scientific reports

OPEN

Machine learning based intratumor heterogeneity related signature for prognosis and drug sensitivity in breast cancer

Hongcai Chen[1,4], Minna Chen[1,4], Cui Yang[2], Tingting Tang[3], Wende Wang[1]✉ & Wenwu Xue[1]✉

Intratumor heterogeneity (ITH) is involved in tumor evolution and drug resistance. Drug sensitivity shows discrepancy in different breast cancer (BRCA) patients due to ITH. The genes mediating ITH in BRCA and their role in predicting prognosis and drug sensitivity is not yet elucidated. An ITH-related signature (IRS) was built by ten methods-based integrative machine learning programs using TCGA, METABRIC and five GEO datasets. Several indicating scores were employed to evaluate the correlation between IRS score and immune microenvironment. The biological role of PINK1 was investigated using CCK-8 assay. The optimal prognostic signature for BRCA cases was the IRS developed using StepCox(both) + Enet(alpha = 0.9) method, which had the highest average C-index of 0.79. IRS acted as a prognostic biomarker and showed good performance in predicting the prognosis of BRCA patients. Lower IRS score indicated high levels of immuno-activated cells, higher TMB score, higher PD1&CTLA4 immunophenoscore, lower ITH score, lower TIDE score and lower tumor escape score in BRCA. The gene set scores correlated with glycolysis, angiogenesis, NOTCH signaling and hypoxia were higher in BRCA with high IRS score. PINK1 knockdown significantly inhibited the proliferation of BRCA cells. Our study developed a novel IRS for BRCA, which could predict the clinical outcome and immunotherapy benefits of BRCA patients.

Breast cancer (BRCA), a heterogeneous cancer with high morbidity rate, results in about 30% of female cancer diagnoses[1]. Depending on the disease stage and pathological features, the current therapeutic approaches for BRCA mostly include surgery, chemotherapy, endocrine therapy, and immunotherapy[2]. Even though the clinical outcome of breast cancer has significantly improved over the past few decades, new techniques are still required to identify patients who are at high risk. However, due to intratumor heterogeneity (ITH), the benefits of chemotherapy and immunotherapy varied throughout BRCA patients and therapy regimens must to be customized for each patient[3]. The use of DNA microarray and next-generation sequencing technology may have been the most important development in the last few decades in the study of cancer heterogeneity[4,5]. Apart from clinicopathological characteristics, unique gene signatures may offer additional data to predict the clinical outcome of BRCA patients.

ITH is defined as several clones present in a single malignancy, resulting in distinct features related to morphology, inflammation, genetics, or transcriptomics[6]. ITH is associated with treatment resistance and promotes tumor evolution and adaptability[7]. ITH is caused by environmental causes and gene mutation[8]. Additionally, a higher ITH may put patients at risk for a worse prognosis, a larger tumor burden, and fewer benefits from immunotherapy[9]. Given the vital role that ITH plays in cancer, it is particularly important to fully understand which genes mediate ITH in BRCA and how important these genes are for forecasting the benefits of treatment.

In our study, we developed a 6 genes-based ITH-related signature (IRS) for BRCA using 10 machine learning methods. Using one TCGA and METABRIC and 5 GEO datasets, the accuracy of IRS in predicting clinical outcome and drug sensitivity was confirmed. The role of key gene PINK1 in BRCA was verified using

[1]Department of Internal Medicine, Cancer Hospital of Shantou University Medical College, Shantou, China. [2]Department of Gynaecology and Obstetrics, Shantou Central Hospital, Shantou, China. [3]Department of Internal Medicine, Jinping District People's Hospital of Shantou, Shantou, China. [4]Hongcai Chen and Minna Chen contributed equally to this work. ✉email: 20wdwang1@stu.edu.cn; vincesdzl@163.com

in vitro experiment. Our findings suggest a potential association between ITH, prognosis, and the tumor microenvironment of BRCA patients, which has hardly ever been reported before.

Materials and methods

Sources of datasets

From the TCGA data portal (https://portal.gdc.cancer.gov/repository), the RNA-seq and expression data of female BRCA patients and the corresponding clinical data were isolated. Moreover, the RNA expression data of BRCA patients with paired clinical information of external validation cohorts (METABRIC + GSE20685 + GSE20711 + GSE42568 + GSE58812 + GSE96058) were downloaded from METABRIC database (https://molonc.bccrc.ca/aparicio-lab/research/metabric/) and GEO database (https://www.ncbi.nlm.nih.gov/geo/). Inclusion criteria included: (1) pathological diagnosis of BRCA, (2) completed clinical data and prognostic information. Exclusion criteria included: (1) followed up for less than six months, (2) accompanied by other malignancies, (3) metastatic BRCA. Three bulk RNA-seq datasets (IMvigor210, GSE91061 and GSE78220) in which patients were treated with PD-1 or CTLA4 inhibitors, were used to verify the role of IRS in predicting the immunotherapy benefits.

ITH score

DEPTH2, an algorithm for evaluating ITH, was used to calculate the ITH score of BRCA patients[10]. BRCA cases were divided into high and low DEPTH2 (ITH) scores for prognosis using the optimal cut-off. To identify the genes causing ITH, we used the "limma" packages to explore the differentially expressed genes (DEGs) between high and low DEPTH2 (ITH) score group with the $|Log_2FC|$ value > 1.5 and the p-value < 0.05 as the threshold value.

Development and assessment of an IRS

To identify potential prognostic biomarkers among DEGs in BRCA, we performed univariate cox analysis. These prognostic biomarkers were submitted to integrative machine learning analysis procedure for constructing a stable prognostic IRS. The machine learning method included random survival forest (RSF), elastic network (Enet), Lasso, Ridge, stepwise Cox, CoxBoost, partial least squares regression for Cox (plsRcox), supervised principal components (SuperPC), generalized boosted regression modelling (GBM), and survival support vector machine (survival-SVM). The run program of IRS was as follows: (1) Prognostic biomarkers was generated using univariate cox regression in the TCGA dataset; (2) Then, 101 algorithm combinations were performed on the prognostic signature to fit prediction models based on the leave-one-out cross-validation (LOOCV) framework in the TCGA dataset; (3) All models were detected in GEO and METABRIC cohorts; (4) For each model, the Harrell's concordance index (C-index) was calculated across all datasets. The prognostic model with the highest average C-index was considered optimal IRS. Previous studies have reported similar machine learning algorithms and the parameter tuning details about the R scripts is available on the Github website (https://github.com/Zaoqu-Liu/IRLS)[11–13]. Based on "surv_cutpoint" function within the R package "survminer", we obtained the best cutoff value and separated BRCA cases into low and high IRS score (risk score) groups in all datasets. The risk factor for the prognosis of BRCA was identified by univariate and multivariate cox analysis. The predictive nomogram was developed using "nomogramEx" R package based on IRS and clinical characters.

Immune infiltration, drug sensitivity and gene set enrichment analyses

Immune cells abundance of BRCA was explored using the R package "immunedeconv"[14]. The "estimate" R program was used to calculate the immunity and ESTIMATE scores[15]. The immune cell gene set score, immune-related functions, cancer-related hallmarks, and immune escape score of BRCA patients was calculated with ssGESA approach with "GSVA" R package. Tumor Immune Dysfunction and Exclusion (TIDE) score from TIDE (https://tide.dfci.harvard.edu/), the immunophenoscore (IPS) from The Cancer Immunome Database (https://tcia.at/home), and the Tumor Mutation Burden (TMB) score from TCGA were used to evaluate the role of IRS in predicting immunotherapy benefits. The "oncoPredict" R program was used to generate the IC50 of BRCA cases using the data of the Genomics of Drug Sensitivity in Cancer. Batch effect was corrected using the "sva" R package.

Cell lines and knockdown of PINK1

The protein level of PINK1 in BRCA and normal breast tissues was explored using The Human Protein Atlas (https://www.proteinatlas.org/)[16]. BRCA cell lines (T47D, BT549, SKBR3, MCF-7, MDA-MB-468) and normal breast cell line (Bst578Bst) from Shanghai Institute of Biochemistry and Cell Biology were cultured in conditions with 5% CO2 and 95% saturated humidity at 37 °C using the ATCC recommended medium. Additionally, we supplemented the medium with 1% penicillin-streptomycin (Sigma-Aldrich) and fetal bovine serum (FBS; Gibco). BRCA cell lines were transfected with PINK1 siRNA or scrambled negative control siRNA using Lipofectamine 3000 transfection reagent (Invitrogen) in accordance with the manufacturer's instructions.

RT-qPCR and proliferation assay

After separating the RNA from the cells using TRIzol (Takara Bio), the RNA was reverse transcribed into cDNA using an oligo (dT) primer. RT-qPCR was performed using Takara Bio's SYBR Premix Ex Taq. BRCA cell lines were grown in 96-well plates (5,000 cells/well in triplicate) for the proliferation assay. At the designated times, cells were supplemented with Cell Counting kit-8 (CCK-8; Beyotime). The OD value of BRCA cell lines at the specified time was computed to evaluate the proliferation of BRCA cell.

Statistical analysis

Utilizing R software (version 4.2.1), statistical studies were conducted. The student T test or the Wilcoxon rank-sum test were used to assess the difference between continuous variables. The correlations between two continuous variables were examined using either Spearman's or Pearson's rank correlation technique. Different Kaplan-Meier survival curves were tested for differences using the two-sided log-rank test.

Results

The prognostic value IRGs of BRCA

The ITH score of each BRCA was shown in Fig. 1A. Compared with BRCA patients in clinical stage III and IV, BRCA patients in clinical stage I and II had a lower ITH score (Fig. 1B). There was no significant difference between different groups based on age, T stage, N stage, ER status, PR status (Supplementary Fig. 1A–E). After separating BRCA patients into high and low ITH score group, we found that high ITH score indicated a lower overall survival rate (Fig. 1C). Next, we then explored DEGs mediating ITH in BRCA, and a total of 598 DEGs were obtained between high and low ITH score group (Fig. 1D). Among these gene, twenty genes were significantly correlated with the prognosis of BRCA patients based on univariate cox analysis, including PFKL, USP33, SDC3. BCL2L10, MT1G, PRKAA2, PINK1, NDRG1, PTPN2, VDAC1, RRP8, CDKN1A, TOP2A, CCNA2, MSH2, CDK6, ISG20, LALBA, ERRFI1 and VHL (Fig. 1E).

A predictive IRS developed by integrative machine learning algorithms

Above prognostic biomarkers were submitted to integrative machine learning analysis procedure for constructing a stable prognostic IRS, and a total of 101 different types of prediction models were obtained. As shown in Fig. 2A, the optimal IRS was the StepCox(both) + Enet(alpha = 0.9) based prognostic model, which had the greatest average C-index of 0.79. The optimal IRS developed by StepCox(both) + Enet(alpha = 0.9) method including six genes. The feature selection was shown in Supplementary Fig. 2A,B. And the IRS score (risk score) of each BRCA patients was calculated with the formula below:

IRS score $= 0.0221 \times TOP2A^{expression} + 0.0311 \times RRP8^{expression} + 0.2356 \times PINK1^{expression} + (-0.0246) \times MSH2^{expression} + 0.0325 \times CCNA2^{expression} + (-0.0753) \times CDK6^{expression}$.

We then separated BRCA into high and low IRS score groups. The results showed that BRCA patients with high IRS scores experienced a poor clinical outcome rate in TCGA, METABRIC and all GEO dataset (Fig. 2B,H). The AUCs of 1-, 3-, and 5-year ROC were, respectively, 0.823, 0.804, and 0.810 in the TCGA cohort; 0.836, 0.824, and 0.815 in the METABRIC cohort; 0.792, 0.790, and 0.768 in the GSE20685 cohort; 0.753, 0.762, and 0.784 in the GSE20711 cohort; 0.844, 0.770, and 0.741 in the GSE42568 cohort; 0.838, 0.812, and 0.792 in the GSE58812 cohort; 0.824, 0.795, and 0.800 in the GSE96058 cohort, respectively (Fig. 2B,H).

IRS acted as prognostic indicator in BRCA

The C-index of IRS was higher than other clinical characteristics, including age, tumor grade, T stage, N stage, M stage, clinical stage, ER status, PR status, and HER2 status, indicating the performance of IRS in predicting
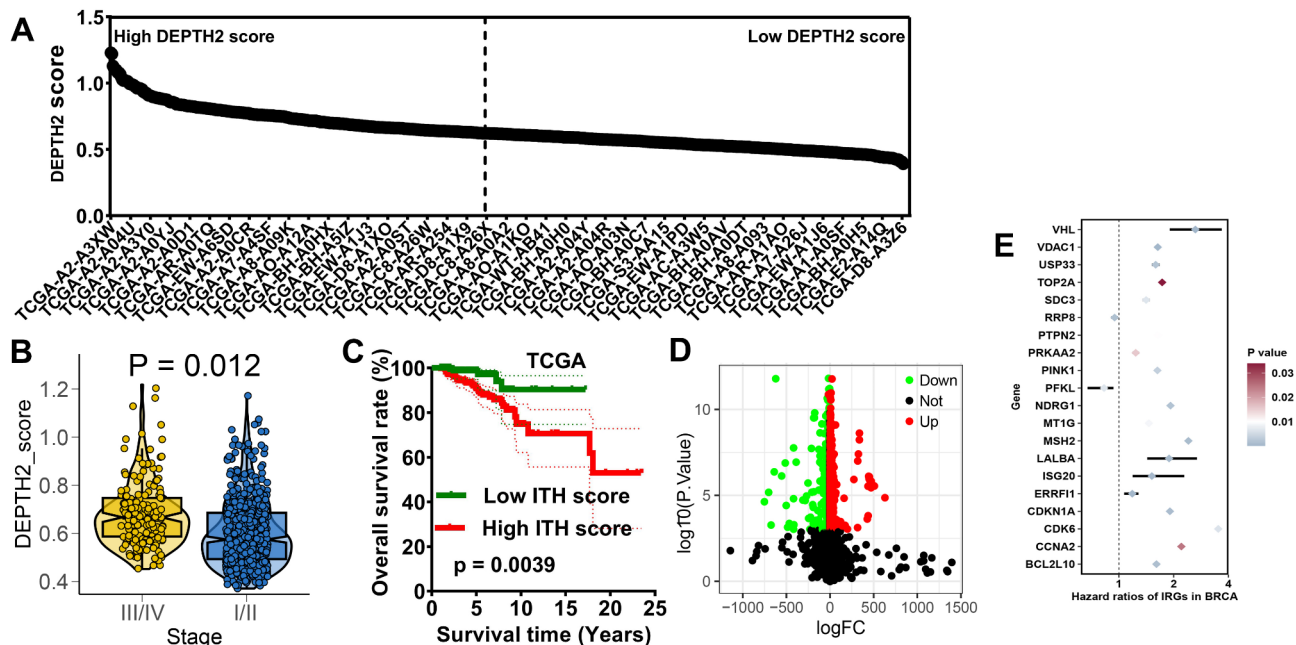

Fig. 1. The intratumor heterogeneity score of BRCA cases. (**A**) The intratumor heterogeneity score of each BRCA patients. (**C**) High intratumor heterogeneity score indicated a poor overall survival rate in BRCA. (**D**) The different expressed genes between in high and low intratumor heterogeneity score group. (**E**) Univariate cox analysis identified potential prognostic biomarkers in BRCA.
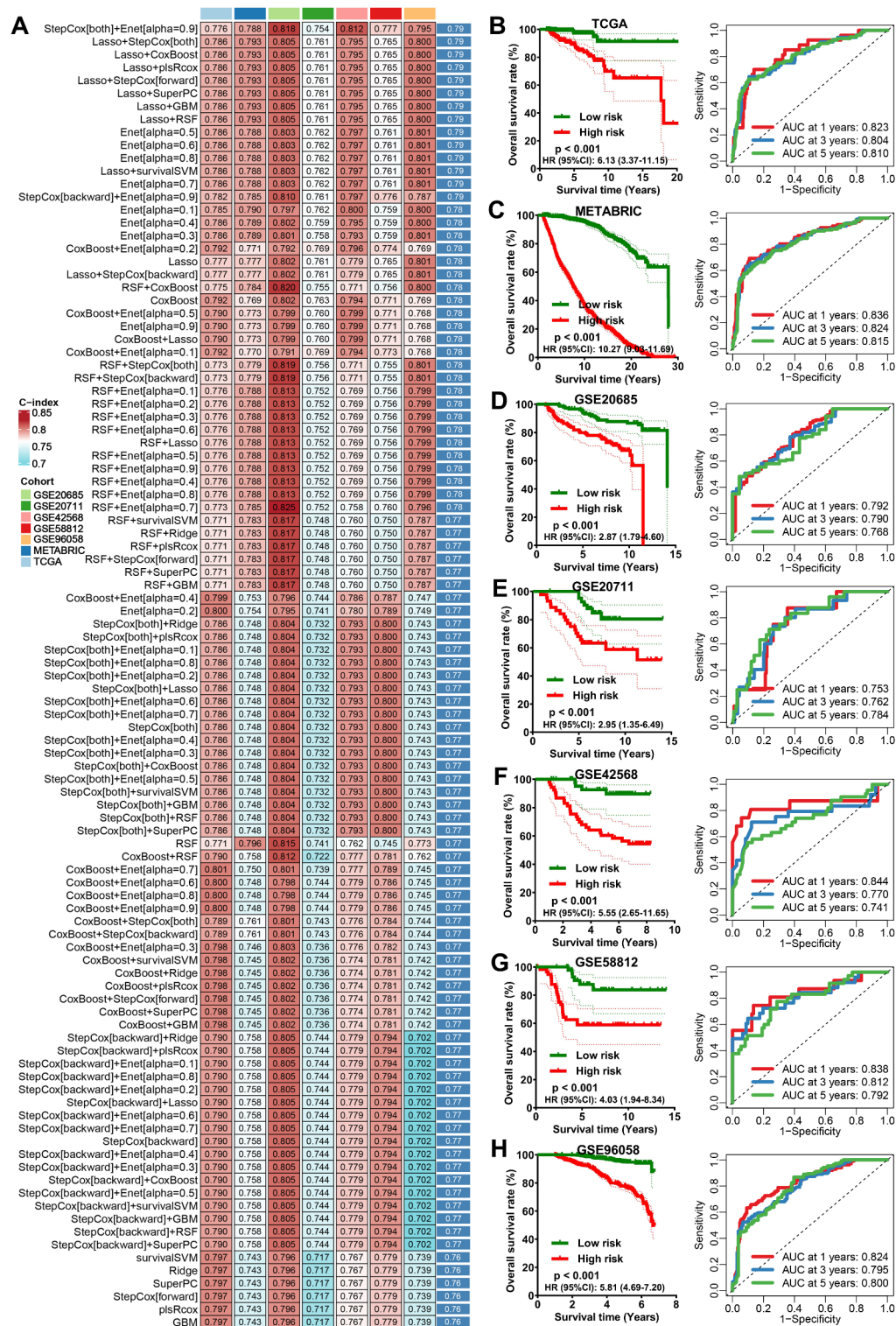
Fig. 2. Machine learning developed a prognostic IRS. (**A**) The C-index of 101 kinds prognostic models of TCGA and all GEO datasets. The survival curve of breast cancer patients with different IRS score and their corresponding ROC curve in TCGA (**B**), METABRIC (**C**), GSE20685 (**D**), GSE20711 (**E**), GSE42568 (**F**), GSE58812 (**G**), and GSE96058 (**H**) datasets.

the clinical outcome of BRCA were better than clinical characteristics (Fig. 3A). The results of the cox regression analysis were shown in Fig. 3B,C, indicating IRS as an independent risk factor for the clinical outcome of BRCA patients in the TCGA, METABRIC and all GEO cohorts. In order to predict the clinical prognosis of BRCA cases more conveniently, a nomogram was also developed based on IRS and clinical characteristics (Fig. 3D). High
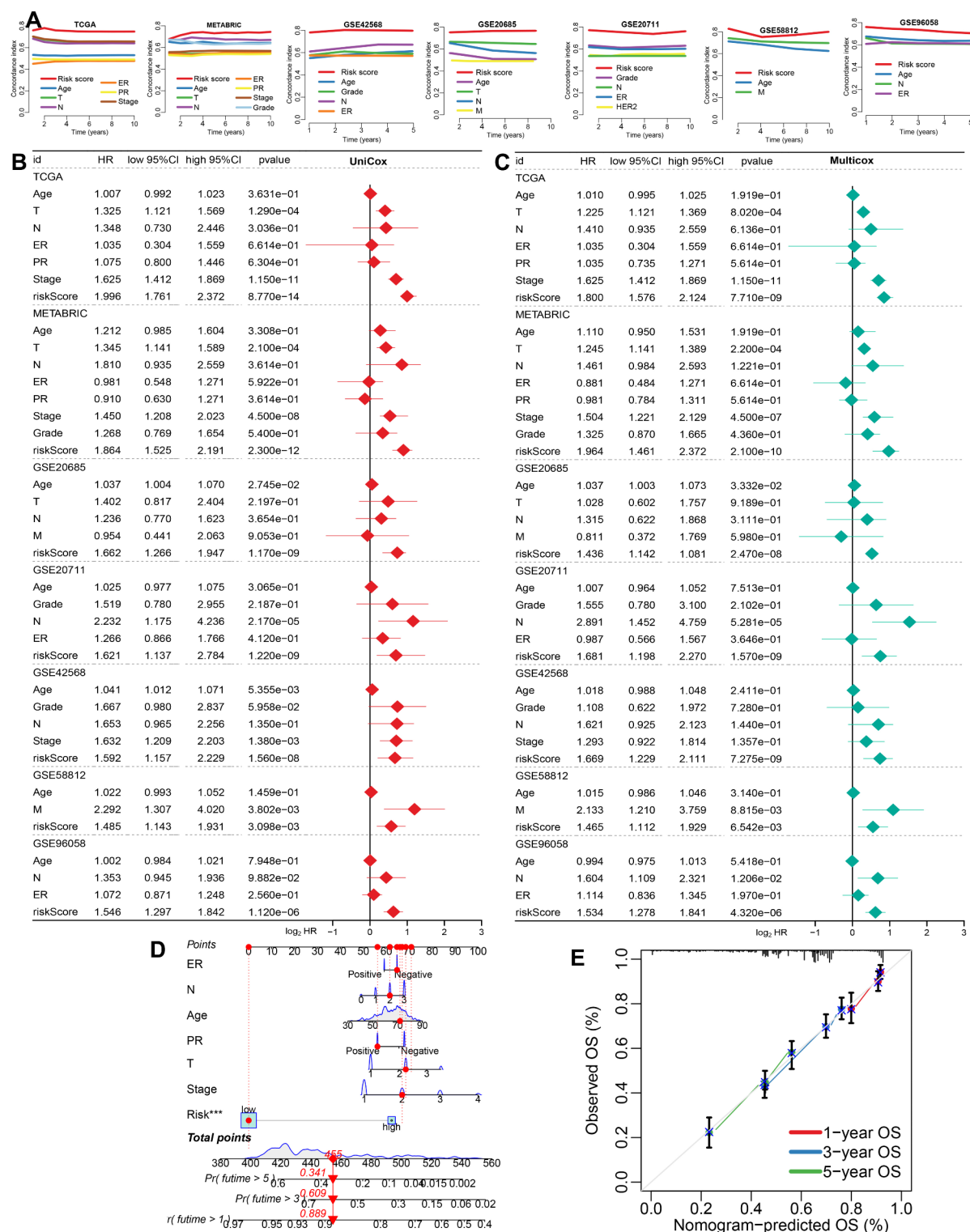
Fig. 3. Evaluation of the role of IRS in BRCA patients. (**A**) The C-index of IRS and clinical characteristics in TCGA, METABRIC and 5 GEO datasets. (**B**, **C**) Prognostic risk factors identified by univariate and multivariate cox regression analysis. (**D**) Nomogram developed based on IRS and clinical characters. (**E**) Calibration evaluated the role of nomogram in evaluating the prognosis of BRCA patients.

coincidence was found between the ideal and actual prediction curves (Fig. 3E), suggesting that nomogram had good potential for clinical use.

Correlation between ITH score and immune microenvironment

We then explored the correlation between ITH score and immune microenvironment. As shown in Fig. 4A, high IRS score indicated higher level of immune score, stromal score and ESTIMAE score (all $p < 0.001$). The overall association between the IRS score and the abundance of immune cell was shown in Fig. 4B, suggesting that the abundance of CD8+ T cells, NK cells, and macrophage M1 cells showed significant negative correlation with risk score in BRCA (Fig. 4C,E). Based on the result of ssGSEA analysis, high IRS score group indicated lower levels of aDCs, B cells, CD8+ T cells, DCs, iDCs, mast cells, neutrophils, NK cells, pDCs, Tfh, and and TILs in BRCA (Fig. 4F). Moreover, BRCA patients with low IRS score had a higher gene set score correlated with APC_co_stimulation, CCR, checkpoint, cytolytic activity, HLA, inflammation promoting, T cell co-stimulation and type II IFN response (Fig. 4G, all $p < 0.05$).

IRS acted as an indicator for predicting therapy benefits in BRCA

A greater variety of antigen presentation was indicated by high HLA-related gene expression, which raised the possibility of presenting more immunogenic antigens and the possibility of immunotherapy success[17]. Low IRS score indicated higher level of HLA-related genes in BRCA (Fig. 5A). Higher checkpoint expression indicated more immunotherapy targets, increasing the likelihood of benefit from immunotherapy. BRCA patients with low IRS score had higher level of checkpoints in BRCA (Fig. 5B). High TMB scores showed correlation with improved immunotherapy outcomes[18]. IPS was a more accurate predictor of immunotherapy outcomes and Higher IPS indicated stronger immunotherapy benefits[19]. As shown in Fig. 5C,D, high IRS score indicated higher TMB score, higher PD1 and CTLA4 IPS. Low TIDE score and ITH score suggested a less likelihood of immune escape and a better response to immunotherapy[20,21]. Compared with high IRS score group, low IRS score group had lower TIDE score, lower immune score and lower ITH score in BRCA (Fig. 5E,G). Therefore, IRS may serve as an indicator for immunotherapy benefit and BRCA with a low IRS score may benefit more from immunotherapy. We then verified these findings with three immunotherapy datasets. In patient receiving Atezolizumab therapy, IRS score of respondents was lower than non-responders (Fig. 5H). Compared with high IRS score group, low IRS score group had a poorer OS rate and lower responder rate (Fig. 5H). Similar outcomes were also obtained in the GSE91061 and GSE78220 cohorts (Fig. 5I,J), further suggesting IRS as an indicator for immunotherapy benefit. As chemotherapy and endocrine therapy were important approaches for BRCA. We then explored the IC50 value of common drugs correlated with chemotherapy and endocrine therapy. As shown in Fig. 6A,B, BRCA patients with low IRS score had lower IC50 value for Oxaliplatin, Gemcitabine, Epirubicin, Docetaxel, Cisplatin, 5-Fluorouracil, Ribociclib, Dinaciclib, Palbociclib, Fulvestrant, Tamoxifen, and Lapatinib (all $p < 0.05$).

The correlation between IRS score and cancer related hallmarks

Functional enrichment analysis was performed to explore why there were substantial differences in the clinical outcome and immunotherapy advantages between high and low IRS score groups. The results found that high IRS score indicated higher angiogenesis related gene set score, coagulation related gene set score, DNA repair related gene set score, glycolysis related gene set score, hypoxia related gene set score, IL2-STAT5 signaling related gene set score, mTORC1 signaling related gene set score, NOTCH signaling related gene set score, EMT signaling related gene set score, P53 pathway related gene set score, E2F targets related gene set score, and peroxisome related gene set score in BRCA (Fig. 7, all $p < 0.05$).

Biological roles of PINK1 in BRCA

As PINK1 had a largest coefficient in risk score calculation formula, we selected PINK1 for further analysis. Compared with normal breast cell line, PINK1 expression was higher in BRCA cell lines (Supplementary Fig. 3A). Typical immunohistochemistry showed that PINK1 expression was higher in BRCA than normal breast tissues (Supplementary Fig. 3B). Further CCK-8 assay demonstrated that PINK1 knockdown significantly inhibited the proliferation of SKBR3 and MCF-7 cells (Supplementary Fig. 3C,D, $p < 0.05$).

Discussion

The current study developed a prognostic IRS for IRS using an integrative machine learning procedure. In the TCGA, METABRIC, and 5 GEO datasets, the prognostic IRS acted as an independent risk factor and demonstrated a powerful performance in predicting the clinical outcome of BRCA patients. Moreover, we also found that IRS acted as a biomarker for predicting the immunotherapy response.

The prognostic IRS was developed based on 6 genes, including TOP2A, RRP8, PINK1, MSH2, CCNA2, CDK6. These genes were reported to play a vital role in BRCA or other types of cancer. Positive TOP2A gene expression on circulating tumor cells was a critical biomarker for predicting outcomes of BRCA patients[22]. TOP2A modulates signaling could promote ovarian cancer cell proliferation AKT/mTOR pathway[23]. PINK1 facilitated breast cancer cell proliferation and affected glycolysis[24]. Moreover, PINK1 displays tissue-specific subcellular location and regulates apoptosis and cell growth in breast cancer[25]. MUC1 promoted tumor progression by promoting PINK1-dependent mitophagy in BRCA[26]. Moreover, targeting ATAD3A-PINK1-mitophagy axis overcomes chemoimmunotherapy resistance by redirecting PD-L1 to mitochondria in BRCA[27]. CCNA2 acted as prognostic biomarker and involved in cell proliferation in BRCA[28]. CCNA2 regulate cell progression by regulating the cell cycle in glioblastoma[29]. In addition to the prognostic markers mentioned above, previous study also identified some biomarkers for BRCA, such as cytokine-induced apoptosis inhibitor 1 and TUBA1B[30,31].
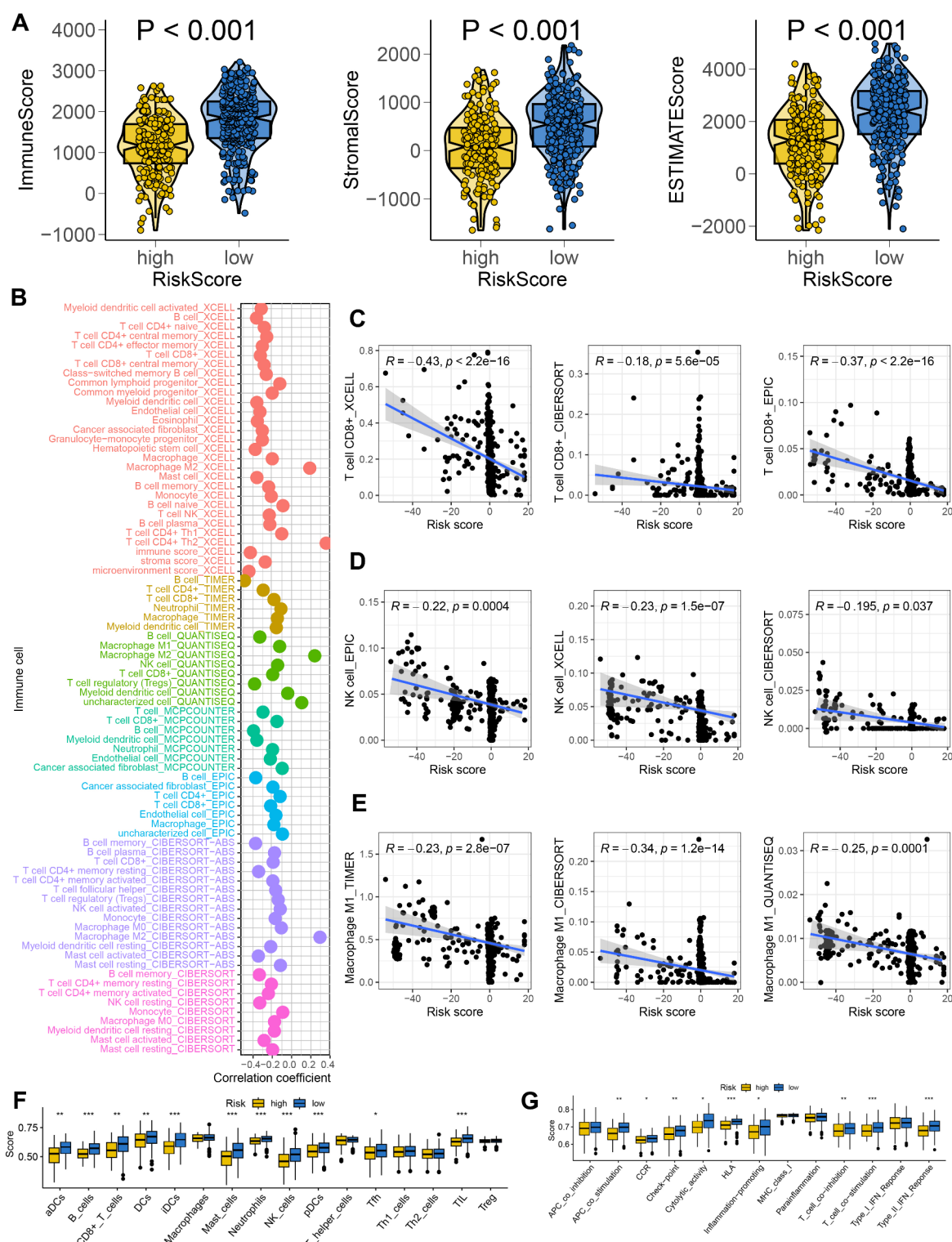
Fig. 4. Correlation between IRS score and immune microenvironment. (**A**) The immune score and ESTIMAE score in BRCA cases with different IRS score. (**B**) The correlation between IRS and the abundance of immune cell based on seven algorithms. (**C–E**) The level of CD8+ T cells, NK cells, and macrophage M1 cells was negatively correlated with IRS score. (**F**, **G**) The level of immune cells and immune-related function in high IRS score group and low IRS score group (High IRS score group VS Low IRS score group). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.
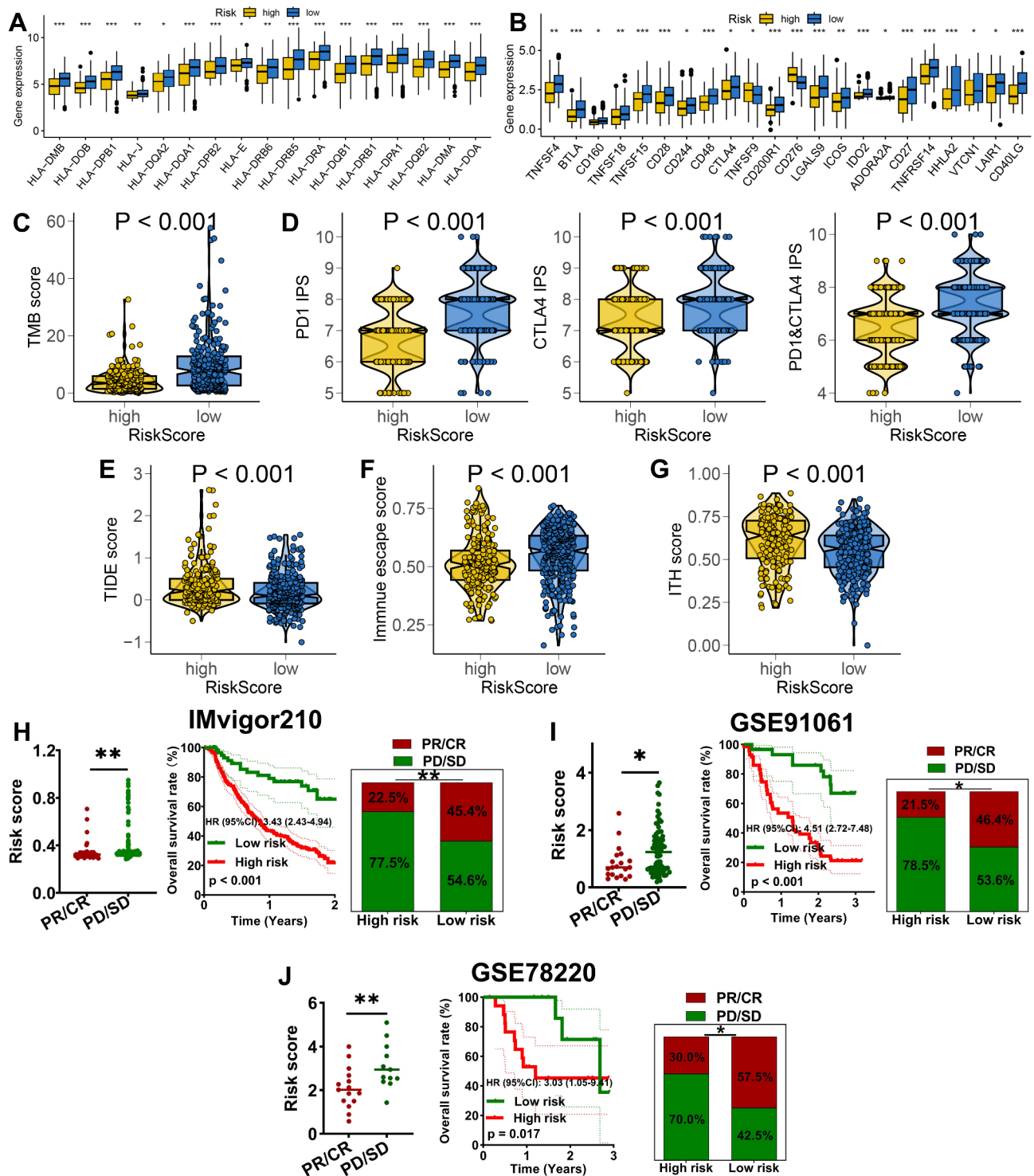
Fig. 5. IRS acted as an indicator for immunotherapy benefits in BRCA. The level of HLA-related genes (**A**), immune checkpoints (**B**), TMB score (**C**) PD1&CTLA4 immunophenoscore (**D**), TIDE score (**E**), immune escape score (**F**) and ITH score (**G**) in different IRS score group. The overall rate and immunotherapy response rate in different IRS score group in IMvigor210 (**H**), GSE91061 (**I**) and GSE78220 (**J**) cohort. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

The tumor microenvironment encompasses a complex network of tumor cells, diverse infiltrating immune cells, stromal cells, and cytokines[32]. As research into this microenvironment deepens, the critical role of infiltrating immune cells in the progression, metastasis, recurrence, and immune evasion of BRCA has become increasingly evident. Consequently, these immune cells represent a promising therapeutic target[33]. Our findings revealed that the low IRS score group exhibited higher levels of CD8 + T cells, NK cells, macrophage M1 cells, and B cells
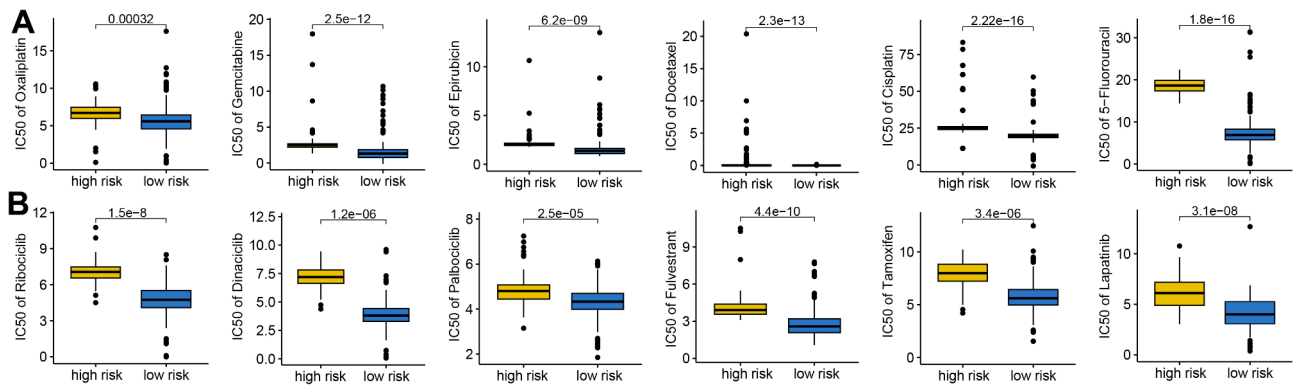
Fig. 6. The IC50 value of drugs in different IRS score group. The IC50 value of chemotherapy (**A**) and endocrinotherapy (**B**) in different IRS score group.
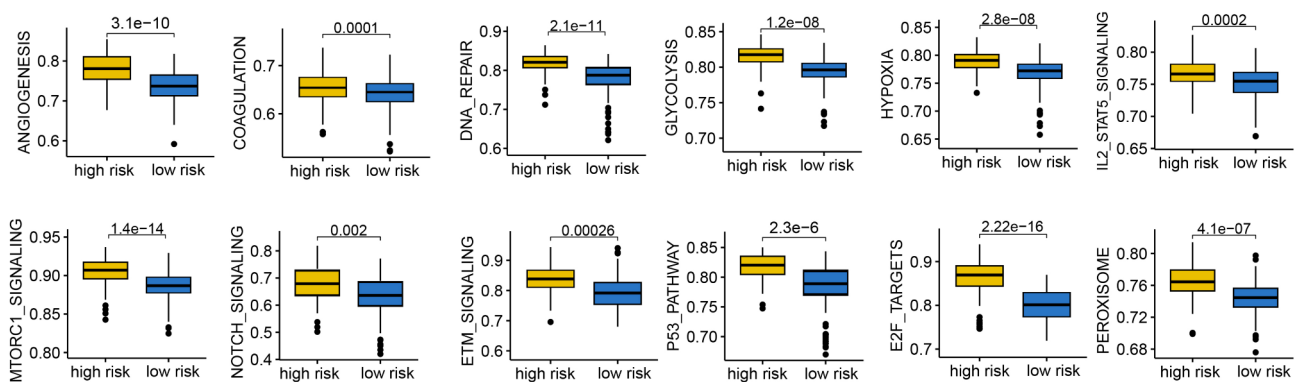

Fig. 7. The correlation between cancer related hallmarks and IRS in BRCA.

compared to the high IRS score group. An elevated density of CD4 + T cells is indicative of a favorable prognosis, which could account for the improved outcomes observed in the low-risk cohort[34,35]. Stromal and immune cells are integral components of the tumor microenvironment, and their respective scores correlate with the prognosis of cancer patients[36]. Our analysis indicated that BRCA patients with low IRS score had markedly higher stromal, immune, and ESTIMATE scores. This suggests that ITH may influence the tumor microenvironment, thereby impacting tumor progression. Immunotherapy offers more life-extending possibilities for cancer patients with unresectable tumors[3]. We then explore the role of IRS in immunotherapy response in BRCA with several indicating scores. High TMB score indicated a better response to immunotherapy[23]. IPS was a superior predictor of response to anti-CTLA4 and anti-PD1 antibodies and high IPS indicated a better response to immunotherapy[24]. Low TIDE score suggested a less likelihood of immune escape and a better response to immunotherapy[25]. The data of our investigation showed that BRCA patients with low IRS score had higher abundance of immuno-activated cells, higher PD1&CTLA4 immunophenoscore, higher TMB score, lower TIDE score, and lower tumor escape score. Thus, IRS may act as an indicator for predicting immunotherapy benefits and IRS patients with low IRS score may had a better immunotherapy benefit.

Early diagnosis is also crucial for extending the overall survival rate of BRCA patients. Previous study identified serum biomarker panels composed of five miRNAs for the early diagnosis of BRCA[37]. Though some diagnostic biomarkers have been used for the early diagnosis of BRCA, the development of new diagnostic markers and new diagnostic methods are also very important for breast cancer or other diseases. For example, Li et al. developed a raman nanosphere based immunochromatographic system for the combined detection of influenza A and B viruses[38]. Moreover, surface-enhanced Raman scattering-lateral flow immunochromatography could be help for the rapid detection of anti-Brucella IgG/IgM[39]. Functional enrichment analysis was performed to explore why there were substantial differences in the clinical outcome and immunotherapy advantages between high and low IRS score groups. The results found that high IRS score indicated higher gene set score involved in angiogenesis, coagulation, DNA repair, glycolysis, hypoxia, IL2-STAT5 signaling, mTORC1 signaling, NOTCH signaling, EMT signaling, P53 pathway, E2F targets, and peroxisome in BRCA. Angiogenesis was involved in the progression of BRCA patients[40]. Glycolysis is essential for tumor cell proliferation and metastasis and drug resistance in BRCA[41,42]. Hypoxia signaling played a vital role in mediating BRCA invasion and metastasis[43].

Our research has certain limitations. Since all of the information was gathered from public sources at the RNA level, it was unable to represent the outcomes at the protein level. The lack of comparative analysis with existing prognosis models for breast cancer makes IRS difficult to highlight the clinical transformation potential.

It would be better to verify the role of IRS using an independent prospective cohort. It would also be preferable to investigate the function and mechanism of PINK1 in BRCA using both in vitro and in vivo research, including animal models or clinical samples. Clarifying the function of PINK1 in the development of BRCA and the associated molecular pathways should be the main goal of future study.

Conclusion

Our study developed a novel IRS for BRCA. The IRS served as an indicator for predicting the clinical outcome and immunotherapy benefits of BRCA patients.

Data availability

The analyzed data sets generated during the study were sourced from the TCGA database (https://portal.gdc.cancer.gov/repository) and GEO database (https://www.ncbi.nlm.nih.gov/geo).

References

1. Safiri, S. et al. Global, regional, and national cancer deaths and disability-adjusted life-years (DALYs) attributable to alcohol consumption in 204 countries and territories, 1990–2019. *Cancer* **128** (9), 1840–1852 (2022).
2. Waks, A. G. & Winer, E. P. Breast cancer treatment: A review. *Jama* **321** (3), 288–300 (2019).
3. Passaro, A., Brahmer, J., Antonia, S., Mok, T. & Peters, S. Managing resistance to immune checkpoint inhibitors in lung cancer: treatment and novel strategies. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **40** (6), 598–610 (2022).
4. Abd-Elnaby, M., Alfonse, M. & Roushdy, M. Classification of breast cancer using microarray gene expression data: A survey. *J. Biomed. Inform.* **117**, 103764 (2021).
5. Hong, M. et al. RNA sequencing: new technologies and applications in cancer research. *J. Hematol. Oncol.* **13** (1), 166 (2020).
6. Angerilli, V. et al. Intratumor morphologic and transcriptomic heterogeneity in (V600E)BRAF-mutated metastatic colorectal adenocarcinomas. *ESMO Open* **6** (4), 100211 (2021).
7. Gerlinger, M. et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366** (10), 883–892 (2012).
8. Andor, N. et al. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat. Med.* **22** (1), 105–113 (2016).
9. Li, J. et al. Tumor cell-intrinsic factors underlie heterogeneity of immune cell infiltration and response to immunotherapy. *Immunity* **49** (1), 178–193e177 (2018).
10. Song, D. & Wang, X. DEPTH2: an mRNA-based algorithm to evaluate intratumoral heterogeneity without reference to normal controls. *J. Transl. Med.* **20** (1), 150 (2022).
11. Liu, Z. et al. Machine learning-based integration develops an immune-derived LncRNA signature for improving outcomes in colorectal cancer. *Nat. Commun.* **13** (1), 816 (2022).
12. Liu, Z. et al. Integrative analysis from multi-center studies identities a consensus machine learning-derived LncRNA signature for stage II/III colorectal cancer. *EBioMedicine* **75**, 103750 (2022).
13. Zhang, H. et al. Machine learning-based tumor-infiltrating immune cell-associated LncRNAs for predicting prognosis and immunotherapy response in patients with glioblastoma. *Brief. Bioinform.* **23**(6). (2022).
14. Li, T. et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* **48** (W1), W509–w514 (2020).
15. Yoshihara, K. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
17. Uhlén, M. et al. Proteomics. Tissue-based map of the human proteome. *Science (New York, NY)* **347** (6220), 1260419. (2015).
17. Lin, A. & Yan, W. H. HLA-G/ILTs targeted solid cancer immunotherapy: opportunities and challenges. *Front. Immunol.* **12**, 698677 (2021).
18. Liu, L. et al. Combination of TMB and CNA stratifies prognostic and predictive responses to immunotherapy across metastatic cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **25** (24), 7413–7423 (2019).
19. Charoentong, P. et al. Pan-cancer Immunogenomic analyses reveal genotype-Immunophenotype relationships and predictors of response to checkpoint blockade. *Cell. Rep.* **18** (1), 248–262 (2017).
20. Fu, J. et al. Large-scale public data reuse to model immunotherapy response and resistance. *Genome Med.* **12** (1), 21 (2020).
21. Vitale, I., Shema, E., Loi, S. & Galluzzi, L. Intratumoral heterogeneity in cancer progression and response to immunotherapy. *Nat. Med.* **27** (2), 212–224 (2021).
22. Ye, J. H., Yu, J., Huang, M. Y. & Mo, Y. M. The correlation study between TOP2A gene expression in circulating tumor cells and chemotherapeutic drug resistance of patients with breast cancer. *Breast Cancer (Tokyo Japan)* **31** (3), 417–425 (2024).
23. Zhang, K. et al. TOP2A modulates signaling via the AKT/mTOR pathway to promote ovarian cancer cell proliferation. *Cancer Biol. Ther.* **25** (1), 2325126 (2024).
24. Li, J. et al. Pink1 promotes cell proliferation and affects glycolysis in breast cancer. *Exp. Biol. Med. (Maywood NJ)* **247** (12), 985–995 (2022).
25. Berthier, A. et al. PINK1 displays tissue-specific subcellular location and regulates apoptosis and cell growth in breast cancer cells. *Hum. Pathol.* **42** (1), 75–87 (2011).
26. Li, Q. et al. The oncoprotein MUC1 facilitates breast cancer progression by promoting Pink1-dependent mitophagy via ATAD3A destabilization. *Cell Death Dis.* **13** (10), 899 (2022).
27. Xie, X. Q. et al. Targeting ATAD3A-PINK1-mitophagy axis overcomes chemoimmunotherapy resistance by redirecting PD-L1 to mitochondria. *Cell Res.* **33** (3), 215–228 (2023).
28. Cao, M. G. et al. The effect of miR-381 on proliferation and prognosis of breast cancer by altering CCNA2 expression. *J. Obstet. Gynaecol. J. Inst. Obstet. Gynecol.* **44** (1), 2360547 (2024).
29. Zhou, H. Y. et al. CCNA2 and NEK2 regulate glioblastoma progression by targeting the cell cycle. *Oncol. Lett.* **27** (5), 206 (2024).
30. Luo, Z. et al. Cytokine-induced apoptosis inhibitor 1: a comprehensive analysis of potential diagnostic, prognosis, and immune biomarkers in invasive breast cancer. *Transl. Cancer Res.* **12** (7), 1765–1786 (2023).
31. Wang, Y. et al. Tubulin alpha-1b chain was identified as a prognosis and immune biomarker in pan-cancer combing with experimental validation in breast cancer. *Sci. Rep.* **14** (1), 8201 (2024).
32. Kessenbrock, K., Plaks, V. & Werb, Z. Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell* **141** (1), 52–67 (2010).
33. Sanz-Pamplona, M. E. I. J., Hermitte, R. & de Miranda, F. Colorectal cancer: A paradigmatic model for cancer immunology and immunotherapy. *Mol. Aspects Med.* **69**, 123–129 (2019).
34. Ahlén Bergman, E. et al. Increased CD4(+) T cell lineage commitment determined by CpG methylation correlates with better prognosis in urinary bladder cancer patients. *Clin. Epigen.* **10** (1), 102 (2018).

35. Schroeder, B. A. et al. CD4 + T cell and M2 macrophage infiltration predict dedifferentiated liposarcoma patient outcomes. *J. Immunother. Cancer* **9** (8). (2021).

37. Malka, D. et al. Immune scores in colorectal cancer: Where are we? *Eur. J. Cancer (Oxford England 1990)* **140** 105–118. (2020).

37. Jing, Y. et al. Diagnostic value of 5 MiRNAs combined detection for breast cancer. *Front. Genet.* **15**, 1482927 (2024).

39. Li, Z, et al. Establishment of a Raman nanosphere based immunochromatographic system for the combined detection of influenza A and B viruses' antigens on a single T-line. *Nanotechnology* **35** (50). (2024).

39. Zhang, Y. et al. Combined and rapid detection of anti-Brucella IgG/IgM in clinical samples based on surface-enhanced Raman scattering-lateral flow immunochromatography. *J. Mater. Chem. B* **12** (42), 11012–11024 (2024).

40. Ribatti, D., Annese, T. & Tamma, R. Controversial role of mast cells in breast cancer tumor progression and angiogenesis. *Clin. Breast Cancer* **21** (6), 486–491 (2021).

41. Lin, J. et al. Glycolytic enzyme HK2 promotes PD-L1 expression and breast cancer cell immune evasion. *Front. Immunol.* **14**, 1189953 (2023).

42. Guo, Z. et al. Hypoxia-induced downregulation of PGK1 crotonylation promotes tumorigenesis by coordinating glycolysis and the TCA cycle. *Nat. Commun.* **15** (1), 6915 (2024).

43. Wang, R. et al. Hypoxia-inducible factor-dependent ADAM12 expression mediates breast cancer invasion and metastasis. *Proceedings of the National Academy of Sciences of the United States of America* **118** (19). (2021).

Author contributions

Hongcai Chen Writing- Original draft preparation, Cui Yang: Investigation, Minna Chen: Study design and Supervision, Tingting Tang: Writing- Reviewing, Wende Wang: Methodology, Conceptualization, Study design and Supervision, Wenwu Xue: Writing- Reviewing, Study design and Supervision. All authors read and approved the final manuscript.

Funding

Declarations

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/s41598-025-92695-1.

Correspondence and requests for materials should be addressed to W.W. or W.X.

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.