

RESEARCH

Open Access



Identification of cell adhesion-related subtypes and construction of risk model to predict breast cancer prognostic and immunological properties

De-Ming Lv¹, Li Yang¹, Chen Fan¹, Ling-Hui Fang¹ and Su-Fen Cheng^{1*}

Abstract

Background Breast invasive carcinoma is the most common form of breast cancer, often resulting in recurrence or metastasis in patients. Cell adhesion molecules play a crucial role in modulating the interactions between tumor cells and surrounding cells. The study aims to identify breast cancer subtypes related to cell adhesion and develop prognostic models that are essential for evaluating the prognostic risk and immunological profile of breast cancer.

Methods Transcriptome and clinical data were obtained from The Cancer Genome Atlas (TCGA) database, while cell adhesion-related genes (CARGs) from the MSigDB database. Molecular subtyping was performed using NMF clustering. Cox regression and Least absolute shrinkage and selection operator (LASSO) regression analyses were employed to construct a risk model for predicting patient prognosis. This model was validated in independent Gene Expression Omnibus (GEO) datasets, specifically GSE20685 and GSE42568. Immune cell infiltration was explored utilizing the CIBERSORT algorithm. Subsequently, we analyzed tumor mutation burden (TMB). Finally, potential drugs and drug sensitivity was evaluated using pRRobhetic algorithm.

Results Based on the expression levels of 39 genes related to cell adhesion, we identified 3 distinct subtypes, and LASSO regression analysis identified 8 genes that could be used as prognostic markers. Receiver operating characteristic (ROC) curves demonstrated that these cell adhesion genes were effective in predicting patient prognosis. Compared to the high-risk group, the low-risk group had a more favorable prognosis and a greater response to immunotherapy. These prognostic genes were found to be closely associated with immune cell infiltration and the response to immunotherapy. Furthermore, their significant associations with breast cancer sensitivities to anti-cancer drugs were revealed.

Conclusion We developed a risk model focused on cell adhesion-related genes. This model accurately predicts the prognosis for breast cancer patients. It may also offer new insights for clinical decisions and immunotherapy.

Highlights

1. This study established molecular subtypes based on cell adhesion-related genes and established a corresponding prognostic model that has strong predictive power.

*Correspondence:
Su-Fen Cheng
csf15958455391@163.com



2. This study further revealed potential associations with immune cell infiltration and patient's responsiveness to immunotherapy.
3. This study evaluated the signature genes of the model and analyzed their mRNA expression in breast cancer.

Keywords Breast cancer, Cell adhesion-related genes, Prognostic markers

Introduction

Breast cancer stands as a leading cause of mortality in postmenopausal women, accounting for 23% of all cancer-related deaths [1]. According to global morbidity and mortality data for 2022, it ranks the second highest incidence rate among cancers [2]. Despite advancements in therapeutic strategies, patients with breast cancer continue to grapple with the risk of recurrence and metastasis, with over 90% of mortality attributed to metastatic progression [3]. Depending on disease stage and pathological characteristics, treatment strategies for breast invasive carcinoma include surgery, chemotherapy, and antibody therapies such as trastuzumab and lapatinib [4]. However, both intrinsic and acquired resistance remains a major obstacle to breast cancer treatment. While conventional molecular subtypes defined by estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) status remain the cornerstone of clinical stratification, there is an urgent need to discover novel predictive biomarkers and molecular subtypes to improve risk stratification for high-risk patient populations. Cell adhesion molecules (CAMs), serving as membrane-associated receptors, facilitate interactions among cells and between cells and the extracellular matrix. They play a pivotal role in the intracellular signaling pathways that regulate essential processes, including adhesion, cellular migration, angiogenesis, and the specific organ tropism of metastatic cells [5]. The majority of CAM components belong to the calcineurin, selectin, and integrin families [6]. β 2 integrins bind intercellular adhesion molecules and are involved in transendothelial migration (TEM) and leukocyte activation [7]. As tumors progress, they become increasingly heterogeneous, with the consequent generation of aggressive subpopulations of tumor cells that subsequently infiltrate surrounding tissues, lymphatic systems, and the bloodstream [8]. The process of successful tumor metastasis is complex and requires a reduction in adhesive interactions between tumor cells and their surrounding cells [9]. Cell adhesion factors are also considered promising targets in pathology. Identifying the adhesion factors that regulate this process is crucial for future therapies aimed at combating breast carcinogenesis and metastasis.

In our research, the primary objective was to develop a prognostic model that leverages cell adhesion-associated

genes to forecast the outcomes for patients with breast cancer. We stratified patients in the Cancer Genome Atlas database (TCGA) dataset into three distinct groups, based on their expression patterns of genes related to cell adhesion. Through the Least Absolute Shrinkage and Selection Operator-Cox Proportional Hazards (LASSO-Cox) regression analysis, we identified a robust 8-gene signature model. Notably, this signature exhibited significant correlations with tumor immune microenvironment (TME) and clinical outcomes, highlighting its potential utility for risk stratification and therapeutic decision-making in breast cancer management. The findings suggest a link between cell adhesion, immune-related signature and clinical outcome, providing value in the prognostic assessment of breast cancer.

Methods

Data acquisition

The overall flowchart of this study was shown in Fig. 1. The following data were retrieved from the TCGA database (<https://portal.gdc.cancer.gov/>): TCGA-BRCA (breast invasive carcinoma) RNA-Seq data, breast cancer variant copy numbers, and clinical information, comprising 103 normal samples and 1104 breast cancer samples. Patients with incomplete survival information were excluded from the study. Additionally, microarray data were sourced from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>), specifically utilizing the GSE42568 [10] and GSE20685 [11] datasets as validation sets, relevant pathological information from these datasets were collated in Supplementary Table S1 and Supplementary Table S2. The Molecular Signatures Database (MSigDB, <https://www.gsea-msigdb.org/gsea/msigdb>) database was consulted to identify cell adhesion-related genes (CARGs).

Analysis of genes related to cell adhesion molecules

The “edgeR” package was utilized to examine the difference between the normal and tumor groups for breast cancer, setting the criteria as ($|\log$ fold change (FC) > 1], false discovery rate (FDR) < 0.05). Subsequently, we utilized enrichment analysis on the differential expression genes (DEGs) obtained. The overlapping genes between the differential genes and those related to cell adhesion molecules were designated as differentially expressed cell adhesion-related genes (DECARGs). To pinpoint genes

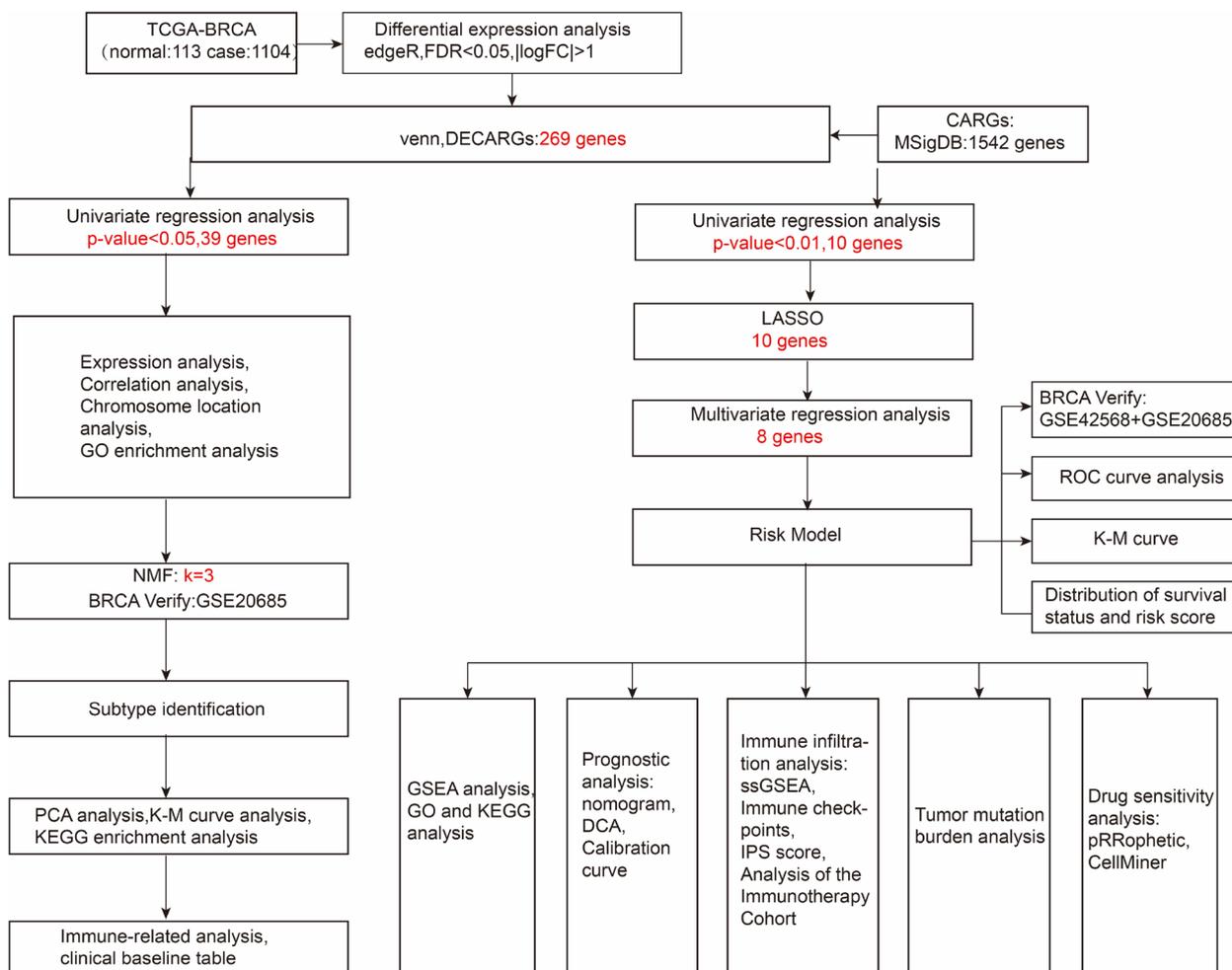


Fig. 1 Overall flowchart of the study. This study used the TCGA-BRCA (breast invasive carcinoma) cohort as the training set and the GSE42568 and GSE20685 datasets as the validation set. A total of 1542 cell adhesion-related genes (CARGs) were included in the study from the Molecular Signatures Database (MSigDB). Of these, 269 DECARGs were identified as differentially expressed genes in Breast invasive carcinoma (breast cancer). 3 cell adhesion related subtypes were identified by NMF algorithm. Cox and LASSO regression ultimately identified 8 genes involved in the construction of prognostic risk models. The GSE42568 and GSE20685 cohort was used for model validation. Kaplan-Meier (K-M) and Receiver operating characteristic (ROC) curves were employed for evaluation. Mechanisms were explored using GSEA, GO, KEGG and tumour mutation load (TMB) analyses. Furthermore, immune landscapes were analysed using algorithms such as ssGSEA. Subsequently, the IMvigor210 cohort was conducted to validate the predictive value of CARGs on the efficacy of immunotherapy. Finally, drug sensitivity analyses were performed using the CellMiner database and the pRRophetic package to identify potential therapeutic agents

linked to prognosis, we conducted univariate analysis using the “survival” package, with a significance threshold of $P < 0.05$. Furthermore, Gene Ontology (GO) enrichment analysis was applied to the prognosis-related genes that exhibited $P < 0.05$ in the univariate analysis. Additionally, we assessed the expression levels and correlations of these prognosis-related genes in both normal and tumor groups.

Identification of cell adhesion-related gene subtypes

We clustered breast cancer tumor samples according to the differential cell adhesion-related gene expression

profile matrix with Non-Negative Matrix Factorization (NMF) algorithm. We applied survival analysis, subtype difference analysis ($|\log FC| > 0.585$, $FDR < 0.05$), and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis on clustered samples. NMF cluster analysis and survival analysis of clustered samples were validated using validation set GSE20685. Subsequently, we explored differences in immune-related function and immune cell infiltration between subtypes using the single-sample Gene Set Enrichment Analysis (ssGSEA), assessing immunity score, stroma score, ESTIMATE score, and tumor purity utilizing ESTIMATE algorithm.

The CIBERSORT algorithm was employed to quantify infiltration levels of immune cells within the different subtypes.

Screening of prognostically relevant features to construct prognostic models

To enhance the specificity of the genes included in our prognostic model, we screened genes using univariate Cox regression analysis with a significance level of $P < 0.01$. To mitigate the risk of overfitting the model, we subsequently applied LASSO regression analysis to the candidate prognostic gene. Through cross-validation, we determined an appropriate penalty parameter lambda, which allowed us to eliminate genes with strong correlations, thereby simplifying the model. We used the median risk score divided samples into high- and low-risk groups. We then calculated the AUC values for 1-, 3-, and 5-year ROC curves. Additionally, we performed survival analysis on the genes included in the model and plotted Kaplan-Meier (K-M) curves for these genes in the training set.

Enrichment analysis and independent prognostic analysis

Utilizing GSEA v4.3.2 software, we conducted pathway enrichment analysis to compare high-risk and low-risk groups. Subsequently, we performed GO and KEGG enrichment analyses of the DEGs between the high- and low-risk groups. By integrating clinical data with risk scores, we conducted both univariate and multivariate regression analyses. To predict patient survival rates at 1, 3, and 5 years, we developed a nomogram. To validate the predictive accuracy of this nomogram and assess its potential as an independent prognostic factor, we plotted corresponding calibration curves. Furthermore, we employed Decision Curve Analysis (DCA) to evaluate the practical utility of our predictive models.

Subgroup analysis of risk model based on clinicopathological features

For TCGA-BRCA patients, we analyzed the correlation between the prognostic risk score and clinical characteristics, and subsequently generated violin plots for various clinical subgroups. Patients were stratified into different subgroups according to factors such as age ($< = 65$ and $>$

65), stage, TNM stage to map the Kaplan-Meier survival curves.

Immune cell infiltration analysis and prediction of response to immunotherapy

We calculated immune infiltration scores for 29 immune cell types or functions using the ssGSEA algorithm. Subsequently, we employed the ESTIMATE algorithm to derive the immune score, stromal score, ESTIMATE score, and tumor purity for both high- and low-risk groups, and compared the differences. We employed the EPIC algorithm to estimate immune cell proportions and quantified the level of immune checkpoints, and we calculated the immunophenoscore (IPS) based on data from The Cancer Immunome Atlas (TCIA, <https://tcia.at>). To assess the performance of the risk score in terms of immunotherapy responsiveness (immune checkpoint blockade), we collected transcriptomic data from patients treated with anti-PD-L1 therapy in the IMvigor210 cohort.

Assessment of tumor mutation burden (TMB)

To quantify the TMB score for each sample, we utilized mutation data from TCGA-BRCA cohort, and employed Wilcoxon test to compare the TMB values between the high-risk and low-risk groups. Then, the mutation data for top20 genes were organized and counted.

Drug sensitivity prediction

To further identify potential therapeutic targets and more effective therapeutic agents, we screened the CellMiner database (accessible at <https://discover.nci.nih.gov/cellminer/>) for targeted antitumor drugs whose sensitivity exhibited a significant correlation with our prognostic signature genes. In addition, we employed the “pRRophic” package to estimate the half maximal inhibitory concentration (IC50) values for different drugs.

Results

Identification of DEGs and CARGs in breast cancer

We utilized differential expression analysis of the tumor and normal groups of TCGA-BRCA cohort and obtained 2248 DEGs, comprising 1060 upregulated and 1188 downregulated genes (Fig. 2A). We intersected DEGs

(See figure on next page.)

Fig. 2 Identification of DEGs and CARGs in breast cancer. **(A)** A volcano plot from TCGA-BRCA data illustrates the DEG analysis, with 1060 genes upregulated and 1188 genes downregulated. **(B)** Intersection of DEGs and CARGs, 267 intersecting genes were obtained. **(C)** Differential gene GO enrichment analysis. **(D)** Differential gene KEGG enrichment analysis. **(E)** Forest plot of univariate regression analyses in breast cancer patients, with risk factors in red and protective factors in green. **(F)** GO enrichment analysis of the 39 prognostic genes. **(G)** Expression box plot of the 39 genes in the tumor group versus the control group. Red represented tumor tissue and green represented normal tissue. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$

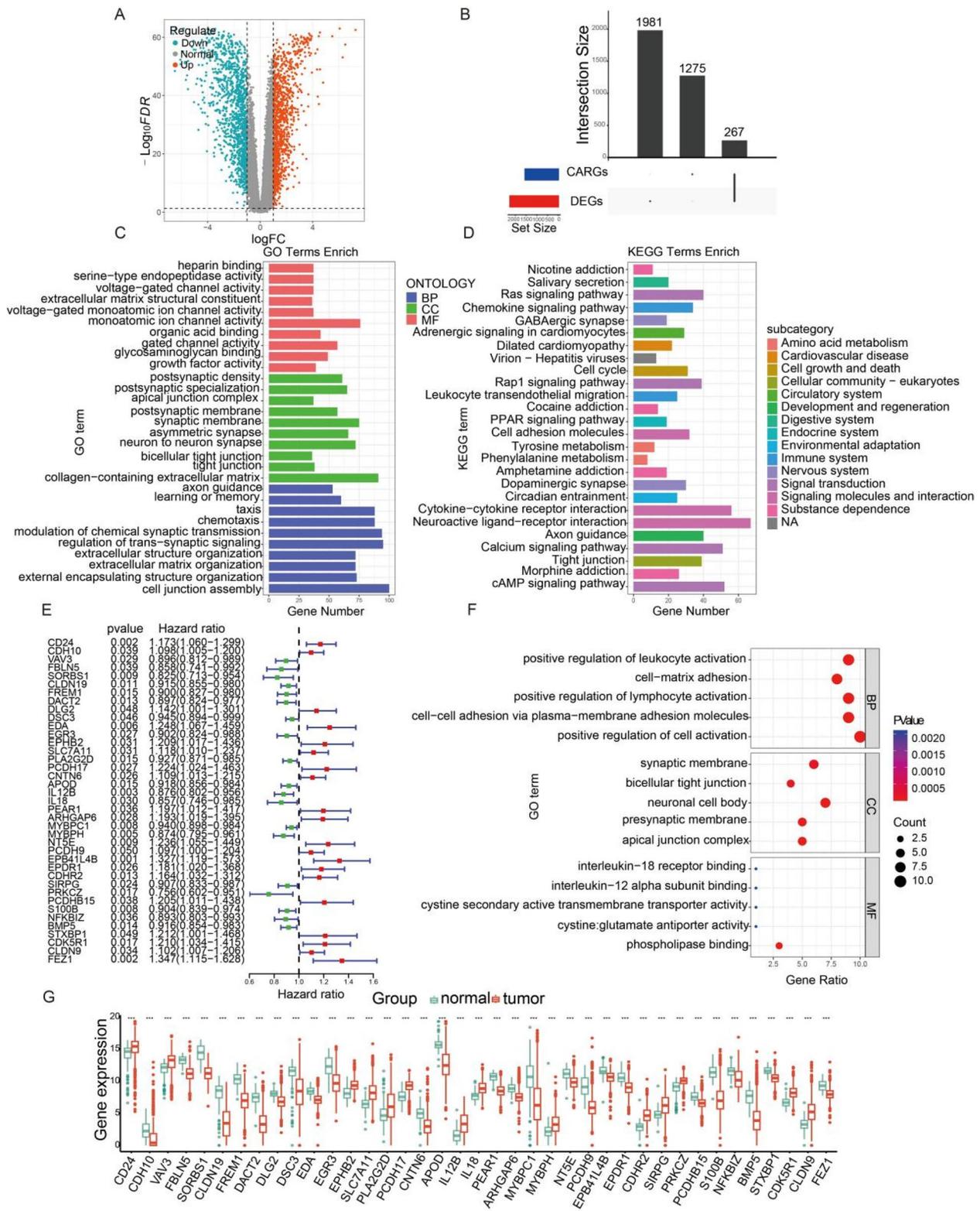


Fig. 2 (See legend on previous page.)

with CARGs and obtained 267 DECARGs (Fig. 2B). GO and KEGG enrichment analysis were conducted on these DEGs (Fig. 2C and D). Biological processes (BP) revealed by GO enrichment analyses included ion channel-related signaling pathways, such as monoatomic ion channel activity, gated channel activity, and glycosaminoglycan binding. KEGG enrichment highlighted intercellular communication pathways, including Neuroactive ligand-receptor interaction and Cytokine-cytokine receptor interactions. We further performed univariate analysis on DECARGs and screened them according to P -value < 0.05, and obtained 39 prognostic genes (Fig. 2E). These 39 prognostic genes were subjected to GO enrichment analysis, focusing on biological processes such as positive regulation of leukocyte, cell-cell and cell-matrix adhesion, and lymphocyte activation (Fig. 2F). This analysis indicated their pivotal roles in cell adhesion and immune responses. In addition, we analyzed the correlation between CARGs and their chromosomal locations, with the results presented in (Supplementary Figure S1). Further, CD24, VAV3, EPHB2, SLC7A11, PLA2G2D, PCDH17, IL12B, IL18, MYBPH, CDHR2, SIRPG, PRKCZ, CDK5R1, and CLDN9 exhibited significantly heightened expression levels in tumor tissues compared to normal tissues (Fig. 2G).

Identification of cell adhesion molecule subtypes

We performed NMF clustering utilizing the expression matrix of 39 prognostic genes (Fig. 3A), and further refined the clustering into three groups after selecting the optimal number of groups 3 (Fig. 3B). Among them, Group 1 encompassed 470 samples, Group 2 included 397 samples, and Group 3 consisted of 237 samples. Optimal clustering indicated reliable and stable differentiation into group 1, group 2, and group 3 (Fig. 3C). Survival analysis showed that group 1 had a diminished survival rate (Fig. 3D). We simultaneously validated the tumor samples in the validation set GSE20685, which were successfully clustered into three classes, and similarly, group 1 had a lower survival rate (Fig. 3E). To further explore the inherent mechanisms of cell adhesion-related molecular subtypes, the different subtypes underwent enrichment analysis (Fig. 3F-H). Inter-cellular communication, such as Cytokine-cytokine receptor interaction processes, were significantly enriched in the three subtypes. Finally, Supplementary Figure S2 demonstrated that the expression levels of the majority of CARGs displayed notable variations among various subtypes.

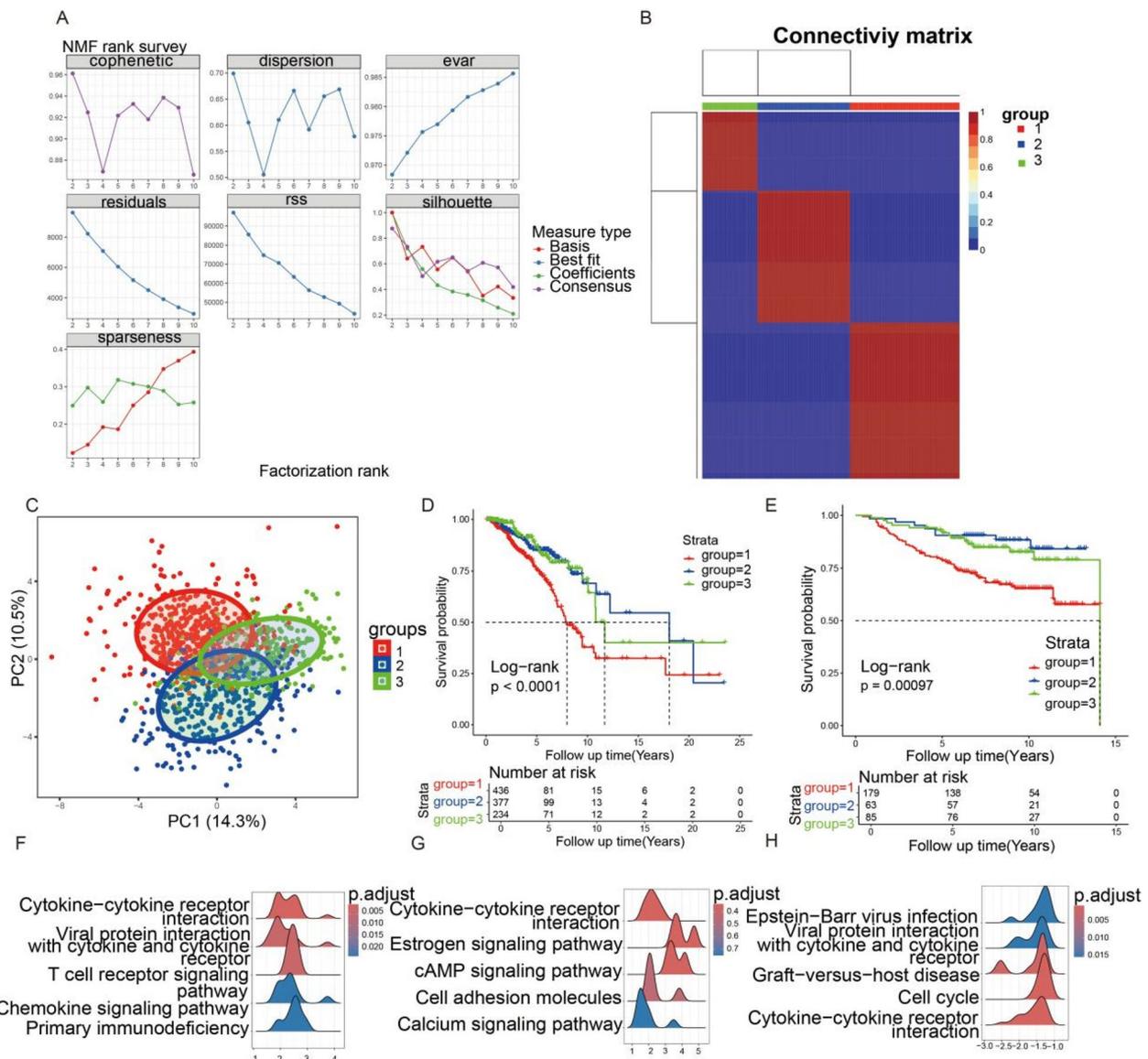
Immunoassay and clinical characterization of subtypes

We analyzed the distribution of 23 immune cell types across various subtypes by ssGSEA, and found that the immune infiltration thermograms highlighted varying

degrees of immune infiltration differences among the three subtypes (Fig. 4A). Subsequently, we evaluated the subtypes grouped based on CARGs using the ESTIMATE algorithm and found that they exhibited significant differences in terms of stromal cell composition, immune cell presence, and tumor purity. The findings indicated that group 2 exhibited the highest ESTIMATE score and immune score (Fig. 4B), while group 1 had higher tumor purity than the other two groups. Group 2 exhibited a markedly higher level of immune infiltration than group 1 and 3 (Fig. 4C). Except for ADORA, NRP1, and VTCN1, the distribution of immune checkpoint factors was the highest in group 2 (Fig. 4D), suggesting that group 2 might have a stronger correlation with immune infiltration. The heatmap and baseline table provide an overview of the clinical features associated with the various subtypes (Fig. 4E and F).

Identification of independent prognostic factors and construction of corresponding prognostic models

We derived 10 prognosis-related candidate genes through univariate Cox analysis, and the LASSO cross-validation process generated 10 genes (Fig. 5A and B). Subsequently, multivariate Cox analysis of these 10 characterized genes resulted in 8 cell adhesion genes, including CD24, SORBS1, EDA, IL12B, MYBPC1, NT5E, EPB41L4B, and FEZ1 (Fig. 5C). ROC curves were generated for 1-, 3-, and 5-year survival predictions, yielding AUC values of 0.638, 0.73, and 0.695, respectively (Fig. 5D). The K-M curves indicated a statistically significant extension in survival time for patients in the low-risk group (Fig. 5E). Survival status plots revealed a notably lower survival rate among individuals in the high-risk group (Fig. 5F). We observed consistent results in the GSE20685 dataset (Fig. 5G-I) and GSE42568 (Fig. 5J-L), further illustrating the robustness and reliability of the 8-gene risk model. Validation of the model using the GSE20685 validation set involved plotting ROC curves and calculating AUC values for 1-, 3-, and 5-year predictions, resulting in AUC values of 0.656, 0.720, and 0.709, respectively (Fig. 5G). Similarly, validation with the GSE42568 validation set involved plotting ROC curves and computing AUC values for the same time points, yielding AUC values of 0.747, 0.707, and 0.711, respectively (Fig. 5J). We identified CD24, EPB41L4B, and NT5E as high-risk genes, while IL12B and SORBS1 were categorized as low-risk genes (Fig. 5M). The other genes were shown in (Supplementary Figure S3). Additionally, we examined the expression patterns of the signature genes in both tumor and normal tissues, as well as in the high-risk versus low-risk groups, as presented in (Supplementary Figure S4).



Enrichment analysis across risk groups

To gain insights into the transcriptional variations between high- and low-risk groups, we conducted GSEA analysis. Our findings revealed the high-risk group showed biological processes including nucleotide-glucose metabolism, nucleotide transport, and response to acetylcholine (Fig. 6A). The low-risk group exhibited biological processes including activation of the immune response, B-cells in the immune response and the

response to antibiotics (Fig. 6B). This results suggested that in the low-risk group, the body's immune system is in a state of relative activity and can be effectively activated in response to tumorigenesis. Meanwhile, in the GO analysis, the enriched biological processes included positive regulation of cell adhesion, positive regulation of leukocyte activation, and acute inflammatory response to antigenic stimuli (Fig. 6C). All of these biological processes were closely related to cell adhesion and immune

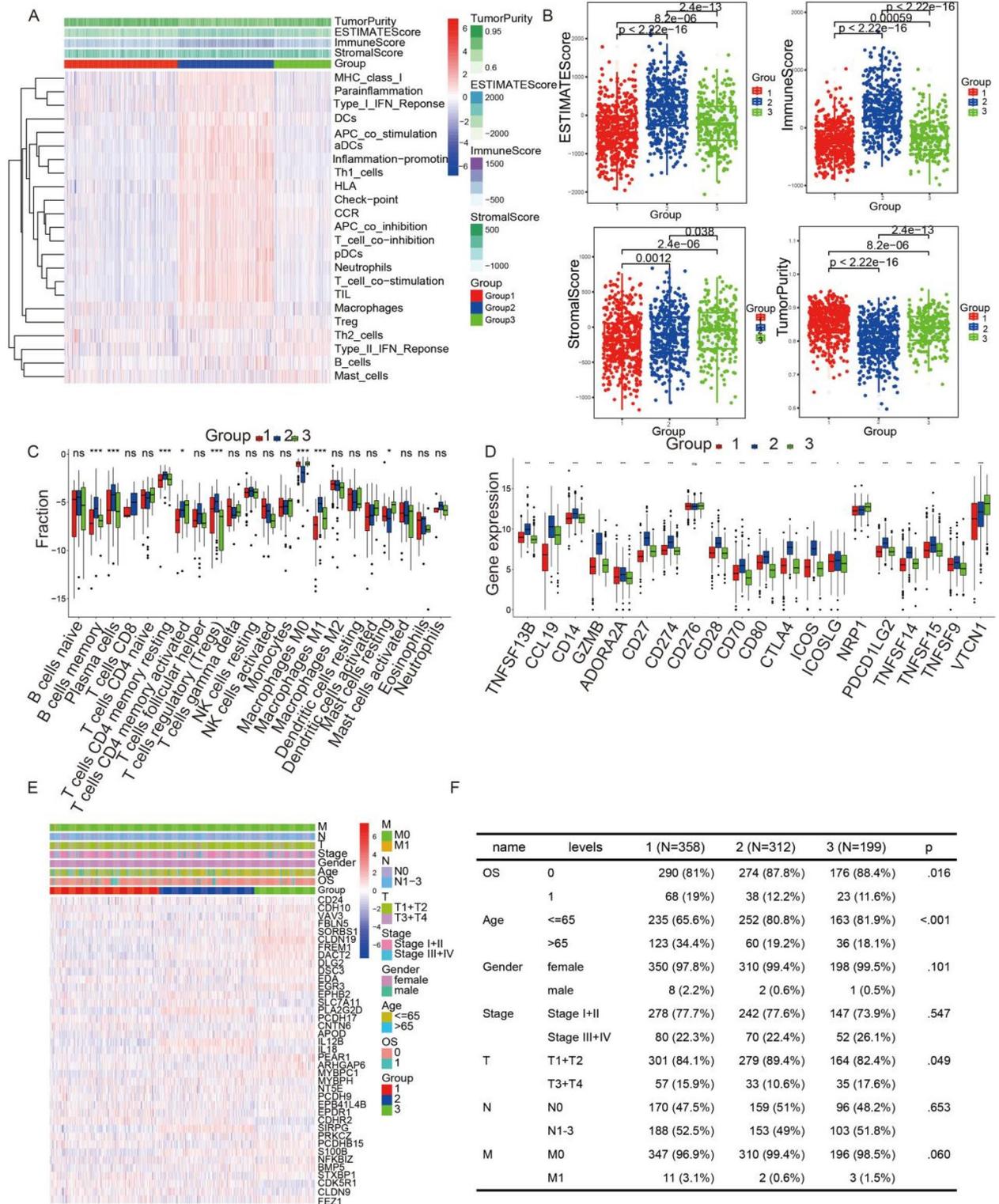


Fig. 4 Immune landscape of the different molecular subtypes. **A** The ssGSEA method for heat mapping of immune cell distribution. The samples were displayed along the horizontal axis, with different immune cell categories denoted on the vertical axis. **B** ESTIMATE score, immune score, stroma score, and tumor purity violin plots across the subtypes. **C** CIBERSORT box plot analysis of immunological differences among 22 immune cell types in three subtypes. **D** Box plot of the differences in the expression levels of 20 immune checkpoint factors among the three subtypes. **E** Heatmap of clinical characterization between subtypes. **F** Baseline table of clinical characterization between subtypes

response processes. The KEGG enrichment analysis, on the other hand, revealed pathways such as cytokine-cytokine receptor interactions, neuroactive ligand-receptor interactions, and cell adhesion molecules (Fig. 6D). These pathways focus on cytokine receptor interactions and immune system regulatory processes.

Construction of a nomogram for independent prediction of prognosis

To determine if the 8-cell adhesion-related gene risk signature was an independent prognostic factors for breast cancer, we conducted both univariate and multivariate Cox regression analyses. The univariate Cox analysis results revealed that the risk score was significantly associated with a hazard ratio (HR) of 1.680 (95% CI: 1.492-1.892; $P < 0.001$), as illustrated in Fig. 7A. In the multivariate analysis, the HR for risk score was 1.609 (95% CI: 1.416-1.829; $P < 0.001$) (Fig. 7B), confirming its independent association with overall survival. We next utilized risk scores, age, and N and M stage to create nomogram predicting the 1-, 3-, and 5-year survival rates of breast cancer patients (Fig. 7C). The Decision Curve Analysis (DCA) demonstrated a robust predictive probability for our nomogram (Fig. 7D). The calibration curves indicated a good agreement between the predicted and actual survival rates for 1-, 3-, and 5-year survival, highlighting the nomogram's reliable predictive accuracy (Fig. 7E). In summary, these findings suggested that the risk profile of these 8-cell adhesion-related genes served as independent prognostic factor for breast cancer patients.

Survival analysis stratified by clinical characteristics and risk models

We conducted an in-depth examination of survival disparities between high- and low-risk patient cohorts across diverse clinical stages, finding these differences to be statistically significant (Fig. 8A). Within patient groups aged 65 years and younger, and across stages N0, N1-3, M0, I-II, and III-IV, the high-risk groups consistently demonstrated lower survival rates compared to low-risk groups (Fig. 8B). Additionally, we assessed the risk scores

based on gender, T1 + T2 and T3 + T4 stages. The data revealed that male had higher risk scores than female, T3 + T4 stages had higher risk score than T1 + T2 stages (Supplementary Figure S5). These findings underscored the robust predictive capability of our model in discerning survival rates among patients with varying clinical characteristics.

Immune infiltration and immunotherapy response analysis based on prognostic characteristics

To further investigate the relationship between infiltrating immune cells and risk models, we utilized the ssGSEA to assess variations in immune functionalities and cellular infiltration between high- and low-risk patient groups. We observed significant differences in tumor immune infiltration between the two risk groups, with the high-risk group characterized by an elevated abundance of Macrophages and a low abundance of aDCs, B cells, DCs, Neutrophils, PDCs, Th1 cells, and TIL cells (Fig. 9A), and the low-risk group exhibited enhanced immune function (Fig. 9B). ESTIMATE algorithm revealed that the low-risk group demonstrated a higher ESTIMATE score and immune score, while tumor purity was significantly lower (Fig. 9C). Furthermore, analysis using the EPIC method revealed notable differences in CD4 T cells, B cells, and other immune cells (Fig. 9D). To further explore immune activity, variations in immune cell populations and checkpoint expression were examined. The low-risk group generally exhibited higher expression levels of immune checkpoint factors (Fig. 9E). Immune infiltration and IPS scores were used to assess the prognosis of breast cancer. The results indicated that IPS scores were significantly higher in the low-risk group, indicating a better response to immunotherapy (Fig. 9F).

PD-L1 blockade immunotherapy is currently one of the most important therapeutic approaches in the field of tumor immunotherapy. Subsequently, we evaluated the IMvigor210 cohort, which revealed four varied responses to anti-PD-L1 receptor antagonists: stable disease (SD), partial response (PR), complete response (CR), and disease progressive disease (PD). Patients in the low-risk group exhibited a superior survival rate (Fig. 9G). We

(See figure on next page.)

Fig. 5 Construction and validation processes of the risk models. **A** Depicted the coefficient distributions derived from $\log(\lambda)$ sequences in the LASSO models. **B** LASSO coefficient spectrum for LASSO cox analysis. **C** Forest plot of multivariate Cox regression analysis. HR is the risk ratio. **D** Training centralized models to predict 1-year, 3-year, and 5-year AUC curves for risk models. All AUC > 0.6. **E** Compared the survival curves for overall survival between the high-risk and low-risk groups within the training set. **F** Plot of the distribution of survival status and risk group scores in the training set. **G** GSE20685 validates the 1-, 3-, and 5-year AUC curves predicted by the centralized model. All AUC > 0.6. (H) GSE20685 validated survival curves for overall survival. **I** Distribution of survival status and distribution of risk group scores in the GSE20685 validation set. (J) GSE42568 validates the 1-, 3-, and 5-year AUC curves predicted by the model. The AUCs are all > 0.7. **K** Survival curves for overall survival in GSE42568 validation, (L) Distribution of GSE42568 validation centralized survival status and distribution of risk group scores. (M) K-M curves of characterized genes CD24, EPB41L4B, IL12B, NT5E, and SORBS1 ($P < 0.05$)

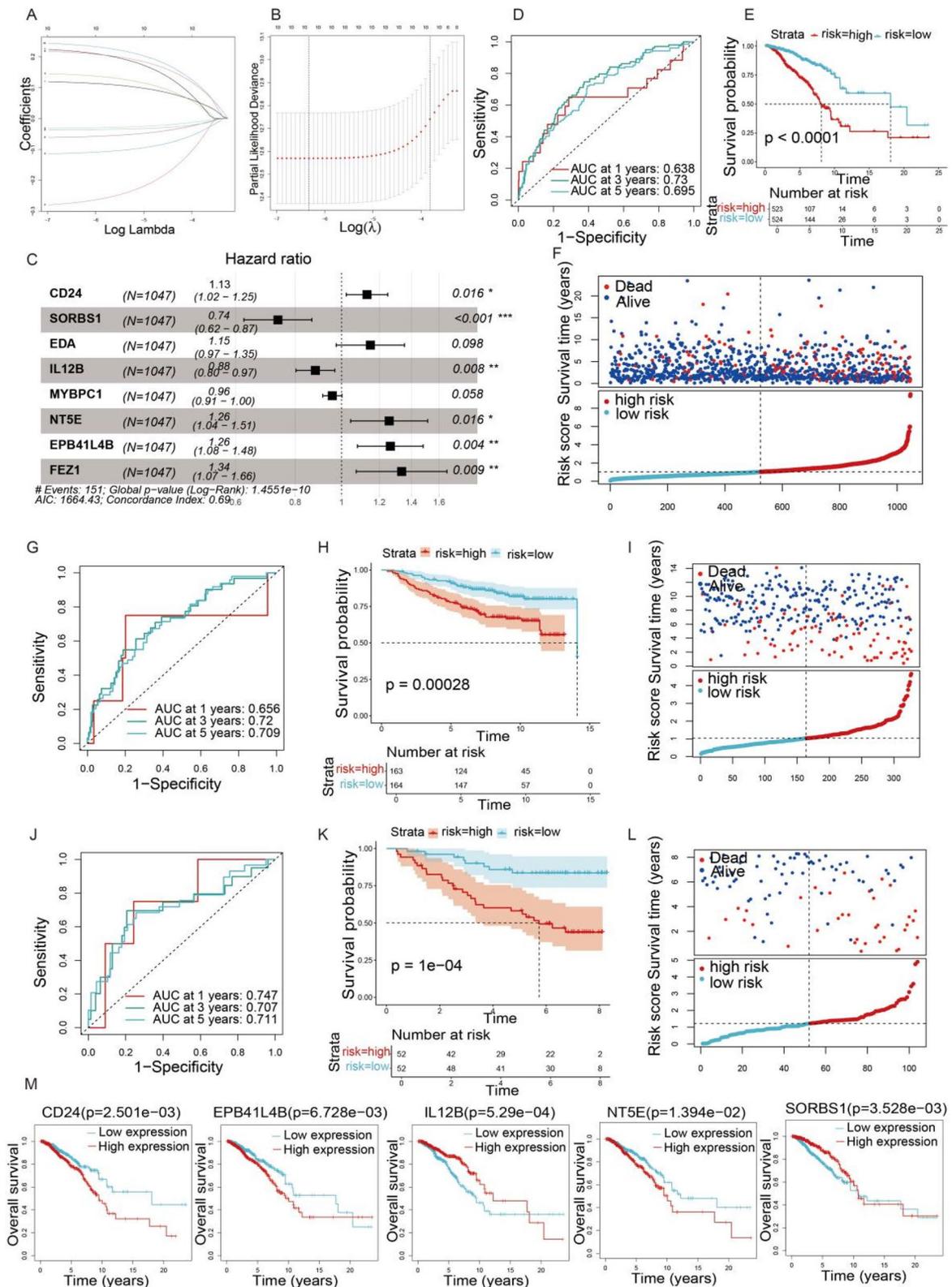
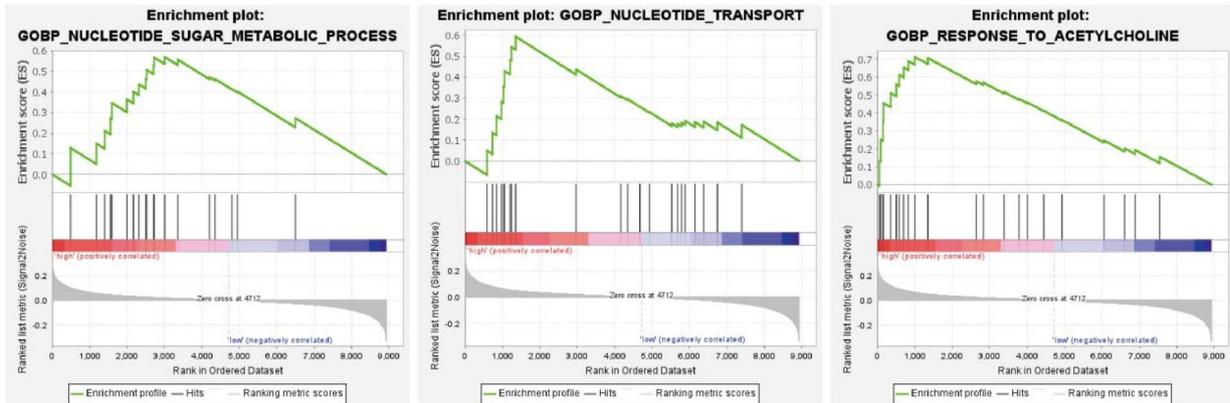
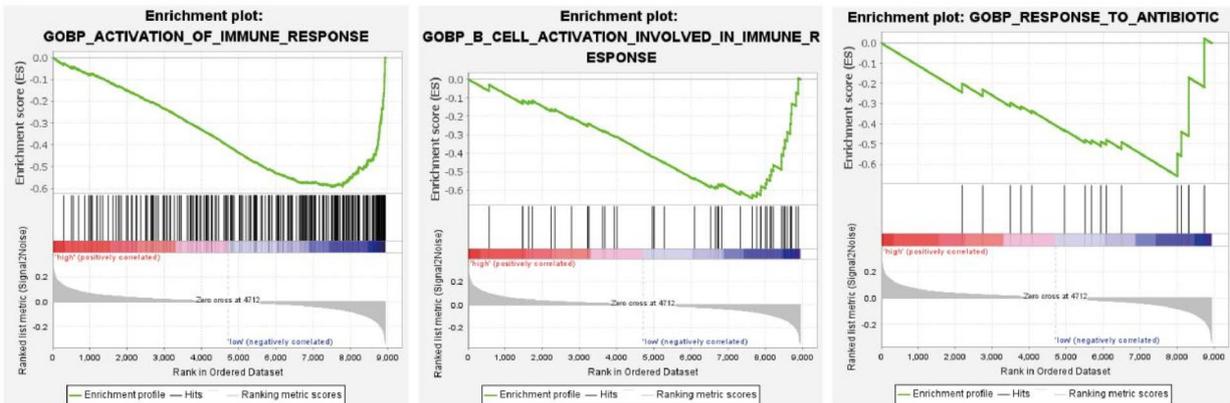


Fig. 5 (See legend on previous page.)

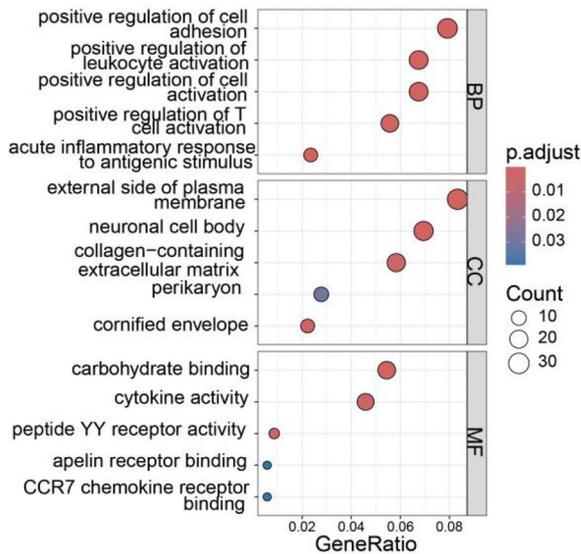
A



B



C



D

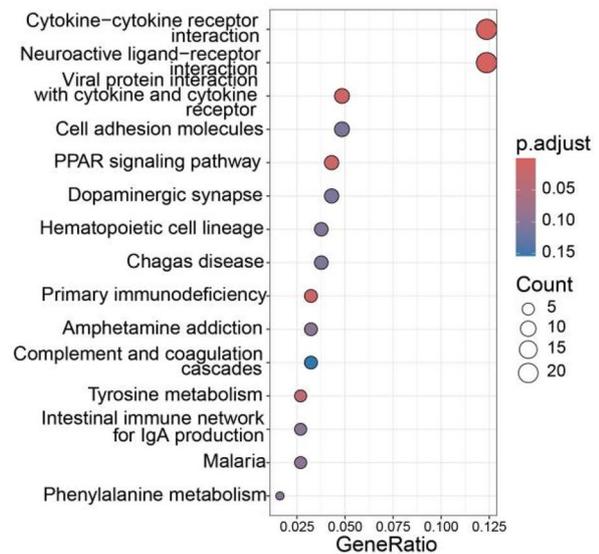


Fig. 6 Functional enrichment of GSEA in high- and low-risk populations. **A** GSEA results for high-risk groups. **B** GSEA results for low-risk groups. The horizontal axis in both panels represented the ranking within the ordered dataset, and the vertical axis signified the enrichment score alongside the ranking metric score. **C** Differential gene GO analysis. The horizontal axis denoted gene proportions and the vertical axis illustrated various gene ontology terms. **D** KEGG analysis of DEGs between the high- and low-risk groups. The horizontal axis was gene proportions and the vertical axis was functional terms. Point size was proportional to the number of genes

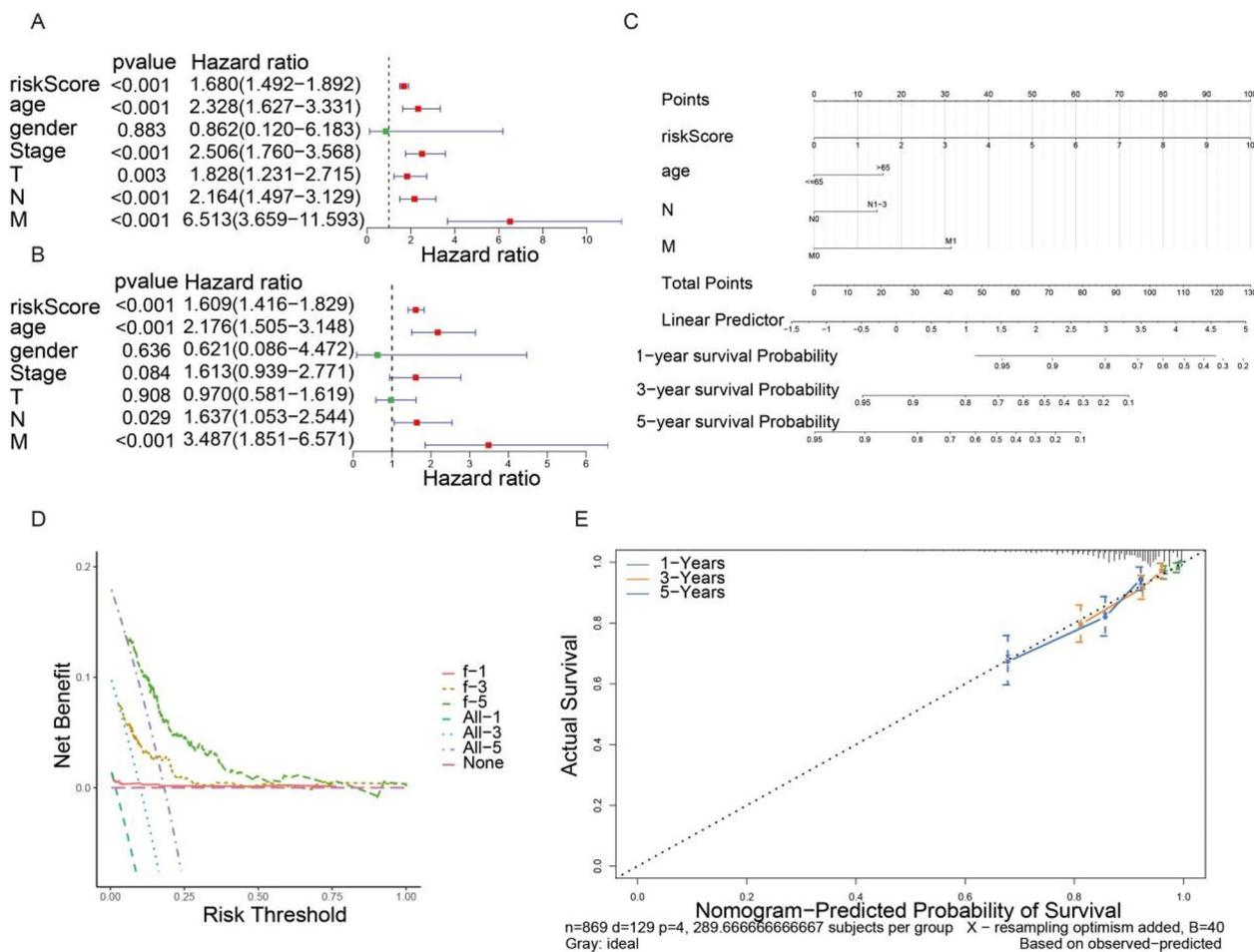


Fig. 7 Analysis of independent prognostic factors. **A** The forest plot illustrated the univariate Cox regression analysis of risk score and clinical factors. **B** The forest plot illustrated the multivariate Cox regression analysis of risk score and clinical factors. **C** The nomogram based on risk score, clinical variables and survival rates at 1, 3, and 5 years. These clinical factors include age, N and T stages. **D** DCA curve for 1-, 3-, and 5-year risk prediction. **E** 1-, 3-, and 5-year risk prediction calibration curves. A line that closely aligns with the ideal dashed line indicated a more reliable result

categorized SD and PD as Not Responders (NR) and CR and PR as Responders (R), the low-risk group demonstrated a significantly higher response rate compared to the high-risk group (Fig. 9H). In contrast, patients with immune response (R) had substantially lower risk scores than those without a response (NR) (Fig. 9I).

Assessment of tumor mutation burden (TMB) and drug responsiveness

To explore the association between gene mutations and risk scores in tumor cells, a TMB analysis was conducted. The results exhibited higher levels of TP53 mutations (24% versus 16%) in the high-risk group (Fig. 10A-B). Furthermore, a statistically significant distinction in TMB was observed between these two groups (Fig. 10C). A summary plot of the overall mutations in both high- and low-risk groups is provided in (Supplementary Figure S6). To analyze drug sensitivity, we screened potentially

effective anti-breast cancer drugs using the “pRRobhetic” package, and we screened four targeted drugs which breast cancer patients showed greater sensitive to treatment: 5-Fluorouracil, Camptothecin, Doxorubicin, and Docetaxel. The low-risk group demonstrated higher drug sensitivity to 5-Fluorouracil, Camptothecin, and Doxorubicin, while the high-risk group presented greater responsiveness to Docetaxel (Fig. 10D). Subsequently, Spearman correlation analysis was conducted to assess the impact of model gene on drug sensitivity using data from the CellMiner database. The results suggested that EPB41L4B exhibited a negative correlation with tepotinib (Cor = -0.498, P<0.001); NT5E was demonstrated a negative association with AFP464 (Cor = -0.541, P<0.001). And CD24 was positively associated with Sapitinib (Cor = 0.473, P<0.001); SORBS1 showed a positive association with PLX-4720 (Cor = 0.430, P<0.001) (Fig. 10E).

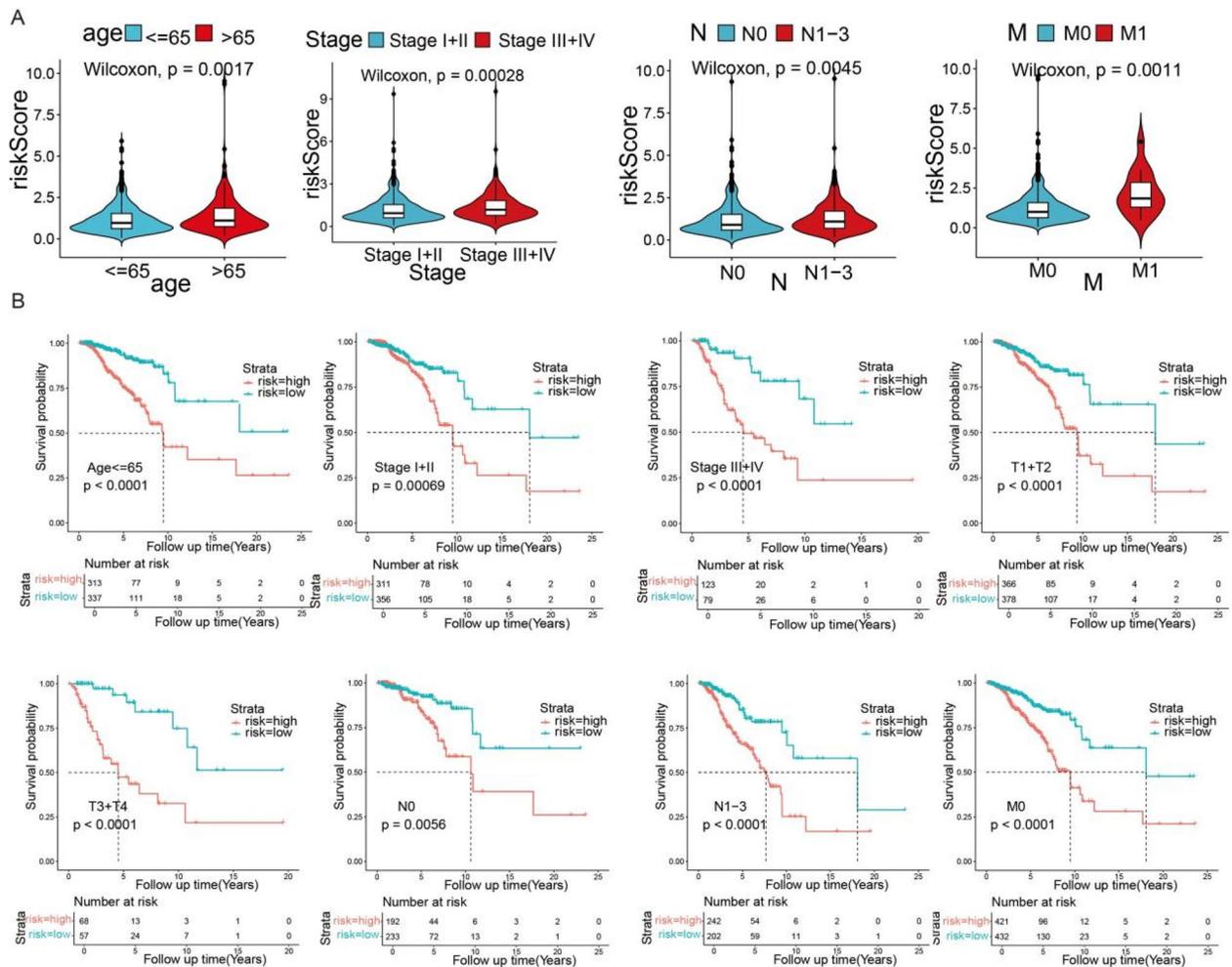


Fig. 8 Survival analysis of risk models based on clinical trait. **A** Violin plots of risk scores versus clinical information for age, stage, N, and M stages. **B** Kaplan-Meier curves demonstrated variations in survival between high-risk and low-risk groups within the patients of age ≤ 65, stages I-II, III+IV, T-stages (T1+T2 versus T3+T4), and N-stages (N0 versus N1-3), M0 stage

Discussion

Breast cancer is a prevalent malignant tumor with strong invasiveness, rapid progression and poor prognosis, etc. In its early stages, the disease often lack distinct symptoms, leading to many patients being diagnosed at intermediate to advanced stages [12]. Consequently, early and accurate diagnosis of breast cancer is of paramount importance. Cell adhesion factors can mediate cell-cell

and cell-extracellular matrix interactions and participate in the apoptosis and invasive migration process of tumor cells [13, 14]. Therefore, further exploration of cell adhesion-related gene in the context of breast cancer is warranted [15].

In our research, we identified 8 prognostic genes using univariate, LASSO, and multivariate regression analysis to establish a predictive model,

(See figure on next page.)

Fig. 9 Assessment of immune infiltration and immune response in high- and low-risk groups. **A** Box plots illustrated ssGSEA immune cell scores. **B** Box plots illustrated ssGSEA immunocompetence scores. **C** ESTIMATE score, Immune score, stroma score, and tumor purity violin plots. **D** Histogram of immune infiltration and percentage heatmap for EPIC high- and low-risk groups. **E** Expression levels of immune checkpoints. **F** IPS scores for high- and low-risk groups, with the high-risk group denoted in red and the low-risk group in blue. **G** Survival curve analysis of the IMvigor210 cohort, $n = 348$. **H** The percentage of alive and dead samples in the two risk groups. Blue represented NR and red represented R. Vertical coordinates represented response rates. **I** Distribution of risk scores in the immune response R and NR groups

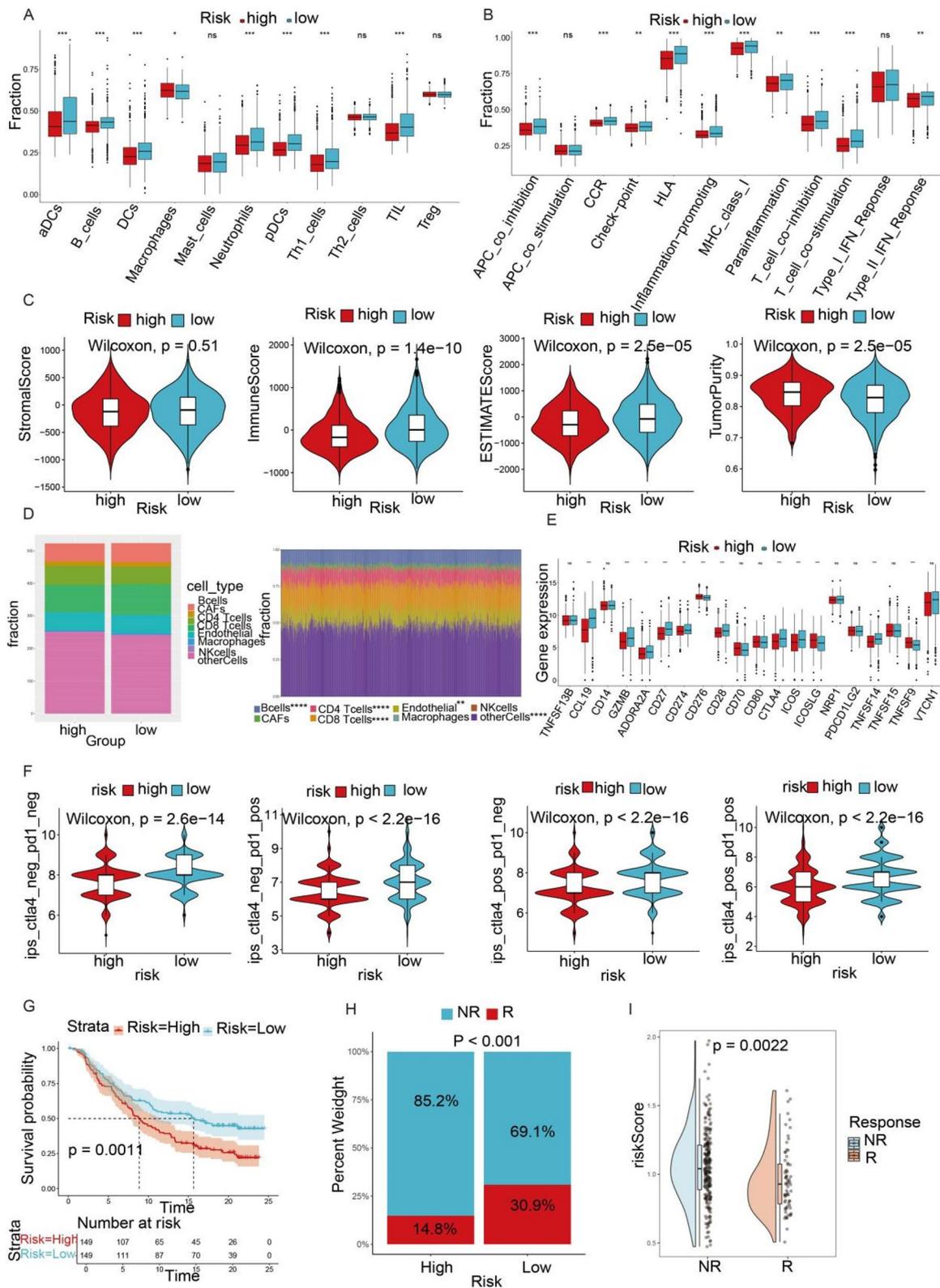


Fig. 9 (See legend on previous page.)

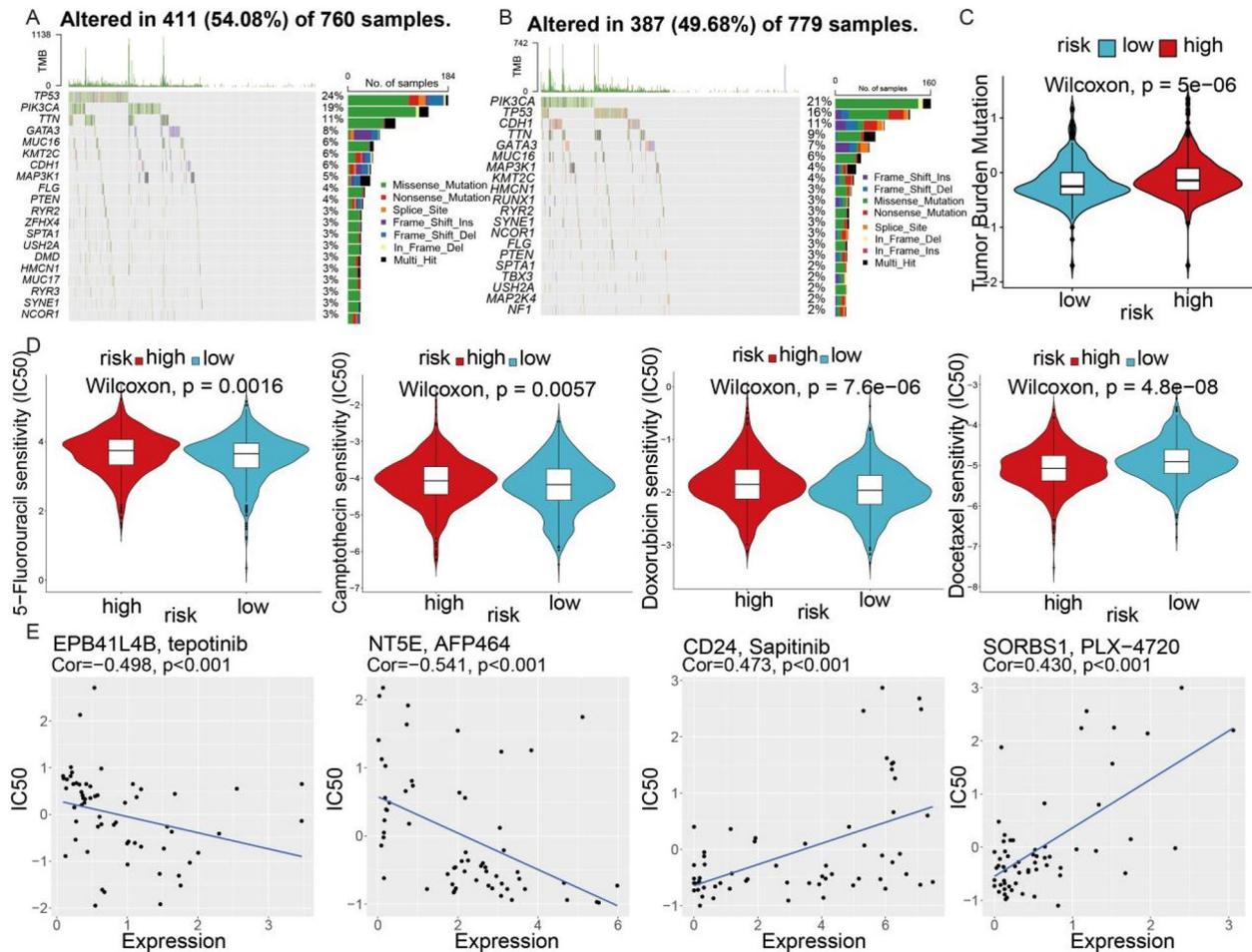


Fig. 10 Assessment of tumor mutation burden (TMB) and drug responsiveness. **A** Waterfall plot illustrating the 20 most frequently mutated genes in breast cancer within a high-risk cohort consisting of 760 samples. **B** Waterfall plot illustrating the top 20 mutated genes in breast cancer for the low-risk cohort consisting of 779 samples. In A and B, the green bar indicated a missense mutation, the purple bar signified a frame shift insertion, the blue bar denoted a frame shift deletion, the yellow bar reflected an in-frame deletion, the orange bar pointed to a splice site, and the black bar represented a multi-hit event. **C** Comparison of TMB between high-risk and low-risk cohorts of breast cancer. **D** IC50 values for 5-Fluorouracil, Camptothecin, Doxorubicin, and Docetaxel in breast cancer high- and low-risk cohorts. In C and D, with the high-risk cohort denoted by red and the low-risk cohort by blue. **E** Relationship between model gene expression levels and drug sensitivity, based on predictions from the CellMiner database. EPB41L4B was negatively correlated with tepotinib (Cor = -0.498), NT5E was negatively correlated with AFP464 (Cor = -0.541). CD24 was positively correlated with Sapitinib (Cor = 0.473), SORBS1 positively correlated with PLX-4720 (Cor = 0.430)

including CD24, SORBS1, EDA, IL12B, MYBPC1, NT5E, EPB41L4B, and FEZ1. The model's predictions of patient survival aligned well with actual survival rates, leading us to conclude that these genes could serve as independent prognostic indicators. Kaarvatn et al. highlighted the association between IL12B gene polymorphisms and breast cancer development [16]. Feng et al. emphasized the role of SORBS1 in inhibiting breast cancer cell invasion and migration through by modulating the PI3K/AKT pathway, promoting M1 macrophage polarization, and inhibiting EMT, suggesting its potential as an anti-metastatic agent [17]. MiR-142-5p, on the other hand, stimulated breast cancer

proliferation, invasion, and migration by targeting SORBS1 [18]. SORBS1 inhibited p53 in breast cancer cells to attenuate sensitivity to cisplatin drugs by inhibiting p53 in breast cancer cells, suggesting that it might serve as a potential inhibitor of cancer metastasis [19]. CD24 had been extensively studied in breast cancer [20]. Qu et al. highlighted ELF5 enhanced macrophage phagocytosis by blocking CD24 treatment and reduced tumor growth in vivo [21]. In addition, CD24 could act as a major innate immune checkpoint in ovarian and breast cancers, and CD24 deficiency mediated breast cancer cell dedifferentiation and chemoresistance [22, 23]. Yin et al. pointed that overexpressing Ehm2/1 in

MCF-7 breast cancer cells hindered cell migration and invasion while enhancing E-calcineurin stability [24]. It was reported high EPB41L4B expression in prostate cancer cells but downregulation in gastric cancer [25, 26]. Previous studies had showed that NT5E sustains cancer stem cell traits by stabilizing SOX9 expression in hepatocellular carcinoma [27]. NT5E KO downregulates genes involved in the cellular stress response, and increases TMB delaying breast tumorigenesis [28]. EDA enhances auto-fibroblast differentiation and promotes breast cancer cell proliferation [29] and promotes an inflammatory environment [30]. MYBPC1 was reported to possibly conduct progesterone in normal and malignant breast tissues, inducing proliferation of mammary epithelial cells [31]. However, studies on this gene remained limited. Previous studies [32] have revealed that FEZ1 acted as a suppressor of cancer cell growth by regulating mitotic process. It was deficiently expressed in cancers including prostate, lung [33], bladder [34] and breast [35]. These genes hold promise as potential biomarkers for predicting prognosis and immunotherapy efficacy in breast cancer patients.

Analysis of immune cell infiltration in breast cancer correlated tumor immunotherapy with immune infiltration, which was essential for achieving personalized treatment. We observed macrophages were increased and all other immune factors were suppressed in the high-risk group. Breast cancer tumor cells can recruit macrophages to form tumor-associated macrophages (TAM), which bind to PD-1, a receptor on the surface of T cells, inhibiting their activation and proliferation and interfering with the immune activation pathway, which prevented other immune cells from performing their immune functions and contributed to immune escape [36]. Tumour infiltrating lymphocytes (TIL) mainly contained immune cells capable of recognizing and attacking tumour cells. We observed a significant decrease in TIL within the high-risk patient cohort, which might have consequently led to an augmented proportion of immunosuppressive cells, such as regulatory T cells, allowing for an increased risk of tumour cell escape [37]. Subsequent analysis of immune-related functions in breast cancer patients revealed a suppression of type II interferon (IFN) responses among high-risk individuals. IFN was capable of disrupting viral replication *in vitro*, and type II IFN constituted a fundamental element of antiviral immunity [38]. II-IFN activation was crucial for the maintenance of immune efficacy, and the suppression of its response might be the major causes of immune escape. However, the intricate relationship between cell adhesion-related genes and immune-related functions remained an area of incomplete understanding. Future endeavors will focus on elucidating the prognostic implications for breast

cancer patients exhibiting immune cell infiltration, as well as exploring the potential contributions of immune cells in targeted therapeutic strategies for this patient subset.

In summary, our findings indicated that CARGs could be utilized as biomarkers for assessing breast cancer prognosis. Despite the prognostic significance of this feature, this study does possess certain limitations. Most of our analyses and conclusions are based on data from public databases, additional clinical cohorts are still needed to support the credibility reliability of our model. Notably, independent clinical cohorts from diverse populations and geographic regions will help ensure that our findings have broad applicability and generalizability. Furthermore, by integrating multi-center and diverse clinical data, we can better assess the performance and robustness of our model in different clinical settings, thereby providing stronger evidence for clinical practice. Further *in vitro* and *in vivo* studies are also important. Moving forward, we aim to continue to delve deeper into the potential mechanisms of the breast cancer signature genes to advance cancer treatment strategies.

Conclusion

In this study, we employed bioinformatics techniques to identify three subtypes associated with cell adhesion and constructed a prognostic risk score model based on CARGs. Based on the risk score, the characteristics of breast cancer patients were evaluated in terms of immunotherapy, TMB and drug sensitivity. The findings indicates that this feature can predict prognosis and therapeutic outcome of immunotherapy in breast cancer patients, presenting a fresh perspective for personalized and tailored treatment approaches.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12957-025-03802-5>.

Supplementary Material 1.
Supplementary Material 2.
Supplementary Material 3.
Supplementary Material 4.
Supplementary Material 5.
Supplementary Material 6.
Supplementary Material 7.
Supplementary Material 8.

Acknowledgments

Not applicable.

Authors' contributions

LDM and CSF contributed to the study design. YL conducted the literature search. YL, FC and FLH acquired the data. CSF performed data analysis. LDM

and CSF drafted. LDM was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Clinical trial number

Not applicable.

Funding

Not applicable.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Breast Thyroid Surgery, Jinhua Maternal And Child Health Care Hospital, Jinhua 321000, China.

Received: 18 February 2025 Accepted: 8 April 2025

Published online: 22 April 2025

References

- Akram M, et al. Awareness and current knowledge of breast cancer. *Biol Res.* 2017;50(1):33.
- Bray F, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2024;74(3):229–63.
- Redig AJ, McAllister SS. Breast cancer as a systemic disease: a view of metastasis. *J Intern Med.* 2013;274(2):113–26.
- Waks AG, Winer EP. Breast Cancer Treatment: A Review. *Jama.* 2019;321(3):288–300.
- Li DM, Feng YM. Signaling mechanism of cell adhesion molecules in breast cancer metastasis: potential therapeutic targets. *Breast Cancer Res Treat.* 2011;128(1):7–21.
- Guerra-Espinosa C, et al. ICAMs in Immunity, Intercellular Adhesion and Communication. *Cells.* 2024;13(4):339.
- Wang J, Springer TA. Structural specializations of immunoglobulin superfamily members for adhesion to integrins and viruses. *Immunol Rev.* 1998;163:197–215.
- Guan X. Cancer metastases: challenges and opportunities. *Acta Pharm Sin B.* 2015;5(5):402–18.
- Calaf GM, et al. Cell Adhesion Molecules Affected by Ionizing Radiation and Estrogen in an Experimental Breast Cancer Model. *Int J Mol Sci.* 2022;23(20):12674.
- Clarke C, et al. Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis. *Carcinogenesis.* 2013;34(10):2300–8.
- Kao KJ, et al. Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization. *BMC Cancer.* 2011;11:143.
- Zhang YN, et al. Review of Breast Cancer Pathological Image Processing. *Biomed Res Int.* 2021;2021:1994764.
- Sivasankar S, Xie B. Engineering the Interactions of Classical Cadherin Cell-Cell Adhesion Proteins. *J Immunol.* 2023;211(3):343–9.
- Läubli H, Borsig L. Altered Cell Adhesion and Glycosylation Promote Cancer Immune Suppression and Metastasis. *Front Immunol.* 2019;10:2120.
- Sousa B, Pereira J, Paredes J. The Crosstalk Between Cell Adhesion and Cancer Metabolism. *Int J Mol Sci.* 2019;20(8):1933.
- Kaarvatn MH, et al. Single nucleotide polymorphism in the interleukin 12B gene is associated with risk for breast cancer development. *Scand J Immunol.* 2012;76(3):329–35.
- Feng K, et al. SORBS1 inhibits epithelial to mesenchymal transition (EMT) of breast cancer cells by regulating PI3K/AKT signaling and macrophage phenotypic polarization. *Aging (Albany NY).* 2024;16(5):4789–810.
- Yu W, et al. MiR-142-5p Acts as a Significant Regulator Through Promoting Proliferation, Invasion, and Migration in Breast Cancer Modulated by Targeting SORBS1. *Technol Cancer Res Treat.* 2019;18:1533033819892264.
- Song L, et al. SORBS1 suppresses tumor metastasis and improves the sensitivity of cancer to chemotherapy drug. *Oncotarget.* 2017;8(6):9108–22.
- Yang P, et al. CD24 is a novel target of chimeric antigen receptor T cells for the treatment of triple negative breast cancer. *Cancer Immunol Immunother.* 2023;72(10):3191–202.
- Qu X, et al. ELF5 inhibits the proliferation and invasion of breast cancer cells by regulating CD24. *Mol Biol Rep.* 2021;48(6):5023–32.
- Bontemps I, et al. Loss of CD24 promotes radiation- and chemo-resistance by inducing stemness properties associated with a hybrid E/M state in breast cancer cells. *Oncol Rep.* 2023;49(1):4.
- Huth HW, et al. Translocation of intracellular CD24 constitutes a triggering event for drug resistance in breast cancer. *Sci Rep.* 2021;11(1):17077.
- Yin X, et al. Ehm2 transcript variant 1 inhibits breast cancer progression and increases E-cadherin stability. *Carcinogenesis.* 2022;43(12):1110–20.
- Schulz WA, et al. Changes in cortical cytoskeletal and extracellular matrix gene expression in prostate cancer are related to oncogenic ERG deregulation. *BMC Cancer.* 2010;10:505.
- Liu HQ, et al. Identifying specific miRNAs and associated mRNAs in CD44 and CD90 cancer stem cell subtypes in gastric cancer cell line SNU-5. *Int J Clin Exp Pathol.* 2020;13(6):1313–23.
- Ma XL, et al. CD73 sustained cancer-stem-cell traits by promoting SOX9 expression and stability in hepatocellular carcinoma. *J Hematol Oncol.* 2020;13(1):11.
- Samain R, et al. CD73 controls Myosin II-driven invasion, metastasis, and immunosuppression in amoeboid pancreatic cancer cells. *Sci Adv.* 2023;9(42):ead0244.
- Kwon A, et al. Extra domain A-containing fibronectin expression in Spin90-deficient fibroblasts mediates cancer-stroma interaction and promotes breast cancer progression. *J Cell Physiol.* 2020;235(5):4494–507.
- Tunali G, et al. A positive feedback loop driven by fibronectin and IL-1 β sustains the inflammatory microenvironment in breast cancer. *Breast Cancer Res.* 2023;25(1):27.
- Hu H, et al. RANKL expression in normal and malignant breast tissue responds to progesterone and is up-regulated during the luteal phase. *Breast Cancer Res Treat.* 2014;146(3):515–23.
- Ishii H, et al. FEZ1/LZTS1 gene at 8p22 suppresses cancer cell growth and regulates mitosis. *Proc Natl Acad Sci U S A.* 2001;98(18):10374–9.
- Nonaka D, et al. Reduced FEZ1/LZTS1 expression and outcome prediction in lung cancer. *Cancer Res.* 2005;65(4):1207–12.
- Vecchione A, et al. FEZ1/LZTS1 is down-regulated in high-grade bladder cancer, and its restoration suppresses tumorigenicity in transitional cell carcinoma cells. *Am J Pathol.* 2002;160(4):1345–52.
- Fang Y, et al. Mitochondrial-related genes as prognostic and metastatic markers in breast cancer: insights from comprehensive analysis and clinical models. *Front Immunol.* 2024;15:1461489.
- Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med.* 2013;19(11):1423–37.
- Zhao Y, et al. Tumor Infiltrating Lymphocyte (TIL) Therapy for Solid Tumor Treatment: Progressions and Challenges. *Cancers (Basel).* 2022;14(17):4160.
- Meyer ML, et al. New promises and challenges in the treatment of advanced non-small-cell lung cancer. *Lancet.* 2024;404(10454):803–22.

Publisher's Note

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