Am J Cancer Res 2021;11(10):5094-5110
www.ajcr.us /ISSN:2156-6976/ajcr0135465

Original Article

Conflicting roles of *EGFR* expression by subtypes in breast cancer

Masanori Oshi[1,2*], Shipra Gandhi[3*], Yoshihisa Tokumaru[1,4], Li Yan[5], Akimitsu Yamada[2], Ryusei Matsuyama[2], Takashi Ishikawa[9], Itaru Endo[2], Kazuaki Takabe[1,2,3,6,7,8,9]

[1]Department of Surgical Oncology, Roswell Park Comprehensive Cancer Center, Buffalo, New York 14263, USA; [2]Department of Gastroenterological Surgery, Yokohama City University Graduate School of Medicine, Yokohama 236-0004, Japan; [3]Department of Medical Oncology, Roswell Park Comprehensive Cancer Center, Elm & Carlton Streets, Buffalo, NY 14263, USA; [4]Department of Surgical Oncology, Graduate School of Medicine, Gifu University, 1-1 Yanagido, Gifu 501-1194, Japan; [5]Department of Biostatistics & Bioinformatics, Roswell Park Comprehensive Cancer Center, Buffalo, New York 14263, USA; [6]Division of Digestive and General Surgery, Niigata University Graduate School of Medical and Dental Sciences, Niigata 951-8520, Japan; [7]Department of Breast Surgery, Fukushima Medical University School of Medicine, Fukushima 960-1295, Japan; [8]Department of Surgery, Jacobs School of Medicine and Biomedical Sciences, State University of New York, Buffalo, New York 14263, USA; [9]Department of Breast Surgery and Oncology, Tokyo Medical University, Tokyo 160-8402, Japan. [*]Equal contributors.

Received May 20, 2021; Accepted August 31, 2021; Epub October 15, 2021; Published October 30, 2021

Abstract: Epidermal growth factor receptor (*EGFR*) is one of the receptors that belong to the epidermal growth factor family of receptor tyrosine kinases (ErbBs). Several malignancies including breast cancer that express EGFR have poor prognosis. Our study examined the EGFR expression among 5176 breast cancer patients from GSE96058 and METABRIC cohorts and the contribution of tumor immune microenvironment in different subtypes. We found that among different breast cancer subtypes, *EGFR* expression in TNBC was the highest compared to other subtypes. *EGFR* high ER-positive/HER2-negative breast cancer had significantly higher survival compared to *EGFR* low ER-positive/HER2-negative breast cancer. It was also associated with high level of intratumor heterogeneity and homologous recombination defects (HRD). This group was also enriched in immune-related gene sets. On the other hand, low *EGFR* tumor was enriched in cell proliferation-related gene sets. However, these findings were not observed in *TNBC.* Interestingly, there was a greater infiltration of anti-cancer immune cells in high *EGFR* ER-positive/HER2-negative breast cancers, while, TNBC with higher EGFR expression had lower fraction of immune cells along with low level of cytolytic activity. Tumor cells have significantly higher *EGFR* expression compared to immune cells in single cell sequencing data. There was higher expression of immune checkpoint molecules in high *EGFR* ER-positive/HER2-negative breast cancer but lower expression in TNBC. High *EGFR* metastatic tumor was significantly associated with worse survival, but no association with infiltrating immune cells was observed. Our study shows that higher *EGFR* expression in ER-positive/HER2-negative breast cancer is associated with improved outcomes and an anti-cancer immune microenvironment.

Keywords: Breast cancer, cytolytic activity, EGFR, subtype, immune cell, tumor immune microenvironment, metastasis, survival analysis, triple negative breast cancer

Introduction

The epidermal growth factor receptor belongs to the epidermal growth factor family of receptor tyrosine kinases (ErbBs). This receptor is responsible for regulating vital functions in epithelial cell physiology [1]. In several malignancies, the EGFR receptor could be overexpressed or mutated. This receptor is also the target of several FDA approved cancer therapies [2]. This ErbB family also includes several other members like HER2, HER3 and HER4. Homo- or heterodimerization of these receptors results in their activation and this initiates further signaling downstream [3].

Overexpression of EGFR is associated with poor clinical outcomes in several malignancies, such

as, head and neck cancer [4], non-small cell lung carcinoma (NSCLC) [5], colorectal carcinoma [6] and endometrial cancer [7]. Other than a higher expression of the receptor, several tumors have other mechanisms whereby somatic mutations of the *EGFR* receptor results in constitutive kinase activity, as observed in NSCLC. First generation tyrosine kinase inhibitors (TKIs) such as erlotinib, gefitinib and second generation TKIs such as afatinib are currently Food and Drug Administration (FDA) approved for the treatment of NSCLC with EGFR somatic activating mutation. This is based on a head-to-head comparison showing improved progression free survival with these TKIs versus chemotherapy alone in prospective clinical trials. United States FDA has approved osimertinib for the first-line treatment of metastatic NSCLC with EGFR exon 19 or exon 21 L858R mutations. Additionally, osimertinib has also shown efficacy in inhibiting T790M mutations which is a secondary activating mutation that develops as a result of use of earlier-generation TKIs [8]. In addition to these activating mutations, FDA has also approved cetuximab and panitumumab, two antibodies blocking EGFR ligand for the treatment of colorectal and head and neck carcinomas.

EGFR has also been studied extensively in breast cancer. *EGFR* overexpression has been associated with an aggressive tumor biology as observed in preclinical models of breast cancer and also with an aggressive phenotype in human breast cancer. In the preclinical models, higher *EGFR* expression predicts for increased tumor proliferation, resistance to apoptosis and also results in epithelial-to-mesenchymal transition. At the same time, *EGFR* overexpression in human breast cancer has been associated with clinical presentation with large tumor size, poor differentiation and worse clinical outcomes [9, 10]. Almost 15-45% breast cancers overexpress *EGFR* with highest expression reported in inflammatory breast cancer and triple negative breast cancer (TNBC). The *EGFR* expression is observed to decrease with increasing hormone receptor expression [11-13].

It has been observed the *EGFR* expression is associated with presence of certain epitopes that are able to induce both adaptive and innate immune response in patients with NSCLC [14]. There is a plethora of data on the direct role of *EGFR* in mediating EMT, however, the interplay between EGFR expression and immune profile in different subtypes of breast cancer has not been investigated. We hypothesized that there is a complex interaction between the *EGFR* expression and immune cells which may potentially play a role in influencing outcomes.

Materials and methods

Data acquisition of breast cancer patients

Data was obtained for 3,273 breast cancer patients from GSE96058 cohort who had transcriptome profiling of resected tumors in an ongoing study. Publicly available clinical data of these patients was obtained from resources that are noted in a recent The Sweden Cancerome Analysis Network-Breast (SCAN-B) study [15]. Molecular Taxonomy of Breast Cancer International Constium (METABRIC) cohort was utlized to obtain both data on clinical information and gene expression from 1,903 breast cancer patients. The Cancer Genome Atlas (TCGA) cohort was utilized to obtain data on EGFR expression and breast cancer subtype from 1,069 female breast cancer patients [16]. Single-cell RNA-sequence data was obtained on tumor cells, stromal cells, immune cells, and myeloid cells in breast cancer from GSE75688 cohort [17], which was obtained from Gene Expression Omnibus.

Other scores

We used several scores, (1) xCell scores as immune cells score calculated by xCell algorithm [18], (2) Cytolytic activity (CYT) was calculated using two genes, granzyme A and perforin, established by Rooney et al. [19], and (3) Intratumor heterogeneity, homologous recombination defects (HRDs), and mutation- and neoantigen-related score; silent and non-silent mutation rate, fraction altered, single-nucleotide variation (SNV) and indel neoantigens, calculated by Thorsson et al. [20] in the TCGA cohort, as we previously reported [21-26].

Gene set enrichment analysis (GSEA)

GSEA [27] algorithm was used to explore the difference in biological function between two levels of *EGFR* expression groups with MSigDB Hallmark gene sets [28]. As recommended by GSEA, the statistical significance was determined using False Discovery Rate (FDR) of less

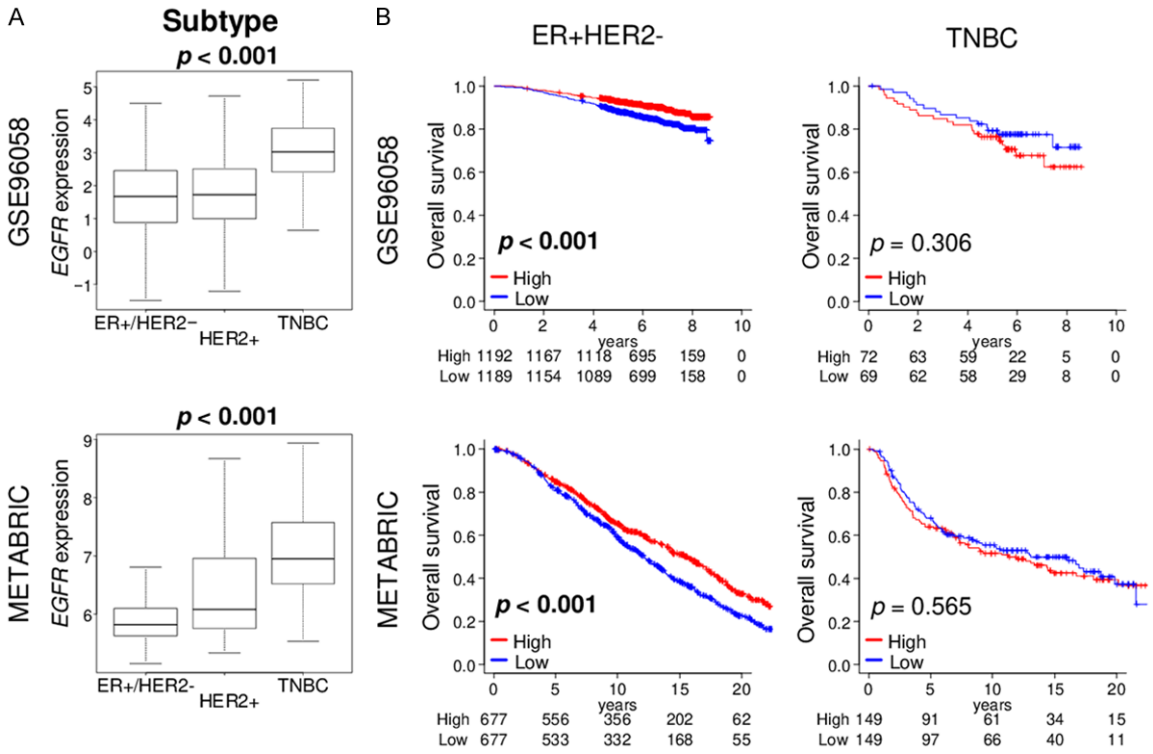

Figure 1. Association of *EGFR* expression with breast cancer subtype in the GSE96058 and METABRIC cohorts. A. Boxplots of *EGFR* expression by breast cancer subtypes; ER-positive/HER2-negative (ER+/HER2-), HER2-positive (HER2+), and TNBC. Kruskal-Wallis test was used to calculate the depicted *P*-values here. B. Kaplan-Meier survival curves with *P*-values calculated using log-rank test shown here for overall survival (OS) both in *EGFR* high and low groups among ER-positive/HER2-negative breast cancer and TNBC subtypes. High and low *EGFR* groups are defined by the median value of *EGFR* expression within each cohort used as the cut-off here. ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; TNBC, triple-negative breast cancer; OS, overall survival.

than 25%, as has been previously reported by our group [29-33].

Other statistical analyses

R software was used to do all statistical analysis in this study. The depicted *P* values were calculated using Kruskal-Wallis, Mann-Whitney U, or Fisher test in group comparison analysis, and log-rank test in survival analysis. Statistical significance was indicated by *P*-value less than 0.05.

Results

Better survival with higher EGFR expression was observed in estrogen receptor (ER)-positive/human epidermal growth factor receptor 2 (HER2)-negative breast cancer, but EGFR expression did not change outcomes in triple-negative breast cancer (TNBC)

Breast cancer is known to have lower *EGFR* level than other cancers. In fact, when comparing *EGFR* expression among cancers using the TCGA data, the *EGFR* expression in breast cancer tended to be lower than in other cancers (Figure S1). Since breast cancer is known to have different biological characteristics among subtypes, we expected that *EGFR* expression would also be different among breast cancer subtypes. GSE96058 ($n = 3273$) and METABRIC ($n = 1904$) cohorts were used to examine the relationship between *EGFR* expression and breast cancer subtypes in this study. As expected and consistent with the known literature, TNBC was associated with the highest *EGFR* expression compared to other breast cancer subtypes consistently in both the cohorts (**Figure 1A**; both $P<0.001$). The role of *EGFR* expression was next tested by examining the survival using the patient data from these cohorts in each breast cancer subtype. High and low EGFR groups within each cohort were defined using the median value as a cut-off. Among ER-positive/HER2-negative breast cancer, high expression of *EGFR* was associated

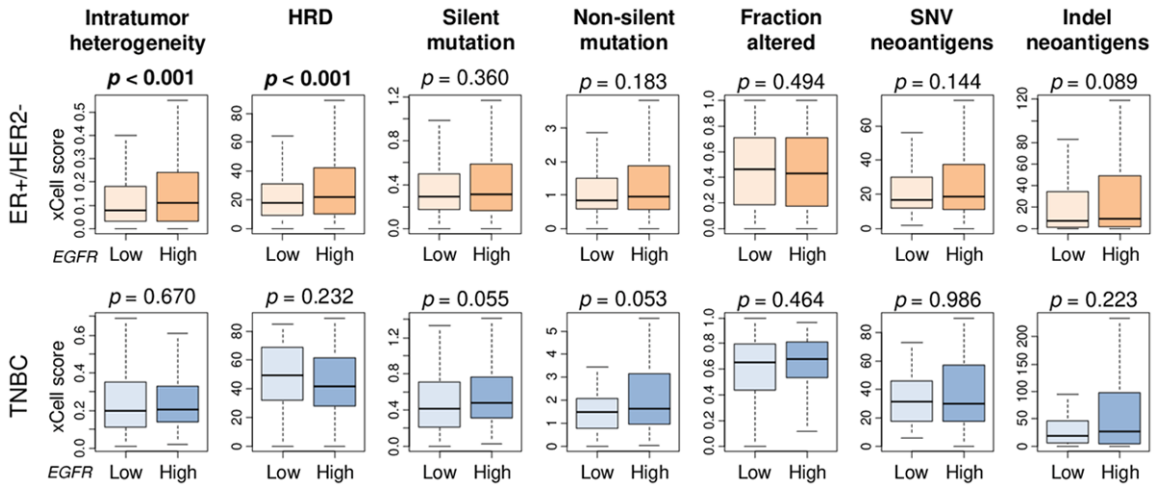

Figure 2. Association of *EGFR* expression with mutation-related score in each breast cancer subtype in TCGA cohort. Boxplots showing homologous recombination deficits (HRD), intratumoral heterogeneity, mutation scores, neoantigens and fraction altered in two breast cancer subtypes, ER-positive/HER2-negative breast cancer (upper) and in TNBC (lower) divided by high and low EGFR expression. The median value of *EGFR* expression within each cohort was used to classify them into high and low groups. Mann-Whitney U test was used to calculate the depicted *P*-values.

with better survival in the GSE96058 cohort, but this difference was not observed in TNBC (**Figure 1B**; *P*<0.001, and *P* = 0.306, respectively). METABRIC cohort was used as the validation cohort and these results were also consistent in that cohort (*P*<0.001 and *P* = 0.565, respectively). Thus, high *EGFR* expression did show statistically significant survival improvement in the ER-positive/HER2-negative breast cancer, but not in TNBC.

High EGFR expression was associated with both homologous recombination deficiency (HRD) and with intratumoral heterogeneity in ER-positive/HER2-negative breast cancer but not in TNBC

We have previously reported that advanced breast cancer has high mutation [34, 35]. Given the knowledge, we expected that high *EGFR* expression, which was associated with better survival as shown in **Figure 1B**, would tend to associate with low mutation level in ER-positive/HER2-negative breast cancer. In contrast to our expectation, we observed that high *EGFR* expression was associated with high score of intratumor heterogeneity and HRD in ER-positive/HER2-negative breast cancer (**Figure 2**, both *P*<0.001), but there was no association observed with other mutation scores, such as, mutation-related score; silent and non-silent mutation (*P* = 0.360 and 0.183), fraction altered (*P* = 0.494), single nucleotide variation (SNV) and indel neoantigens (*P* = 0.144 and 0.089). On the other hand, among TNBC, we did not observe any significant association between *EGFR* expression or intratumoral heterogeneity, HRD, mutation or neoantigen scores (**Figure 2**).

A high EGFR tumor enriched several pro-cancer-related gene sets and a low EGFR tumor enriched cell proliferation-related gene sets in ER-positive/HER2-negative breast cancer, but not in TNBC

Since EGFR overexpression is associated with different clinical outcomes in different breast cancer subtypes, therefore, we expected that the biological function of EGFR expression would vary by subtypes. We found that high EGFR ER-positive/HER2-negative tumor was enriched in several hallmark gene sets, such as hypoxia, angiogenesis, notch signaling, KRAS signaling, and epithelial mesenchymal transition (EMT) in the GSE96058 cohort (**Figure 3A**). Consistent results were observed in the validation cohort, METABRIC (Figure S3A). On the other hand, high EGFR TNBC was enriched in protein secretion, androgen response, estrogen response early, glycolysis, PI3K/AKT/MTOR signaling, and cholesterol homeostasis gene sets (**Figure 3A**). Interestingly, these results were also consistently validated by METABRIC cohort (Figure S3A).

Clinical relevance of *EGFR* expression in breast cancer subtypes

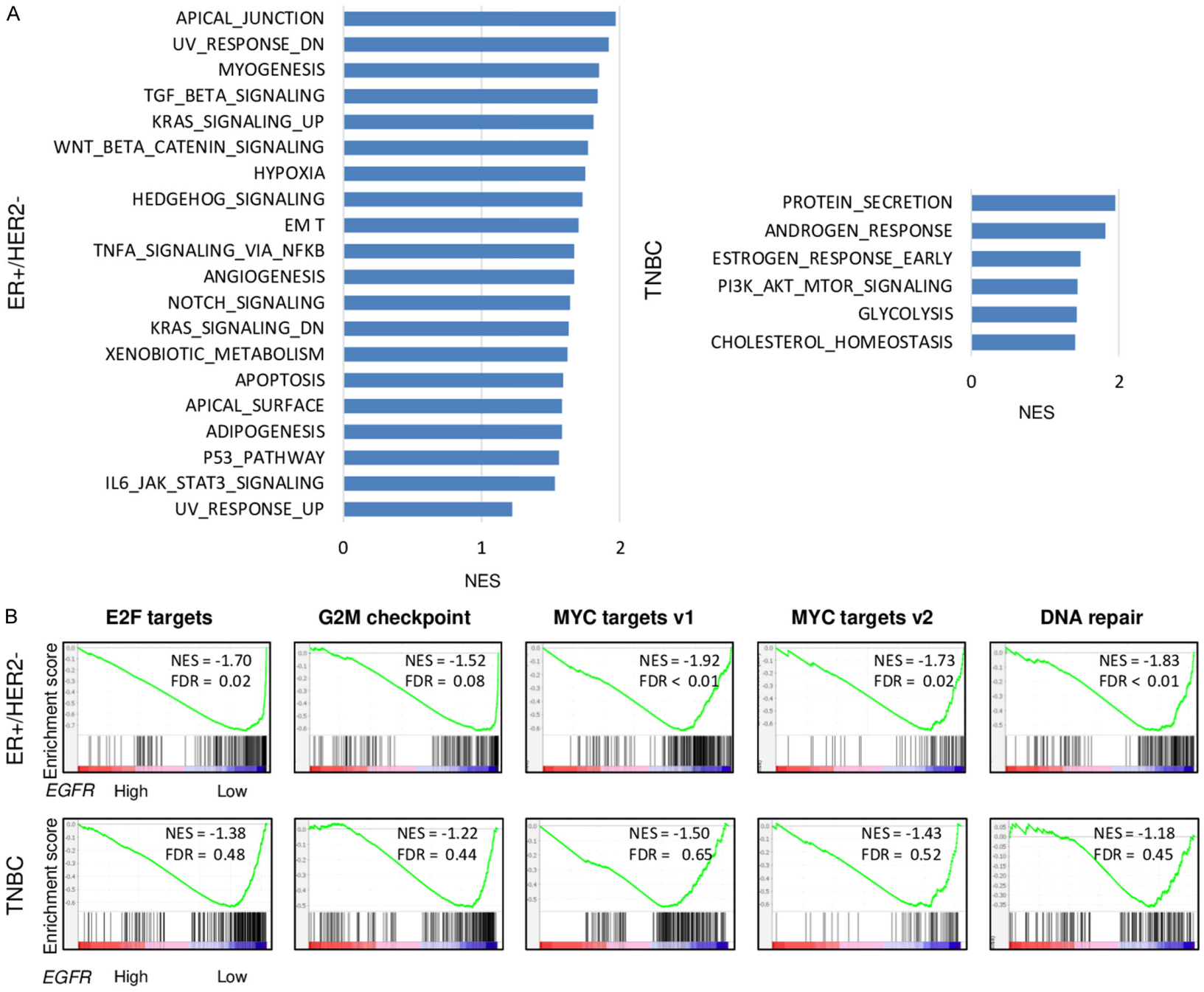

Figure 3. Association of *EGFR* expression with biological function in each breast cancer subtype in the GSE96058 cohort. A. Bar plots showing significantly enriched gene sets in the high *EGFR* groups in each subtype consistently in both GSE96058 and METABRIC cohorts. B. Enrichment plots of cell proliferation-related gene sets; G2M checkpoint, E2F targets, and MYC targets v1 and v2, and DNA repair in ER+/HER2- and TNBC subtype. The median value of *EGFR* expression within each cohort was used to classify them into high and low groups.

Furthermore, low EGFR tumor was enriched in cell proliferation-related gene sets; G2M checkpoint, E2F targets, MYC target v1 and v2, and DNA repair in ER-positive/HER2-negative breast cancer, but not in TNBC (**Figure 3B**). Particulary the association of EGFR with E2F targets, MYC targets v2, and DNA repair gene sets, were also validated in the METABRIC cohort (Figure S2B). EGFR expression was confirmed to be significantly correlated with EGFR signaling pathway in GSE96058 cohort (Figure S2; Spearman's rank correlation coefficient (r) = 0.518, P<0.01). Acitivity of EGFR signaling pathway was quantified by gene set variation analysis (GSVA) algorithm utilizing the similar approach we have previously reported [36-41]. These results suggest that the association between low EGFR and poor clinical outcomes in ER-positive/HER2-negative breast cancer may be influenced by cell proliferation.

A high EGFR tumor enriched immune-related gene sets significantly in ER-positive/HER2-negative breast cancer, but not in TNBC

Since there is data showing that certain epitopes within *EGFR* expressed tumors are able to elicit both cellular and immune response, we investigated the distribution of immune related gene sets in EGFR high breast cancer [14]. As expected, a high *EGFR* ER+/HER2- tumor was enriched in immune-related gene sets, including allograft rejection (**Figure 4**, normal enrichment score (NES) = 1.40, false discovery rate (FDR) = 0.11), inflammatory response (NES = 1.61, FDR = 0.04), IL6/JAK/STAT3 signaling (NES = 1.53, FDR = 0.06), IL2/STAT5 signaling (NES = 1.70, FDR = 0.02), complement (NES = 1.60, FDR = 0.04) and coagulation (NES = 1.93, FDR<0.01), in ER-positive/HER2-negative breast cancer in GSE96058 cohort. METABRIC cohort was used as the validation cohort where consistent results were observed (Figure S4). On the other hand, low *EGFR* TNBC were enriched in immune-related gene sets, although this was not statistically significant (**Figure 4**). Consistent results were observed in the METABRIC cohort which was used to validate the results (Figure S4). Therefore, we observe that high *EGFR* ER-positive/HER2-negative tumor is enriched in immune-related gene sets, but similar results were not observed in TNBC.

High EGFR was signifiantly associated with high level of cytolytic activity in ER-positive/HER2-negative breast cancer, but with lower cytolytic activity in TNBC

Next, it was interesting to see which types of immune cells infiltrated the high *EGFR* tumor. We found that there was a significantly higher fraction of anti-cancer immune cells, namely, $CD8^+$ T cells, $CD4^+$ T cells, dendritic cells (DC), and lower fraction of T helper type 1 (Th1) cells and M1 macrophages, and pro-cancer immune cells; regulatory T cells (Tregs), and Th2 cells, in ER-positive/HER2-negative breast cancer (**Figure 5A** and **5B**). These results were validated consistently in METABRIC cohort (Figure S5). On the other hand, there were lower fraction of anti-cancer immune cells; $CD8^+$ T cells, $CD4^+$ T cells, and M1 macrophages and pro-cancer immune cells; Th2 and M2 macrophages, in *EGFR* high TNBC (**Figure 5A** and **5B**). Consistent results were observed in the validation cohort, METABRIC (Figure S5). Furthermore, high *EGFR* ER-positive/HER2-negative breast cancer was significantly associated with high level of cytolytic activity (CYT), on the other hand, opposite results were observed in high *EGFR* TNBC (**Figure 5C**, both *P*<0.001). These results were validated consistently in METABRIC cohort (Figure S5C). Furthermore, *EGFR* was highly expressed by tumor cells, but not immune cells (**Figure 5D**).

High EGFR expression showed significant association with high expression of immune checkpoint molecules in ER-positive/HER2-negative breast cancer, while, lower expression of these immune checkpoint molecules was observed in TNBC

There was higher expression of several immune checkpoint molecules, including Programmed cell death protein 1 (PD-1), Programmed death-

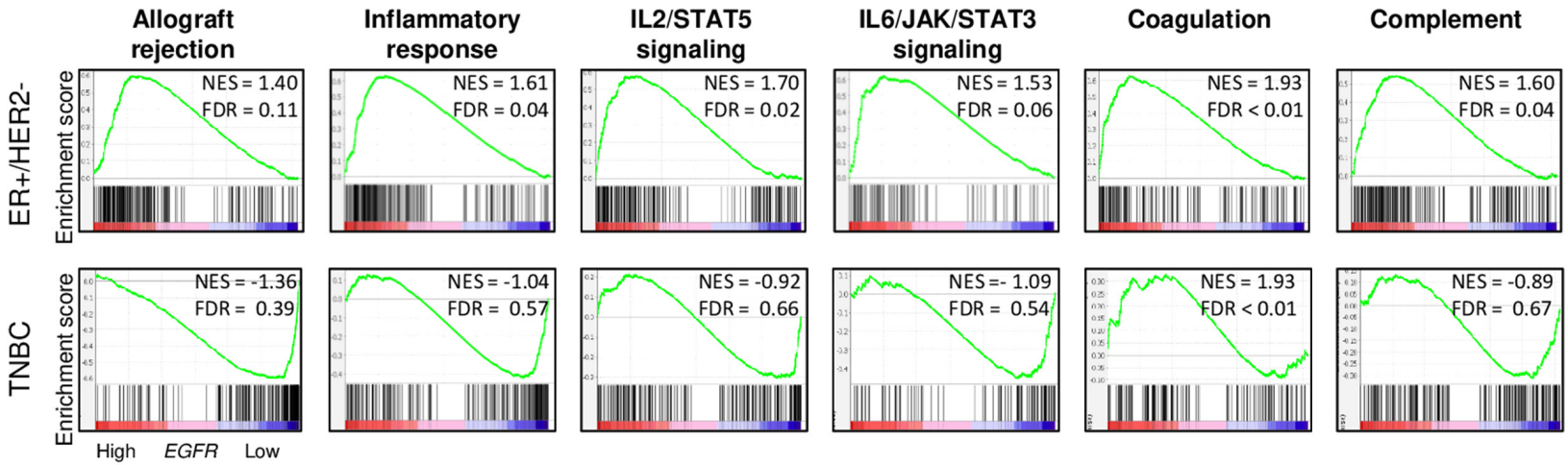

Figure 4. Association of *EGFR* expression with immune-related biological function in each breast cancer subtype in the GSE96058 cohort. Enrichment plots for immune-related gene sets for EGFR high and low expression within the breast cancer subtypes, ER-positive/HER2-negative breast cancer and TNBC. Immune-related gene sets consisting of allograft rejection, inflammatory response, IL6/JAK/STAT3 signaling, IL2/STAT5 signaling, complement and coagulation pathways. Median value of *EGFR* expression was used to divide the two groups within each cohort.

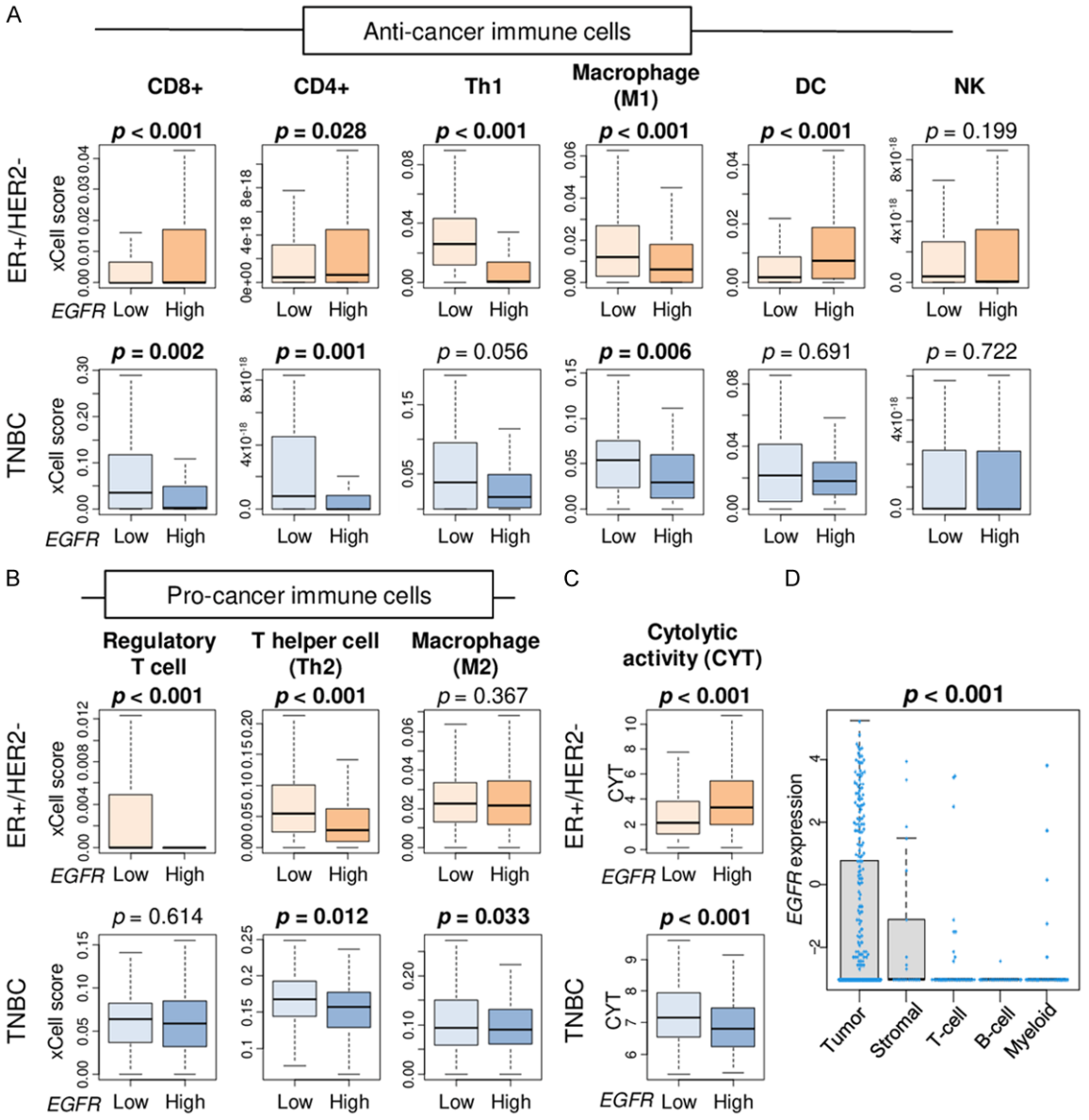

Figure 5. Association of *EGFR* expression with immune fraction and cytolytic activity in each breast cancer subtype in the GSE96058 cohort. (A) Boxplots showing the distribution of anti-cancer immune cells; $CD8^+$ T cells, $CD4^+$ T cells, Th1 cells, M1 macrophages, DC, and NK cells, and (B) pro-cancer immune cells; regulatory T cells, Th2 cells, and M2 macrophages, and (C) CYT score divided by high and low *EGFR* expression in each breast cancer subtype, namely, ER+/HER2- and TNBC. The median value of *EGFR* expression is used to divide into EGFR high and low groups within each cohort. Mann-Whitney U test is used to calculate the depicted *P*-values. (D) Relative *EGFR* expression by cells including tumor, stromal, T-cell, B-cell, and Myeloid in the GSE7688 cohort shown by boxplots. Kruskal-Wallis test is used to calculate the depicted *P*-values. CYT, cytolytic activity; DC, dendritic cells; NK, natural killer; Th1, T helper type1; Th2, T helper type2.

ligand 1 (PD-L1) and PD-L2, Cytotoxic T-lymphocyte-associated protein 4 (CTLA4), Indoleamine 2,3-dioxygenase 1 (IDO1), and B- and T-lymphocyte attenuator (BTLA), as well as inhibitory checkpoint index (IICI), in high EGFR ER-positive/HER2-negative breast cancer in GSE96058 cohort (**Figure 6**; all P<0.001). On the other hand, among TNBC, high EGFR expression was significantly associated with low expression of immune checkpoint molecules, and the level of IICI. These results of the association of EGFR with immune checkpoint molecules were validated in METABRIC cohort (Figure S6).

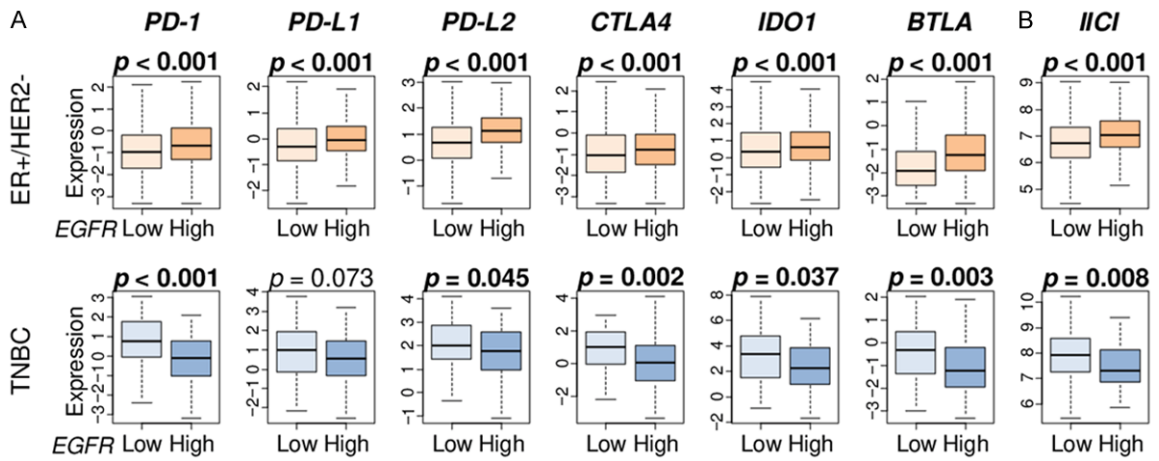

Figure 6. Association between *EGFR* expression and immune checkpoint molecules expression in estrogen-receptor (ER)-positive/human epidermal growth factor receptor (HER2)-negative and triple negative breast cancer subtype in the GSE96058 cohort. Boxplots with median and interquartile ranges shown for (A) expression of immune checkpoint molecules; *PD-1, PD-L1, PD-L2, CTLA4, IDO1,* and *BTLA,* and (B) immune checkpoint index by high and low *EGFR* groups in ER-positive/HER2-negative breast cancer and TNBC.

A high EGFR metastatic tumor was signficantly associated with worse survival, but there was no association observed with the fraction of infiltrating immune cells

Finally, we examined the role of *EGFR* overexpression in influencing outcomes for metastatic breast cancer. We found that high *EGFR* expression was associated with a trend towards worse survival in metastatic tumor (**Figure 7A**). Especially, high *EGFR* expression was significantly associated with worse survival in liver metastatic tumor. In order to investigate the contribution of immune cells to outcomes, the association between EGFR expression and immune cell infiltration was analyzed in metastatic tumor and no significant association was observed (**Figure 7C** and **7D**). Furthermore, *EGFR* was not associated with CYT in metastatic tumor (**Figure 7E**). These results suggest that high *EGFR* in metastatic tumor was associated with worse survival, but immune cells did not seem to be involved in influencing the observed outcomes.

Discussion

In this study, the role of *EGFR* expression in influencing clinical outcomes in breast cancer was investigated using large number of samples in multiple cohorts. We found that among breast cancer subtypes, *EGFR* expression in TNBC was the highest when compared to other subtypes, which supports the existing literature. High *EGFR* expression was significantly associated with improved survival in ER-positive/HER2-negative breast cancer and also significantly associated with high level of intratumor heterogeneity and HRD. Furthermore, they were enriched not only in several pro-cancer-related gene sets, but also in immune-related gene sets. On the other hand, low *EGFR* tumor was enriched in cell proliferation-related gene sets which may explain the observed worse survival. However, these findings were not observed in TNBC. Interestingly, high *EGFR* ER-positive/HER2-negative breast cancers were infiltrated with anti-cancer immune cells. On the other hand, a lower fraction of immune cells was observed in high *EGFR* TNBC along with low level of cytolytic activity. In the single cell sequence data, tumor cells have significantly higher *EGFR* expression compared to immune cells. High *EGFR* expression among ER-positive/HER2-negative breast cancer was significantly associated with higher expression of immune checkpoint molecules, while, on the other hand, there was lower immune checkpoint molecule expression in high *EGFR* TNBC. Finally, high *EGFR* metastatic tumor was significantly associated with worse survival, but there was no association with infiltrating immune cells.

Even though we did not observe any survival difference between high vs. low *EGFR* expres-

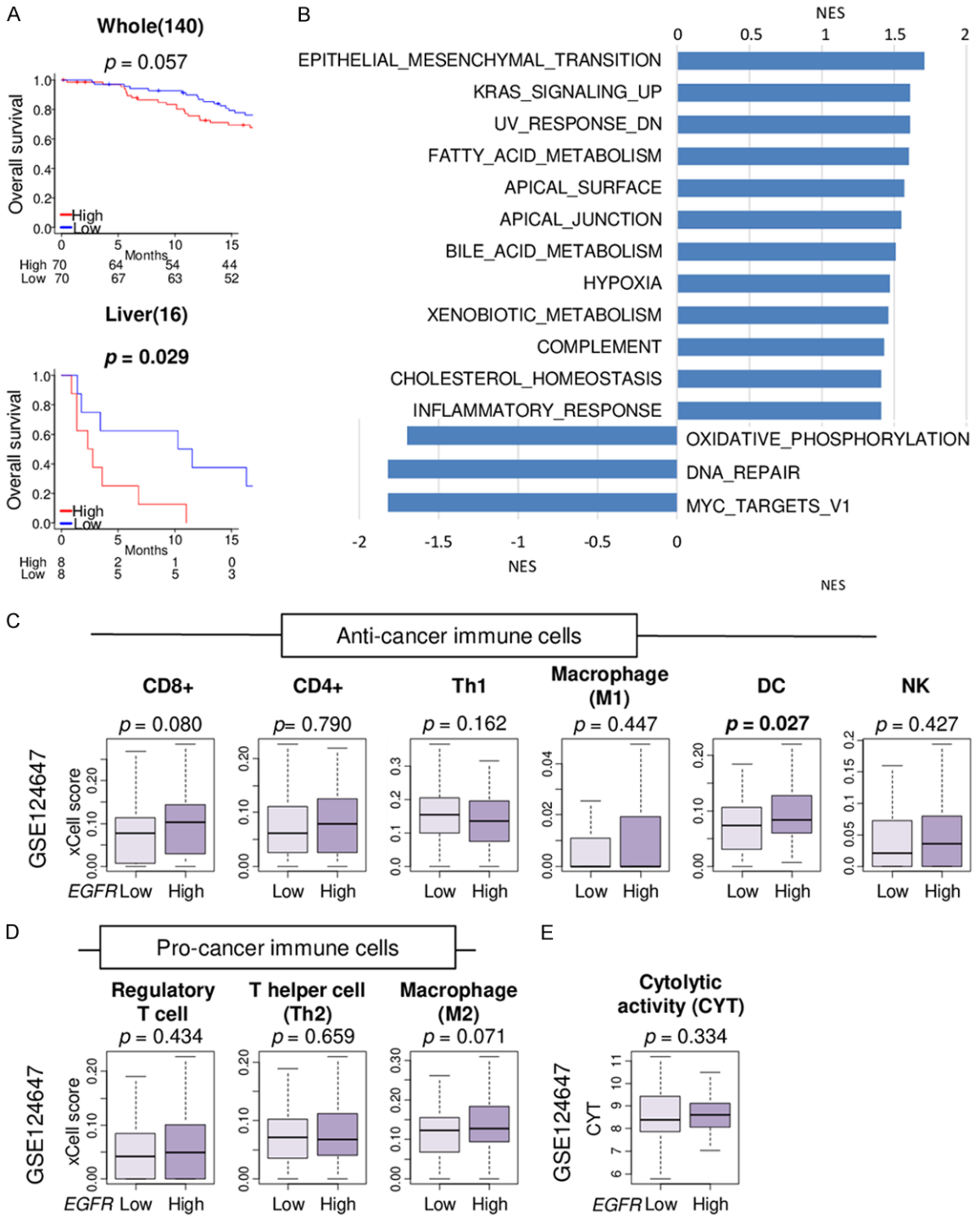

Figure 7. Association of *EGFR* expression with patient survival and anti-cancer and pro-cancer immune cells in metastatic tumor using the GSE96058 cohort. (A) Kaplan-Meier survival curves with *P*-values calculated using the log-rank test for overall survival (OS) in metastatic breast cancer within the low and high *EGFR* groups. (B) Bar plots of hallmark gene sets with significant enrichment within high and low *EGFR* metastatic tumor. Boxplots with medians and interquartile ranges of (C) anti-cancer immune cells; $CD8^+$ T cells, $CD4^+$ T cells, Th1 cells, M1 macrophages, DC, and NK cells, and (D) pro-cancer immune cells; regulatory T cells, Th2 cells, and M2 macrophages, and (E) cytolytic activity by low and high *EGFR* tumor. Mann-Whitney U test was used to calculate the depicted *P*-values. The median value of *EGFR* expression within each cohort was used as a cut-off to divide into high and low *EGFR* groups.

sion in TNBC in our study, but, it is evident in literature that enhanced *EGFR* expression in TNBC within the primary tumor has been associated with EMT resulting in increase in metastases and therefore, could decrease the survival of TNBC patients. Since there are no targeted treatments for TNBC, therefore, *EGFR* expression presents a great opportunity to study targeted approaches. *EGFR* inhibitors including erlotinib [42] and gefitinib [43] were studied in an unselected population of metastatic breast cancer, and no correlation between the expression of *EGFR* with treatment response was observed and overall results were disappointing. Similarly, *EGFR* monoclonal antibodies (panitumumab, cetuximab) also have had minimal efficacy in treatment of TNBC [44, 45]. It has also been observed in breast cancer tissues from patients that once the tumor metastasizes, EGFR expression is down-regulated via increased *EGFR* internalization, decreased expression as part of an adaptive response, and this could correlate with resistance to *EGFR* inhibitors in the breast cancer setting as opposed to NSCLC where *EGFR* inhibitors are approved for activating *EGFR* mutation [46].

It is also important to note the rarity of *EGFR* mutations in breast cancer, which is in sharp contrast to lung cancer [47]. *EGFR* mutations were only present in 11.4% (8/70) TNBC tumor tissues as revealed by mutational analysis [48]. Similar mutations like exon 19 deletions and exon 21 mutations in the *EGFR* kinase domain were reported in breast cancer, that have been previously reported in lung cancer [48]. However, the percentage of TNBC harboring mutations has varied from zero to varying percentages [49]. Although our findings did not show any significant association between TNBC and survival, low *EGFR* expression has been shown to increase cell proliferation in TNBC. We have previously reported the poor prognosis for breast cancer observed with high cell proliferation [50-52]. In addition, high tumor immunity has been reported to be associated with better prognosis in breast cancer [53, 54]. Our current findings suggest that differences in the clinical outcomes of *EGFR* expression in subtypes were probably due to differences in the relationship between *EGFR* expression and anti-cancer immunity. Furthermore, the finding that tumor cells expressed *EGFR* significantly higher than immune cells, as indicated by single cell sequence data, suggests that *EGFR* expression in tumor cells may attract immune cells.

As already established, certain epitopes of *EGFR* are able to elicit both a humoral and cellular immune response. This is consistent with our data showing higher cytolytic activity in ER+/HER2 negative breast cancer [14]. This was also seen in our analysis where higher *EGFR* expression correlated with higher immune checkpoint molecule expression in ER-positive/HER2-negative breast cancer. This is consistent with prior data in human bronchial epithelial cell lines, showing that mutation in the EGFR signaling pathway could result in increased PD-L1 expression, or if we are able to induce mutant EGFR expression, it drives increased PD-L1 expression [55], although our study focuses on *EGFR* expression and not mutation. Higher checkpoint expression due to greater immune cell infiltration may make higher *EGFR* expression on tumors an attractive target for testing immunotherapeutic strategies for ER-positive/HER2-negative breast cancer. These findings were further confirmed when higher immune related gene sets were observed in the ER-positive/HER2-negative breast cancer subtype. Similar findings in TNBC were not observed in our study. Better outcomes in ER-positive/HER2-negative breast cancer with higher *EGFR* expression could be explained by a favorable immune profile with higher cytolytic activity, more anti-tumor immune cells, higher immune related gene sets. On the other hand, higher expression of *EGFR* in TNBC was associated with a worse immune profile, however, no impact on survival was observed within TNBC in our study, but the findings of an unfavorable anti-cancer immune response may explain the worse outcomes observed in TNBC in literature.

EGFR is a well-known cancer driver gene. Aberrant activation of *EGFR* signaling is associated with progression of breast cancer and with a worse prognosis [9, 10]. Interestingly, we found that high *EGFR* gene expression was associated with better survival in ER-positive/HER2-negative breast cancer, but not in TNBC. One of the reasons for this observed discrepancy may be related to the fact that *EGFR* gene expression was so high in TNBC overall that low

expression in that subtype is not low enough compared with ER-positive/HER2-negative breast cancer to show the survival difference as it did in the latter. In agreement with our findings, TNBC has been reported to express higher *EGFR* gene compared with the other subtypes [11-13]. In ER-positive/HER2-negative subtype, we found that low *EGFR* expression is associated with high cell proliferation that contributes to worse biology, and high *EGFR* expression with enhanced immune response and anti-cancer immune cell infiltration that contributes to better biology.

Further, it is known that expression of *EGFR* does not necessarily cause activation of *EGFR* signaling pathway, thus clinically *EGFR* inhibitors suppress the activity of *EGFR* signaling primarily by inhibiting *EGFR* phosphorylation. Generally *EGFR* signaling is assessed by protein expression using western blotting, and there are only a few studies using gene expression. We showed that there was no positive association between EGFR-mutation and EGFR gene expression in Figure S1. In agreement, Lee et al. reported that overexpression of *EGFR* gene, but not *EGFR* copy number, was associated with patient prognosis although *EGFR* gene expression correlated with *EGFR* protein expression. They suggested that besides the increased *EGFR* gene copy number, *EGFR* overexpression may be induced by other mechanisms, such as mutation, aberrant transcription or translational modification [56]. Several genes are involved in *EGFR* signaling, and it is reasonable to consider that the expression of a single *EGFR* gene alone may not reflect the entire signaling cascade. We previously reported that ER response signaling score comprised of 200 gene expression was significantly associated with clinical outcome [33], but *ESR1* and *ESR2* expression were not. Furthermore, there was no correlation of ER response signaling with *ESR1* and *ESR2* expression. To this end, we analyzed the correlation between *EGFR* expression and activation of *EGFR* signaling pathway utilizing GSVA algorithm that we have shown in multiple previous studies [36-41], and found a significant correlation with Spearman correlation coefficient (r) = 0.518, $P<0.01$.

Utilizing the same approach, we confirmed that *EGFR* gene expression significantly correlated with *EGFR* signaling pathway (Figure S2). This is in agreement with the previous reports that treatment of cancer cells with *EGFR* inhibitors would usually decrease *EGFR* expression due to negative feedback. This finding seems somewhat contradictory with our other finding that cell proliferation-related gene sets were enriched in low *EGFR* gene expression, and high expression was associated with high level of anti-immune cell infiltration and better survival in ER-positive/HER2-negative breast cancer, whereas that was not the case in TNBC. On the other hand, we did see that *EGFR* gene expression was associated with epithelial mesenchymal transition, which is a known cancer progression marker [57], and with worse prognosis in metastatic breast cancer. In addition, Bhargava et al. suggested that differences in the relevance of *EGFR* overexpression by different studies may be due to variation in techniques including the type of antibodies used, definition of overexpression and interobserver variability [47]. Taken together, the role of *EGFR* expression in breast cancer progression may be context dependent.

Recently, the role of photochemistry-based cancer therapy near-infrared (NIR) photoimmunotherapy (PIT) has been focused on. This is a molecularly targeted phototherapy for cancer where a conjugate of a near-infrared, water-soluble, silicon-phothalocyanine derivative called IRdye700DX (IR700), and a monoclonal antibody (mAb) are injected that specifically target an antigen that is expressed on the cancer cell surface. Kobayashi et al. used an antibody that binds to EGFR in cancer cell membranes, which are known to be present in lung, head and neck, and breast cancer. A photochemical “death” switch is activated on local exposure to NIR light that results in immunogenic cell death (ICD) of targeted cancer cells. By the combined use of cancer-targeting NIR-PIT with other immunotherapies, NIR-PIT of a local tumor, could lead to responses in distant metastases, also known as abscopal effect that may result in activation of a systemic anti-cancer immunity and inhibit recurrences by induction of long-term immune memory without the systemic autoimmune adverse effects that are generally observed with immune checkpoint inhibitors. As we have observed in our study, high *EGFR* tumors attract immune cells (anti-cancer immune cells) and are associ-

ated with high expression of immune checkpoint molecules, especially the ER-positive/HER2-negative breast cancer subtype. Therefore, we cannot help but speculate that PIT may be expected to have a high therapeutic effect on ER-positive/HER2-negative breast cancer for the ones with high *EGFR* tumors given the favorable immune milieu.

We have discovered interesting features of intratumor *EGFR* expression in different breast cancer subtypes and also shown that the tumor microenvironment plays unique role in high versus low *EGFR* expressing tumors in ER-positive/HER2-negative breast cancer and TNBC. However, this study has some limitations. First, this study is retrospective in nature utilizing a large amount of publicly available clinical and genetic data, however, it is important to note that the data on co-morbidity and therapeutic intervention are not available and therefore, the influence of treatment on outcomes is unclear. Another limitation of the current study is that the conclusions are drawn based on data analysis alone without any experiments to validate the underlying mechanism, although the analyses of transcriptome were validated by multiple independent large patient cohorts. We chose to report this way for two reasons. First, the novelty of the current study is to clarify the clinical relevance of *EGFR* gene expression, where the mechanistic concepts that we investigated have already been previously published and conducting these experiments is simply repetitive. Second, we analyzed cancer tissue in human patients' body that cannot be completely replicated by any animal models or even cell culture system. As a matter of fact, we have studied numerous syngeneic [58-62], genetic [63-65], and even patient-derived xenograft models [66-68] and organoids [69] using human patient cancer samples; however, none of these completely replicate human tumor. To this end, one can argue that the different results between human samples and animal models may be due to difference between the species. Finally, our study focused on studying *EGFR* expression and not *EGFR* mutation and its clinical relevance in the setting when approval of *EGFR* inhibitors based on the activating mutation status is also limited.

In conclusion, *EGFR* expression was found to have different characteristics depending on breast cancer subtype. Especially, high *EGFR* ER-positive/HER2-negative breast cancer was significantly associated with better survival and immunity in tumor immune microenvironment.

Acknowledgements

This work was supported by National Institutes of Health, USA grant R01CA160688, R01CA251545, R01CA250412, R37CA248-018, as well as US Department of Defense BCRP grant W81XWH-19-1-0674, W81XWH-19-1-0111 to K.T., and National Cancer Institute, USA cancer center support grant P30-CA016056 to Roswell Park Comprehensive Cancer Center (RPCCC). Research reported in this publication was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under award Number KL2TR001413 and UL1TR001412. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Kazuaki Takabe, Breast Surgery, Department of Surgical Oncology, Roswell Park Comprehensive Cancer Center, Elm & Carlton Streets, Buffalo, New York 14263, USA. Tel: 1-716-845-2918; Fax: 1-716-845-1668; E-mail: kazuaki.takabe@roswellpark.org

References

[1] Lemmon MA, Schlessinger J and Ferguson KM. The EGFR family: not so prototypical receptor tyrosine kinases. Cold Spring Harb Perspect Biol 2014; 6: a020768.

[2] Yarden Y and Pines G. The ERBB network: at last, cancer therapy meets systems biology. Nat Rev Cancer 2012; 12: 553-563.

[3] Citri A and Yarden Y. EGF-ERBB signalling: towards the systems level. Nat Rev Mol Cell Biol 2006; 7: 505-516.

[4] Rubin Grandis J, Melhem MF, Barnes EL and Tweardy DJ. Quantitative immunohistochemical analysis of transforming growth factor-alpha and epidermal growth factor receptor in patients with squamous cell carcinoma of the head and neck. Cancer 1996; 78: 1284-1292.

[5] Brabender J, Danenberg KD, Metzger R, Schneider PM, Park J, Salonga D, Hölscher AH and Danenberg PV. Epidermal growth factor receptor and HER2-neu mRNA expression in

non-small cell lung cancer is correlated with survival. Clin Cancer Res 2001; 7: 1850-1855.
[6] Spano JP, Lagorce C, Atlan D, Milano G, Domont J, Benamouzig R, Attar A, Benichou J, Martin A, Morere JF, Raphael M, Penault-Llorca F, Breau JL, Fagard R, Khayat D and Wind P. Impact of EGFR expression on colorectal cancer patient prognosis and survival. Ann Oncol 2005; 16: 102-108.
[7] Konecny GE, Santos L, Winterhoff B, Hatmal M, Keeney GL, Mariani A, Jones M, Neuper C, Thomas B, Muderspach L, Riehle D, Wang HJ, Dowdy S, Podratz KC and Press MF. HER2 gene amplification and EGFR expression in a large cohort of surgically staged patients with non-endometrioid (type II) endometrial cancer. Br J Cancer 2009; 100: 89-95.
[8] Jänne PA, Yang JC, Kim DW, Planchard D, Ohe Y, Ramalingam SS, Ahn MJ, Kim SW, Su WC, Horn L, Haggstrom D, Felip E, Kim JH, Frewer P, Cantarini M, Brown KH, Dickinson PA, Ghiorghiu S and Ranson M. AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. N Engl J Med 2015; 372: 1689-1699.
[9] Di Fiore PP, Pierce JH, Fleming TP, Hazan R, Ullrich A, King CR, Schlessinger J and Aaronson SA. Overexpression of the human EGF receptor confers an EGF-dependent transformed phenotype to NIH 3T3 cells. Cell 1987; 51: 1063-1070.
[10] Radisky DC. Epithelial-mesenchymal transition. J Cell Sci 2005; 118: 4325-4326.
[11] Masuda H, Zhang D, Bartholomeusz C, Doihara H, Hortobagyi GN and Ueno NT. Role of epidermal growth factor receptor in breast cancer. Breast Cancer Res Treat 2012; 136: 331-345.
[12] Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y and Pietenpol JA. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 2011; 121: 2750-2767.
[13] Rakha EA, El-Sayed ME, Green AR, Lee AH, Robertson JF and Ellis IO. Prognostic markers in triple-negative breast cancer. Cancer 2007; 109: 25-32.
[14] Shomura H, Shichijo S, Komatsu N, Matsueda S, Mine T, Rikimaru T, Sato Y, Todo S and Itoh K. Identification of epidermal growth factor receptor-derived peptides recognised by both cellular and humoral immune responses in HLA-A24+ non-small cell lung cancer patients. Eur J Cancer 2004; 40: 1776-1786.
[15] Brueffer C, Gladchuk S, Winter C, Vallon-Christersson J, Hegardt C, Häkkinen J, George AM, Chen Y, Ehinger A, Larsson C, Loman N, Malmberg M, Rydén L, Borg Å and Saal LH. The mutational landscape of the SCAN-B real-world primary breast cancer transcriptome. EMBO Mol Med 2020; 12: e12118.
[16] Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, Kovatich AJ, Benz CC, Levine DA, Lee AV, Omberg L, Wolf DM, Shriver CD, Thorsson V and Hu H. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. Cell 2018; 173: 400-416, e411.
[17] Chung W, Eum HH, Lee HO, Lee KM, Lee HB, Kim KT, Ryu HS, Kim S, Lee JE, Park YH, Kan Z, Han W and Park WY. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. Nat Commun 2017; 8: 15081.
[18] Aran D, Hu Z and Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol 2017; 18: 220.
[19] Rooney MS, Shukla SA, Wu CJ, Getz G and Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell 2015; 160: 48-61.
[20] Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA, Ziv E, Culhane AC, Paull EO, Sivakumar IKA, Gentles AJ, Malhotra R, Farshidfar F, Colaprico A, Parker JS, Mose LE, Vo NS, Liu J, Liu Y, Rader J, Dhankani V, Reynolds SM, Bowlby R, Califano A, Cherniack AD, Anastassiou D, Bedognetti D, Mokrab Y, Newman AM, Rao A, Chen K, Krasnitz A, Hu H, Malta TM, Noushmehr H, Pedamallu CS, Bullman S, Ojesina AI, Lamb A, Zhou W, Shen H, Choueiri TK, Weinstein JN, Guinney J, Saltz J, Holt RA, Rabkin CS, Lazar AJ, Serody JS, Demicco EG, Disis ML, Vincent BG and Shmulevich I. The immune landscape of cancer. Immunity 2019; 51: 411-412.
[21] Oshi M, Tokumaru Y, Patel A, Yan L, Matsuyama R, Endo I, Katz MHG and Takabe K. A novel four-gene score to predict pathologically complete (R0) resection and survival in pancreatic cancer. Cancers (Basel) 2020; 12: 3635.
[22] Oshi M, Tokumaru Y, Asaoka M, Yan L, Satyananda V, Matsuyama R, Matsuhashi N, Futamura M, Ishikawa T, Yoshida K, Endo I and Takabe K. M1 macrophage and M1/M2 ratio defined by transcriptomic signatures resemble only part of their conventional clinical characteristics in breast cancer. Sci Rep 2020; 10: 16554.
[23] Oshi M, Newman S, Tokumaru Y, Yan L, Matsuyama R, Kalinski P, Endo I and Takabe K. Plasmacytoid dendritic cell (pDC) infiltration correlate with tumor infiltrating lymphocytes, cancer immunity, and better survival in triple negative breast cancer (TNBC) more strongly than conventional dendritic cell (cDC). Cancers (Basel) 2020; 12: 3342.
[24] Oshi M, Newman S, Murthy V, Tokumaru Y, Yan L, Matsuyama R, Endo I and Takabe K. ITPKC as a prognostic and predictive biomarker of

neoadjuvant chemotherapy for triple negative breast cancer. Cancers (Basel) 2020; 12: 2758.

[25] Oshi M, Angarita FA, Tokumaru Y, Yan L, Matsuyama R, Endo I and Takabe K. High expression of NRF2 is associated with increased tumor-infiltrating lymphocytes and cancer immunity in ER-positive/HER2-negative breast cancer. Cancers (Basel) 2020; 12: 3856.

[26] Oshi M, Asaoka M, Tokumaru Y, Angarita FA, Yan L, Matsuyama R, Zsiros E, Ishikawa T, Endo I and Takabe K. Abundance of regulatory T cell (Treg) as a predictive biomarker for neoadjuvant chemotherapy in triple-negative breast cancer. Cancers (Basel) 2020; 12: 3038.

[27] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES and Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005; 102: 15545-15550.

[28] Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP and Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. Cell Syst 2015; 1: 417-425.

[29] Oshi M, Katsuta E, Yan L, Ebos JML, Rashid OM, Matsuyama R, Endo I and Takabe K. A novel 4-gene score to predict survival, distant metastasis and response to neoadjuvant therapy in breast cancer. Cancers (Basel) 2020; 12: 1148.

[30] Oshi M, Kim TH, Tokumaru Y, Yan L, Matsuyama R, Endo I, Cherkassky L and Takabe K. Enhanced DNA repair pathway is associated with cell proliferation and worse survival in hepatocellular carcinoma (HCC). Cancers (Basel) 2021; 13: 323.

[31] Oshi M, Newman S, Tokumaru Y, Yan L, Matsuyama R, Endo I, Katz MHG and Takabe K. High G2M pathway score pancreatic cancer is associated with worse survival, particularly after margin-positive (R1 or R2) resection. Cancers (Basel) 2020; 12: 2871.

[32] Oshi M, Newman S, Tokumaru Y, Yan L, Matsuyama R, Endo I, Nagahashi M and Takabe K. Intra-tumoral angiogenesis is associated with inflammation, immune reaction and metastatic recurrence in breast cancer. Int J Mol Sci 2020; 21: 6708.

[33] Oshi M, Tokumaru Y, Angarita FA, Yan L, Matsuyama R, Endo I and Takabe K. Degree of early estrogen response predict survival after endocrine therapy in primary and metastatic ER-positive breast cancer. Cancers (Basel) 2020; 12: 3557.

[34] Takahashi H, Asaoka M, Yan L, Rashid OM, Oshi M, Ishikawa T, Nagahashi M and Takabe K. Biologically aggressive phenotype and anticancer immunity counterbalance in breast cancer with high mutation rate. Sci Rep 2020; 10: 1852.

[35] Takahashi H, Oshi M, Asaoka M, Yan L, Endo I and Takabe K. Molecular biological features of nottingham histological grade 3 breast cancers. Ann Surg Oncol 2020; 27: 4475-4485.

[36] Tokumaru Y, Oshi M, Katsuta E, Yan L, Satyananda V, Matsuhashi N, Futamura M, Akao Y, Yoshida K and Takabe K. KRAS signaling enriched triple negative breast cancer is associated with favorable tumor immune microenvironment and better survival. Am J Cancer Res 2020; 10: 897-907.

[37] Oshi M, Satyananda V, Angarita FA, Kim TH, Tokumaru Y, Yan L, Matsuyama R, Endo I, Nagahashi M and Takabe K. Angiogenesis is associated with an attenuated tumor microenvironment, aggressive biology, and worse survival in gastric cancer patients. Am J Cancer Res 2021; 11: 1659-1671.

[38] Gandhi S, Oshi M, Murthy V, Repasky EA and Takabe K. Enhanced thermogenesis in triple-negative breast cancer is associated with protumor immune microenvironment. Cancers (Basel) 2021; 13: 2559.

[39] Oshi M, Tokumaru Y, Angarita FA, Lee L, Yan L, Matsuyama R, Endo I and Takabe K. Adipogenesis in triple-negative breast cancer is associated with unfavorable tumor immune microenvironment and with worse survival. Sci Rep 2021; 11: 12541.

[40] Oshi M, Patel A, Le L, Tokumaru Y, Yan L, Matsuyama R, Endo I and Takabe K. G2M checkpoint pathway alone is associated with drug response and survival among cell proliferation-related pathways in pancreatic cancer. Am J Cancer Res 2021; 11: 3070-3084.

[41] Murthy V, Oshi M, Tokumaru Y, Endo I and Takabe K. Increased apoptosis is associated with robust immune cell infiltration and cytolytic activity in breast cancer. Am J Cancer Res 2021; 11: 3674-3687.

[42] Dickler MN, Cobleigh MA, Miller KD, Klein PM and Winer EP. Efficacy and safety of erlotinib in patients with locally advanced or metastatic breast cancer. Breast Cancer Res Treat 2009; 115: 115-121.

[43] von Minckwitz G, Jonat W, Fasching P, du Bois A, Kleeberg U, Lück HJ, Kettner E, Hilfrich J, Eiermann W, Torode J and Schneeweiss A. A multicentre phase II study on gefitinib in taxane- and anthracycline-pretreated metastatic breast cancer. Breast Cancer Res Treat 2005; 89: 165-172.

[44] Carey LA, Rugo HS, Marcom PK, Mayer EL, Esteva FJ, Ma CX, Liu MC, Storniolo AM, Rimawi MF, Forero-Torres A, Wolff AC, Hobday TJ, Ivanova A, Chiu WK, Ferraro M, Burrows E, Ber-

nard PS, Hoadley KA, Perou CM and Winer EP. TBCRC 001: randomized phase II study of cetuximab in combination with carboplatin in stage IV triple-negative breast cancer. J Clin Oncol 2012; 30: 2615-2623.
[45] Yardley DA, Ward PJ, Daniel BR, Eakle JF, Lamar RE, Lane CM and Hainsworth JD. Panitumumab, gemcitabine, and carboplatin as treatment for women with metastatic triple-negative breast cancer: a sarah cannon research institute phase II trial. Clin Breast Cancer 2016; 16: 349-355.
[46] Dittadi R, Donisi PM, Brazzale A, Cappellozza L, Bruscagnin G and Gion M. Epidermal growth factor receptor in breast cancer. Comparison with non malignant breast tissue. Br J Cancer 1993; 67: 7-9.
[47] Bhargava R, Gerald WL, Li AR, Pan Q, Lal P, Ladanyi M and Chen B. EGFR gene amplification in breast cancer: correlation with epidermal growth factor receptor mRNA and protein expression and HER-2 status and absence of EGFR-activating mutations. Mod Pathol 2005; 18: 1027-1033.
[48] Teng YH, Tan WJ, Thike AA, Cheok PY, Tse GM, Wong NS, Yip GW, Bay BH and Tan PH. Mutations in the epidermal growth factor receptor (EGFR) gene in triple negative breast cancer: possible implications for targeted therapy. Breast Cancer Res 2011; 13: R35.
[49] Nakai K, Hung MC and Yamaguchi H. A perspective on anti-EGFR therapies targeting triple-negative breast cancer. Am J Cancer Res 2016; 6: 1609-1623.
[50] Schulze A, Oshi M, Endo I and Takabe K. MYC targets scores are associated with cancer aggressiveness and poor survival in ER-positive primary and metastatic breast cancer. Int J Mol Sci 2020; 21: 8127.
[51] Oshi M, Asaoka M, Tokumaru Y, Yan L, Matsuyama R, Ishikawa T, Endo I and Takabe K. CD8 T cell score as a prognostic biomarker for triple negative breast cancer. Int J Mol Sci 2020; 21: 6968.
[52] Oshi M, Newman S, Tokumaru Y, Yan L, Matsuyama R, Endo I and Takabe K. Inflammation is associated with worse outcome in the whole cohort but with better outcome in triple-negative subtype of breast cancer patients. J Immunol Res 2020; 2020: 5618786.
[53] Oshi M, Takahashi H, Tokumaru Y, Yan L, Rashid OM, Nagahashi M, Matsuyama R, Endo I and Takabe K. The E2F pathway score as a predictive biomarker of response to neoadjuvant therapy in ER+/HER2- breast cancer. Cells 2020; 9: 1643.
[54] Oshi M, Takahashi H, Tokumaru Y, Yan L, Rashid OM, Matsuyama R, Endo I and Takabe K. G2M cell cycle pathway score as a prognostic biomarker of metastasis in estrogen receptor (ER)-positive breast cancer. Int J Mol Sci 2020; 21: 2921.
[55] Akbay EA, Koyama S, Carretero J, Altabef A, Tchaicha JH, Christensen CL, Mikse OR, Cherniack AD, Beauchamp EM, Pugh TJ, Wilkerson MD, Fecci PE, Butaney M, Reibel JB, Soucheray M, Cohoon TJ, Janne PA, Meyerson M, Hayes DN, Shapiro GI, Shimamura T, Sholl LM, Rodig SJ, Freeman GJ, Hammerman PS, Dranoff G and Wong KK. Activation of the PD-1 pathway contributes to immune escape in EGFR-driven lung tumors. Cancer Discov 2013; 3: 1355-1363.
[56] Lee HJ, Seo AN, Kim EJ, Jang MH, Kim YJ, Kim JH, Kim SW, Ryu HS, Park IA, Im SA, Gong G, Jung KH, Kim HJ and Park SY. Prognostic and predictive values of EGFR overexpression and EGFR copy number alteration in HER2-positive breast cancer. Br J Cancer 2015; 112: 103-111.
[57] Moussa RA, Khalil EZI and Ali AI. Prognostic role of epithelial-mesenchymal transition markers "E-cadherin, β-catenin, ZEB1, ZEB2 and p63" in bladder carcinoma. World J Oncol 2019; 10: 199-217.
[58] Katsuta E, DeMasi SC, Terracina KP, Spiegel S, Phan GQ, Bear HD and Takabe K. Modified breast cancer model for preclinical immunotherapy studies. J Surg Res 2016; 204: 467-474.
[59] Rashid OM and Takabe K. Does removal of the primary tumor in metastatic breast cancer improve survival? J Womens Health (Larchmt) 2014; 23: 184-188.
[60] Nagahashi M, Yamada A, Aoyagi T, Allegood J, Wakai T, Spiegel S and Takabe K. Sphingosine-1-phosphate in the lymphatic fluid determined by novel methods. Heliyon 2016; 2: e00219.
[61] Katsuta E, Oshi M, Rashid OM and Takabe K. Generating a murine orthotopic metastatic breast cancer model and performing murine radical mastectomy. J Vis Exp 2018.
[62] Nagahashi M, Yamada A, Miyazaki H, Allegood JC, Tsuchida J, Aoyagi T, Huang WC, Terracina KP, Adams BJ, Rashid OM, Milstien S, Wakai T, Spiegel S and Takabe K. Interstitial fluid sphingosine-1-phosphate in murine mammary gland and cancer and human breast tissue and cancer determined by novel methods. J Mammary Gland Biol Neoplasia 2016; 21: 9-17.
[63] Nagahashi M, Kim EY, Yamada A, Ramachandran S, Allegood JC, Hait NC, Maceyka M, Milstien S, Takabe K and Spiegel S. Spns2, a transporter of phosphorylated sphingoid bases, regulates their blood and lymph levels, and the lymphatic network. FASEB J 2013; 27: 1001-1011.
[64] Donoviel MS, Hait NC, Ramachandran S, Maceyka M, Takabe K, Milstien S, Oravecz T and Spiegel S. Spinster 2, a sphingosine-1-phos-

phate transporter, plays a critical role in inflammatory and autoimmune diseases. FASEB J 2015; 29: 5018-5028.
[65] Huang WC, Liang J, Nagahashi M, Avni D, Yamada A, Maceyka M, Wolen AR, Kordula T, Milstien S, Takabe K, Oravecz T and Spiegel S. Sphingosine-1-phosphate phosphatase 2 promotes disruption of mucosal integrity, and contributes to ulcerative colitis in mice and humans. FASEB J 2016; 30: 2945-2958.
[66] Okano M, Oshi M, Butash A, Okano I, Saito K, Kawaguchi T, Nagahashi M, Kono K, Ohtake T and Takabe K. Orthotopic implantation achieves better engraftment and faster growth than subcutaneous implantation in breast cancer patient-derived xenografts. J Mammary Gland Biol Neoplasia 2020; 25: 27-36.
[67] Oshi M, Okano M, Maiti A, Rashid OM, Saito K, Kono K, Matsuyama R, Endo I and Takabe K. Novel breast cancer brain metastasis patient-derived orthotopic xenograft model for preclinical studies. Cancers (Basel) 2020; 12: 444.
[68] Kawaguchi T, Foster BA, Young J and Takabe K. Current update of patient-derived xenograft model for translational breast cancer research. J Mammary Gland Biol Neoplasia 2017; 22: 131-139.
[69] Tokumaru Y, Oshi M, Patel A, Tian W, Yan L, Matsuhashi N, Futamura M, Yoshida K and Takabe K. Organoids are limited in modeling the colon adenoma-carcinoma sequence. Cells 2021; 10: 488.

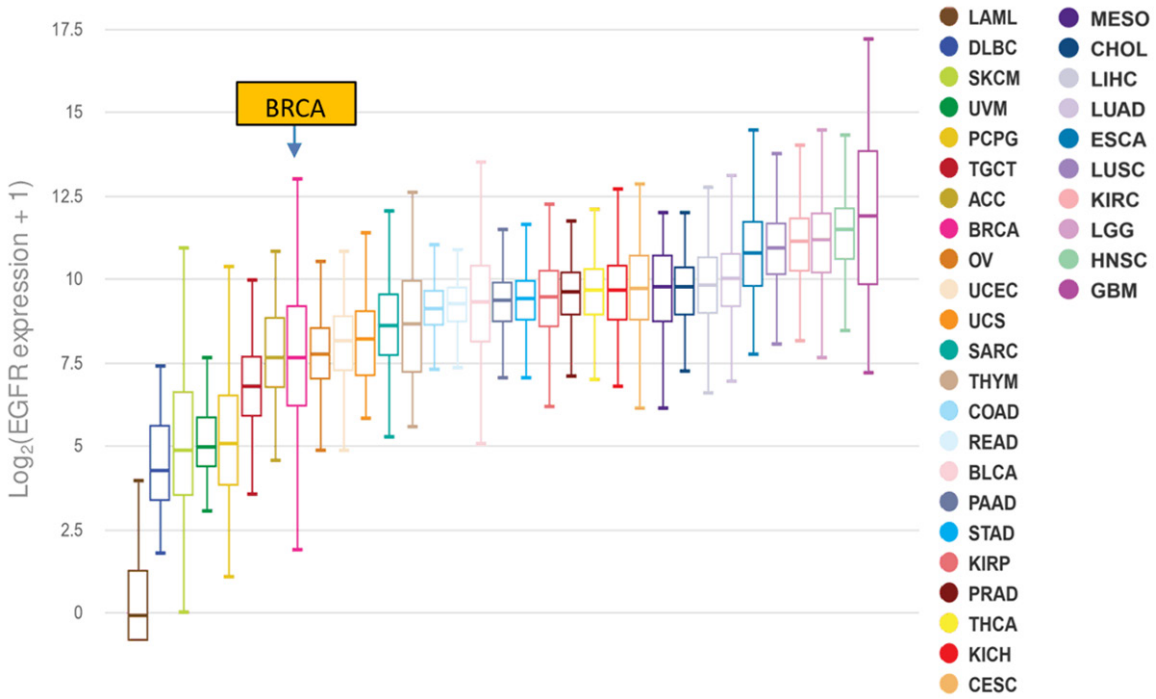

Figure S1. Association of *EGFR* expression with cancers. Boxplots of *EGFR* expression by several cancers in the TCGA cohort. The figure was made through University of California Santa Cruz (UCSC) Xena (xena.ucsc.edu). ACC: adrenocortical carcinoma; BLCA: bladder urothelial carcinoma; BRCA: breast invasive carcinoma; CESC: cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL: cholangiocarcinoma; COAD: colon adenocarcinoma; DLBC: lymphoid neoplasm, diffuse large B-cell lymphoma; ESCA: esophageal carcinoma; GBM: glioblastoma multiforme; HNSC: head and neck squamous cell carcinoma; KICH: kidney chromophobe; KIRC: kidney renal clear-cell carcinoma; KIRP: kidney renal papillary cell carcinoma; LGG: brain lower-grade glioma; LIHC: liver hepatocellular carcinoma; LUAD: lung adenocarcinoma; LUSC: lung squamous cell carcinoma; MESO: mesothelioma; OV: ovarian serous cystadenocarcinoma; PAAD: pancreatic adenocarcinoma; PCPG: pheochromocytoma and paraganglioma; PRAD: prostate adenocarcinoma; READ: rectum adenocarcinoma; SARC: sarcoma; SKCM: skin cutaneous melanoma; STAD: stomach adenocarcinoma; TGCT: Testicular Germ Cell Tumors; THCA: thyroid carcinoma; THYM: thymoma; UCEC: uterine corpus endometrial carcinoma; UCS: uterine carcinosarcoma; UVM: uveal melanoma.

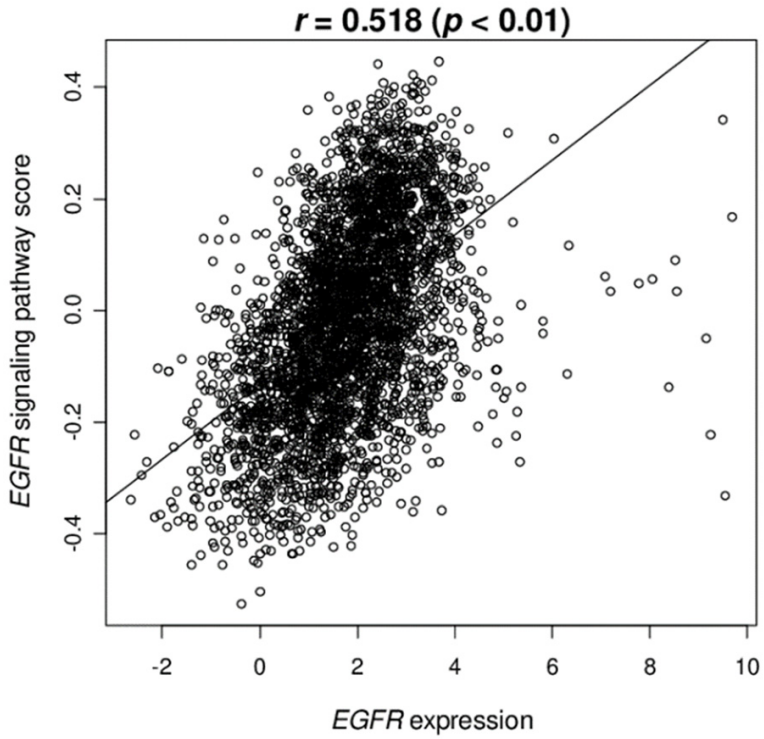

Figure S2. Association between EGFR expression and EGFR signaling pathway score in breast cancer. Correlation plot of EGFR expression with EGFR signaling pathway score, which was calculated by GSVA with KEGG pathway in the GSE96058 cohort. Spearman's rank correlation coefficient (*r*) was used to perform the analysis.

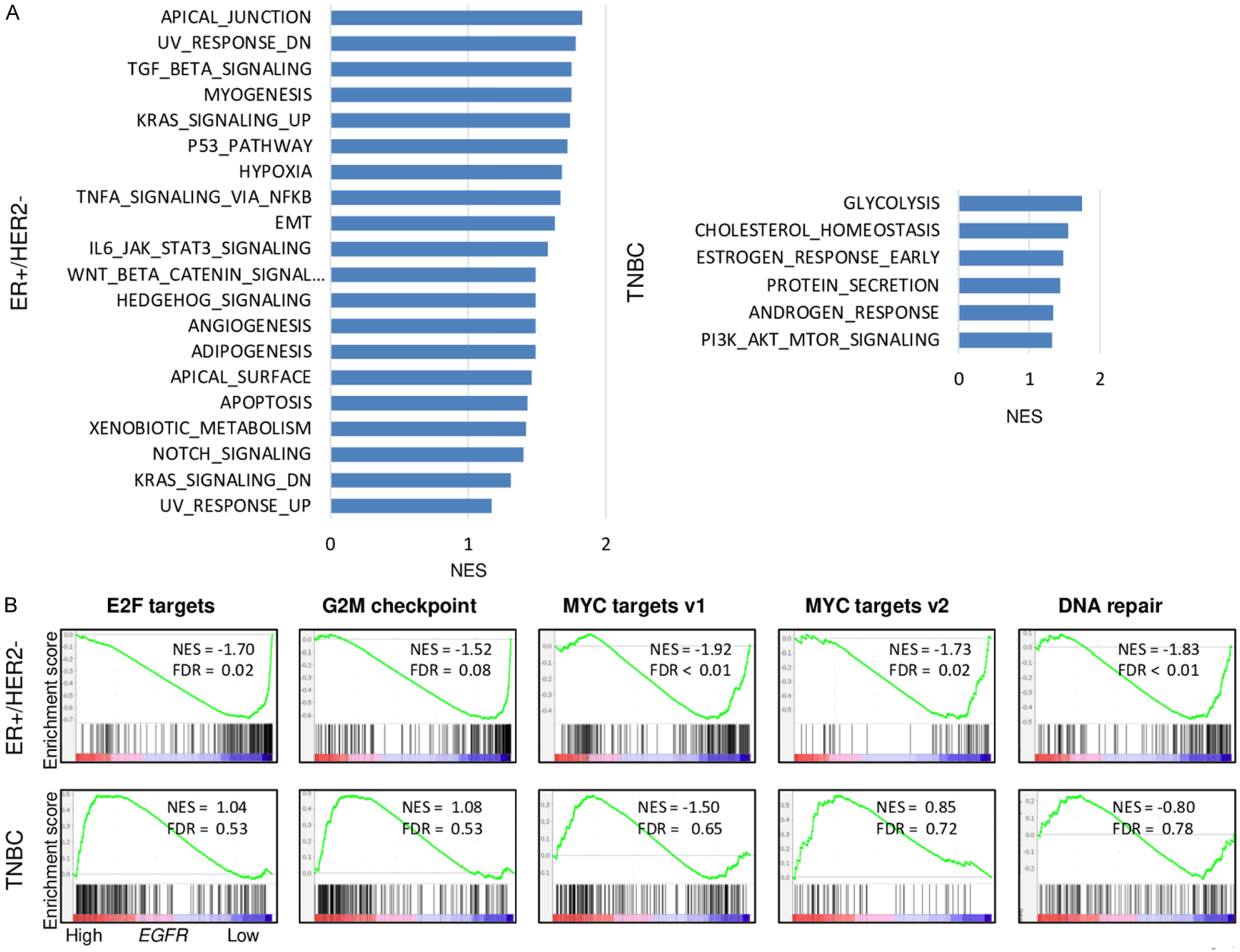
A
ER+/HER2-
APICAL_JUNCTION
UV_RESPONSE_DN
TGF_BETA_SIGNALING
MYOGENESIS
KRAS_SIGNALING_UP
P53_PATHWAY
HYPOXIA
TNFA_SIGNALING_VIA_NFKB
EMT
IL6_JAK_STAT3_SIGNALING
WNT_BETA_CATENIN_SIGNAL...
HEDGEHOG_SIGNALING
ANGIOGENESIS
ADIPOGENESIS
APICAL_SURFACE
APOPTOSIS
XENOBIOTIC_METABOLISM
NOTCH_SIGNALING
KRAS_SIGNALING_DN
UV_RESPONSE_UP
0
1
2
NES
TNBC
GLYCOLYSIS
CHOLESTEROL_HOMEOSTASIS
ESTROGEN_RESPONSE_EARLY
PROTEIN_SECRETION
ANDROGEN_RESPONSE
PI3K_AKT_MTOR_SIGNALING
0
1
2
NES
B
E2F targets
G2M checkpoint
MYC targets v1
MYC targets v2
DNA repair
ER+/HER2-
TNBC
Enrichment score
NES = -1.70
FDR = 0.02
NES = -1.52
FDR = 0.08
NES = -1.92
FDR < 0.01
NES = -1.73
FDR = 0.02
NES = -1.83
FDR < 0.01
NES = 1.04
FDR = 0.53
NES = 1.08
FDR = 0.53
NES = -1.50
FDR = 0.65
NES = 0.85
FDR = 0.72
NES = -0.80
FDR = 0.78
High
EGFR
Low

Figure S3. Association of *EGFR* expression with biological function in each breast cancer subtype in METABRIC cohort. A. Bar plots of gene sets that significantly enriched in high *EGFR* groups in each subtype consistently in both GSE96058 and METABRIC cohorts. B. Enrichment plots of cell proliferation-related gene sets; E2F targets, G2M checkpoint, and MYC targets v1 and v2, and DNA repair in each breast cancer subtype. The two groups were divided by median value of *EGFR* expression within each cohort.

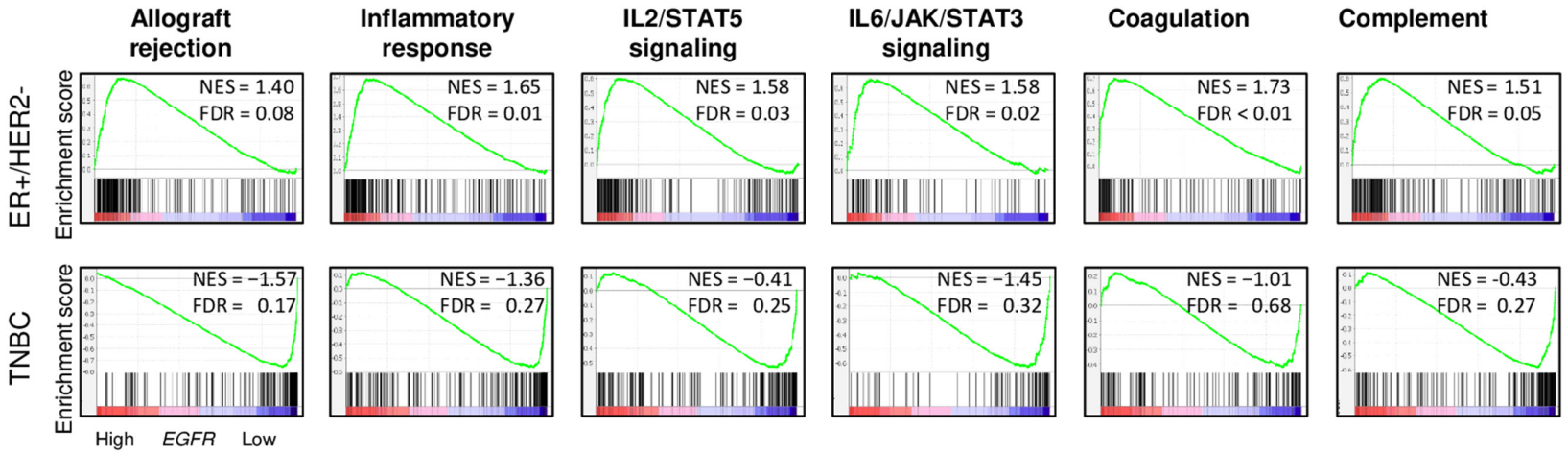

Figure S4. Association of *EGFR* expression with immune-related biological function in each breast cancer subtype in METABRIC cohort. Enrichment plots of immune-related gene sets; allograft rejection, inflammatory response, IL2/STAT5 signaling IL6/JAK/STAT3 signaling, coagulation, and complement in ER-positive/HER2-negative breast cancer and TNBC within *EGFR* high and low groups. The two groups were divided by median value of *EGFR* expression within each cohort.

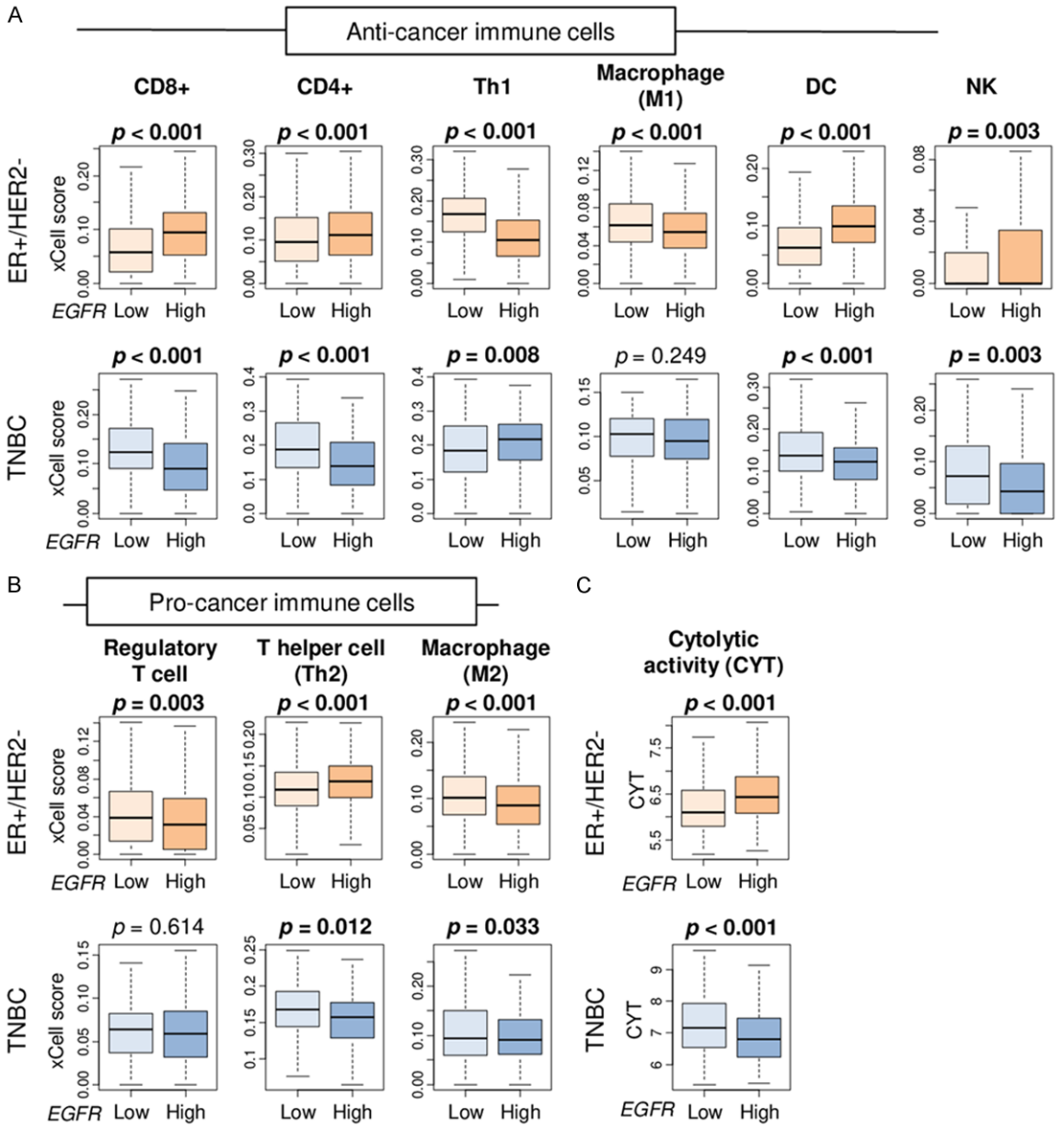

Figure S5. Association of *EGFR* expression with immune fraction and cytolytic activity in each breast cancer subtype in the METABRIC cohort. (A) Boxplots of the infiltrating fraction of anti-cancer immune cells; $CD8^+$ T cells, $CD4^+$ T cells, T helper type1 (Th1) cells, M1 macrophages, dendritic cells (DC), and natural killer (NK) T cells, and (B) pro-cancer immune cells; regulatory T cells, T helper type2 (Th2) cells, and M2 macrophages, and (C) cytolytic activity (CYT) score by high and low *EGFR* expression in each breast cancer subtype. The two groups were divided by median value of *EGFR* expression within each cohort. *P*-values were calculated by Mann-Whitney U test.

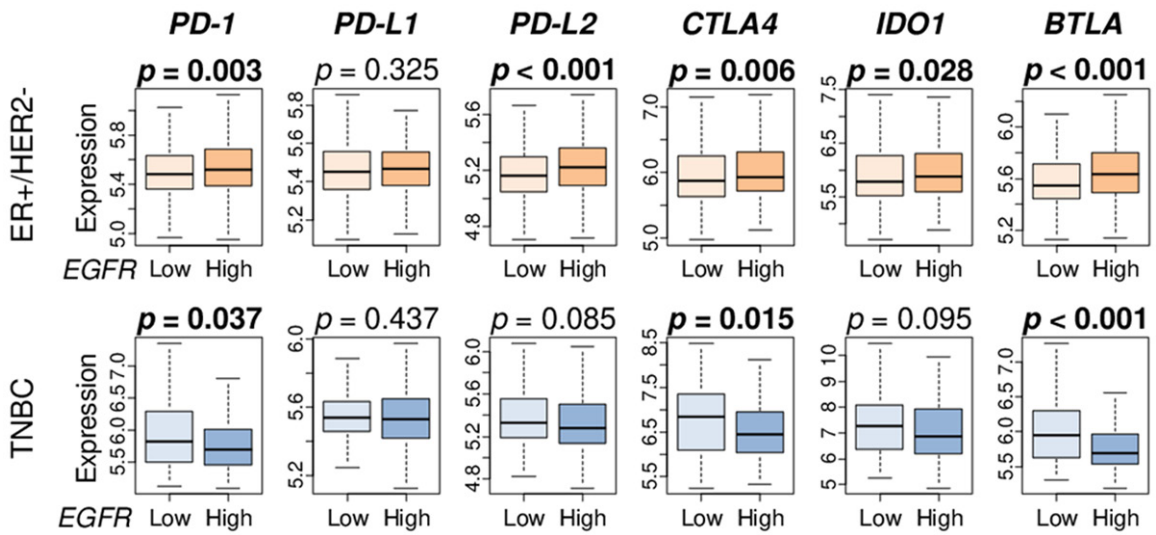

Figure S6. Association between *EGFR* and immune checkpoint molecules expression in each breast cancer subtype in METABRIC cohort. Boxplots of expression of immune checkpoint molecules; *PD-1, PD-L1, PD-L2, CTLA4, IDO1,* and *BTLA.*