Novel Immune Subtypes Identification of HER2 Positive Breast Cancer Based On Immunogenomic Landscape

Lingli Huang

Jiangsu Institute of Cancer Research

Xin Liu

Jiangsu Institute of Cancer Research

Li Li

Jiangsu Institute of Cancer Research

Lei Wang

Jiangsu Institute of Cancer Research

Nan Wu

Jiangsu Institute of Cancer Research

Zhixian Liu (✉ liuzhixian@njmu.edu.cn)

Jiangsu Institute of Cancer Research <https://orcid.org/0000-0003-4710-1442>

Research Article

Keywords: HER2 positive breast cancer, tumor immunity, immunogenomic profiling, classification, machine learning, mutations

Posted Date: December 6th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1088499/v1>

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

[Read Full License](#)

Abstract

Background: HER2 positive BC is heterogeneous. But few studies discussed the classification of HER2 positive BC based on immune-related signatures.

Methods: Using two publicly BC genomics datasets, we classified HER2 positive BC based on 33 immune-related signatures and used unsupervised machine learning methods to predict and perform the classification.

Results: We grouped three HER2 positive BC subtypes that we called Immune-High (IM-H), Immune-Medium (IM-M), and Immune-Low (IM-L), and manifested this categorization was predictable, duplicable and reliable by analyzing another dataset. Compared to other subtypes, IM-H had a higher immune cell infiltration level and stronger anti-tumor immune activities, as well as better clinical survival outcome. Besides these signatures, there were some cancer-related pathways which were hyperactivated in IM-H, including cytokine-cytokine receptor interactions, antigen processing and presentation pathways, natural killer cell-mediated cytotoxicity, Th1 and Th2 cell differentiation, chemokine signaling pathway, Th17 cell differentiation, B and T cell receptor signaling, NF-kappa B signaling, PD-L1 expression and PD-1 checkpoint pathway in cancer, TNF signaling, IL-17 signaling, NOD-like receptor signaling and Toll-like receptor signaling. By contrast, IM-L showed depressed immune-related signatures and enhanced activation of lysosylphosphatidylinositol

-anchor biosynthesis and mismatch repair. Moreover, we discovered a gene co-expression network focused on eight transcription factor genes (EOMES, TBX21, GFI1, IRF4, POU2AF1, CIITA, FOXP3 and TOX) and one tumor suppress gene (PRF1), which were closely related with tumor immune.

Conclusion: We identified three HER2 positive BC subtypes based on immune-related signatures, which had potential clinical implications and promoted the optimal stratification of HER2 positive BC responsive to immunotherapy.

1. Introduction

Human epidermal growth factor receptor 2 (HER2) positive BC (BC) patients account for about 15 to 30 percent of the total number of BC [1]. HER2 overexpression plays a key role in tumorigenesis and metastasis [2, 3]. HER2 positive BC is associated with aggressive biology, reduced responses to traditional therapies, and has a poor prognosis with a median survival of only 2-3 years [4]. Several studies even have shown that HER2 positivity is associated with a biological propensity to metastasize to the brain [5, 6].

There have been some targeted-therapy-based approaches to HER2 positive BC treatment such as trastuzumab, which can specifically bind to the extracellular domain of the HER2 receptor and interfere with the heterodimers formation between HER2 and other ErbB family members, thereby inhibiting tumor cell proliferation and promoting tumor cell apoptosis [7–9]. The presence of trastuzumab and other anti-

HER2 therapeutics has significantly altered the treatment of HER2 positive BC, and HER2 positivity is no longer an absolute symbol of poor prognosis [10]. However, as trastuzumab can't pass through the blood-brain barrier, central nervous system (CNS) becomes a "shelter" for the metastasis of HER2 positive BC, increasing the risk of brain metastases in patients treated with trastuzumab. Although the anti-HER2-targeted therapy has dramatically improved the prognosis of HER2 positive BC patients, the metastasis and recurrence of HER2 positive metastatic BC is still a problem in the clinical. Thus, it is urgent to improve the prognosis and reduce the risk of recurrence in HER2 positive BC patients.

Recently, cancer immunotherapy has achieved huge successes in treating various types of malignant tumors [11]. Especially, the use of immune checkpoints inhibitors has achieved great clinical successes in multiple tumors [12]. Yet, present immunotherapies are unfortunately only benefit to a few patients [13]. Currently, some biomarkers that respond to cancer immunotherapy have been explored, such as neoantigen, tumor mutation burden (TMB), mismatch repair (MMR) deficiency, and PD-1/PD-L1 or cytolytic markers expression in the tumor immune microenvironment (TIM) which positively respond to immune checkpoint blockade in various cancers [14–16]. But there were very few studies which were correlated the HER2 status with tumor immunotherapy response. In the research, we divided HER2 positive BC into three subtypes based on the immunogenomic signatures: Immune-High (IM-H), Immune-Medium (IM-M), and Immune-Low (IM-L). We confirmed the reliability and reproducibility of this classification in three independent and different large-scale datasets with machine learning approach. In addition, we identified the specific clinical features and molecular characters of each subtype, including genes, gene ontology and networks, copy number alterations, signaling pathways. And we believed that the identification of HER2 positive BC subtypes based on immune-related signatures may promote the optimal stratification of HER2 positive BC patients responsive to immunotherapy and provide more suitable therapy selection for HER2 positive BC patients.

2. Materials And Methods

2.1. Unsupervised hierarchical clustering.

Based on the expression levels of the 33 immune-related signatures[17], we performed unsupervised hierarchical clustering analysis by the single-sample gene-set enrichment analysis (ssGSEA) [18] and grouped the datasets. Euclidean distance was used to recognize and analyze the clustering results. Eventually, HER2 positive BC patients were grouped into Immune-High (IM-H), Immune-Medium (IM-M), and Immune-Low (IM-L) subtypes.

2.2. Evaluation of immune cell infiltration level, tumor purity, stromal content, tumor aneuploid and TMB in HER2 positive BC.

ESTIMATE [19] and ABSOLUTE [20] algorithm was used to estimate immune infiltration, stromal levels and tumor purity, aneuploidy in the two datasets (TCGA and METABRIC) respectively among three groups of HER2 positive BC patients. TIMER [21] and CIBERSORT [22] were used to calculate the 6 major immune

cells and proportions of 22 human leukocyte cell subsets. In addition, we defined the TMB as the total count of somatic mutations in the tumor.

2.3. Gene-set enrichment analysis (GSEA) and KEGG analysis.

We performed gene-set enrichment analysis by GSEA (R packages “clusterprofiler”) [23]. KEGG [24] analysis were used to identify the activated pathways in IM-H and IM-L (false discovery rate, FDR < 0.05). We selected the common pathways which were identified in both datasets.

2.4. Somatic copy number alteration (SCNA) analysis.

The copy number of HER2 positive BC from TCGA was analyzed using GISTIC 2.0 [25]. Then we compared the SCNA levels between Immunity_H and Immunity_L subtypes.

2.5. Weighted correlation network analysis (WGCNA) identified subtype-specific gene modules.

WGCNA [26] was applied to character the gene modules and hub genes which were closely correlated with immunogenicity. We built the heatmap of correlation between gene modules which were identified and the clinical trait. The highest correlation module was visualized using String (<https://string-db.org/>). The connectedness weight between genes was > 0.25.

2.6. Classification accuracy prediction.

We transformed each ssGSEA score x_i into x_i' by the equation $x_i' = (x_i - x_{\min}) / (x_{\max} - x_{\min})$, in which $x_{\min}$ means the minimum of the scores of HER2 positive BC samples, and $x_{\max}$ means the maximum of the scores. Then we used Random Forest (RF) to classify the HER2 positive BC subtypes, and assessed the classification by the accuracy and the weighted F-score. We lastly performed the classification in Weka (<https://www.cs.waikato.ac.nz/ml/>).

3. Results

3.1. Immunogenomic profiling identifies three HER2-positive BC subtypes.

We first analyzed 33 immune-related gene sets representing different immune cell types and functions (Table S1). Then we quantified the activity of these gene sets in the BC samples with the ssGSEA score [18]. Based on the ssGSEA scores of these 33 gene-sets, we clustered HER2 positive BC in two large-scale datasets (TCGA and METABRIC). Notably, the clustering results were similar in the two datasets that the patients were clearly divided into three clusters, and we termed these three clusters as IM-H, IM-M, and IM-L (Figure 1A). Then we compared the differences of tumor microenvironment between the three subtypes in the two datasets by evaluating the immune score, the stromal score and the tumor purity score using

ESTIMATE [19]. We found that the immune score and the stromal score showed the significantly higher level, and tumor purity showed the significantly lower level in IM-H compared to IM-L subtype in both two datasets (Kruskal-Wallis test, immune score: $P = 2.2e-16$ (TCGA), $P = 2.2e-16$ (METABRIC); stromal score: $P = 3.78e-11$ (TCGA), $P = 3.79e-10$, (METABRIC); tumor purity: $P = 2.2e-16$ (TCGA), $P = 2.2e-16$ (METABRIC)), with the immune score and the stromal score followed the decreasing trend from IM-H to IM-L ($IM-H > IM-M > IM-L$) and tumor purity followed the increasing trend from IM-H to IM-L ($IM-H < IM-M < IM-L$) (Kruskal-Wallis test, $P < 0.001$) (Figure 1B). Furthermore, we also evaluated the tumor purity and the tumor ploidy through ABSOLUTE [20], and the result displayed that there was a higher tumor purity in IM-L than IM-H subtype (Kruskal-Wallis test, $P = 3.66e-12$ (TCGA)), with the trend that tumor purity decreasing from IM-H to IM-L ($IM-H < IM-M < IM-L$), but there was no significance in the tumor ploidy among these three subtypes. These results indicated that IM-H had the highest levels of immune score and stromal score than other subtypes, while IM-L subtype had the highest number of tumor cells.

Then, we further found that these three HER2 positive BC subtypes had distinct clinical prognosis by survival analysis. The IM-H subtype showed the best prognosis than the IM-M and IM-L subtype, and IM-L subtype had the worst prognosis (Figure 1C), indicating that enhanced immune activity in HER2 positive BC were related with more beneficial clinical prognosis.

The expression level of PD-L1 can be a biomarker for the response to PD-1/PD-L1 pathway inhibition, and its expression on tumor cells is important in tumor immune escape [27, 28]. We further examined the expression of PD-L1 in the three HER2 positive BC subtypes, and we found that PD-L1 showed the highest expression levels in IM-H compared to IM-M and IM-L subtype (ANOVA test, $P < 0.05$) (Figure 1D), suggesting that the IM-H subtype might have the best response to anti-PD-L1 immunotherapy among the HER2 positive BC subtypes, as PD-L1 expression tends to have positive correlation with immunotherapeutic responsiveness.

3.2. The immune microenvironment in IM-H subtype was more active compared to other subtypes.

To compare the differences of immune components between three subtypes, we analyzed the lymphocyte infiltration percentage based on the BC pathological slides data in TCGA. We found that there was a significantly higher lymphocyte infiltration percentage in IM-H than in IM-M and IM-L (Kruskal-Wallis test, $P = 0.017$, Figure 2A). We then used TIMER [21] and CIBERSORT [22] to evaluated the differences of major immune cells and proportions of 22 leukocyte cell subsets. The TIMER results showed a much higher density of immune cells in IM-H subtype, such as B cells, CD4+ T cells, CD8+ T cells, neutrophil cells, and myeloid dendritic cell (Figure 2B). Furthermore, the CIBERSORT results showed that IM-H had significantly higher numbers of CD4+ T cells, CD8+ T cells, NK cells, and M1 macrophages (Mann-Whitney U test, $FDR < 0.05$);. On the contrary, IM-L had significantly higher numbers of M2 macrophages (Mann-Whitney U test, $FDR < 0.05$) (Figure 2C). This indicated that IM-H subtype had more elevated immune activity and more active immune microenvironment. Furthermore, M1 macrophages that boost inflammation had much higher ratios in IM-H, while M2 macrophages which suppress

inflammation and had higher ratios in IM-L. This finding reminded that the anti-tumor immune response in IM-L subtype was suppressed, thus leading to aggressive, and associated with the worse prognosis.

Strikingly, all of HLA genes were highly expressed in IM-H than in other subtypes with the expression level decreasing from IM-H to IM-L (IM-H > IM-M > IM-L, ANOVA test, $P < 0.05$, Figure 2D). Moreover, we also compared the expression level of various immune cell subpopulation marker genes [29] in diverse subtypes, which also showed much higher expression level in IM-H subtype than in IM-L subtype, e.g. *CD8A* (cytotoxic T cell), *CD45RO* (memory T cell), *CD20* (B cell), *CXCR5* (Tfh cells), *FOXP3* (Treg), *CD3* (T cells), *CD68* (macrophages), *FOXP3* (Treg cells), *CD57* (activated T or NK cells), *ENG* (blood vessels), *IL-17* (Th17 cell), *CD1A* (iDC), and *IL3RA* (pDC) [29] (Figure 2E).

3.3. Subtype-specific pathways, gene ontology, and networks were identified in HER2 positive BC.

We identified plenty of KEGG [24] pathways which were enriched in IM-H and IM-L subtype with GSEA (Figure 3A). Notably, in IM-H subtype, the immune-associated pathways were more enriched, including cytokine-cytokine receptor interactions, antigen processing and pathways, natural killer cell-mediated cytotoxicity, Th1 and Th2 cell differentiation, chemokine signaling pathway, Th17 cell differentiation, B and T cell receptor signaling, NF-kappa B signaling, PD-L1 expression and PD-1 checkpoint pathway in cancer, TNF signaling, IL-17 signaling, NOD-like receptor signaling and Toll-like receptor signaling (Figure 3A). These results suggested that IM-H subtype had enhanced immune activity. Additionally, we further identified the hyperactivated cancer-related pathways in IM-H, such as ECM-receptor interaction, cell adhesion molecules (CAMs), calcium signaling, MAPK signaling, PI3K-Akt signaling, and RAS signaling (Figure 3A). Conversely, IM-L subtype showed hyperactivated mismatch repair and glycosylphosphatidylinositol (GPI)-anchor biosynthesis pathways (Figure 3A), indicating that higher genomic instability in IM-L and the activities of mismatch repair and GPI-anchor biosynthesis pathways negatively correlated with HER2-positive BC immunity.

Moreover, WGCNA [26] was used to identify a gene modules (dark blue color) that was significant in IM-H subtype (Figure 3B). Among them, the immune response was markedly elevated in IM-H ($P = 5.0e-19$, $R = 0.64$), while it was depressed in IM-L ($P = 4.0e-31$, $R = -0.77$), suggesting that high immune response could be used to predict better survival outcomes in HER2 positive BC patients (OS: $P = 0.01$, $R = 0.2$, DFS: $P=0.002$, $R=0.22$). This finding is in accordance with our earlier result that the IM-H subtype is related with better clinical prognosis than the other subtypes. The trend is highly similar with the results in METABRIC.

In addition, we got 139 hub genes from the dark blue module, including one tumor suppress gene (*PRF1*) and eight transcription factor (TF) genes (*EOMES*, *TBX21*, *GFI1*, *IRF4*, *POU2AF1*, *CIITA*, *FOXP3* and *TOX*). These 8 TFs interact with each other and form a sub-network with the genes they regulate (Figure 3C).

3.4. Identification of HER2 positive BC subtype-specific genomic features which were closed related with clinical features.

We further identified the genes mutated > 10% to distinguish the genomic differences between IM-H and IM-L, and identified top 10 genes mutated in IM-H and IM-L subtypes (Figure 3D, 3E). Among them, *TP53* mutation rate was the highest in HER2 positive BC, and IM-H had the significantly higher *TP53* mutation rate compared to IM-L. *PIK3CA* was the second highest mutated gene and IM-H also had the significantly higher *PIK3CA* mutation rate compared to IM-L. Furthermore, we identified several genes which were significantly higher mutated in IM-H, e.g. *SPEN*, *TNRC6C*, *SPTA1*, and *PKD1L1*. In addition, in pathological level, patients who were in advanced stage (pStage was "stage iii/iv") have a significantly higher proportion than patients in early stage in IM-H subtype (pStage was "stage i / ii") (Fisher's exact test, $P = 0.038$, Figure 3D), suggesting that patients in advanced stage had higher immunogenicity, and may be more suitable for immunotherapy. TMB predicted immunotherapy response in clinical. We further calculated the differences of TMB among three subtypes. The result showed there was no significant difference in IM-H and IM-L (Wilcoxon signed-rank test, $P = 0.1$), suggesting that the differences of immunogenicity among the HER2 positive BC subtypes were not related with TMB and neoantigens.

Moreover, we also compared SCNA levels among three subtypes, and the results showed that the SCNAs levels were significantly lower in IM-H than in IM-L (Wilcoxon signed-rank test, $P = 0.19$, 0.009 , 0.03 (TCGA), $P = 2.56e-06$, $2.71e-08$, $1.51e-09$ (METABRIC), for comparisons of amplification, deletion, and total alteration frequencies, respectively). Our previous study showed that pathways related to MMR and GPI-anchor biosynthesis were enriched in IM-L, indicating that the chromosome instability was much higher in IM-L subtype compared to IM-H. It was in agreement with previous results that tumor aneuploidy was negatively correlated with tumor immune infiltration [30]. Consistent with prior research, high clonal heterogeneity significantly was significantly correlated with reduced tumor immunity [31].

3.5. Class prediction of HER2 positive BC subtypes based on immunogenomic profiling.

To predict classification accuracy, we first employed 10-fold cross validation (CV) to estimate the classification property by METABRIC dataset which was regarded as the training set, and then the TCGA and another BC dataset were regarded as the verification set. In the METABRIC dataset, the accuracy of 10-fold CV was 95%. In TCGA and GSE75688, the accuracy of classification accuracies were 70% and 84%. In these three classifications, the weighted F-scores were 95%, 75% and 83% for the datasets (METABRIC, TCGA, and GSE75688, respectively) (Figure 4A). Lastly, we separately predicted the dataset GSE75688 and grouped the BC patients into three subtypes: IM-H, IM-M and IM-L. The results of ESTIMATE indicated that there was a significantly higher immune score in IM-H subtype and a significantly lower immune score in IM-L subtype (Mann-Whitney U test, $P < 0.001$, Figure 4B). These findings suggested that classification of HER2-positive BC based on the immunogenomic profiling was stable and predictable.

4. Discussion

BC subtypes have been identified by analyzing genomic data in some studies [32]. However, few studies have classified HER2-positive BC specifically based on immune signatures. In this study, we focused on identifying immune-related HER2 positive BC subtypes using immune-related gene sets and verifying classification reliability and accuracy to fill this knowledge gap of HER2 positive BC. Our studies showed that HER2 positive BC could be classified into three subtypes: IM-H, IM-M, and IM-L. We then verified that this classification was reproducible and predictable. IM-H subtype was enriched in immune-related signatures, as well as in many cancer-related pathways, including cytokine-cytokine receptor interactions, antigen processing and presentation pathways, natural killer cell-mediated cytotoxicity, Th1 and Th2 cell differentiation, chemokine signaling pathway, Th17 cell differentiation, B and T cell receptor signaling, NF-kappa B signaling, PD-L1 expression and PD-1 checkpoint pathway in cancer, TNF signaling, IL-17 signaling, NOD-like receptor signaling and Toll-like receptor signaling. In comparison, the IM-L subtype was impoverished in immune signatures, but enriched in GPI-anchor biosynthesis and mismatch repair. These indicate that negative correlation exists between mismatch repair pathway activity and immune activity in HER2 positive BC, because mismatch repair deficiency often leads to increased tumor immunity [33], consisting with our results that mismatch repair pathway was more enriched in IM-L. In general, these results showed that there are potential positive or negative correlations between pathway activity and immune activity in HER2 positive BC.

IM-H subtype showed significantly stronger immune cell infiltration and immune activity than other subtypes: e.g., stromal score and immune score were significantly lower in IM-L subtype, while tumor purity was significantly lower for IM-L subtype (Figure 1B). IM-H subtype also had significantly higher lymphocyte infiltration percentages and immune cell densities (Figure 2A, 2B). By further calculating the proportions of 22 immune cell subsets in HER2 positive BC, we found that the numbers of CD4⁺ T cells, CD8⁺ T cells, NK cells, and M1 macrophages were significantly increased in IM-H compared with that in IM-M and IM-L (Figure 2C). By comparison, IM-L had significantly higher numbers of M2 macrophages (Figure 2C). The results reinforced elevated immune activity and more active immune microenvironment in IM-H subtype. Furthermore, M1 macrophages can result in inflammation, while M2 macrophages suppress inflammation. This is in concordance with our research demonstrating that IM-H has more beneficial clinical outcomes compared to other subtypes (Figure 1C), consisting with previous studies that the density of tumor-infiltrating lymphocytes and survival prognosis have a positive correlation in various cancers [34].

In addition, all of HLA genes and various immune cell subpopulation marker genes were highly expressed in IM-H than in other subtypes. Nonetheless, TMB or neoantigen load in IM-H subtype was not higher than that in other subtypes, suggesting that the immunogenicity differences between each HER2 positive BC subtype cannot be ascribe to TMB and neoantigens.

Then we identified a gene module which was significantly related with IM-H subtype, and all the genes in the module were extracted and analyzed. We subsequently found 9 important genes, including 8 TFs (*EOMES*, *TBX21*, *GFI1*, *IRF4*, *POU2AF1*, *CIITA*, *FOXP3* and *TOX*) and one tumor suppress gene (*PRF1*), which were closely related with tumor immune. The T-box transcription factors *EOMES* and *TBX21* were

identified as pivotal drivers of immune cell development and cytolytic function [35]. *EOMES* was reported to be conducive to the high expression of PD-1 and CTLA4 in tumor-infiltrated T cells [36]. *TBX21* and *TOX* were both proved to be associated with cancer progression by regulating immune response [37, 38]. *GFI1*, a zinc-finger transcriptional repressor, not only controlled the development of T- or B-lymphoid, but also played a critical role in their maturation, activation and effector function [39, 40]. *IRF4* was shown to regulate immune responses by participating in lymphocyte activation and plasma cells to secrete immunoglobulin [41, 42]. *POU2AF1* was a B-cell-specific coactivator to stimulate immunoglobulin gene transcription, which may affect the humoral immune response [43]. Similarly, *CIITA* was also involved in immune response by regulating the expression of MHC II genes [44]. *FOXP3* played an important role in the control of immune homeostasis mediated by T regulatory cells as a multifaceted transcription factor [45]. *PRF1* belonging to the membrane-attack-complex/PRF protein family, was essentially related to immune cell infiltration [46], as well as involved in the granule-dependent killing activities of cytotoxic T lymphocytes and NK cells [47].

Moreover, we compared the SCNA levels in IM-H subtype with that in IM-L subtype. IM-H had significantly lower SCNAs levels than IM-L, supporting the correlation between high tumor aneuploidy and lessened tumor immune infiltration. Besides, patients in advanced stage had higher immunogenicity in pathological level than patients in early stage in IM-H subtype. These results suggest patients in advanced stage may be more suitable for immunotherapy.

Currently, immunotherapy for BC is an active research field [48]. However, clinical trials of preliminary immunotherapy have not been significantly improved in HER2 positive BC patients. Expression level of PD-L1 was known as a biomarker to predict the response of PD-1/PD-L1-directed immunotherapy. Our results showed that, PD-L1 expression in IM-H subtype was significantly higher than that in other subtypes, suggesting that IM-H subtype can be more responsive to immune checkpoint inhibitors, thus providing evidence for anti-HER2 positive BC immunotherapy based on immune checkpoint inhibitors. Therefore, IM-H subtype would be a “hot” tumor which could have positive response to immunotherapy.

5. Conclusion

We have identified three HER2 positive BC subtypes based on immune-related signatures and validated it, which may help to stratify HER2 positive BC patients to identify those who respond to immunotherapy.

Declarations

Funding

This work was supported by Natural Science Foundation of Jiangsu Province (BK20201090) and China Postdoctoral Science Foundation (2021M691338) to ZL.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

All authors contributed to the study conception and design. ZL conceived the project. LH and XL researched and analyzed the data. NW and LW contributed to the discussion. ZL wrote the manuscript. XL, LL, NW, and LW revised the manuscript. All authors contributed to the article and approved the submitted version.

Data Availability

Three publicly HER2 positive BC genomic datasets were used: TCGA, METABRIC, and GSE75688. The TCGA dataset was from the TCGA website (<https://portal.gdc.cancer.gov/>). The METABRIC dataset was obtained from http://www.cbioportal.org/study?id=brca_metabricsummary. The GSE75688 datasets were downloaded from <https://www.ncbi.nlm.nih.gov/geo/>. 33 immunesignatures (immune gene sets) were from the publications [17, 29] (Table S1).

References

1. Badowska-Kozakiewicz, A.M., M. Sobol, J. Patera and W. Kozłowski, *Immunohistochemical evaluation of human epidermal growth factor receptor 2 and estrogen and progesterone receptors in invasive breast cancer in women*. Arch Med Sci, 2013. **9**(3): p. 466–71.
2. Benz, C.C., G.K. Scott, J.C. Sarup, R.M. Johnson, D. Tripathy, E. Coronado, et al., *Estrogen-dependent, tamoxifen-resistant tumorigenic growth of MCF-7 cells transfected with HER2/neu*. Breast Cancer Res Treat, 1992. **24**(2): p. 85–95.
3. Chazin, V.R., M. Kaleko, A.D. Miller and D.J. Slamon, *Transformation mediated by the human HER-2 gene independent of the epidermal growth factor receptor*. Oncogene, 1992. **7**(9): p. 1859–66.
4. Pegram, M.D., G. Pauletti and D.J. Slamon, *HER-2/neu as a predictive marker of response to breast cancer therapy*. Breast Cancer Research and Treatment, 1998. **52**(1-3): p. 65–77.
5. Gabos, Z., R. Sinha, J. Hanson, N. Chauhan, J. Hugh, J.R. Mackey, et al., *Prognostic significance of human epidermal growth factor receptor positivity for the development of brain metastasis after newly diagnosed breast cancer*. J Clin Oncol, 2006. **24**(36): p. 5658–63.
6. Hedayatzadeh-Omran, A., A. Rafiei, R. Alizadeh-Navaei, M. Tehrani, R. Valadan, K. Moradzadeh, et al., *Role of HER2 in brain metastasis of breast cancer: a systematic review and meta-analysis*. Asian Pac J Cancer Prev, 2015. **16**(4): p. 1431–4.
7. Molina, M.A., J. Codony-Servat, J. Albanell, F. Rojo, J. Arribas and J. Baselga, *Trastuzumab (herceptin), a humanized anti-Her2 receptor monoclonal antibody, inhibits basal and activated Her2 ectodomain cleavage in breast cancer cells*. Cancer Res, 2001. **61**(12): p. 4744–9.
8. Cuello, M., S.A. Ettenberg, A.S. Clark, M.M. Keane, R.H. Posner, M.M. Nau, et al., *Down-regulation of the erbB-2 receptor by trastuzumab (herceptin) enhances tumor necrosis factor-related apoptosis-*

- inducing ligand-mediated apoptosis in breast and ovarian cancer cell lines that overexpress erbB-2*. Cancer Research, 2001. **61**(12): p. 4892–4900.
9. Junttila, T.T., R.W. Akita, K. Parsons, C. Fields, G.D.L. Phillips, L.S. Friedman, et al., *Ligand-Independent HER2/HER3/PI3K Complex Is Disrupted by Trastuzumab and Is Effectively Inhibited by the PI3K Inhibitor GDC-0941 (vol 15, pg 429, 2009)*. Cancer Cell, 2011. **20**(6): p. 818-818.
 10. Dawood, S., K. Broglio, A.U. Buzdar, G.N. Hortobagyi and S.H. Giordano, *Prognosis of women with metastatic breast cancer by HER2 status and trastuzumab treatment: an institutional-based review*. J Clin Oncol, 2010. **28**(1): p. 92–8.
 11. Del Paggio, J.C., *Immunotherapy: Cancer immunotherapy and the value of cure*. Nat Rev Clin Oncol, 2018. **15**(5): p. 268–270.
 12. Li, X., C. Shao, Y. Shi and W. Han, *Lessons learned from the blockade of immune checkpoints in cancer immunotherapy*. J Hematol Oncol, 2018. **11**(1): p. 31.
 13. Braun, D.A., K.P. Burke and E.M. Van Allen, *Genomic Approaches to Understanding Response and Resistance to Immunotherapy*. Clin Cancer Res, 2016. **22**(23): p. 5642–5650.
 14. Chan, T.A., J.D. Wolchok and A. Snyder, *Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma*. N Engl J Med, 2015. **373**(20): p. 1984.
 15. Rizvi, N.A., M.D. Hellmann, A. Snyder, P. Kvistborg, V. Makarov, J.J. Havel, et al., *Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer*. Science, 2015. **348**(6230): p. 124–8.
 16. Taube, J.M., A. Klein, J.R. Brahmer, H. Xu, X. Pan, J.H. Kim, et al., *Association of PD-1, PD-1 ligands, and other features of the tumor immune microenvironment with response to anti-PD-1 therapy*. Clin Cancer Res, 2014. **20**(19): p. 5064–74.
 17. Liu, Z., M. Li, Z. Jiang and X. Wang, *A Comprehensive Immunologic Portrait of Triple-Negative Breast Cancer*. Transl Oncol, 2018. **11**(2): p. 311–329.
 18. Hanzelmann, S., R. Castelo and J. Guinney, *GSVA: gene set variation analysis for microarray and RNA-seq data*. BMC Bioinformatics, 2013. **14**: p. 7.
 19. Yoshihara, K., M. Shahmoradgoli, E. Martinez, R. Vegesna, H. Kim, W. Torres-Garcia, et al., *Inferring tumour purity and stromal and immune cell admixture from expression data*. Nat Commun, 2013. **4**: p. 2612.
 20. Carter, S.L., K. Cibulskis, E. Helman, A. McKenna, H. Shen, T. Zack, et al., *Absolute quantification of somatic DNA alterations in human cancer*. Nat Biotechnol, 2012. **30**(5): p. 413–21.
 21. Li, T., J. Fan, B. Wang, N. Traugh, Q. Chen, J.S. Liu, et al., *TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells*. Cancer Res, 2017. **77**(21): p. e108-e110.
 22. Newman, A.M., C.L. Liu, M.R. Green, A.J. Gentles, W. Feng, Y. Xu, et al., *Robust enumeration of cell subsets from tissue expression profiles*. Nat Methods, 2015. **12**(5): p. 453–7.
 23. Subramanian, A., P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles*.

- Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545–50.
24. Kanehisa, M., M. Furumichi, M. Tanabe, Y. Sato and K. Morishima, *KEGG: new perspectives on genomes, pathways, diseases and drugs*. Nucleic Acids Res, 2017. **45**(D1): p. D353-D361.
 25. Mermel, C.H., S.E. Schumacher, B. Hill, M.L. Meyerson, R. Beroukhi and G. Getz, *GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers*. Genome Biol, 2011. **12**(4): p. R41.
 26. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis*. BMC Bioinformatics, 2008. **9**: p. 559.
 27. Zhou, Z.Q., J.J. Zhao, Q.Z. Pan, C.L. Chen, Y. Liu, Y. Tang, et al., *PD-L1 expression is a predictive biomarker for CLK cell-based immunotherapy in postoperative patients with breast cancer*. J Immunother Cancer, 2019. **7**(1): p. 228.
 28. Juneja, V.R., K.A. McGuire, R.T. Manguso, M.W. LaFleur, N. Collins, W.N. Haining, et al., *PD-L1 on tumor cells is sufficient for immune evasion in immunogenic tumors and inhibits CD8 T cell cytotoxicity*. J Exp Med, 2017. **214**(4): p. 895–904.
 29. Bindea, G., B. Mlecnik, M. Tosolini, A. Kirilovsky, M. Waldner, A.C. Obenauf, et al., *Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer*. Immunity, 2013. **39**(4): p. 782–95.
 30. Davoli, T., H. Uno, E.C. Wooten and S.J. Elledge, *Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy*. Science, 2017. **355**(6322).
 31. Thorsson, V., D.L. Gibbs, S.D. Brown, D. Wolf, D.S. Bortone, T.H. Ou Yang, et al., *The Immune Landscape of Cancer*. Immunity, 2018. **48**(4): p. 812-830 e14.
 32. He, Y., Z. Jiang, C. Chen and X. Wang, *Classification of triple-negative breast cancers based on Immunogenomic profiling*. J Exp Clin Cancer Res, 2018. **37**(1): p. 327.
 33. Le, D.T., J.N. Durham, K.N. Smith, H. Wang, B.R. Bartlett, L.K. Aulakh, et al., *Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade*. Science, 2017. **357**(6349): p. 409–413.
 34. Garnelo, M., A. Tan, Z. Her, J. Yeong, C.J. Lim, J. Chen, et al., *Interaction between tumour-infiltrating B cells and T cells controls the progression of hepatocellular carcinoma*. Gut, 2017. **66**(2): p. 342–351.
 35. Knox, J.J., G.L. Cosma, M.R. Betts and L.M. McLane, *Corrigendum: Characterization of T-bet and Eomes in Peripheral Human Immune Cells*. Front Immunol, 2016. **7**: p. 337.
 36. He, H., Y. Yi, X. Cai, J. Wang, X. Ni, Y. Fu, et al., *Down-regulation of EOMES drives T-cell exhaustion via abolishing EOMES-mediated repression of inhibitory receptors of T cells in liver cancer*. J Cell Mol Med, 2021. **25**(1): p. 161–169.
 37. Reppert, S., I. Boross, M. Koslowski, O. Tureci, S. Koch, H.A. Lehr, et al., *A role for T-bet-mediated tumour immune surveillance in anti-IL-17A treatment of lung cancer*. Nat Commun, 2011. **2**: p. 600.
 38. Yu, X. and Z. Li, *TOX gene: a novel target for human cancer gene therapy*. Am J Cancer Res, 2015. **5**(12): p. 3516–24.

39. Hock, H., M.J. Hamblen, H.M. Rooke, D. Traver, R.T. Bronson, S. Cameron, et al., *Intrinsic requirement for zinc finger transcription factor Gfi-1 in neutrophil differentiation*. Immunity, 2003. **18**(1): p. 109–20.
40. Möröy, T. and C. Khandanpour, *Growth factor independence 1 (Gfi1) as a regulator of lymphocyte development and activation*. Semin Immunol, 2011. **23**(5): p. 368–78.
41. Shaffer, A.L., N.C. Emre, L. Lamy, V.N. Ngo, G. Wright, W. Xiao, et al., *IRF4 addiction in multiple myeloma*. Nature, 2008. **454**(7201): p. 226–31.
42. Nam, S. and J.S. Lim, *Essential role of interferon regulatory factor 4 (IRF4) in immune cell development*. Arch Pharm Res, 2016. **39**(11): p. 1548–1555.
43. Salas, M. and L.A. Eckhardt, *Critical role for the Oct-2/OCA-B partnership in Ig-secreting cells*. J Immunol, 2003. **171**(12): p. 6589–98.
44. Drozina, G., J. Kohoutek, N. Jabrane-Ferrat and B.M. Peterlin, *Expression of MHC II genes*. Curr Top Microbiol Immunol, 2005. **290**: p. 147–70.
45. Lozano, T., N. Casares and J.J. Lasarte, *Searching for the Achilles Heel of FOXP3*. Front Oncol, 2013. **3**: p. 294.
46. Fan, C., H. Hu, Y. Shen, Q. Wang, Y. Mao, B. Ye, et al., *PRF1 is a prognostic marker and correlated with immune infiltration in head and neck squamous cell carcinoma*. Transl Oncol, 2021. **14**(4): p. 101042.
47. Trapani, J.A., K.Y. Thia, M. Andrews, I.D. Davis, C. Gedye, P. Parente, et al., *Human perforin mutations and susceptibility to multiple primary cancers*. Oncoimmunology, 2013. **2**(4): p. e24185.
48. Setordzi, P., X. Chang, Z. Liu, Y. Wu and D. Zuo, *The recent advances of PD-1 and PD-L 1 checkpoint signaling inhibition for breast cancer immunotherapy*. Eur J Pharmacol, 2021. **895**: p. 173867.

Figures

Figure 1

Immune-genomic analysis characterized three different subtypes which showed different phenotypes in TCGA and METABRIC. (A) Unsupervised hierarchical clustering based on immune signature of HER2 positive BC yields three stable subtypes in TCGA and METABRIC: Immune-high (IM-H), Immune-Medium (IM-M) and Immune-low (IM-L). ESTIMATE was used to evaluate the Tumor purity, Stromal score, and Immune score. (B) Comparison of the immune cell infiltration, stromal levels, and tumor purity among HER2 positive BC subtypes. (C) Comparison of clinical survival outcomes among HER2 positive BC subtypes. (D) IM-H had much higher PD-L1 expression level than other HER2 positive BC subtypes (ANOVA test). *P < 0.05, **P < 0.01, ***P < 0.001.

Figure 2

Comparison of the proportions of immune cell subsets among HER2 positive BC subtypes. (A) IM-H had significantly higher degree of lymphocyte infiltration than IM-L based on pathological slides data in TCGA. (B) Comparison of abundance of five major immune cell types by using TIMER among HER2 positive BC subtypes. (C) Compared to other subtypes, IM-H had significantly higher leukocyte cell subset infiltrates which were evaluated with CIBERSORT. (D) Comparison of HLA genes expression level among HER2 positive BC subtypes. (E) Comparison of genes expression level of immune cell subpopulation marker among HER2 positive BC subtypes. *P < 0.05, **P < 0.01, ***P < 0.001.

Figure 3

Characterization of subtype-specific signaling pathways, genomic features, clinical features, gene ontology features, and sub-networks in HER2 positive BC. (A) The enrichment of immune related and cancer related signaling pathways in IM-H and IM-L subtype. (B) Gene modules distinguished HER2 positive BC significantly based on subtype, survival time, or survival status. (C) A subnetwork which was significantly activated in IM-H, focused on eight TFs. (D) Mutation heatmap revealed distribution of highly mutated genes, the differences of genomic and clinical features between IM-H and IM-L subtypes. (E) The status of the top 10 mutated genes in IM-H and IM-L subtypes.

Figure 4

Figure 4

The prediction of HER2 positive BC phenotypes by using the machine learning method. (A) Property in the classification of HER2 positive BC subtypes based on immune signatures. (B) Validation and comparison of the immune score in HER2 positive subtypes (IM-H and IM-L) in the dataset GSE75688 predicted with the machine learning method. *: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1.pdf](#)