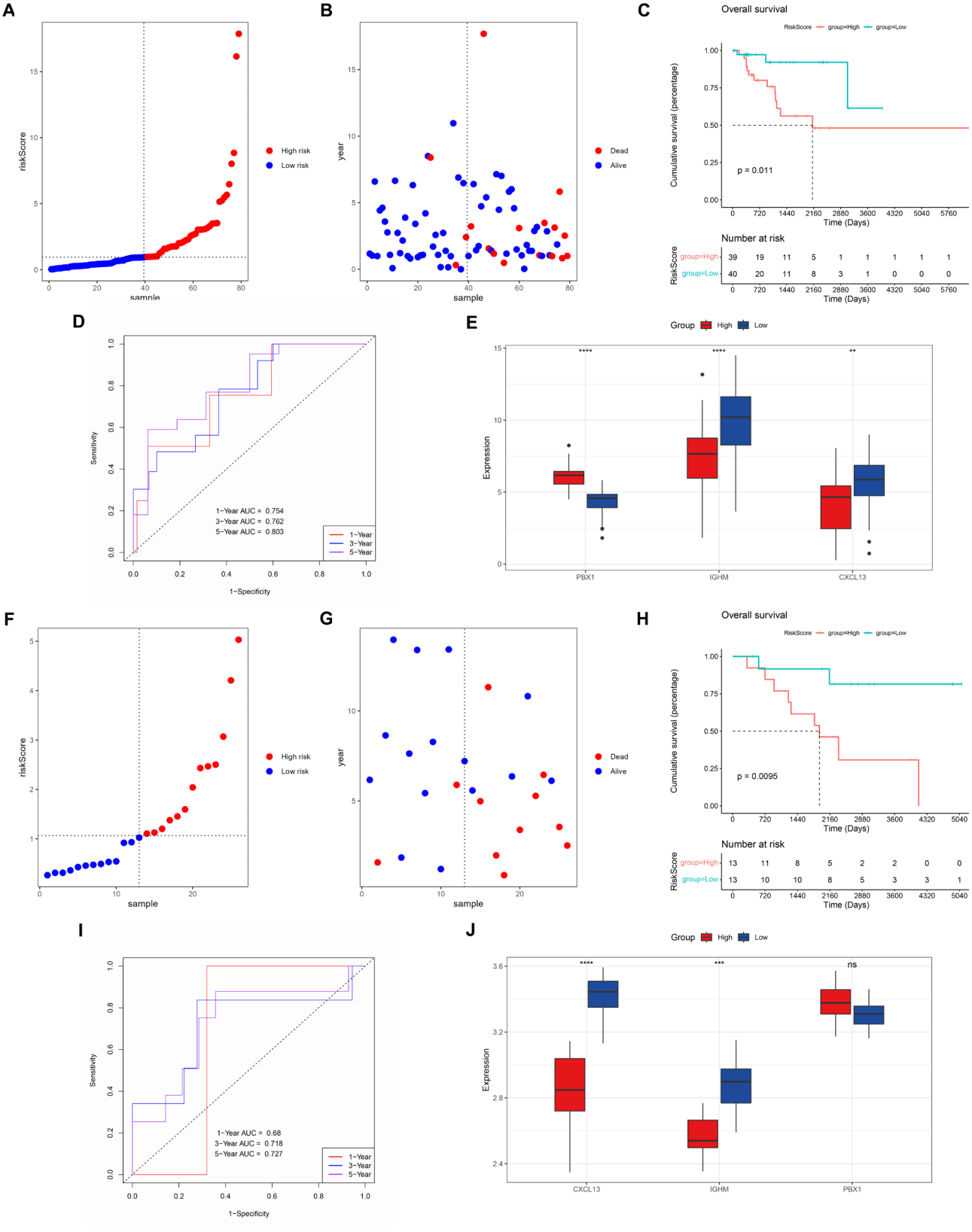

Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Construction of risk models, survival status and mRNA expression levels in the training and validation set. In the training set, **A-B** Risk score plot (**A**) and survival state scatterplot (**B**) of HER2-positve BC patients. **C** Kaplan-Meier OS analysis of high- and low-risk model groups. **D** AUC predicting 1-, 3-, and 5-year survival in HER2-positve BC patients. **E** Box plots of expression levels of PBX1, IGHM, and CXCL13 genes between high- and low-risk model groups. In the validation set, **F-G** Risk score plot (**F**) and survival state scatterplot (**G**) of HER2-positve BC patients. **H** Kaplan-Meier OS analysis of high- and low-risk model groups. **I** AUC predicting 1-, 3-, and 5-year survival in HER2-positve BC patients. **J** Box plots of expression levels of PBX1, IGHM, and CXCL13 genes between high- and low-risk model groups. BC, breast cancer; OS, overall survival; AUC, area under the curve.*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and "ns" represents non-significant

Macrophages M0. On the contrary, CXCL13 showed a negative correlation with macrophages M2, while IGHM and CXCL13 exhibited positive correlations with all other remaining immune cells (Fig. 5H, Supplementary Fig. 3). This indicates that IGHM and CXCL13 may engage with various immune cells to enhance the prognosis of HER2-positive BC. Furthermore, we delved into the correlational analysis between these genes and diverse immune cell types. The results revealed significant correlations within the low-risk group for certain immune cells. For instance, IGHM showed a significant correlation with cells including naive B cells, resting NK cells, and T cells CD8 (Fig. 5I). Similarly, CXCL13 exhibited a notable correlation with cells like T follicular helper cells (Fig. 5J, Supplementary Figs. 4–6), suggesting that these cells have a crucial role in the prognosis of the low-risk group.

Validation of signature genes expression in BC cells

To ascertain the expression levels of PBX1, IGHM, and CXCL13, we conducted culturing of breast cancer cell lines HCC1954 and BT474. The results of the bioinformatics analysis were corroborated by the observation of a significant elevation in mRNA and protein expression of PBX1 in the HCC1954 and BT474 cell lines relative to normal breast epithelial cells MCF-10 A (Fig. 6A-C). mRNA expression of IGHM showed a tendency to increase in the HCC1954H and BT474 cell lines, but was not statistically different (Fig. 6A). Similarly, in comparison to MCF-10 A cells, although to a lesser extent, the mRNA and protein levels of CXCL13 were found to be significantly elevated in HCC1954 (Fig. 6A-C). The discrepancy in PBX1 expression was the most pronounced, and thus was subjected to further investigation in subsequent experiments.

Interference with PBX1 inhibits HER2-positive BC cell proliferation, migration, and epithelial-mesenchymal transition (EMT) process

To explore the role of PBX1 in HER2-positive BC cells, we interfered with PBX1 expression. By qRT-PCR and western blot analysis, PBX1 expression was successfully knocked down in HCC1954 cell lines (Fig. 7A-C). In light of these findings, sh-PBX1-1 was selected for subsequent experimentation. Functional studies showed significant inhibition of proliferation of HCC1954 cells at 24 h, 48 h and 72 h following PBX1 knockdown (Fig. 7D). These findings were corroborated by Edu staining, which revealed a reduction in proliferating cells in the sh-PBX1 group relative to the control and sh-NC groups (Fig. 7E). Furthermore, we assessed the migration and invasion ability of the cells. Our results indicated that PBX1 knockdown resulted in a significant reduction in both wound healing and transwell migration in HCC1954 cells (Fig. 7F, G). To investigate the potential mechanism underlying these effects, we analyzed the expression levels of EMT-related proteins by western blot. In HCC1954 cell lines, PBX1 knockdown led to an increase in E-cadherin expression and a decrease in N-cadherin, vimentin, MMP2, and MMP9 expression, suggesting a reversal of the EMT process (Fig. 7H-J). These findings provide evidence that PBX1 plays a crucial role in the proliferation, migration, and EMT of HER2-positive BC cells and highlight its potential as a therapeutic target for BC treatment.

Discussion

BC is classified into four types based on molecular typing: luminal A, luminal B, HER2-positive, and triple-negative [25, 26]. With the advent of personalized medicine, traditional prognostic factors such as tumor size, grade, and lymph node involvement are no longer adequate for optimal management of early-stage BC patients [27]. Therefore, it becomes imperative to identify and select molecular biomarkers that can reliably predict therapeutic outcomes [28, 29]. ERBB3 plays a pivotal role in HER2-mediated transformation, tumor progression, and drug resistance. In HER2-dependent cells, deletion of ERBB3 results in reduced signaling through PI3K and cell proliferation [30–32], suggesting that HER2 may be dependent on ERBB3 to drive breast cancer cell growth and survival. In this study, we identified three prognostic molecular biomarkers, namely PBX1, IGHM, and CXCL13, through the analysis of data pertaining to ERBB3 expression levels and survival status in patients with HER2-positive BC.

The comparative analysis of ERBB3 expression levels in high and low groups among HER2-positive BC patients revealed the identification of 129 DEGs. The GO and KEGG enrichment analyses revealed a notable clustering of the ERBB3-associated DEGs within signaling pathways related to immunoglobulins, chemokines, antigen binding, as well as the NF-κB and IL-17 pathways. Subsequently, a differential analysis of survival state groups yielded 80 DEGs. GO and KEGG enrichment analyses

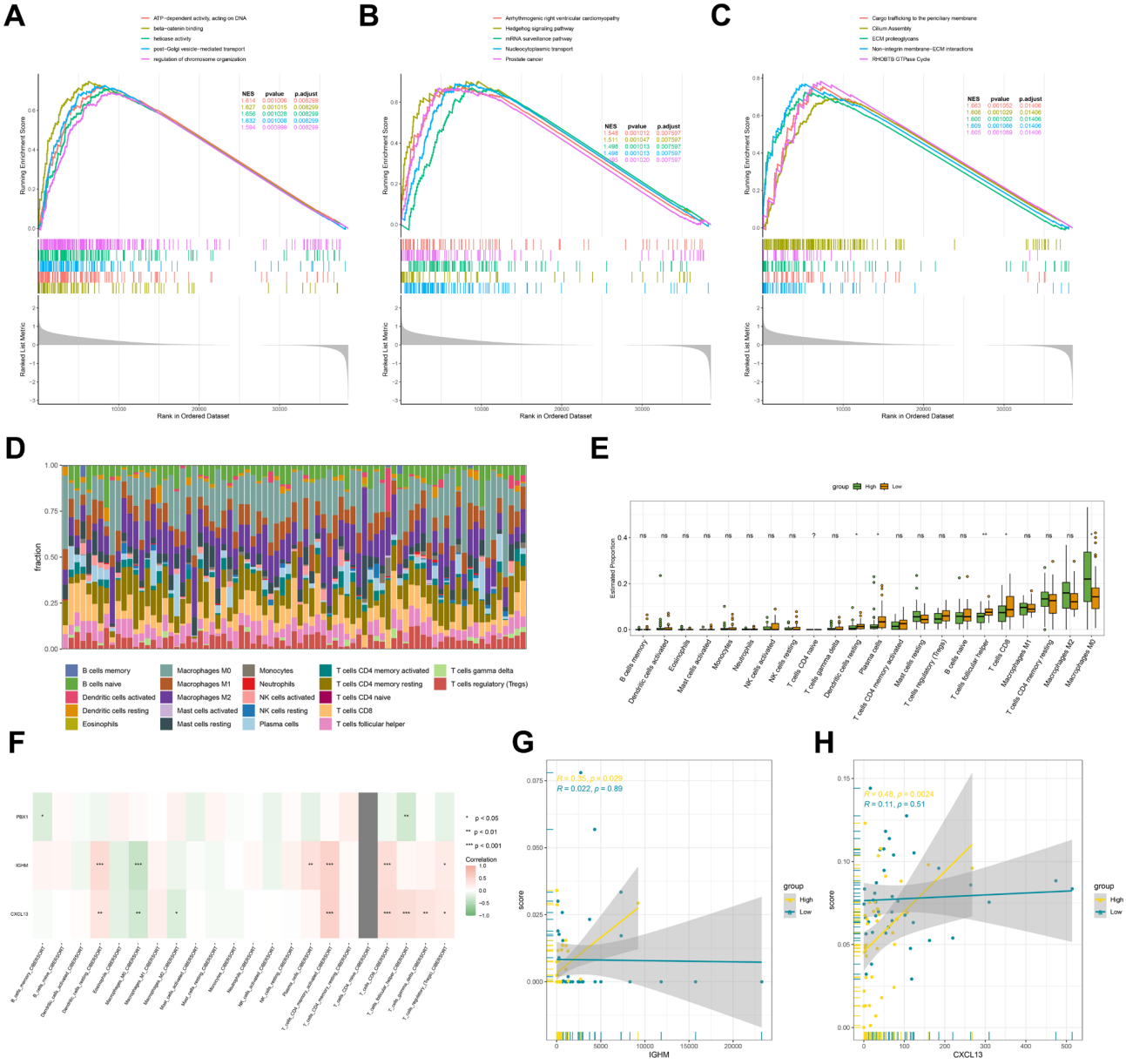

Fig. 5 Biological characteristics and different immune statuses of risk models. **A-C** GSEA analysis for Top5 biological characteristics of GO (**A**), KEGG (**B**) and REACTOME (**C**) between high- and low-risk groups. **D-E** Histograms of relative abundance of 22 immune cells analyzed by the CIBERSORT algorithm in the training set (**D**), and box plots of immune cell infiltration enrichment analysis in the low-risk group (yellow box) and high-risk group (green box) (**E**). **F** Correlation analysis of PBX1, IGHM and CXCL13 genes with immune cells. **G** Scatter plot of correlation between IGHM gene expression levels and NK_cells_resting cells. **H** Correlation scatter plot of CXCL13 gene expression level with T_cells_follicular_helper cells. GSEA, Gene Set Enrichment Analysis; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and "ns" represents non-significant

showed that survival-related DEGs were predominantly associated with the signaling pathways of immunoglobulin, antigen binding, AMPK and thyroid hormones. Chemokines, a class of small cytokines or signaling proteins secreted by cells, serve as crucial active mediators in the immune response. They induce the proliferation of immune cells, facilitate their involvement in the immune response, activate various enzyme activities, and stimulate the killing capability of natural killer cells [33]. The IL-17 signaling pathway has been demonstrated to trigger the activation of inflammatory transcription factors, which in turn lead to the upregulation of gene expression and activation of the MAPK pathway via NF-κB [34]. The cytokines and chemokines induced by IL-17 orchestrate the mobilization of myeloid-derived suppressor cells, which in turn foster angiogenesis while simultaneously suppressing anti-tumor immune responses [35–37]. We found that cell proliferation-related pathways were also

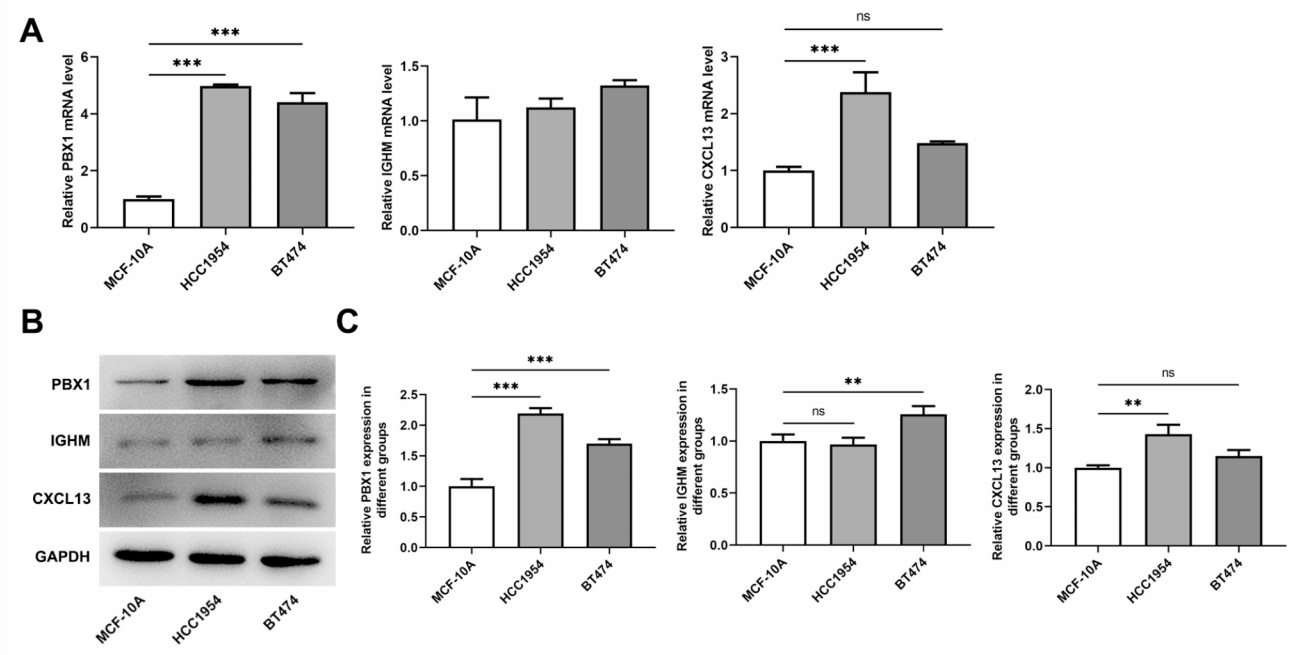

Fig. 6 Validation of key gene expression in HCC1954 and BT474 cells. **A** The mRNA expression levels of PBX1, IGHM, and CXCL13 were detected by qRT-PCR. **B-C** The protein expression levels of PBX1, IGHM, and CXCL13 were detected by western blot. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and "ns" represents non-significant

significantly altered in BC tissues, but not as dramatically as immune-related pathways, including protein tyrosine kinase activator activity and protein kinase activator activity-related processes. In addition, ERBB3 has been shown to play a role in regulating the immune response [6], which is consistent with our observation of some significantly enriched pathways. These findings indicate a robust correlation between ERBB3 expression levels and the prognosis of HER2-positive BC patients, suggesting that the associated genes may play a pivotal role in the complex immune-inflammatory regulatory mechanisms underlying the pathogenesis of BC [38]. Consequently, the intersection of ERBB3-related DEGs and survival-related DEGs yielded 25 genes with correlated ERBB3 expression levels and prognostic implications.

For the above overlapping genes, LASSO, RF, and XGBoost machine learning algorithms were used to screen 3 signature genes, PBX1, IGHM, and CXCL13. These genes exhibit a strong correlation with the prognosis of BC patients and thus have significant diagnostic value and clinical implications. Furthermore, our analysis revealed that the PBX1 gene had higher expression levels in both the high ERBB3 expression and Dead groups, whereas the IGHM and CXCL13 genes had higher expression levels in the low ERBB3 expression and Alive groups, respectively. The expression of these genes was significantly elevated in HCC1954 and BT474 cell lines compared to MCF-10 A cells, further confirming the accuracy of the bioinformatics analysis. PBX1 belongs to the PBX homeobox family of transcription factors and encodes a nuclear protein. It has been reported that PBX1 is correlated with ERα and promotes the expression of genes associated with the aggressive progression of ERα-positive BC by guiding estrogen-induced ERα recruitment to its target chromatin [39]. Furthermore, upregulation of PBX1 has been associated with an increased risk of metastatic progression of ERα-positive BC [40]. In a loss-of-function assay conducted in vitro, we confirmed the biological role of PBX1 in HER2-positive BC cells. Our findings revealed that knockdown PBX1 markedly suppressed the proliferation and metastatic capabilities of HER2-positive BC cells. Furthermore, this knockdown also impeded the EMT process in vitro. IGHM gene is a constant region of the immunoglobulin heavy chain, which is correlated with RFS and distal metastasis-free survival in triple-negative breast cancer and general BC [41], and its role in HER2-positive BC is unclear. CXCL13, a chemokine ligand expressed by stromal cells in β-cell follicles, was found to be overexpressed in tumor tissues and peripheral blood of patients with BC [42]. The CXCL13/CXCR5 axis is related to improved outcomes of HER2-positive BC [43]. Our study demonstrated that 3 signature genes, PBX1, IGHM, and CXCL13, are prognostic molecular biomarkers with prognostic significance in HER2-positive BC.

In addition, all HER2-positive BC samples were divided into high- and low-risk groups based on the calculated risk scores, and we observed that the high-risk group

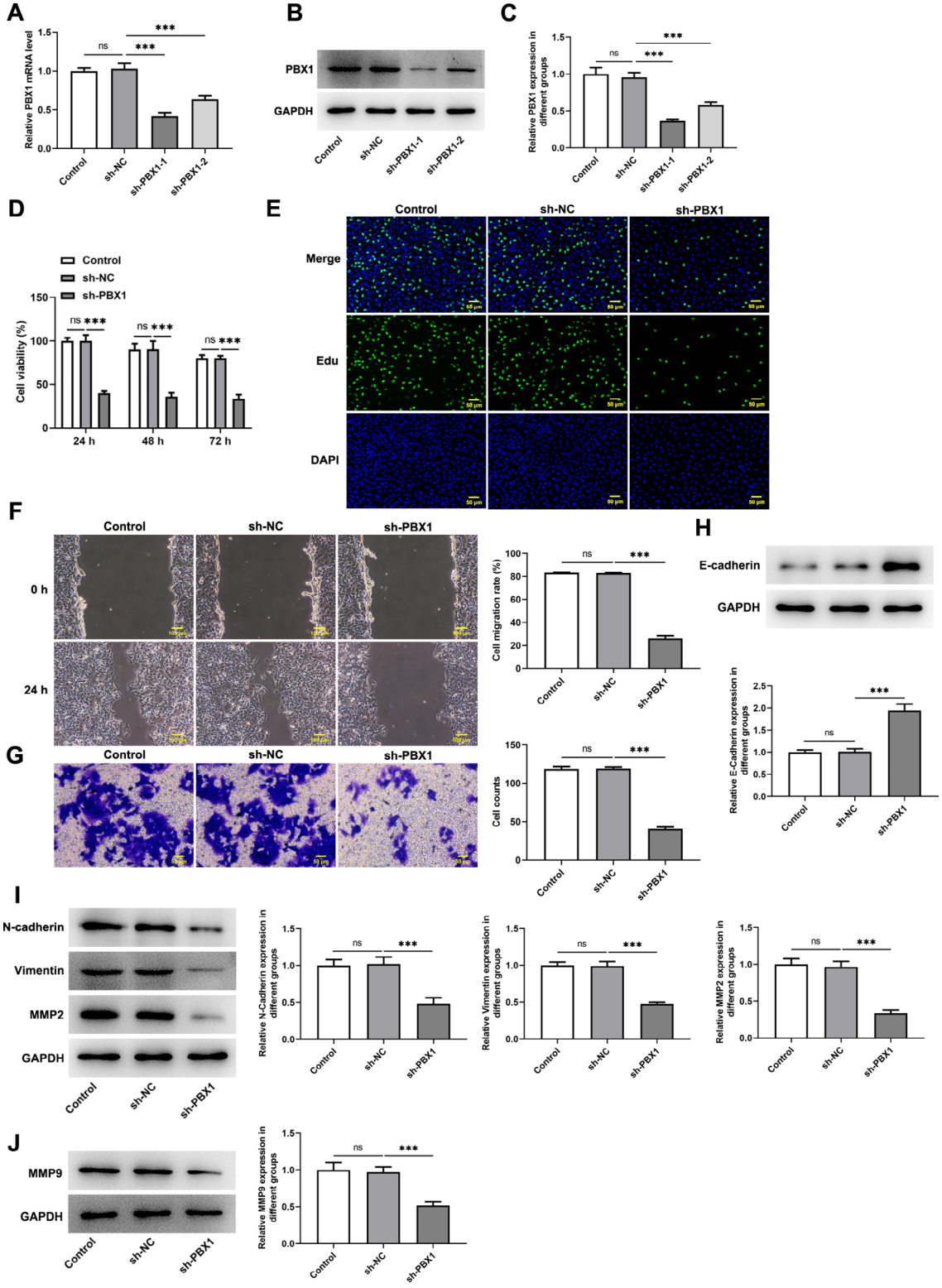

Fig. 7 Interference with PBX1 inhibits proliferation and migration of HER2-positive BC cells and the process of EMT. **A-C** Detection of PBX1 interference levels by qRT-PCR (**A**) and western blot (**B**, **C**). **D** CCK8 detects the proliferation level of HCC1954 cells. **E** Edu staining (green) to observe the proliferation level of HCC1954 cells. DAPI (blue) restaining of cell nuclei. **F-G** wound healing (**F**) and transwell (**G**) assays for cell invasive migration levels. **H-J** western blot detection of migration and EMT-related protein expression levels in cells. BC, breast cancer; EMT, epithelial-mesenchymal transition. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and "ns" represents non-significant

was predominantly enriched for energy transport, ECM, nucleoplasmic transport, and mRNA-associated pathways, and thus hypothesizing that the risk scores may be a potential predictor of HER2-positive BC patients receiving immunotherapy, we conducted an in-depth analysis to assess the immune cell infiltration patterns between the two distinct groups. With regards to the immune cell distribution, no significant variations were observed in the majority of immune cell types, but there were significant differences between the two groups in dendritic cells activated, plasma cells, T cells follicular helper, T cells CD8 and macrophages M0. Breast tumors sites had higher numbers of CD8 T cells, lower numbers of CD4 T cells, and increased OS [44]. CD4 T cells, one of the most important subtypes of T cells and a key player in tumor immunity, include the T regulatory cell, which impairs the function of effector immune cells [45, 46]. Furthermore, our study revealed a strong correlation between IGHM and CXCL13 with various immune cells and their functionalities within the low-risk group, suggesting that they may influence BC cell proliferation via immune cell modulation. This indicates a significant role for these factors in the prognosis of the low-risk group. However, further experiments are still needed to verify these findings. In addition, the sample size of the validation set GSE20711 was relatively small, and to address this challenge, future studies should incorporate diverse validation sets and larger sample sizes, to further validate our results and establish more accurate clinical models.

Although our study successfully identified three key genes PBX1, IGHM and CXCL13 associated with ERBB3 expression levels by in-depth analyses of TCGA data and initially explored their prognostic value in HER2-positive BC, there are still some limitations to the work. Firstly, the sample size of the validation set was relatively small, limiting the broad applicability of the findings, and future studies should include more samples to improve the generality and reliability of the results. Second, the experimental validation was mainly limited to in vitro experiments and a few cell lines, and lacked in vivo experiments and wider cell line validation, which limits our understanding of the mechanisms of these genes' roles in complex biological environments. Finally, differences in statistical and computational methods may lead to uncertainty in conclusions, and optimisation of data pre-processing, parameter settings and algorithm selection should be enhanced in the future to improve the accuracy and reliability of results.

In summary, genes associated with ERBB3 expression levels and their impact on prognosis were intensively investigated by analyzing TCGA data, and three characterized genes, PBX1, IGHM and CXCL13, were identified by risk assessment modeling, external dataset and experimental validation. PBX1 has been shown to promote proliferation, invasion and metastasis of HER2-positive BC cells. Therefore, PBX1, IGHM and CXCL13 genes associated with ERBB3 gene can be used as prognostic markers for HER2 typing and influence the prognosis of HER2 typing.

Abbreviations

GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
FC	Fold difference
DEGs	Differentially expressed genes
OS	Overall survival
RF	Random Forest
BC	Bladder cancer
AUC	Area under the curve
ROC	Receiver operating characteristic
GEO	Gene Expression Omnibus
GSEA	Gene set enrichment analysis
LASSO	Least absolute shrinkage and selection operator
XGBoost	eXtreme Gradient Boosting
CC	Cellular components
MF	Molecular functions
BP	Biological pathways

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12863-024-01292-0.

Supplementary Material 1

Acknowledgements
Not applicable.

Author contributions
All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by SM, HZ, WD, YL, BQ, TL and YC. The first draft of the manuscript was written by SM. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding
This work was supported by the Competitive allocation project of special funds for science and technology development in Zhanjiang City (2022A01200).

Data availability
All data are downloaded from TCGA database (https://portal.gdc.cancer.gov/) and GEO database (https://www.ncbi.nlm.nih.gov/geo/).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All the authors consent for the manuscript to be published.

Competing interests
The authors declare no competing interests.

Received: 31 October 2024 / Accepted: 19 December 2024
Published online: 10 January 2025

References

1. Sung H, Ferlay J, Siegel RL et al. Global Cancer Statistics. 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 71(3):209–249. https://doi.org/10.3322/caac.21660
2. Loibl S, Gianni L. HER2-positive breast cancer. Lancet. 2017;389(10087):2415–2429. https://doi.org/10.1016/s0140-6736(16)32417-5.
3. Bianchini G, De Angelis C, Licata L, et al. Treatment landscape of triple-negative breast cancer - expanded options, evolving needs. Nat Rev Clin Oncol. 2022;19(2):91–113. https://doi.org/10.1038/s41571-021-00565-2.
4. Oh DY, Bang YJ. HER2-targeted therapies - a role beyond breast cancer. Nat Rev Clin Oncol. 2020;17(1):33–48. https://doi.org/10.1038/s41571-019-0268-3.
5. Swain SM, Shastry M, Hamilton E. Targeting HER2-positive breast cancer: advances and future directions. Nat Rev Drug Discov. 2023;22(2):101–126. https://doi.org/10.1038/s41573-022-00579-0.
6. Kumagai S, Koyama S, Nishikawa H. Antitumour immunity regulated by aberrant ERBB family signalling. Nat Rev Cancer. 2021;21(3):181–197. https://doi.org/10.1038/s41568-020-00322-0.
7. Drago JZ, Ferraro E, Abuhadra N, et al. Beyond HER2: targeting the ErbB receptor family in breast cancer. Cancer Treat Rev. 2022;109:102436. https://doi.org/10.1016/j.ctrv.2022.102436.
8. Kemmer S, Berdiel-Acer M, Reinz E, et al. Disentangling ERBB signaling in breast Cancer Subtypes-A model-based analysis. Cancers (Basel). 2022;14(10). https://doi.org/10.3390/cancers14102379.
9. Diwanji D, Trenker R, Thaker TM, Wang F, Agard DA, Verba KA, et al. Structures of the HER2-HER3-NRG1β complex reveal a dynamic dimer interface. Nature. 2021;600(7888):339–343. https://doi.org/10.1038/s41586-021-04084-z.
10. Dey P, Gadewal N, De A. Pathogenic HER3 dimerization domain mutations create a structural bias towards un-conventional EGFR-HER3 signalling axis in breast cancer. Int J Biol Macromol. 2023;242(Pt 2):124765. https://doi.org/10.1016/j.ijbiomac.2023.124765.
11. Hanker AB, Brown BP, Meiler J, et al. Co-occurring gain-of-function mutations in HER2 and HER3 modulate HER2/HER3 activation, oncogenesis, and HER2 inhibitor sensitivity. Cancer Cell. 2021;39(8):1099–e11141098. https://doi.org/10.1016/j.ccell.2021.06.001.
12. Mezynski MJ, Farrelly AM, Cremona M, et al. Targeting the PI3K and MAPK pathways to improve response to HER2-targeted therapies in HER2-positive gastric cancer. J Transl Med. 2021;19(1):184. https://doi.org/10.1186/s12967-021-02842-1.
13. Vlacich G, Coffey RJ. Resistance to EGFR-targeted therapy: a family affair. Cancer Cell. 2011;20(4):423–425. https://doi.org/10.1016/j.ccr.2011.10.006.
14. Vaught DB, Stanford JC, Young C, Hicks DJ, Wheeler F, Rinehart C, et al. HER3 is required for HER2-induced preneoplastic changes to the breast epithelium and tumor formation. Cancer Res. 2012;72(10):2672–2682. https://doi.org/10.1158/0008-5472.Can-11-3594.
15. Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. https://doi.org/10.1093/nar/gkv007.
16. Gustavsson EK, Zhang D, Reynolds RH, Garcia-Ruiz S, Ryten M. Ggtranscript: an R package for the visualization and interpretation of transcript isoforms using ggplot2. Bioinformatics. 2022;38(15):3844–3846. https://doi.org/10.1093/bioinformatics/btac409.
17. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023;51(D1):D587–592. https://doi.org/10.1093/nar/gkac963.
18. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005;102(43):15545–15550. https://doi.org/10.1073/pnas.0506580102.
19. Wang Q, Qiao W, Zhang H, et al. Nomogram established on account of Lasso-Cox regression for predicting recurrence in patients with early-stage hepatocellular carcinoma. Front Immunol. 2022;13:1019638. https://doi.org/10.3389/fimmu.2022.1019638.
20. Asadi S, Roshan S, Kattan MW. Random forest swarm optimization-based for heart diseases diagnosis. J Biomed Inf. 2021;115:103690. https://doi.org/10.1016/j.jbi.2021.103690.
21. Huang G, Jin Q, Mao Y. Predicting the 5-Year risk of nonalcoholic fatty liver Disease using machine learning models: prospective cohort study. J Med Internet Res. 2023;25:e46891. https://doi.org/10.2196/46891.
22. Guan M, Jiao Y, Zhou L. Immune Infiltration Analysis with the CIBERSORT Method in Lung Cancer. Dis Markers. 2022;2022(3186427). https://doi.org/10.1155/2022/3186427.
23. Lai Y, Lin P, Lin F, et al. Identification of immune microenvironment subtypes and signature genes for Alzheimer's disease diagnosis and risk prediction based on explainable machine learning. Front Immunol. 2022;13:1046410. https://doi.org/10.3389/fimmu.2022.1046410.
24. Sui S, An X, Xu C, Li Z, Hua Y, Huang G, et al. An immune cell infiltration-based immune score model predicts prognosis and chemotherapy effects in breast cancer. Theranostics. 2020;10(26):11938–11949. https://doi.org/10.7150/thno.49451.
25. Feng Y, Spezia M, Huang S, et al. Breast cancer development and progression: risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. Genes Dis. 2018;5(2):77–106. https://doi.org/10.1016/j.gendis.2018.05.001.
26. Yeo SK, Guan JL. Breast Cancer: multiple subtypes within a Tumor? Trends Cancer. 2017;3(11):753–760. https://doi.org/10.1016/j.trecan.2017.09.001.
27. Duffy MJ, O'Donovan N, McDermott E, et al. Validated biomarkers: the key to precision treatment in patients with breast cancer. Breast. 2016;29:192–201. https://doi.org/10.1016/j.breast.2016.07.009.
28. Dey P, Rathod M, De A. Targeting stem cells in the realm of drug-resistant breast cancer. Breast Cancer (Dove Med Press). 2019;11:115–135. https://doi.org/10.2147/bctt.S189224.
29. Dey P, Joshi M, Mujawar A, Malhotra R, De A. Direct knockdown of phospho-PTM targets mediated by TRIM21 can improve personalized treatment in breast cancer. Cell Oncol (Dordr). 2022;45(5):873–891. https://doi.org/10.1007/s13402-022-00693-6.
30. Garrett JT, Olivares MG, Rinehart C, Granja-Ingram ND, Sánchez V, Chakrabarty A, et al. Transcriptional and posttranslational up-regulation of HER3 (ErbB3) compensates for inhibition of the HER2 tyrosine kinase. Proc Natl Acad Sci USA. 2011;108(12):5021–5026. https://doi.org/10.1073/pnas.1016140108.
31. Lee-Hoeflich ST, Crocker L, Yao E, Pham T, Munroe X, Hoeflich KP, et al. A central role for HER3 in HER2-amplified breast cancer: implications for targeted therapy. Cancer Res. 2008;68(14):5878–5887. https://doi.org/10.1158/0008-5472.Can-08-0380.
32. Chakrabarty A, Sánchez V, Kuba MG, Rinehart C, Arteaga CL. Feedback upregulation of HER3 (ErbB3) expression and activity attenuates antitumor effect of PI3K inhibitors. Proc Natl Acad Sci USA. 2012;109(8):2718–2723. https://doi.org/10.1073/pnas.1018001108.
33. Hwaiz R, Hasan Z, Rahman M, et al. Rac1 signaling regulates sepsis-induced pathologic inflammation in the lung via attenuation of Mac-1 expression and CXC chemokine formation. J Surg Res. 2013;183(2):798–807. https://doi.org/10.1016/j.jss.2013.02.045.
34. Amatya N, Garg AV, Gaffen SL. IL-17 signaling: the Yin and the Yang. Trends Immunol. 2017;38(5):310–322. https://doi.org/10.1016/j.it.2017.01.006.
35. Jin C, Lagoudas GK, Zhao C, et al. Commensal microbiota promote Lung Cancer Development via γδ T cells. Cell. 2019;176(5):998–e10131016. https://doi.org/10.1016/j.cell.2018.12.040.
36. Coffelt SB, Kersten K, Doornebal CW, Weiden J, Vrijland K, Hau CS, et al. IL-17-producing γδ T cells and neutrophils conspire to promote breast cancer metastasis. Nature. 2015;522(7556):345–348. https://doi.org/10.1038/nature14282.
37. Chung AS, Wu X, Zhuang G, Ngu H, Kasman I, Zhang J, et al. An interleukin-17-mediated paracrine network promotes tumor resistance to anti-angiogenic therapy. Nat Med. 2013;19(9):1114–1123. https://doi.org/10.1038/nm.3291.
38. Zhu M, Yu M, Meng Y, et al. HER3 receptor and its role in the therapeutic management of metastatic breast cancer. J Transl Med. 2024;22(1):665. https://doi.org/10.1186/s12967-024-05445-8.
39. Magnani L, Ballantyne EB, Zhang X, et al. PBX1 genomic pioneer function drives ERα signaling underlying progression in breast cancer. PLoS Genet. 2011;7(11):e1002368. https://doi.org/10.1371/journal.pgen.1002368.
40. Magnani L, Patten DK, Nguyen VT, Hong SP, Steel JH, Patel N, et al. The pioneer factor PBX1 is a novel driver of metastatic progression in ERα-positive breast cancer. Oncotarget. 2015;6(26):21878–21891. https://doi.org/10.18632/oncotarget.4243.
41. Hsu HM, Chu CM, Chang YJ, et al. Six novel immunoglobulin genes as biomarkers for better prognosis in triple-negative breast cancer by gene co-expression network analysis. Sci Rep. 2019;9(1):4484. https://doi.org/10.1038/s41598-019-40826-w.
42. Panse J, Friedrichs K, Marx A, Hildebrandt Y, Luetkens T, Barrels K, et al. Chemokine CXCL13 is overexpressed in the tumour tissue and in the peripheral blood of breast cancer patients. Br J Cancer. 2008;99(6):930–938. https://doi.org/10.1038/sj.bjc.6604621.

43. Razis E, Kalogeras KT, Kotoula V, Eleftheraki AG, Nikitas N, Kronenwett R, et al. Improved outcome of high-risk early HER2 positive breast cancer with high CXCL13-CXCR5 messenger RNA expression. Clin Breast Cancer. 2012;12(3):183–193. https://doi.org/10.1016/j.clbc.2012.03.006.
44. Galluzzi L, Buqué A, Kepp O, et al. Immunological effects of Conventional Chemotherapy and targeted Anticancer agents. Cancer Cell. 2015;28(6):690–714. https://doi.org/10.1016/j.ccell.2015.10.012.
45. de la Cruz-Merino L, Chiesa M, Caballero R, et al. Breast Cancer Immunology and Immunotherapy: current status and future perspectives. Int Rev Cell Mol Biol. 2017;331:1–53. https://doi.org/10.1016/bs.ircmb.2016.09.008.
46. Li C, Jiang P, Wei S, et al. Regulatory T cells in tumor microenvironment: new mechanisms, potential therapeutic strategies and future prospects. Mol Cancer. 2020;19(1):116. https://doi.org/10.1186/s12943-020-01234-1.

Publisher's note