scientific reports

OPEN

Establishing a prognostic model with immune-related genes and investigating *EPHB6* expression pattern in breast cancer

Hui Lyu[1,2], Tao Zhou[3], Xiaoqin Sun[4], Hui Chen[4], Jing Li[4], Mingxiu Shao[2], Jianmei Li[2], Quanmei Zhang[5], Guosheng Jiang[1✉] & Xin Zhou[6✉]

In breast cancer, the behavior of genes linked to the immune system and their interaction with the tumor's microenvironment suggest new paths for tailored therapies. Utilizing the TCGA-BRCA cohort, we established a robust overall survival prediction model through LASSO regression and Gaussian mixture model based on risk group. We found that low-risk patients responded better to chemotherapy. Single-cell analysis further confirmed expression patterns of signature genes in both healthy and malignant breast samples. Our study, the first to use immunohistochemistry (IHC) to assess *EPHB6* expression in benign and malignant breast samples, revealed higher *EPHB6* levels in benign tissue and triple-negative cancer. In axillary lymph nodes, *EPHB6* was predominantly expressed in stroma cells, with diminished expression in cancerous cells upon infiltration. These insights highlight the significance of immune-related genes in breast cancer.

Keywords Breast cancer, Tumor microenvironment, Prognostic model, *EPHB6*, Single-cell RNA-seq

Abbreviations

AUC	Area Under the Curve
BEST	Bioinformatics Evaluation System for Tumor Immunity and Microenvironment Status
B cells	B Lymphocytes
DCIS	Ductal Carcinoma In Situ
DEGs	Differentially Expressed Genes
EDTA	Ethylenediaminetetraacetic Acid
ER	Estrogen Receptor
GDSC2	Genomics of Drug Sensitivity in Cancer
GEPIA2	Gene Expression Profiling Interactive Analysis 2
GEO	Gene Expression Omnibus
GMM	Gaussian Mixture Model
GO	Gene Ontology
Her2	Human Epidermal Growth Factor Receptor 2
IHC	Immunohistochemistry
ILC	Invasive Lobular Carcinoma
IOD	Integrated Optical Density
KEGG	Kyoto Encyclopedia of Genes and Genomes
LASSO	Least Absolute Shrinkage and Selection Operator
NS	Not Significant
OS	Overall Survival
ROC	Receiver Operating Characteristic
SD	Standard Deviation

[1]Department of Immunology, College of Basic Medicine, Binzhou Medical University, Yantai, Shandong, China. [2]Clinical Laboratory, Zibo Maternal and Child Health Hospital, Zibo, Shandong, China. [3]Department of General Surgery, Rizhao People's Hospital, Rizhao, Shandong, China. [4]Department of Pathology, Zibo Maternal and Child Health Hospital, Zibo, Shandong, China. [5]Department of Ultrasound, Zibo Maternal and Child Health Hospital, Zibo, Shandong, China. [6]Department of Breast and Thyroid Surgery, Zibo Maternal and Child Health Hospital, Zibo, Shandong, China. ✉email: jianggsh@sohu.com; zhouxin2286@163.com

TCGA-BRCA   The Cancer Genome Atlas Breast Invasive Carcinoma
TME   Tumor Microenvironment
TNBC   Triple Negative Breast Cancer
TPM   Transcripts Per Kilobase Million

Breast cancer is the most frequently diagnosed cancer among women globally, and its high mortality rate highlights the urgent need for ongoing research and advancements in treatment[1]. While significant progress has been made in treatment options, such as surgery, chemotherapy, radiotherapy, endocrine therapy, targeted therapy, and immunotherapy[2], the complexity of breast cancer remains a challenge. Recent research has focused on the tumor microenvironment (TME), particularly the interactions between tumor cells and the immune microenvironment, which includes surrounding immune cells, stromal cells, extracellular matrix molecules, and cytokines[3,4]. The TME's impact on breast cancer prognosis and treatment response is increasingly evident, with the immune microenvironment playing a crucial role in immunotherapy efficacy and patient outcomes[5,6]. Evidence shows that the immune microenvironment significantly affects immunotherapy effectiveness and the overall survival of breast cancer patients[7,8]. The presence of infiltrating immune cells like T cells and B cells is linked to better clinical outcomes in breast cancer[9,10]. Additionally, alterations in the expression of immune-related genes and pathways have been implicated in tumor immune evasion and resistance to therapy[11]. These findings have driven intensive research to understand the interplay between the immune system and breast cancer cells. For instance, recent studies have identified immune-related genes that may serve as biomarkers for breast cancer prognosis and treatment response, shedding light on this critical area[12,13].

Based on this background, our study analyzed publicly available breast cancer datasets and employed various bioinformatics tools to identify a series of immune-related prognostic genes. Among these, *EPHB6*, a prominent member of the receptor tyrosine kinase superfamily (comprising *EPHAs* and *EPHBs*)[14,15], emerged as the most significant prognostic gene for overall survival (OS) in breast cancer. *EPH* receptors and their ligands are crucial in cell processes like migration, interactions, and vascular development[16]. Humans have nine *EPHA* receptors (*EPHA1-EPHA8*, *EPHA10*) that bind to five *ephrin-A* ligands (*ephrin-A1-A5*), and five *EPHB* receptors (*EPHB1-EPHB4*, *EPHB6*) that bind to three ephrin-B ligands (*ephrin-B1-B3*)[17]. Ephrin-As link to *EPHAs* via a plasma membrane anchor, and *ephrin-Bs* connect to *EPHBs* through a transmembrane domain[18].

We also constructed a gene set consisting of seven key genes that represent the immune characteristics within the TME. Functional analysis showed that these genes are closely associated with various immune cells, such as T cells and B cells, and are involved in a wide range of immune processes, emphasizing their significant role in the breast cancer immune environment. Furthermore, our research introduces a new focus on the EPHB6 gene, providing the first evidence of its differential expression in benign and malignant breast tissues. These findings not only offer new insights for immunotherapy in breast cancer but also lay the foundation for the development of future personalized treatment strategies.
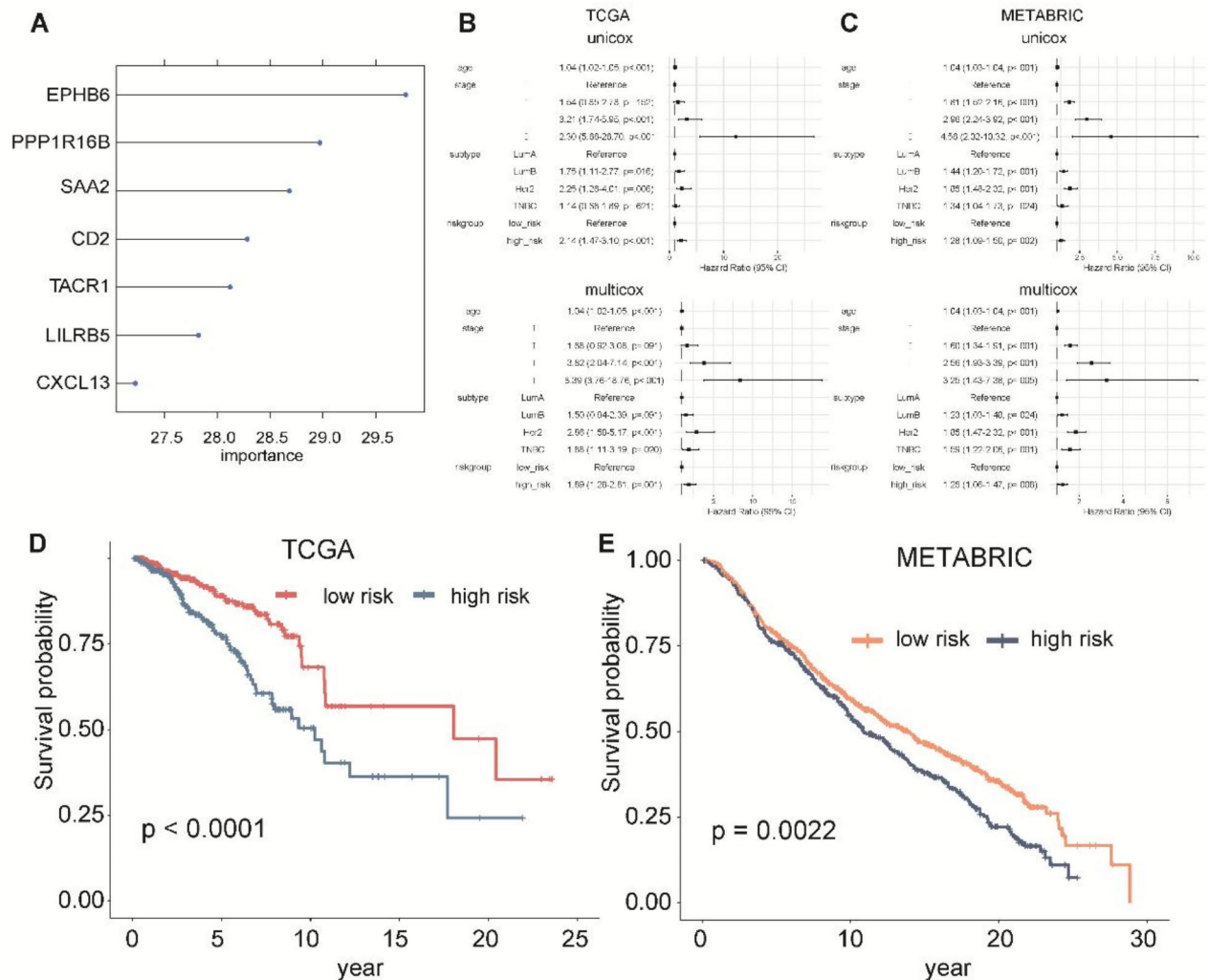
Results
Identification of immune-related prognostic genes
In this study, we first obtained the immune and stromal scores of the TCGA breast cancer dataset from the ESTIMATE website. Then, we classified patients into four distinct groups based on their scores: high/low immune score and high/low stromal score groups. After this classification, we performed a survival analysis on these categorized cohorts. We found that patients in the high immune score group had significantly better OS compared to those in the low immune score group (Supplementary Fig. 1A). However, there was no statistically significant difference in OS between patients with high and low stromal scores (Supplementary Fig. 1B). The findings of this study indicate that immune scores may serve as a valuable tool for identifying patients with a better prognosis, potentially guiding more personalized management strategies for breast cancer. Future research should focus on validating these results in larger cohorts and exploring the biological mechanisms underlying the association between immune scores and patient outcomes.

Next, we employed the 'limma'[19] package to identify differentially expressed genes (DEGs) between the high and low immune score groups. Following this, we intersected these genes with the DEGs between breast cancer and normal breast tissue obtained from the GEPIA2 website, resulting in 88 potential immune-related prognostic genes. Through Lasso regression analysis, we further pinpointed 8 key genes (Supplementary Fig. 2A-B). The best combination of prognostic genes was selected based on area under the curve (AUC) values using the GMM method. Then we identified an immune signature gene set that consists of seven genes (Supplementary Fig. 2C), including *CD2*, *CXCL13*, *PPP1R16B*, *LILRB5*, *EPHB6*, *TACR1*, and *SAA2*. Using the BEST platform for tumor immune infiltration, GO (gene ontology), and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analyses, we found that these genes held strong correlations with various immune cells such as T cells, B cells (Supplementary Fig. 3). GO analysis revealed that these genes participated in multiple immune processes, such as the immune system process and immune response (Supplementary Fig. 4). KEGG analysis demonstrated that these genes were related to numerous immune-related pathways, such as cytokine-cytokine receptor interaction, natural killer cell-mediated cytotoxicity (Supplementary Fig. 5). Consequently, these genes were selected for subsequent analyses. Notably, the random forest algorithm confirmed the pivotal role of the *EPHB6* in predicting overall survival (OS) prognosis (Fig. 1A).

Construction and validation of the prognostic model
First, we evaluated the effect of the risk group on patient outcomes in both the TCGA and METABRIC datasets. Consistently, both univariable and multivariable COX regression analyses identified the risk group as an independent predictor of OS (Fig. 1B-C). Further survival analysis confirmed that patients in the low-risk group had significantly longer OS compared to those in the high-risk group (Fig. 1D-E). Using the TCGA cohort as the

Fig. 1. Prognostic significance of risk group. Application of the Random Forest algorithm for prioritizing seven prognostic genes (**A**). Forest plot visualizing the results of univariable and multivariable Cox analysis of the risk group and clinicopathological factors in TCGA (**B**) and METABRIC (**C**) datasets. Kaplan–Meier survival curves of OS between high- and low-risk groups in in TCGA (**D**) and METABRIC (**E**) datasets.

training dataset, we developed a nomogram prognostic model incorporating four variables: age, stage, subtype, and risk group (Fig. 2A). This model achieved impressive AUC values of 0.857, 0.788, and 0.745 for predicting 1-year, 3-year, and 5-year OS, respectively, in the TCGA cohort (Fig. 2B). Upon validation in the METABRIC cohort, the corresponding AUC values were 0.832, 0.653, and 0.679 (Fig. 2B). Calibration and DCA curves further supported the model's robust predictive power, providing additional evidence of its reliability (Fig. 2C-D). Notably, the model demonstrated the highest net benefit in predicting 5-year OS (Fig. 2D), indicating that our prognostic tool is dependable and can accurately guide treatment decisions and estimate patient survival.

The risk score predicts therapeutic benefits

To explore the response of breast cancer patients in different risk groups to commonly used chemotherapy drugs, we utilized the GDSC2 database as a reference and employed the Oncopredict package for predicting drug sensitivity. Among the six chemotherapy medications we selected—taxanes (including docetaxel), epirubicin, cisplatin, gemcitabine, and vinorelbine—taxanes and epirubicin are key components in breast cancer chemotherapy regimens. Our results showed that TCGA patients in the low-risk group exhibited greater responsiveness, especially to taxanes and epirubicin, as well as the other medications (Fig. 3A). This finding suggests that patients in this group may gain more substantial benefits from these therapies, potentially guiding the development of personalized treatment strategies in the future. We validated these findings using the METABRIC dataset, further strengthening the credibility of our research (Fig. 3B).

Subgroup analysis revealed distinct drug response patterns. In the TCGA cohort, the low-risk group demonstrated significantly enhanced sensitivity to gemcitabine across Luminal A, Luminal B, and HER2-positive subtypes (Supplementary Fig. 6A). However, the METABRIC validation cohort exhibited a broader
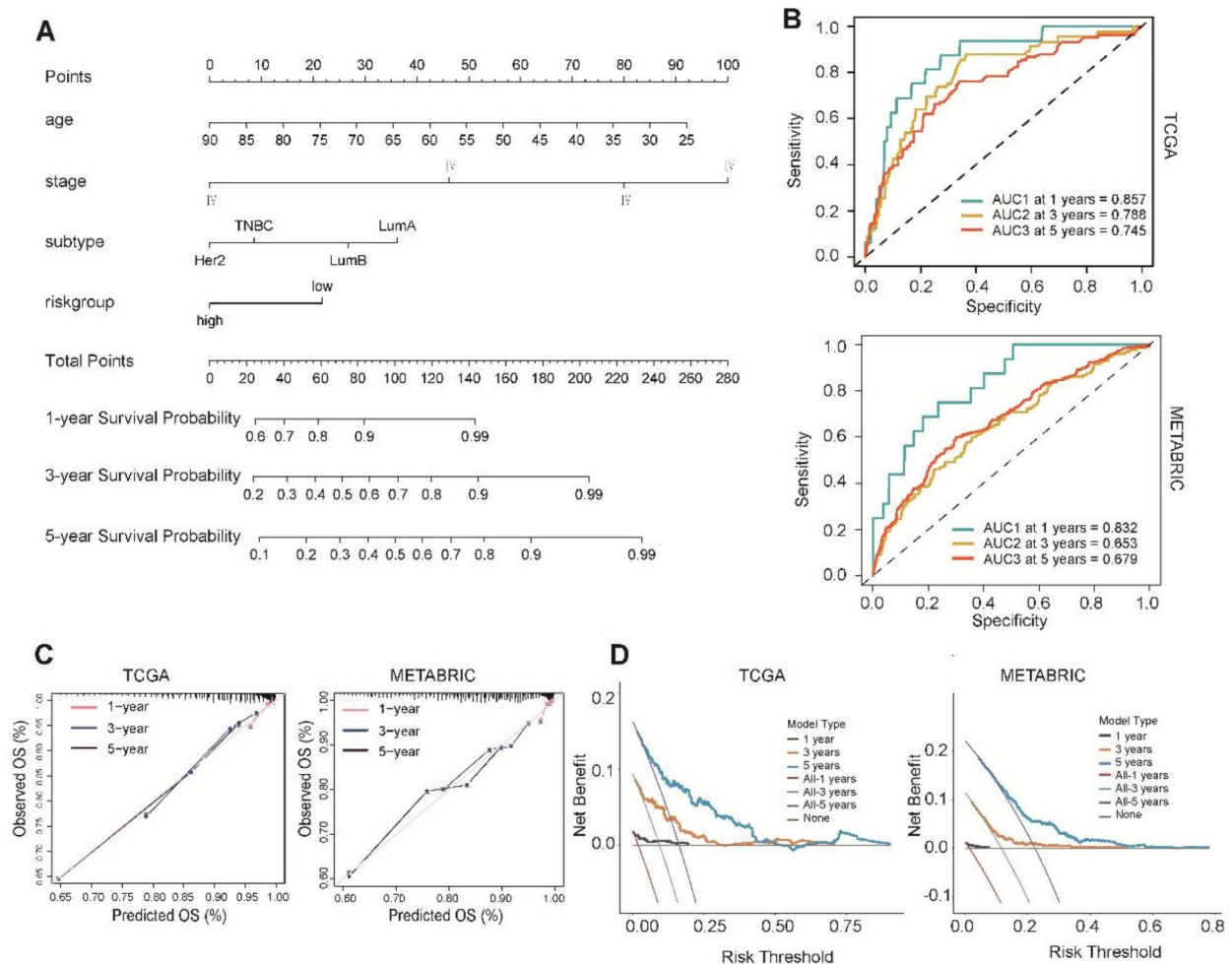
Fig. 2. Establishment of the prognostic nomogram in breast cancer with the risk group combining clinicopathological factors. (**A**) A nomogram for predicting 1-, 3-, and 5-year survival possibilities of individual patients with breast cancer. (**B**) Time-dependent ROC curves at 1 year, 3 years, and 5 years in TCGA and METABRIC datasets. (**C**) The calibration curves of 1-year, 3-year, and 5-year survival in TCGA and METABRIC datasets. The 45∘ dashed line represented a perfect uniformity between nomogram-predicted and real possibilities. (**D**) The DCA curves for 1-, 3-, and 5-year OS in TCGA and METABRIC datasets.

response profile, with the low-risk group showing superior sensitivity to nearly all six selected chemotherapeutic agents, except in the TNBC subgroup (Supplementary Fig. 6B). This discrepancy between cohorts may be attributed to inherent variations in patient characteristics, or methodological approaches. Further investigations are warranted to elucidate the underlying mechanisms and validate the generalizability of these findings across diverse clinical settings.

In the GSE20685 dataset, we observed that breast cancer patients with distant metastases had higher risk scores compared to those without (Fig. 3C). Furthermore, deceased patients exhibited higher risk scores than those who were still alive (Fig. 3D). In the GSE35640 dataset, patients with metastatic melanoma who responded to MAGE-A3 immunotherapy had significantly lower risk scores than non-responders (Fig. 3E). Similarly, within the IMvigor210 cohort, bladder cancer patients who responded to anti-PD-1/PD-L1 therapy had lower risk scores than non-responders (Fig. 3F). Collectively, these results indicate that patients with lower risk scores experience greater therapeutic benefits.

Single-cell analysis of immune-related prognostic gene expression

When comparing normal breast samples with different breast cancer subtypes, we applied harmony batch correction to mitigate batch effects (Supplementary Fig. 7). As shown in Fig. 4A, normal breast tissue and cancer cells share several cell subgroups, including epithelial cells, endothelial cells, T cells, myeloid cells, plasma cells, fibroblasts, and pericytes. In normal breast samples, epithelial cells are further subdivided into mature luminal epithelial cells, luminal progenitor cells, and basal cells. Different types of cancer samples each contained a subgroup of B cells. Specifically, in both ER-positive and Her2-positive cancer samples, we identified a group of normal basal cell subpopulations[20]. The stacked bar chart illustrates the proportion of cell types in each sample
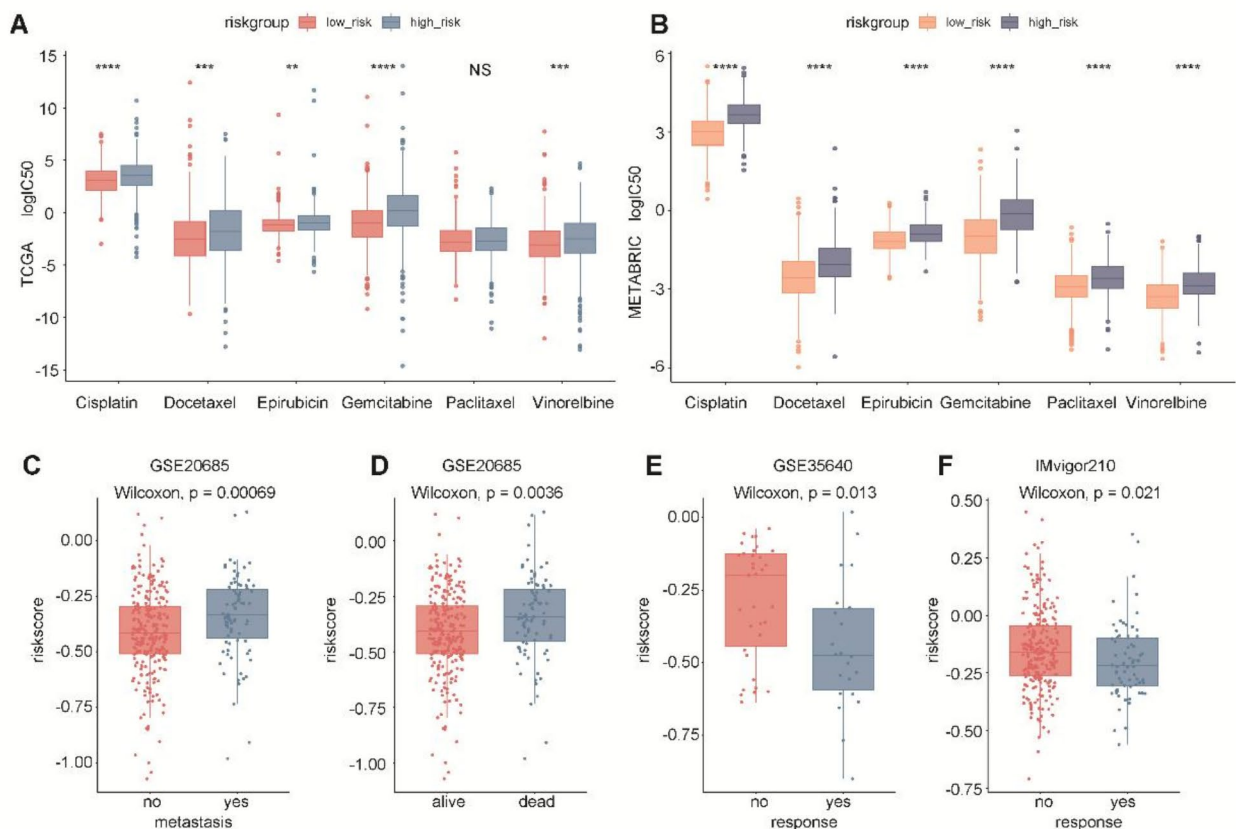
Fig. 3. The risk score predicts therapeutic benefits. Log IC50 (The half maximal inhibitory concentration) values of various chemotherapy drugs between high- and low-risk groups in in TCGA (**A**) and METABRIC (**B**) datasets. NS, not significant. In the GSE20685 dataset, breast cancer patients with distant metastases showed higher risk scores than those without metastases (**C**); deceased patients had higher risk scores than survivors (D). In the GSE35640 dataset, metastatic melanoma patients who responded to MAGE A3 immunotherapy had notably lower risk scores than non-responders (**E**). The same trend was observed in the IMvigor210 cohort, where bladder cancer patients responding to anti-PD-1/PD-L1 therapy had lower risk scores compared to those who did not respond (**F**). *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$.

(Supplementary Fig. 7). Figure 4B lists the marker genes associated with these subpopulations. Single-cell analysis revealed that the immune-related genes *CD2* and *CXCL13* are primarily expressed in T cells (Fig. 4C), which supports the accuracy of our findings. *SAA2* was mainly expressed in normal breast basal cells, while *LILRB5* was expressed at low levels, mainly in myeloid cells. *EPHB6* and *PPP1R16B* showed limited expression, mainly in T cells, whereas *TACR1* was detected in only a minority of cells (Fig. 4C). However, it's worth noting that the previous sections of this study reported *EPHB6* as one of the differentially expressed genes between breast cancer and benign breast tissues, with a prominent contribution to the OS of breast cancer patients. In contrast, the current single-cell analysis failed to uncover a distinct expression pattern for *EPHB6*. Given the potential for technical artifacts and data quality issues in single-cell sequencing, we plan to further validate EPHB6 expression in both benign and malignant breast tissues in subsequent sections of this study.
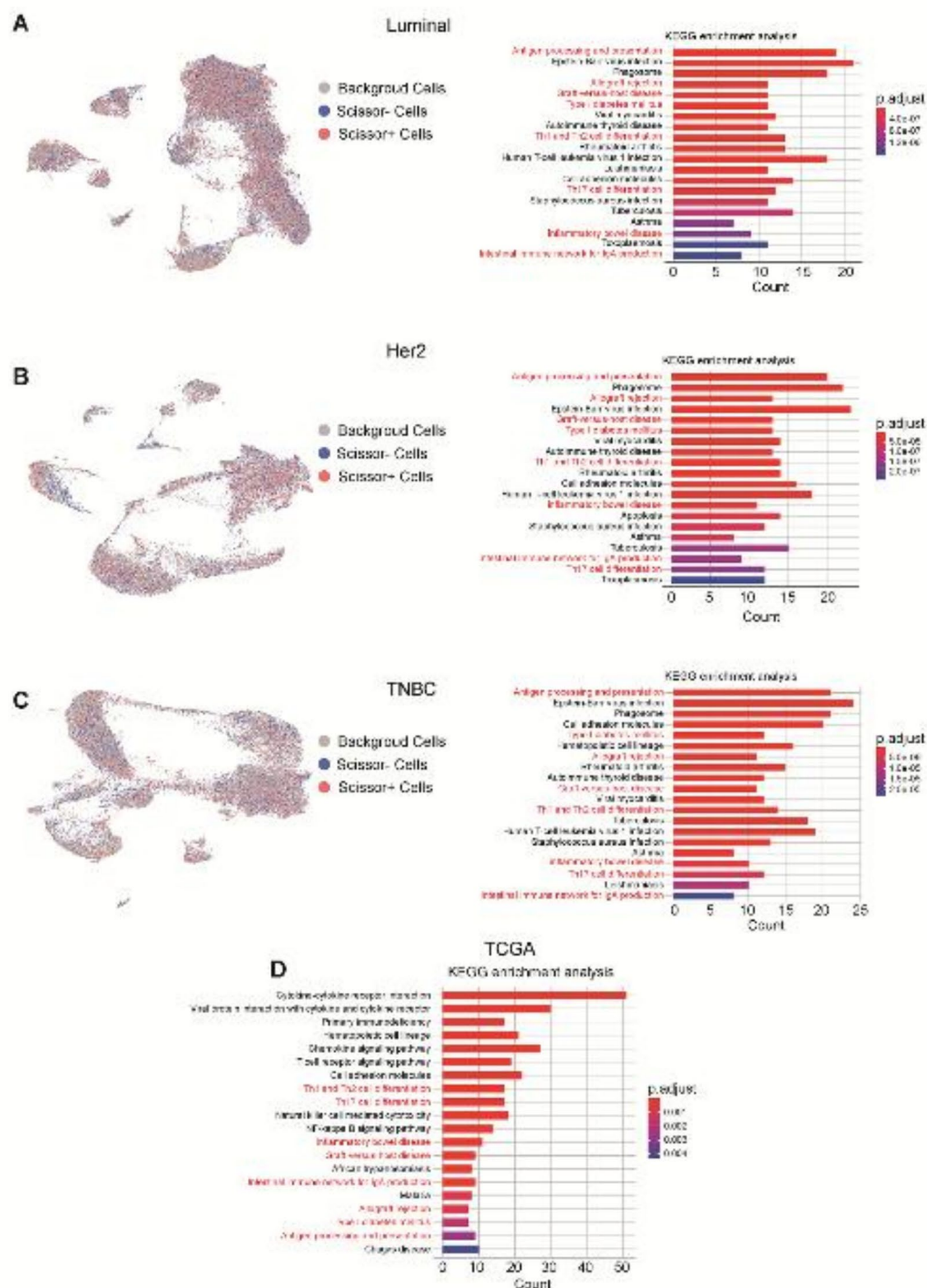
Functional analysis
To better understand the genetic differences between high-risk and low-risk groups, we conducted the KEGG pathway analysis. Using the advanced Scissor algorithm, we identified cell subgroups associated with each risk category and analyzed single-cell data across three subtypes of breast cancer. Specifically, Scissor (+) cells are associated with the high-risk group, while Scissor (-) cells are linked to the low-risk group. KEGG analysis of DEGs between Scissor (+) and Scissor (-) cells indicated that these genes were associated with several immune-related pathways (Fig. 5A-C). This pattern was consistently observed in the TCGA cohort as well, where DEGs between distinct risk groups similarly exhibited correlations with immune pathways (Fig. 5D). Importantly, we identified overlapping pathways at both the bulk and single-cell levels, which include Th1 and Th2 cell differentiation, Th17 cell differentiation, inflammatory bowel disease, graft-versus-host disease, IgA production within the intestinal immune network, allograft rejection, type I diabetes, and, finally, antigen processing and presentation. These findings underscore a robust relationship between risk groups and the involvement of the immune system in breast cancer. After integrating single-cell data from all cancer cells, we observed differences
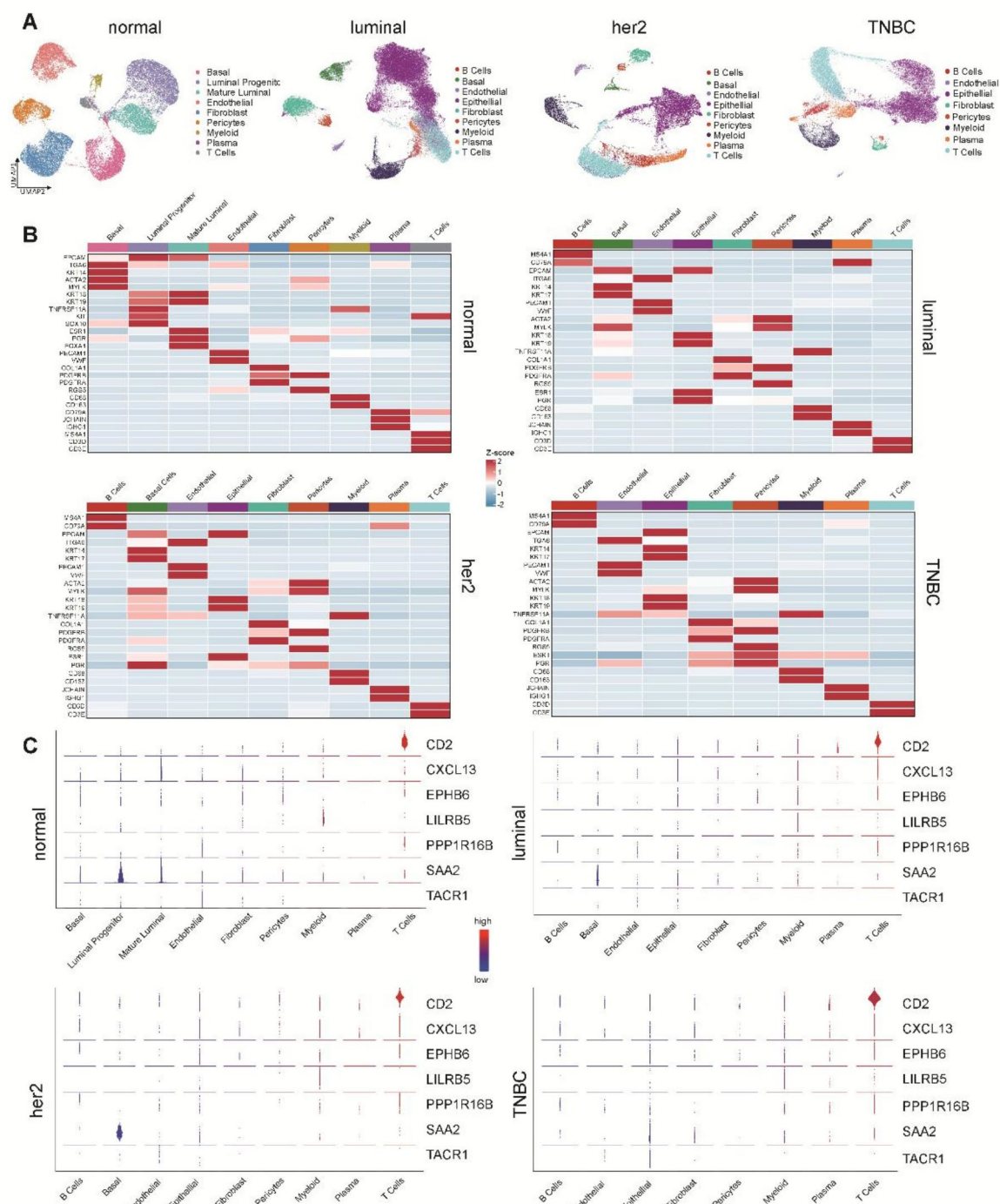
Fig. 4. Expressions of Prognostic Genes at the Single-Cell Level. (**A**) UMAP visualizations of cell-type-specific annotation in scRNA-seq dataset. (**B**) Heat maps show marker genes for each cell type. (**C**) Beeswarm plots display the expressions of prognostic genes in different cell types.

in the composition ratios of Scissor + and Scissor- cell populations within the immune microenvironment (Supplementary Fig. 8A-C). This pattern was also observed in Luminal, HER2-positive, and TNBC, respectively (Supplementary Fig. 8D-F). This finding offers additional context for the KEGG pathway enrichment analysis results (Table 1).

Fig. 5. Results of Scissor algorithm and functional analysis. At the single-cell level, UMAP visualizations of Scissor + and Scissor- cells and the results of KEGG pathway analysis in different subtypes of breast cancer samples, including Luminal (**A**), Her2 (**B**), and TNBC (**C**). (**D**) At the bulk level, the result of KEGG pathway analysis in TCGA dataset. The common pathways are highlighted in red font.

EPHB6 expression in benign and malignant breast tissues

In this section, we used IHC to assess *EPHB6* expression across a variety of breast tissues, including fibroadenomas, cancer tissues, and adjacent non-cancerous samples (Fig. 6A-F). Our IHC analysis revealed the presence of *EPHB6* in both epithelial cells and the surrounding tissue. Specifically, strong expression (3+) of *EPHB6* was

	level	Overall
n		58
age (mean SD)		53.76 (9.09)
type (%)	DCIS	9 (15.5)
	IDC	45 (77.6)
	ILC	4 (6.9)
T (%)	Tis	9 (15.5)
	1	25 (43.1)
	2	24 (41.4)
N (%)	0	40 (69.0)
	1	11 (19.0)
	2	3 (5.2)
	3	4 (6.9)
subtype (%)	Luminal A	21 (36.2)
	Luminal B	14 (24.1)
	Her2	12 (20.7)
	TNBC	11 (19.0)
EPHB6 (%)	1+	21 (36.2)
	2+	17 (29.3)
	3+	20 (34.5)

Table 1. Clinical and histopathologic characteristics of the patient cohort. DCIS: ductal carcinoma in situ; IDC: invasive ductal carcinoma; ILC: invasive lobular carcinoma; Her2, human epidermal growth factor receptor 2; N, lymph node metastasis; T, tumor size; TNBC, triple negative breast cancer.

observed in both adjacent normal tissues and fibroadenoma tissues (Fig. 6E-F). Among different types of breast cancer, TNBC stood out with the highest levels of *EPHB6* expression, scored as 3+, whereas other types generally displayed lower (1+) or moderate (2+) expression levels (Table 2). It is also worth mentioning that high EPHB6 expression in non-TNBC was exclusively associated with in situ carcinoma (Table 2). To further investigate *EPHB6* expression at the protein level, we performed an analysis of the integrated optical density (IOD) per area, which revealed a significant pattern in the expression of *EPHB6* (Fig. 6G): Firstly, no significant difference was observed in the expression of *EPHB6* between fibroadenomas and adjacent normal tissues. Secondly, the expression of *EPHB6* was significantly lower in malignant tumors compared to benign breast tissues. Lastly, and perhaps most notably, the expression of *EPHB6* was significantly higher in in situ carcinomas than in invasive cancers. Additionally, *EPHB6* expression was conspicuously higher in triple-negative breast cancer (TNBC) compared to non-TNBC cases (Fig. 6G). On the other hand, *EPHB6* mRNA levels were higher in normal breast tissue than in cancerous tissue, with TNBC samples showing significantly elevated EPHB6 expression compared to non-TNBC cases (Fig. 6H-J). These RNA-level findings are consistent with our protein-level results.

We also evaluated *EPHB6* expression in axillary lymph nodes. Notably, in nodes without cancer metastasis, prominent *EPHB6* expression was detected in the interstitium, indicating its importance for preserving the normal structure and function of the lymph nodes (Fig. 7A-C). In contrast, upon cancer cell invasion into the lymph nodes, *EPHB6* expression was virtually absent within the cancer cells (Fig. 7A-C). This finding strengthens the hypothesis that reduced *EPHB6* expression may be associated with the tumor's invasive nature and metastatic capacity. Interestingly, in a section of metastatic lymph node, both well-differentiated and poorly differentiated carcinomas were found. Notably, *EPHB6* expression was present in well-differentiated carcinomas but absent in poorly differentiated ones. (Fig. 7D). Nevertheless, further research is needed to validate this observation and explore the underlying mechanisms.

We then focused our efforts on a more in-depth exploration of *EPHB6* expression in distant breast cancer metastases. In the GSE46141 dataset, we analyzed the expression of *EPHB6* at six metastatic sites: liver, breast, lymph nodes, skin, bone, and lung. Our pairwise comparisons revealed a significant finding: the expression level of *EPHB6* was significantly lower in hepatic metastatic lesions compared to pulmonary lesions and was also markedly decreased in lymph nodes when compared to the lungs (Supplementary Fig. 9A). Conversely, upon pairwise examination, no significant differences were detected among the remaining metastatic sites (Supplementary Fig. 9A). In the GSE56493 dataset, which also included six metastatic sites of breast cancer (liver, breast, lymph nodes, skin, skeletal muscle, and lung/pleura), we found that *EPHB6* expression was significantly lower in hepatic metastatic lesions compared to those in the breast, skin, and lung/pleura (Supplementary Fig. 9B). This observation highlights significant variations in *EPHB6* expression across different metastatic sites, suggesting potential differences in tumor behavior or microenvironment. However, pairwise comparisons among the remaining metastatic sites failed to reveal any statistically significant differences (Supplementary Fig. 9B).

The GSE56493 dataset provides PAM50 subtype information of metastatic lesions. The metastatic lesions in skin and lymph nodes contain both basal and non-basal samples. Surprisingly, we found that in lymph node metastatic cancer, the expression level of *EPHB6* was significantly elevated in basal samples compared to non-
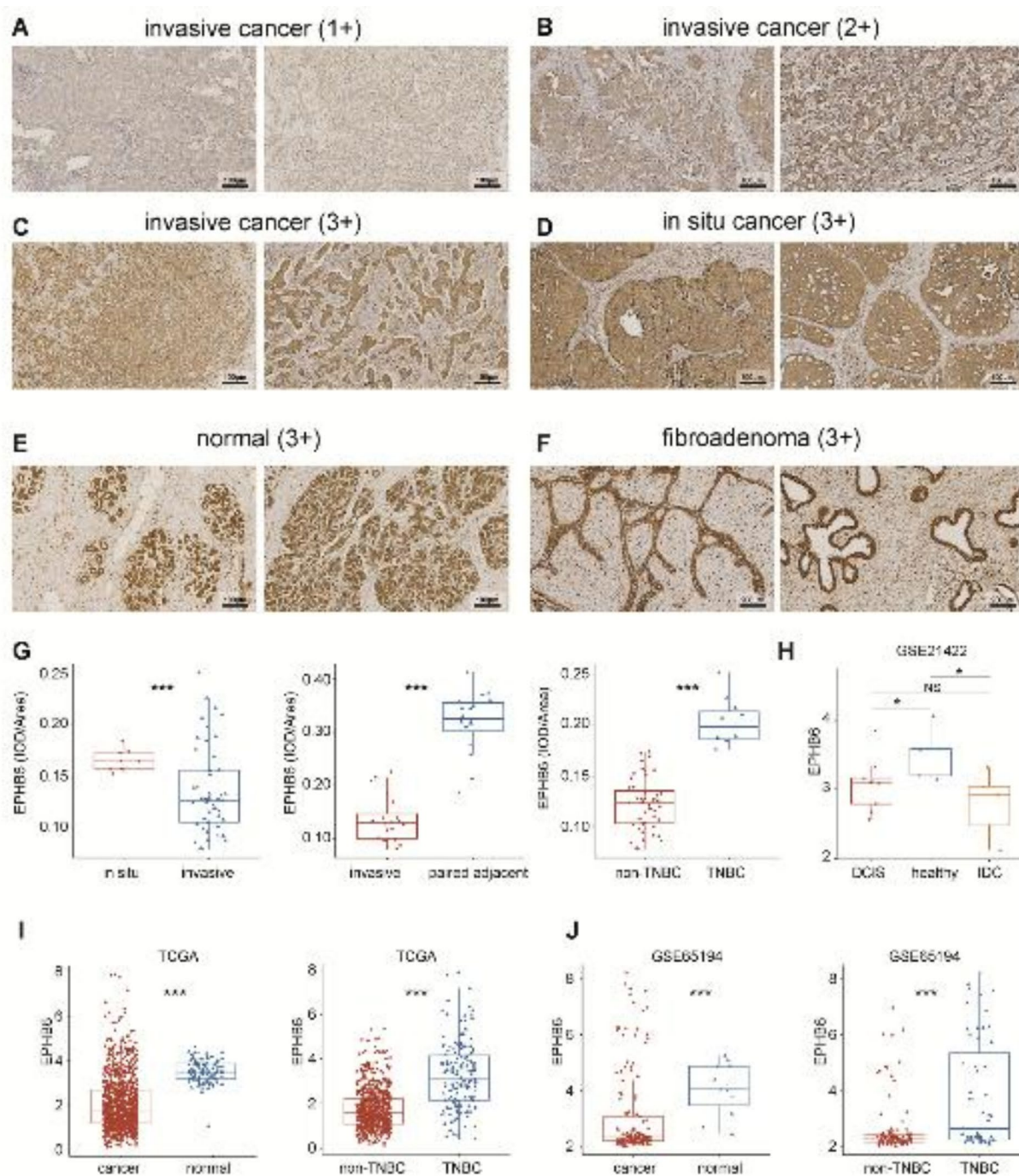
Fig. 6. *EPHB6* expression in benign and cancerous breast tissues. (**A**-**F**) Representative images of benign and cancerous breast samples stained for *EPHB6* expression by IHC: (**A**-**C**) invasive cancer samples were classified as weak (1+), intermediate (2+), and strong (3+). (**D**) In situ tumor samples with *EPHB6* overexpression (Strong 3+). (**E**) Adjacent non-cancerous tissues with *EPHB6* overexpression (Strong 3+). (**F**) Fibroadenoma samples with *EPHB6* overexpression (Strong 3+). (**G**) Quantification of *EPHB6* expression using integrated optical density/specimen area (IOD/area) in different breast tissues. (**H**-**J**) *EPHB6* mRNA expressions in publicly available datasets: GSE21422 (**H**), TCGA-BRCA(**I**), GSE65194(**J**). NS, not significant; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

basal ones (Supplementary Fig. 9C). This intriguing discovery may shed light on the potential role of *EPHB6* in specific tumor subtypes. In contrast, no significant difference in *EPHB6* expression was observed between basal and non-basal samples in skin metastatic cancer (Supplementary Fig. 9D). However, given the limitation of the sample size, these preliminary results warrant further validation through larger-scale studies.

	Weak (1+) (%)	Intermediate (2+) (%)	Strong (3+) (%)	total
Luminal A	10 (47.6)	6 (35.3)	5 (25.0)	21
Luminal B	4 (19.0)	7 (41.2)	3 (15.0)	14
Her2	7 (33.3)	4 (23.5)	1 (5.0)	12
TNBC	0 (0.0)	0 (0.0)	11 (55.0)	11
total	21	17	20	58

Table 2. *EPHB6* expression in breast carcinoma.

Discussion

Using bioinformatics tools, we thoroughly analyzed public breast cancer datasets to identify immune-related prognostic genes. We identified a gene set comprising seven pivotal genes that form an "immune signature". These genes exhibited strong correlations with different types of immune cells, including T cells and B cells, and were involved in a range of immune processes, suggesting a substantial impact on the immune microenvironment of breast cancer. Based on these genes, we calculated risk scores to categorize patients into high and low-risk groups. Both univariable and multivariable analyses confirmed that the classification into these risk groups serves as an independent prognostic factor. Functional analysis revealed that the differentially expressed genes (DEGs) between these groups are primarily involved in several immune-related pathways, which was further supported by a separate analysis of single-cell data. Remarkably, we discovered that patients in the low-risk group exhibited increased sensitivity to commonly prescribed chemotherapy drugs. Moreover, those who responded positively to immunotherapy had notably lower risk scores compared to non-responders. Although immunotherapy is primarily used for triple-negative breast cancer (TNBC), our findings indirectly suggest the potential for applying immunotherapy to non-TNBC. Pioneering studies conducted by researchers at Fudan University Shanghai Cancer Center have provided valuable insights in this regard. In HER2-positive breast cancer, they identified the immunomodulatory subtype (HER2-IM), which exhibits immune-activating characteristics and is considered suitable for immunotherapy and may respond well to such treatments[21]. For luminal breast cancer, the team discovered that supplementing with tyramine can reshape the immune microenvironment and enhance sensitivity to immunotherapy, opening up a new direction for the immunotherapy of this subtype by enhancing the immune response[22]. These studies have laid a solid foundation for precise immunotherapy in breast cancer, highlighting the significance of formulating individualized treatment plans based on molecular characteristics to improve patient outcomes and are expected to further optimize the treatment strategies for breast cancer.

Our research has shown that *EPHB6* contributes most significantly to breast cancer OS among immune-related signature genes. However, our knowledge of *EPHB6* expression in human tissues is still limited, given that only a handful of studies have been undertaken thus far. In studies focusing on colorectal cancer, it was found that *EPHB6* was highly expressed in normal colon tissue, whereas its expression decreased in cancer tissue, particularly in samples exhibiting lymph node metastasis, indicating its potential role in tumor invasiveness and metastasis[23,24]. Similarly, in prostate cancer research, it was reported that *EphB6* expression was moderate to strong in normal tissue but generally diminished in cancer tissue, often appearing negative or weak[25]. Moreover, in melanoma research, observations revealed that *EphB6* was highly expressed in benign nevi but underwent a significant decrease in both melanoma and metastatic tumors, with its expression being nearly absent in the latter[26]. As for gastric cancer, although some studies have demonstrated a positive correlation between *EphB6* expression and tumor differentiation, as well as a negative correlation with lymph node metastasis and tumor stage, this relationship has not been universally observed across all studies[27,28]. Regarding breast cancer, a previous investigation reported that *EphB6* protein was undetectable in three invasive breast cancer cell lines: MDA-MB-231, MDA-MB-435, and BT549[29]. In contrast, substantial *EphB6* protein levels were noted in non-invasive cell lines (MCF-7, BT-20, and SkBr3) and normal breast cell lines (MCF-10 A)[29].

Previous research on *EPHB6* expression in breast cancer had focused primarily on cellular models, but little is known about *EPHB6* expression in human breast tissues. Our study, therefore, for the first time, using IHC to bridge this gap, assessing *EPHB6* expression in both benign and malignant breast tissues. Notably, we observed overexpression of *EPHB6* protein in fibroadenomas and significantly higher expression levels in adjacent non-cancerous tissues compared to cancer tissues. Intriguingly, *EPHB6* expression in IDC was markedly elevated compared to other invasive cancer types, which is consistent with RNA-level analysis from public bulk data. Our findings hint at a potential association between decreased *EPHB6* expression and breast cancer progression. Of particular interest, triple-negative breast cancer, renowned for its aggressive nature[30], exhibited significantly higher *EPHB6* expression than other breast cancer types, suggesting that increased *EPHB6* expression in breast cancer might correlate with enhanced invasiveness. However, this observation stands in contrast to earlier cell line studies, which suggested that reduced expression of *EPHB6* could enhance the invasiveness of breast cancer cells[29,31,32].

These contradictory conclusions may stem from several sources. Firstly, in vitro cell lines might not comprehensively reflect the intricate nature of primary tumor tissues, resulting in discrepancies with human tissue studies. Secondly, the multifaceted roles of *EPHB6* in different cell types and stages of tumor development cannot be overlooked; it may inhibit invasion in some contexts while promoting it in others, depending on its interaction with signaling pathways and the cellular context. Additionally, the dynamic variation of *EPHB6* expression during tumor progression, which current cell line models might not fully capture, contributes to the complexity. Importantly, our immunohistochemical analysis emphasizes the intricacy of the human tumor microenvironment, demonstrating *EPHB6* expression in both epithelial cells and their surrounding
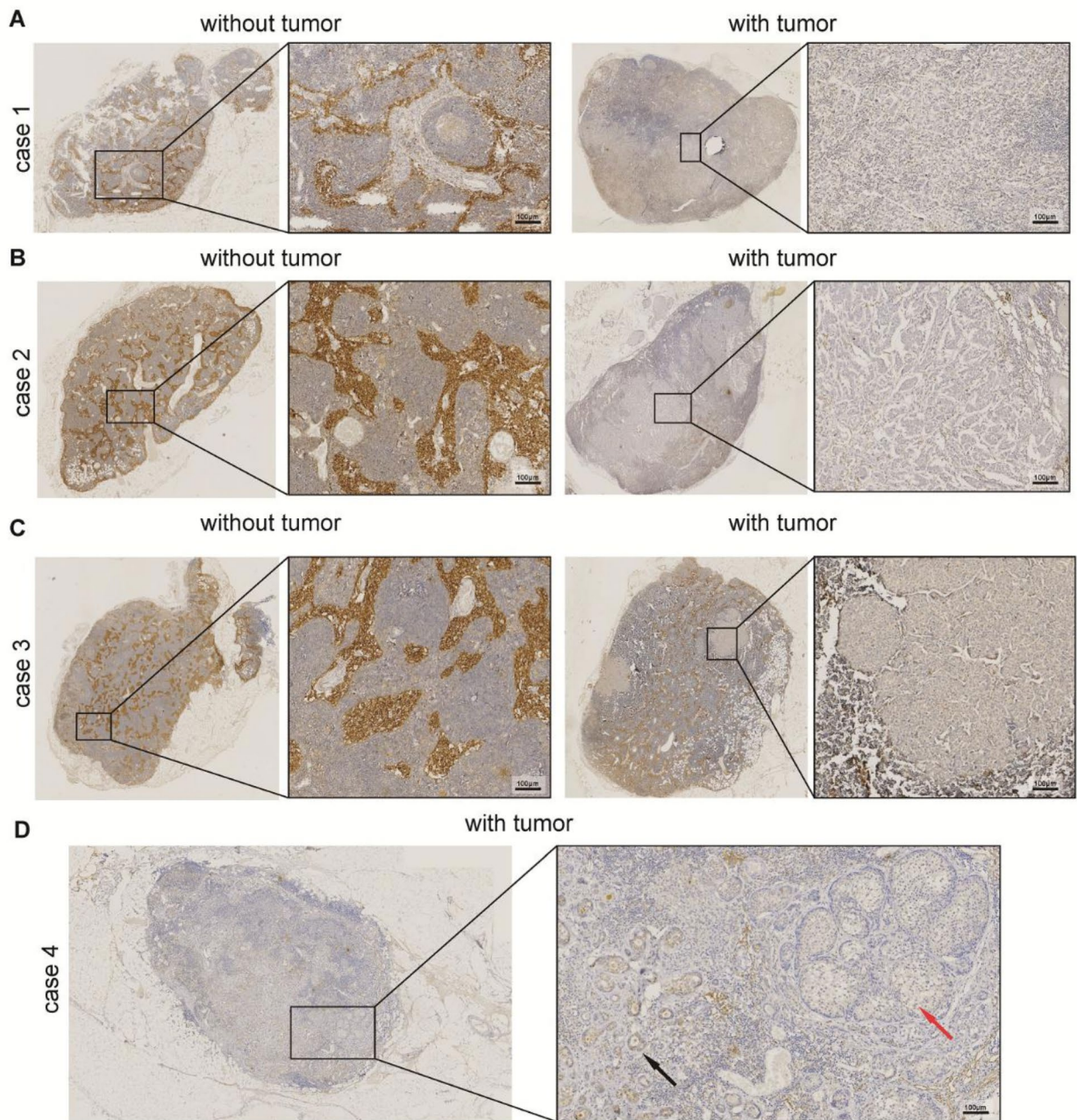
Fig. 7. *EPHB6* expression in lymph nodes. (**A**-**C**) The figures compare *EPHB6* protein levels in lymph nodes from three patients, each providing one node without cancer invasion and one with metastatic cancer. (**D**) The figure depicts *EPHB6* expression in a lymph node from the fourth patient, which had metastatic cancer. The node contains a mix of high-grade (black arrow) and low-grade cancer cells (red arrow). Notably, *EPHB6* is evident in the high-grade cells, whereas it is absent in the low-grade cells.

stroma, further highlighting potential contributors to the inconsistent findings regarding its relationship with invasiveness. Thus, to comprehensively elucidate *EPHB6*'s role in breast cancer, future studies should focus on its specific mechanisms within this complex microenvironment.

Additionally, we pioneered the confirmation of *EPHB6* expression in axillary lymph nodes. In these nodes, *EPHB6* is primarily expressed in stroma cells, which are crucial not only for the structural framework of the lymph nodes but also for guiding immune cells to their designated locations. Furthermore, stroma cells play a pivotal role in maintaining immune balance, fostering immune tolerance, and regulating the immune response to potential threats[33]. Interestingly, our study also revealed that upon cancer cell infiltration into the lymph nodes, *EPHB6* expression was absent in these cancerous cells. This finding could suggest that the level of *EPHB6*

expression is linked to the differentiation state and metastatic potential of tumor cells. Its downregulation may be associated with enhanced invasiveness and metastatic capabilities. Regarding the potential of *EPHB6* as a therapeutic target for breast cancer, it is noteworthy that previous reports have indicated that *EPHB6* enhances tumor sensitivity to DNA-damaging treatments in triple-negative breast cancer (TNBC), potentially by driving tumor-initiating cells into a more active division phase, making them more susceptible to treatment[34]. Consequently, *EPHB6* appears to be a promising potential therapeutic target for breast cancer, particularly in the context of TNBC.

In conclusion, the study identified a prognostic "immune signature" in breast cancer, with *EPHB6* as a key gene linked to tumor behavior and treatment response, suggesting its potential as a therapeutic target for personalized cancer treatment.

Limitations of the study

This study also has several limitations. Firstly, despite the use of multiple public databases and bioinformatics tools, the results are still contingent on existing datasets, which may contain selection bias and heterogeneity and may limit the generalizability of our findings. Secondly, our research was primarily based on retrospective analysis, which may introduce bias, and future prospective studies will help validate our findings. Additionally, while IHC detection offers insights into the expression pattern of *EPHB6* in benign and cancerous breast tissues, a larger sample size and multi-center studies are essential to further substantiate the results.

Materials and methods

Data acquisition

In our study, we utilized the TCGAbiolinks package to access TCGA-BRCA TPM (Transcripts Per Kilobase Million) data along with corresponding patient clinical profiles. Furthermore, we retrieved the METABRIC breast cancer dataset via cBioPortal (https://www.cbioportal.org)[35,36]. We applied the following data exclusion criteria: (1) genes with low expression, defined as those having an expression level of zero in more than 10% of the samples; (2) cases with incomplete clinical information; (3) patients with an OS time of less than 30 days; (4) male cases. We downloaded gene expression profile data (GSE65194, GSE21422, GSE20685, and GSE35640) and corresponding clinical information from the public Gene Expression Omnibus database (GEO, http://www.ncbi.nlm.nih.gov/geo/). Additionally, we obtained a single-cell dataset (GSE161529) from the GEO database, selecting 13 normal breast tissue samples, 6 Her2-positive breast cancer tissue samples, 17 ER-positive breast cancer tissue samples, and 8 triple-negative breast cancer tissue samples. We also downloaded the IMvigor210 cohort, accessible via the provided link (http://research-pub.gene.com/IMvigor210CoreBiologies/)[37]. To further enrich our dataset, we downloaded two additional GEO datasets: GSE46141 and GSE56493, which contain metastatic breast cancer data from multiple sites.

Identification of immune-related prognostic genes in breast cancer

We obtained immune and stromal scores for the TCGA-BRCA dataset from the ESTIMATE website (https://bioinformatics.mdanderson.org/estimate)[38]. We used the "surv_cutpoint" function within the "survminer" package to determine the optimal threshold, dividing breast cancer patients into high and low immune score groups, and high and low stromal score groups, for survival analysis. Subsequently, we identified differentially expressed genes (DEGs) between the high and low groups for both immune and stromal scores. Additionally, we retrieved genes differentially expressed between breast cancer and normal breast tissue from the GEPIA2 database (http://gepia2.cancer-pku.cn/)[39]. To discover the most predictive gene combination, we applied Lasso regression and the Gaussian mixture model (GMM). To explore the relationship of these genes with immune functions, we utilized the BEST platform (https://rookieutopia.com/app_direct/BEST), enabling us to perform tumor immune infiltration, gene ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. These analyses helped to elucidate the role of the identified gene set in immune-related processes[40].

Establishment and validation of the prognostic model

We developed a 'risk score' by integrating gene expression levels with their respective Lasso regression coefficients. The risk score is calculated as: Risk score = $(\beta_1 * X_1) + (\beta_2 * X_2) + \ldots + (\beta_i * X_i)$, where X_i is the expression level of each gene, and β_i is the corresponding Lasso regression coefficient. To categorize patients into high- and low-risk groups, we utilized the 'survival' package in R to determine the optimal cutoff. TCGA data served as our training dataset, while METABRIC data was employed for validation. Furthermore; we developed a nomogram that integrates the risk score and clinical characteristics to predict overall survival (OS). We assessed the predictive accuracy of our model using calibration plots, receiver operating characteristic (ROC) curves, and decision curve analysis (DCA) curves. These statistical methods evaluate the accuracy, discriminatory power, and clinical utility of our model, respectively.

Drug sensitivity prediction

We utilized the 'oncoPredict' package[41] to investigate whether patients in different risk groups have different responses to breast cancer chemotherapy medications. This specific package facilitated our assessment of the sensitivity of agents listed in the GDSC2 (Genomics of Drug Sensitivity in Cancer) database[42]. We employed a ridge regression model, specifically designed for breast cancer transcriptomic data, to evaluate the sensitivity and predict the half-maximal inhibitory concentration (IC50) of drugs for the two risk groups. A significance threshold of $p < 0.05$ was applied.

Transcriptome and clinical datasets were analyzed for information on chemotherapy and immune checkpoint Blockade therapies

We analyzed data from the GEO (Gene Expression Omnibus) repository to explore how risk scores correlate with treatment responses in patients. The GSE20685 dataset comprised samples from 327 individuals with breast cancer, among whom 268 received adjuvant chemotherapy and 91 relapsed. Furthermore, the GSE35640 dataset documented the treatment outcomes of metastatic melanoma patients undergoing MAGE A3 immunotherapy. Additionally, the IMvigor210 trial assessed the effectiveness of atezolizumab, a drug targeting the PD-L1 protein, in patients with advanced or metastatic urothelial bladder cancer.

Single-cell analysis

In this part, we utilized Seurat v4[43], adhering to the data quality control guidelines outlined by the 'scCancer' package[44]. The "DoubletFinder" package[45] assisted in eliminating doublets from our dataset, while the "Harmony" package[46] facilitated the integration of data from multiple samples. Drawing upon markers identified in previous studies[20,47], we were able to delineate distinct cell subgroups. To distinguish cell subpopulations associated with bulk sample phenotypes, we employed the scissor algorithm (Single-cell identification of subpopulations with bulk sample phenotype correlation)[48]. For this study, the input data consisted of GSE161529 scRNA-seq data and TCGA-BRCA data, with the risk group serving as the phenotype feature that corresponded to the TCGA samples. Consequently, we classified the single cells into Scissor (+) and Scissor (-) groups, representing cells associated with high- and low-risk groups, respectively. By leveraging the "Findmarkers" function, we identified differentially expressed genes (DEGs) between these two cell subgroups, applying the following parameters: min.pct = 0.1, logfc.threshold = 0.25, p-adjust < 0.05. Subsequently, we subjected these DEGs to KEGG pathway analysis using the "clusterProfiler" package[49].

Clinical sample collection

In our study, we collected 12 fibroadenoma samples from patients at the Breast and Thyroid Surgery Department of Zibo Maternal and Child Health Hospital. Additionally, 58 clinical breast cancer samples were gathered. Among the breast cancer patients, from 20 of them, we obtained paired paracancerous tissues, ensuring a distance of at least 5 cm from the tumor margin (Table 1, Supplementary Table S1). Furthermore, we sampled one cancerous and one non-cancerous lymph node from each of six patients with axillary lymph node metastasis. To guarantee the quality of our data, we applied stringent criteria for selecting breast cancer patients for the study. Eligible patients were those who did not previously receive any treatment and underwent a modified radical mastectomy for breast cancer. The study was approved by the Ethics Review Committee of Zibo Maternal and Child Health Hospital, all methods were performed in accordance with the relevant guidelines and regulations, and all participants provided written informed consent.

Immunohistochemical (IHC) staining and scoring

Immunohistochemical experiments were conducted following a standard protocol. Tissue sections embedded in paraffin were first treated with xylene to remove the wax, then rehydrated using a series of ethanol solutions. Antigen retrieval was performed with an EDTA buffer at a pH of 9.0 in a DAKO PT Link device, heated to 97 °C for 20 min. After cooling to 65 °C, the samples were removed and washed with Tris-buffered saline. Endogenous peroxidase activity was inactivated with 3% hydrogen peroxide. The primary antibody used was a mouse polyclonal anti-EphB6 antibody sourced from Abnova, diluted to a 1:1000 ratio. The sections were then incubated with this antibody overnight at 4 °C, followed by a wash in Tris-buffered saline. Subsequently, the sections were treated with a secondary antibody, anti-mouse IgG from Dako, for 20 min at room temperature. The final step was color development using a DAB chromogen for one minute. The slides were evaluated independently by three pathologists for the intensity and distribution of the staining signal. The intensity of IHC staining was scored as negative (score 0), weak (1+), intermediate (2+), and strong (3+). The average Integrated Optical Density (IOD) of three areas randomly selected from the acquired images was analyzed using Image-Pro Plus 6 (Supplementary Table S2-4).

Statistical analysis

All statistical analyses and graphs in our study were carried out using R software, version 4.2.2, with the help of specific R packages designed for our analytical needs. We determined overall survival (OS) rates using the Kaplan-Meier method and evaluated differences in these rates with the log-rank test. For nonparametric comparisons, we applied the Wilcoxon test for pairwise analysis and the Kruskal-Wallis test when dealing with multiple groups. This thorough approach enabled us to precisely assess statistical significance within our data set.

Data availability

The dataset of TCGA-BRCA is available at the TCGA database (https://cancergenome.nih.gov/). We obtained the dataset by using the TCGAbiolinks package in R. The datasets generated and/or analyzed during the current study are available at GEO: GSE65194, GSE21422, GSE20685, GSE35640, and GSE161529. The dataset of IMvigor210 is available at http://research-pub.gene.com/IMvigor210CoreBiologies/#transcriptome-wide-gene-expression-data. The IHC images originally generated during this study cannot be publicly disclosed due to ethical and privacy concerns related to the patient samples involved. Nevertheless, they may be obtained upon a reasonable request to the corresponding author.

References

1. Han, L. et al. LncRNA HOTTIP facilitates the stemness of breast cancer via regulation of miR-148a-3p/WNT1 pathway. *J. Cell. Mol. Med.* **24**, 6242–6252 (2020).
2. Zuo, S., Yu, J., Pan, H. & Lu, L. Novel insights on targeting ferroptosis in cancer therapy. *Biomark. Res.* **8**, 1–11 (2020).
3. Hanahan, D. & Coussens, L. M. Accessories to the crime: Functions of cells recruited to the tumor microenvironment. *Cancer Cell.* **21**, 309–322 (2012).
4. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).
5. Tekpli, X. et al. An independent poor-prognosis subtype of breast cancer defined by a distinct tumor immune microenvironment. *Nat. Commun.* **10**, (2019).
6. Baxevanis, C. N., Fortis, S. P. & Perez, S. A. The balance between breast cancer and the immune system: Challenges for prognosis and clinical benefit from immunotherapies. *Semin Cancer Biol.* **72**, 76–89 (2021).
7. Chen, D. S. & Mellman, I. Oncology Meets immunology: The cancer-immunity cycle. *Immunity* **39**, 1–10 (2013).
8. Al, B. M. E. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* **24**, 541–550 (2018).
9. Loi, S. et al. Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: Results from the FinHER trial. *Ann. Oncol.* **25**, 1544–1550 (2014).
10. Adams, S. et al. Prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancers from two phase III randomized adjuvant breast cancer trials: ECOG 2197 and ECOG 1199. *J. Clin. Oncol.* **32**, 2959–2966 (2014).
11. Sharma, P. & Allison, J. P. The future of immune checkpoint therapy. *Sci. (80-).* **348**, 56–61 (2015).
12. Pei, S. et al. Integrating single-cell RNA-seq and bulk RNA-seq to construct prognostic signatures to explore the role of glutamine metabolism in breast cancer. *Front. Endocrinol. (Lausanne).* **14**, 1–17 (2023).
13. Pei, S. et al. Exploring the role of sphingolipid-related genes in clinical outcomes of breast cancer. *Front. Immunol.* **14**, 1–18 (2023).
14. Wei, Q. et al. Structures of an Eph receptor tyrosine kinase and its potential activation mechanism. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70**, 3135–3143 (2014).
15. Kania, A. & Klein, R. Mechanisms of ephrin-Eph signalling in development, physiology and disease. *Nat. Rev. Mol. Cell. Biol.* **17**, 240–256 (2016).
16. Rudno-Rudzińska, J. et al. A review on Eph/ephrin, angiogenesis and lymphangiogenesis in gastric, colorectal and pancreatic cancers. *Chin. J. Cancer Res.* **29**, 303–312 (2017).
17. Shiuan, E. & Chen, J. Eph receptor tyrosine kinases in tumor immunity. *Cancer Res.* **76**, 6452–6457 (2016).
18. Yamada, T. et al. After repeated division, bone marrow stromal cells express inhibitory factors with osteogenic capabilities, and EphA5 is a primary candidate. *Bone* **57**, 343–354 (2013).
19. Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
20. Pal, B. et al. A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic States in the human breast. *EMBO J.* **40**, 1–23 (2021).
21. Li, Y. W. et al. Molecular characterization and classification of HER2-Positive breast Cancer inform tailored therapeutic strategies. *Cancer Res.* **21**, 3669–3683 (2024).
22. Cai, Y. W. et al. MAP3K1 mutations confer tumor immune heterogeneity in hormone receptor–positive HER2-negative breast cancer. *J. Clin. Invest.* **135** (2), e183656 (2024).
23. Xu, D. et al. EphB6 overexpression and apc mutation together promote colorectal cancer. *Oncotarget* **7**, 31111–31121 (2016).
24. Mateo-Lozano, S. et al. Loss of the EPH receptor B6 contributes to colorectal cancer metastasis. *Sci. Rep.* **7**, 1–12 (2017).
25. Mohamed, E. R. et al. Reduced expression of erythropoietin-producing hepatocyte B6 receptor tyrosine kinase in prostate cancer. *Oncol. Lett.* **9**, 1672–1676 (2015).
26. Hafner, C. et al. Loss of EphB6 expression in metastatic melanoma. *Int. J. Oncol.* **23**, 1553–1559 (2003).
27. Liu, J. et al. Reduced EphB6 protein in gastric carcinoma and associated lymph nodes suggests EphB6 as a gastric tumor and metastasis inhibitor. *Cancer Biomarkers.* **19**, 241–248 (2017).
28. Nakagawa, M. et al. Erythropoietin-Producing hepatocellular A1 is an independent prognostic factor for gastric Cancer. *Ann. Surg. Oncol.* **22**, 2329–2335 (2015).
29. Fox, B. P. & Kandpal, R. P. EphB6 receptor significantly alters invasiveness and other phenotypic characteristics of human breast carcinoma cells. *Oncogene* **28**, 1706–1713 (2009).
30. Bardia, A. et al. Sacituzumab Govitecan-hziy in refractory metastatic Triple-Negative breast Cancer. *N Engl. J. Med.* **380**, 741–751 (2019).
31. Bhushan, L. et al. Modulation of liver-intestine Cadherin (Cadherin 17) expression, ERK phosphorylation and WNT signaling in EPHB6 Receptor-expressing MDA-MB-231 cells. *Cancer Genomics Proteom.* **11**, 239–250 (2014).
32. Akada, M., Harada, K., Negishi, M. & Katoh, H. EphB6 promotes Anoikis by modulating EphA2 signaling. *Cell. Signal.* **26**, 2879–2884 (2014).
33. Krishnamurty, A. T. & Turley, S. J. Lymph node stromal cells: Cartographers of the immune system. *Nat. Immunol.* **21**, 369–380 (2020).
34. Toosi, B. M. et al. EPHB6 augments both development and drug sensitivity of triple-negative breast cancer tumours. *Oncogene* **37**, 4073–4093 (2018).
35. Gao, J. et al. Integrative analysis of complex Cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1 (2013).
36. Cerami et al. The cBio Cancer genomics portal: An open platform for exploring multidimensional Cancer genomics data. *Cancer Discov.* **32**, 736–740 (2017).
37. Mariathasan, S. et al. TGF-β attenuates tumour response to PD-L1 Blockade by contributing to exclusion of T cells. **554**, 544–548 (2018).
38. Yoshihara, K. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, (2013).
39. Tang, Z., Kang, B., Li, C., Chen, T. & Zhang, Z. GEPIA2: An enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* **47**, W556–W560 (2019).
40. Liu, Z. et al. BEST: a web application for comprehensive biomarker exploration on large-scale data in solid tumors. *J. Big Data* **10**, (2023).
41. Maeser, D., Gruener, R. F. & Huang, R. S. OncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Brief. Bioinform.* **22**, 1–7 (2021).
42. Yang, W. et al. Genomics of drug sensitivity in Cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41**, 955–961 (2013).
43. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587e29 (2021).
44. Guo, W. et al. ScCancer: A package for automated processing of single-cell RNA-seq data in cancer. *Brief. Bioinform.* **22**, 10–11 (2021).
45. McGinnis, C. S., Murrow, L. M., Gartner, Z. J. & DoubletFinder Doublet detection in Single-Cell RNA sequencing data using artificial nearest neighbors. *Cell. Syst.* **8**, 329–337e4 (2019).

46. Korsunsky, I. et al. Fast, sensitive, and accurate integration of single cell data with harmony. *Nat. Methods.* **16**, 1289–1296 (2019).
47. Wu, S. Z. et al. A single-cell and spatially resolved atlas of human breast cancers. *Nat. Genet.* **53**, 1334–1347 (2022).
48. Sun, D. et al. Identifying phenotype-associated subpopulations by integrating bulk and single-cell sequencing data. *Nat. Biotechnol.* **40**, 527–538 (2022).
49. Yu, G., Wang, L. G., Han, Y., He, Q. Y. & ClusterProfiler An R package for comparing biological themes among gene clusters. *Omi J. Integr. Biol.* **16**, 284–287 (2012).

Acknowledgements

Author contributions

Hui Lyu: Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Investigation, Data curation, Conceptualization. Tao Zhou: Writing – original draft, Resources, Investigation, Formal analysis. Xiaoqin Sun: Writing – original draft, Methodology, Investigation, Data curation. Hui Chen: Validation. Jing Li: Resources. Mingxiu Shao: Investigation. Jianmei Li: Investigation. Quanmei Zhang: Data curation. Guosheng Jiang: Writing – review & editing, Validation, Conceptualization. Xin Zhou: Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Investigation, Data curation, Conceptualization.

Funding

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval and consent to participate

This study was reviewed and approved by the Ethics Review Committee of Zibo Maternal and Child Health Hospital (approval no. 202106073, data: 2022-06-23). Patient informed consent was obtained as part of surgical consent at the time of surgery for scientific research. The patient's information was kept confidential.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-91318-z.

Correspondence and requests for materials should be addressed to G.J. or X.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.