



# Estrogen receptor—positive breast cancer survival prediction and analysis of resistance-related genes introduction

Chen Shuai<sup>1</sup>, Fengyan Yuan<sup>2</sup>, Yu Liu<sup>3</sup>, Chengchen Wang<sup>3</sup>, Jiansong Wang<sup>3</sup> and Hongye He<sup>4</sup>

<sup>1</sup> Department of Breast and Thyroid Surgery, Yiyang Central Hospital, Yiyang, Hunan, China

<sup>2</sup> Hunan Normal University of Medicine, Changsha, Hunan, China

<sup>3</sup> Hunan Provincial People's Hospital, Changsha, Hunan, China

<sup>4</sup> Department of General Surgery, The Second Xiangya Hospital, Central South University, Changsha, Hunan, China

## ABSTRACT

**Background.** In recent years, ER+ and HER2- breast cancer of adjuvant therapy has made great progress, including chemotherapy and endocrine therapy. We found that the responsiveness of breast cancer treatment was related to the prognosis of patients. However, reliable prognostic signatures based on ER+ and HER2- breast cancer and drug resistance-related prognostic markers have not been well confirmed, This study in aimed to establish a drug resistance-related gene signature for risk stratification in ER+ and HER2- breast cancer.

**Methods.** We used the data from The Cancer Genoma Atlas (TCGA) breast cancer dataset and gene expression database (Gene Expression Omnibus, GEO), constructed a risk profile based on four drug resistance-related genes, and developed a nomogram to predict the survival of patients with I-III ER+ and HER2- breast cancer. At the same time, we analyzed the relationship between immune infiltration and the expression of these four genes or risk groups.

**Results.** Four drug resistance genes (AMIGO2, LGALS3BP, SCUBE2 and WLS) were found to be promising tools for ER+ and HER2- breast cancer risk stratification. Then, the nomogram, which combines genetic characteristics with known risk factors, produced better performance and net benefits in calibration and decision curve analysis. Similar results were validated in three separate GEO cohorts. All of these results showed that the model can be used as a prognostic classifier for clinical decision-making, individual prediction and treatment, as well as follow-up.

Submitted 25 February 2021

Accepted 2 September 2021

Published 26 October 2021

Corresponding author  
Hongye He, hhyhongye@csu.edu.cn,  
qingyunshuimo@sina.com

Academic editor  
Ulrich Pfeiffer

Additional Information and  
Declarations can be found on  
page 15

DOI 10.7717/peerj.12202

© Copyright  
2021 Shuai et al.

Distributed under  
Creative Commons CC-BY 4.0

OPEN ACCESS

**Subjects** Bioinformatics, Oncology, Women's Health

**Keywords** ER+ and HER2- breast cancer, Drug resistance-related gene, Risk signature

## INTRODUCTION

Breast cancer is a common diagnosed malignant tumor that lead to cancer-related mortality in women worldwide; it can mainly classified into invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC) by histopathological classification (*Dianatinasab et al., 2019; Li, Uribe & Daling, 2005; Viale, 2012*). The basic status of the estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2)

play a crucial role in molecular subtypes of breast cancer. Previous molecular studies demonstrated that breast cancer can be classified into Luminal A (ER+/PR+/HER2-/low Ki-67); Luminal B (ER+/PR+/HER2-/+/high Ki-67), HER2-overexpression (ER-/PR-/HER2+) and triple negative breast cancers (ER-/PR-/HER2-) (*Al-Thoubaity, 2020; Goldhirsch et al., 2011*). Different molecular subtypes exhibit different responsiveness to therapies. For instance, most patients of breast cancer are present as ER+ and HER2- and can benefit from endocrine therapy.

Although surgery-based comprehensive therapy can significantly improve the prognosis of breast cancer, the endocrine resistance may occur in approximately 33% of patients with endocrine therapy (*Achinger-Kawecka et al., 2020*). When endocrine-resistance appears, chemotherapy is needed to control the disease progression (*Zhu et al., 2018*). Nevertheless, a small subset of patients are resistant to both endocrine therapy and chemotherapy, so that lead to poor prognosis. In clinical practice, tamoxifen is one of the first-line endocrine therapies that competes with estrogen for ER and counteracts the estrogen signal pathway, which retards breast cancer growth (*Swaby, Sharma & Jordan, 2007; Ye et al., 2019*). Similarly, Doxorubicin is the most widely used DNA-damaging chemotherapeutic drug (*Pfizer et al., 2019*). Many studies have been investigated to explore the mechanism of drug resistance. Several pathways including the PI3K/Akt/mTOR pathway, fibroblast growth factor receptor pathway, etc. are activated to reverse the growth suppression induced by tamoxifen and doxorubicin (*Guo et al., 2016; Li et al., 2020a; Li et al., 2019; Lu et al., 2019; Shah et al., 2015; Turner et al., 2010*). In addition, several studies showed that tumor infiltrating lymphocytes (TILs) play a pivotal role in chemotherapy response and associate with clinical outcome in all subtypes of breast cancer (*Denkert et al., 2010; Guo et al., 2016; Stanton & Disis, 2016; Turner et al., 2010*).

The clinicopathological risk factors, such as TNM stage and histological grade, are usually used to assess the survival risk of breast cancer. However, as a result of molecular differences, clinical prognoses are mainly different even in patients with histologically similar tumors, which indicates that only using these risk factors as evaluation indicators cannot accurately predict the prognoses of breast cancer patients (*Tang et al., 2019*). Therefore, it is necessary to identify novel prognostic biomarkers to improve the prognosis evaluation ability of breast cancer.

In this study, we used the data obtained from The Cancer Genome Atlas (TCGA) breast cancer dataset and the Gene Expression Omnibus (GEO) database, then built a four resistance-related genes based risk signature and developed a nomogram to predict the survival of patients with ER+ and HER2- breast cancer in stage I-III. Finally, we analyzed the association between immune infiltrates and the four genes' expression or risk groups.

## MATERIALS & METHODS

### Data download

The raw counts of the RNA sequence data of the TCGA-BRCA dataset were downloaded from the TCGA website (<https://portal.gdc.cancer.gov/>). The clinical features and the disease-free survival (DFS) status of these patients were downloaded from the cBioportal

**Table 1** Baseline data sheet for patients in TCGA ( $n = 473$ ).

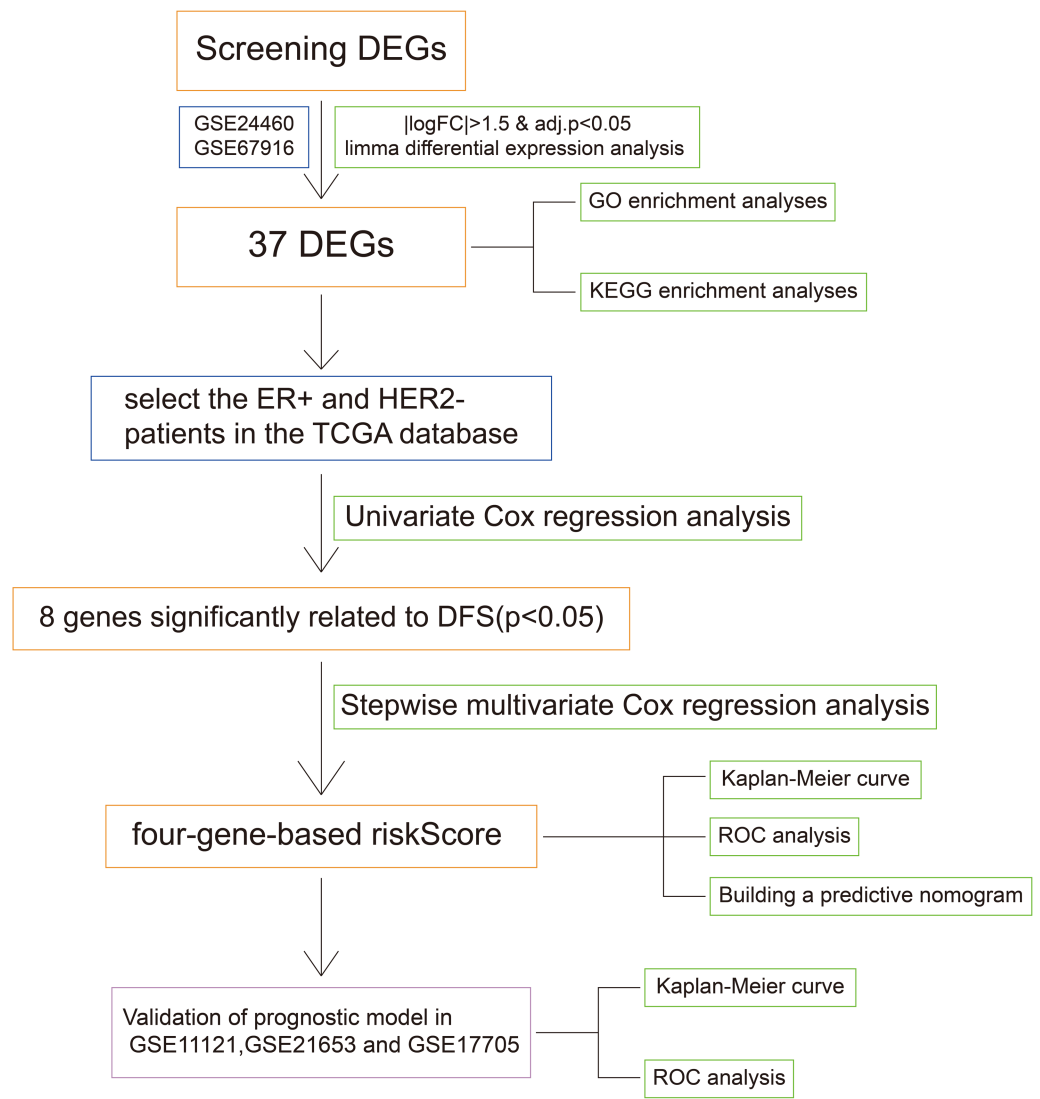
	NO.	%
Age		
$\geq 50$ year	348	73.57%
$< 50$ year	125	26.43%
T		
T1	139	29.39%
T2	259	54.76%
T3	69	14.59%
T4	6	1.27%
N		
N0	214	45.24%
N1	168	35.52%
N2	53	11.21%
N3	35	7.40%
NX	3	0.63%
Pathologic Stage		
Stage I	95	20.08%
Stage II	265	56.03%
Stage III	113	23.89%
PR Status		
positive	411	86.89%
negative	60	12.68%
unknown	2	0.42%

database (<http://www.cbioportal.org>). We selected 473 ER+ and HER2- breast cancer patients in stage I-III or the subsequent analysis (Table 1). The R software (version 3. 6. 3) was used to standardize and process data.

GSE21653 (100 samples of ER+ and HER2- early invasive breast cancer), GSE17705 (298 samples of ER+ breast cancer) and GSE11121 (200 patients) datasets were acquired from GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) as the validation cohorts. GSE21653 has DFS as the endpoint, GSE17705 has relapse-free survival (RFS) as the endpoint and GSE11121 has distant metastases-free survival (DMFS) as the endpoint.

### Screening of differentially expressed genes

The resistance-related differentially expressed genes (DEGs) were detected with the “limma” package in R software from GSE67916 (DEGs between MCF7 tamoxifen resistant (MCF7/TamR) cell lines and MCF7 parental cell lines) and GSE24460 (DEGs between MCF7 doxorubicin resistant (MCF/ADR) cell lines and MCF7 parental cell lines). Defined genes with adjusted  $p$  values  $< 0.05$  and  $|\log_2\text{foldchange}|$  values  $> 1.5$  as DEGs. The genes both up-regulated (16 genes) and down-regulated (21 genes) in the two GEO datasets were finally screened (Fig. S1B). Volcano plots and heatmaps were performed using the R packages “ggpubr”, “ggthemes” and “pheatmap” to exhibit the differentially expressed genes (Figs. S1A, Fig. 1C).



**Figure 1** The workflow of the identification of breast cancer resistance-related four gene signatures. [Full-size !\[\]\(5f471a71b78d7676bc356df190b88ab4\_img.jpg\) DOI: 10.7717/peerj.12202/fig-1](https://doi.org/10.7717/peerj.12202/fig-1)

## Identification and validation of a risk signature associated with breast cancer

The data downloaded from TCGA datasets were used as the training cohort to analyze the correlation of genes' expression with the DFS time based on univariate Cox regression analysis. Patients without survival for less than 0 days were removed to identify prognostic genes, only the genes with the  $p$  values  $< 0.05$  were selected as candidates (Fig. 2A). Then, the candidate genes were selected for stepwise multivariate Cox regression analysis (Fig. 2B). The Akaike information criterion (AIC) was used to avoid over-fitting, we chose the best-fit predictive model with the lowest AIC (AIC = 359.44). Finally, a four-gene risk signature was established and calculated their risk coefficients use the formula: The riskScore =  $(-0.175 \times \text{expression of AMIGO2}) + (-0.210 \times \text{expression of}$

LGALS3BP) + (−0.106\* expression of SCUBE2) + (−0.293\* expression of WLS). Using the optimal riskScore cut-off value (−10.09) calculated by the R package “survminer”, we divided patients in both the training cohort and the validation cohort into high-risk and low-risk group. Then, Kaplan–Meier curve and log-rank test were used to analyze the different survival time between high-risk group and low-risk group. Subsequently, a time-dependent receiver operating characteristic (ROC) curve was calculated to compare the sensitivity and specificity of risk factors.

### Development of the prognostic nomogram

To identify independent predictors of DFS, we tested clinical risk factors and the gene-based riskScore. Then we used the R package “rms” to establish the prognostic nomogram. The discrimination and calibration abilities of the nomogram were assessed with a concordance index (C-index) and calibration plot. For further evaluation of the predictive performance, we also used the time-dependent ROC curves.

### Function and pathway enrichment analyses

The R package “clusterProfiler” was used to classify the 37 DEGs’ GO terms or KEGG pathways. Functional enrichment analyses were carried out for GO terms and KEGG pathways through the significance threshold of  $p$  value < 0.05.

### Immune infiltration analyses

The 22 kinds of immune cell infiltration levels of TCGA-BRCA dataset were quantified by CIBERSORT algorithm with 1,000 permutations (<http://cibersort.stanford.edu/>) (Chen *et al.*, 2018a). CIBERSORT calculates a  $p$  value to confirm the accuracy of the results, only the samples with the  $p$  value < 0.05 were selected for analyzing the immune infiltration levels between the high- and low- risk groups. Moreover, to explore the correlation between the four genes’ expression and the abundance of immune cell infiltrates (B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages, and dendritic cells), the Tumor Immune Estimation Resource (TIMER, <https://cistrome.shinyapps.io/timer/>) was utilized.

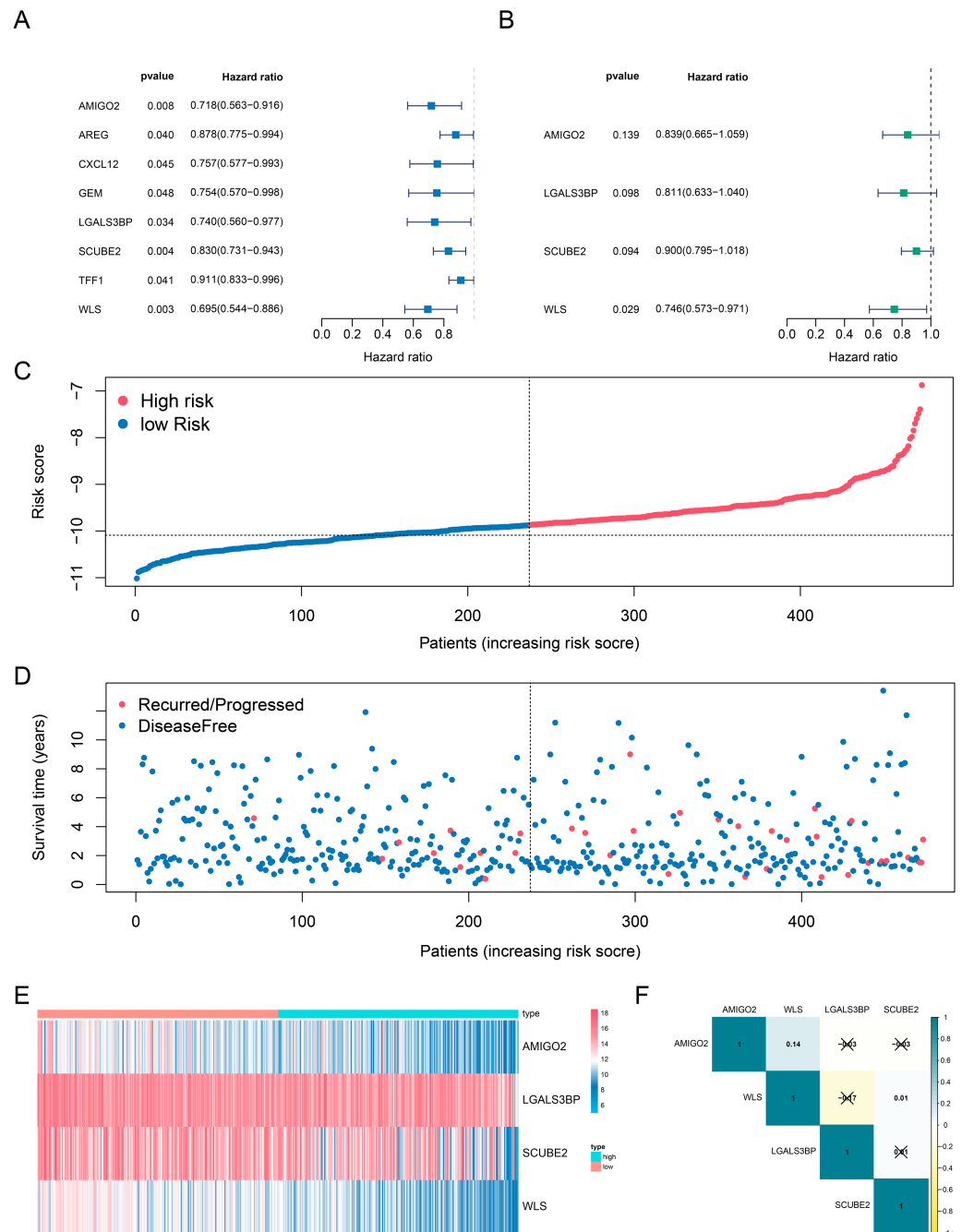
### Statistical analyses

All the statistical analyses were performed with R software (version 3.6.3). All the statistical analyses included univariate and multivariate Cox regression analysis, ROC curve analysis, Kaplan–Meier survival analysis and correlation analysis. The correlation analysis was based on the R package “corrplot”. Statistical differences were compared by Wilcoxon test between two groups and Kruskal-Wallis H test for multigroup comparison.  $P < 0.05$  was considered statistically significant.

## RESULTS

### Screening DFS-related resistance-related DEGs

The flow chart was shown in Fig. 1. To screen the DFS-related resistance-related DEGs, two datasets was downloaded from GEO database. After “limma” filtering between MCF7 doxorubicin or tamoxifen resistant cell lines and MCF parental cell lines (Fig. S1A), the intersection of up-regulated and down-regulated genes was taken respectively (Fig. S1B).



**Figure 2** Determination and analysis of the eight genes signature in the training cohort. (A, B) Univariate Cox (A) and multivariate Cox (B) regression analysis results. (C) The four-gene signature riskScore distribution plot. (D) The vital status plot of patients. (E) The heatmap of the expression profiles of the four genes. (F) The correlation of four genes in TCGA.

Full-size DOI: [10.7717/peerj.12202/fig-2](https://doi.org/10.7717/peerj.12202/fig-2)

A total of 37 resistance-related DEGs were obtained. The expression profile of the 37 DEGs in GSE24460 and GSE67916 were shown in Fig. S1C.

Next, we investigated whether genes could be used as prognostic candidates in ER positive (ER+) and HER2 negative (HER2-) breast cancer patients in stage I-III. The univariate Cox regression analysis found that eight genes were associated with DFS time in ER+ and HER2- breast cancer, all of these genes were protective factors (Fig. 2A). The eight candidates were further analyzed by multivariate Cox regression analysis and examined four genes (*AMIGO2*, *LGALS3BP*, *SCUBE2* and *WLS*) as the prognostic candidates (Table 2, Fig. 2B). In addition, we verified the prognosis of these four genes in the Kaplan–Meier Plotter website (<https://kmplot.com/analysis/>). Consistent with our results, *WLS* and *SCUBE2* were protective factors, high expression of the two genes had longer RFS time (Fig. S2C). The results of other two genes were not statistically significant. A risk score formula (the formula shown in methods) derived from the expression levels of the four prognostic candidates was constructed by the multivariate Cox regression analysis.

### Identification and validation of a four-gene based risk signature

All enrolled patients were divided into high- and low-risk groups with its optimal cut-off value among the training cohort by using the riskScore. The survival status plot and heatmap for the expression of the four genes were showed in Figs. 2C–2D. The prognostic capacity of the four-gene based riskScore was estimated by calculating the AUC of a time-dependent ROC curve (Fig. 3A). The AUCs of the four-gene risk signature were 0.686 and 0.714 for 3- and 5-year DFS respectively, suggesting that the risk signature had certain sensitivity and specificity. Kaplan–Meier curve showed that patients in the high-risk group had a lower DFS time than those in the low-risk group (Fig. 3A).

We also analyzed the correlation of the four genes in ER+ and HER2- breast cancer patients in TCGA-BRCA dataset. The correlation heatmap showed that the maximum R values was 0.14, indicating that these four genes were independent of each other (Fig. 2F).

To validate the prognostic prediction performance of riskScore, the same as training cohort, patients in the validation cohorts were classified into high- and low-risk groups using the same formula obtained from the training cohort. In all validation cohorts, patients with a high riskScore had a shorter survival time and a high riskScore was associated with worse prognosis (Figs. 3B–3D). Additionally, the time-dependent ROC analyses for the survival prediction of the risk signature showed AUCs of 0.624 at 3 years, 0.613 at 5 years in GSE21653 and 0.618 at 3 years, 0.616 at 5 years in GSE17705, demonstrating that this risk signature was capable of predicting survival in ER+ and HER2- or ER+ and HER2 equivocal patients (Figs. 3B–3C). Moreover, we also used the time-dependent ROC analysis to explore the prognostic prediction performance if the subtype of breast cancer was unknown. The AUCs were 0.730 at 3 years and 0.669 at 5 years in GSE11121 (Fig. 3D), suggesting that our risk signature also has good predictive ability when breast cancer subtype is unknown.

### Stratified analyses with clinical factors

As shown in Fig. 4, the four-gene based risk signature could serve as a promising candidate for predicting the survival of breast cancer in different subtypes stratified by the clinical



**Table 2** Four resistance-related genes associated with DFS.

Gene	Coef	HR	Group
AMIGO2	-0.17541	0.839117	Down-regulated
LGALS3BP	-0.20954	0.810961	Up-regulated
SCUBE2	-0.10588	0.899535	Down-regulated
WLS	-0.29345	0.745685	Up-regulated

factors, including N0-1 type ( $p = 0.0014$ ), N2-3 ( $p = 0.026$ ), T1-2 ( $p = 0.0006$ ) T3-4 ( $p = 0.04$ ), stage I-II ( $p = 0.0022$ ), stage III ( $p = 0.0078$ ), PR positive (PR+,  $p < 0.0001$ ) patients, respectively. But the PR negative (PR-) subtype was not statistically significant. As shown in Fig. 5, the riskScore value was higher expression in PR- subtype than in PR+ subtype, *SCUBE2* and *WLS* were higher expression in PR+ subtype than in PR-subtype, the expression of *SCUBE2* was negatively correlated with T and Stage, these results were consistent with the clinical prognostic significance.

### Functional enrichment analyses

To explain the function of the resistance-related DEGs, we performed enrichment analyses of the GO terms and KEGG pathways (Figs. S2A–S2B).

GO analyses showed that these resistance-related DEGs were enriched in several biological processes (BP), including antigen processing and presentation of peptide antigen *via* MHC class I, antigen processing and presentation of exogenous peptide antigen *via* MHC class I and epithelial structure maintenance (Fig. S2A).

KEGG analysis showed that the resistance-related DEGs were related to endocytosis, cellular senescence, antigen processing and presentation and adherens junction (Fig. S2B).

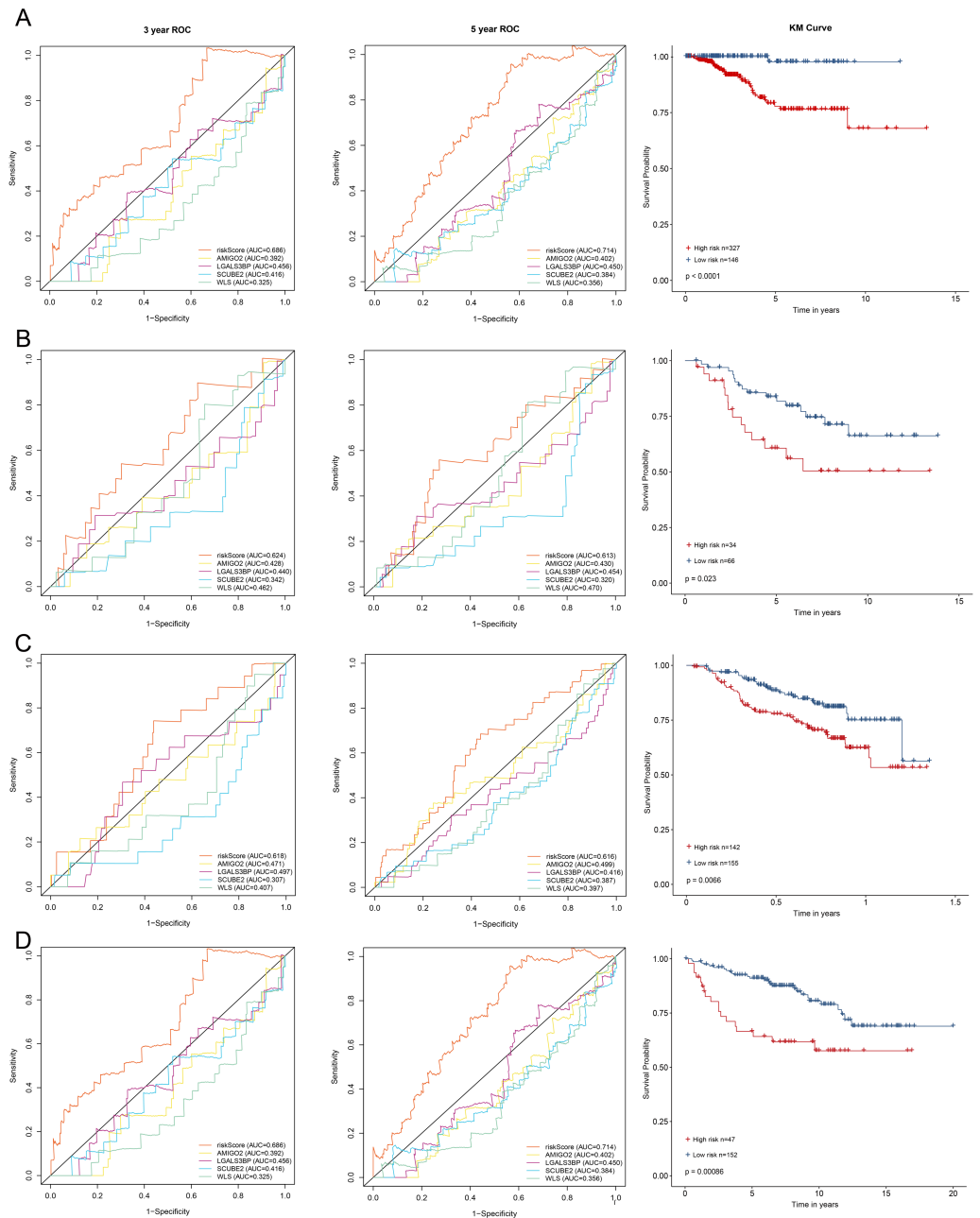
### Revealing the correlation between the risk signature and tumor-infiltrating immune cells

Because TILs play a key role in chemotherapy response and our enrichment analyses results showed that the DEGs associated with antigen presentation, first, we evaluated the abundance of the 22 kinds of immune cells in each TCGA-BRCA patient (Fig. 6A). Next, we utilized the Wilcoxon test to compare the difference immune cell infiltration levels between the high- and low groups and found that B cells memory ( $p = 0.042$ ) and T cells follicular helper ( $p = 0.016$ ) were high infiltrated in the high-risk group, and B cells naive ( $p = 0.047$ ) and mast cells resting ( $p = 0.042$ ) were low infiltrated in the high-risk group (Fig. 6B). Unfortunately, due to the limitation of sample size, the relationship between the immune cell infiltration levels and survival is unknown. Whereafter, we assessed the correlation between the four resistance-related DEGs and six infiltrating immune cells by using the TIMER database. The results revealed that the genes had good correlation with immune cells (Fig. S3).

### Building a prognostic nomogram combining riskScore with clinical risk factors

Nomograms are widely used in predicting cancer patients' prognoses, principally because they have the ability to decrease the statistical prediction models into a single numerical



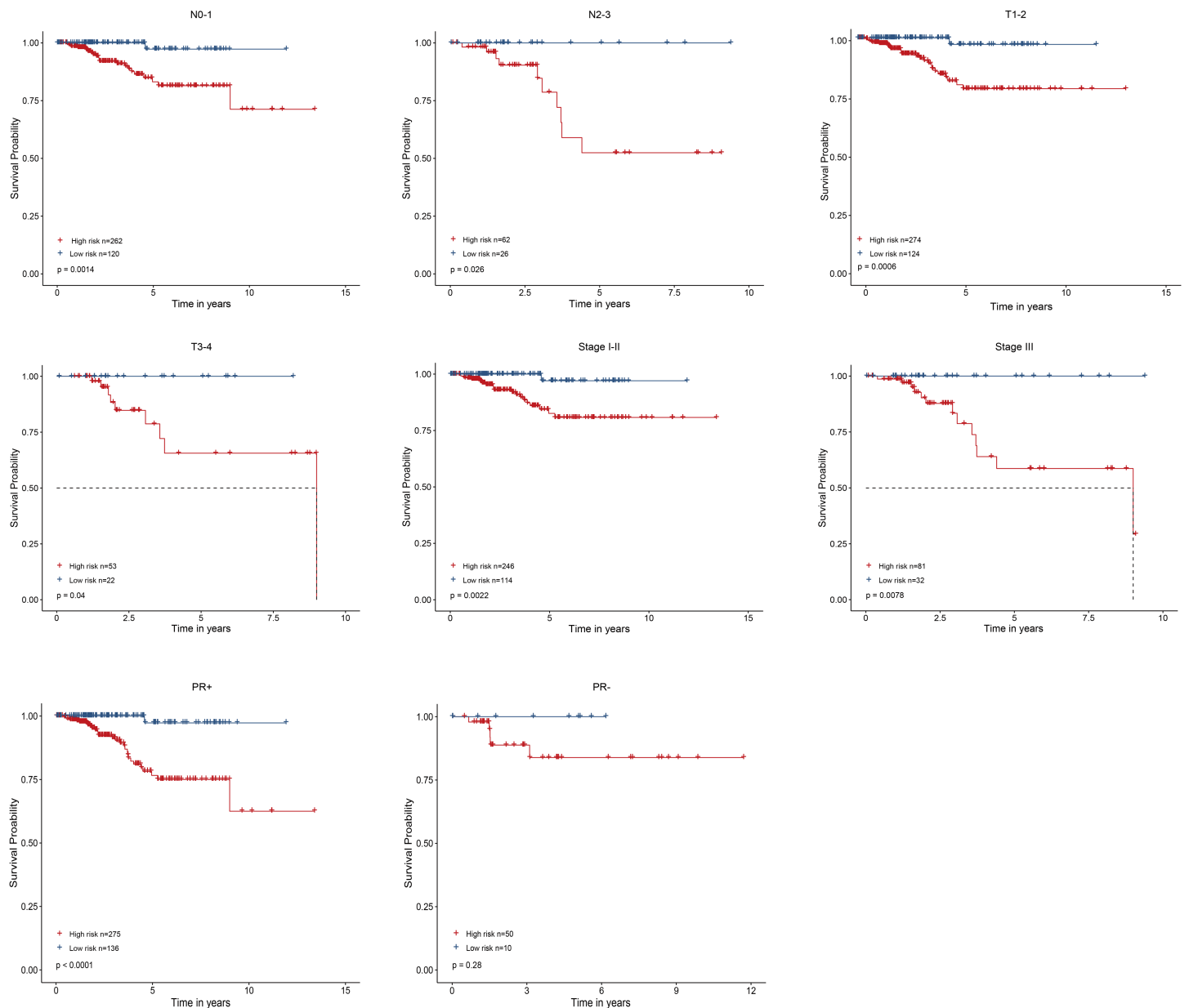


**Figure 3** Kaplan–Meier survival curves of DFS between high-risk and low-risk patients and ROC curves in (A) TCGA-BRCA, (B) GSE21653, (C) GSE17705 and (D) GSE11121.

Full-size [DOI: 10.7717/peerj.12202/fig-3](https://doi.org/10.7717/peerj.12202/fig-3)

assessment, which is tailored to the profile of an individual patient (*Iasonos et al., 2008; Long et al., 2018*).

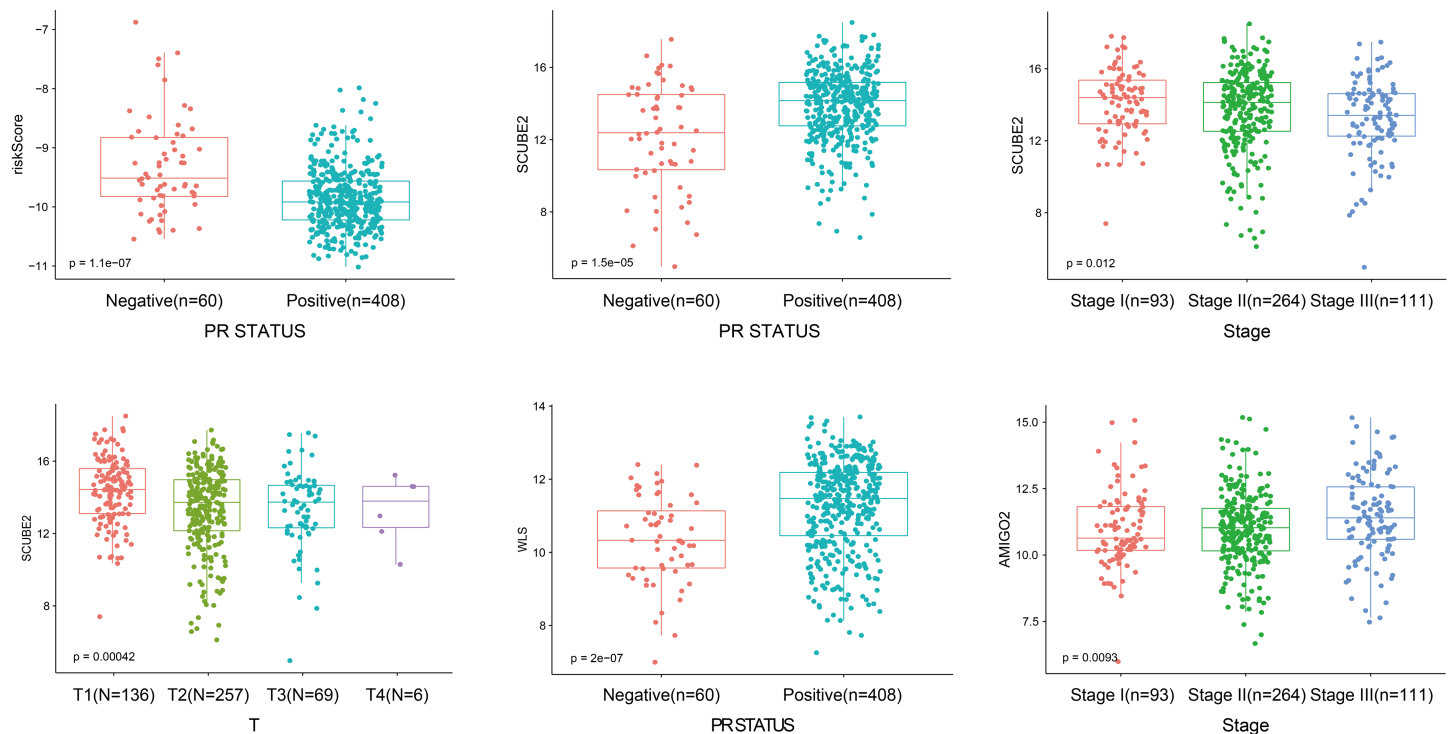
In this article, we combined the four-gene based riskScore with clinical risk factors to construct a nomogram to assess the probability of 3- and 5- year DFS for patients with ER+ and HER2- early stage invasive breast cancer in TCGA-BRCA dataset. After univariate and



**Figure 4** Performed the risk stratified subgroup analysis according to clinical factors.

Full-size DOI: [10.7717/peerj.12202/fig-4](https://doi.org/10.7717/peerj.12202/fig-4)

multivariate Cox regression analyses filtering, the riskScore, T and Stage remained in the final Cox model for DFS (Figs. 7A–7B). A nomogram was established based on the final Cox model for DFS to evaluate of the 3- or 5-year DFS (Fig. 7C). The clinical stage and riskScore of the patient can be matched with this Nomogram to obtain the total points and estimate the survival rate of the patient. The nomogram achieved a C-index of 0.740 (95% CI, 0.653–0.826), and the calibration plots demonstrated that the nomogram has good predictive power (Fig. 7D). Time-dependent ROC analyses also indicated that the nomogram was accurate in prediction of DFS compared to use riskScore or clinical risk



**Figure 5** The expression of the four genes in different clinicopathological risk factors.

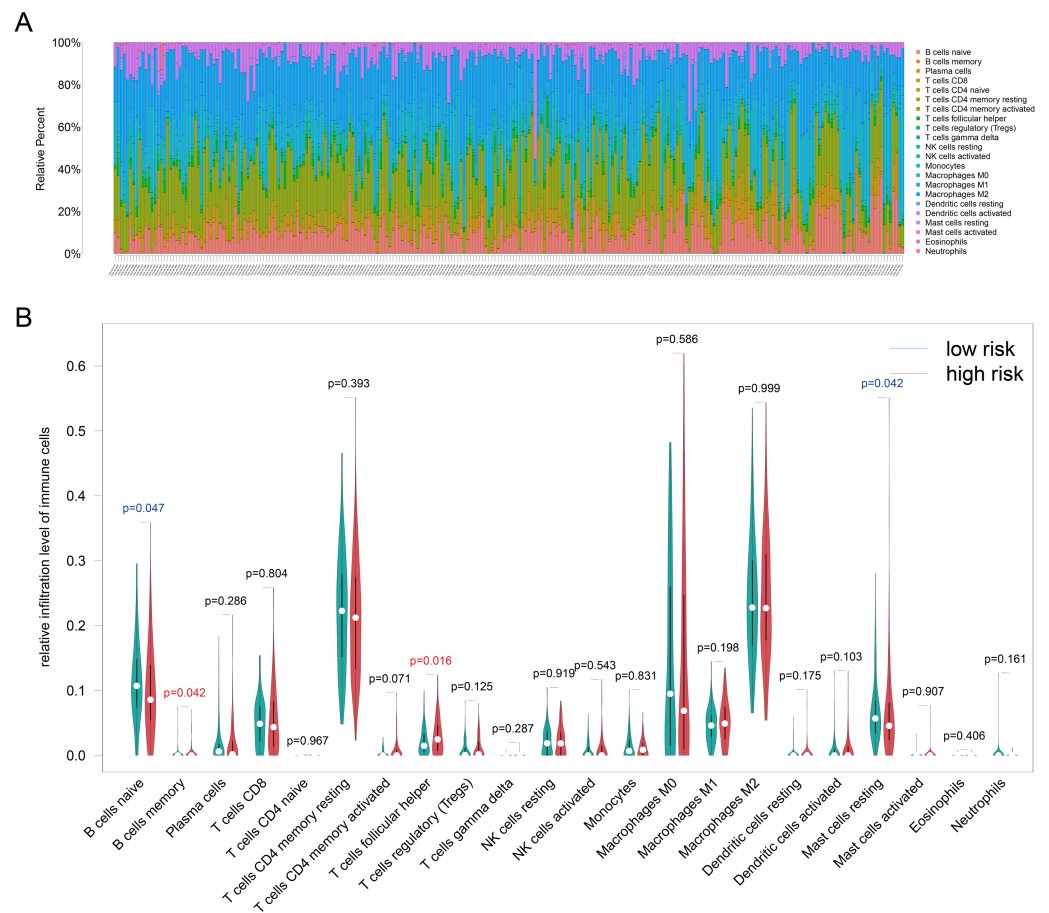
Full-size DOI: [10.7717/peerj.12202/fig-5](https://doi.org/10.7717/peerj.12202/fig-5)

factors alone (Fig. 7E). In conclusion, all the results suggest the clinical practical use of the nomogram for prediction of DFS.

## DISCUSSION

Breast cancer is highly heterogeneous in molecular and cellular, the morbidity of it increases every year. In recent years, molecular biomarkers of breast cancer become more and more important. Understanding the biological functions of breast cancer may offer new methods that can be used to predict and treat the disease. However, there are still few specific biomarkers with significant prognostic to predict survival of breast cancer. Therefore, in order to improve the prognosis of breast cancer, molecular screening of biomarkers is urgently needed.

In this study, we developed and validated a novel risk signature based on four resistance-related genes to predict patients with stage I-III ER+ and HER2- breast cancer. This article showed that our risk signature could effectively distinguish the breast cancer patients between low and high risk of survival groups. In addition, the four-gene based model was demonstrate to be independent from other clinicopathological features. Then, a comprehensive nomogram was constructed to predict the individualized DFS, which performed better than clinical factors and riskScore to classify the patients' survival risk. Finally, we explored the correlation between the four resistance-related DEGs and immune infiltrates.

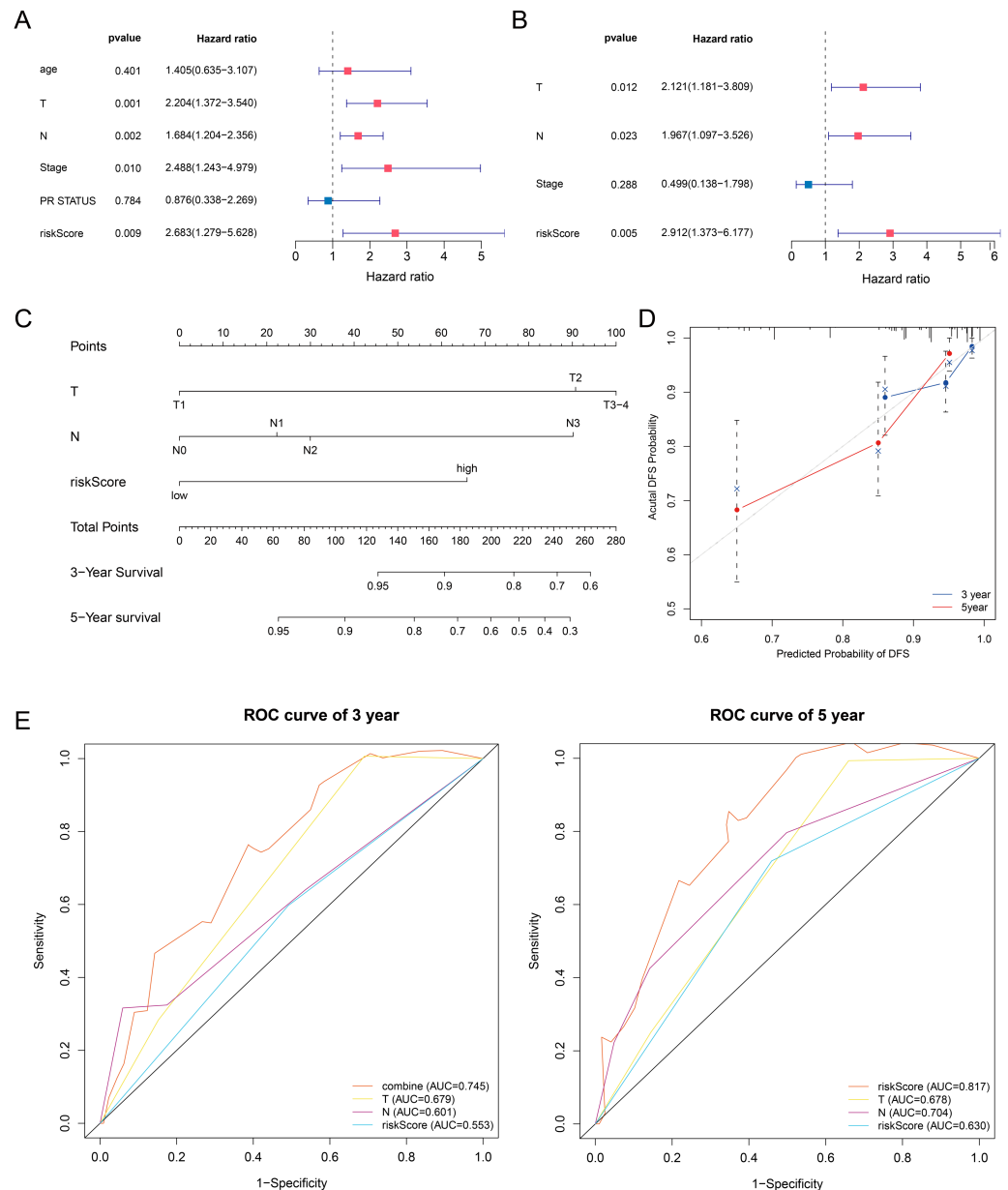


**Figure 6** Immune infiltration analyses results. (A) The relative infiltration level of immune cells in patients of TCGA-BRCA dataset, the 22 immune cells were annotated by various colors under the legend. (B) The difference immune cell infiltration levels between the high- and low groups.

Full-size [DOI: 10.7717/peerj.12202/fig-6](https://doi.org/10.7717/peerj.12202/fig-6)

Previous investigation proved that protein-based signature could predict tamoxifen treatment outcome in recurrent breast cancer and Long Noncoding RNA based signature could predict recurrence among ER+ breast cancer patients treated with tamoxifen (De Marchi et al., 2016; Wang et al., 2018). In reality, the expression profile-based prognostic biomarkers signature has been already investigated in several types of cancer, including glioblastoma multiforme, colon cancer, hepatocellular carcinoma, gastric cancer and non-small cell lung cancer (Long et al., 2018; Qiu et al., 2020; Wang et al., 2020; Zuo et al., 2019). But there is no research on the application of predictive models of resistance-related genes in ER+ and HER2- breast cancer.

There are four genes involved in building the risk signature in this study, they are *AMIGO2*, *LGALS3BP*, *SCUBE2* and *WLS*. All of these genes are protective markers in our study, there are no previous studies demonstrated their role in drug resistance. These genes play different roles in different diseases.



**Figure 7** Building a prognostic nomogram combining riskScore with clinical risk factors. (A,B) Univariate Cox (A) and multivariate Cox (B) association of the risk signature and clinicopathological characteristics with DFS in TCGA-BRCA. (C) Nomogram predicting 3- and 5-year DFS for patients with ER+ and HER2- breast cancer patients in stage I-III. (D) The calibration plots of the nomogram in prediction of the 3-year and 5-year DFS in the training cohort. (E) Time-dependent ROC analyses evaluate the accuracy of the nomogram and riskScore or clinical risk factors.

Full-size DOI: 10.7717/peerj.12202/fig-7

Adhesion molecule with IgG-like domain 2 (*AMIGO2*) is a multifunctional amphoterin-induced genes (*Kuja-Panula et al., 2003*). *AMIGO2* controls cell survival and angiogenesis in endothelial cells *via* Akt activation, knockout of *AMIGO2* can reduce cell survival and induce apoptosis (*Park et al., 2015*). *AMIGO2* also can promote liver metastasis by

regulating the tumor cell adhesion to liver endothelial cells and its function as a pro-proliferation factor in human melanoma *in vitro* and *in vivo* (Fontanals-Cirera et al., 2017; Kanda et al., 2017).

The lectin galactoside-binding soluble 3 binding protein (LGALS3BP), a large oligomeric glycoprotein, is overexpressed in many cancers and always associated with cancer growth and migration (Woo et al., 2017). LGALS3BP can induce vascular endothelial growth factor in human MDA-MB-231 cells and promotes angiogenesis, it also acts as a E-selectin ligand that is associated with metastasis and poor prognosis in ER- breast cancer (Piccolo et al., 2013; Traini et al., 2014). But in colon cancer, different from previous studies, LGALS3BP suppresses tumor proliferation in colon cancer cells, high expression of LGALS3BP has a good prognosis (Piccolo et al., 2015). There are no researches to explore the function of LGALS3BP in ER+ and HER2- invasive breast cancer.

The findings of Signal Peptide-CUB-EGF Domain-Containing Protein 2 (SCUBE2) are controversial. Chen et al. (2018b) proved that upregulated SCUBE2 expression in breast cancer stem cells can promote triple negative breast cancer aggression. But Lin et al. (2014) demonstrated that SCUBE2 inhibited MDA-MB-231 cell proliferation. In MCF-7 breast cancer cell line, overexpression of SCUBE2 protein suppresses proliferation, this result is consistent with our study (Cheng et al., 2009). SCUBE2 also performed as a protective factor in glioma, non-small cell lung cancer and colorectal cancer and inhibited the proliferation and migration (Guo, Liu & Liu, 2017; Song et al., 2015; Yang, Miao & Li, 2018).

The Wnt secretion protein Wntless (WLS) has been proved to be involved in the development of several human cancers. WLS mostly plays the role of oncogene in most tumors. In triple negative breast cancer, glioma and colon cancer, WLS promotes the cell proliferation (Augustin et al., 2012; Voloshanenko et al., 2013). In lung adenocarcinoma cell lines, WLS is upregulated by small nucleolar RNA host gene 17 (SNHG17), the up-regulation of WLS could reverse the inhibition of proliferation and promote apoptosis effects of SNHG17 down-regulation (Li et al., 2020b). But in melanoma cell, WLS inhibits proliferation through the  $\beta$ -catenin signaling pathway and reduces spontaneous metastasis (Yang et al., 2012).

Tumor-infiltrating immune cells in the TME play a crucial role in antitumor and chemotherapeutic effects (Denkert et al., 2010; Stanton & Disis, 2016). Therefore, we used the CIBERSORT algorithm and the TIMER database to detect the relationship between risk signature and immune cells. The results proved that B cells memory and T cells follicular helper (Tfh) were positive correlated infiltrated in the high-risk group, but B cells naive and mast cells resting were negative correlated infiltrated in the high-risk group. Tfh cells are crucial for maintenance of humoral memory that assist B cells in producing germinal centers (Crotty, 2019). Mast cells may play protumoral or antitumoral roles in different tumor types, for instance, it could inhibit prostate adenocarcinoma development but promote the tumorigenesis of highly malignant neuroendocrine cancers (Pittoni et al., 2011) discovered.

This study also has some limitations. This is a retrospective research study and is susceptible to inherent biases. First, the study has potential selection bias and limited sample size. Then, the riskScore and nomogram cannot guide treatment strategies of breast

cancer, so that further verification and prospective trials are needed. Moreover, all of the data were obtained from the database, the population ethnicities in the TCGA database are mainly composed of White patients and Black patients; therefore, the findings of other ethnicities still require proof, and some other clinicopathological features including histological grade, are not contained in the nomogram. Because of the lack of clinical data in database, the nomogram has not been verified, so further researches are needed to verify the nomogram. Finally, the precise mechanism of resistance-related genes involved in the risk signature need further *in vitro* and *in vivo* validation experiment.

## CONCLUSIONS

In this study, we used the data from TCGA-BRCA dataset and GEO database, constructed a risk signature based on four drug resistance-related genes and developed a nomogram to predict the survival of patients with stage I-III in ER+ and HER2- breast cancer. The results were validated in three GEO cohorts. All the results indicated that this risk signature might serve as novel molecular biomarkers for predicting breast cancer survival and provide guidance for therapeutic strategies.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

The authors received no funding for this work.

### Competing Interests

The authors declare there are no competing interests.

### Author Contributions

- Chen Shuai and Hongye He conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Fengyan Yuan conceived and designed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Yu Liu performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Chengchen Wang and Jiansong Wang performed the experiments, prepared figures and/or tables, and approved the final draft.

### Data Availability

The following information was supplied regarding data availability:

This data is available at TCGA-BRCA and NCBI: [GSE24460](https://portal.gdc.cancer.gov/) and [GSE67916](https://portal.gdc.cancer.gov/).

<https://portal.gdc.cancer.gov/>

### Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.12202#supplemental-information>.



## REFERENCES

- Achinger-Kawecka J, Valdes-Mora F, Luu PL, Giles KA, Caldon CE, Qu W, Nair S, Soto S, Locke WJ, Yeo-Teh NS, Gould CM, Du Q, Smith GC, Ramos IR, Fernandez KF, Hoon DS, Gee JMW, Storzaker C, Clark SJ. 2020. Epigenetic reprogramming at estrogen-receptor binding sites alters 3D chromatin landscape in endocrine-resistant breast cancer. *Nature Communications* 11:320 DOI 10.1038/s41467-019-14098-x.
- Al-Thoubaity FK. 2020. Molecular classification of breast cancer: a retrospective cohort study. *Annals of Medicine and Surgery* 49:44–48 DOI 10.1016/j.amsu.2019.11.021.
- Augustin I, Goidts V, Bongers A, Kerr G, Vollert G, Radlwimmer B, Hartmann C, Herold-Mende C, Reifenberger G, Von Deimling A, Boutros M. 2012. The Wnt secretion protein Evi/Gpr177 promotes glioma tumorigenesis. *EMBO Molecular Medicine* 4:38–51 DOI 10.1002/emmm.201100186.
- Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. 2018a. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods in Molecular Biology* 1711:243–259 DOI 10.1007/978-1-4939-7493-1\_12.
- Chen JH, Kuo KT, Bamodu OA, Lin YC, Yang RB, Yeh CT, Chao TY. 2018b. Upregulated SCUBE2 expression in breast cancer stem cells enhances triple negative breast cancer aggression through modulation of notch signaling and epithelial-to-mesenchymal transition. *Experimental Cell Research* 370:444–453 DOI 10.1016/j.yexcr.2018.07.008.
- Cheng CJ, Lin YC, Tsai MT, Chen CS, Hsieh MC, Chen CL, Yang RB. 2009. SCUBE2 suppresses breast tumor cell proliferation and confers a favorable prognosis in invasive breast cancer. *Cancer Research* 69:3634–3641 DOI 10.1158/0008-5472.CAN-08-3615.
- Crotty S. 2019. T follicular helper cell biology: a decade of discovery and diseases. *Immunity* 50:1132–1148 DOI 10.1016/j.immuni.2019.04.011.
- De Marchi T, Liu NQ, Stingl C, Timmermans MA, Smid M, Look MP, Tjoa M, Braakman RB, Opdam M, Linn SC, Sweep FC, Span PN, Kliffen M, Luider TM, Foekens JA, Martens JW, Umar A. 2016. 4-protein signature predicting tamoxifen treatment outcome in recurrent breast cancer. *Molecular Oncology* 10:24–39 DOI 10.1016/j.molonc.2015.07.004.
- Denkert C, Loibl S, Noske A, Roller M, Muller BM, Komor M, Budczies J, Darb-Esfahani S, Kronenwett R, Hanusch C, Von Torne C, Weichert W, Engels K, Solbach C, Schrader I, Dietel M, Von Minckwitz G. 2010. Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. *Journal of Clinical Oncology* 28:105–113 DOI 10.1200/JCO.2009.23.7370.
- Dianatinasab M, Fararouei M, Daneshi N, Rezaian S, Mohammadianpanah M, Chaman R, Ghiasvand R. 2019. Heterogeneity in risk factors for ductal and lobular breast carcinomas: a case-control study. *International Journal of Cancer* 145:2917–2925 DOI 10.1002/ijc.32182.

- Fontanals-Cirera B, Hasson D, Vardabasso C, Di Micco R, Agrawal P, Chowdhury A, Gantz M, De Pablos-Aragoneses A, Morgenstern A, Wu P, Filipescu D, Valle-Garcia D, Darvishian F, Roe JS, Davies MA, Vakoc CR, Hernando E, Bernstein E. 2017.** Harnessing BET inhibitor sensitivity reveals a melanoma survival gene. *Molecular Cell* **68**:731–744 DOI [10.1016/j.molcel.2017.11.004](https://doi.org/10.1016/j.molcel.2017.11.004).
- Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ, Panel M. 2011.** Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Annals of Oncology* **22**:1736–1747 DOI [10.1093/annonc/mdr304](https://doi.org/10.1093/annonc/mdr304).
- Guo Y, Ding Y, Zhang T, An H. 2016.** Sinapine reverses multi-drug resistance in MCF-7/dox cancer cells by downregulating FGFR4/FRS2alpha-ERK1/2 pathway-mediated NF-kappaB activation. *Phytomedicine* **23**:267–273 DOI [10.1016/j.phymed.2015.12.017](https://doi.org/10.1016/j.phymed.2015.12.017).
- Guo E, Liu H, Liu X. 2017.** Overexpression of SCUBE2 inhibits proliferation, migration, and invasion in glioma cells. *Oncology Research* **25**:437–444 DOI [10.3727/096504016X14747335734344](https://doi.org/10.3727/096504016X14747335734344).
- Iasonos A, Schrag D, Raj GV, Panageas KS. 2008.** How to build and interpret a nomogram for cancer prognosis. *Journal of Clinical Oncology* **26**:1364–1370 DOI [10.1200/JCO.2007.12.9791](https://doi.org/10.1200/JCO.2007.12.9791).
- Kanda Y, Osaki M, Onuma K, Sonoda A, Kobayashi M, Hamada J, Nicolson GL, Ochiya T, Okada F. 2017.** Amigo2-upregulation in tumour cells facilitates their attachment to liver endothelial cells resulting in liver metastases. *Scientific Reports* **7**:43567 DOI [10.1038/srep43567](https://doi.org/10.1038/srep43567).
- Kuja-Panula J, Kiiltomaki M, Yamashiro T, Rouhiainen A, Rauvala H. 2003.** AMIGO, a transmembrane protein implicated in axon tract development, defines a novel protein family with leucine-rich repeats. *Journal of Cell Biology* **160**:963–973 DOI [10.1083/jcb.200209074](https://doi.org/10.1083/jcb.200209074).
- Li D, Ji H, Niu X, Yin L, Wang Y, Gu Y, Wang J, Zhou X, Zhang H, Zhang Q. 2020a.** Tumor-associated macrophages secrete CC-chemokine ligand 2 and induce tamoxifen resistance by activating PI3K/Akt/mTOR in breast cancer. *Cancer Science* **111**:47–58 (c) 2019 The Authors. Cancer Science published by John Wiley & Sons Australia, Ltd on behalf of Japanese Cancer Association DOI [10.1111/cas.14230](https://doi.org/10.1111/cas.14230).
- Lin YC, Lee YC, Li LH, Cheng CJ, Yang RB. 2014.** Tumor suppressor SCUBE2 inhibits breast-cancer cell migration and invasion through the reversal of epithelial-mesenchymal transition. *Journal of Cell Science* **127**:85–100 DOI [10.1242/jcs.132779](https://doi.org/10.1242/jcs.132779).
- Li Z, Qian J, Li J, Zhu C. 2019.** Knockdown of lncRNA-HOTAIR downregulates the drug-resistance of breast cancer cells to doxorubicin via the PI3K/AKT/mTOR signaling pathway. *Experimental and Therapeutic Medicine* **18**:435–442 DOI [10.3892/etm.2019.7629](https://doi.org/10.3892/etm.2019.7629).
- Li CI, Uribe DJ, Daling JR. 2005.** Clinical characteristics of different histologic types of breast cancer. *British Journal of Cancer* **93**:1046–1052 DOI [10.1038/sj.bjc.6602787](https://doi.org/10.1038/sj.bjc.6602787).

- Li W, Zheng Y, Mao B, Wang F, Zhong Y, Cheng D. 2020b.** SNHG17 upregulates WLS expression to accelerate lung adenocarcinoma progression by sponging miR-485-5p. *Biochemical and Biophysical Research Communications* **533**:1435–1441 DOI [10.1016/j.bbrc.2020.09.130](https://doi.org/10.1016/j.bbrc.2020.09.130).
- Long J, Zhang L, Wan X, Lin J, Bai Y, Xu W, Xiong J, Zhao H. 2018.** A four-gene-based prognostic model predicts overall survival in patients with hepatocellular carcinoma. *Journal of Cellular and Molecular Medicine* **22**:5928–5938 DOI [10.1111/jcmm.13863](https://doi.org/10.1111/jcmm.13863).
- Lu PW, Li L, Wang F, Gu YT. 2019.** Inhibitory role of large intergenic noncoding RNA-ROR on tamoxifen resistance in the endocrine therapy of breast cancer by regulating the PI3K/Akt/mTOR signaling pathway. *Journal of Cellular Physiology* **234**:1904–1912 DOI [10.1002/jcp.27066](https://doi.org/10.1002/jcp.27066).
- Park H, Lee S, Shrestha P, Kim J, Park JA, Ko Y, Ban YH, Park DY, Ha SJ, Koh GY, Hong VS, Mochizuki N, Kim YM, Lee W, Kwon YG. 2015.** AMIGO2, a novel membrane anchor of PDK1, controls cell survival and angiogenesis via Akt activation. *Journal of Cell Biology* **211**:619–637 DOI [10.1083/jcb.201503113](https://doi.org/10.1083/jcb.201503113).
- Pfitzer L, Moser C, Gegenfurtner F, Arner A, Foerster F, Atzberger C, Zisis T, Kubisch-Dohmen R, Busse J, Smith R, Timinszky G, Kalinina OV, Muller R, Wagner E, Vollmar AM, Zahler S. 2019.** Targeting actin inhibits repair of doxorubicin-induced DNA damage: a novel therapeutic approach for combination therapy. *Cell Death & Disease* **10**:302 DOI [10.1038/s41419-019-1546-9](https://doi.org/10.1038/s41419-019-1546-9).
- Piccolo E, Tinari N, D'Addario D, Rossi C, Iacobelli V, La Sorda R, Lattanzio R, D'Egidio M, DiRisio A, Piantelli M, Natali PG, Iacobelli S. 2015.** Prognostic relevance of LGALS3BP in human colorectal carcinoma. *Journal of Translational Medicine* **13**:248 DOI [10.1186/s12967-015-0606-x](https://doi.org/10.1186/s12967-015-0606-x).
- Piccolo E, Tinari N, Semeraro D, Traini S, Fichera I, Cumashi A, La Sorda R, Spinella F, Bagnato A, Lattanzio R, D'Egidio M, Di Risio A, Stampolidis P, Piantelli M, Natoli C, Ullrich A, Iacobelli S. 2013.** LGALS3BP, lectin galactoside-binding soluble 3 binding protein, induces vascular endothelial growth factor in human breast cancer cells and promotes angiogenesis. *Journal of Molecular Medicine* **91**:83–94 DOI [10.1007/s00109-012-0936-6](https://doi.org/10.1007/s00109-012-0936-6).
- Pittoni P, Tripodo C, Piconese S, Mauri G, Parenza M, Rigoni A, Sangaletti S, Colombo MP. 2011.** Mast cell targeting hampers prostate adenocarcinoma development but promotes the occurrence of highly malignant neuroendocrine cancers. *Cancer Research* **71**:5987–5997 DOI [10.1158/0008-5472.Can-11-1637](https://doi.org/10.1158/0008-5472.Can-11-1637).
- Qiu J, Sun M, Wang Y, Chen B. 2020.** Identification and validation of an individualized autophagy-clinical prognostic index in gastric cancer patients. *Cancer Cell International* **20**:178 DOI [10.1186/s12935-020-01267-y](https://doi.org/10.1186/s12935-020-01267-y).
- Shah KN, Wilson EA, Malla R, Elford HL, Faridi JS. 2015.** Targeting ribonucleotide reductase M2 and NF-kappaB activation with didox to circumvent tamoxifen resistance in breast cancer. *Molecular Cancer Therapeutics* **14**:2411–2421 DOI [10.1158/1535-7163.MCT-14-0689](https://doi.org/10.1158/1535-7163.MCT-14-0689).

- Song Q, Li C, Feng X, Yu A, Tang H, Peng Z, Wang X. 2015.** Decreased expression of SCUBE2 is associated with progression and prognosis in colorectal cancer. *Oncology Reports* **33**:1956–1964 DOI [10.3892/or.2015.3790](https://doi.org/10.3892/or.2015.3790).
- Stanton SE, Disis ML. 2016.** Clinical significance of tumor-infiltrating lymphocytes in breast cancer. *Journal for ImmunoTherapy of Cancer* **4**:59 DOI [10.1186/s40425-016-0165-6](https://doi.org/10.1186/s40425-016-0165-6).
- Swaby RF, Sharma CG, Jordan VC. 2007.** SERMs for the treatment and prevention of breast cancer. *Reviews in Endocrine and Metabolic Disorders* **8**:229–239 DOI [10.1007/s11154-007-9034-4](https://doi.org/10.1007/s11154-007-9034-4).
- Tang J, Ren J, Cui Q, Zhang D, Kong D, Liao X, Lu M, Gong Y, Wu G. 2019.** A prognostic 10-lncRNA expression signature for predicting the risk of tumour recurrence in breast cancer patients. *Journal of Cellular and Molecular Medicine* **23**:6775–6784 DOI [10.1111/jcmm.14556](https://doi.org/10.1111/jcmm.14556).
- Traini S, Piccolo E, Tinari N, Rossi C, La Sorda R, Spinella F, Bagnato A, Lattanzio R, D'Egidio M, Di Risio A, Tomao F, Grassadonia A, Piantelli M, Natoli C, Iacobelli S. 2014.** Inhibition of tumor growth and angiogenesis by SP-2, an anti-lectin, galactoside-binding soluble 3 binding protein (LGALS3BP) antibody. *Molecular Cancer Therapeutics* **13**:916–925 DOI [10.1158/1535-7163.MCT-12-1117](https://doi.org/10.1158/1535-7163.MCT-12-1117).
- Turner N, Pearson A, Sharpe R, Lambros M, Geyer F, Lopez-Garcia MA, Natrajan R, Marchio C, Iorns E, Mackay A, Gillett C, Grigoriadis A, Tutt A, Reis-Filho JS, Ashworth A. 2010.** FGFR1 amplification drives endocrine therapy resistance and is a therapeutic target in breast cancer. *Cancer Research* **70**:2085–2094 DOI [10.1158/0008-5472.CAN-09-3746](https://doi.org/10.1158/0008-5472.CAN-09-3746).
- Viale G. 2012.** The current state of breast cancer classification. *Annals of Oncology* **23**(Suppl 10):x207–x210 DOI [10.1093/annonc/mds326](https://doi.org/10.1093/annonc/mds326).
- Voloshanenko O, Erdmann G, Dubash TD, Augustin I, Metzger M, Moffa G, Hundsrucker C, Kerr G, Sandmann T, Anchang B, Demir K, Boehm C, Leible S, Ball CR, Glimm H, Spang R, Boutros M. 2013.** Wnt secretion is required to maintain high levels of Wnt activity in colon cancer cells. *Nature Communications* **4**:2610 DOI [10.1038/ncomms3610](https://doi.org/10.1038/ncomms3610).
- Wang K, Li J, Xiong Y-F, Zeng Z, Zhang X, Li H-Y. 2018.** A potential prognostic long noncoding RNA signature to predict recurrence among ER-positive breast cancer patients treated with tamoxifen. *Scientific Reports* **8**:3179 DOI [10.1038/s41598-018-21581-w](https://doi.org/10.1038/s41598-018-21581-w).
- Wang Y, Zhao W, Xiao Z, Guan G, Liu X, Zhuang M. 2020.** A risk signature with four autophagy-related genes for predicting survival of glioblastoma multiforme. *Journal of Cellular and Molecular Medicine* **24**:3807–3821 DOI [10.1111/jcmm.14938](https://doi.org/10.1111/jcmm.14938).
- Woo JK, Jang JE, Kang JH, Seong JK, Yoon YS, Kim HC, Lee SJ, Oh SH. 2017.** Lectin, galactoside-binding soluble 3 binding protein promotes 17-N-Allylamino-17-Demethoxygeldanamycin resistance through PI3K/Akt pathway in lung cancer cell line. *Molecular Cancer Therapeutics* **16**:1355–1365 DOI [10.1158/1535-7163.MCT-16-0574](https://doi.org/10.1158/1535-7163.MCT-16-0574).

- Yang PT, Anastas JN, Toroni RA, Shinohara MM, Goodson JM, Bosserhoff AK, Chien AJ, Moon RT. 2012.** WLS inhibits melanoma cell proliferation through the beta-catenin signalling pathway and induces spontaneous metastasis. *EMBO Molecular Medicine* 4:1294–1307 DOI [10.1002/emmm.201201486](https://doi.org/10.1002/emmm.201201486).
- Yang B, Miao S, Li Y. 2018.** SCUBE2 inhibits the proliferation, migration and invasion of human non-small cell lung cancer cells through regulation of the sonic hedgehog signaling pathway. *Gene* 672:143–149 DOI [10.1016/j.gene.2018.06.012](https://doi.org/10.1016/j.gene.2018.06.012).
- Ye L, Lin C, Wang X, Li Q, Li Y, Wang M, Zhao Z, Wu X, Shi D, Xiao Y, Ren L, Jian Y, Yang M, Ou R, Deng G, Ouyang Y, Chen X, Li J, Song L. 2019.** Epigenetic silencing of SALL2 confers tamoxifen resistance in breast cancer. *EMBO Molecular Medicine* 11:e10638 DOI [10.15252/emmm.201910638](https://doi.org/10.15252/emmm.201910638).
- Zhu Y, Liu Y, Zhang C, Chu J, Wu Y, Li Y, Liu J, Li Q, Li S, Shi Q, Jin L, Zhao J, Yin D, Efroni S, Su F, Yao H, Song E, Liu Q. 2018.** Tamoxifen-resistant breast cancer cells are resistant to DNA-damaging chemotherapy because of upregulated BARD1 and BRCA1. *Nature Communications* 9:1595 DOI [10.1038/s41467-018-03951-0](https://doi.org/10.1038/s41467-018-03951-0).
- Zuo S, Wei M, Zhang H, Chen A, Wu J, Wei J, Dong J. 2019.** A robust six-gene prognostic signature for prediction of both disease-free and overall survival in non-small cell lung cancer. *Journal of Translational Medicine* 17:152 DOI [10.1186/s12967-019-1899-y](https://doi.org/10.1186/s12967-019-1899-y).