

CD8⁺ T Lymphocyte Co-expression Genes Correlates With Immune Microenvironment and Overall Survival in Breast Cancer

Jialing Jiang

China Medical University

Yutao Wang

China Medical University

Yi Zhao (✉ zhao1969yi@163.com)

China Medical University

Research Article

Keywords: CD8⁺ T lymphocyte, Antigen presentation, Response to interferon- γ , Weighted gene co-expression network analysis (WGCNA), Tumor microenvironment, Breast cancer

Posted Date: January 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-146843/v1>

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

[Read Full License](#)

Abstract

Purpose: To identify CD8⁺ T lymphocyte related co-expressed genes that increase CD8⁺ T lymphocyte proportions in breast cancer, and to elucidate the underlying mechanisms among relevant genes in the tumor microenvironment.

Method: We obtained breast cancer expression matrix data, patient phenotype following information from TCGA–BRCA FPKM. Tumor purity, immune score, stromal score, and estimate score were calculated using the ‘estimate’ package in R. The CD8⁺ T lymphocyte proportions in each breast carcinoma sample were estimated using the CIBERSORT algorithm. The samples with $P < 0.05$ were considered to be significant and were taken into the weighted gene co-expression network analysis. Based on the CD8⁺ T lymphocyte proportion and tumor purity, we generated CD8⁺ T lymphocyte co-expression networks and selected the most CD8⁺ T lymphocyte related module as our interested co-expression modules. We constructed a Cox regression risk model based on CD8⁺ T cell-related factors and explored their correlations with microenvironmental indicators.

Results: A breast carcinoma CD8⁺ T lymphocyte proportion co-expression yellow module was determined. Four CD8⁺ T lymphocyte proportion co-expressed genes (CD74, HLA-DMA, HCST and GIMAP4) were determined to increase the CD8⁺ T lymphocyte proportion levels in breast cancer patients. The yellow module was significantly enriched in the antigen presentation process, cellular response to interferon–gamma, and leukocyte proliferation. Subsequently, we generated risk scores based on CD74, HLA-DMA, HCST, and GIMAP4. The AUC of the CD8⁺ T lymphocyte related risk score was 0.66. The risk score showed significant prognostic ability in various subgroups. Expression levels of proteins encoded by CD74, HLA-DMA, HCST and GIMAP4 were lower in the breast carcinoma samples than in normal tissue, suggesting that expression differences both at the mRNA and the protein levels.

Conclusion: These four CD8⁺ T lymphocyte proportion co-expression genes increase CD8⁺ T lymphocyte in breast cancer by an antigen presentation process. The mechanism might suggest new pathways to improve outcomes in patients who do not benefit from immune therapy.

Introduction

Breast cancer has entered the era of individualized treatment by way of molecular classification. In addition to traditional surgery, chemotherapy, and radiotherapy, breast cancer treatment includes endocrine therapy, molecular targeted therapy, and immunotherapy [1]. Of these, immunotherapy has achieved remarkable effects in the treatment of a variety of malignant tumors. However, immunotherapy for breast cancer is a recent development. Breast cancer is a ‘cold’ tumor in terms of immunotherapy, preliminary exploratory studies of PD-1/PD-L1 inhibitors using monotherapy have not been not, and the population that benefits is very limited [1]. At present, the most critical issue in breast cancer immunotherapy is the issue of selection of the appropriate population and reasonable treatment mode,

so that more patients can benefit from immunotherapy, survival can be extended, and quality of life can be improved.

Lack of CD8+ tumor-infiltrating lymphocytes, low PD-1 expression, and tumor mutation burden factors are thought to be the primary influencing factors leading to insensitivity to immunotherapy in advanced breast cancer. In triple-negative breast cancer, the positive expression rate of PD-L1 is about 20%, which is higher than that of other subtypes of breast cancer [2]. An advanced breast cancer analysis suggested that PD-L1 is not only related to advanced breast cancer prognosis, and it is also a biomarker for screening suitable populations for immunotherapy [3]. A meta-analysis involving 8583 breast cancer patients of various subtypes suggested that PD-L1 overexpression is significantly negatively correlated with overall survival of patients, and the mechanism may be that high PD-L1 expression promotes breast cancer immune escape [4]. Although clinical immunohistochemistry can be used to assess tumor PD-L1 expression levels, there remain limitations to using PD-L1 expression as a biomarker of immunotherapy sensitivity. Tumor PD-L1 expression is heterogeneous [5] and is affected by previous treatments such as chemotherapy and radiotherapy [6].

Tumor-infiltrating lymphocytes are polymorphic, mainly existing in the microenvironment of tumor tissues; they include CD4+, CD8+ T cells, B cells, and NK cells [7]. Studies have shown that there are more TILs in the tumor microenvironment of HER2-positive breast cancer and this is related to prognosis [8]. In a meta-analysis of non-small cell lung cancer, more CD8+ TILs were associated with improvement in overall survival [9]. In patients with advanced melanoma treated with pembrolizumab, the density of CD8+ T cells in the invasion margin and tumor center of the tissue specimens of responders were higher than those of non-responders [10].

In the present study, we hypothesized that increasing the content of CD8+ T lymphocytes would improve outcomes after immunotherapy. By constructing a co-expression network of CD8+ T lymphocyte content, we explored the biological functions and related co-expression factors that are most related to CD8+ lymphocyte content.

Method

CD8+ T cell proportion, tumor purity, and tumor mutation burden

We downloaded The Cancer Genome Atlas TCGA–BRCA FPKM data (<http://cancergenome.nih.gov/>) containing 1097 samples. GSE78220 [11] was also download from the GEO database. We evaluated CD8+ T lymphocyte cell proportions based on the LM22 matrix using the CIBERSORT [12] algorithm. Breast tissue samples with $P < 0.05$ were considered to be significant, and were taken into the subsequent analysis. The Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data (ESTIMATE) is a method that infers the fraction of stromal and immune cells using gene expression signatures [13]. Using the ESTIMATE package, we calculated tumor purity in each breast cancer sample. Tumor mutation burden (TMB) per megabyte was calculated by dividing the total number of mutations by the size of the target coding region [14, 15].

Co-expression network generation

Weighted gene co-expression network analysis (WGCNA) is a system biology method that transforms correlations into connection weights or topology overlap values [16]. We used this method to generate a CD8+ T lymphocyte proportion co-expressing network. The expression patterns are similar for genes with the same pathway and biological effect [17, 18]. In this study, we built a scale-free topology network, set the soft threshold at 5, R square = 0.93, slope = -2.09, and we set the number of genes in the minimum module at 30. The CD8+ T lymphocyte proportion was considered for phenotype files in the WGCNA analysis. In this manner, a cluster of CD8+ T lymphocyte proportion-related genes with similar pathway process were determined in the same module. We identified factors with CD8+ T lymphocyte correlation greater than 0.4, and module correlation greater than 0.6 in the most co-expression modules.

Function enrichment and protein-protein network of co-expression module

The Database for Annotation, Visualization and Integrated Discovery (DAVID, v6.8) is an open source database that performs function enrichment [19]. The Kyoto Encyclopedia of Genes and Genomes (KEGG) [20] (<https://www.genome.jp/kegg/>) and Gene Ontology (GO)[21] (<http://geneontology.org/>) analysis were applied to determine the biological function, cellular component and molecular function in each co-expression module. Cytoscape was used to conduct the protein-protein interaction network for the co-expression genes.

CD8+ T lymphocyte genes prognostic value

Kaplan–Meier analysis was used to calculate the clinical outcome significance of these CD8+ T lymphocyte co-expression genes. Subsequently, a Cox regression hazard model was built based on the CD8+ T lymphocyte co-expression genes. We evaluated the CD8+ T lymphocyte co-expression genes prognosis model in various subgroups. Finally, we calculated the difference of these co-expression genes in various subgroups, including tumor purity, survival status, and TMB.

Gene Set Enrichment Analysis and the Human Protein Atlas database

Gene Set Enrichment Analysis (GSEA) [22] was used to calculate the most involved pathway to these co-expression genes. The Human Protein Atlas (HPA) (<http://www.proteinatlas.org>) [23] was used to demonstrate differences in co-expressing genes at the protein level.

Results

The flow chart of the experimental protocol is shown in Figure 1.

CD8+ T lymphocyte, tumor purity, and tumor mutation burden evaluation

We obtained the tumor purity, matrix score, immune score, and tumor mutation burden corresponding to each sample. Using the screening principle of $p < 0.05$, we obtained 860 breast cancer samples accurately

evaluated by CD8+ T lymphocytes (Figure 2A). By integrating the immune microenvironment scoring file with the CD8+ T lymphocyte content samples, we determined WGCNA's phenotype entry files.

CD8+ T lymphocyte co-expression network construction in TCGA

Weighted gene co-expression network analysis (WGCNA) analysis was performed using TCGA-BRCA. A hierarchical clustering tree was built using the dynamic hybrid cutting method (Figure 2B); 22 co-expression models were identified (Figure 2C). The correlation coefficients among CD8+ T lymphocyte proportion, tumor purity, TMB, and co-expression modules are shown in Figure 2C. The yellow module had the strongest correlation with CD8+ T lymphocyte proportion in the TCGA - BRCA cohort (Cor = -0.41; P = 1e-28) (Figure 2C). Based on these findings, we supplemented the heat map of the correlation between the factors in the yellow module (Figure 2D-G). The yellow module showed a significant correlation with CD8+ T cell (Cor = 0.78, p = 9.7e-59), tumor purity (Cor = 0.86, p = 1.7e-83), immune score (Cor = 0.98, p = 1.2e-197) and stromal score (Cor = 0.28, p = 1.9e-06)

CD8+ T lymphocyte co-expression module functional enrichment

We determined 28 CD8+ T lymphocyte proportions positively co-expressing mRNA with coefficient > 0.4 in the TCGA - BRCA yellow module (Table 1). The 28 CD8+ T lymphocyte proportion positively co-expressing mRNA were most significantly enriched in the antigen processing and presentation and response to interferon-gamma, suggesting that these biological processes might promote CD8+ T lymphocyte infiltration in the breast cancer microenvironment (Figure 3A). The CD8+ T lymphocyte negatively co-expressing module was most significantly enriched in extracellular matrix organization (Figure 3B). The protein-protein interaction network of yellow module and green module is shown in Figure 3.

Clinical outcome of CD8+ T lymphocyte infiltration-related genes

To demonstrate their significance on clinical outcomes, we performed survival analysis. The patients in low expression groups for GZMA (TCGA: P = P < 0.001), CD74 (TCGA: P < 0.001), IL2RG (TCGA: P = 0.009), CD3E (TCGA: P < 0.001), CCL5 (TCGA: P < 0.001), CD3D (TCGA: P < 0.001), CORO1A (TCGA: P < 0.001), HLA-DMA (TCGA: P = 0.003), SELPLG (TCGA: P = 0.002), HCST (TCGA: P < 0.001), HLA-DPB (TCGA: P = 0.001), GZMK (TCGA: P = 0.001), CD48 (TCGA: P < 0.001), PAMB9 (TCGA: P = 0.005), CD2 (TCGA: P = 0.003), CD27 (TCGA: P = 0.003), IRF1 (TCGA: P = 0.003), CD8A (TCGA: P = 0.005), GBP4 (TCGA: P = 0.048), TNFRSF1B (TCGA: P = 0.011), GMFG (TCGA: P = 0.006), CST7 (TCGA: P = 0.001), GZMB (TCGA: P = 0.049), PSMB10 (TCGA: P = 0.002) and HLA-E (TCGA: P = 0.046) showed survival risk against high expression groups (Figure 4). These results suggest that these CD8+ T lymphocyte infiltration-related genes act in protective roles in breast cancer.

Cox regression hazard model of CD8+ T lymphocyte co-expression genes

A CD8+ T lymphocyte co-expression gene Cox regression hazard model was conducted based on these breast cancer prognosis protective factors.

$$\text{Risk} = -0.003 * \text{CD74} + 0.045 * \text{HLADMA} - 0.107 * \text{HCST} + 0.032 * \text{GIMAP4}$$

The samples in high risk level samples for breast cancer patients (TCGA: $P < 0.001$; HR = 2.75) (Figure 5) showed survival risk against low risk groups, with the area under curve (AUC) = 0.66 (Figure 5). The risk score was evaluated in various subgroups, including age, gender, stage, tumor purity, and tumor mutation burden, metastasis status, Ki-67, and EGFR. The results were significant in these subgroups.

Clinical phenotype and immunophenotype

Having defined a clinical prognostic risk propensity weighted score consisting of four factors, we then found that these factors were co-expressed with one another and were closely related to the level of CD8+ T lymphocyte infiltration. This factors affect outcomes. Then, to demonstrate the relationship between these factors and clinical phenotype and immunophenotype more specifically, we drew multiple sets of box plots. The content of CD8+ T lymphocytes in the high expression group of these four factors showed a higher level of infiltration, suggesting that our four factors and related biological processes promoted the infiltration of CD8+ T lymphocytes in tumor tissues (Figure 6A). The expression levels of genes in the 5-year mortality group were lower than those of the 5-year survival group, suggesting their protective effect on outcomes. This trend was the same as that of CD8+ T lymphocytes (Figure 6B). Then, we found that expression levels of these factors were low in the high tumor purity group, and these factors in the high immune score group were low (Figure 6CD). These directly or indirectly indicate that these four factors promote the CD8+ T lymphocyte infiltration. We also drew a scatter plot of correlations with clinical stages (Figure 7A), CD8+ T lymphocytes (Figure 7B) and M2 macrophages (Figure 7C) to further illustrate the clinical phenotypic correlation of these factors.

GSEA and HPA

Antigen processing and presentation, the chemokine signaling pathway, B cell receptor signaling pathway, and the T cell receptor signaling pathway were related to the high expression group in CD74, GIMAP4, HCST and HLA-DMA (Figure 8).

We compared the various expression levels of these genes between normal and tumor tissues. Labeling with HPA010592, an antibody against CD74, showed higher intensity in the tumor tissue than in normal tissue (Figure 9).

Discussion

In 2000, immune checkpoint inhibitors were used to enhance the anti-tumor ability of the immune system and achieved better curative effect. In 2011, the world's first immune checkpoint inhibitor, cytotoxic T lymphocyte-associated antigen 4 (CTLA-4) antibody, was approved by the U.S. Food and Drug Administration, marking a new era in tumor immunotherapy. Checkpoint inhibitors and programmed cell death receptor-1 (PD-1) antibodies increase the 5-year survival rate of relapsed/refractory non-small cell lung cancer from 5% to 16% [24]. tumor immunotherapy is further promoted to the status of a research

hotspot, such that immunotherapy has become the main treatment for some tumors. Biomarkers closely related to immunotherapy include PD-L1 IC (IC immune cell), PD-L1 TC (tumor cell), CD8+ T cells, tumor mutational burden, and others. Their expression levels and roles differ in various stages of particular tumors. It has been reported that immunotherapy is more effective in the context of T cell proliferation [25]. For this reason, we intended to identify the most relevant biological processes and co-expression networks in the tumor microenvironment with CD8+ T lymphocyte infiltration, and to determine whether these biological processes and related factors would improve outcomes by promoting CD8+ T lymphocyte infiltration.

We obtained 1,097 breast cancer samples through TCGA-BRCA, and constructed a CD8+t-related co-expression network by estimating the proportion of CD8+ T lymphocytes in each sample. WGCNA identifies gene modules with similar expression patterns. By calculating the correlation between multiple modules and CD8 content, the most relevant biological functions and co-expression networks of CD8+ T lymphocytes can be screened out. We determined that the 25 factors in the yellow module were most closely related to the content of CD8+ T lymphocytes, and were most related to the antigen presentation process and interferon responses. Studies showed that the drug resistance of PD-1 or PD-L1 monoclonal antibody after immunotherapy was closely related to tumor antigen mutation and antigen presentation process [26]. In addition, the tumor interferon signaling pathway was related to the multi-gene resistance program that blocks immune checkpoints [27].

Twenty-five factors can be used as independent prognostic protective factors in breast cancer patients. This conclusion is consistent with the trend of the influence of CD8+ T lymphocytes on outcomes. Then, we constructed a Cox regression model consisting of four factors, based on the 25 independent prognostic factors. CD74, GIMAP4, HCST and HLA-DMA not only showed good clinical phenotype correlation, but also correlated with M2 type macrophages, tumor purity, and immune score.

CD74 encodes a protein related to major histocompatibility complex (MHC) class II that regulates immune response antigen presentation. As the cell surface receptor of the cytokine macrophage migration inhibitory factor, CD74 initiates survival pathways and cell proliferation when combined with the encoded protein [28]. Basha et al. demonstrated the CD74-dependent MHC class I cross-presentation pathway in dendritic cells, which is thought to be related to the response of MHC class I restricted cytolytic T lymphocytes (CTL) to cell-related antigens [29]. GIMAP4, also known as GTPase, encodes proteins belonging to the Immune-related nucleotide (IAN) subfamily of GTP-binding superfamily and nucleotide-binding proteins. Mégarbané et al. found that the lack of GIMAPs may play a tumor-suppressive role against breast cancer [30]. The HCST encodes a transmembrane signaling adapter that encodes a protein that forms part of the immune recognition receptor complex with type C lectin-like receptor NKG2D [31]. Gilfillan et al. found, that in dap10-deficient mice, CD8+ T cells lack NKG2D expression and cannot mediate tumor-specific responses [32]. HLA-DMA is also called MHC class II, DM alpha [33]. Both HLA-DMA and -DMB genes are needed for the formation of MHC class II/peptide complexes in antigen presenting cells [34].

To determine whether co-expression of CD8+ T lymphocytes improve the inference of immunotherapy, we attempted to identify the cohort with immunotherapy outcomes. Unfortunately, we did not find such breast cancer results. We only found that the four factors in a follow-up cohort of immunotherapy for melanoma, and found that GIMAP4 can be used as an independent prognostic factor after immunotherapy (Figure 10).

This article has certain limitations. First, only samples from two cohorts were included, and joint analysis is still needed for more cohorts. In addition, this article only explains the co-expression network that promotes the infiltration of CD8+ T lymphocytes from the perspective of computational biology. More in-depth cell labeling experiments need to be performed. In the end, we did not obtain enough immunotherapy follow-up samples, and statistical systematic errors are inevitable.

These four CD8+ T lymphocyte proportion co-expression genes increase CD8+ T lymphocyte in breast cancer by an antigen presentation process. The mechanism might provide new ideas to improve the curative effect in patients who do not benefit from immune therapy.

Declarations

Ethics approval and consent to participate

No ethics approval was required for this work. All utilized public data sets were generated by others who obtained ethical approval.

Consent for publication

Not applicable.

Availability of data and materials

The datasets TCGA-BRCA for this study can be found in the [The Cancer Genome Atlas] [<http://cancergenome.nih.gov/>]. The datasets GSE78220 in this study can be found in the [GEO] [<http://www.ncbi.nlm.nih.gov/geo/>].

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This article is supported by the Research on the molecular mechanism of MAC30 promoting breast cancer cell proliferation and invasion 2015020544

Author Contributions

Jialing Jiang conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Yutao Wang performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Yi Zhao conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Acknowledgments

We want to thank TCGA, GEO for free use.

References

1. Nanda R, Chow LQ, Dees EC, Berger R, Gupta S, Geva R, Pusztai L, Pathiraja K, Aktan G, Cheng JD, Karantza V, Buisseret L. Pembrolizumab in Patients With Advanced Triple-Negative Breast Cancer: Phase Ib KEYNOTE-012 Study. *J Clin Oncol*. 2016;34:2460-7.
2. Mittendorf EA, Philips AV, Meric-Bernstam F, Qiao N, Wu Y, Harrington S, Su X, Wang Y, Gonzalez-Angulo AM, Akcakanat A, Chawla A, Curran M, Hwu P, Sharma P, Litton JK, Molldrem JJ, Alatrash G. PD-L1 expression in triple-negative breast cancer. *Cancer Immunol Res*. 2014;2:361-70.
3. Zhang M, Sun H, Zhao S, Wang Y, Pu H, Wang Y, Zhang Q. Expression of PD-L1 and prognosis in breast cancer: a meta-analysis. *Oncotarget*. 2017;8:31347-54.
4. Wang C, Zhu H, Zhou Y, Mao F, Lin Y, Pan B, Zhang X, Xu Q, Huang X, Sun Q. Prognostic Value of PD-L1 in Breast Cancer: A Meta-Analysis. *Breast J*. 2017;23:436-43.
5. Mansfield AS, Aubry MC, Moser JC, Harrington SM, Dronca RS, Park SS, Dong H. Temporal and spatial discordance of programmed cell death-ligand 1 expression and lymphocyte tumor infiltration between paired primary lesions and brain metastases in lung cancer. *Ann Oncol*. 2016;27:1953-8.
6. Peng J, Hamanishi J, Matsumura N, Abiko K, Murat K, Baba T, Yamaguchi K, Horikawa N, Hosoe Y, Murphy SK, Konishi I, Mandai M. Chemotherapy Induces Programmed Cell Death-Ligand 1 Overexpression via the Nuclear Factor- κ B to Foster an Immunosuppressive Tumor Microenvironment in Ovarian Cancer. *Cancer Res*. 2015;75:5034-45.
7. Rosenberg SA, Packard BS, Aebersold PM, Solomon D, Topalian SL, Toy ST, Simon P, Lotze MT, Yang JC, Seipp CA. Use of tumor-infiltrating lymphocytes and interleukin-2 in the immunotherapy of patients with metastatic melanoma. A preliminary report. *N Engl J Med*. 1988;319:1676-80.
8. Loi S, Sirtaine N, Piette F, Salgado R, Viale G, Van Eenoo F, Rouas G, Francis P, Crown JP, Hitre E, de Azambuja E, Quinaux E, Di Leo A, Michiels S, Piccart MJ, Sotiriou C. Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98. *J Clin Oncol*. 2013;31:860-7.

9. Yu Y, Zeng D, Ou Q, Liu S, Li A, Chen Y, Lin D, Gao Q, Zhou H, Liao W, Yao H. Association of Survival and Immune-Related Biomarkers With Immunotherapy in Patients With Non-Small Cell Lung Cancer: A Meta-analysis and Individual Patient-Level Analysis. *JAMA Netw Open*. 2019;2:e196879.
10. Tumei PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJ, Robert L, Chmielowski B, Spasic M, Henry G, Ciobanu V, West AN, Carmona M, Kivork C, Seja E, Cherry G, Gutierrez AJ, Grogan TR, Mateus C, Tomasic G, Glaspy JA, Emerson RO, Robins H, Pierce RH, Elashoff DA, Robert C, Ribas A. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature*. 2014;515:568-71.
11. Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, Berent-Maoz B, Pang J, Chmielowski B, Cherry G, Seja E, Lomeli S, Kong X, Kelley MC, Sosman JA, Johnson DB, Ribas A, Lo RS. Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell*. 2016;165:35-44.
12. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol Biol*. 2018;1711:243-59.
13. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB, Verhaak RG. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612.
14. Li Y, Chen Z, Wu L, Tao W. Novel tumor mutation score versus tumor mutation burden in predicting survival after immunotherapy in pan-cancer patients from the MSK-IMPACT cohort. *Ann Transl Med*. 2020;8:446.
15. Yang Z, Wei S, Deng Y, Wang Z, Liu L. Clinical significance of tumour mutation burden in immunotherapy across multiple cancer types: an individual meta-analysis. *Jpn J Clin Oncol*. 2020;50:1023-31.
16. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
17. Jiang J, Sun X, Wu W, Li L, Wu H, Zhang L, Yu G, Li Y. Construction and application of a co-expression network in *Mycobacterium tuberculosis*. *Sci Rep*. 2016;6:28422.
18. Jiang J, Sun X, Wu W, Li L, Wu H, Zhang L, Yu G, Li Y. Corrigendum: Construction and application of a co-expression network in *Mycobacterium tuberculosis*. *Sci Rep*. 2017;7:40563.
19. Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC, Lempicki RA. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol*. 2007;8:R183.
20. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 2017;45:D353-353D361.
21. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25:25-9.

22. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102:15545-50.
23. Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhorji G, Benfeytas R, Arif M, Liu Z, Edfors F, Sanli K, von Feilitzen K, Oksvold P, Lundberg E, Hober S, Nilsson P, Mattsson J, Schwenk JM, Brunnström H, Glimelius B, Sjöblom T, Edqvist PH, Djureinovic D, Micke P, Lindskog C, Mardinoglu A, Ponten F. A pathology atlas of the human cancer transcriptome. *Science*. 2017;357.
24. Saka H, Nishio M, Hida T, Nakagawa K, Sakai H, Nogami N, Atagi S, Takahashi T, Horinouchi H, Takenoyama M, Katakami N, Tanaka H, Takeda K, Satouchi M, Isobe H, Maemondo M, Goto K, Hirashima T, Minato K, Yada N, Tamura T. Five-year follow-up results from phase II studies of nivolumab in Japanese patients with previously treated advanced non-small cell lung cancer: pooled analysis of the ONO-4538-05 and ONO-4538-06 studies. *Jpn J Clin Oncol*. 2020.
25. Kallies A, Zehn D, Utzschneider DT. Precursor exhausted T cells: key to successful immunotherapy. *Nat Rev Immunol*. 2020;20:128-36.
26. Wang Z, Wu X. Study and analysis of antitumor resistance mechanism of PD1/PD-L1 immune checkpoint blocker. *Cancer Med*. 2020.
27. Benci JL, Xu B, Qiu Y, Wu TJ, Dada H, Twyman-Saint Victor C, Cucolo L, Lee D, Pauken KE, Huang AC, Gangadhar TC, Amaravadi RK, Schuchter LM, Feldman MD, Ishwaran H, Vonderheide RH, Maity A, Wherry EJ, Minn AJ. Tumor Interferon Signaling Regulates a Multigenic Resistance Program to Immune Checkpoint Blockade. *Cell*. 2016;167:1540-54.e12.
28. Wang ZQ, Milne K, Webb JR, Watson PH. CD74 and intratumoral immune response in breast cancer. *Oncotarget*. 2017;8:12664-74.
29. Basha G, Omilusik K, Chavez-Steenbock A, Reinicke AT, Lack N, Choi KB, Jefferies WA. A CD74-dependent MHC class I endolysosomal cross-presentation pathway. *Nat Immunol*. 2012;13:237-45.
30. Mégarbané A, Piquemal D, Rebillat AS, Stora S, Pierrat F, Bruno R, Noguier F, Mircher C, Ravel A, Vilaire-Meunier M, Durand S, Lefranc G. Transcriptomic study in women with trisomy 21 identifies a possible role of the GTPases of the immunity-associated proteins (GIMAP) in the protection of breast cancer. *Sci Rep*. 2020;10:9447.
31. Quatrini L, Molfetta R, Zitti B, Peruzzi G, Fionda C, Capuano C, Galandrini R, Cippitelli M, Santoni A, Paolini R. Ubiquitin-dependent endocytosis of NKG2D-DAP10 receptor complexes activates signaling and functions in human NK cells. *Sci Signal*. 2015;8:ra108.
32. Gilfillan S, Ho EL, Cella M, Yokoyama WM, Colonna M. NKG2D recruits two distinct adapters to trigger NK cell activation and costimulation. *Nat Immunol*. 2002;3:1150-5.
33. Álvaro-Benito M, Wieczorek M, Sticht J, Kipar C, Freund C. HLA-DMA polymorphisms differentially affect MHC class II peptide loading. *J Immunol*. 2015;194:803-16.
34. Fling SP, Arp B, Pious D. HLA-DMA and -DMB genes are both required for MHC class II/peptide complex formation in antigen-presenting cells. *Nature*. 1994;368:554-8.

Tables

Table 1. The Module and Gene significance for CD8 ⁺ T Cells related genes			
ID	moduleColor	GS.TcellsCD8	p.GS.TcellsCD8
CD8A	yellow	0.637280634	3.56E-77
GZMA	yellow	0.615288086	1.27E-70
NKG7	yellow	0.596372475	2.19E-65
CST7	yellow	0.543848238	1.57E-52
CD2	yellow	0.523524185	3.85E-48
CD3D	yellow	0.505128651	2.05E-44
GZMK	yellow	0.493345337	3.81E-42
GZMB	yellow	0.489263621	2.22E-41
CD3E	yellow	0.485176437	1.27E-40
CCL5	yellow	0.484214853	1.91E-40
CXCL9	yellow	0.466110209	3.17E-37
IRF1	yellow	0.456695053	1.26E-35
HLA-E	yellow	0.449975369	1.64E-34
HLA-DPB1	yellow	0.447710011	3.83E-34
CD74	yellow	0.446926288	5.13E-34
IL2RG	yellow	0.443336997	1.94E-33
PSMB9	yellow	0.429897672	2.47E-31
CD27	yellow	0.426640739	7.73E-31
GBP4	yellow	0.422872734	2.85E-30
CORO1A	yellow	0.42090121	5.61E-30
PSMB10	yellow	0.419193325	1.00E-29
GIMAP4	yellow	0.416341962	2.63E-29
CD48	yellow	0.416335453	2.64E-29
HLA-DMA	yellow	0.4146621	4.63E-29
SELPLG	yellow	0.414330462	5.17E-29
TNFRSF1B	yellow	0.413004893	8.05E-29
HCST	yellow	0.406386032	7.12E-28

GMFG	yellow	0.403523457	1.80E-27
GS: Gene significance			

Figures

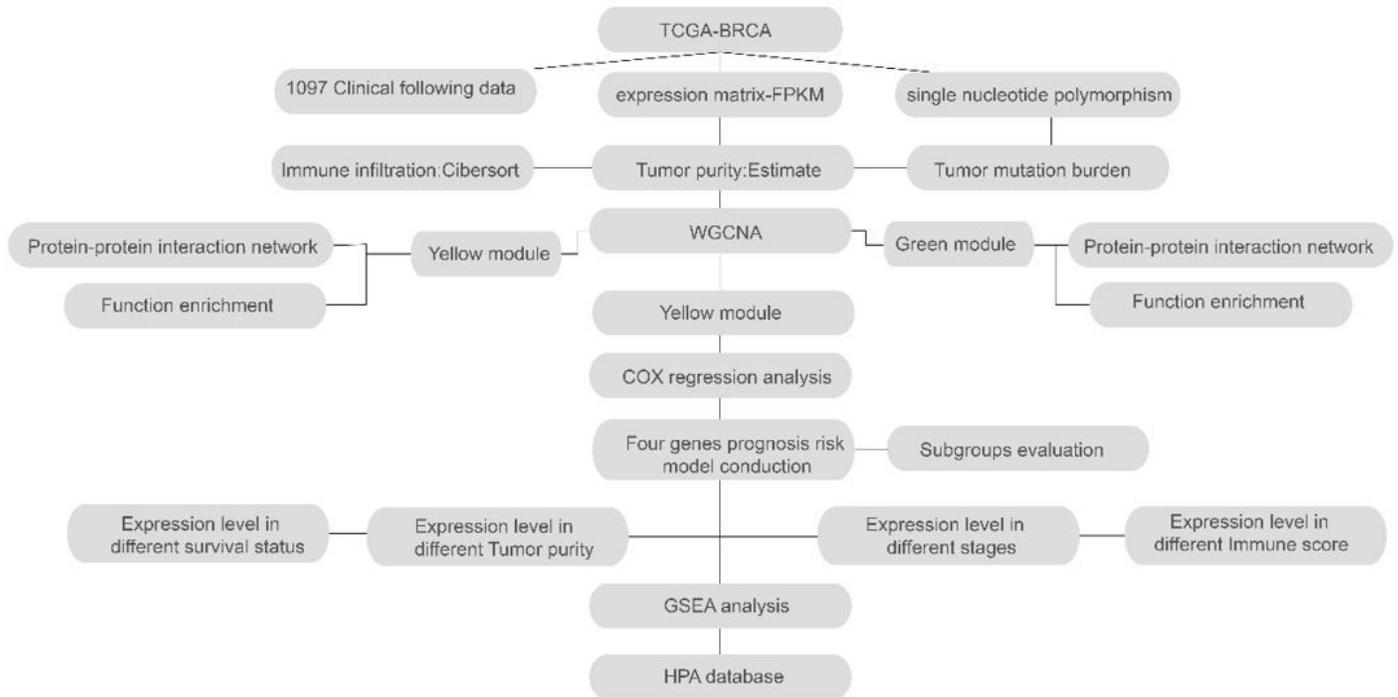


Figure 1

The flow chart of experimental sequence.

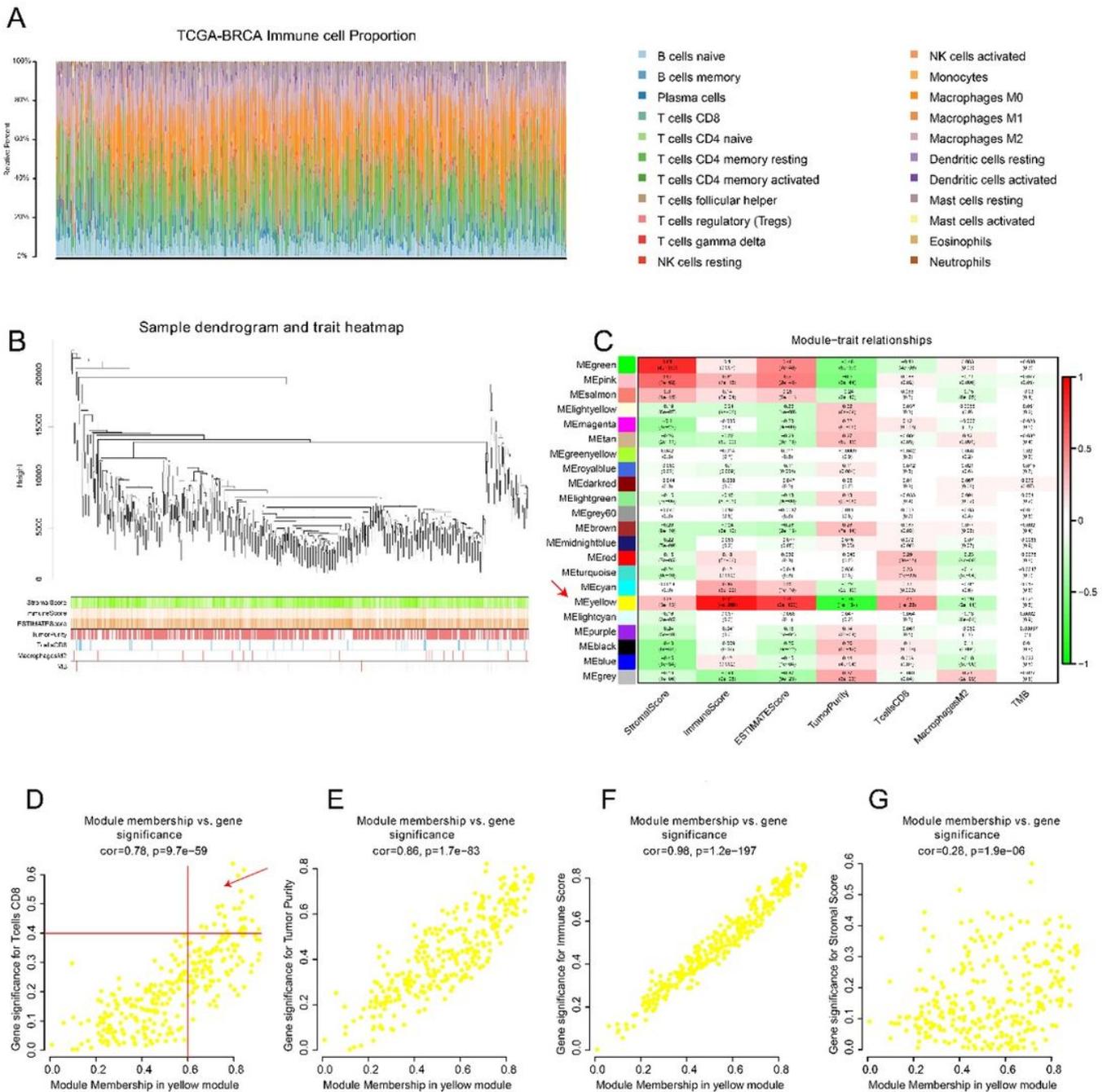


Figure 2

(A) We evaluated 860 breast cancer samples accurately using CD8+ T lymphocytes. (B) Hierarchical clustering tree was built using the dynamic hybrid cutting method. (C) Twenty-two co-expression models were identified. the yellow module had the strongest correlation with CD8+ T lymphocyte proportion in the TCGA - BRCA cohort (Cor = -0.41; P = 1e-28). (D) The yellow module showed significantly correlation to CD8+ T Cell (Cor = 0.78, p = 9.7e-59), (E) tumor purity (Cor = 0.86, p = 1.7e-83), (F) immune score (Cor= 0.98, p = 1.2e-197) and (G) stomal score(Cor = 0.28, p = 1.9e-06).

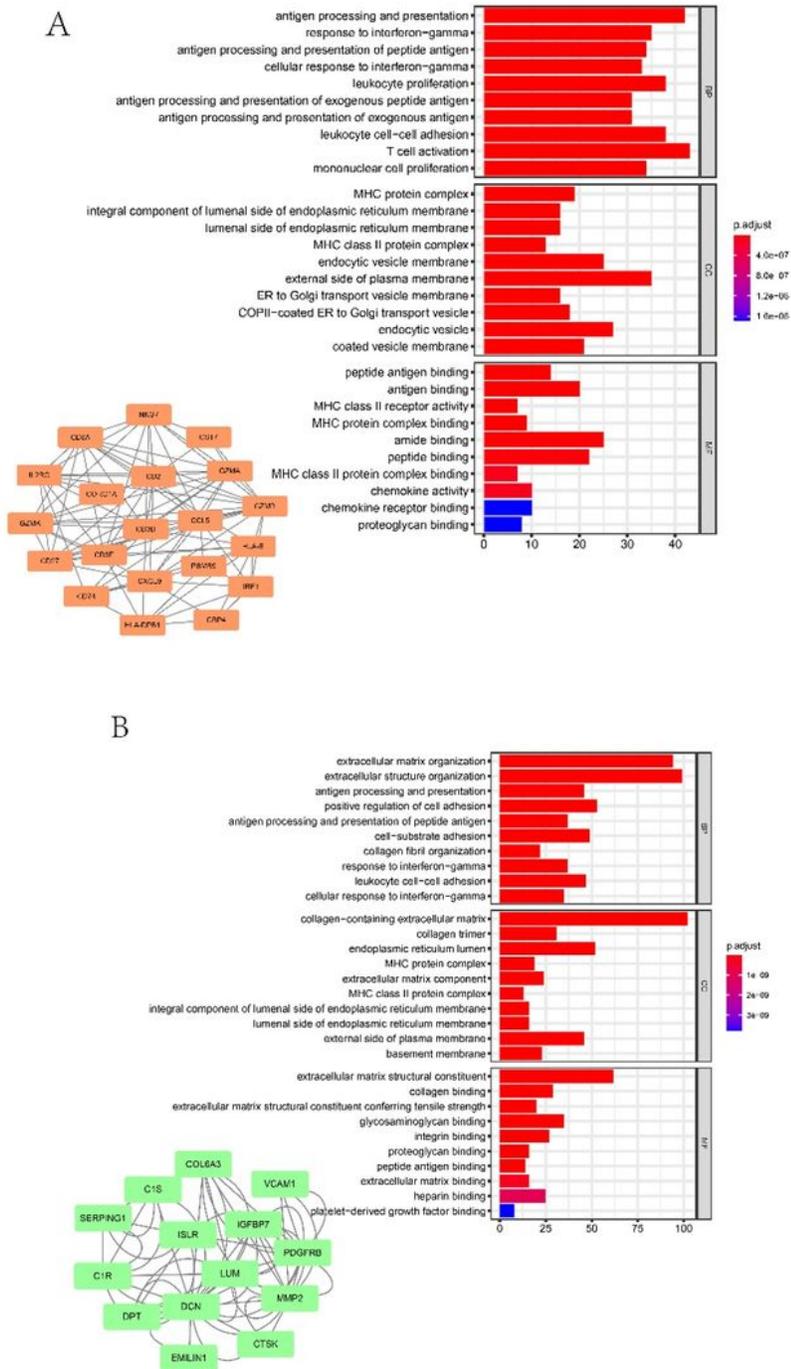


Figure 3

(A) The 28 CD8+ T lymphocyte proportion positively co-expressing mRNA were most significantly enriched in the antigen processing and presentation and response to interferon-gamma, which suggested these biological processes might promote CD8+ T lymphocyte infiltration in breast cancer microenvironment. (B) The CD8+ T lymphocyte negatively co-expressing module was most significantly enriched in extracellular matrix organization.

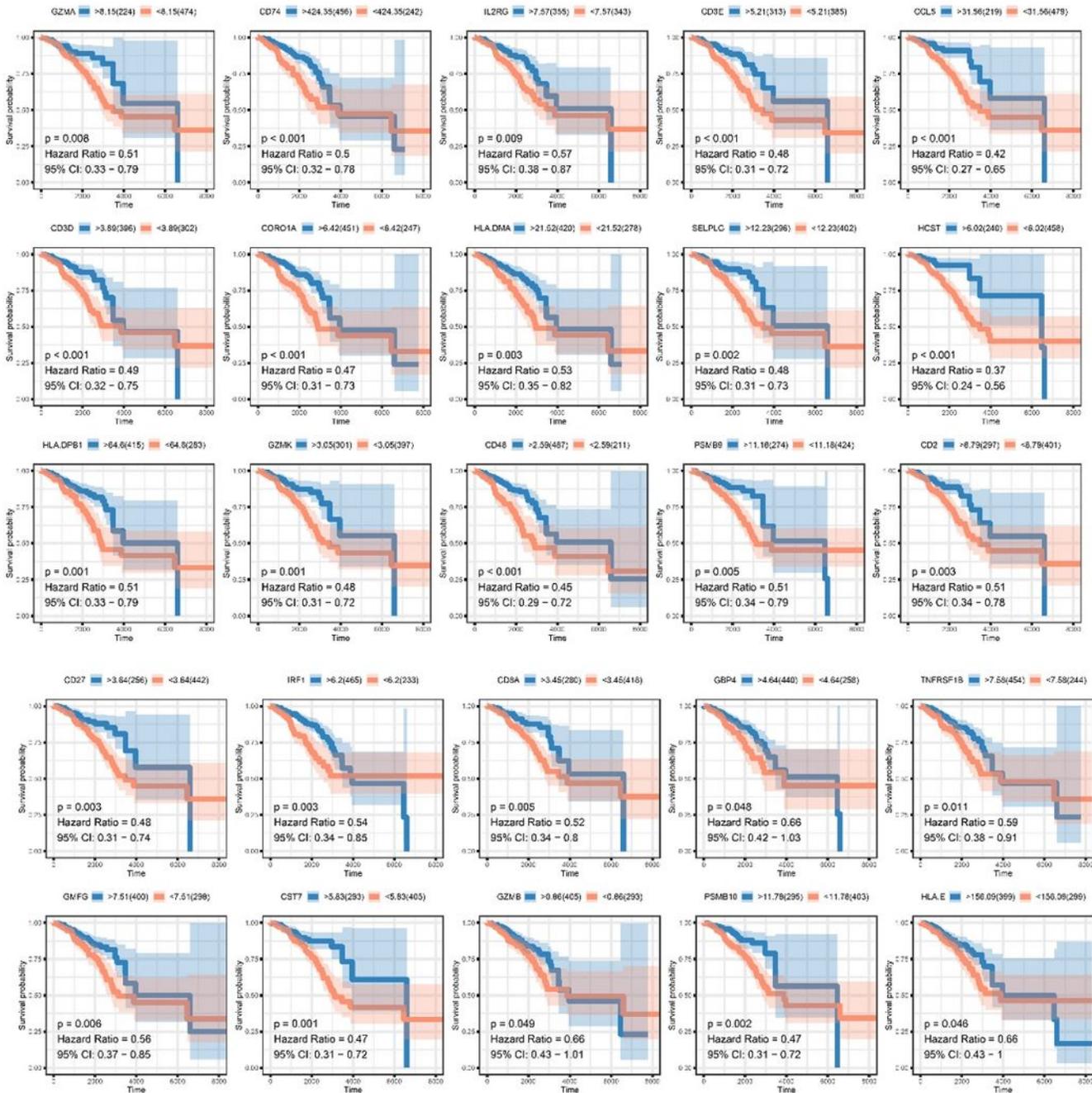


Figure 4

The patients in low expression groups for GZMA (TCGA: P = P < 0.001), CD74 (TCGA: P < 0.001), IL2RG (TCGA: P = 0.009), CD3E (TCGA: P < 0.001), CCL5 (TCGA: P < 0.001), CD3D (TCGA: P < 0.001), CORO1A (TCGA: P < 0.001), HLA-DMA (TCGA: P = 0.003), SELPLG (TCGA: P = 0.002), HCST (TCGA: P < 0.001), HLA-DPB (TCGA: P = 0.001), GZMK (TCGA: P = 0.001), CD48 (TCGA: P < 0.001), PAMB9 (TCGA: P = 0.005), CD2 (TCGA: P = 0.003), CD27 (TCGA: P = 0.003), IRF1 (TCGA: P = 0.003), CD8A (TCGA: P = 0.005),

GBP4 (TCGA: P = 0.048), TNFRSF1B (TCGA: P = 0.011), GMFG (TCGA: P = 0.006), CST7 (TCGA: P = 0.001), GZMB (TCGA: P = 0.049), PSMB10 (TCGA: P = 0.002) and HLA-E (TCGA: P = 0.046) showed survival risk against high expression groups.

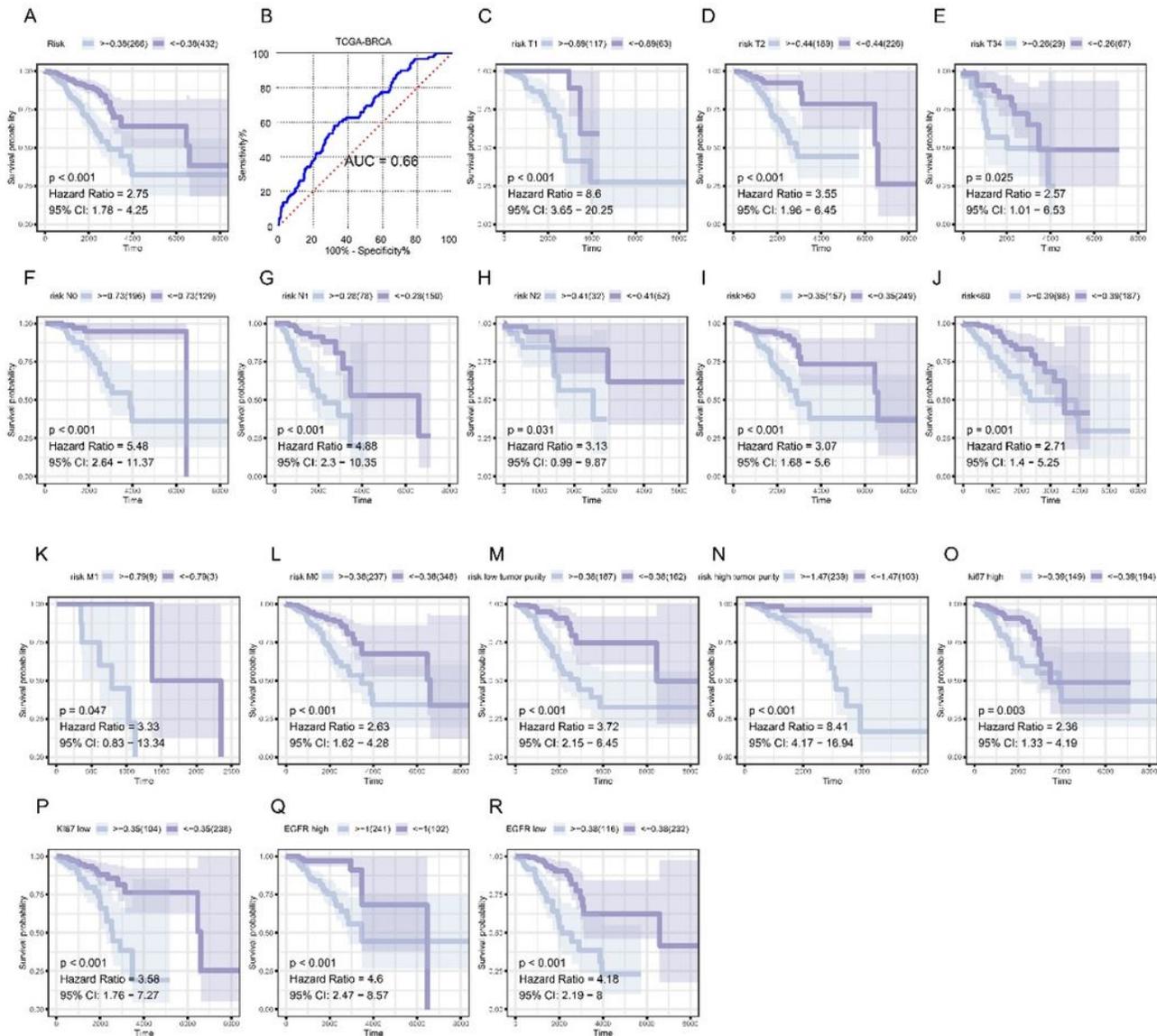


Figure 5

(A-B) The samples in high risk level samples for breast cancer patients (TCGA: P < 0.001; HR = 2.75) showed survival risk against low risk groups, with the area under curve (AUC) = 0.66. (C-R) The risk score was evaluated in various subgroups, including age, gender, stage, tumor purity, and tumor mutation burden, metastasis status, Ki-67 and EGFR. The results were the same significant in these subgroups.

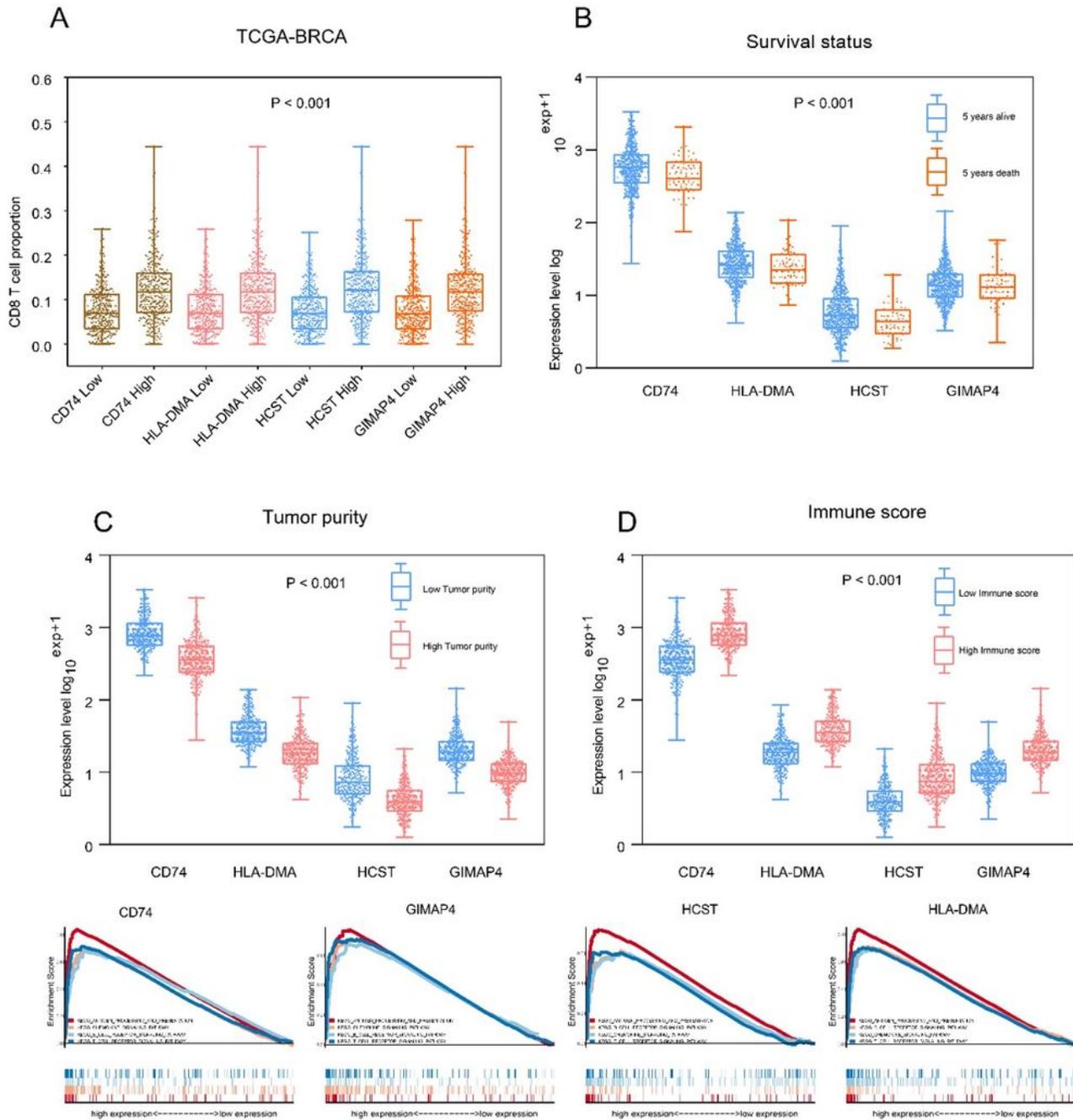


Figure 6

(A) The CD8+ T cell proportion level in different gene expression pattern. (B) The gene expression level in different survival status. (C) The gene expression level in different tumor purity. (D) The gene expression level in different immune score.

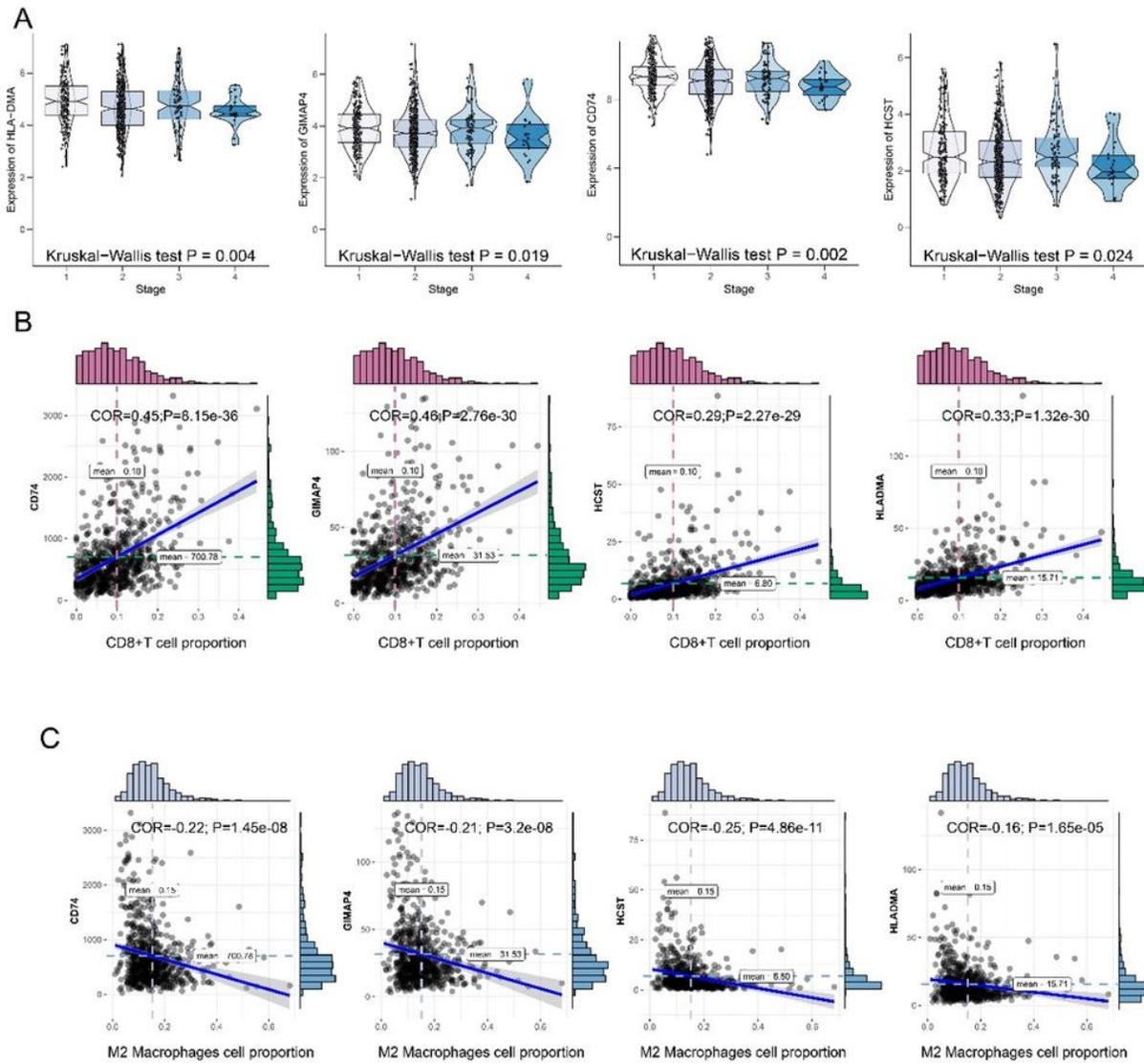


Figure 7

The correlation between CD8+ T cell co expression genes to clinical stages (A), CD8+ T lymphocytes (Figure B) and M2 macrophages (C).

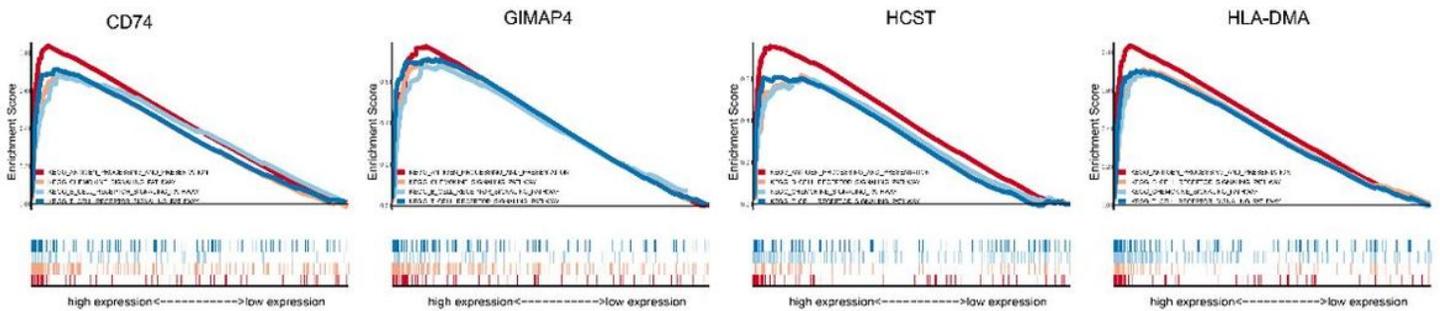


Figure 8

Antigen processing and presentation, the chemokine signaling pathway, B cell receptor signaling pathway, and the T cell receptor signaling pathway were related to the high expression group in CD74, GIMAP4, HCST and HLA-DMA.

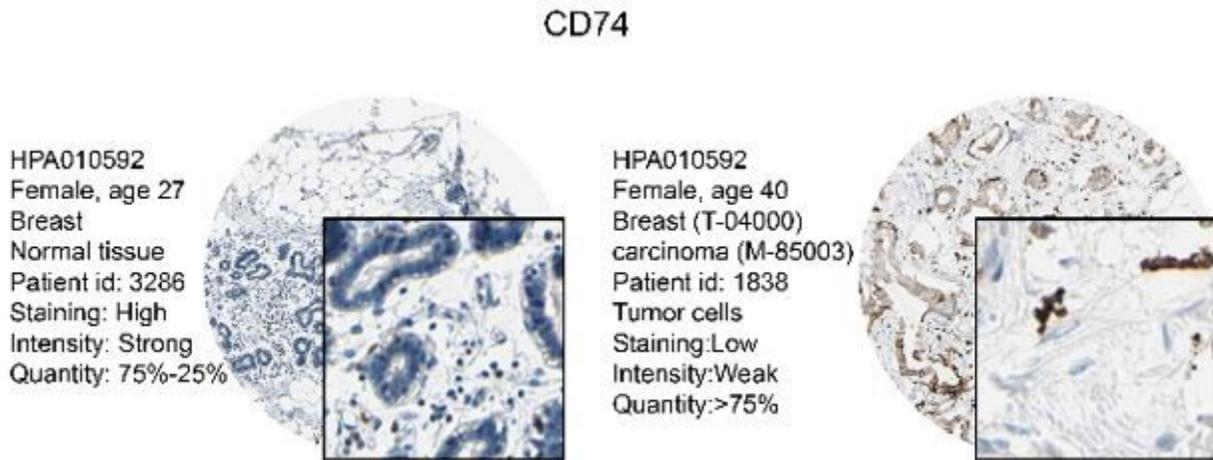


Figure 9

We compared the various expression levels of these genes between normal and tumor tissues. HPA010592 was the antibody of CD74, which showed higher intensity in the tumor tissue against normal tissue.

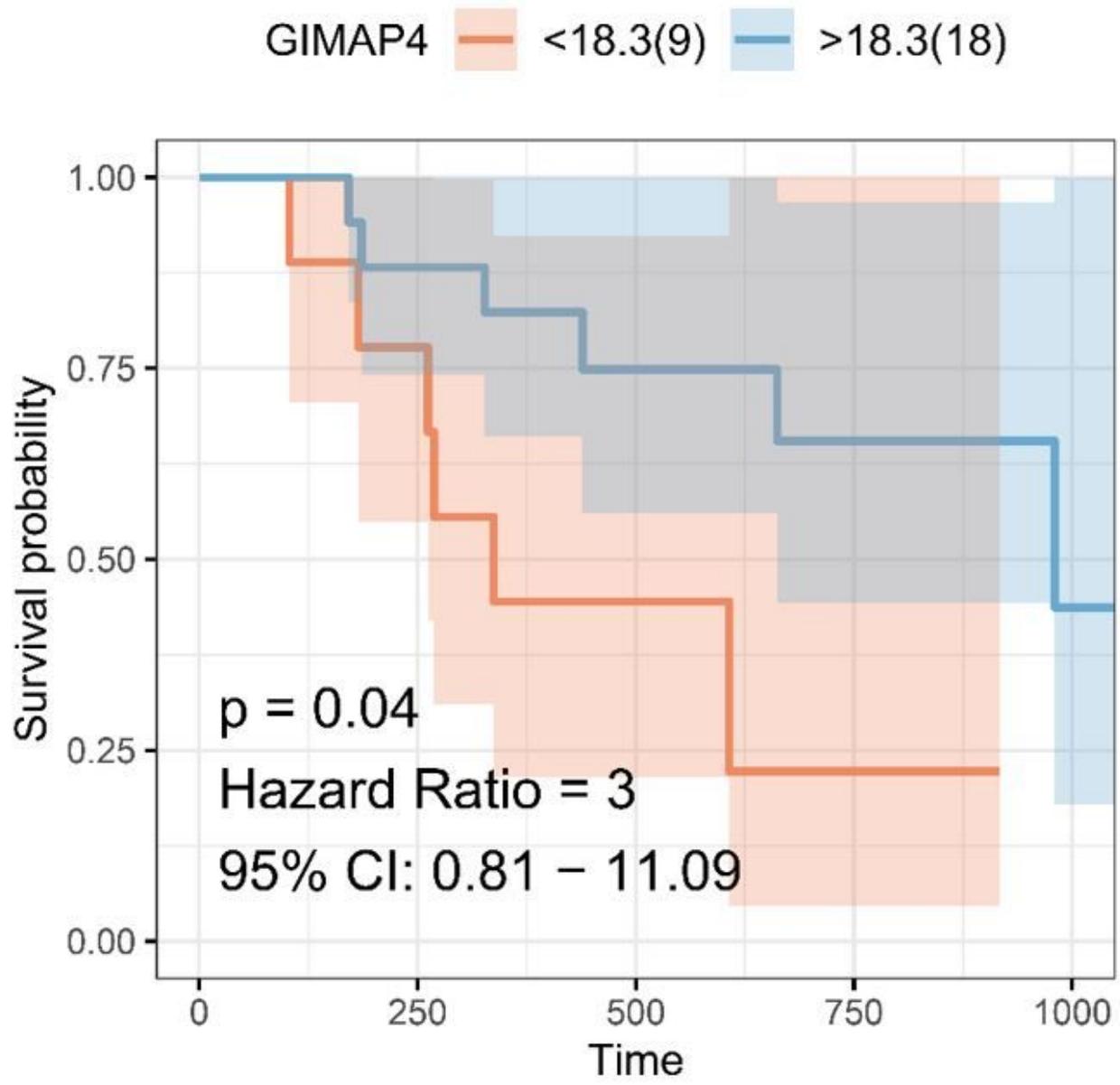


Figure 10

GIMAP4 can be used as an independent prognostic factor after immunotherapy.