Check for updates

Immune Landscape of Invasive Ductal Carcinoma Tumor Microenvironment Identifies a Prognostic and Immunotherapeutically Relevant Gene Signature

Xuanwen Bao [1,2*†], Run Shi [3†], Kai Zhang [4], Shan Xin [1], Xin Li [5], Yanbo Zhao [4] and Yanfang Wang [6*]

[1] Helmholtz Center Munich, German Research Center for Environmental Health, Institute of Radiation Biology, Munich, Germany, [2] Technical University Munich, Munich, Germany, [3] Department of Radiation Oncology, University Hospital, Ludwig Maximilian University of Munich, Munich, Germany, [4] Department of Cardiology, School of Medicine, Sir Run Run Shaw Hospital, Zhejiang University, Hangzhou, China, [5] College of Life Sciences, Nanjing University, Nanjing, China, [6] Department of Pharmacy, Pharmaceutical Biotechnology, Center of Nanoscience (CeNS), Ludwig Maximilian University of Munich, Munich, Germany

Background: Invasive ductal carcinoma (IDC) is a clinically and molecularly distinct disease. Tumor microenvironment (TME) immune phenotypes play crucial roles in predicting clinical outcomes and therapeutic efficacy.

Method: In this study, we depict the immune landscape of IDC by using transcriptome profiling and clinical characteristics retrieved from The Cancer Genome Atlas (TCGA) data portal. Immune cell infiltration was evaluated via single-sample gene set enrichment (ssGSEA) analysis and systematically correlated with genomic characteristics and clinicopathological features of IDC patients. Furthermore, an immune signature was constructed using the least absolute shrinkage and selection operator (LASSO) Cox regression algorithm. A random forest algorithm was applied to identify the most important somatic gene mutations associated with the constructed immune signature. A nomogram that integrated clinicopathological features with the immune signature to predict survival probability was constructed by multivariate Cox regression.

Results: The IDC were clustered into low immune infiltration, intermediate immune infiltration, and high immune infiltration by the immune landscape. The high infiltration group had a favorable survival probability compared with that of the low infiltration group. The low-risk score subtype identified by the immune signature was characterized by T cell-mediated immune activation. Additionally, activation of the interferon-α response, interferon-γ response, and TNF-α signaling via the NFκB pathway was observed in the low-risk score subtype, which indicated T cell activation and may be responsible for significantly favorable outcomes in IDC patients. A random forest algorithm identified the most important somatic gene mutations associated with the constructed immune signature. Furthermore, a nomogram that integrated clinicopathological features with the immune signature to predict survival probability was constructed, revealing that

the immune signature was an independent prognostic biomarker. Finally, the relationship of VEGFA, PD1, PDL-1, and CTLA-4 expression with the immune infiltration landscape and the immune signature was analyzed to interpret the responses of IDC patients to immunotherapy.

Conclusion: Taken together, we performed a comprehensive evaluation of the immune landscape of IDC and constructed an immune signature related to the immune landscape. This analysis of TME immune infiltration landscape has shed light on how IDC respond to immunotherapy and may guide the development of novel drug combination strategies.

INTRODUCTION

Invasive ductal carcinoma (IDC) is a clinically and molecularly distinct disease. IDCs are typically of high histologic grade and high mitotic index. HER2 overexpression or amplification is detected in 20% of these tumors (1). IDC tends to metastasize to bone, liver, and lung, whereas invasive lobular carcinoma (ILC) has a higher tendency to metastasize in gastrointestinal and genital tracts, serosal cavities, and meninges (2). IDCs usually form glandular structures in contrast to the small clusters formed by ILCs. The loss of CDH1 leads to the discohesive morphology in ILCs, whereas IDCs maintain intact cell adhesion (3). Furthermore, the frequency of recurrently mutated genes and recurrent copy-number alterations often differs significantly between IDCs and ILCs (3). These features are generally associated with a poor prognosis. Taken together, these differences suggest that ILCs and IDCs are distinct cancer types and progress along different pathways.

Genetic and epigenetic changes contribute to the progression of tumor progression and recurrence in different cancer types. However, accumulated evidence indicates that the tumor microenvironment (TME) has clinicopathological significance in predicting survival outcomes and assessing therapeutic efficacy factors (4, 5). TME cells constitute a vital element of cellular and non-cellular components in the tumoural niche, including extracellular matrix and cellular components, such as fibroblasts, adipose cells, immune-inflammatory cells, and neuroendocrine cells. Previous studies have revealed that immune cells in the TME modulate cancer progression and are attractive therapeutic targets (6, 7). To date, the comprehensive landscape of immune cells infiltrating the TME of IDCs has not yet been elucidated. We propose that IDCs have a distinct immune landscape and that the immune landscape might lead to different prognoses and treatment responses. In this study, by applying several computational algorithms, we estimated the abundance of immune cells in the TME of IDCs and analyzed the correlation of the immune landscape with genomic characteristics and pathological features of IDCs. Furthermore, we built an immune signature, which is a robust prognostic biomarker and predictive factor for the response to immunotherapy.

METHODS

Data Download

TCGA RNA-seq datasets and clinical data for IDCs were downloaded by UCSC Xena browser (https://xenabrowser.net/). GSE20685 and GSE86948 were downloaded from the Gene Expression Omnibus (GEO) database.

Implementation of Single-Sample Gene Set Enrichment Analysis (ssGSEA)

We obtained the marker gene set for immune cell types from Bindea et al. (8). MDSC gene set was imported from MSIGDB gmt file from Broad institute. We used the ssGSEA program to derive the enrichment scores of each immune-related term. In brief, the infiltration levels of immune cell types were quantified by ssGSEA in the R package gsva (9). The ssGSEA applies gene signatures expressed by immune cell populations to individual cancer samples. The computational approach used in our study included immune cells types that are involved in innate immunity and adaptive immunity. Tumors with qualitatively different immune cell infiltration patterns were grouped using hierarchical agglomerative clustering (based on Euclidean distance and Ward's linkage).

The T cell infiltration score (TIS) was defined as the average of the standardized values for $CD8^+$ T, central memory $CD4^+$ T, effector memory $CD4^+$ T, central memory $CD8^+$ T, effector memory $CD8^+$ T, Th1, Th2, Th17, and Treg cells. The obtained cytotoxic activity scores (CYT) score was calculated by the geometrical mean of PRF1 and GZMA (10). The $CD8^+$ T/Treg ratio was the digital ratio of the ssGSEA scores for these two cell types. The correlation between risk score and immune cell ssGSEA score was calculated by Pearson correlation.

LASSO Regularization

LASSO (least absolute shrinkage and selection operator) is an important regularization in many regression analysis methods (e.g., COX regression and logistic regression) (10–12). The idea

behind LASSO is that a L1-norm is used to penalize the weight of the model parameters. Assuming a model has a set of parameters, the LASSO regularization can be defined as:

$$\lambda \cdot \sum_{i=0}^{n} \|w_i\|_1$$

It can also be expressed as a constraint to the targeted objective function:

$$\sum \left\| Y - Y^* \right\|_2 , \quad s.t. \quad \|w_i\|_1 < t$$

An important property of the LASSO regularization term is that it can force the parameter values to be 0, thus generating a sparse parameter space, which is a desirable characteristic for feature selection. In our analysis, 19 genes which were highly associated with OS were used as the input. QRSL1, TIMM8A, IGHA1, BATF, KLRB1, SPIB, and FLT3LG were picked after the penalizing process. A risk score (RS) formula was established by including individual normalized gene expression values weighted by their LASSO Cox coefficients:

$$\sum_i Coefficient(mRNA_i) \times Expression(mRNA_i)$$

Risk score = (0.210 * expression level of QRSL1) + (0.092 * expression level of TIMM8A) + (−0.046 * expression level of IGHA1) + (−0.066 * expression level of BATF) + (−0.110 * expression level of KLRB1) + (−0.139 * expression level of SPIB) + (−0.262 * expression level of FLT3LG).

Differentially Expressed Gene (DEG) Analysis

DEG analysis was performed by the Limma package (13). The samples were separated into a high-risk score group and a low-risk score group. An empirical Bayesian approach was applied to estimate the gene expression changes using moderated t-tests. The Q-value (adjusted p-value) for multiple testing was calculated using the Benjamini-Hochberg correction. The DEGs were defined as genes with a Q-value < 0.05. The clusterProfiler R package was applied for the GO analysis (14). GSEA was applied with the GSEA software.

Co-expression Gene Network Based on RNA-seq Data

The Weighted correlation network analysis (WGCNA) was used to construct the gene co-expression network (15, 16). The co-expression similarity $s_{i,j}$ was defined as the absolute value of the correlation coefficient between the profiles of nodes i and j:

$$s_{i,j} = |cor(x_i, x_j)|$$

where, x_i and x_j are expression values of for genes i and j, and $s_{i,j}$ represent Pearson's correlation coefficients of genes i and j, respectively.

A weighed network adjacency was defined by raising the co-expression similarity to a power β:

$$a_{i,j} = s_{i,j}^{\beta}$$

with $\beta \geq 1$. We selected the power of $\beta = 5$ and scale-free $R^2 = 0.95$ as the soft-thresholding parameters to ensure a signed scale-free co-expression gene network. Briefly, network construction, module detection, feature selection, calculations of topological properties, data simulation, and visualization were performed. Modules were identified via hierarchical clustering of the weighting coefficient matrix. The module membership of node i in module q was defined as:

$$K_{cor,i}^{(q)} := cor(x_i, E^{(q)})$$

where, x_i is the profile of node i, and $E(q)$ is the module eigengene (the first principal component of a given module) of module q. The module membership measure $K_{cor,i}^{(q)}$, lies in $[-1, 1]$ and specifies how close node i is to module q, $q = 1, \cdots, Q$.

By evaluating the correlations between the immune infiltration status, immune signature of IDCs and the module membership of each module, a brown module was selected for further analysis.

Data Processing and Integration

The mutation datasets were download by R package TCGAbiolinks. The expression profiles of the most powerful prognostic features (QRSL1, TIMM8A, IGHA1, BATF, KLRB1, SPIB, and FLT3LG) were extracted from the whole transcriptome datasets. The immune infiltration status was calculated by the deconvolution algorithm and grouped using hierarchical agglomerative clustering. We summarized the clinic datasets, mutation datasets, expression profiles and immune infiltration status into an integrated dataset (**Supplementary File 1**).

Statistical Analysis

A random forest algorithm was applied to find the most important somatic mutation associated with the immune signature. Survival outcome analysis modeled the results in reference to the patient OS and RFS. P-values and Hazard ratios were obtained from univariate Cox proportional-hazards regression models using the R package survival. Multivariate Cox regression was used to calculate the coefficients in the nomogram. The nomogram was plotted by the rms package. The time-dependent AUC value was calculated by the survivalROC package.
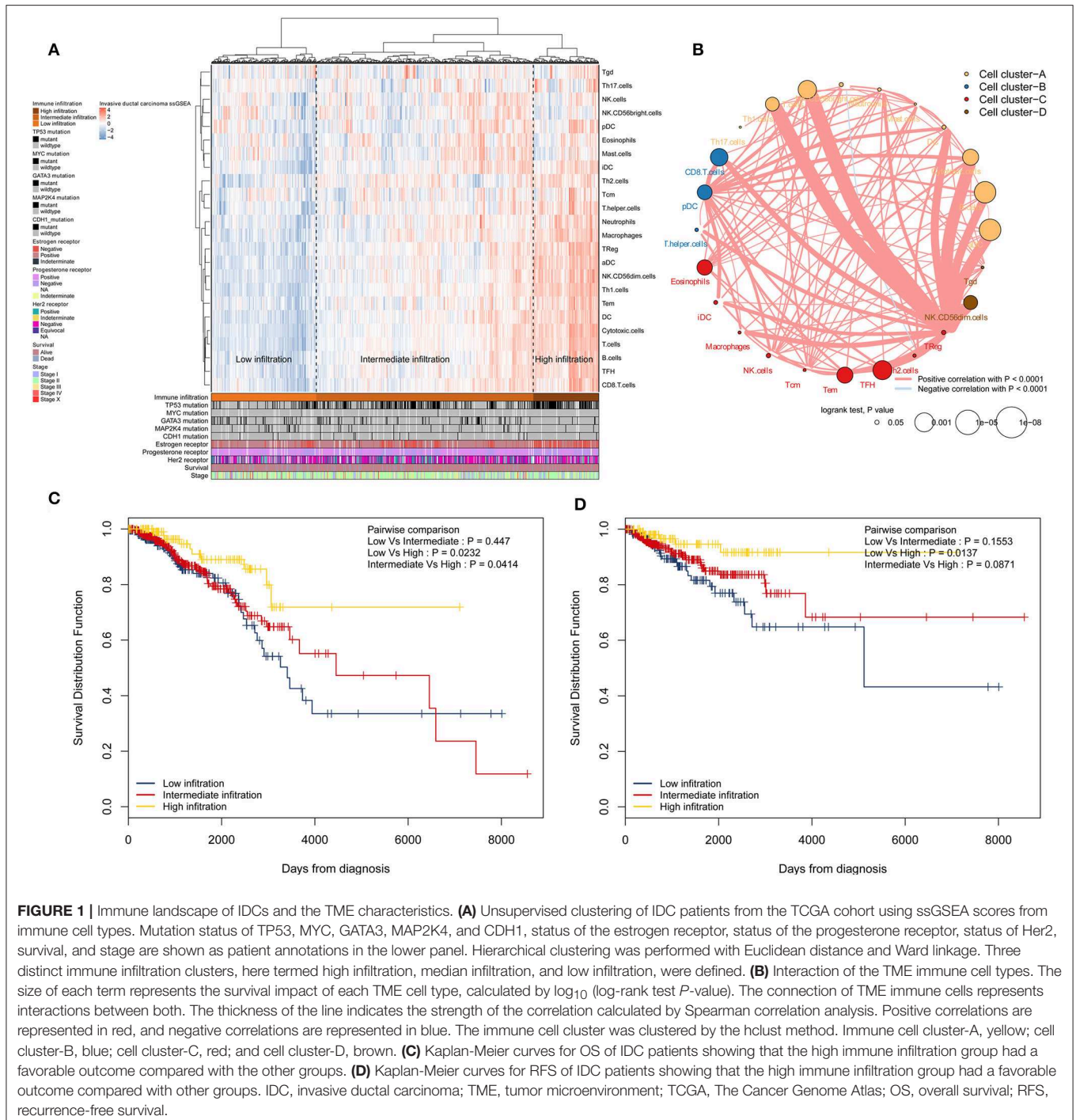
RESULTS

Immune Phenotype Landscape in the TME of IDC

Immune cell populations modulate diverse immune responses and lead to anti-tumour effects by infiltrating the IDC TME. The immune cell infiltration status was assessed by applying the ssGSEA approach to the transcriptomes of IDCs. Twenty-four immune-related terms were incorporated to deconvolve

the abundance of diverse immune cell types in IDCs. The IDCs were clustered into 3 clusters (low infiltration: 208; intermediate infiltration: 430; and high infiltration: 130) in terms of immune infiltration by applying an unsupervised hierarchical clustering algorithm (**Figure 1A**). By applying hierarchical cluster analysis and K-means clustering analysis, we constructed a TME cell network, depicting a comprehensive landscape of tumor-immune cell interactions and their effects

on the OS of patients with IDC (**Figure 1B** and **Figures S1, S2**). The TME immune cells were clustered into 4 clusters, and the correlation among different immune cell types is shown in **Figure 1B**. The association of OS and RFS with different clusters of IDCs was analyzed by a pairwise log-rank test. The results indicated that the high infiltration group had a favorable survival probability compared with that of the low infiltration group (**Figures 1C,D**).

FIGURE 1 | Immune landscape of IDCs and the TME characteristics. **(A)** Unsupervised clustering of IDC patients from the TCGA cohort using ssGSEA scores from immune cell types. Mutation status of TP53, MYC, GATA3, MAP2K4, and CDH1, status of the estrogen receptor, status of the progesterone receptor, status of Her2, survival, and stage are shown as patient annotations in the lower panel. Hierarchical clustering was performed with Euclidean distance and Ward linkage. Three distinct immune infiltration clusters, here termed high infiltration, median infiltration, and low infiltration, were defined. **(B)** Interaction of the TME immune cell types. The size of each term represents the survival impact of each TME cell type, calculated by $\log_{10}$ (log-rank test P-value). The connection of TME immune cells represents interactions between both. The thickness of the line indicates the strength of the correlation calculated by Spearman correlation analysis. Positive correlations are represented in red, and negative correlations are represented in blue. The immune cell cluster was clustered by the hclust method. Immune cell cluster-A, yellow; cell cluster-B, blue; cell cluster-C, red; and cell cluster-D, brown. **(C)** Kaplan-Meier curves for OS of IDC patients showing that the high immune infiltration group had a favorable outcome compared with the other groups. **(D)** Kaplan-Meier curves for RFS of IDC patients showing that the high immune infiltration group had a favorable outcome compared with other groups. IDC, invasive ductal carcinoma; TME, tumor microenvironment; TCGA, The Cancer Genome Atlas; OS, overall survival; RFS, recurrence-free survival.

Construction of the Immune Signature

A total of 413 genes were involved in the 24 immune-related terms. We applied the univariate COX regression based on the survival datasets of patients with IDC and the expression profiles of the 413 genes. The 19 most significant genes were selected with the criteria of a p-value < 0.0005 (**Figure 2A**). The expression profiles of the 19 genes are shown in **Figure 2B**. LASSO Cox regression was performed on the 19 genes to identify the most important features in terms of predicting the survival of IDC patients (**Figures 2C–E**). By forcing the sum of the absolute value of the regression coefficients to be less than a fixed value, certain coefficients were reduced to exactly zero, and the most powerful prognostic features (QRSL1, TIMM8A, IGHA1, BATF, KLRB1, SPIB, and FLT3LG) were identified with relative regression coefficients. Cross-validation was applied to prevent over-fitting. A 7-gene immune signature was constructed according to the individual coefficients of the genes. Then, we calculated the risk score for each IDC patient and ranked them (**Figure 2F**). **Figure 2G** shows the survival overview in the IDC patients. A heatmap showed that patients in the high-risk group tended to have increased QRSL1 and TIMM8A expression levels, as well as decreased expression levels of IGHA1, BATF, KLRB1, SPIB, and FLT3LG (**Figure 2H**). The Kaplan-Meier curve and Cox regression suggested that patients with high risk scores had significantly worse OS and RFS than those with low risk scores (HR $= 2.94$, $p < 0.0001$ and HR $= 2.28$, $p = 0.001$, respectively) (**Figures 2I,J**). The effect of the seven genes on the OS and RFS of IDC patients is shown in **Figures S3**, **S4**, respectively. To confirm our findings in the IDC cohort, we validated the prognostic function of the immune signature in two independent GEO cohorts (GSE20685 and GSE86948). The risk score was calculated for each patient by using the same formula as in the IDC cohort. The GSE20685 and GSE86948 cohorts were used to predict the OS of BRCA patients based on our immune signature model. Consistent with our previous findings, the Kaplan-Meier curve suggested a significantly better overall survival in the low-risk group than in the high-risk group (**Figures S5A,B**).

The Low Risk Score Was Associated With Active Infiltration Status and High Cytotoxic Potential

High infiltration status showed a lower risk score than the intermediate infiltration status and low infiltration status showed (**Figure 3A**). Similarly, patients with a low risk score had a higher proportion of high immune infiltration than patients with a high risk score (**Figure 3B**). The presence of high immune infiltration in patients was linked to a low risk score and was associated with a favorable outcome (**Figure 3C**). To compare cytotoxic function with the immune landscape and immune signature that we constructed, the associated signatures were identified for each patient. IDCs with high infiltration status and low risk score were associated with increased levels of immune activation. The TIS ($p < 0.0001$ and $p < 0.0001$, respectively) (**Figures 3D,G**), interferon-γ signature ($p < 0.0001$ and $p < 0.0001$, respectively) (**Figures 3E,H**), and CYT ($p < 0.0001$ and

$p < 0.0001$, respectively) (**Figures 3F,I**) were increased in the low-risk score group and high infiltration group. The ssGSEA score of DCs was higher in the low-risk score group than in the high-risk score group. The Kaplan-Meier curve showed that in the low-risk score group, the ssGSEA score of DC cells affected survival but did not affect the high-risk score group (**Figures S6A–C**). Furthermore, the correlation between MDSCs and risk score was analyzed (**Figure S7A**). The ssGSEA score for MDSCs was positively associated with the OS of IDC patients in whole cohorts ($p = 0.017$) (**Figure S7B**). When we stratified the patients into low-risk score and high-risk score groups, the ssGSEA score of MDSCs showed opposite association with the survival of IDC patients (HR $= 2.42$ and 0.63, respectively) (**Figures S7C,D**). These data indicate that compared with high-risk score tumors, low-risk score tumors have a distinct immune phenotype, characterized by increased immune infiltration and increased levels of immune activation.

The Low-Risk Score Was Associated With Increased T Cell Infiltration

The association of risk score and immune-related cells was analyzed by Pearson correlation. Cytotoxic cells, CD8[+] T cells, T cells and the 6 other most significant immune-related cell types are shown in **Figure 4**. A high level of correlation was found between the risk score and the T cell-mediated immune response. The ssGSEA scores of 24 immune-related terms in the low, intermediate, and high immune status and low- and high-risk score groups are shown in **Figures S8A,C**. The p-value and difference in the mean ssGSEA score from the high- and low-infiltration status and low- and high-risk score groups are shown in **Figures S8B,D**. The proportions of low, intermediate, and high immune infiltration status in different pathological subtypes and different AJCC stages of IDC are shown in **Figures S8E,F**. The triple-negative subtype of IDCs had a higher proportion of high infiltration status IDCs than other pathological subtypes, indicating an active immune response in triple-negative IDCs. The risk score distribution in different pathological subtypes and different AJCC stages of IDC are shown in **Figures S8G,H**. The luminal A subtype had a lower risk score than the other pathological subtypes.

Functional Annotation and WGNCA of the Transcriptomes of IDC Patients

To identify the underlying biological characteristics of the constructed immune signature, DEG analysis was performed based on the high-risk score group and low-risk score group. The heatmap depicts the significant DEGs between the two groups (**Figure 5A**). The GO analysis indicated that T cell activation, positive regulation of leukocyte cell-cell adhesion, and regulation of lymphocyte activation were the most significantly enriched biological processes between the high-risk score group and the low-risk score group (**Figure 5B**). The GSEA results showed that allograft rejection, IL-6/JAK/STAT3 signaling, the inflammatory response, interferon-α response, interferon-γ response, and TNF-α signaling via the NFκB pathway were the most predominantly upregulated pathways in the low-risk score group. In contrast, the E2F targets, G2M checkpoints, MTORC1 signaling, and protein secretion pathways were significantly
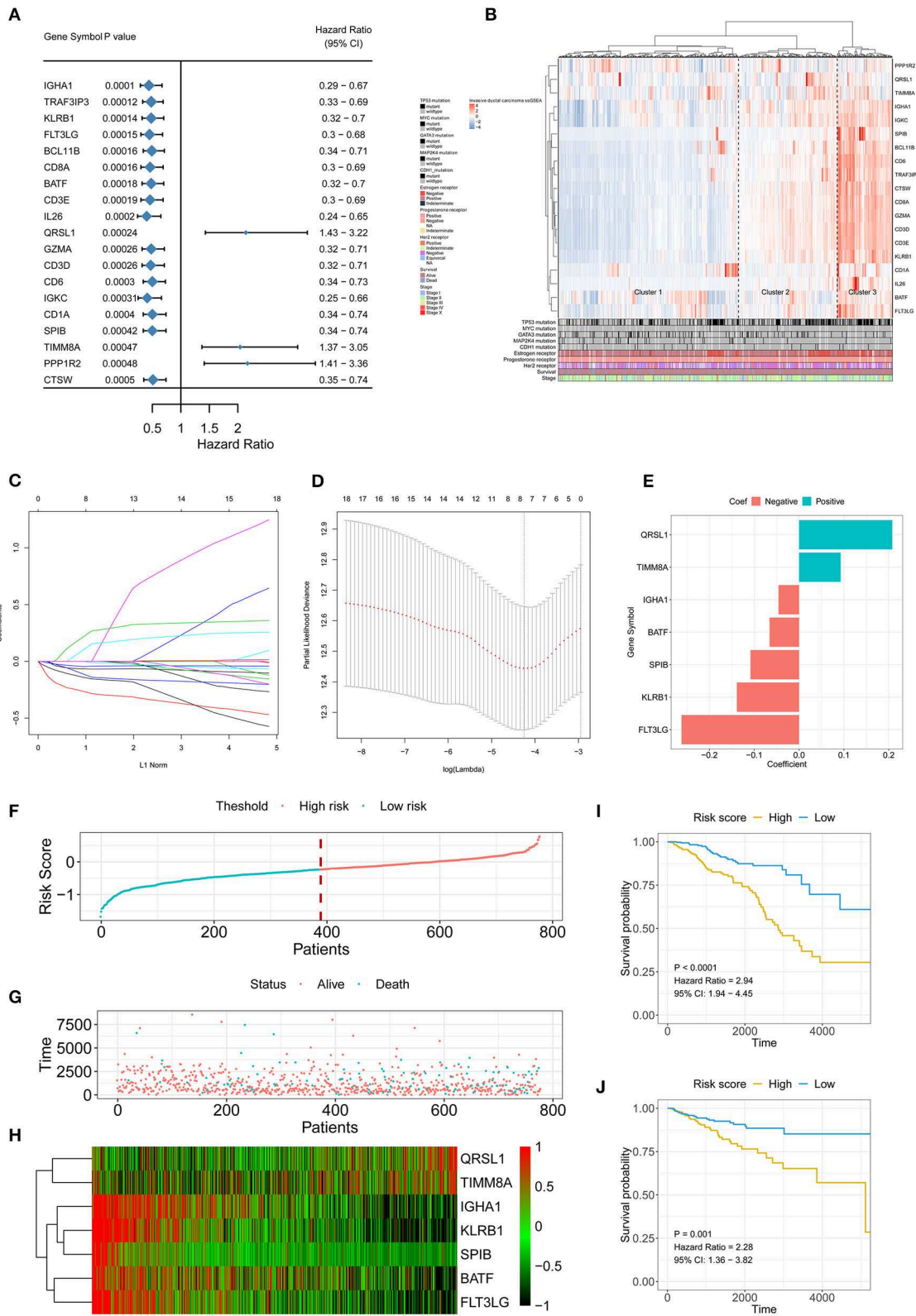
FIGURE 2 | Signature-based risk score is a promising marker of survival in IDC patients. **(A)** The HR and P-value from the univariable Cox HR regression of selected genes in the immune terms (Criteria: P-value < 0.001). **(B)** The expression of the selected genes shown by heatmap. Mutation status of TP53, MYC, GATA3,

(Continued)

FIGURE 2 | MAP2K4, and CDH1, status of the estrogen receptor, status of the progesterone receptor, status of Her2, survival, and stage are shown as patient annotations in the lower panel. Hierarchical clustering was performed with Euclidean distance and Ward linkage. **(C,D)** LASSO Cox analysis identified seven genes most correlated with overall survival, and 10-round cross validation was performed to prevent overfitting. **(E)** Coefficient distribution of the gene signature. **(F)** Risk score distribution. **(G)** Survival overview. **(H)** Heatmap showing the expression profiles of the signature in the low- and high-risk groups. **(I)** Patients in the high-risk group exhibited worse OS than those in the low-risk group. **(J)** Patients in the high-risk group exhibited worse RFS than those in the low-risk group. IDC, invasive ductal carcinoma; OS, overall survival; RFS, recurrence-free survival.
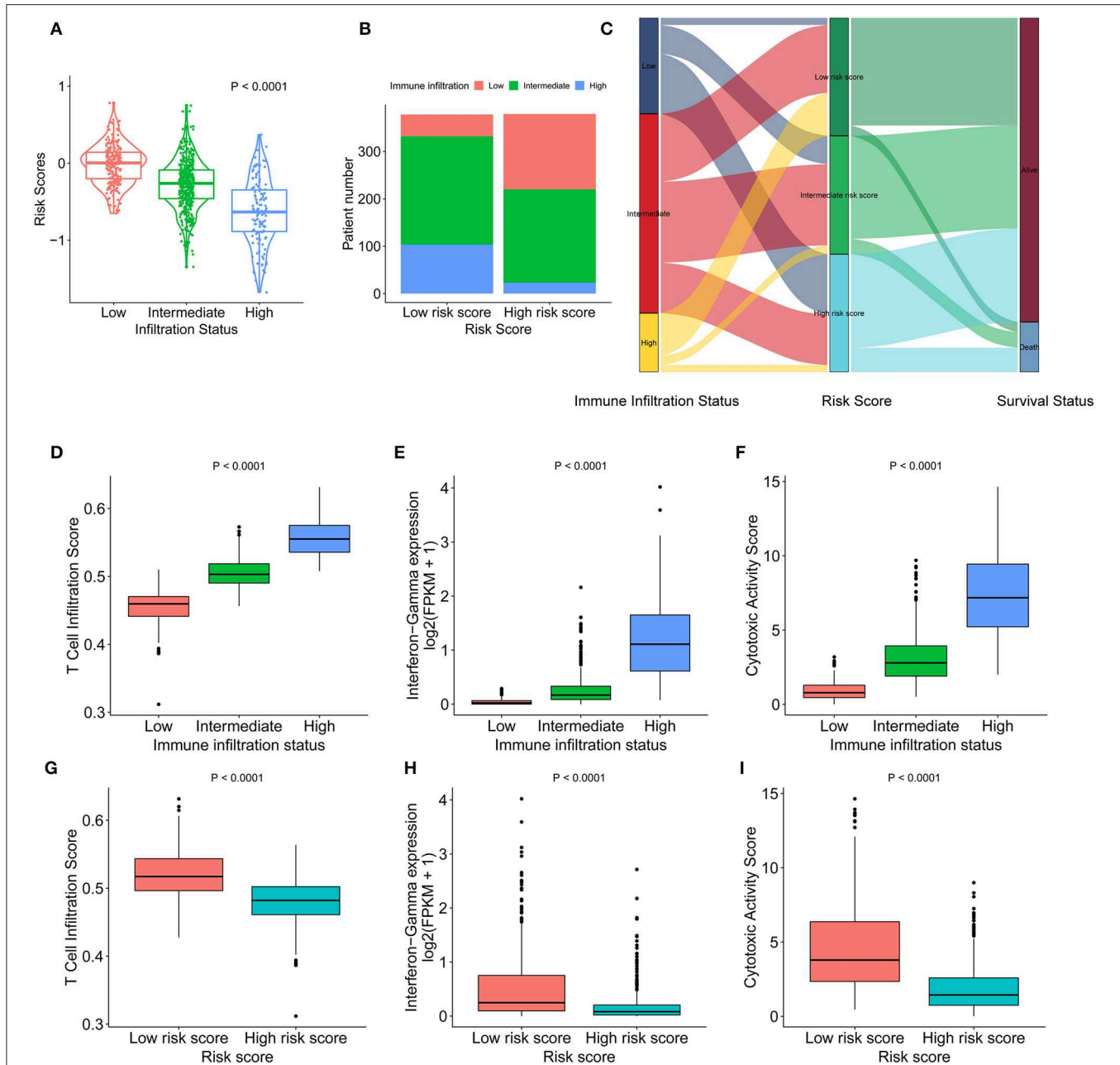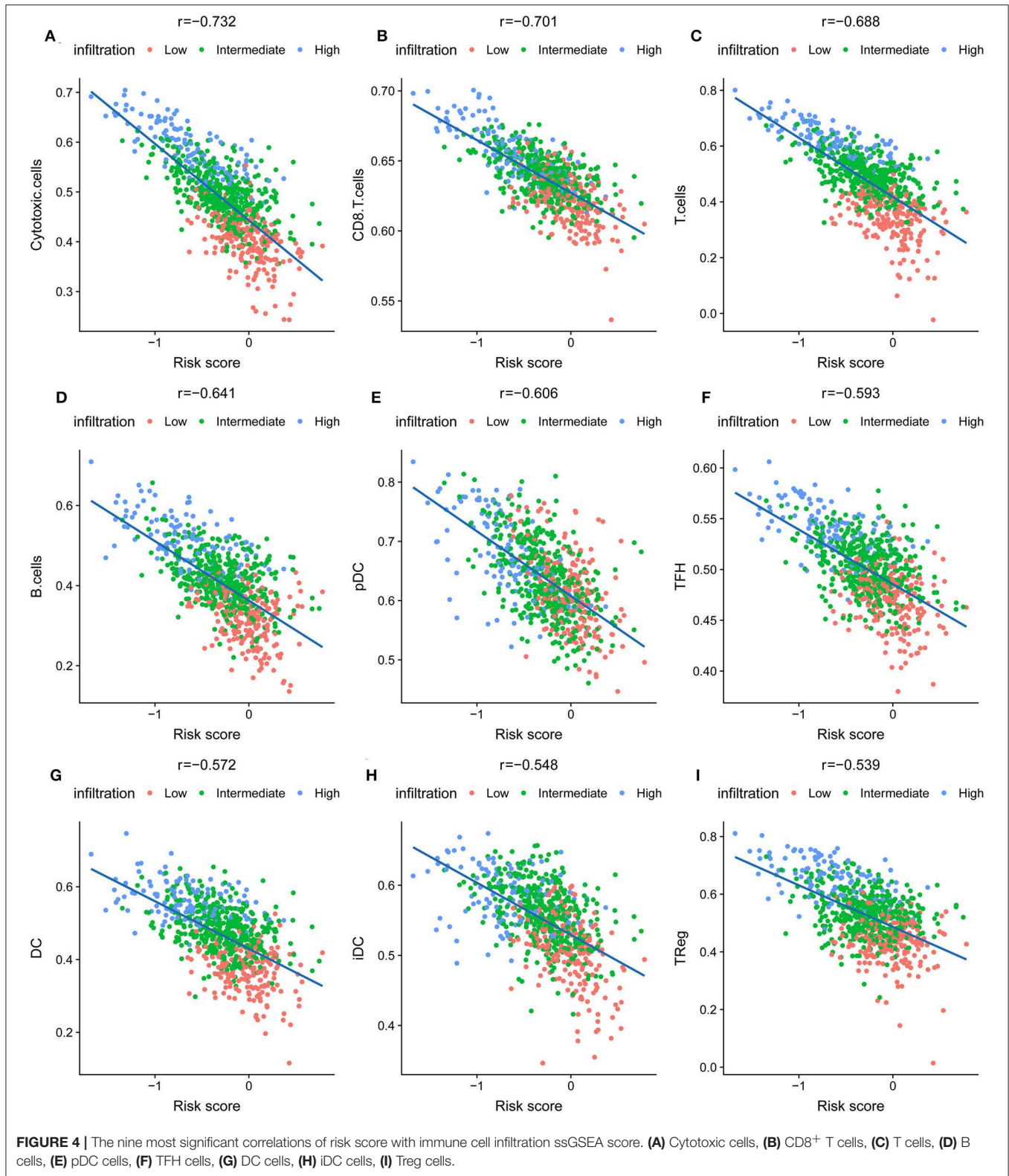

FIGURE 3 | Heterogeneous immune cell infiltration in the low- and high-risk score groups. **(A)** The distribution of risk scores in low, mediate, and high immune infiltration patterns. **(B)** The distribution of immune infiltration patterns in the low- and high-risk score groups. **(C)** Alluvial diagram of immune infiltration patterns in groups with different risk scores and survival outcomes. **(D)** TIS in low, mediate, and high immune infiltration patterns. **(E)** Relative interferon-γ signature in low, mediate, and high immune infiltration patterns. **(F)** Comparison of relative CYT in low, mediate, and high immune infiltration patterns. **(G)** Relative TIS in the low- and high-risk score groups. **(H)** Relative interferon-γ signature in the low- and high-risk score groups. **(I)** Comparison of relative CYT in the low- and high-risk score groups. TIS, T cell infiltration score; CYT, cytotoxic activity scores.

FIGURE 4 | The nine most significant correlations of risk score with immune cell infiltration ssGSEA score. **(A)** Cytotoxic cells, **(B)** CD8[+] T cells, **(C)** T cells, **(D)** B cells, **(E)** pDC cells, **(F)** TFH cells, **(G)** DC cells, **(H)** iDC cells, **(I)** Treg cells.

downregulated in the low-risk score group (**Figures 5C,D**). To further identify the underlying biological characteristics in the immune signature, WGCNA was performed, and the

correlation of risk score and immune infiltration status with module membership were analyzed. The soft threshold selection is shown in **Figure S9**. The module-trait heatmap illustrates
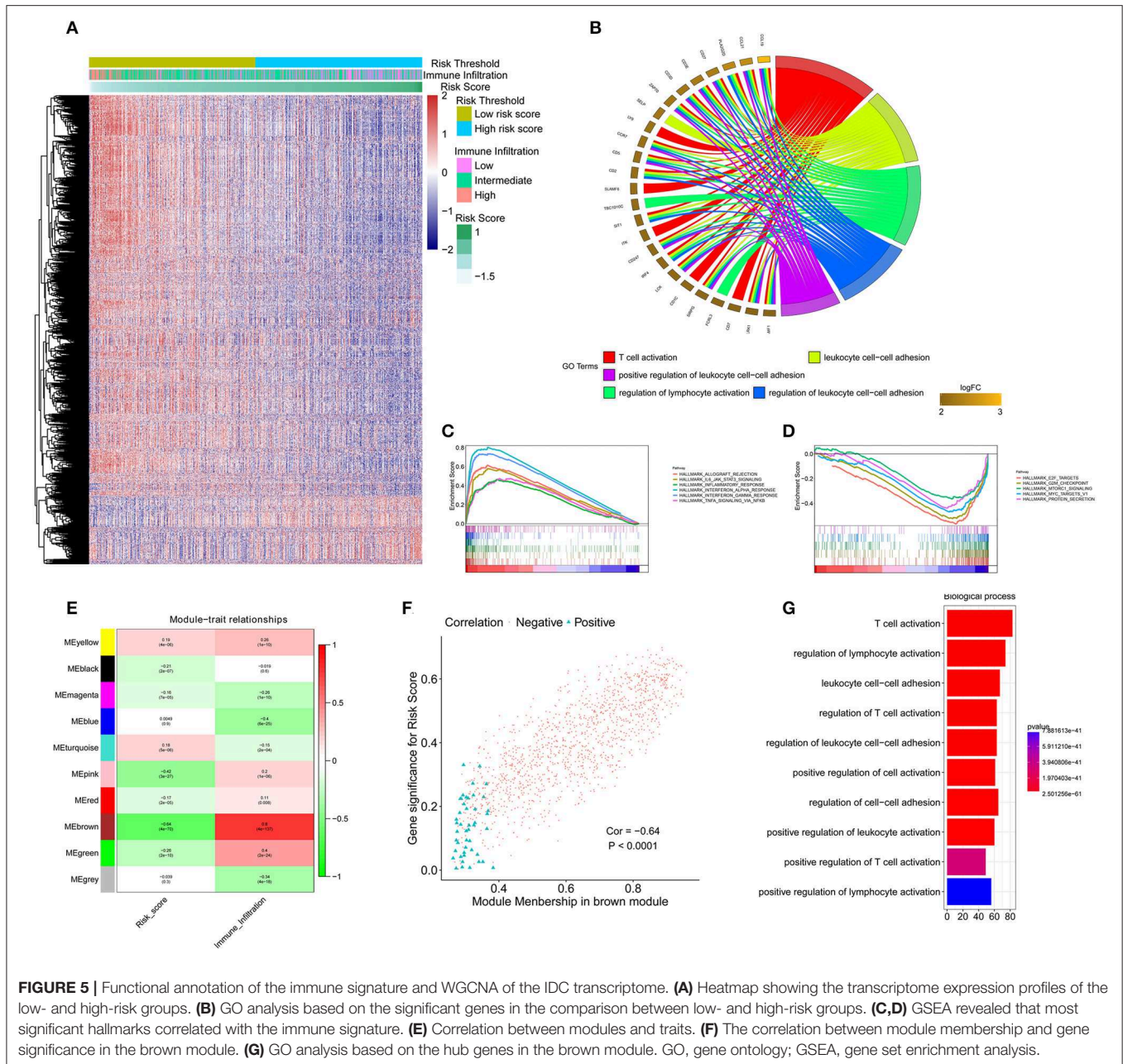
FIGURE 5 | Functional annotation of the immune signature and WGCNA of the IDC transcriptome. **(A)** Heatmap showing the transcriptome expression profiles of the low- and high-risk groups. **(B)** GO analysis based on the significant genes in the comparison between low- and high-risk groups. **(C,D)** GSEA revealed that most significant hallmarks correlated with the immune signature. **(E)** Correlation between modules and traits. **(F)** The correlation between module membership and gene significance in the brown module. **(G)** GO analysis based on the hub genes in the brown module. GO, gene ontology; GSEA, gene set enrichment analysis.

that the brown module had a significant *p*-value with both immune signature and immune infiltration status (**Figure 5E**); the coefficients were −0.64 and 0.8, respectively. The association between module membership and gene significance for each gene in the brown module is shown in **Figure 5F**. The genes from the brown module with a coefficient >0.5 were selected as hub genes, and GO enrichment analysis revealed that T cell activation and lymphocyte activation were the most significantly enriched biological processes, which further confirmed the results from the DEG analysis (**Figure 5G**).

Mutation Load and Immune Signature

The spectrum of somatic mutations in patients with IDCs is known to be varied. We next investigated the distributions of somatic mutations and observed different patterns among IDCs in terms of total mutations. The risk score from the immune signature had a positive correlation with total mutations in IDC patients (**Figure 6A**). By applying a random forest algorithm, we identified 35 highly variable mutated genes that were associated with the immune signature (**Figure 6B**). TP53 was the predominant gene of the 35 identified genes.
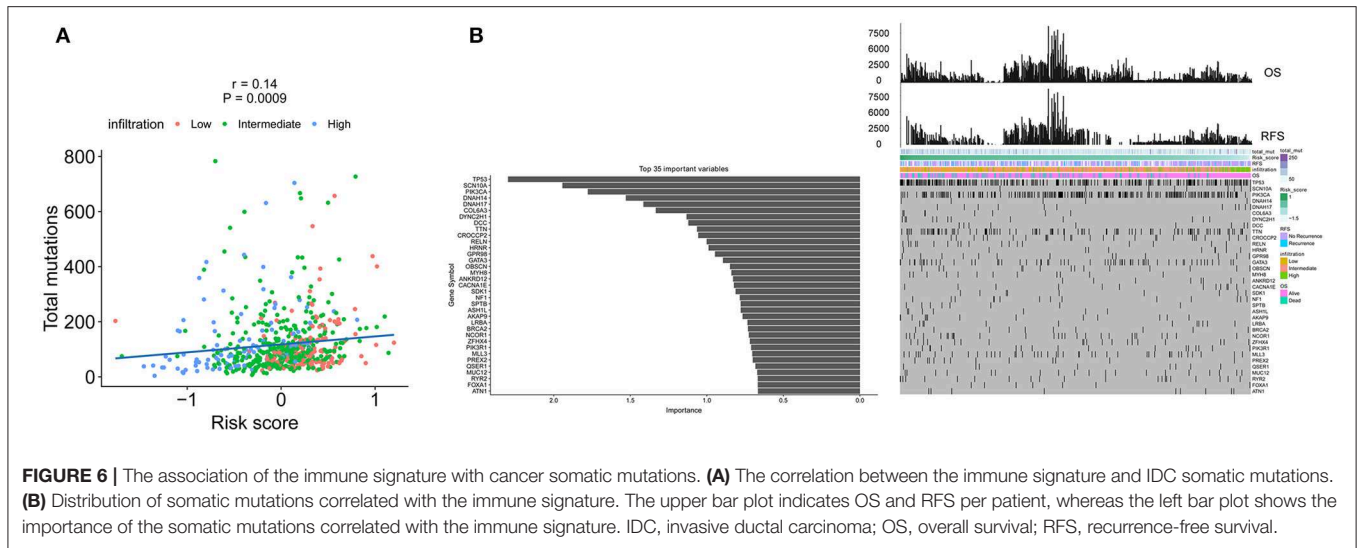
FIGURE 6 | The association of the immune signature with cancer somatic mutations. **(A)** The correlation between the immune signature and IDC somatic mutations. **(B)** Distribution of somatic mutations correlated with the immune signature. The upper bar plot indicates OS and RFS per patient, whereas the left bar plot shows the importance of the somatic mutations correlated with the immune signature. IDC, invasive ductal carcinoma; OS, overall survival; RFS, recurrence-free survival.
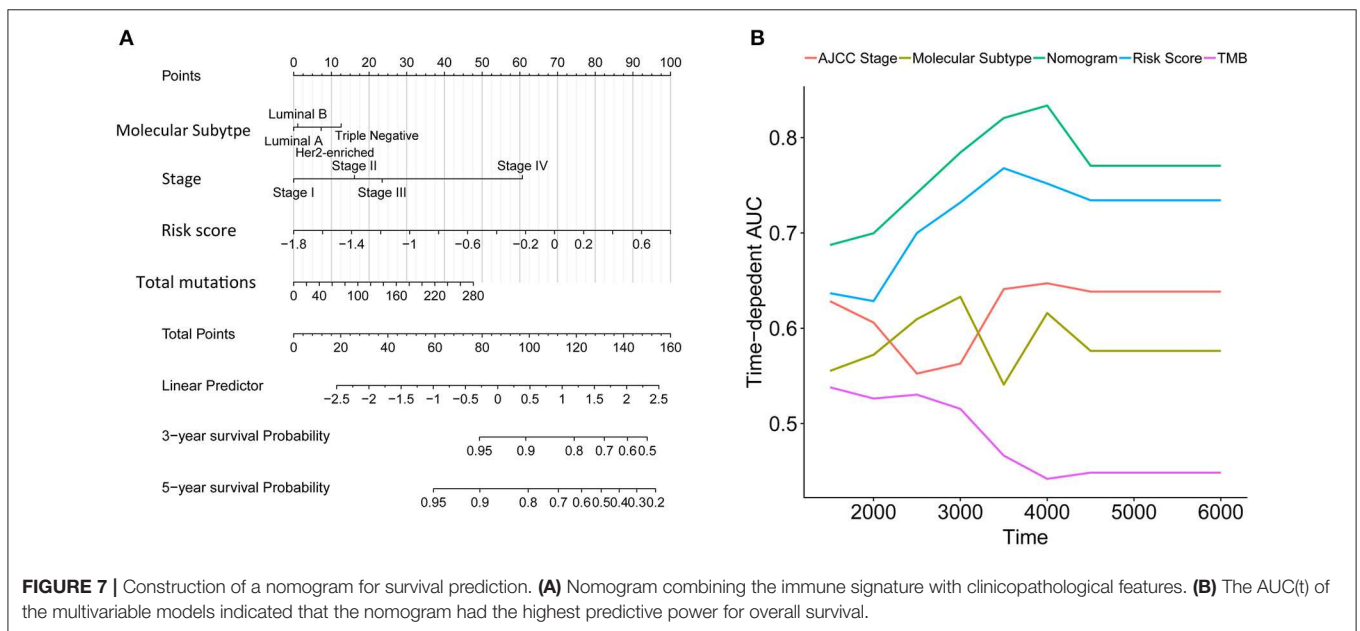

FIGURE 7 | Construction of a nomogram for survival prediction. **(A)** Nomogram combining the immune signature with clinicopathological features. **(B)** The AUC(t) of the multivariable models indicated that the nomogram had the highest predictive power for overall survival.

Construction of a Nomogram to Predict Overall Survival in IDC Patients

We constructed a nomogram that integrated clinicopathological features with the immune signature to predict the survival probability of IDC patients (**Figure 7A**). The AUC(t) functions of the multivariable models were developed to indicate how well these features serve as prognostic markers. Compared to other features, such as signature-based risk score, AJCC-TNM stage, and total mutation burden, the nomogram consistently showed the highest predictive power for overall survival in the follow-up period (**Figure 7B**).

The Immune Signature Predicted the Immunotherapeutic Benefits in IDC Patients

VEGF-A, the main mediator in tumor angiogenesis, hinders T cell infiltration in the tumor microenvironment. Hence, we explored the correlation between VEGF-A expression and the T cell immune response in IDC tumors. Interestingly, the increased VEGFA expression significantly correlated with both decreased levels of activated $CD8^+$ T cells and Th1 cell infiltration in the high immune infiltration tumor microenvironment but not in the low immune infiltration tumor microenvironment (**Figures 8A,B**). Furthermore, perforin, the molecular effector
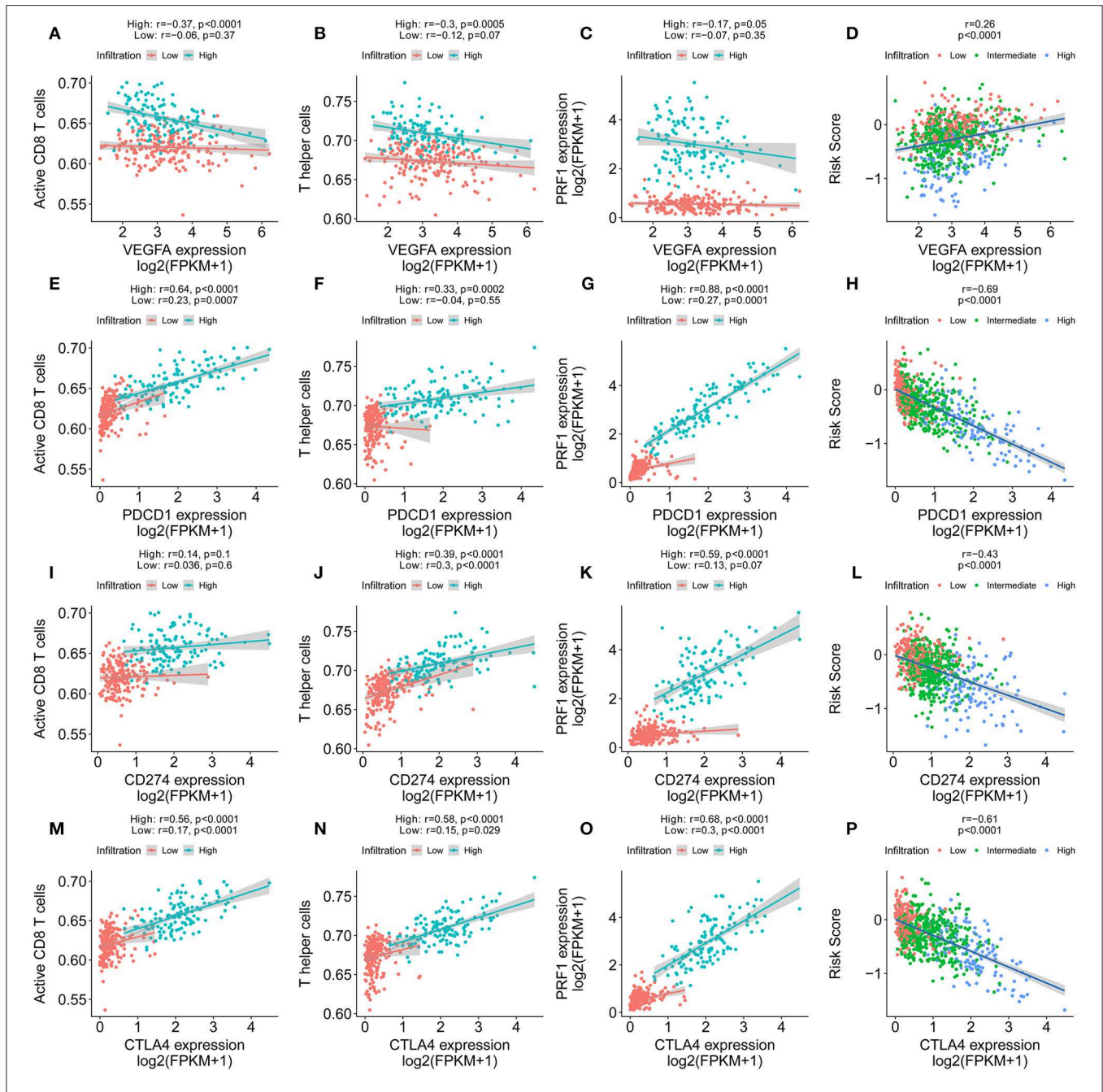
FIGURE 8 | Immune signature predicts immunotherapeutic benefits. **(A–C)** The correlation of VEGFA expression with T cell infiltration, Th1 cells, and PRF1 expression in high and low immune infiltration conditions. **(D)** The correlation of VEGFA expression with the immune signature. **(E–G)** The correlation of PD-1 expression with T cell infiltration, Th1 cells, and PRF1 expression in high and low immune infiltration conditions. **(H)** The correlation of PD-1 expression with the immune signature. **(I–K)** The correlation of PDL-1 expression with T cell infiltration, Th1 cells, and PRF1 expression in high and low immune infiltration conditions. **(L)** The correlation of PDL-1 expression with the immune signature. **(M–O)** The correlation of CTLA-4 expression with T cell infiltration, Th1 cells, and PRF1 expression in high and low immune infiltration conditions. **(P)** The correlation of CTLA-4 expression with the immune signature.

found in the granules of cytotoxic T lymphocytes and natural killer cells, also showed a negative correlation with VEGF-A expression (**Figure 8C**). Finally, the positive correlation of VEGF-A and the risk score was identified (**Figure 8D**). PD-1, PDL-1, and cytotoxic T lymphocyte antigen-4 (CTLA-4) are

promising targets for the treatment of patients with breast and non-small cell lung cancer. PD-1, PDL-1, and CTLA-4 antibodies are undergoing studies for the treatment of breast cancer. We analyzed the correlation of PD-1, PDL-1, and CTLA-4 expression in the high- and low-infiltration groups (**Figures 8E–P**). The

expression of PD-1, PDL-1, and CTLA-4 was more significantly correlated with CD8$^+$ T cells, Th1 cell ssGSEA score, and perforin expression in the high-infiltration group than in the low-infiltration group. Furthermore, the mean expression of PD-1, PDL-1, and CTLA-4 was significantly increased in the high-infiltration group, indicating a potentially enhanced response to the corresponding anticancer antibody for IDCs with high immune infiltration status. In our constructed immune signature, the risk score showed a negative correlation with PD-1, PDL-1, and CTLA-4 expression, which implies a potentially enhanced effect of PD-1, PDL-1, and CTLA-4 antibodies in patients with low risk score. Lastly, we checked the correlation of the expression profiles of several immune checkpoint proteins, e.g., CD160, CD274, CD276, CTLA-4, LAG3, and PDCD1, risk score, and VEGF-A in the TCGA and GSE20685 cohorts (**Figure S10**).

DISCUSSION

In this study, we depicted the immune landscape of IDC using a large cohort. The immune landscape might explain the differences in prognoses of patients with IDC and responses to PD1, PDL-1, and CTLA-4 antibodies. Based on the immune landscape, we constructed an immune signature that calculated the risk score per patient. The correlation of signature and immune landscape revealed that the T cell-mediated immune response played a crucial role in the signature. Patients with low risk scores had increased T cell infiltration scores, interferon-γ signatures, and cytotoxic activity scores, indicating active T cell immune responses and favorable survival probability. A random forest algorithm was applied to find the most important somatic mutation correlated with the immune signature. A nomogram was constructed based on the immune signature and other clinicopathological properties of IDCs. A time-dependent ROC analysis showed high accuracy of the immune signature and nomogram in terms of predicting the survival of IDC patients. Lastly, PD-1, PDL-1, and CTLA-4 expression was found to be highly associated with the risk score. The patients with low risk scores had increased expression levels of PD-1, PDL-1, and CTLA-4, indicating a potentially high response rate to PD-1, PDL-1, and CTLA-4 antibodies.

In our analysis, the IDCs were clustered into three main clusters (low immune infiltration, intermediate immune infiltration, and high immune infiltration). The patients in the high-infiltration cluster had the best survival probability compared with patients in the low- and intermediate-infiltration clusters. The T cell immune response is the central event in antitumour immunity (17). T cells are divided into CD4$^+$ (helper T cells, Th) and CD8$^+$ (cytotoxic T cells, Tc) T cells. Their secretomes include IFN-γ, TNF-α, and IL17, which have antitumour effects. Hence, the increased T cell infiltration score, interferon-γ signature, and cytotoxic activity score may lead to an anti-tumor effect in the high-infiltration group. This finding could explain the different OS and RFS in the high- and low-infiltration groups.

From the immune landscape in IDCs, we built an immune signature that included seven features (QRSL1, TIMM8A,

IGHA1, BATF, KLRB1, SPIB, and FLT3LG). FLT3LG is a crucial cytokine that controls the development of DCs and is particularly important for CD8-positive classical DCs and their CD103-positive tissue counterparts. A clinical trial is currently underway to treat melanoma patients with a combination of immunostimulatory FLT3LG and a peptide-based vaccine targeting DCs (18). KLRB1, which encodes CD161, a surface marker on several T cell subsets and NK cells, has been found to be most frequently associated with favorable outcomes in many cancer types by enhancing innate immune characteristics (19). SPIB is a member of the ETS family and profoundly affects B cell functions. B cells that lack SPIB fail to proliferate in response to IgM cross-linking, exhibit limited capacity to respond to T-dependent antigens, and produce low levels of IgG1, IgG2a, and IgG2b (20). In addition, SPIB can activate enhancer elements in both Ig-λ and Ig-κ genes, increasing the expression of these two genes. BATF is an inhibitor of AP-1-driven transcription. Recent studies have revealed that BATF can regulate positive transcriptional activity in dendritic cells, B cells and T cells (21). BATF leucine zipper motifs interact with interferon-regulatory factor 4 (IRF4) and IRF8 at AP-1–IRF consensus elements (AICEs), adding additional flexibility to the actions of IRF4 and IRF8, which were previously considered to interact with SPIB and PU.1 (22). The interaction of IRF4 and BATF in T helper 17 cells increases the production of IL-17, IL-21, IL-22, and IL-23 receptor. TIMM8A is involved in the import and insertion of hydrophobic membrane proteins from the cytoplasm to the mitochondrial inner membrane. The Bax/Bak complex mediates the release of DDP/TIMM8a and activates Drp1-mediated fission to promote mitochondrial fragmentation and subsequent elimination during programmed cell death (23). From the expression profiles of the seven genes above, we calculated the risk score for each patient and predicted the survival of IDC patients.

The risk score from the immune signature was most significantly correlated with the ssGSEA score of cytotoxic cells, CD8 T cells and T cells, indicating the important roles of the T cell immune response in the immune signature. Interestingly, DCs in the low-risk group played a more important role than DCs in the high-risk group. The increased proportion of DCs significantly correlated with favorable survival in the low-risk group but did not correlate with favorable survival of patients in the high-risk group. Th innate inflammatory cytokines, such as IL-1, IL-12, and IL-23 expressed by DCs, promote IFN-γ-secreting CD4$^+$ T cell and cytotoxic T lymphocyte responses (24). The high proportion of DCs and T cells cooperate to achieve the antitumour effect in IDC patients with low risk scores. MDSCs were immunosuppressive population. Patients in the high-risk score group had lower infiltration status and poor survival compared with that of patients in the low-risk score group. This might explain why the patients in the high-risk score group with high MDSC score had a poor survival compared with that of patients with low MDSC score. Furthermore, the GSEA results revealed high levels of IFN-γ, TNF-α, and TNF-α secretion in the low-risk group, which contribute to the antitumour activity in IDC patients with low risk scores.

WGCNA revealed opposing directions of the risk score (cor = −0.64) and immune infiltration (cor = 0.8) with the brown module, indicating the high level of correlation of risk score (calculated by immune signature) and immune infiltration. The hub gene in the brown module plays an essential role in regulating immune infiltration. The GO analysis revealed that T cell activation was the most significantly enriched biological process, indicating that the T cell-mediated immune response is the central event in both immune infiltration and the immune signature.

The spectrum of somatic mutations varied in IDC patients. The different mutation burdens in IDCs led us to analyse whether the landscape of immune cells and the immune signature were associated with somatic mutations. The total mutations showed a positive correlation with the risk score in IDC patients. Furthermore, a random forest algorithm was performed to identify the most important variables correlated with the immune signature. TP53, SCN10A, PIK3CA, and 32 other genes were the most significant variables in the analysis. TP53 and PIK3CA mutations are the most common gene mutations in IDCs (44 and 33%, respectively). In the 35 gene variables, GATA3, a key regulator of ER activity, is a newly identified gene that is mutated in IDCs (5% in ILC vs. 13% in IDC, $q = 0.03$) (3). Mutations in GATA3 are more frequent in luminal A IDC and are mutually exclusive with FOXA1 events. The differential expression level and enrichment for mutations of GATA3 in IDCs and of FOXA1 in ILC indicates a preferential requirement for the distinct regulation of ER activity in ILC and IDC (3). Previous studies revealed that the GATA3 mutation correlates with increased expression, which is associated with the immune response (25, 26). Our analysis further confirms the correlation of the GATA3 mutation with immune infiltration. In addition, we constructed a nomogram that integrated clinicopathological features with the immune signature to predict the survival probability of IDC patients. Compared with other clinicopathological features, the immune signature showed the best accuracy in predicting the survival of IDC patients at any time point and would therefore be helpful for the diagnosis and precise treatment of IDC patients.

There have been several studies for the treatment of breast cancer with immunotherapeutic antibodies. PD-1 is expressed by exhausted T cells. PD-1 and PD-L1 exhibit inhibitory receptor–ligand interactions, which are involved in the negative regulation of T cell activation and peripheral tolerance during immune responses by cancer cells. Despite demonstrated successes, only a proportion of patients benefit from PD-1 and PDL-1 antibody treatment. Hence, it is important to determine the mechanism that leads to the varied therapeutic effect of PD-1 and PD-L1 antibody treatment and thus improve individual diagnosis and precision medicine. PD-L1 expression, microsatellite instability and deficient mismatch repair are important biomarkers that predict the response to anti-PD-1/PD-L1 therapies (27–29). Among the three biomarkers, PD-L1 expression has been validated in nearly all tumor types for all approved anti-PD-1/PD-L1 therapies. In our analysis, the expression of PD1, PDL-1, and CTLA-4 was significantly increased in the high-infiltration group. Furthermore, the expression of PD1, PDL-1, and CTLA-4 had a significant correlation with CD8$^+$ T cells, Th1 cell

ssGSEA score, and perforin expression in the high-infiltration group, which provides a basis for PD-1/PD-L1 and CTLA-4 treatment. Similarly, the immune signature we constructed also indicated that high expression levels of PD1, PDL-1, and CTLA-4 correlated with low risk score. Therefore, patients with a low risk score could derive more benefit from immunotherapy than patients with a high risk score.

Some limitations should be acknowledged. First, this is a retrospective study, so the robustness of predictive value of the gene signature should be further validated in large prospective clinical trials. Second, experimental studies are required to further elucidate the biological functions underlying the gene signature in IDC.

In the current study, we performed a comprehensive evaluation of the immune landscape of IDC and constructed an immune signature related to the immune landscape. This analysis of TME immune infiltration patterns has shed light on how IDC respond to immunotherapy and may guide the development of novel drug combination strategies.

DATA AVAILABILITY

The datasets supporting the conclusions of this article are available in the Xena browser (https://xenabrowser.net/) repository.

AUTHOR CONTRIBUTIONS

XB and RS conceived and designed the experiments. XB, SX, and KZ analyzed the data. XB and YW wrote the paper. YZ, KZ, and RS reviewed the draft. All authors read and approved the final manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.00903/full#supplementary-material

Figure S1 | The correlation between different infiltrating immune cells.

Figure S2 | The correlation between the ssGSEA scores of infiltrating immune cells and the OS probability of IDC patients.

Figure S3 | The correlation between the expression level of seven genes in the immune signature and the OS probability of IDC patients. **(A)** BATF, **(B)** KLRB1, **(C)** TIMM8A, **(D)** FLT3LG, **(E)** QRSL1, **(F)** IGHA1, **(G)** SPIB.

Figure S4 | The correlation between the expression of seven genes in the immune signature and the RFS probability of IDC patients. **(A)** BATF, **(B)** KLRB1, **(C)** TIMM8A, **(D)** FLT3LG, **(E)** QRSL1, **(F)** IGHA1, **(G)** SPIB.

Figure S5 | Validation of the immune signature in two external cohorts, GSE20685 **(A)** and GSE86948 **(B)**.

Figure S6 | The correlation between the ssGSEA scores of DCs and the OS probability of IDC patients in the high- and low-risk score groups. **(A)** The ssGSEA scores were higher in the high- and low-risk score groups. **(B)** The correlation between the ssGSEA scores of DCs and the OS probability of IDC patients in the low-risk score group. **(C)** The correlation between the ssGSEA scores of DCs and the OS probability of IDC patients in the high-risk score group.

Figure S7 | The correlation between the ssGSEA scores of MDSCs and the OS probability of IDC patients in the high- and low-risk score groups. **(A)** The correlation between MDSC ssGSEA scores and risk scores. **(B)** The correlation between the ssGSEA scores of DCs and the OS probability of IDC patients in the whole cohort. **(C)** The correlation between the ssGSEA scores of DCs and the OS probability of IDC patients in the low-risk score group. **(D)** The correlation between

the ssGSEA scores of DCs and the OS probability of IDC patients in the high-risk score group.

Figure S8 | The ssGSEA score distribution in the low, intermediate, and high immune infiltration patterns and in the low- and high-risk score groups. **(A)** The ssGSEA score distribution in low, intermediate and high immune infiltration patterns. **(B)** The difference and *P*-value from the comparison between the ssGSEA score from low and high immune infiltration patterns. **(C)** The ssGSEA score distribution in the low- and high-risk score groups. **(D)** The difference and *P*-value from the comparison between the ssGSEA score from the low- and high-risk score group. **(E)** The distribution of immune infiltration patterns in different pathological subtypes. **(F)** The distribution of risk scores in different pathological subtypes. **(G)** The distribution of immune infiltration patterns at different pathological stages. **(H)** The distribution of risk scores at different pathological stages.

Figure S9 | The selection of the soft threshold in the WGCNA. **(A)** Scale free topology model fit, **(B)** mean connectivity.

Figure S10 | The correlation of the expression profiles of several immune checkpoint proteins, risk score, and VEGF-A in the TCGA **(A)** cohort and GSE20685 cohort **(B)**.

REFERENCES

1. Eble JN, Tavassoli FA, Devilee P. *Pathology and Genetics of Tumours of the Breast and Female Genital Organs*. Iarc (2003).
2. Arpino G, Bardou VJ, Clark GM, Elledge RM. Infiltrating lobular carcinoma of the breast: tumor characteristics and clinical outcome. *Breast Cancer Res.* (2004) 6:R149. doi: 10.1186/bcr767
3. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, et al. Comprehensive molecular portraits of invasive lobular breast cancer. *Cell.* (2015) 163:506–19. doi: 10.1016/j.cell.2015.09.033
4. Smid M, Rodríguez-González FG, Sieuwerts AM, Salgado R, Prager-Van der Smissen WJ, van Der Vlugt-Daane M, et al. Breast cancer genome and transcriptome integration implicates specific mutational signatures with immune cell infiltration. *Nat Commun.* (2016) 7:12910. doi: 10.1038/ncomms12910
5. Pages F, Galon J, Dieu-Nosjean M, Tartour E, Sautes-Fridman C, Fridman W. Immune infiltration in human tumors: a prognostic factor that should not be ignored. *Oncogene.* (2010) 29:1093. doi: 10.1038/onc.2009.416
6. Keir ME, Butte MJ, Freeman GJ, Sharpe AH. PD-1 and its ligands in tolerance and immunity. *Annu Rev Immunol.* (2008) 26:677–704. doi: 10.1146/annurev.immunol.26.021607.090331
7. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 blockade in tumors with mismatch-repair deficiency. *New Engl J Med.* (2015) 372:2509–20. doi: 10.1056/NEJMoa1500596
8. Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity.* (2013) 39:782–95. doi: 10.1016/j.immuni.2013.10.003
9. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* (2013) 14:7. doi: 10.1186/1471-2105-14-7
10. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell.* (2015) 160:48–61. doi: 10.1016/j.cell.2014.12.033
11. Wang Y, Deng H, Xin S, Zhang K, Shi R, Bao X. Prognostic and predictive value of three DNA methylation signatures in lung adenocarcinoma. *Front Genet.* (2019) 10:349. doi: 10.3389/fgene.2019.00349
12. Wang Y, Zhang Q, Gao Z, Xin S, Zhao Y, Zhang K, et al. A novel 4-gene signature for overall survival prediction in lung adenocarcinoma patients with lymph node metastasis. *Cancer Cell Int.* (2019) 19:100. doi: 10.1186/s12935-019-0822-1
13. Smyth GK. limma: linear models for microarray data. In: Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S, editors. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Statistics*

for Biology and Health. New York, NY: Springer (2005). p. 397–420. doi: 10.1007/0-387-29362-0_23
14. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* (2012) 16:284–7. doi: 10.1089/omi.2011.0118
15. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* (2008) 9:559. doi: 10.1186/1471-2105-9-559
16. Wang Y, Xin S, Zhang K, Shi R, Bao X. Low GAS5 Levels as a predictor of poor survival in patients with lower-grade gliomas. *J Oncol.* (2019) 2019:1–15. doi: 10.1155/2019/1785042
17. Baeuerle PA, Reinhardt C. Bispecific T-cell engaging antibodies for cancer therapy. *Cancer Res.* (2009) 69:4941–4. doi: 10.1158/0008-5472.CAN-09-0547
18. Vacchelli E, Aranda F, Bloy N, Buqué A, Cremer I, Eggermont A, et al. Trial watch—immunostimulation with cytokines in cancer therapy. *Oncoimmunology.* (2016) 5:e1115942. doi: 10.1080/2162402X.2015.1115942
19. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med.* (2015) 21:938–45. doi: 10.1038/nm.3909
20. Gallant S, Gilkeson G. ETS transcription factors and regulation of immunity. *Arch Immunol Ther Exp (Warsz).* (2006) 54:149–63. doi: 10.1007/s00005-006-0017-z
21. Murphy TL, Tussiwand R, Murphy KM. Specificity through cooperation: BATF–IRF interactions control immune-regulatory networks. *Nat Rev Immunol.* (2013) 13:499–509. doi: 10.1038/nri3470
22. Li P, Spolski R, Liao W, Wang L, Murphy TL, Murphy KM, et al. BATF–JUN is critical for IRF4-mediated transcription in T cells. *Nature.* (2012) 490:543–6. doi: 10.1038/nature11530
23. Arnoult D, Rismanchi N, Grodet A, Roberts RG, Seeburg DP, Estaquier J, et al. Bax/Bak-dependent release of DDP/TIMM8a promotes Drp1-mediated mitochondrial fission and mitoptosis during programmed cell death. *Curr Biol.* (2005) 15:2112–8. doi: 10.1016/j.cub.2005.10.041
24. Butt A, Mills K. Immunosuppressive networks and checkpoints controlling antitumor immunity and their blockade in the development of cancer immunotherapeutics and vaccines. *Oncogene.* (2014) 33:4623–31. doi: 10.1038/onc.2013.432
25. Usary J, Llaca V, Karaca G, Presswala S, Karaca M, He X, et al. Mutation of GATA3 in human breast tumors. *Oncogene.* (2004) 23:7669–78. doi: 10.1038/sj.onc.1207966
26. Mantel P-Y, Kuipers H, Boyman O, Rhyner C, Ouaked N, Rückert B, et al. GATA3-driven TH2 responses inhibit TGF-β1–induced FOXP3 expression and the formation of regulatory T cells. *PLoS Biol.* (2007) 5:e329. doi: 10.1371/journal.pbio.0050329

27. Lee V, Le DT. Efficacy of PD-1 blockade in tumors with MMR deficiency. *Future Med.* (2016) 8:1–3. doi: 10.2217/imt.15.97

28. Diaz LA, Marabelle A, Delord J-P, Shapira-Frommer R, Geva R, Peled N, et al. Pembrolizumab therapy for microsatellite instability high (MSI-H) colorectal cancer (CRC) and non-CRC. *J Clin Oncol.* (2017) 35(15 Suppl.):3071. doi: 10.1200/JCO.2017.35.15_suppl.3071

29. Topalian SL, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF, et al. Safety, activity, and immune correlates of anti–PD-1 antibody in cancer. *New Engl J Med.* (2012) 366:2443–54. doi: 10.1056/NEJMoa1200690

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