Original Article

Identification and validation of a novel prognostic signature based on transcription factors in breast cancer by bioinformatics analysis

Yingmei Yang[1], Zhaoyun Li[1], Qianyi Zhong[1], Lei Zhao[1], Yichao Wang[1], Hongbo Chi[2]

[1]Department of Clinical Laboratory Medicine, Taizhou Central Hospital (Taizhou University Hospital), Taizhou, China; [2]Department of Clinical Laboratory Medicine, Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, Linhai, China

Contributions: (I) Conception and design: Y Yang, Y Wang, H Chi; (II) Administrative support: Z Li; (III) Provision of study materials or patients: Q Zhong, L Zhao; (IV) Collection and assembly of data: Y Yang, Q Zhong, L Zhao, H Chi; (V) Data analysis and interpretation: Y Yang, Y Wang, H Chi; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Hongbo Chi. Department of Clinical Laboratory Medicine, Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, 150 Ximen Street, Linhai 317000, China. Email: chihb@enzemed.com; Yichao Wang. Department of Clinical Laboratory Medicine, Taizhou Central Hospital (Taizhou University Hospital), No. 999 Donghai Road, Jiaojiang District, Taizhou 318000, China. Email: wangyc9359@tzzxyy.com.

Background: Breast cancer (BRCA) is the leading cause of cancer mortality among women, and it is associated with many tumor suppressors and oncogenes. There is increasing evidence that transcription factors (TFs) play vital roles in human malignancies, but TFs-based biomarkers for BRCA prognosis were still rare and necessary. This study sought to develop and validate a prognostic model based on TFs for BRCA patients.

Methods: Differentially expressed TFs were screened from 1,109 BRCA and 113 non-tumor samples downloaded from The Cancer Genome Atlas (TCGA). Univariate Cox regression analysis was used to identify TFs associated with overall survival (OS) of BRCA, and multivariate Cox regression analysis was performed to establish the optimal risk model. The predictive value of the TF model was established using TCGA database and validated using a Gene Expression Omnibus (GEO) data set (GSE20685). A gene set enrichment analysis was conducted to identify the enriched signaling pathways in high-risk and low-risk BRCA patients. Gene Ontology (GO) function and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of the TF target genes were also conducted separately.

Results: A total of 394 differentially expressed TFs were screened. A 9-TF prognostic model, comprising PAX7, POU3F2, ZIC2, WT1, ALX4, FOXJ1, SPIB, LEF1 and NFE2, was constructed and validated. Compared to those in the low-risk group, patients in the high-risk group had worse clinical outcomes ($P<0.001$). The areas under the curve of the prognostic model for 5-year OS were 0.722 in the training cohort and 0.651 in the testing cohort. Additionally, the risk score was an independent prediction indicator for BRCA patients both in the training cohort (HR =1.757, $P<0.001$) and testing cohort (HR =1.401, $P=0.001$). It was associated with various cancer signaling pathways. Ultimately, 9 overlapping target genes were predicted by 3 prediction nomograms. The GO and KEGG enrichment analyses of these target genes suggested that the TFs in the model may regulate the activation of some classical tumor signaling pathways to control the progression of BRCA through these target genes.

Conclusions: Our study developed and validated a novel prognostic TF model that can effectively predict 5-year OS for BRCA patients.

Keywords: Breast cancer (BRCA); transcription factor (TF); prognostic model; risk score

Submitted Mar 28, 2022. Accepted for publication May 18, 2022.

doi: 10.21037/gs-22-267

View this article at: https://dx.doi.org/10.21037/gs-22-267

Introduction

Breast cancer (BRCA) is the most commonly diagnosed malignancy and the leading cause of cancer-related death in women worldwide (1,2). According to the Global Cancer Statistics 2020, there were about 2.3 million new BRCA cases and 685,000 deaths in 2020, which accounted for 24.5% of all new female malignant tumors and 15.5% of all cancer mortalities (2). Despite advancements in BRCA screening, diagnosis, and therapeutic strategies, the survival outcome of BRCA patients is not entirely satisfactory due to metastasis (3,4). A majority of patients are diagnosed at the local advanced or metastatic stage. The prognosis of these patients, who have a 5-year survival rate of 26%, is poor (5). Tumor-node-metastasis (TNM) staging system and molecular subtypes were traditional prognostic factors for BRCA patients (6). The TNM staging system has been universally used for cancer treatment, but not individualized survival prediction (7). In addition, some patients with particular subtypes have distinct clinical outcomes (8). Thus, the identification of novel signatures to improve the prognosis and clinical outcomes of such patients is essential.

Transcription factors (TFs) are critical regulatory deoxyribonucleic acid (DNA)-binding proteins, which can recognize specific DNA sequences in the promotion of different genes to control transcription and thus regulate various physiological functions, such as cell development, cell cycle controls, responses to environmental stresses, and carcinogenesis (9). There is growing evidence that TFs, as tumor suppressors or oncogenes, play a vital role in the progression, recurrence, and metastasis of human cancers (10-12). For example, nuclear factor-κB (NF-κB) is a family of critical TFs that activate the inflammatory signaling pathways, which in turn contribute to cell proliferation, invasion, angiogenesis, and metastases in various cancers, such as breast, liver, prostate, and bladder cancer (13). NF-κB is correlated with the pathogenesis of BRCA and mediates the poor prognosis of patients (14). Metastasis-associated protein 2 (MTA2) is a TF belonging to the metastasis tumor-associated family and plays a significant role in promoting the metastatic potential of tumor cells (15). A previous study has shown that MTA2 is highly expressed in triple-negative breast cancer (TNBC) tissues and serves as an oncogene to promote cell migration, invasion, and epithelial-mesenchymal transition (EMT) in TNBC (16). Although it is known that TFs participate in tumor occurrence and progression, the potential mechanism of TFs in BRCA has not been fully explored.

During the past decades, various prognostic models based on gene chips, high-throughput sequencing technology and bioinformatics techniques have been developed for BRCA. Studies have developed a microRNA (miRNA) -based signature for BRCA (17) and an immune-related long non-coding RNAs (lncRNAs) prognostic signature for TNBC (18), which provided the basis for predicting prognosis. Another study identified a novel prognostic TF signature for predicting disease-free survival (DFS) of BRCA patients (19). Nevertheless, only a few studies have systematically investigated the expression pattern, potential mechanism, and prognostic ability of TFs in BRCA. There is still a lack of prognostic models based on TFs that can predict the OS of BRCA patients.

To further expand the current knowledge of the pathogenesis and potential prognostic biomarkers for BRCA, we examined the associations between the expression profiles of TFs and clinical outcomes of BRCA patients using The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases. We developed a TF prognostic signature that we then showed to be an independent predictor of OS in BRCA patients. Our study provides an effective multi-dimensional biomarker strategy for predicting the prognosis of BRCA patients. We present the following article in accordance with the TRIPOD reporting checklist (available at https://gs.amegroups.com/article/view/10.21037/gs-22-267/rc).

Methods

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Data download and preprocessing

Public transcriptomic data and clinical data were downloaded from TCGA data portal (https://portal.gdc.cancer.gov/). TCGA data set contained 1,109 BRCA and 113 non-tumor samples. The clinical data and transcriptomic data did not correspond precisely, as the clinical data were incomplete, which led to their exclusion from subsequent analyses. A total of 1,039 BRCA patients (the training cohort) with a survival time of over 1 month were selected for further research. The GSE20685 data set (n=327) was obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=) and was used for validation. A total of 327 BRCA patients (the testing cohort) with a survival time >1 month were used for the further

Table 1 Clinical characteristics in the training and testing cohort

Characteristic	Training cohort (N=1,097)	Testing cohort (N=327)
Age or age at diagnosis (years), n (%)		
≤65	776 (70.74)	305 (93.27)
>65	321 (29.26)	22 (6.73)
Gender, n (%)		
Male	12 (1.09)	0 (0.00)
Female	1,085 (98.91)	327 (100.00)
Stage, n (%)		
Stage I	183 (16.68)	–
Stage II	621 (56.61)	–
Stage III	249 (22.70)	–
Stage IV	20 (1.82)	–
Stage X	24 (2.19)	–
T, n (%)		
T1	281 (25.62)	101 (30.89)
T2	635 (57.89)	188 (57.49)
T3	138 (12.58)	26 (7.95)
T4	40 (3.65)	12 (3.67)
TX	3 (0.27)	–
M, n (%)		
M0	912 (83.14)	319 (97.55)
M1	22 (2.01)	8 (2.45)
MX	163 (14.86)	–
N, n (%)		
N0	516 (47.04)	137 (41.90)
N1	364 (33.18)	67 (20.49)
N2	120 (10.94)	83 (25.38)
N3	77 (7.02)	40 (12.23)
NX	20 (1.82)	–

analyses. Detailed clinical information of the 2 cohorts is displayed in *Table 1*.

Identification of differentially expressed TFs

In this study, we first collected 1,639 TFs with official annotation from the previously published literature for further exploration (9). Next, the overlapping TFs between TCGA data set and the collected list of TFs were identified. An analysis of the differentially expressed TFs was then conducted using R-Limma package in R software (R version 4.0.2). An adjusted P value <0.05 and a log2 |fold change| (log2|FC|) >1 were set as the selection criteria. The pheatmap package and ggplot2 package in R software were used to make heat and volcano maps.

Construction of the prognostic risk model based on the TFs in the training cohort

A univariate Cox regression analysis was conducted to identify the TFs that were significantly correlated with OS in the training cohort. A multivariate Cox regression analysis was performed to determine the optimal TFs and construct the best prognostic model. The risk score of each patient was calculated using the following formula:

$$\text{Risk score} = \sum_{k=1}^{n} coefficient(k) \times expression(k) \quad [1]$$

where *expression* (*k*) is the expression of gene *k*, *coefficient* (*k*) is the regression coefficient of gene *k* in the multivariate Cox regression analysis, and *n* is the number of independent indicators. The median risk score of the training cohort was set as the cut-off value to divide the BRCA patients into a high-risk group and a low-risk group. Kaplan-Meier survival curves were generated to assess the difference in the OS rates between the 2 groups. The risk score and survival status curves were used to reflect the distribution of the risk scores in the high-risk and low-risk groups, and the relationship between the risk score and the survival status. The heat map shows the changes in the expression of various significant prognostic TFs in the high-risk and low-risk groups. Receiver operating characteristic (ROC) curves were used to evaluate the accuracy of the prognostic model. An area under the curve (AUC) >0.60 was considered an acceptable prediction value, and an AUC >0.75 was considered an excellent prediction value (20).

Validation of the prognostic risk model in the testing cohort

We used the GEO testing cohort to verify the accuracy of the prognostic risk model. Survival and ROC analyses were conducted to validate the model. Risk score distribution plots, survival status scatter plots, and a heat map were also used to evaluate the model.

Independent prognostic value of the model in the training cohort

Univariate and multivariate Cox proportional hazards regression analyses were conducted to assess the independence of the prognostic model. Multiple clinical characteristics, including age (the training cohort) or age at diagnosis (the testing cohort), gender, T stage, M stage, N stage, and the risk score of the prognostic model were analyzed. Factors with a P value <0.05 in both the univariate and multivariate analyses were identified as independent prognostic variables.

Correlation analysis between prognosis-related TFs and clinical characteristics in BRCA

The relationships between the TFs in the risk model and clinical characteristics were conducted using the R software. The clinical factors included age (≤65/>65 years), gender (female/male), tumor-node-metastasis (TNM) stage (I/II/III/IV), T stage (T1/T2/T3/T4), M stage (M0/M1), and N stage (N0/N1/N2/N3). Wilcoxon rank-sum testings were used for 2 sample comparisons. Kruskal-Wallis testings were used to assess differences between 3 or more groups.

Comprehensive analysis of TFs in the risk model

The prognostic value of a single TF was evaluated with the online Gene Expression Profiling Interactive Analysis (GEPIA) database (http://gepia.cancer-pku.cn/detail.php?gene=&clicktag=survival). We also analyzed the genetic alterations of the TFs using cBioPortal (http://www.cbioportal.org/). The Breast Invasive Carcinoma (TCGA, Firehose Legacy) data set containing 1,108 samples was selected. The genomic profiles included mutations, putative copy-number alterations from Genomic Identification of Significant Targets in Cancer (GISTIC), messenger ribonucleic acid (mRNA) expression z-scores relative to diploid samples (RNA Seq V2 RSEM), and protein expression z-scores (RPPA). The Human Protein Atlas (http://www.proteinatlas.org) database was used to validate the protein levels of each TFs by immunohistochemistry.

Gene set enrichment analysis (GSEA) of the TF signature

A GSEA was conducted to determine the related Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and molecular mechanisms of the high-risk and low-risk groups in the training cohort. The corresponding normalized enrichment scores (NES) in each KEGG enriched pathway were examined to determine whether the pathway was active in the high-risk group or the low-risk group. Gene sets with a | NES | >1, a P value <0.05, and a false discovery rate (FDR) <0.25 were considered significantly enriched. The representative 8 enriched KEGG pathways in the high-risk and low-risk patients with BRCA were visualized. The gene-set network was visualized using Cytoscape software (Cytoscape version 3.8.0).

Prediction of TF target genes and functional analysis

To further study the potential regulatory genes and mechanisms of the 9 TFs in the prognostic model, a series of target genes of the TFs were predicted with the algorithms from Harmonizome (https://maayanlab.cloud/Harmonizome/), including CHEA, ENCODE, JASPAR, MotifMap and TRANSFAC, and TRRUST (https://www.grnpedia.org/trrust/). We chose to integrate the results of 3 bioinformatics prediction programs. Next, the clusterProfiler package and enrichplot package in R software were used to perform statistical analysis and visualize the targets' functional profiles, including Gene Ontology (GO) and KEGG analysis. A P value <0.05 was considered statistically significant.

Statistical analysis

All the statistical analyses were performed by R software (R version 4.0.2). Univariate and multivariate Cox regression analyses were conducted to evaluate the prognostic value of the risk scores and clinical features. Survival curves were generated using the Kaplan-Meier method and compared using the log-rank testing. ROC curves and AUCs were used to evaluate the performance of the risk model. A P value <0.05 was considered statistically significant, and P<0.05 is two-sided.

Results

Identification of differentially expressed TFs

The flow chart of our study is illustrated in *Figure 1*. Among 1,639 TFs from the previously published literature, 1,564 TFs with gene expression data in TCGA data set were included in this study. In total, 394 differentially expressed TFs with an adjusted P value <0.05 and log2|FC| >1

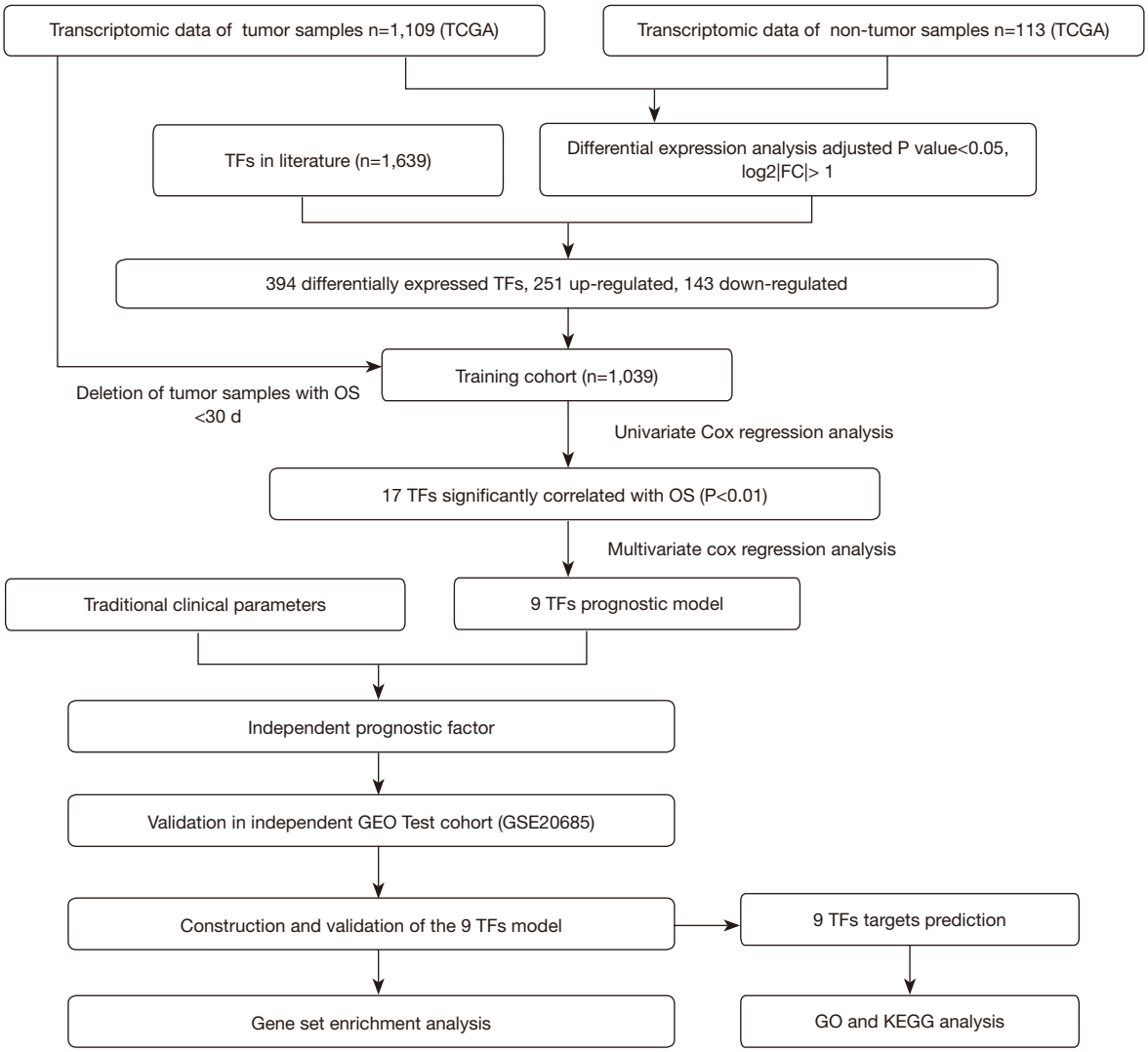

Figure 1 Flow chart of the research study. TCGA, The Cancer Genome Atlas; TF, transcription factor; OS, overall survival; GEO, Gene Expression Omnibus; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

were identified between the BRCA patients and normal controls. Of the 394 TFs, 251 were upregulated, and 143 were downregulated. The 40 representative differentially expressed TFs (20 up-regulated and 20 down-regulated) are shown in the cluster heat map (see *Figure 2A*) and the volcano map (see *Figure 2B*).

Construction and validation of a 9-TF prognostic risk model

A total of 1,039 BRCA patients with a survival time >1 month and RNA-sequencing expression profiles downloaded from TCGA were selected as the training cohort. Based on the training cohort, the univariate Cox regression analysis revealed that 17 TFs were significantly associated with the OS of the BRCA patients, including tumor protein p63 (TP63), aristaless-like homeobox4 (ALX4), forkhead box J1 (FOXJ1), odd-skipped related transcription factor 1 (OSR1), basonuclin 1 (BNC1), paired box 7 (PAX7), mesenchyme homeobox 1 (MEOX1), LIM homeobox 1 (LHX1), Spi-B transcription factor (SPIB), lymphoid enhancer binding factor 1 (LEF1), POU class 3 homeobox 2 (POU3F2), Zic family member 2 (ZIC2), wilms' tumor gene (WT1), Zic family member 5 (ZIC5), NK2 homeobox 3 (NKX2-3), nuclear factor, erythroid 2 (NFE2), and AT-rich interaction domain 5A (ARID5A)

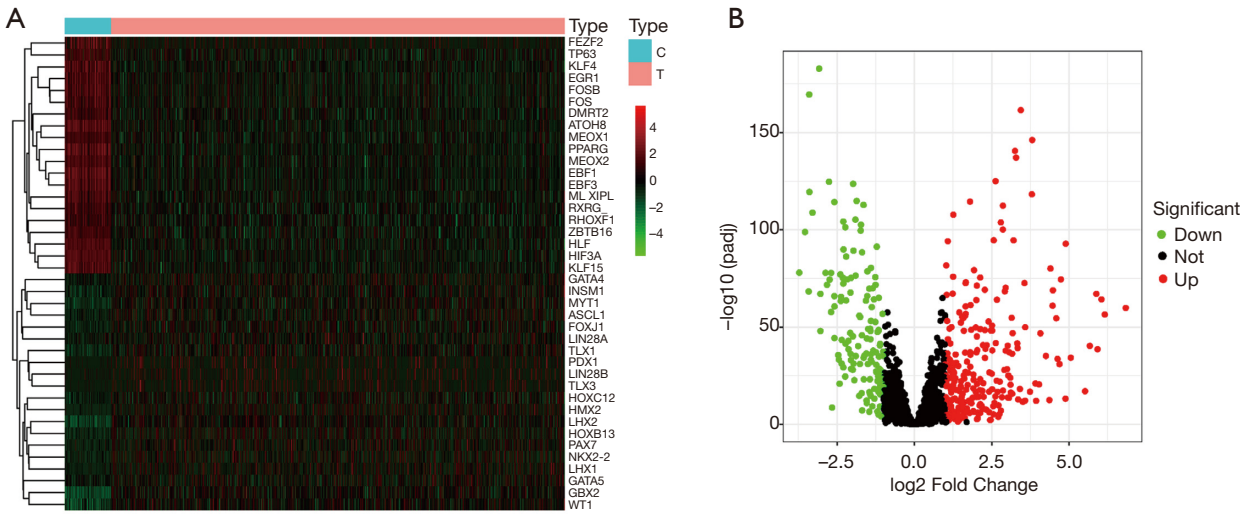

Figure 2 Differentially expressed TFs between BRCA and normal breast tissues. (A) Heat map of the top 40 significant differentially expressed TFs; (B) volcano map of differentially expressed TFs. The red dots represent the high expression of TFs in BRCA; the green dots represent the low expression of TFs in BRCA; the black dots represent TFs that are not differentially expressed. TFs, transcription factors; BRCA, breast cancer.

	P value	Hazard ratio
TP63	0.001	0.879 (0.814–0.951)
ALX4	0.006	0.898 (0.832–0.970)
FOXJ1	0.006	0.922 (0.870–0.977)
OSR1	0.005	0.896 (0.830–0.967)
BNC1	0.008	0.887 (0.812–0.969)
PAX7	<0.001	1.089 (1.037–1.144)
MEOX1	0.006	0.897 (0.830–0.969)
LHX1	0.003	1.094 (1.031–1.160)
SPIB	0.002	0.894 (0.833–0.959)
LEF1	<0.001	0.783 (0.678–0.903)
POU3F2	<0.001	1.187 (1.088–1.295)
ZIC2	0.002	1.090 (1.033–1.151)
WT1	0.002	1.105 (1.036–1.179)
ZIC5	0.009	1.082 (1.020–1.149)
NKX2-3	0.010	1.112 (1.026–1.205)
NFE2	<0.001	0.848 (0.773–0.931)
ARID5A	0.005	0.782 (0.659–0.927)

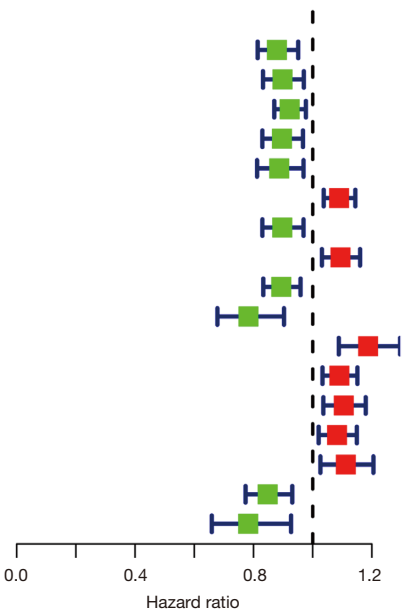

Figure 3 Forest plot of the prognosis-related TFs in BRCA. A univariate Cox regression analysis identified 17 TFs correlated with the OS of BRCA patients. The red squares indicate high-risk TFs; the green squares indicate low-risk TFs. TFs, transcription factors; BRCA, breast cancer; OS, overall survival.

(see *Figure 3*). Next, a multivariate Cox regression analysis was performed to identify the 9 optimal TFs and construct the prognostic model. In the model, the high expression of PAX7, POU3F2, ZIC2, and WT1 were associated with poor OS, while the high expression of ALX4, FOXJ1, SPIB, LEF1 and NFE2 were associated with improved

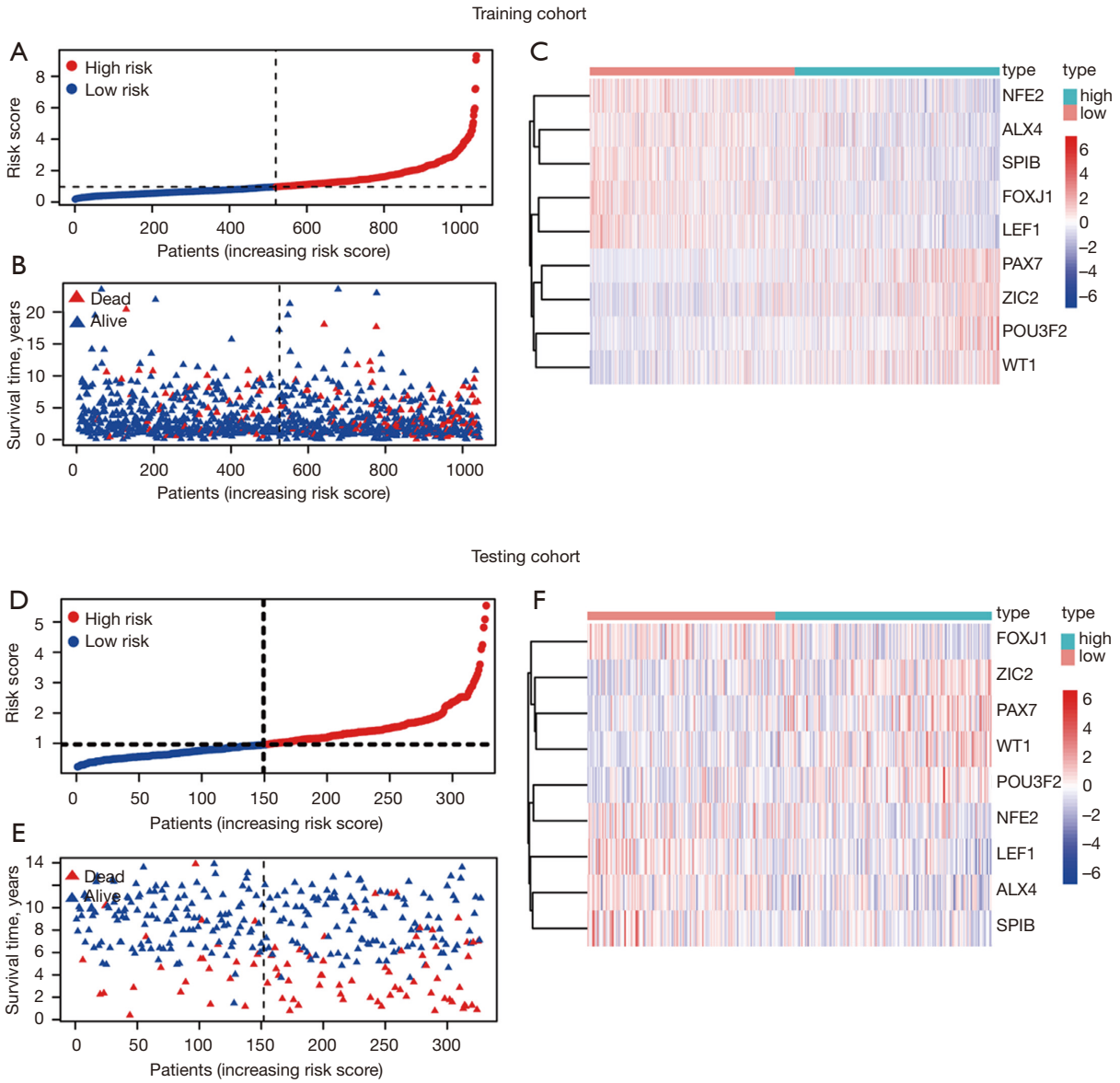

Figure 4 Construction and validation of a 9 TF-related prognostic model for BRCA. (A-C) The distributions of the risk score, survival status, and the heat map of the expression levels of the 9 TFs in the BRCA patients in the training cohort. (D-F) The distributions of the risk score, survival status, and the heat map of the expression levels of the 9 TFs in the BRCA patients in the testing cohort. TF, transcription factor; BRCA, breast cancer.

OS. The risk score was calculated as follows: risk score =0.053481 × expression value of PAX7 + 0.121020 × expression value of POU3F2 + 0.055321 × expression value of ZIC2 + 0.068781 × expression value of WT1 + (–0.074270) × expression value of ALX4 + (–0.098588) × expression value of FOXJ1 + (–0.107956) × expression value of SPIB + (–0.160061) × expression value of LEF1 + (–0.104653) × expression value of NFE2. The distribution of the risk scores and the correlations between the risk scores and the survival data are illustrated in scatterplots (see *Figure 4A,4B*). Notably, in the training cohort, the number of deaths increased and the survival time decreased as the risk score increased. The BRCA patients were divided into high-risk and low-risk groups according to the median risk score of the training cohort. The expression profiles of the prognosis-related TFs between the high-risk and low-

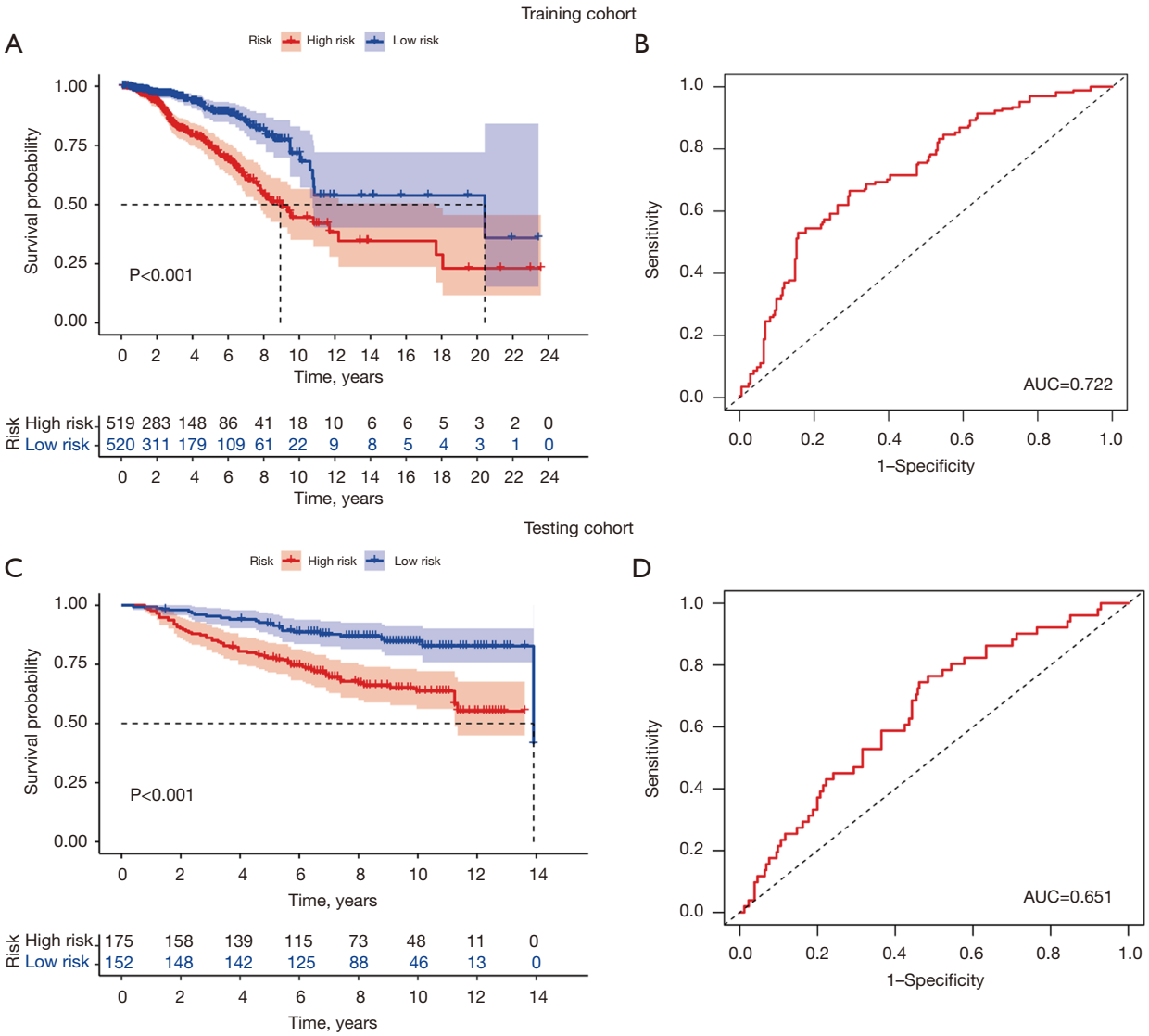

Figure 5 Kaplan-Meier survival analysis and ROC analysis of the TF-related prognostic model. (A) Kaplan-Meier survival analysis of the OS of the high-risk and low-risk patients with BRCA in the training cohort. (B) The ROC curve analysis of the risk model at 5-year time points in the training cohort. (C) Kaplan-Meier survival analysis of the OS of the high-risk and low-risk patients with BRCA in the testing cohort. (D) The ROC curve analysis of the risk model at 5-year time points in the testing cohort. AUC, area under the curve; ROC, receiver operating characteristic; TF, transcription factor; OS, overall survival; BRCA, breast cancer.

risk groups are displayed in a heat map (see *Figure 4C*). A total of 327 BRCA patients with a survival time >1 month and RNA-sequencing expression profiles from the GEO database were selected as the testing cohort. To validate the prognostic value of the risk scores, we divided the testing cohort into high-risk and low-risk groups, using the same cut-off value of the training cohort. The distribution of the risk scores and the correlation between the risk scores and the survival data are illustrated in *Figure 4D,4E*. The expression profile of the testing cohort is visualized in *Figure 4F*.

In the training cohort, the Kaplan-Meier survival analysis revealed that the survival probability in the low-risk group was significantly higher than that in the high-risk group (P<0.001; see *Figure 5A*). The AUC at 5 years of OS reached 0.722, which indicated good sensitivity and

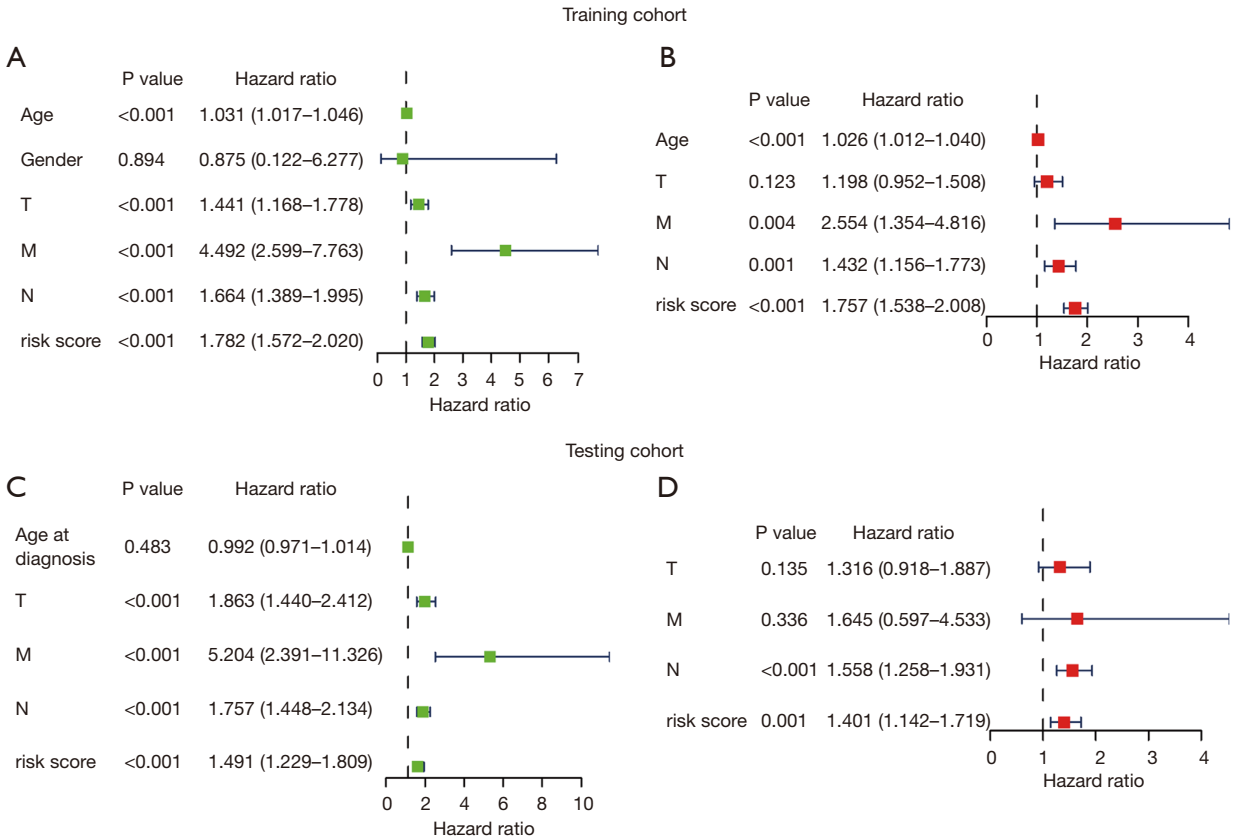

Figure 6 Independent prognosis analysis of BRCA patients. (A-B) Univariate Cox regression analysis (A) and Multivariate Cox regression analysis (B) of the clinical characteristics or risk score with the prognosis of BRCA patients in the training cohort. (C-D) Univariate Cox regression analysis and (C) and Multivariate Cox regression analysis of the clinical characteristics or risk score with the prognosis of BRCA patients in the testing cohort (D). BRCA, breast cancer.

specificity (see *Figure 5B*). In the testing cohort, patients in the high-risk group exhibited a significantly lower OS rate than low-risk group (see *Figure 5C*). The AUC of the ROC curve for 5 years of OS was 0.651 (see *Figure 5D*).

Independent prognostic value of the risk model

Univariate and multivariate Cox regression analyses were performed to evaluate the prognostic value of the risk score. The univariate Cox regression analysis revealed that the risk score [P<0.001, hazard ratio (HR) =1.782, 95% confidence interval (CI): 1.572–2.020] and clinicopathological parameters, including age (P<0.001, HR =1.031, 95% CI: 1.017–1.046), T stage (P<0.001, HR =1.441, 95% CI: 1.168–1.778), M stage (P<0.001, HR =4.492, 95% CI: 2.599–7.763), and N stage (P<0.001, HR =1.664, 95% CI: 1.389–1.995) were significantly associated with OS in the training cohort (see *Figure 6A*). The multivariate Cox regression analysis proved that the risk score was an independent prognostic variable in the training cohort (P<0.001, HR =1.757, 95% CI: 1.538–2.008; see *Figure 6B*). In the testing cohort, risk score (P<0.001, HR =1.491, 95% CI: 1.229–1.809), T stage (P<0.001, HR =1.863, 95% CI: 1.440–2.412), M stage (P<0.001, HR =5.204, 95% CI: 2.391–11.326), and N stage (P<0.001, HR =1.757, 95% CI: 1.448–2.134) were significantly associated with OS (see Figure 6C). Moreover, the risk score was also an independent prognostic variable (P=0.001, HR =1.401, 95% CI: 1.142–1.719; see *Figure 6D*).

Relationships between the 9 prognosis-related TFs and the clinicopathological factors of patients with BRCA

We investigated the correlation between the TFs in the predictive model and the clinical factors in the TCGA-BRCA patients. As *Figure 7A-7D* show, 4 of the 9 TFs were

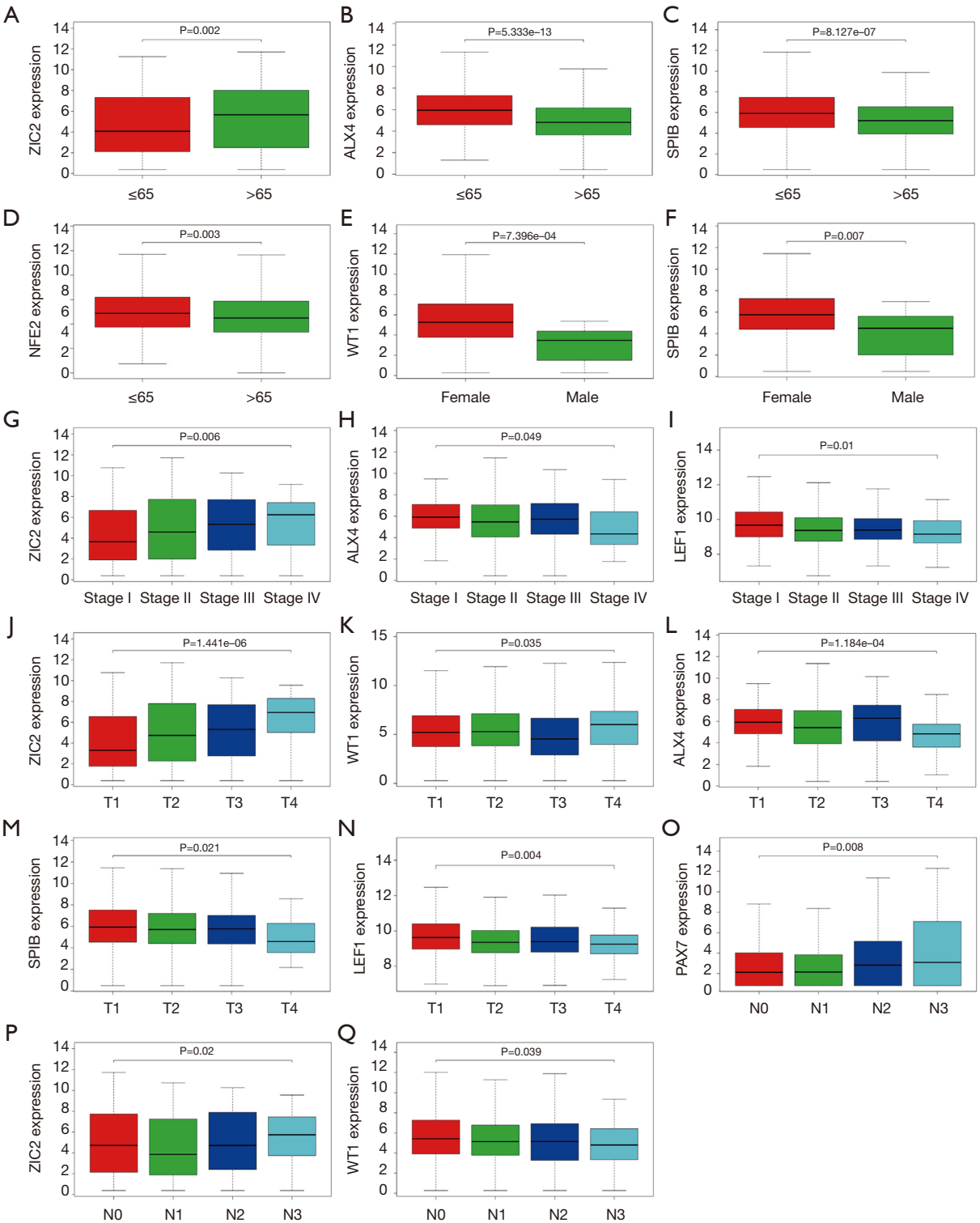

Figure 7 The relationships between the TFs in the prognostic model and clinical indicators. (A-D) Age, (E,F) gender, (G-I) stage, (J-N) T stage, (O-Q) N stage. The P values were calculated using the Wilcox-testing (for the comparison of 2 groups) and the Kruskal-Wallis testing (for the comparison of 3 or more groups). TFs, transcription factors.

Table 2 Relationships between the 9 TFs and the clinicopathological characteristics in BRCA

Gene	P value					
	Age (years) (>65/≤65)	Gender (male/female)	Stage (I/II/III/IV)	T stage (T1/T2/T3/T4)	M stage (M0/M1)	N stage (N0/N1/N2/N3)
PAX7	0.905	0.131	0.134	0.125	0.170	0.008
POU3F2	0.316	0.190	0.109	0.067	0.058	0.088
ZIC2	0.002	0.930	0.006	0.000	0.177	0.020
WT1	0.209	0.001	0.173	0.035	0.219	0.039
ALX4	5.333E-13	0.208	0.049	0.000	0.186	0.835
FOXJ1	0.295	0.626	0.441	0.717	0.689	0.642
SPIB	8.127E-07	0.007	0.426	0.021	0.623	0.260
LEF1	0.126	0.607	0.010	0.004	0.242	0.371
NFE2	0.003	0.694	0.886	0.125	0.473	0.786

The P values were calculated using the Wilcox-testing (for the comparison of 2 groups) and the Kruskal-Wallis testing (for the comparison of 3 or more groups). TFs, transcription factors; BRCA, breast cancer.

closely related to age indicators (P<0.01); that is, ZIC2, ALX4, SPIB, and NFE2. The expression levels of WT1 and SPIB were significantly more upregulated in female BRCA patients than male BRCA patients (P<0.01; see *Figure 7E*,*7F*). There were also significant correlations between the expression of TFs and the TNM stage. The results showed that ZIC2, ALX4, and LEF1 were significantly correlated to BRCA stage (P<0.05) (see *Figure 7G-7I*). Differences in the expression of ZIC2, WT1, ALX4, SPIB, and LEF1 were significantly correlated with the pathological T stage (P<0.05) (see *Figure 7J-7N*). Additionally, PAX7, ZIC2 and WTI were closely related to the pathological N stage (P<0.05) (see *Figure 7O-7Q*). However, there was no association between the expression of TFs and the pathological M stage. The relationships between the 9 TFs and all the clinicopathological factors in BRCA are set out in *Table 2*.

Comprehensive analysis of TFs in the prognostic model

The correlation between the expression of the 9 TFs and the OS of patients in BRCA was explored using the public online GEPIA database. The results showed that PAX7, POU3F2, and ZIC2 were negatively correlated with OS in BRCA, while the high expression of SPIB, LEF1 and NFE2 indicate improved OS (P<0.05; see *Figure 8*). Next, we analyzed the genetic alterations of PAX7, POU3F2, ZIC2, WT1, ALX4, FOXJ1, SPIB, LEF1, and NFE2, and found that these were altered in 36% of BRCA patients. The most prevalent alterations were mRNA high (19.67%) and amplification (10.25%) (see *Figure 9A*). Further, we investigated the expression of these TFs at the protein level through immunochemistry. As *Figure 9B* shows, the expression of WT1 and SPIB was more elevated in the BRCA tissues than the normal breast tissues. However, there were no differences in LEF1 and NFE2 at the protein level. Unfortunately, no data on other TFs were acquired from the Human Protein Atlas database.

Exploration of the signaling pathways associated with the 9-TF prognostic risk model

The GSEA revealed that the 18 KEGG pathways were enriched in the high-risk group. The top 4 pathways were cell cycle (NES =2.095, P=0.004), aminoacyl transfer RNA (tRNA) biosynthesis (NES =1.939, P=0.002), RNA degradation (NES =1.937, P<0.001), and oocyte meiosis (NES =1.929, P=0.002) (see *Figure 10A-10D*). Conversely, 12 KEGG pathways were enriched in the low-risk group. The representative 4 pathways were hematopoietic cell lineage (NES =−1.923, P=0.002), cytokine-cytokine receptor interaction (NES =−1.898, P=0.002), asthma (NES =−1.869, P=0.006), and intestinal immune network for immunoglobulin A (IgA) production (NES =−1.859, P=0.010) (see *Figure 10E-10H*). All the KEGG pathways enriched in the high-risk and low-risk groups are shown in *Table 3*. The interaction of the gene sets and the regulatory

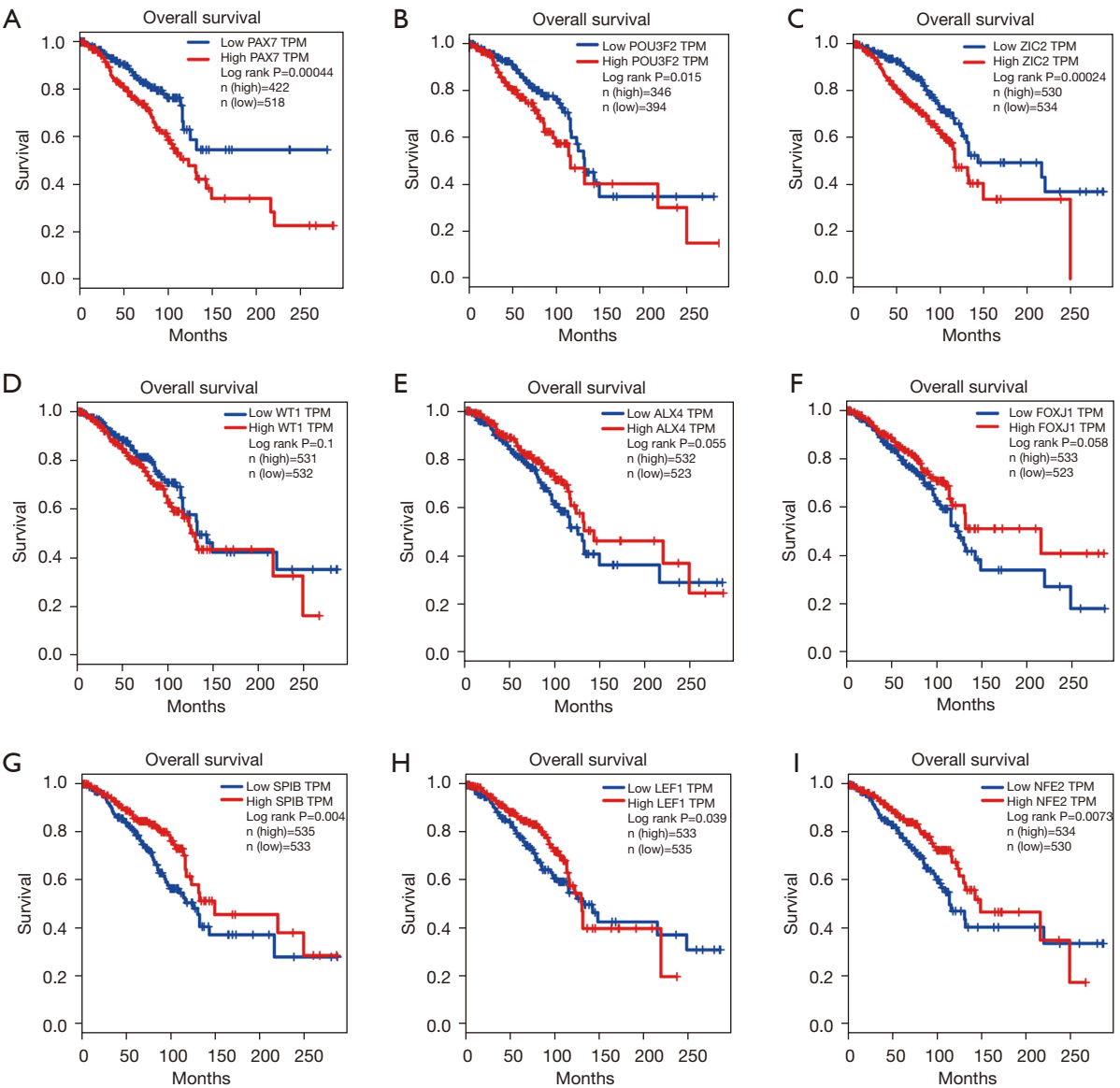

Figure 8 Kaplan-Meier survival analysis of the TFs in the prognostic model (GEPIA). Kaplan-Meier analysis of (A) PAX7, (B) POU3F2, (C) ZIC2, (D) WT1, (E) ALX4, (F) FOXJ1, (G) SPIB, (H) LEF1, and (I) NFE2. TPM, Transcripts Per Kilobase per Million mapped reads; TFs, transcription factors; GEPIA, Gene Expression Profiling Interactive Analysis.

network of the KEGG pathways in the high-risk and low-risk groups are shown in *Figure 11*.

The target-gene prediction of TFs in the prognostic model and functional annotation

A total of 9 overlapping targets of the 9 TFs in the risk model were predicted using 3 bioinformatics prediction programs. To explore the targets' potential functions and possible regulatory pathways, including PDZRN4, PODXL, FOXP2, JUN, MITF, PDGFA, CX3CL1, IGF1R and IGF2, we conducted further GO and KEGG pathway analyses. A total of 423 GO items, including 367 biological process (BP) items, 15 cellular component (CC) items, and 41 molecular function (MF) items, were identified as significantly enriched. Similarly, 22 significantly enriched KEGG items were detected. The top 30 GO items (see *Figure 12A*) and 22 KEGG items (see Figure 12B) are

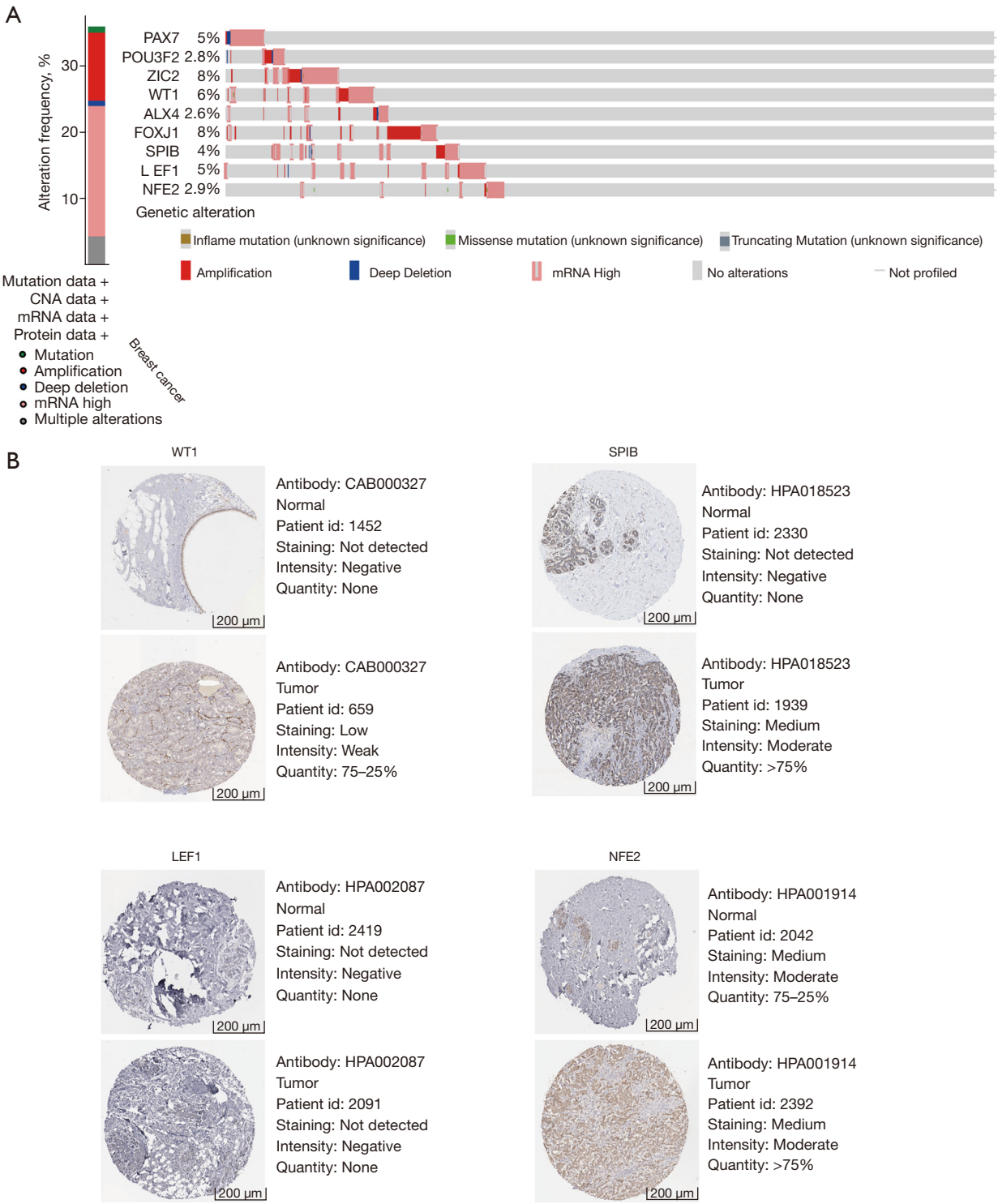

Figure 9 Genetic alterations landscape and protein levels of TFs in the prognostic model for BRCA. (A) Genetic alteration analysis of the 9 TFs in BRCA (cBioPortal). (B) The protein expression of WT1, SPIB, LEF1, and NFE2 were determined by immunohistochemistry using the indicated antibodies in the Human Protein Atlas database. TFs, transcription factors; BRCA, breast cancer; CNA, copy number variations.

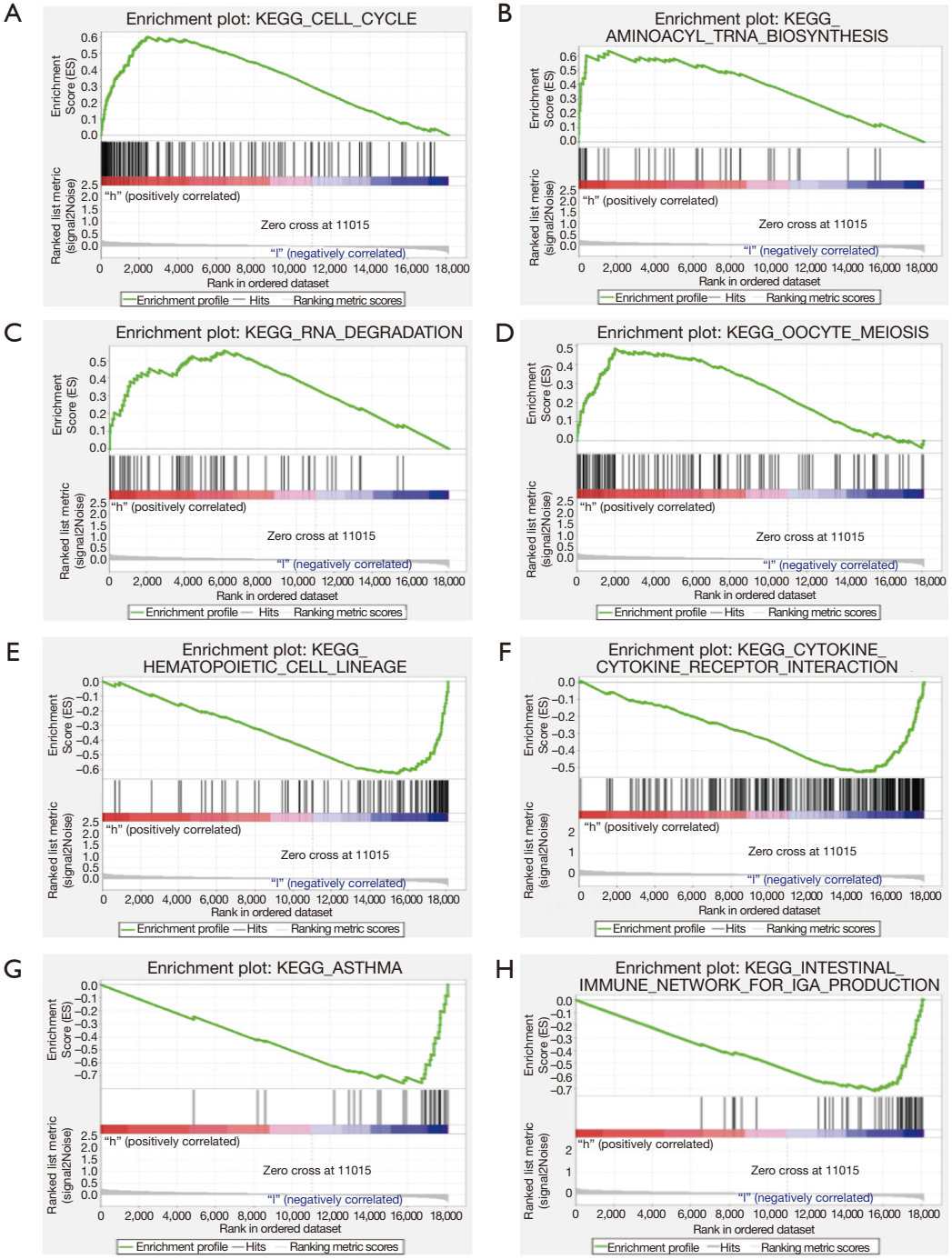

Figure 10 GSEA analysis of the representative pathways enriched in the model. (A-D) The top 4 enriched KEGG pathways in the high-risk group. (E-H) The top 4 enriched KEGG pathways in the low-risk group. KEGG, Kyoto Encyclopedia of Genes and Genomes; GSEA, Gene Set Enrichment Analysis.

Table 3 The representative enriched KEGG pathways in the high- and low-risk patients

Names	Size	ES	NES	NOM P value	FDR q-value
High-risk group					
KEGG_CELL_CYCLE	124	0.603	2.095	0.004	0.023
KEGG_AMINOACYL_TRNA_BIOSYNTHESIS	41	0.637	1.939	0.002	0.095
KEGG_RNA_DEGRADATION	57	0.556	1.937	0.000	0.066
KEGG_OOCYTE_MEIOSIS	109	0.484	1.929	0.002	0.053
KEGG_PROTEIN_EXPORT	23	0.675	1.796	0.014	0.155
KEGG_DNA_REPLICATION	36	0.670	1.775	0.024	0.155
KEGG_ONE_CARBON_POOL_BY_FOLATE	17	0.597	1.756	0.004	0.154
KEGG_UBIQUITIN_MEDIATED_PROTEOLYSIS	133	0.415	1.739	0.006	0.154
KEGG_AMINO_SUGAR_AND_NUCLEOTIDE_SUGAR_METABOLISM	44	0.511	1.724	0.010	0.153
KEGG_CYSTEINE_AND_METHIONINE_METABOLISM	33	0.508	1.724	0.006	0.138
KEGG_BASAL_TRANSCRIPTION_FACTORS	34	0.508	1.716	0.016	0.133
KEGG_CITRATE_CYCLE_TCA_CYCLE	29	0.611	1.714	0.030	0.124
KEGG_PYRIMIDINE_METABOLISM	97	0.446	1.690	0.019	0.138
KEGG_N_GLYCAN_BIOSYNTHESIS	46	0.486	1.669	0.018	0.147
KEGG_LYSINE_DEGRADATION	44	0.453	1.661	0.006	0.145
KEGG_HOMOLOGOUS_RECOMBINATION	28	0.557	1.639	0.033	0.158
KEGG_GLYOXYLATE_AND_DICARBOXYLATE_METABOLISM	16	0.562	1.563	0.031	0.227
KEGG_GLYCOLYSIS_GLUCONEOGENESIS	60	0.422	1.558	0.037	0.222
Low-risk group					
KEGG_HEMATOPOIETIC_CELL_LINEAGE	81	–0.630	–1.923	0.002	0.125
KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION	243	–0.530	–1.898	0.002	0.085
KEGG_ASTHMA	27	–0.757	–1.869	0.006	0.074
KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION	45	–0.720	–1.859	0.010	0.060
KEGG_AUTOIMMUNE_THYROID_DISEASE	37	–0.728	–1.743	0.033	0.145
KEGG_CELL_ADHESION_MOLECULES_CAMS	129	–0.513	–1.736	0.024	0.129
KEGG_ARACHIDONIC_ACID_METABOLISM	52	–0.485	–1.730	0.006	0.117
KEGG_PRIMARY_IMMUNODEFICIENCY	35	–0.711	–1.700	0.036	0.129
KEGG_GRAFT_VERSUS_HOST_DISEASE	37	–0.721	–1.676	0.048	0.136
KEGG_CYTOSOLIC_DNA_SENSING_PATHWAY	41	–0.503	–1.641	0.037	0.158
KEGG_CHEMOKINE_SIGNALING_PATHWAY	185	–0.423	–1.619	0.049	0.154
KEGG_JAK_STAT_SIGNALING_PATHWAY	137	–0.429	–1.578	0.038	0.173

KEGG, Kyoto Encyclopedia of Genes and Genomes; ES, enrichment scores; NES, normalized enrichment scores; NOM, nominal; FDR, false discovery rate.

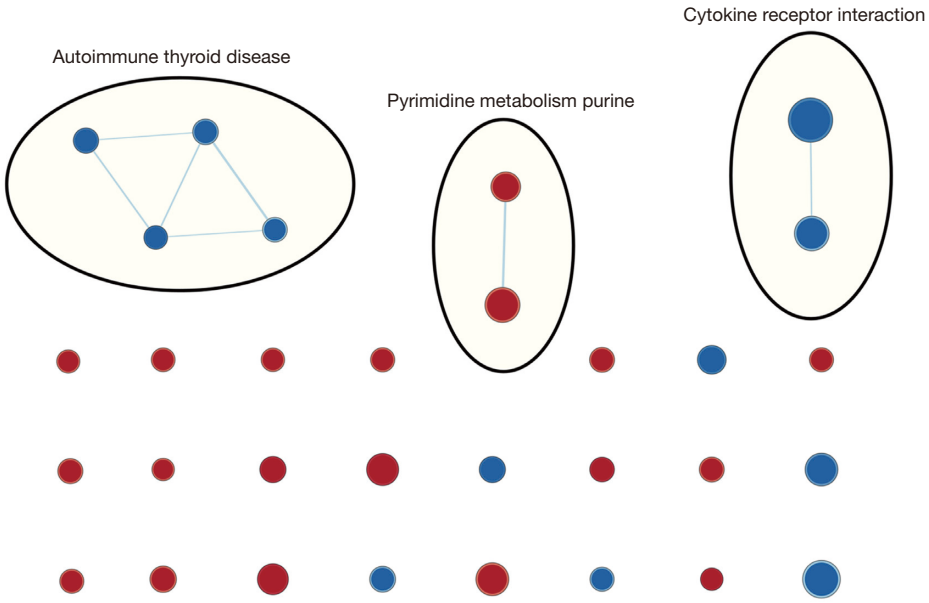

Figure 11 The regulatory network of the KEGG pathways in the high-risk and low-risk groups. The red dots indicate the KEGG pathways enriched in the high-risk patients with BRCA; the blue dots indicate the KEGG pathways enriched in the low-risk patients. The lines between the different dots indicate the same gene among them. KEGG, Kyoto Encyclopedia of Genes and Genomes; BRCA, breast cancer.

shown in *Figure 12*.

Discussion

The prognosis of BRCA varies widely due to its high genetic heterogeneity (21,22). Traditional prognostic factors, such as the TNM staging system, histological grades, and molecular subtypes, still fail to accurately predict the survival of BRCA patients (6-8). Thus, it is urgent to identify new molecular markers or establish a prognostic model for BRCA patients that is better able to predict patient prognosis and can serve as a reliable resource that explains the mechanism of prognosis, which in turn will enable the more precise treatment or lead to the cure of BRCA.

TFs play vital roles in various BPs, including tumorigenesis and drug resistance. TFs may serve as biomarkers to predict the prognosis of BRCA (23,24). Several TF-related signatures have been reported to predict the clinical outcomes of many cancers (19,25-27); however, few studies have focused on TF-related signatures that can predict OS in BRCA. In our study, we established and validated a 9-TF signature risk model that could predict OS in BRCA via a comprehensive analysis of TF data. In the training cohort, the Kaplan-Meier survival analysis revealed that patients in the high-risk group exhibited significantly worse outcomes than those in the low-risk group. The 5-year OS was quite different between the high-risk and low-risk groups. The 9-TF prognostic model showed high prediction performance and had a high AUC value of 0.722, and was validated with the GEO testing cohort (AUC =0.651). Together, these results show the practicability and reliability of the 9-TF model at predicting the OS of BRCA patients. Indeed, the predictive power of the model is promising.

The 9-TF prognostic model comprised PAX7, POU3F2, ZIC2, WT1, ALX4, FOXJ1, SPIB, LEF1, and NFE2. These genes may contribute to BRCA progression and may be potential biomarkers or therapeutic targets of BRCA. Most of these genes have been shown to be related to the tumorigenesis and metastasis of various cancers in previous studies as described below. PAX7 is a member of the paired-box (PAX) family of TFs, which contains a PAX domain that plays vital roles in postnatal skeletal muscle development, fetal development, and cancer growth (28). The function of PAX7 is context and tumor-specific. It has been reported that the high expression of PAX7 indicates shorter OS and event-free survival in pediatric and adolescent acute myeloid leukemia (AML) patients (29). PAX7 has also been shown to be an independent prognostic factor for AML patients (29).

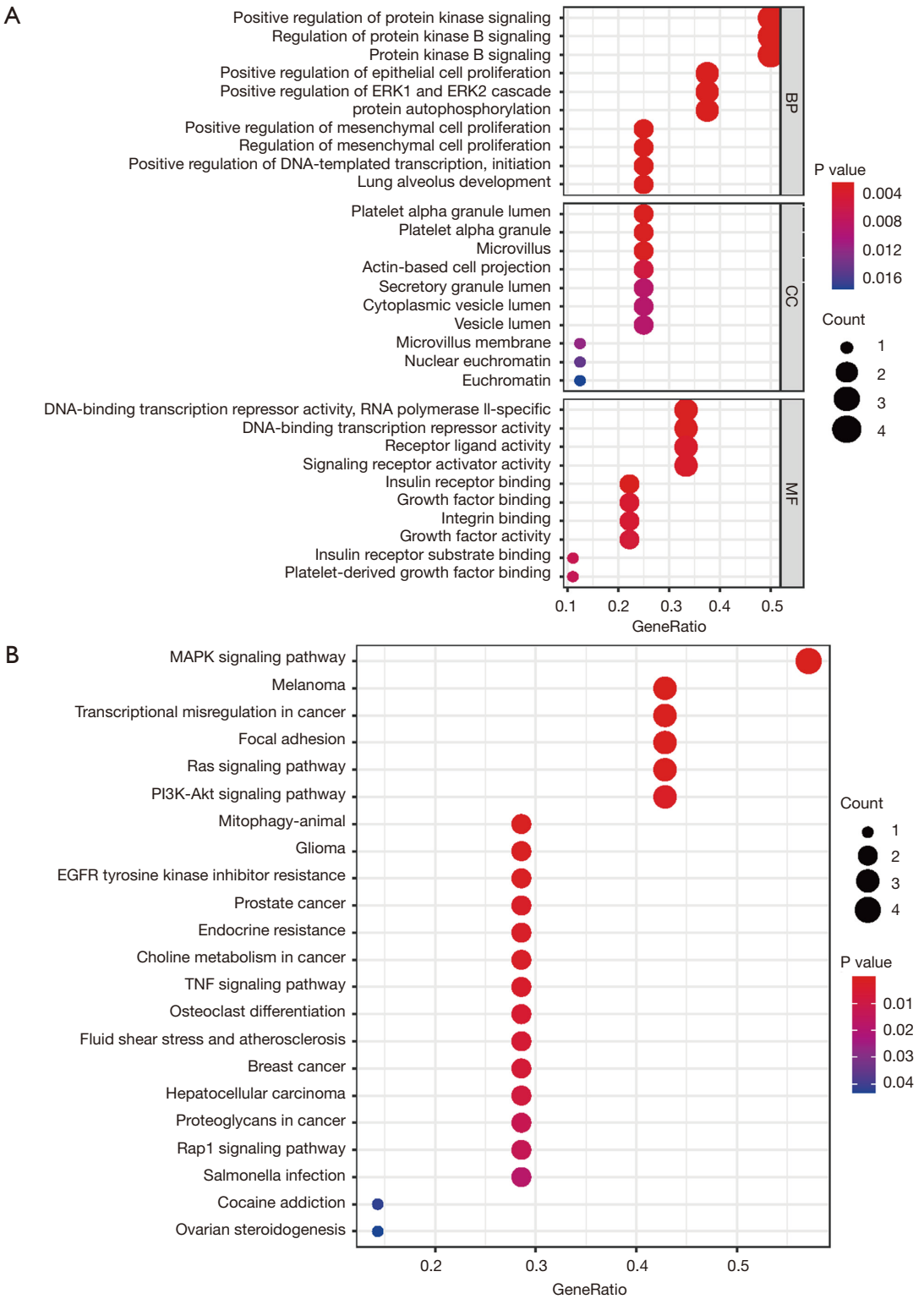

Figure 12 GO and KEGG enrichment analysis of the target genes of the 9 TFs in the model. (A) A bubble chart of the top 10 GO items, including BP, CC, and MF. (B) A bubble chart of the KEGG pathways enriched in BRCA. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; TFs, transcription factors; BP, biological process; CC, cellular component; MF, molecular functions; BRCA, breast cancer.

However, PAX7 was overexpressed in Ewing Sarcoma (ES) biopsies and cell lines and associated with the good prognosis of ES patients (30).

POU3F2 belongs to the class III POU factors, and has been reported to contribute to tumor progression and metastasis (31). Additionally, the high expression of POU3F2 in hepatocellular carcinoma patients was shown to indicate an unfavorable prognosis (32). Research suggests that ZIC2 functions as an oncogene and a prognostic biomarker for multiple solid cancers, including lung adenocarcinoma (33), hepatocellular carcinoma (34), nasopharyngeal carcinoma (35), and BRCA (36). Conversely, the downregulation of ZIC2 has been identified in BRCA and has been shown to be associated with the poor outcome of BRCA patients (37). These findings are inconsistent with our results; however, this may be due to the genetic heterogeneity of BRCA. Research has shown that WT1 is highly expressed in hematological malignancies and various solid tumors (38). The abnormal expression of WT1 was also found to be correlated with the short survival of BRCA patients (39).

ALX4 is an epigenetically downregulated tumor suppressor that inhibits BRCA progression. Additionally, it has been shown to be a favorable independent prognostic factor in BRCA (40). The high expression of FOXJ1 was found to be associated with better disease-free survival, and it may be a good prognostic factor for BRCA (41). Little is known about the role of SPIB in BRCA, but research has shown that it exerts anti-cancer effects in colorectal cancer (42). LEF1 is regarded as a potential biomarker for the metastasis of BRCA (43). Zhang *et al.* reported that NFE2 could potentially contribute to BRCA cell survival in the bone microenvironment (44). Currently, the roles of PAX7, POU3F2, and SPIB in BRCA are largely unknown; however, research suggests that they may be novel promising biomarkers for the clinical diagnosis, prognosis, and individual therapy of BRCA. Additionally, the roles of ZIC2 in BRCA are controversial and thus need to be verified by *in vitro* and *in vivo* experiments.

To further explore the function and mechanism of the prognostic signature in BRCA, we performed a GSEA analysis and constructed a regulatory network. Notably, several pathways were activated in patients with high-risk scores, such as cell cycle, aminoacyl tRNA biosynthesis, RNA degradation and oocyte meiosis. These pathways have been previously shown to be related to the progression and prognosis of tumors. Cell cycle-related genes (i.e., *CDK4*, *CCND1*, *CDKN1A*, *CDKN1C*, and *CHEK2*), which are closely related to cell cycle signaling, have been reported to be independent prognostic factors in hormone-receptor positive/human epidermal growth factor receptor 2 negative BRCA patients (45).

Aminoacyl-tRNA synthesis has been shown to be regulated by aminoacyl-tRNA synthetase and responsible for cellular protein synthesis and cell viability (46). Several studies have reported that aberrant aminoacyl tRNA synthetases (ARS)-mediated catalysis is involved in various processes of tumorigenesis (47,48). In contrast, immune-related pathways, such as asthma, the intestinal immune network for IgA production, and cytokine-cytokine receptor interactions, were activated in patients with low-risk scores. There is increasing evidence that the inflammatory reaction plays a critical role in the BPs of various cancers, such as progression, recurrence, and prognosis (49). Based on our GSEA analysis and previous studies, we speculated that TFs might be involved in the synthesis of aminoacyl-tRNA RNA damage, the immune response, and the inflammatory reactions that regulate tumor cell cycle and metabolism during the development of BRCA. Further, we predicted the target genes of 9 TFs in the prognostic model. The GO and KEGG pathway analyses provide a basis for further studies on the regulatory mechanism of hub TFs in BRCA.

It should be noted that the present study had some limitations. First, as a retrospective study, our research has some inherent bias. Second, the prognostic model must be further validated in more independent cohorts and verified by prospective clinical trials. Third, functional experiments *in vivo* and *in vitro* need to be conducted in the future to reveal the underlying mechanisms of the different expression TFs.

Conclusions

In summary, we constructed and validated a novel 9-TF prognostic model to predict the OS of patients with BRCA. The predictive model was shown to be an accurate independent prognostic factor complementary to clinicopathological factors. Additionally, our results also provide novel insights into the TFs that mediate the progression and malignant biological behavior of BRCA. In sum, our study identified novel potential prognostic biomarkers and therapeutic targets for BRCA.

Acknowledgments

Funding: This study was supported by the Natural Science

Foundation of China (grant No. 81902138), the Zhejiang Medical and Health Science and Technology Project (grant No. 2021KY1218), and the Science and Technology Plan Project of Taizhou (grant Nos. 20ywa29, 21ywb27).

Footnotes

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. available at https://gs.amegroups.com/article/view/10.21037/gs-22-267/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://gs.amegroups.com/article/view/10.21037/gs-22-267/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work, including ensuring that any questions related to the accuracy or integrity of any part of the work have been appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013)

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin 2020;70:7-30.
2. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-49.
3. Miller KD, Nogueira L, Mariotto AB, et al. Cancer treatment and survivorship statistics, 2019. CA Cancer J Clin 2019;69:363-85.
4. Esteva FJ, Hubbard-Lucey VM, Tang J, et al. Immunotherapy and targeted therapy combinations in metastatic breast cancer. Lancet Oncol 2019;20:e175-86.
5. DeSantis CE, Ma J, Gaudet MM, et al. Breast cancer statistics, 2019. CA Cancer J Clin 2019;69:438-51.
6. Nicolini A, Ferrari P, Duffy MJ. Prognostic and predictive biomarkers in breast cancer: Past, present and future. Semin Cancer Biol 2018;52:56-73.
7. Li J, Chen Z, Su K, et al. Clinicopathological classification and traditional prognostic indicators of breast cancer. Int J Clin Exp Pathol 2015;8:8500-5.
8. Rakha EA, Reis-Filho JS, Baehner F, et al. Breast cancer prognostic classification in the molecular era: the role of histological grade. Breast Cancer Res 2010;12:207.
9. Lambert SA, Jolma A, Campitelli LF, et al. The Human Transcription Factors. Cell 2018;172:650-65.
10. Sirotkin AV. Transcription factors and ovarian functions. J Cell Physiol 2010;225:20-6.
11. Safe S, Abbruzzese J, Abdelrahim M, et al. Specificity Protein Transcription Factors and Cancer: Opportunities for Drug Development. Cancer Prev Res (Phila) 2018;11:371-82.
12. Imran QM, Hussain A, Lee SU, et al. Transcriptome profile of NO-induced Arabidopsis transcription factor genes suggests their putative regulatory role in multiple biological processes. Sci Rep 2018;8:771.
13. Shin MK, Sasaki F, Ki DW, et al. Identification of Ophiocordyceps gracilioides by Its Anti-tumor Effects through Targeting the NFκB-STAT3-IL-6 Inflammatory Pathway. Biol Pharm Bull 2021;44:686-90.
14. Papila KB, Sozer V, Cigdem KP, et al. Circulating nuclear factor-kappa B mediates cancer-associated inflammation in human breast and colon cancer. J Med Biochem 2021;40:150-9.
15. Covington KR, Fuqua SA. Role of MTA2 in human cancer. Cancer Metastasis Rev 2014;33:921-8.
16. Chu J. MicroRNA-589 serves as a tumor suppressor microRNA through directly targeting metastasis-associated protein 2 in breast cancer. Oncol Lett 2019;18:2232-9.
17. Tang J, Ma W, Zeng Q, et al. Identification of miRNA-Based Signature as a Novel Potential Prognostic Biomarker in Patients with Breast Cancer. Dis Markers 2019;2019:3815952.
18. Li YX, Wang SM, Li CQ. Four-lncRNA immune prognostic signature for triple-negative breast cancer Running title: Immune lncRNAs predict prognosis of TNBC. Math Biosci Eng 2021;18:3939-56.
19. Fan C, Du J, Liu N. Identification of a Transcription Factor Signature That Can Predict Breast Cancer Survival. Comput Math Methods Med 2021;2021:2649123.
20. Chen H, Luo J, Guo J. Development and validation of a

five-immune gene prognostic risk model in colon cancer. BMC Cancer 2020;20:395.
21. Kalinowski L, Saunus JM, McCart Reed AE, et al. Breast Cancer Heterogeneity in Primary and Metastatic Disease. Adv Exp Med Biol 2019;1152:75-104.
22. Holm J, Eriksson L, Ploner A, et al. Assessment of Breast Cancer Risk Factors Reveals Subtype Heterogeneity. Cancer Res 2017;77:3708-17.
23. Wang L, Wang Y, Su B, et al. Transcriptome-wide analysis and modelling of prognostic alternative splicing signatures in invasive breast cancer: a prospective clinical study. Sci Rep 2020;10:16504.
24. Sun CC, Li SJ, Hu W, et al. Comprehensive Analysis of the Expression and Prognosis for E2Fs in Human Breast Cancer. Mol Ther 2019;27:1153-65.
25. Liu J, Dong C, Jiang G, et al. Transcription factor expression as a predictor of colon cancer prognosis: a machine learning practice. BMC Med Genomics 2020;13:135.
26. Li H, Wu N, Liu ZY, et al. Development of a novel transcription factors-related prognostic signature for serous ovarian cancer. Sci Rep 2021;11:7207.
27. Yang X, Cheng Y, Li X, et al. A Novel Transcription Factor-Based Prognostic Signature in Endometrial Cancer: Establishment and Validation. Onco Targets Ther 2021;14:2579-98.
28. Dulak J. Many roles for Pax7. Cell Cycle 2017;16:21-2.
29. Yan T, Naren D, Gong Y. Positive expression of PAX7 indicates poor prognosis of pediatric and adolescent AML patients. Expert Rev Hematol 2020;13:289-97.
30. Ribeiro-Dantas MDC, Oliveira Imparato D, Dalmolin MGS, et al. Reverse Engineering of Ewing Sarcoma Regulatory Network Uncovers PAX7 and RUNX3 as Master Regulators Associated with Good Prognosis. Cancers (Basel) 2021;13:1860.
31. Zhao G, Wei Z, Guo Y. MicroRNA-107 is a novel tumor suppressor targeting POU3F2 in melanoma. Biol Res 2020;53:11.
32. Ding S, Jin Y, Hao Q, et al. LncRNA BCYRN1/miR-490-3p/POU3F2, served as a ceRNA network, is connected with worse survival rate of hepatocellular carcinoma patients and promotes tumor cell growth and metastasis. Cancer Cell Int 2020;20:6.
33. Wei-Hua W, Ning Z, Qian C, et al. ZIC2 promotes cancer stem cell traits via up-regulating OCT4 expression in lung adenocarcinoma cells. J Cancer 2020;11:6070-80.
34. Lu SX, Zhang CZ, Luo RZ, et al. Zic2 promotes tumor growth and metastasis via PAK4 in hepatocellular carcinoma. Cancer Lett 2017;402:71-80.
35. Yi W, Wang J, Yao Z, et al. The expression status of ZIC2 as a prognostic marker for nasopharyngeal carcinoma. Int J Clin Exp Pathol 2018;11:4446-60.
36. Zhang P, Yang F, Luo Q, et al. miR-1284 Inhibits the Growth and Invasion of Breast Cancer Cells by Targeting ZIC2. Oncol Res 2019;27:253-60.
37. Liu ZH, Chen ML, Zhang Q, et al. ZIC2 is downregulated and represses tumor growth via the regulation of STAT3 in breast cancer. Int J Cancer 2020;147:505-18.
38. Hastie ND. Wilms' tumour 1 (WT1) in development, homeostasis and disease. Development 2017;144:2862-72.
39. Zhang Y, Yan WT, Yang ZY, et al. The role of WT1 in breast cancer: clinical implications, biological effects and molecular mechanism. Int J Biol Sci 2020;16:1474-80.
40. Yang J, Han F, Liu W, et al. ALX4, an epigenetically down regulated tumor suppressor, inhibits breast cancer progression by interfering Wnt/β-catenin pathway. J Exp Clin Cancer Res 2017;36:170.
41. Zhou X, Xiao C, Han T, et al. Prognostic biomarkers related to breast cancer recurrence identified based on Logit model analysis. World J Surg Oncol 2020;18:254.
42. Zhao X, Li L, Yuan S, et al. SPIB acts as a tumor suppressor by activating the NFkB and JNK signaling pathways through MAP4K1 in colorectal cancer cells. Cell Signal 2021;88:110148.
43. Zheng T, Wang A, Hu D, et al. Molecular mechanisms of breast cancer metastasis by gene expression profile analysis. Mol Med Rep 2017;16:4671-7.
44. Zhang D, Iwabuchi S, Baba T, et al. Involvement of a Transcription factor, Nfe2, in Breast Cancer Metastasis to Bone. Cancers (Basel) 2020;12:3003.
45. Lai J, Chen B, Li Y, et al. Integrated analysis of cell cycle-related genes in HR+/HER2- breast cancer. Breast Cancer 2022;29:121-30.
46. Gao X, Guo R, Li Y, et al. Contribution of upregulated aminoacyl-tRNA biosynthesis to metabolic dysregulation in gastric cancer. J Gastroenterol Hepatol 2021;36:3113-26.
47. He Y, Gong J, Wang Y, et al. Potentially functional polymorphisms in aminoacyl-tRNA synthetases genes are associated with breast cancer risk in a Chinese population. Mol Carcinog 2015;54:577-83.
48. Zhou Z, Sun B, Nie A, et al. Roles of Aminoacyl-tRNA Synthetases in Cancer. Front Cell Dev Biol

2020;8:599765.
49. Singh R, Mishra MK, Aggarwal H. Inflammation, Immunity, and Cancer. Mediators Inflamm 2017;2017:6027305.

(English Language Editor: L. Huleatt)

Cite this article as: Yang Y, Li Z, Zhong Q, Zhao L, Wang Y, Chi H. Identification and validation of a novel prognostic signature based on transcription factors in breast cancer by bioinformatics analysis. Gland Surg 2022;11(5):892-912. doi: 10.21037/gs-22-267