Hindawi

Research Article

Establishment and Analysis of an Individualized EMT-Related Gene Signature for the Prognosis of Breast Cancer in Female Patients

Wei Xue [ID],[1,2,3] **Chenyu Sun** [ID],[4] **Hui Yuan,**[5] **Xin Yang,**[2] **Qiuping Zhang,**[2,6] **Yunnuo Liao,**[1,2] **and Hongwei Guo** [ID][1,2]

[1]*Guangxi Key Laboratory for Research and Evaluation of Bioactive Molecules& College of Pharmacy, Guangxi Medical University, Nanning 530021, China*
[2]*Key Laboratory of Longevity and Aging-Related Diseases of Chinese Ministry of Education & Center for Translational Medicine, Guangxi Medical University, Nanning 530021, China*
[3]*Department of Pharmacy, Ruikang Hospital Affiliated to Guangxi University of Chinese Medicine, Nanning 530011, China*
[4]*AMITA Health Saint Joseph Hospital Chicago, 2900 N. Lake Shore Drive, Chicago, 60657 Illinois, USA*
[5]*Public Health Center, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, 710061, China*
[6]*The First Affiliated Hospital of Guangxi Medical University, Guangxi Medical University, Nanning 530021, China*

Correspondence should be addressed to Hongwei Guo; hongweiguo@gxmu.edu.cn

Background. The current high mortality rate of female breast cancer (BC) patients emphasizes the necessity of identifying powerful and reliable prognostic signatures in BC patients. Epithelial-mesenchymal transition (EMT) was reported to be associated with the development of BC. The purpose of this study was to identify prognostic biomarkers that predict overall survival (OS) in female BC patients by integrating data from TCGA database. *Method*. We first downloaded the dataset in TCGA and identified gene signatures by overlapping candidate genes. Differential analysis was performed to find differential EMT-related genes. Univariate regression analysis was then performed to identify candidate prognostic variables. We then developed a prognostic model by multivariate analysis to predict OS. Calibration curves, receiver operating characteristics (ROC) curves, *C*-index, and decision curve analysis (DCA) were used to test the veracity of the prognostic model. *Result*. In this study, we identified and validated a prognostic model integrating age and six genes (CD44, P3H1, SDC1, COL4A1, TGFβ1, and SERPINE1). *C*-index values for BC patients were 0.672 (95% CI 0.611–0.732) and 0.692 (95% CI 0.586–0.798) in the training cohort and test set, respectively. The calibration curve and the DCA curve show the good predictive performance of the model. *Conclusion*. This study offered a robust predictive model for OS prediction in female BC patients and may provide a more accurate treatment strategy and personalized therapy in the future.

1. Introduction

Breast cancer is one of the most prevalent malignancies in women worldwide and the leading cause of most cancer-related deaths, although early-stage BC is considered curable [1, 2]. In 2018, BC was the most commonly diagnosed cancer (24.2% of all cancer cases) and the leading cause of cancer-related deaths (15% of all cancer deaths) in women worldwide. Among these, metastatic BC accounted for more than 90% of BC-related deaths [3]. At present, the main treatment strategies for BC include surgery, chemotherapy, radiotherapy, immunotherapy, and hormonal therapy [4]. Although nanomedicine has been developed this year to target progesterone and estrogen receptors (PR and ER), human epidermal growth factor receptor 2 (HER2), and microRNA (miRNAs) and long chain non-coding RNAs, the incidence of BC remains high, with previous studies suggesting that the number of new cases worldwide will be
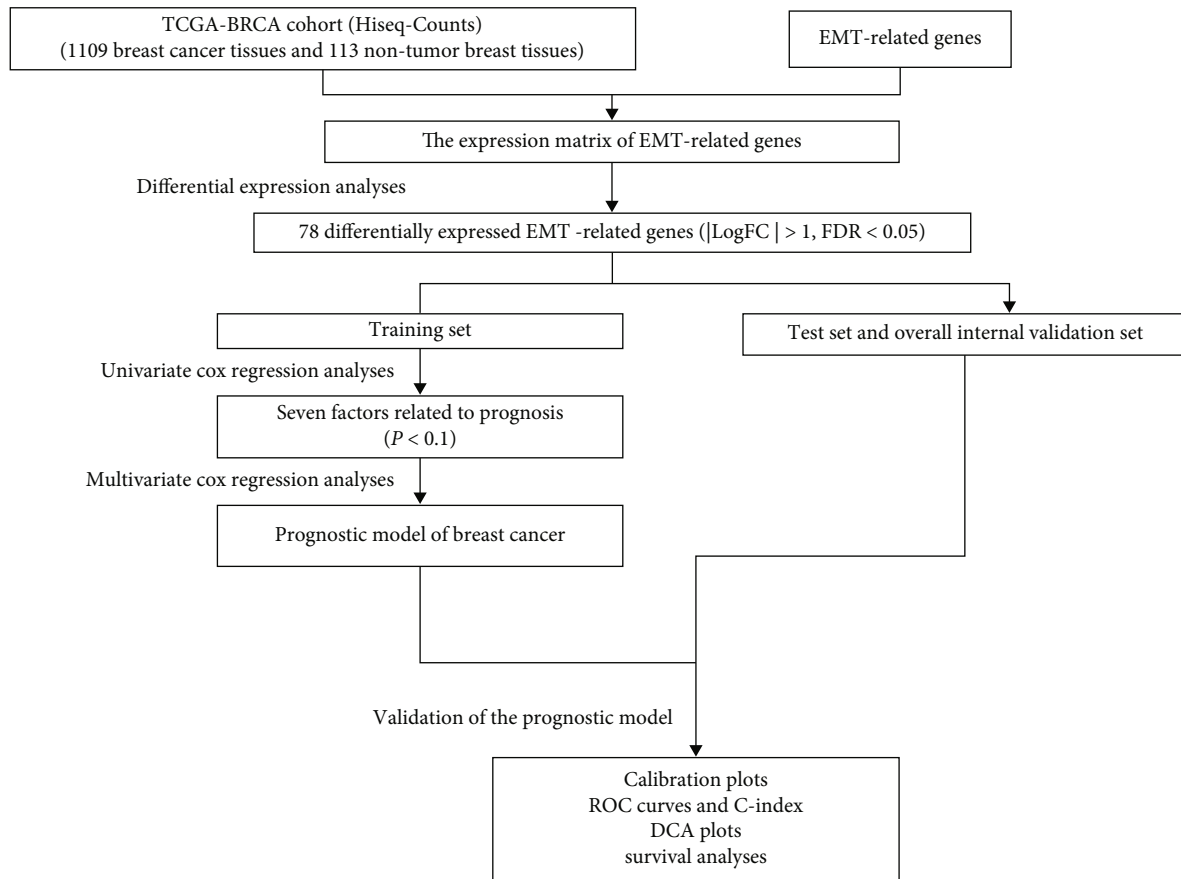
Figure 1: Flow chart of this study.

2,261,419 women in 2020, and this number is expected to increase to 30.2 million by 2040 [5, 6].

Epithelial-mesenchymal transition (EMT) is widely known to occur during mammalian development, wound healing, and cancer metastasis [7]. In recent years, EMT has received increasing attention for its role in cancer drug resistance [8]. Many studies have shown that EMT is associated with tumorigenesis, invasion, metastasis, and resistance to treatment, especially in BC [9, 10]. Saotome et al. demonstrated that GATA3 truncation mutants affected ductal BC development by altering EMT-related gene expression through partial motif recognition in luminal BC cells [11]. Parthasarathi and his colleagues found that EMT-related genes were associated with dysregulated ion channels in BC-associated tumorigenesis and could potentially be used to determine the prognosis of BC patients. Therefore, in this study, we evaluated the relevance of the EMT genes in female BC patients to explore the mechanisms of EMT in BC [12].

The new 8th edition of a related Union for International Cancer Control (UICC) and American Joint Committee on Cancer (AJCC) publication updates the description of BC staging for tumor lymph node metastasis (TNM) [13]. Yet, it is not sufficient to simply predict the prognosis of BC based on the TNM staging system. Some of the factors that influence BC include age, genes, reproductive factors, estrogen, and lifestyle [14]. Hence, a multifactorial predictive

model is essential. Predictive modeling is a more advanced approach as it can be visualized using a nomogram and it can estimate individualized risk based on a more comprehensive set of gene signatures and clinical characteristics. In previous studies, we constructed a clinical prediction model based on the clinical data of metastatic colon cancer patients extracted from the SEER database. The nomogram developed with high prognosis prediction accuracy to evaluate the 1-, 3-, and 5-year survival of metastatic colon cancer patients, which will help clinical decision-making of metastatic colon cancer patients after surgery and individualized treatment [15].

In this study, we identified EMT-related genes with independent prognostic value to establish a prognostic model for predicting the overall survival (OS) at 1-, 3-, and 5-year of female BC patients and generating new insights about BC progression.

2. Materials and Methods

2.1. Data Collection.
We downloaded gene expression data from The Cancer Genome Atlas (TCGA) database (https://cancergenome.nih.gov) of 1109 BC patients and 113 nontumor breast tissues. Clinical data were also acquired, but the clinical data of 12 male BC patients were removed because the study population in this paper was female. EMT-related genes were collected from the Molecular Signature
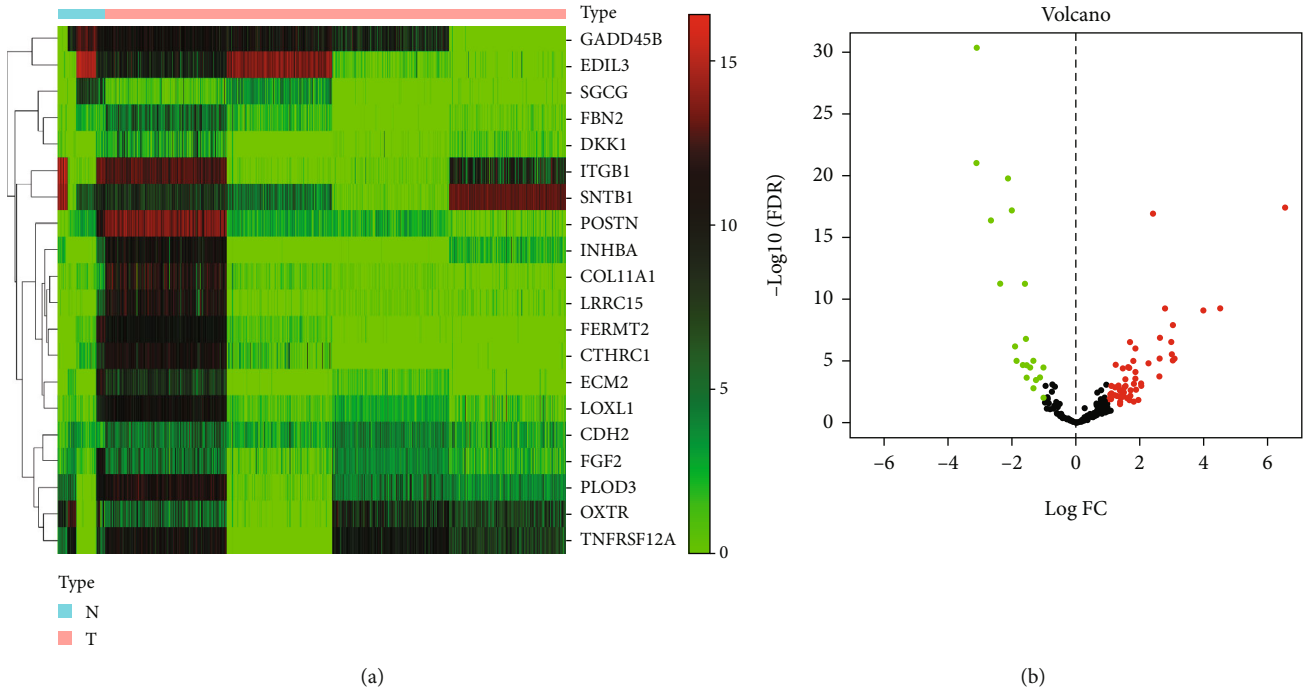
(a)

(b)

FIGURE 2: Heat map (a) and volcano map (b) of differentially expressed gene related to EMT.

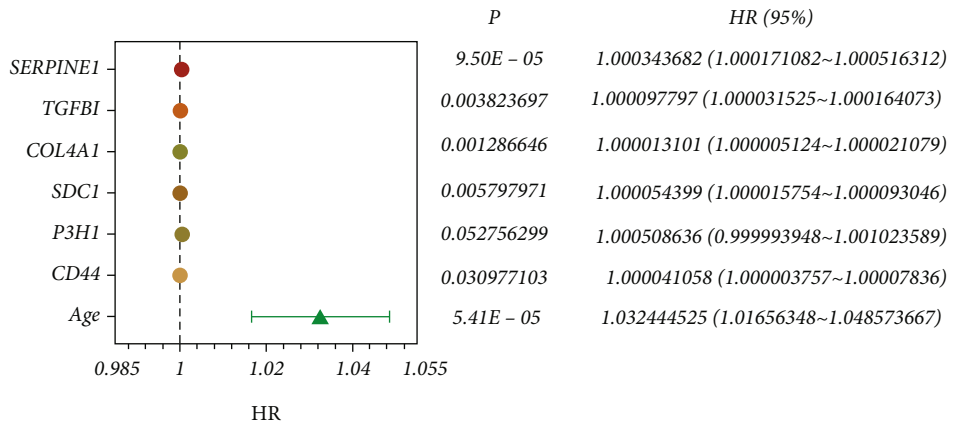|	P	HR (95%)
SERPINE1	9.50E − 05	1.000343682 (1.000171082~1.000516312)
TGFBI	0.003823697	1.000097797 (1.000031525~1.000164073)
COL4A1	0.001286646	1.000013101 (1.000005124~1.000021079)
SDC1	0.005797971	1.000054399 (1.000015754~1.000093046)
P3H1	0.052756299	1.000508636 (0.999993948~1.001023589)
CD44	0.030977103	1.000041058 (1.000003757~1.00007836)
Age	5.41E − 05	1.032444525 (1.01656348~1.048573667)

FIGURE 3: Forest plot analyzed by univariate Cox regression.

TABLE 1: Genes contained in the prognostic model of breast cancer.

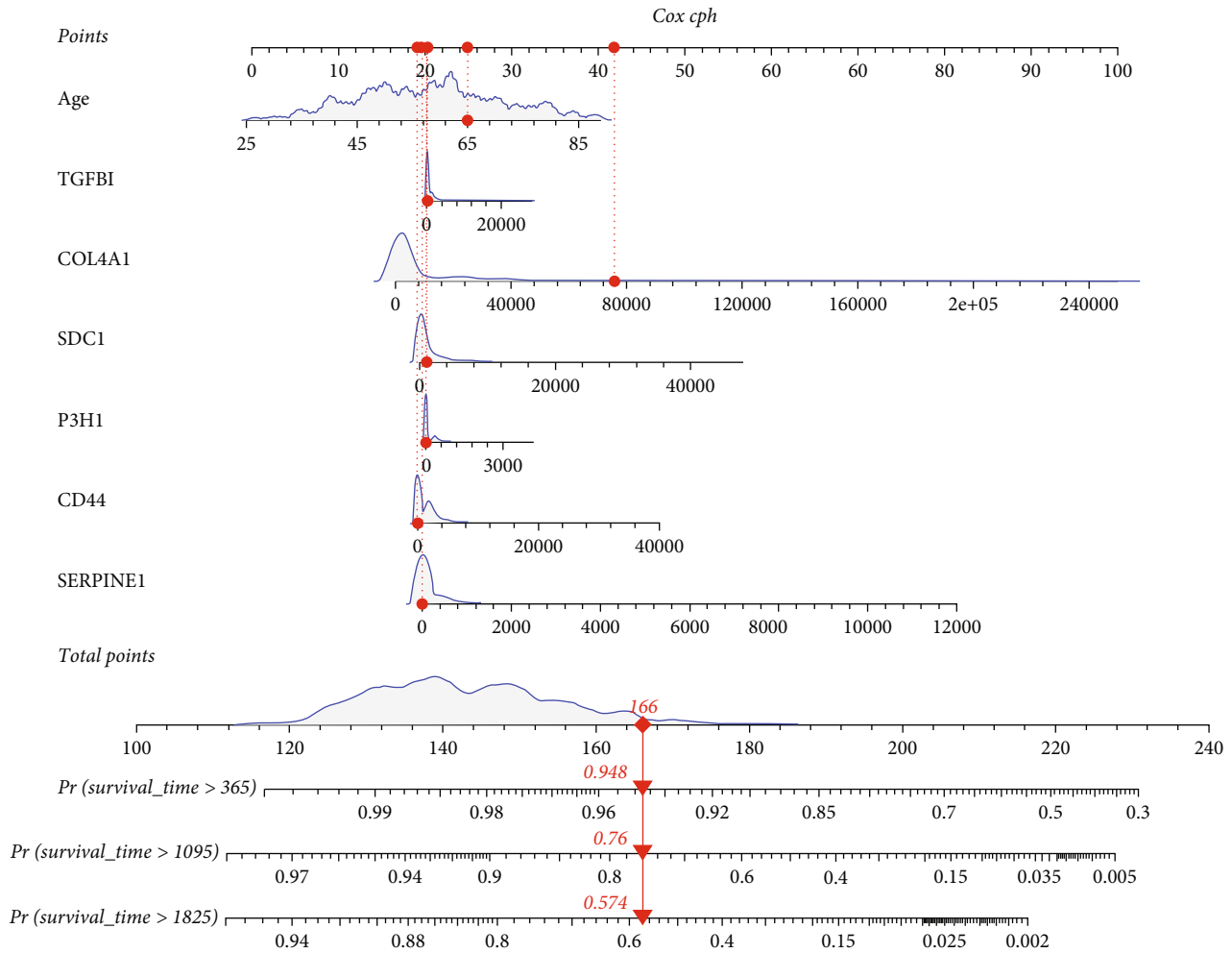| Factors	coef	HR	HR_95L	HR_95U	P
Age	0.030655	1.031129	1.015051	1.047463	0.000132
CD44	3.35E − 05	1.000033	0.999994	1.000073	0.098155
P3H1	0.000142	1.000142	0.999283	1.001002	0.74605
SDC1	3.76E − 05	1.000038	0.999966	1.000109	0.304272
COL4A1	1.60E − 05	1.000016	1.000008	1.000024	0.000149
TGFBI	2.04E − 05	1.00002	0.999884	1.000156	0.768605
SERPINE1	0.000247	1.000247	0.999911	1.000583	0.149765

FIGURE 4: Nomogram for predicting 1-, 3-, and 5-year overall survival (OS) for BC patients in the training cohort.

database v7.1 (MSigDB) (http://www.broad.mit.edu/gsea/msigdb/).

2.2. Identification of Differentially Expressed EMT-Related Genes.

Combining the gene expression data obtained from TCGA database with EMT-related genes by using the "edgeR" package of R software, the expression data of the target genes could be obtained. After that, the "limma" package was utilized to derive differentially expressed EMT-related genes according to False Discovery Rate (FDR) values less than 0.05 and the absolute value of fold change above 1.

2.3. Statistical Analysis

2.3.1. Univariate Cox Regression Analysis for Independent Prognostic Factors.

The expression matrix of the obtained EMT-related genes was further analyzed by incorporating the matrix with the survival time and survival status. Based on previous studies, age had an impact on the prognosis of female BC patients, so we included age as a study variable [16]. Using the "caret" package in R software (version 4.1.0) to randomly divide the overall cohort into two groups in the ratio of 7 : 3. The subgroup containing 70% of female

BC patients was used to construct the prediction model, while the remaining 30% of patients were examined for the accuracy and reliability of the model. Also, the whole cohort was used as the overall internal validation set. The basic values of patients were listed (Table S1).

Univariate Cox regression analysis was used to screen for independent prognostic factors. Factors with a cutoff value of $P < 0.1$ were defined as candidates associated with OS.

2.3.2. Prognostic Nomogram Construction.

The genes filtered by univariate Cox regression were then analyzed in the multivariate Cox regression for the risk scoring model. The risk score for each patient can be calculated by the following formula: risk score $= \text{Exp}(x1) * \beta1 + \text{Exp}(x2) * \beta2 + \cdots + \text{Exp}(xn) * \beta n$, where n is the number of selected variables, Exp is the expression level of the variable, and β is the regression coefficient of the variable. Then, the nomogram was developed using R software. According to the scores calculated from the nomogram, the patient's OS at 1, 3, and 5 years can be predicted. Subsequently, according to the median risk scores, patients with risk scores greater than the median value were divided into the high-risk group and otherwise into the low-risk group.
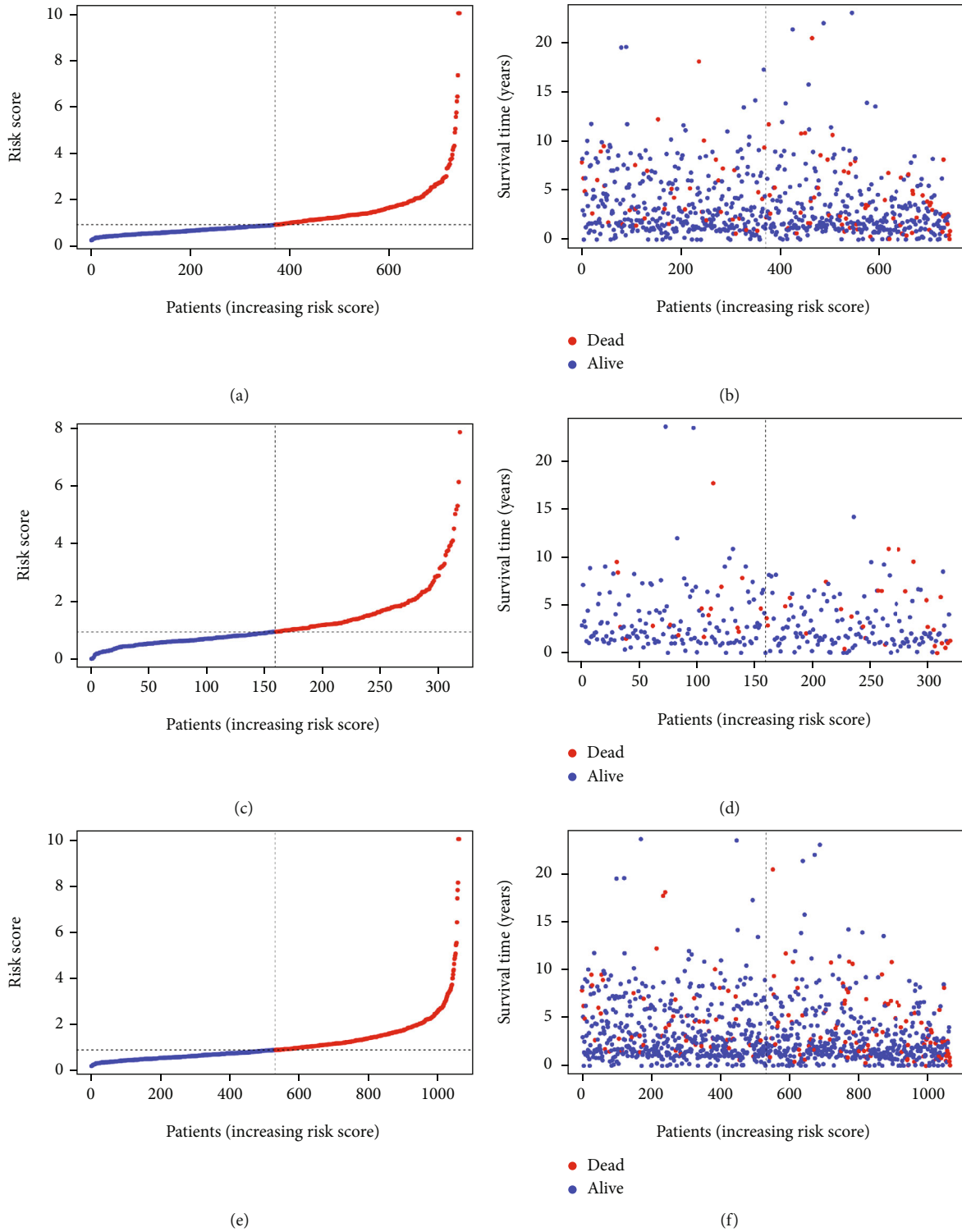
(a)

(b)

(c)

(d)

(e)

(f)

FIGURE 5: (a, c, e) Distribution of risk score in patients with BC. The black dotted line serves as the dividing line between the high-risk group and the low-risk group. (b, d, f) Diagram of the relationship between risk score and patient survival time. The result of (a, b) is based on training set, the result of (c, d) is based on test set, and the result of (e, f) is based on the overall internal validation set.

2.3.3. Prognostic Nomogram Evaluation and Validation. In order to improve the reliability of the prediction model and thus its clinical application, 30% of the patients and the overall cohort were used as an internal validation cohort to test the validity of the prediction model.

The discriminative power of the nomogram was calculated using the concordance index (C-index). We also measured the area under the curve (AUC) at 1, 3, and 5 years, which was derived from a ROC analysis. The C-indexes and AUCs take values ranging from 0.5 to 1.0, where 1.0
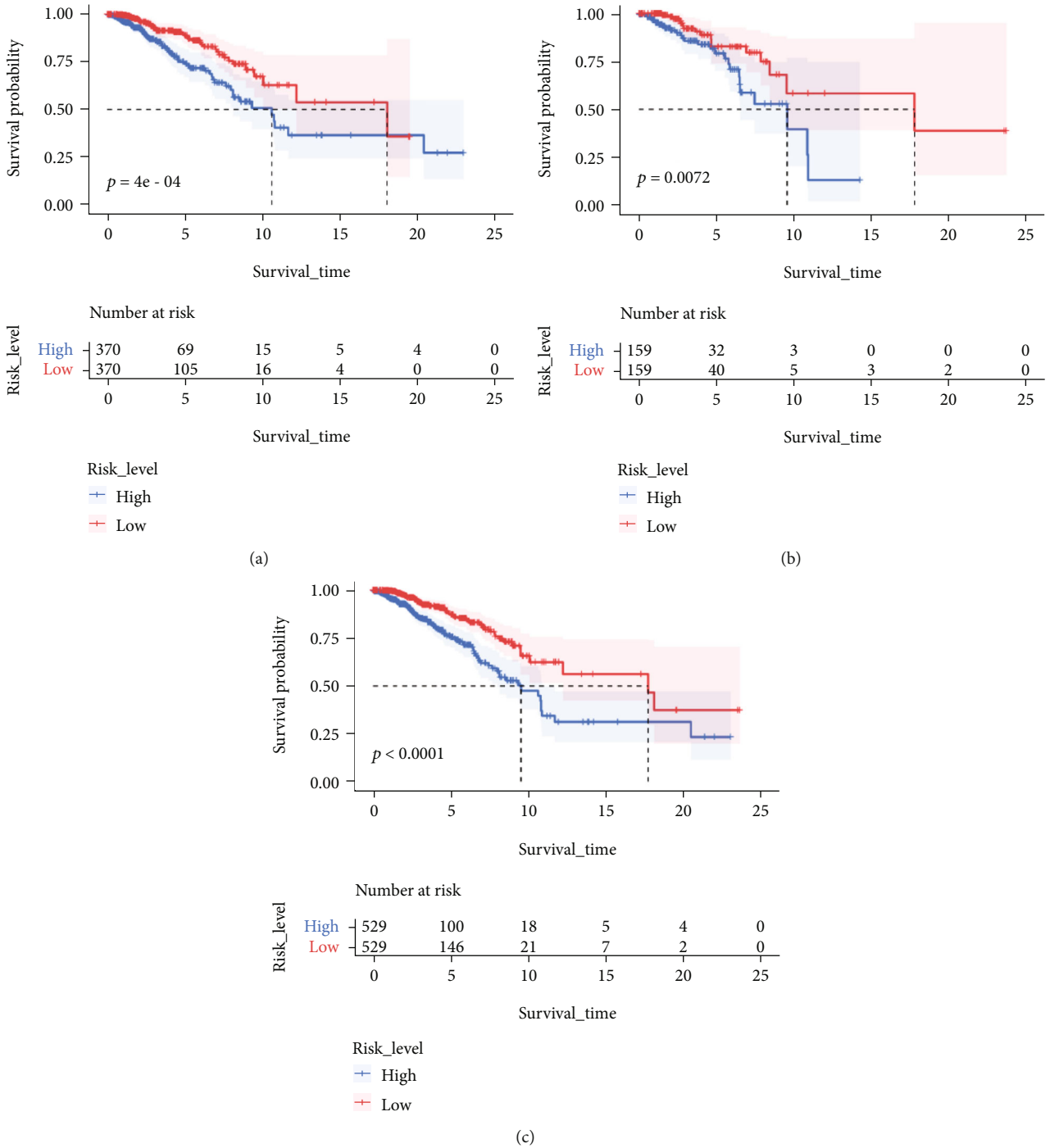
(a)

(b)

(c)

FIGURE 6: Overall survival (OS) Kaplan-Meier curves for patients in the low- and high-risk groups: (a) training set; (b) test set; (c) overall internal validation set.

represents the perfect ability to correctly distinguish the results from the model and 0.5 represents random chance. The calibration curve of the nomogram was evaluated graphically by plotting the ratio of the predicted probability to the observed ratio of the nomogram. Overlapping with the reference line indicated that the model was perfectly consistent. Finally, decision curve analysis was performed to evaluate the clinical benefits. A flow chart of the study process of this article was presented (Figure 1).

3. Results

3.1. Identification of Differentially Expressed EMT-Related Genes.
To describe our study more clearly, we developed a flowchart of the analysis procedure. First, we obtained data from TCGA database for 1109 tumor tissues and 113 nontumor tissues. After taking intersection with EMT-related genes, a matrix of 200 EMT-related genes (Table S2) expression values was acquired. Then, after differential
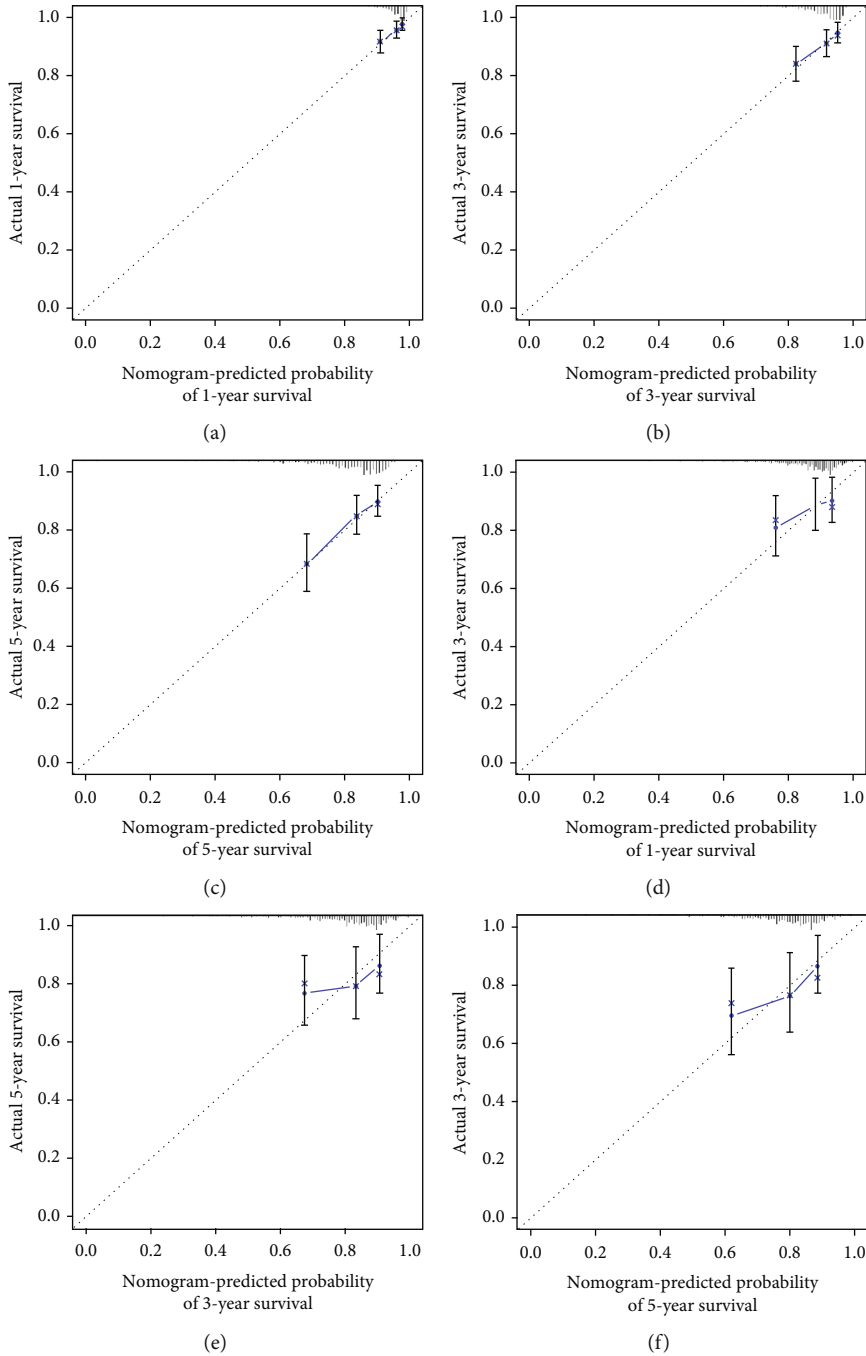
(a)

(b)

(c)
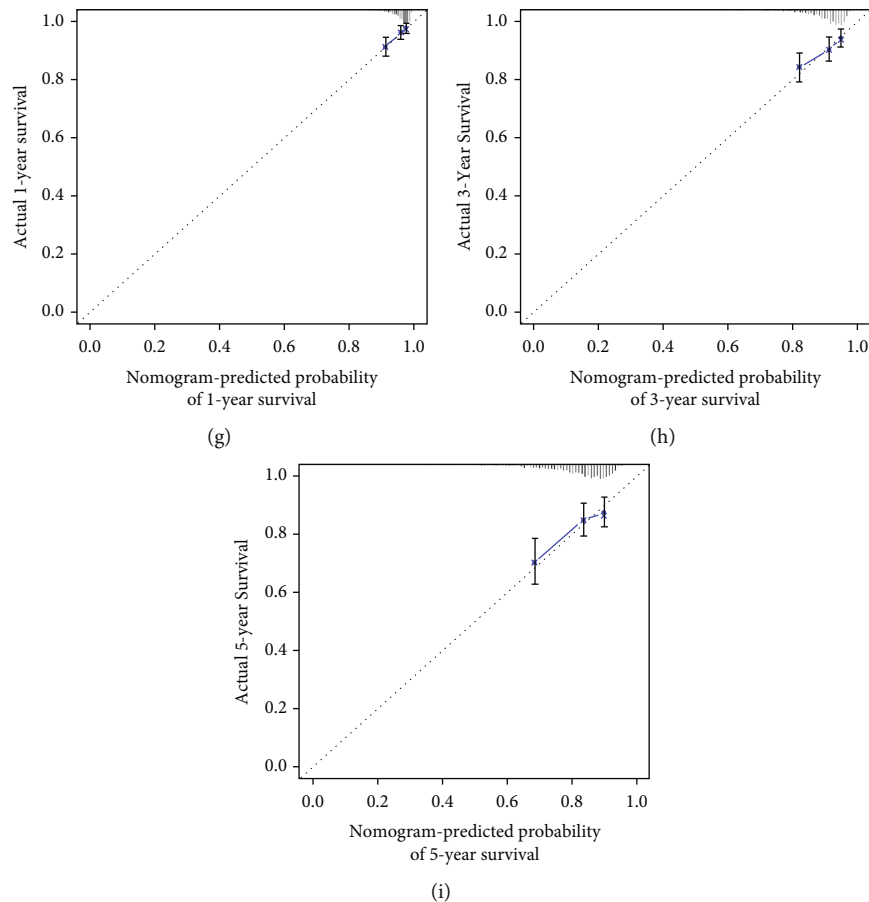
(d)

(e)

(f)

Figure 7: Continued.

(g)

(h)

(i)

Figure 7: (a–c) Calibration plots to predict 1-, 3-, and 5-year overall survival (OS) in the training set; (d–f) calibration plots to predict 1-, 3-, and 5-year; overall survival (OS) in the test set; (g–i) calibration plots to predict 1-, 3-, and 5-year overall survival (OS) in the overall internal validation set.

analyses, a total of 78 differentially expressed EMT-related genes were identified. (logFC > 1 or logFC < −1, FDR < 0.05). The results were expressed in heat maps and volcano plots (Figures 2(a) and 2(b)).

3.2. Prognostic Nomogram Construction. Since age is associated with prognosis in female BC patients, we included age and 78 differentially expressed genes in univariate Cox regression to investigate the correlation between the included variables and prognostic value in BC patients and finally identified seven variables significantly associated with OS in BC patients at *P* value < 0.1 (Figure 3). The model was then constructed with age, CD44, P3H1, SDC1, COL4A1, TGFβ1, and SERPINE1 by multivariate Cox regression: risk score = $(0.030655 * \text{age level}) + (3.35E − 05 * \text{expression level of CD44}) + (0.000142 * \text{expression level of P3H1}) + (3.76E − 05 * \text{expression level of SDC1}) + (1.60E − 05 * \text{COL4A1 expression level}) + (2.04E − 05 * \text{TGF}\beta\text{1 expression level}) + (0.000247 * \text{SERPINE1 expression level})$ (Table 1).

The nomogram was then constructed and consisted of a total of seven variables (Figure 4), and the total score could be obtained by summing the scores of each variable. The total score can be used to predict the survival rate of individual patients at 1, 3, and 5 years. For example, a BC patient

aged 65 years (20 points) with CD44 expression of 0 (20 points), P3H1 expression of 0 (21 points), SDC1 expression of 0 (20 points), COL4A1 expression of 80000 (43 points), TGFβ1 expression of 0 (21 points), and SERPINE1 expression of 0 (21 points) gets a sum-point of 166, corresponding to predicted 1-, 3-, and 5-year OS of 94.8%, 76.0%, and 57.4%, respectively.

Patients in TCGA group were divided into a low-risk group and a high-risk group using the median risk score as the threshold value. Figures 5(a), 5(c), and 5(e) show the distribution of the risk scores of BC patients from high to low in the training set, the internal validation set, and the overall internal validation set. The relationship between risk score and patient survival time in the training set, test set, and overall internal validation set is also shown (Figures 5(b), 5(d), and 5(f)). Patients with high-risk scores tended to have poorer clinical outcomes compared with those with low-risk scores. The survival analyses indicated the high-risk group had worse OS than that of the high-risk group (*P* < 0.05) (Figures 6(a)–6(c)).

3.3. Nomogram Calibration and Validation. The small angle between the survival probability and the actual survival outcome in the calibration curve indicates a strong agreement between them (Figure 7).

(a)

AUC at 1 year: 0.78

(b)

AUC at 3 year: 0.63

(c)

AUC at 5 year: 0.69

(d)

AUC at 1 year: 0.90

(e)

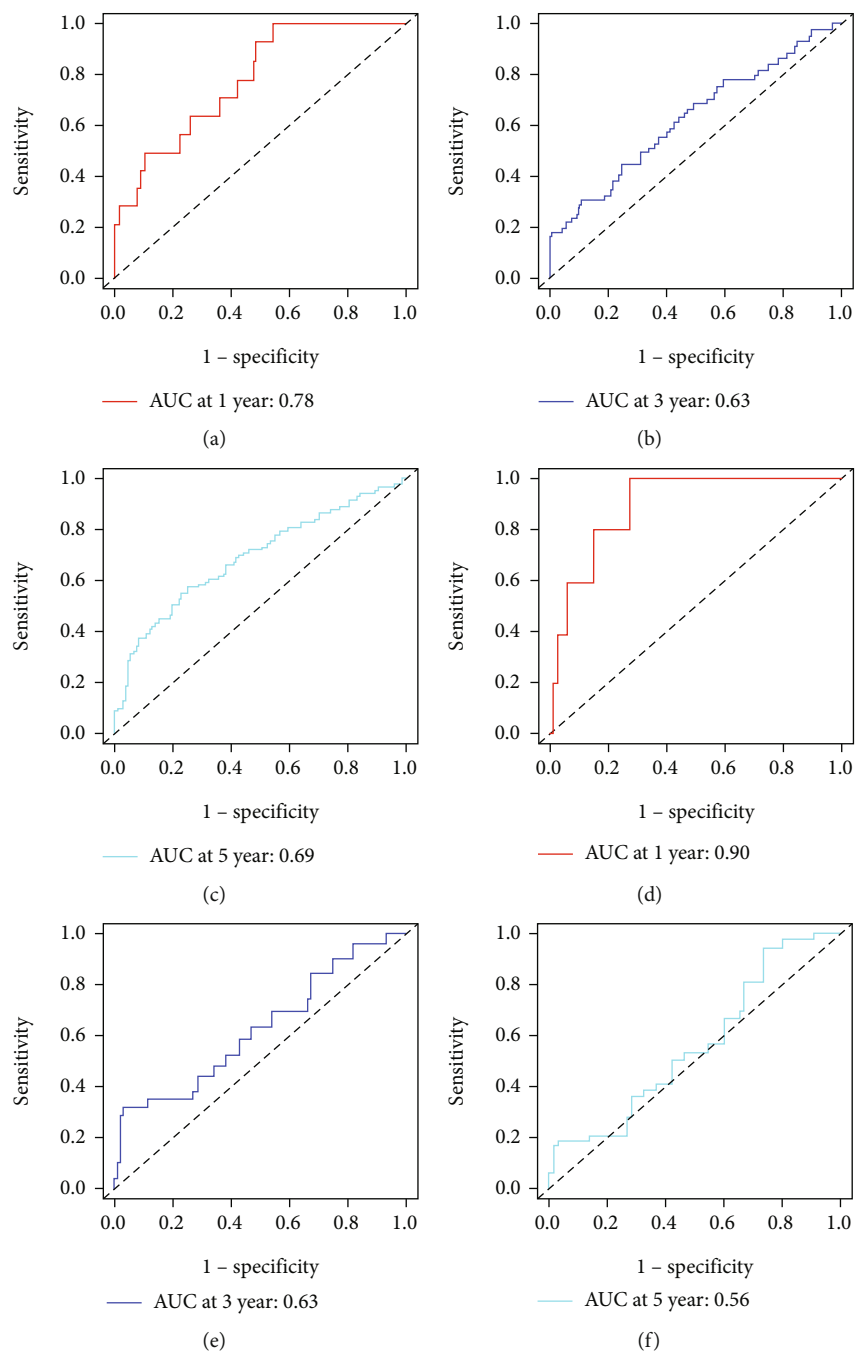AUC at 3 year: 0.63

(f)

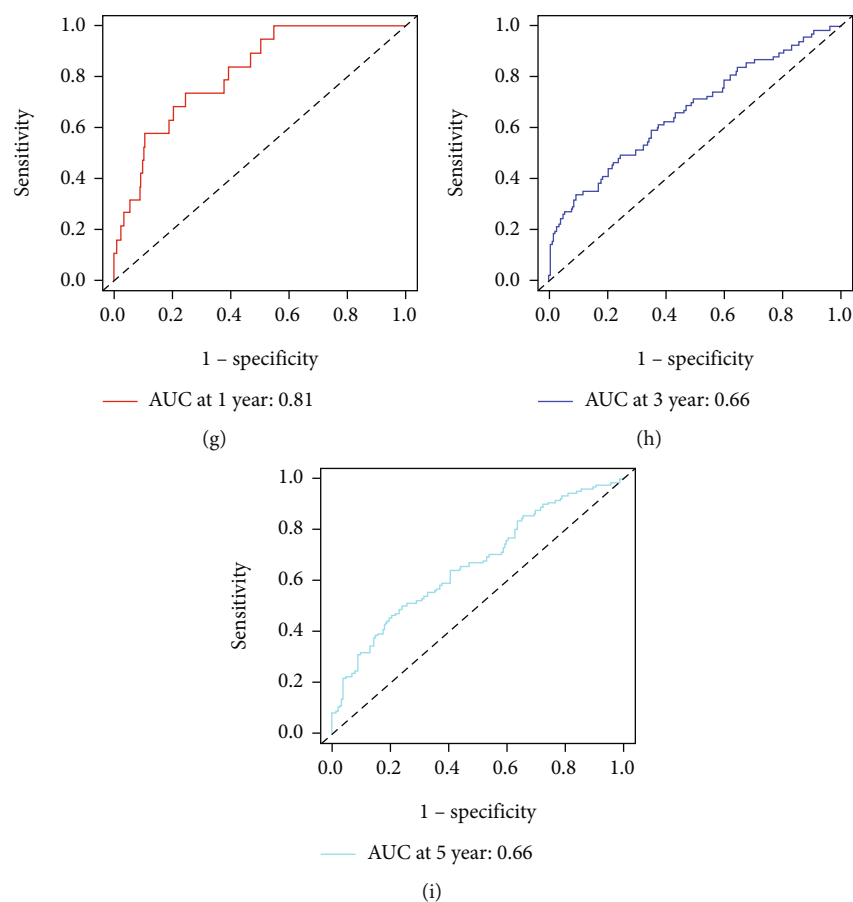AUC at 5 year: 0.56

Figure 8: Continued.

FIGURE 8: (a–c) ROC curves to predict 1-, 3-, and 5-year overall survival (OS) in the training set; (d–f) ROC curves to predict 1-, 3-, and 5-year; overall survival (OS) in the test set; (g–i) ROC curves to predict 1-, 3-, and 5-year overall survival (OS) in the overall internal validation set.

The C-index values for BC patients were 0.672 (95% CI 0.611–0.732), 0.692 (95% CI 0.586–0.798), and 0.679 (95% CI 0.626–0.732) in the training cohort, test set, and overall internal validation set, respectively. The time-dependent ROC curves were used to measure the sensitivity and specificity of the prediction model for predicting OS. Significantly, the AUC values were all greater than 0.63, except for the overall internal validation set with a 5-year predicted survival rate of 0.56, indicating that the model has high survival outcome prediction performance (Figure 8). The DCA curves also revealed better clinical applications for the risk scoring model (Figure 9).

The results based on C-index, ROC curves, calibration curves, and DCA curves indicated that the nomogram in our study demonstrated favorable predictive accuracy for the survival prognosis of female BC patients.

4. Discussion

BC is one of the most common cancers in females, with over 1,300,000 new cases and 450,000 deaths worldwide each year [16, 17]. Treatment of BC has advanced considerably, mainly through surgery, neoadjuvant chemotherapy, adjuvant chemotherapy, radiotherapy, systemic therapy, targeted therapy, and so on, with initial conventional surgery no lon-

ger being the best option for all patients [1]. However, BC remains one of the leading causes of cancer deaths in women worldwide, largely due to delayed diagnosis and unsuccessful treatment strategies [18]. Therefore, it is crucial to diagnose BC at an early stage and propose a personalized treatment plan based on the characteristics of the women patient's condition to predict their prognosis.

The TNM staging system is still the most widely used prognosis method to predict the survival of patients with BC. Although the American Joint Committee on Cancer (AJCC) updated BC staging in 2016 to include T, N, M, tumor grade, and expression of estrogen and progesterone receptors and HER2 [19], the current TNM staging system still has its undeniable deficiencies. For instance, it does not take into account other pathophysiological characteristics of the patient that have an impact on the prognosis of the tumor: age, gender, exercise, and overweight [20–22]. In addition, gene signature is an important factor in determining the prognosis of BC patients, as BC is a highly heterogeneous disease with different subtypes with different biological, molecular, and clinical processes. Gene expression profiling can identify genetic features to predict prognosis and guide the use of adjuvant therapy [23]. Among others, EMT genes regulate tumor proliferation, invasion, and metastasis [24, 25]. There are many prognostic models
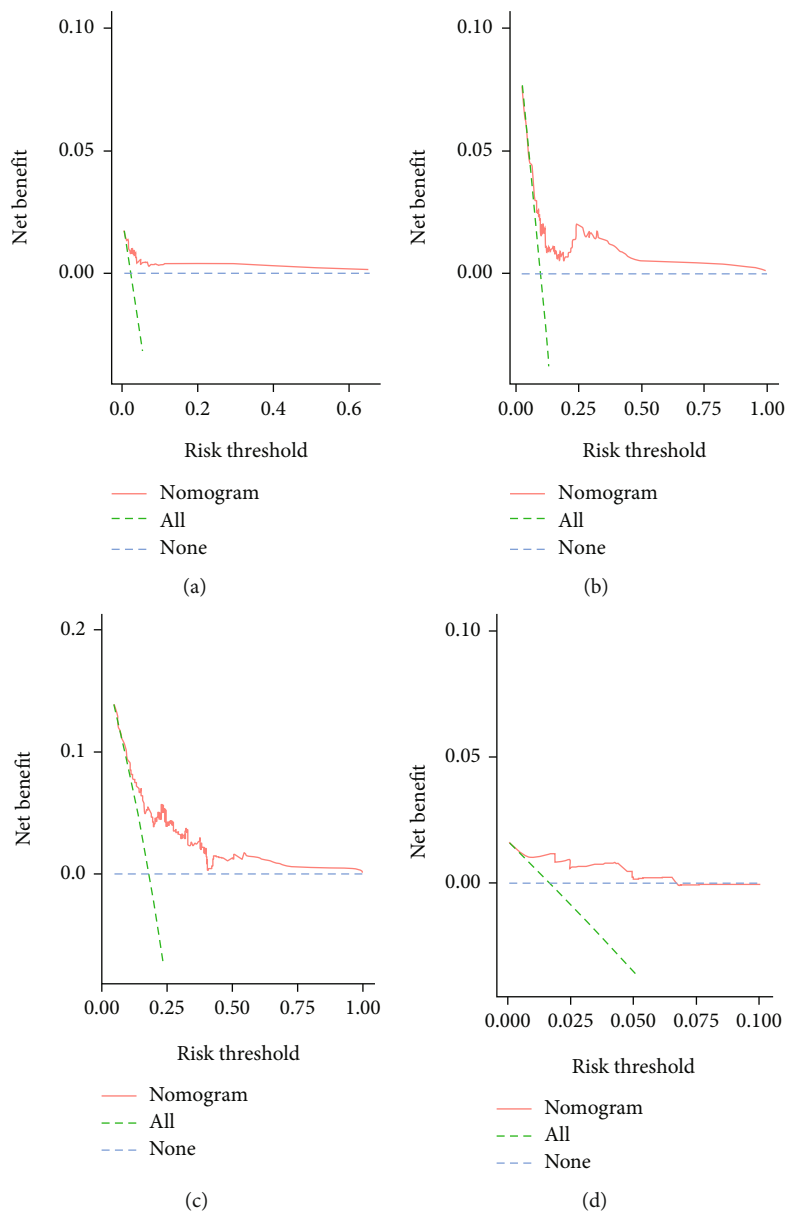
(a)

(b)

(c)

(d)

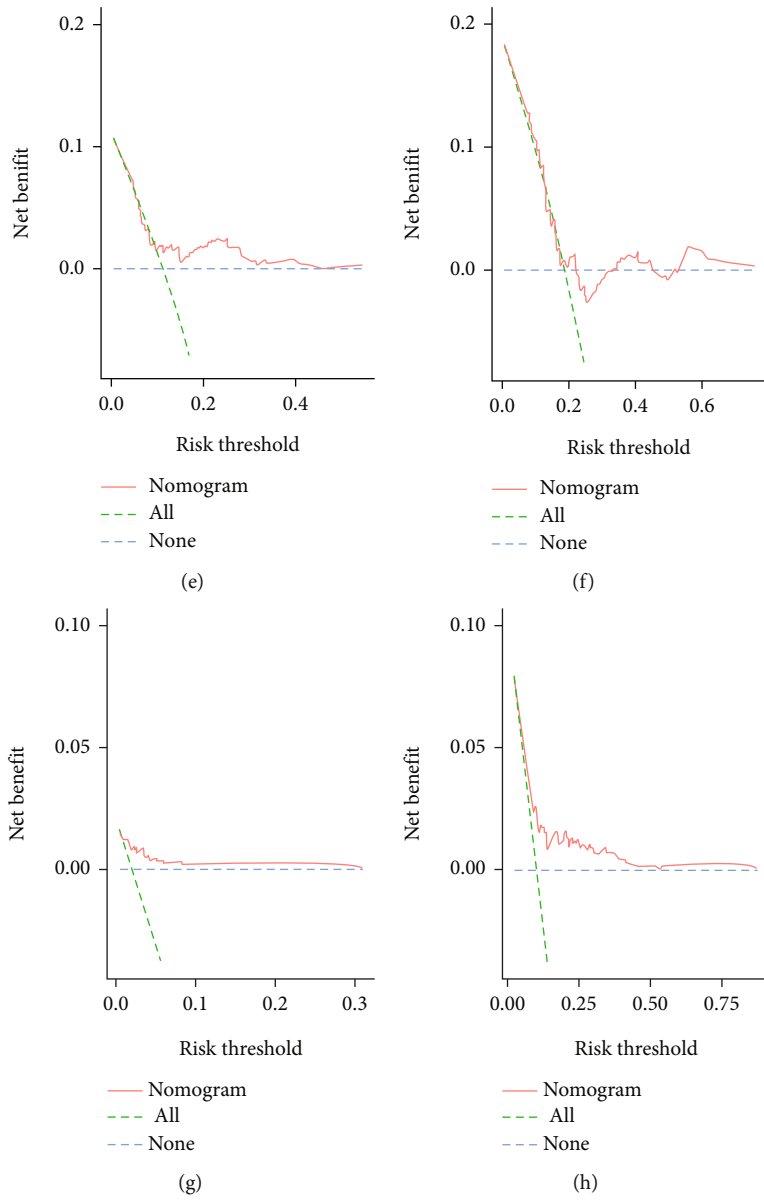FIGURE 9: Continued.

(e)

(f)
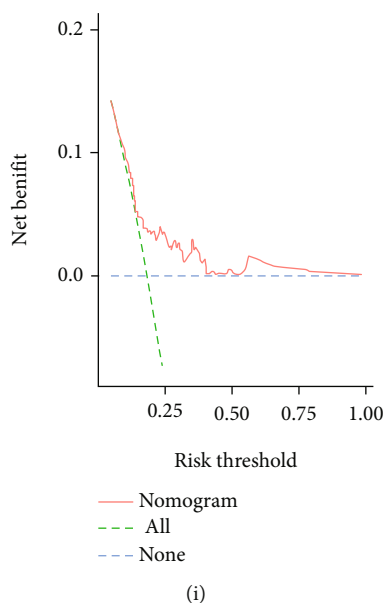

(g)

(h)

Figure 9: Continued.

FIGURE 9: (a–c) DCA analysis predicting 1-, 3-, and 5-year overall survival (OS) in the training set; (d–f) DCA analysis predicting 1-, 3-, and 5-year; overall survival (OS) in the test set; (g–i) DCA analysis predicting 1-, 3-, and 5-year overall survival (OS) in the overall internal validation set.

on BC; however, this is the first gene signature constructed by EMT-related genes. Moreover, considering the role of age and gender in the onset and progression of BC, we chose to study the prognosis of patients in women and used age as one of the predictors. Compared to previous studies, this nomogram was more accurate.

EMT is a cellular process in which cells lose their epithelial characteristics and acquire mesenchymal characteristics, such as quiescent adnexal cells gaining the ability to migrate [26]. EMT has been associated with a variety of tumor functions, including tumor initiation, malignant progression, tumor stemness, tumor cell migration, intravascular infiltration, metastasis, and resistance to therapy [9]. Most notably in this context, previous studies have shown that both cancer stem cell-like properties and drug resistance are associated with EMT [27]. Given the close link between oncogenic signaling and EMT blockers, EMT has emerged as a therapeutic target or goal in cancer therapy [28]. The relationship between EMT-related genes and breast cancer is also increasingly being investigated by researchers. The major focus of current studies is the regulatory mechanisms and therapeutic approaches of EMT for breast cancer in metastasis and invasion, mainly including miRNA and signaling pathways such as Wnt, Notch, TNF-α, NF-κB, and RTK. Investigators suppress breast cancer by attempting to therapeutically target or inhibit key/auxiliary players in these pathways [8, 29–31]. Most notably, upregulation of programmed death ligand 1 (PD-L1) expression is associated with EMT cell phenotype activation, and the control of the interaction between p53 and EMT master regulators is of importance in breast cancer. These two mechanisms have also been studied in other types of cancer and play a key role in the development and metastasis of cancer [30, 32].

This study was based on TCGA database. Differential analysis was firstly performed to find differential EMT-related genes. Univariate regression analysis was then conducted to identify candidate prognostic variables. We then developed a prognostic model by multivariate analysis to predict OS. Calibration curves, receiver operating characteristics (ROC) curves, C-index, and decision curve analysis (DCA) were used to test the veracity of the prognostic model. In addition to the training cohort of 70% BC patients, the remained cohort was treated as the test set. In the end, we derived that patient's age, CD44, P3H1, SDC1, COL4A1, TGFβ1, and SERPINE1 were independent prognostic factors for overall survival in female BC patients and constructed predictive models. The accuracy of the model has also been verified using various methods.

In accordance with our findings, in stage I and IV BC tumors, excess mortality increased linearly with age [33]. Recent studies have shown that a novel positive feedback loop between IL1β and CD44 promoted malignant progression in triple-negative BC (TNBC) and that CD44 was a potential target for inhibiting PD-L1 function in TNBC [34, 35]. Sayyad et al. demonstrated the role of Sdc1 in promoting brain metastasis in BC [36]. Several studies have demonstrated that COL4A1 expression could be used as a biomarker for superior prognosis in BC patients receiving neoadjuvant chemotherapy [37], while epigallocatechin-3-gallate (EGCG) exerted antitumor effects by restoring nine key genes, including COL4A1, in myeloid-derived suppressor cells (MDSCs) [37]. TGFβ1-activated cancer-associated fibroblasts (CAFs) promote BC growth and metastasis in part through autophagy [38]. The evolutionary branch E member 1 (SERPINE1) is a molecule involved in a variety of human malignancies. Zhang et al. showed that SERPINE1 served as an oncogene for PTX resistance in BC, and Xu et al. identified a functional pathway linking miR-1185-2-3p, GOLPH3L, and SERPINE1, which played an essential role in glucose metabolism in BC. Both of their studies

revealed that it may serve as a possible target for the treatment of BC [39, 40]. No studies have yet explored the mechanisms by which P3H1 affects BC development, progression, and metastasis, but an algorithm-based meta-analysis of genome-wide and proteomic data identified P3H1 as a potential biomarker for CRC. Our study indicates a direction of research for subsequent basic studies [41].

In this endeavor, some limitations need to be acknowledged. To begin with, the population races in TCGA database are primarily limited to whites and blacks, and extrapolation of findings to other racial groups needs to be validated. Second, a robust nomogram should be externally validated across cohorts; therefore, our nomogram needs to be further validated in multicenter clinical trials and prospective studies. Finally, some of the genes identified in this paper are relatively rarely reported in the academic literature. Therefore, more evidence including sample collection with complete experimental and clinical information should be performed for future validation is needed to elucidate the intrinsic association between age and six-gene signature and prognosis of BC patients.

However, our study also has some advantages. To our knowledge, this is the first study to additionally combine age as a prognostic variable with EMT-related genes to predict the prognosis of BC patients. Prognostic models may predict patient prognosis more accurately than conventional indicators.

In conclusion, we have developed and validated a relatively effective predictive model to predict the survival outcome of female BC patients at 1, 3, and 5 years. The accuracy and reliability of the prognostic model have also been verified. The results of our research need to be further validated in subsequent clinical practice.

Data Availability

The research article data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Wei Xue and Chenyu Sun performed the study, collected data, and wrote the manuscript. Hui Yuan, Xin Yang, Qiuping Zhang, and Yunnuo Liao participated in analyzing data and prepared the tables and figures. Hongwei Guo designed the research study and revised the manuscript. All authors read and approved the final manuscript. Wei Xue and Chenyu Sun contributed equally to this work.

Acknowledgments

Supplementary Materials

Supplementary Table S1: basic values of patients. Supplementary Table S2: EMT-related gene list. (Supplementary Materials)

References

[1] N. Harbeck and M. Gnant, "Breast cancer," *Lancet*, vol. 389, no. 10074, article S0140673616318918, pp. 1134–1150, 2017.

[2] Y. Liang, H. Zhang, X. Song, and Q. Yang, "Metastatic heterogeneity of breast cancer: molecular mechanism and potential therapeutic targets," *Seminars in Cancer Biology*, vol. 60, article S1044579X1930063X, pp. 14–27, 2020.

[3] X. Xu, M. Zhang, F. Xu, and S. Jiang, "Wnt signaling in breast cancer: biological mechanisms, challenges and opportunities," *Molecular Cancer*, vol. 19, no. 1, article 1276, p. 165, 2020.

[4] G. Cui, J. Wu, J. Lin et al., "Graphene-based nanomaterials for breast cancer treatment: promising therapeutic strategies," *Journal of Nanobiotechnology*, vol. 19, no. 1, p. 211, 2021.

[5] P. Boix-Montesinos, P. M. Soriano-Teruel, A. Armiñán, M. Orzáez, and M. J. Vicent, "The past, present, and future of breast cancer models for nanomedicine development," *Advanced Drug Delivery Reviews*, vol. 173, article S0169409X2100096X, pp. 306–330, 2021.

[6] S. Yardım-Akaydin, B. Karahalil, and S. N. Baytas, "New therapy strategies in the management of breast cancer," *Drug Discovery Today*, vol. 27, no. 6, pp. S1359–S6446, 2022.

[7] B. Bakir, A. M. Chiarella, J. R. Pitarresi, and A. K. Rustgi, "EMT, MET, plasticity, and tumor metastasis," *Trends in Cell Biology*, vol. 30, no. 10, article S0962892420301446, pp. 764–776, 2020.

[8] J. Luo, J. F. Yao, X. F. Deng et al., "14, 15-EET induces breast cancer cell EMT and cisplatin resistance by up-regulating integrin $\alpha v \beta 3$ and activating FAK/PI3K/AKT signaling," *Journal of Experimental & Clinical Cancer Research: CR*, vol. 37, no. 1, article 694, p. 23, 2018.

[9] I. Pastushenko and C. Blanpain, "EMT transition states during tumor progression and metastasis," *Trends in Cell Biology*, vol. 29, no. 3, article S0962892418302010, pp. 212–226, 2019.

[10] M. Saotome, D. B. Poduval, R. Nair, M. Cooper, and M. Takaku, "GATA3 truncation mutants Alter EMT related gene expression via partial motif recognition in luminal breast cancer cells," *Frontiers in Genetics*, vol. 13, p. 820532, 2022.

[11] K. Parthasarathi, S. Mandal, S. Singh et al., "In silico analysis of ion channels and their correlation with epithelial to mesenchymal transition in breast cancer," *Cancers*, vol. 14, no. 6, p. 1444, 2022.

[12] G. Cserni, E. Chmielik, B. Cserni, and T. Tot, "The new TNM-based staging of breast cancer," *Virchows Archiv: an International Journal of Pathology*, vol. 472, no. 5, article 2301, pp. 697–703, 2018.

[13] Y. S. Sun, Z. Zhao, Z. N. Yang et al., "Risk factors and preventions of breast cancer," *International Journal of Biological Sciences*, vol. 13, no. 11, pp. 1387–1397, 2017.

[14] L. Cao, C. W. Towe, R. Shenk, N. Stabellini, A. L. Amin, and A. J. Montero, "A comparison of local therapy alone with local plus systemic therapy for stage I pT1aN0M0 HER2+ breast cancer: a National Cancer Database analysis," *Cancer*, vol. 128, no. 13, pp. 2433–2440, 2022.

[15] Q. Tai, W. Xue, M. Li et al., "Survival nomogram for metastasis colon cancer patients based on SEER database," *Frontiers in Genetics*, vol. 13, p. 832060, 2022.

[16] N. A. Bagegni and L. L. Peterson, "Age-related disparities in older women with breast cancer," *Advances in Cancer Research*, vol. 146, pp. 23–56, 2020.

[17] Y. Huang, S. Du, J. Liu et al., "Diagnosis and prognosis of breast cancer by high-performance serum metabolic fingerprints," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 119, no. 12, article e2122245119, 2022.

[18] N. Jokar, I. Velikyan, H. Ahmadzadehfar et al., "Theranostic approach in breast cancer," *Clinical Nuclear Medicine*, vol. 46, no. 8, pp. e410–e420, 2021.

[19] S. B. Edge, G. N. Hortobagyi, and A. E. Giuliano, "New and important changes in breast cancer TNM: incorporation of biologic factors into staging," *Expert Review of Anticancer Therapy*, vol. 19, no. 4, pp. 309–318, 2019.

[20] N. R. Latha, A. Rajan, R. Nadhan et al., "Gene expression signatures: a tool for analysis of breast cancer prognosis and therapy," *Critical Reviews in Oncology/Hematology*, vol. 151, article S1040842820301025, p. 102964, 2020.

[21] M. de Roon, A. M. May, A. McTiernan et al., "Effect of exercise and/or reduced calorie dietary interventions on breast cancer-related endogenous sex hormones in healthy postmenopausal women," *Breast Cancer Research: BCR*, vol. 20, no. 1, article 1009, p. 81, 2018.

[22] S. Jiralerspong and P. J. Goodwin, "Obesity and breast cancer prognosis: evidence, challenges, and opportunities," *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, vol. 34, no. 35, pp. 4203–4216, 2016.

[23] M. Kwa, A. Makris, and F. J. Esteva, "Clinical utility of gene-expression signatures in early stage breast cancer," *Clinical Oncology*, vol. 14, no. 10, pp. 595–610, 2017.

[24] L. Sun, C. Shi, S. Liu et al., "Overexpression of NuSAP1 is predictive of an unfavourable prognosis and promotes proliferation and invasion of triple-negative breast cancer cells via the Wnt/β-catenin/EMT signalling axis," *Gene*, vol. 747, article S0378111920303267, p. 144657, 2020.

[25] H. Wang, S. Guo, S. J. Kim et al., "Cisplatin prevents breast cancer metastasis through blocking early EMT and retards cancer growth together with paclitaxel," *Theranostics*, vol. 11, no. 5, pp. 2442–2459, 2021.

[26] N. M. Aiello and Y. Kang, "Context-dependent EMT programs in cancer metastasis," *The Journal of Experimental Medicine*, vol. 216, no. 5, pp. 1016–1026, 2019.

[27] G. Pan, Y. Liu, L. Shang, F. Zhou, and S. Yang, "EMT-associated microRNAs and their roles in cancer stemness and drug resistance," *Cancer communications (London, England)*, vol. 41, no. 3, pp. 199–217, 2021.

[28] E. S. Cho, H. E. Kang, N. H. Kim, and J. I. Yook, "Therapeutic implications of cancer epithelial-mesenchymal transition (EMT)," *Archives of Pharmacal Research*, vol. 42, no. 1, article 1108, pp. 14–24, 2019.

[29] S. Kotiyal and S. Bhattacharya, "Breast cancer stem cells, EMT and therapeutic targets," *Biochemical and Biophysical Research Communications*, vol. 453, no. 1, article S0006291X14016957, pp. 112–116, 2014.

[30] S. S. Messeha, N. O. Zarmouh, and K. Soliman, "Polyphenols modulating effects of PD-L1/PD-1 checkpoint and EMT-mediated PD-L1 overexpression in breast cancer," *Nutrients*, vol. 13, no. 5, article nu13051718, p. 1718, 2021.

[31] Y. Li, Y. W. Wang, X. Chen et al., "MicroRNA-4472 promotes tumor proliferation and aggressiveness in breast cancer by targeting RGMA and inducing EMT," *Clinical Breast Cancer*, vol. 20, no. 2, article S152682091930669X, pp. e113–e126, 2020.

[32] S. Parfenyev, A. Singh, O. Fedorova, A. Daks, R. Kulshreshtha, and N. A. Barlev, "Interplay between p53 and non-coding RNAs in the regulation of EMT in breast cancer," *Cell Death & Disease*, vol. 12, no. 1, article 3327, p. 17, 2021.

[33] C. Cluze, M. Colonna, L. Remontet et al., "Analysis of the effect of age on the prognosis of breast cancer," *Breast Cancer Research and Treatment*, vol. 117, no. 1, article 222, pp. 121–129, 2009.

[34] J. H. Jang, D. H. Kim, J. M. Lim et al., "Breast cancer cell-derived soluble CD44 promotes tumor progression by triggering macrophage IL1β production," *Cancer Research*, vol. 80, no. 6, pp. 1342–1356, 2020.

[35] T. Kong, R. Ahn, K. Yang et al., "CD44 promotes PD-L1 expression and its tumor-intrinsic function in breast and lung cancers," *Cancer Research*, vol. 80, no. 3, pp. 444–457, 2020.

[36] M. R. Sayyad, M. Puchalapalli, N. G. Vergara et al., "Syndecan-1 facilitates breast cancer metastasis to the brain," *Breast Cancer Research and Treatment*, vol. 178, no. 1, article 5347, pp. 35–49, 2019.

[37] S. M. Wang, P. M. Chen, Y. W. Sung, W. C. Huang, H. S. Huang, and P. Y. Chu, "Effect of COL4A1 expression on the survival of neoadjuvant chemotherapy breast cancer patients," *Journal of Oncology*, vol. 2020, Article ID 5209695, 8 pages, 2020.

[38] M. Huang, M. Fu, J. Wang et al., "TGF-β1-activated cancer-associated fibroblasts promote breast cancer invasion, metastasis and epithelial-mesenchymal transition by autophagy or overexpression of FAP-α," *Biochemical Pharmacology*, vol. 188, article S0006295221001234, p. 114527, 2021.

[39] Q. Zhang, L. Lei, and D. Jing, "Knockdown of SERPINE1 reverses resistance of triple-negative breast cancer to paclitaxel via suppression of VEGFA," *Oncology Reports*, vol. 44, no. 5, pp. 1875–1884, 2020.

[40] Y. Xu, W. Chen, J. Liang et al., "The miR-1185-2-3p-GOLPH3L pathway promotes glucose metabolism in breast cancer by stabilizing p53-induced SERPINE1," *Journal of Experimental & Clinical Cancer Research: CR*, vol. 40, no. 1, article 1767, p. 47, 2021.

[41] D. R. Gawel, E. J. Lee, X. Li et al., "An algorithm-based meta-analysis of genome- and proteome-wide data identifies a combination of potential plasma biomarkers for colorectal cancer," *Scientific Reports*, vol. 9, no. 1, article 51999, p. 15575, 2019.