A framework to predict the molecular classification and prognosis of breast cancer patients and characterize the landscape of immune cell infiltration

Kun Zheng[1†], Zhiyong Luo[2†], Yilu Zhou[3,4†], Lili Zhang[1], Yali Wang[1], Xiuqiong Chen[1], Shuo Yao[1], Huihua Xiong[1], Xianglin Yuan[1], Yanmei Zou[1], Yihua Wang[3,4*] and Hua Xiong[1*]

[1] Department of Oncology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China.
[2] Department of Thyroid and Breast Surgery, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China.
[3] Biological Sciences, Faculty of Environmental and Life Sciences, University of Southampton, Southampton SO17 1BJ, UK.
[4] Institute for Life Sciences, University of Southampton, Southampton SO17 1BJ, UK.

Correspondence should be addressed to Yihua Wang yihua.wang@soton.ac.uk and Hua Xiong cnhxiong@tjh.tjmu.edu.cn.

[†] These authors have contributed equally to this work and share first authorship.

Abstract

It is known that all current cancer therapies can only benefit a limited proportion of patients, thus molecular classification and prognosis evaluation are critical for correctly classifying breast cancer patients and selecting the best treatment strategy. These processes usually involve the disclosure of molecular information like mutation, expression, and immune microenvironment of a breast cancer patient, which are not been fully studied untill now. Therefore, there is an urgent clinical need to identify potential markers to enhance molecular classification, precision prognosis and therapy stratification for breast cancer patients. In this study, we explored the gene expression profiles of 1,721 breast cancer patients through CIBERSORT and ESTIMATE algorithms, then we obtained a comprehensive intra-tumoral immune landscape. The immune cell infiltration (ICI) patterns of breast cancer were classified into 3 separate subtypes according to the infiltration levels of 22 immune cells. The differentially expressed genes between these subtypes were further identified and ICI scores were calculated to assess the immune landscape of BRCA patients. Importantly, we demonstrated that ICI scores correlate with patients' survival, tumor mutation burden, neoantigens, and sensitivity to specific drugs. Based on these ICI scores, we were able to predict the prognosis of patients and their response to immunotherapy. Together, these findings provide a realistic scenario to stratify breast cancer patients for precision medicine.

Introduction

Breast cancer (BRCA) has now risen to become the most common malignant tumor throughout the world and the second leading cause of cancer-related death in women. The US added 270,000 new diagnosed cases and more than 40,000 deaths in 2020 [1, 2]. Due to its considerable influence on public health worldwide, the molecular mechanisms of breast cancer-like associated genes and pathways [3], metastasis [4], and drug responses [5, 6] have

been widely studied. Over recent years, there have been great advances in treatment strategies for BRCA including surgical resection, chemotherapy, radiotherapy, targeted therapy, and endocrine therapy. However, due to factors such as local recurrence, distant metastasis, and high tumor heterogeneity, the prognosis of BRCA patients is still unsatisfactory [7, 8].

The tumor microenvironment (TME) includes tumor cells, tumor-infiltrating lymphocytes (TILs), and stromal components, which can serve as a key mediator of cancer progression and treatment outcome [9, 10]. Over the past few years, numerous studies have shown that TILs play key roles in tumor extension, recurrence, metastasis, and therapeutic response to cancer immunotherapy [11, 13]. For example, naive $CD8^+$ T cells, when bound and activated by antigen-presenting dendritic cells, would become effector T cells, which could then recognize and kill tumor cells by releasing granzymes to induce apoptosis [14]. Chemokines secreted by tumor cells, such as C-C Motif Chemokine Ligand 2, C-C Motif Chemokine Ligand 5, and colony stimulating factor 1, can recruit M2-type tumor-associated macrophages, and their abundance in TME correlates with a poor prognosis [15].

Cancer immunotherapy, including immune checkpoint inhibitors, has provided clinical benefit to the treatment of many BRCA patients through direct or indirect effects on TILs, reversing the TMEs to immune-permitted environments from immunosuppressive ones [16]. Promising outcomes in response to antibodies to programmed cell death 1(PD-1) or antibodies to programmed cell death ligand 1(PD-L1) therapy for BRCA have been reported in recent years [17]-[19]. However, the immune microenvironment of BRCA remains poorly understood and this treatment can only benefit a limited proportion of patients [20],[21]. Therefore, identification of potential biomarkers is in urgent clinical need to enhance precision prognosis and therapy stratification for BRCA patients.

In our study, the gene expression profiles of 1,721 BRCA patients were analyzed by CIBERSORT and ESTIMATE algorithm, by which we obtained a comprehensive intra-tumoral immune landscape. The immune cell infiltration (ICI) patterns of BRCA were classified into 3 separate subtypes according to the infiltration levels of 22 immune cells. The differentially expressed genes (DEGs) between these subtypes were further identified and ICI scores were calculated to assess the immune landscape of BRCA patients. Importantly, we demonstrated that ICI scores correlate with patients' survival, tumor mutation burden (TMB), neoantigens, and sensitivity to specific drugs. Based on these ICI scores, we were able to predict the prognosis of patients and their response to immunotherapy. Together, these findings provide a realistic scenario to stratify BRCA patients for precision medicine.

Materials and Methods

Source of cohort datasets and immune-related data and preprocessing

The training datasets of BRCA for this study were integrated from two separate cohorts (TCGA-BRCA and Yau-cohort), with only tumor samples retained. The expression profile data of TCGA-BRCA cohort (considering only protein-coding mRNA) were downloaded from the Cancer Genome Atlas (TCGA) database by Genomic Data tools (https://portal.gdc.cancer.gov/projects/TCGA-BRCA). The fragments per kilobase million values were downloaded via TCGAbiolinks [22] package and transformed to transcripts per million, with the ensemble id matrix converted to a gene symbol matrix and other forms for subsequent analysis. The Yau-cohort dataset [23], integrated by Dr. Yau from four studies (GSE2034, GSE5327, GSE7390, and NKI295), was downloaded from the online database

University of California Santa Cruz (UCSC) Xena browser (https://xenabrowser.net/). It contains the gene expression matrix along with clinical information of 682 breast cancer patients. At last, the batch effects caused by nonbiological technical bias were reduced through "Combat" algorithm [24].

The clinical information for the TCGA-BRCA cohort was extracted from the pan-cancer data, which included age, sex, clinicopathological stage, TNM stage, and PAM50 subtype, and only overall survival (OS) was considered. The Yau-cohort cohort considered OS, age, and PAM50 subtype. A total of 1721 breast cancer samples were generated after kicking out the samples with incomplete clinical information and survival time and male breast cancer samples. To analyze the efficiency of immunotherapy, the R package IMvigor210CoreBiologies [25] obtained from the work of Snyder *et al.* was used as a validation dataset, which included expression profiles, survival outcomes and immunotherapy response results in metastatic uroepithelial cancer patients treated with anti-PD-L1 agents atezolizumab.

Consensus clustering of TME immune cell infiltration

The CIBERSORT and ESTIMATE algorithms were combined to reckon the abundance and infiltration levels of 22 immune cell species of the integrated BRCA cohort [26, 27]. LM22 signature matrix, which provided a gene expression signature set of 22 immune cell subtypes and CIBERSORT source code were downloaded from the CIBERSORT website (https://cibersortx.stanford.edu/). Unsupervised clustering analysis of ICI of each sample was performed using R package "ConsensusClusterPlus" [28], which classified the tumor ICI pattern of BRCA patients into different subtypes (maxK, the maximum number of classifications K = 3). 90% of the samples have been repeated 500 times, ensuring stability of the classification. Calculation of distances using Spearman's distance measure and Ward's linkage.

Identification of DEGs between ICI subtypes and gene signature generation

To identify genes associated with ICI patterns, we applied the R package “limma” [29] to determine the DEGs between different ICI subtypes and plotted the DEGs heatmap using the “ComplexHeatmap” R package [30]. The significance cutoff criteria used to distinguish DEGs were set as fold-change (FC) > 1.5, adjusted false discovery rate (FDR) < 0.05.

To quantify the ICI pattern of a single tumor patient, we established a scoring system, ICI gene signature, to confirm the ICI pattern for each BRCA patient, and we termed it ICIscore. The steps to establish an ICI gene signature are as follows: Firstly, the DEGs were analyzed by unsupervised cluster analysis using the R package "ConsensusClusterPlus". The maximum number of classifications is 3 and the distances were calculated using the Pearson distance measure and complete linkage, which divided the TCGA-BRCA cohort into 3 genomic clusters, namely, ICI gene clusters A, B, and C. And then, Pearson correlation analysis was done on the mRNA expression values of all TCGA samples with the three gene clusters, and the DEGs with positive and negative correlation with clustering features were, respectively, defined as ICI signature genes A and B. Then Boruta algorithm was used to reduce the dimensionality of different ICI signature genes. Finally, two total scores were calculated using single sample gene set enrichment analysis (ssGSEA): (1) ICI score A which is from ICI signature gene A; (2) ICI score B which is from ICI signature gene B. $\boldsymbol{ICI\ score = ICI\ score\ A - ICI\ score B}$, with median as cutoff value to determine high ICI group and

low ICI group. When survival analyses were performed with ICI score groups, we only picked out genes whose P-value < 0.05 in the univariate survival analysis. Principal component analysis (PCA) was used to calculate the ICI score for each patient, and PCA1 was calculated as the signature score using PCA: $\boldsymbol{ICI\ score} = |\boldsymbol{PCA1\ positive}| + |\boldsymbol{PCA1\ negative}|$. Patients were reclassified as high and low ICI score groups using median as the cutoff value.

Collection and analysis of somatic mutation data

The copy number variant (CNV) data of TCGA-BRCA cohort were obtained from the firehose database (http://gdac.broadinstitute.org/), and mutant annotation format (MAF) files downloaded from the cBioportal database (http://www.cbioportal.org/). To determine the TMB of BRCA, we matched TCGA-BRCA MAF files with ICI-related expression profiles and used the R package "maftools" to calculate the TMB [31]. Based on the OncodriveCLUST algorithm [32], we used the positional information of the somatic mutation sites to cluster the driver genes from different ICI score groups, and used the "maftools" package to draw a waterfall map of the top 25 driver genes in the two groups. The CNV analysis was performed with the GenePattern online analysis tool (https://www.genepattern.org/) and visualized with the “maftools” package.

Identification of sensitive drugs and other biological processes correlated with ICI gene signatures

The drug.txt is a dataset for the sensitivity and response of cancer cells to therapeutic drugs obtained from the online database Genomics of Drug Sensitivity in Cancer (GDSC), used to predict IC50 with R package “pRRophetic” [33, 34]. We assessed the IC50 values in both ICI score groups using Wilcoxon test, then compared the differences in sensitivity between ICI score groups on more than 100 drugs, and graphed the top 12 (according to P-value) differentially response drugs. Wilcoxon test was also used to compare the differential expression of neoantigen between ICI score groups. Sample data used for predicting neoantigen number in the TCGA-BRCA cohort were from a research already published in 2015 by Rooney MS *et al.* [35].

Additionally, we separately performed Gene Ontology (GO) enrichment analysis of ICI gene signatures A and B via the “org.Hs.eg.db” R package to explore the biological process, cellular composition and molecular function that they may participate in. After differential expression analysis with the “limma” package for high or low ICI score groups, the differentially expressed genes were subjected to Gene Set Enrichment Analysis (GESA), and the gene sets “h.all.v7.2.symbol” were downloaded from the Molecular Signatures Database (MSigDB) (http://www.gsea-msigdb.org/gsea/msigdb) for running GESA analysis. To confirm the difference in the efficacy of anti-PD-L1 immunotherapy between the two ICI score groups in the validation cohort, IMvigor210, objective remission rate bar graphs were plotted for the ICI score groups using "GSVA" R package [36].

Statistical analysis

All statistical analyses were performed using R software (version 3.6.2). The Wilcoxon test was used to compare the differences between two groups and the Kruskal-Wallis test was used to compare the differences between more than two groups. The Kaplan-Meier survival curves were plotted using the R package “survminer” for different subgroups, such as ICI

clusters, ICI gene clusters, ICI gene signatures, and TMB subgroups, in relation to survival. Log-rank test was used for statistically significant differences. The R packages "ComplexHeatmap" and "ggplot" were used to draw heatmaps, scatter plots, violin plots, and other plots. Correlation coefficients were calculated by using Spearman analysis. Two-tailed $P < 0.05$ was considered a statistically significant difference.

Results

The immune cell infiltration (ICI) landscape in BRCA immune microenvironment

We first performed PCA of integrated gene expression profiles of 1,721 BRCA patients from the training cohort consisting of the TCGA-BRCA and Yau cohorts by using Combat algorism to eliminate batch effects across cohorts (Figure 1a). Subsequently, we performed the CIBERSORT algorithm combined with ESTIMATE algorithm to determine the abundances of 22 immune cells as well as the enrichment scores of stromal cells (Stromal Score) and immune cells (Immune Score) in BRCA patients in this cohort (Supplementary Table 1). We performed an unsupervised cluster analysis of this cohort by ConsensusClusterPlus R package to divide BRCA patients into 3 separate subtypes based on ICI patterns, referred to as ICI clusters I, II and III, respectively (Figure 1b). A hotspot matrix of correlation coefficients was created to demonstrate the overall landscape of interactions among immune cells in the TME of BRCA patients, including their immune scores and stromal scores (Figure 1c).

To explore the inherent biological differences between the different ICI subtypes, we compared the composition of immune cells in the 3 ICI clusters. As shown in Figure 1d, ICI cluster I was characterized by high level M2 macrophages, neutrophils, resting mast cells, activated natural killer (NK) cells, resting $CD4^+$ T cells and gamma delta T cells infiltration; patients from ICI cluster II had a higher density of memory B cells, activated dendritic cells, resting dendritic cells, M1 macrophages, monocytes, memory activated $CD4^+$ T cells, $CD8^+$ T cell, , follicular helper T cells, plasma cells and regulatory T cells; while ICI cluster III displayed an increase in naïve B cells, naïve $CD4^+$ T cells, resting NK cells, M0 macrophages and activated mast cells infiltration. Survival analysis conducted on these 3 ICI subtypes showed significant differences among them, with ICI clusters I and II being associated with better prognosis and patients in ICI cluster III having a poorer OS (log-rank test, $P = 0.007$; Figure 1e). In addition, we analyzed the expression of PD-1 and PD-L1 in each ICI subtype (Figures 1f and 1g). The results of Kruskal-Wallis test showed higher expression of PD-1 and PD-L1 in ICI cluster II, while their expressions were lowest in ICI cluster III.

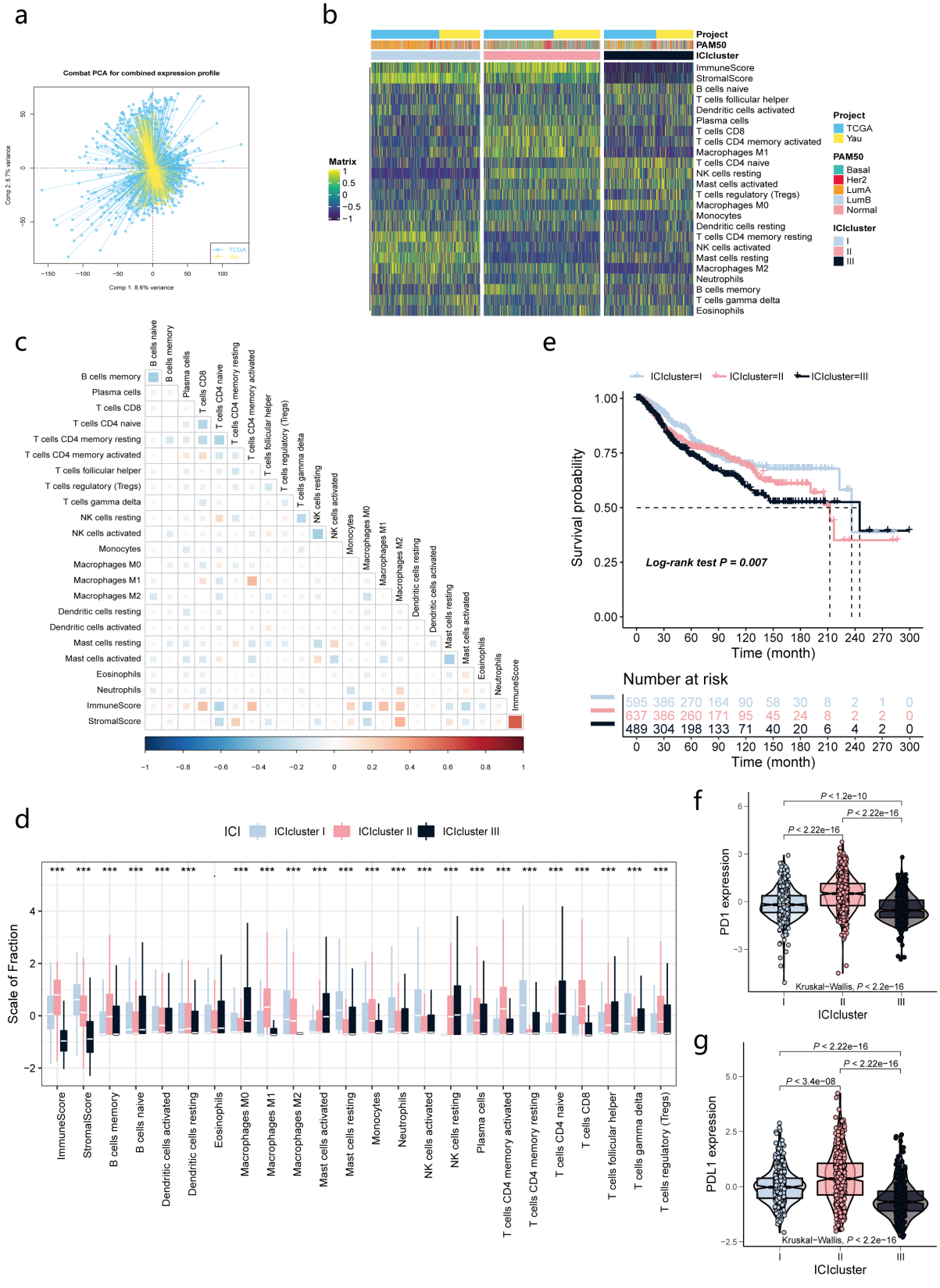

Figure 1: The immune-cell infiltration (ICI) landscape in BRCA immune microenvironment. (a) PCA of integration of expression profiles of TCGA-BRCA and Yau cohorts by Combat algorism to eliminate batch effects of different cohorts. (b) Heatmap with unsupervised clustering analysis of tumor-infiltrating immune cells in TCGA-BRCA and Yau cohorts. (c) Hotspot plot for correlation matrix of immune cells in three ICI clusters, including their immune scores and stromal scores. Red indicates positive correlation and blue indicates

negative correlation. (d) Box plot for abundance of each immune infiltrating cells in the three ICI clusters. The asterisks represented the statistical P-value (Kruskal-Wallis test, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$). (e) Survival analysis for three ICI clusters of 1721 breast cancer patients from TCGA-BRCA and Yau cohorts using Kaplan-Meier curves. The log-rank test showed that $P = 0.007$. (f and g) Violin plots of the differential expression of PD1 (f) and PD-L1 (g) (only for TCGA-BRCA cohort) among the three ICI clusters. The statistical differences among ICI clusters were compared by Kruskal-Wallis test ($P < 0.001$).

Identification and comprehensive analysis of immunogenic gene clusters

To elucidate the potential characteristics of the different immunophenotypes, we conducted the limma package to identify DEGs among ICI clusters I, II, and III (FC = 1.5, FDR = 0.05). Based on the above cutoffs, we identified 665 DEGs (213 in ICI cluster I; 239 in ICI cluster II; and 213 in ICI cluster III; Supplementary Table 2) and used the ComplexHeatmap package to generate a heat map of all DEGs. Hereafter, we focused our analysis on the TCGA-BACR cohort as it had comprehensive information on clinical aspects. We performed an unsupervised clustering analysis of these DEGs and divided the TCGA-BRCA cohort into 3 distinct ICI genomic phenotypes, named ICI gene cluster A, B and C, respectively (Figure 2a). We defined all above DEGs with positive association with these 3 ICI gene clusters as ICI signature genes A, while the rest DEGs were termed as ICI signature genes B. By down-dimensioning the ICI signature genes using Boruta algorithm to reduce redundant genes, we finally obtained 216 genes in ICI signature gene A and 164 in ICI signature gene B (Supplementary Table 3).

In Figure 2b, we figured out the prognostic differences among these ICI gene clusters, and we confirmed that ICI gene clusters A and B had a better prognosis, and the prognosis of ICI gene cluster C was poorer (log-rank test, $P = 0.04$). Figures 2c and 2d showed the results of gene ontology (GO) enrichment analysis of both ICI signature gene groups in the 3 functional groups, biological process, cellular component, and molecular function, respectively, which were significantly enriched in items related to immunity. Given that the immune system can exert both antitumor and protumor activities [37,38], we next explored the level of immune infiltration cells among different gene clusters, and the box plot showed that gene clusters A and B with favorable prognosis had higher immune and stromal scores (Figure 2e). Besides, there were the highest infiltrations of M1 macrophages, $CD8^+$ T cells, memory activated $CD4^+$ T cells, memory B cells, activated dendritic cells, and plasma cells *etc.* within ICI gene cluster B, showing the active immune phenotype. In contrast, the level of infiltration of these TILs was very low in the poorly prognosed ICI gene cluster C. The three ICI gene clusters also showed significant differences in the expression levels of PD-1 and PD-L1. There were relatively high expression levels of PD-1 and PD-L1 in ICI gene clusters A and B, while they had the lowest expression levels in ICI gene cluster C (Figures 2f and 2g). From the above comprehensive analysis of immunogenic gene clusters, we demonstrated that there is a significant correlation between the level of ICI and prognosis in different gene clusters.

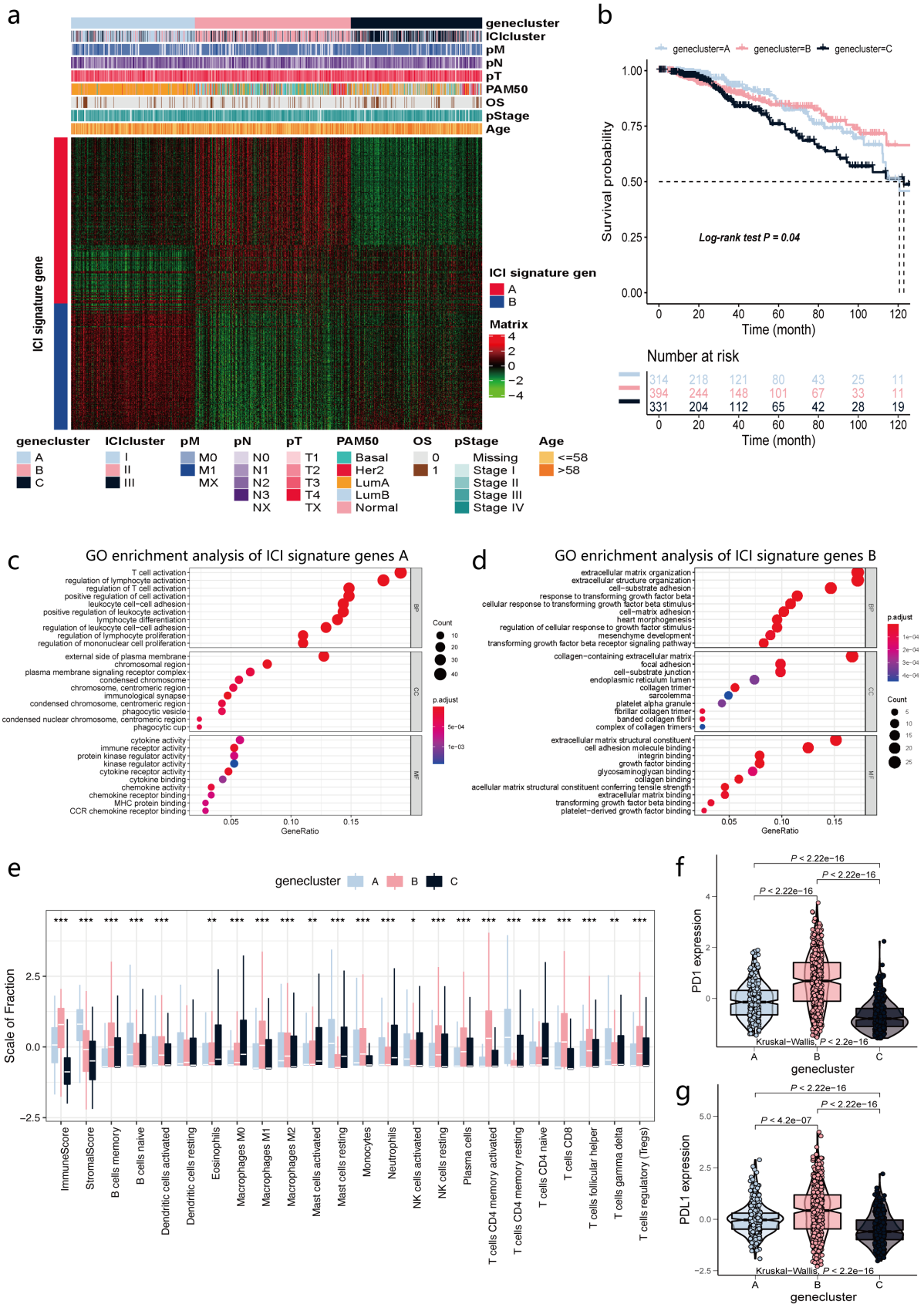

Figure 2. Identification and comprehensive analysis of immunogenic gene clusters. (a) Heat map with unsupervised clustering analysis of all DEGs in the three ICI patterns, dividing TCGA-BRCA patients into three genomic clusters, defined as ICI gene clusters A-C. Rows represent genes and columns represent samples. (b) Survival analysis for the three ICI gene clusters in TCGA-BRCA patients using Kaplan-Meier curves. The log-

rank test showed that P = 0.04. (c and d) Functional annotation of ICI gene clusters A (c) and B (d) using GO enrichment analysis. The circle size of the bubble plots represented the number of enriched genes. (e) Box plot for abundance of each immune infiltrating cells in the three ICI gene clusters. The asterisks represented the statistical P-value (Kruskal-Wallis test, *P < 0.05, **P < 0.01, ***P < 0.001). (f and g) Violin plots of the differential expression of PD1 (f) and PD-L1 (g) among the three ICI gene clusters. The statistical differences among ICI gene clusters were compared by Kruskal-Wallis test (P < 0.001).

Immune-cell infiltration (ICI) score construction

Given the individual heterogeneity of the TME, we quantified the ICI pattern of BRCA patients. We calculated 2 summary scores, that is, ICI score A from ICI signature gene A and ICI score B from ICI signature gene B, using ssGSEA. The ICI score of each patient of TCGA-BRCA cohort was determined using the difference between ICI scores A and B. The high ICI score group and low ICI score group were defined using median as the cutoff value. The distribution of ICI scores and survival of patients in ICI gene clusters were shown in Figure 3a and Supplementary Table 4.

We further analyzed the differences in the expression of immunoreactive-related genes in the high or low ICI score groups to determine the status of immune activity or tolerance in each group. Among them, *CD274*, *HAVCR2*, *CTLA4*, *LAG3*, *PDCD1*, and *IDO1* were chosen as immune inhibitory genes [39], while *CD8A*, *GZMA*, *PRF1*, *CXCL10*, *CXCL9*, *TNF* and *TBX2* as immune stimulatory genes [40]. As we can observe in Figure 3b, the expression levels of all immunoreactive-related genes were significantly elevated in the high ICI score group. We performed the differential expression analysis of genes in the high or low ICI score groups using the limma package (FC = 1.5, FDR = 0.05) and obtained 890 DEGs. Our subsequent GSEA analysis of these DEGs showed that the high ICI score group was significantly enriched in allograft rejection, E2F targets, G2M checkpoint, interferon gamma response and MYC targets V2 pathways; while the low ICI score group was mainly enriched in epithelial mesenchymal transition, estrogen response early, protein secretion, TGF-β signaling and UV response pathways. (Figures 3c and 3d and Supplementary Table 5). In addition, when we compared the relationship between ICI scores and prognosis, we only selected genes with P-value < 0.05 in the univariate survival analysis. We then used PCA to calculate the ICI score for each patient. Patients were redivided into high and low ICI score groups using the median value as the cutoff. The Kaplan-Meier curves in Figure 3e indicated that patients of the high ICI score group have significantly longer survival than those of the low ICI score group (log-rank test, P = 0.033).

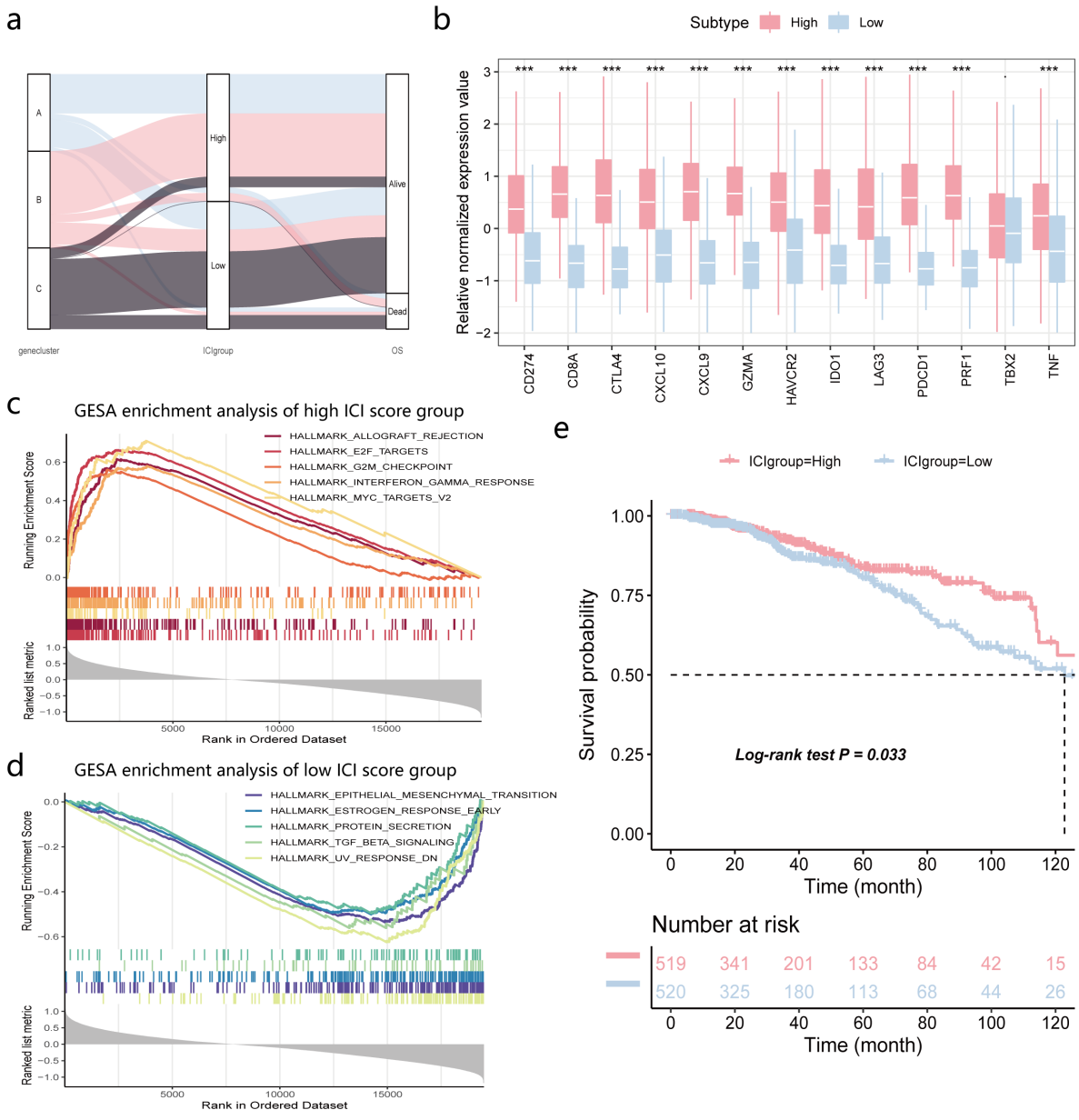

Figure 3. Immune-cell infiltration (ICI) scores construction. (a) Alluvial diagram showing the distribution of ICI gene clusters in different ICI score groups and survival status of TCGA-BRCA patients. (b) Box plot for the relative expression of immune checkpoint-associated genes in different ICI score groups. Among them, *CD274*, *HAVCR2*, *CTLA4*, *LAG3*, *PDCD1* and *IDO1* are inhibitory genes, *CD8A*, *GZMA*, *PRF1*, *CXCL10*, *CXCL9*, *TNF* and *TBX2* are stimulatory genes. The asterisks represented the statistical P-value (Kruskal-Wallis test, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$). (c and d) GESA enrichment maps for high (c) and low (d) ICI score groups. Allograft rejection, E2F targets, G2M checkpoint, interferon gamma response and MYC targets V2 pathways pathways were enriched in the high ICI score group. Epithelial mesenchymal transition, estrogen response early, protein secretion, TGF-β signaling and UV response pathways were enriched in the low ICI score group. (e) Survival analysis for high or low ICI score groups (calculated by using PCA) in TCGA-BRCA patients using Kaplan-Meier curves. The log-rank test showed that $P = 0.033$.

Correlation between immune cell infiltration (ICI) scores and tumor mutation burden (TMB)

Numerous studies have suggested that the immune phenotype may be associated with alterations in the tumor genome [41, 42]. To validate this hypothesis, we tested the

relationship between TMB and ICI scores in the TCGA-BRCA cohort and found that patients in the high ICI score group had more TMB (Kruskal test, $P = 0.002$; Figure 4a). Besides, the scatter plot of the association between TMB and ICI scores also showed a positive association (Pearson correlation = 0.324, $P < 0.001$; Figure 4b). In our stratified survival analysis, which divided patients into different subgroups according to TMB and ICI scores (calculated by using PCA), we found that patients with high level of TMB and low ICI scores had the worst prognosis (log-rank test, $P = 0.039$; Figure 4c). We also performed clustering analysis by using the position information of somatic mutations to identify mutation driver genes in different ICI score subgroups, and mapped the waterfall of the top 25 most significant mutation driver genes using the maftools package (Figure 4d). Expression profiles of patients in distinct ICI score groups in the TCGA-BRCA cohort were matched with CNV data downloaded from the firehose database, and the GISTIC2.0 module of the GenePattern online tool was used to analyze the status of CNV in different groups. The results from the analysis were visualized using the maftools package and are presented in Figures 4e (high ICI score group) and 4f (low ICI score group). We found that both high and low ICI score groups had many copy number variations, but high ICI score groups had more CNVs. The regions significantly amplified in the high ICI score group of patients involved 11q13.3, 17q12 and 8q24.21, while 11q13.3 spanned the *CCND1* gene. Significantly deleted regions in the high ICI score group included 9p21.3, which spans the tumor suppressor genes *CDKN2A* and *CDKN2B*.

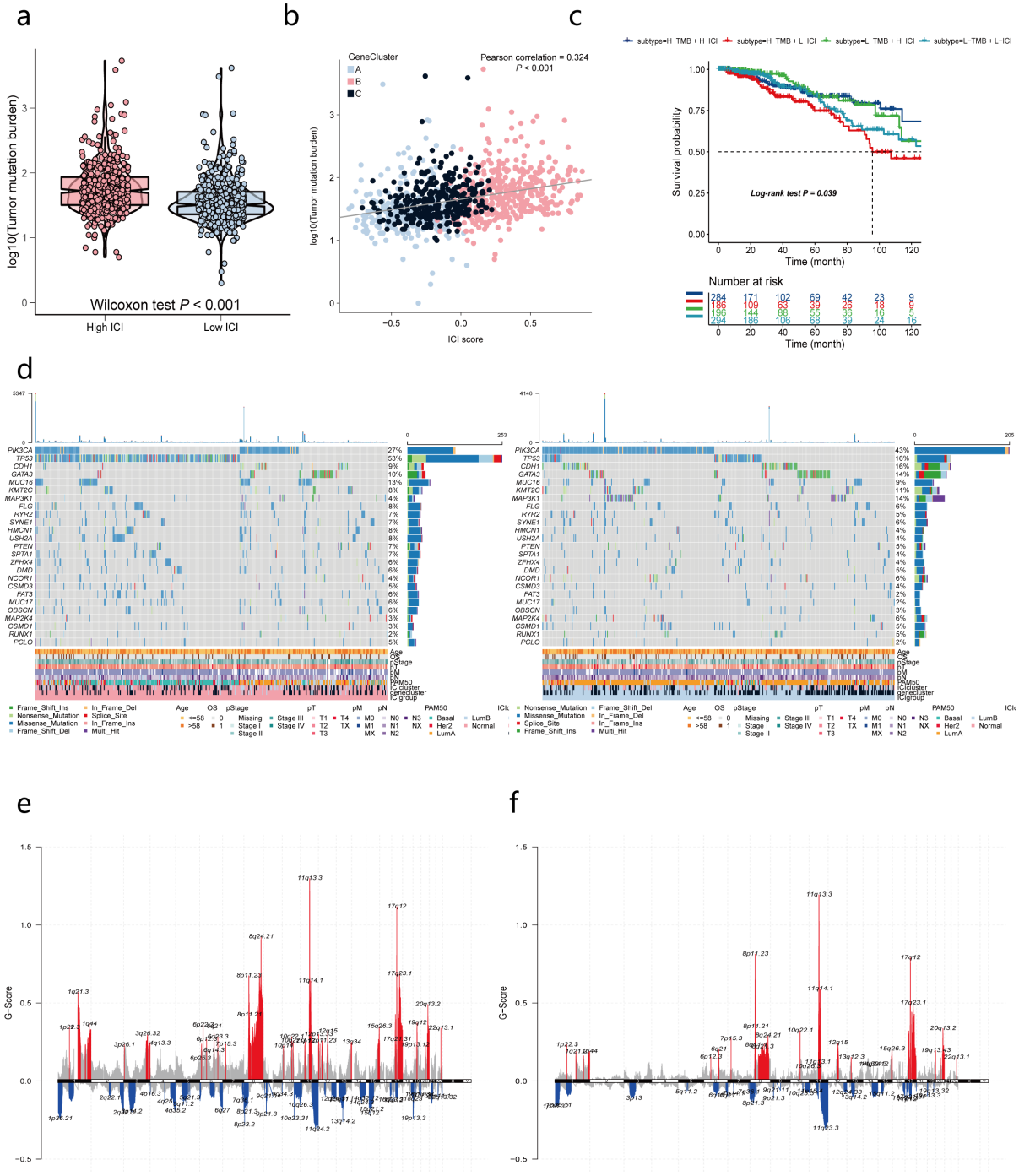

Figure 4. Correlations between immune-cell infiltration (ICI) scores and tumor mutation burden (TMB). (a) Differences in TMB between high or low ICI score groups (Kruskal test, $P < 0.001$). (b) Scatter plot of correlation between ICI scores and the mutational burden in the TCGA-BRCA cohort (Pearson correlation = 0.133, $P < 0.001$). (c) Stratified survival analysis of TCGA-BRCA patients stratified by both TMB and ICI scores (calculated by using PCA) using Kaplan Meier curves. The log-rank test showed that $P = 0.039$ (d) Waterfall plots of the top 25 significantly driver mutated genes in the high (left) or low (right) ICI score groups. Each column represented for individual patients, and the bar plot on top showed the TMB. (e and f) GISTIC2.0-based copy number variant (CNV) analysis of high (e) or low (f) ICI score groups visualized by maftools.

Integrative analysis of immune cell infiltration (ICI) scores on drug response

Furthermore, we selected genes from various pathways related to tumor immune processes and classified the immune-related genes of our interest into gene set 1 (Figure 5a) and gene

set 2 (Figure 5b), and then created a heat map of these genes in the high or low ICI score groups. From Figure 5a, we found that most of the genes related to diverse immune pathways were upregulated in the high ICI score group and downregulated in the low ICI score group. Figure 5b showed a significant increase of genes related to pathways such as cytotoxic cells, effector memory CD8, macrophages, and T cells in the high ICI score group. Both Figures 5a and 5b showed that patients with Luminal A BRCA are mostly enriched in the low ICI score group. In addition to TMB, we also noted the positive correlation between ICI scores and neoantigens (Wilcoxon test, $P < 0.001$) (Figure 5c). We downloaded a dataset of drugs sensitive to the treatment of cancer from the GDSC website, from which we compared the differences in the sensitivity of high or low ICI groups to more than 100 drugs used to treat tumors (Supplementary Table 6). The top 12 drugs with differential treatment responses according to P-value ranking were illustrated in Figure 5d, from which it was clear that high ICI scores may lead to increased sensitivity of BRCA to drugs such as imatinib, CCT007093, MK-2206, CHIR-99021, FH535 and KIN001-135 and so on. All above results may provide new perspectives for investigating the role of individual gene mutations in the immune microenvironment and immunotherapy of cancer.

In recent years, blockade therapy targeting immune checkpoints has emerged as a mainstream immunotherapy with the potential to significantly improve the survival of cancer patients, but only small numbers of patients have responded to this treatment [17, 18]. Markers that can effectively predict the effect of immunotherapy are limited; therefore, to validate the role of ICI scores constructed in BRCA patients in predicting patients' response to immunotherapy, we selected the IMvigor210 cohort of metastatic uroepithelial cancer patients with immunotherapy received as a validation cohort to test the potential to forecast immunotherapy benefit of the ICI scores we established. Encouragingly, in the IMvigor210 cohort, we found that ICI scores were in a significantly positive correlation with the objective response rate (ORR) for anti-PD-L1 therapy (Wilcoxon test, $P = 0.002$; Figure 5e). Moreover, in this cohort, patients with high ICI scores had significantly longer survival (log-rank test, $P = 0.02$; Figure 5f). What is more, we found the high ICI score group had a higher ORR after anti-PD-L1 treatment (Figure 5g). In conclusion, these data suggest that the ICI scores can predict the responses to immunotherapy.

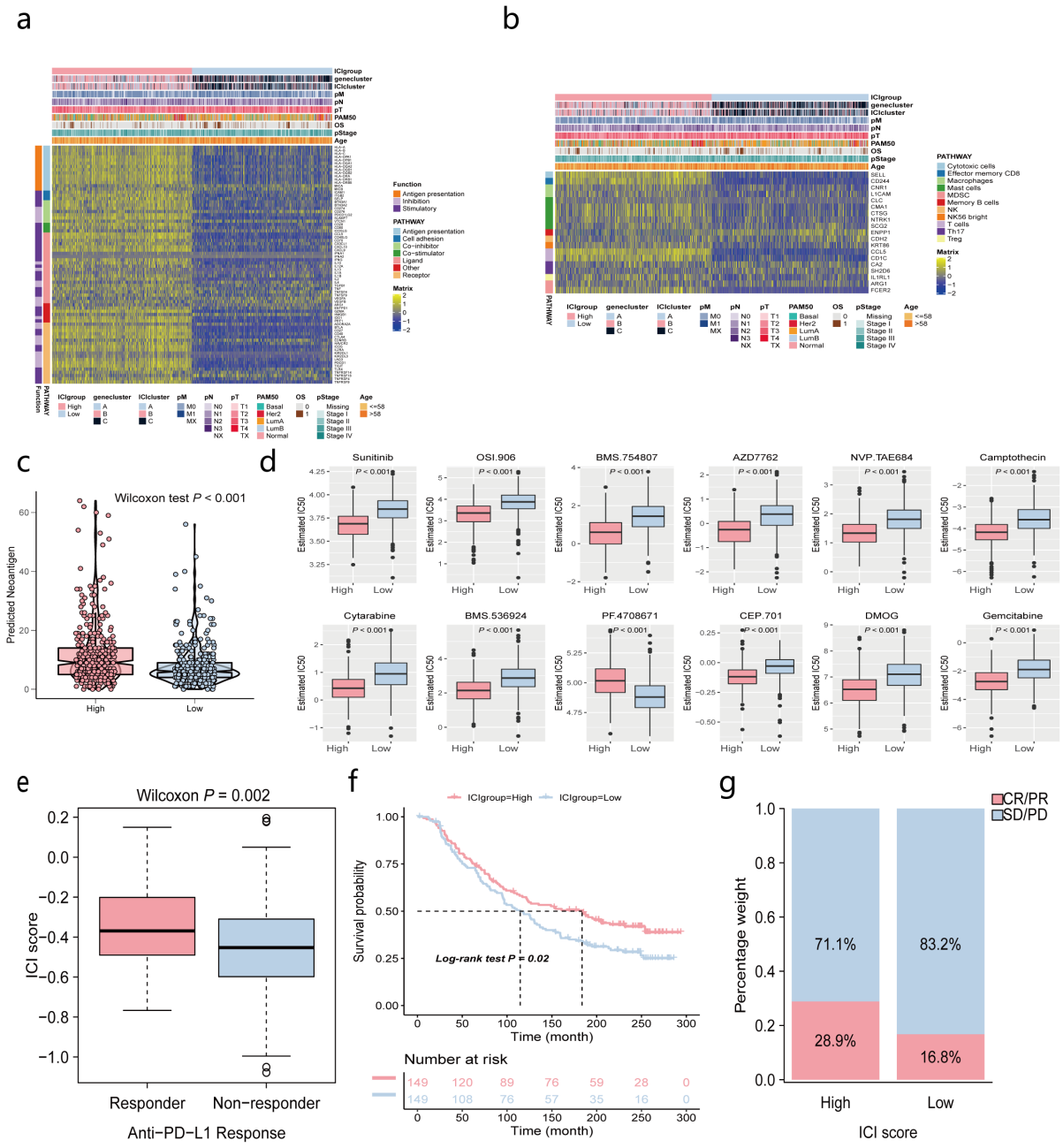

Figure 5. Integrative analysis of immune-cell infiltration (ICI) scores on drug response. (a and b) Heatmap of immune-related genes in different ICI score groups. The immune-related genes were divided into two groups: gene set 1 (a) and gene set 2 (b). The ICI cluster, ICI gene cluster, PAM50 subtype, age, tumor stage and survival status were used as patient annotations. (c) Differences in neoantigen between high or low ICI score groups. (d) Comparison of drug sensitivity in high or low ICI score groups (Wilcoxon test, $P < 0.001$). The box plots show the differences in IC50 values for the top 12 drugs sorted by P-value. (e) Boxplot for the distribution of ICI scores of patients in different anti-PD-L1 therapeutic responses in IMvigor210 cohort (Wilcoxon test, $P = 0.002$). (f) Survival analysis for high or low ICI score groups in IMvigor210 cohort patients using Kaplan-Meier curves. The log-rank test showed that $P = 0.02$. (g) Bar graph showing the proportion of patients with various clinical responses (responder: complete response (CR)/ partial response (PR); non-responder: stable disease (SD)/progressive disease (PD)) to anti-PD-L1 immunotherapy in the high or low ICI score groups of the IMvigor210 cohort.

Discussion

Immunotherapy has changed the treatment and prognosis for many malignancies. In recent years, immunotherapy using checkpoint blockades have proven to generate unprecedented and durable responses in patients suffering from diverse cancers [43, 45]. In BRCA, building on the favorable results of the Phase III IMpassion130 trial [46] and the Phase III KEYNOTE-355 trial [47], the U.S. FDA has accelerated approval for the PD-L1 inhibitor atezolizumab as well as the PD-1 inhibitor pembrolizumab, combined with chemotherapy, for the treatment of locally advanced or metastatic PD-L1-positive triple-negative BRCA (TNBC) patients in 2019 and 2020, respectively [48]. In recent years, there has been increasing evidence that patterns of the immune system play a key role in determining both the response to treatment and survival of BRCA patients [49]. These data and the clinical use of immune checkpoint blockers in a variety of solid tumors have also demonstrated striking success [50, 51]. Stromal TILs concentration shows a linear relationship with clinical outcome in different clinical subtypes of BRCA [49]. For example, it has been shown that HER-2+ BRCA and TNBC have higher levels of TILs and PD-L1 expression in TME at diagnosis than luminal BRCA, which can be predicted to benefit more from adjuvant and neoadjuvant chemotherapy, are more likely to respond to PD-1/PD-L1 blockade and have longer survival [52-54]. However, the use of immunotherapy in BRCA remains limited and only a minority of patients would benefit from it. Poor immunogenicity, T-cell infiltration in TME, and enhanced immunosuppression have been identified as potential challenges to successful immunotherapy for BRCA [55]. Therefore, the development of more efficient biomarkers for predicting response and resistance to therapy, as well as the recognition of environmental modifiers to immunity (mutational load, neoantigens, and sensitive combination therapeutics) is important to improve the efficacy of immunotherapy. It will be of great help to choose the appropriate timing and patients for immunotherapy, patients should be detected markers of immunotherapy response when initial diagnosis and immunotherapy should be used in treatment as early as possible [56]. In this study, we developed a method to quantify ICI in TME of BRCA patients - ICI score, and our results demonstrated that this score can be used as a predictor to assess the effectiveness and prognosis of immunotherapy.

Many studies have demonstrated the importance of an abundant and active BRCA TME in forecasting the response of tumor patients to immunotherapy [57, 58]. For example, tumors with increased TILs, positive PD-L1, and elevated tumor-infiltrating $CD8^+$ T cells exhibit a higher response rate to immunotherapy. Such tumors are considered "inflamed" or "hot" tumors. In contrast, "non-inflamed" or "cold" tumors with lower TILs, PD-L1 expression, and $CD8^+$ T cell infiltration are less likely to respond to immunotherapy [56]. In this study, we analyzed the ICI patterns of 1721 BRCA samples from the integrated cohort and classified BRCA into three separate immune subtypes, ICI clusters I, II, and III. The results of our analysis suggested that patients in ICI clusters I and II with higher TILs infiltration, PD-L1 expression, and high immune scores had longer survival. This is consistent with previous studies [59, 60]. These findings illustrated that the preexistent immune responses in TME can have an impact on the prognosis of BRCA patients as well as on the degree of benefit from immunotherapy. However, it is not sufficient to rely solely on the immune phenotype of the tumor to project the response to immunotherapy. Alterations of certain molecules during tumor progression may also interfere with the interaction between immune cells or between immune cells and tumor cells, thereby disrupting the balance of immune resistance and activation in tumors [38]. However, how the genomic landscape in BRCA shapes and influences antitumor immunity is not yet clear.

Systematic analysis of tumor immune-related gene expression profiles can shed further light on the relationship between tumor genetics and TME. Genetic characterization may also assist in identifying suitable BRCA patients for immunotherapy. We clustered the cohort again based on DEGs between the previous ICI clusters, divided TCGA-BRCA patients into new ICI gene clusters, and defined ICI signature genes. Among these different ICI gene clusters, we discovered that ICI gene cluster C with the lowest levels of activated TILs, immune score, stromal score, exhibited an immune exhausted phenotype. On the contrary, ICI gene clusters A and B had higher inflammatory cell infiltration, immune scores, and stromal scores. And we also observed that ICI gene cluster B had a promising immune activation phenotype because of the highest content of macrophages, resting NK cells, memory activated $CD4^+$ T cells, plasma cells, $CD8^+$ T cells, *etc*. [61,62]. Meanwhile, patients in ICI gene cluster B had the highest expression of PD-1 and PD-L1 and a more optimistic prognosis. We speculated that patients with ICI gene cluster B may be more likely to benefit from immunotherapy. The opposite is the case for patients with ICI gene cluster C, probably because their immune exhausted phenotype may lead to tumor cells evading the immune system and not responding to immunotherapy. Our study is following previous studies [63]. These findings suggested that combining the synthetic features of ICI profiles with expression patterns of immune-related genes in TME may become a promising approach to developing more precise immunotherapy regimens for BRCA patients.

Due to the high individual heterogeneity of TME, we used the ssGSEA method to establish ICI scores for patients in the TCGA-BRCA cohort and to quantify the ICI pattern for each patient. We found that the expression of most of the immune-related genes was higher in the group that had high ICI scores. GSEA analysis of the low ICI score group showed significantly enriched in TGF-β signaling pathway, epithelial mesenchymal transition, *etc.* Notably, TGF-β is a gene that is known to be involved in immunosuppressive pathways [37]. In addition, we found that our constructed ICI scores also correlated with TMB and neoantigens that also could predict response to immunotherapy [42, 64]. TMB levels and the number of neoantigens were significantly higher in the high ICI score group. Both Survival analysis and stratified analysis indicated that higher ICI scores conferred a better prognosis for patients. Moreover, patients with high TMB and low ICI scores had the shortest survival. We also observed that different ICI score groups were associated with altered tumor driver genes (*e.g.*, *PI3KCA, TP53, CDH1*) and gene copy number. High or low ICI scores also showed significant differences in sensitivity to certain other target drugs.

The ability of our established ICI score to predict response to immunotherapy in tumor patients was validated in a cohort of metastatic uroepithelial cancer patients treated with anti-PD-L1 agents (IMvigor210) [65]. Our results showed that ICI scores were significantly higher in patients who had a response to immunotherapy than in those who did not. Patients from the high ICI score group had longer survival and higher ORR. However, lacking data from a cohort of BRCA patients receiving immunotherapy, additional prospective trials are needed to validate these predictors that we constructed in the TCGA-BRCA cohort. In summary, our analysis has revealed environmental and genetic mechanisms affecting tumor-immune interactions in BRCA, and our constructed ICI score may serve as a powerful marker for predicting patient prognosis and the extent of benefit from immunotherapy.

Conclusions

In summary, our analysis has revealed environmental and genetic mechanisms affecting tumor-immune interactions in BRCA, and our constructed ICI score may serve as a powerful marker for predicting patient prognosis and the extent of benefit from immunotherapy.

Data Availability

The datasets and original contributions used to support the findings of this study are included within the article and supplementary information files. Further inquiries can be directed to the corresponding author.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding Statement

This work was supported by the National Natural Science Foundation of China [grant number 81772827].

Supplementary Materials

Supplementary Table 1. Relative Fractions of tumor-infiltrating immune cells of 1721 BRCA patients.

Supplementary Table 2. Differentially expressed genes of ICI clusters in the TCGA-BRCA cohort.

Supplementary Table 3. ICI signature genes A and B.

Supplementary Table 4. Clinical and survival information with ICI cluster and ICI gene cluster and ICI score of TCGA-BRCA patients.

Supplementary Table 5. Gene set enrichment analysis (GSEA) of differentially expressed genes of ICI score groups.

Supplementary Table 6. Drug estimated IC50 between high or low ICI score groups.

References

1. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2020.* CA Cancer J Clin, 2020. **70**(1): p. 7-30.
2. Liu, H., et al., *Evaluating DNA Methylation, Gene Expression, Somatic Mutation, and Their Combinations in Inferring Tumor Tissue-of-Origin.* Front Cell Dev Biol, 2021. **9**: p. 619330.
3. Zhang, Y., et al., *Identifying breast cancer-related genes based on a novel computational framework involving KEGG pathways and PPI network modularity.* Frontiers in Genetics, 2021. **12**: p. 876.
4. He, B., et al., *TOOme: A Novel Computational Framework to Infer Cancer Tissue-of-Origin by Integrating Both Gene Mutation and Expression.* Front Bioeng Biotechnol, 2020. **8**: p. 394.
5. Liu, C., et al., *An Improved Anticancer Drug-Response Prediction Based on an Ensemble Method Integrating Matrix Completion and Ridge Regression.* Mol Ther Nucleic Acids, 2020. **21**: p. 676-686.

6. Liu, X., et al., *A systematic study on drug-response associated genes using baseline gene expressions of the Cancer Cell Line Encyclopedia.* Sci Rep, 2016. **6**: p. 22811.
7. Gonzalez-Angulo, A.M., F. Morales-Vasquez, and G.N. Hortobagyi, *Overview of resistance to systemic therapy in patients with breast cancer.* Adv Exp Med Biol, 2007. **608**: p. 1-22.
8. Wang, X., et al., *Exploring the concepts and practices of advanced breast cancer treatment: a narrative review.* Ann Transl Med, 2021. **9**(8): p. 721.
9. Puram, S.V., et al., *Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer.* Cell, 2017. **171**(7): p. 1611-1624.e24.
10. Song, Z., et al., *Evaluating the Potential of T Cell Receptor Repertoires in Predicting the Prognosis of Resectable Non-Small Cell Lung Cancers.* Mol Ther Methods Clin Dev, 2020. **18**: p. 73-83.
11. Yang, L., et al., *Clinical significance of the immune microenvironment in ovarian cancer patients.* Mol Omics, 2018. **14**(5): p. 341-351.
12. Erdag, G., et al., *Immunotype and immunohistologic characteristics of tumor-infiltrating immune cells are associated with clinical outcome in metastatic melanoma.* Cancer Res, 2012. **72**(5): p. 1070-80.
13. Galon, J., et al., *Type, density, and location of immune cells within human colorectal tumors predict clinical outcome.* Science, 2006. **313**(5795): p. 1960-4.
14. Banchereau, J. and R.M. Steinman, *Dendritic cells and the control of immunity.* Nature, 1998. **392**(6673): p. 245-52.
15. Chen, Y., et al., *Tumor-associated macrophages: an accomplice in solid tumor progression.* J Biomed Sci, 2019. **26**(1): p. 78.
16. Nelson, M.A., et al., *Prognostic and therapeutic role of tumor-infiltrating lymphocyte subtypes in breast cancer.* Cancer Metastasis Rev, 2021.
17. Domchek, S.M., et al., *Olaparib and durvalumab in patients with germline BRCA-mutated metastatic breast cancer (MEDIOLA): an open-label, multicentre, phase 1/2, basket study.* Lancet Oncol, 2020. **21**(9): p. 1155-1164.
18. Tolaney, S.M., et al., *Effect of Eribulin With or Without Pembrolizumab on Progression-Free Survival for Patients With Hormone Receptor-Positive, ERBB2-Negative Metastatic Breast Cancer: A Randomized Clinical Trial.* JAMA Oncol, 2020. **6**(10): p. 1598-1605.
19. Brufsky, A., et al., *A phase II randomized trial of cobimetinib plus chemotherapy, with or without atezolizumab, as first-line treatment for patients with locally advanced or metastatic triple-negative breast cancer (COLET): primary analysis.* Ann Oncol, 2021. **32**(5): p. 652-660.
20. Rugo, H.S., et al., *Safety and Antitumor Activity of Pembrolizumab in Patients with Estrogen Receptor-Positive/Human Epidermal Growth Factor Receptor 2-Negative Advanced Breast Cancer.* Clin Cancer Res, 2018. **24**(12): p. 2804-2811.
21. Zou, W., J.D. Wolchok, and L. Chen, *PD-L1 (B7-H1) and PD-1 pathway blockade for cancer therapy: Mechanisms, response biomarkers, and combinations.* Sci Transl Med, 2016. **8**(328): p. 328rv4.
22. Colaprico, A., et al., *TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data.* Nucleic Acids Research, 2015. **44**(8).
23. Yau, C., et al., *A multigene predictor of metastatic outcome in early stage hormone receptor-negative and triple-negative breast cancer.* Breast Cancer Res, 2010. **12**(5): p. R85.
24. Newman, A.M., et al., *Robust enumeration of cell subsets from tissue expression profiles.* Nature Methods, 2015. **12**(5): p. 453-457.
25. Alexandra, S., et al., *Contribution of systemic and somatic factors to clinical response and resistance to PD-L1 blockade in urothelial cancer: An exploratory multi-omic analysis.* PLOS Medicine, 2017. **14**(5).
26. Chao, C., et al., *Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods.* PLoS ONE, 2011. **6**(2): p. e17238.
27. Yoshihara, K., et al., *Inferring tumour purity and stromal and immune cell admixture from expression data.* Nature Communications, 2013. **4**.
28. Hayes, D.N., *ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking.* Bioinformatics, 2010. **26**(12): p. 1572-1573.
29. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic acids research, 2015. **43**(7): p. e47.
30. Gu, Z., E. Roland, and S. Matthias, *Complex heatmaps reveal patterns and correlations in multidimensional genomic data.* Bioinformatics, 2016. **32**(18): p. 2847.
31. Mayakonda, A., et al., *Maftools: efficient and comprehensive analysis of somatic variants in cancer.* Genome Research, 2018. **28**(11).
32. David, T., G.P. Abel, and L.B. Nuria, *OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes.* Bioinformatics, 2013(18): p. 2238-2244.
33. Yang, W., et al., *Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells.* Nucleic Acids Research, 2013(D1): p. D955.

34. Paul, G., et al., *pRRophetic: An R Package for Prediction of Clinical Chemotherapeutic Response from Tumor Gene Expression Levels.* Plos One, 2014. **9**(9): p. e107468-e107468.
35. Rooney, M., et al., *Molecular and genetic properties of tumors associated with local immune cytolytic activity.* Cell, 2015. **160**(1-2): p. 48-61.
36. Hnzelmann, S., R. Castelo, and J. Guinney, *GSVA: gene set variation analysis for microarray and RNA-Seq data.* Bmc Bioinformatics, 2013. **14**(1): p. 7-7.
37. Chen, Y.P., et al., *Identification and validation of novel microenvironment-based immune molecular subgroups of head and neck squamous cell carcinoma: implications for immunotherapy.* Ann Oncol, 2019. **30**(1): p. 68-75.
38. Chen, D.S. and I. Mellman, *Elements of cancer immunity and the cancer-immune set point.* Nature, 2017. **541**(7637): p. 321-330.
39. Ayers, M., et al., *IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade.* J Clin Invest, 2017. **127**(8): p. 2930-2940.
40. Hugo, W., et al., *Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma.* Cell, 2017. **168**(3): p. 542.
41. Rizvi, N.A., et al., *Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer.* Science, 2015. **348**(6230): p. 124-8.
42. McGranahan, N., et al., *Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade.* Science, 2016. **351**(6280): p. 1463-9.
43. Pinato, D.J., et al., *PRIME-HCC: phase Ib study of neoadjuvant ipilimumab and nivolumab prior to liver resection for hepatocellular carcinoma.* BMC Cancer, 2021. **21**(1): p. 301.
44. Sacco, A.G., et al., *Pembrolizumab plus cetuximab in patients with recurrent or metastatic head and neck squamous cell carcinoma: an open-label, multi-arm, non-randomised, multicentre, phase 2 trial.* Lancet Oncol, 2021. **22**(6): p. 883-892.
45. Schmid, P., et al., *Pembrolizumab plus chemotherapy as neoadjuvant treatment of high-risk, early-stage triple-negative breast cancer: results from the phase 1b open-label, multicohort KEYNOTE-173 study.* Ann Oncol, 2020. **31**(5): p. 569-581.
46. Schmid, P., et al., *Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer.* N Engl J Med, 2018. **379**(22): p. 2108-2121.
47. Cortes, J., et al., *Pembrolizumab plus chemotherapy versus placebo plus chemotherapy for previously untreated locally recurrent inoperable or metastatic triple-negative breast cancer (KEYNOTE-355): a randomised, placebo-controlled, double-blind, phase 3 clinical trial.* Lancet, 2020. **396**(10265): p. 1817-1828.
48. Lin, H., et al., *T Cell Receptor Repertoire Sequencing.* Methods Mol Biol, 2020. **2204**: p. 3-12.
49. Savas, P., et al., *Clinical relevance of host immunity in breast cancer: from TILs to the clinic.* Nat Rev Clin Oncol, 2016. **13**(4): p. 228-41.
50. Lipson, E.J., et al., *Antagonists of PD-1 and PD-L1 in Cancer Treatment.* Semin Oncol, 2015. **42**(4): p. 587-600.
51. Emens, L.A., et al., *Cancer immunotherapy: Opportunities and challenges in the rapidly evolving clinical landscape.* Eur J Cancer, 2017. **81**: p. 116-129.
52. Cimino-Mathews, A., et al., *PD-L1 (B7-H1) expression and the immune tumor microenvironment in primary and metastatic breast carcinomas.* Hum Pathol, 2016. **47**(1): p. 52-63.
53. Li, X., et al., *Prognostic Role of Programmed Death Ligand-1 Expression in Breast Cancer: A Systematic Review and Meta-Analysis.* Target Oncol, 2016. **11**(6): p. 753-761.
54. Topalian, S.L., et al., *Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy.* Nat Rev Cancer, 2016. **16**(5): p. 275-87.
55. Basu, A., et al., *Immunotherapy in breast cancer: Current status and future directions.* Adv Cancer Res, 2019. **143**: p. 295-349.
56. Heeke, A.L. and A.R. Tan, *Checkpoint inhibitor therapy for metastatic triple-negative breast cancer.* Cancer Metastasis Rev, 2021: p. 1-11.
57. Woo, S.R., L. Corrales, and T.F. Gajewski, *The STING pathway and the T cell-inflamed tumor microenvironment.* Trends Immunol, 2015. **36**(4): p. 250-6.
58. Loi, S., et al., *Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: results from the FinHER trial.* Ann Oncol, 2014. **25**(8): p. 1544-50.
59. Rooney, M.S., et al., *Molecular and genetic properties of tumors associated with local immune cytolytic activity.* Cell, 2015. **160**(1-2): p. 48-61.
60. He, Y., et al., *Classification of triple-negative breast cancers based on Immunogenomic profiling.* J Exp Clin Cancer Res, 2018. **37**(1): p. 327.

61. Biswas, S.K. and A. Mantovani, *Macrophage plasticity and interaction with lymphocyte subsets: cancer as a paradigm.* Nat Immunol, 2010. **11**(10): p. 889-96.
62. Hwang, M.L., J.R. Lukens, and T.N. Bullock, *Cognate memory CD4+ T cells generated with dendritic cell priming influence the expansion, trafficking, and differentiation of secondary CD8+ T cells and enhance tumor control.* J Immunol, 2007. **179**(9): p. 5829-38.
63. Zhang, X., et al., *Characterization of the Immune Cell Infiltration Landscape in Head and Neck Squamous Cell Carcinoma to Aid Immunotherapy.* Molecular Therapy - Nucleic Acids, 2020. **22**: p. 298-309.
64. Goodman, A.M., et al., *Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers.* Mol Cancer Ther, 2017. **16**(11): p. 2598-2608.
65. Mariathasan, S., et al., *TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells.* Nature, 2018. **554**(7693): p. 544-548.