

Comprehensive multi-omics analysis identifies an immune related predicting model in early stage breast cancer

Xuan Li (✉ 772032017@qq.com)

Fudan University Shanghai Cancer Center <https://orcid.org/0000-0003-2851-8963>

Wei Jin

Fudan University Shanghai Cancer Center

Primary research

Keywords: breast cancer, immune infiltration, macrophages, prognosis, gene expression

Posted Date: March 30th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-19691/v1>

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

[Read Full License](#)

Abstract

Background

Breast cancer is known the highest incidence cancer in women. Tumor-infiltrating immune cells were reported closely related to cancers' fate but it's function still not very clear in breast cancer. The aim of our study is to build a infiltrating immune cells based nomogram model to better predict patients survival and explore its relationship with immune features.

Methods

We first use CIBERSORT to analyze 22 immune cells' status in two unrelated breast cancer cohorts (TCGA and METABRIC). The univariate and multivariate Cox analyses were used to analyze the prognostic ability of immune cells and other clinicopathological factors. Nomogram model were built by using independent prognostic factors. Different immune signatures and gene enrichment analysis were performed in different nomogram risk groups.

Results

Multivariate cox analysis showed that Macrophages M2 with HR of 1.733 (95% CI: 1.013-2.966) in TCGA cohort and 1.334 (95% CI: 1.125-1.581) in METABRIC cohort is the only independent prognostic factor among the 22 immune cells in early stage breast cancer. Macrophages M2, age, TNM stage and molecular types were used to build the nomogram model. The AUC of the ROC of nomogram reached 0.732 in TCGA cohort and 0.702 in METABRIC cohort. Using the nomogram score can better classified patients to low and high risk group. High risk group showed higher malignant signatures and predicted immunotherapy response rates than low risk group which consistent with the gene entrenchment analysis.

Conclusion

In this study, we revealed that M2 macrophages could predict the OS of breast cancer patients. Based on M2 and other clinical features we established a nomogram model which were significantly different in immune features that can better assess the OS risk or be a predictor for the immunotherapy response in breast cancer for further research.

Background

Breast cancer has the highest incidence rate and second highest mortality rate among female in US[1, 2]. The overall 5-year survival rate are better for Stage I and II with 98% and 92% respectively. While the survival rates are decreased to 75% in stage III and 27% for stage IV. The treatment pattern choice were different in different TNM stage patients as in stage I and II breast-conserving surgery plus radiotherapy account for the highest proportion, while in stage III and IV chemotherapy plus radiotherapy were important in America patients[2]. Breast cancer's outcome improved in recent years contributed to the

advents of more and more treatment especially the chemotherapy. In most cases, immunotherapy included in the chemotherapy, but their effect still controversy in breast cancer[3, 4]. Found the immune condition of the cancer and better classify the breast cancer patients into different risk stages can not only help manage patients better but also help found some patients that may benefit most from immunotherapies.

In breast cancer, tumor-infiltrating lymphocytes (TILs) are the most studied immune cells and it has been shown to be potentially predict the prognosis in specific subtypes of breast cancer especially in HER2 positive and triple negative patients. Large retrospective studies have shown that higher levels of TILs can increase patients' response to neoadjuvant treatment and benefits more from adjuvant treatment. What's more, higher TILs in breast cancer tissues are associated with higher overall survival (OS) and fewer recurrences rates regardless of therapy[5–8]. Low quantity of TILs in breast cancers were identified as a reason for little efficacy in treatments against PD-1 and PD-L1[9]. However, higher TILs do not always predict good response to treatment or survival, that may relate to the various other immune cells in tumor microenvironment (TME). Besides TILs, tumor-associated macrophages (TAMs) were also ubiquity in cancer. In breast cancer patients, the TAMs have been found associated with different outcomes of treatment response or survival mostly because TAMs contained three subtypes that are M0, M1 and M2, those different subtypes have different inflammation and immunologic functions. The interconversion of them can result in different fate in breast cancer[10–12]. Other immune cells like NK cells, mast cells and several immune genes have also been found related to breast cancer outcomes[13–16]. As more and more immunotherapy being invented, there is growing interest in developing the immunotherapy with other therapies such as molecularly targeted therapies in breast cancer[9]. Therefore, better estimate the status of immune cells in the TME and probe the distribution and function of various immune cells are important in cancers that may improve the efficacy of immunotherapy or found new treatment methods.

CIBERSORT, the abbreviation of “Cell type Identification By Estimating Relative Subsets Of RNA Transcripts”, is a gene-base algorithm using the method of deconvolution to transform the signature matrix of 547 marker genes into 22 infiltration immune cell types' fraction in tumor tissues that has been successfully validated in some cancers like breast, lung and liver cancers[17–20]. The Cancer Genome Atlas (TCGA) database is an America project and METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) database is a Canada-UK project[21, 22]. Both two databases contents large amount of transcriptome data and correspondent clinicopathologic information of breast cancer and make researchers more convenient in implementing researches. By using the information from TCGA and METABRIC database, the CIBERSORT can count the infiltration immune cells of each breast cancer patients. The study of us aimed to explore the infiltrating immune cells status in early stage breast cancer and their relationship with clinical information and survival outcomes. A nomogram model will be built by using immune cells and clinical factors to better distinguish patients into good and bad risk groups. Furthermore, we also want to explore the relationship of nomogram risk and immune features and found some hub genes and possible network that may related with patients' risk and immunotherapy response.

Methods

Data acquisition

The RNA-seq data of TCGA breast cancer and normal control were downloaded from the TCGA data portal (<https://tcga-data.nci.nih.gov/tcga/>), and were normalized by the “limma” R package. The corresponding survival and clinicopathological information of TCGA cohort were also downloaded from the TCGA data portal. Normalized gene expression data, clinical data and overall survival data of breast cancer patients in METABRIC cohort were downloaded from cBioPortal website (<https://www.cbioportal.org/>). We excluded patients in both TCGA and METABRIC cohorts with primary metastasis and those with information such as estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2(HER2), lymph node status and tumor size missing and the patients with overall survival (OS) follow-up time less than 30 days were also excluded in the clinical associated analysis.

Tumor infiltration immune cells (TIICs) count

The subpopulations of TIICs were identified by CIBERSORT algorithm. With the help of gene matrix build by 547 marker genes, the normalized gene expression data can be figured out to the fractions of 22 infiltrated immune cells which included seven types of T cells, three types of macrophages, two types of NK cells, B cells, dendritic cells, mast cells and one type of plasma cells, monocytes, eosinophils and neutrophils. Samples with CIBERSORT p-value < 0.05 were considered with high accuracy and selected for further study. Patients were classified as either low or high level of TIICs by using the median of their TIICs proportion. The six major classification of TIICs as Lymphocytes, Macrophages, Dendritic cells, Mast cells, Neutrophils and Eosinophils were according to the instruction of Thorsson[23].

Nomogram model built and immune signatures

By using the independent risk factors that identified by multivariate Cox regression analysis, we construct a nomogram for the prediction of the OS. The prognostic value of the nomogram in determining OS were assessed based on the area under the curve (AUC) values of receiver operating characteristic (ROC) curves. The calibration plots and the concordance index (C-index) were used to evaluate the actual outcomes and the probabilities of the model (rms R package in R software). The best cut-off value of nomogram was selected by the highest Youden’s index in ROC and stratified patients into nomogram low and high risk groups.

The immune signatures such as SNV Neoantigens, Aneuploidy score BCR Shannon and TCR Shannon of each TCGA breast cancer patients got from the study of Thorsson[23]. To explore the relationship between the nomogram risk groups and the immunotherapeutic response, a web named Tumor Immune Dysfunction and Exclusion (TIDE) (<http://tide.dfci.harvard.edu>) was used to predict the response of each patients to anti-PD1 or anti-CTLA4 immunotherapeutic based on the transcriptome information of the TCGA cohort[24].

Identification of hub genes and transcription factor-immune gene networks

The selection of different expression genes (DEGs) between nomogram high risk group and low risk group patients in TCGA cohort were used edgeR package in R software. The cutoff values of false discovery rate (FDR) < 0.05 and a \log_2 |fold change| > 1 were used. Differentially expressed genes were then use Online analysis tool STRING (<http://string-db.org>) and Cytoscape software to found the hub genes. By the help of package CytoHubba in Cytoscape, five different algorithm (Maximal Clique Centrality (MCC), Maximum Neighborhood Component (MNC), Edge Percolated Component (EPC), Density of Maximum Neighborhood Component (DMNC) and Degree) were used and the top20 important hub genes in each algorithm were selected. The overlapped hub genes from these 5 algorithm methods were showed by Venn diagram. The immune genes and transcription factor genes were got from the web IMMPORT(<https://www.immport.org/home>) and Cistrome Project (<http://www.cistrome.org/>) respectively and the network were built by Cytoscape. The gene ontology (GO) of Biological processes enrichment analysis were used the analysis function of STRING.

Statistical analyses

Overall survival (OS) was defined as the time of pathologically confirmed disease until death for any reason or the last follow-up. T test or one-way ANOVA was used to compare the immune signatures in different nomogram risk groups. The analyses were performed in SPSS 25.0 or R software 3.6.1, and two-tailed p values < 0.05 were considered significant.

Results

Survival prognosis analyze of TIICs and clinicopathology factors

We first investigated the fractions of tumor infiltrated immune cells (TIICs) between tumor and non-tumorous tissues in TCGA cohort and METABRIC cohort. After the exclusion of clinicopathology data missing cases, a total of 602 patients diagnosed with breast cancer were registered in the TCGA cohort as well as 1082 breast cancer patients from METABRIC cohort were included in our study for analysis. The clinical information of both two cohorts are listed in Table 1. To assess the predictive value of TIICs in breast cancer, we examined patient overall survival according to different infiltrated immune cells in TCGA and METABRIC cohort. In TCGA cohort, as shown in Table 2, univariate Cox analysis showed that the Macrophages M2 and Monocyte were act as risk factors with p value lower than 0.05, while other factors showed no significant difference in univariate Cox analysis (Macrophages M2(HR 1.771, 95% CI: 1.155–2.717); Monocyte (HR 1.602, 95% CI: 1.038–2.472)). Kaplan-Meier survival analyses showed that high fraction of Macrophages M2 and Monocyte cells were significantly associated with worse OS in breast cancer (Fig. 1A and B). Univariate Cox analysis also showed that age, TNM type and molecular type of patients were associated with OS (Table 2). The statistical difference prognostic factors for OS

identified by univariate Cox analysis were further analyzed by multivariate Cox analysis. The results indicated that the Macrophages M2 was an independent protective factor in TCGA cohort, with a hazard ratio (HR) of 1.733 (95% CI: 1.013–2.966) as well as the age, TNM type and molecular types (Table 2). In METABRIC cohort, univariate Cox analysis showed that Macrophages M2 (HR = 1.387, 95% CI: 1.181–1.629), T cells CD4 memory activated (HR = 0.820, 95% CI: 0.699–0.963), T cells regulatory Tregs (HR = 1.245, 95% CI: 1.060–1.463), T cells gamma delta (HR = 0.751, 95% CI: 0.640–0.882) and NK cells resting (HR = 0.790, 95% CI: 0.673–0.928) were act as risk factors for OS (Table 2). Kaplan-Meier survival analysis results showed in Fig. 1C-G. In multivariate Cox analysis, only Macrophages M2 (HR = 1.334, 95% CI: 1.125–1.581) along with age, TNM type and molecular type were independent factors for prognosis (Table 2).

Construction of nomogram model based on Macrophages M2

As Macrophages M2 performed as an independent prognostic factor in breast cancer, we explored its potential to improve the prognostic accuracy with other clinicopathological prognostic factors by building a nomogram predictive model using the TCGA cohort and verified by METABRIC cohort (Fig. 2A). After incorporating with Macrophages M2, the nomogram's accuracy to predict OS of breast cancer patients were improved as the AUC of the ROC curve reached to 0.732 (Fig. 2B, table 3) and the C-index reached to 0.751 (Table 3). The calibration plot indicated that the nomogram showed good performance for OS (Fig. 2C). By using the ROC curve and best Youden index, we found the best cut-off of the risk is 112.5. Patients with total points lower than 112.5 were included in low risk group and those with higher than 112.5 were included in high risk group. The METABRIC cohort were used as validation cohort and the accuracy to predict OS of breast cancer patients were also improved by using the nomogram model. The AUC of ROC curve was 0.702 (Fig. 2D, Table 3) and the C-index was 0.715 (Table 3) and the calibration plot also demonstrate good consistency of prediction with reality (Fig. 2E). The low and high risk group showed significant difference survival in breast cancer patients (Fig. 3A (TCGA), Fig. 3B (METABRIC)), and it can further distinguish each TNM stage patients into a low or high risk level which were proved in both TCGA (Fig. 3(C, E and G)) and METABRIC cohort (Fig. 3(D, F and H)).

Nomogram model shows significant association with immune features

We compared the TCGA nomogram low and high risk groups according to immune subtypes and other dominant sample characteristics. The proportion of Lymphocytes were higher in low risk group while the proportion of Macrophages were higher in high risk group with p value less than 0.05. Other immune subtype showed no difference between the low and high risk groups (Fig. 4A). High risk group showed higher tumor mutation burden (TMB) (Fig. 4B, $p = 0.016$) and SNV Neoantigens (Fig. 4C, $p = 0.023$). The aneuploidy score which is an indicator for DNA damage were significantly higher in high risk group (Fig. 4D, $p < 0.0001$). T cell receptor (TCR) and B cell receptor (BCR) analysis showed that low risk group

had higher TCR Shannon score than high risk group while the BCR Shannon score showed no difference between high and low risk groups (BCR Shannon: Fig. 4E, $P = 0.486$; TCR Shannon: Fig. 4F, $p = 0.024$). Because the low and high risk group were highly associated with the immune signatures, we then tested whether they could predict the immunotherapeutic response. By using the TIDE web program, we found that patients in high risk group had a lower TIDE score (Fig. 4G, $p = 0.001$) and higher predict clinical response to immune checkpoint blockade than low risk group patients (Fig. 4H, $p = 0.002$).

Identification of the hub genes and transcription factor-immune gene networks

To explore the differential expressed genes (DEGs) within nomogram high and low risk groups in breast cancer, we examined the transcriptional microarray in the 602 patients from TCGA cohort. Based on risk level, 310 DEGs were found with 192 genes up-regulated and 118 genes down-regulated in the nomogram high risk group than the low risk group (Fig. 5A). We then choice the top20 hub genes by using five methods in Cytohubba module by using the 310 DEGs (Supplementary Table 1) and after overlapping analysis, twenty genes were existed in at least 3 methods as shown in Venn picture in Fig. 5B. The expression of the 20 hub genes of each patients were shown in Fig. 5C and fourteen out of the twenty hub genes were immune related genes as framed in red line. We then analyzed the transcription factor and immune genes regulatory network in low and high risk group separately. The survival related transcription factors and immune genes with correlation p value less than 0.05 were selected to build the network. In low risk group, 3 transcription factors BCL11A, AFF4 and POLR3D were positive regulate the protective immune gene MAPK3 and negative regulate 3 hazard immune genes MC4R, IL23R and SFTPD (Fig. 5D). In high risk group, two transcription factors TAT and PPARD with statistical significance in survival were oppositely regulate some hazard immune genes such as CTLA4 and protective immune genes such as ITGAV respectively (Fig. 5E). Biological processes enrichment analysis with gene ontology (GO) in high risk group are showed in Fig. 5F. The most significant biological process was the immune response, and others such as response to external stimulus and cell population proliferation were also significantly enriched in high risk group.

Discussion

Though breast cancer was not thought to be immunogenic cancer types in previous compared with other cancer types, the role of specific immune cells in clinical and prognostic of breast cancer still in controversial. In the present research, we used two different breast cancer cohort TCGA and METABRIC with confirmed clinicopathology and RNA-seq expression data to analysis the immune cells expression pattern and their relationship with overall survival. The density of 22 main TIICs in breast cancer were evaluated by CIBERSORT algorithm and we found that though some immune cells shows prognostic ability in TCGA and METABRIC cohorts, only macrophage M2 were testified as an independent prognostic factor by multivariate Cox analysis in both two cohort. Based on the independent prognostic factors including M2 macrophages we build a nomogram model to further improve the prognostic ability in

breast cancer. By using the nomogram risk, we can best separate patients into high and low risk groups in all patients as well as in patients in different TNM stages. The nomogram low and high risk groups showed significant difference in some immune features and the predicted clinical response to immune checkpoint blockade were higher in high risk group patients.

The TME comprised by various cells and the predominant cells recruited to and activated in it are immune cells[25, 26]. Immune cells in TME are closely related to tumor processes such as growth, angiogenesis and metastasis[27–29]. The tumor progression may be effected by imbalanced host immune response, and inverse, the quantity and phenotype of immune cells can be also effected by tumor cells[30, 31]. Therefore, testing the immune status and found the possible way to change the infiltrated immune cells type in the TME may help us dope out strategies to improve the response rate of immunotherapy. Tumor infiltrating lymphocytes (TILs) account for most of immune cells in TME which included T cells, B cells and NK cells are highly heterogeneous in tumors and in some cases showed opposite functions and effects on survival[32]. In TCGA cohort, those TILs showed no significant difference in overall survival, while in METABRIC cohort the T cells CD4 memory activated, T cells regulatory, NK cells resting and another TILs T cells gamma delta showed prognostic ability. The T cells, which reported to produce some regulatory cytokines such as TGF- β and IL-10 and can be recruited to the cancer environment via CCL22/CCR4, were able to escape from immune surveillance, result in tumor immune tolerance and ultimately reduced tumor survival[33–35]. TILs play different functions in breast cancer and more research should be done in the future to help us take advantage of it.

Macrophages are also important in infiltration immune cells in cancer and it can be divided into M0, M1 and M2 subtypes. M0 subtype, which is inactivated and not have inflammatory or tumor-related function, can through different activation pathways polarized to two distinctly immunoregulatory imparity subtypes M1(pro-inflammatory) and M2(anti-inflammatory). M2 macrophages can secrete cytokines to inhibit inflammation and inhibit the proliferation and differentiation of T cells, it can also promote tumor proliferation and some cancer associated process such as angiogenesis[36–40]. TCGA cohort exhibit that the density fraction of macrophage M0 and M1 were higher in tumor tissues than normal tissues, while M2 were on the opposite. In breast cancer tissues the fraction of M2 were higher than M1, which means in breast cancer the macrophage M0 tend to differentiate towards M2 subtype. We found that only M2 macrophage neither M0 nor M1 macrophages is an independent risk factor to breast cancer patients. Many studies also report macrophages were associated with shorter survival in breast cancer patients and that may due to high M2 subtype polarization[11, 12, 41]. Based on the shift of macrophages, several approaches such as eliminate the macrophages, blockade the M2 subtype shift, reprogramming the macrophages shift to M1 subtype and inhibit the recruitment of monocyte into cancer may try to as possible anti-tumor therapies in cancer[42].

We build a nomogram model including the density level of macrophages M2 to better predict the OS in breast cancer. The patients were appropriate separated into high risk group and low risk group according to the nomogram score. We found that the low and high risk group were different in some immune signatures. The content of immune cell subtypes were different between low and high risk groups and as

Thorsson demonstrate that the tumor with macrophage dominated especially with higher M2 macrophage content and low lymphocytes infiltrate had poor prognosis[23]. DNA aneuploidy, as reported were related to cell proliferation and poor tumor differentiation[43]. In our study, the high risk group patients had higher Aneuploidy score, proliferation and worse prognosis which consistent with the results from a large meta-analysis that aneuploid breast cancer patients had poor DFS and OS[44]. The immunotherapy is an attempt in breast cancer patients when there is no other way out and the current immunotherapy response were relatively poor[3]. It is very important to seek out the potential molecular mechanism of immunotherapy responsiveness and those who are suitable to receive immunotherapy[45]. Tumor mutation burden (TMB) which was calculated as total count of variants divided by the whole length of exons is a novel biomarker for predicting immunotherapy effect. In small-cell lung cancer patients, high TMB means better treatment prognosis either in single or combined immunotherapy[46]. In breast cancer TMB were associated with prognostic immune subclasses and those patients with high TMB and favorable immune-infiltrate subtypes had better prognosis[47]. Neoantigens are a kind of peptides that are generated from somatic mutations and it's immunotherapy response were still contradict in tumors[48]. In order to release tumor-specific immune responses, the Neoantigens can result T cells recognizing cancer cells through T-cell receptors (TCR) interacted with major histocompatibility complex (MHC)[48, 49]. Our high risk group patients had higher Neoantigens but lower TCR score than low risk group patients so to improve specificity and sensitivity of neoantigen–MHC complexes may more helpful in those patients. The TIDE score that can indicate tumor immune escape and predict response to immune checkpoint blockade such as anti-PD1 and anti-CTLA4 is mainly used for melanoma and non-small cell lung cancer, but can still try to applied in other tumors[24]. The high risk group patients had lower TIDE score means they experience less immune evasion and may more likely to benefit from immune checkpoint blockade therapy. From the above, we found that though the high risk group patients had worse prognosis they may benefit more from immunotherapy.

Twenty gene selected by multiple methods were supposed to be hub genes between high risk and low risk groups. Fourteen out of the twenty genes were immune genes and among the 14 immune hub genes, CXCR2 is the most studied one in cancer and immunology. Knockdown CXCR2 in mammary tumor cells reduced cell invasion, metastasis and enhanced the toxicity of chemotherapies[50]. CXCR2 positive regulate neutrophil recruitment to cancer and can promote tumor metastasis through TNF α -activated mesenchymal stromal cells (MSCs)[51]. Macrophages also reported can increase expression of inflammatory chemokines and promoting tumor cell migration and invasion through activation of CXCR2[52]. Therefore, targeting CXCR2 may represent a novel therapeutic strategy for breast cancers. Other genes, such as vasoactive intestinal polypeptide (VIP) could inhibit the cytotoxicity of NK cells and could inhibit its expressions of some immune-associated genes like NKG2D and NF- κ B[53]. Neurotensin (NTS) belong to Neuropeptide G protein-coupled receptors (GPCRs) are over-expressed in numerous cancer cells and were reported to stimulate tumor growth[54]. Further study could try to focus on these hub genes to regulate tumor microenvironment and immune status, which may help to found some new therapeutic methods for breast cancer. transcription factors and immune genes network in low and high risk group showed significant difference. In high risk group, higher transcription factor TAT expression co-

upregulated with hazard immune genes and co-downregulated with protective immune genes while the transcription factor PPARG were in adverse. Methods that inhibit the gene expression of TAT or improve the gene expression of PPARG may change the immune gene expression and improve the high risk group patients survival as well. Further GO analysis of these survival related immune genes in high risk group showed that they were enriched in cell population proliferation and immune-relevant biological pathways, such as immune response and response to stimulus, this can partly explain the high risk group had relatively poor survival but may have better immune response than low risk group patients.

Though we have used two different cohort and large patients to get and verify our results, there are still some limitations in our study. The first is data missing, as the immune cells infiltration data were calculated by CIBERSORT algorithm and some patients were excluded because of the p value of predictive accuracy higher than 0.05. The second is the population race imbalance. The TCGA is an America project and the METABRIC is a Canada-UK project, as most of the patients included in our study were white, the results should be further researched in people of other races. The last is the study failed to use experiment to prove the results and exploring the underlying mechanisms of the results. Therefore, future research needs to do to better testify our results.

Conclusions

In conclusion, the study of us demonstrate that higher M2 macrophages infiltration in breast cancer tissues were the poor prognostic factor for OS. The nomogram model incorporated in macrophages M2 level, age, TNM stage and molecular types can help to better prognoses the survival of patients. The nomogram risk groups significantly different in immune features and can be a predictor for the prognosis and immunotherapy response in breast cancer for further research.

Abbreviations

HR

hazard ratio

AUC

the area under the curve values of

ROC

receiver operating characteristic

OS

overall survival

TILs

tumor-infiltrating lymphocytes

TME

tumor microenvironment

TAMs

tumor-associated macrophages

ER
estrogen receptor
PR
progesterone receptor
HER2
human epidermal growth factor receptor 2
C- index
the concordance index
SNV
single-nucleotide variant
BCR
B cell receptor
TCR
T cell receptor
TIDE
Tumor Immune Dysfunction and Exclusion
DEGs
different expression genes
FDR
false discovery rate
MCC
Maximal Clique Centrality
MNC
Maximum Neighborhood Component
EPC
Edge Percolated Component
DMNC
Density of Maximum Neighborhood Component
GO
gene ontology

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the TCGA repository [<https://www.cancer.gov/>] and METABRIC [<https://www.cbioportal.org/>].

Competing interests

The authors declare that they have no competing interests

Funding

This work was supported by the grant from National Natural Science Foundation of China (81972727, 81773093) and a Municipal Human Resources Development Program for Outstanding Leaders in Medical Disciplines in Shanghai (2017BR028).

Author Contributions

XL and HS contributed to the study design. XL, XY and YZ contributed to data collection. XL, QL and YZ performed statistical analysis, interpretation, and drafted the manuscript. YH, WJ and HS revised the manuscript. All authors contributed to critical revision of the final manuscript and approved the final version of the manuscript. XL, WJ, YH, and HS provided financial support and study supervision.

Acknowledgments

The authors declare that they have no acknowledgment to make.

References

1. DeSantis CE, Ma J, Gaudet MM, Newman LA, Miller KD, Goding Sauer A, Jemal A, Siegel RL: **Breast cancer statistics, 2019**. *CA Cancer J Clin* 2019, **69**(6):438-451.
2. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, Jemal A, Kramer JL, Siegel RL: **Cancer treatment and survivorship statistics, 2019**. *CA Cancer J Clin* 2019, **69**(5):363-385.
3. Emens LA: **Breast Cancer Immunotherapy: Facts and Hopes**. *Clin Cancer Res* 2018, **24**(3):511-520.
4. Lipson EJ, Forde PM, Hammers HJ, Emens LA, Taube JM, Topalian SL: **Antagonists of PD-1 and PD-L1 in Cancer Treatment**. *Semin Oncol* 2015, **42**(4):587-600.
5. Loi S, Sirtaine N, Piette F, Salgado R, Viale G, Van Eenoo F, Rouas G, Francis P, Crown JP, Hitre E *et al*: **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**(7):860-867.
6. Loi S, Michiels S, Salgado R, Sirtaine N, Jose V, Fumagalli D, Kellokumpu-Lehtinen PL, Bono P, Kataja V, Desmedt C *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**(8):1544-1550.

7. Salgado R, Denkert C, Campbell C, Savas P, Nuciforo P, Aura C, de Azambuja E, Eidtmann H, Ellis CE, Baselga J *et al*: **Tumor-Infiltrating Lymphocytes and Associations With Pathological Complete Response and Event-Free Survival in HER2-Positive Early-Stage Breast Cancer Treated With Lapatinib and Trastuzumab: A Secondary Analysis of the NeoALTTO Trial.** *JAMA Oncol* 2015, **1**(4):448-454.
8. Denkert C, von Minckwitz G, Brase JC, Sinn BV, Gade S, Kronenwett R, Pfitzner BM, Salat C, Loi S, Schmitt WD *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**(9):983-991.
9. Esteva FJ, Hubbard-Lucey VM, Tang J, Pusztai L: **Immunotherapy and targeted therapy combinations in metastatic breast cancer.** *Lancet Oncol* 2019, **20**(3):e175-e186.
10. Leek RD, Lewis CE, Whitehouse R, Greenall M, Clarke J, Harris AL: **Association of macrophage infiltration with angiogenesis and prognosis in invasive breast carcinoma.** *Cancer Res* 1996, **56**(20):4625-4629.
11. Campbell MJ, Tonlaar NY, Garwood ER, Huo D, Moore DH, Khramtsov AI, Au A, Baehner F, Chen Y, Malaka DO *et al*: **Proliferating macrophages associated with high grade, hormone receptor negative breast cancer and poor clinical outcome.** *Breast Cancer Res Treat* 2011, **128**(3):703-711.
12. Zhang Y, Cheng S, Zhang M, Zhen L, Pang D, Zhang Q, Li Z: **High-infiltration of tumor-associated macrophages predicts unfavorable clinical outcome for node-negative breast cancer.** *PLoS One* 2013, **8**(9):e76147.
13. Desmedt C, Haibe-Kains B, Wirapati P, Buyse M, Larsimont D, Bontempi G, Delorenzi M, Piccart M, Sotiriou C: **Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes.** *Clin Cancer Res* 2008, **14**(16):5158-5165.
14. Teschendorff AE, Miremadi A, Pinder SE, Ellis IO, Caldas C: **An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer.** *Genome Biol* 2007, **8**(8):R157.
15. Perez EA, Thompson EA, Ballman KV, Anderson SK, Asmann YW, Kalari KR, Eckel-Passow JE, Dueck AC, Tenner KS, Jen J *et al*: **Genomic analysis reveals that immune function genes are strongly linked to clinical outcome in the North Central Cancer Treatment Group n9831 Adjuvant Trastuzumab Trial.** *J Clin Oncol* 2015, **33**(7):701-708.
16. Gu-Trantien C, Loi S, Garaud S, Equeter C, Libin M, de Wind A, Ravoet M, Le Buanec H, Sibille C, Manfouo-Foutsop G *et al*: **CD4(+) follicular helper T cell infiltration predicts breast cancer survival.** *J Clin Invest* 2013, **123**(7):2873-2892.
17. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA: **Robust enumeration of cell subsets from tissue expression profiles.** *Nat Methods* 2015, **12**(5):453-457.
18. Bense RD, Sotiriou C, Piccart-Gebhart MJ, Haanen J, van Vugt M, de Vries EGE, Schroder CP, Fehrmann RSN: **Relevance of Tumor-Infiltrating Immune Cell Composition and Functionality for Disease Outcome in Breast Cancer.** *J Natl Cancer Inst* 2017, **109**(1).

19. Rohr-Udilova N, Klinglmüller F, Schulte-Hermann R, Stift J, Herac M, Salzmann M, Finotello F, Timelthaler G, Oberhuber G, Pinter M *et al*: **Deviations of the immune cell landscape between healthy liver and hepatocellular carcinoma.** *Sci Rep* 2018, **8**(1):6220.
20. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA: **Profiling Tumor Infiltrating Immune Cells with CIBERSORT.** *Methods Mol Biol* 2018, **1711**:243-259.
21. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y *et al*: **The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups.** *Nature* 2012, **486**(7403):346-352.
22. Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM: **The Cancer Genome Atlas Pan-Cancer analysis project.** *Nat Genet* 2013, **45**(10):1113-1120.
23. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA *et al*: **The Immune Landscape of Cancer.** *Immunity* 2018, **48**(4):812-830.e814.
24. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, Li Z, Traugh N, Bu X, Li B *et al*: **Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response.** *Nat Med* 2018, **24**(10):1550-1558.
25. Bremnes RM, Al-Shibli K, Donnem T, Sirera R, Al-Saad S, Andersen S, Stenvold H, Camps C, Busund LT: **The role of tumor-infiltrating immune cells and chronic inflammation at the tumor site on cancer development, progression, and prognosis: emphasis on non-small cell lung cancer.** *J Thorac Oncol* 2011, **6**(4):824-833.
26. Senbabaoglu Y, Gejman RS, Winer AG, Liu M, Van Allen EM, de Velasco G, Miao D, Ostrovnaya I, Drill E, Luna A *et al*: **Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures.** *Genome Biol* 2016, **17**(1):231.
27. Quail DF, Joyce JA: **Microenvironmental regulation of tumor progression and metastasis.** *Nat Med* 2013, **19**(11):1423-1437.
28. Joyce JA, Pollard JW: **Microenvironmental regulation of metastasis.** *Nat Rev Cancer* 2009, **9**(4):239-252.
29. Lim YZ, South AP: **Tumour-stroma crosstalk in the development of squamous cell carcinoma.** *Int J Biochem Cell Biol* 2014, **53**:450-458.
30. Taddei ML, Giannoni E, Comito G, Chiarugi P: **Microenvironment and tumor cell plasticity: an easy way out.** *Cancer Lett* 2013, **341**(1):80-96.
31. Galon J, Angell HK, Bedognetti D, Marincola FM: **The continuum of cancer immunosurveillance: prognostic, predictive, and mechanistic signatures.** *Immunity* 2013, **39**(1):11-26.
32. Fridman WH, Pages F, Sautes-Fridman C, Galon J: **The immune contexture in human tumours: impact on clinical outcome.** *Nat Rev Cancer* 2012, **12**(4):298-306.
33. Curiel TJ, Coukos G, Zou L, Alvarez X, Cheng P, Mottram P, Evdemon-Hogan M, Conejo-Garcia JR, Zhang L, Burow M *et al*: **Specific recruitment of regulatory T cells in ovarian carcinoma fosters**

- immune privilege and predicts reduced survival. *Nat Med* 2004, **10**(9):942-949.
34. Zou W: **Regulatory T cells, tumour immunity and immunotherapy.** *Nat Rev Immunol* 2006, **6**(4):295-307.
35. Lin YC, Mahalingam J, Chiang JM, Su PJ, Chu YY, Lai HY, Fang JH, Huang CT, Chiu CT, Lin CY: **Activated but not resting regulatory T cells accumulated in tumor microenvironment and correlated with tumor progression in patients with colorectal cancer.** *Int J Cancer* 2013, **132**(6):1341-1350.
36. Mantovani A, Sica A, Locati M: **Macrophage polarization comes of age.** *Immunity* 2005, **23**(4):344-346.
37. Hume DA: **The Many Alternative Faces of Macrophage Activation.** *Front Immunol* 2015, **6**:370.
38. Pollard JW: **Tumour-educated macrophages promote tumour progression and metastasis.** *Nat Rev Cancer* 2004, **4**(1):71-78.
39. Bingle L, Brown NJ, Lewis CE: **The role of tumour-associated macrophages in tumour progression: implications for new anticancer therapies.** *J Pathol* 2002, **196**(3):254-265.
40. Xiong Y, Wang K, Zhou H, Peng L, You W, Fu Z: **Profiles of immune infiltration in colorectal cancer and their clinical significant: A gene expression-based study.** *Cancer Med* 2018, **7**(9):4496-4508.
41. Mantovani A, Sica A: **Macrophages, innate immunity and cancer: balance, tolerance, and diversity.** *Curr Opin Immunol* 2010, **22**(2):231-237.
42. Rubio C, Munera-Maravilla E, Lodewijk I, Suarez-Cabrera C, Karaivanova V, Ruiz-Palomares R, Paramio JM, Duenas M: **Macrophage polarization as a novel weapon in conditioning tumor microenvironment for bladder cancer: can we turn demons into gods?** *Clin Transl Oncol* 2019, **21**(4):391-403.
43. Silvestrini R: **Relevance of DNA-ploidy as a prognostic instrument for solid tumors.** *Ann Oncol* 2000, **11**(3):259-261.
44. Xu J, Huang L, Li J: **DNA aneuploidy and breast cancer: a meta-analysis of 141,163 cases.** *Oncotarget* 2016, **7**(37):60218-60229.
45. Gibney GT, Weiner LM, Atkins MB: **Predictive biomarkers for checkpoint inhibitor-based immunotherapy.** *Lancet Oncol* 2016, **17**(12):e542-e551.
46. Hellmann MD, Callahan MK, Awad MM, Calvo E, Ascierio PA, Atmaca A, Rizvi NA, Hirsch FR, Selvaggi G, Szustakowski JD *et al*: **Tumor Mutational Burden and Efficacy of Nivolumab Monotherapy and in Combination with Ipilimumab in Small-Cell Lung Cancer.** *Cancer Cell* 2018, **33**(5):853-861.e854.
47. Thomas A, Routh ED, Pullikuth A, Jin G, Su J, Chou JW, Hoadley KA, Print C, Knowlton N, Black MA *et al*: **Tumor mutational burden is a determinant of immune-mediated survival in breast cancer.** *Oncoimmunology* 2018, **7**(10):e1490854.
48. Efremova M, Finotello F, Rieder D, Trajanoski Z: **Neoantigens Generated by Individual Mutations and Their Role in Cancer Immunity and Immunotherapy.** *Front Immunol* 2017, **8**:1679.
49. Chandran SS, Klebanoff CA: **T cell receptor-based cancer immunotherapy: Emerging efficacy and pathways of resistance.** *Immunol Rev* 2019, **290**(1):127-147.

50. Sharma B, Nawandar DM, Nannuru KC, Varney ML, Singh RK: **Targeting CXCR2 enhances chemotherapeutic response, inhibits mammary tumor growth, angiogenesis, and lung metastasis.** *Mol Cancer Ther* 2013, **12**(5):799-808.
51. Yu PF, Huang Y, Han YY, Lin LY, Sun WH, Rabson AB, Wang Y, Shi YF: **TNFalpha-activated mesenchymal stromal cells promote breast cancer metastasis by recruiting CXCR2(+) neutrophils.** *Oncogene* 2017, **36**(4):482-490.
52. Bohrer LR, Schwertfeger KL: **Macrophages promote fibroblast growth factor receptor-driven tumor cell migration and invasion in a CXCR2-dependent manner.** *Mol Cancer Res* 2012, **10**(10):1294-1305.
53. Wang C, Zhou XJ, Li YY, Wan J, Yang LY, Li GH: **Effect of vasoactive intestinal peptide (VIP) on NKG2D signal pathway and its contribution to immune escape of MKN45 cells.** *ScientificWorldJournal* 2013, **2013**:429545.
54. Moody TW, Ramos-Alvarez I, Jensen RT: **Neuropeptide G Protein-Coupled Receptors as Oncotargets.** *Front Endocrinol (Lausanne)* 2018, **9**:345.

Tables

Due to technical limitations, Tables 1-3 are provided in the Supplementary Files section.

Figures

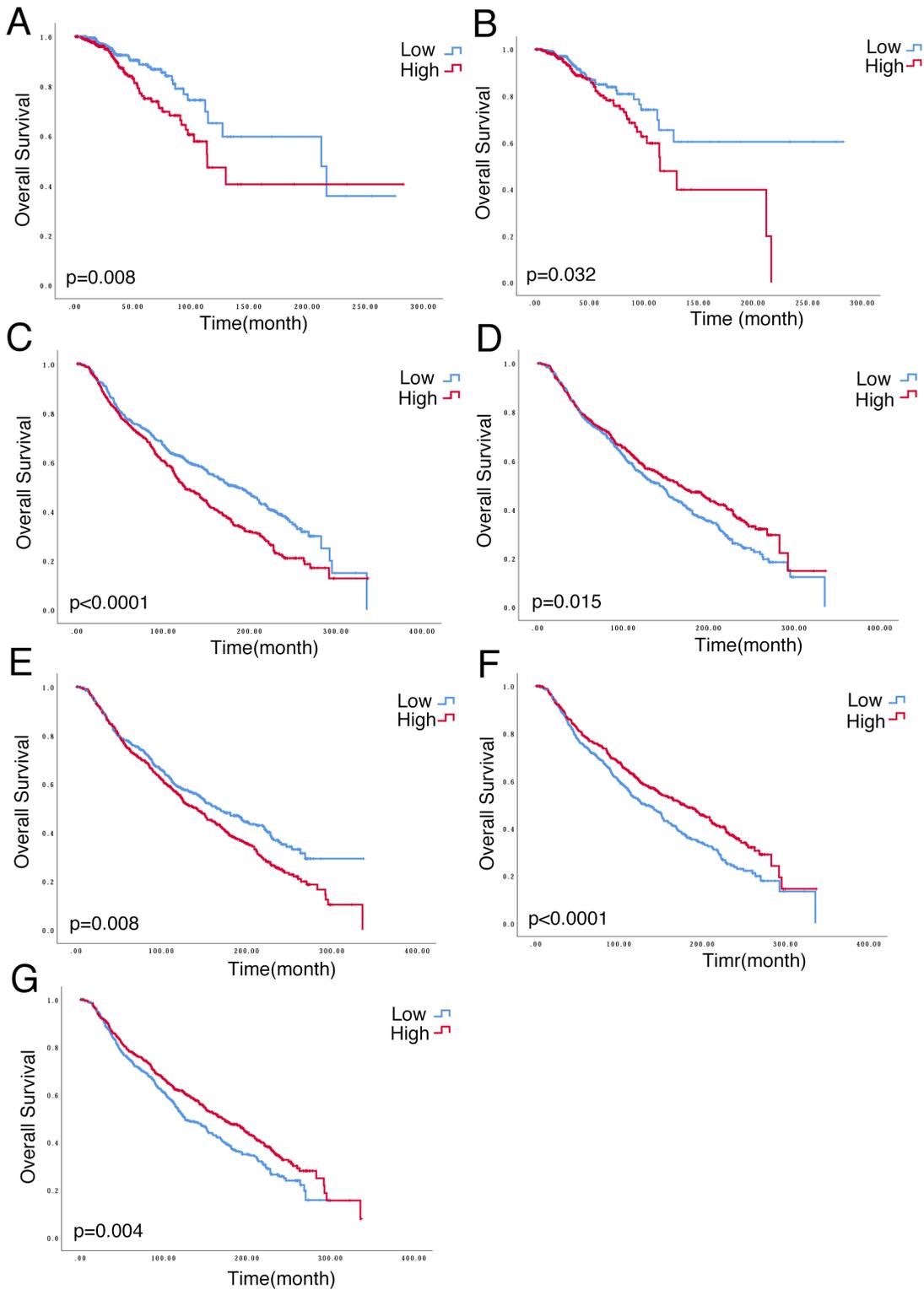


Figure 1

TIICs significantly associated with the prognosis of patients' OS in the TCGA and METABRIC group. Kaplan-Meier plots visualized that high level of M2 macrophages (A) and high level of Monocyte(B) were associated with poor OS in TCGA cohort. Kaplan-Meier plots visualized that high level of M2 macrophages (C), low level of T cells CD4 memory activated(D), high level of T cells regulatory Tregs (E),

low level of T cells CD4 memory activated(F) and low level of NK cells resting (G) were associated with poor OS in METABRIC cohort.

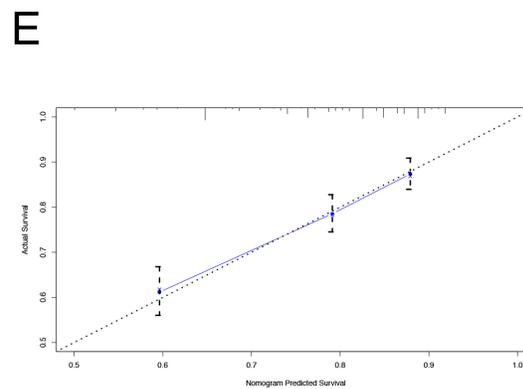
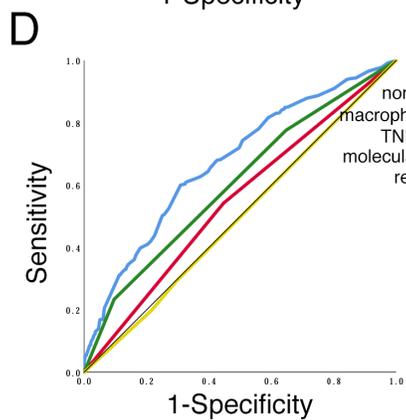
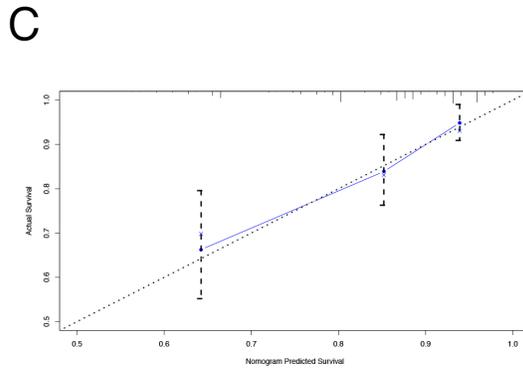
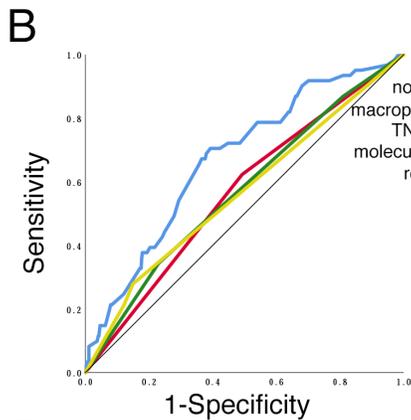
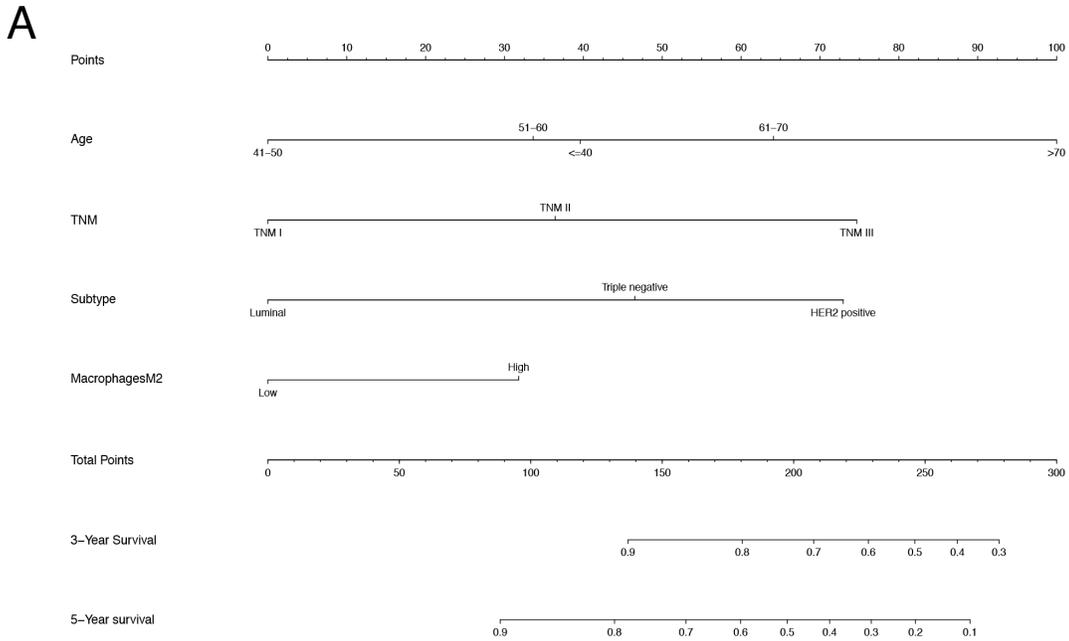


Figure 2

Nomogram and its prediction of OS in breast cancer. Nomograms for OS (A) based on macrophage M2 and other well-recognized prognosticators in TCGA cohort. The (ROC) curves for predicting patients' overall survival using the nomogram score, macrophage M2, TNM stage and molecular types in TCGA

cohort (B) and METABRIC cohort (D). The calibration of nomograms for OS in TCGA cohort (C) and METABRIC cohort (E) demonstrated a great prediction accuracy of this risk signature nomograms.

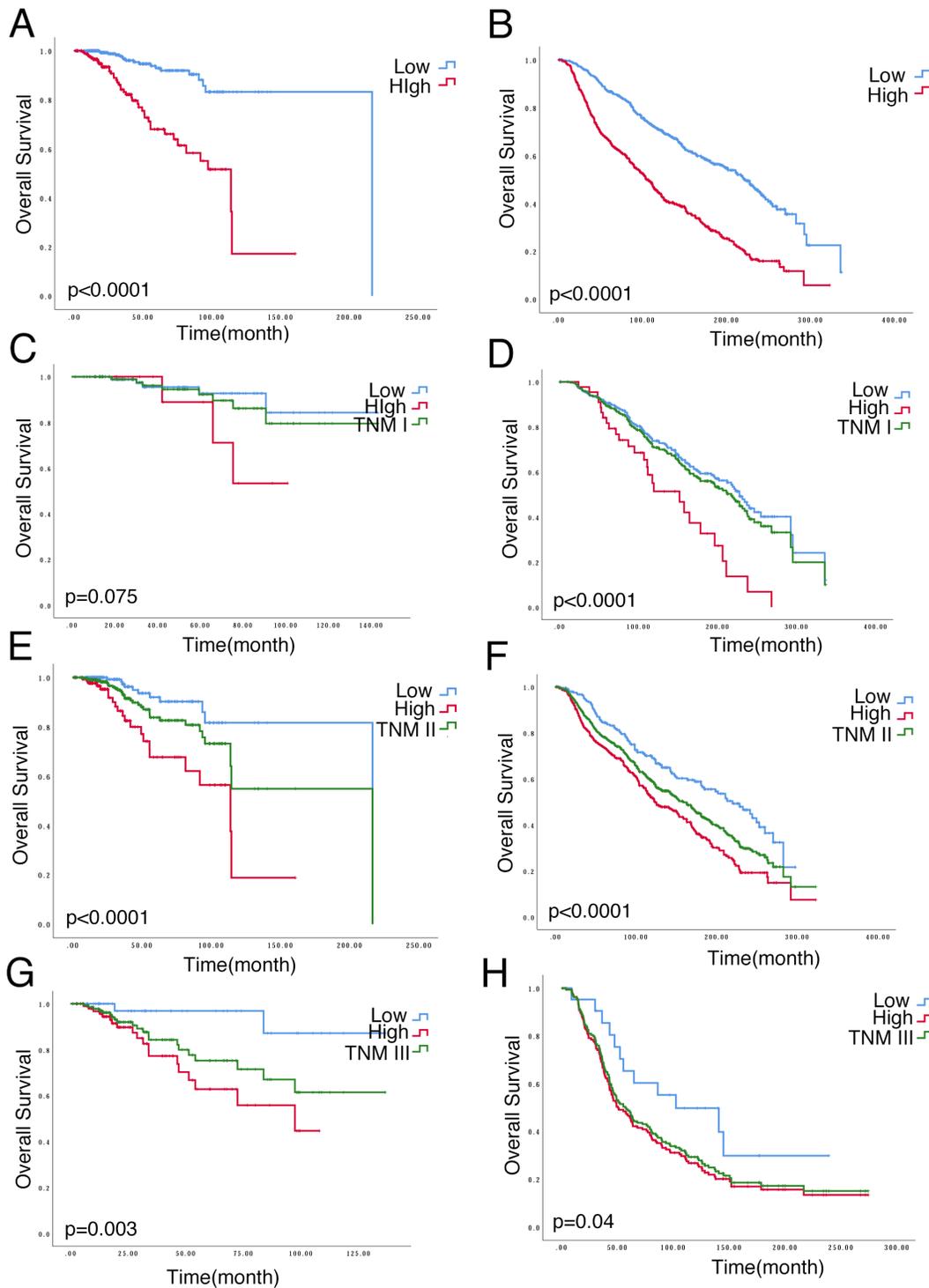


Figure 3

Nomogram risk group significantly associated with the patients' OS in breast cancer. Kaplan-Meier plots visualized nomogram high risk group in TCGA (A) and METABRIC (B) cohort were associated with poor OS in breast cancer. Kaplan-Meier plots visualized that the nomogram risk classification can separate

different TNM stages reclassified into a good or poor risk group in TCGA (TNM I (C), TNM II (E) and TNM III (G)) cohort and METABRIC (TNM I (D), TNM II (F) and TNM III (H)) cohort.

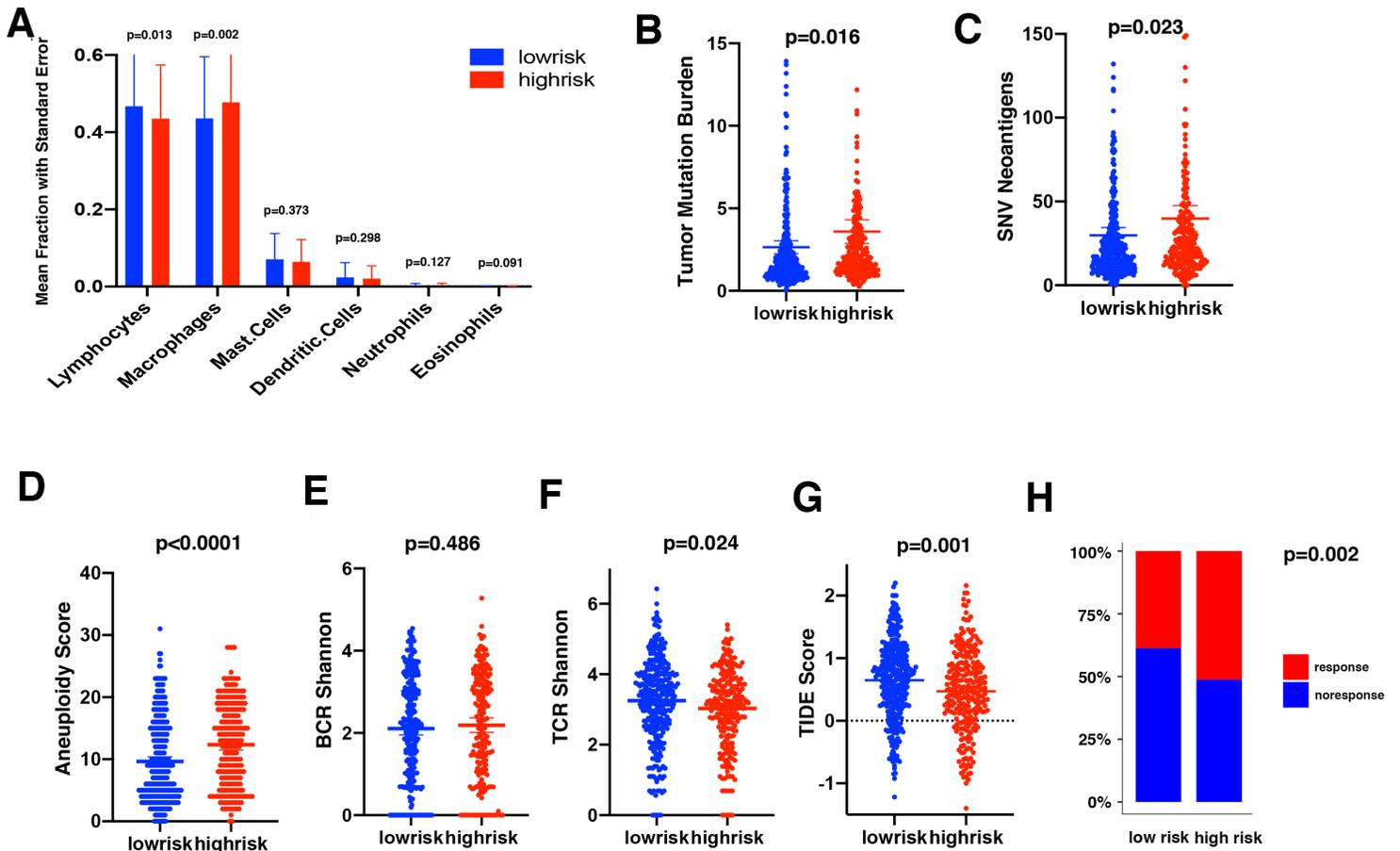


Figure 4

The association of nomogram risk groups with other immune features. (A) The proportion of six major classes of immune cells (from CIBERSORT) in nomogram low and high risk groups (Error bars show the standard error of the mean). Comparison the values of tumor mutation burden (B), SNV Neoantigens (C), Aneuploidy score (D), BCR Shannon (E), TCR Shannon (F) and TIDE Score (G) in low and high risk groups (Independent sample T-test). (H) Rates of the different anti-PD1 and anti-CTLA4 responses of patients from the TCGA cohort predicted by the TIDE web program, the predicted response was 51,2% in high risk group while 38.6% in low risk group (Chi-square test, P =0.002).

means hazard immune genes and the genes in green circles means protective immune genes. The red line means positive regulation between the two genes and the green line means negative regulation between the two genes. (F) Survival related genes' biological processes enrichment analysis in high risk group.

Supplementary Files

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

- [table2.xlsx](#)
- [table3.xlsx](#)
- [stable1.xlsx](#)
- [table1.xlsx](#)