Subgrouping breast cancer patients based on immune evasion mechanisms unravels a high involvement of transforming growth factor-beta and decoy receptor 3

Mayassa J. Bou-Dargham[1‡]*, **Yuhang Liu**[2‡], **Qing-Xiang Amy Sang**[1,3]*, **Jinfeng Zhang**[2]*

1 Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida, United States of America, **2** Department of Statistics, Florida State University, Tallahassee, Florida, United States of America, **3** Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida, United States of America

‡ These authors are shared first authors on this work.
* jinfeng@stat.fsu.edu (JZ); qxsang@chem.fsu.edu (QXAS); mdargham@chem.fsu.edu (MJBD)

Abstract

In the era of immunotherapy and personalized medicine, there is an urgent need for advancing the knowledge of immune evasion in different cancer types and identifying reliable biomarkers that guide both therapy selection and patient inclusion in clinical trials. Given the differential immune responses and evasion mechanisms in breast cancer, we expect to identify different breast cancer groups based on their expression of immune-related genes. For that, we used the sequential biclustering method on The Cancer Genome Atlas RNA-seq breast cancer data and identified 7 clusters. We found that 77.4% of the clustered tumor specimens evade through transforming growth factor-beta (TGF-β) immunosuppression, 57.7% through decoy receptor 3 (DcR3) counterattack, 48.0% through cytotoxic T-lymphocyte-associated protein 4 (CTLA4), and 34.3% through programmed cell death-1 (PD-1). TGF-β and DcR3 are potential novel drug targets for breast cancer immunotherapy. Targeting TGF-β and DcR3 may provide a powerful approach for treating breast cancer because 57.7% of patients overexpressed these two molecules. Furthermore, triple-negative breast cancer (TNBC) patients clustered equally into two subgroups: one with impaired antigen presentation and another with high leukocyte recruitment but four different evasion mechanisms. Thus, different TNBC patients may be treated with different immunotherapy approaches. We identified biomarkers to cluster patients into subgroups based on immune evasion mechanisms and guide the choice of immunotherapy. These findings provide a better understanding of patients' response to immunotherapies and shed light on the rational design of novel combination therapies.

Introduction

Cancer cells express markers that differentiate them from normal cells and allow for their detection through immune surveillance and subsequent destruction [1–3]. Antigen-presenting

cells (APCs) phagocytose and present tumor antigens on their cell surface together with class I major histocompatibility complex (MHC-I): HLA-A and HLA-B. The recognition of antigen-MHC-I complexes by naïve cytotoxic T lymphocytes (CTLs; *aka* CD8[+] T cells) results in their activation. Then, trafficking and infiltration of CTLs from the bloodstream to the tumor microenvironment follow, orchestrated by adhesion molecules. Activated CTLs kill cancer cells via (i) granular exocytosis: perforins and granzymes or (ii) death ligand-death receptor-mediated apoptosis. The latter occurs through the binding of death ligand cluster of differentiation 95 (CD95L; *aka* FasL) to the death receptor CD95 (Fas) on tumor cell surface [4,5]. This process from antigen presentation to apoptosis induction by CTL is known as the cancer-immunity cycle [6].

Despite immune surveillance, cancer cells manage to evade immune destruction [7]. There are multitudes of evasion mechanisms that govern the suppressive nature of the tumor micro-environment and understanding these mechanisms is an active area of study. Several clinical trials targeting the molecules that inhibit anti-tumor immune response in breast cancer are in progress (S1 File). Given the heterogeneity of breast cancer, several studies have been conducted to investigate the molecular and prognostic differences between the different hormone receptor subtypes and the ductal and lobular subgroups. Triple-negative breast cancer (TNBC) studies have identified intra-subtype heterogeneity where certain patients showed poor prognosis while others responded well to anthracycline-based treatments [8]. Scientific evidence suggested that the clinical outcome for TNBC is affected by tumor-infiltrating immune cells [8,9]. Studies on invasive breast cancer showed that invasive lobular carcinoma (ILC) is 3 times more likely to metastasize and less responsive to neoadjuvant chemotherapy compared to invasive ductal carcinoma (IDC) [10,11]. Furthermore, an "immune-related" subgroup of ILC was identified, characterized by upregulated mRNA expression of PD-1, it's ligand PD-L1, and CTLA4 [12].

Currently, immune checkpoints, CTLA4 and PD-1/PD-L1, are the most intensively studied immune evasion molecules in cancer, and immunotherapies targeting them are probably the most successful ones. However, substantial proportions of PD-L1 positive or CTLA4 positive patients do not respond to the corresponding immunotherapies [13,14]. Thus, these molecules are not reliable biomarkers for the prediction of treatment response [15].

Furthermore, clinical trials are shifting towards combination immunotherapy under the assumption that multiple immune evasion mechanisms may be utilized by a tumor. However, the choice of immunotherapy and combined treatments is poorly guided as they are generally given indiscriminately even though different patients may have different evasion mechanisms.

To address the current knowledge gaps, we sought to gain further insight into immune evasion mechanisms and heterogeneity in breast cancer and provide a better diagnostic to help guide treatments. We used the sequential biclustering method and classification tree algorithm to analyze the expression pattern of The Cancer Genome Atlas (TCGA) RNA-seq data of immune-related genes and identify putative biomarkers, respectively. We identified seven distinct breast cancer groups that represent different combinations of evasion mechanisms (M) and reveal molecular features that provide a better understanding of the evasion mechanisms in breast cancer and suggest potential therapeutic strategies.

Materials and methods

Generating the list of immune genes

Based on the current knowledge of the mechanism of tumor evasion from immune system destruction, a list of 86 genes was generated manually based on the available literature (S2 File). The generated list included genes involved in the cancer-immunity cycle, and tolerance

and immunosuppression-inducing genes (Fig 1). To make sure no important genes were missing, the list was expanded from 86 genes to 1,356 by adding all interacting proteins determined using bioGRID database (https://thebiogrid.org/) [16].

TCGA breast cancer dataset

RNA-seq data of 1,065 breast cancer (BRCA) tumor samples were obtained from TCGA (http://cancergenome.nih.gov/) together with 111 non-malignant adjacent normal tissue samples. So, we created two data matrices: cancer matrix (1356*1065) and normal matrix (1356*111). The patients' clinical information was obtained as well from TCGA and matched with their genomic information.

Sequential biclustering

Biclustering, also known as block clustering, co-clustering, or two-way clustering, is the technique of simultaneously clustering the rows and columns of a matrix. In our study, biclustering shuffles rows (genes) and columns (patients) of the data matrix to generate clusters with a minimum variation of gene expression amongst a group of patients (intra-cluster variation) and maximum variation with other groups of patients (inter-cluster variation). We used the *biclust* package available in R on the Log2-transformed gene expression data of cancer patients to divide patients into subgroups based on their expression of immune-related genes. The BCPlaid algorithm was used as it can cluster patients (columns) based on their similarity in gene expression (row-based) rather than by the similar gene expression per patient (column-based) [17] (Fig 2A).

Running the BCPlaid algorithm generated overlapping clusters of patients and genes, where the clusters are sorted by layers output by the program. Each layer is searched based on residuals given all the previous layers, so by nature, earlier layers contain more information about the data and tend to be more coherent. Since our goal is to divide patients into non-overlapping groups, which means that one patient should fall in only one group, we adopted a sequential approach. In this procedure, we run the BCPlaid algorithm multiple times sequentially. After each run, we take the earliest cluster with at least 5% of the cancer patients and remove the patients in this cluster from the whole dataset. The remaining dataset will be clustered in the next run. The constraint of each candidate cluster having at least 5% of the cancer patients guarantees that finally, each immune evasion cluster is a representative set of patients from the population.

Verification of robustness of the biclustering procedure

The robustness (reproducibility) of the proposed sequential biclustering procedure was assessed by running the whole procedure multiple times with five different random seeds. We reran the procedure with each of these seeds and checked the agreement, measured by the random index, between the clusters obtained by these runs and the original ones [18]. Specifically, each seed leads to a set of patient group labels $C(i), i = 1,2,...,n$, where n is the total number of patients. Denote the original group labels as $C_0(i)$ and the group labels from the other five runs as $C_k(i), k = 1,2,3,4,5$. For each pair of patients $1 \leq i < j \leq n$, define

$$\delta_k^{ij} = \begin{cases} 1, if C_0(i) = C_0(j), C_k(i) = C_k(j) \lor C_0(i) \neq C_0(j), C_k(i) \neq C_k(j) \\ 0, otherwise \end{cases}$$

$and\ \gamma_k = \dfrac{\sum_{1 \leq i < j \leq n} \delta_k^{ij}}{\binom{n}{2}}, k = 1,2,3,4,5.$ This is an intuitive measure of consistency between two
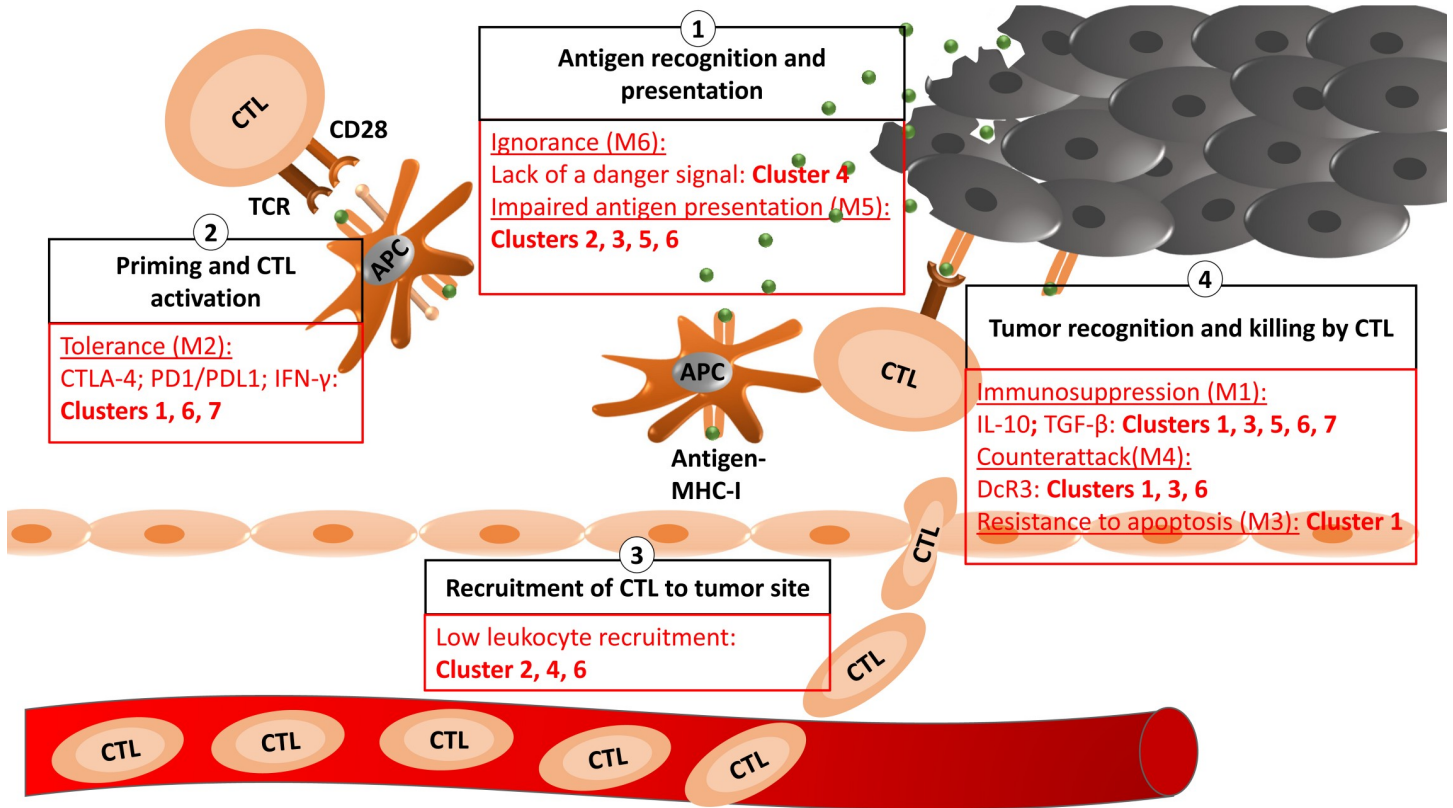
Fig 1. Evasion at different levels of the cancer-immunity cycle in each cluster.

different sequences of biclustering runs by investigating if any two patients are clustered consistently. Values of agreement with the five random number seeds were all in the range of 90–93% (S3 File). We also compared the results of each seed with the classification tree mentioned below, and error rates were relatively stable too (S3 File).

Fisher exact test

In our study, we wanted to determine whether the clustering was dependent on breast cancer receptor status and invasive ductal and lobular subtypes. Accordingly, the receptor status information was obtained from TCGA and a Fisher exact test was used (Tables 1 and 2). The p-values were calculated by comparing the number of patients in a cluster belonging to a specific subtype to the total number of patients in the cluster. A p-value ≤ 0.05 indicates that distribution of a number of patients in that cluster is significantly different from the overall pattern.
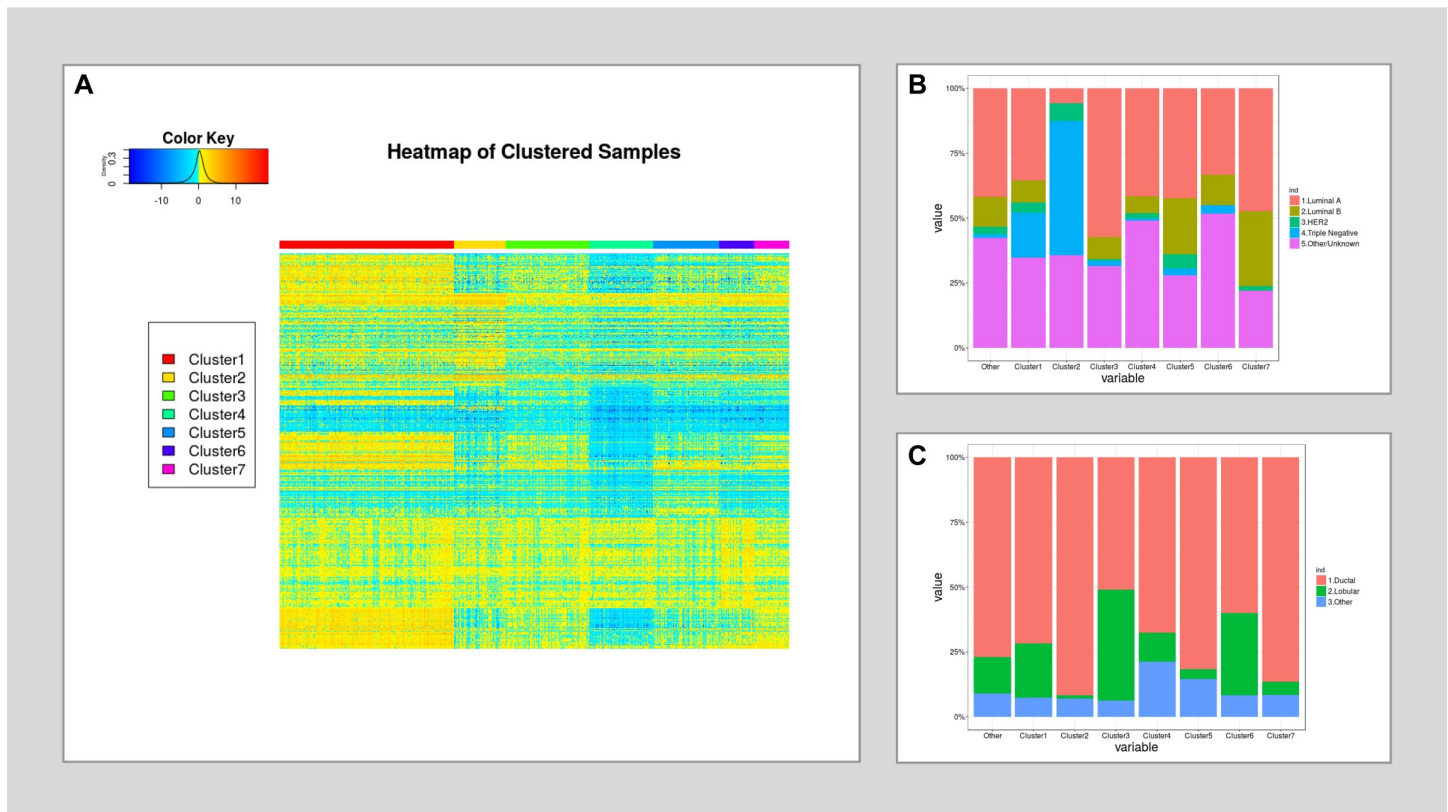
Fig 2. The clustering results for the sequential biclustering. (A) The heatmap representing the level of gene expression (rows) in different clusters of patients (columns). (B) The percentage of different breast cancer receptor subtype and (C) lobular and ductal carcinomas, ILC and IDC, per cluster.

https://doi.org/10.1371/journal.pone.0207799.g002

Differential gene expression analysis

Comparison of gene expression between any two groups is done by a combination of p-value from t-test and log2 fold change cutoff. Both comparisons for tumor versus tumor and tumor versus normal were performed. T-tests were implemented to compare the mean expression of a gene within a cluster to the mean expression of other cancer patients (tumor versus tumor) and to that of normal samples (tumor versus normal) (S4 File). Comparing tumor to tumor can be misleading as a gene may be up-regulated for example in one tumor cluster compared to others but is still less than normal. Thus, it is important to use the normal as a reference. The number of genes per cluster may not always be enough to help determine the mechanism of evasion. Knowing that each cluster of patients will have different levels of gene expression, we unified all the genes (rows) from the seven clusters and checked for their mean expression per cluster of patients.

Table 1. The association of clusters with breast cancer subtypes (receptor status).

Cluster	Number of patients	HER2	Luminal A	Luminal B	TNBC	Fisher exact test p-value
Cluster 1	193	12 (6.2%)	105 (54.4%)	25 (13.0%)	51 (26.4%)	8.46E-03
Cluster 2	56	6 (10.7%)	5 (8.9%)	0	45 (80.4%)	4.95E-26
Cluster 3	98	1 (1%)	82 (83.7%)	12 (12.2%)	3 (3%)	3.10E-05
Cluster 4	55	2 (3.6%)	45 (81.8%)	7 (12.7%)	1 (1.8%)	4.49E-03
Cluster 5	80	6 (7.5%)	47 (58.75%)	24 (30%)	3 (3.75%)	9.30E-04
Cluster 6	29	0	20 (68.9%)	7 (24%)	2 (6.9%)	0.315
Cluster 7	46	1 (2%)	28 (60.9%)	17 (37%)	0	1.94E-04
Other	201	6 (3.0%)	84 (41.8%)	23 (11.4%)	3 (1.5%)	2.16E-04

The results of the Fisher exact test show a significant association between cluster 3, cluster 4, cluster 5, cluster 7 and Luminal A, and CL2 and triple negative breast cancer (TNBC) (P<0.05). Some patients had no information on their receptor status in TCGA, hence the lower number of patients in column 2 compared to the identified. HER2: Human epidermal growth factor receptor 2.

https://doi.org/10.1371/journal.pone.0207799.t001

Pathway analysis

To find out whether an immune pathway is altered in each of the 7 clusters, we compared the expression of genes in a cluster to those outside that cluster and normal samples (figures names are Tum_Tum and Tum_Nor, respectively). The significantly differentially expressed genes and the corresponding KEGG pathways were plotted using the R/Bioconductor package, *pathview* for visualization [19]. We chose the immune-related pathways and showed only the pathways for antigen processing and presentation (hsa04612), leukocyte transendothelial migration (hsa04670), and cell adhesion molecules (hsa04514) that also shows the fold change of PD-1 (PDCD1), PD-L2 (PDCD1LG2), and CTLA4 (S5 File and Fig 3). KEGG however lacks immune evasion pathways. Thus, we categorized the genes into different groups based on their role in the cancer-immunity cycle (S6 File). Using the p-values generated by the t-test for the tumor-tumor and tumor-normal mean comparisons, we were able to understand at which level evasion was happening and using which molecules.

Classification tree

After obtaining the 7 clusters, we wanted to identify a small set of biomarker genes which could classify a subset of patients into their corresponding clusters. A classification tree was used to build a model to predict the cluster into which a patient sample belongs. This was achieved using the *rpart* package in R [20,21]. The selected biomarker genes with their cutoff values are displayed in Fig 4.

Table 2. The association of clusters with invasive lobular and ductal carcinoma subgroups, ILC and IDC, respectively.

Cluster	Number of patients	IDC	ILC	Fisher exact test p-value
Cluster 1	274	212 (77.37%)	62 (22.63%)	3.5030E-01
Cluster 2	79	78 (98.73%)	1 (1.27%)	1.4315E-06
Cluster 3	134	73 (54.48%)	61 (45.52%)	6.2561E-10
Cluster 4	85	73 (85.88%)	12 (14.12%)	2.5155E-01
Cluster 5	93	89 (95.7%)	4 (4.30%)	6.2294E-05
Cluster 6	55	36 (65.45%)	19 (34.55%)	1.5219E-02
Cluster 7	54	51 (94.44%)	3 (5.56%)	6.8520E-03
Other	182	154 (84.62%)	28 (15.38%)	1.8154E-01

Clusters 2, 5 and 7 were significantly associated with IDC and clusters 3 and 6 with ILC.

https://doi.org/10.1371/journal.pone.0207799.t002
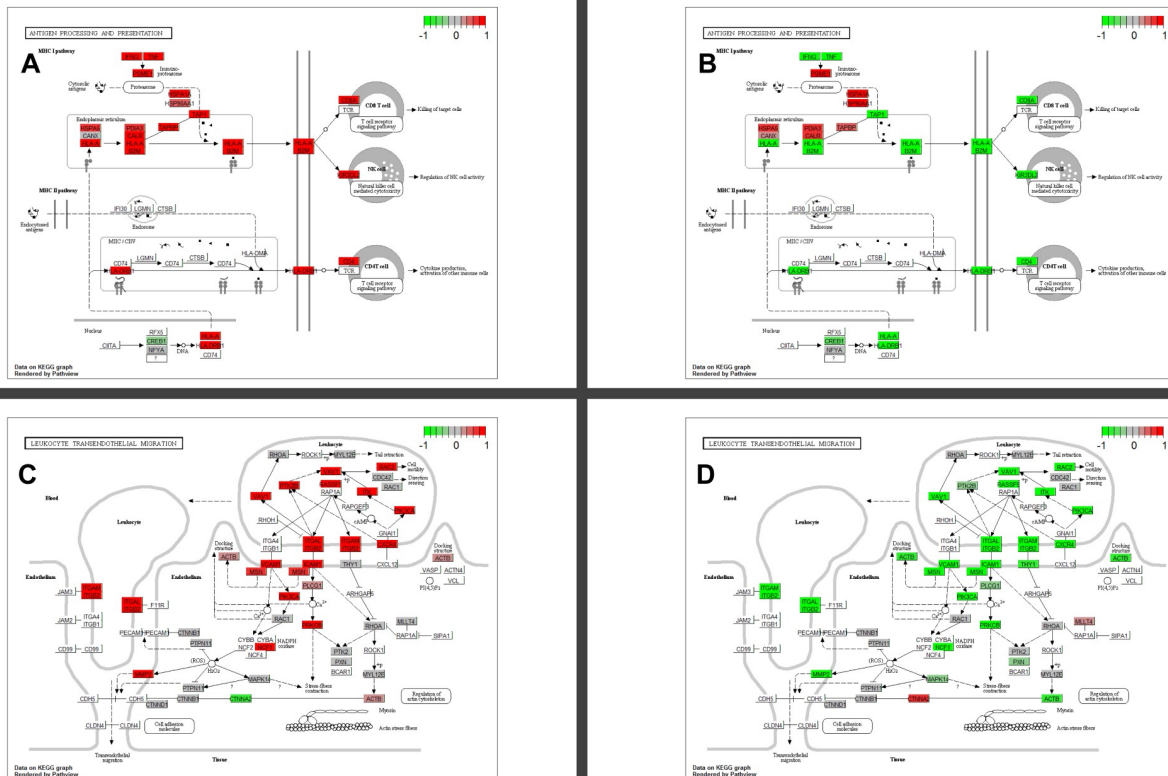
Fig 3. Pathway analysis based on the log2 fold change for clusters 1 and 4 compared to normal. Fold change level of molecules involved in antigen processing and presentation molecules in cluster 1 (A) and cluster 4 (B). Fold change level of molecules involved in leukocyte recruitment in cluster 1 (C) and cluster 4 (D). These results show how cluster 1 genes for the first 2 steps of the cancer-immunity cycle are up-regulated while those of cluster 4 are mostly downregulated. These pathways and others for other clusters and steps of the cancer-immunity pathway are in S5 File. The color scale ranges from the downregulated expression in green (-1 fold), to the non-differential expression in grey, to the up-regulated expression in red (+1 fold).

https://doi.org/10.1371/journal.pone.0207799.g003

Results

Cohort and patient clustering

To investigate the different evasion mechanisms in breast cancer, we first compiled a list of 1,356 genes involved in immune activation and immune evasion (S2 File) as described in the Methods section. Then the RNA-seq expression data of these genes in breast cancer patients were obtained from TCGA database and used to categorize patients into different groups using a sequential biclustering algorithm based on BCPlaid [17]. Eighty-one percent of TCGA breast cancer patients were clustered into 7 groups with non-overlapping patients, whereas the other nineteen percent fell into much smaller groups whose specific expression patterns were
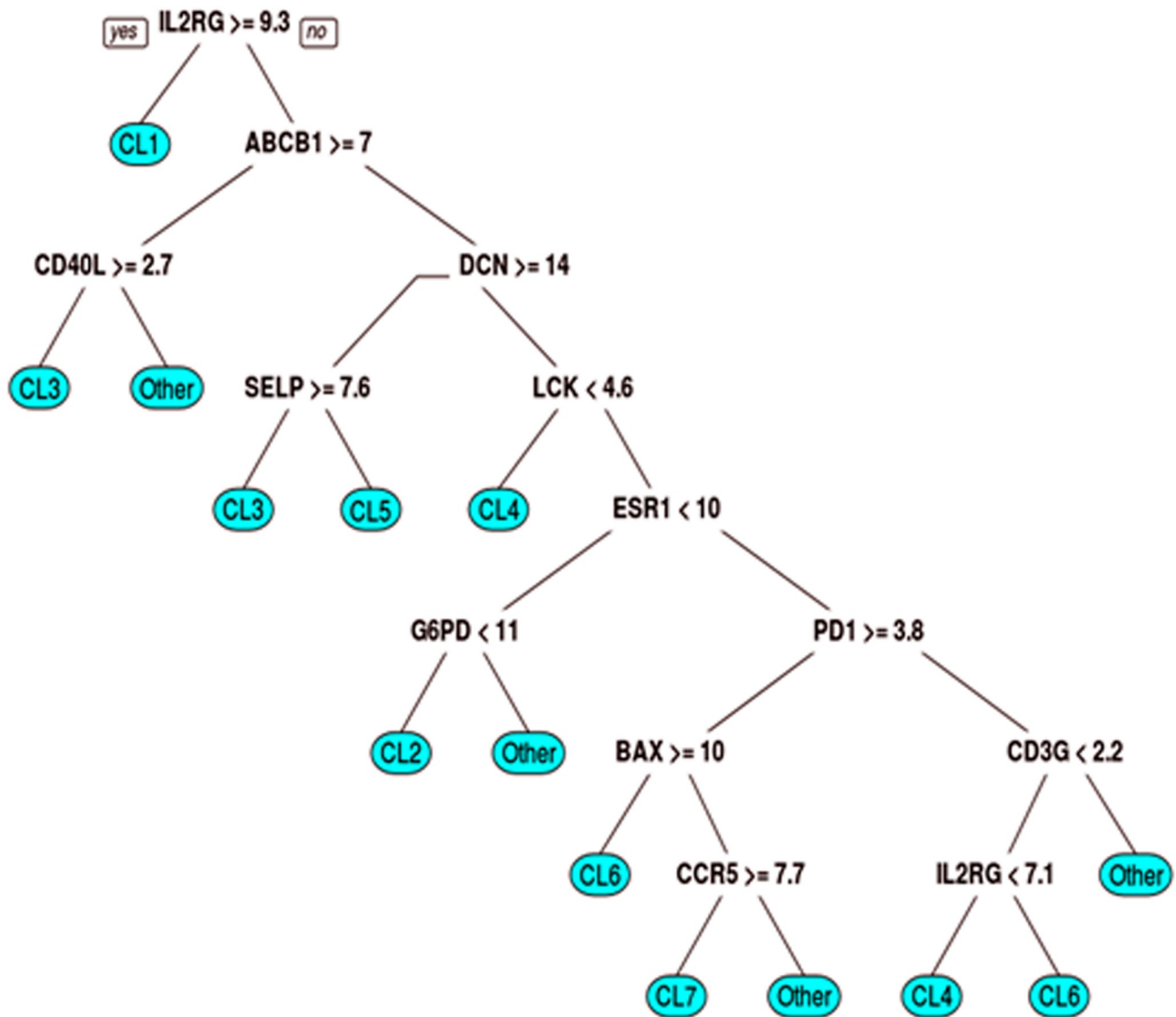
Fig 4. Classification tree with 12 biomarkers and their log2 gene expression cutoffs for the identified clusters (CL). Interleukin-2 receptor subunit gamma (IL2RG), ATP-binding cassette sub-family B member 1 (ABCB1), cluster of differentiation-40 ligand (CD40LG), decorin (DCN), lymphocyte-specific protein tyrosine kinase (LCK), selectin-P (SELP), estrogen receptor-1 (ESR1), glucose-6-phosphate dehydrogenase (G6PD), programmed cell death-1 (PDCD1), cluster of differentiation-3 subunit gamma.

not characterized in this study (S2 File). TCGA's 111 nonmalignant adjacent breast cancer samples were used as a normal reference for gene expression.

Identifying evasion mechanisms in the seven clusters

To understand the mechanisms of evasion of patients falling into a specific group, the mean expression and fold change of genes involved in different steps of cancer-immunity cycle were

Table 3. The rationale for deciding the immune evasion mechanisms (M).

Mechanism	Genes	Expression compared to normal
Immunosuppression (M1)	IL-10 or TGF-β1 or TGF-β2	Up-regulated
Tolerance (M2)	CTLA4 or (PD-1 and PD-L1/2) or IFN-γ	Up-regulated
Anti-apoptosis (M3)	At least 2 out of {BIRC3, TNFAIP3, TRAF1, TRAILR4}	Up-regulated
Counterattack (M4)	DcR3	Up-regulated
Impaired antigen presentation (M5)	B2M or HLA-A or HLA-B	Not up-regulated
	and CD4 or CD8A	Not up-regulated
	and at least 1 out of {GZMA, GZMB, PRF1}	Up-regulated
	or TGF-β1	Up-regulated
Ignorance (M6)	B2M and HLA-A and HLA-B	Not up-regulated
	and CD4 and CD8A	Not up-regulated
	and GZMA and GZMB and PRF1	Not up-regulated
	and TGF-β1	Not up-regulated

Interleukin 10 (IL-10), transforming growth factor-beta (TGF-β1/2), cytotoxic T-lymphocyte associated protein 4 (CTLA4); programmed cell death-1 (PD-1) and ligand (PD-L1/2), interferon gamma (IFN-γ), Baculoviral IAP repeat-containing protein 3 (BIRC3), Tumor necrosis factor alpha-induced protein 3 (TNFAIP3), TNF receptor-associated factor 1 (TRAF1), TNF-related apoptosis-inducing ligand receptor 4 (TRAILR4), decoy receptor 3 (DcR3); beta 2 microglobulin (B2M), human leukocyte antigen A and B (HLA-A/B), cluster of differentiation 4 (CD4), cluster of differentiation 8A (CD8A), granzyme A and B (GZMA/B), and perforin 1 (PRF1).

compared to those of normal samples (S4 and S6 Files). The evasion mechanisms (M) were identified based on the rationale summarized in Table 3 and discussed in S7 File.

Evident immunosuppression with TGF-β1 and TGF-β1 + DcR3

Although several clusters had different combinations of evasion mechanisms, the majority of breast cancer patient groups shared and upregulated expression TGF-β1. Around 77.4% of the clustered TCGA breast cancer patients' evasion was through TGF-β1-mediated immunosuppression, 57.7% with DcR3, 48.0% with CTLA4, and 34.3% with PD-1. All clusters with upregulated DcR3 (57.7%) had upregulated TGF-β1 expression as well and all 34.3% with upregulated PD-1 had upregulated CTLA4 expression (Table 4). Thus, the most prevalent evasion mechanisms are through TGF-β1 and TGF-β1 and DcR3 combined.

DcR3 and TGF-β1 were shown to work in concert to induce epithelial to mesenchymal transitioning in colorectal cancer [22]. Thus, it is possible that these molecules work together in breast cancer as well to aid in tumor progression and immune evasion and thus, serve as potential immunotherapy targets.

ILC and IDC show distinctive evasion combinations

Forty-two percent of TCGA's ILC were significantly associated with clusters 3 and 6 while only 28.5% of IDC patients were significantly associated with clusters 2, 5 and 7 (Table 2). Although ILC is less common than IDC, studies have suggested that overall long-term outcomes of patients with ILC may be worse than those with stage-matched IDC and ILC patients are less responsive to neoadjuvant chemotherapy [10,11]. An obvious distinction between the 2 subgroups was that IDC clusters have fewer combinations of mechanisms compared to ILC clusters and the counterattack via DcR3 (M4) was exclusive to ILC. Since certain chemotherapy treatments have immune-stimulating properties, they can help sensitize the immune response by promoting antigen presentation and tumor sensitization to T-cell mediated killing by Treg depletion [8,23]. Thus, M1, M2, and M5 could be diminished by chemotherapy. This might be the reason why IDC derives greater benefit from chemotherapy compared to ILC which also evades via DcR3 and induces T cell apoptosis. However, this hypothesis requires further validation.

Table 4. The mechanism of evasion in each cluster and the potential immunotherapies for future clinical trials.

Cluster	Mechanism of Evasion	Potential Immunotherapies
Cluster 1 (TNBC)	M1: Immunosuppression: IL-10, TGF-β1	Anti-TGF-β1
	M2: Tolerance: CTLA4, PD-1/PD-L1, IFN-γ	Anti-PD-1; anti-CTLA4; anti-IFN-γ*
	M3: Apoptosis resistance: Anti-apoptotic molecules¥	-
	M4: Counterattack: DcR3	Anti-DcR3
Cluster 2 (TNBC, IDC)	M5: Impaired antigen presentation	DC vaccine + Chemotherapy
Cluster 3 (Luminal A, ILC)	M1: Immunosuppression: TGF-β1	Anti-TGF-β1
	M4: Counterattack: DcR3	Anti-DcR3
	M5: Impaired antigen presentation	DC vaccine
Cluster 4 (Luminal A)	M6: Ignorance: No danger signals	DC vaccine + Chemotherapy
Cluster 5 (Luminal B, IDC)	M1: Immunosuppression: TGF-β1	Anti-TGF-β1
	M5: Impaired antigen presentation	DC vaccine
Cluster 6 (ILC)	M1: Immunosuppression: TGF-β1	Anti-TGF-β1
	M2: Tolerance: CTLA4	Anti-CTLA4
	M4: Counterattack: DcR3	Anti-DcR3
	M5: Impaired antigen presentation	DC vaccine
Cluster 7 (Luminal B, IDC)	M1: Immunosuppression: TGF-β1	Anti-TGF-β1
	M2: Tolerance: CTLA4, IFN-γ	Anti-CTLA4; anti-IFN-γ*

*: require further investigation

¥: These include BIRC3, TRAF1, TNFAIP3 and TRAILR4; IDC: invasive ductal carcinoma; ILC: invasive lobular carcinoma.

https://doi.org/10.1371/journal.pone.0207799.t004

TNBC falls into 2 evasion clusters

TNBC was significantly associated with clusters 1 and 2, Luminal A with clusters 3 and 4, and Luminal B with clusters 5 and 7 (Table 1). TNBC patients were mostly distributed between cluster 1 which was identified with 4 different evasion mechanisms (47%) and cluster 2 where evasion is mainly driven by impaired antigen presentation (42%). Thus, there are 2 immune evasion subgroups for TNBC: one with only impaired antigen presentation and another with a combination of 4 evasion mechanisms.

Although TNBC is associated with poor prognosis, certain patients seemed to respond well to anthracycline-based chemotherapies [8]. Previous studies have shown that objective complete response was significantly associated with immune modules only in luminal breast cancer subgroups and that the tumor-infiltrating lymphocytes in TNBC had no significant interaction with paclitaxel plus non-pegylated liposomal doxorubicin-based therapies in neoadjuvant settings [9,24]. In postneoadjuvant setting, tumor-infiltrating lymphocytes were associated with better prognosis in TNBC [25]. Since chemotherapy may induce immunogenic cell death [23], it is possible that TNBC patients falling into cluster 2 can overcome the impaired antigen presentation after chemotherapy-induced sensitization, resulting in the recruitment of TIL. However, more investigations are required to validate this possibility.

Furthermore, cluster 1 immune gene expression seems to overlap with the immunomodulatory subtype of TNBC, which is characterized by upregulated expression of genes involved in T cell function, interferon response, and antigen presentation [26].

Identification of biomarkers for the immune evasion clusters

The discovery of seven immune evasion groups for breast cancer helps provide guidance for the choice of treatment, if we can identify for any given patient the cluster they belong to and hence, their evasion mechanisms. Using the molecules directly related to the immune evasion mechanisms, such as PD-L1, CTLA4, etc., may not be optimal especially that our results

showed that evasion mostly occurs by a combination of several mechanisms and several molecules. Instead, we generated a decision tree and identified 12 biomarkers along with their gene expression thresholds (Fig 4). Interestingly, the identified immune evasion genes were not all part of this set of putative biomarkers, which indicates that evasion genes may not be powerful in terms of identifying patients' evasion mechanisms. S8 File summarizes the rationale for using classifier genes for each cluster and their corresponding functions.

Discussion

Tumors evade immune surveillance using 6 different mechanisms, which may occur simultaneously in the same tumor (Fig 1). Thus, there are 63 possible combinations of mechanisms in BRCA. This heterogeneity even at the level of a single cancer hallmark explains the challenge of identifying effective treatments.

Using the sequential biclustering method on BRCA RNA-seq data from TCGA, we identified seven clusters of BRCA patients with different evasion mechanisms and combinations of mechanisms. The mechanisms of evasion were determined based on the expression of immune evasion genes and the genes involved in the cancer-immunity cycle. To make it easier to identify patients' evasion group and thus, the choice of immunotherapy, we narrowed down the list of immune-related genes to 12 biomarkers using a classification tree algorithm. Fitting a single tree may not be able to achieve the maximum degree of accuracy, however, the classification tree was used due to its intuitive output with straightforward clinical interpretations. To further improve the classification accuracy, we applied the random forest model to the same data set (S9 File).

Our results indicate that there are different combinations of evasion mechanisms involved in breast cancer. We found that 77.4% of the clustered patients evade with TGF-β1-induced immunosuppression and 57.7% with a DcR3-induced counterattack. All clustered patients with upregulated DcR3 expression had TGF-β1 upregulated as well. Thus, targeting TGF-β1 alone or with DcR3 may be promising for breast cancer treatment. The immunosuppressive nature of TGF-β has been well studied. It was shown that TGF-β disrupts antigen presentation and T cell activation, induces a Treg transformation from naïve T cells, and causes epithelial to mesenchymal transitioning [27–31]. The blockade of TGF-β in colon cancer unleashed a potent and endured cytotoxic T-cell response against cancer cells, inhibited metastases, and rendered metastatic colon cancer more susceptible to anti-PD-1-PD-L1 therapy [32]. DcR3 was shown to inhibit cytotoxicity against tumor cells and its expression was positively associated with cancer progression, angiogenesis, and metastasis [33,34]. Furthermore, DcR3 was suggested as a prognostic factor for early tumor detection and a predictor of recurrence after resection in breast cancer, specifically [35,36].

Examining the invasive breast cancer subgroup, we found that, compared to IDC, ILC-associated clusters had an exclusive upregulation of DcR3. This led us to hypothesize that the reason IDC benefits more from chemotherapy is that the latter helps diminish evasion by aiding in antigen presentation and killing Tregs [8,23]. However, the upregulated expression of DcR3 in the ILC-enriched clusters results in further cytotoxic T cell death. Our results contradict with a previous study on a northeastern Chinese population where they showed increased expression of DcR3 in IDC [36], however, our population is representative of a large racial and ethnic population.

Due to the higher immune system-tumor interactions involved, TNBC and HER2 were thought to be more immunogenic than Luminal A [37]. However, our results showed that TNBC patients split into two groups: the highly immunogenic cluster1 (TNBC: 51/105) and the less immunogenic (low leukocyte infiltration) cluster 2 (TNBC: 45/105) (Table 1). Based

on our results, we hypothesized that cluster 1 TNBCs correspond to the immunomodulatory TNBC subtype and that cluster 2 TNBC's treatment with chemotherapy may trigger an immune response [23,26,38]. Furthermore, the comparison between the different clusters showed that only cluster 1 was significantly associated with a higher proliferation gene expression signature (S10 File). However, given the strong immune response in cluster 1, this may be caused by the highly proliferating immune cells recruited to the tumor microenvironment rather than cancer cells' proliferation.

Currently, there are no clinical trials on DcR3 however, there are several ongoing in breast cancer for blockading CTLA4, TGF-β, PD-1/PD-L1, CLTA4 and PD-1, and chimeric antigen receptor T-cell therapy (CAR-T). The list is summarized in S1 File. The response rate of blockading PD-1/PD-L1 in breast cancer patients has been tested in several hormone receptor subtypes. The overall response rate to PD-1 blockade in PD-L1 positive patients with advanced TNBC BRCA was 18.5% [39] and 12% in ER+ and HER 2- advanced BRCA [40]. The overall response rate to anti-PD-L1 was 24% in TNBC [41], and only 3% in metastatic BRCA—62.5% of whom were PD-L1+ [42]. In this latter study patients with TNBC experienced an overall response rate of 5.2% [42]. No drugs targeting CTLA4 alone are currently in clinical trial. Combining anti-PD-1 anti-CTLA4 in eighteen patients with refractory metastatic BRCA resulted in an overall response rate of 17%: 0% for the 11 ER+ patients and 43% for the 7 TNBC patients [43]. However, the small TNBC sample size makes the results inconclusive. Overall, these results show a poor overall response rate when the treatment criteria are based on breast cancer receptor subtypes and the expression of PD-1/PD-L1 and CTLA4, stressing the need for new criteria and more reliable biomarkers. Our results showed that there is no one hormone receptor subtype that falls into one immune evasion cluster, rendering receptor subtyping a weak guide to the choice of immunotherapy. We also showed that the expression of PD-1 and CTLA4 is not a good biomarker for immunotherapy and hence the need for more reliable biomarkers.

Despite the large breast cancer sample size in TCGA, the absence of matched normal samples for about 90% of the breast cancer patients prevented us from taking individual genetic variation into account in this study. Moreover, the TCGA normal samples are not exactly normal but non-malignant adjacent samples.

Supporting information

S1 File. Potential immunotherapies for breast cancer patients. This table summarizes potential treatment methods, drug names, FDA status, and whether or not it is in a clinical trial for breast cancer.
(XLSX)

S2 File. Immune gene lists and the number of patients and genes in the identified clusters.
(XLSX)

S3 File. Consistency measure and cross validated error rates by classification tree model.
(XLSX)

S4 File. T-test results for tumor versus tumor and tumor versus normal comparisons for the 7 clusters.
(XLSX)

S5 File. Immune pathways generated by Pathview package.
(ZIP)

S6 File. Categorized gene expression results into the different steps of the cancer-immunity cycle and different immune evasion mechanisms.
(XLSX)

S7 File. A detailed explanation of the evasion mechanisms in each cluster.
(DOCX)

S8 File. Identified biomarkers using the classification tree.
(XLSX)

S9 File. Random forest results.
(DOCX)

S10 File. Enrichment analysis for the proliferation GO term.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Qing-Xiang Amy Sang, Jinfeng Zhang.

Data curation: Yuhang Liu.

Formal analysis: Mayassa J. Bou-Dargham, Yuhang Liu.

Investigation: Mayassa J. Bou-Dargham, Qing-Xiang Amy Sang, Jinfeng Zhang.

Methodology: Mayassa J. Bou-Dargham, Yuhang Liu, Jinfeng Zhang.

Project administration: Qing-Xiang Amy Sang, Jinfeng Zhang.

Software: Yuhang Liu.

Supervision: Qing-Xiang Amy Sang, Jinfeng Zhang.

Writing – original draft: Mayassa J. Bou-Dargham, Yuhang Liu.

Writing – review & editing: Mayassa J. Bou-Dargham, Qing-Xiang Amy Sang, Jinfeng Zhang.

References

1. Ehrlich P. Ueber den jetzigen Stand der Karzinomforschung. Vortrag gehalten vor den Studenten der Amsterdamer Universitaet, Vereinigung fuer wissenschaftliche Arbeit 1 June 1908. Printed in: P. Ehrlich. Beitraege Zur Exp Pathol Und Chemother Verlagsgesellschaft, Leipzig 1909:118–64.

2. Igney FH, Krammer PH. Immune escape of tumors: apoptosis resistance and tumor counterattack. J Leukoc Biol 2002; 71:907–20. PMID: 12050175

3. Wang M, Zhang C, Song Y, Wang Z, Wang Y, Luo F, et al. Mechanism of immune evasion in breast cancer. Onco Targets Ther 2017; 10:1561. https://doi.org/10.2147/OTT.S126424 PMID: 28352189

4. Walsh CM, Glass AA, Chiu V, Clark WR. The role of the Fas lytic pathway in a perforin-less CTL hybridoma. J Immunol 1994; 153:2506–14. PMID: 7521365

5. Li JH, Rosen D, Ronen D, Behrens CK, Krammer PH, Clark WR, et al. The regulation of CD95 ligand expression and function in CTL. J Immunol 1998; 161:3943–9. PMID: 9780162

6. Chen DS, Mellman I. Oncology meets immunology: The cancer-immunity cycle. Immunity 2013; 39:1–10. https://doi.org/10.1016/j.immuni.2013.07.012 PMID: 23890059

7. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell 2011; 144:646–74. https://doi.org/10.1016/j.cell.2011.02.013 PMID: 21376230

8. Stagg J, Allard B. Immunotherapeutic approaches in triple-negative breast cancer: Latest research and clinical prospects. Ther Adv Med Oncol 2013; 5:169–81. https://doi.org/10.1177/1758834012475152 PMID: 23634195

9. Denkert C, Von Minckwitz G, Brase JC, Sinn B V., Gade S, Kronenwett R, et al. Tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy with or without carboplatin in human epidermal growth factor receptor 2-positive and triple-negative primary breast cancers. J Clin Oncol 2015; 33:983–91. https://doi.org/10.1200/JCO.2014.58.1967 PMID: 25534375

10. Korkola JE, DeVries S, Fridlyand J, Hwang ES, Estep ALH, Chen YY, et al. Differentiation of Lobular versus Ductal Breast Carcinomas by Expression Microarray Analysis. Cancer Res 2003; 63:7167–75. https://doi.org/10.1038/sj.bjc.6600556 PMID: 14612510

11. Barroso-Sousa R, Metzger-Filho O. Differences between invasive lobular and invasive ductal carcinoma of the breast: Results and therapeutic implications. Ther Adv Med Oncol 2016; 8:261–6. https://doi.org/10.1177/1758834016644156 PMID: 27482285

12. Michaut M, Chin SF, Majewski I, Severson TM, Bismeijer T, De Koning L, et al. Integration of genomic, transcriptomic and proteomic data identifies two biologically distinct subtypes of invasive lobular breast cancer. Sci Rep 2016; 6. https://doi.org/10.1038/srep18517 PMID: 26729235

13. Vonderheide RH, LoRusso PM, Khalil M, Gartner EM, Khaira D, Soulieres D, et al. Tremelimumab in combination with exemestane in patients with advanced breast cancer and treatment-associated modulation of inducible costimulator expression on patient T cells. Clin Cancer Res 2010; 16:3485–94. https://doi.org/10.1158/1078-0432.CCR-10-0505 PMID: 20479064

14. Mittendorf EA, Hunt KK. Breast Cancer Immunotherapy: Is It Ready for Prime Time? Am J Hematol 2015; 11.

15. Socinski M, Creelan B, Horn L, Reck M, Paz-Ares L, Steins M, et al. NSCLC, metastaticCheckMate 026: A phase 3 trial of nivolumab vs investigator's choice (IC) of platinum-based doublet chemotherapy (PT-DC) as first-line therapy for stage iv/recurrent programmed death ligand 1 (PD-L1)−positive NSCLC. Ann Oncol 2016; 27. https://doi.org/10.1093/annonc/mdw435.39

16. Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers Mike. BioGRID: a general repository for interaction datasets. Nucleic Acids Res 2006; 34:D535–9. https://doi.org/10.1093/nar/gkj109 PMID: 16381927

17. Lazzeroni L, Owen A. Plaid Models for Gene Expression Data. Stat Sin 2002; 12:61–86.

18. Rand WM. Objective Criteria for the Evaluation of Clustering Methods. J Am Stat Assoc 1971; 66:846. https://doi.org/10.2307/2284239

19. Luo W, Brouwer C. Pathview: An R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics 2013; 29:1830–1. https://doi.org/10.1093/bioinformatics/btt285 PMID: 23740750

20. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. vol. 19. 1984.

21. Therneau TM, Atkinson EJ. An introduction to recursive partitioning using the RPART routines. Mayo Found 1997.

22. Liu Y, Zhu H-F, Liu D, Hu Z-Y. DcR3 induces epithelial-mesenchymal transition through activation of the TGF- β 3 / SMAD signaling pathway in CRC. Oncotarget 2016; 7:77306–18. https://doi.org/10.18632/oncotarget.12639 PMID: 27764793

23. van der Most RG, Currie AJ, Cleaver AL, Salmons J, Nowak AK, Mehendran S, et al. Cyclophosphamide chemotherapy sensitizes tumor cells to TRAIL-dependent CD8 T cell-mediated immune attack resulting in suppression of tumor growth. PLoS One 2009; 4. https://doi.org/10.1371/journal.pone.0006982 PMID: 19746156

24. Gianni L, Zambetti M, Clark K, Baker J, Cronin M, Wu J, et al. Gene expression profiles in paraffin-embedded core biopsy tissue predict response to chemotherapy in women with locally advanced breast cancer. J Clin Oncol 2005; 23:7265–77. https://doi.org/10.1200/JCO.2005.02.0818 PMID: 16145055

25. Dieci M V., Criscitiello C, Goubar A, Viale G, Conte P, Guarneri V, et al. Prognostic value of tumor-infiltrating lymphocytes on residual disease after primary chemotherapy for triple-negative breast cancer: A retrospective multicenter study. Ann Oncol 2014; 25:611–8. https://doi.org/10.1093/annonc/mdt556 PMID: 24401929

26. Lehmann BDB, Bauer J a J, Chen X, Sanders ME, Chakravarthy a B, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 2011; 121:2750–67. https://doi.org/10.1172/JCI45014 PMID: 21633166

27. Thomas D a, Massagué J. TGF-beta directly targets cytotoxic T cell functions during tumor evasion of immune surveillance. Cancer Cell 2005; 8:369–80. https://doi.org/10.1016/j.ccr.2005.10.012 PMID: 16286245

28. Di Bari MG, Lutsiak MEC, Takai S, Mostbock S, Farsaci B, Tolouei Semnani R, et al. TGF-β modulates the functionality of tumor-infiltrating CD8 + T cells through effects on TCR signaling and Spred1 expression. Cancer Immunol Immunother 2009; 58:1809–18. https://doi.org/10.1007/s00262-009-0692-9 PMID: 19319531

29. Li Y, Han B, Li K, Jiao LR, Habib N, Wang H. TGF-β 1 inhibits HLA-DR and β 2-microglobulin expression in HeLa cells induced with r-IFN. Transplant. Proc., vol. 31, Elsevier; 1999, p. 2143–5. PMID: 10455994

30. Chen W, Jin W, Hardegen N, Lei K-J, Li L, Marinos N, et al. Conversion of peripheral CD4+CD25- naive T cells to CD4+CD25+ regulatory T cells by TGF-beta induction of transcription factor Foxp3. J Exp Med 2003; 198:1875–86. https://doi.org/10.1084/jem.20030152 PMID: 14676299

31. MIYAZONO K. Transforming growth factor-β signaling in epithelial-mesenchymal transition and progression of cancer. Proc Japan Acad Ser B 2009; 85:314–23. https://doi.org/10.2183/pjab.85.314 PMID: 19838011

32. Tauriello DVF, Palomo-Ponce S, Stork D, Berenguer-Llergo A, Badia-Ramentol J, Iglesias M, et al. TGFβ drives immune evasion in genetically reconstituted colon cancer metastasis. Nature 2018; 554:538–43. https://doi.org/10.1038/nature25492 PMID: 29443964

33. Wu Q, Zheng Y, Chen D, Li X, Lu C, Zhang Z. Aberrant expression of decoy receptor 3 in human breast cancer: Relevance to lymphangiogenesis. J Surg Res 2014; 188:459–65. https://doi.org/10.1016/j.jss.2014.01.058 PMID: 24612949

34. Ge Z, Sanders AJ, Ye L, Jiang WG. Aberrant expression and function of death receptor-3 and death decoy receptor-3 in human cancer (Review). Exp Ther Med 2011; 2:167–72. https://doi.org/10.3892/etm.2011.206 PMID: 22977485

35. Kanbayashi C, Koyama Y, Ichikawa H, Sakata E, Hasegawa M. Amplification of Genomic DNA for Decoy Receptor 3 Predicts Post-Resection Disease Recurrence in Breast Cancer Patients. World J Oncol 2014; 5:14–23. http://dx.doi.org/10.14740/wjon764w. PMID: 29147372

36. Fu Z, Chen S, Liu S, Han S, Gao X, Li D, et al. DcR3 gene polymorphisms are associated with sporadic breast infiltrating ductal carcinoma in Northeast Chinese women. Oncotarget 2016; 7. https://doi.org/10.18632/oncotarget.11153 PMID: 27517320

37. García-Teijido P, Cabal ML, Fernández IP, Pérez YF. Tumor-infiltrating lymphocytes in triple negative breast cancer: The future of immune targeting. Clin Med Insights Oncol 2016; 10:31–9. https://doi.org/10.4137/CMO.S34540 PMID: 27081325

38. Zitvogel L, Apetoh L, Ghiringhelli F, André F, Tesniere A, Kroemer G. The anticancer immune response: Indispensable for therapeutic success? J Clin Invest 2008; 118:1991–2001. https://doi.org/10.1172/JCI35180 PMID: 18523649

39. Nanda R, Pathiraja K, Aktan G, Cheng JD, Karantza V, Co M&, et al. JOURNAL OF CLINICAL ONCOLOGY O R I G I N A L R E P O R T Pembrolizumab in Patients With Advanced Triple-Negative Breast Cancer: Phase Ib KEYNOTE-012 Study. J Clin Oncol 2016; 34:2460–7. https://doi.org/10.1200/JCO.2015.64.8931 PMID: 27138582

40. Rugo HS, Delord J-P, Im S-A, Ott PA, Piha-Paul SA, Bedard PL, et al. Safety and Antitumor Activity of Pembrolizumab in Patients with Estrogen Receptor-Positive/Human Epidermal Growth Factor Receptor 2-Negative Advanced Breast Cancer. Clin Cancer Res 2018; 24:2804–11. https://doi.org/10.1158/1078-0432.CCR-17-3452 PMID: 29559561

41. Emens LA, Braiteh FS, Cassier P, Delord J-P, Eder JP, Fasso M, et al. Abstract 2859: Inhibition of PD-L1 by MPDL3280A leads to clinical activity in patients with metastatic triple-negative breast cancer (TNBC). Cancer Res 2015; 75:2859–2859. https://doi.org/10.1158/1538-7445.AM2015-2859

42. Dirix LY, Takacs I, Jerusalem G, Nikolinakos P, Arkenau HT, Forero-Torres A, et al. Avelumab, an anti-PD-L1 antibody, in patients with locally advanced or metastatic breast cancer: a phase 1b JAVELIN

Solid Tumor study. Breast Cancer Res Treat 2017; 167:1–16. https://doi.org/10.1007/s10549-017-4490-3

43.   Santa-maria CA, Kato T, Park J, Kiyotani K, Rademaker A, Shah AN, et al. A pilot study of durvalumab and tremelimumab and immunogenomic dynamics in metastatic breast cancer 2018; 9:18985–96. https://doi.org/10.18632/oncotarget.24867 PMID: 29721177