

Construction and validation of an immunity-related prognostic signature for breast cancer

Tao Zhu^{1,2}, Juyan Zheng^{1,2}, Shuo Hu³, Wei Zhang^{1,2}, Honghao Zhou^{1,2}, Xi Li^{1,2}, Zhao-Qian Liu^{1,2}

¹Department of Clinical Pharmacology, Hunan Key Laboratory of Pharmacogenetics, and National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha 410008, P.R. China

²Institute of Clinical Pharmacology, Engineering Research Center for Applied Technology of Pharmacogenomics of Ministry of Education, Central South University, Changsha 410078, P.R. China

³Department of Nuclear Medicine, Key Laboratory of Biological Nanotechnology of National Health Commission, Xiangya Hospital, Central South University, Changsha 410008, P.R. China

Correspondence to: Zhao-Qian Liu, Xi Li; **email:** zqliu@csu.edu.cn, bayern@csu.edu.cn

Keywords: tumor immunology, breast cancer, prognostic signature, immunity-related gene, transcription factor

Received: June 9, 2020

Accepted: August 8, 2020

Published: November 7, 2020

Copyright: © 2020 Zhu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Breast cancer is one of the most lethal malignancies among women, and understanding the effects of host immunity on disease progression offers the potential to improve immunotherapies against it. Here, we constructed an immunity-related gene (IRG)-based prognostic signature to stratify breast cancer patients and predict their survival. We identified differentially-expressed genes by analyzing the breast cancer transcriptome data from The Cancer Genome Atlas. Univariate Cox regression revealed 179 survival-correlated IRGs, 12 of which we used to construct an immunity-based prognostic signature that stratified breast cancer patients into high- and low-risk groups. The signature was an independent predictor for survival and was validated in an independent dataset. We also investigated the correlations between our prognostic signature and immune infiltrates and found that signature-derived risk scores correlated negatively with infiltration of B cells, CD4+ T cells, CD8+ T cells, neutrophils and dendritic cells. Our results show that the proposed prognostic signature reflects the tumor immune microenvironment, which makes it a potential indicator for survival that warrants further research to assess its clinical utility.

INTRODUCTION

Breast cancer is the most frequently diagnosed malignancy and one of the leading causes of cancer-related mortality among women. It is estimated that in 2020 breast cancer will account for ~30% of all new female cancer diagnoses, and for ~15% of all female cancer deaths [1]. Patient survival is still threatened by resistance to treatment, relapse and metastasis, despite advances in multidimensional treatment including surgery, chemotherapy and targeted therapy. Breast cancer is a heterogeneous disease that can be differently subtyped based on disparate classifications [2–4]. Identification of specific biomarkers of breast cancer

can help to predict and monitor disease progression, and to reduce aggressive cases through early intervention.

The interplay between tumors and host immunity plays a critical role in breast cancer biology [5]. Immunoediting has emerged as a relevant hallmark of cancer, which enables tumor cells to escape from immune attack by concealing their surface antigens or switching off functions of immune cell effectors [6, 7]. Cancer immunoediting can be described by three phases termed elimination, equilibrium, and escape [8]. In the elimination phase, tumor cells are recognized and killed by innate immune cells. If tumor cells evolve and reduce their immunogenicity, resistance will be

acquired as a result, which means they can survive host immune responses and equilibrium is reached. During the escape phase, nascent cancer cells can't be eradicated by immune cells, giving rise to rapid progression [8]. Hence, breast cancer progression is subject to the activity of the immune system, which varies among individuals. A recent study showed that based on immune activity, basal-like triple-negative breast cancer (TNBC) can be subclassified into immunosuppressed and immune-activated subtypes with distinct prognoses [3], suggesting a potential immunological gene-based signature in breast cancer that may be exploited for prognostic benefit.

Recent studies have improved our understanding of the mechanisms underlying carcinogenesis and have propelled the development of anti-cancer drugs, such as cancer vaccines [9–11]. Genome-wide profiling has helped to gain insights into disease progression and is an efficient way to deepen our understanding of cancer biology [12–16]. In this study, our aim was to identify high-risk breast cancer patients based on a prognostic signature of immunity-related genes (IRGs), which we constructed by integrating transcriptomic data with clinical information to screen for survival-associated IRGs. We also investigated potential regulatory mechanisms by analyzing associated transcription factors and validated the reliability of the IRGs-based prognostic signature in an independent cohort. Lastly, to investigate if the signature could reflect the immune microenvironment of breast cancer, we analyzed the correlations between risk scores derived from the IRGs-based signature and immune cell infiltration.

RESULTS

Identification and functional annotation of differentially-expressed IRGs

We retrospectively analyzed the TCGA breast cancer dataset, which included 1104 tumor samples and 113 normal adjacent samples, to identify differentially-expressed genes. A total of 2459 genes were identified as differentially expressed, with 1231 being upregulated and 1228 downregulated (Figure 1A, 1B). Since host immunity contributes to determining neoplasm destiny, here we focused on immunity-related genes. From these differentially-expressed genes, we extracted IRGs and obtained 69 upregulated and 110 downregulated IRGs (Figure 1C, 1D), which are listed in Supplementary Table 1. To gain insight into the biological roles of the 179 differentially-expressed IRGs, we performed GO and KEGG pathway enrichment analysis, which identified 64 relevant GO terms and five relevant KEGG pathways (Supplementary Table 2). As expected, GO analysis indicated that major terms

enriched in the biological process category were signal transduction, inflammatory response and immune response. The most enriched cellular component terms were “extracellular region” and “extracellular space”. For the molecular function category, the primary enriched terms were growth factor activity, cytokine activity and chemokine activity (Figure 2A). KEGG pathway analysis revealed that the differentially-expressed IRGs were involved in cytokine-cytokine receptor interactions, chemokine signaling, and neuroactive ligand-receptor interactions (Figure 2B).

Screening of survival-associated IRGs

We next asked whether and which of the aforementioned 179 differentially-expressed IRGs could aid in disease management in breast cancer patients. Univariate Cox analyses revealed that 19 IRGs were associated with overall survival. Most of the 19 survival-associated IRGs were protective factors, as indicated by their hazard ratios (Figure 3).

Transcription factors involved in regulation of survival-associated IRGs

Dynamic alterations in the expression level of transcription factors can affect immune cell fates and tumor development and progression [17–19]. To investigate potential regulatory mechanisms of the survival-associated IRGs, we extracted candidate transcription factors from all differentially-expressed genes to analyze the associations in gene expression between transcription factors and survival-associated IRGs. Of the 36 differentially-expressed transcription factors we identified (Figure 4A), E2F1 and MYBL2 were negatively associated with LIFR expression, while CBX2, ELF5, FOXM1, MYBL2, MYH11 and TP63 were in positive associations with at least one of the survival-associated IRGs. Figure 4B shows a modulatory network illustrating the regulation between the above seven transcription factors and survival-associated IRGs.

Construction of the immunity-related prognostic model

The 19 survival-associated IRGs were subject to further analysis in order to construct an immunity gene-based prognostic signature. According to results of multivariate Cox regression analysis, 12 hub IRGs were eligible and we used them to construct a prognostic signature, which we in turn used to derive a risk score for each patient. This risk score was the sum of products of the expression value and the regression coefficient of each gene, per the following formula: risk score = (0.2699 * expression level of ULBP2) + (0.1645 *

expression level of FABP6) + [(-0.1348) * expression level of FABP7] + (0.1831 * expression level of LALBA) + (0.2911 * expression level of IGHE) + [(-0.3570) * expression level of NRG1] + (0.1907 * expression level of SCG2) + [(-0.8889) * expression level of TSLP] + [(-0.2905) * expression level of ADRB1] + (0.1408 * expression level of NPR3) + [(-0.1646) * expression level of TNFRSF18] + [(-0.6666) * expression level of TNFRSF8]. As expected, we could see in the formula that risky genes with a hazard ratio > 1 had a positive coefficient, while genes with a hazard ratio < 1 had a negative coefficient (Figure 3). With the median risk score as the cutoff value, all breast cancer patients were assigned to either a high-risk or a low-risk group (Figure 5A). Figure 5B, 5C show an overview of the overall survival and hub IRG expression of these

groups. Patients in the high risk group had poorer overall survival compared with that of the low risk group (Figure 5D). Receiver operating characteristic (ROC) curve analysis suggested good sensitivity and specificity of the prognostic model (Figure 5E). We next validated the 12 hub IRGs-based prognostic model in an independent breast cancer cohort, for which we calculated patient risk scores to group the patients into low- and high-risk groups using the median risk score as a cutoff. In line with the survival analysis result of the TCGA BRCA cohort, patients with a higher risk score showed poorer overall survival than those with a lower risk score (Figure 5F). These data collectively revealed that the 12 hub IRGs-based prognostic model might be useful in evaluating the risk degree of a breast cancer patient while predicting overall survival time.

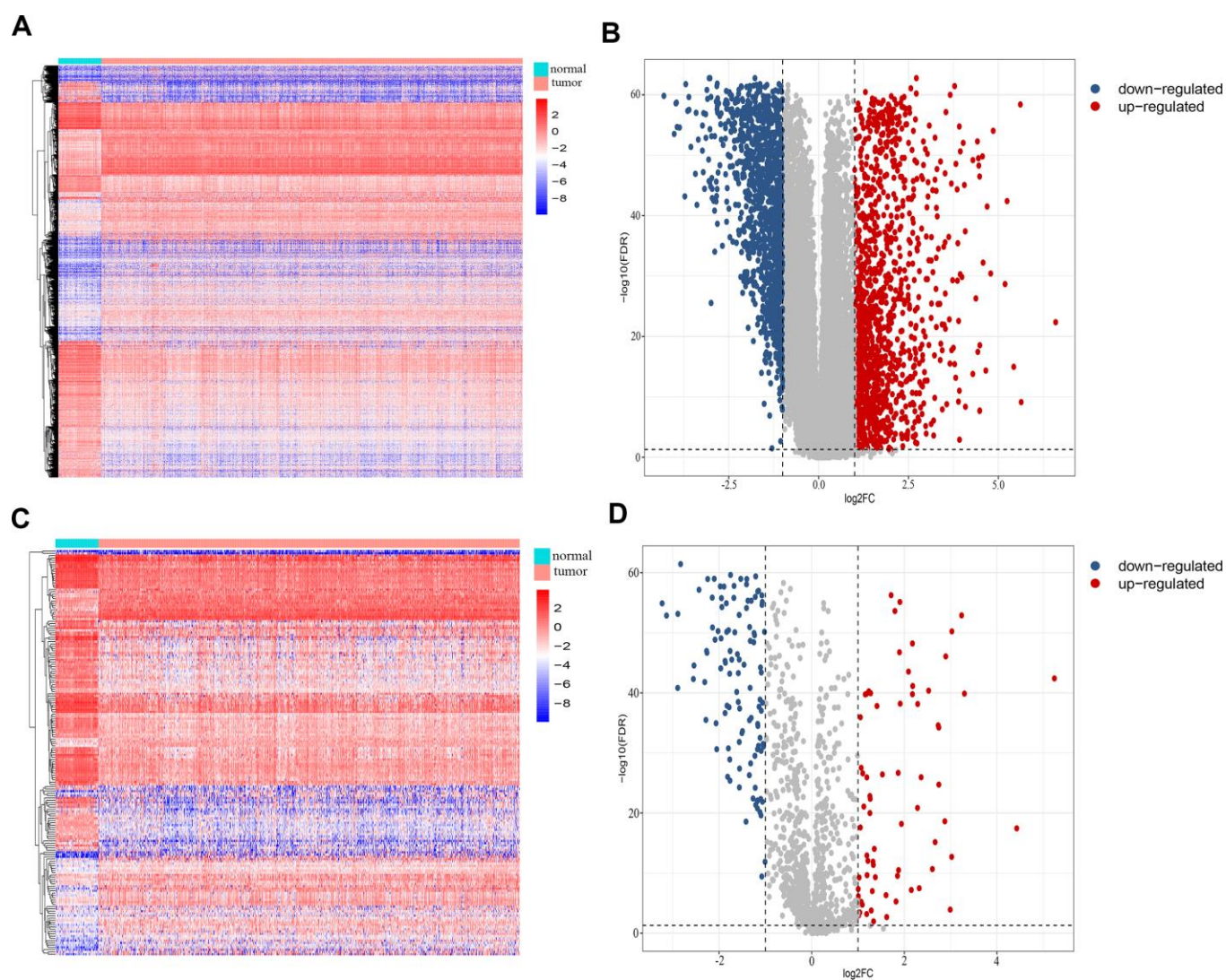


Figure 1. Identification of differentially-expressed immunity-related genes in breast cancer. (A) Heatmap and (B) volcano plot showing differentially-expressed genes between breast and non-malignant tissues. (C) Heatmap and (D) volcano plot of differentially-expressed immunity-related genes in breast cancer.

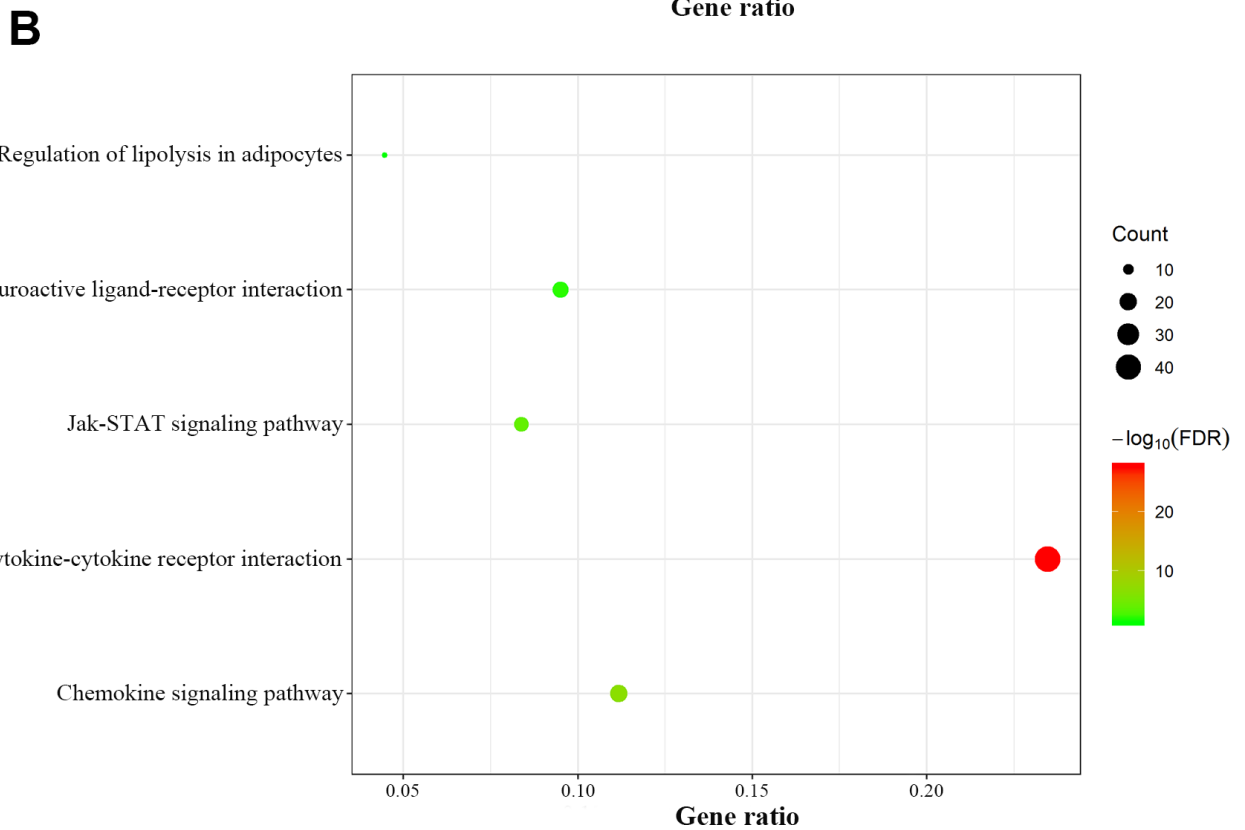
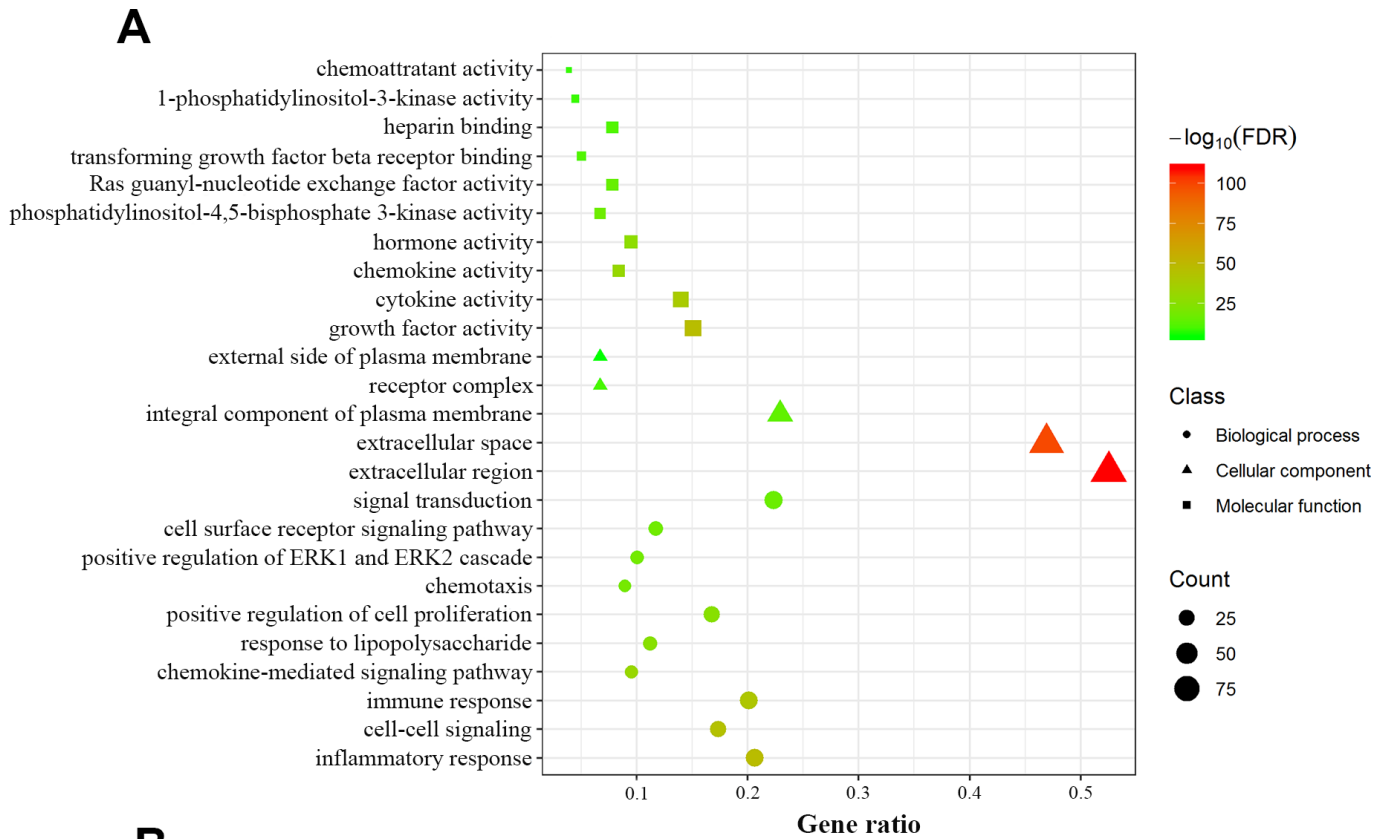


Figure 2. Functional annotation of differentially expressed immunity-related genes. (A) Enriched Gene Ontology terms including biological process (dots), cellular component (triangles) and molecular function (squares). **(B)** Enriched KEGG pathways.

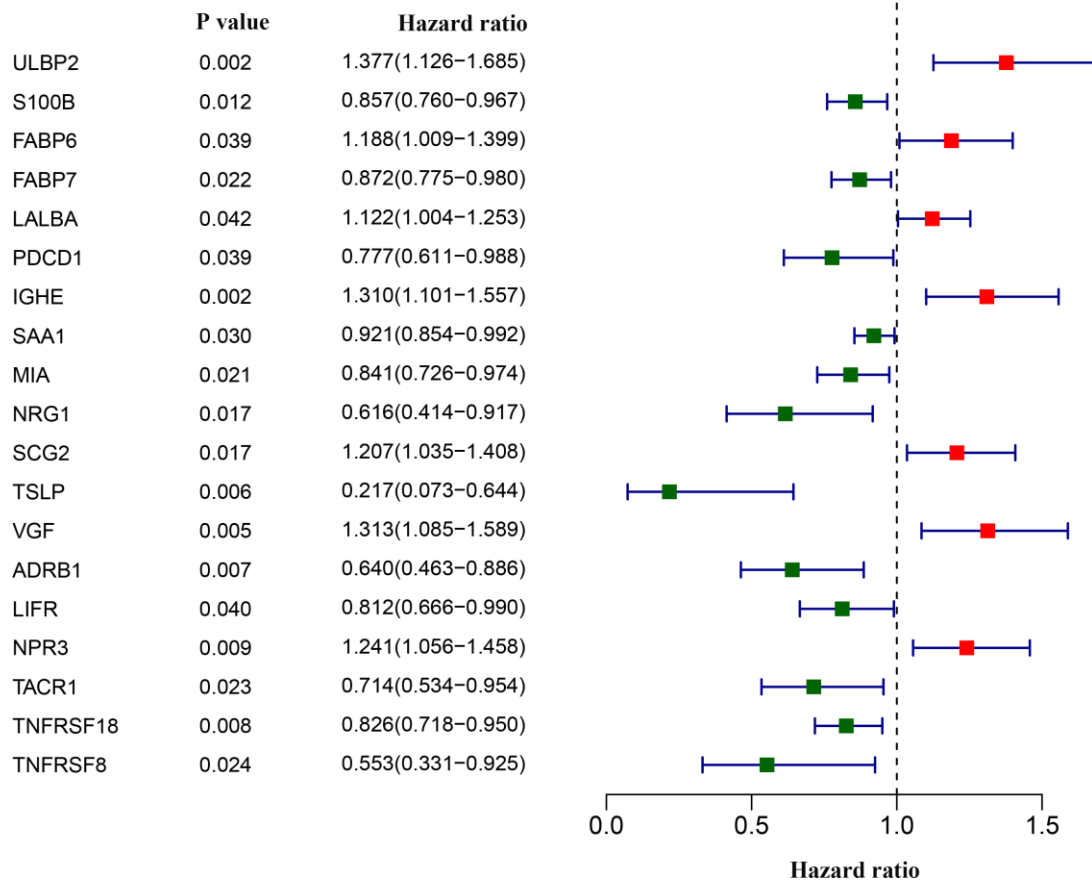


Figure 3. Screening of differentially expressed immunity-related genes correlated with overall survival in breast cancer. The forest plot shows hazard ratios of each gene.

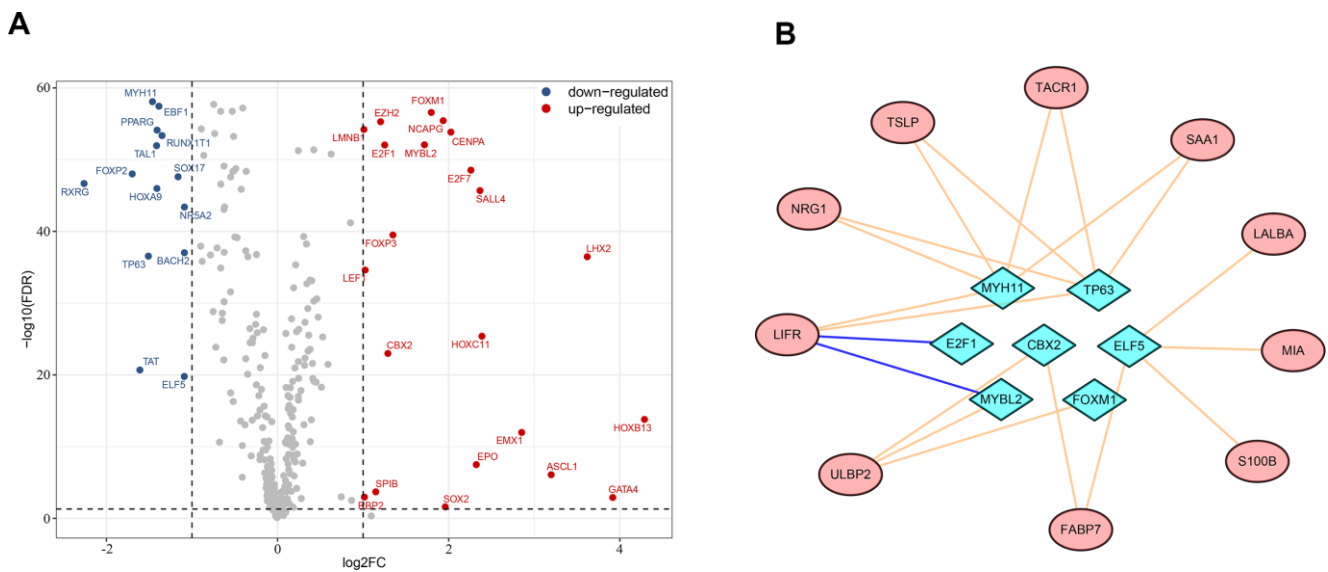


Figure 4. Regulatory network of survival-associated immunity-related genes based on differentially-expressed transcription factors. (A) Highlight of up- and down-regulated transcription factors in breast cancer. (B) The regulatory network between transcription factors and prognostic immunity-related genes. Survival-associated immunity-related genes are shown in ellipses and differentially-expressed transcription factors in diamonds. A pink line represents positive regulation and a blue line represents negative regulation.

IRGs as an independent risk factor for breast cancer patients

We next investigated whether the power of the immunity-related prognostic model was influenced by confounding factors. In addition to the risk score derived from the prognostic model, relevant clinical characteristics including age, tumor stage and TNM classification were used for univariate analysis. As expected, all of these factors were associated with a poor prognosis, with a metastatic disease being the most hazardous factor (Figure 6A). According to multivariate regression, age (HR = 1.036, 95% CI:

1.020–1.051, $P < 0.001$) and risk score (HR = 1.447, 95% CI:1.331–1.574, $P < 0.001$) were independent factors correlated with shorter overall survival (Figure 6B). However, the hazard ratios indicated that the prognostic model-derived risk score was a more powerful factor to assess patient prognosis. We also excluded patients without tumor stage information to investigate the correlation of the derived risk score with tumor stage. We compared the risk scores of the Stage i-ii group to those of the Stage iib-iv group and found that risk scores were higher in the latter than in the former (Supplementary Table 3).

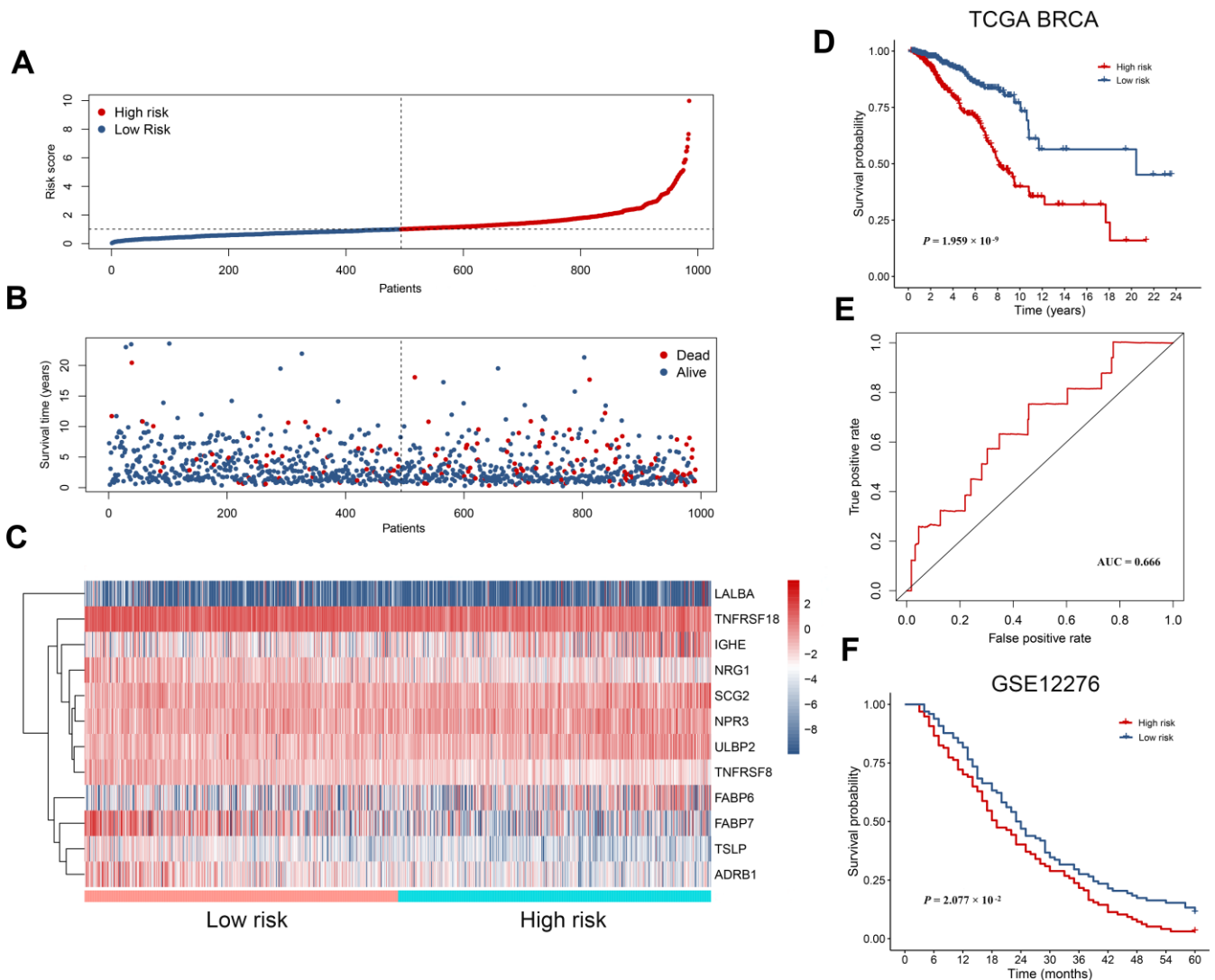


Figure 5. The prognostic immunity-based signature in breast cancer. (A) Distribution of risk score derived from the signature. Patients are ranked according to the corresponding risk score. (B) Survival status of breast cancer patients. They are ranked in the same way as in (A). (C) Heatmap showing expression of the 12 hub immunity-related genes in different risk groups. (D) Kaplan-Meier survival curves for patients in the TCGA BRCA dataset. Patients are assigned into high and low risk groups according to the median risk score. (E) The receiver operating characteristic (ROC) curve showing a prognostic value of the immunity-based signature. (F) Kaplan-Meier survival curves for patients in the validation dataset (GSE12276).

Associations of immunity-related signature with tumor-infiltrating immune cells

To explore whether the immunity-related signature could reflect infiltration degree in the tumor microenvironment of immune cells, we further analyzed the associations between the immune prognostic model-derived risk score and immune infiltrates. As shown in Figure 7A–7E, infiltration of B cells, CD4+ T cells, CD8+ T cells, neutrophils and dendritic cells was negatively correlated with risk score, suggesting that there are fewer tumor infiltrates in high risk patients. These results are consistent with previous findings [20–24]. However, we didn't find an association between macrophage infiltration and risk score (Figure 7F). We also investigated whether differences in the immune

phenotype could be observed between high- and low-risk groups. According to GSEA results, a total of 44 immunity-related GO terms were enriched (Figure 8A), indicating varied intensities of immune response between these two groups. Eight representative immunity-related terms correlated positively with low risk (Figure 8B), which included humoral immune response, regulation of immune effector process, T cell activation and differentiation involved in immune response, cytokine production involved in immune response, regulation of adaptive immune response, activation of innate immune response, and B cell activation involved in immune response. These data suggest that the low risk estimated for many breast cancer patients using our prognostic signature may be attributed to a hyperactivated immune response in the tumor microenvironment.

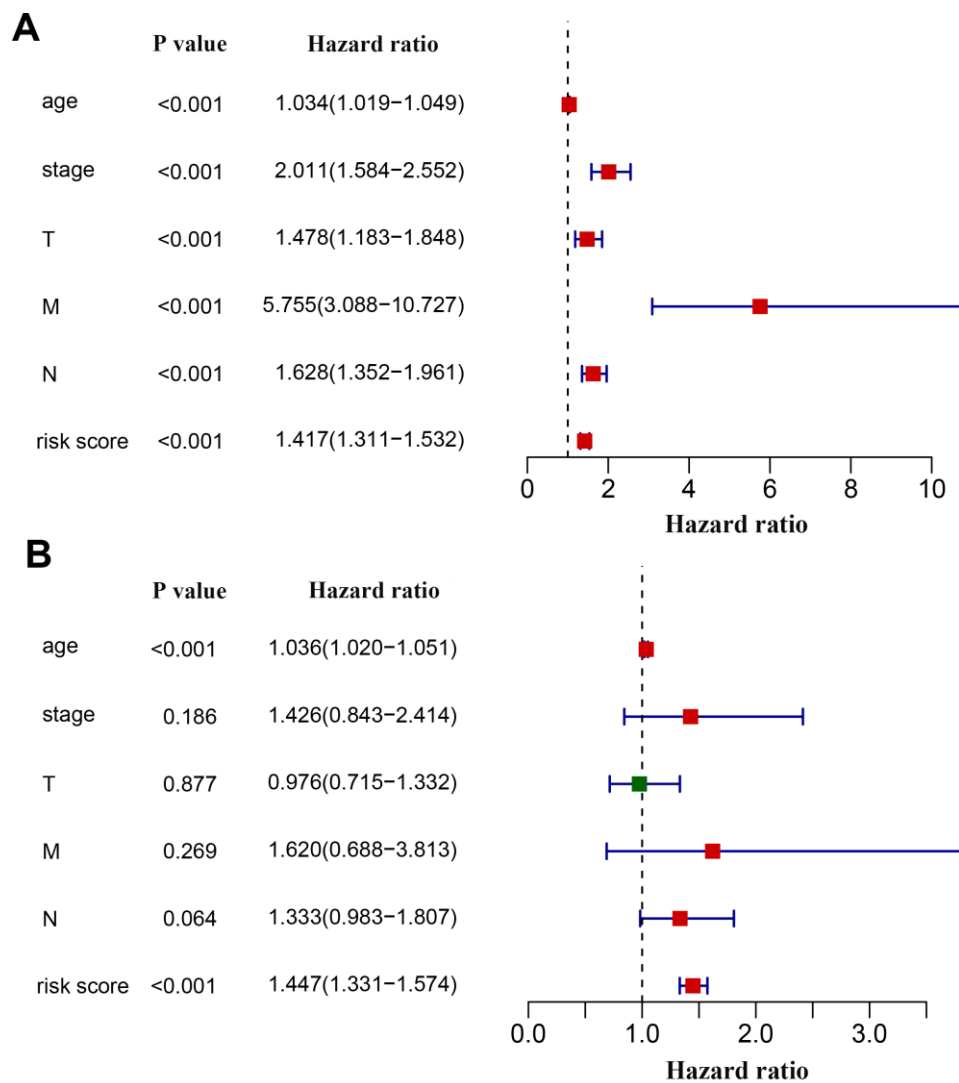


Figure 6. Effects of critical clinical characteristics on patients' overall survival in breast cancer. (A) Forest plot showing prognostic values of age, stage, TNM staging and the immunity-based signature-derived risk score. **(B)** Forest plot showing the prognostic value of age, stage, TNM staging and risk score as an independent factor.

DISCUSSION

Breast cancer is a heterogeneous disease that severely threatens the health of females. Although advances in early diagnosis and multiple therapies have contributed to large improvements in survival rate, no specific risk factors have been identified yet for the majority of breast cancer patients [5], and patients afflicted with metastatic disease will eventually succumb to it [25]. The immune system is actively involved in the development and progression of many solid tumors, including those of breast cancer. Accumulating evidence highlights that the response to antitumor therapy and overall survival of breast cancer patients is subject to host immunity [26]. In this regard, an immunity-based prognostic signature can be rationally applicable to identify patients with poor survival in advance. In this study, using two breast cancer cohorts containing over 1000 patients, we constructed and validated a prognostic signature based on the expression of 12 immunity-related genes. According to the risk score yielded from the prognostic signature, patients could be stratified into subgroups with distinct survival outcomes. Moreover, we found that this immune signature was negatively associated with immune infiltrates in the tumor microenvironment. These findings suggest that our immunity-related signature could be used effectively to evaluate prognosis and

provide potential immunotherapeutic targets for breast cancer patients.

The current study identified 179 differentially-expressed IRGs between tumoral and normal tissues, which might be involved in breast cancer initiation and progression. Accumulating studies establish that chronic inflammation can trigger tumor progression and an inflammatory surrounding microenvironment is indispensable to all tumors [27–29]. Cytokines, regardless of their source, play opposing roles in cancer progression, either facilitating or suppressing it. A delicate balance between antitumor immunity and tumor-promoting inflammation is achieved through cytokines and chemokines released by immune cells, tumor cells and other components in the tumor microenvironment [27, 30]. Our functional enrichment analysis revealed that the differentially-expressed IRGs participated in extracellular region components, inflammatory response, immune response, cytokine activity, interaction of cytokines with their receptors, and chemokine signaling pathways, which reflected the intricate relationship between inflammation, immunity and cancer, and the involvement of the differentially-expressed IRGs in breast cancer initiation and progression. Increasing evidence demonstrates a critical role of signal transducer and activator of transcription (STAT) proteins in determining whether immune

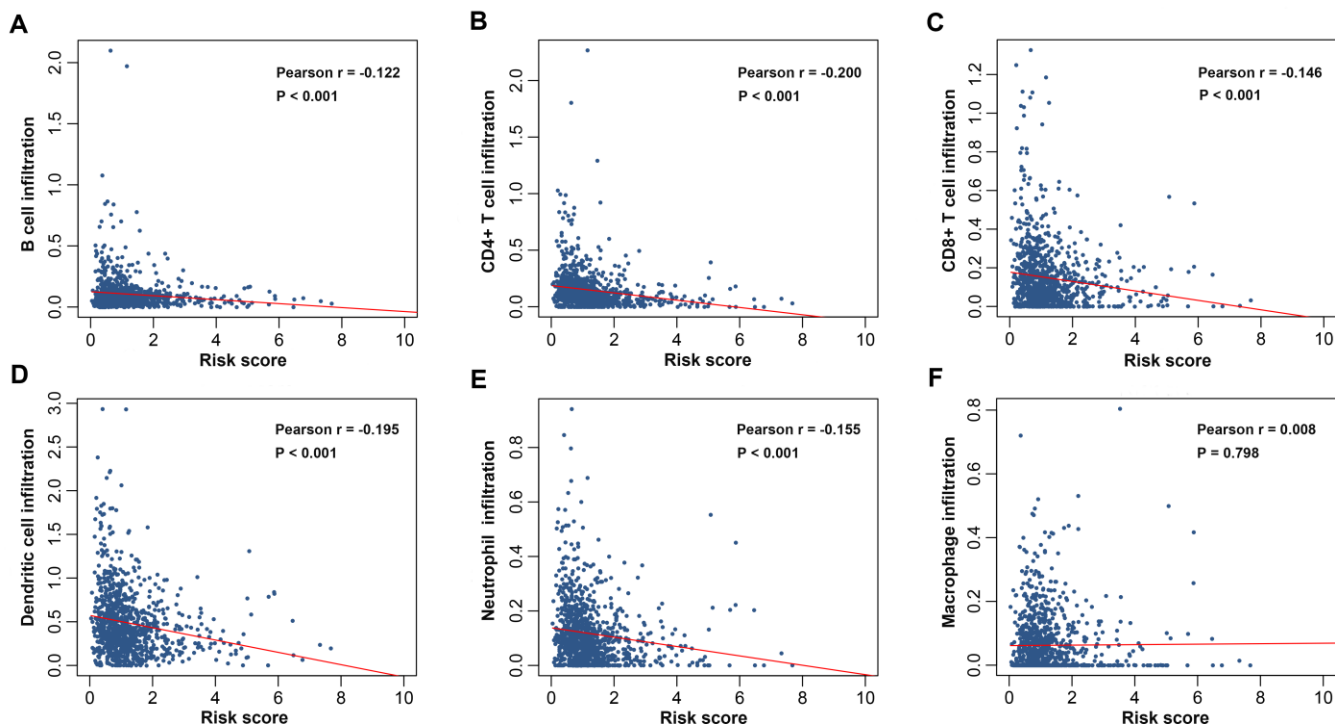


Figure 7. Correlations between the prognostic signature-derived risk score and infiltration abundances of multiple immune cells. (A) B, (B) CD4+ T, (C) CD8+ T, and (D) dendritic cells; (E) neutrophils; (F) macrophages.

responses in the surrounding milieu are pro-carcinogenic or anti-carcinogenic [31, 32]. In particular, cytokine-activated STAT3 can promote cancer proliferation and invasion while suppressing anti-tumor immunity by regulating the expression of immune

checkpoint proteins PD-1, PD-L1 and CTLA4 [31, 33–35]. According to the KEGG pathway analysis in our study, enrichment of Jak-STAT signaling suggested that differentially-expressed IRGs disrupt the balance between anti-tumor immunity and tumorigenesis.

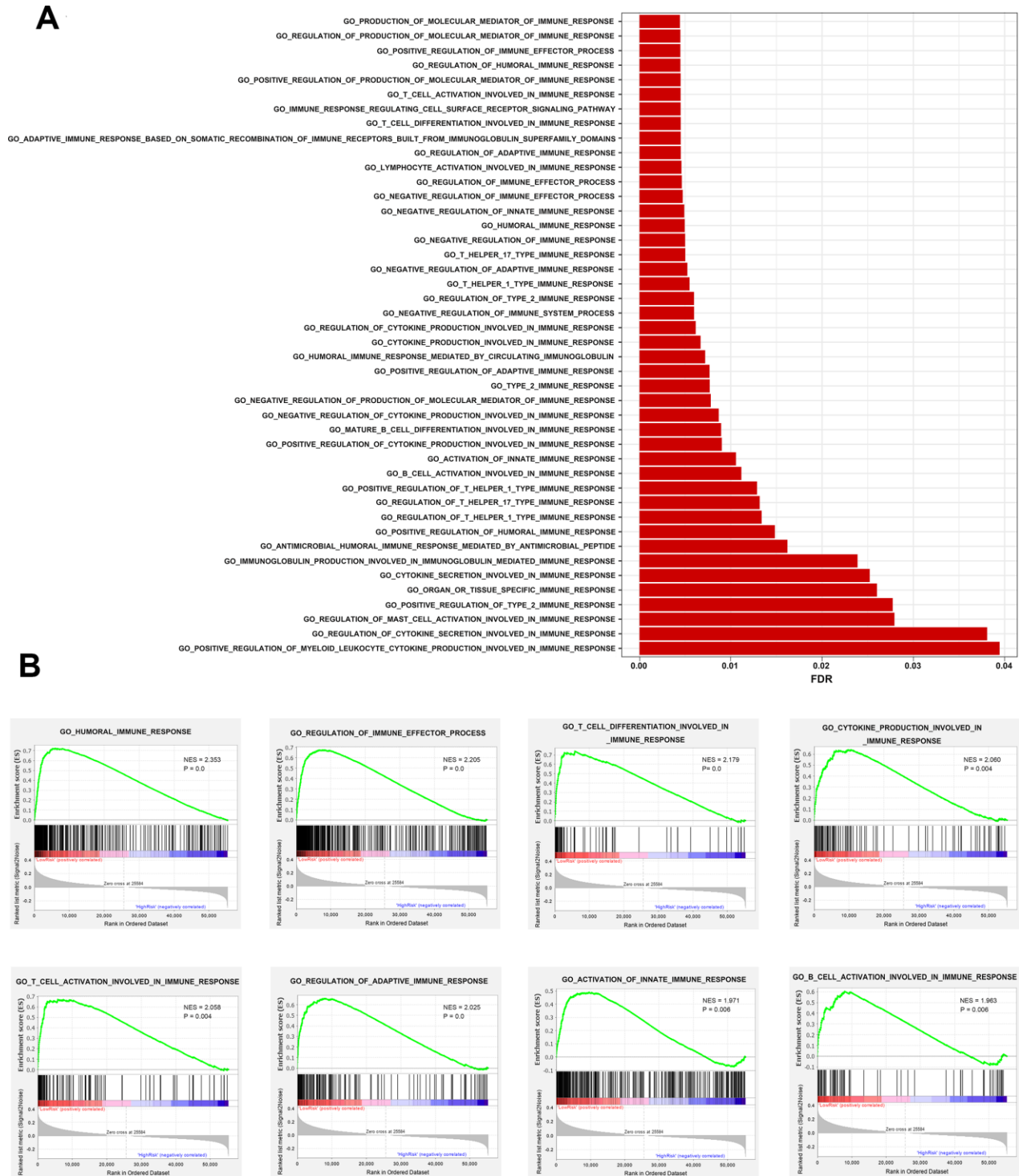


Figure 8. Comparison of immune traits between high-risk and low-risk groups via gene set enrichment analysis. (A) Enriched immunity-related GO terms. (B) Representative immunity-related terms positively correlated with low risk.

In order to answer whether and which of the differentially-expressed IRGs have an effect on the clinical outcomes of breast cancer patients, we assessed the associations between expression levels of these IRGs and patient overall survival. Out of the 179 IRGs, 7 were positively and 12 were negatively associated with favorable prognosis. The TFs-based regulatory network we built for these 19 survival-associated IRGs showed that MYH11, TP63 and ELF5, were critical regulators of the survival-associated IRGs. To date, the function of MYH11 in breast cancer remains elusive. Increasing evidence demonstrates that TP63 proteins are tumor suppressors which block the metastatic potential of tumor cells [36, 37], although an oncogenic TP63 isoform has also been identified [38]. ELF5 was established as a suppressor of epithelial-mesenchymal transition that inhibited metastasis in breast cancer [39]. We found that MYH11, TP63 and ELF5 were downregulated in breast cancer tissues, which is consistent with previous findings highlighting TP63 and ELF5 as tumor suppressors. Moreover, targets of MYH11 and TP63 (LIFR, NRG1, TSLP, TACR1), and of ELF5 (LALBA, MIA, S100B, FABP7) in the network were also downregulated (except LALBA), which were indicators of better overall survival according to univariate Cox regression. Further, positive associations between the hub TFs (MYH11, TP63 and ELF5) and their targets were found and showed in the network. Collectively, our data suggest that altered expression of certain IRGs correlated with survival in breast cancer patients, perhaps owing to changes in expression of tumor-related TFs, and that the tumor suppressors TP63 and ELF5 can exert their anti-tumor effects by disrupting immune responses in the tumor microenvironment. Further evidence is needed to experimentally elucidate underlying regulatory mechanisms of these immunity-related genes.

Our immunity-based prognostic signature consists of 12 survival-associated IRGs. Among them, ULBP2, TSLP, TNFRSF8 and TNFRSF18 are involved in intercellular cytokine-mediated communications. ULBP2 is one of the NKG2D ligands to cause release of multiple cytokines and chemokines that help to activate NK cells. However, soluble ULBP2 secreted by tumor cells also contributes to evasion of immunosurveillance. Recent studies have found that elevated expression of ULBP2 is an indicator of poor prognosis in ovarian and pancreatic cancer [40, 41]. Similarly, our analysis revealed that ULBP2 is upregulated in breast cancer with detrimental effects on patient overall survival. TNFRSF8 is a tumor necrosis factor with unclear contributions to breast cancer. Here, we found a decreased expression of TNFRSF8, similar to a recent immunohistochemistry-based study that reported a lack of CD30 expression in breast cancer [42]. Moreover, its downregulation was found to be associated with poor

prognosis. Interestingly, here we found that TSLP expression is reduced in breast cancer patients and correlates positively with a favorable prognosis, both in disagreement [43] and agreement [44] with previous reports. Further efforts are required to better understand TSLP functions in breast cancer. In line with previous research [45], we confirmed here that NRG1 is downregulated in breast cancer tissues, with detrimental effects on long-term survival. FABP7 was previously associated with lower lymph node stage and a longer disease-free survival [46], which supports our finding that high FABP7 expression contributes to a low risk score. Serum alpha-lactalbumin (LALBA) was previously identified as a marker for breast cancer [47]. We found that high LALBA expression was a hazardous prognostic factor. The roles of other IRGs in our signature in breast cancer, including IGHE, SCG2, NPR3 and FABP6, require further clarification.

We also investigated the relationships between risk score derived from IRGs-based signature and immune cell infiltration to figure out if the signature could reflect the immune microenvironment of breast cancer. Our data revealed that patients with a higher risk score had decreased infiltration degrees of B cells, CD4+ T cells, CD8+ T cells, neutrophils, and dendritic cells. The negative associations of risk score with immune cell infiltration suggested that the signature might serve as a predictor for local immune responses in the tumor bed. Tumor infiltrating B cells express and secrete antibodies to facilitate lysis and apoptosis of tumor cells [48]. The direct and indirect cytotoxic effects of CD8+ T cells and CD4+ T cells support the finding that their infiltration is correlated with favorable survival outcomes [20, 49]. Neutrophils play a critical role in activating and regulating immune cells as well as their effector functions [50]. A recent study has shown that infiltration by myeloperoxidase-positive neutrophils is an independent prognostic factor associated with a better overall survival in breast cancer [51]. Dendritic cells-mediated priming of T cells is a crucial step in antitumor immunity. It was found that a defective function of dendritic cells in patients with early breast cancer might be important for tumor progression [52]. Taken together, high risk score may be ascribed to, at least in part, unbalanced immune infiltrates and dysfunctional immune responses in the milieu surrounding tumors. It is proposed here that strategies attempting to enrich immune cells in the tumor microenvironment may have positive effects in breast cancer immunotherapy.

Although an immunity-related gene signature constructed based on IRGs differentially expressed between high- and low-immune-score groups in breast cancer was previously reported [53], our study differs

from it in that the current prognostic signature was computed on the basis of differentially-expressed IRGs between breast cancer tissues and adjacent normal tissues. This is a well-established approach to identify candidate genes involved in tumorigenesis. Some limitations exist in this study. First, although we validated the signature in an independent cohort, a larger sample size would allow for stronger validation. Second, the prognostic signature should be applied in clinical contexts to test its efficacy. Third, since our immunity-based prognostic signature is constructed using a set of specific genes, further experimental studies are warranted to reveal their functions in breast cancer initiation and progression. Fourth, breast cancer is a heterogeneous disease; thus, whether our prognostic signature is applicable to different histological types of breast cancer remains to be further evaluated in subsequent studies with samples from various histopathologic types.

MATERIALS AND METHODS

Data acquisition and processing

RNA-sequencing data of TCGA BRCA and relevant clinical information were downloaded from UCSC Xena (<https://xena.ucsc.edu/>). A validation cohort (GSE12276) was downloaded from Gene Expression Omnibus (GEO). In these two breast cancer datasets, only patients with complete survival information and a follow-up longer than three months were included for survival analysis. In addition, patients with prior treatment such as neoadjuvant chemotherapy and surgery were excluded.

Differentially-expressed, immunity-related gene analysis

To identify IRGs involved in breast cancer progression, we first screened differentially expressed genes (DEGs) between tumoral and normal tissue. FPKM data were transformed in the form of $\log_2(\text{FPKM}+0.001)$, $|\log_2 \text{fold change}| > 1$ and false discovery rates (FDR) < 0.05 were set as the cutoff values. A list of 2498 immunity-related genes was obtained from the Immunology Database and Analysis Portal (ImmPort) [54]. Then, the immunity-related gene list and DEGs were intersected to extract differentially-expressed IRGs. We performed hierarchical clustering analysis to visualize the expression patterns of the differentially-expressed IRGs between normal and breast cancer tissues.

Functional enrichment analysis

To further annotate the biological function of IRGs of interest, we carried out enrichment analysis of Kyoto

Encyclopedia of Genes and Genomes (KEGG) pathway [55] and Gene Ontology (GO) including biological process (BP), molecular function (MF) and cellular component (CC) genes by using the well-established DAVID Bioinformatics Resources [56]. *P* values were adjusted by the False-Discovery Rate (FDR) method for multiple comparisons and pathways with $\text{FDR} < 0.05$ were considered enriched.

Construction of the IRGs-based prognostic signature

Univariate Cox regression analysis was performed using the “survival package” of R to investigate the prognostic value of differentially-expressed IRGs in breast cancer patients. Ones with $P < 0.05$ were considered survival-associated IRGs, which were selected for multivariate Cox regression analyses to obtain independent prognostic IRGs and construct an immunity-based prognostic model. The risk score of each patient was the sum of products of the expression value and the regression coefficient of each independent prognostic IRG, $\text{risk score} = \sum_{i=1}^n \beta_i \text{Exp}_i$. Patients were then divided into high- and low-risk groups according to the median risk score. Survival analysis was performed by the Kaplan-Meier method and the log-rank test was utilized to compare the statistical significance of the difference between high- and low-risk groups.

Additionally, univariate and multivariate analyses were conducted to evaluate the effects of risk score and other critical clinical features on patient overall survival.

Regulatory network analysis between transcription factors and survival-associated IRGs

The Cistrome Cancer web resource enables us to investigate regulatory links between TFs and transcriptomes in cancer by integrating public cancer genomics data with chromatin profiling data [57]. We downloaded a total of 318 TFs from Cistrome Cancer, according to which we screened differentially-expressed TFs from the DEGs. Correlation tests were conducted between differentially-expressed TFs and survival-associated IRGs, and the regulatory network between them was visualized by using Cytoscape.

Association analysis between immunity-related signature and immune infiltrates

The correlation of the risk score of each patient with immune cell infiltration was analyzed by virtue of Tumor Immune Estimation Resource (TIMER), which is an online resource for systematical analysis in diverse cancers of immune infiltrating cells, including B cells,

CD4+ T cells, CD8+ T cells, neutrophils, macrophages and dendritic cells [58]. We used TIMER-estimated abundances of tumor-infiltrating immune cells to analyze their associations with IRGs-based prognostic signature in breast cancer.

Gene set enrichment analysis

To investigate immune response processes affected by the hub IRGs in the prognostic signature, we performed gene set enrichment analysis (GSEA) between high- and low-risk groups. Immunity-related gene sets in GO terms were downloaded from the Molecular Signatures Database (MSigDB) for GSEA analysis. A gene set was considered enriched for $P < 0.05$ and $FDR < 0.25$.

Statistical analysis

All analyses were performed using R software (version 3.4.4), except for GSEA analysis. Unless otherwise indicated, differences were considered statistically significant when $P < 0.05$.

AUTHOR CONTRIBUTIONS

TZ, XL, ZL designed this study. TZ and JZ analyzed the data. HZ, SH and WZ offered theoretical guidance. TZ drafted the manuscript. All authors approved the final manuscript.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

FUNDING

This work was supported by the National Natural Science Foundation of China (grant No. 81874327), the Key Research and Development Program of Hunan Province (grant No. 2019SK2251), and the Fundamental Research Funds for the Central Universities of Central South University (grant No. 2019zzts342).

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020; 70:7–30. <https://doi.org/10.3322/caac.21590> PMID:31912902
2. Provenzano E, Ulaner GA, Chin SF. Molecular classification of breast cancer. *PET Clin.* 2018; 13:325–38. <https://doi.org/10.1016/j.cpet.2018.02.004> PMID:30100073
3. Burstein MD, Tsimelzon A, Poage GM, Covington KR, Contreras A, Fuqua SA, Savage MI, Osborne CK, Hilsenbeck SG, Chang JC, Mills GB, Lau CC, Brown PH. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. *Clin Cancer Res.* 2015; 21:1688–98. <https://doi.org/10.1158/1078-0432.CCR-14-0432> PMID:25208879
4. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature.* 2012; 490:61–70. <https://doi.org/10.1038/nature11412> PMID:23000897
5. Emens LA. Breast cancer immunotherapy: facts and hopes. *Clin Cancer Res.* 2018; 24:511–20. <https://doi.org/10.1158/1078-0432.CCR-16-3001> PMID:28801472
6. Schreiber RD, Old LJ, Smyth MJ. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. *Science.* 2011; 331:1565–70. <https://doi.org/10.1126/science.1203486> PMID:21436444
7. de la Cruz-Merino L, Chiesa M, Caballero R, Rojo F, Palazón N, Carrasco FH, Sánchez-Margalet V. Breast cancer immunology and immunotherapy: current status and future perspectives. *Int Rev Cell Mol Biol.* 2017; 331:1–53. <https://doi.org/10.1016/bs.ircmb.2016.09.008> PMID:28325210
8. Bates JP, Derakhshandeh R, Jones L, Webb TJ. Mechanisms of immune evasion in breast cancer. *BMC Cancer.* 2018; 18:556. <https://doi.org/10.1186/s12885-018-4441-3> PMID:29751789
9. Kou Y, Koag MC, Lee S. Promutagenicity of 8-chloroguanine, a major inflammation-induced halogenated DNA lesion. *Molecules.* 2019; 24:3507. <https://doi.org/10.3390/molecules24193507> PMID:31569643
10. Koag MC, Kou Y, Ouzon-Shubeita H, Lee S. Transition-state destabilization reveals how human DNA polymerase β proceeds across the chemically unstable lesion N7-methylguanine. *Nucleic Acids Res.* 2014; 42:8755–66. <https://doi.org/10.1093/nar/gku554> PMID:24966350
11. Wen R, Umeano AC, Kou Y, Xu J, Farooqi AA. Nanoparticle systems for cancer vaccine. *Nanomedicine (Lond).* 2019; 14:627–48. <https://doi.org/10.2217/nnm-2018-0147> PMID:30806568
12. Zhang J, Zhang S, Zuo L, Yue W, Li S, Xin S, Liu L, Lu J. Differential expression profiling of lncRNAs related to

- epstein-barr virus infection in the epithelial cells. *J Med Virol.* 2019; 91:1845–55.
<https://doi.org/10.1002/jmv.25516> PMID:[31199512](https://pubmed.ncbi.nlm.nih.gov/31199512/)
13. Li G, Liu Y, Liu C, Su Z, Ren S, Wang Y, Deng T, Huang D, Tian Y, Qiu Y. Genome-wide analyses of long noncoding RNA expression profiles correlated with radioresistance in nasopharyngeal carcinoma via next-generation deep sequencing. *BMC Cancer.* 2016; 16:719.
<https://doi.org/10.1186/s12885-016-2755-6> PMID:[27599611](https://pubmed.ncbi.nlm.nih.gov/27599611/)
 14. Zhu T, Gao YF, Chen YX, Wang ZB, Yin JY, Mao XY, Li X, Zhang W, Zhou HH, Liu ZQ. Genome-scale analysis identifies GJB2 and ERO1LB as prognosis markers in patients with pancreatic cancer. *Oncotarget.* 2017; 8:21281–89.
<https://doi.org/10.18632/oncotarget.15068> PMID:[28177904](https://pubmed.ncbi.nlm.nih.gov/28177904/)
 15. Gao YF, Mao XY, Zhu T, Mao CX, Liu ZX, Wang ZB, Li L, Li X, Yin JY, Zhang W, Zhou HH, Liu ZQ. COL3A1 and SNAP91: novel glioblastoma markers with diagnostic and prognostic value. *Oncotarget.* 2016; 7:70494–503.
<https://doi.org/10.18632/oncotarget.12038> PMID:[27655637](https://pubmed.ncbi.nlm.nih.gov/27655637/)
 16. Gao S, Zhang H, Zuo X, Xiao Y, Liu D, Zhu H, Luo H. Integrated comparison of the miRNAome and mRNAome in muscles of dermatomyositis and polymyositis reveals common and specific miRNA-mRNAs. *Epigenomics.* 2019; 11:23–33.
<https://doi.org/10.2217/epi-2018-0064> PMID:[30523707](https://pubmed.ncbi.nlm.nih.gov/30523707/)
 17. Yu B, Zhang K, Milner JJ, Toma C, Chen R, Scott-Browne JP, Pereira RM, Crotty S, Chang JT, Pipkin ME, Wang W, Goldrath AW. Epigenetic landscapes reveal transcription factors that regulate CD8⁺ T cell differentiation. *Nat Immunol.* 2017; 18:573–82.
<https://doi.org/10.1038/ni.3706> PMID:[28288100](https://pubmed.ncbi.nlm.nih.gov/28288100/)
 18. Sizemore GM, Pitarresi JR, Balakrishnan S, Ostrowski MC. The ETS family of oncogenic transcription factors in solid tumours. *Nat Rev Cancer.* 2017; 17:337–51.
<https://doi.org/10.1038/nrc.2017.20> PMID:[28450705](https://pubmed.ncbi.nlm.nih.gov/28450705/)
 19. Yang L, Liu L, Gao H, Pinnamaneni JP, Sanagasetti D, Singh VP, Wang K, Mathison M, Zhang Q, Chen F, Mo Q, Rosengart T, Yang J. The stem cell factor SALL4 is an essential transcriptional regulator in mixed lineage leukemia-rearranged leukemogenesis. *J Hematol Oncol.* 2017; 10:159.
<https://doi.org/10.1186/s13045-017-0531-y> PMID:[28974232](https://pubmed.ncbi.nlm.nih.gov/28974232/)
 20. Mahmoud SM, Paish EC, Powe DG, Macmillan RD, Grainge MJ, Lee AH, Ellis IO, Green AR. Tumor-infiltrating CD8⁺ lymphocytes predict clinical outcome in breast cancer. *J Clin Oncol.* 2011; 29:1949–55.
<https://doi.org/10.1200/JCO.2010.30.5037> PMID:[21483002](https://pubmed.ncbi.nlm.nih.gov/21483002/)
 21. Loi S, Sirtaine N, Piette F, Salgado R, Viale G, Van Eenoo F, Rouas G, Francis P, Crown JP, Hitre E, de Azambuja E, Quinaux E, Di Leo A, et al. Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98. *J Clin Oncol.* 2013; 31:860–67.
<https://doi.org/10.1200/JCO.2011.41.0902> PMID:[23341518](https://pubmed.ncbi.nlm.nih.gov/23341518/)
 22. Ali HR, Provenzano E, Dawson SJ, Blows FM, Liu B, Shah M, Earl HM, Poole CJ, Hiller L, Dunn JA, Bowden SJ, Twelves C, Bartlett JM, et al. Association between CD8⁺ t-cell infiltration and breast cancer survival in 12,439 patients. *Ann Oncol.* 2014; 25:1536–43.
<https://doi.org/10.1093/annonc/mdu191> PMID:[24915873](https://pubmed.ncbi.nlm.nih.gov/24915873/)
 23. Loi S, Michiels S, Salgado R, Sirtaine N, Jose V, Fumagalli D, Kellokumpu-Lehtinen PL, Bono P, Kataja V, Desmedt C, Piccart MJ, Loibl S, Denkert C, 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.* 2014; 25:1544–50.
<https://doi.org/10.1093/annonc/mdu112> PMID:[24608200](https://pubmed.ncbi.nlm.nih.gov/24608200/)
 24. Adams S, Gray RJ, Demaria S, Goldstein L, Perez EA, Shulman LN, Martino S, Wang M, Jones VE, Saphner TJ, Wolff AC, Wood WC, Davidson NE, 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.* 2014; 32:2959–66.
<https://doi.org/10.1200/JCO.2013.55.0491> PMID:[25071121](https://pubmed.ncbi.nlm.nih.gov/25071121/)
 25. García-Aranda M, Redondo M. Immunotherapy: a challenge of breast cancer treatment. *Cancers (Basel).* 2019; 11:1822.
<https://doi.org/10.3390/cancers11121822> PMID:[31756919](https://pubmed.ncbi.nlm.nih.gov/31756919/)
 26. Savas P, Salgado R, Denkert C, Sotiriou C, Darcy PK, Smyth MJ, Loi S. Clinical relevance of host immunity in breast cancer: from TILs to the clinic. *Nat Rev Clin Oncol.* 2016; 13:228–41.
<https://doi.org/10.1038/nrclinonc.2015.215> PMID:[26667975](https://pubmed.ncbi.nlm.nih.gov/26667975/)
 27. Lin WW, Karin M. A cytokine-mediated link between innate immunity, inflammation, and cancer. *J Clin Invest.* 2007; 117:1175–83.

- <https://doi.org/10.1172/JCI31537>
PMID:17476347
28. Balkwill F, Charles KA, Mantovani A. Smoldering and polarized inflammation in the initiation and promotion of Malignant disease. *Cancer Cell*. 2005; 7:211–17.
<https://doi.org/10.1016/j.ccr.2005.02.013>
PMID:15766659
29. Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. *Cell*. 2010; 140:883–99.
<https://doi.org/10.1016/j.cell.2010.01.025>
PMID:20303878
30. Ben-Baruch A. Inflammation-associated immune suppression in cancer: the roles played by cytokines, chemokines and additional mediators. *Semin Cancer Biol*. 2006; 16:38–52.
<https://doi.org/10.1016/j.semcancer.2005.07.006>
PMID:16139507
31. Yu H, Pardoll D, Jove R. STATs in cancer inflammation and immunity: a leading role for STAT3. *Nat Rev Cancer*. 2009; 9:798–809.
<https://doi.org/10.1038/nrc2734> PMID:19851315
32. Kortylewski M, Xin H, Kujawski M, Lee H, Liu Y, Harris T, Drake C, Pardoll D, Yu H. Regulation of the IL-23 and IL-12 balance by Stat3 signaling in the tumor microenvironment. *Cancer Cell*. 2009; 15:114–23.
<https://doi.org/10.1016/j.ccr.2008.12.018>
PMID:19185846
33. Huynh J, Chand A, Gough D, Ernst M. Therapeutically exploiting STAT3 activity in cancer - using tissue repair as a road map. *Nat Rev Cancer*. 2019; 19:82–96.
<https://doi.org/10.1038/s41568-018-0090-8>
PMID:30578415
34. Celada LJ, Kropski JA, Herazo-Maya JD, Luo W, Creecy A, Abad AT, Chioma OS, Lee G, Hassell NE, Shaginurova GI, Wang Y, Johnson JE, Kerrigan A, et al. PD-1 up-regulation on CD4⁺ T cells promotes pulmonary fibrosis through STAT3-mediated IL-17A and TGF-β1 production. *Sci Transl Med*. 2018; 10:eaar8356.
<https://doi.org/10.1126/scitranslmed.aar8356>
PMID:30257954
35. Austin JW, Lu P, Majumder P, Ahmed R, Boss JM. STAT3, STAT4, NFATc1, and CTCF regulate PD-1 through multiple novel regulatory regions in murine T cells. *J Immunol*. 2014; 192:4876–86.
<https://doi.org/10.4049/jimmunol.1302750>
PMID:24711622
36. Barbieri CE, Tang LJ, Brown KA, Pietenpol JA. Loss of p63 leads to increased cell migration and up-regulation of genes involved in invasion and metastasis. *Cancer Res*. 2006; 66:7589–97.
<https://doi.org/10.1158/0008-5472.CAN-06-2020>
PMID:16885358
37. Su X, Chakravarti D, Cho MS, Liu L, Gi YJ, Lin YL, Leung ML, El-Naggar A, Creighton CJ, Suraokar MB, Wistuba I, Flores ER. TAp63 suppresses metastasis through coordinate regulation of dicer and miRNAs. *Nature*. 2010; 467:986–90.
<https://doi.org/10.1038/nature09459> PMID:20962848
38. Gatti V, Bongiorno-Borbone L, Fierro C, Annicchiarico-Petruzzelli M, Melino G, Peschiaroli A. P63 at the crossroads between stemness and metastasis in breast cancer. *Int J Mol Sci*. 2019; 20:2683.
<https://doi.org/10.3390/ijms20112683>
PMID:31159154
39. Chakrabarti R, Hwang J, Andres Blanco M, Wei Y, Lukačičin M, Romano RA, Smalley K, Liu S, Yang Q, Ibrahim T, Mercatali L, Amadori D, Haffty BG, et al. Elf5 inhibits the epithelial-mesenchymal transition in mammary gland development and breast cancer metastasis by transcriptionally repressing Snail2. *Nat Cell Biol*. 2012; 14:1212–22.
<https://doi.org/10.1038/ncb2607>
PMID:23086238
40. Li K, Mandai M, Hamanishi J, Matsumura N, Suzuki A, Yagi H, Yamaguchi K, Baba T, Fujii S, Konishi I. Clinical significance of the NKG2D ligands, MICA/B and ULBP2 in ovarian cancer: high expression of ULBP2 is an indicator of poor prognosis. *Cancer Immunol Immunother*. 2009; 58:641–52.
<https://doi.org/10.1007/s00262-008-0585-3>
PMID:18791713
41. Kegasawa T, Tatsumi T, Yoshioka T, Suda T, Ikezawa K, Nakabori T, Yamada R, Kodama T, Shigekawa M, Hikita H, Sakamori R, Takehara T. Soluble UL16-binding protein 2 is associated with a poor prognosis in pancreatic cancer patients. *Biochem Biophys Res Commun*. 2019; 517:84–88.
<https://doi.org/10.1016/j.bbrc.2019.07.020>
PMID:31303272
42. Nassif S, El-Zaatari ZM, Attieh M, Hijazi M, Fakhreddin N, Aridi T, Boulos F. Lack of expression of ALK and CD30 in breast carcinoma by immunohistochemistry irrespective of tumor characteristics. *Medicine (Baltimore)*. 2019; 98:e16702.
<https://doi.org/10.1097/MD.00000000000016702>
PMID:31393373
43. Kuan EL, Ziegler SF. A tumor-myeloid cell axis, mediated via the cytokines IL-1α and TSLP, promotes the progression of breast cancer. *Nat Immunol*. 2018; 19:366–74.
<https://doi.org/10.1038/s41590-018-0066-6>
PMID:29556001
44. Ghirelli C, Sadacca B, Reyat F, Zollinger R, Michea P, Sirven P, Pattarini L, Martínez-Cingolani C, Guillot-Delost M, Nicolas A, Scholer-Dahirel A, Soumelis V. No

- evidence for TSLP pathway activity in human breast cancer. *Oncoimmunology*. 2016; 5:e1178438.
<https://doi.org/10.1080/2162402X.2016.1178438>
PMID:[27622057](https://pubmed.ncbi.nlm.nih.gov/27622057/)
45. Chua YL, Ito Y, Pole JC, Newman S, Chin SF, Stein RC, Ellis IO, Caldas C, O'Hare MJ, Murrell A, Edwards PA. The NRG1 gene is frequently silenced by methylation in breast cancers and is a strong candidate for the 8p tumour suppressor gene. *Oncogene*. 2009; 28:4041–52.
<https://doi.org/10.1038/onc.2009.259> PMID:[19802002](https://pubmed.ncbi.nlm.nih.gov/19802002/)
46. Zhang H, Rakha EA, Ball GR, Spiteri I, Aleskandarany M, Paish EC, Powe DG, Macmillan RD, Caldas C, Ellis IO, Green AR. The proteins FABP7 and OATP2 are associated with the basal phenotype and patient outcome in human breast cancer. *Breast Cancer Res Treat*. 2010; 121:41–51.
<https://doi.org/10.1007/s10549-009-0450-x>
PMID:[19590950](https://pubmed.ncbi.nlm.nih.gov/19590950/)
47. Thean ET, Toh BH. Serum human alpha-lactalbumin as a marker for breast cancer. *Br J Cancer*. 1990; 61:773–75.
<https://doi.org/10.1038/bjc.1990.173> PMID:[2337516](https://pubmed.ncbi.nlm.nih.gov/2337516/)
48. Ruffell B, Au A, Rugo HS, Esserman LJ, Hwang ES, Coussens LM. Leukocyte composition of human breast cancer. *Proc Natl Acad Sci USA*. 2012; 109:2796–801.
<https://doi.org/10.1073/pnas.1104303108>
PMID:[21825174](https://pubmed.ncbi.nlm.nih.gov/21825174/)
49. Matsumoto H, Thike AA, Li H, Yeong J, Koo SL, Dent RA, Tan PH, Iqbal J. Increased CD4 and CD8-positive T cell infiltrate signifies good prognosis in a subset of triple-negative breast cancer. *Breast Cancer Res Treat*. 2016; 156:237–47.
<https://doi.org/10.1007/s10549-016-3743-x>
PMID:[26960711](https://pubmed.ncbi.nlm.nih.gov/26960711/)
50. Mantovani A, Cassatella MA, Costantini C, Jaillon S. Neutrophils in the activation and regulation of innate and adaptive immunity. *Nat Rev Immunol*. 2011; 11:519–31.
<https://doi.org/10.1038/nri3024> PMID:[21785456](https://pubmed.ncbi.nlm.nih.gov/21785456/)
51. Zeindler J, Angehrn F, Droeser R, Däster S, Piscuoglio S, Ng CK, Kilic E, Mechera R, Meili S, Isaak A, Weber WP, Muenst S, Soysal SD. Infiltration by myeloperoxidase-positive neutrophils is an independent prognostic factor in breast cancer. *Breast Cancer Res Treat*. 2019; 177:581–89.
<https://doi.org/10.1007/s10549-019-05336-3>
PMID:[31267330](https://pubmed.ncbi.nlm.nih.gov/31267330/)
52. Saththaporn S, Robins A, Vassanasiri W, El-Sheemy M, Jibril JA, Clark D, Valerio D, Eremin O. Dendritic cells are dysfunctional in patients with operable breast cancer. *Cancer Immunol Immunother*. 2004; 53:510–18.
<https://doi.org/10.1007/s00262-003-0485-5>
PMID:[14740176](https://pubmed.ncbi.nlm.nih.gov/14740176/)
53. Xie P, Ma Y, Yu S, An R, He J, Zhang H. Development of an immune-related prognostic signature in breast cancer. *Front Genet*. 2020; 10:1390.
<https://doi.org/10.3389/fgene.2019.01390>
PMID:[32047513](https://pubmed.ncbi.nlm.nih.gov/32047513/)
54. Bhattacharya S, Andorf S, Gomes L, Dunn P, Schaefer H, Pontius J, Berger P, Desborough V, Smith T, Campbell J, Thomson E, Monteiro R, Guimaraes P, et al. ImmPort: disseminating data to the public for the future of immunology. *Immunol Res*. 2014; 58:234–39.
<https://doi.org/10.1007/s12026-014-8516-1>
PMID:[24791905](https://pubmed.ncbi.nlm.nih.gov/24791905/)
55. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res*. 2012; 40:D109–14.
<https://doi.org/10.1093/nar/gkr988> PMID:[22080510](https://pubmed.ncbi.nlm.nih.gov/22080510/)
56. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; 4:44–57.
<https://doi.org/10.1038/nprot.2008.211>
PMID:[19131956](https://pubmed.ncbi.nlm.nih.gov/19131956/)
57. Mei S, Meyer CA, Zheng R, Qin Q, Wu Q, Jiang P, Li B, Shi X, Wang B, Fan J, Shih C, Brown M, Zang C, Liu XS. Cistrome cancer: a web resource for integrative gene regulation modeling in cancer. *Cancer Res*. 2017; 77:e19–22.
<https://doi.org/10.1158/0008-5472.CAN-17-0327>
PMID:[29092931](https://pubmed.ncbi.nlm.nih.gov/29092931/)
58. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, Liu XS. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res*. 2017; 77:e108–10.
<https://doi.org/10.1158/0008-5472.CAN-17-0307>
PMID:[29092952](https://pubmed.ncbi.nlm.nih.gov/29092952/)

SUPPLEMENTARY MATERIALS

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Table 2.

Supplementary Table 1. Differentially expressed immunity-related genes identified between breast cancer and normal tissues.

Regulation	Gene symbol
Up-regulated (69)	<i>CST4, CGA, VGF, CCL11, SCT, FABP6, S100A7A, TUBB3, CCL7, CCR8, IGHE, CAMP, PRLH, ORM1, TFR2, ULBP1, EPO, TNFRSF9, UMODL1, S100P, PROC, ARTN, OPRD1, IL21R, AMH, CXCL11, BIRC5, BMP8A, IL20, IL9R, CD1A, ORM2, ESM1, INHBA, PCSK1, PYY, IL27, TNFRSF18, BMPR1B, LTA, TRH, CCL20, ICOS, TMPRSS6, KCNH2, TNFSF4, EDN2, IL11, CCR4, ADM2, PDCD1, LCN12, ANGPTL6, WFIKKN1, CD19, OLR1, CXCL10, SCG2, SLC11A1, MMP12, TRAV26-1, OASL, PIK3R2, FGFR3, THPO, ULBP2, RBP2, TRAV39, INHA</i>
Down-regulated (110)	<i>ANGPTL7, ACVR1C, DEFB132, LEP, TSLP, LALBA, PENK, GDF10, RNASE7, EDN3, GFAP, RXRG, CCL14, SSTR1, PAK3, ADCYAP1R1, GPR17, CCL24, NTF4, NRG2, CXCL2, CMA1, ADIPOQ, AVPR2, TACR1, KL, PLXNA4, IL17B, BMP5, ANGPTL1, RBP4, OXTR, PROK1, NRG1, ANGPT1, DES, CXCL3, FAM3D, ALB, IGF1, ADRB2, PTGFR, FGF2, LEPR, SEMA3D, GREM2, BMP2, PTGS2, SAA2, S100B, CNTFR, IL6, TAC1, NTS, CD209, NPR1, CCL23, ADRB1, PTH1R, IL22RA2, PPARG, NR3C2, IL33, FABP4, PII5, PTX3, ROBO3, SEMA3A, SEMA6D, PI3, NTF3, GHR, FABP7, BMP6, TNFRSF10D, OGN, CTSG, LGR6, EGFR, LIFR, NR4A3, CCL13, SCGB3A1, SPINK5, CXCL5, MASP2, EDNRB, CCL28, IL17D, TGFBR3, SHC3, NPR3, NGFR, FGF10, FGF7, CXCL6, MIA, PTH2R, BACH2, NR5A2, FGF16, SEMA3G, FGF1, GNAI1, MARCO, CCL21, DEFB1, TNFRSF8, SAA1, S100A12</i>

Supplementary Table 2. Functional enrichment analysis for differentially expressed immunity-related genes in breast cancer.

Supplementary Table 3. Correlation of risk score with tumor stage in breast cancer.

Tumor stage	Number of patients	Risk score	P
i-ii	497	1.205 ± 1.011	0.044
iiib-iv	466	1.338 ± 1.054	