

Article

Elucidating the Prognostic and Therapeutic Implications of Insulin Resistance Genes in Breast Cancer: A Machine Learning-Powered Analysis

Lengyun Wei, Dashuai Li, Hongjin Chen, Yajing Pu, Qun Wang, Jintao Li, Meng Zhou, Chenfeng Liu * and Pengpeng Long *

School of Life Science, Anhui Medical University, Hefei 230032, China; weilengyun@ahmu.edu.cn (L.W.)

* Correspondence: chenfengliu0607@163.com (C.L.); longpp@mail.ustec.edu.cn (P.L.)

Simple Summary: The study investigated the association between insulin resistance and breast cancer through the application of various machine learning techniques. By analyzing data from a large patient cohort, seven key genes (LIFR, EZR, TBC1D4, NSF, RPL5, SAA1, and PGK1) were identified. Based on these genes, a prognostic model was constructed, which demonstrated a strong correlation with overall survival, clinical outcomes, and treatment responses. These findings suggest that the proposed model may serve as a valuable tool for clinicians in predicting treatment responses, thereby facilitating more personalized therapeutic strategies.

Abstract: Breast cancer (BC) is among the most prevalent malignancies and remains the leading cause of cancer-related mortality in women worldwide. While prior studies have highlighted the associations between insulin resistance (IR) and both tumorigenesis and cancer progression, the prognostic relevance of IR in BC has not been fully elucidated. In this study, we employed a suite of machine learning algorithms and statistical methods to construct a robust prognostic model for BC based on insulin resistance-related genes (IRGs). The model's prognostic value was subsequently validated in four independent validate cohorts, including METABRIC and three GSE datasets. The resulting IR signature, comprising seven hub IRGs (LIFR, EZR, TBC1D4, NSF, RPL5, SAA1, and PGK1), demonstrated high predictive power for overall survival (OS) across public datasets. Notably, a lower insulin resistance risk score (IRRS) was significantly associated with more favorable clinical outcomes, including enhanced responses to neoadjuvant therapy. Based on single-cell RNA sequencing data, we found that the hub genes were more enriched in T cells, B cells, and epithelial cells. Furthermore, we used machine learning methods to perform feature selection and reduction, which generated a clinically applicable scoring system consisting of the seven hub genes for predicting clinical outcomes in BC patients. This novel IR-based prognostic signature offers a valuable tool for stratifying BC patients by risk and tailoring personalized therapeutic strategies, thus enhancing precision oncology in breast cancer care.

Keywords: breast cancer; insulin resistance; risk score; overall survival; prognosis



Academic Editor: Ahmed Elbediwy

Received: 26 March 2025

Revised: 3 May 2025

Accepted: 9 May 2025

Published: 13 May 2025

Citation: Wei, L.; Li, D.; Chen, H.; Pu, Y.; Wang, Q.; Li, J.; Zhou, M.; Liu, C.; Long, P. Elucidating the Prognostic and Therapeutic Implications of Insulin Resistance Genes in Breast Cancer: A Machine Learning-Powered Analysis. *Biology* **2025**, *14*, 539. <https://doi.org/10.3390/biology14050539>

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Breast cancer (BC) is the most frequently diagnosed cancer and the second leading cause of cancer-related death among women worldwide [1,2]. Although adjuvant chemotherapy and hormonal therapies have significantly reduced BC mortality, accurately

pre-dicting individual patient prognosis, identifying those likely to benefit from specific treat-ments, and recognizing individuals at risk of adverse reactions remains a challenge [3]. Therefore, the discovery of novel and reliable prognostic biomarkers, as well as tools for monitoring treatment response, is of critical importance in improving outcomes for BC patients [4].

Insulin resistance (IR), defined as a systemic condition in which cells and tissues exhibit reduced responsiveness to the hormone insulin, affects multiple organs and regulatory pathways involved in metabolic homeostasis [5]. Numerous studies have documented a strong link between IR, visceral adiposity dysfunction, and systemic inflammation, all of which contribute to a carcinogenic environment [6]. Moreover, therapeutic strategies targeting key genes involved in IR have shown promising therapeutic effects in both preclinical models and clinical settings [7]. Recent evidence increasingly indicates that IR plays a significant role in the development and progression of breast cancer (BC) [8,9].

In this study, we developed a robust prognostic signature based on insulin resistance-related genes (IRGs) within a training cohort of BC patients. The prognostic value of this IR-based signature was subsequently validated across multiple independent cohorts. We further examined the correlation between the derived risk scores and various clinical parameters, including treatment outcomes, BC molecular subtypes, and immune check-point expressions. Furthermore, the predictive capacity of the IR signature was assessed in BC patients receiving chemotherapy, providing insight into its potential application in personalized treatment planning.

2. Materials and Methods

2.1. Data Collection and Processing

Transcriptome expression data and corresponding clinicopathological data for BC samples were obtained from the Cancer Genome Atlas (TCGA) and the Molecular Taxonomy of BC International Consortium (METABRIC). The distribution of major clinical features of TCGA-BRCA and METABRIC were visualized (Supplementary Figure S1). Additionally, eight BC cohorts were collected from the Gene Expression Omnibus (GEO) and the Single Cell Portal. After excluding patients lacking overall survival (OS) data, a comprehensive dataset comprising 6907 BC samples across 10 cohorts was compiled (Table 1). For the RNA sequencing data, fragments per kilobase of transcript per million mapped reads (FPKM) values were transformed into transcripts per million (TPM) values. Both RNA sequencing and microarray datasets underwent \log_2 transformation and Z-score normalization to normalize the data distribution.

Table 1. Cohorts involved in this study.

Cohort	Data Type	Source	Sample Size	Reference
TCGA-BRCA	RNASeq	TCGA	1095	[10]
METABRIC	microarray	cBioPortal	1906	[11]
GSE20685	microarray	GEO	327	[12]
GSE96058	RNASeq	GEO	3069	[13]
GSE7390	microarray	GEO	198	[14]
GSE191127	RNAseq	GEO	40	[15]
GSE20181	microarray	GEO	176	[16]
GSE18728	microarray	GEO	61	[17]
GSE225078	RNAseq	GEO	30	[18]
SCP1106	scRNA-seq	Single-cell portal	5	[19]

The training cohort for the prognostic model included 1095 BC patients from TCGA-BRCA. The validate cohort consisted of 5500 BC patients from four cohorts (METABRIC, GSE96058, GSE20685, and GSE7390). To evaluate the prognostic model's predictive accu-

racy regarding therapeutic response, four additional BC cohorts (GSE191127, GSE20181, GSE18728, and GSE225078) consisting of 307 patients treated with neoadjuvant therapy were analyzed.

Insulin resistance-related genes (IRGs) were retrieved from the GeneCards database using the keyword “insulin resistance” and default search parameters. Only protein-coding genes with a relevance score greater than 3.0 were included in further analysis (Supplementary Table S1).

2.2. Construction and Validation of an IRG Prognostic Signature for Breast Cancer Patients

To identify significant differentially expressed genes (DEGs) between breast cancer (BC) samples and normal controls, the TCGA-BRCA cohort were analyzed using the R package limma (version 4.1.0). DEGs were identified by a threshold of absolute log2 fold change (FC) > 0.5 and an adjusted p value < 0.01. Among these DEGs, insulin resistance-related genes (IRGs) were subjected to univariate Cox regression analysis in both the TCGA-BRCA and METABRIC cohorts (p value < 0.01 in each). The overlapping significant IRGs from both cohorts were selected to develop a prognostic model. Feature selection was conducted using least absolute shrinkage and selection operator (LASSO) Cox regression analysis, a widely applied method for penalized regression and variable selection. This process identified seven hub genes with non-zero coefficients for inclusion in the prognostic model. The model calculates an insulin resistance risk score (IRRS) for each BC patient using the formula: $IRRS = \sum E_i * \beta_i$, where E_i represents the expression level of the i th hub gene, and β_i is its corresponding LASSO regression coefficient. The expression levels of 7 hub genes in TCGA-BRCA were visualized using a heatmap generated by the R package ggplot2 (version 4.1.3) [20].

The BC patients were stratified into two groups based on the median IRRS value. To assess survival outcomes across these groups, Kaplan–Meier (K–M) survival analyses were conducted to compare overall survival (OS) between the two groups in both the training and validation cohorts. In the training cohort, additional K–M analyses were performed to evaluate differences in disease-free survival (DFS), disease-specific survival (DSS), progression-free interval (PFI), and progression-free survival (PFS) between the risk groups.

2.3. Development and Evaluation of an Insulin Resistance-Related Clinicopathological Nomogram

A clinicopathological nomogram was developed by integrating IRRS with other key clinical variables, including T stage, M stage, overall stage, and PAM50 molecular subtype. Univariate and multivariate Cox regression analyses were conducted for variable selection, resulting in the retention of two significant predictors—IRRS and PAM50 (p < 0.05)—for inclusion in the final model. The regression coefficients from the final model were converted into a point-based scoring system using the R package rms. A graphical nomogram was constructed using the R package regplot (version 1.1), with scales for each predictor and a total points scale to estimate 1-year, 3-year, 5-year, and 10-year survival probabilities. Calibration curve analysis and decision curve analysis (DCA) were performed to confirm the satisfactory predictive discrimination of the nomogram.

2.4. Tumor Microenvironment Characterization Analysis

To evaluate the tumor microenvironment (TME), including its stromal and immune components, the R package ESTIMATE (version 1.0.13) [21] was applied to the TCGA-BRCA dataset. This package calculated tumor purity, immune scores, and stromal scores, providing insight into the extent of immune and stromal cell infiltration in each breast cancer (BC) sample. Additionally, CIBERSORT was used to deconvolute the gene expression profiles and estimate the relative abundance of 22 immune cell types under default parameters [22].

The distribution of these immune cell populations was compared between high- and low-IRRS groups was investigated. Furthermore, four major immune cell categories—total lymphocytes, dendritic cells, macrophages, and mast cells—were analyzed to access their relative abundance across IRRS groups.

2.5. Differential Gene Expression of Low- and High-IRRS Groups and Functional Enrichment Analysis

Differential gene expression analysis between the high- and low-IRRS patient groups within the TCGA-BRCA cohort was conducted using the R package limma (version 4.1.0). Gene set enrichment analysis (GSEA) was performed using the R package clusterProfiler (version 4.2.2) to identify pathways significantly associated with IRRS groups. Additionally, Gene Ontology (GO) [23] and Kyoto Encyclopedia of Genes and Genomes (KEGG) term [24–26] enrichment analyses were carried out through the gseGO and gseKEGG functions in the R package clusterProfiler (version 4.2.2). The gseaplot2 function from the R package enrichplot (version 1.14.2) was used to visualize the GSEA results.

2.6. Single-Cell RNA Sequencing Data Analysis

The raw count matrix and associated metadata that for five BC samples (SCP1106 [19]) were downloaded from the Single Cell Portal (https://singlecell.broadinstitute.org/single_cell, accessed on 21 March 2024). All samples subjected to single-cell RNA sequencing originated from triple-negative breast cancer (TNBC) cases. Low quality cells were filtered out based on the following criteria: (1) fewer than 200 unique molecular identifiers (UMIs); (2) more than 6000 or fewer than 100 expressed genes; and (3) the percentage of mitochondrial UMI being less than 10%. Subsequently, the gene expression matrices were normalized using the Normalization function, and variable gene features were identified using the FindVariableFeatures function with default parameters. To integrate data across samples, the RunFastMNN function was applied following the developer's guidelines [27]. The top 15 principal components were used for subspace alignment and dimensionality reduction via uniform manifold approximation and projection (UMAP). Cell type annotation was conducted as previously described [19], and the analysis of single-cell RNA sequencing data was carried out using the R package Seurat (version 4.1) [28].

2.7. Ucell Analysis

The R package Ucell (version 2.0) was utilized to assess the enrichment of the IR signature at the single-cell level. An IR Ucell signature score, representing the enrichment of the seven hub genes in each cell, was calculated using the AddModuleScore_UCell function. Subsequently, cells were stratified into high and low groups based on their UCell scores. GSEA using a hallmark gene set was conducted to identify DEGs between the high- and low-Ucell-score groups, using the R package clusterProfiler (version 4.2.2). Normalized enrichment scores were visualized by R package ggplot2 (version 4.1.3), and Ucell score distributions were illustrated via violin plots.

2.8. Construction and Validation of the Machine Learning Model

The BC patients from TCGA-BRCA were randomly divided into training (70%) and testing (30%) sets using the sample.split function in the R package caTools. Pivotal IRGs involved in the prognostic model were taken as the initial features, and all data were standardized using z-score normalization. To develop a robust machine learning model, two algorithms—XGBoost and SVM—were employed.

For the XGBoost model, recursive feature elimination (RFE) with a preliminary XGBoost model was employed to refine the feature set. The final model was configured with the following parameters: (1) a learning rate (eta) of 0.4, (2) a maximum depth of trees of 2,

(3) 50 boosting rounds, and (4) a binary objective function. Hyperparameter optimization was performed using grid search with 5-fold cross-validation, and the optimal parameters were selected based on the highest average ROC-AUC score. The model was trained using the `xgb.train` function in the R package `xgboost`, and feature importance was identified using `xgb.importance` function.

In parallel, an SVM model was built using the `svm` function from the R package `e1071`. The radial basis function (RBF) kernel was used, and hyperparameters including the cost (C) and kernel coefficient (gamma) were optimized via grid search with 5-fold cross-validation, based on ROC-AUC performance. The SVM model was trained on the same set of features selected through RFE to ensure comparability with the XGBoost model.

External validation of the XGBoost and SVM model was performed using the METABRIC and GSE96058 cohorts. Expression data for the final feature set in these validation cohorts were standardized using z-score normalization and used to predict patient groupings via the trained XGBoost model. All model training and validation procedures were executed on a system equipped with an Intel i7 processor and 32 GB of RAM.

2.9. Patients and Specimens

Formalin-fixed, paraffin-embedded (FFPE) tissue sections from 10 patients for immunohistochemistry (IHC) staining were obtained from the First People's Hospital of Changzhou. Clinical samples included ER-positive, HER2-positive, and triple-negative cases across different tumor stages (I–III).

2.10. Immunohistochemistry

Immunohistochemistry (IHC) staining was performed as described previously [29]. Primary antibodies against CD163, CD8 α , and PGK1 (Zen-Bioscience Co., Ltd., Chengdu, China) were used in this study. All primary antibodies were incubated at 4 °C overnight, followed by incubation with a horseradish peroxidase-conjugated anti-rabbit secondary antibody. All sections were visualized using a DAB substrate. For imaging, slides were scanned with a Panoramic 250 Flash III scanner (3DHISTECH Ltd, Budapest, Hungary), and images were obtained using Panoramic Viewer 1.1 Software (3DHISTECH Ltd.).

2.11. Statistical Analysis

All statistical analyses were performed using R software (Version 4.1.0). Kaplan–Meier survival curves and univariate Cox proportional hazard regression analysis were conducted with the R packages `survival` (version 3.7.0) and `survivalminer` (version 3.4.0). The LASSO algorithm was executed using the R package `glmnet` (version 4.1.8). Multivariate binary logistic regression, nomogram construction, and calibration curve plotting were performed using the R package `rms` (version 7.0.0). Differences in IRG expression levels between normal and tumor tissues, as well as changes in IRRS before and after therapy, were assessed using Student's *t*-test. Additionally, associations between IRRS and clinicopathological features—including PAM50 subtypes, AJCC staging (T and N stages), and lymph node status—were evaluated using the Wilcoxon rank-sum test.

3. Results

3.1. Construction and Validation of an Insulin Resistance-Relevant Prognostic Signature for Breast Cancer Patients

Initially, 2828 insulin resistance-related genes (IRGs) were retrieved from GeneCards after data processing. Differential expression analysis comparing 1095 BC samples with 113 normal breast samples in the TCGA-BRCA cohort, revealed 6324 significant differentially expressed genes (DEGs). Among these, 1115 genes were recognized as significant IRGs and retained for further analysis. Univariate Cox hazard regression analysis then

isolated 54 and 290 significant IRGs within the TCGA-BRCA and METABRIC cohorts, respectively. A subsequent LASSO Cox regression analysis of the 24 IRGs common to both cohorts yielded a seven-gene prognostic signature comprising LIFR, EZR, TBC1D4, NSF, RPL5, SAA1, and PGK1 (Figure 1A,B). The expression levels of these genes, correlated with patient overall survival (OS), were visualized (Supplementary Figure S2). Using the established prognostic model, an insulin-related risk score (IRRS) was calculated for each patient by applying the following formula: $\text{IRRS} = (0.040 \times \text{expression of EZR}) + (-0.046 \times \text{expression of LIFR}) + (-0.138 \times \text{expression of TBC1D4}) + (-0.0105 \times \text{expression of SAA1}) + (0.0218 \times \text{expression of NSF}) + (-0.0566 \times \text{expression of RPL5}) + (0.464 \times \text{expression of PGK1})$. Patients were subsequently classified into low- or high-IRRS groups, based on the median IRRS value. A forest plot illustrating the relationship between gene expressions and risk of BC (Figure 1C). The analysis of hazard ratios within this plot identified the EZR, NSF, and PGK1 genes as prognostic markers in BC, with elevated expression (highlighted in red) linked to poorer outcomes. Notably, the combined IRRS factor demonstrated a significantly higher hazard ratio ($\text{HR} = 4.368$, 95% CI: 2.810–6.792; p value < 0.0001) when compared to any individual gene. A heatmap analysis further demonstrated the correlation between the expression levels of each IRG and clinical parameters within the TCGA-BRCA and METABRIC cohorts (Figure 1D, Supplementary Figure S3A).

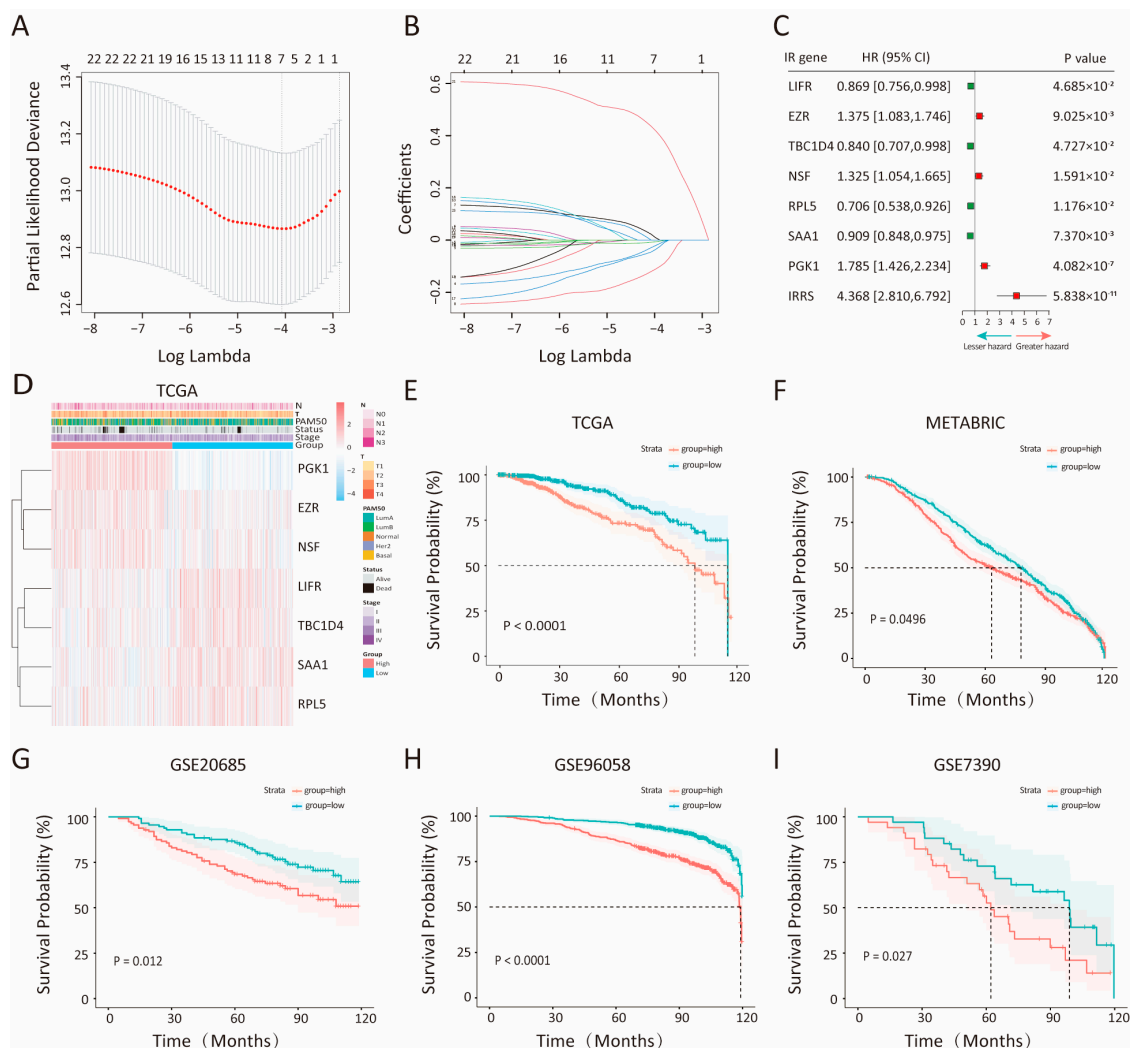


Figure 1. Construction and validation of the insulin resistance risk score (IRRS) model. (A) Partial likelihood deviance displayed by the LASSO regression model during 10-fold cross-validation. Vertical

dotted lines represent optimal values determined using minimum and 1-SE criteria. **(B)** LASSO coefficient profiles of 24 insulin resistance genes selected during 10-fold cross-validation. **(C)** Multivariate Cox analysis assessing the independent predictive capability of 7 hub IRGs and IRRS for overall survival (OS). Unadjusted hazard ratios (HRs) are provided with 95% confidence intervals. A lesser hazard direction signifies better survival, while a greater hazard direction indicates lower survival. **(D)** Heatmap illustrating the expression levels of 7 hub genes in the training cohort (TCGA-BRCA). The color scale ranges from red (indicating high-IRRS group) to blue (indicating low-IRRS group) to denote magnitude. The heatmap was generated by the R package ggplot2 (version 4.1.3). **(E–I)** Kaplan–Meier curves for high- and low-IRRS groups in the TCGA-BRCA **(E)**, METABRIC **(F)**, and cohorts GSE20658 **(G)**, GSE96058 **(H)**, and GSE7390 **(I)**.

To evaluate the prognostic value of the IRRS model, Kaplan–Meier (K-M) survival analysis was employed to compare overall survival (OS) between the low- and high-IRRS groups across the training cohort and four independent validation cohorts (Figure 1E–I). In all datasets, patients classified in the high-IRRS group exhibited significantly reduced OS compared to those in the low-IRRS group, as evidenced by log-rank test results (TCGA: $p < 0.0001$; METABRIC: $p < 0.048$; GSE20685: $p = 0.012$; GSE96058: $p = 1.8 \times 10^{-11}$; GSE7390: $p = 0.027$). Further validation of the IRRS model's robustness was conducted through its performance assessment in predicting disease-specific survival (DSS), disease-free interval (DFI), progression-free interval (PFI), and progression-free survival (PFS) within the TCGA-BRCA cohort. Consistently, high IRRS was significantly associated with poorer survival outcomes (DSS of TCGA: p value = 1.4×10^{-4} ; DFI of TCGA: p value = 5.8×10^{-3} ; PFI of TCGA: p value = 8.1×10^{-5} ; PFS of TCGA: p value < $8.1 \times 10^{-0.5}$; log-rank test), as demonstrated in Supplementary Figure S3B–E. In addition, the association between IRRS and overall survival (OS) across different breast cancer subtypes was explored in Supplementary Figure S3F–J. Collectively, these findings underscore the IRRS model as a significant predictive marker for survival outcomes in BC patients.

3.2. Integrated Assessment of the Prognostic Model and Clinical Parameters in Patients with Breast Cancer

To further evaluate the clinical relevance of the IRRS model, IRRS values were computed for each patient in the TCGA-BRCA, METABRIC, and GSE96058 cohorts, and their associations with key clinical parameters were analyzed. In the training cohort, a significant association was observed between the IRRS and critical clinical outcomes, such as PAM50 subtype, survival status, tumor stage, and clinicopathological T stage. Notably, an increase in the IRRS corresponded to a worsening of these clinical indicators (Figure 2A). These findings were validated in the METABRIC cohort, where IRRS was significantly associated with PAM50 subtype, tumor stage, tumor size, and tumor grade (Figure 2B). Similar trends were observed in the GSE96058 cohort, where higher IRRS values were predictive of more aggressive PAM50 subtypes, shorter survival, larger tumor size, and greater lymph node involvement (Figure 2C). Collectively, these findings highlight the strong alignment between IRRS and established clinical parameters, indicating that higher IRRS values are associated with adverse clinical outcomes in BC patients.

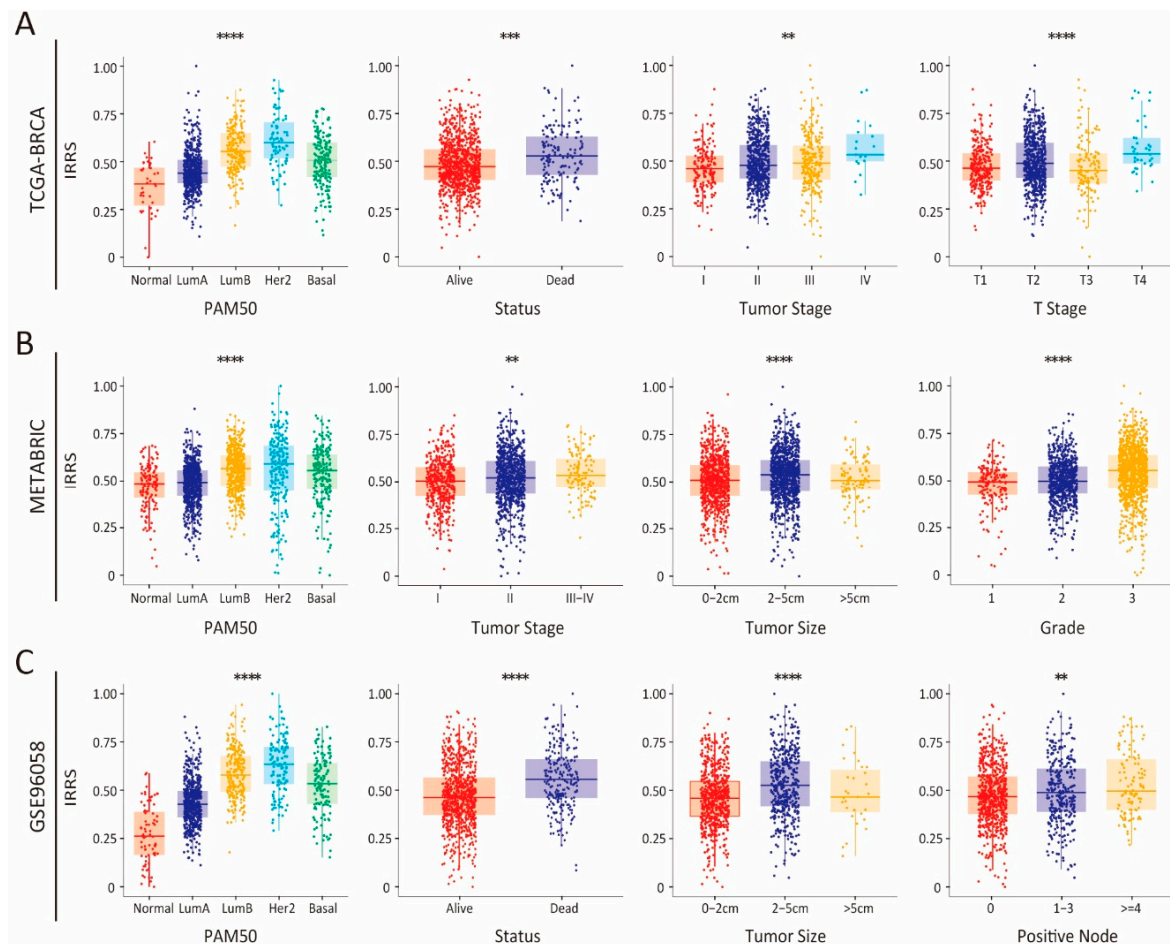


Figure 2. Comprehensive evaluation of the IRRS and clinical parameters in breast cancer (BC) patients. Correlation between IRRS levels and diverse clinical parameters in the TCGA-BRCA (A), METABRIC (B), and GSE96058 (C) cohorts. ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

3.3. Development and Evaluation of an Insulin Resistance-Related Clinicopathologic Nomogram

To enhance individualized prognostication, nomograms incorporating both IRRS and PAM50 subtype were developed for BC samples within the TCGA-BRCA and METABRIC cohorts [30]. These nomograms were designed to estimate 1-, 3-, 5-, and 10-year overall survival (OS) for individual patients. The score for each factor reflects its respective contribution to survival outcome prediction, with results indicating a more significant contribution from IRRS compared to PAM50 subtype (Figure 3A, Supplementary Figure S4A). To evaluate the predictive accuracy of the nomograms, calibration curves were generated by comparing predicted survival probabilities with actual outcomes. Except for the 10-year OS predictions, the calibration curves showed good agreement with the ideal reference line for both the TCGA-BRCA and METABRIC cohorts (Figure 3B–E, Supplementary Figure S4B–E). Furthermore, decision curve analysis for 1-, 3-, 5-, and 10-year survival demonstrated that the nomograms offer a greater net benefit over standard treat-none or treat-all approaches (Figure 3F). The predictive performance of the nomogram was further quantified using the concordance index (C-index), which was 0.64 in the TCGA-BRCA cohort and 0.59 in the METABRIC cohort. Notably, our IRRS-based gene signature outperformed several previously published breast cancer biomarkers, as evidenced by its superior C-index values (Table 2).

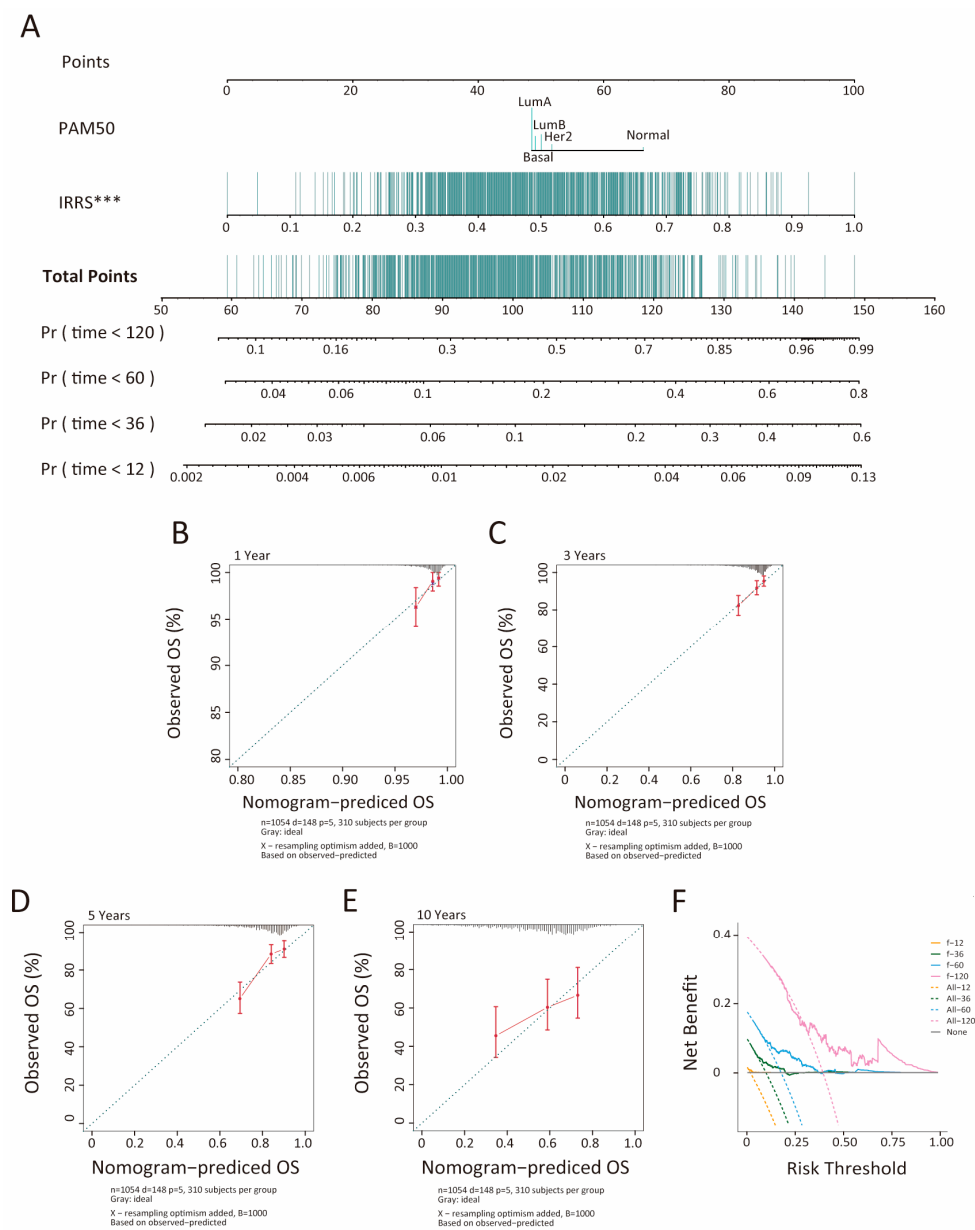


Figure 3. Nomogram developed to predict the probability of 1-, 3-, 5-, and 10-year overall survival in the TCGA-BRCA cohort. (A) Nomogram constructed for overall survival prediction in the TCGA-BRCA cohort. Calibration curves depict agreement between predicted and observed 1- (B), 3- (C), 5- (D), and 10-year (E) survival outcomes. Decision curve analysis (DCA) assesses 1-, 3-, 5-, and 10-year risks (F). *** $p < 0.001$.

Table 2. C-index of nomogram constructed by IRRS and published signatures.

Signature	C-Index	Reference
Yang.FG	0.56	[31]
Zhang.FG	0.53	[32]
Wang.FG	0.58	[33]
Wang.BB	0.63	[34]
IRRS	0.64	[35]

3.4. Association of TME Subcomponents with IRRS and Outcome in Patients with Breast Cancer

The tumor microenvironment (TME) plays a critical role in cancer progression and significantly impacts patient survival outcomes [36]. To explore differences in TME composition between high- and low-IRRS groups, we utilized the ESTIMATE algorithm to

quantify stromal and immune components. As shown in Figure 4A, the tumor purity of the low-IRRS group was significantly lower compared to the high-IRRS group, corroborating the observation that higher IRRS is associated with poorer survival outcomes. This observation supports the association between elevated IRRS and poorer prognosis. Furthermore, the results demonstrated that the low-IRRS group had markedly higher immune and stromal scores than their high-IRRS counterparts.

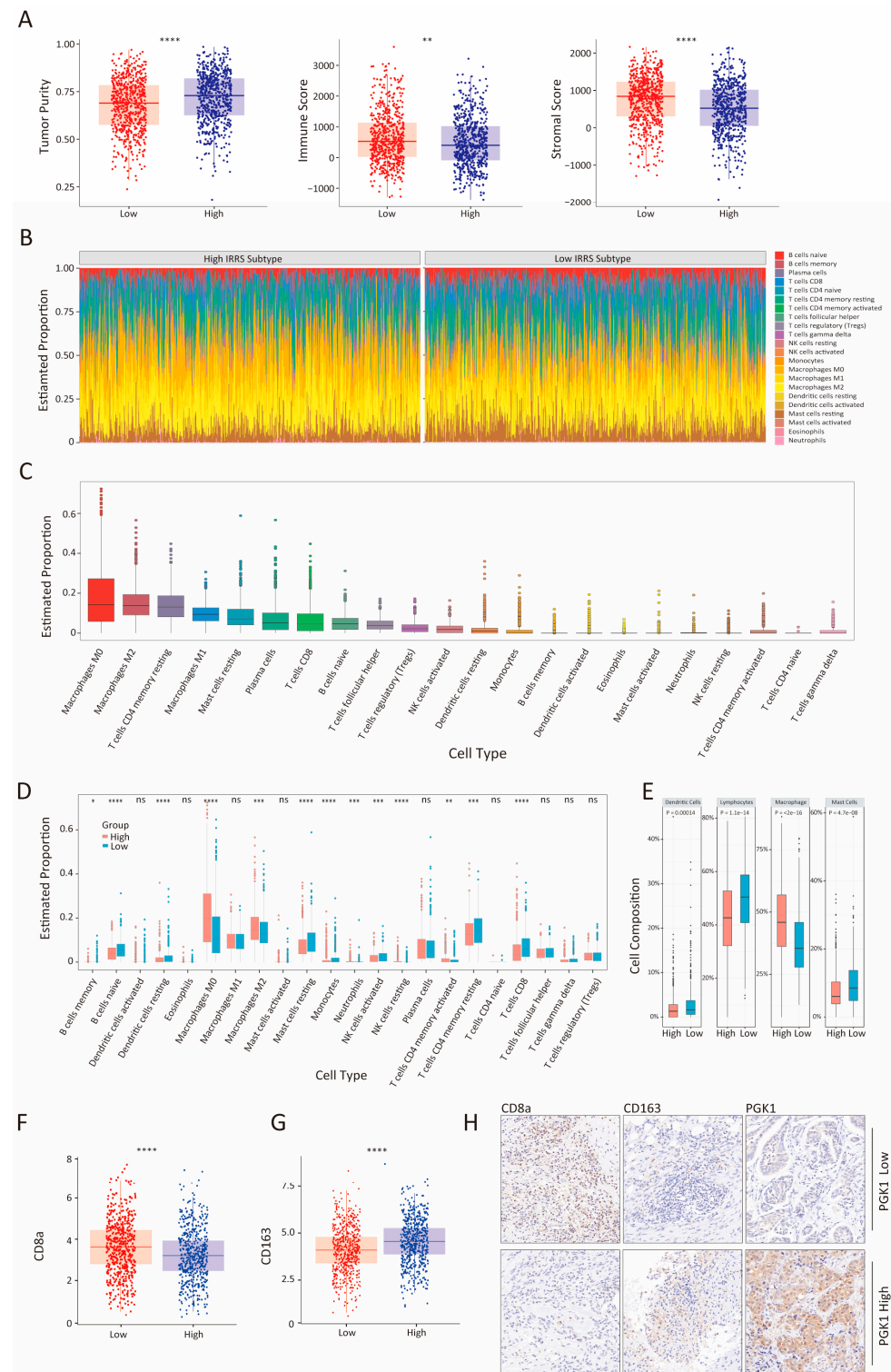


Figure 4. Analysis of immune cell proportions in high- and low-IRRS patients within the TCGA-BRCA cohort. (A) Boxplots illustrating tumor purity, immune score, and stromal score differences between

low- and high-IRRS groups. **(B)** Proportions of 22 immune cell types infiltrating low- and high-IRRS groups. **(C)** Boxplot illustrating proportions of 22 immune cell types within the TCGA-BRCA cohort. **(D)** Distribution of proportions of 22 immune cell types between low- and high-IRRS groups. **(E)** Boxplots representing cell composition of total dendritic cells, total lymphocytes, total macrophages, and mast cells within the TCGA-BRCA cohort. Statistical significance: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$; ns (not significant), $p > 0.05$. **(F,G)** mRNA expression level of CD8a and CD163 in low- and high-IRRS groups in TCGA. **(H)** CD8a, CD163, and PGK1 protein levels in BC tissues and adjacent normal tissues were examined by IHC. Representative images were shown (magnification at $\times 40$). Scale bar: 20 μm .

To further characterize immune infiltration differences between low- and high-IRRS groups, we analyzed the distribution of immune cell types across both groups. Figure 4B presents the relative proportions of immune cell populations, while Figure 4C displays these distributions as boxplots. A comparative analysis comparing the proportions of each cell type between the groups is shown in Figure 4D. Notably, the high-IRRS group exhibited significantly higher proportions of M0 macrophages, M2 macrophages, neutrophils, resting NK cells, and activated memory CD4+ T cells compared to the low-IRRS group. Conversely, the low-IRRS group had higher proportions of memory B cells, naive B cells, resting dendritic cells, resting mast cells, monocytes, activated NK cells, resting memory CD4+ T cells, and CD8 T cells (p value < 0.05 ; Wilcoxon test).

Additionally, to provide a broader perspective, four aggregated immune cell categories were also compared between groups, including total lymphocytes, total dendritic cells (sum of activated and resting), total macrophages (sum of M0, M1 and M2), and total mast cells (sum of activated and resting) between the two groups, as illustrated in Figure 4E. The analysis revealed that the low-IRRS group had significantly higher levels of total dendritic cells, total lymphocytes, and total mast cells, while the high-IRRS group had a greater total macrophage count. Furthermore, gene expression analysis of the TCGA-BRCA dataset indicated a significant reduction in CD8a expression within the high-IRRS group (Figure 4F), suggesting a diminished cytotoxic CD8+ T cell response, which is critical for anti-tumor immunity. Conversely, CD163 expression, a marker of immunosuppressive M2 macrophages, was significantly upregulated in the high-IRRS group (Figure 4G), indicating an immune-evasive microenvironment (Figure 4G). Immunohistochemistry (IHC) staining on human BC tissues for CD8a, CD163, and PGK1 further detailed the correlation between PGK1 expression and immune cell infiltration. High PGK1-expressing tumors were infiltrated more by CD163+ macrophages and less by CD8+ T cells. Contrastingly, low PGK1 expression areas showed increased CD8+ T cell infiltration and fewer CD163+ macrophages (Figure 4H). This inverse relationship between PGK1 expression and CD8+ T/CD163+ cell infiltration suggests that PGK1 may play a role in modifying the BC immune environment, indicating a significant immunosuppressive TME in the high-IRRS group that may promote immune escape and worsen BC prognosis. These findings highlight a strong immunosuppressive tumor microenvironment (TME) in the high-IRRS group, potentially facilitating tumor cell immune escape and leading to a poorer prognosis in BC patients.

3.5. Bioinformatic Analysis of the Characteristics and Signaling Pathways Among Patients in Different Risk Groups

To elucidate the biological distinctions between the low- and high-IRRS groups, we conducted a comprehensive analysis of signaling pathways enriched among differentially expressed genes (DEGs). Initially, 1470 significant DEGs were pinpointed using criteria of absolute $\log_2\text{FC} > 1$ and $\text{FDR} < 0.01$, of which 501 were upregulated and 969 were downregulated in BC samples (Figure 5A). KEGG pathway analysis indicated that the high-IRRS group was associated with pathways including PI3k-Akt signaling, cell cycle, focal adhesion, extracellular matrix (ECM)-receptor interaction, HIF-1 signaling, PPAR signaling,

and regulation of lipolysis in adipocytes (Figure 5B). GO analysis focusing on the biological process (BP) highlighted nuclear division as a primary activity in the high-IRRS group, while cellular component (CC) analysis showed enrichment in collagen-containing ECM. Analysis of molecular function (MF) revealed a significant role in signaling receptor activator activity (Figure 5C). Further, gene set enrichment analysis (GSEA) of GO, KEGG, and hallmark gene sets revealed additional pathways significantly enriched in the high-IRRS group, including chromosome segregation, DNA replication, endothelial cell proliferation, and diverse metabolic processes (Figure 5D). Specifically, GSEA of KEGG terms indicated that the high-IRRS group activated pathways like nucleotide sugars biosynthesis, cell cycle, and DNA replication (Figure 5E), along with hallmark pathways such as adipogenesis, bile acid metabolism, and glycolysis (Figure 5F). To further investigate the relationship between the high-IRRS group and insulin resistance, we performed GSEA on the insulin-related pathway. As shown in Supplementary Figure S5A–D, the pathways, including insulin signaling pathway, response to insulin, and PI3K-AKT signaling, were down-regulated in the high-IRRS group. Together, these findings highlight the distinct molecular landscapes of the low- and high-IRRS groups. The high-IRRS subtype is characterized by upregulation of cell proliferation, metabolic reprogramming, and ECM remodeling pathways, underscoring the complex and multifaceted biological processes that may drive poorer prognosis in these patients.

3.6. Therapeutic Benefit of the IRG Prognostic Signature

To evaluate the predictive utility of the IRG-based prognostic signature in the context of neoadjuvant chemotherapy, we analyzed pre- and post-treatment IRRS values across four breast cancer (BC) cohorts: GSE191127, GSE20181, GSE18728, and GSE225078. In the GSE191127 cohort, pretreatment IRRS was elevated compared to post-treatment IRRS (Figure 6A, p value = 1.8×10^{-2} ; T test), indicating a reduction in IRRS following chemotherapy. Additionally, comparisons between responders and nonresponders, as well as between recurrent and nonrecurrent cases, confirmed that lower IRRS values were associated with better treatment response and reduced recurrence risk (Figure 6B,C, nonresponse vs. response: p value = 3.8×10^{-2} ; nonrecurrent vs. recurrent: p value = 1.5×10^{-3} ; T test). In the GSE20181 cohort, significant IRRS reduction was observed when comparing prechemotherapy levels with those on Day 14 and Day 90 post-treatment (Figure 6D,E, p value = 4.3×10^{-1} for before vs. Day 14; p value = 3.2×10^{-4} for before vs. Day 90; T test), with the 14-day treatment group displaying a notably higher IRRS than the 90-day treatment group (Figure 6F, p value = 1.7×10^{-5} for Day 14 vs. Day 90; T test). In the GSE18728 cohort, a significant reduction in IRRS was also observed after one cycle of adjuvant chemotherapy (C1) and in the definitive surgical specimen (OR) among patients treated with docetaxel (T) and capecitabine (X) (Figure 6G,H, p value = 2.3×10^{-2} for C1 vs. BL; p value = 6.9×10^{-3} for OR vs. BL; T test). Similarly, in the GSE225078 cohort, patients treated with pembrolizumab plus paclitaxel exhibited a substantial decrease in IRRS compared to prechemotherapy levels (Figure 6I, p value = 1.5×10^{-3}). These findings collectively demonstrate that a lower IRRS is associated with increased therapeutic efficacy, suggesting that IRRS may serve as a useful biomarker for predicting responses to neoadjuvant chemotherapy in BC patients.

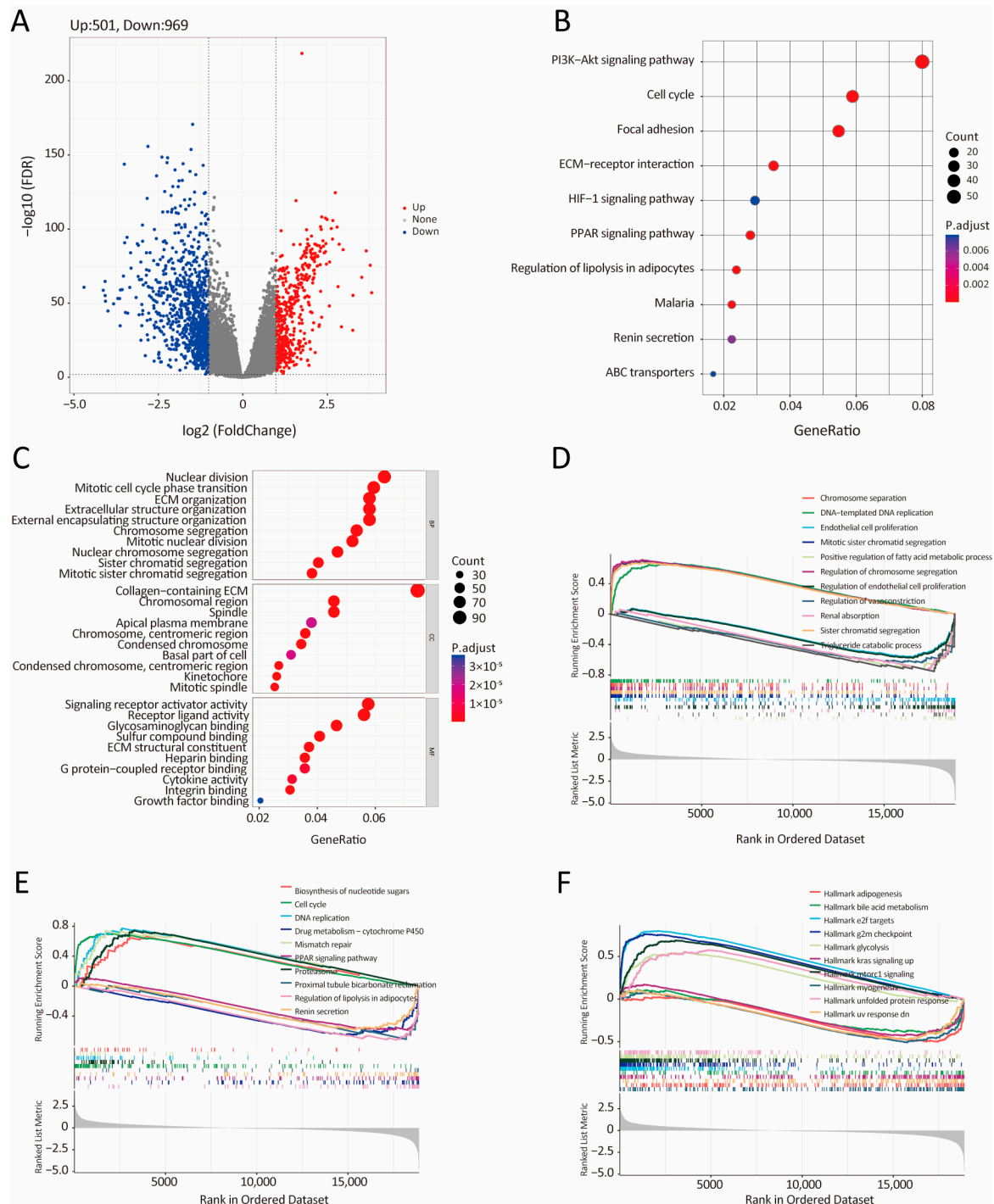


Figure 5. Bioinformatics analysis of characteristics and signaling pathways among patients in distinct risk groups. (A) Volcano plot illustrating differentially expressed genes (DEGs) in the high-IRRS group compared to the low-IRRS group within the TCGA-BRCA cohort. Genes labeled in red or green indicate significant up- or downregulation, respectively. FC: fold change; FDR: false discovery rate. (B) Kyoto Encyclopedia of Genes and Genomes (KEGG) [24–26] analysis and (C) Gene Ontology (GO) analysis of significant DEGs. Gene set enrichment analysis (GSEA) enriched in GO (D), KEGG (E), and hallmark (F) gene sets.

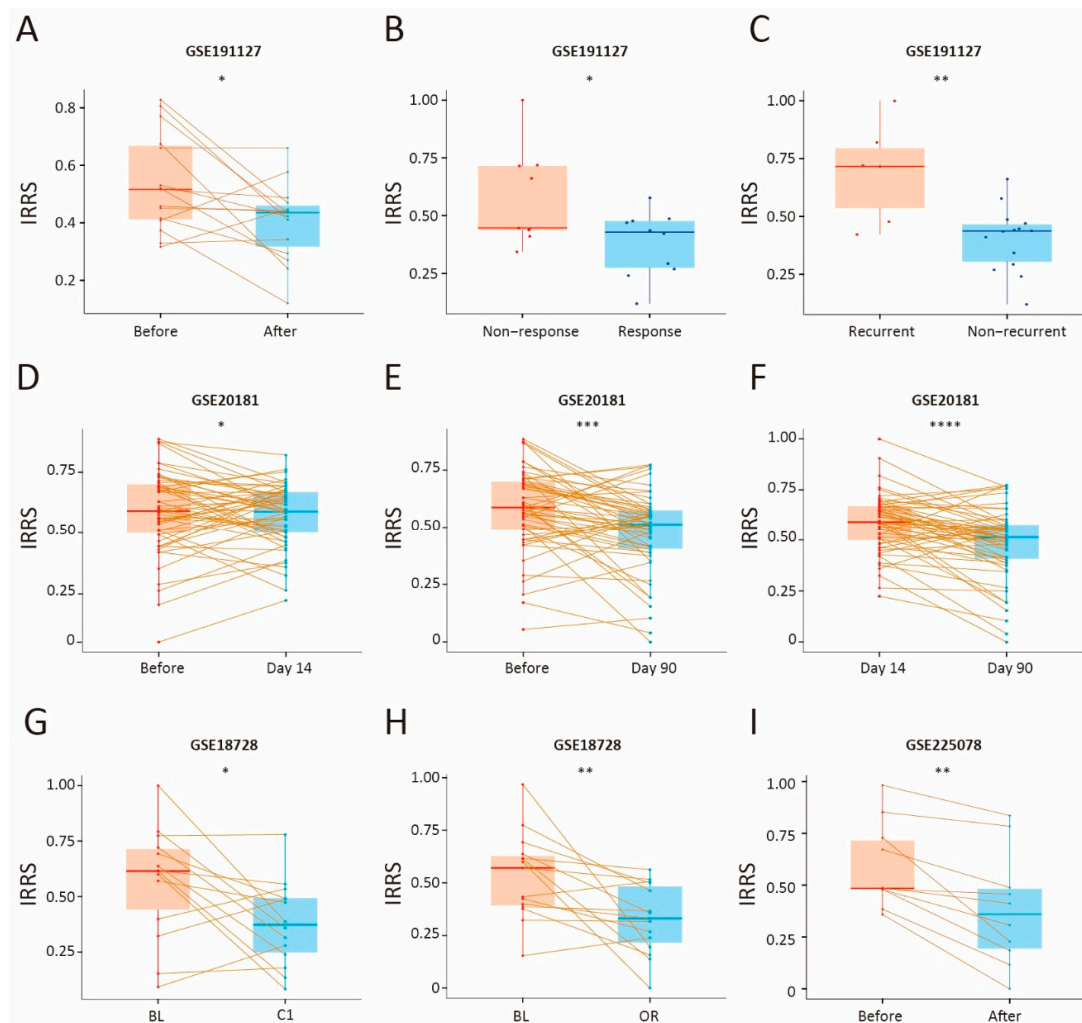


Figure 6. Therapeutic benefit assessment of the IRRS value. Pairwise comparison of IRRS between pre- and post-chemotherapy patients in GSE191127 (A–C), GSE20181 (D–F), GSE18728 (G,H), and GSE225078 (I). Significance assessed using pairwise t-test. Statistical significance denoted as * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$.

3.7. Single-Cell RNA Sequencing Data Analysis

To further dissect the cellular landscape underlying the IRRS signature, we performed single-cell RNA sequencing (scRNA-seq) analysis on five breast cancer samples, resulting in 24,271 high-quality cells after rigorous quality control. Unsupervised clustering revealed 19 transcriptionally distinct clusters (Figure 7A). Additionally, cell type annotation identified eight major cell populations: T cells, myeloid cells, epithelial cells, plasma cells, B cells, cancer-associated fibroblasts (CAFs), endothelial cells, and perivascular-like cells (PVLs) (Figure 7B). The expression levels of the seven IRRS hub genes across each cell type are illustrated in Figure 7C,D. To quantify the activity of these genes at the single-cell level, an IR Ucell score was calculated for each cell based on seven hub genes using the R package UCell (Figure 7E). Subsequent analysis divided the cells into two categories based on the IR Ucell score: top 25% (high IR Ucell score) and bottom 25% (low IR Ucell Score), as demonstrated in Figure 7F. Interestingly, the high-IR-Ucell-score group showed a higher proportion of T cells, epithelial cells, and B cells compared to the low-score group (Figure 7G). Furthermore, GSEA based on the hallmark gene set revealed that, within epithelial clusters, high-IR-Ucell-score cells exhibited significant enrichment of cancer-associated pathways such as epithelial homeostasis and the unfolded protein response (Figure 7H). These single-cell data reinforce the relevance of the IRRS signature in defining functional heterogeneity

within the tumor microenvironment and highlight its potential role in shaping immune and epithelial cell dynamics in breast cancer.

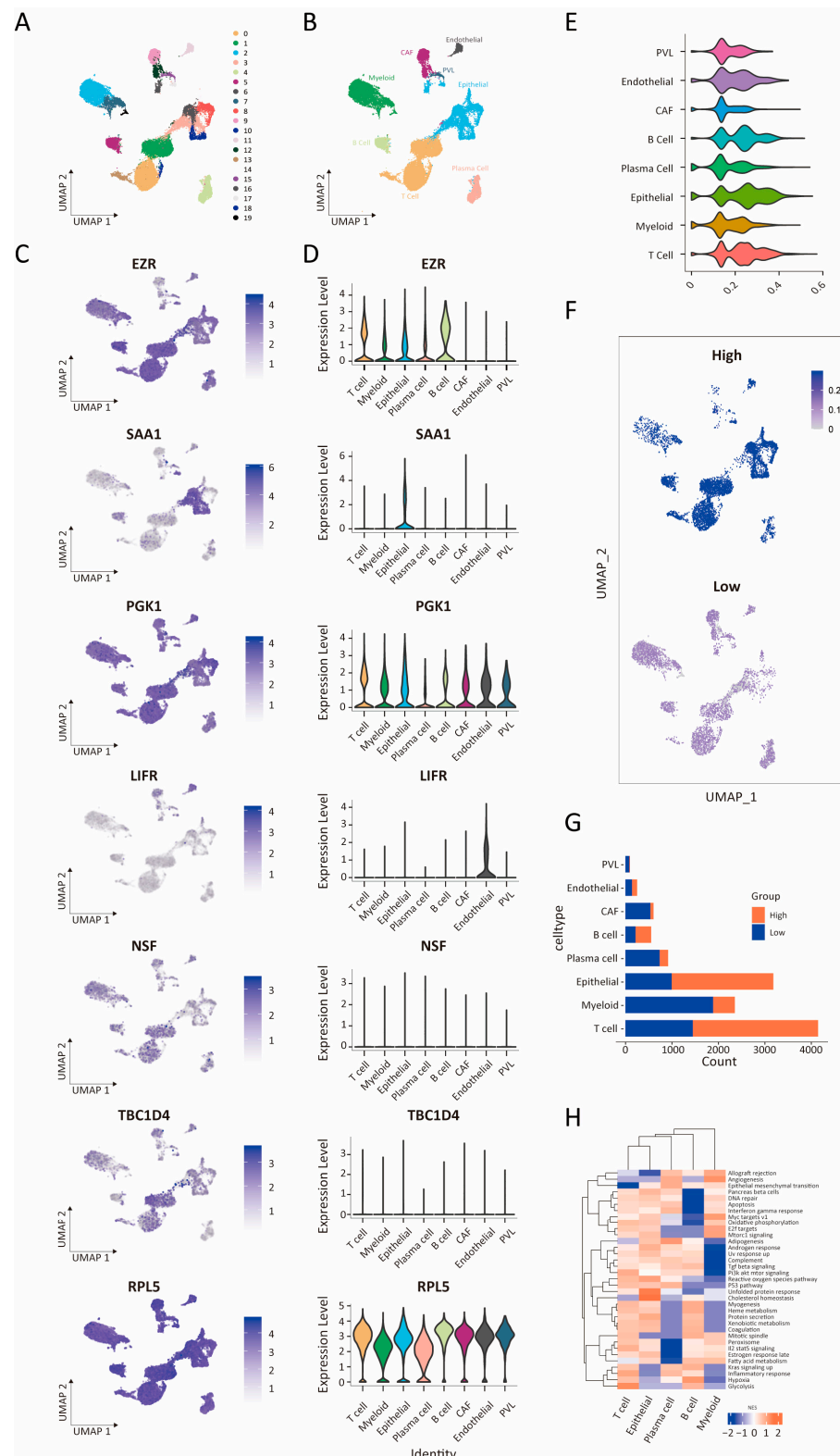


Figure 7. Analysis of 7 IRGs using single-cell RNA sequencing (scRNA-seq). **(A)** Clustering of cells from 5 breast cancer samples into 19 clusters. **(B)** Subclustering of aggregated scRNA-seq data from 5 breast cancer samples into T cell, myeloid cell, epithelial cell, plasma cell, B cell, cancer-associated fibroblast (CAF), endothelial cell, and perivascular-like cell (PVL). **(C,D)** Feature plots and violin plots

displaying expression level distributions of 7 IRGs across different cell types. (E) Violin plot showing IR Ucell scores across various cell types. (F) Distribution of IR Ucell scores among cells, divided into a high-IR-Ucell-score group (top 25%) and a low-IR-Ucell-score group (bottom 25%). (G) Distribution of cells in the high-IR-Ucell-score or low-IR-Ucell-score groups. (H) Heatmap displaying pathways enriched by GSEA analysis in the hallmark gene set. The heatmap was generated by the R package ggplot2 (version 4.1.3, <https://github.com/tidyverse/ggplot2>, accessed on 15 May 2024).

3.8. Prediction of the Low- and High-IRRS Subtypes by the XGBoost Algorithm

To develop a robust classifier for distinguishing between low- and high-IRRS groups in BC samples, we implemented and compared two machine learning algorithms: extreme gradient boosting (XGBoost) and a support vector machine (SVM). Both models utilized the seven hub genes from the IRRS signature as input features. For the XGBoost classifier, Shapley additive explanations (SHAP) were used to interpret the marginal contributions of each gene during a 10-fold cross-validation process (Figure 8A). The TCGA-BRCA cohort was randomly split into training (80%) and validation (20%) sets. SHAP analysis identified PGK1 and TBC1D4 as the most influential genes in the mode (Figure 8B). Performance evaluation of the validation set showed an overall accuracy of 0.86 and an area under the receiver operating characteristic curve (AUC) of 0.94 (Figure 8C). To assess the model's capacity to classify new data accurately, we also analyzed the GSE96058 and METABRIC cohorts. Figure 8D presents the overall accuracy and AUC of GSE96058, which were determined to be 0.88 and 0.83, respectively. However, the overall accuracy and AUC of METABRIC, as shown in Figure 8E, were 0.74 and 0.70, respectively, and were comparatively lower than those of GSE96058.

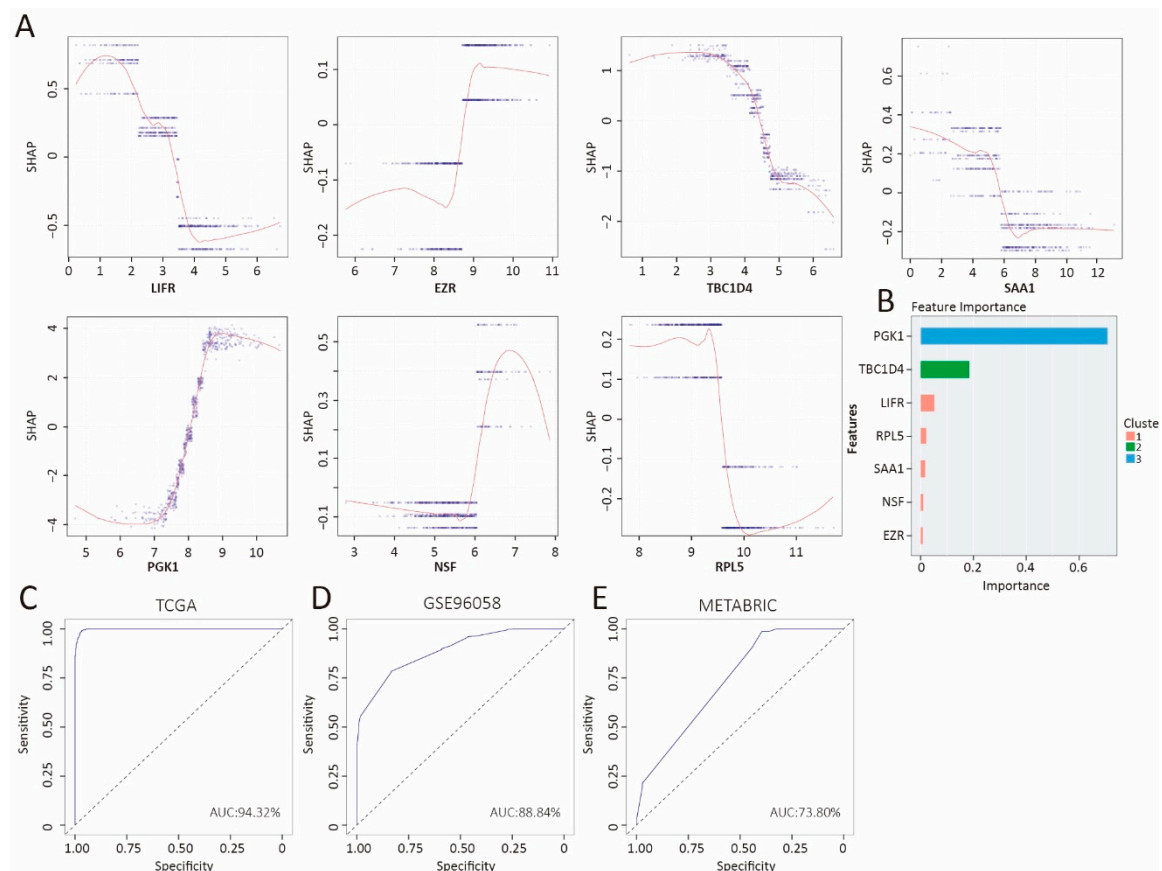


Figure 8. Prediction outcomes of the XGBoost algorithm. (A) SHAP contribution dependency plots depicting the seven crucial features. The x-axis represents the log2(TPM) of each feature. (B) Contributions

of the seven crucial features for the TCGA-BRCA cohort. ROC curves generated by the XGBoost algorithm for predicting high- and low-IRRS subtypes in (C) the TCGA-BRCA, (D) GSE96058, and (E) METABRIC cohorts.

In parallel, we constructed an SVM classifier using the R package *e1071* with a radial basis function (RBF) kernel. In the TCGA-BRCA dataset, the SVM model achieved an accuracy of 0.80 and an AUC of 0.91—slightly lower than XGBoost (Supplementary Figure S5E). SVM also showed worse performance than XGBoost in the validation cohorts GSE96058 (0.80 for accuracy and 0.77 for AUC) and METABRIC (0.69 for accuracy and 0.74 for AUC) (Supplementary Figure S5F–G). Compared to SVM, XGBoost consistently demonstrated higher predictive accuracy across the datasets.

4. Discussion

Mounting evidence indicates that insulin resistance (IR) within both cancer cells and the tumor microenvironment (TME) plays a pivotal role in tumor progression, recurrence, metastasis, and resistance to various therapies, including chemotherapy, endocrine therapy, and radiotherapy [37–41]. However, the critical molecular markers linking IR to breast cancer (BC) outcomes remain insufficiently characterized. In this study, we developed and validated a prognostic signature based on seven insulin resistance-related genes (IRGs) that is predictive of overall survival (OS) and therapeutic response in BC patients.

Among the seven IRGs, several have been reported in previous studies [42–44]. These IRGs not only play a role in regulating insulin levels but also partake in the intricate regulation of BC progression, particularly influencing cell proliferation and metastasis [45–47]. For instance, studies have shown that inhibition of PGK1 (phosphoglycerate kinase 1) phosphorylation suppresses glycolysis, tumorigenesis, and cell proliferation in various cancers [48,49]. Moreover, PGK1 silencing has been demonstrated to increase paclitaxel sensitivity in triple-negative BC through XAF1-mediated apoptosis [50]. Feedback activation of the leukemia inhibitory factor receptor (LIFR) signaling pathway in BC can limit the inhibitory response to HDAC inhibitors, marking it as a potential therapeutic target [51]. Furthermore, *in vitro* overexpression of ribosomal protein L5 (RPL5) has been shown to significantly inhibit BC cell proliferation, the G1-S cell cycle transition, and induce apoptosis [52]. TBC1D4 (AS160), a key regulator of GLUT4 translocation, remains underexplored in BC but may contribute to insulin-mediated metabolic alterations in tumors. Despite their individual contributions, prognostic prediction based solely on single IRG expression remains limited due to the complexity of tumor biology. Our findings emphasize that cancer progression is likely influenced by a coordinated network of genes, reinforcing the need for integrated models such as the IRRS signature.

The IRRS model effectively stratified patients by survival risk and showed strong predictive performance across multiple cohorts. Nomogram analysis further validated that the IRG signature possessed high precision in forecasting 1-year, 3-year, 5-year, and 10-year OS rates for BC patients, demonstrating its robust potential for clinical application. A primary challenge in current BC treatment is overcoming therapeutic resistance and metastasis, with many patients experiencing disease progression after initial treatment success in clinical settings [53,54]. Consequently, this study assessed the prognostic value of the IRG signature in predicting patient response to neoadjuvant chemotherapy. Findings indicated that the IRG signature was an effective biomarker for assessing treatment efficacy in various patient groups, including those pre- and post-therapy (GSE191127 and GSE225078), with differing survival outcomes (GSE20181), and those categorized by treatment response or nonresponse, and recurrence or nonrecurrence (GSE191127). Consistently, patients exhibiting a low IRRS posttreatment, those with long-term survival, and those

who responded to treatment or did not experience recurrence were more likely to benefit significantly from therapeutic interventions.

Nonetheless, this study has several limitations. First, while data normalization (z-score normalization and log2 transformation) was applied to account for platform-specific differences (RNA-seq vs. microarray), the use of multiple public datasets introduces potential batch effects and clinical heterogeneity. Additionally, intratumoral and interpatient heterogeneity, known to affect treatment responses, were not fully accounted for in our analyses. Third, the single-cell RNA sequencing analysis was based on a relatively small sample size with limited genetic diversity, which may constrain the generalizability of our immune landscape findings. Lastly, the biological functions and mechanistic roles of the seven IRGs were inferred through bioinformatic analyses rather than experimentally validated. Future studies should focus on functional validation of these genes through *in vitro* and *in vivo* experiments. Investigate the potential of small molecules or biologics that specifically target the identified IRGs, screening for modulators of their expression or activity, and evaluating their therapeutic potential in preclinical models are important next steps. Additionally, assessing the synergistic effects of combining IRG-targeted therapies with existing treatment modalities, such as chemotherapy, endocrine therapy, and immunotherapy, will help determine if these combinations can enhance therapeutic outcomes.

5. Conclusions

In conclusion, this study identified and validated a robust prognostic signature based on insulin resistance-related genes, capable of predicting survival outcomes and therapeutic responses in breast cancer patients. This model holds promise for clinical application, potentially enhancing OS rates and facilitating the development of personalized treatment strategies incorporating targeted therapeutics.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biology14050539/s1>, Table S1: The insulin resistance-related genes downloaded from GeneCards. Figure S1: The distribution of clinical features in TCGA-BRCA (A–E) and METABRIC (F–I). Figure S2: (A–G) mRNA expression levels of 7 hub IRGs; (H–N) Kaplan–Meier curves for 7 hub IRGs. Figure S3: (A) Heatmap illustrating the expression of 7 hub IRGs in the METABRIC cohort. (B–E) Kaplan–Meier curve estimates for DSS, DFI, PFS, and DFI in the TCGA-BRCA. (F–G) Kaplan–Meier curve estimates of OS in different PAM50 subtype. Figure S4: (A) Nomogram for predicting overall survival in the METABRIC cohort. (B–D) Calibration plots demonstrating agreement between predicted and observed 3-, 5-, and 10-year overall survival outcomes. (E) Decision curve analysis (DCA) assessing 3-, 5-, and 10-year risks. Figure S5: (A–D) The GO enrichment results on PI3K-AKT signaling in cancer (A), response to insulin (B), insulin signaling pathway (C), and insulin resistance (D). (I, J) The ROC curve of SVM on TCGA-BRCA (I), GSE96058 (J), and METABRIC (G).

Author Contributions: L.W., C.L. and P.L. contributed to the conception, design, analyzed the data, prepared the figures and wrote the manuscript; P.L. and M.Z. contributed to the software and data collection; D.L., Y.P., J.L. and Q.W. prepared the figures; M.Z. and H.C. are responsible for the execution and data analysis; L.W. supervised the study. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Natural Science Major Research Project of Anhui Educational Committee (2022AH040105) and the Natural Science Key Research Project of Anhui Educational Committee (2022AH050695).

Institutional Review Board Statement: This study was approved by the Ethics Committee of Anhui Medical University (Approval code: 82240070, approved on 05-08-2024). All research was performed in accordance with relevant guidelines and regulations, and written informed consent was obtained

from all patients. Research involving human research participants was conducted in accordance with the Declaration of Helsinki.

Informed Consent Statement: Informed consent for participation was obtained from all subjects involved in the study.

Data Availability Statement: All the RNA sequencing data of breast cancer were acquired from the Gene Expression Omnibus database (GEO, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>, accessed on 24 January 2024).

Acknowledgments: We would like to thank Yang Xu from the University of Science and Technology of China and Panpan Zhu from Anhui YiXinXing Bio for their valuable technical consultation.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Loibl, S.; Poortmans, P.; Morrow, M.; Denkert, C.; Curigliano, G. Breast cancer. *Lancet* **2021**, *397*, 1750–1769. [\[CrossRef\]](#)
- Trapani, D.; Ginsburg, O.; Fadelu, T.; Lin, N.U.; Hassett, M.; Ilbawi, A.M.; Anderson, B.O.; Curigliano, G. Global challenges and policy solutions in breast cancer control. *Cancer Treat. Rev.* **2022**, *104*, 102339. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kerr, A.J.; Dodwell, D.; McGale, P.; Holt, F.; Duane, F.; Mannu, G.; Darby, S.C.; Taylor, C.W. Adjuvant and neoadjuvant breast cancer treatments: A systematic review of their effects on mortality. *Cancer Treat. Rev.* **2022**, *105*, 102375. [\[CrossRef\]](#)
- Jokar, N.; Velikyan, I.; Ahmadzadehfah, H.; Rekapour, S.J.; Jafari, E.; Ting, H.H.; Biersack, H.J.; Assadi, M. Theranostic Approach in Breast Cancer: A Treasured Tailor for Future Oncology. *Clin. Nucl. Med.* **2021**, *46*, e410–e420. [\[CrossRef\]](#)
- Li, M.; Chi, X.; Wang, Y.; Setrerrahmane, S.; Xie, W.; Xu, H. Trends in insulin resistance: Insights into mechanisms and therapeutic strategy. *Signal Transduct. Target Ther.* **2022**, *7*, 216. [\[CrossRef\]](#) [\[PubMed\]](#)
- Dev, R.; Bruera, E.; Dalal, S. Insulin resistance and body composition in cancer patients. *Ann. Oncol.* **2018**, *29*, ii18–ii26. [\[CrossRef\]](#) [\[PubMed\]](#)
- Singh, P.; Alex, J.M.; Bast, F. Insulin receptor (IR) and insulin-like growth factor receptor 1 (IGF-1R) signaling systems: Novel treatment strategies for cancer. *Med. Oncol.* **2014**, *31*, 805. [\[CrossRef\]](#)
- Chiefari, E.; Mirabelli, M.; La Vignera, S.; Tanyolac, S.; Foti, D.P.; Aversa, A.; Brunetti, A. Insulin Resistance and Cancer: In Search for a Causal Link. *Int. J. Mol. Sci.* **2021**, *22*, 11137. [\[CrossRef\]](#)
- Chou, P.C.; Choi, H.H.; Huang, Y.; Fuentes-Mattei, E.; Velazquez-Torres, G.; Zhang, F.; Phan, L.; Lee, J.; Shi, Y.; Bankson, J.A.; et al. Impact of diabetes on promoting the growth of breast cancer. *Cancer. Commun.* **2021**, *41*, 414–431. [\[CrossRef\]](#)
- Comprehensive molecular portraits of human breast tumours. *Nature* **2012**, *490*, 61–70. [\[CrossRef\]](#)
- Curtis, C.; Shah, S.P.; Chin, S.F.; Turashvili, G.; Rueda, O.M.; Dunning, M.J.; Speed, D.; Lynch, A.G.; Samarajiwa, S.; Yuan, Y.; et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **2012**, *486*, 346–352. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kao, K.J.; Chang, K.M.; Hsu, H.C.; Huang, A.T. Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: Implications for treatment optimization. *BMC Cancer* **2011**, *11*, 143. [\[CrossRef\]](#) [\[PubMed\]](#)
- Brueffer, C.; Vallon-Christersson, J.; Grabau, D.; Ehinger, A.; Häkkinen, J.; Hegardt, C.; Malina, J.; Chen, Y.; Bendahl, P.O.; Manjer, J.; et al. Clinical Value of RNA Sequencing-Based Classifiers for Prediction of the Five Conventional Breast Cancer Biomarkers: A Report from the Population-Based Multicenter Sweden Cancerome Analysis Network-Breast Initiative. *JCO Precis Oncol.* **2018**, *2*, 1–18. [\[CrossRef\]](#)
- Desmedt, C.; Piette, F.; Loi, S.; Wang, Y.; Lallemand, F.O.; Haibe-Kains, B.; Viale, G.; Delorenzi, M.; Zhang, Y.; d’Assignies, M.S.; et al. Strong Time Dependence of the 76-Gene Prognostic Signature for Node-Negative Breast Cancer Patients in the TRANSBIG Multicenter Independent Validation Series. *Clin. Cancer Res.* **2007**, *13*, 3207–3214. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hoogstraat, M.; Lips, E.H.; Mayayo-Peralta, I.; Mulder, L.; Kristel, P.; van der Heijden, I.; Annunziato, S.; van Seijen, M.; Nederlof, P.M.; Sonke, G.S.; et al. Comprehensive characterization of pre- and post-treatment samples of breast cancer reveal potential mechanisms of chemotherapy resistance. *Npj Breast Cancer* **2022**, *8*, 60. [\[CrossRef\]](#)
- Miller, W.R.; Larionov, A. Changes in expression of oestrogen regulated and proliferation genes with neoadjuvant treatment highlight heterogeneity of clinical resistance to the aromatase inhibitor, letrozole. *Breast Cancer Res.* **2010**, *12*, R52. [\[CrossRef\]](#)
- Korde, L.A.; Lusa, L.; McShane, L.; Lebowitz, P.F.; Lukes, L.; Camphausen, K.; Parker, J.S.; Swain, S.M.; Hunter, K.; Zujewski, J.A. Gene expression pathway analysis to predict response to neoadjuvant docetaxel and capecitabine for breast cancer. *Breast Cancer Res. Treat.* **2010**, *119*, 685–699. [\[CrossRef\]](#)
- Page, D.B.; Pucilowska, J.; Chun, B.; Kim, I.; Sanchez, K.; Moxon, N.; Mellinger, S.; Wu, Y.; Koguchi, Y.; Conrad, V.; et al. A phase Ib trial of pembrolizumab plus paclitaxel or flat-dose capecitabine in 1st/2nd line metastatic triple-negative breast cancer. *Npj Breast Cancer* **2023**, *9*, 53. [\[CrossRef\]](#)

19. Wu, S.Z.; Roden, D.L.; Wang, C.; Holliday, H.; Harvey, K.; Cazet, A.S.; Murphy, K.J.; Pereira, B.; Al-Eryani, G.; Bartonicek, N.; et al. Stromal cell diversity associated with immune evasion in human triple-negative breast cancer. *EMBO J.* **2020**, *39*, e104063. [\[CrossRef\]](#)
20. Ginestet, C. ggplot2: Elegant Graphics for Data Analysis. *J. R. Stat. Soc. Ser. A Stat. Soc.* **2011**, *174*, 245–246. [\[CrossRef\]](#)
21. Yoshihara, K.; Shahmoradgoli, M.; Martínez, E.; Vegesna, R.; Kim, H.; Torres-Garcia, W.; Treviño, V.; Shen, H.; Laird, P.W.; Levine, D.A.; et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **2013**, *4*, 2612. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Thorsson, V.; Gibbs, D.L.; Brown, S.D.; Wolf, D.; Bortone, D.S.; Ou Yang, T.H.; Porta-Pardo, E.; Gao, G.F.; Plaisier, C.L.; Eddy, J.A.; et al. The Immune Landscape of Cancer. *Immunity* **2018**, *48*, 812–830.e4. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Aleksander, S.A.; Balhoff, J.; Carbon, S.; Cherry, J.M.; Drabkin, H.J.; Ebert, D.; Feuermann, M.; Gaudet, P.; Harris, N.L.; Hill, D.P.; et al. The Gene Ontology knowledgebase in 2023. *Genetics* **2023**, *224*, iyad031. [\[PubMed\]](#)
24. Kanehisa, M.; Furumichi, M.; Sato, Y.; Kawashima, M.; Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **2023**, *51*, D587–D592. [\[CrossRef\]](#)
25. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **2019**, *28*, 1947–1951. [\[CrossRef\]](#)
26. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [\[CrossRef\]](#)
27. Haghverdi, L.; Lun, A.T.L.; Morgan, M.D.; Marioni, J.C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **2018**, *36*, 421–427. [\[CrossRef\]](#)
28. Hao, Y.; Hao, S.; Andersen-Nissen, E.; Mauck, W.M., 3rd; Zheng, S.; Butler, A.; Lee, M.J.; Wilk, A.J.; Darby, C.; Zager, M.; et al. Integrated analysis of multimodal single-cell data. *Cell* **2021**, *184*, 3573–3587.e3529. [\[CrossRef\]](#)
29. Wu, H.; Wang, W.; Zhang, Y.; Chen, Y.; Shan, C.; Li, J.; Jia, Y.; Li, C.; Du, C.; Cai, Y.; et al. Establishment of patient-derived organoids for guiding personalized therapies in breast cancer patients. *Int. J. Cancer* **2024**, *155*, 324–338. [\[CrossRef\]](#)
30. Balachandran, V.P.; Gonen, M.; Smith, J.J.; DeMatteo, R.P. Nomograms in oncology: More than meets the eye. *Lancet Oncol.* **2015**, *16*, e173–e180. [\[CrossRef\]](#)
31. Yang, Y.; Liu, H.L.; Liu, Y.J. A Novel Five-Gene Signature Related to Clinical Outcome and Immune Microenvironment in Breast Cancer. *Front. Genet.* **2022**, *13*, 912125. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Zhang, Z.; Fang, T.; Lv, Y. A novel lactate metabolism-related signature predicts prognosis and tumor immune microenvironment of breast cancer. *Front. Genet.* **2022**, *13*, 934830. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Wang, Z.; Bao, A.; Liu, S.; Dai, F.; Gong, Y.; Cheng, Y. A Pyroptosis-Related Gene Signature Predicts Prognosis and Immune Microenvironment for Breast Cancer Based on Computational Biology Techniques. *Front. Genet.* **2022**, *13*, 801056. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Wang, S.; Xiong, Y.; Zhang, Q.; Su, D.; Yu, C.; Cao, Y.; Pan, Y.; Lu, Q.; Zuo, Y.; Yang, L. Clinical significance and immunogenomic landscape analyses of the immune cell signature based prognostic model for patients with breast cancer. *Brief Bioinform.* **2021**, *22*, bbaa311. [\[CrossRef\]](#)
35. Kanehisa, M.; Furumichi, M.; Sato, Y.; Ishiguro-Watanabe, M.; Tanabe, M. KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Res.* **2021**, *49*, D545–D551. [\[CrossRef\]](#)
36. Kohsaka, S.; Hinohara, K.; Wang, L.; Nishimura, T.; Urushido, M.; Yachi, K.; Tsuda, M.; Tanino, M.; Kimura, T.; Nishihara, H.; et al. Epiregulin enhances tumorigenicity by activating the ERK/MAPK pathway in glioblastoma. *Neuro-Oncology* **2014**, *16*, 960–970. [\[CrossRef\]](#)
37. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov.* **2022**, *12*, 31–46. [\[CrossRef\]](#)
38. Gallagher, E.J.; Greco, G.; Lin, S.; Yagnik, R.; Feldman, S.M.; Port, E.; Friedman, N.B.; Boolbol, S.K.; Killelea, B.; Pilewskie, M.; et al. Insulin resistance and racial disparities in breast cancer prognosis: A multi-center cohort study. *Endocr. Relat. Cancer* **2022**, *29*, 693–701. [\[CrossRef\]](#)
39. Dieli-Conwright, C.M.; Courneya, K.S.; Demark-Wahnefried, W.; Sami, N.; Lee, K.; Buchanan, T.A.; Spicer, D.V.; Tripathy, D.; Bernstein, L.; Mortimer, J.E. Effects of Aerobic and Resistance Exercise on Metabolic Syndrome, Sarcopenic Obesity, and Circulating Biomarkers in Overweight or Obese Survivors of Breast Cancer: A Randomized Controlled Trial. *J. Clin. Oncol.* **2018**, *36*, 875–883. [\[CrossRef\]](#)
40. Yin, D.T.; He, H.; Yu, K.; Xie, J.; Lei, M.; Ma, R.; Li, H.; Wang, Y.; Liu, Z. The association between thyroid cancer and insulin resistance, metabolic syndrome and its components: A systematic review and meta-analysis. *Int. J. Surg.* **2018**, *57*, 66–75. [\[CrossRef\]](#)
41. Argirion, I.; Weinstein, S.J.; Männistö, S.; Albanes, D.; Mondul, A.M. Serum Insulin, Glucose, Indices of Insulin Resistance, and Risk of Lung Cancer. *Cancer Epidemiol. Biomark. Prev.* **2017**, *26*, 1519–1524. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Manousaki, D.; Kent, J.W., Jr.; Haack, K.; Zhou, S.; Xie, P.; Greenwood, C.M.; Brassard, P.; Newman, D.E.; Cole, S.; Umans, J.G.; et al. Toward Precision Medicine: TBC1D4 Disruption Is Common Among the Inuit and Leads to Underdiagnosis of Type 2 Diabetes. *Diabetes Care* **2016**, *39*, 1889–1895. [\[CrossRef\]](#) [\[PubMed\]](#)

43. Nie, H.; Ju, H.; Fan, J.; Shi, X.; Cheng, Y.; Cang, X.; Zheng, Z.; Duan, X.; Yi, W. O-GlcNAcylation of PGK1 coordinates glycolysis and TCA cycle to promote tumor growth. *Nat. Commun.* **2020**, *11*, 36. [\[CrossRef\]](#)
44. Xiao, G.; Cheng, F.; Yuan, J.; Lu, W.; Wang, P.; Fan, H. Integrative multiomics analysis identifies a metastasis-related gene signature and the potential oncogenic role of EZR in breast cancer. *Oncol. Res.* **2022**, *30*, 35–51. [\[CrossRef\]](#)
45. Chen, D.; Sun, Y.; Wei, Y.; Zhang, P.; Rezaeian, A.H.; Teruya-Feldstein, J.; Gupta, S.; Liang, H.; Lin, H.K.; Hung, M.C.; et al. LIFR is a breast cancer metastasis suppressor upstream of the Hippo-YAP pathway and a prognostic marker. *Nat. Med.* **2012**, *18*, 1511–1517. [\[CrossRef\]](#)
46. Ma, X.; Li, Y.; Zhao, B. Ribosomal protein L5 (RPL5)/ E2F transcription factor 1 (E2F1) signaling suppresses breast cancer progression via regulating endoplasmic reticulum stress and autophagy. *Bioengineered* **2022**, *13*, 8076–8086. [\[CrossRef\]](#)
47. Niu, X.; Yin, L.; Yang, X.; Yang, Y.; Gu, Y.; Sun, Y.; Yang, M.; Wang, Y.; Zhang, Q.; Ji, H. Serum amyloid A 1 induces suppressive neutrophils through the Toll-like receptor 2-mediated signaling pathway to promote progression of breast cancer. *Cancer Sci.* **2022**, *113*, 1140–1153. [\[CrossRef\]](#)
48. Zhang, Y.; Yu, G.; Chu, H.; Wang, X.; Xiong, L.; Cai, G.; Liu, R.; Gao, H.; Tao, B.; Li, W.; et al. Macrophage-Associated PGK1 Phosphorylation Promotes Aerobic Glycolysis and Tumorigenesis. *Mol. Cell.* **2018**, *71*, 201–215.e7. [\[CrossRef\]](#)
49. Zhang, T.; Wang, Y.; Yu, H.; Zhang, T.; Guo, L.; Xu, J.; Wei, X.; Wang, N.; Wu, Y.; Wang, X.; et al. PGK1 represses autophagy-mediated cell death to promote the proliferation of liver cancer cells by phosphorylating PRAS40. *Cell Death Dis.* **2022**, *13*, 68. [\[CrossRef\]](#)
50. Sun, S.; Liang, X.; Zhang, X.; Liu, T.; Shi, Q.; Song, Y.; Jiang, Y.; Wu, H.; Jiang, Y.; Lu, X.; et al. Phosphoglycerate kinase-1 is a predictor of poor survival and a novel prognostic biomarker of chemoresistance to paclitaxel treatment in breast cancer. *Br. J. Cancer* **2015**, *112*, 1332–1339. [\[CrossRef\]](#)
51. Zeng, H.; Qu, J.; Jin, N.; Xu, J.; Lin, C.; Chen, Y.; Yang, X.; He, X.; Tang, S.; Lan, X.; et al. Feedback Activation of Leukemia Inhibitory Factor Receptor Limits Response to Histone Deacetylase Inhibitors in Breast Cancer. *Cancer Cell* **2016**, *30*, 459–473. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Tong, D.; Zhang, J.; Wang, X.; Li, Q.; Liu, L.Y.; Yang, J.; Guo, B.; Ni, L.; Zhao, L.; Huang, C. MeCP2 facilitates breast cancer growth via promoting ubiquitination-mediated P53 degradation by inhibiting RPL5/RPL11 transcription. *Oncogenesis* **2020**, *9*, 56. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Bai, X.; Ni, J.; Beretov, J.; Graham, P.; Li, Y. Cancer stem cell in breast cancer therapeutic resistance. *Cancer Treat. Rev.* **2018**, *69*, 152–163. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Zhang, Y. The root cause of drug resistance in HER2-positive breast cancer and the therapeutic approaches to overcoming the resistance. *Pharmacol. Ther.* **2021**, *218*, 107677. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.