



OPEN

Evaluating the tumor immune profile based on a three-gene prognostic risk model in HER2 positive breast cancer

Jianqing Lin¹, Aiyue Zhao² & Deqiang Fu²✉

To date, there have not been great breakthroughs in immunotherapy for HER2 positive breast cancer (HPBC). This study aimed to build a risk model that might contribute to predicting prognosis and discriminating the immune landscape in patients with HPBC. We analyzed the tumor immune profile of HPBC patients from the TCGA using the ESTIMATE algorithm. Thirty survival-related differentially expressed genes were selected according to the ImmuneScore and StromalScore. A prognostic risk model consisting of PTGDR, PNOC and CCL23 was established by LASSO analysis, and all patients were classified into the high- and low-risk score groups according to the risk scores. Subsequently, the risk model was proven to be efficient and reliable. Immune related pathways were the dominantly enriched category. ssGSEA showed stronger immune infiltration in the low-risk score group, including the infiltration of TILs, CD8 T cells, NK cells, DCs, and so on. Moreover, we found that the expression of immune checkpoint genes, including PD-L1, CTLA-4, TIGIT, TIM-3 and LAG-3, was significantly upregulated in the low-risk score group. All the results were validated with corresponding data from the GEO database. In summary, our investigation indicated that the risk model composed of PTGDR, PNOC and CCL23 has potential to predict prognosis and evaluate the tumor immune microenvironment in HPBC patients. More importantly, HPBC patients with a low-risk scores are likely to benefit from immune treatment.

Breast cancer is one of the most common carcinomas worldwide. Approximately 2.1 million new patients were reportedly diagnosed with breast cancer in 2018, and new patients with breast cancer accounted for approximately 24.2% of all cancer cases among women¹. HPBC, a subtype of breast cancer, is characterized by the over-expression of HER2 protein², shows drug resistance and highly aggressive biology, and is associated with poor survival^{3–5}. Published data show that HPBC accounts for one-fourth to one-fifth of breast cancers⁶.

According to NCCN guidelines, the combination regimen of chemotherapy plus anti-HER2 drugs is regarded as the standard treatment regimen for HPBC patients. Although patient survival has been markedly prolonged by treatment with anti-HER2 drugs⁷, HPBC patients still face lack of optimal subsequent treatment after the development of drug resistance or tumor recurrence. In recent years, tumor immunotherapy, represented by PD-1/PD-L1 pathway blocking antibodies, has shown great success in various malignancies^{8–12}, including triple-negative breast cancer¹³. Unfortunately, to date, there has been no great improvement in tumor immunotherapy for HPBC, except for a few reports in animal experiments¹⁴.

However, in the past few years, accumulating evidence has demonstrated a positive association between the tumor immune microenvironment (TIME) and the prognosis of HPBC patients. An increased number of stromal tumor-infiltrating lymphocytes (TILs) denotes a good prognosis in HPBC patients who received chemotherapy¹⁵. Coincidentally, high PD-L1 expression and marked CD8 TIL infiltration were regarded as markers of better outcome in HPBC patients treated with conventional chemotherapy and HER2-blocking therapy¹⁶. Recently, assessment of the TIME based on gene expression has attracted increasing attention in the HPBC field^{17,18}. However, a recent phase II clinical trial showed that combination treatment with anti-Her2 and anti-PD-L1 agents did not improve progression-free survival (PFS) in HPBC patients¹⁹. Hence, it is crucial to accurately evaluate the TIME in HPBC patients.

¹Thyroid & Breast Surgery, The Second Affiliated Hospital of Fujian Medical University, Quanzhou, China. ²Department of Oncology, The Second Affiliated Hospital of Fujian Medical University, Quanzhou, China. ✉email: honglai1999@fjmu.edu.cn

In this study, we attempted to construct a risk model for predicting prognosis and evaluating the tumor immune profile in HPBC patients. Initially, we explored the immune-infiltrating profile of HPBC patients based on the ESTIMATE algorithm using data from the TCGA and GEO databases. Then, 30 survival-related genes were selected from 606 differentially expressed genes that were previously identified according to the ESTIMATE score. Finally, we created a gene-based risk model that may be helpful for accurately predicting prognosis and assessing the tumor-infiltrating immune profile in patients with HPBC. With this model, we hope to provide personalized tumor immunotherapy for appropriate patients with HPBC in the future.

Results

Characteristics of the cohorts. A total of 280 HER2-positive patients from the TCGA, GSE20711, GSE45255, GSE162228 and GSE1456 databases were selected for our study. The training cohort contained 132 patients from the TCGA. The validation cohort comprised 26 patients from GSE20711, 61 from GSE45255, 46 from GSE162228 and 15 from GSE1456. The accuracy of the constructed model was verified by a validation cohort. The clinical characteristics of the HPBC patients are shown in Table S1.

Immune infiltration is a survival-associated factor in HPBC patients. To evaluate immune cell infiltration in HPBC patients, the ESTIMATE algorithm was utilized to calculate the ESTIMATE scores for the TCGA cohort, reflecting the TIME landscape²⁰. The median scores, including the StromalScore, ImmuneScore and ESTIMATEScore, were assigned as the cutoff values. All patients were classified into the corresponding low- and high-risk groups. The results of the Kaplan–Meier survival analysis showed that the overall survival (OS) in the high ImmuneScore group was better than that in the low ImmuneScore group ($p = 0.03$), but no survival difference was found between the groups based on the StromalScore ($p = 0.224$) or the ESTIMATEScore ($p = 0.106$) (Fig. 1A–C). We further analyzed the correlations of OS with ImmuneScore, age and stage by univariate and multivariate Cox analyses. Unexpectedly, the results showed that the ImmuneScore could not be regarded as an independent prognosis-related factor in the TCGA cohort (Fig. S1).

Subsequently, we explored differential gene expression between groups based on the StromalScore and ImmuneScore. In the StromalScore groups, 1140 differentially expressed genes (DEGs) were identified ($|\log FC| \geq 1$, $p < 0.05$). Among them, 1110 genes were significantly upregulated and 30 genes were significantly downregulated (Fig. 1D). Similarly, among 1268 DEGs identified in the ImmuneScore groups, 1244 were significantly upregulated and 24 were significantly downregulated ($|\log FC| \geq 1$, $p < 0.05$) (Fig. 1E). Furthermore, 606 DEGs (604 upregulated genes and 2 downregulated genes) were identified in both the StromalScore groups and ImmuneScore groups (Fig. 1F,G). Next, we found that 30 of the 606 DEGs were strongly associated with OS (Table S2). Among the 30 survival-related DEGs, 8 (CXCR2, PTGDR, GPR171, CD3E, CD3D, P2RX1, PNOC and CCL23) were involved in the immune-infiltrating evaluation of CIBERSORT²¹.

Construction of a prognosis-related risk model based on survival-associated genes. To precisely evaluate the correlation between immune infiltration and OS, 8 survival-related DEGs were analyzed based on a least absolute shrinkage and selection operator (LASSO) algorithm, and 3 survival-related DEGs (PTGDR, PNOC and CCL23) were selected. Next, the corresponding LASSO coefficients and the optimum value of the penalty parameter λ ($\lambda = 0.04003709$) were acquired (Fig. 2A).

Thus, a prognosis-related risk model was established, and the survival of HPBC patients was assessed according to the risk score calculated based on the expression of the three genes. Risk score = $(-3.372 \times \text{PTGDR}) + (-0.948 \times \text{PNOC}) + (-0.570 \times \text{CCL23})$. The median risk scores in the different cohorts served as the corresponding cutoff values. Then, all patients were divided into the low- and high-risk score groups. The survival analysis results showed that patients with a low-risk score had a better OS than those with a high-risk score, with p values of 0.009 in the training cohort and 0.0037 in the validation cohort (Fig. 2B). The predictive accuracy of this risk model was assessed by time-dependent receiver operating characteristic (ROC) analysis. The area under the curve (AUC) values were 0.932, 0.890, and 0.841 at 1 year, 2 years, and 3 years, respectively, in the training cohort (Fig. 2C). Similarly, the AUC values were 0.843, 0.517 and 0.574 at the same time points in the validation cohort (Fig. 2C).

All patients were classified based on risk score, OS and living status, as shown in Fig. 2D. Principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) analysis were applied to assess the interrelationship between patients in different risk groups (Fig. 2E,F). The results suggested that the established risk model contributed to the accurate classification of HPBC patients.

To judge the independent prognostic ability of the risk model, several clinical features, including age, tumor stage (T stage or N stage), histologic grade and ER/PR status, were regarded as binary variables and evaluated by univariate and multivariate analyses. The results suggested a strong association between the risk score and the OS of HPBC patients in the training cohort. The p values for the univariate and multivariate analyses were 0.019 and 0.029, respectively (Fig. 3A,B). Similar results were found in the validation cohort; the p values for the univariate and multivariate analyses were 0.046 and 0.03, respectively (Fig. 3C,D). These data provided evidence that the risk score can function as an independent prognostic factor to evaluate the OS of patients with HPBC.

Risk scores correlate with the TIME of HPBC patients. To investigate the relationship between the risk score and the TIME of HPBC patients, a single-sample gene set enrichment analysis (ssGSEA) algorithm was used to discriminate immune infiltration in HPBC patients. The low-risk score group exhibited more significant immune cell infiltration, including of CD8 T cells, TILs, T helper cells, DCs, aDCs, iDCs, neutrophils, macrophages, NK cells, pDCs and Tregs (Fig. 4A,B). The low-risk score group also demonstrated a more powerful immune response, including checkpoint activity, T-cell coinhibition, T-cell costimulation, APC costimu-

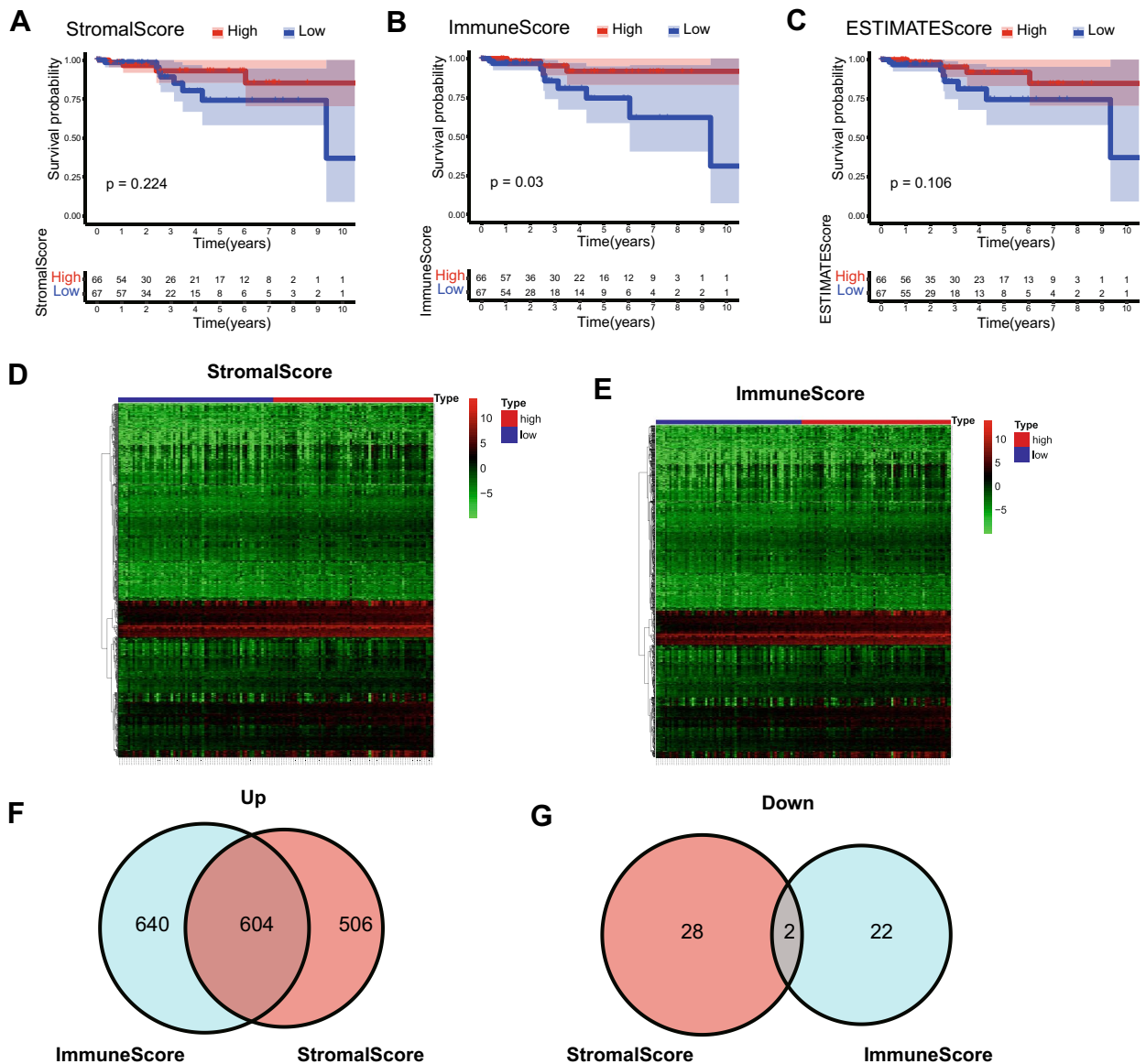


Figure 1. Survival analysis and differentially expressed genes based on the ESTIMATE algorithm in the training cohort. (A–C) Kaplan–Meier survival curves of patients grouped by the median of StromalScore, ImmuneScore and ESTIMATEScore. (D,E) Heatmap of differentially expressed genes between the high score group and the low score group from StromalScore data and ImmuneScore data. (F) Venn diagram of the upregulated differentially expressed genes from StromalScore data and ImmuneScore data. (G) Venn diagram of the downregulated differentially expressed genes from StromalScore data and ImmuneScore data.

lation, cytolytic activity, CCR, inflammation promotion, parainflammation and APC coinhibition. Based on the ssGSEA score, the CD8 T cell/Treg ratio and CD8 T cell/neutrophil ratio were calculated. As shown in Fig. S3, the CD8 T cell/Treg ratio ($p < 0.001$) and CD8 T cell/neutrophil ratio ($p < 0.001$) were both higher in the low-risk group than in the high-risk group. Furthermore, unsupervised clustering analysis indicated that the patients with a low-risk score almost always harbored greater immune infiltration (Fig. 4C). In addition, the ESTIMATE analysis showed that the low-risk score group possessed a higher StromalScore ($p < 0.001$), ImmuneScore ($p < 0.001$) and ESTIMATEScore ($p < 0.001$) than the high-risk score group in both cohorts (Fig. 4D).

Identification and functional enrichment of DEGs grouped by risk scores. Next, we analyzed the DEGs between different risk score groups in both cohorts. Our results showed that 3338 DEGs were identified in the training cohort ($p < 0.05$, Fig. 5A), and 1205 DEGs were identified in the validation cohort ($p < 0.05$, Fig. 5B). Among these genes, 322 overlapping DEGs were found with the cutoff $p < 0.01$ (Fig. S2).

Furthermore, the potential functions of 322 DEGs were evaluated according to Kyoto Encyclopedia of Genes and Genomes (KEGG) and gene ontology (GO) enrichment analyses. KEGG analysis showed that immune-related pathways (including cytokine–cytokine receptor interaction pathway, antigen processing and presentation, B cell receptor signaling pathway and chemokine signaling pathway) are potentially the most important

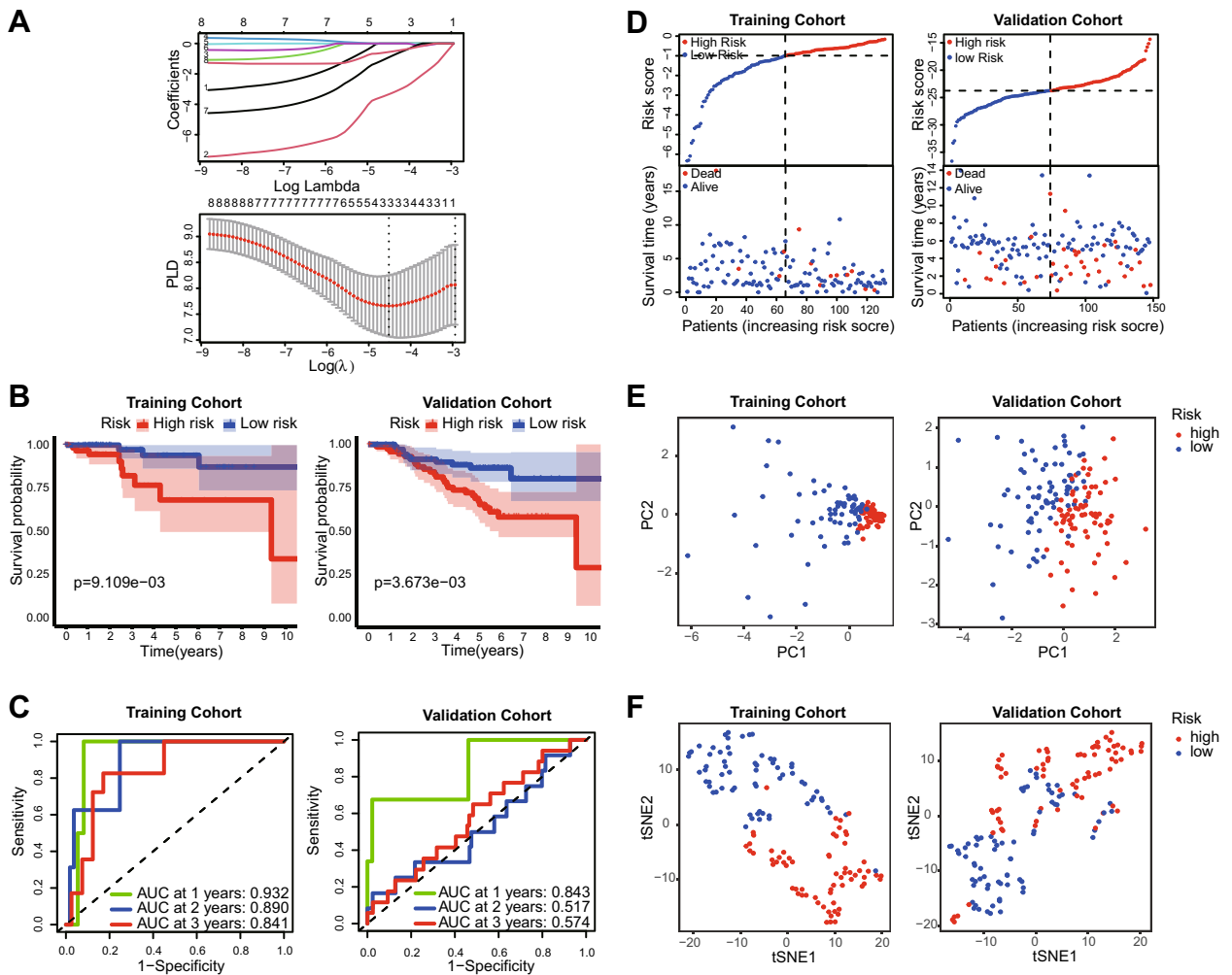


Figure 2. LASSO analysis based on survival-related genes and validation of the established risk model. **(A)** LASSO coefficient profiles (top) of survival-related genes and tenfold cross-validation results (bottom), which identified optimal values of the penalty parameter λ in the training cohort. **(B)** Kaplan–Meier survival curves of patients grouped by the risk score in the training cohort and validation cohort. **(C)** ROC curves of the risk model in training cohort and validation cohort. **(D)** HPBC patients were divided into high- and low-risk score groups according to the risk score model. **(E, F)** PCA and tSNE of the risk model in the training cohort and validation cohort.

mechanisms. (Fig. 6A). GO analysis suggested that these overlapping DEGs primarily mediated immune-related biological functions (Fig. 6B). These data showed that the important biological processes included T cell activation, lymphocyte proliferation, regulation of T cell activation, lymphocyte differentiation and positive regulation of lymphocyte activation. In addition, the enriched cellular component and molecular function categories are in Fig. 6B, these included external side of plasma membrane, plasma membrane signaling receptor complex, immunological synapse, immune receptor activity, cytokine receptor activity and MHC protein complex binding. Furthermore, GSEA was performed to map all DEGs. As shown in Fig. S4, in both the training and validation cohorts, immune-related pathways and biological functions were the crucial factors, and were enriched in the low-risk group. To observe interactions among these 322 overlapping DEGs, the STRING protein–protein interaction (PPI) database was used with an interaction score greater than 0.95. As shown in Fig. 7A, all the DEGs were divided into four groups: immune cell migration (green), immune cell activation and maturation (red), immune cell binding (blue) and intracellular signal transduction (yellow). The top 30 genes ranked by interaction count are shown in Fig. 7B; these genes included CD4, LCP2, LCK, PTPRC and FYN. Taken together, the evidence indicates that these DEGs based on the risk model may play a critical role in the tumor immune profile in HPBC patients.

Finally, we evaluated the importance of immune checkpoint-related genes, such as PD-L1, LAG-3, CTLA-4, TIM-3 and TIGIT. The results demonstrated that the expression of these 5 genes was significantly upregulated in the low-risk score group of the training cohort (all p values < 0.001 , Fig. 8A). As expected, similar differences were found in the validation cohort ($p < 0.001$, $p = 0.026$, $p < 0.001$, $p = 0.012$, and $p < 0.001$, respectively; Fig. 8B).

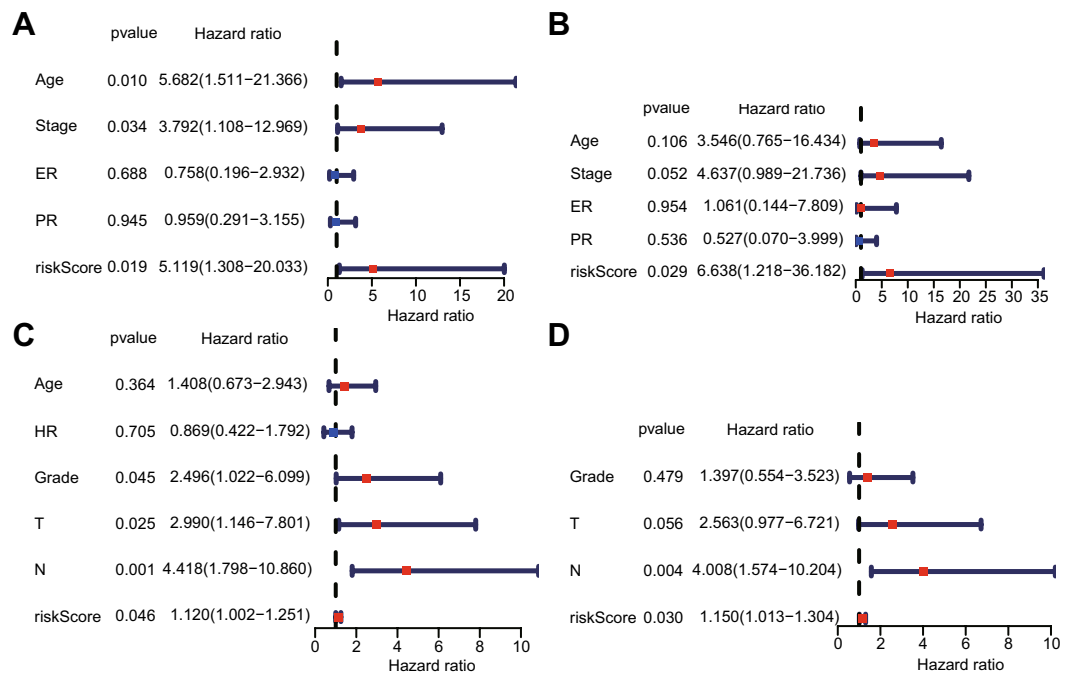


Figure 3. The results of the univariate and multivariate Cox regression analyses. In the training cohort, the forest plots for univariate Cox regression analysis (A) show that risk score (high risk vs. low risk), AJCC stage (stages I vs. II, III and IV) and age (≤ 50 vs. > 50) were prognostic risk-related variables. The forest plots for multivariate Cox regression analysis (B) show that the risk score (high risk vs. low risk) was an independent prognostic factor. In the validation cohort, the forest plots of univariate Cox regression analysis (C) and multivariate Cox regression analysis (D) indicated that the risk score (high risk vs. low risk) was an independent prognostic factor.

Discussion

At present, tumor immunotherapy is an important treatment option, that is widely utilized to treat various malignancies. However, this revolutionary therapeutic regimen has not achieved great breakthroughs in patients with HPBC. Therefore, it is crucial to distinguish the tumor immune-infiltrating profile in HPBC patients, and in the future, we will likely provide immunotherapy for appropriate patients according to TIME analysis. In our study, we initially utilized the ESTIMATE algorithm to assess tumor immune infiltration in patients with HPBC. Our results indicated a strong association between tumor immune infiltration and the prognosis of HPBC patients based on the ImmuneScore. Unfortunately, the ImmuneScore could not serve as an independent marker with predictive prognostic ability.

In recent years, risk models with the ability to predict prognosis have gained increasing attention. Many risk models are established based on gene expression profiles and are widely utilized in studies of various malignancies, such as breast cancer²² and lung adenocarcinoma²³. Hence, after 8 differentially expressed genes based on ESTIMATE scores were verified to be significantly associated with survival and the immune-infiltrating assessment of CIBERSORT, the LASSO algorithm was employed to select optimal genes. Finally, 3 genes including PTGDR (prostaglandin D2 receptor), PNOC (prepronociceptin) and CCL23 (C–C motif chemokine ligand 23), were selected as core entities of this model. These genes are classified as immune-related genes in the Immunology Database and Analysis Portal (Imm-Port) (<https://immport.niaid.nih.gov>)²⁴. PTGDR, known as the receptor of prostaglandin D2, contains 2 subtypes: prostaglandin D2 receptor 1 (DP1) and prostaglandin D2 receptor 2 (DP2). PTGDR may serve as an immune inhibition-related gene. A few studies have shown that PTGDR might suppress the function of natural killer cells²⁵ and promote an increase in Foxp3⁺ CD4⁺ regulatory T cells²⁶. Accumulating data suggest that hypermethylation of PTGDR might play a role in the tumorigenesis and development of various cancers, including bladder cancer²⁷, colon cancer²⁸, lung adenocarcinoma²⁹, cervical cancer³⁰, endometrial carcinoma³¹ and gastric cancer³². PNOC encodes a preproprotein that is proteolytically processed to generate multiple protein products, including nociceptin, nocistatin, and orphanin FQ2 (OFQ2)³³. It has been reported that nociception/orphanin FQ2 (N/OFQ) might be activated by the proinflammatory cytokines, interleukin-1 beta (IL-1beta) and tumor necrosis factor-alpha (TNF-alpha) via the ERK 1/2 and p38 MAP kinases pathways³⁴. Further, N/OFQ interacts with nociceptin receptor (NOP) in cancer, and promotes the proliferation and invasion of tumor cells via the PI3K/Akt signaling pathway³⁵, it also functions as an immune inhibitor by binding the N/OFQ peptide receptor on the surface of circulating immune cells³⁶ or inhibiting the activation of DCs³⁷. Published clinical data showed that NOP overexpression implied a poor prognosis in non-small-cell lung cancer³⁵. However, until now, few studies have verified the direct association between PNOC and cancer. Thus, we hypothesize that PNOC may indirectly mediate immune inhibition and promote tumor growth by

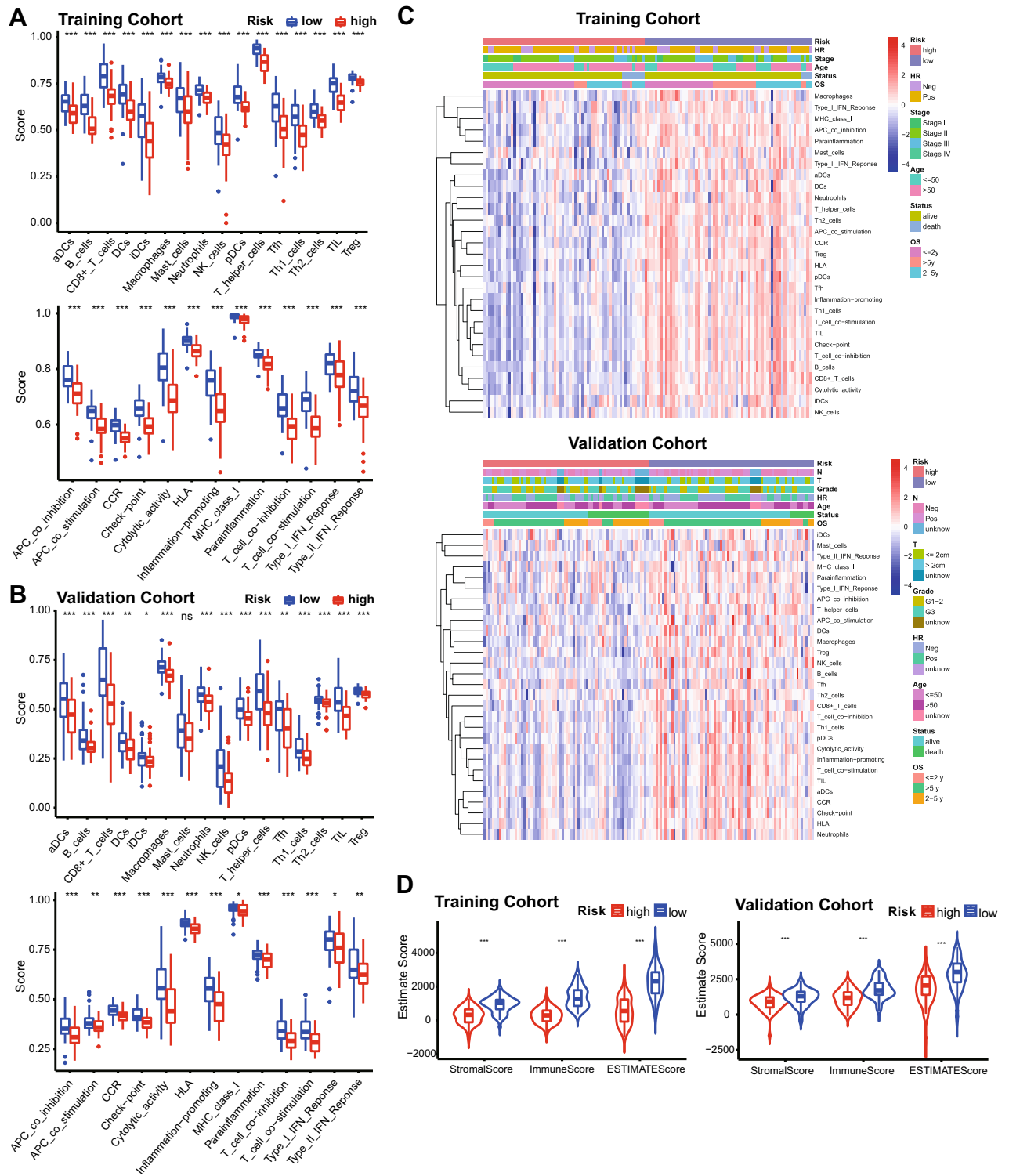


Figure 4. Tumor immune landscape in the high-risk score group and low-risk score group. **(A)** Immune infiltration (top) and immune function (bottom) in the training cohort. **(B)** Immune infiltration (top) and immune function (bottom) in the validation cohort. **(C)** Unsupervised clustering heat map showing the association between immune infiltration and survival-related clinical characteristic parameters, including risk score, age, hormone receptor status, histology grade, tumor stage, OS and survival status, in the training cohort (top) and validation cohort (bottom). **(D)** Comparison of the Estimate scores between the high-risk score group and the low-risk score group in the training cohort and validation cohort.

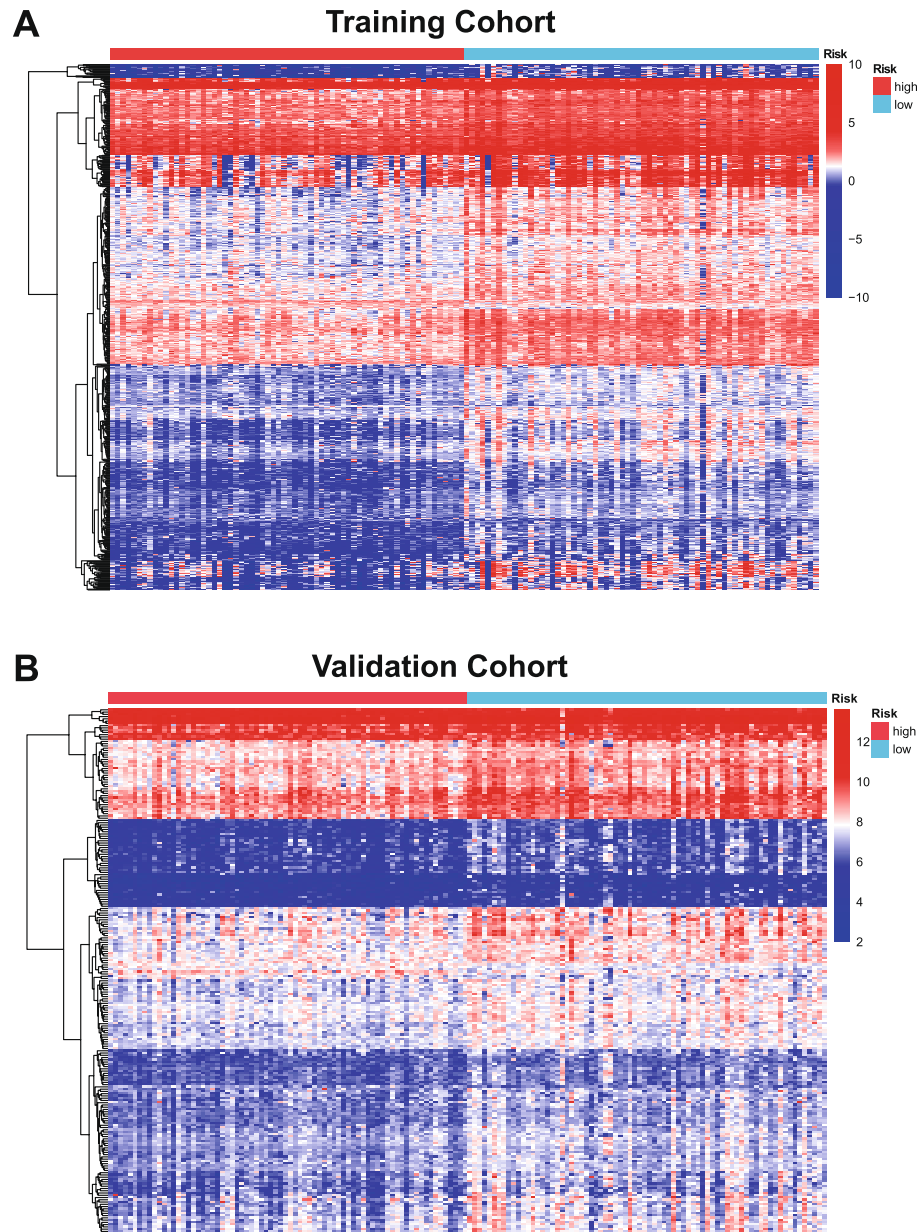


Figure 5. Differentially expressed genes in the training cohort and validation cohort. **(A)** Heatmap of differentially expressed genes between the high-risk score group and the low-risk score group in the training cohort. **(B)** Heatmap of differentially expressed genes between the high-risk score group and the low-risk score group in the validation cohort.

N/OFG. CCL23 is a member of the CC chemokine family and demonstrates chemotactic activity on resting T lymphocytes, monocytes, neutrophils³⁸ and dendritic cells³⁹ but is not expressed on activated T lymphocytes. In addition, human neutrophils orchestrate the recruitment of different cell types to inflamed sites by releasing CCL23, which then controls the immune response⁴⁰. There is much evidence that CCL23 expression is significantly upregulated in various cancers, including lung cancer⁴¹, colorectal cancer^{42,43}, and ovarian cancer⁴⁴. CCL23 could promote ovarian cancer migration by activating the ERK1/2 and PI3K pathways⁴⁴. Furthermore, CCL23 expression exhibited positive correlation with PD-L1 expression⁴⁵. Hence, we hypothesize that CCL23 chiefly acts as an immune inhibitor and promotes tumor growth. However, a liver cancer study showed that CCL23 could suppress tumor progression by promoting the immune infiltration of CD8 T cells⁴⁶.

According to the median risk score, all patients were classified into the high- and low-risk score groups. The predictive capacity of the risk model was verified to be precise and powerful. The patients in different cohorts were grouped by the respective median risk score, because there was a great difference in the range of the risk scores between the two cohorts. This difference in median risk score might be caused by detection platform, detection batch, and geographic area.

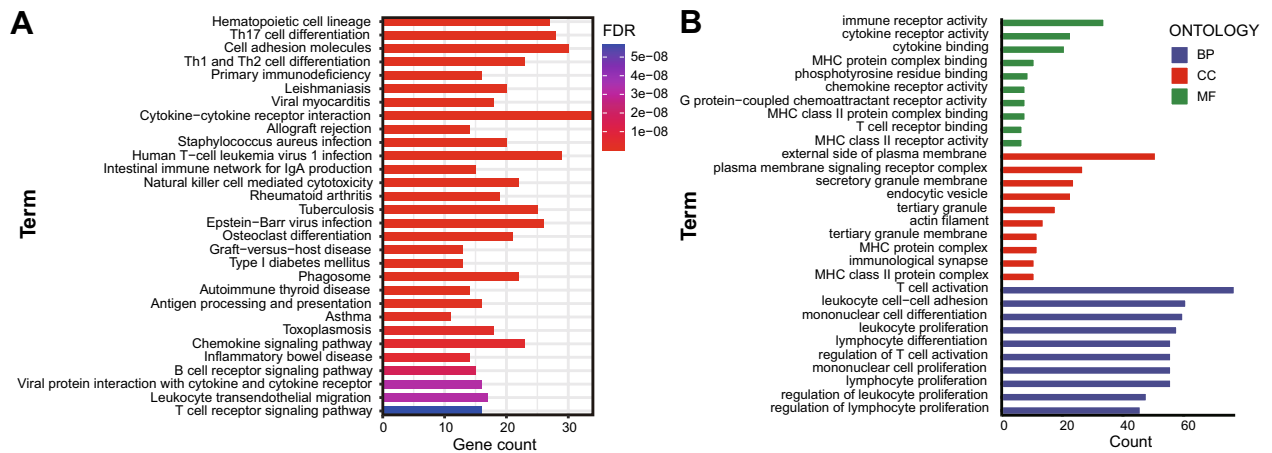


Figure 6. KEGG and GO analysis of differentially expressed genes co-existing in the training cohort and validation cohort. **(A)** Representative results of KEGG pathway analysis of differentially expressed genes. **(B)** Representative results of GO enrichment analysis of differentially expressed genes.

The ssGSEA algorithm was broadly employed to assess the TIME. A higher ssGSEA score indicated a better immune-infiltrating profile and potentially a better prognosis. In our study, the low-risk score group demonstrated strong immune infiltration, including of CD8 T cells, TILs, T helper cells, DCs, aDCs, iDCs, neutrophils, macrophages, NK cells, pDCs and Tregs. Numerous studies have confirmed that many subtypes of immune cells contribute to promoting the antitumor immune response. DCs play a crucial role during T-cell priming by presenting processed MHC antigenic peptides⁴⁷. In addition, macrophages and NK cells act as part-time antigen presenting cells and contribute to T-cell priming⁴⁸. CD8 T cells and TILs professionally execute cytotoxic effects on tumor cells, and T helper cells were confirmed to kill tumor cells via antigen cross-presentation⁴⁹. Hence, we confirmed the presence of a better tumor-infiltration profile in the TIME of patients with a low-risk score.

Unexpectedly, the proportions of neutrophils and Tregs, which are regarded as immunosuppressive cells, were higher in the patients with a low-risk score. Most studies provide evidence that an increased neutrophil or Treg frequency would imply poor outcome or survival. Nonetheless, contrasting evidence has been published. Some research found that high infiltration of Tregs could improve overall survival in Hodgkin's lymphoma^{50,51}, follicular lymphoma⁵², head and neck cancer⁵³ and breast cancer⁵⁴. Several authors proposed that Tregs are not a homogenous population, and can be divided into several subgroups that differ in function according to surface antigen expression⁵⁵. Therefore, Tregs appear to play dual roles in cancer and cannot accurately predict prognosis⁵⁶. Recently, the TIL/Treg ratio and TIL/neutrophil ratio were reported to be powerful prognostic indicators of survival or treatment outcome in breast cancer^{57,58} and other cancers^{59,60}. In our study, the low-risk group had a higher CD8 T cell/Treg ratio and CD8 T cell/neutrophil ratio. This result suggested that the low-risk group might have a better prognosis and was consistent with the survival analysis results.

As expected, the low-risk score group showed signs of a more powerful immune response, including checkpoint activity, T-cell coinhibition, T-cell costimulation, APC costimulation, cytolytic activity, CCR, increased inflammation, parainflammation and APC coinhibition. Publication of the theory of the T-cell costimulatory pathway was an important milestone in the field of immunology⁶¹, and this pathway includes both coinhibitory and costimulatory signaling pathways⁶². Accumulating evidence has proven that costimulatory molecules (either coinhibitory receptors or costimulatory receptors) direct T-cell function and determine T-cell fate after T cells are activated by the TCR signaling pathway⁶². In essence, costimulatory molecules are immune checkpoints. Previous studies have reported that enhanced cytolytic activity is a feature of ADCC mediated by activated T cells⁶³ or NK cells⁶⁴. Activated T cells migrate directly into tumor tissue along the chemokine gradient and kill tumor cells. Chemokine receptors (CCRs) expressed on the surface of T cells mediate this migration of T cells by binding the corresponding chemokine⁶⁵. Overall, the immune function profile in HPBC patients with a low-risk score might suggest an increased antitumor immune response.

Tumor tissue composition including of stromal cells and immune cells, is strongly correlated with the OS of cancer patients⁶⁶. Coincidentally, the ESTIMATE algorithm could assess the overall constitution of tumor tissue, including stromal cells and immune cells. Furthermore, several studies have validated that the ESTIMATE score might accurately map the TIME and precisely predict the OS of patients with cancer⁶⁷. Similar to the ssGSEA results, our ESTIMATE analysis results support that patients with a low-risk score harbored a better immune-infiltration profile.

Subsequently, we tried to reveal the potential mechanism that promotes antitumor immune infiltration and improves the OS of patients with HPBC. A total of 322 DEGs overlapped between the two cohorts. We thought that these genes might play crucial roles during tumorigenesis and progression. Further KEGG analysis indicated that these DEGs were involved in immune-related pathways. Meanwhile, the top GO terms suggested that these genes might mediate the entire T-cell immune response, including priming, activation, proliferation, migration, recognition and killing. In addition, the gene-gene interaction analysis results strongly supported the importance of immune-related signaling pathways. Moreover, GSEA revealed an overview of the TIME in HPBC patients.

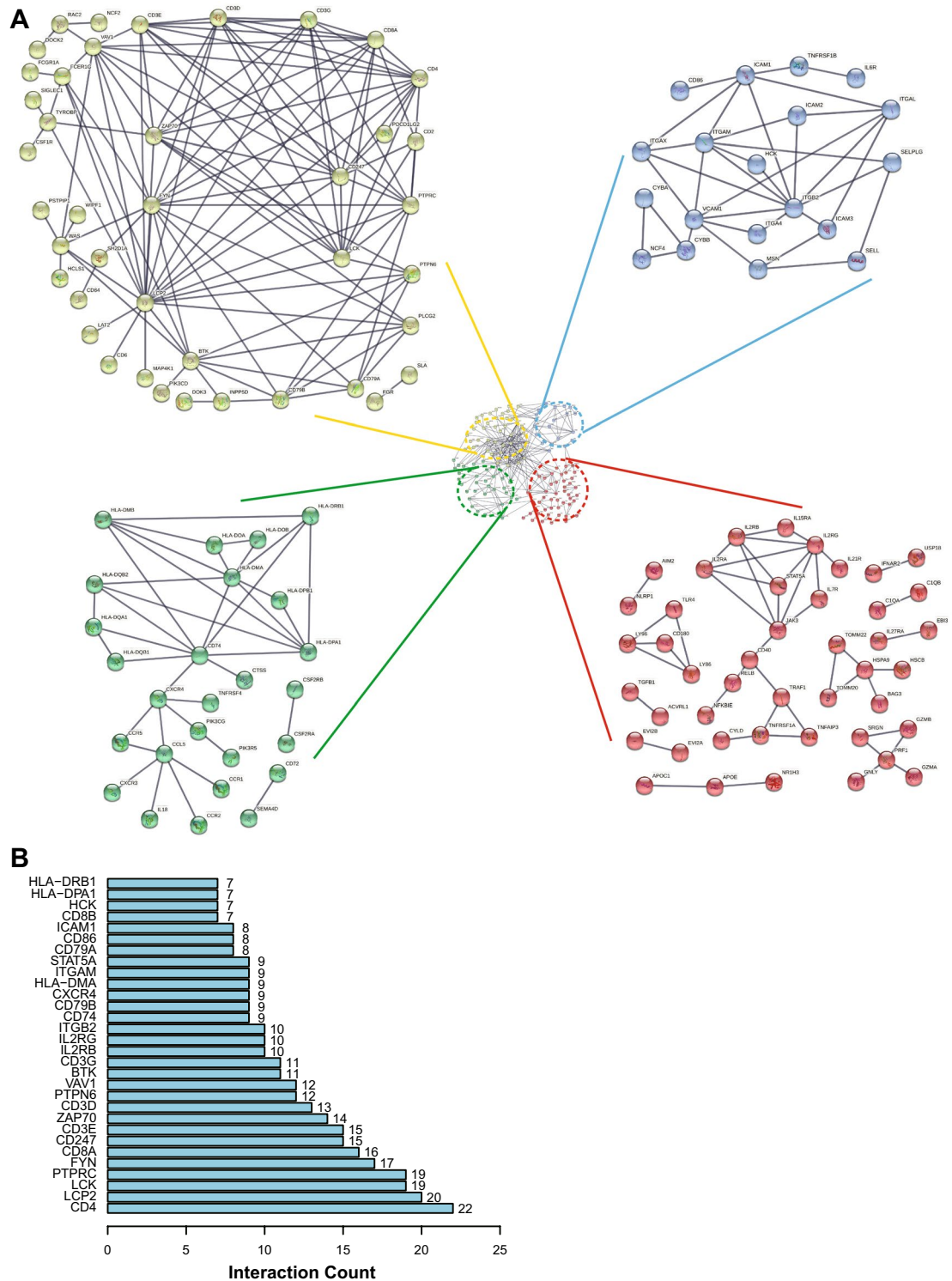


Figure 7. The results of protein–protein interactions of differentially expressed genes co-existing in the training cohort and validation cohort. A total of 322 differentially expressed genes were analyzed in the STRING database, a network (A) of PPIs was acquired (the confidence was 0.95), the different colors represented different subgroups after PPI analysis, and the top 30 genes (B) were ranked by interaction count.

We further verified the difference in the TIME between the low- and high-risk groups and hypothesized that this difference contributes to survival in the low-risk score group.

At present, immune checkpoints including PD-L1, CTLA-4, LAG3, TIGIT and TIM-3, are thought to contribute to tumor immune escape. High expression of these immune checkpoints indicates poor prognosis; however,

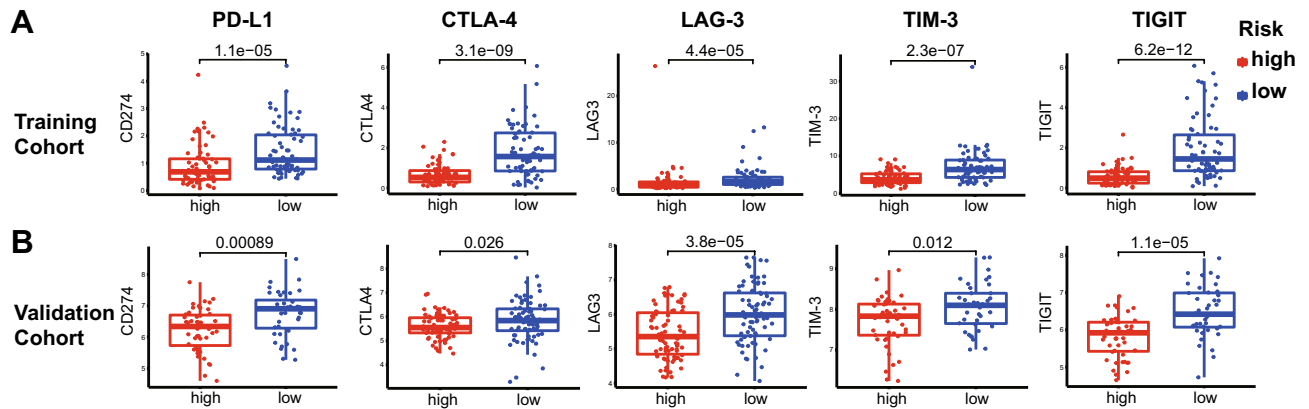


Figure 8. Expression of immune checkpoint genes in different risk groups. **(A)** The expression of immune checkpoint genes, including PD-L1, CTLA-4, LAG-3, TIM-3 and TIGIT, in the training cohort. **(B)** The expression of the same immune checkpoint genes in the validation cohort.

some different opinions have been published recently. Shang et al.⁶⁸ reported that HPBC patients with high PD-L1 expression of tumor cells had a better prognosis after neoadjuvant chemotherapy. Ni et al.⁶⁹ found that PD-L1 expression of TILs implied a good prognosis in breast cancer patients. Lee et al.⁷⁰ reported that high LAG-3 mRNA levels were associated with high levels of TILs in HPBC. High infiltration of TILs is a marker of favorable prognosis. In this study, PD-L1, LAG3, TIGIT, TIM-3 and CTLA-4 were upregulated in the low-risk group. We speculated that high expression of these immune checkpoints might be associated with TIL-mediated antitumor inflammatory responses⁷¹. However, our study indicated that multiple coinhibited signaling pathways might be involved in the tumorigenesis of HPBC patients with a low-risk score, and combination treatment with multiple immune checkpoint inhibitors may achieve better outcomes in these patients.

In our study, we established and validated a prognosis-related risk model composed of three survival-related genes that were strongly associated with immune infiltration. More importantly, we found that immune escape caused by multiple immune cosignaling pathways might underlie oncogenesis and tumor progression in some patients with HPBC. Hence, we speculated that combination regimens of multiple immune checkpoint inhibitors might achieve breakthroughs in the treatment of HPBC patients in the future. However, there were some limitations of our study. First, this study lacked validation with experimental data. Moreover, some analytical biases were inevitable due to the absence of some clinical data. Therefore, an analysis of clinical samples was necessary to consolidate the conclusions of this study.

Methods

Data acquisition and preprocessing. We obtained all data (including clinical characteristics and mRNA expression) of HPBC patients from the TCGA database (<https://portal.gdc.cancer.gov/>) and GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). The TCGA dataset was considered the training cohort, and the GSE20711, GSE45255, GSE162228 and GSE1456 GEO datasets were regarded as the validation cohorts. Patients with an OS of less than 1 month were excluded from the analysis. Thus, 132 HPBC patients in the TCGA and 148 HPBC patients in GEO were accepted for subsequent study. The procedures involving with the data collection and analyses abided by all provisions of the TCGA and GEO.

Assessment of the TIME in HPBC patients. The ESTIMATE algorithm is appropriate for assessing the profile of the TIME, which contains stromal cells and immune cells²⁰. A higher score indicates more significant infiltration of the corresponding component in the tumor tissue. The ssGSEA algorithm was applied to judge the level of immune cell infiltration and the immune response mechanism based on published data from Bindea et al.⁷². Thus, these analyses provided an overview of the tumor-infiltrating landscape.

Identification of DEGs and the analysis of potential functions. The DEGs between different groups were mapped by the “limma” R package. Genes with a p value < 0.05 were regarded as DEGs.

Functional GO and KEGG analyses⁷³ of the DEGs were performed utilizing the “clusterProfiler” package. Then, the potential mechanisms were elucidated according to FDR < 0.05. All the data were visualized by the “ggplot2” and “topGO” packages.

We carried out gene set enrichment analysis (GSEA) using GSEA software (V4.1.0) to determine whether DEGs exhibited significant and consistent differences between the high- and low-risk groups⁷⁴.

Establishment of a risk model with prognostic ability. The optimal survival-related genes in the training cohort were selected by LASSO analysis based on the “glmnet” R package⁷⁵. Tenfold cross-validation was used for filtering, and the λ value was obtained. Meanwhile, the coefficients of the selected genes were acquired. Then, the risk score of each sample was regarded as the summation of the product of each selected gene

and the corresponding coefficient. Thus, all patients were classified into the high-risk score or low-risk score group based on the respective median risk scores in both cohorts.

Subsequently, the survival difference between the two groups was analyzed and plotted by the “survival” and “survminer” R packages. The “survival ROC” R package was applied to determine the AUC value and plot the ROC curve. PCA and t-SNE were utilized to analyze the clustering of each sample in the model. Furthermore, the risk score was regarded as a prognosis-related variable, and the independent predictive power of this variable was assessed by univariate and multivariate analyses using the “survival” package.

Statistical analysis. All statistical analyses were performed using R software (version 4.1.1) (<https://www.r-project.org/>). The differences between variables in two groups were examined by using Wilcoxon’s test. Kaplan–Meier curves were utilized to assess the survival data. Independent prognostic factors were judged via univariate and multivariate Cox regression analyses. $p < 0.05$ was considered to indicate statistical significance ($*p < 0.05$; $**p < 0.01$; $***p < 0.001$).

Ethics approval and consent to participate. Not applicable. All data in this study are publicly available.

Data availability

All data obtained for this study can be found in the TCGA (<https://portal.gdc.cancer.gov/>) and Gene Expression Omnibus (GEO) repository (<https://www.ncbi.nlm.nih.gov/geo/>).

Received: 14 January 2022; Accepted: 25 May 2022

Published online: 03 June 2022

References

- Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424. <https://doi.org/10.3322/caac.21492> (2018).
- Advani, P. P., Crozier, J. A. & Perez, E. A. HER2 testing and its predictive utility in anti-HER2 breast cancer therapy. *Biomark. Med.* **9**, 35–49. <https://doi.org/10.2217/bmm.14.95> (2015).
- Pegram, M. D., Pauletti, G. & Slamon, D. J. HER-2/neu as a predictive marker of response to breast cancer therapy. *Breast Cancer Res. Treat.* **52**, 65–77. <https://doi.org/10.1023/a:1006111117877> (1998).
- Seshadri, R. *et al.* Clinical significance of HER-2/neu oncogene amplification in primary breast cancer. The South Australian Breast Cancer Study Group. *J. Clin. Oncol.* **11**, 1936–1942. <https://doi.org/10.1200/jco.1993.11.10.1936> (1993).
- Slamon, D. J. *et al.* Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* **244**, 707–712. <https://doi.org/10.1126/science.2470152> (1989).
- Ross, J. S. *et al.* The HER-2 receptor and breast cancer: Ten years of targeted anti-HER-2 therapy and personalized medicine. *Oncologist* **14**, 320–368. <https://doi.org/10.1634/theoncologist.2008-0230> (2009).
- Hurvitz, S. A. *et al.* Neoadjuvant trastuzumab, pertuzumab, and chemotherapy versus trastuzumab emtansine plus pertuzumab in patients with HER2-positive breast cancer (KRISTINE): A randomised, open-label, multicentre, phase 3 trial. *Lancet Oncol.* **19**, 115–126. [https://doi.org/10.1016/s1470-2045\(17\)30716-7](https://doi.org/10.1016/s1470-2045(17)30716-7) (2018).
- Reck, M. Pembrolizumab as first-line therapy for metastatic non-small-cell lung cancer. *Immunotherapy* **10**, 93–105. <https://doi.org/10.2217/imt-2017-0121> (2018).
- Ribas, A. *et al.* Pembrolizumab versus investigator-choice chemotherapy for ipilimumab-refractory melanoma (KEYNOTE-002): A randomised, controlled, phase 2 trial. *Lancet Oncol.* **16**, 908–918. [https://doi.org/10.1016/s1470-2045\(15\)00083-2](https://doi.org/10.1016/s1470-2045(15)00083-2) (2015).
- Shanbhag, S. & Ambinder, R. F. Hodgkin lymphoma: A review and update on recent progress. *CA Cancer J. Clin.* **68**, 116–132. <https://doi.org/10.3322/caac.21438> (2018).
- Kono, K., Nakajima, S. & Mimura, K. Current status of immune checkpoint inhibitors for gastric cancer. *Gastr. Cancer* **23**, 565–578. <https://doi.org/10.1007/s10120-020-01090-4> (2020).
- Yamamoto, S. & Kato, K. Pembrolizumab for the treatment of esophageal cancer. *Expert Opin. Biol. Ther.* <https://doi.org/10.1080/14712598.2020.1792881> (2020).
- Schmid, P. *et al.* Atezolizumab and nab-paclitaxel in advanced triple-negative breast cancer. *N. Engl. J. Med.* **379**, 2108–2121. <https://doi.org/10.1056/NEJMoa1809615> (2018).
- Kodumudi, K. N. *et al.* Sequential anti-PD1 therapy following dendritic cell vaccination improves survival in a HER2 mammary carcinoma model and identifies a critical role for CD4 T cells in mediating the response. *Front. Immunol.* **10**, 1939. <https://doi.org/10.3389/fimmu.2019.01939> (2019).
- Loi, S. *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.* **31**, 860–867. <https://doi.org/10.1200/jco.2011.41.0902> (2013).
- Hou, Y. *et al.* PD-L1 expression and CD8-positive T cells are associated with favorable survival in HER2-positive invasive breast cancer. *Breast J.* **24**, 911–919. <https://doi.org/10.1111/tbj.13112> (2018).
- Wang, L. *et al.* AMPD1 is associated with the immune response and serves as a prognostic marker in HER2-positive breast cancer. *Front. Oncol.* **11**, 749135. <https://doi.org/10.3389/fonc.2021.749135> (2021).
- Li, H. *et al.* From cellular infiltration assessment to a functional gene set-based prognostic model for breast cancer. *Front. Immunol.* **12**, 751530. <https://doi.org/10.3389/fimmu.2021.751530> (2021).
- Emens, L. A. *et al.* Trastuzumab emtansine plus atezolizumab versus trastuzumab emtansine plus placebo in previously treated, HER2-positive advanced breast cancer (KATE2): A phase 2, multicentre, randomised, double-blind trial. *Lancet Oncol.* **21**, 1283–1295. [https://doi.org/10.1016/s1470-2045\(20\)30465-4](https://doi.org/10.1016/s1470-2045(20)30465-4) (2020).
- Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612. <https://doi.org/10.1038/ncomms3612> (2013).
- Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457. <https://doi.org/10.1038/nmeth.3337> (2015).
- Luo, D. *et al.* The nomogram based on the 6-lncRNA model can promote the prognosis prediction of patients with breast invasive carcinoma. *Sci. Rep.* **11**, 20863. <https://doi.org/10.1038/s41598-021-00364-w> (2021).
- Liu, Y. *et al.* A novel prognostic signature of immune-related lncRNA pairs in lung adenocarcinoma. *Sci. Rep.* **11**, 16794. <https://doi.org/10.1038/s41598-021-96236-4> (2021).
- Bhattacharya, S. *et al.* ImmPort: Disseminating data to the public for the future of immunology. *Immunol. Res.* **58**, 234–239. <https://doi.org/10.1007/s12026-014-8516-1> (2014).

25. Oguma, T., Asano, K. & Ishizaka, A. Role of prostaglandin D(2) and its receptors in the pathophysiology of asthma. *Allergol. Int.* **57**, 307–312. <https://doi.org/10.2332/allergolint.08-RAI-0033> (2008).
26. Kagawa, S. *et al.* Role of prostaglandin D2 receptor CRTH2 in sustained eosinophil accumulation in the airways of mice with chronic asthma. *Int. Arch. Allergy Immunol.* **155**(Suppl 1), 6–11. <https://doi.org/10.1159/000327257> (2011).
27. Reinert, T. *et al.* Comprehensive genome methylation analysis in bladder cancer: Identification and validation of novel methylated genes and application of these as urinary tumor markers. *Clin. Cancer Res.* **17**, 5582–5592. <https://doi.org/10.1158/1078-0432.Ccr-10-2659> (2011).
28. Spisák, S. *et al.* Genome-wide screening of genes regulated by DNA methylation in colon cancer development. *PLoS ONE* **7**, e46215. <https://doi.org/10.1371/journal.pone.0046215> (2012).
29. Pradhan, M. P., Desai, A. & Palakal, M. J. Systems biology approach to stage-wise characterization of epigenetic genes in lung adenocarcinoma. *BMC Syst. Biol.* **7**, 141. <https://doi.org/10.1186/1752-0509-7-141> (2013).
30. Chen, Y. C. *et al.* Methyloomics analysis identifies epigenetically silenced genes and implies an activation of β -catenin signaling in cervical cancer. *Int. J. Cancer* **135**, 117–127. <https://doi.org/10.1002/ijc.28658> (2014).
31. Lai, H. C. *et al.* DNA methylation as a biomarker for the detection of hidden carcinoma in endometrial atypical hyperplasia. *Gynecol. Oncol.* **135**, 552–559. <https://doi.org/10.1016/j.ygyno.2014.10.018> (2014).
32. Kim, H. J. *et al.* Whole genome MBD-seq and RRBS analyses reveal that hypermethylation of gastrointestinal hormone receptors is associated with gastric carcinogenesis. *Exp. Mol. Med.* **50**, 1–14. <https://doi.org/10.1038/s12276-018-0179-x> (2018).
33. Mollereau, C. *et al.* Structure, tissue distribution, and chromosomal localization of the prepronociceptin gene. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8666–8670. <https://doi.org/10.1073/pnas.93.16.8666> (1996).
34. Buzas, B., Rosenberger, J., Kim, K. W. & Cox, B. M. Inflammatory mediators increase the expression of nociceptin/orphanin FQ in rat astrocytes in culture. *Glia* **39**, 237–246. <https://doi.org/10.1002/glia.10106> (2002).
35. Wang, K. *et al.* Nociceptin receptor is overexpressed in non-small cell lung cancer and predicts poor prognosis. *Front. Oncol.* **9**, 235. <https://doi.org/10.3389/fonc.2019.00235> (2019).
36. Kadhim, S., Bird, M. F. & Lambert, D. G. N/OFQ-nop system in peripheral and central immunomodulation. *Handb. Exp. Pharmacol.* **254**, 297–311. https://doi.org/10.1007/164_2018_203 (2019).
37. Spaziano, G. *et al.* Nociceptin reduces the inflammatory immune microenvironment in a conventional murine model of airway hyperresponsiveness. *Clin. Exp. Allergy* **47**, 208–216. <https://doi.org/10.1111/cea.12808> (2017).
38. Patel, V. P. *et al.* Molecular and functional characterization of two novel human C–C chemokines as inhibitors of two distinct classes of myeloid progenitors. *J. Exp. Med.* **185**, 1163–1172. <https://doi.org/10.1084/jem.185.7.1163> (1997).
39. Nardelli, B. *et al.* Dendritic cells and MPIF-1: Chemotactic activity and inhibition of endogenous chemokine production by IFN- γ and CD40 ligation. *J. Leukoc. Biol.* **65**, 822–828. <https://doi.org/10.1002/jlb.65.6.822> (1999).
40. Arruda-Silva, F. *et al.* Human neutrophils produce CCL23 in response to various TLR-agonists and TNF α . *Front. Cell. Infect. Microbiol.* **7**, 176. <https://doi.org/10.3389/fcimb.2017.00176> (2017).
41. Kullmann, T., Barta, L., Csiszér, E., Antus, B. & Horváth, I. Differential cytokine pattern in the exhaled breath of patients with lung cancer. *Pathol. Oncol. Res.* **14**, 481–483. <https://doi.org/10.1007/s12253-008-9046-8> (2008).
42. Miyoshi, H. *et al.* Expression profiles of 507 proteins from a biotin label-based antibody array in human colorectal cancer. *Oncol. Rep.* **31**, 1277–1281. <https://doi.org/10.3892/or.2013.2935> (2014).
43. McLean, M. H. *et al.* The inflammatory microenvironment in colorectal neoplasia. *PLoS ONE* **6**, e15366. <https://doi.org/10.1371/journal.pone.0015366> (2011).
44. Krishnan, V. *et al.* Omental macrophages secrete chemokine ligands that promote ovarian cancer colonization of the omentum via CCR1. *Commun. Biol.* **3**, 524. <https://doi.org/10.1038/s42003-020-01246-z> (2020).
45. Franzén, B. *et al.* Multiplex immune protein profiling of fine-needle aspirates from patients with non-small-cell lung cancer reveals signatures associated with PD-L1 expression and tumor stage. *Mol. Oncol.* **15**, 2941–2957. <https://doi.org/10.1002/1878-0261.12952> (2021).
46. Meng, J. *et al.* CCL23 suppresses liver cancer progression through the CCR1/AKT/ESR1 feedback loop. *Cancer Sci.* **112**, 3099–3110. <https://doi.org/10.1111/cas.14995> (2021).
47. Palucka, K. & Banchereau, J. Cancer immunotherapy via dendritic cells. *Nat. Rev. Cancer* **12**, 265–277. <https://doi.org/10.1038/nrc3258> (2012).
48. Jhunjhunwala, S., Hammer, C. & Delamarre, L. Antigen presentation in cancer: Insights into tumour immunogenicity and immune evasion. *Nat. Rev. Cancer* **21**, 298–312. <https://doi.org/10.1038/s41568-021-00339-z> (2021).
49. Marzo, A. L. *et al.* Tumor-specific CD4+ T cells have a major “post-licensing” role in CTL mediated anti-tumor immunity. *J. Immunol.* **165**, 6047–6055. <https://doi.org/10.4049/jimmunol.165.11.6047> (2000).
50. Tzankov, A. *et al.* Correlation of high numbers of intratumoral FOXP3+ regulatory T cells with improved survival in germinal center-like diffuse large B-cell lymphoma, follicular lymphoma and classical Hodgkin’s lymphoma. *Haematologica* **93**, 193–200. <https://doi.org/10.3324/haematol.11702> (2008).
51. Alvaro, T. *et al.* Outcome in Hodgkin’s lymphoma can be predicted from the presence of accompanying cytotoxic and regulatory T cells. *Clin. Cancer Res.* **11**, 1467–1473. <https://doi.org/10.1158/1078-0432.Ccr-04-1869> (2005).
52. Carreras, J. *et al.* High numbers of tumor-infiltrating FOXP3-positive regulatory T cells are associated with improved overall survival in follicular lymphoma. *Blood* **108**, 2957–2964. <https://doi.org/10.1182/blood-2006-04-018218> (2006).
53. Badoual, C. *et al.* Prognostic value of tumor-infiltrating CD4+ T-cell subpopulations in head and neck cancers. *Clin. Cancer Res.* **12**, 465–472. <https://doi.org/10.1158/1078-0432.Ccr-05-1886> (2006).
54. Müller, P. *et al.* Trastuzumab emtansine (T-DM1) renders HER2+ breast cancer highly susceptible to CTLA-4/PD-1 blockade. *Sci. Transl. Med.* **7**, 315. <https://doi.org/10.1126/scitranslmed.aac4925> (2015).
55. Miyara, M. & Sakaguchi, S. Human FoxP3(+)/CD4(+) regulatory T cells: Their knowns and unknowns. *Immunol. Cell Biol.* **89**, 346–351. <https://doi.org/10.1038/icb.2010.137> (2011).
56. Frydrychowicz, M., Boruczkowski, M., Kolecka-Bednarczyk, A. & Dworacki, G. The dual role of Treg in cancer. *Scand. J. Immunol.* <https://doi.org/10.1111/sji.12615> (2017).
57. Lu, J. *et al.* Breast cancer chemo-immunotherapy through liposomal delivery of an immunogenic cell death stimulus plus interference in the IDO-1 pathway. *ACS Nano* **12**, 11041–11061. <https://doi.org/10.1021/acsnano.8b05189> (2018).
58. Ivars Rubio, A. *et al.* Neutrophil-lymphocyte ratio in metastatic breast cancer is not an independent predictor of survival, but depends on other variables. *Sci. Rep.* **9**, 16979. <https://doi.org/10.1038/s41598-019-53606-3> (2019).
59. McCoy, M. J., Lake, R. A., van der Most, R. G., Dick, I. M. & Nowak, A. K. Post-chemotherapy T-cell recovery is a marker of improved survival in patients with advanced thoracic malignancies. *Br. J. Cancer* **107**, 1107–1115. <https://doi.org/10.1038/bjc.2012.362> (2012).
60. Bigot, F. *et al.* Prospective validation of a prognostic score for patients in immunotherapy phase I trials: The Gustave Roussy Immune Score (GRIm-Score). *Eur. J. Cancer* **84**, 212–218. <https://doi.org/10.1016/j.ejca.2017.07.027> (2017).
61. Bretscher, P. & Cohn, M. A theory of self-nonsel self discrimination. *Science* **169**, 1042–1049. <https://doi.org/10.1126/science.169.3950.1042> (1970).
62. Chen, L. & Flies, D. B. Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nat. Rev. Immunol.* **13**, 227–242. <https://doi.org/10.1038/nri3405> (2013).

63. Gertner-Dardenne, J. *et al.* Bromohydrin pyrophosphate enhances antibody-dependent cell-mediated cytotoxicity induced by therapeutic antibodies. *Blood* **113**, 4875–4884. <https://doi.org/10.1182/blood-2008-08-172296> (2009).
64. Anikeeva, N. *et al.* Integrin receptors on tumor cells facilitate NK cell-mediated antibody-dependent cytotoxicity. *Eur. J. Immunol.* **44**, 2331–2339. <https://doi.org/10.1002/eji.201344179> (2014).
65. Balkwill, F. Cancer and the chemokine network. *Nat. Rev. Cancer* **4**, 540–550. <https://doi.org/10.1038/nrc1388> (2004).
66. Turley, S. J., Cremasco, V. & Astarita, J. L. Immunological hallmarks of stromal cells in the tumour microenvironment. *Nat. Rev. Immunol.* **15**, 669–682. <https://doi.org/10.1038/nri3902> (2015).
67. Motakis, E., Ivshina, A. V. & Kuznetsov, V. A. Data-driven approach to predict survival of cancer patients: Estimation of microarray genes' prediction significance by Cox proportional hazard regression model. *IEEE Eng. Med. Biol. Mag.* **28**, 58–66. <https://doi.org/10.1109/memb.2009.932937> (2009).
68. Shang, M. *et al.* The therapeutic effectiveness of neoadjuvant trastuzumab plus chemotherapy for HER2-positive breast cancer can be predicted by tumor-infiltrating lymphocytes and PD-L1 expression. *Front. Oncol.* **11**, 706606. <https://doi.org/10.3389/fonc.2021.706606> (2021).
69. Ni, Y. *et al.* Combining analysis of tumor-infiltrating lymphocytes (TIL) and PD-L1 refined the prognostication of breast cancer subtypes. *Oncologist* **27**, e313–e327. <https://doi.org/10.1093/oncolo/oyab063> (2022).
70. Lee, S. *et al.* Lymphocyte-activating gene-3 expression is associated with tumor-infiltrating lymphocyte levels in HER2-positive breast cancers. *Medicine* **100**, e28057. <https://doi.org/10.1097/md.00000000000028057> (2021).
71. AiErken, N. *et al.* High PD-L1 expression is closely associated with tumor-infiltrating lymphocytes and leads to good clinical outcomes in chinese triple negative breast cancer patients. *Int. J. Biol. Sci.* **13**, 1172–1179. <https://doi.org/10.7150/ijbs.20868> (2017).
72. Bindea, G. *et al.* Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* **39**, 782–795. <https://doi.org/10.1016/j.immuni.2013.10.003> (2013).
73. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
74. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550. <https://doi.org/10.1073/pnas.0506580102> (2005).
75. Wang, H., Lengerich, B. J., Aragam, B. & Xing, E. P. Precision Lasso: Accounting for correlations and linear dependencies in high-dimensional genomic data. *Bioinformatics* **35**, 1181–1187. <https://doi.org/10.1093/bioinformatics/bty750> (2019).

Acknowledgements

This work was supported by the Science and Technology Bureau of Quanzhou (Grant No. 2020N041s, Grant No. 2018N015S), the Natural Science Foundation of Fujian Province (Grant No. 2020J01217), the Second Affiliated Hospital of Fujian Medical University (Grant No. BS202111) and WU JIEPING Medical Foundation (Grant No. 320.6750.2021-10-9). All the funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

D.Q.F. and J.Q.L. conceived and designed the experiments. D.Q.F., J.Q.L. and A.Y.Z. analyzed and acquired data. D.Q.F. interpreted the results and prepared Figure. J.Q.L. drafted the manuscript. All authors revised and reviewed this work, and all authors gave their final approval of the submitted manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-13499-1>.

Correspondence and requests for materials should be addressed to D.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022