Construction and Validation of an Immune-related Gene Pairs Prognostic Model for Breast Cancer

Ying Zhong

peking union medical college hospital

Zhe Wang

PUMCH: Peking Union Medical College Hospital

Yidong Zhou

peking union medical college hospital

Feng Mao

PUMCH: Peking Union Medical College Hospital

Yan Lin

PUMCH: Peking Union Medical College Hospital

Jinghong Guan

PUMCH: Peking Union Medical College Hospital

Songjie Shen

PUMCH: Peking Union Medical College Hospital

Yali Xu

PUMCH

Xin Huang

PUMCH: Peking Union Medical College Hospital

Xuefei Wang

PUMCH: Peking Union Medical College Hospital

Qiang Sun (✉ Sunqiang_pumch@163.com)

PUMCH: Peking Union Medical College Hospital

Research

Keywords: breast cancer, immune-related genes, prognostic model, immune cells, biomarkers

Posted Date: January 29th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-153832/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Immunotherapy plays an increasingly important role in the treatment of advanced female breast cancer, which has the highest mortality rate among malignant tumors. The purpose of this study was to identify immune-related genes associated with breast cancer prognosis as possible targets of immunotherapy, and their related biological processes and signaling pathways.

Methods: Clinical data and gene expression profiles of patients with breast cancer were extracted from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases and divided into training (n = 1053) and verification (n = 508) groups. CIBERSORT was used to predict differences in immune cell infiltration in patient subsets stratified according to risk. Gene Ontology (GO) enrichment analysis was used to identify pathways associated with immune-related genes in patient subsets stratified according to risk.

Results: The prognostic model composed of 27 immune-related gene pairs significantly distinguished between high- and low-risk patients. Univariate and multivariate analyses indicated that the model was an independent prognostic factor for breast cancer. Among the identified genes, *APOBEC3G*, *PLXNB1*, and *C3AR1* had not been previously studied in breast cancer and warrant further exploration. CCR chemokine receptor binding, regulation of leukocyte-mediated cytotoxicity, T cell migration, T cell receptor complex, and other pathways were significantly enriched in low-risk patients. M2 and M0 macrophages were more highly expressed in high-risk than in low-risk patients. CD8+ T cells and naïve B cells were more abundant in low-risk than in high-risk patients.

Conclusion: The immune-related gene pairs prognostic model developed in the current study can help assess breast cancer prognosis and provides a potential target and research direction for breast cancer immunotherapy in the future.

Introduction:

Breast cancer is the most serious malignant tumor threatening the health of women worldwide. It is the leading global cause of cancer deaths in women and remains incurable in advanced stage¹. The American Cancer Association estimated that 279,100 Americans would be diagnosed with breast cancer, and 42,690 would die from the disease in 2020².

Approximately 3–10% of new breast cancer patients are diagnosed with distant metastasis³. Metastatic breast cancer remains an almost incurable disease, with an overall survival (OS) of approximately 3 years and a 5-year survival rate of approximately 25%⁴. For treatment of patients with hormone receptor (HR) + advanced breast cancer, a variety of new drugs have been developed in recent years, such as fulvestrant, and CDK4/6 inhibitors^{5,6}.

The mutation burden of breast cancer is lighter than that of other malignant tumors, although that of human epidermal growth factor receptor 2 (HER2) + tumors and triple-negative breast cancer (TNBC) is

heavier than that of luminal-type tumors^{7,8}. Similarly, the proportion of tumor-infiltrating lymphocytes (TILs) in HER2 + tumors and TNBC is much higher than that in HR + tumors, in which the increased proportion of TILs indicates better prognosis^{9,10}. In the TRYPHAENA trial, TILs were not associated with the PCR but with improved event-free survival (EFS)¹¹. These data indicate the immunogenicity of breast cancer.

Although breast cancer is not considered an inflammatory tumor, the immunogenicity of TNBC and HER2 + tumors is higher than that of luminal A-subtype breast cancer¹². Immunotherapy has become the first treatment choice for patients diagnosed with advanced programmed death ligand (PD-L)1+ (TNBC)¹³.

Increasing numbers of studies have investigated immunotherapy treatments for breast cancer. Traditional immunotherapy for HER2 + breast cancer comprises mainly the HER2-blocking antibody trastuzumab. Although immune checkpoint inhibitors (ICIs) have been successfully used in the treatment of melanoma and lung cancer¹⁴, the first effective ICI in the treatment of metastatic TNBC was the PD-L1 antibody atezolizumab combined with nab-paclitaxel¹⁵. In neoadjuvant therapy, patients treated with atezolizumab had a higher pathological complete response rate (PCR) than controls¹⁶. However, only PD-L1 was found to predict the efficacy of atezolizumab treatment¹³, suggesting that the role of tumor immunity needs to be better understood to identify additional immune biomarkers and potential therapeutic targets. Single-center or single-ICI studies provide limited clinical and genetic data; more data are needed to provide evidence for combination treatments with ICIs. High-throughput technology and large-scale gene databases can enable the discovery of unique immune-related genes, resulting in more flexible immunotherapy treatments for each tumor type.

PD-L1 expression in immunocytes was used to evaluate the efficacy of a combination of atezolizumab plus nab-paclitaxel in the Impassion130 trial¹³. However, another study revealed no clear correlation between PD-L1 expression and clinical efficacy¹⁶. To date, a predictive biomarker for the efficacy of ICI in breast cancer therapy has not been identified.

Therefore, we aimed to identify powerful biomarkers for the prediction of immune-checkpoint blockade (ICB) responsiveness using data extracted from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases. Further, we combined these data with those in the immunology database and analysis portal ImmPort to investigate relevant molecular mechanisms and immune cell relationships.

Materials And Methods:

Obtaining breast cancer gene expression data

This was a retrospective study of gene expression and corresponding clinical data based on two independent datasets obtained from publicly available databases. In total, data from 1561 patients were analyzed. Expression of 56,737 genes and survival outcome data of 1053 patients were obtained from TCGA (<https://portal.gdc.cancer.gov/repository>). Gene expression and disease-free survival (DFS) of 508

patients were obtained from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE25066>).

Construction of immune-related gene pairs prognostic model

In order to construct an immune-related gene prognostic model, 2498 immune-related genes were obtained from the ImmPort database (<https://www.immport.org/home>) on May 30, 2020. This gene platform includes a list of immunologically relevant genes curated with functions and Gene Ontology terms. The ImmuneRegulation web-based tool identified regulators of immune system-specific genes of interest, and the Immcantation framework analyzed high-throughput adaptive immune receptor repertoire sequencing (AIRR-Seq) datasets characterizing B cell and T cell receptors. In this study, we retained only immune-related genes identified in both the GEO and TCGA datasets with a median absolute deviation > 0.5 ¹⁷. Relative expression within each immune-related gene pair was compared for each patient in the TCGA dataset. In each pair, if expression of one gene was larger than that of the other, the value of the gene pair was considered 1; otherwise, the value was considered 0. After removing immune-related gene pairs with relatively small variations in expression within the pair ($< 20\%$), least absolute shrinkage and selection operator (Lasso) regression was performed for 1000 simulations, and a prognostic model containing 27 immune-related gene pairs was obtained. The risk value of each patient in the TCGA dataset was calculated using the model. A receiver operating characteristic (ROC) curve was established using the risk values, and an optimal cut-off value was determined to distinguish between low- and high-risk patients.

Verification of immune-related gene pairs prognostic model

To further verify the immune-related gene pairs prognostic model, the GEO dataset was used as the validation group. The risk value of each patient in the GEO dataset was calculated using the model, and the cut-off value obtained for the training group was used to stratify the GEO patients into high- and low-risk groups. Univariate and multivariate Cox proportional hazards analyses were used to verify whether the model could be used as an independent prognostic factor relative to other clinical features such as age, HR, HER2, and American Joint Committee on Cancer (AJCC) stage in the GEO and TCGA datasets.

Immune cell infiltration associated with immune-related gene pairs prognostic model

The CIBERSORT algorithm was used to estimate differences in immune cell infiltration using gene expression data in the high- and low-risk TCGA groups¹⁸. This algorithm can predict the proportion of 22 types of tumor-infiltrating immune cells, such as T cells, B cells, macrophages, and natural killer cells, using gene expression data.

Enrichment analysis by Gene Ontology (GO)

Enrichment analysis of the identified immune-related genes was performed using g:Profiler^{19,20}. All GO gene sets were downloaded from the Gene Set Enrichment Analysis (GSEA) website (<https://www.gsea->

msigdb.org/gsea/index.jsp), and gene sets in high- and low-risk TCGA groups were compared using the Bioconductor "fgsea" package in R. After 10,000 cycles, significant enrichment pathways were obtained and sequenced. Gene sets with statistical significance were selected with a false discovery rate (FDR)-adjusted $P < 0.05$.

Statistical analysis

All statistical analyses were performed using R Statistical Software version 3.6.3. Comparison between groups was performed using a t-test. The Kaplan–Meier method was used for survival analysis, and "survival" package in R was used for survival curve analysis. Cox proportional hazards regression analysis was used for univariate and multivariate analyses of OS. The Wilcoxon test was used to compare differences in immune cell infiltration. $P < 0.05$ was considered statistically significant. Statistical differences were recorded as follows: $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

Results:

Construction of immune-related gene pairs prognostic model

A total of 56,735 genes and 2498 unique immune-related genes were obtained from the TCGA and ImmPort databases, respectively. Among them, 1653 immune-related genes were shared in the data obtained from both databases. Then, the immune-related genes from ImmPort and the genes obtained from the GEO dataset were intersected to find the same genes. Among 606 shared immune-related genes, 31,896 immune-related gene pairs were found after removing gene pairs with relatively small variations within the pair. The immune-related gene pairs from the TCGA dataset were combined with the corresponding clinical data, revealing 69 immune-related gene pairs significantly associated with prognosis. Then, the Lasso method for Cox proportional hazards regression analysis was used to construct the immune-related gene pairs prognostic model for the training group. Finally, 27 immune-related gene pairs using 43 immune-related genes were selected in the model (Table 1). The model was used to calculate the risk value of each patient in the training group. A time-dependent ROC curve was employed to distinguish between high- and low-risk patients using the optimal cut-off value (Fig. 1). The prognostic model significantly distinguished between high- and low-risk patients related to OS in the TCGA dataset ($P < 0.001$); the OS of high-risk patients was significantly lower than that of low-risk patients (Fig. 2a). Univariate and multivariate Cox proportional hazards regression analyses were used to analyze the corresponding clinical data in the TCGA dataset. The immune-related gene pairs prognostic model and AJCC stage were independent prognostic factors in the TCGA dataset (Figs. 3a, 3b).

Table 1
immune-related gene pairs prognostic model

IRG1	IRG2	Coefficient
CD74	CRABP2	-0.129381388033907
HSPA2	NEDD4	-0.431472955510512
CIITA	TLR7	-0.0861166676605254
CIITA	PLXNB3	-0.1318512759267
MICA	PLXNB1	0.397483055209339
RELB	CCR1	-0.126764197388246
RFXAP	IGLV6-57	0.0399268657548744
TAPBPL	IGF2R	-0.266631594480481
CXCL14	HMOX1	-0.0424869848483334
CCL8	CD3D	0.123645099256706
S100B	PLXNB3	-0.0745978715504017
APOBEC3G	PLXNB3	-0.349449105911909
TRIM5	IL27RA	0.187761415761893
TYK2	PTK2	-0.237158486425192
MSR1	IL18	0.2957793442697
PPARG	PLXNB3	-0.128237814963708
VAV1	ITGAL	0.318667102339933
RAC2	C3AR1	-0.319397161742718
IGHD	BTC	-0.00256266119415776
IGHD	SCG2	-0.019365921132164
IGHD	NPR3	-0.252893524646672
IGHD	ZAP70	-0.194457565172438
SEMA3B	SEMA6C	-0.192591720762394
SEMA3B	BTC	-0.18492459694902
ACVRL1	IL27RA	0.116702510407455
IGF1R	IGF2R	-0.280680737393679

Value of the immune-related gene pairs prognostic model in disease-free survival (DFS)

In order to verify the predictive value of the immune-related gene pairs prognostic model, we applied the model to the GEO dataset and stratified the patients into high- and low-risk groups. The DFS values of the two validation groups were similar to those of the training groups; the DFS of the high-risk group was significantly lower than that of the low-risk group (Fig. 2b, $P = 0.026$). Univariate and multivariate Cox proportional hazards regression analyses were used to analyze the corresponding clinical data in the GEO dataset. The immune-related gene pairs prognostic model was an independent prognostic factor in the GEO dataset (Fig. 3c).

Immune cell infiltration in different risk groups

CIBERSORT, which has been applied in many tumor microenvironments^{21,22}, was used to predict the infiltration of 21 different immune cell types in the high- and low-risk TCGA groups (Fig. 4a), including M0 and M2 macrophages, CD8 T cells, and resting dendritic cells. M0 (Fig. 4b, $P < 0.001$) and M2 (Fig. 4c, $P < 0.001$) macrophages were highly expressed in the high-risk group, while CD8 T cells (Fig. 4d, $p < 0.001$) and naive B cells (Fig. 4e, $p < 0.001$) were highly expressed in the low-risk group.

Functional evaluation of immune-related gene pairs

In order to determine the biological processes and signaling pathways related to the immune-related gene pairs in the prognostic model, we used GO enrichment to analyze the identified immune-related genes in the TCGA dataset and identified pathways with significant differences between high- and low-risk patients (Fig. 5). We found that CCR chemokine receptor binding, regulation of leukocyte-mediated cytotoxicity, T cell migration, T cell receptor complex, and other pathways were significantly enriched in low-risk patients (Fig. 6a-6f). The enrichment of these pathways in low-risk patients confirmed the importance of immune cells in the treatment and prognosis of breast cancer.

Discussion:

In the current study, we demonstrated that a prognostic model constructed with 27 immune-related gene pairs from 43 independent immune-related genes predicted OS and DFS in breast cancer. Many of the identified immune-related genes have been previously investigated in breast cancer research. The application of gene expression profiles to construct prognostic models has been studied; however, owing to the heterogeneity of organisms and differences in technology platforms, the accuracy of the analysis method requires verification²³. Relative sequencing and pairing of genes to preprocess an immune-related gene pairs prognostic model has provided reliable results for many tumors^{24,25}.

Breast cancer stem-like cells were rendered as sensitive to $\gamma\delta$ Tc cytotoxicity as non-stem-like cells by preventing increased *MICA* shedding via *ADAM* inhibitor GW280264X²⁶. This approach, using *MICA* shedding to enhance $\gamma\delta$ Tc targeting of cancer types, was considered an immune evasion mechanism²⁶. Genetic knockdown and pharmacological blockade of *Rac1/Rac2* upregulated claudin-1 and prohibited

the strong metastatic potential of TNBC characterized by a mesenchymal-like phenotype²⁷. *RNF144A* is known to be downregulated in a subset of primary breast tumors and suppresses breast cancer growth and metastasis through, at least in part, targeting *HSPA2* for ubiquitination and degradation in a ubiquitin ligase activity-dependent manner²⁸. *HSPA2* has been shown to be upregulated in various types of human cancers and promotes cancer cell growth, angiogenesis, migration, invasion, and metastasis through distinct mechanisms²⁸. *PPRP4* via *NEDD4* downregulates *ROBO1*, which is implicated as a tumor suppressor in various cancers, resulting in activation of *SRC* and *FAK*, thus promoting breast cancer metastasis²⁹. *PLXNB3* is reportedly overexpressed in breast cancer tissues³⁰, and *CD74* is associated with poor breast cancer prognosis³¹. *MIF* can activate *NF-κB* through *CD74* to control the dynamics and stability of mitochondria, thus promoting carcinogenesis by preventing apoptosis³². *CD74* expression in the membrane and cytoplasm of breast cancer cells was found to be higher than that in normal breast tissue. Downregulation of *CD74* was shown to decrease the invasion and migration of MDA-MB-231 breast cancer cells³³. Finally, in ER + breast cancer cells, interaction between *CRABP2* and *LATS1* promotes *LATS1* ubiquitination and inactivates the Hippo signaling pathway, thus promoting invasion and metastasis of ER + breast cancer³⁴.

Other identified immune-related genes in the current study have been investigated in relation to immunotherapy for other cancers. As an agonist of Toll-like receptors *TLR7* and *TLR8*, R848 is the driving factor of the M1 anti-tumor macrophage phenotype *in vitro*. R848-loaded nanoparticles effectively delivered drugs to tumor-associated macrophages *in vivo*³⁵. A *CCR1* antagonist has been shown to reduce T cell transportation to the omentum and liver in obesity-related cancer³⁶. *CXCL 14* inhibits human papillomavirus-associated head and neck cancer by restoring expression of *MHC-I* in tumor cells and promoting antigen-specific CD8 + T cell response³⁷. *IL 18* can inhibit CD4 + CD25 + Foxp3 + T cells through accumulation of pre-mNK cells and memory-type CD8 + T cells and enhanced the therapeutic effects of ICIs on the peritoneal dissemination of malignant tumors or melanoma metastasis via tail vein injection³⁸. The *CXCL 14* / *ACKR2* pathway of autocrine fibroblasts is a clinically relevant stimulator for epithelial-mesenchymal transition (EMT), tumor cell invasion, and metastasis. *ACKR2* is a newly identified mediator of *CXCL 14* function, which is a potential pathway related to drug targets³⁹. Taken together, these studies support that immune-related genes identified in the current study play important roles in the occurrence and development of tumors through immune cells. Immune-related genes have thus gradually become targets of immunotherapy or the means to enhance the effects of immunotherapy. However, some of the identified immune-related genes, such as *APOBEC3G*, *PLXNB1*, and *C3AR1*, have not been previously studied in breast cancer and warrant further exploration.

Additionally, the study findings revealed that M2 and M0 macrophages were highly expressed in high-risk breast cancer patients, while CD8 T cells and naïve B cells were highly expressed in low-risk breast cancer patients. M2 and M0 macrophages are associated with poor prognosis in colon cancer²². Some studies have demonstrated that M2 macrophages might impart outgrowth and M1 macrophages may contribute to dormancy behaviors in metastatic breast cancer cells. Research has also demonstrated that EMT and

secondary metastatic sites are regulated by selected macrophage phenotypes in the liver metastatic microenvironment. These results indicate that macrophages could be a potential therapeutic target for limiting death due to malignant metastases in patients with breast cancer⁴⁰. The presence of CD8 + T cells in breast cancer is associated with a significant reduction in the relative risk of death from disease in both ER- and ER + subtypes of HER2 + breast cancer. Thus, inclusion of TILs may improve risk stratification in patients with breast cancer classified into these subtypes⁴¹. Intratumor infiltration of B cells is significantly impaired during hepatocellular carcinoma progression. High densities of tumor-infiltrating B cells imply a better clinical outcome. Therapies designed to target B cells may be a novel strategy for hepatocellular carcinoma⁴². Therefore, different immune cells have different effects on breast cancer prognosis. Macrophages may promote the metastasis of breast cancer, while CD8 + T cells and B cells may inhibit the growth of breast cancer. Their individual effects and related molecular mechanisms need to be further investigated.

In this study, the identified immune-related genes were associated with multiple pathways related to immune cell infiltration and migration and immune checkpoint enhancement. Among them, regulation of the leukocyte-mediated cytotoxicity pathway is associated with tumor progression and decreased CD8 + infiltration in pancreatic cancer⁴³. The T cell migration pathway can enhance tumor immunity and increase the efficacy of ICI in preclinical breast cancer models⁴⁴. This pathway can enhance T cells into tumors and intratumoral T-cell diversity⁴⁵. A previous study identified the chemokine signal regulator *FROUNT* as a target to control tumor-associated macrophages by the chemokine receptor binding pathway⁴⁶. Anti-PD-L1 combined with Liposomal-AMD3100, a *CXCR4* antagonist, exerted an increased antitumor effect and prolonged survival in a murine TNBC model compared with monotherapies, measured by chemokine response⁴⁷. CD3 bispecific antibody promoted tumor cell killing by cross-linking the CD3 component of the T cell receptor complex pathway with a tumor-associated antigen on the surface of the target cell⁴⁸. These results indicate that immune-related pathways may play an important role in the treatment of breast cancer in the future. Consistent with these conclusions, our study determined that upregulation of these pathways was closely associated with low-risk patients. CCR chemokine receptor binding, immunoglobulin complex circulation, and regulation of cell killing pathways have not been studied in tumors but warrant further *in vivo* and *in vitro* studies.

It is necessary to gather information regarding the expression of multiple immune-related genes in the application of our prognostic model. The reliability of our model requires further verification in a large clinical patient population. Our two datasets originated from retrospective studies; thus, our prognostic model needs to be more widely validated in prospective cohort studies. In addition, the immune-related genes and related pathways associated with our prognostic model, and their molecular mechanisms, require further verification *in vitro* and *in vivo*.

Conclusion:

We established a prediction model based on immunogenomics that reliably predicted the prognosis of breast cancer patients using retrospective data. We also identified several key immune-related genes, including *MICA*, *RAC2*, *HSPA2*, *NEDD4*, *PLXNB3*, *APOBEC3G*, *PLXNB1*, and *C3AR1*, which may play important roles in the treatment of breast cancer in the future and provide new targets for immunotherapy. Further, we determined that M2 and M0 macrophages were highly expressed in high-risk breast cancer patients, while CD8 T cells and naïve B cells were highly expressed in low-risk breast cancer patients. Specific immune cell infiltration combined with ICI is the future direction of breast cancer treatment research.

Abbreviations:

TCGA: The Cancer Genome Atlas; GEO: Gene Expression Omnibus; GO: Gene Ontology; OS: overall survival; HR: hormone receptor; HER2: human epidermal growth factor receptor 2; TNBC: triple-negative breast cancer; TILs: tumor-infiltrating lymphocytes; EFS: event-free survival; PD-L: programmed death ligand; ICIs: immune checkpoint inhibitors; PCR: pathological complete response rate; ICB: immune-checkpoint blockade; DFS: disease-free survival; AIRR-Seq: adaptive immune receptor repertoire sequencing; Lasso: least absolute shrinkage and selection operator; ROC: receiver operating characteristic; AJCC: American Joint Committee on Cancer; GSEA Gene Set Enrichment Analysis; FDR: false discovery rate; EMT: epithelial-mesenchymal transition;

Declarations:

AUTHOR CONTRIBUTIONS

Dr. Ying Zhong and Dr. Zhe Wang wrote the main manuscript text; Dr. Yidong Zhou, Dr. Feng Mao, Dr. Yan Lin and Dr. Jinghong Guan collected patients' data; Dr. Songjie Shen, Dr Yali Xu, Dr. Xin Huang, and Dr.Xuefei Wang conducted statistical analysis; Dr. Qiang Sun supported quality management and directed team.

ACKNOWLEDGEMENTS

The authors would like to thank the TCGA, GEO, and GO databases for the availability of the data.

CONFLICTS OF INTEREST

The authors have no conflicts of interest.

FUNDING

This study was Supported by the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (2019XK320033).

Ethics approval and consent to participate

Not applicable

Availability of supporting data

The data of TCGA, GEO, GO and ImmPort databases are available on line.

Consent for publication

All the authors are consent for publication.

References:

1. Lord, S. J. *et al.* Metastatic breast cancer incidence, site and survival in Australia, 2001-2016: a population-based health record linkage study protocol. *BMJ Open* **9**, e026414, doi:10.1136/bmjopen-2018-026414 (2019).
2. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA Cancer J Clin* **70**, 7-30, doi:10.3322/caac.21590 (2020).
3. Gonzalez-Angulo, A. M., Morales-Vasquez, F. & Hortobagyi, G. N. Overview of resistance to systemic therapy in patients with breast cancer. *Adv Exp Med Biol* **608**, 1-22, doi:10.1007/978-0-387-74039-3_1 (2007).
4. Cardoso, F. *et al.* Global analysis of advanced/metastatic breast cancer: Decade report (2005-2015). *Breast* **39**, 131-138, doi:10.1016/j.breast.2018.03.002 (2018).
5. Sledge, G. W., Jr. & Frenzel, M. Analysis of Overall Survival Benefit of Abemaciclib Plus Fulvestrant in Hormone Receptor-Positive, ERBB2-Negative Breast Cancer-Reply. *JAMA Oncol* **6**, 1122-1123, doi:10.1001/jamaoncol.2020.1518 (2020).
6. Howell, A. *et al.* Fulvestrant, formerly ICI 182,780, is as effective as anastrozole in postmenopausal women with advanced breast cancer progressing after prior endocrine treatment. *J Clin Oncol* **20**, 3396-3403, doi:10.1200/JCO.2002.10.057 (2002).
7. Budczies, J. *et al.* Classical pathology and mutational load of breast cancer - integration of two worlds. *J Pathol Clin Res* **1**, 225-238, doi:10.1002/cjp2.25 (2015).
8. Luen, S., Virassamy, B., Savas, P., Salgado, R. & Loi, S. The genomic landscape of breast cancer and its interaction with host immunity. *Breast* **29**, 241-250, doi:10.1016/j.breast.2016.07.015 (2016).
9. 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, doi:10.1200/JCO.2013.55.0491 (2014).
10. Dieci, M. V. *et al.* Prognostic and predictive value of tumor-infiltrating lymphocytes in two phase III randomized adjuvant breast cancer trials. *Ann Oncol* **26**, 1698-1704, doi:10.1093/annonc/mdv239 (2015).
11. Ignatiadis, M. *et al.* Tumor-Infiltrating Lymphocytes in Patients Receiving Trastuzumab/Pertuzumab-Based Chemotherapy: A TRYPHAENA Substudy. *J Natl Cancer Inst* **111**, 69-77,

- doi:10.1093/jnci/djy076 (2019).
12. Nagarajan, D. & McArdle, S. E. B. Immune Landscape of Breast Cancers. *Biomedicines* **6**, doi:10.3390/biomedicines6010020 (2018).
 13. Schmid, P. *et al.* Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer. *N Engl J Med* **379**, 2108-2121, doi:10.1056/NEJMoa1809615 (2018).
 14. Nathan, M. R. & Schmid, P. The emerging world of breast cancer immunotherapy. *Breast* **37**, 200-206, doi:10.1016/j.breast.2017.05.013 (2018).
 15. Schmid, P. *et al.* Atezolizumab plus nab-paclitaxel as first-line treatment for unresectable, locally advanced or metastatic triple-negative breast cancer (IMpassion130): updated efficacy results from a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol* **21**, 44-59, doi:10.1016/S1470-2045(19)30689-8 (2020).
 16. Mittendorf, E. A. *et al.* Neoadjuvant atezolizumab in combination with sequential nab-paclitaxel and anthracycline-based chemotherapy versus placebo and chemotherapy in patients with early-stage triple-negative breast cancer (IMpassion031): a randomised, double-blind, phase 3 trial. *Lancet* **396**, 1090-1100, doi:10.1016/S0140-6736(20)31953-X (2020).
 17. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat Med* **21**, 1350-1356, doi:10.1038/nm.3967 (2015).
 18. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **12**, 453-457, doi:10.1038/nmeth.3337 (2015).
 19. Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* **47**, W191-W198, doi:10.1093/nar/gkz369 (2019).
 20. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).
 21. Cole, C. *et al.* Inhibition of FGFR2 and FGFR1 increases cisplatin sensitivity in ovarian cancer. *Cancer Biol Ther* **10**, 495-504, doi:10.4161/cbt.10.5.12585 (2010).
 22. Edin, S. *et al.* The distribution of macrophages with a M1 or M2 phenotype in relation to prognosis and the molecular characteristics of colorectal cancer. *PLoS One* **7**, e47045, doi:10.1371/journal.pone.0047045 (2012).
 23. Eddy, J. A., Sung, J., Geman, D. & Price, N. D. Relative expression analysis for molecular cancer diagnosis and prognosis. *Technol Cancer Res Treat* **9**, 149-159, doi:10.1177/153303461000900204 (2010).
 24. Zhou, X. *et al.* Development and Validation of an Individualized Immune-Related Gene Pairs Prognostic Signature in Papillary Renal Cell Carcinoma. *Front Genet* **11**, 569884, doi:10.3389/fgene.2020.569884 (2020).
 25. Zhang, L. *et al.* An immune-related gene pairs signature predicts overall survival in serous ovarian carcinoma. *Onco Targets Ther* **12**, 7005-7014, doi:10.2147/OTT.S200191 (2019).

26. Dutta, I. *et al.* ADAM protease inhibition overcomes resistance of breast cancer stem-like cells to gammadelta T cell immunotherapy. *Cancer Lett* **496**, 156-168, doi:10.1016/j.canlet.2020.10.013 (2021).
27. Chiang, S. K., Chang, W. C., Chen, S. E. & Chang, L. C. DOCK1 Regulates Growth and Motility through the RRP1B-Claudin-1 Pathway in Claudin-Low Breast Cancer Cells. *Cancers (Basel)* **11**, doi:10.3390/cancers11111762 (2019).
28. Yang, Y. L. *et al.* RNF144A functions as a tumor suppressor in breast cancer through ubiquitin ligase activity-dependent regulation of stability and oncogenic functions of HSPA2. *Cell Death Differ* **27**, 1105-1118, doi:10.1038/s41418-019-0400-z (2020).
29. Zhang, L. *et al.* PRRG4 promotes breast cancer metastasis through the recruitment of NEDD4 and downregulation of Robo1. *Oncogene* **39**, 7196-7208, doi:10.1038/s41388-020-01494-7 (2020).
30. Valladares, A. *et al.* Genetic expression profiles and chromosomal alterations in sporadic breast cancer in Mexican women. *Cancer Genet Cytogenet* **170**, 147-151, doi:10.1016/j.cancergencyto.2006.06.002 (2006).
31. Wang, Z. Q., Milne, K., Webb, J. R. & Watson, P. H. CD74 and intratumoral immune response in breast cancer. *Oncotarget* **8**, 12664-12674, doi:10.18632/oncotarget.8610 (2017).
32. De, R. *et al.* Macrophage migration inhibitory factor regulates mitochondrial dynamics and cell growth of human cancer cell lines through CD74-NF-kappaB signaling. *J Biol Chem* **293**, 19740-19760, doi:10.1074/jbc.RA118.003935 (2018).
33. Liu, Z. *et al.* CD74 interacts with CD44 and enhances tumorigenesis and metastasis via RHOA-mediated cofilin phosphorylation in human breast cancer cells. *Oncotarget* **7**, 68303-68313, doi:10.18632/oncotarget.11945 (2016).
34. Feng, X. *et al.* CRABP2 regulates invasion and metastasis of breast cancer through hippo pathway dependent on ER status. *J Exp Clin Cancer Res* **38**, 361, doi:10.1186/s13046-019-1345-2 (2019).
35. Rodell, C. B. *et al.* TLR7/8-agonist-loaded nanoparticles promote the polarization of tumour-associated macrophages to enhance cancer immunotherapy. *Nat Biomed Eng* **2**, 578-588 (2018).
36. Conroy, M. J. *et al.* CCR1 antagonism attenuates T cell trafficking to omentum and liver in obesity-associated cancer. *Immunol Cell Biol* **94**, 531-537, doi:10.1038/icb.2016.26 (2016).
37. Westrich, J. A. *et al.* CXCL14 suppresses human papillomavirus-associated head and neck cancer through antigen-specific CD8(+) T-cell responses by upregulating MHC-I expression. *Oncogene* **38**, 7166-7180, doi:10.1038/s41388-019-0911-6 (2019).
38. Ma, Z. *et al.* Augmentation of Immune Checkpoint Cancer Immunotherapy with IL18. *Clin Cancer Res* **22**, 2969-2980, doi:10.1158/1078-0432.CCR-15-1655 (2016).
39. Sjoberg, E. *et al.* A Novel ACKR2-Dependent Role of Fibroblast-Derived CXCL14 in Epithelial-to-Mesenchymal Transition and Metastasis of Breast Cancer. *Clin Cancer Res* **25**, 3702-3717, doi:10.1158/1078-0432.CCR-18-1294 (2019).
40. Yang, M., Ma, B., Shao, H., Clark, A. M. & Wells, A. Macrophage phenotypic subtypes diametrically regulate epithelial-mesenchymal plasticity in breast cancer cells. *BMC Cancer* **16**, 419,

doi:10.1186/s12885-016-2411-1 (2016).

41. Ali, H. R. *et al.* Association between CD8+ T-cell infiltration and breast cancer survival in 12,439 patients. *Ann Oncol* **25**, 1536-1543, doi:10.1093/annonc/mdu191 (2014).
42. Zhang, Z. *et al.* Landscape of infiltrating B cells and their clinical significance in human hepatocellular carcinoma. *Oncoimmunology* **8**, e1571388, doi:10.1080/2162402X.2019.1571388 (2019).
43. Zhuang, H., Zhang, C. & Hou, B. GTF2IRD1 overexpression promotes tumor progression and correlates with less CD8+ T cells infiltration in pancreatic cancer. *Biosci Rep* **40**, doi:10.1042/BSR20202150 (2020).
44. Hollande, C. *et al.* Inhibition of the dipeptidyl peptidase DPP4 (CD26) reveals IL-33-dependent eosinophil-mediated control of tumor growth. *Nat Immunol* **20**, 257-264, doi:10.1038/s41590-019-0321-5 (2019).
45. Juric, V. *et al.* MMP-9 inhibition promotes anti-tumor immunity through disruption of biochemical and physical barriers to T-cell trafficking to tumors. *PLoS One* **13**, e0207255, doi:10.1371/journal.pone.0207255 (2018).
46. Terashima, Y. *et al.* Targeting FROUNT with disulfiram suppresses macrophage accumulation and its tumor-promoting properties. *Nat Commun* **11**, 609, doi:10.1038/s41467-020-14338-5 (2020).
47. Lu, G., Qiu, Y. & Su, X. Targeting CXCL12-CXCR4 Signaling Enhances Immune Checkpoint Blockade Therapy Against Triple Negative Breast Cancer. *Eur J Pharm Sci*, 105606, doi:10.1016/j.ejps.2020.105606 (2020).
48. Clynes, R. A. & Desjarlais, J. R. Redirected T Cell Cytotoxicity in Cancer Therapy. *Annu Rev Med* **70**, 437-450, doi:10.1146/annurev-med-062617-035821 (2019).

Figures

Figure 1

ROC curve validation of immune-related gene pairs model

Figure 2

a Overall survival between high- and low-risk patients in the TCGA datasets. b Disease free survival between high- and low-risk patients in the GEO datasets.

Figure 3

a Univariate Cox proportional hazards regression analysis of breast cancer patients in TCGA datasets. b Multivariate Cox proportional hazards regression analysis of breast cancer patients in TCGA datasets. c Univariate Cox proportional hazards regression analysis of breast cancer patients in GEO datasets. d Multivariate Cox proportional hazards regression analysis of breast cancer patients in GEO datasets.

Figure 4

a Infiltration of 21 different immune cell types in the high- and low-risk TCGA groups. b M0 macrophages infiltration in the high- and low-risk TCGA groups. c M2 macrophages infiltration in the high- and low-risk TCGA groups. d CD8 T cells infiltration in the high- and low-risk TCGA groups. e Naive B cells infiltration in the high- and low-risk TCGA groups.

Figure 5

Identified pathways with significant differences between high- and low-risk TCGA patients.

Figure 6

a CCR chemokine receptor binding pathway was significantly enriched in low-risk patients b Regulation of leukocyte-mediated cytotoxicity pathway was significantly enriched in low-risk patients. c T cell migration pathway was significantly enriched in low-risk patients. d T cell receptor complex pathway was significantly enriched in low-risk patients. e Cytokine receptor binding pathway was significantly enriched in low-risk patients. f Response to chemokine pathway was significantly enriched in low-risk patients.