Original Article

A novel 12-gene prognostic signature in breast cancer based on the tumor microenvironment

Jiujun Zhu[1#], Yong Shen[2#], Lina Wang[1], Jianghua Qiao[1], Yajie Zhao[1], Qiming Wang[3]

[1]Department of Breast Disease, Henan Breast Cancer Center, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China; [2]Department of Clinical Laboratory, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China; [3]Department of Medical Oncology, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China

Contributions: (I) Conception and design: J Zhu; (II) Administrative support: Q Wang; (III) Provision of study materials or patients: All authors; (IV) Collection and assembly of data: Y Shen; (V) Data analysis and interpretation: Y Shen, Y Zhao; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Qiming Wang. Department of Medical Oncology, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China. Email: qimingwang1006@126.com; Jiujun Zhu. Department of Breast Disease, Henan Breast Cancer Center, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China. Email: zlyyzhujiujun3033@zzu.edu.cn.

Background: The progression of breast cancer (BC) is highly dependent on the tumor microenvironment. Inflammation, stromal cells, and the immune landscape have been identified as significant drivers of BC in multiple preclinical studies. Therefore, this study aimed to clarify the predictive relevance of stromal and immune cell-associated genes in patients suffering from BC.

Methods: We employed the estimation of stromal and immune cells in malignant tumor tissues using expression data (ESTIMATE) algorithm to calculate the stromal and immunological scores, which were then used to evaluate differentially expressed genes (DEGs) in BC samples using The Cancer Genome Atlas (TCGA) database. Univariate analyses were conducted to identify the DEGs linked to survival in BC patients. Next, the prognostic DEGs (with a log-rank P<0.05) were used to create a risk signature, and the least absolute shrinkage and selection operator (LASSO) regression method was used to analyze and optimize the risk signature. The following formula was used to compute the prognostic risk score values: Risk score = Gene 1 * β1 + Gene 2 * β2 +… Gene n * βn. The median prognostic risk score values were used to divide BC patients into the low-risk (LR) and high-risk (HR) groups. The patient samples of the validation cohort were then assessed using this formula. We used principal component analysis (PCA) to determine the expression patterns of the different patient groups. Gene Set Enrichment Analysis (GSEA) was used to determine whether there were significant variations between the groups in the evaluated gene sets.

Results: The present study revealed that DEGs linked with survival were closely associated with immunological responses. A prognostic signature was constructed that consisted of 12 genes (*ASCL1*, *BHLHE22*, *C1S*, *CLEC9A*, *CST7*, *EEF1A2*, *FOLR2*, *KLRB1*, *MEOX1*, *PEX5L*, *PLA2G2D*, and *PPP1R16B*). According to their survival, BC patients were separated into LR and HR groups using the identified 12-gene signature. The immunological status and immune cell infiltration were observed differently in the LR and HR groups.

Conclusions: Our results provide novel insights into several microenvironment-linked genes that influence survival outcomes in patients with BC, which suggests that these genes could be candidate therapeutic targets.

Keywords: Breast cancer (BC); tumor microenvironment (TME); gene signature; bioinformatics analyses

Page 2 of 13

Zhu et al. Gene signature based on TME in BC

Introduction

Globally, breast cancer (BC) is a highly prevalent form of carcinoma. Over the last two decades, advances in early diagnostic tools and treatments have reduced BC mortality rates by a factor of three when compared to the incidence in 1990 (1,2). However, BC is still a major cause of mortality and threatens women's health throughout the world.

Growing evidence shows that BC tumors are highly heterogeneous. The local tumor microenvironment (TME) is critical for cancer progression, and it is becoming clear that the local TME plays a crucial role in tumor growth (3,4), including BC initiation, progression, metastasis as well as drug resistance(5). This dynamic TME includes endothelial progenitor cells (EPCs), stromal and immunological cells, complex extracellular matrix (ECM), and a wide range of growth factors (GFs) and cytokines (6). Stromal and immunological cells are key cells that promote tumor progression and metastasis. The crosstalk between stromal and immunological cells in the TME has been considered as another key factor in promoting tumor progression. For example, cancer-associated fibroblasts interact with tumor-infiltrating immune cells as well as other immune components over time by secreting various cytokines, growth factors, chemokines, ectoplasmic and other immune molecules, thus creating an immunosuppressive TME that allows cancer cells to evade surveillance by the immune system(7).Various tumor cell characteristics, such as chemotaxis and survival, can be influenced by proteins and the microenvironment of the cells. Recently, several immunomodulatory pharmacological methods have been reported for the treatment of BC (8,9). In order to improve the prognosis of BC and provide reliable information to guide individual treatment strategies, there is an urgent need to screen for reliable TME-related prognostic indicators that can be used clinically for patient management.

The estimation of stromal and immune cells in malignant tumor tissues using expression data (ESTIMATE) algorithm uses datasets of gene expressions to evaluate stromal and immunological cell infiltration in tumor tissues (10). This approach has previously been used to assess the microenvironment composition of colon cancer (11), prostate cancer (12), and glioblastoma (13). Although, there were several reports about the immune cell-associated genes or immune and stroma related genes in patients suffering from BC (14-16), study focus on signature based on both stromal and immunological cell infiltration in BC is still absent to date.

In this study, we used The Cancer Genome Atlas (TCGA) genome expression profiles and the ESTIMATE algorithm to determine immune/stromal scores for BC patients, resulting in the discovery of a group of microenvironment-related genes linked to the overall survival (OS) of BC patients. The genes were then used to create a gene signature associated with patient survival outcomes and investigate the importance of this profile to the immunological response and immune cell infiltration. In conclusion, our findings shed new light on the BC microenvironment, also suggest possible prognostic and therapeutically important gene targets in these cancers.

We present the following article in accordance with the TRIPOD reporting checklist (available at https://atm.amegroups.com/article/view/10.21037/atm-21-6748/rc).

Methods

Sample datasets

TCGA database was used to obtain raw mRNA expression data as well as clinical information from BC patients. A total of 1,069 BC samples (combined set) were randomly separated into training and validation sets of equal size. The validation set was used for validating the findings of the training set. Because all of the data in this report was derived from public sources, no ethical oversight was necessary. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Evaluation of differentially expressed genes (DEGs) according to the immune/stromal scores

The sample stromal and immunological scores were estimated using the ESTIMATE algorithm. The ESTIMATE R program (www.bioconductor.org/packages/release/BiocViews.html) was used to generate stromal, immune, and ESTIMATE scores by Gene Set Enrichment Analysis (GSEA). Patients were separated into low-risk (LR) and high-risk (HR) groups based on whether their immune/stromal scores were above or below the median value. To discover the immune/stromal score-related DEGs, the expression profiles of BC patients with a high immunological/stromal score were compared to those of BC patients with a low score. The data were then analyzed using the "edgeR" package (www.bioconductor.org/packages/release/bioc/html/edgeR.html), with DEGs classified as those that met the following criteria: false

discovery rate (FDR) <0.05 and |log2 fold change| >1.

Functional enrichment analysis of DEGs

Functional enrichment analyses were carried out to evaluate key cascades in which the identified genes were enriched. The "org.Hs.eg.db" package (www.bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html) was initially used to convert gene symbols into entrez IDs, after which Gene Ontology (GO) analyses were conducted using the "clusterProfiler", "ggplot2", and "enrichplot" packages (www.bioconductor.org/packages/release/BiocViews.html), with FDR <0.05 as the significance threshold. A Kyoto Gene and Genome Encyclopedia (KEGG) pathway enrichment analysis was also conducted. A P<0.05 was considered statistically significant.

Development and validation of a DEG (survival-associated) prognostic signature

Univariate Cox regression analyses were carried out to find DEGs associated with the survival of BC patients by using the "survival" R package (www.bioconductor.org/packages/survival). Next, the prognostic DEGs (with a log-rank P<0.05) were used to create a risk signature, and the least absolute shrinkage and selection operator (LASSO) regression method was used to analyze and optimize the risk signature. The following formula was used to compute the prognostic risk score values: Risk score = Gene 1 * β1 + Gene 2 * β2 +… Gene n * βn. The β symbol represents a regression coefficient calculated from the training dataset for each gene of interest. After that, the median prognostic risk score values were used to divide BC patients into the LR and HR groups. The patient samples of the validation cohort were then assessed using this formula. The prognostic relevance of the signature risk score was determined using univariate analyses of patients in both cohorts, followed by multivariate analysis to determine its independent prognostic value in these two cohorts.

Bioinformatics analysis

We used principal component analysis (PCA) to determine the expression patterns of the different patient groups. Gene set enrichment analysis (GSEA) was used to determine whether there were significant variations between the groups in the evaluated gene sets. We focused on two gene sets (immune system process, M13664, and immunological

response, M19817) from the Molecular Signatures Database v4.0 (http://www.broadinstitute.org/gsea/msigdb/index.jsp). We also used the Molecular Signatures Database (MSigDB) derived from the KEGG gene sets. To establish significance, we used a normalized enrichment score (NES) <0.05 and an FDR <0.05.

Immune infiltration analysis

The cell-type identification by estimating relative subsets of RNA transcripts (CIBERSORT) R v.12 (www.bioconductor.org/packages/release/BiocViews.html) was utilized to transform the gene expression profiling of BC into the fraction of 22 tumor-infiltrating immune cells in the given samples. One-tailed *t*-tests were used to compare the characteristics between groups.

Statistical analysis

Statistical evaluations were conducted using R v3.6.3 (http://www.Rproject.org). The "glmnet" software was used to run the LASSO regression analysis. After weighting by gene-specific regression coefficient (β) values, the final gene signature was specified as the accumulative individual gene expression. Kaplan-Meier (K-M) curves and log-rank tests were used to evaluate the rate of OS. Uni- and multivariate Cox regression analyses were carried out to evaluate the prognostic markers. The survival receiver operator curve (ROC) program was used for the ROC analysis (time-dependent). A P value <0.05 was considered statistically significant.

Results

Cohort characteristics

TCGA gene expression data were obtained from 1,069 patients suffering from BC, as represented in *Table 1*. A total of 17.12%, 56.88%, 22.45%, and 2.05% of patients had stage I, II, III, and IV tumors, respectively, and the median patient age was 58.08 years.

Association of clinical features with stromal/immune scores

The ESTIMATE algorithm was employed to evaluate the stromal, immune, and ESTIMATE scores ranging between –2,065.59–2,109.48, –1,277.91–3,672.57, and –2,916.86–5,355.63, respectively (*Figure 1* and Table S1).

Page 4 of 13

Zhu et al. Gene signature based on TME in BC

Table 1 Clinical characteristics of the combined, training, and validation cohorts

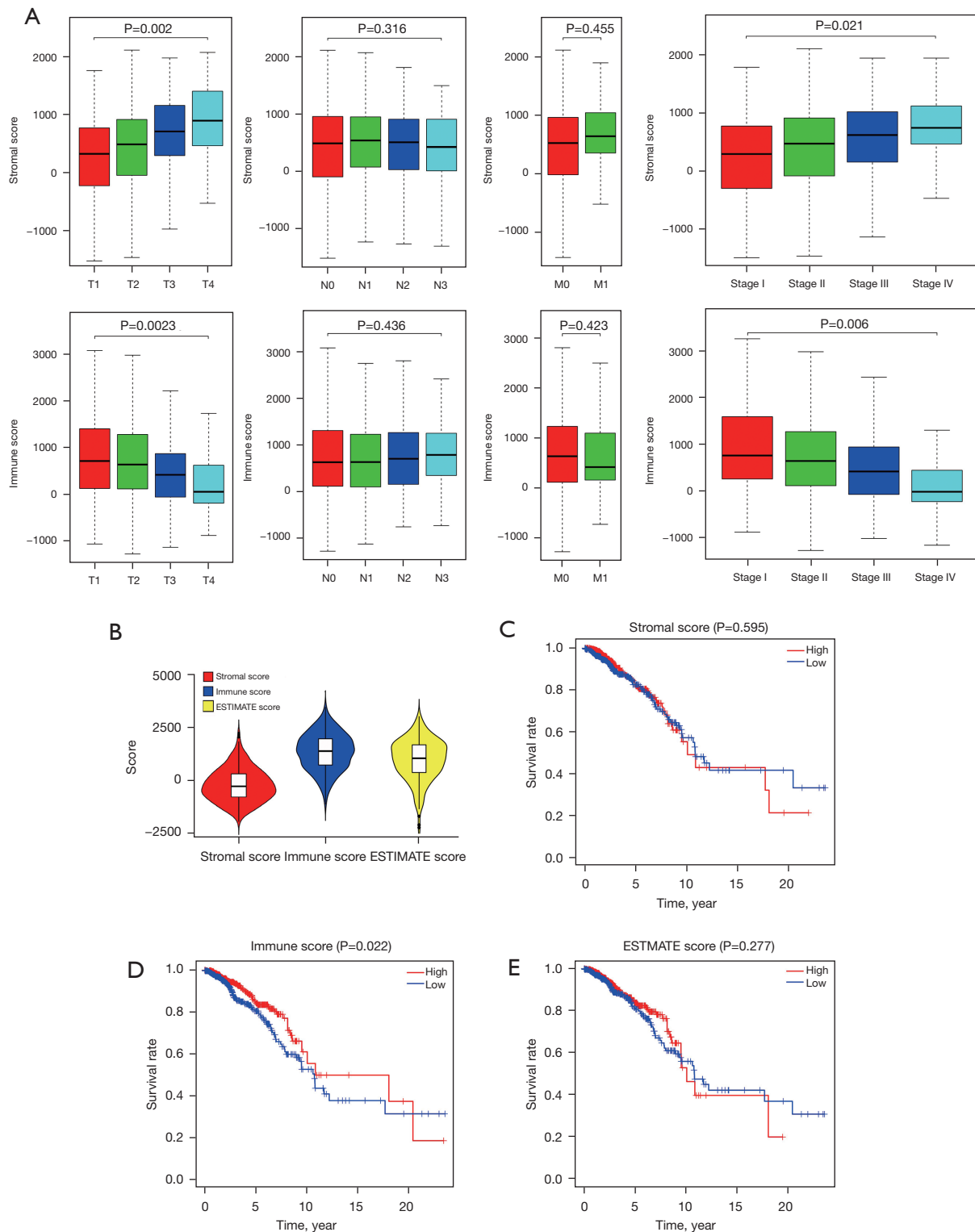| Characteristics	Combined cohort, n=1,069	Training cohort, n=535	Validation cohort, n=534
Age (year)	58.08±12.75	57.71±14.51	58.45±11.72
T stage			
T1	279 (26.10)	146 (27.29)	133 (24.91)
T2	617 (57.72)	305 (57.01)	312 (58.43)
T3	132 (12.35)	65 (12.15)	67 (12.55)
T4	38 (3.55)	18 (3.36)	20 (3.75)
Tx	3 (0.28)	1 (0.19)	2 (0.37)
N stage			
N0	502 (46.96)	257 (48.04)	245 (45.88)
N1	357 (33.40)	178 (33.27)	179 (33.52)
N2	120 (11.23)	51 (9.53)	69 (12.92)
N3	73 (6.83)	43 (8.34)	30 (5.62)
Nx	17 (1.59)	6 (1.12)	11 (2.06)
M stage			
M0	1,036 (96.91)	517 (96.64)	519 (97.19)
M1	22 (2.06)	14 (2.62)	8 (1.50)
Mx	11 (1.03)	4 (0.75)	7 (1.31)
TNM stage			
Stage I	183 (17.12)	93 (17.38)	90 (16.85)
Stage II	608 (56.88)	307 (57.38)	301 (56.37)
Stage III	240 (22.45)	113 (21.12)	127 (23.78)
Stage IV	22 (2.05)	14 (2.62)	8 (1.50)
Other	16 (1.50)	8 (1.50)	8 (1.50)

Next, score distributions were evaluated among the various tumor stages, including T, N, M, and TNM, to identify the relationship between stromal/immune scores and the pathologic features of patients with BC. The results revealed that the stromal/immune scores were significantly associated with the T and TNM stages. Elevated stromal scores were significantly correlated with higher T (P<0.05) and TNM stages, while elevated immune scores were associated with lower T (P<0.05) and TNM stages (all P<0.05). However, both scores (stromal and immune) were not significantly associated with the N or M stages (all P>0.05), as depicted in *Figure 1A*.

The patients were categorized into LR and HR groups according to the median score values (median stromal score =529.92; median immune score =634.34; median ESTIMATE score =1,206.08) to identify the prognostic relevance of the immune and stromal scores. According to the K-M analyses, the OS rate was longer in patients with high immune scores than those with low immune scores (P=0.022), as depicted in *Figure 1C-1E*.

Evaluation of DEGs according to the immune/stromal scores

We used the RNA-seq data from the identified BC patients in TCGA database to see if there was any correlation between the gene expression levels and the immune/stromal scores calculated earlier. To find the immune (or stromal)

Figure 1 The relationship between patient clinical characteristics and stromal/immune scores in the combined cohort. (A) Stromal and immune scores among BC patients with different stages (T, N, M, and TNM) of the disease. (B) The violin plot of the stromal, immune, and ESTIMATE scores. (C-E) Kaplan-Meier plots demonstrating OS outcomes between low- and high-score groups. BC, breast cancer; ESTIMATE, estimation of stromal and immune cells in malignant tumor tissues using expression data; OS, overall survival.

Page 6 of 13

Zhu et al. Gene signature based on TME in BC
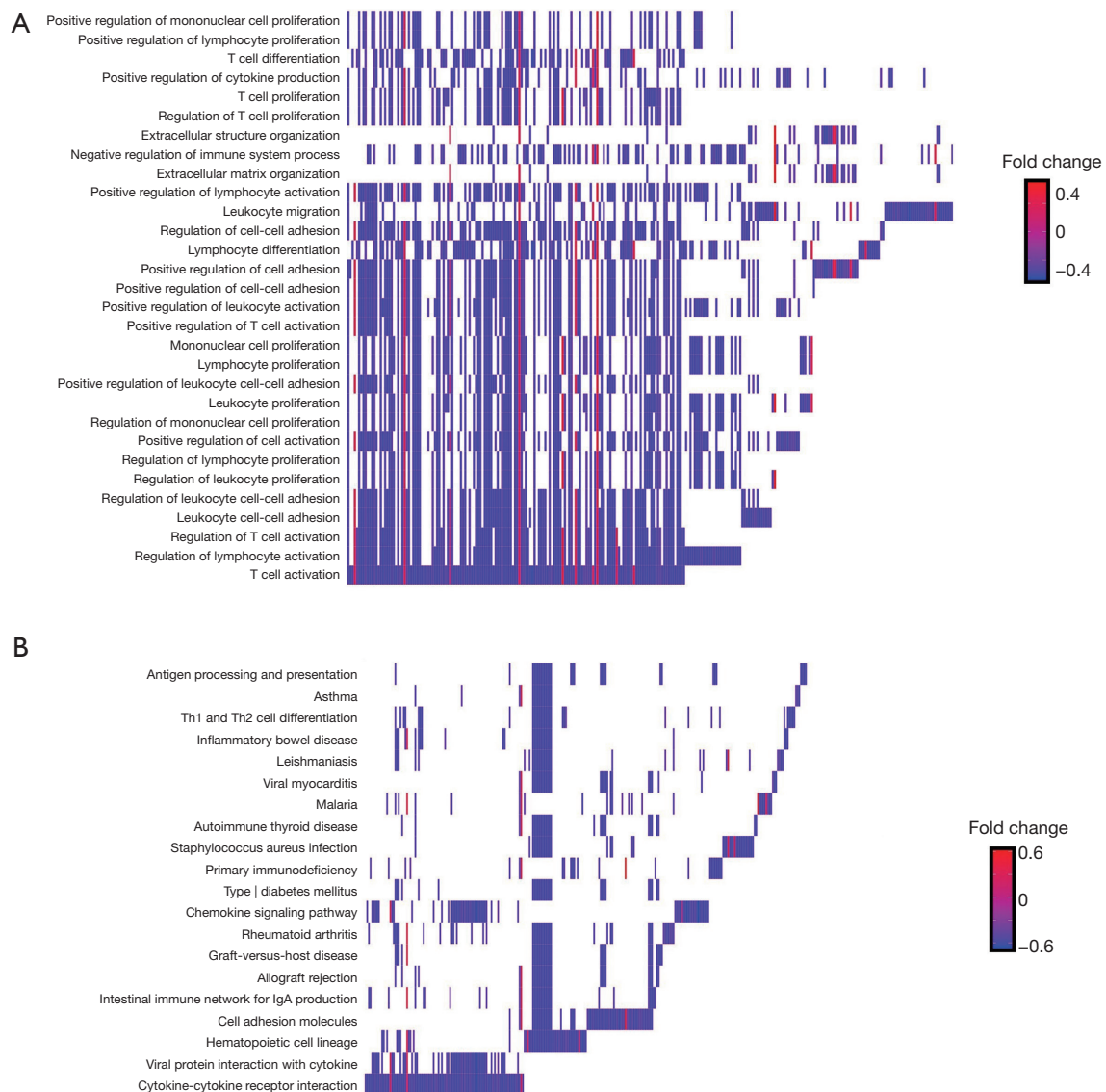
Figure 2 GO and KEGG pathway enrichment analysis results for DEGs. (A) and (B) show GO and KEGG pathway enrichment analysis of DEGs, respectively. The FDR approach was employed to alter the P values. GO, gene ontology; KEGG, kyoto gene and genome encyclopedia; DEGs, differentially expressed genes; FDR, false discovery rate.

score-related DEGs, researchers compared the expression patterns of BC patients with high immune (or stromal) scores to those with low scores. We found 764 stromal score-related DEGs (n=626; n=138) and 829 immune score-related DEGs (n=726; n=103) in the high- and low-score groups, respectively (|log2 fold change|>1 and FDR <0.05) (Tables S2,S3).

Functional enrichment analyses highlight the roles of the DEGs

GO and KEGG analyses were performed to evaluate the immune and stromal score-related DEGs that were enriched (*Figure 2*). With respect to GO, DEGs were primarily enriched in T cell activation, regulation of T cell activation, lymphocyte differentiation, and regulation of
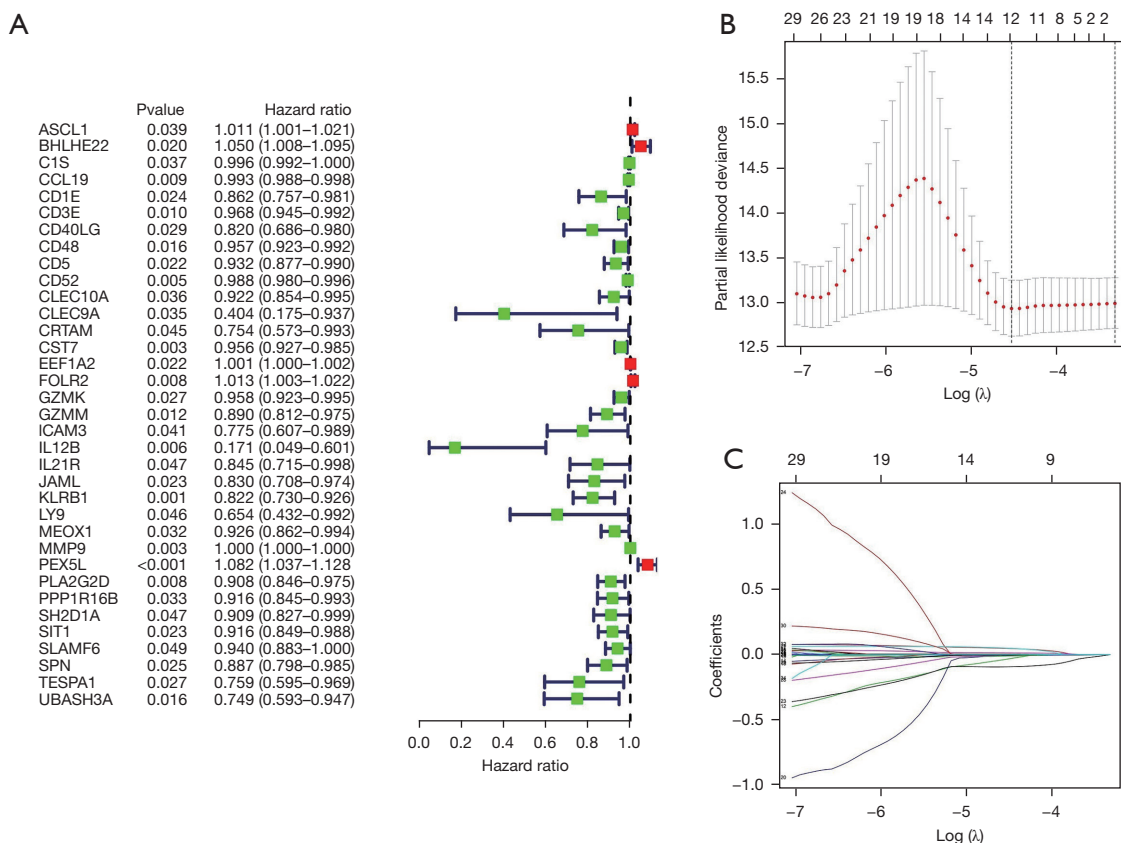
Figure 3 Generation and validation of a 12-gene signature risk score model. (A) Survival-associated DEGs. (B) A coefficient distribution map for a logarithmic (λ) sequence was generated as appropriate. (C) Selection of the optimal BC-related parameters in the LASSO model (λ). DEGs, differentially expressed genes; BC, breast cancer; LASSO, least absolute shrinkage and selection operator.

lymphocyte activation. We also conducted a KEGG analysis that revealed that these DEGs were primarily associated with cytokine-cytokine receptor interaction, chemokine signaling pathway, Th1 and Th2 cell differentiation, and primary immunodeficiency.

Construction of a prognostic signature based on DEGs

We first identified 35 prognostic DEGs as candidates for the development of a gene risk signature based on the BC patient survival data from the training cohort (*Figure 3A*). Ultimately, 12 genes were chosen for inclusion in this signature using a LASSO Cox regression technique (*Table 2, Figure 3B,3C*). The predictive significance of the signature-derived risk scores was next investigated for the training cohort by stratifying patients into LR and HR groups using the cohort's median risk score (0.92985) (*Figure 4A,4B*).

LR BC patients had a considerably longer OS than HR BC patients (16.4±2.07 *vs.* 11.16±1.23 years, P<0.001), which

was in line with our predictions. Importantly, we discovered that in the validation cohort, LR patients had a longer OS than HR patients (16.57±1.32 *vs.* 12.62±1.32 years; P=0.002) (*Figure 4C,4D*). Multivariate Cox regression models were used to evaluate the independent risk factors in the two cohorts. Multivariate analysis included several clinicopathological factors as well as the 12-gene signature scores, revealing that age (Training set: HR =1.028, 95 % CI: 1.008–1.049, P=0.007; Validation set: HR =1.046, 95% CI: 1.025–1.068, P<0.001) and the 12-gene signature score (Training set: HR =1.192, 95% CI: 1.111–1.28, P<0.001; Validation set: HR =1.512, 95% CI: 1.286–1.779, P<0.001) were independent prognostic indicators both in the training and validation cohorts, as depicted in *Figure 4E,4F*.

We used a ROC analysis to test the predictive value of our 12-gene signature risk model against that of other clinicopathologic parameters, such as age, T, N, M, and TNM stage. Compared to the other clinicopathologic feature curves in the training and validation cohorts, the 12-

Page 8 of 13

Zhu et al. Gene signature based on TME in BC

Table 2 12-gene signature

Gene symbol	Gene ID	Description	Coefficient
ASCL1	429	Achaete-Scute family bHLH transcription factor 1	0.00349
BHLHE22	27319	Basic Helix-Loop-Helix Family Member E22	0.04779
C1S	716	Complement component C1s activity	−0.00067
CLEC9A	283420	C-Type Lectin Domain Containing 9A	−0.03476
CST7	8530	Cystatin F	−0.00614
EEF1A2	1917	Eukaryotic Translation Elongation Factor 1 Alpha 2	0.00033
FOLR2	2350	Folate Receptor Beta	0.01399
KLRB1	3820	Killer Cell Lectin Like Receptor B1	−0.09338
MEOX1	4222	Mesenchyme Homeobox 1	−0.01082
PEX5L	51555	Peroxisomal Biogenesis Factor 5 Like	0.05008
PLA2G2D	26279	Phospholipase A2 Group IID	−0.00362
PPP1R16B	26051	Protein Phosphatase 1 Regulatory Subunit 16B	−0.004

gene signature curve had the highest AUC value (Training set: AUC =0.806; Validation set: AUC =0.776). As a result, this 12-gene signature risk score could be more accurate than other clinical factors in identifying BC patients (*Figure 4G,4H*).

### HR and LR BC patients exhibit differences in immune status and infiltration

A PCA analysis was used to determine the variations in distribution patterns between the LR and HR groups using the 12-gene signature risk scores. The LR and HR groups were distributed into two separate clusters, as depicted in *Figure 5A*. Additionally, a GSEA functional annotation technique was used, revealing that LR samples were highly enriched for immune response pathway-related genes compared with HR samples (*Figure 5B*). When tumor samples from LR and HR BC patients were classified according to their 12-gene signature risk scores, variations were observed in the immunological status of these tumor samples. Additionally, samples from the LR group were enriched for immunological cascades, including the cytokine-cytokine receptor interaction, Notch signaling, primary immunodeficiency, JAK-STAT, Toll-like receptor, and T cell receptor signaling. On the other hand, the enriched cascades were primarily connected with the TGF-signaling cascades in the HR group, as depicted in *Figure 5C*. An immune cell-based study was performed, which

suggested an elevation in the numbers of M1 macrophages, plasma cells, naive/memory B cells, activated CD4[+] memory T cells, CD8[+] T cells, resting NK cells, and follicular helper T cells in the LR group patients. In comparison, HR patients showed elevated levels of M2 macrophages, M0 macrophages, and regulatory T cells (Tregs), as shown in *Figure 5D,5E*.

Discussion

BC is a form of rapidly progressing cancer with a poor prognosis that can be profoundly impacted by the TME. According to previous research, both immunological and stromal cells are key components of the TME and can have a significant impact on tumor growth and proliferation, as well as therapeutic responsiveness (17). A growing body of evidence suggests that the TME promotes BC growth and development, as well as influencing tumor invasion and metastasis. While alterations in immune cells, soluble molecules, and the ECM have all been shown to promote cancer growth, the link between TME-related genes and BC prognosis remains unknown. Therefore, extensive research is needed to assess the genome profiling correlated with existing tumor sequencing datasets to better understand the interaction between BC cells and the TME. In present study, we, as the first, constructed a 12-gene signature (*ASCL1, BHLHE22, C1S, CLEC9A, CST7, EEF1A2, FOLR2, KLRB1, MEOX1, PEX5L, PLA2G2D,* and
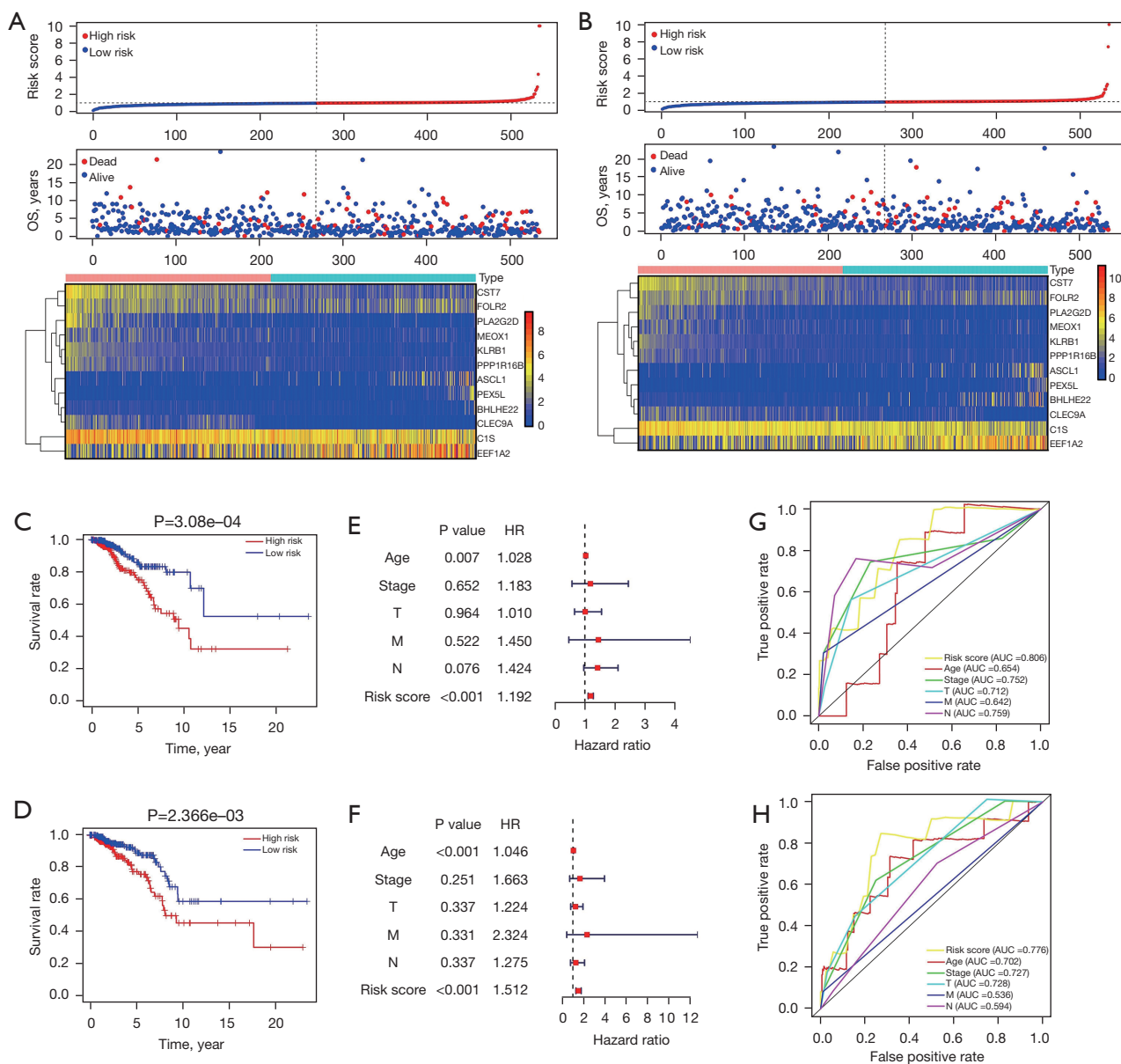
Figure 4 The risk score, survival time, and 12-gene expression (heatmap) distributions in the training (A) and validation cohorts (B). In the heatmap, rows represent genes, while columns correspond to individual patients. Survival analysis of LR and HR BC patients as shown by the Kaplan-Meier analysis of BC patients' OS in the HR and LR subgroups of the training (C) and validation cohorts (D). Multivariate independent prognostic analysis of 12-gene signature risk score and other clinical features in the training (E) and validation cohorts (F). Time-dependent ROC analysis comparing the 12-gene signature risk model and other clinicopathologic features as tools for predicting the OS of BC patients in the training (G) and validation cohorts (H). HR, high risk; LR, low risk; BC, breast cancer; OS, overall survival; ROC, receiver operating characteristic curve, AUC, area under the cure.

PPP1R16B) based on both stromal and immunological cell infiltration in BC. According to their survival, BC patients were separated into LR and HR groups using the identified 12-gene signature. The immunological status and immune cell infiltration were observed differently in the LR and HR groups.

We assessed the prognostic value of TME-associated genes using TCGA database. We discovered that BC patients
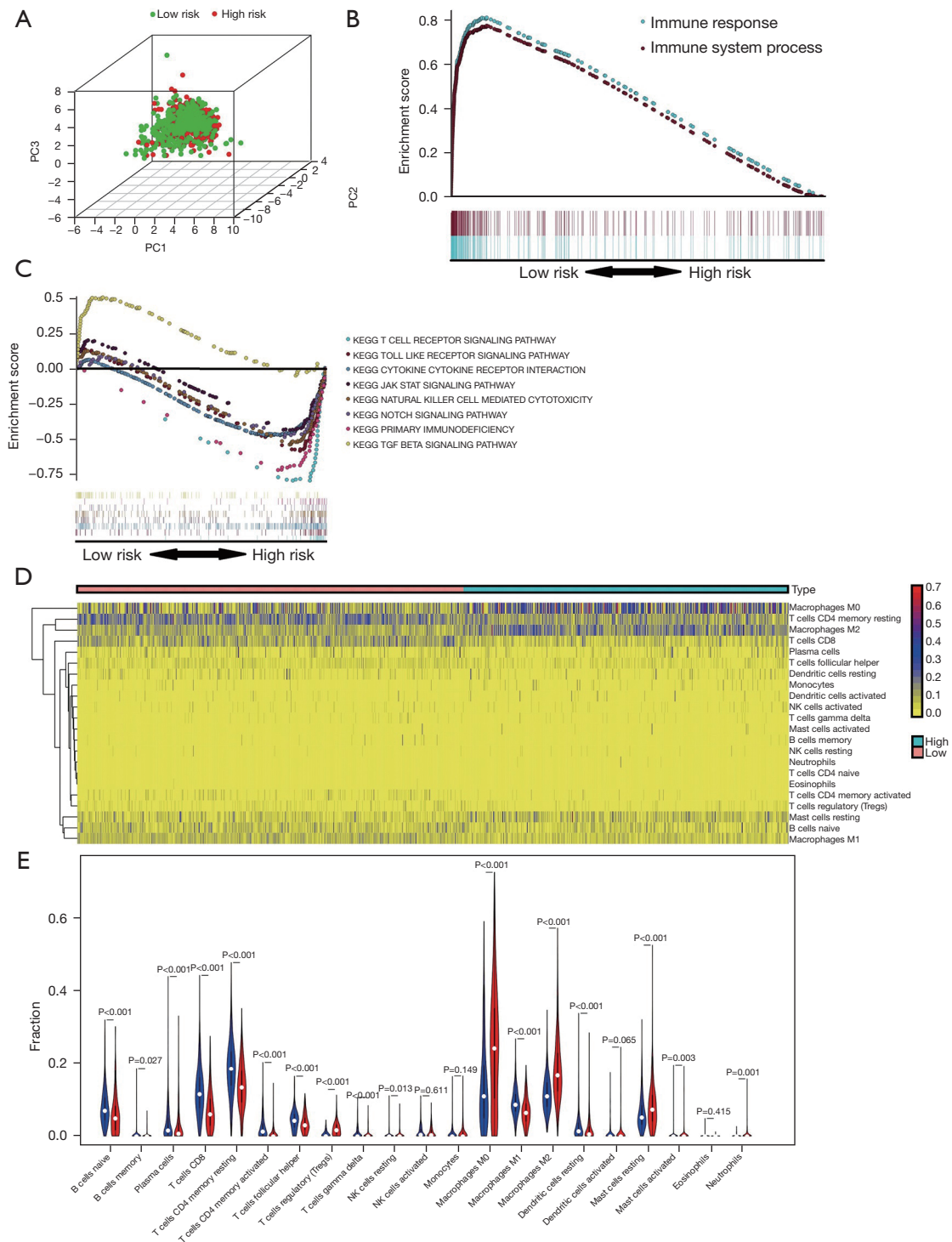
Figure 5 HR and LR BC patients demonstrate differences in immune status. (A) Principal components analysis between LR and HR groups based on the 12-gene signature risk score model. (B) Functional annotation of the 12-gene signature in the Immune response set and the Immune system process set in a GSEA analysis. (C) KEGG enrichment analysis. (D) Estimated immune cell populations as determined using the CIBERSORT algorithm. (E) Violin plots demonstrating the immune cell populations in the LR (blue) and HR (red) groups. P values have been identified based on the Wilcoxon Test. HR, high risk; LR, low risk; BC, breast cancer; GSEA, gene set enrichment analysis; KEGG, kyoto gene and genome encyclopedia; CIBERSORT, cell-type identification by estimating relative subsets of RNA transcripts.

with elevated immune scores had a considerably longer OS rate than those with lower immune scores. DEGs were found to be linked to immune and stromal scores as well as survival. The above genes were used to create a predictive risk profile that assessed differences in OS and immunological cell infiltration between LR and HR BC patients. Our work provides a comprehensive and accurate tool for assessing the TME and survival outcomes of BC patients. Our results thus hold great promise for identifying innovative molecular targets for immunotherapy and, hence, the improvement of treatment strategies available to BC patients.

TME-related genes have a significant impact on clinical outcomes in a variety of solid tumors. In this study, we used a gene signature risk model to incorporate a subset of survival-associated DEGs. Patients were divided into LR and HR groups using risk scores obtained from this model, with additional analyses revealing considerable variations in the OS rates of the patients in these two risk cohorts that corresponded to an elevated AUC value. Among the 12 genes included in this prognostic risk signature, some have previously been shown to be correlated with BC tumorigenesis. For example, eukaryotic translation elongation factor 1-alpha 2 (EEF1A2) has been reported as a putative oncogene owing to its pronounced expression in BC (18). One prior report indicated that EEF1A2 induced Akt-dependent actin remodeling and enhanced the invasion of BC cells (19). Furthermore, Sun et al. confirmed that mesenchyme homeobox 1 (MEOX1) is a critical molecular target involved in regulating breast cancer stem cells and mesenchymal-like cell proliferation. MEOX1 has been linked with poor survival, lymph node metastasis, and tumor stage in BC patients (20). And others have been reported to be associated with other tumors or diseases. ASCL1 is a transcription factor and required for neural differentiation, which is also essential for the development of pulmonary neuroendocrine cells (21,22). BHLHE22 was found to be involved in the development of psychiatric disease (23). C1S is part of the complement system family, while CLEC9A is a surface marker on tumor-associated macrophage. FOLR2 is expressed in macrophages, placental and hematopoietic cells (24). Expressing KLRB1 (encoding CD161, a surface marker on NK cells and several T cell subsets) was reported to be associated with favorable outcomes in pan-cancer (25). PLA2G2D, a metabolism- and immune-related molecule, was identified to be a biomarker for immune cell infiltration and patient survival in cervical squamous cell carcinoma (26). While upregulated expression of PLA2G2D was reported to be associated with adaptive resistance to immune

checkpoint blockade (pembrolizumab) (27).

In addition, the immunological phenotypes identified in the HR and LR groups were shown to be considerably varied. According to the GSEA, LR patient samples had gene expression patterns that were enriched for genes related to the immune system. While TME composition is of great interest to researchers, immune cells in the BC TME vary greatly and are linked to patient outcomes. Tumor-infiltrating lymphocytes (TILs), especially CD8+ T cells, are important mediators of tumor immune surveillance. Normally, antigen and co-stimulatory molecule exposure activate the effector CD8+ T cells, allowing them to lyse tumor cells and inhibit tumor growth. In patients with BC, T cell infiltration plays an important role. CD8+ T cell monitoring is useful in tracking the course of cancer and the prognosis of patients. Our results are consistent with these prior findings, as LR patients exhibited an increased CD8+ T cell profile. Although the involvement of Tregs in BC prognosis has often been the subject of research, controversy remains regarding the specific prognostic impact of these cells and whether they are beneficial or harmful. According to Shang et al. (28), Treg infiltration was reportedly linked to a favorable prognosis in estrogen receptor (ER)- but not ER+ BC patients. We observed significant Treg enrichment in HR patients. The role of B cells in BC remains poorly understood. Our results demonstrated enrichment of B cells in LR patients, indicating a negative correlation between these cells and risk. However, the precise mechanistic role of B cells remains to be defined in this prognostic context.

Tumor-associated macrophages (TAMs) also play a regulatory role in tumor cell-immune system interactions (29,30). TAMs have traditionally been divided into two groups: M1 and M2. Because of their anti-inflammatory, angiogenic, and ECM-remodeling actions, M2 macrophage infiltration within tumors has been associated with a poorer prognosis. In samples from HR BC patients, M2 macrophages were shown to be abundant.

However, we only have a rudimentary understanding of the balance between M1 and M2 phenotypes. The JAK/STAT signaling cascades are among the most significant polarization regulators. According to the GSEA data, signaling cascades, such as JAK/STAT signaling were shown to be elevated in the LR group. IFN-γ triggers STAT1 and produces M1 macrophage polarization, with the IFN-γ/JAK/STAT1 signaling cascade serving as a critical regulator of the M1 phenotype. According to previous research, the activation of the IL-6/JAK/STAT3 signaling cascade is

Page 12 of 13

Zhu et al. Gene signature based on TME in BC

thought to be the primary mediator of macrophage M2 polarization.

The current study has some shortcomings. For one, this was an in silico analysis and as such, all conclusions were correlative. Additional and *in vivo* functional analyses will thus be needed to validate and expand upon these results. In addition, while we made efforts to verify the microenvironment-associated DEGs and subsequently assess their prognostic significance in BC patients, further large-scale prospective studies will be necessary to validate this hypothesis. It is also worth noting that the immune signature discovered in this study is based on an assessment of immunological cell infiltration within tumors derived from algorithmic analyses of RNA-sequenced data. To establish the accuracy of the immune infiltration data and examine the cell-to-cell interactions that may emerge within these tumors and affect BC development or prognosis, more in-depth studies will be required in the future.

Conclusions

To assess the BC TME, we performed an extensive bioinformatics analysis for gene expression in BC patients based on TCGA. Using this method, we were able to identify microenvironment-associated DEGs and subsequently assess their prognostic significance in BC patients. To fully validate their association with patient OS, future clinical studies of the functional roles of the identified genes will be required. Taken together, our findings shed light on the complicated interplay between BC stromal cells and immunological cells in the TME, possibly pointing to novel therapeutic possibilities for future clinical trials.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://atm.amegroups.com/article/view/10.21037/atm-21-6748/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://atm.amegroups.com/article/view/10.21037/atm-21-6748/coif).

The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

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;71:209-49.
2. DeSantis CE, Ma J, Goding Sauer A, et al. Breast cancer statistics, 2017, racial disparity in mortality by state. CA Cancer J Clin 2017;67:439-48.
3. Yang L, Wang S, Zhang Q, et al. Clinical significance of the immune microenvironment in ovarian cancer patients. Mol Omics 2018;14:341-51.
4. Ayala F, Dewar R, Kieran M, et al. Contribution of bone microenvironment to leukemogenesis and leukemia progression. Leukemia 2009;23:2233-41.
5. Mehraj U, Ganai RA, Macha MA, et al. The tumor microenvironment as driver of stemness and therapeutic resistance in breast cancer: New challenges and therapeutic opportunities. Cell Oncol (Dordr) 2021;44:1209-29.
6. Liu CC, Steen CB, Newman AM. Computational approaches for characterizing the tumor immune microenvironment. Immunology 2019;158:70-84.
7. Mao X, Xu J, Wang W, et al. Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: new findings and future perspectives. Mol Cancer 2021;20:131.
8. Romaniuk A, Lyndin M. Immune microenvironment as a factor of breast cancer progression. Diagn Pathol 2015;10:79.

9.   Reisfeld RA. The tumor microenvironment: a target for combination therapy of breast cancer. Crit Rev Oncog 2013;18:115-33.

10. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 2013;4:2612.

11. Alonso MH, Aussó S, Lopez-Doriga A, et al. Comprehensive analysis of copy number aberrations in microsatellite stable colon cancer in view of stromal component. Br J Cancer 2017;117:421-31.

12. Shah N, Wang P, Wongvipat J, et al. Regulation of the glucocorticoid receptor via a BET-dependent enhancer drives antiandrogen resistance in prostate cancer. Elife 2017;6:27861.

13. Jia D, Li S, Li D, et al. Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. Aging (Albany NY) 2018;10:592-605.

14. Ding S, Sun X, Zhu L, et al. Identification of a novel immune-related prognostic signature associated with tumor microenvironment for breast cancer. Int Immunopharmacol 2021;100:108122.

15. Zhang Y, Di X, Chen G, et al. An immune-related signature that to improve prognosis prediction of breast cancer. Am J Cancer Res 2021;11:1267-85.

16. Xu M, Li Y, Li W, et al. Immune and Stroma Related Genes in Breast Cancer: A Comprehensive Analysis of Tumor Microenvironment Based on the Cancer Genome Atlas (TCGA) Database. Front Med (Lausanne) 2020;7:64.

17. Sturm G, Finotello F, Petitprez F, et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. Bioinformatics 2019;35:i436-45.

18. Joseph P, O'Kernick CM, Othumpangat S, et al. Expression profile of eukaryotic translation factors in human cancer tissues and cell lines. Mol Carcinog 2004;40:171-9.

19. Amiri A, Noei F, Jeganathan S, et al. eEF1A2 activates Akt and stimulates Akt-dependent actin remodeling, invasion and migration. Oncogene 2007;26:3027-40.

20. Sun L, Burnett J, Gasparyan M, et al. Novel cancer stem cell targets during epithelial to mesenchymal transition in PTEN-deficient trastuzumab-resistant breast cancer. Oncotarget 2016;7:51408-22.

21. Borromeo MD, Savage TK, Kollipara RK, et al. ASCL1 and NEUROD1 Reveal Heterogeneity in Pulmonary Neuroendocrine Tumors and Regulate Distinct Genetic Programs. Cell Rep 2016;16:1259-72.

22. Ito T, Udaka N, Yazawa T, et al. Basic helix-loop-helix transcription factors regulate the neuroendocrine differentiation of fetal mouse pulmonary epithelium. Development 2000;127:3913-21.

23. Subaran RL, Odgerel Z, Swaminathan R, et al. Novel variants in ZNF34 and other brain-expressed transcription factors are shared among early-onset MDD relatives. Am J Med Genet B Neuropsychiatr Genet 2016;171B:333-41.

24. Sabharanjak S, Mayor S. Folate receptor endocytosis and trafficking. Adv Drug Deliv Rev 2004;56:1099-109.

25. Gentles AJ, Newman AM, Liu CL, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med 2015;21:938-45.

26. Liu H, Xu R, Gao C, et al. Metabolic Molecule PLA2G2D Is a Potential Prognostic Biomarker Correlating With Immune Cell Infiltration and the Expression of Immune Checkpoint Genes in Cervical Squamous Cell Carcinoma. Front Oncol 2021;11:755668.

27. Cindy Yang SY, Lien SC, Wang BX, et al. Pan-cancer analysis of longitudinal metastatic tumors reveals genomic alterations and immune landscape dynamics associated with pembrolizumab sensitivity. Nat Commun 2021;12:5137.

28. Shang B, Liu Y, Jiang SJ, et al. Prognostic value of tumor-infiltrating FoxP3+ regulatory T cells in cancers: a systematic review and meta-analysis. Sci Rep 2015;5:15179.

29. Mahmoud SM, Lee AH, Paish EC, et al. The prognostic significance of B lymphocytes in invasive carcinoma of the breast. Breast Cancer Res Treat 2012;132:545-53.

30. Mohammed ZM, Going JJ, Edwards J, et al. The role of the tumour inflammatory cell infiltrate in predicting recurrence and survival in patients with primary operable breast cancer. Cancer Treat Rev 2012;38:943-55.

(English Language Editor: D. Fitzgerald)