Original Article

Applying gene set analysis to characterize the activities of immune cells in estrogen receptor positive breast cancer

Yi-Hsuan Chang[1*], Yu-Chiao Chiu[1*], Yu-Ching Hsu[2], Hui-Mei Tsai[2], Eric Y. Chuang[1,3], Tzu-Hung Hsiao[2]

[1]Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei, Taiwan; [2]Department of Medical Research, Taichung Veterans General Hospital, Taichung, Taiwan; [3]Bioinformatics and Biostatistics Core, Center of Genomic Medicine, National Taiwan University, Taipei, Taiwan

Contributions: (I) Conception and design: YH Chang, YC Chiu; (II) Administrative support: EY Chuang, TH Hsiao, HM Tsai, YC Hsu; (III) Provision of study materials or patients: YH Chang; (IV) Collection and assembly of data: YH Chang; (V) Data analysis and interpretation: YH Chang, YC Chiu, TH Hsiao; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors;

*These authors contributed equally to the work.

Correspondence to: Eric Y. Chuang. Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei, Taiwan. Email: chuangey@ntu.edu.tw; Tzu-Hung Hsiao. Department of Medical Research, Taichung Veterans General Hospital, Taichung, Taiwan. Email: d93921032@gmail.com.

Background: Estrogen receptor (ER) is a crucial biomarker for subtyping breast cancer. The present study aimed to understand the influence of infiltrated immune cells to patients' outcome in estrogen receptor positive (ER+) breast cancer.

Methods: Gene expression profiles of three breast cancer cohorts downloaded from Gene Expression Omnibus (GEO) were used in this study. We utilized gene set enrichment analysis (GSEA) to estimate the activities of immune cell infiltration based on 31 published immune gene sets. Each gene set was tested for ER+ associated prognostic value. GSEA was applied to identify biological functions associated with prognostic immune gene sets.

Results: Nine subtypes of immune cells showed ER+ specific association with patient survival; seven of them formed two co-activation clusters, including: (I) activated CD4, CD8, effector memory CD4, and (II) regulatory T cell, dendritic cell, eosinophil, and mast cell, substantially representing innate and adaptive immunity. Among them, activated CD8 and mast cell were independent prognostic factors in multivariate Cox regression. Functional annotation analysis revealed their involvement in breast cancer subtyping, relapse, and metastasis.

Conclusions: We devised a gene set analysis to comprehensively investigate the involvement of ER specific immune cell activities and prognosis in breast cancer. Our work provides hints of the interaction between infiltrated immune cells and activated oncogene in ER+ breast cancer and may contribute to the biological basis for the development of immunotherapy.

Keywords: Gene set; tumor infiltrating lymphocyte (TIL); breast cancer; estrogen receptor (ER); gene expression; survival analysis

Submitted Apr 01, 2016. Accepted for publication Apr 07, 2016.
doi: 10.21037/tcr.2016.04.09
View this article at: http://dx.doi.org/10.21037/tcr.2016.04.09

Introduction

Breast cancer is one of the prevalent cancer types in female in the United States and world-wide (1). Three major molecular subtypes of breast cancer have been identified by the genetic profiles in tumors: estrogen receptor positive (ER+), human epidermal growth factor receptor 2 positive (HER2+), and triple negative breast cancer (2,3). About 70% of human breast cancers are ER+, and most ER+

breast cancer patients benefited from hormone therapies, such as tamoxifen (4-6). However, still 30% of ER+ tumors do not respond well to hormone therapy (7).

Previous studies showed the tumor infiltrating lymphocytes (TILs) were associated with prognosis and treatment response of breast cancer (8-12). For example, large amount of TILs were prone to decreased recurrence in triple negative breast cancer (8). High levels of TILs also increased trastuzumab benefit in HER2+ disease (8). However, most of these studies focused on the HER2+ and triple negative subtypes. The role of TILs in ER+ breast cancer was rarely discussed.

Recently, FDA approved ipilimumab (anti-CTLA4 antibody), nivolumab, and pembrolizumab (anti PD-1 antibody) for cancer immunotherapies to inhibit the immune checkpoint blockade, PD-1 and CTLA-4 (13-15). One of the determining factors of responses to immunotherapy is the presentation of neoantigens and the activation of immune system. Although the roles of different immune cells in cancer progression have been individually studied and reviewed (16,17), a systematic study which comprehensively investigates all various types of immune cells in specific subtype of breast cancer remains an unchartered territory.

In this study, we utilized a gene set enrichment approach to analyze subtypes of TILs. Using genome-wide expression profiles derived from breast tumors, the landscape of immunology in breast cancer and its association with tumor prognosis were explored. The findings could provide a biological basis of tumor infiltrating immune cells in ER+ tumors and benefit the development of immunotherapy.

Methods

Immune gene sets and microarray data sets

A total of thirty-one immune cell associated gene sets, collected from a previous research (18), were used in the study. We analyzed three data sets of breast cancer, GSE4922 (19), GSE2034 (20) and GSE2990 (21). The gene expression profiles were downloaded from Gene Expression Omnibus (GEO) (22). A total of 714 samples were used in the study, and the summary of the three data set was list in *Table S1*. Patients without survival information or ER status were excluded, and all samples were divided into subgroups based on ER status (ER+ and ER–). We took GSE4922 as the training data set and GSE2034 and GSE2990 as the validation data sets.

Gene set enrichment score

In gene set analysis, we used gene set enrichment scores to represent gene activities of immunological gene sets. The gene set enrichment score was defined as the averaged normalized expression value of the member genes in a given data set. Suppose a gene set s consist N genes and the log2-transformed expression level of gene j in sample i be $x_{j,i}$. The enrichment score is defined as below:

$$s_i = \frac{1}{N}\sum_{j=1}^{N} z_{j,i} \quad [1]$$

$$z_{j,i} = \left(x_{j,i} - \mu_j\right)/\sigma_j \quad [2]$$

where μ_j is the mean value of gene j in a gene set among samples, and σ_j is the standard deviation.

Statistical analysis

Patient survival information in the data sets was retrieved from the GEO database. Both Kaplan-Meier estimator method (23) and Cox hazard proportional model (24) were applied. In Kaplan-Meier analysis, patients were divided into two groups based on the median of gene set enrichment score. For multivariate analysis, the analyzed risk factors in the Cox hazard proportional model were the status of lymph node (LN) involvement, enrichment score of cell cycle indexing genes (25), consensus ER+ prognostic genes (26), and the immune cell associated gene sets. Association between each pair of immune gene sets was calculated by Pearson correlation.

Annotation of biological functions by gene set enrichment analysis (GSEA)

GSEA (27) was applied to identify the gene sets enriched in the sub-clusters of 211 breast cancer samples. The gene sets are collected from the Molecular Signatures Database (MSigDB) v5.0 (28). A total of 3,951 gene sets including the gene set of C2-curated chemical or genetic perturbations (CGP), C3-transcription factor targets (TFT), C5-gene ontology biological process (GO-BP), and C6-oncogenic signatures were analyzed in this study.

Results

Study overview

To explore the association of TILs and patient's survival in

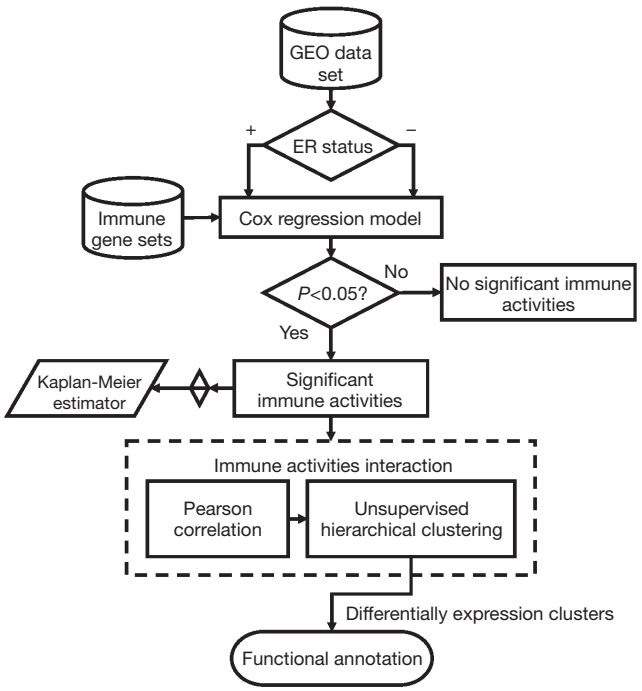

Figure 1 Flowchart for identifying prognostic immune biomarkers associated with survival outcomes in breast cancer.

ER+ breast cancer, 31 previously defined gene sets were used to present the subpopulation of TILs. The immune cell gene sets were applied for survival analysis. The overall analysis flowchart is showed in *Figure 1*. The samples were first divided into two groups according to the ER status, and the enrichment score of each gene set was computed for each sample. Then, for each group, the enrichment scores were applied for Cox regression analysis. Statistically significant gene sets were identified based on the threshold of Cox P value <0.05. Moreover, to validate the results from survival analysis, we included two other datasets, GSE2990 and GSE2034. The results were consistent to what we found in the discovery dataset. The survival analysis results for the three datasets were visualized by Kaplan-Meier plots.

Further investigation of the immunological interactions between each survival associated immune cell gene sets was performed by Pearson correlation on the enrichment scores. Gene sets that showed strong interactions were grouped. Based on the enrichment scores of these gene sets we performed the unsupervised hierarchical clustering with respect to samples. Two clusters were chosen for subsequent analysis. One of them was patients with better survival and the other was the opposite. Finally, GSEA software was executed to provide functional interpretations of these two clusters.

Table 1 Univariate Cox regression of the 9 immune cell gene sets in GSE4922

Gene sets	Overall survival in ER+		Overall survival in ER–	
	Hazard ratio	P value	Hazard ratio	P value
Activated CD8	2.12	<0.001	1.92	0.220
Activated CD4	1.73	<0.001	1.52	0.271
Mast cells	0.33	0.003	0.80	0.836
Effector memory CD4	2.20	0.004	2.24	0.156
Th2	0.38	0.022	0.68	0.697
Treg	0.52	0.026	0.73	0.607
Eosinophil	0.49	0.030	0.92	0.921
NK56 dim	2.32	0.032	13.01	0.041
DC	0.55	0.044	0.91	0.875

ER, estrogen receptor; Th, T-helper; NK, natural killer; DC, dendritic cell.

Predictive immune signatures for breast cancer prognosis

To identify the activities of immune cell subtypes that may affect the clinical outcome of breast cancer patients, enrichment score of each gene set was calculated for each sample and was then applied to univariate Cox regression model. The yielded hazard ratios and P values are listed in *Table 1*. Nine of the 31 immune cell gene sets were found with statistical significance (P<0.05) in the ER+ group while only one of them in the ER– group. For the ER+ group, the significant gene sets were assigned to two groups, the protective and the risk group, according to the hazard ratio. The risk group included four gene sets: activated CD8, activated CD4, effector memory CD4, and NK56 dim, and the protective group contained five gene sets: mast cells, Th2, Treg, eosinophil, and DC. Among the nine gene sets, the activated CD8, activated CD4, effector memory CD4, Th2 and Treg gene set were related to adaptive immune system, and the NK56 dim, mast cells, eosinophil, and DC gene sets are associated with innate immune system. This implied that certain immunological signaling pathways and the cooperation of the innate and adaptive immune systems may affect the prognosis of the breast cancer patient. Previous studies also showed some relation between immune system and cancer microenvironment (29,30). Although the mechanisms underlying the result were still unclear, this may be an interesting topic for future research.

Validation for the identified predictive TIL subtypes

We validated the findings in the discovery dataset in two validation data sets (*Table S2*). Eight and five out of the nine gene sets showed concordant significance in the validation data sets, GSE2034 and GSE2990. This demonstrates the potential importance of the immune activities to ER+ breast cancer patients' survival. Kaplan-Meier plot of the three most significant gene sets are activated CD4 (Cox P value <0.001), activated CD8 (Cox P value <0.001), and mast cell (Cox P value =0.003) (*Figure 2*). All of them had associations with patient survival in the ER+ group across the three datasets. The Cox P value of the CD4, CD8 and Mast cell gene set were P<0.001, P=0.002, and P=0.019 for GSE2034; P<0.001, P=0.003, and P=0.025 for GSE2990, respectively. Breast cancer patients with higher expression in both activated CD4 and activated CD8 gene sets had poor survival while upregulated mast cell gene set was associated with better prognosis. The result was in consistence with the earlier study which suggested the correlation between mast cell activities and lower grade tumors in stromal cells (31,32).

Multi-variate Cox analysis with other prognostic features

In addition to univariate Cox regression model for immune-related gene sets, we explored the potential to cooperate the immune cell gene sets with other well-known prognostic features. One prognostic features used in clinic (LN status) and two factors identified in the previous studies (cell cycle gene signature and ER+ consensus prognostic genes) were considered. The results showed that two tumor infiltration leukocytes can be an independent predictor comparing with the prognostic features (*Table S3*). Activated CD8 and mast cell signatures showed the statistical significance (P=0.03 and P=0.04), comparing with the variable of "lymph node status" which was one of the factors of current staging system used in clinic.

Construction the TIL interaction network

We assessed connectivity of the 7 out of 9 immune cell gene sets identified from ER+ group by Pearson correlation. The results are shown in *Figure 3*. In the ER+ group, two groups of gene sets showed strong correlations. One group contains effector memory CD4, activated CD4 and activated CD8 gene set while the other group includes mast cells, Treg, eosinophil and DC gene set. In the ER– group, similar correlation can be observed. Although not all these gene sets were correlated to patient survival in the ER– group, association between these immune signaling still appeared in both ER status groups. In addition, the highly correlated gene sets all fell into the same group based on the hazard ratio. The effecter memory CD4, activated CD4, and activated CD8 gene set all belonged to the risk group and the mast cells, Treg, eosinophil, DC gene set were all in protective group. This supports the implication of the cooperative effect of these immune cell gene sets as we found in univariate Cox regression model. Previous studies also revealed that complicated network between the immune-regulatory processes may be related to breast cancer progression (17,33).

Seven identified immune cell gene sets were analyzed by two-way unsupervised hierarchical clustering for the purpose of testing the predictive power of these gene sets. The clustering result is shown in *Figure 4*. There were apparently two clusters in the dendrogram. On the left is the cluster of patients with poor survival, which tends to have higher expression in the three gene sets of risk group and lower expression in the four gene sets of protective group. On the other hand, on the right is the cluster of patients with better survival, and their expression in the seven gene sets is the opposite of the one on the left (higher expression in the protective group and lower expression in the risk group). Such result revealed the potential of these seven gene sets as predictive signature.

Functional relevance in the integrative subgroups

We conducted GSEA software to investigate the underlying mechanisms that affect the survival of the two clusters of patients. After computing the enrichment scores, we set the filtering threshold as FDR <0.001 and absolute enrichment score >0.6. A total of 27 gene sets were identified in cluster A (*Table S4*) and 51 gene sets were identified in cluster B (*Table S5*). Among these gene sets, several gene sets have been reported to be associated with breast cancer. We further interpreted two gene sets, "SHEN SMARCA2 TARGETS UP" and "SMID BREAST CANCER NORMAL LIKE UP" (*Figure 5*). In cluster A, "SHEN SMARCA2 TARGETS UP" is a set of genes whose expressions are positively correlated with the *SMARCA2* gene. It has been reported that the *SMARCA2* gene is associated with the transcription of certain genes related to poor survival of breast cancer patients (34,35). As for cluster B, an identified gene set, "SMID BREAST

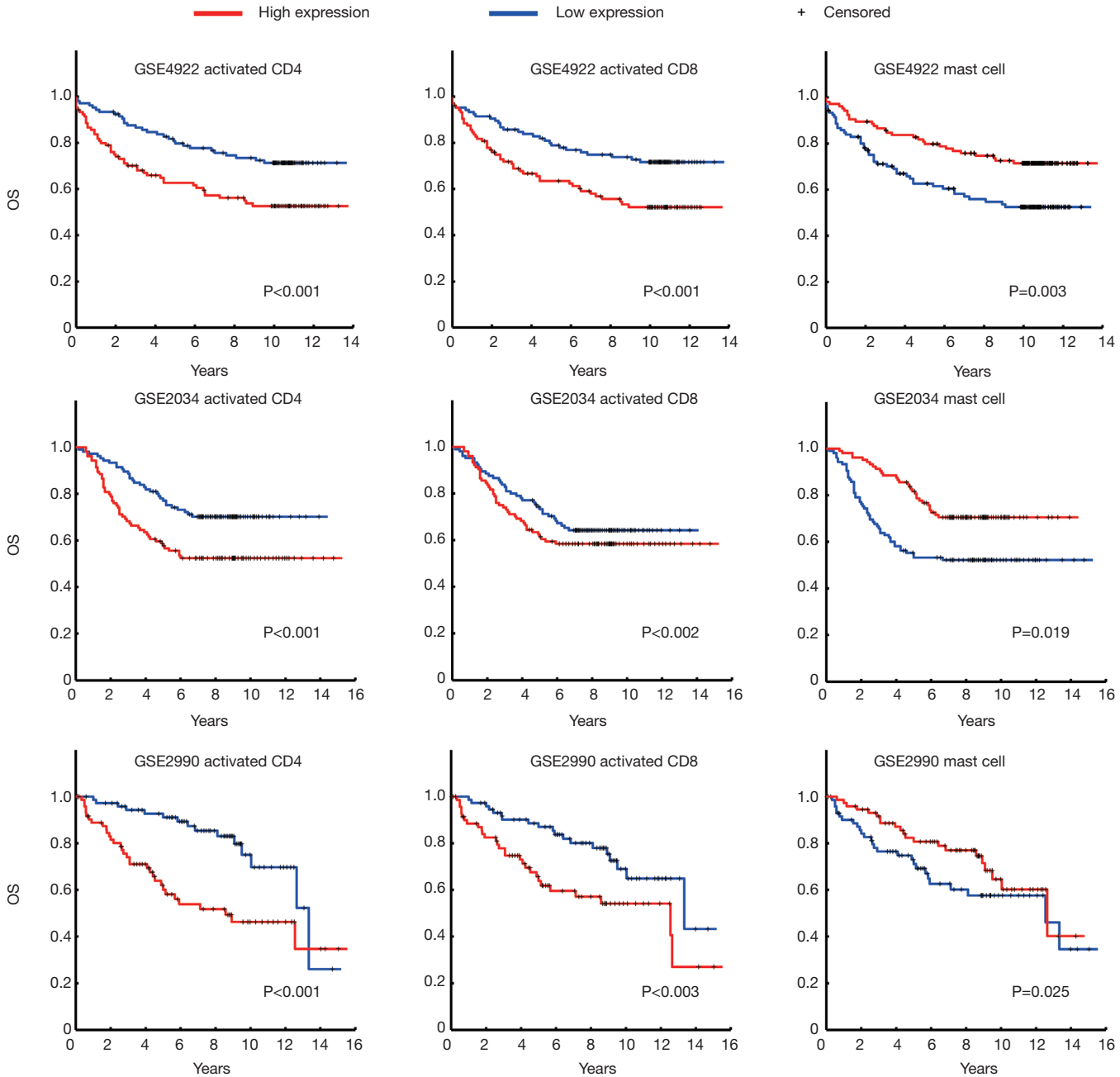

Figure 2 Kaplan-Meier survival curves of activated CD4, CD8 and mast cell activities of three GEO data sets. The three immune signatures were analyzed with overall survival, and the survival relevance of the three immune signatures were confirmed in validation data sets. Activated CD4 and activated CD8 were poor prognostic factors, while mast cells were favorable. All P values were estimated in Cox regression model.

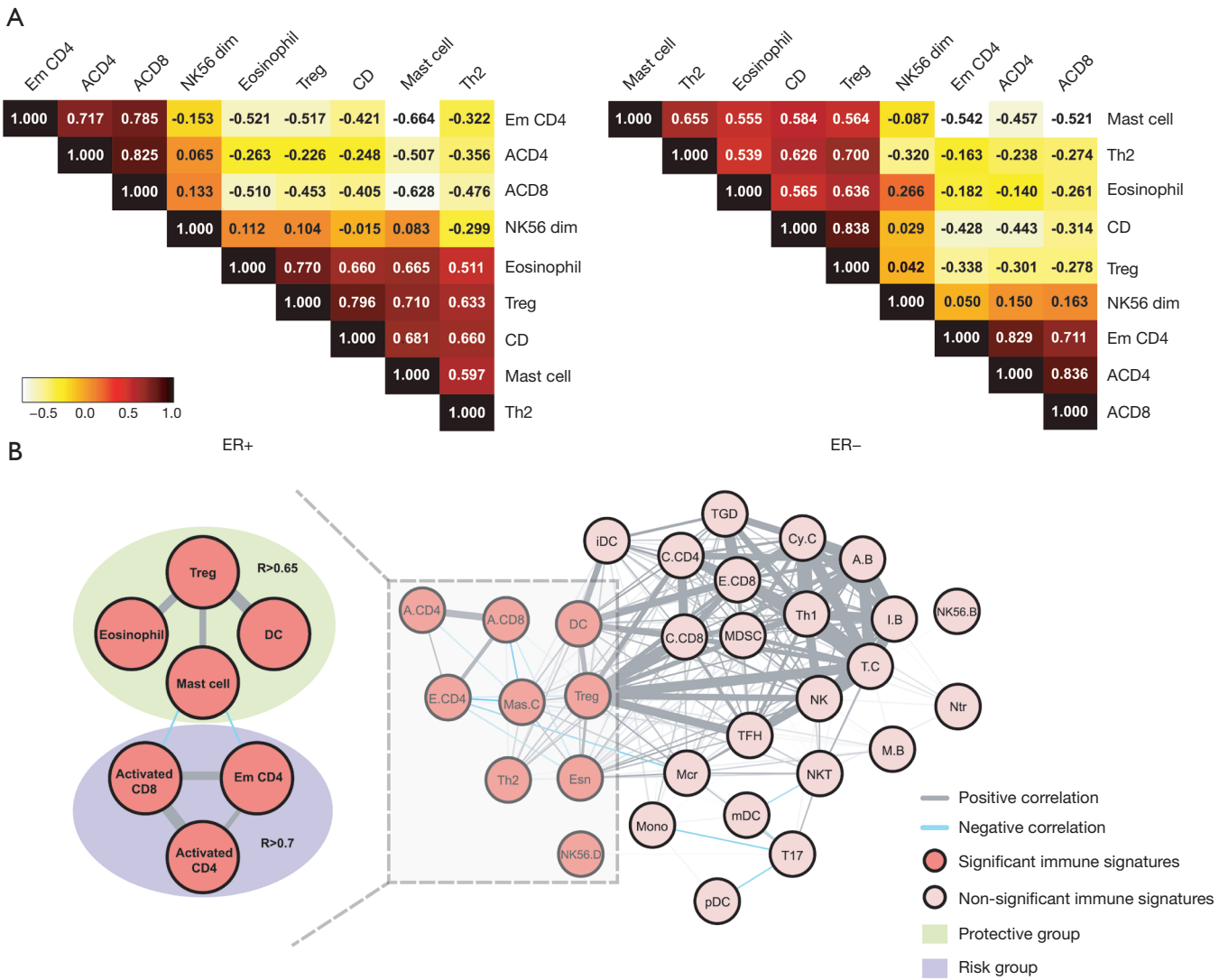

Figure 3 Correlation of immune signatures and networks. Upper panel, Pearson correlation coefficient of each pair of immunological gene sets in both ER groups. Colormap denoted the relationship of each pair of immune signatures. We set the filter based on the significance ($P<0.05$) and thresholds on positive correlation coefficient as higher than 0.65, negative correlation coefficient lower than –0.6 in ER positive; positive correlation coefficient higher than 0.53 or negative correlation coefficient lower than –0.52 was set in the ER negative group. Lower panel, networks constructed in the correlation coefficient to investigate the connection between immune activities. High correlation ($R>0.7$) was identified between the activated CD4, CD8 and the effector memory CD4. Another cluster of highly correlated gene sets was composed of mast cell, Treg, DC and eosinophil ($R>0.65$). Noteworthy, we also focused on the negative regulation between immune activities and found significant negative correlation coefficients between mast cell and effector memory CD4, activated CD8 ($R<-0.6$). ER, estrogen receptor; Treg, regulatory T cell; DC, dendritic cell; AB, activated B cell; IB, immature B cell; Esn, eosinophil; Mono, monocytes; TC, T cells; MB, memory B cell; Mcr, macrophages; CyC, cytotoxic cells.

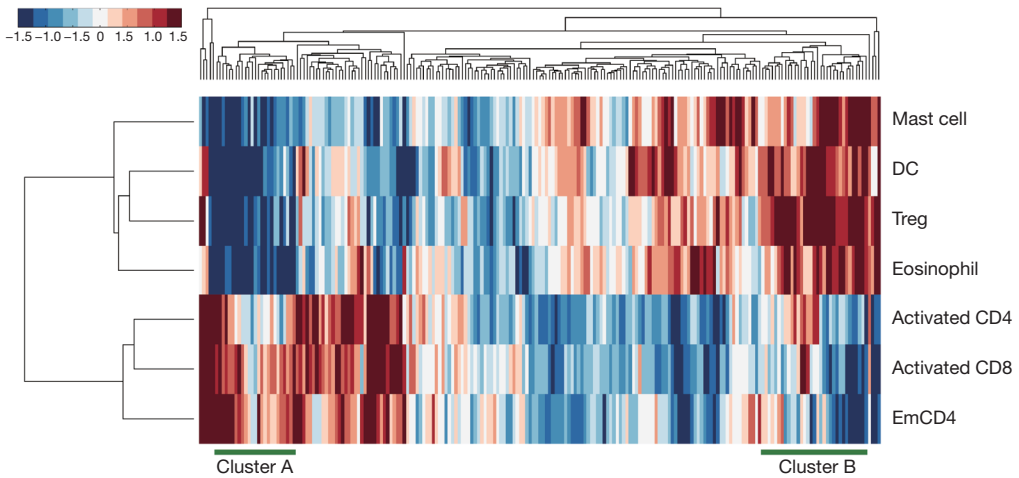

Figure 4 Expression profiles of seven highly correlated immune gene sets. Red indicates high expression and blue indicates low expression profiles. Input data were normalized and presented in the log-2 scale. Two clusters of samples were highlighted to represent the differential expression of protective group immune activities, which comprised mast cells, Treg, DC and eosinophil. Cluster A was low expression profile in selected immune activities and cluster B was high expression. These two clusters were further analyzed for associated functional annotations. Treg, regulatory T cell; DC, dendritic cell.

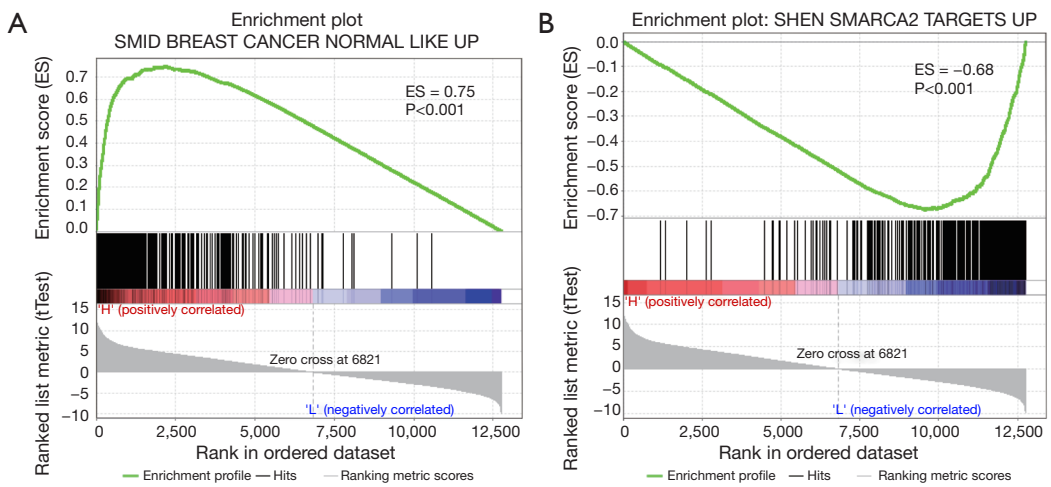

Figure 5 Selected enrichment plots of two clusters of immune activities. Two clusters of expression profiles were analyzed by the analytical software GSEA to rank all 12,754 genes in GSE4922 according to the Student's *t*-test for the differences between two clusters. With a calculation of the enrichment score (ES), enrichment plot illustrates specific gene set associated with the difference between two cohorts. Significance of each enrichment score was calculated by 1,000 permutation tests. (A) Enrichment plot of the SMID_BREAST_CANCER_NORMAL_LIKE_UP gene set identified in cluster B. Protective immune signature was associated with normal-like breast cancer subtype-upregulating genes reported by Smid *et al.*; (B) enrichment plot of the SHEN_SMARCA2_TARGETS_UP function found in cluster A.

CANCER NORMAL LIKE UP", has been reported to have differential expression in normal-like molecular subtypes of breast cancer (36).

Discussion

In this study, we utilized the gene set approach to identify the activities of TILs subtypes in ER+ breast cancer.

Nine of them were associated with patients' survival. Nevertheless, only one of the nine immune cell gene sets showed correlation to prognosis in ER– breast cancer. The results indicated the characteristics of different breast cancer molecular subtypes could influence the interaction of immune system during tumor development and thus affect patient's survival. Through the analysis of Pearson correlation, the nine gene sets were separated into two groups. One contains the immune cells associated with adaptive immunity (activated CD4, Activated CD8, and Em CD4), and the other belongs to the innate immunity (Treg, DC, eosinophil, and mast cell). We speculated this may arise with tumor progression. A previous study (17) proposed that adaptive leukocytes played contrast roles in acute and chronic inflammation. Two types of inflammation were directly regulated by the cytotoxic, Th1 cells and the myeloid suppressors (Th2, CD4+ T cell and Treg). With accumulation of innate leukocytes regulated by chronic activation of B cells, tumor promotion was activated and transforming tumor rejection properties of innate immune cells. This immune surveillance dominated tumor development in many cancer types. In view of that, malignant tumor could induce the acute inflammation and adaptive immunity by proliferation and invasion. For innate immunity, favorable prognosis showed in the group of TILs concluded owing to trigger acute inflammation response while immune system encounter neo-antigens. Two groups of TILs depicted reverse survival by triggering different inflammation response, and these inter and intra-cluster connections can be the evidence of immunity contexture in tumor development.

Immune cells cooperate in immune response with a complicated mechanism; how to parse these immune subtypes networks in a specific cancer type cohort still remains a challenging task. We used Cox regression model and Kaplan-Meier estimator to investigate the prognosis trend with several TIL subtypes. Some of them are associated with adverse prognosis, such as regulatory T cells (37,38). However, several literatures suggested that Treg appear to play a dual role depending on tumor microenvironment (39), which can regulate the inflammation and trigger different signal pathways to oncogenesis, acting as antitumor suppressor or tumor suppressor. In this study, Treg accumulation led to favorable prognosis. With variety of immunosuppressive antigens, immune system signal transduction and cytokines secretion were interfered, causing immune dysfunction (40), that may be a reason of Treg diversity. The activated CD4 and CD8 immune signatures, analogous to regulatory T cell, were not expected in survival analysis. We inferred that this was caused by the divergence in CD4 and CD8 expression on the T cells in different clinical variation. Data source of gene collection was multi-faceted, which contained various clinical conditions. Breast cancer cohorts may have specificity of these two gene sets but with different prognosis expectancy.

In the survival analysis, we demonstrated some of the TILs subtypes were associated with overall survival, inclusive of activated CD8 and mast cells. TILs resulting in favorable and poor prognosis were different from the literature of immune subtype gene sets source. In Angelova's research (18), few survival associated immune subtypes showed inverse trend to our study. For instance, regulatory T cell was a poor prognostic factor among TILs subtypes and also an independent survival predictor, but that was not found in our study, reflecting the molecular subtype diversities in cancer development and patient heterogeneity. Here, we focus subsequently analysis on ER+ cohort. Based on the concordant results achieved in two validation data sets, we confirmed the TILs and patient's outcome modulated by ER status.

We pooled nine survival correlated immune cell gene sets to evaluate the independence of prognostic predictors. Cell cycle and LN status were known to predict patient survival. In our multivariate analysis LN status remained a dominant factor among all predictors in other survival associated TIL subtypes, overtaking the cell cycle genes. Thus, we can take activated CD8, mast cell and LN status as a group of powerful prognostic predictors in ER+ patients.

Functional annotations suggested that differentially activated immune signatures play important roles on other molecular signatures upon breast cancer oncogenesis. In large retrospective studies, mast cell was differentially correlated to prognostic in various cancer types (18,41,42). It was reported with the capability of defense against allergic pathogens. We analyzed biological functions associated with two clusters of protective immune signatures expression. Because in protective immune signatures, mast cell was a significant prognostic factor in multivariate Cox regression, we further investigated the mast cell functioning as principle on breast tumor. In two clusters of functional annotation, there appeared inverse regulation of a gene set modulate *SMARCA2* in prostate cancer (43). Upregulation of *SMARCA2* genes was seen in low expression of protective immune signatures, and downregulation of *SMARCA2* genes in high expression of protective immune activities. *SMARCA2* encodes the SWI/SNF chromatin remodeling complex which is an effector of DNA repair transcription.

It has been demonstrated a functional association between *BRCA1* and the SWI/SNF-related complex; with their combination and participation in transcriptional regulatory mechanism, mutation occurring in any or both of them can be a potential factor causing breast cancers (44). Among lower enrichment score annotated functions (ES =–0.58), low expression of protective group showed connectivity with the genes down regulated in the brain relapse of breast cancer (36). Based on the findings of this study, future works may build up more accurate predicting criteria and clear modulation network, contributing a genetic framework to promote immunotherapy.

Acknowledgments

Funding: This work was supported by the Taichung Veterans General Hospital (TCVGH-YM1040205, TCVGH-YM1050204, and TCVGH-VTA105A1-2).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/tcr.2016.04.09). Eric Y. Chuang serves as the Editor-in-Chief of Translational Cancer Research. The other author has no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA 2015;65:5-29.
2. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature 2000;406:747-52.
3. Malhotra GK, Zhao X, Band H, et al. Histological, molecular and functional subtypes of breast cancers. Cancer Biol Ther 2010;10:955-60.
4. Masood S. Estrogen and progesterone receptors in cytology: a comprehensive review. Diagn Cytopathol 1992;8:475-91.
5. Mohibi S, Mirza S, Band H, et al. Mouse models of estrogen receptor-positive breast cancer. J Carcinog 2011;10:35.
6. Trichopoulos D, MacMahon B, Cole P. Menopause and breast cancer risk. J Natl Cancer Inst 1972;48:605-13.
7. Fisher B, Costantino JP, Wickerham DL, et al. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P1 Study. J Natl Cancer Inst 1998;90:1371-88.
8. Loi S, Michiels S, Salgado R, et al. Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: results from the FinHER trial. Ann Oncol 2014;25:1544-50.
9. Carbognin L, Pilotto S, Nortilli R, et al. Predictive and Prognostic Role of Tumor-Infiltrating Lymphocytes for Early Breast Cancer According to Disease Subtypes: Sensitivity Analysis of Randomized Trials in Adjuvant and Neoadjuvant Setting. Oncologist 2016;21:283-91.
10. Adams S, Gray RJ, Demaria S, et al. Prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancers from two phase III randomized adjuvant breast cancer trials: ECOG 2197 and ECOG 1199. J Clin Oncol 2014;32:2959-66.
11. Ibrahim EM, Al-Foheidi ME, Al-Mansour MM, et al. The prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancer: a meta-analysis. Breast Cancer Res Treat 2014;148:467-76.
12. Pruneri G, Vingiani A, Bagnardi V, et al. Clinical validity of tumor-infiltrating lymphocytes analysis in patients with triple-negative breast cancer. Ann Oncol 2016;27:249-56.
13. Ansell SM, Lesokhin AM, Borrello I, et al. PD-1 blockade with nivolumab in relapsed or refractory Hodgkin's lymphoma. N Engl J Med 2015;372:311-9.
14. Topalian SL, Hodi FS, Brahmer JR, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. N Engl J Med 2012;366:2443-54.
15. Brahmer JR, Tykodi SS, Chow LQ, et al. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. N Engl J Med 2012;366:2455-65.
16. de Visser KE, Eichten A, Coussens LM. Paradoxical roles of the immune system during cancer development. Nat Rev Cancer 2006;6:24-37.
17. DeNardo DG, Coussens LM. Inflammation and breast

cancer. Balancing immune response: crosstalk between adaptive and innate immune cells during breast cancer progression. Breast Cancer Res 2007;9:212.
18. Angelova M, Charoentong P, Hackl H, et al. Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. Genome Biol 2015;16:64.
19. Ivshina AV, George J, Senko O, et al. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. Cancer Res 2006;66:10292-301.
20. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005;365:671-9.
21. Sotiriou C, Wirapati P, Loi S, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst 2006;98:262-72.
22. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30:207-10.
23. Kaplan EL, Meier P. Nonparametric-Estimation from Incomplete Observations. J Am Stat Assoc 1958;53:457-81.
24. Cox DR. Regression Models and Life-Tables. Journal of the Royal Statistical Society. Series B (Methodological) 1972;34:187-220.
25. Mizuno H, Nakanishi Y, Ishii N, et al. A signature-based method for indexing cell cycle phase distribution from microarray profiles. BMC Genomics 2009;10:137.
26. Teschendorff AE, Naderi A, Barbosa-Morais NL, et al. A consensus prognostic gene expression classifier for ER positive breast cancer. Genome Biol 2006;7:R101.
27. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545-50.
28. Liberzon A, Subramanian A, Pinchback R, et al. Molecular signatures database (MSigDB) 3.0. Bioinformatics 2011;27:1739-40.
29. Gajewski TF, Schreiber H, Fu YX. Innate and adaptive immune cells in the tumor microenvironment. Nat Immunol 2013;14:1014-22.
30. Vesely MD, Kershaw MH, Schreiber RD, et al. Natural innate and adaptive immunity to cancer. Annu Rev Immunol 2011;29:235-71.
31. Dabiri S, Huntsman D, Makretsov N, et al. The presence of stromal mast cells identifies a subset of invasive breast cancers with a favorable prognosis. Mod Pathol 2004;17:690-5.
32. Rajput AB, Turbin DA, Cheang MC, et al. Stromal mast cells in invasive breast cancer are a marker of favourable prognosis: a study of 4,444 cases. Breast Cancer Res Treat 2008;107:249-57.
33. Spörri R, Reis e Sousa C. Inflammatory mediators are insufficient for full dendritic cell activation and promote expansion of CD4+ T cell populations lacking helper function. Nat Immunol 2005;6:163-70.
34. Bertucci F, Nasser V, Granjeaud S, et al. Gene expression profiles of poor-prognosis primary breast cancer correlate with survival. Hum Mol Genet 2002;11:863.
35. Wilson BG, Roberts CW. SWI/SNF nucleosome remodellers and cancer. Nat Rev Cancer 2011;11:481-92.
36. Smid M, Wang Y, Zhang Y, et al. Subtypes of breast cancer show preferential site of relapse. Cancer Res 2008;68:3108-14.
37. Liu F, Lang R, Zhao J, et al. CD8+ cytotoxic T cell and FOXP3+ regulatory T cell infiltration in relation to breast cancer survival and molecular subtypes. Breast Cancer Res Treat 2011;130:645-55.
38. Bates GJ, Fox SB, Han C, et al. Quantification of regulatory T cells enables the identification of high-risk breast cancer patients and those at risk of late relapse. J Clin Oncol 2006;24:5373-80.
39. Whiteside TL. What are regulatory T cells (Treg) regulating in cancer and why? Semin Cancer Biol 2012;22:327-34.
40. Ferrone S, Whiteside TL. Tumor microenvironment and immune escape. Surg Oncol Clin N Am 2007;16:755-74, viii.
41. Conti P, Castellani ML, Kempuraj D, et al. Role of mast cells in tumor growth. Ann Clin Lab Sci 2007;37:315-22.
42. Imada A, Shijubo N, Kojima H, et al. Mast cells correlate with angiogenesis and poor outcome in stage I lung adenocarcinoma. Eur Respir J 2000;15:1087-93.
43. Shen H, Powers N, Saini N, et al. The SWI/SNF ATPase Brm is a gatekeeper of proliferative control in prostate cancer. Cancer Res 2008;68:10154-62.
44. Bochar DA, Wang L, Beniya H, et al. BRCA1 Is Associated with a Human SWI/SNF-Related Complex: Linking Chromatin Remodeling to Breast Cancer. Cell 2000;102:257-65.

Cite this article as: Chang YH, Chiu YC, Hsu YC, Tsai HM, Chuang EY, Hsiao TH. Applying gene set analysis to characterize the activities of immune cells in estrogen receptor positive breast cancer. Transl Cancer Res 2016;5(2):176-185. doi: 10.21037/tcr.2016.04.09

Supplementary

Table S1 Summary of microarray datasets used in this study

Data set accession num.	GSE4922	GSE2990	GSE2034
Usage	Discovery	Validation	Validation
Ratio of samples ER status (ER+/ER–)	211/34	149/34	209/77
Microarray platform	Affymetrix human genome U133A array		
Reference	(19)	(21)	(20)

Patients with missing survival index and the ER status are excluded from the study.

Table S2 Validation analysis in GSE2034 and GSE2990 data sets

Gene set	GSE2034		GSE2990	
	HR	P value	HR	P value
Activated CD8	2.14	0.002	2.07	0.003
Activated CD4	2.02	<0.001	2.36	<0.001
Mast cells	0.40	0.019	0.33	0.025
Effector memory CD4	2.52	0.003	1.89	0.072
Th2	0.27	0.001	0.30	0.026
Treg	0.37	0.002	0.58	0.131
Eosinophil	0.39	0.007	0.94	0.906
NK56 dim	0.48	0.043	2.90	0.024
DC	0.71	0.233	0.67	0.255

Th, T-helper; Treg, regulatory T cell; NK, natural killer; DC, dendritic cell.

Table S3 Multivariate Cox hazard analysis of immune infiltration and clinical variables

Gene sets	Hazard ratio	P value
Activated CD8	1.90 (1.06–4.29)	0.028
LN status	1.88 (1.17–3.02)	0.007
Cell cycle	0.82 (0.33–2.01)	0.660
CPGs	1.49 (0.65–3.45)	0.339
Mast cells	0.41 (0.17–0.92)	0.040
LN status	1.98 (1.24–3.16)	0.004
Cell cycle	0.92 (0.40–2.11)	0.848
CPGs	1.67 (0.76–3.68)	0.196

CPGs, the ER+ consensus prognostic genes.

Table S4 The 27 enriched gene sets in cluster A

Gene set	Size	ES
YANG_BREAST_CANCER_ESR1_BULK_UP	25	–0.71
YANG_BREAST_CANCER_ESR1_UP	36	–0.70
YANG_BREAST_CANCER_ESR1_LASER_UP	31	–0.69
VANTVEER_BREAST_CANCER_ESR1_UP	133	–0.69
MOOTHA_VOXPHOS	84	–0.68
ABRAMSON_INTERACT_WITH_AIRE	42	–0.68
SHEN_SMARCA2_TARGETS_UP	405	–0.68
ZHAN_MULTIPLE_MYELOMA_SUBGROUPS	30	–0.67
SCHAEFFER_PROSTATE_DEVELOPMENT_AND_CANCER_BOX4_DN	27	–0.67
SMID_BREAST_CANCER_RELAPSE_IN_BRAIN_DN	80	–0.65
YAO_TEMPORAL_RESPONSE_TO_PROGESTERONE_CLUSTER_10	56	–0.65
WONG_PROTEASOME_GENE_MODULE	46	–0.64
GOLGI_VESICLE_TRANSPORT	43	–0.64
WONG_MITOCHONDRIA_GENE_MODULE	190	–0.63
CREIGHTON_AKT1_SIGNALING_VIA_MTOR_UP	34	–0.63
SCHLOSSER_MYC_TARGETS_REPRESSED_BY_SERUM	156	–0.62
MOREAUX_MULTIPLE_MYELOMA_BY_TACI_DN	137	–0.62
TRANSLATIONAL_INITIATION	33	–0.62
DOANE_BREAST_CANCER_ESR1_UP	106	–0.61
LIEN_BREAST_CARCINOMA_METAPLASTIC_VS_DUCTAL_DN	77	–0.61
KAMMINGA_EZH2_TARGETS	39	–0.61
PROTEIN_RNA_COMPLEX_ASSEMBLY	55	–0.61
DING_LUNG_CANCER_EXPRESSION_BY_COPY_NUMBER	85	–0.61
CELLULAR_PROTEIN_CATABOLIC_PROCESS	51	–0.60
BIDUS_METASTASIS_UP	189	–0.60
GKCGCNNNNNNNTGAYG_UNKNOWN	43	–0.60
RHODES_CANCER_META_SIGNATURE	64	–0.60

All FDR q-values <0.001. ES, enrichment score.

Table S5 The 51 enriched gene sets in cluster B

Gene set	Size	ES
WINTER_HYPOXIA_DN	43	0.80
BOYAULT_LIVER_CANCER_SUBCLASS_G5_DN	26	0.78
ONO_AML1_TARGETS_DN	35	0.77
WORSCHECH_TUMOR_EVASION_AND_TOLEROGENICITY_UP	27	0.76
WILENSKY_RESPONSE_TO_DARAPLADIB	27	0.76
FARMER_BREAST_CANCER_CLUSTER_1	42	0.75
SMID_BREAST_CANCER_NORMAL_LIKE_UP	464	0.75
BOSCO_TH1_CYTOTOXIC_MODULE	87	0.74
HAHTOLA_SEZARY_SYNDROM_DN	41	0.73
VILIMAS_NOTCH1_TARGETS_UP	47	0.73
WORSCHECH_TUMOR_REJECTION_UP	48	0.72
KUROZUMI_RESPONSE_TO_ONCOCYTIC_VIRUS	42	0.72
TARTE_PLASMA_CELL_VS_B_LYMPHOCYTE_DN	37	0.70
BERTUCCI_INVASIVE_CARCINOMA_DUCTAL_VS_LOBULAR_DN	41	0.70
GAURNIER_PSMD4_TARGETS	67	0.70
WALLACE_PROSTATE_CANCER_RACE_UP	284	0.69
TIAN_TNF_SIGNALING_VIA_NFKB	28	0.69
T_CELL_ACTIVATION	39	0.68
NAKAYAMA_SOFT_TISSUE_TUMORS_PCA1_UP	74	0.67
ONO_FOXP3_TARGETS_DN	37	0.67
DUNNE_TARGETS_OF_AML1_MTG8_FUSION_UP	46	0.66
MORI_LARGE_PRE_BII_LYMPHOCYTE_DN	54	0.65
SHEN_SMARCA2_TARGETS_DN	344	0.64
FLECHNER_BIOPSY_KIDNEY_TRANSPLANT_REJECTED_VS_OK_UP	86	0.64
SANA_RESPONSE_TO_IFNG_UP	59	0.64
POOLA_INVASIVE_BREAST_CANCER_UP	284	0.64
WIELAND_UP_BY_HBV_INFECTION	99	0.63
MCLACHLAN_DENTAL_CARIES_UP	244	0.63
KRAS.BREAST_UP.V1_UP	134	0.63
LIM_MAMMARY_LUMINAL_MATURE_DN	83	0.62
LEE_EARLY_T_LYMPHOCYTE_DN	49	0.62
HADDAD_T_LYMPHOCYTE_AND_NK_PROGENITOR_DN	63	0.62
JAATINEN_HEMATOPOIETIC_STEM_CELL_DN	197	0.62
KIM_GLIS2_TARGETS_UP	81	0.62

Table S5 *(continued)*

Table S5 *(continued)*

Gene set	Size	ES
WONG_ENDMETRIUM_CANCER_DN	63	0.62
KLEIN_PRIMARY_EFFUSION_LYMPHOMA_DN	58	0.62
CELL_ACTIVATION	66	0.62
BASSO_CD40_SIGNALING_UP	99	0.62
REGULATION_OF_IMMUNE_SYSTEM_PROCESS	56	0.61
LEE_DIFFERENTIATING_T_LYMPHOCYTE	154	0.61
G_PROTEIN_SIGNALING_COUPLED_TO_CAMP_NUCLEOTIDE_SECOND_MESSENGER	63	0.61
LINDSTEDT_DENDRITIC_CELL_MATURATION_A	64	0.61
KRAS.LUNG.BREAST_UP.V1_UP	136	0.61
LOCOMOTORY_BEHAVIOR	88	0.61
SCHUETZ_BREAST_CANCER_DUCTAL_INVASIVE_UP	343	0.60
LEUKOCYTE_ACTIVATION	61	0.60
LYMPHOCYTE_ACTIVATION	55	0.60
LIEN_BREAST_CARCINOMA_METAPLASTIC_VS_DUCTAL_UP	74	0.60
LINDGREN_BLADDER_CANCER_CLUSTER_2B	332	0.60
MCLACHLAN_DENTAL_CARIES_DN	238	0.60
IMMUNE_RESPONSE	215	0.60

All FDR q-values <0.001. ES, enrichment score.