CANCER GENOMICS & PROTEOMICS *21*: 316-326 (2024)
doi:10.21873/cgp.20450

Identification of a Gene Expression Signature to Predict the Risk of Early Recurrence and the Degree of Immune Cell Infiltration in Triple-negative Breast Cancer

KEIKO SATO[1], KENTARO MIURA[1], SHOMA TAMORI[2] and KAZUNORI AKIMOTO[2]

[1]*Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science, Chiba, Japan;*
[2]*Department of Medicinal and Life Sciences, Faculty of Pharmaceutical Sciences, Tokyo University of Science, Chiba, Japan*

Abstract. *Background/Aim: Patients with triple-negative breast cancer (TNBC) have a high rate of recurrence within 3 years of diagnosis and a high rate of death within 5 years compared to other subtypes. The number of clinical trials investigating various new agents and combination therapies has recently increased; however, current strategies benefit only a minority of patients. This study aimed to identify specific genes that predict patients at high risk of recurrence and the immune status of the tumor microenvironment at an early stage, thereby providing insight into potential therapeutic targets to improve clinical outcomes in TNBC patients. Materials and Methods: We evaluated the prognostic significance of microarray mRNA expression of 20,603 genes in 233 TNBC patients from the METABRIC dataset and further validated the results using RNA-seq mRNA expression data in 143 TNBC patients from the GSE96058 dataset. Results: Eighteen differentially expressed genes (AKNA, ARHGAP30, CA9, CD3D, CD3G, CD6, CXCR6, CYSLTR1, DOCK10, ENO1, FLT3LG, IFNG, IL2RB, LPXN, PRKCB, PVRIG, RASSF5, and STAT4) identified in both datasets were found to be reliable biomarkers for predicting TNBC recurrence and progression. Notably, the genes whose low expression was associated with increased risk of recurrence and death were immune-related genes, with significant differences in levels of immune cell infiltration in the tumor microenvironment between high- and low- expression groups. Conclusion: Genes reported herein may be effective biomarkers to identify TNBC patients who will and will not benefit from immunotherapy and may be particularly important genes for developing future treatment strategies, including immunotherapy.*

Correspondence to: Keiko Sato, 2641 Yamazaki, Noda, Chiba 278-8510, Japan. Tel: +81 471241501, e-mail: keiko@is.noda.tus.ac.jp

Key Words: TNBC, recurrence, immune cell infiltration, biomarker, gene expression signature.

Triple-negative breast cancer (TNBC) is a subtype of breast cancer that is pathologically defined by lack of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor-2 (HER2). TNBC accounts for approximately 15-20% of all newly diagnosed breast tumors (1-4) and is associated with an earlier age of onset, higher histological grade and poorer clinical outcome compared to other breast cancer subtypes (1, 2, 5). Patients with TNBC have a high rate of recurrence within 3 years of diagnosis and a high rate of death within 5 years (6, 7). This disease progression is attributed to increased aggressiveness, molecular heterogeneity, and resistance to chemotherapy (8). The urgent need for improved survival in TNBC remains unmet despite the approval of several targeted therapies (9, 10). TNBC is the breast cancer subtype that has the highest incidence of patients with a robust tumor immune infiltrate (11). Increasing levels of tumor-infiltrating lymphocytes have been associated with favorable outcome in TNBC (11, 12). The molecular features that cause higher or lower levels of immune infiltration in TNBC are unclear. Biomarkers that can predict the immune status of the tumor microenvironment and identify patients at high risk of recurrence at an early stage are essential for the discovery of potential therapeutic targets and the implementation of the most appropriate immunotherapy, chemotherapy, targeted therapy, and combination therapy. The aim of the study is to identify specific genes associated with TNBC recurrence and death by comprehensively investigating the molecular features that promote progression of TNBC. Furthermore, our aim is to investigate the effect of genes

involved in the clinical outcome of TNBC on immunogenicity and to provide directions for potential therapeutic targets to improve clinical outcome.

Materials and Methods

Dataset. The METABRIC, Nature 2012 & Nat Commun 2016 dataset was downloaded from cBioPortal (13, 14) on June 22, 2023. Clinical information for 2,509 breast cancer patients was obtained from the data_clinical_patient.txt and data_clinical_sample.txt files. Microarray mRNA expression levels of 20,603 genes, expressed as log intensity levels, were obtained from the data_mrna_illumina_microarray.txt file. Based on available clinical data, we defined ER negative as both "Negative" in ER_IHC and "Negative" in ER_STATUS, PR negative as "Negative" in PR_STATUS, and HER2 negative as "Negative" in HER2_STATUS after excluding "GAIN" in HER2_SNP6. Of 2,509 patients, 233 were classified as TNBC. We used a total of 233 TNBC patients with clinical and mRNA expression data.

Overall survival (OS) was defined as the time from initial diagnosis until death or last follow-up. Of the 233 patients, 90 patients died of their cancer, 32 patients died of other causes, and 111 patients were alive (censored) at last follow-up. Relapse-free status (RFS) was defined as the time from initial diagnosis until relapse or last follow-up. Of the 233 patients, 97 patients relapsed. Age and Nottingham Prognostic Index (NPI) were available for all 233 patients, but stage and grade were not available for 64 and 2 patients, respectively. The median age at diagnosis and the median NPI were 54 years [interquartile range (IQR)=43-65] and 4.10 (IQR=4.04-5.07), respectively.

Differential gene expression analysis and survival analysis. We examined the clinical outcomes of 233 patients with TNBC in the METABRIC dataset. The Kaplan-Meier method was used to estimate survival rates. Spearman correlation was used to measure the strength of association between RFS and OS.

For each of the 20,603 genes, we calculated the mean mRNA expression level, excluding outliers, in all 233 TNBC patients, and then divided the patients into high and low expression groups based on the mean mRNA expression level. Outliers were defined as values greater than the third quartile + 1.5 × IQR or less than the first quartile − 1.5 × IQR. The generalized Wilcoxon test was applied to select genes with highly significant differences (p<0.005) in both RFS and OS between the two groups for each gene. Patients who died from causes other than breast cancer were treated as censored at the time of their death. Cox proportional hazards analysis was used to further evaluate their effect on recurrence and breast cancer-specific death and identify genes associated with clinical outcomes in TNBC patients.

The Chi-square and p-values for the Wilcoxon test were calculated using Python lifelines (version 0.26.0) and Spearman correlation analysis, Kaplan-Meier survival analysis and Cox proportional hazards analysis were performed using JMP Pro 17. Statistical significance was set at p<0.05 unless otherwise noted.

Validation analysis. We validated the genes identified in the METABRIC dataset using RNA-seq mRNA expression data from 143 TNBC patients in the GSE96058 dataset, which was downloaded from the Gene Expression Omnibus database on July 3, 2023. The GSE dataset included age at diagnosis, tumor size, lymph node status, histological grade, and ER, PR and HER2 status. OS information was available, but RFS information was not available. Of 143 patients, 26 patients died, and 117 patients were alive (censored) at last follow-up. The NPI was calculated for each TNBC patient in the GSE dataset based on tumor size, lymph node status and histologic grade (15). The median age at diagnosis and the median NPI were 61 years (IQR=51-72) and 4.44 (IQR=4.24-5.40), respectively. Cox proportional hazards analysis was used to determine whether the genes that were differentially expressed between TNBC patients with good and poor outcomes in the METABRIC dataset were also associated with survival in TNBC patients in the GSE dataset. As in the METABRIC dataset, the 143 patients in the GSE dataset were divided into high and low groups based on the mean mRNA expression level of each gene.

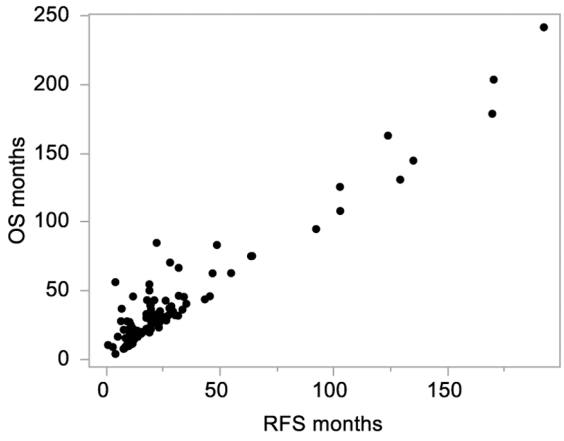

Figure 1. *Scatter plot showing correlation between RFS and OS for TNBC patients. RFS, Relapse-free status; OS, overall survival; TNBC: triple-negative breast cancer.*

In addition, using a set of genes consisting of differentially expressed genes identified in both datasets, we calculated a gene expression signature score for each respective patient as follows. For both the METABRIC and GSE datasets, after performing z-score normalization on the mRNA expression level of each gene in the set, the gene expression signature score of each patient was defined as the mean of the z-scores of all genes in the set. To obtain the mean, the z-score was multiplied by (–1) for genes whose low expression was associated with an increased risk of recurrence and death, and by (+1) for genes whose high expression was associated with an increased risk of recurrence and death. Each patient was divided into high (>0) and low (≤0) groups based on the gene expression signature score. We used the generalized Wilcoxon test to test for statistically significant differences in OS and RFS between the two groups in the METABRIC dataset, or in OS between the two groups in the GSE dataset.

Correlation analysis of differentially expressed genes with immune checkpoint molecules. Spearman's correlation was used to examine the correlation of expression levels between differentially expressed genes involved in clinical outcomes, and their correlation with immune checkpoint molecules, including *CD274*, *CTLA4*, *HAVCR2*, *IDO1*, *LAG-3*, *PDCD1*, *PDCD1LG2*, and *TIGIT*. Spearman's correlation was considered positive if it was greater than 0.4, and negative if it was less than 0.4. Correlation analysis was performed using Python pandas (version 1.0.5).

Table I. *Genes with highly significant differences in both RFS and OS between the high and low expression groups in METABRIC TNBC patients.*

88 genes whose low expression was associated with high risk							35 genes whose high expression was associated with high risk		
ADARB2-AS1	*AGAP2*	*AKNA*	*APOBEC3G*	*ARHGAP30*	*BTN3A2*	*C8orf31*	*CA9*	*CAPN14*	*CYTH2*
CCDC28A	*CD226*	*CD300LF*	*CD3D*	*CD3G*	*CD52*	*CD53*	*DCAF4*	*DNAAF3*	*DNAJB7*
CD6	*CD83*	*CLEC12A*	*CTSS*	*CXCL9*	*CXCR6*	*CYSLTR1*	*DNM3*	*DOLK*	*DONSON*
DEF6	*DOCK10*	*DOCK2*	*EVI2B*	*FBXO47*	*FLT3LG*	*GBP5*	*ENO1*	*EXD2*	*FAM204A*
GZMA	*GZMH*	*HCLS1*	*HHEX*	*HLA-DMA*	*HLA-DMB*	*HLA-DOA*	*FFAR1*	*GRWD1*	*H2BC9*
HLA-DQA1	*HLA-DRA*	*IFNG*	*IGFALS*	*IL10RA*	*IL2RB*	*INSRR*	*LBX2*	*LINC00943*	*MAP3K9*
ITK	*KCNE1*	*KIR2DL3*	*KIR3DL1*	*KIR3DL3*	*KLRD1*	*LAPTM5*	*MRGBP*	*NCAPG*	*NCBP2*
LAT2	*LCP1*	*LPXN*	*P2RY13*	*PARVG*	*PCED1B*	*PITPNM2*	*PDCD6IPP2*	*PHOX2A*	*PLIN3*
PLBD1	*PLCL2*	*PLEK*	*PPM1M*	*PRAMEF17*	*PREX1*	*PRKCB*	*PRADC1*	*PTRHD1*	*RCAN3*
PRKCH	*PSMB10*	*PTPRO*	*PVRIG*	*RASGEF1A*	*RASSF5*	*RCBTB2*	*SCAF4*	*SKA1*	*SNORA65*
REM1	*RFTN1*	*S1PR4*	*SASH3*	*SEL1L*	*SLAMF7*	*SLC52A1*	*TMEM190*	*TMEM59L*	*TSEN34*
SLCO2B1	*SPCS3*	*STAT4*	*TBXAS1*	*TPK1*	*TRAT1*	*VAV1*	*USP30*	*ZNF271P*	
VCAM1	*WNT2*	*ZBED2*	*ZNF831*						

RFS, Relapse-free status; OS, overall survival; TNBC: triple-negative breast cancer.

Enrichment analysis and immune infiltration analysis. Gene Ontology (GO) enrichment analysis was performed using Metascape (16) to further investigate the biological function of the differentially expressed genes identified in this study. GO analysis was included biological processes, cellular components, and molecular functions. In addition, the relationship between the expression of these genes and the infiltration levels of 22 different immune cells was examined using the CIBERSORTx algorithm (17). The immune cell types were B cells naïve, B cells memory, Plasma cells, T cells CD8, T cells CD4 naïve, T cells CD4 memory resting, T cells CD4 memory activated, T cells follicular helper, T cells regulatory (Tregs), T cells gamma delta, NK cells resting, NK cells activated, Monocytes, Macrophages M0, Macrophages M1, Macrophages M2, Dendritic cells resting, Dendritic cells activated, Mast cells resting, Mast cells activated, Eosinophils and Neutrophils.

Results

Genes associated with clinical outcome in TNBC. In the METABRIC dataset, the mean OS was 115.2 months for a total of 233 TNBC patients and 44.1 months for 90 TNBC patients who died of breast cancer (median 94.9 and 31.1 months, respectively). Of the 233 patients, 97 (41.6%) relapsed (77 relapsed within 3 years of diagnosis and 20 relapsed beyond 3 years), and of those who relapsed, 90 died of breast cancer (73 died within 5 years of diagnosis and 17 died beyond 5 years), 4 died of other causes, and 3 remained alive. The 5-year survival rates were 67.6% for all 233 patients and only 24.0% for the 97 recurrent patients. There was a strong positive correlation (Spearman correlation 0.802) between time to recurrence and time to death for the 90 patients with recurrence and death (Figure 1). Of the 77 patients who relapsed within 3 years of diagnosis, 74 died of breast cancer (71 died within 5 years and 3 died beyond 5 years), and of the 20 patients who relapsed beyond 3 years of diagnosis, 16 died of breast cancer (2 died within 5 years and 14 died beyond 5 years). The results showed that all TNBC patients who died of breast cancer had evidence of recurrence, and that more than 30% of TNBC patients died within 5 years of their diagnosis.

In the TNBC population, the differences in RFS and OS between the high- and low-expression groups for each of a total of 20,603 genes were assessed using the generalized Wilcoxon test. Table I shows 123 genes with highly significant differences ($p<0.005$) in both RFS and OS. The 88 genes whose low expression was associated with a high risk of recurrence and death are listed on the left side of Table I, and the 35 genes whose high expression was associated with a high risk of recurrence and death are listed on the right side of Table I.

Univariable Cox proportional hazards analysis showed that NPI was significantly associated with RFS and OS (both $p<0.0001$), whereas age at diagnosis was not associated with RFS and OS ($p=0.605$ and $p=0.656$, respectively). Multivariable Cox proportional hazards analysis was performed to estimate the hazard ratios of the high- and low-expression groups for recurrence and breast cancer-specific death, adjusted for NPI and age at diagnosis (Supplementary Table I and Supplementary Table II). Low expression of the 88 genes was associated with an increased risk of recurrence and death (adjusted hazard ratios=1.665-2.667 for recurrence and 1.689-3.024 for death, both $p<0.05$ for all genes). In contrast, high expression of the 35 genes was associated with an increased risk of recurrence and death (adjusted hazard ratios=1.482-2.173 for recurrence, $p<0.05$ for all genes except *SKA1* gene; adjusted hazard ratios=1.601-2.106 for death, $p<0.05$ for all genes).

Validation of identified genes. Out of the 123 genes associated with clinical outcome in the METABRIC (microarray)

Table II. *Eighteen genes associated with clinical outcome in TNBC in both the METABRIC and GSE datasets.*

	GSE95058 dataset			METABRIC dataset		
Gene	HR for death	95% CI of HR	*p*-Value	HR for death	95% CI of HR	*p*-Value
HR of low to high expression						
AKNA	2.713	1.129-7.535	0.0246	2.048	1.340-3.182	0.0009
ARHGAP30	2.979	1.297-7.674	0.0093	2.279	1.486-3.557	0.0001
CD3D	2.476	1.035-6.861	0.0411	2.563	1.678-3.950	<0.0001
CD3G	3.053	1.270-8.184	0.0118	2.361	1.544-3.670	<0.0001
CD6	2.407	1.027-6.294	0.0429	2.274	1.487-3.524	0.0001
CXCR6	2.649	1.131-6.963	0.0240	2.316	1.502-3.647	0.0001
CYSLTR1	2.480	1.092-6.150	0.0295	2.060	1.339-3.237	0.0009
DOCK10	2.926	1.211-8.150	0.0159	2.293	1.499-3.559	0.0001
FLT3LG	2.641	1.174-6.357	0.0185	2.172	1.417-3.382	0.0003
IFNG	2.451	1.029-6.437	0.0426	2.588	1.640-4.223	<0.0001
IL2RB	3.570	1.504-9.830	0.0032	2.106	1.376-3.269	0.0006
LPXN	2.896	1.252-7.308	0.0124	2.294	1.495-3.577	0.0001
PRKCB	2.998	1.199-9.081	0.0174	2.251	1.466-3.512	0.0002
PVRIG	3.151	1.315-8.736	0.0089	2.255	1.461-3.531	0.0002
RASSF5	3.219	1.245-7.942	0.0045	3.024	1.951-4.796	<0.0001
STAT4	3.293	1.388-9.068	0.0059	2.495	1.620-3.902	<0.0001
HR of high to low expression						
CA9	4.910	1.757-17.658	0.0015	1.802	1.185-2.729	0.0062
ENO1	2.650	1.038-8.153	0.0411	1.664	1.099-2.529	0.0163

HR, Hazard ratio; CI, confidence interval; TNBC: triple-negative breast cancer. Hazard ratio was adjusted for age at diagnosis (continuous) and NPI (continuous). The corresponding 95% CI and *p*-value was based on likelihood ratio test. Sample size n=233 for METABRIC. Sample size n=143 for GSE, except *IFNG* and *CA9*, which was n=119 and n=134, respectively. Genes were listed in alphabetical order.

dataset, 18 genes (*AKNA*, *ARHGAP30*, *CA9*, *CD3D*, *CD3G*, *CD6*, *CXCR6*, *CYSLTR1*, *DOCK10*, *ENO1*, *FLT3LG*, *IFNG*, *IL2RB*, *LPXN*, *PRKCB*, *PVRIG*, *RASSF5*, and *STAT4*) were significantly associated with survival in TNBC patients in the GSE (RNA-seq) dataset (Table II). As in METABRIC, multivariable Cox proportional hazards analysis in GSE revealed that low expression levels of 16 genes, *AKNA*, *ARHGAP30*, *CD3D*, *CD3G*, *CD6*, *CXCR6*, *CYSLTR1*, *DOCK10*, *FLT3LG*, *IFNG*, *IL2RB*, *LPXN*, *PRKCB*, *PVRIG*, *RASSF5*, and *STAT4*, were associated with poor survival, whereas high expression levels of *CA9* and *ENO1* were associated with poor survival. *H2BC9* and *ZNF271P* identified in the METABRIC dataset were not validated because they were not included in the GSE gene list. In addition, 11 genes identified in the METABRIC dataset (*ADARB2-AS1*, *FBXO47*, *FFAR1*, *KIR2DL3*, *KIR3DL1*, *KIR3DL3*, *LINC00943*, *PHOX2A*, *PRAMEF17*, *SNORA65*, and *TMEM190*) were excluded from validation due to insufficient number of GSE TNBC patients with available expression levels. Furthermore, when TNBC progression was compared between the high and low signature score groups using a gene set consisting of the 18 differentially expressed genes identified in both datasets, there were significant differences in OS and RFS between these two groups in METABRIC and OS between these two groups in GSE, which did not have RFS information (Figure 2).

Correlation of genes associated with clinical outcomes with immune checkpoint molecules. Among 18 differentially expressed genes identified in both datasets, the expression levels of 16 genes whose low expression was associated with poor survival were positively correlated in all pairs (Figure 3). Each of the 16 genes was also positively correlated with immune checkpoint molecules. Interestingly, in the GSE dataset, these 16 genes were positively correlated with all 8 immune checkpoint molecules (*CD274*, *CTLA4*, *HAVCR2*, *IDO1*, *LAG-3*, *PDCD1*, *PDCD1LG2*, *TIGIT*). In the METABRIC dataset, they were positively correlated with 6 immune checkpoint molecules *CTLA4*, *HAVCR2*, *IDO1*, *LAG-3*, *PDCD1*, and *TIGIT*. Of the 16 genes in the METABRIC dataset, 8 genes (*CD3D*, *CD3G*, *CXCR6*, *DOCK10*, *IFNG*, *IL2RB*, *LPXN*, and *STAT4*) were also positively correlated with *CD274* (also known as *PD-L1*), while the remaining 8 genes had a correlation coefficient of less than 0.4. Except for *DOCK10*, the 15 genes had a correlation coefficient of less than 0.4 with *PDCD1LG2* (also known as *PD-L2*), indicating no correlation. On the other hand, two genes, *CA9* and *ENO1*, had no correlation in the

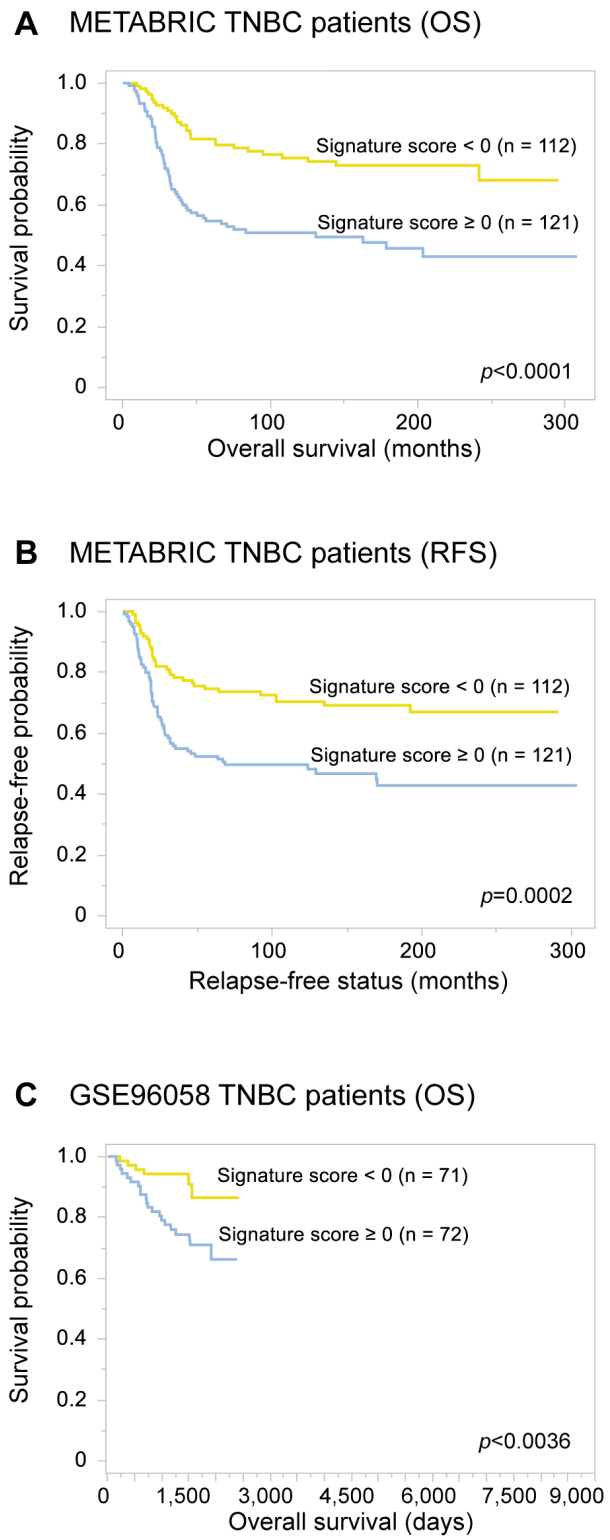

Figure 2. *Comparison of Kaplan-Meier curves for OS (A) and RFS (B) in METABRIC TNBC patients and for OS (C) in GSE TNBC patients according to high and low scores of the 18-gene expression signature. RFS, Relapse-free status; OS, overall survival; TNBC: triple-negative breast cancer.*

METABRIC dataset, but had a positive correlation in the GSE dataset (Figure 4). Also, none of these two genes were found to correlate with the immune checkpoint molecules in both datasets.

Furthermore, regarding the 123 genes (88 and 35 above) identified in the METABRIC dataset, of the 88 genes whose low expression was associated with early recurrence and poor survival, almost all of the 71 genes except for 17 genes (*ADARB2-AS1*, *AGAP2*, *C8orf31*, *CCDC28A*, *FBXO47*, *IGFALS*, *INSRR*, *KCNE1*, *KIR3DL3*, *PLBD1*, *PRAMEF17*, *RASGEF1A*, *RCBTB2*, *REM1*, *SLC52A1*, *SPCS3*, and *WNT2*) showed positive correlations between their expression levels. Spearman's correlation coefficients for the 2,485 pairs of the 71 genes ranged from 0.16 to 0.93 (all $p<0.05$), and 2,369 of the 2,485 pairs (95.3%) had correlation coefficients of 0.4 or greater. Most of these 71 genes were positively correlated with immune checkpoint molecules, particularly *CTLA4*, *HAVCR2*, *IDO1*, *LAG-3*, *PDCD1*, and *TIGIT*, in terms of expression levels. The 17 genes that were not correlated among the 88 genes were also not correlated with all eight immune checkpoint molecules. As for the 35 genes whose high expression was associated with early recurrence and poor survival, there was a positive correlation in expression levels between some genes, such as *NCAPG* and *SKA1*, but no correlation between most of the genes. Also, none of these 35 genes had any correlation with immune checkpoint molecules.

Gene ontology enrichment analysis of differentially expressed genes. We performed GO enrichment analysis for the above 71 genes that were differentially expressed between TNBC patients with good and poor outcomes. The enrichment analysis showed that almost all the 71 genes were immune-related genes. These genes were enriched in the GO biological processes of lymphocyte activation, positive regulation of immune response, regulation of cell activation, immune effector process, and regulation of antigen receptor-mediated signaling pathway, *etc*.; the GO cellular components of side of membrane and immunological synapse; and the GO molecular functions of immune receptor activity and T cell receptor binding (Figure 5). Out of the 71 genes, 65 were included in the complete list of enriched terms (Supplementary Data).

Evaluation of immune cell infiltration. For each of the 16 genes whose low expression was associated with poor survival identified in both the METABRIC and GSE datasets, and for each of the eight immune checkpoint molecules, the proportions of 22 tumor-infiltrating immune cells in the high- and low-expression groups were estimated using the CIBERSORTx algorithm. In addition, for the 88 genes whose low expression was associated with poor outcome identified in the METABRIC dataset (71 genes, mostly immune-related genes, including the aforementioned 16 genes, and 17 genes

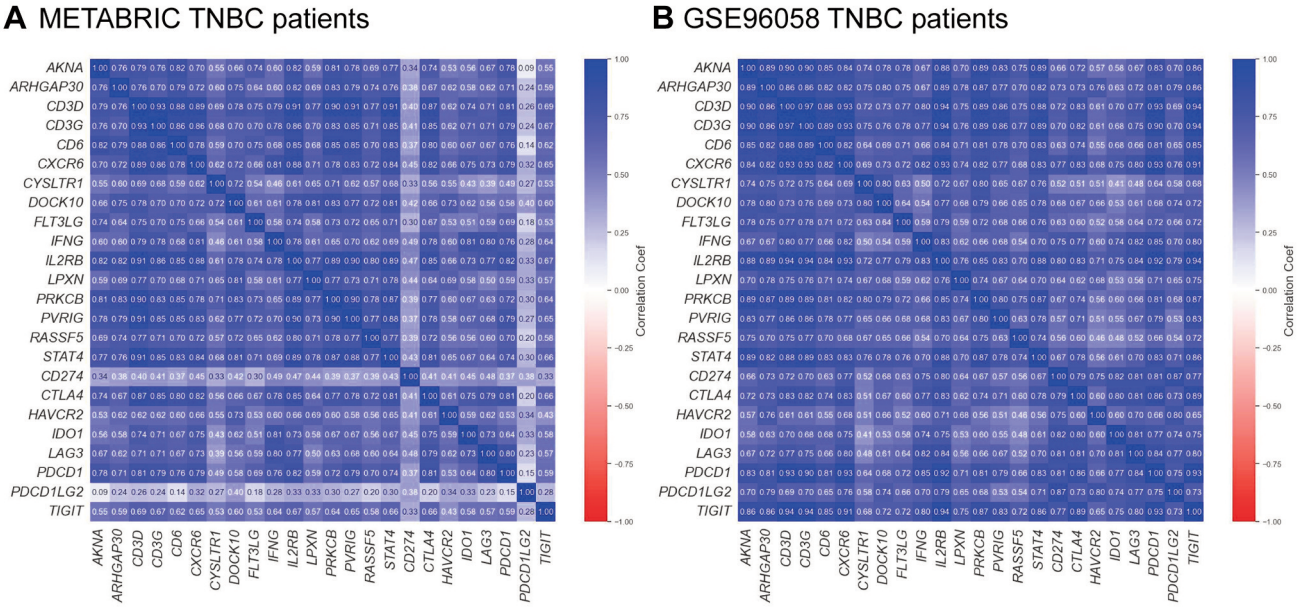

Figure 3. *Spearman correlation heatmap of the selected immune checkpoint molecules and the 16 genes whose low expression was associated with early recurrence and poor survival. Positive correlations are shown in blue and negative correlations are shown in red.*

not correlated with the eight immune checkpoint molecules), the proportions of 22 tumor-infiltrating immune cells in the high- and low-expression groups were estimated. The 17 genes were added to examine differences from the immune-related genes. TNBC patients in both datasets had higher infiltration levels of $CD8^+$ T cells, resting memory $CD4^+$ T cells, and M0-M2 macrophages when compared among their 22 cells; when restricted to TNBC patients in the METABRIC dataset, they also had higher levels of follicular helper T cell infiltration (Supplementary Figure 1 and Supplementary Figure 2). In both datasets, for each of the immune-related genes and immune checkpoint molecules, such as *CTLA4*, *IDO1*, *LAG-3*, *PDCD1*, and *TIGIT*, the infiltration levels of $CD8^+$ T cells, activated memory $CD4^+$ T cells, activated NK cells, and M1 macrophages were significantly higher in the high expression group, whereas the infiltration levels of M0 macrophages and M2 macrophages, were significantly higher in the low expression group (Figure 6). In contrast, for each of the 17 genes not correlated with immune checkpoint molecules, there was no statistically significant difference in the infiltration levels of these cells between the high- and low-expression groups.

The flow chart outlining the data analysis steps and results is shown in Supplementary Figure 3.

Discussion

In the METABRIC dataset (microarray data), low expression of 88 genes and high expression of 35 genes were associated with early recurrence and shorter survival in TNBC patients. Multivariable Cox proportional hazards analysis indicated that the 88 and 35 genes were independent prognostic biomarkers for TNBC recurrence and progression. Most of the up-regulated genes had no correlation with each other and none of them had any correlation with immune checkpoint molecules. This means that it is also important to develop new therapies that are not immune checkpoint inhibitors. Most of the down-regulated genes were immune-related genes that were positively correlated not only with each other regarding their expression levels but also with key immune checkpoint molecules. In addition, these immune-related genes showed significant differences in the level of immune cell infiltration in the tumor microenvironment between their high- and low-expression groups. Thus, there is a difference in the level of infiltration between TNBC patients at low and high risk of recurrence and progression. Also, most of immune checkpoint molecules including *PDCD1* showed significant differences in the level of immune cell infiltration in the tumor microenvironment between their high- and low-expression groups. Additional multivariable Cox proportional hazards analysis adjusted for NPI and age at diagnosis (results not shown) showed that the low *PDCD1*, low *CTLA4* , low *HAVCR2*, and low *IDO1* expression groups had a higher risk of recurrence and death compared to the respective high expression groups (*PDCD1*, HR for death 1.848 [1.205-2.869], HR for recurrence 1.633 [1.086-2.483]; *CTLA-4*, HR for death 2.139 [1.387-3.355], HR for recurrence 2.061 [1.357-3.179]; *HAVCR2*, HR for death 1.702 [1.110-2.634],

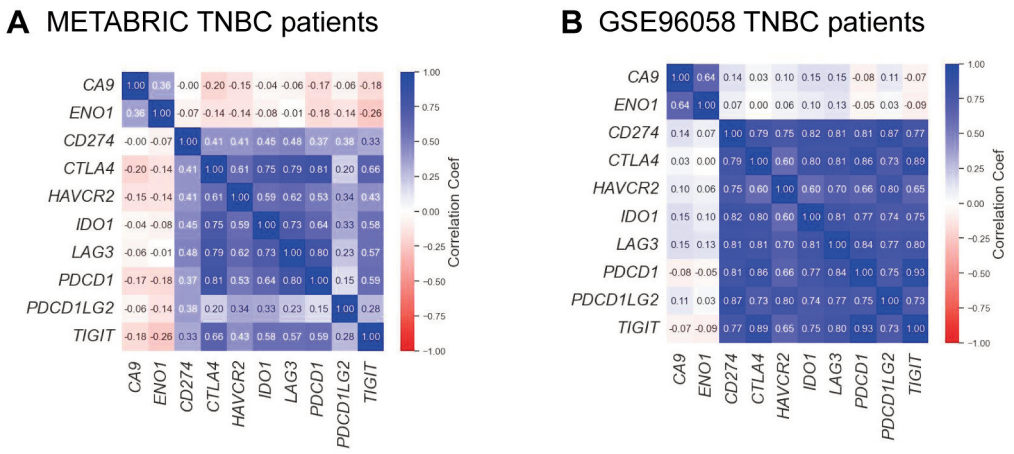

Figure 4. *Spearman correlation heatmap of the selected immune checkpoint molecules and the 2 genes whose high expression was associated with early recurrence and poor survival. Positive correlations are shown in blue and negative correlations are in red.*

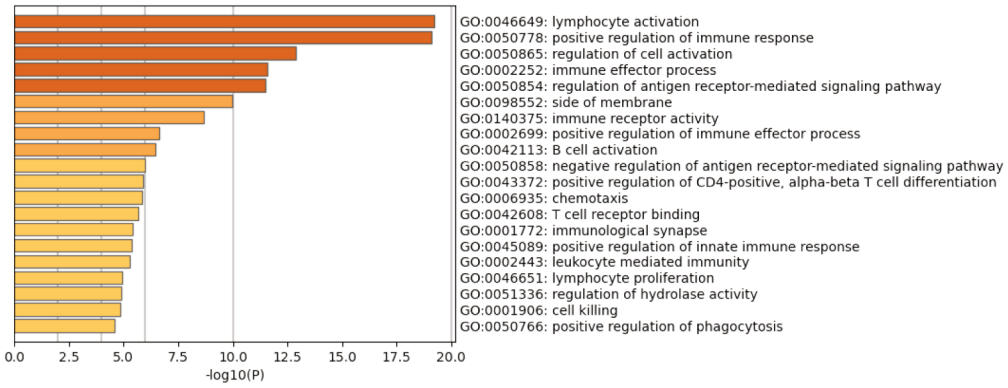

Figure 5. *Bar graph of 20 enriched terms across the 71 genes that were differentially expressed between TNBC patients with good and poor outcomes. TNBC: Triple-negative breast cancer.*

HR for recurrence 1.632 [1.084-2.477]; IDO1, HR for death 1.995 [1.303-3.104], HR for recurrence 1.946 [1.290-2.977]; $p<0.05$ for all cases). *PDCD1* and *CTLA-4* are up-regulated on activated T cells and are considered markers of activation (18). Our results suggest that T cells from high-risk patients are not activated, while T cells from low-risk patients are activated. Activated T cells infiltrate the tumor microenvironment and destroy tumor cells (19). Tumor-infiltrating CD8 lymphocytes have anti-tumor activity and have a favorable impact on patient survival in a variety of cancer types, including breast cancer (20, 21). Our study also showed that the high expression group consisting of patients at low risk of recurrence or death had greater $CD8^+$ T cell infiltration than the low expression group consisting of patients at high risk, not only for immune checkpoint molecules, such as *PDCD1* and *CTLA4*, but also for many of the differentially expressed genes. These genes are as follows: *AKNA*, *APOBEC3G*, *ARHGAP30*, *BTN3A2*, *CD3D*, *CD3G*, *CD6*, *CD52*, *CD53*, *CD83*, *CD226*, *CD300LF*, *CLEC12A*, *CTSS*, *CXCL9*, *CXCR6*, *CYSLTR1*, *DEF6*, *DOCK2*, *DOCK10*, *EBI2V*, *FLT3LG*, *GBP5*, *GZMA*, *GZMH*, *HCLS1*, *HHEX*, *HLA-DMA*, *HLA-DMB*, *HLA-DOA*, *HLA-DQA1*, *HLA-DRA*, *IFNG*, *IL2RB*, *IL10RA*, *ITK*, *KIR2DL3*, *KIR3DL1*, *KLRD1*, *LAPTM5*, *LAT2*, *LCP1*, *LPXN*, *P2RY13*, *PARVG*, *PCED1B*, *PITPNM2*, *PLBD1*, *PLCL2*, *PLEK*, *PPM1M*, *PREX1*, *PRKCB*, *PRKCH*, *PSMB10*, *PVRIG*, *RASSF5*, *RFTN1*, *S1PR4*, *SASH3*, *SEL1L*, *SLAMF7*, *STAT4*, *TRAT1*, *VAV1*, *VCAM1*, *ZBED2*, and *ZNF831*. Therefore, it is thought that patients with low risk TNBC have stronger immunogenicity compared to patients with high risk TNBC. If patients with low $CD8^+$ T-cell infiltration in the tumor microenvironment do not respond to immune checkpoint inhibitors (22, 23), then our results suggest that patients with high risk TNBC who have low expression of the genes mentioned above at the time of

diagnosis may not benefit from immune checkpoint inhibitors as they are currently available. Also, it is unclear whether patients with low-risk TNBC will have an improved OS with immune checkpoint inhibitors or whether they will have a higher survival rate than patients with high-risk TNBC because they are low-risk patients to begin with. In any case, patients with low $CD8^+$ T cell infiltration are molecularly characterized by low expression of these genes.

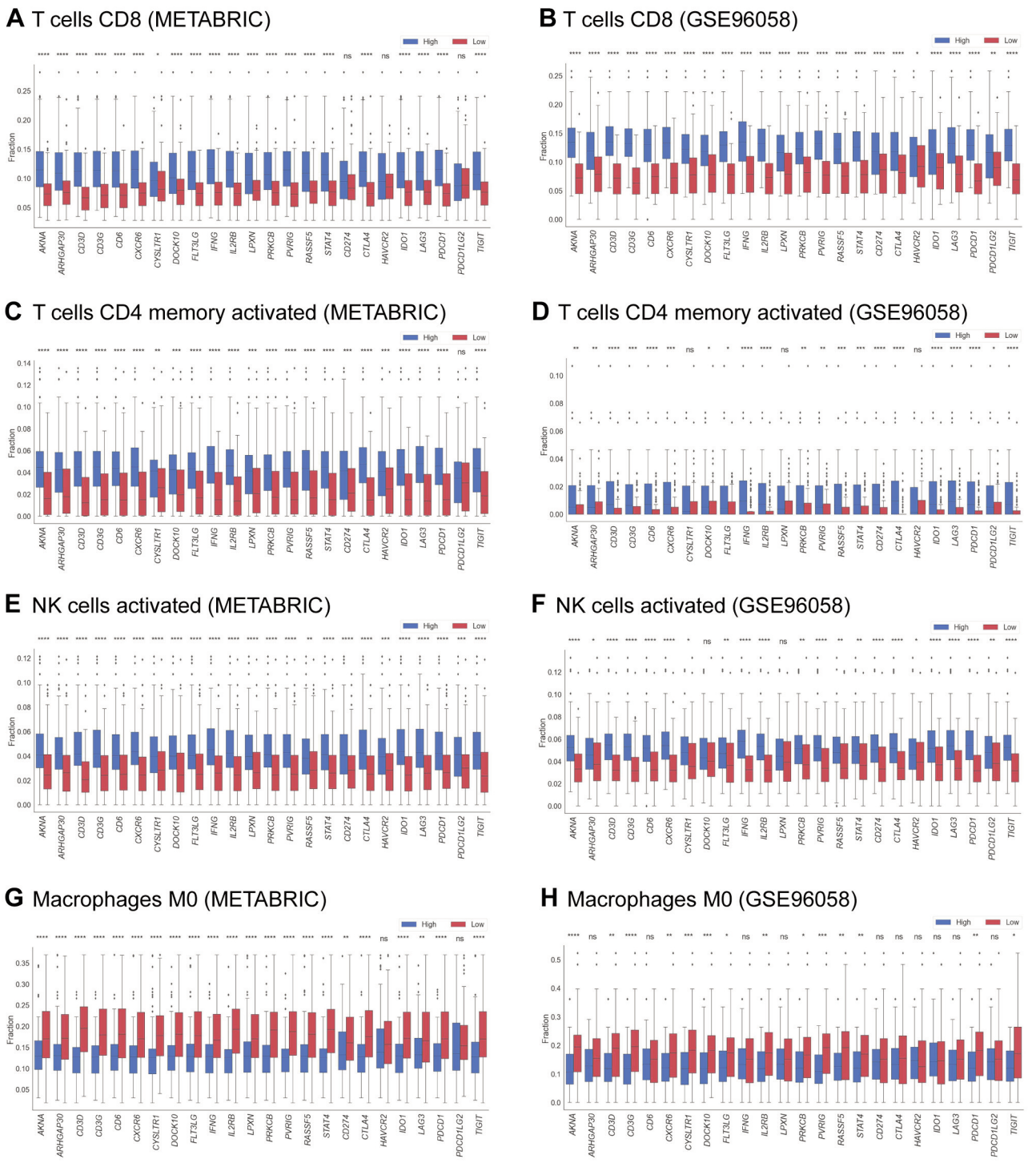

Figure 6. *Continued*

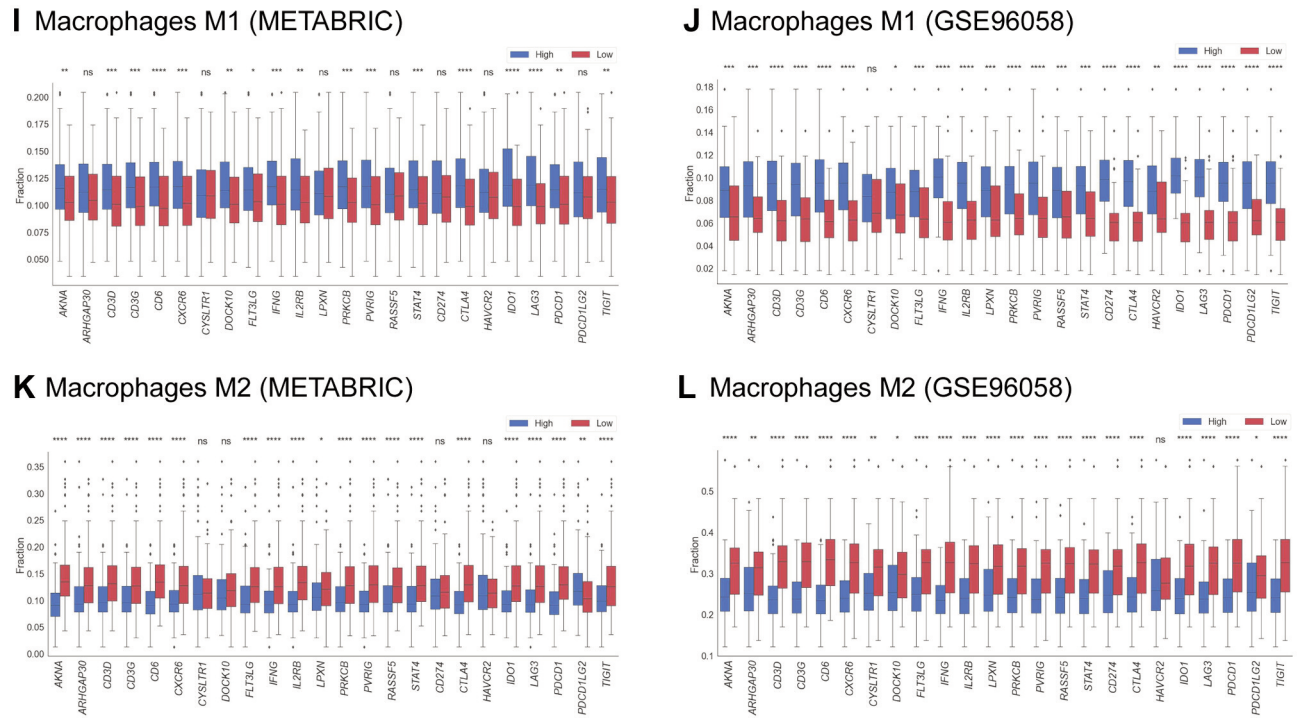

Figure 6. *Comparison of immune cell infiltration between the high- and low-expression groups for each gene in METABRIC (A, C, E, G, I, K) and GSE (B, D, F, H, J, L). Statistical significance was evaluated using the Mann-Whitney U-test (ns, not significant; *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001).*

Within the tumor microenvironment, macrophages exist in an immunosuppressed state, preventing T cells from removing the tumor (24). Macrophages enter the tumor microenvironment in the naïve state (M0) (24). Within the tumor microenvironment, macrophages are called tumor-associated macrophages (TAMs) and are associated with poor outcome for cancer patients (25, 26). Macrophages are generally defined as two extremes: classically activated M1 (immune-promoting) and alternatively activated M2 (immunosuppressive) (24, 27). Most TAMs are in an M2-like state and further promote tumor growth (24). *PDCD1* expression tends to support M2 polarization of TAMs, which may enhance the immune escape of tumor cells (28). We found that M0 and/or M2 infiltration was higher and M1 infiltration was lower in the low expression groups of genes, such as *AKNA*, *APOBEC3G*, *ARHGAP30*, *etc*. mentioned in the CD8 T cell infiltration section than in the high expression groups. In addition to M2 macrophages, M0 macrophages also appear to be important in promoting TNBC progression and may be responsible for the poor clinical outcome of the low expression group of the above genes. These genes are at least somehow involved in causing higher or lower immune cell infiltrate in TNBC. For TNBC patients who are at high risk of recurrence and death, it is necessary to activate these genes, reduce the immunosuppressive ability of macrophages, and promote $CD8^+$ T cell function.

In particular, 18 genes (*AKNA*, *ARHGAP30*, *CA9*, *CD3D*, *CD3G*, *CD6*, *CXCR6*, *CYSLTR1*, *DOCK10*, *ENO1*, *FLT3LG*, *IFNG*, *IL2RB*, *LPXN*, *PRKCB*, *PVRIG*, *RASSF5*, and *STAT4*) differentially expressed in both the microarray and RNA-seq datasets are strong biomarkers for predicting TNBC progression. Based on the results of the METABRIC dataset comparing OS and RFS between the high and low scoring groups for the 18-gene expression signature and the low survival rate of patients with early recurrent TNBC, it is highly likely that not only OS but also RFS will differ between these two groups in the GSE dataset. In addition, 16 of these genes showed significant differences in the level of immune cell infiltration in the tumor microenvironment between the high- and low-expression groups. The risk of early recurrence and the degree of immune cell infiltration in patients with TNBC can be predicted by this gene expression signature. In both data sets, TNBC patients with high expression of *CA9* and *ENO1* had a higher risk of death than those with low expression. A hypoxic tumor microenvironment induces the expression of the *CA9* gene (29, 30). Carbonic anhydrase IX (CAIX), encoded by *CA9*, is known to be a hypoxia-induced enzyme that regulates tumor pH and facilitates tumor cell migration and invasion (31, 32). Facilitation of cancer cell survival by tumor hypoxia can reduce the effectiveness of other tumor therapies, such as radiotherapy, chemotherapy, and immunotherapy and

increases metastasis risk that may facilitate patient mortality (33, 34). Since CAIX is almost exclusively expressed in cancer cells, it has attracted much attention as a cancer-specific therapeutic target (35). Alpha enolase (ENO1), encoded by *ENO1*, is a glycolytic enzyme that catalyzes the conversion of 2-phosphoglyceric acid to phosphoenolpyruvic acid during glycolysis (36). ENO1 promotes cellular functions associated with tumor progression, including enhanced glycolysis, cancer cell proliferation, migration, invasion, drug resistance, and activation of oncogenic signaling pathways (36). Overexpression of ENO1 has been established in a wide range of human cancers and is often associated with poor clinical outcomes (36-38). It has also been reported that down-regulation of ENO1 activity suppresses the glycolytic activity of tumor-infiltrating $CD8^+$ lymphocytes, leading to their functional depletion (39). Due to its localization at the tumor surface, ENO1 is a useful prognostic and diagnostic cancer biomarker as well as a potential cancer therapeutic target (36). On the other hand, as mentioned in the first half of the discussion, the 16 genes excluding *CA9* and *ENO1* are immune-related genes. These genes, whose low expression is associated with a higher risk of death than high expression, may be more effective biomarkers for identifying TNBC patients who will and will not benefit from immunotherapy, and may be particularly important genes for future treatment strategies, including immunotherapy. The identification of molecular factors associated with clinical outcomes in patients with TNBC could allow early detection of high-risk patients and provide potential therapeutic targets for effective clinical management of TNBC.

Supplementary Material

Available at: https://github.com/satolab-tus/Supplementary-Material

Conflicts of Interest

The Authors declare no potential conflicts of interest.

Authors' Contributions

KS contributed to the study concept and design, data acquisition and analysis, and data interpretation and discussion of results, and drafting and critical revision of the manuscript. KM contributed to data acquisition and analysis, and data interpretation and discussion of the results, and critical revision of the manuscript. ST and KA contributed to study concept and design, data interpretation and discussion of the results, and critical revision of the manuscript. All Authors approved the final version of the manuscript.

References

1 Won KA, Spruck C: Triple negative breast cancer therapy: Current and future perspectives (Review). Int J Oncol 57(6): 1245-1261, 2020. DOI: 10.3892/ijo.2020.5135

2 Núñez Abad M, Calabuig-Fariñas S, Lobo de Mena M, José Godes Sanz de Bremond M, García González C, Torres Martínez S, García-García JÁ, Iranzo González-Cruz V, Camps Herrero C: Update on systemic treatment in early triple negative breast cancer. Ther Adv Med Oncol 13: 1758835920986749, 2021. DOI: 10.1177/1758835920986749

3 Cravero K, Pantone MV, Shin DH, Bergman R, Cochran R, Chu D, Zabransky DJ, Karthikeyan S, Waters IG, Hunter N, Rosen DM, Kyker-Snowman K, Dalton WB, Button B, Shinn D, Wong HY, Donaldson J, Hurley PJ, Croessmann S, Park BH: NOTCH1 PEST domain variants are responsive to standard of care treatments despite distinct transformative properties in a breast cancer model. Oncotarget 13: 373-386, 2022. DOI: 10.18632/oncotarget.28200

4 Maqbool M, Bekele F, Fekadu G: Treatment strategies against triple-negative breast cancer: an updated review. Breast Cancer (Dove Med Press) 14: 15-24, 2022. DOI: 10.2147/BCTT.S348060

5 Wang Z, Jiang Q, Dong C: Metabolic reprogramming in triple-negative breast cancer. Cancer Biol Med 17(1): 44-59, 2020. DOI: 10.20892/j.issn.2095-3941.2019.0210

6 Boyle P: Triple-negative breast cancer: epidemiological considerations and recommendations. Ann Oncol 23 Suppl 23: vi7-vi12, 2012. DOI: 10.1093/annonc/mds187

7 Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, Lickley LA, Rawlinson E, Sun P, Narod SA: Triple-negative breast cancer: Clinical features and patterns of recurrence. Clin Cancer Res 13(15 Pt 1): 4429-4434, 2007. DOI: 10.1158/1078-0432.CCR-06-3045

8 Ferrari P, Scatena C, Ghilli M, Bargagna I, Lorenzini G, Nicolini A: Molecular mechanisms, biomarkers and emerging therapies for chemotherapy resistant TNBC. Int J Mol Sci 23(3): 1665, 2022. DOI: 10.3390/ijms23031665

9 Gupta GK, Collier AL, Lee D, Hoefer RA, Zheleva V, Siewertsz van Reesema LL, Tang-Tan AM, Guye ML, Chang DZ, Winston JS, Samli B, Jansen RJ, Petricoin EF, Goetz MP, Bear HD, Tang AH: Perspectives on triple-negative breast cancer: current treatment strategies, unmet needs, and potential targets for future therapies. Cancers (Basel) 12(9): 2392, 2020. DOI: 10.3390/cancers12092392

10 Li L, Zhang F, Liu Z, Fan Z: Immunotherapy for triple-negative breast cancer: combination strategies to improve outcome. Cancers (Basel) 15(1): 321, 2023. DOI: 10.3390/cancers15010321

11 Disis ML, Stanton SE: Triple-negative breast cancer: immune modulation as the new treatment paradigm. Am Soc Clin Oncol Educ Book (35): e25-e30, 2015. DOI: 10.14694/EdBook_AM.2015.35.e25

12 Lotfinejad P, Asghari Jafarabadi M, Abdoli Shadbad M, Kazemi T, Pashazadeh F, Sandoghchian Shotorbani S, Jadidi Niaragh F, Baghbanzadeh A, Vahed N, Silvestris N, Baradaran B: Prognostic role and clinical significance of tumor-infiltrating lymphocyte (TIL) and programmed death ligand 1 (PD-L1) expression in triple-negative breast cancer (TNBC): a systematic review and meta-analysis study. Diagnostics (Basel) 10(9): 704, 2020. DOI: 10.3390/diagnostics10090704

13 Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N: The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2(5): 401-404, 2012. DOI: 10.1158/2159-8290.CD-12-0095

14 Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N: Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 6(269): pl1, 2013. DOI: 10.1126/scisignal.2004088
15 Lee AHS, Ellis IO: The Nottingham Prognostic Index for invasive carcinoma of the breast. Pathol Oncol Res 14(2): 113-115, 2008. DOI: 10.1007/s12253-008-9067-3
16 Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK: Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 10(1): 1523, 2019. DOI: 10.1038/s41467-019-09234-6
17 Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, Khodadoust MS, Esfahani MS, Luca BA, Steiner D, Diehn M, Alizadeh AA: Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol 37(7): 773-782, 2019. DOI: 10.1038/s41587-019-0114-2
18 Shirinbak S, Chan RY, Shahani S, Muthugounder S, Kennedy R, Hung LT, Fernandez GE, Hadjidaniel MD, Moghimi B, Sheard MA, Epstein AL, Fabbri M, Shimada H, Asgharzadeh S: Combined immune checkpoint blockade increases $CD8^+CD28^+PD\text{-}1^+$ effector T cells and provides a therapeutic strategy for patients with neuroblastoma. Oncoimmunology 10(1): 1838140, 2021. DOI: 10.1080/2162402X.2020.1838140
19 Kumar S, Singh SK, Rana B, Rana A: Tumor-infiltrating CD8(+) T cell antitumor efficacy and exhaustion: molecular insights. Drug Discov Today 26(4): 951-967, 2021. DOI: 10.1016/j.drudis.2021.01.002
20 Mahmoud SM, Paish EC, Powe DG, Macmillan RD, Grainge MJ, Lee AH, Ellis IO, Green AR: Tumor-infiltrating $CD8^+$ lymphocytes predict clinical outcome in breast cancer. J Clin Oncol 29(15): 1949-1955, 2011. DOI: 10.1200/JCO.2010.30.5037
21 Maimela NR, Liu S, Zhang Y: Fates of $CD8^+$ T cells in tumor microenvironment. Comput Struct Biotechnol J 17: 1-13, 2018. DOI: 10.1016/j.csbj.2018.11.004
22 Lao Y, Shen D, Zhang W, He R, Jiang M: Immune checkpoint inhibitors in cancer therapy-how to overcome drug resistance? Cancers (Basel) 14(15): 3575, 2022. DOI: 10.3390/cancers14153575
23 Zheng Z, Wieder T, Mauerer B, Schäfer L, Kesselring R, Braumüller H: T cells in colorectal cancer: unravelling the function of different T cell subsets in the tumor microenvironment. Int J Mol Sci 24(14): 11673, 2023. DOI: 10.3390/ijms241411673
24 Cess CG, Finley SD: Multi-scale modeling of macrophage-T cell interactions within the tumor microenvironment. PLoS Comput Biol 16(12): e1008519, 2020. DOI: 10.1371/journal.pcbi.1008519
25 Yang L, Zhang Y: Tumor-associated macrophages: from basic research to clinical application. J Hematol Oncol 10(1): 58, 2017. DOI: 10.1186/s13045-017-0430-2
26 Chanmee T, Ontong P, Konno K, Itano N: Tumor-associated macrophages as major players in the tumor microenvironment. Cancers (Basel) 6(3): 1670-1690, 2014. DOI: 10.3390/cancers6031670
27 Najafi M, Hashemi Goradel N, Farhood B, Salehi E, Nashtaei MS, Khanlarkhani N, Khezri Z, Majidpoor J, Abouzaripour M, Habibi M, Kashani IR, Mortezaee K: Macrophage polarity in cancer: A review. J Cell Biochem 120(3): 2756-2765, 2019. DOI: 10.1002/jcb.27646
28 Li W, Wu F, Zhao S, Shi P, Wang S, Cui D: Correlation between PD-1/PD-L1 expression and polarization in tumor-associated macrophages: A key player in tumor immunotherapy. Cytokine Growth Factor Rev 67: 49-57, 2022. DOI: 10.1016/j.cytogfr.2022.07.004
29 de la Cruz-López KG, Castro-Muñoz LJ, Reyes-Hernández DO, García-Carrancá A, Manzo-Merino J: Lactate in the regulation of tumor microenvironment and therapeutic approaches. Front Oncol 9: 1143, 2019. DOI: 10.3389/fonc.2019.01143
30 Sarnella A, Ferrara Y, Auletta L, Albanese S, Cerchia L, Alterio V, De Simone G, Supuran CT, Zannetti A: Inhibition of carbonic anhydrases IX/XII by SLC-0111 boosts cisplatin effects in hampering head and neck squamous carcinoma cell growth and invasion. J Exp Clin Cancer Res 41(1): 122, 2022. DOI: 10.1186/s13046-022-02345-x
31 Kajanova I, Zatovicova M, Jelenska L, Sedlakova O, Barathova M, Csaderova L, Debreova M, Lukacikova L, Grossmannova K, Labudova M, Golias T, Svastova E, Ludwig A, Muller P, Vojtesek B, Pastorek J, Pastorekova S: Impairment of carbonic anhydrase IX ectodomain cleavage reinforces tumorigenic and metastatic phenotype of cancer cells. Br J Cancer 122(11): 1590-1603, 2020. DOI: 10.1038/s41416-020-0804-z
32 Daunys S, Petrikaitė V: The roles of carbonic anhydrases IX and XII in cancer cell adhesion, migration, invasion and metastasis. Biol Cell 112(12): 383-397, 2020. DOI: 10.1111/boc.201900099
33 Zhuang Y, Liu K, He Q, Gu X, Jiang C, Wu J: Hypoxia signaling in cancer: Implications for therapeutic interventions. MedComm (2020) 4(1): e203, 2023. DOI: 10.1002/mco2.203
34 Liao C, Liu X, Zhang C, Zhang Q: Tumor hypoxia: From basic knowledge to therapeutic implications. Semin Cancer Biol 88: 172-186, 2023. DOI: 10.1016/j.semcancer.2022.12.011
35 McDonald PC, Chafe SC, Supuran CT, Dedhar S: Cancer therapeutic targeting of hypoxia induced carbonic anhydrase IX: from bench to bedside. Cancers (Basel) 14(14): 3297, 2022. DOI: 10.3390/cancers14143297
36 Huang CK, Sun Y, Lv L, Ping Y: ENO1 and cancer. Mol Ther Oncolytics 24: 288-298, 2022. DOI: 10.1016/j.omto.2021.12.026
37 Chelakkot C, Chelakkot VS, Shin Y, Song K: Modulating glycolysis to improve cancer therapy. Int J Mol Sci 24(3): 2606, 2023. DOI: 10.3390/ijms24032606
38 Almaguel FA, Sanchez TW, Ortiz-Hernandez GL, Casiano CA: Alpha-Enolase: Emerging tumor-associated antigen, cancer biomarker, and oncotherapeutic target. Front Genet 11: 614726, 2021. DOI: 10.3389/fgene.2020.614726
39 Gemta LF, Siska PJ, Nelson ME, Gao X, Liu X, Locasale JW, Yagita H, Slingluff CL Jr, Hoehn KL, Rathmell JC, Bullock TNJ: Impaired enolase 1 glycolytic activity restrains effector functions of tumor-infiltrating CD8(+) T cells. Sci Immunol 4(31): eaap9520, 2019. DOI: 10.1126/sciimmunol.aap9520

Received December 29, 2023
Revised February 26, 2024
Accepted March 6, 2024