Discover Oncology

Analysis

Development of a prognostic model for breast cancer patients based on intratumoral tumor-infiltrating lymphocytes using machine learning algorithms

Xinyi Wu[1,2] · **Chun Li**[1,2]

Received: 5 December 2024 / Accepted: 6 May 2025
Published online: 14 May 2025

Abstract

Background Breast cancer remains a formidable global health challenge, with tumor-infiltrating lymphocytes (TILs) serving as pivotal biomarkers associated with disease progression, therapeutic response, and survival. While research typically focused on stromal TILs (sTILs), we hypothesize that intratumoral TILs (iTILs), which are in direct contact with tumor cells, have a more profound role in the immune-tumor interactions. In light of this, we have developed an iTIL-centric model for breast cancer patient stratification and prognostic prediction.

Methods We sourced RNA-seq data and clinical profiles of breast cancer patients from The Cancer Genome Atlas (TCGA) and the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) to form our training dataset. Testing datasets, including GSE20685, GSE42568, GSE48390, and GSE88770, were retrieved from Gene Expression Omnibus (GEO). Employing consensus clustering and Weighted Correlation Network Analysis (WGCNA), we identified iTIL-associated hub genes. Our iTIL-centric signature was developed using a machine learning framework integrating 101 algorithms, validated across independent testing sets. Kaplan–Meier analysis and a nomogram model were utilized to evaluate the prognostic accuracy and clinical correlation of our model. GO and KEGG analyses elucidated the biological processes and pathways related to the iTIL signature. The immune profiling provided a comprehensive assessment of the immunological landscape. Moreover, potential drugs for high-risk patients were identified using CTRP v.2.0 and PRISM databases.

Results Our study constructed a pioneering prognostic model based on iTIL-centric signature via a machine learning framework that evaluated 101 algorithm combinations. This model revealed significant differences in the immune landscape among stratified patient cohorts, and demonstrated robust predictive capabilities across multiple datasets. The model showed excellent predictive performance with area under the curve (AUC) values of 0.940, 0.959, and 0.973 for 3-, 5-, and 10-year survival predictions, respectively. Additionally, it was identified as a significant risk factor for overall survival (OS) in the univariate analysis, with a hazard ratio (HR) > 1 and a p-value < 0.001.

Conclusions Our prognostic model, founded on machining learning algorithms and anchored by an iTIL-centric signature, stands out as an invaluable tool for breast cancer patients, offering advanced prognostic insights and facilitating the development of personalized therapeutic strategies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12672-025-02585-1.

✉ Chun Li, chunli186@tongji.edu.cn | [1]Key Laboratory of Spine and Spinal Cord Injury Repair and Regeneration of Ministry of Education, Tongji Hospital, School of Medicine, Tongji University, 389 Xincun Road, Putuo District, Shanghai 200065, China. [2]Tongji University Cancer Center, Shanghai Tenth People's Hospital, School of Medicine, Tongji University, Shanghai 200092, China.

Keywords Breast cancer · Tumor-infiltrating lymphocytes (TILs) · Machine learning · Prognostic model · Immune landscape

1 Introduction

Cancer remains a formidable global public health challenge, with breast cancer representing a leading cause of morbidity and mortality among women. In 2023, an estimated 948,000 women in the United States were projected to be diagnosed with cancer for the first time, of which breast cancer accounted for 31% of cases [1]. Concurrently, breast cancer was responsible for 15% of the 287,740 cancer-related deaths anticipated among women in the same year [1]. These statistics underscore the urgent need for improved prognostic tools and therapeutic strategies to address this pervasive disease.

Tumor-infiltrating lymphocytes (TILs) have emerged as critical determinants of breast cancer progression, treatment response, and survival outcomes [2–5]. TILs are broadly categorized into stromal TILs (sTILs), which reside within the tumor stroma, and intratumoral TILs (iTILs), which directly interact with tumor cells. This direct contact positions iTILs as key mediators of tumor-immune interactions, suggesting a more immediate role in modulating anti-tumor immunity. While the International TILs Working Group has highlighted the reproducibility and reliability of sTILs as a prognostic marker in breast cancer [6], emerging evidence underscores the unique predictive value of iTILs. Specifically, iTILs have been shown to independently predict complete pathological response (pCR) following neoadjuvant therapy in breast cancer [7, 8]. Notably, within the luminal subtypes, iTILs have demonstrated the ability to independently predict pCR, surpassing the prognostic value of sTILs. Furthermore, higher iTIL densities are associated with significantly prolonged disease-free survival (DFS) [9], emphasizing their prognostic relevance. Despite these promising findings, the study of iTILs remains challenging due to their low abundance, high heterogeneity, and limitations in detection through conventional Hematoxylin and Eosin (H&E) staining [6].

Recent advancements have sought to address these challenges. Wu et al. [10] have demonstrated that tumor tissues from the TCGA-BRCA cohort, collected under stringent standards, can be reliably characterized as iTILs. Their work validated the xCell algorithm as a suitable tool for quantifying iTIL scores, based on its strong correlation with lymphocyte markers.

In our study, we employed this methodology to ascertain iTIL scores in our training dataset and performed consensus clustering analysis to stratify samples into two distinct groups. Through the application of Weighted Correlation Network Analysis (WGCNA), we identified hub genes associated with the iTIL-enriched cluster. Building on these findings, we developed a machine learning model to predict outcomes of breast cancer patients based on iTIL scores. This model enabled the stratification of patients into high- and low-risk groups, which were further analyzed to uncover intrinsic molecular and clinical differences. Additionally, we explored potential therapeutic agents that could offer clinical benefits specifically to patients with elevated iTIL values, providing a foundation for personalized treatment strategies.

2 Materials and methods

2.1 Datasets collection and processing

The RNA-seq data of breast cancer and corresponding clinical information were obtained from The Cancer Genome Atlas (TCGA), Molecular Taxonomy of Breast Cancer International Consortium (METABRIC), and Gene Expression Omnibus (GEO) databases. TCGA and METABRIC datasets were merged to form the training set (n = 3046) for developing a predictive breast cancer risk score, while GSE20685 (n = 327), GSE42568 (n = 104), GSE48390 (n = 81), and GSE88770 (n = 117) were used as the testing set to verify the risk score. Somatic mutation data sorted in mutation annotation format (MAF) were downloaded from TCGA.

The Fragments Per Kilobase of exon model per Million mapped fragments (FPKM) data of TCGA-BRCA was downloaded from UCSC Xena database (https://xenabrowser.net/datapages/), and further subjected to log-2 transformation. The METABRIC cohort was downloaded from the cBioPortal (https://www.cbioportal.org/datasets). Datasets from GEO database were sourced from the Affymetrix® GPL570 platform. The ComBat function in the R package sva was utilized to process each breast cancer cohort separately, thereby eliminating potential multicenter batch effects across different experiments.

2.2 Estimation of intratumoral TILs

Referring to the method proposed by Wu et al. [10] for identifying iTILs, and the TCGA criteria requiring a tumor cell nuclei content of at least 60%—80% for sample collection (TCGA criteria: https://www.cancer.gov/ccg/research/genome-sequencing/tcga), we defined TILs derived from TCGA samples as iTILs. The R package immunedeconv was used to estimate immune cell components in TCGA-BRCA cohort with deconvolution algorithms, yielding results from 8 algorithms. Table 1 shows TILs cell types defined by various algorithms. Spearman correlation coefficients were then calculated to assess the relationships between TILs components from different algorithms and genes that serve as lymphocyte markers.

2.3 Consensus clustering and weighted correlation network analysis (WGCNA)

The TILs scores calculated by the CONSENSUS_TME were subjected to a rigorous consensus clustering analysis using the R package ConsensusClusterPlus. The ideal number of clusters ($k = 2$) was identified based on a comprehensive assessment of the consensus score matrix, consensus cumulative distribution function (CDF), and PAC score. To elucidate the correlation between gene networks and iTIL clusters, the R package WGCNA was utilized to identify co-expressed gene modules with high biological significance, specifically for pinpointing genes that were most relevant to the cluster defined by the TILs score. Initially, the optimal soft-thresholding power, conducive to achieving scale-free topology, was determined based on gene expression matrices from TCGA-BRCA and METABRIC datasets. Subsequently, a weighted co-expression network was constructed, yielding 25 modules. The correlation between these modules and iTIL clusters were analyzed to identify the module with the most pronounced correlation. Genes with Gene Significance (GS) > 0.6 and Module Membership (MM) > 0.8 were chosen as hub genes.

2.4 Constructing a prognostic signature with integrated machine learning methods

Initially, hub genes identified by WGCNA were subjected to univariate Cox regression analysis in a training set consisting of TCGA-BRCA and METABRIC datasets. Genes with a significance level of $p < 0.05$ were select for further inclusion in the construction of a prognostic signature. To minimize bias and improve prediction accuracy, we implemented a comprehensive machine learning framework by integrating 10 representative algorithms: Lasso, Ridge regression, stepwise Cox regression (both forward and backward), CoxBoost, random survival forest (RSF), elastic net (Enet), partial least squares regression for Cox (plsRcox), supervised principal components (SuperPC), generalized boosted regression modeling (GBM), and survival support vector machine (survival-SVM). These algorithms were systematically combined in a pairwise or sequential manner (e.g., variable selection using Lasso followed by model building with RSF), resulting in a total of 101 unique model construction strategies. Each combination was applied to the training dataset using a tenfold cross-validation procedure to mitigate overfitting and assess internal performance. The Harrell's concordance index (C-index) was computed for each model, and the model with the highest average C-index was identified as the optimal predictive model. Finally, the optimal model was utilized to construct a prognostic model related to TILs.

Table 1 Definitions of whole lymphocyte cells estimated with 8 algorithms

Algorithm Name	Lymphocyte Definition
ABIS	NK cell, T cell CD8 + memory, T cell CD4 + naïve, T cell CD8 + naïve, B cell naïve, T cell CD4 + memory, T cell MAIT, T cell gamma delta VD2, T cell gamma delta non-VD2, B cell memory, B cell plasma immature
xCell	B-cells, CD4 + memory T-cells, CD4 + naive T-cells, CD4 + T-cells, CD4 + Tcm, CD4 + Tem, CD8 + naive T-cells, CD8 + T-cells, CD8 + Tcm, CD8 + Tem, Class-switched memory B-cells, Memory B-cells, naive B-cells, NK cells, NKT, pro B-cells, Th1 cells, Th2 cells, Tregs
CIBRSORT	B cells naïve, B cells memory, T cells CD8, T cells CD4 naïve, T cells CD4 memory resting, T cells CD4 memory activated, T cells follicular helper, T cells regulatory (Tregs),T cells gamma delta, NK cells resting, NK cells activated
TIMER	B cell, T cell CD4 +, T cell CD8 +
CONSENSUS_TME	B cell, NK cell, T cell CD4 +, T cell CD8 +, T cell gamma delta, T cell regulatory (Tregs)
EPIC	B cell, T cell CD4 +, T cell CD8 +, NK cell
QUANTISEQ	B cell, NK cell, T cell CD4 + (non-regulatory), T cell CD8 +, T cell regulatory (Tregs)
DANAHER-TIL	B-cells, CD8 T cells, NK CD56 dim cells, NK cells, T-cells, Th1 cells, Treg

2.5 Kaplan–Meier analysis and construction of Nomogram model

R package survminer was utilized to ascertain the optimal cut-off value for stratifying risk score and facilitate the delineation of Kaplan–Meier survival curves. Performing univariate Cox regression analysis on various clinical factors and risk scores, we identified factors with statistical significance ($p < 0.05$) that were subsequently incorporated into the multivariate Cox regression analysis. Employing the R packages rms and ggDCA, we delineated the Nomogram plot, calibration curve and decision curve. Additionally, utilizing the R package timeROC, we generated ROC curves for the Nomogram model.

2.6 Functional enrichment analyses

Differential gene expression analysis on TCGA-BRCA read count data was performed using the R package limma. Then we proceeded with Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses for the identified differentially expressed genes using the R package clusterProfiler.

2.7 Immune-related analysis

Utilizing the R package genefu, we performed PAM50 subtype classification on TCGA-BRCA dataset. Further, leveraging the R package immunedeconv [11], we calculated the proportions of different immune cell components in CONSENSUS_TME score. Additionally, we assessed the scores of various immune features in TCGA-BRCA dataset using the R package EaSIeR [12]. The assessment of immune activity levels was conducted using Cytolytic activity score (CYT), which is defined as the arithmetic mean of the expression of Granzyme A (GZMA) and Perforin (PRF1) genes. The collection of tertiary lymphoid structure (TLS) related genes was informed by gene sets compiled by Sautès-Fridman et al. [13] and Cabrita et al. [14].

Employing the R package maftools, we generated waterfall plots to depict the top 20 genes exhibiting the highest mutation rates in the high and low-risk groups. Furthermore, we obtained the score of intratumor heterogeneity, homologous recombination defects, silent mutation rate, nonsilent mutation rate, single nucleotide variation neoantigens, and indel neoantigens from the previous research by Thorsson et al. [15] for the TCGA cohort.

2.8 Predicting potential drugs for high-risk scores breast cancer patients

To identify candidate drugs with therapeutic efficacy for breast cancer patients exhibiting high immune-related tumor-infiltrating lymphocyte (ITILS) scores, we integrated gene expression and drug sensitivity data from multiple public pharmacogenomic resources. Firstly, we employed the Genomics of Drug Sensitivity in Cancer (GDSC) database to estimate the half-maximal inhibitory concentration (IC50) values of drugs in breast cancer patients. Spearman correlation analysis was performed between predicted IC50 values and ITILS scores across samples, and the top 50 compounds with significantly positive correlation coefficients ($p < 0.05$) were identified. Additionally, we specifically examined the correlation between IC50 values of standard-of-care endocrine and targeted therapies and ITILS scores to explore differences in predicted sensitivity between high- and low-risk subgroups. To further screen potential agents effective in the high-risk subgroup, we obtained drug response data from the Cancer Therapeutics Response Portal (CTRP v2.0) and the PRISM Repurposing dataset (19Q4 release). After removing duplicate compounds and drugs with excessive missing values, 356 compounds from CTRP and 1291 from PRISM were retained. Predicted AUC values for breast cancer patients were computed using ridge regression models trained on Cancer Cell Line Encyclopedia (CCLE) gene expression and drug sensitivity data. To identify high-risk specific candidate compounds, we stratified patients based on their ITILS scores and conducted differential drug response analysis between the top decile (high ITILS) and bottom decile (low ITILS) groups. Compounds with significantly lower predicted AUC values in the high-risk group ($p < 0.05$) were considered as potentially more sensitive. To further ensure consistency and robustness, Spearman correlation analysis was performed to assess the relationship between AUC values and ITILS scores across all samples. Compounds with negative correlation coefficients ($R < -0.2$) were retained as high-confidence candidates.

This integrative analysis identified 6 candidate compounds from the CTRP dataset and 5 from the PRISM dataset. To comprehensively evaluate the therapeutic potential of these 11 compounds, we performed the following additional analyses:

(1) Literature mining using PubMed to retrieve prior experimental and clinical studies supporting their efficacy in breast cancer;

(2) Connectivity Map (CMap) analysis, which assesses transcriptional signatures of compound-induced perturbations in cancer cells. Drugs with CMap scores < -85 were considered to have potential therapeutic relevance for high ITILS patients.

2.9 Statistical analysis

Data processing, statistical computation, and graphical illustration were conducted using R software (version 4.3.0). Supplementary Table 1 delineates the versions of the R packages that were integral to the analytical procedures undertaken in this study. The Spearman correlation coefficient was utilized to assess the correlation between two continuous variables. Comparisons of continuous data were performed using t-tests or Wilcoxon rank-sum tests. Unless otherwise specified, statistical significance was set at $p < 0.05$.

3 Results

3.1 Establishment and validation of immune infiltration consensus clusters in breast cancer

The flowchart outlining the overview of this study is depicted in Fig. 1. Drawing on the method proposed by Wu et al. [10] for calculating intratumoral tumor infiltrating lymphocytes (iTILs), we aimed to enhance the accuracy in evaluating the quantity of iTILs. To this end, we employed eight deconvolution algorithms (ABIS, xCell, CIBRSORT, TIMER, CONSENSUS_TME, EPIC, QUANTISEQ, and DANAHER-TIL [16]) from the R package immunedeconv to estimate the proportion of immune cells within the TCGA-BRCA cohort (Table 1). By defining the iTILs proportion as defined by each algorithm, we conducted a thorough assessment of the correlation between lymphocyte marker genes (PDCD1, CD8 A, CD4, CD3D) and estimated TILs scores using these different algorithms (Fig. 2A). The TILs score, as calculated by the CONSENSUS_TME algorithm, displayed the strongest correlation with TILs parameters and was therefore selected as the iTILs score for our research. Subsequently, based on this score, consensus cluster analysis was conducted on a training set (TCGA-BRCA and METABRIC, $n = 3046$). The cumulative distribution function (CDF) curves of consensus score and the proportion of ambiguous clustering (PAC) statistics collectively suggested that an optimal solution was attained with $k = 2$ (Fig. 2B, C). Upon scrutinizing the disparities in iTILs between cluster1 and cluster2, we observed a significantly higher level of iTIL infiltration in cluster2 compared to cluster1 (Fig. 2D). Furthermore, in breast cancer stratified by various PAM50 subtypes, cluster2 exhibited significantly higher immune scores and CYT scores than cluster1 (Fig. 2E, F).

3.2 Construction of a prognosis signature based on integrative machine learning

To identify genes with the highest correlation to iTILs, we performed Weighted Gene Co-expression Network Analysis (WGCNA). A rigorous selection process for the soft threshold identified the optimal value of 5 (scale-free $R^2 = 0.90$), ensuring the establishment of a robust scale-free topological network. By optimizing the minModuleSize parameter to a threshold of 30 in our WGCNA analysis, we delineated 25 distinct modules. We then proceeded to calculate the correlations of these modules with key variables, including Overall Survival (OS), age, and cluster, to explore the potential prognostic associations. Notably, the red module exhibited the strongest correlation with the immune cluster (Fig. 3A). Within the red module, the correlation coefficient between gene significance (GS) and module membership (MM) achieved a high value of 0.86, suggesting the superior construction quality of the module. 285 genes with $MM > 0.8$ and $GS > 0.6$ were defined as hub genes related to the iTILs score (Fig. 3B).

Subsequently, we conducted univariate Cox regression analysis on these 285 genes in our training set, identifying 29 prognostic-associated genes (Fig. 3C). To establish an optimal prognostic model correlated with iTILs, these genes were integrated into 101 prognostic models within the machine-learning framework. Using the Concordance Index (C-index) as the metric, the performance of each model was evaluated across multiple test sets (GSE20685, GSE42568, GSE48390, GSE88770). The model composed of CoxBoost + RSF exhibited the highest average C-index

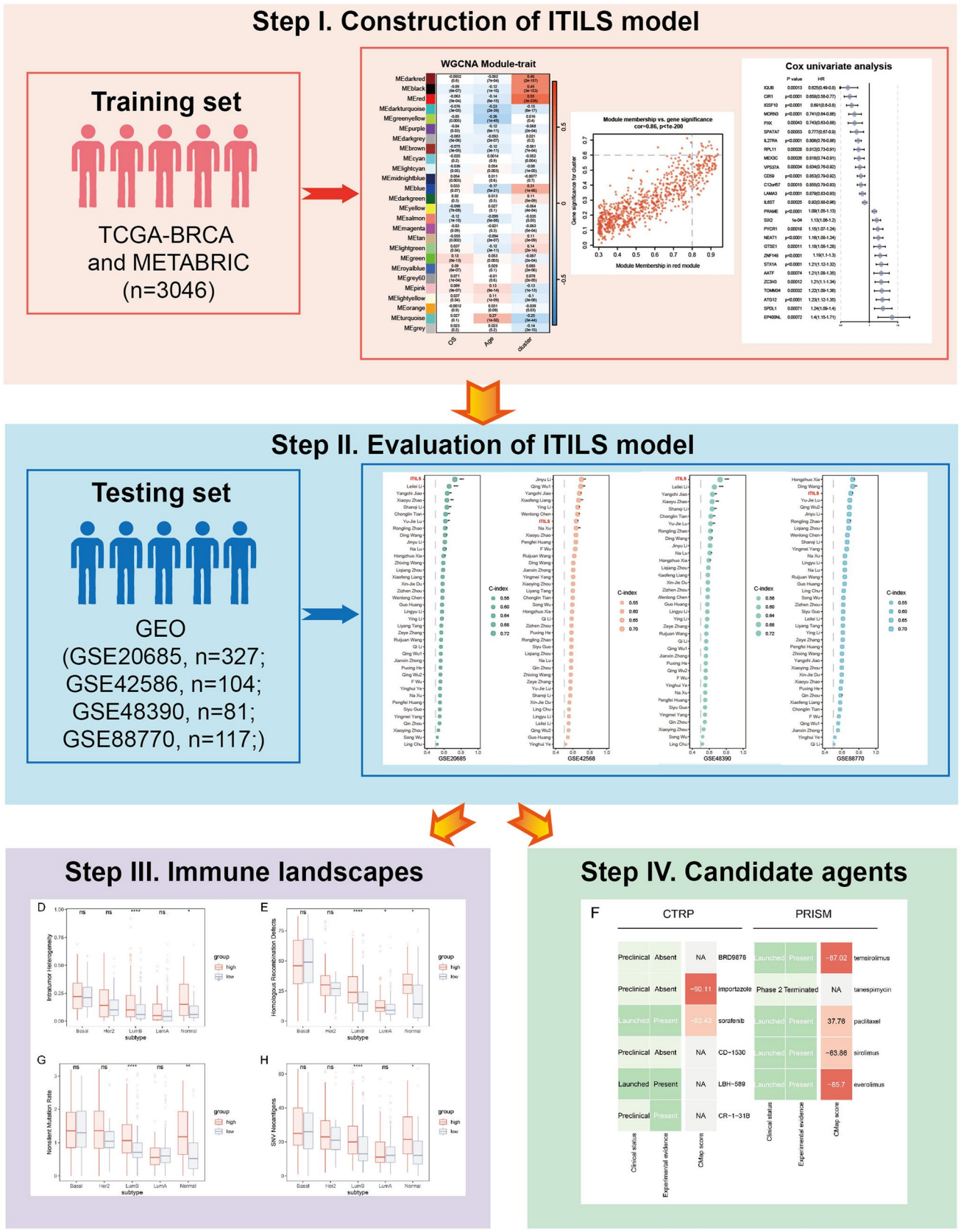


Fig. 1 The flowchart of our study

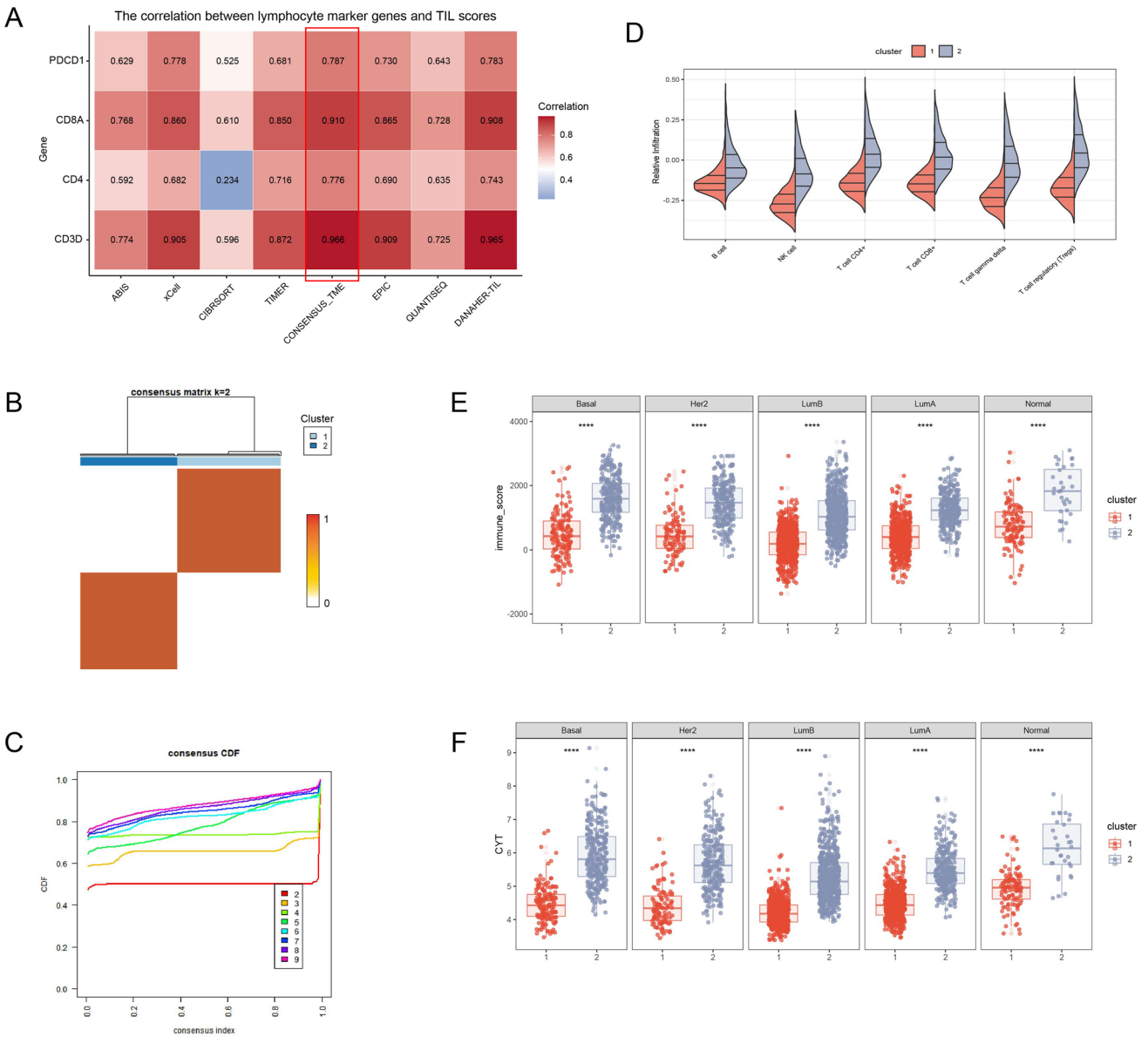


Fig. 2 Establishment of the iTILs score using the CONSENSUS_TME algorithm. **A** Correlation between lymphocyte marker genes and TIL scores calculated using eight different algorithms. **B** Consensus score matrix for all samples (k = 2). The heatmap shows a clear division between the two clusters, with the majority of the high consensus values (indicated by dark red) confined within the boundaries of each cluster. This pattern suggests strong intra-cluster similarity, supporting the validity of the two-cluster solution. **C** Cumulative distribution function (CDF) curves of the consensus matrix for each k values (color-coded by k). **D** The levels of tumor lymphocyte immune infiltration in two clusters identified by the CONSENSUS_TME algorithm. **E** Immune scores in cluster 1 and cluster 2 across different PAM50 subtypes (****p < 0.0001). **F** CYT scores in cluster 1 and cluster 2 across different PAM50 subtypes (****p < 0.0001)

(0.667), emerging as the best-performing model, which we designate as the ITILS model (Fig. 3D). The model commenced with a preliminary screening for genes with non-zero coefficients in the CoxBoost model under a tenfold cross-validation framework (Fig. 3E). Subsequently, the model was refined using random survival forest (RSF) to identify genes with a variable importance > 0 (Fig. 3F, G). Ultimately, the analysis culminated in the derivation of an iTIL-centric immune landscape signature composed of 19 pivotal genes. Supplementary Fig. 1 A illustrated the partial dependency plot analysis, depicting the relationship between the 19 pivotal genes and the predictive model. In addition, leveraging the GEPIA2.0 database, we elucidated the differential mRNA expression profiles of these genes, contrasting the levels observed in breast cancer tissues against those in normal tissues (Supplementary Fig. 1B).

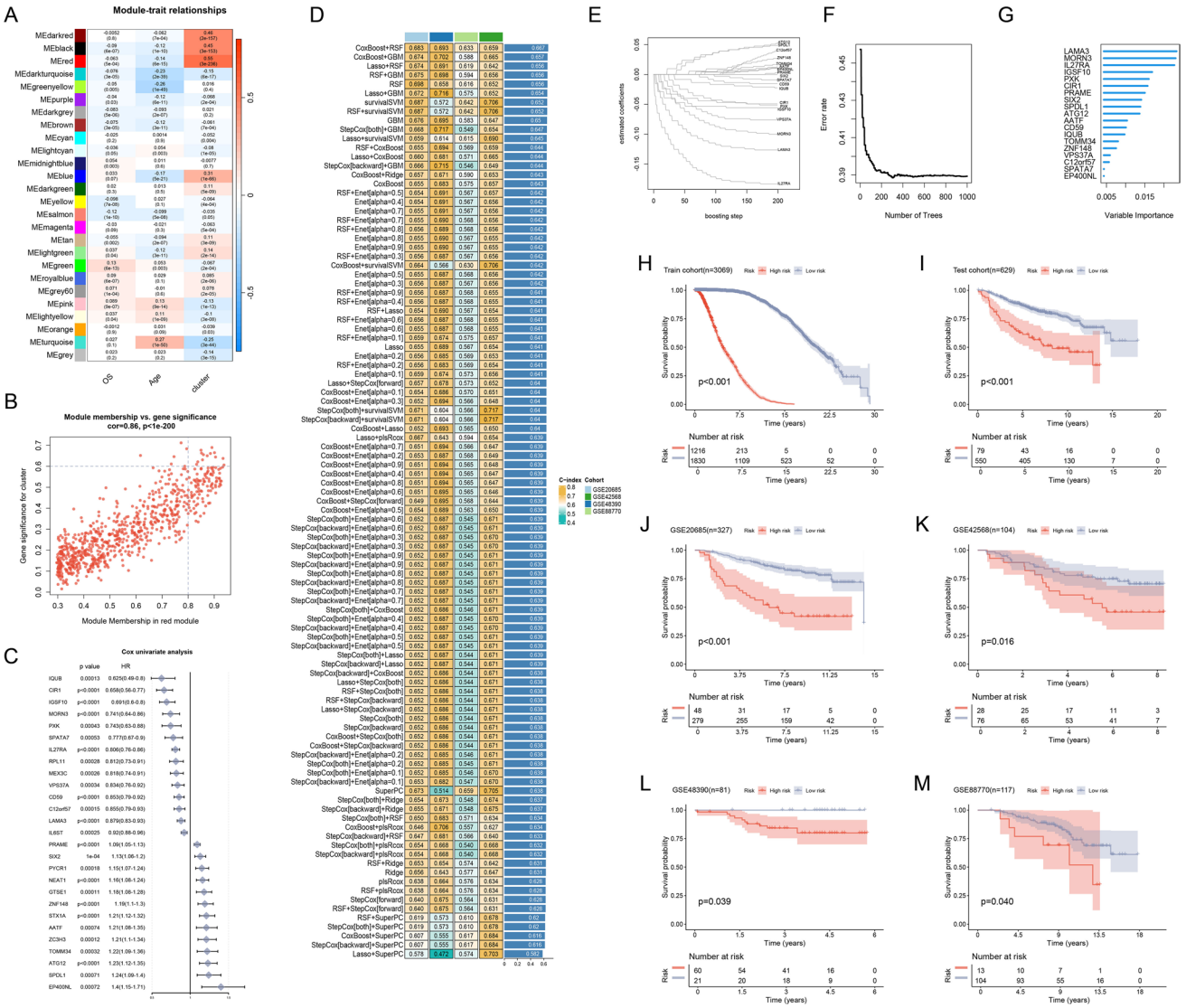


Fig. 3 Development of an iTILs-based prognostic signature using integrated machine learning program. **A** Correlations between co-expression modules (rows) and phenotypic/clinical traits (columns). Each row represents a gene module, identified by its unique color through WGCNA. Cells display Pearson correlation coefficients (top value) and p-values (parentheses) between module eigengenes and traits. Positive correlations are shown in red, indicating direct associations, while negative correlations are shown in blue, representing inverse relationships. **B** Correlation between gene significance (GS) and module membership (MM) within the MEred module. **C** Cox univariate analysis of hub genes in the training set. **D** Performance of 101 prediction models, evaluated by the concordance index (C-index) across all testing sets, along with the average C-index indicated. **E** CoxBoost model results. **F** Relationship between the number of trees and the error rate in a Random Survival Forest (RSF) model. **G** The variable importance in the RSF model. (**H**-**M**) Kaplan–Meier curves of overall survival (OS) stratified by ITILS in the training set (n = 3,069), testing set (n = 629), and external validation cohorts: GSE20685 (n = 327), GSE42568 (n = 104), GSE48390 (n = 81), and GSE88770 (n = 117)

To assess the predictive accuracy of the model for the prognosis of breast cancer patients, we stratified the patients into high- and low-risk groups based on the ITILS score which was calculated by the model and calibrated with the optimal cutoff value determined by the package survminer. Kaplan–Meier (KM) curves were generated for these stratified groups, revealing that the high-risk group exhibited a significantly diminished OS in both the training and testing cohorts ($p < 0.05$) (Fig. 3H–M).

3.3 Evaluation of the ITILS model and comparison with other prognostic models

Initially, the receiver operating characteristic (ROC) curve analysis was performed to assess the predictive efficacy of the ITILS model in both the training and testing sets. The results indicated that in the training cohort, the model achieved area under the curve (AUC) values of 0.940 (95% confidence interval [CI]: 0.929 to 0.950), 0.959 (95% CI: 0.951 to 0.966), and 0.973 (95% CI: 0.967 to 0.979) for the 3-, 5-, and 10-year survival predictions, respectively. In the overall testing cohort, the model attained AUC values of 0.773 (95% CI 0.619 to 0.926), 0.691 (95% CI 0.624 to 0.757), and 0.684 (95% CI 0.628 to 0.740) for the 1-, 3-, and 5-year survival predictions, respectively (Fig. 4A, B). These outcomes underscored the high predictive accuracy of the ITILS model. In an effort to compare the performance of ITILS model against other prognostic models based on distinct signatures, we undertook a comprehensive retrieval of published prognostic models for breast cancer [17–25]. These models were linked to a variety of biological processes, including immune response, super-enhancer activity, lactylation, autophagy, and ferroptosis. Subsequently, we engaged in a univariate Cox regression analysis across a diverse array of datasets, including TCGA-BRCA, METABRIC, GSE20685, GSE42568, GSE48390, and GSE88770 cohorts. The results indicated that among all models, ITILS exhibited the highest prognostic relevance across all evaluated datasets (Fig. 4C). Furthermore, by calculating the C-index of each model in these cohorts, we found that ITILS demonstrated the highest aggregate C-index among all models (Fig. 4D-I), thereby showcasing its preeminent prognostic accuracy.

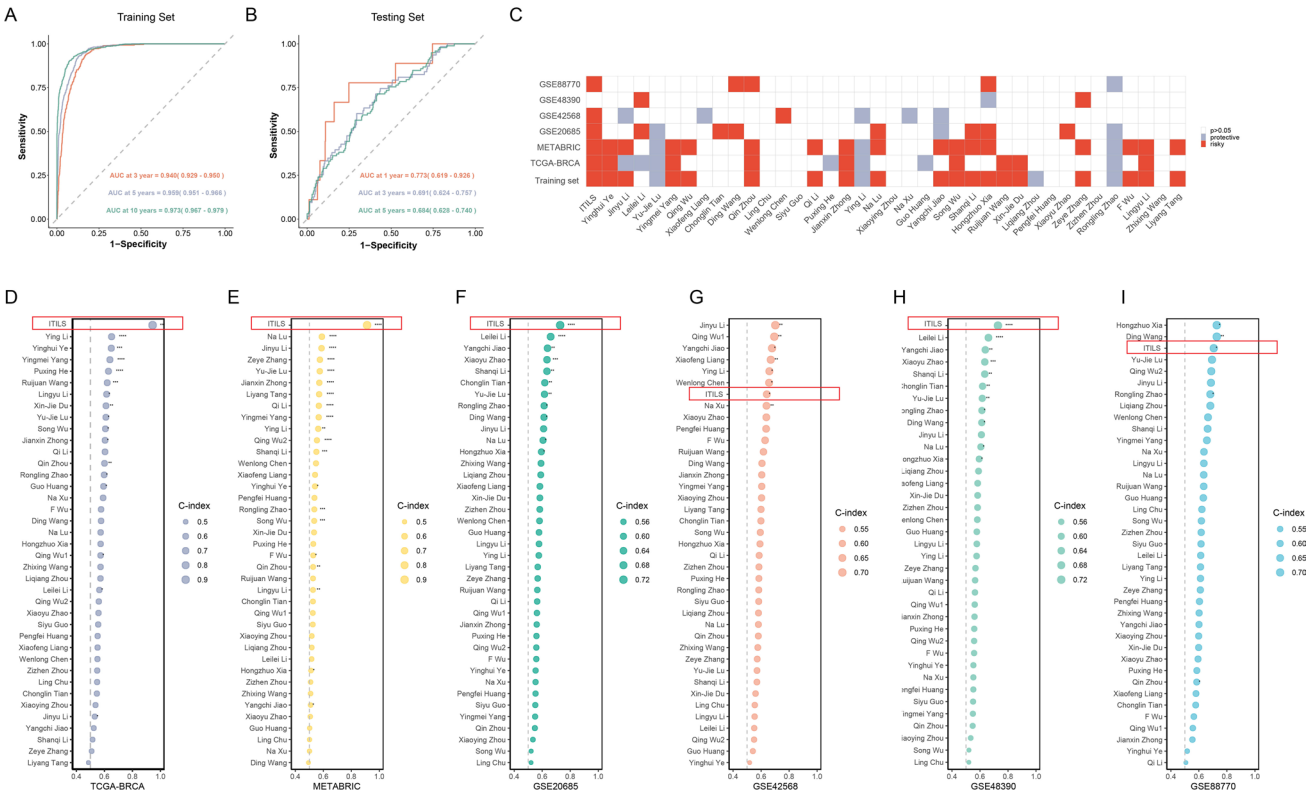


Fig. 4 Evaluation of the ITILS and comparation with 38 published models. **A** Receiver operating characteristic (ROC) curve analysis of ITILS for predicting 3-, 5-, and 10-year OS in the training set. **B** ROC curve analysis of ITILS for predicting 1-, 3-, and 5-year OS in the testing set. **C** Cox univariate analysis of ITILS and 38 published signatures in the training set, TCGA-BRCA, METABRIC, GSE20685, GSE42568, GSE48390 and GSE88770. **D**–**I** Comparison of the C-index values for ITILS and 38 published signatures across six independent datasets: TCGA-BRCA, METABRIC, GSE20685, GSE42568, GSE48390 and GSE88770. The red box indicates the performance of our model, ITILS. Due to the variability in model rankings across datasets, each panel retains its own y-axis to ensure accurate representation of the performance rankings. Significance levels: *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001

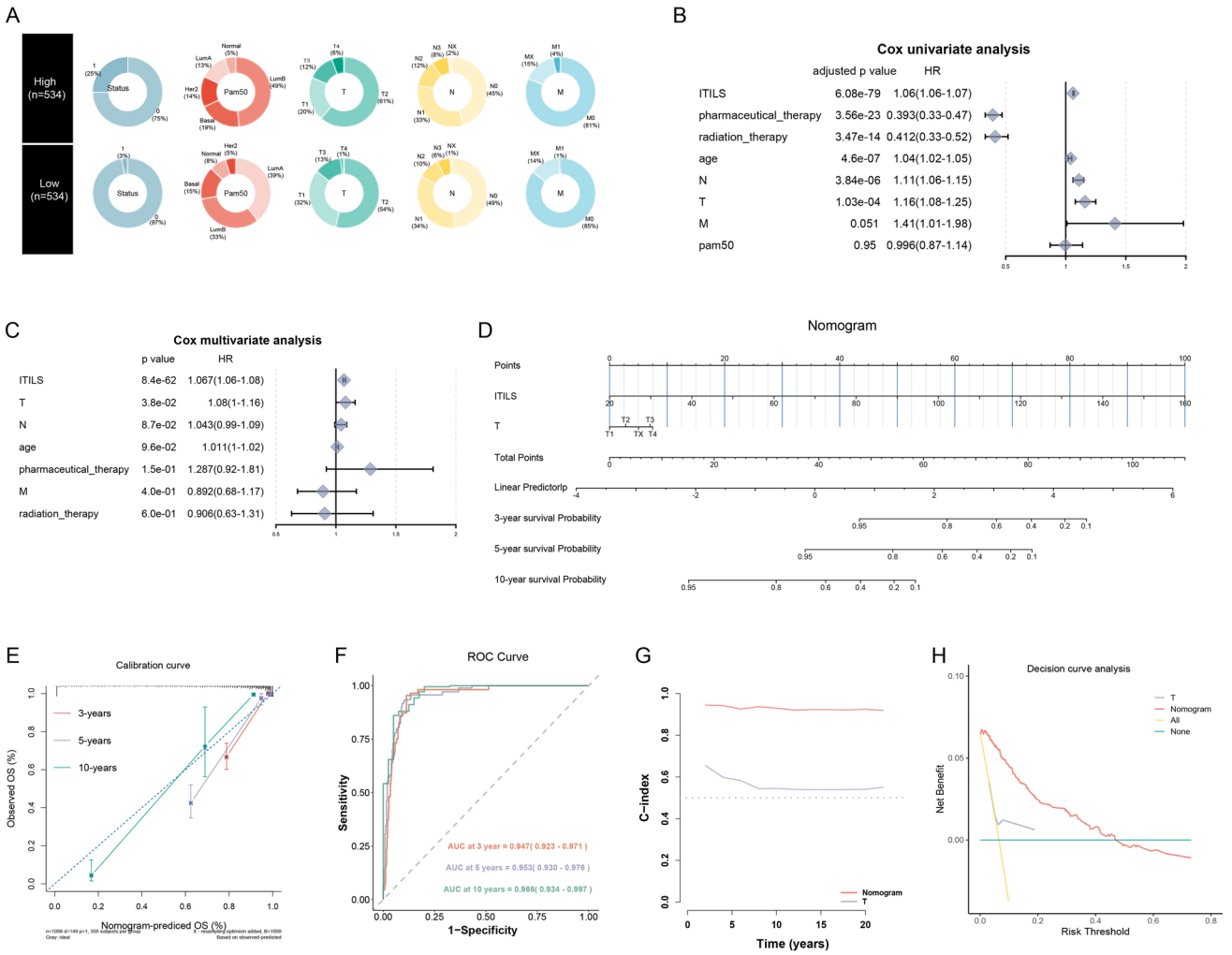


Fig. 5 Relationship between ITILS and clinical characteristics. **A** Association of ITILS low- and high-risk groups with survival status, Pam50 subtype, and TNM staging (T stage, N stage and M stage). **B** Cox univariate analysis of ITILS and clinical characteristics. **C** Cox multivariate analysis of ITILS and clinical characteristics. **D** Nomogram integrating ITILS and T stage for survival prediction. **E** Calibration curve evaluating the nomogram's predictive accuracy. **F** ROC curve analysis of the nomogram for predicting 1-, 3-, and 5-year OS in the TCGA-BRCA cohort. **G** Time-dependent C-index comparison between the nomogram and T stage. **H** Decision curve analysis (DCA) assessing the clinical utility of the nomogram

3.4 Construction and evaluation of the ITILS-related Nomogram model

To further assess the predictive capability of ITILS for patient prognosis, we analyzed its relationship with various clinical factors in the TCGA-BRCA cohort. Our results disclosed significant differences in survival status, Pam50 subtypes, and T, N, M stages between high- and low-risk patients (Fig. 5A). Notably, ITILS scores were significantly reduced in patients at earlier TNM stage, specifically T1-T2, N0-N1, and M0 stages, compared to those at advanced stages T3-T4, N2-N3, and M1 stages (Supplementary Fig. 2 A-C), suggesting a correlation between higher ITILS scores and a poorer prognosis in breast cancer patients. Additionally, KM analysis confirmed the robust prognostic predictive ability of ITILS model across various TNM stage subgroups (Supplementary Fig. 2D-I).

Following our initial assessment, we then performed univariate Cox regression analysis to ascertain the prognostic impact of ITILS alongside key clinical variables such as pharmaceutical therapy, radiotherapy, age, T, N, M stages, and pam50 subtype. Our findings identified ITILS as a significant prognostic indicator, with a high hazard ratio (HR > 1) and a p-value significantly below the threshold of 0.0001 (Fig. 5B), indicating a strong correlation with an adverse outcome. Variables significantly associated with prognosis were then incorporated in a multivariate Cox regression

analysis. ITILS retained as a pronounced prognostic risk factor (HR > 1, p < 0.0001), confirming its substantial prognostic ability in breast cancer patients (Fig. 5C). In light of these findings, we developed a nomogram integrating ITILS scores with clinical factors to predict the 3-, 5-, and 10-year survival probability of breast cancer patients. (Fig. 5D). The calibration curve of the nomogram demonstrated a good consistency between predicted and observed outcomes (Fig. 5E). The AUC of the nomogram reached 0.947 (95% CI 0.923 to 0.971), 0.953 (95% CI 0.930 to 0.976), and 0.966 (95% CI 0.934 to 0.997) at the 3-, 5-, and 10-year intervals, respectively, signifying its high predictive accuracy (Fig. 5F). Furthermore, the time-dependent C-index illustrated that ITILS outperforms T stage in predicting OS from 1 to 20 years (Fig. 5G). Decision curve analysis (DCA) further verified that the nomogram provided a superior net clinical benefit over the T stage (Fig. 5H). Collectively, these findings suggested that the ITILS-based nomogram was capable of delivering a more accurate and personalized prognostic assessment for breast cancer patients.

Metastasis is a crucial determinant of poor prognosis in breast cancer, with substantial evidence supporting the role of radiation and pharmacotherapy treatments in improving survival outcomes for affected patients [26, 27]. To further elucidate the relationship between ITILS and prognosis in breast cancer patients, we analyzed the associations between ITILS, metastasis, and the impact of radiation and pharmacotherapy. Patients with metastasis exhibited significantly diminished OS compared to those without metastasis (Fig. 6A). Importantly, our findings indicated that ITILS scores were markedly elevated in breast cancer patients with metastasis compared to the non-metastatic group (Fig. 6B). Additionally, patients underwent radiation or pharmacotherapy showed significantly reduced ITILS scores and concomitantly improved OS, compared to those who did not benefit from such treatments (Fig. 6C–F).

3.5 The immune landscape and ITILS correlation in breast cancer risk stratification

To further investigate the significance of ITILS in breast cancer patients, we stratified the TCGA-BRCA cohort into high- and low-risk groups based on ITILS. Subsequent analysis revealed a significant difference in OS between these two groups (Fig. 7A). We extended our investigation to delve into the disparities between these stratified high- and low-risk patient cohorts. Following this, differentially expressed genes (DEGs) between these two patient cohorts were subjected to Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses (Fig. 7B, C) to underlying the biological functions. GO analysis highlighted substantial enrichment of DEGs in immune-related functional pathways, specifically in immunoglobulin production (BP: Biological Processes), immunoglobulin complexes (CC: Cellular Components), and antigen binding (Molecular Functions: MF). Additionally, KEGG pathways analysis underscored a predominant enrichment in pathways governing neuroactive ligand-receptor interaction and cytokine-cytokine receptor interaction. Building on these insights, we advanced our investigation with a comprehensive assessment of the immune profiles between high- and low-risk patients. We applied the consensus_TME algorithm to evaluate the distribution of various immune cell within high- and low-risk patient cohorts across diverse breast cancer subtypes (Fig. 7D). Our findings revealed that in Her2, LumB and LumA subtypes, patients in the low-risk cohort displayed a markedly greater abundance of immune infiltration across all immune cell types compared to the high-risk cohort.

Previous studies have implicated that tumors with lymphatic endothelial cell infiltration [28] or adipocyte infiltration [29] may be associated with an enhanced immune responses. Guided by these findings, we analyzed the abundance of adipocytes, endothelial cells, and lymphatic endothelial cells in the high- and low-risk patient cohorts across various breast cancer subtypes (Fig. 7E–G). Our results demonstrated that in Her2, LumB, and LumA subtypes, the low-risk cohort exhibited higher infiltration levels of endothelial cells, lymphatic endothelial cells, and adipocytes compared to the high-risk cohort. These observations suggested that tumors in the low-risk patient cohort may indeed possess a stronger immune response.

Next, we employed the R package EaSIeR to provide a comprehensive overview of the immune response in the TCGA-BRCA cohort, with a specific focus on the correlation between ITILS scores and various immune features (Fig. 7H). Notably, the most robust absolute correlation coefficients were identified as tertiary lymphoid structure (TLS) features and immune resistance programs (resF_down, resF), with correlation coefficients of −0.26, −0.21, and 0.21 respectively (p < 0.05). Consequently, we proceeded our investigation to explore the correlation between ITILS scores and TLS-related genes, generating scatter plots that depicted the relationship between the top 8 genes with the highest absolute correlation coefficients and ITILS scores (Fig. 7I). The results represented a predominantly negative correlation, suggesting a potential inverse relationship between ITILS scores and TLS-related genes.

Fig. 6 Association of ITILS with treatment modalities and tumor metastasis. **A** Kaplan–Meier survival curves comparing metastatic (1) and non-metastatic (0) breast cancer patients. **B** Comparison of ITILS scores between metastatic and non-metastatic breast cancer patients. **C** Kaplan–Meier survival curves for patients receiving (1) and not receiving (0) radiotherapy. **D** Comparison of ITILS scores between patients receiving and not receiving radiotherapy. **E** Kaplan–Meier survival curves for patients receiving (1) and not receiving (0) pharmaceutical therapy. **F** Comparison of ITILS scores between patients receiving and not receiving pharmaceutical therapy

3.6 Immunogenomic analysis in different ITILS subgroups

Given the pivotal role of genetic variation in cancer initiation and progression [15], we undertook the investigation to delineate the genomic mutational profiles between high- and low-risk ITILS subgroups (Fig. 8A, B). We discerned a marked disparity in genomic mutational landscape between these two subgroups. Specifically, within the high-risk group, there was a significantly elevated incidence ($p < 0.05$) of mutations in the tumor suppressor gene TP53 compared to the low-risk group (Fig. 8C). We have also identified significant differences in immune infiltration between ITILS subgroups (Fig. 7G). Considering the established correlation between immune infiltration and DNA damage response [15], we conducted an immunogenomic analysis from this perspective across both subgroups. Our results showed that, in the Luminal B and Normal subtypes, the high-risk subgroup exhibited substantially elevated levels of Intratumor Heterogeneity (ITH), Homologous Recombination Defects (HRD), Silent Mutation Rate, Nonsilent Mutation Rate, and SNV Neoantigens, in contrast to the low-risk subgroup (Fig. 8D-I).

3.7 Identification of potential compounds for breast cancer patients with high ITILS scores

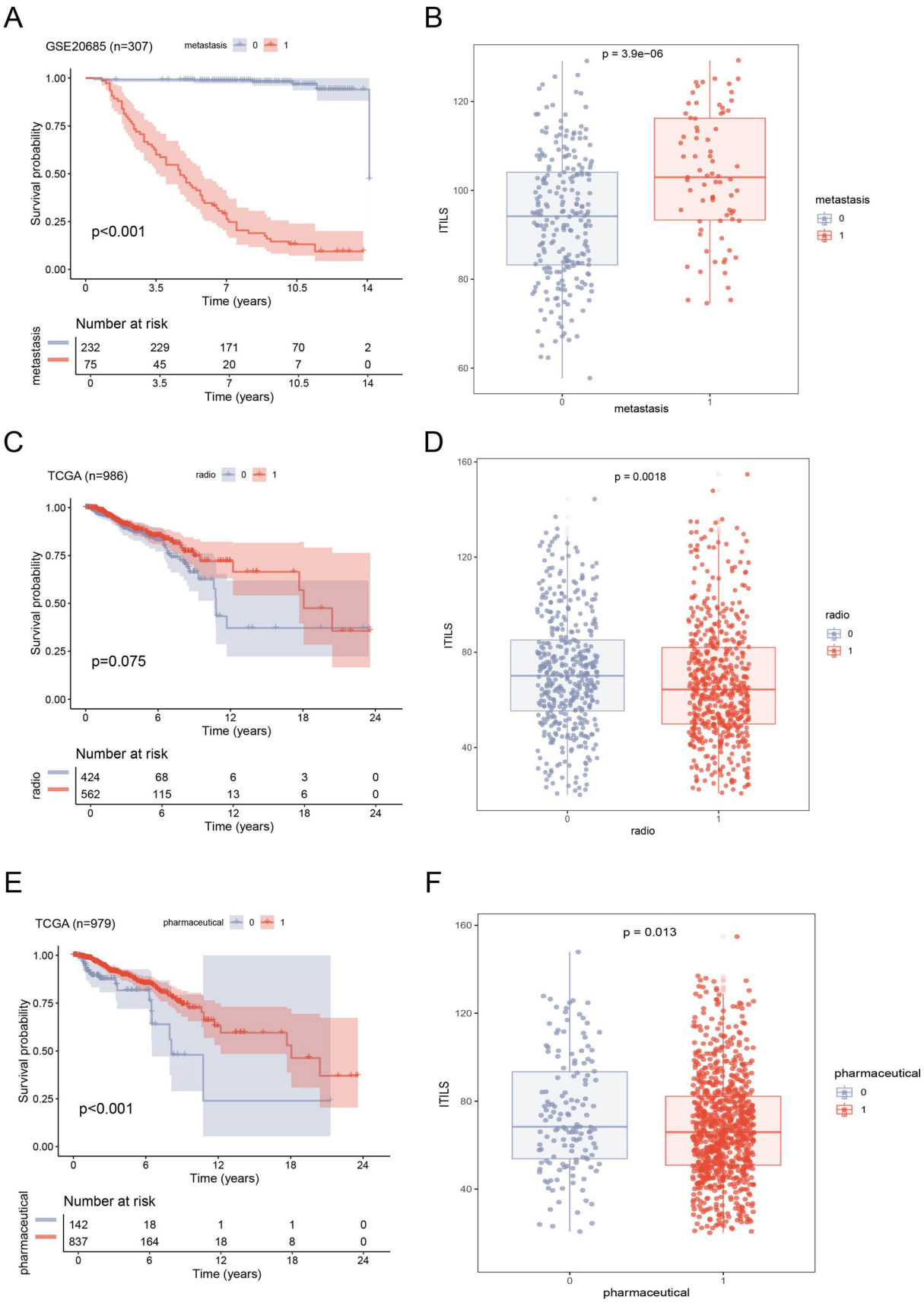
We utilized data from the Genomics of Drug Sensitivity in Cancer (GDSC) to predict the drug half-maximal inhibitory concentration (IC50) for breast cancer patients. We identified the top 50 drugs with the strongest positive correlation coefficients to ITILS scores ($p < 0.05$) (Fig. 9A). Further analysis of the relationship between IC50 values of standard endocrine and targeted therapies for breast cancer and ITILS scores disclosed a markedly disparity, with drugs showing significantly higher IC50 values in the high-risk subgroup compared to the low-risk subgroup (Fig. 9B, C). This finding underscored the necessity to identify and develop therapeutic drugs that are effective for high-risk patients.

We then focused on identifying such efficacious drugs by conducting a thorough analysis of drug response data from the Cancer Therapeutics Response Portal (CTRP) and the PRISM Repurposing dataset. After removing duplicate and null data entries, we retained 356 compounds from the CTRP dataset and 1291 compounds from the PRISM dataset. We executed a differential drug response analysis between the top-decile group (with high ITILS scores) and the bottom-decile group (with low ITILS scores) to pinpoint compounds with diminished predicted AUC values in the high ITILS group. Following this, a refined filtration process incorporated Spearman correlation analysis to assess the relationship between AUC values and ITILS scores, selecting compounds with negative correlation coefficients ($R < -0.2$). This systematic approach identified 6 sensitive drugs for high-risk group in the CTRP dataset (Fig. 9D) and 5 in the PRISM dataset (Fig. 9E).

Subsequently, we sought to evaluate the therapeutic potential of these 11 drugs for breast cancer patients with high ITILS scores from multiple perspectives. Firstly, we performed an exhaustive search via PubMed to gather experimental and clinical evidence regarding their therapeutic efficacy in breast cancer treatment. Next, we employed the Connectivity Map (CMap) database for an advanced analysis of these drugs, with the aim of discerning those exhibiting enhanced sensitivity in the high ITILS group. Our findings indicated that Importazole, Temsirolimus, and Everolimus had CMap scores < -85, indicative of a potentially significant therapeutic effects on breast cancer. Upon synthesizing these comprehensive analyses, we concluded that Temsirolimus and Everolimus are the most promising candidates for treating high-risk breast cancer patients (Fig. 9F).

4 Discussion

Extensive research has demonstrated that TILs are remarkably associated with key clinical outcomes, encompassing survival times, chemotherapy responses, and immunotherapy efficacy in breast cancer patients. While most studies have focused on sTILs, theoretically, iTILs are in more direct contact with tumor cells and thus suggesting a potentially stronger correlation with the intrinsic characteristics of tumor cells [6]. Consequently, quantitative assessment of iTILs could offer


A
GSE20685 (n=307)
metastasis 0 1
Survival probability
p<0.001
Time (years)
Number at risk
metastasis
232 229 171 70 2
75 45 20 7 0
B
p = 3.9e−06
ITILS
metastasis
C
TCGA (n=986)
radio 0 1
p=0.075
Number at risk
radio
424 68 6 3 0
562 115 13 6 0
D
p = 0.0018
radio
E
TCGA (n=979)
pharmaceutical 0 1
p<0.001
Number at risk
pharmaceutical
142 18 1 1 0
837 164 18 8 0
F
p = 0.013
pharmaceutical

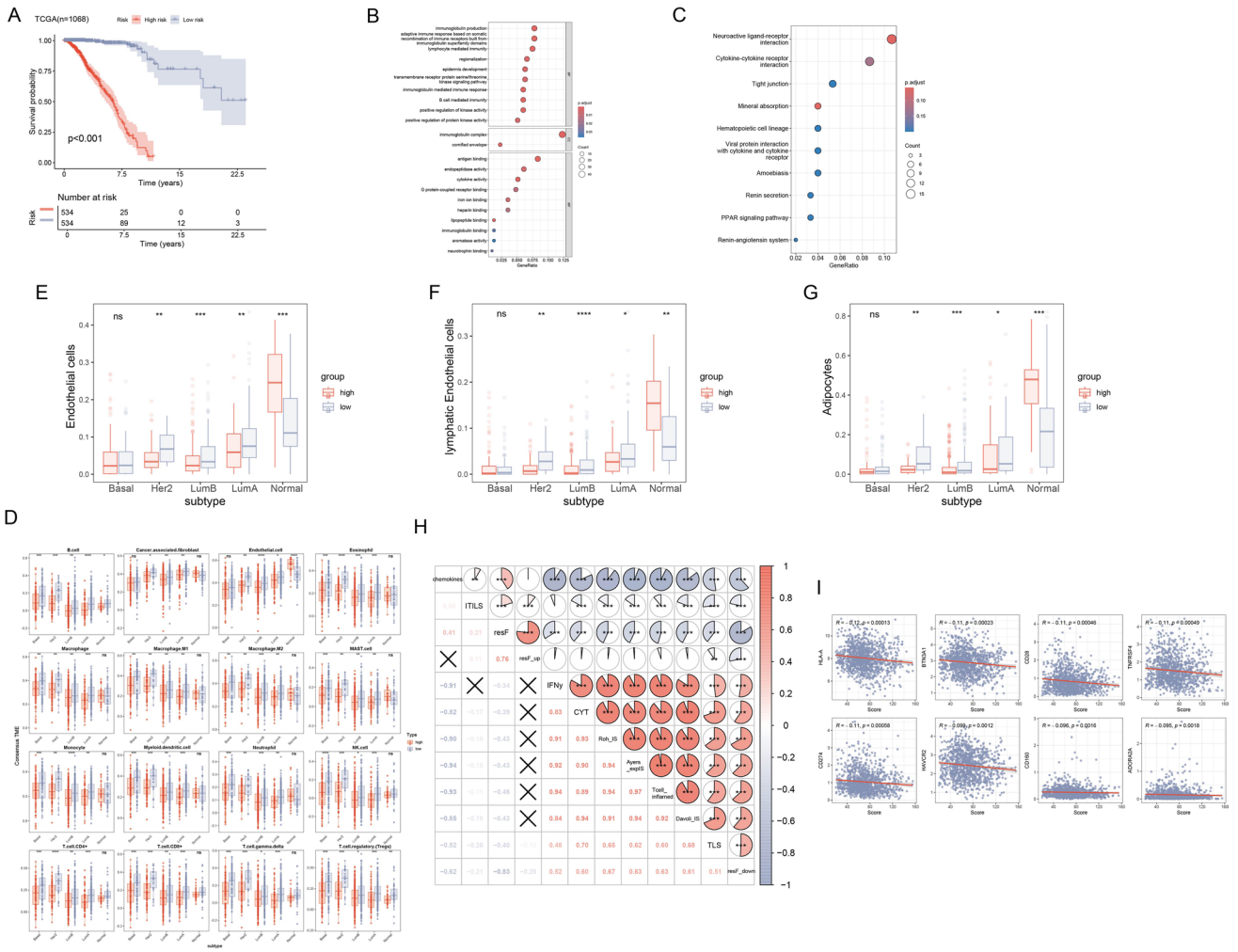


Fig. 7 Immune landscape of high- and low-risk ITILS groups. **A** Kaplan–Meier survival analysis comparing high- and low-risk ITILS groups in the TCGA-BRCA cohort. **B**, **C** Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of differentially expressed genes between ITILS high- and low-risk groups. **D** The difference in immune cell composition between high- and low-risk ITILS groups, analyzed using the CONSENSUS_TME algorithm across PAM50 subtypes. **E**–**G** Quantification of endothelial cells, lymphatic endothelial cells, and adipocytes in high- and low-risk ITILS groups across PAM50 subtypes. **H** Correlation analysis between ITILS score and various immune characteristics. **I** Correlation analysis between ITILS score and tertiary lymphoid structure-related genes. Significance levels: *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001

new opportunities for better predicting stratification and prognosis in breast cancer patients. However, the challenges in assessing the quantity of iTILs have constrained research advancements in this area. Recently, Wu et al. [10] developed a new methodology to define iTILs, employing a TIL scoring algorithm to estimate iTIL counts from the transcriptomes of breast cancers, thereby enabling the precise quantification of even minimal cellular populations.

Building on this approach, we employed eight deconvolution algorithms to compute TIL score and analyzed the correlation of each algorithm with lymphocyte markers. The TIL score derived from the CONSENSUS_TME algorithm, which displayed the strongest correlation with TILs parameters, was subsequently defined as the iTIL score. Using consensus clustering analysis in conjunction with WGCNA, we identified hub genes that are highly correlated with the iTIL score. Through univariate Cox regression analysis, we then screened for genes significantly associated with prognosis. These genes were incorporated into a framework comprising 101 machine learning methods. The optimal machine learning model, which synergistically combined CoxBoost and RSF, was selected based on the average C-index across four testing datasets, and has been designated as ITILS model. We evaluated the prognostic stratification capabilities of ITILS in both training and testing datasets using KM analysis. The findings indicated that across all cohorts, patients classified in the high-risk subgroup by the ITILS score exhibited significantly reduced OS compared to those in the low-risk subgroup. These results underscored the potential of iTILs as a powerful prognostic tool in breast cancer.

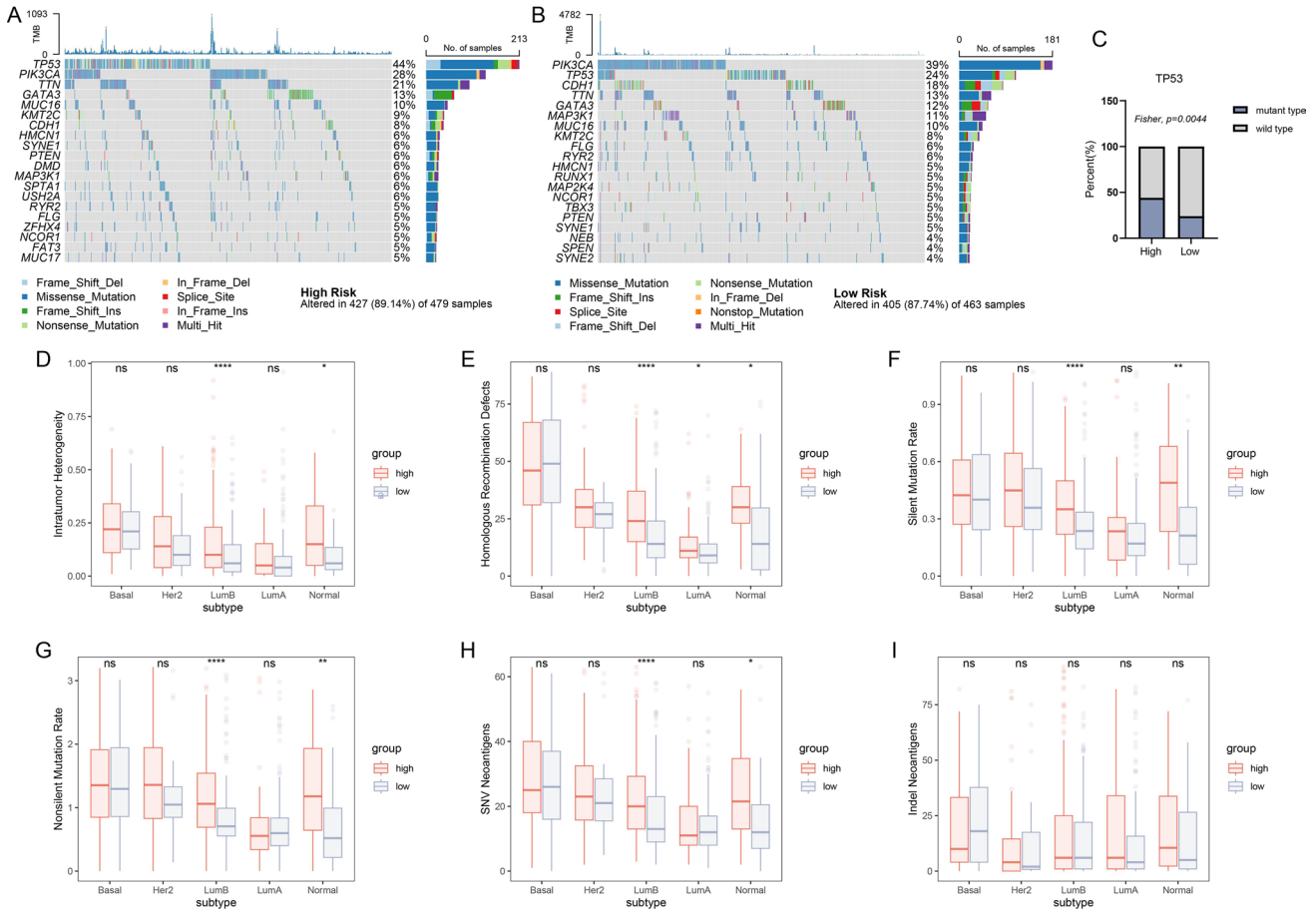


Fig. 8 Immunogenomic landscape in high- and low-risk ITILS subgroups. **A** Top 20 most frequently mutated genes in the high-risk ITILS group. **B** Top 20 most frequently mutated genes in the low-risk ITILS group. **C** Comparison of TP53 mutation rates between high-risk and low-risk ITILS groups. **D**–**I** Differences in intratumor heterogeneity, homologous recombination defects, silent mutation rate, nonsilent mutation rate, SNV neoantigens and indel neoantigens between high-risk and low-risk ITILS groups across different PAM50 subtypes. Significance levels: *p < 0.05, **p < 0.01, ****p < 0.0001

To further evaluate the prognostic predictive capability of the ITILS model, we calculated its AUC index in both the training and testing sets, with a high AUC index signifying superior predictive accuracy. Additionally, we executed an extensive literature search on PubMed to identify published prognostic models for breast cancer patients and compared the C-index of ITILS with those of other models across both training and testing datasets. Remarkably, ITILS demonstrated superior performance across all datasets, outperforming nearly all other existing models. It was noted that the majority of existing predictive models, while performing well within their own training sets, frequently underperformed when applied to external datasets [21]. In this context, the ITILS model stands out for its exceptional performance on external datasets, showcasing its advantageous utility across various data cohorts.

Clinical factors, such as TNM stage, age, and treatment modalities like radiotherapy and pharmacotherapy, are wildly known to significantly influence the survival outcomes of breast cancer patients [30, 31]. To assess the correlation between ITILS and these clinical factors, we constructed a nomogram model that integrated ITILS with relevant clinical parameters. Our analysis illustrated that ITILS can serve as an independent prognostic factor which indicated poor OS in breast cancer patients. The nomogram model was rigorously validated, demonstrating its robust predictive capabilities with AUC values of 0.947, 0.953, and 0.966 for 3-, 5-, and 10-year survival predictions, respectively. Metastasis, radiotherapy and pharmacotherapy are considered as pivotal factors influencing the prognosis of breast cancer patients [26, 32, 33]. Therefore, we further analyzed the differences in iTILs score between metastatic and non-metastatic patients, as well as between those who underwent radiotherapy or pharmacotherapy and those who did not. Our findings suggested that patients perceived to have a better prognosis (non-metastatic, undergoing radiotherapy or pharmacotherapy) exhibited significantly lower ITILS score and higher survival rates. This demonstrated that ITILS is not only predictive of

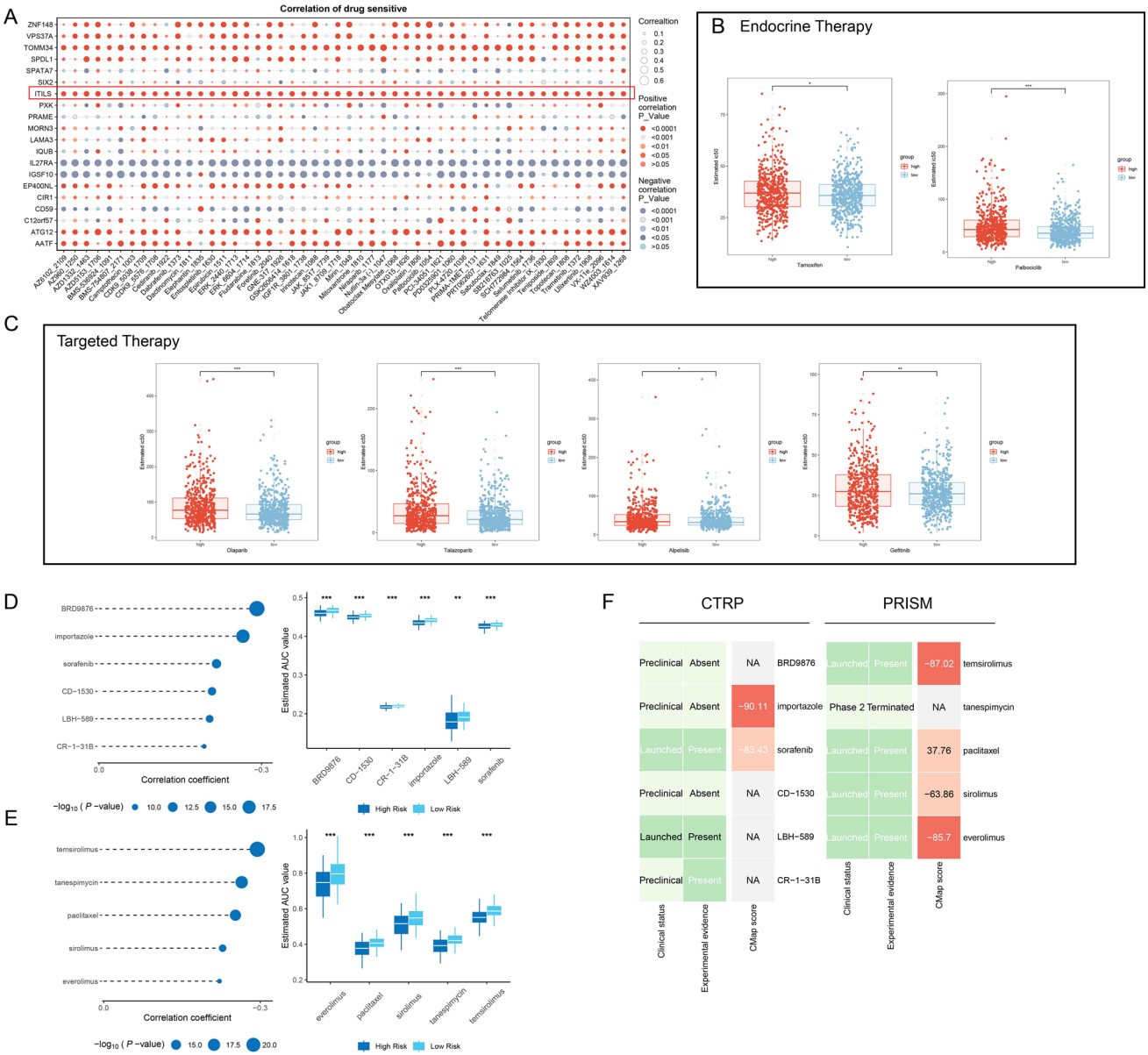


Fig. 9 Identification of promising compounds with enhanced drug sensitivity in high-risk ITILS patients. **A** The relationship between ITILS, genes within the ITILS model, and drug sensitivity. **B** Estimated IC50 values of Tamoxifen and Palbociclib in high- versus low-risk ITILS groups. **C** Estimated IC50 values of Olaparib, Talazoparib, Alpelisib, and Gefitinib in high- versus low-risk ITILS groups. **D** Correlation and differential drug response analysis of CTRP-derived agents. **E** Correlation and differential drug response analysis of PRISM-derived agents. **F** Clinical status, experimental evidence, and Connectivity Map (CMap) scores of promising compounds. Significance levels: *p < 0.05, **p < 0.01, ***p < 0.001

the prognosis for patients with primary breast cancer but also for those with metastases or those receiving radiotherapy and pharmacotherapy.

Having confirmed the stratification capability of ITILS in predicting patient outcomes, we proceeded to explore the differences among patients stratified by ITILS. We divided the TCGA-BRCA cohort into high- and low-risk groups based on the median ITILS score and performed a differential gene expression analysis between these groups. GO and KEGG pathway enrichment analyses revealed that the differentially expressed genes were primarily enriched in immune-related pathways, such as immunoglobulin production and cytokine-cytokine receptor interactions. Subsequent analyses were conducted to investigate immunological distinctions between the high- and low-risk groups from various perspectives. From the viewpoint of stromal cell infiltration, with the exception of the Basal subtype, low-risk patients exhibited a higher abundance of stromal cell infiltrates, implying a more robust immune response. From the perspective of immune

cell infiltration, our analysis showed that in the Her2, LumB, and LumA subtypes, low-risk patients had a significantly greater abundance of immune cell infiltrates compared to high-risk patients. Further analysis correlating ITILS scores with immune features disclosed a significant negative correlation with the formation of tertiary lymphoid structures (TLS), indicating a potential immunosuppressive microenvironment associated with higher ITILS scores [13, 34].

Somatic mutations are recognized as pivotal determinants linked with cancer development [35], tumor immunity [36], and patient prognosis [37]. Therefore, we concentrated on discerning the genomic disparities between high- and low-risk groups based on somatic mutations. We discovered a markedly elevated mutation rate of TP53 in the high-risk group compared to the low-risk group. This observation is corroborated by previous studies, which posited that an increased TP53 mutation rate is indicative of a poorer prognosis [37].

Furthermore, considering the relevance of intratumor heterogeneity (ITH) to the heterogeneity of immune cell infiltration [38], the association of homologous recombination deficiencies (HRD) with both innate immune responses and immune evasion mechanisms [39], and the correlation of silent and non-silent mutations, as well as tumor neoantigens (SNV Neoantigens, Indel Neoantigens) with immune infiltration [15], we proceeded to analyze the disparities in these aspects between high-risk and low-risk groups. Our results indicated that in the LumB subtype, the high-risk group exhibited significantly higher levels of ITH, HRD, silent mutations, non-silent mutations, and SNV neoantigens compared to the low-risk group. Integrating our findings with the seminal study of Thorsson et al. [15], who classified cancer into six immune subtypes, we noted their observations that C1 subtype displayed the highest level of ITH, while the C3 subtype exhibited the lowest. On this basis, we inferred that the high-risk group in our study presented an ITH landscape similar to the C1 subtype, whereas the low-risk group was analogous to the C3 subtype. Additionally, their research corroborated that the C3 subtype is linked to a superior prognosis, aligning with our observation that the low-risk group experienced more favorable outcomes.

To ascertain efficacious therapeutic drugs for breast cancer patients, we undertook a stratified analysis based on ITILS scores. Subsequently, leveraging the GDSC2 database, we estimated the IC50 values for a spectrum of drugs within the stratified TCGA-BRCA patient cohort. Our results demonstrated that the top 50 drugs, which were significantly linked to the ITILS score ($p < 0.05$), displayed a uniform positive correlation with the ITILS score, implying that high-risk patients may lack sensitive to these drugs. In pursuit of identifying efficacious drugs for high-risk patients, we initially focused on several commonly used drugs in breast cancer treatment, encompassing the endocrine therapy drugs such as Tamoxifen [40] and Palbociclib [41], as well as the targeted therapy drugs including Olaparib [42], Talazoparib [43], Alpelisib [44], and Gefitinib [45]. Our findings elucidated a diminished sensitivity among high-risk patients to these widely prescribed drugs, thus necessitating the identification of therapeutic drugs that exhibit efficacy in high-risk patients. Utilizing the CTRP and PRISM datasets, we performed a differential drug response analysis, specifically targeting drugs with AUC values that negatively correlated with their ITILS scores. Then we combined the current experimental and clinical research on these drugs with their corresponding CMap scores, thereby facilitating the identification of drugs with potential therapeutic benefits for high-risk patients.

Intriguingly, our analysis identified Importazole, Temsirolimus, and Everolimus as drugs with CMap scores less than −85, suggesting that high-risk patients might display an increased sensitivity to these drugs. Importazole characterized as an inhibitor of the nuclear transport receptor importin-β [46], has demonstrated its efficacy in inhibiting the growth of prostate tumors in both in vitro assays and mouse models. However, no related studies have yet explored its implications in breast cancer. Temsirolimus, an mTOR inhibitor, earned FDA approval in 2007 for the treatment of renal cell carcinoma [47] and has shown promise in the treatment of breast cancer, as evidenced by a series of preclinical and clinical trials [48–51]. Everolimus is an mTOR1 inhibitor that has received FDA approval for the treatment of HER2-negative breast cancer [47]. It is currently a second-line treatment for patients who are HR-positive, HER2-negative, and without BRCA mutations or visceral crisis [52]. Given these comprehensive analyses, we concluded that Temsirolimus and Everolimus are the most promising candidates for the treatment of patients with high-risk ITILS scores.

5 Limitations

Our study has successsfully established a stratified and prognostic ITILS signature for breast cancer patients based on iTILs, elucidating its clinical significance and identifying potential therapeutic drugs for high-risk patients. Nonetheless, there are inherent limitations to acknowledge. Our findings are predicated on retrospective cohorts, necessitating prospective studies to validate the clinical validity of our findings. Although our investigation has thoroughly characterized the differential expression patterns of the 19-gene signature within the TCGA-BRCA cohort, further functional studies

are required to elucidate the mechanistic roles of these genes in breast cancer initiation and progression. Specifically, experimental validation—such as in vitro and in vivo models—is needed to determine whether these genes act as drivers, suppressors, or bystanders in tumorigenesis and to identify their downstream molecular pathways.

6 Conclusions

Overall, our study has developed a robust predictive model for the stratification and prognostic prediction of breast cancer patients based on iTILs by using a combination of various machine learning frameworks. We conducted an in-depth analysis to discern the clinical and molecular variances between patient subgroups stratified by this model and predicted therapeutic drugs for high-risk patients. This research provided valuable insights into prognostic prediction and precision treatment for breast cancer patients.

Acknowledgements We are very grateful to Professor Siguang Li from Tongji University School of Medicine for financial support and scientific supervision of this project.

Author contributions Xinyi Wu contributed to the design of the study, the collection, verification, and analysis of the data, the visualization and generation of figures, and writing of the original draft. Chun Li contributed to the conception of the study, supervision of the work, administration of the project, analysis of the data, and revising/editing of the article. All authors have read and approved the final version of the article.

Funding This study was supported by grants from the National Natural Science Foundation of China (Grant numbers: 82171387 and 31830111) and Key Research and Innovation Program of Shanghai Municipal Education Commission (Grant number: 2019–01-07–00-07-E00040).

Data availability The datasets used in this study are from public databases and can be acquired through the following websites. UCSC Xena database: https://xenabrowser.net/datapages/; cBioPortal for Cancer Genomics: https://www.cbioportal.org/datasets; Gene Expression Omnibus: http://www.ncbi.nlm.nih.gov/geo/. Other data supporting the findings of this study are available from the corresponding author upon reasonable request.

Code availability The source code utilized for our machine learning analyses is publicly available on GitHub. This repository can be accessed via the following link: https://github.com/uniwu/uni_ml.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin. 2023;73(1):17–48.
2. Denkert C, Loibl S, Noske A, Roller M, Müller BM, Komor M, Budczies J, Darb-Esfahani S, Kronenwett R, Hanusch C, von Törne C, Weichert W, Engels K, Solbach C, Schrader I, Dietel M, von Minckwitz G. Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. J Clin Oncol. 2010;28(1):105–13.
3. Loi S, Sirtaine N, Piette F, Salgado R, Viale G, Van Eenoo F, Rouas G, Francis P, Crown JP, Hitre E, de Azambuja E, Quinaux E, Di Leo A, Michiels S, Piccart MJ, Sotiriou C. Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02–98. J Clin Oncol. 2013;31(7):860–7.

4. Loi S, Drubay D, Adams S, Pruneri G, Francis PA, Lacroix-Triki M, Joensuu H, Dieci MV, Badve S, Demaria S, Gray R, Munzone E, Lemonnier J, Sotiriou C, Piccart MJ, Kellokumpu-Lehtinen PL, Vingiani A, Gray K, Andre F, Denkert C, Salgado R, Michiels S. Tumor-infiltrating lymphocytes and prognosis: a pooled individual patient analysis of early-stage triple-negative breast cancers. J Clin Oncol. 2019;37(7):559–69.
5. Denkert C, von Minckwitz G, Darb-Esfahani S, Lederer B, Heppner BI, Weber KE, Budczies J, Huober J, Klauschen F, Furlanetto J, Schmitt WD, Blohmer JU, Karn T, Pfitzner BM, Kümmel S, Engels K, Schneeweiss A, Hartmann A, Noske A, Fasching PA, Jackisch C, van Mackelenbergh M, Sinn P, Schem C, Hanusch C, Untch M, Loibl S. Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. Lancet Oncol. 2018;19(1):40–50.
6. Salgado R, Denkert C, Demaria S, Sirtaine N, Klauschen F, Pruneri G, Wienert S, Van den Eynden G, Baehner FL, Penault-Llorca F, Perez EA, Thompson EA, Symmans WF, Richardson AL, Brock J, Criscitiello C, Bailey H, Ignatiadis M, Floris G, Sparano J, Kos Z, Nielsen T, Rimm DL, Allison KH, Reis-Filho JS, Loibl S, Sotiriou C, Viale G, Badve S, Adams S, Willard-Gallo K, Loi S. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. Ann Oncol. 2015;26(2):259–71.
7. Ruan M, Tian T, Rao J, Xu X, Yu B, Yang W, Shui R. Predictive value of tumor-infiltrating lymphocytes to pathological complete response in neoadjuvant treated triple-negative breast cancers. Diagn Pathol. 2018;13(1):66.
8. Khoury T, Nagrale V, Opyrchal M, Peng X, Wang D, Yao S. Prognostic significance of stromal versus intratumoral infiltrating lymphocytes in different subtypes of breast cancer treated with cytotoxic neoadjuvant chemotherapy. Appl Immunohistochem Mol Morphol. 2018;26(8):523–32.
9. Li S, Zhang Y, Zhang P, Xue S, Chen Y, Sun L, Yang R. Predictive and prognostic values of tumor infiltrating lymphocytes in breast cancers treated with neoadjuvant chemotherapy: a meta-analysis. Breast. 2022;66:97–109.
10. Wu R, Oshi M, Asaoka M, Yan L, Benesch MGK, Khoury T, Nagahashi M, Miyoshi Y, Endo I, Ishikawa T, Takabe K. Intratumoral tumor infiltrating lymphocytes (TILs) are associated with cell proliferation and better survival but not always with chemotherapy response in breast cancer. Ann Surg. 2023;278(4):587–97.
11. Sturm G, Finotello F, List M. Immunedeconv: an R package for unified access to computational methods for estimating immune cell fractions from bulk RNA-sequencing data. Methods Mol Biol. 2020;2120:223–32.
12. Lapuente-Santana Ó, van Genderen M, Hilbers PAJ, Finotello F, Eduati F. Interpretable systems biomarkers predict response to immune-checkpoint inhibitors. Patterns. 2021;2(8):100293.
13. Sautès-Fridman C, Petitprez F, Calderaro J, Fridman WH. Tertiary lymphoid structures in the era of cancer immunotherapy. Nat Rev Cancer. 2019;19(6):307–25.
14. Cabrita R, Lauss M, Sanna A, Donia M, Skaarup Larsen M, Mitra S, Johansson I, Phung B, Harbst K, Vallon-Christersson J, van Schoiack A, Lövgren K, Warren S, Jirström K, Olsson H, Pietras K, Ingvar C, Isaksson K, Schadendorf D, Schmidt H, Bastholt L, Carneiro A, Wargo JA, Svane IM, Jönsson G. Tertiary lymphoid structures improve immunotherapy and survival in melanoma. Nature. 2020;577(7791):561–5.
15. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA, Ziv E, Culhane AC, Paull EO, Sivakumar IKA, Gentles AJ, Malhotra R, Farshidfar F, Colaprico A, Parker JS, Mose LE, Vo NS, Liu J, Liu Y, Rader J, Dhankani V, Reynolds SM, Bowlby R, Califano A, Cherniack AD, Anastassiou D, Bedognetti D, Mokrab Y, Newman AM, Rao A, Chen K, Krasnitz A, Hu H, Malta TM, Noushmehr H, Pedamallu CS, Bullman S, Ojesina AI, Lamb A, Zhou W, Shen H, Choueiri TK, Weinstein JN, Guinney J, Saltz J, Holt RA, Rabkin CS, Lazar AJ, Serody JS, Demicco EG, Disis ML, Vincent BG, Shmulevich I. The immune landscape of cancer. Immunity. 2018;48(4):812–30.
16. Danaher P, Warren S, Dennis L, D'Amico L, White A, Disis ML, Geller MA, Odunsi K, Beechem J, Fling SP. Gene expression markers of tumor infiltrating leukocytes. J Immunother Cancer. 2017;5:18.
17. Yang Y, Li Z, Zhong Q, Zhao L, Wang Y, Chi H. Identification and validation of a novel prognostic signature based on transcription factors in breast cancer by bioinformatics analysis. Gland Surg. 2022;11(5):892–912.
18. Wu Q, Zheng S, Lin N, Xie X. Comprehensive research into prognostic and immune signatures of transcription factor family in breast cancer. BMC Med Genomics. 2023;16(1):87.
19. Lu YJ, Gong Y, Li WJ, Zhao CY, Guo F. The prognostic significance of a novel ferroptosis-related gene model in breast cancer. Ann Transl Med. 2022;10(4):184.
20. Ye Z, Zou S, Niu Z, Xu Z, Hu Y. A novel risk model based on lipid metabolism-associated genes predicts prognosis and indicates immune microenvironment in breast cancer. Front Cell Dev Biol. 2021;9: 691676.
21. Ye Y, Luo Y, Guo T, Zhang C, Sun Y, Xu A, Ji L, Ou J, Wu SY. Leveraging senescence-oxidative stress co-relation to predict prognosis and drug sensitivity in breast invasive carcinoma. Front Endocrinol (Lausanne). 2023;14:1179050.
22. Li L, Yang W, Jia D, Zheng S, Gao Y, Wang G. Establishment of a N1-methyladenosine-related risk signature for breast carcinoma by bioinformatics analysis and experimental validation. Breast Cancer. 2023;30(4):666–84.
23. Li J, Huang G, Ren C, Wang N, Sui S, Zhao Z, Li M. Identification of differentially expressed genes-related prognostic risk model for survival prediction in breast carcinoma patients. Aging (Albany NY). 2021;13(12):16577–99.
24. Tian C, Wang Y, Song X. Prognostic characteristics of immune-related genes and the related regulatory axis in patients with stage N+M0 breast cancer. Front Oncol. 2022;12: 878219.
25. Liang X, Peng Z, Lin Z, Lin X, Lin W, Deng Y, Yang S, Wei S. Identification of prognostic genes for breast cancer related to systemic lupus erythematosus by integrated analysis and machine learning. Immunobiology. 2023;228(5): 152730.
26. Gennari A, Conte P, Rosso R, Orlandini C, Bruzzi P. Survival of metastatic breast carcinoma patients over a 20-year period: a retrospective analysis based on individual patient data from six consecutive studies. Cancer. 2005;104(8):1742–50.
27. Qi G, Zhang X, Gai X, Yan X. Retrospective analysis of estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor-2 (HER2), Ki67 changes and their clinical significance between primary breast cancer and metastatic tumors. PeerJ. 2024;12: e17377.
28. Wu R, Sarkar J, Tokumaru Y, Takabe Y, Oshi M, Asaoka M, Yan L, Ishikawa T, Takabe K. Intratumoral lymphatic endothelial cell infiltration reflecting lymphangiogenesis is counterbalanced by immune responses and better cancer biology in the breast cancer tumor microenvironment. Am J Cancer Res. 2022;12(2):504–20.
29. Tokumaru Y, Oshi M, Katsuta E, Yan L, Huang JL, Nagahashi M, Matsuhashi N, Futamura M, Yoshida K, Takabe K. Intratumoral adipocyte-high breast cancer enrich for metastatic and inflammation-related pathways but associated with less cancer cell proliferation. Int J Mol Sci. 2020. https://doi.org/10.3390/ijms21165744.

30. Yang P, Zhang G, Zhang Y, Zhao W, Tang J, Zeng S, Lv X, Lv L. Effect of adjuvant radiotherapy on overall survival and breast cancer-specific survival of patients with malignant phyllodes tumor of the breast in different age groups: a retrospective observational study based on SEER. Radiat Oncol. 2024;19(1):59.
31. Lu YS, Mahidin EI, Azim H, Eralp Y, Yap YS, Im SA, Rihani J, Gokmen E, El Bastawisy A, Karadurmus N, Lim YN. Final results of RIGHT choice: ribociclib plus endocrine therapy vs combination chemotherapy in premenopausal women with clinically aggressive HR+/HER2- advanced breast cancer. J Clin Oncol. 2024;42:2812.
32. Bradley JA, Mendenhall NP. Novel radiotherapy techniques for breast cancer. Annu Rev Med. 2018;69:277–88.
33. Anthracycline-containing and taxane-containing chemotherapy for early-stage operable breast cancer: a patient-level meta-analysis of 100 000 women from 86 randomised trials, Lancet 401(10384) (2023) 1277–1292.
34. Fridman WH, Meylan M, Petitprez F, Sun CM, Italiano A, Sautès-Fridman C. B cells and tertiary lymphoid structures as determinants of tumour immune contexture and clinical outcome. Nat Rev Clin Oncol. 2022;19(7):441–57.
35. Martínez-Jiménez F, Muiños F, Sentís I, Deu-Pons J, Reyes-Salazar I, Arnedo-Pac C, Mularoni L, Pich O, Bonet J, Kranas H, Gonzalez-Perez A, Lopez-Bigas N. A compendium of mutational cancer driver genes. Nat Rev Cancer. 2020;20(10):555–72.
36. Wellenstein MD, de Visser KE. Cancer-cell-intrinsic mechanisms shaping the tumor immune landscape. Immunity. 2018;48(3):399–416.
37. Silwal-Pandit L, Vollan HK, Chin SF, Rueda OM, McKinney S, Osako T, Quigley DA, Kristensen VN, Aparicio S, Børresen-Dale AL, Caldas C, Langerød A. TP53 mutation spectrum in breast cancer is subtype specific and has distinct prognostic relevance. Clin Cancer Res. 2014;20(13):3569–80.
38. Marusyk A, Janiszewska M, Polyak K. Intratumor heterogeneity: the rosetta stone of therapy resistance. Cancer Cell. 2020;37(4):471–84.
39. Marzio A, Kurz E, Sahni JM, Di Feo G, Puccini J, Jiang S, Hirsch CA, Arbini AA, Wu WL, Pass HI, Bar-Sagi D, Papagiannakopoulos T, Pagano M. EMSY inhibits homologous recombination repair and the interferon response, promoting lung cancer immune evasion. Cell. 2022;185(1):169-183.e19.
40. Buijs SM, Koolen SLW, Mathijssen RHJ, Jager A. Tamoxifen dose de-escalation: an effective strategy for reducing adverse effects? Drugs. 2024;84(4):385–401.
41. Wekking D, Leoni VP, Lambertini M, Dessì M, Pretta A, Cadoni A, Atzori L, Scartozzi M, Solinas C. CDK4/6 inhibition in hormone receptor-positive/HER2-negative breast cancer: Biological and clinical aspects. Cytokine Growth Factor Rev. 2024;75:57–64.
42. Tutt ANJ, Garber JE, Kaufman B, Viale G, Fumagalli D, Rastogi P, Gelber RD, de Azambuja E, Fielding A, Balmaña J, Domchek SM, Gelmon KA, Hollingsworth SJ, Korde LA, Linderholm B, Bandos H, Senkus E, Suga JM, Shao Z, Pippas AW, Nowecki Z, Huzarski T, Ganz PA, Lucas PC, Baker N, Loibl S, McConnell R, Piccart M, Schmutzler R, Steger GG, Costantino JP, Arahmani A, Wolmark N, McFadden E, Karantza V, Lakhani SR, Yothers G, Campbell C, Geyer CE Jr. Adjuvant olaparib for patients with BRCA1- or BRCA2-mutated breast cancer. N Engl J Med. 2021;384(25):2394–405.
43. Litton JK, Rugo HS, Ettl J, Hurvitz SA, Gonçalves A, Lee KH, Fehrenbacher L, Yerushalmi R, Mina LA, Martin M, Roché H, Im YH, Quek RGW, Markova D, Tudor IC, Hannah AL, Eiermann W, Blum JL. Talazoparib in patients with advanced breast cancer and a germline BRCA mutation. N Engl J Med. 2018;379(8):753–63.
44. André F, Ciruelos E, Rubovszky G, Campone M, Loibl S, Rugo HS, Iwata H, Conte P, Mayer IA, Kaufman B, Yamashita T, Lu YS, Inoue K, Takahashi M, Pápai Z, Longin AS, Mills D, Wilke C, Hirawat S, Juric D. Alpelisib for PIK3CA-mutated, hormone receptor-positive advanced breast cancer. N Engl J Med. 2019;380(20):1929–40.
45. Chemmalar S, Intan Shameha AR, Che Abdullah CA, Ab Razak NA, Yusof LM, Ajat M, Chan KW, Abu Bakar Zakaria MZ. Busting the breast cancer with AstraZeneca's Gefitinib. Adv Pharmacol Pharm Sci. 2023;1:8127695.
46. Soderholm JF, Bird SL, Kalab P, Sampathkumar Y, Hasegawa K, Uehara-Bingen M, Weis K, Heald R. Importazole, a small molecule inhibitor of the transport receptor importin-β. ACS Chem Biol. 2011;6(7):700–8.
47. Roskoski R Jr. Properties of FDA-approved small molecule protein kinase inhibitors: A 2024 update. Pharmacol Res. 2024;200: 107059.
48. Chan S, Scheulen ME, Johnston S, Mross K, Cardoso F, Dittrich C, Eiermann W, Hess D, Morant R, Semiglazov V, Borner M, Salzberg M, Ostapenko V, Illiger HJ, Behringer D, Bardy-Bouxin N, Boni J, Kong S, Cincotta M, Moore L. Phase II study of temsirolimus (CCI-779), a novel inhibitor of mTOR, in heavily pretreated patients with locally advanced or metastatic breast cancer. J Clin Oncol. 2005;23(23):5314–22.
49. Fleming GF, Ma CX, Huo D, Sattar H, Tretiakova M, Lin L, Hahn OM, Olopade FO, Nanda R, Hoffman PC, Naughton MJ, Pluard T, Conzen SD, Ellis MJ. Phase II trial of temsirolimus in patients with metastatic breast cancer. Breast Cancer Res Treat. 2012;136(2):355–63.
50. Ma CX, Suman VJ, Goetz M, Haluska P, Moynihan T, Nanda R, Olopade O, Pluard T, Guo Z, Chen HX, Erlichman C, Ellis MJ, Fleming GF. A phase I trial of the IGF-1R antibody Cixutumumab in combination with temsirolimus in patients with metastatic breast cancer. Breast Cancer Res Treat. 2013;139(1):145–53.
51. Wolff AC, Lazar AA, Bondarenko I, Garin AM, Brincat S, Chow L, Sun Y, Neskovic-Konstantinovic Z, Guimaraes RC, Fumoleau P, Chan A, Hachemi S, Strahs A, Cincotta M, Berkenblit A, Krygowski M, Kang LL, Moore L, Hayes DF. Randomized phase III placebo-controlled trial of letrozole plus oral temsirolimus as first-line endocrine therapy in postmenopausal women with locally advanced or metastatic breast cancer. J Clin Oncol. 2013;31(2):195–202.
52. Moreau-Bachelard C, Robert M, Gourmelon C, Bourbouloux E, Patsouris A, Frenel JS, Campone M. Evaluating everolimus for the treatment of breast cancer. Expert Opin Pharmacother. 2023;24(10):1105–11.