Research Article

Developing ZNF Gene Signatures Predicting Radiosensitivity of Patients with Breast Cancer

Derui Yan,[1,2] Mingjing Shen,[3] Zixuan Du,[1,2] Jianping Cao,[4] Ye Tian,[5] Ping Zeng [1D],[6] and Zaixiang Tang [1D][1,2]

[1]*Department of Biostatistics, School of Public Health, Medical College of Soochow University, Suzhou, China*
[2]*Jiangsu Key Laboratory of Preventive and Translational Medicine for Geriatric Diseases, Medical College of Soochow University, Suzhou, China*
[3]*Department of Thoracic Surgery, The Second Affiliated Hospital of Soochow University, Suzhou, China*
[4]*School of Radiation Medicine and Protection and Collaborative Innovation Center of Radiation Medicine of Jiangsu Higher Education Institutions, Soochow University, Suzhou, China*
[5]*Department of Radiotherapy & Oncology, The Second Affiliated Hospital of Soochow University, Suzhou, China*
[6]*Department of Biostatistics, School of Public Health, Xuzhou Medical University, Xuzhou, China*

Correspondence should be addressed to Ping Zeng; zpstat@xzhmu.edu.cn and Zaixiang Tang; tangzx@suda.edu.cn

Adjuvant radiotherapy is one of the main treatment methods for breast cancer, but its clinical benefit depends largely on the characteristics of the patient. This study aimed to explore the relationship between the expression of zinc finger (ZNF) gene family proteins and the radiosensitivity of breast cancer patients. Clinical and gene expression data on a total of 976 breast cancer samples were obtained from The Cancer Genome Atlas (TCGA) database. ZNF gene expression was dichotomized into groups with a higher or lower level than the median level of expression. Univariate and multivariate Cox regression analyses were used to evaluate the relationship between ZNF gene expression levels and radiosensitivity. The Molecular Taxonomy Data of the International Federation of Breast Cancer (METABRIC) database was used for validation. The results revealed that 4 ZNF genes were possible radiosensitivity markers. High expression of ZNF644 and low expression levels of the other 3 genes (ZNF341, ZNF541, and ZNF653) were related to the radiosensitivity of breast cancer. Hierarchical cluster, Cox, and CoxBoost analysis based on these 4 ZNF genes indicated that patients with a favorable 4-gene signature had better overall survival on radiotherapy. Thus, this 4-gene signature may have value for selecting those patients most likely to benefit from radiotherapy. ZNF gene clusters could act as radiosensitivity signatures for breast cancer patients and may be involved in determining the radiosensitivity of cancer.

1. Introduction

Breast cancer ranks as the fifth cause of cancer death overall and is the most commonly diagnosed cancer and the leading cause of cancer death in women [1, 2]. In 2018, there were about 2.1 million newly diagnosed female breast cancer patients worldwide, accounting for nearly a quarter of female cancer cases [2]. Treatment for breast cancer includes surgery, chemotherapy, and radiotherapy. A series of studies have confirmed that radiotherapy can effectively reduce the risk of local and overall recurrence and can improve overall survival [3–6]. Thus, after breast-conserving surgery, radiotherapy reduced the recurrence rate by half and mortality by about one-sixth [7]. However, not all breast cancer patients benefit from radiotherapy. Although well tolerated by most breast cancer patients, some patients may experience breast or chest wall pain during and after radiotherapy, and the known late side effects of radiotherapy include rib fractures, heart disease, and lymphedema [8, 9]. The absolute benefit of radiotherapy depends largely on the characteristics of the patient. Identifying factors related to individual radiosensitivity will optimize adjuvant radiotherapy and

reduce the severity of normal tissue responses. Predicting tumor response to radiotherapy is one of the main problems of cancer treatment.

Until now, multigene signatures constructed based on mRNA [10, 11], lncRNA [12, 13], or miRNA [14] expression levels have been extensively studied to provide prognostic and predictive information for breast cancer treatment. For instance, the 9-mRNA signature can distinguish patients with different risks for overall survival and further predict the survival probability of patients [10]. The genome instability-derived lncRNA signature was used as an independent prognostic marker to stratify risk subgroups for patients with breast cancer [13]. A GI-derived three miRNA signature was used as a minimally invasive predictor of risk and unfavorable prognosis in breast cancer [14]. Most of the existing signatures can predict overall survival or local recurrence, but few signatures can predict the benefit of treatment. Although prognostic biomarkers can determine which patients will have a poor outcome and may require intensive treatment, they cannot determine which specific treatment should be given or whether the desired results can be obtained from these treatments. The development of predictive models has more clinical value [15].

Nowadays, many methods have been developed to combine gene therapy or molecular therapy with radiotherapy to predict which patients may be sensitive to and benefit from radiotherapy. Previous studies have shown that gene expression patterns can predict the intrinsic radiosensitivity of cells. Gene signatures have been identified and validated for predicting the radiosensitivity of many cancer types including gastric cancer [16], breast cancer [17], and glioma [18]. A 31-gene cluster related to cell cycle, cell adhesion, and cell junctions was found to have the potential to predict the radiosensitivity of cancer cells [19]. A radiosensitivity molecular signature [radiosensitivity index (RSI)] was clinically validated in 4 independent datasets (breast, rectal, esophageal, and head and neck) [20, 21]. Gene signatures can be used as predictive biomarkers to identify radiosensitive patients and optimize adjuvant radiotherapy.

Alterations in the transcriptome are a common feature of human cancer. In many cases, transcriptional disorders are caused by changes in expression levels or transcription factor activity [22]. Zinc finger (ZNF) proteins are the largest family of transcriptional regulators in mammals. The classic zinc finger protein family includes three categories, which are known as the BTB domain (Broad-complex, Tramtrack, and Bric-a-brac), the Krüppel-associated box (KRAB), and the newly defined SCAN domain, also known as the leucine-rich domain [23]. The functions of zinc finger proteins are extremely diverse, such as DNA recognition, RNA packaging, protein folding and assembly, lipid binding, transcriptional activation, and regulation of apoptosis [24–26]. Several family members mediate various different functions in tumor pathobiology. ZNF750 and ZNF545 have been identified as tumor suppressors [27, 28]. ZNF750 induces cell cycle arrest in the $G0/G1$ phase and regulates the tumor vascular microenvironment to inhibit oral squamous cell carcinoma's malignant progression. ZNF545 inhibits breast tumor cell proliferation by inducing apoptosis.

Downregulation of ZNF121 and ZNF259 expression may inhibit cell proliferation and increase the proportion of apoptotic cells [29, 30]. The overexpression of ZNF genes is related to poor prognosis and lymph node metastasis [31–34]. Taken together, these studies indicate that ZNF genes may function as oncogenes involved in the occurrence and progression of cancer.

So far, no reports have appeared on correlations between ZNF genes and the radiosensitivity of breast cancer patients. In the present study, we used The Cancer Genome Atlas (TCGA) database to analyze the relationship between ZNF protein family genes and the radiosensitivity of breast cancer. Data from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) database were employed for independent external validation. This study suggests new targets for the subsequent development of therapies for breast cancer.

2. Materials and Methods

2.1. Study Samples. Clinical information and ZNF gene expression data of patients with breast cancer were downloaded from The Cancer Genome Atlas (TCGA; http://cancergenome.nih.gov/) via the R package TCGA assembler [35]. First, we eliminated patients without survival time data or survival status to select only those with clear survival information. In addition, we excluded samples of male breast cancer, patients with no radiotherapy data records, and those with less than 5 days of survival. Second, normalized read counts for the mRNA sequence data were downloaded. We eliminated genes with a maximum expression value <10 because they were barely expressed. Genes with zero expression rates in >50% of cases were also removed. Third, we merged the expression dataset and clinical dataset according to the patient's barcode number to obtain 511 gene expression profiles of 976 patients for the next phase of the analysis. The specific process of cleaning the clinical data is shown in Figure 1. Data from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) database (http://www.cbioportal.org/) were obtained via the cBioPortal to perform the validation analysis. The GSE31863 dataset was available from the GEO database (https://www.ncbi.nlm.nih.gov/geo/).

2.2. Analysis Method. In the present study, radiosensitive patients were defined as a group of patients who had better overall survival after receiving radiotherapy compared with those who did not receive radiotherapy. Radiosensitivity genes were defined as those identifying patients responding to radiotherapy [16, 36]. ZNF genes were screened for further analysis by using two-step Cox regression analysis. The effect of radiotherapy on the overall survival of breast cancer was related to the expression levels of these ZNF genes. For example, if a gene related to radiosensitivity was expressed at a high level, patients who received radiotherapy had longer survival than patients who did not receive radiotherapy. For the low expression level of this gene, there were no significant survival differences between the two
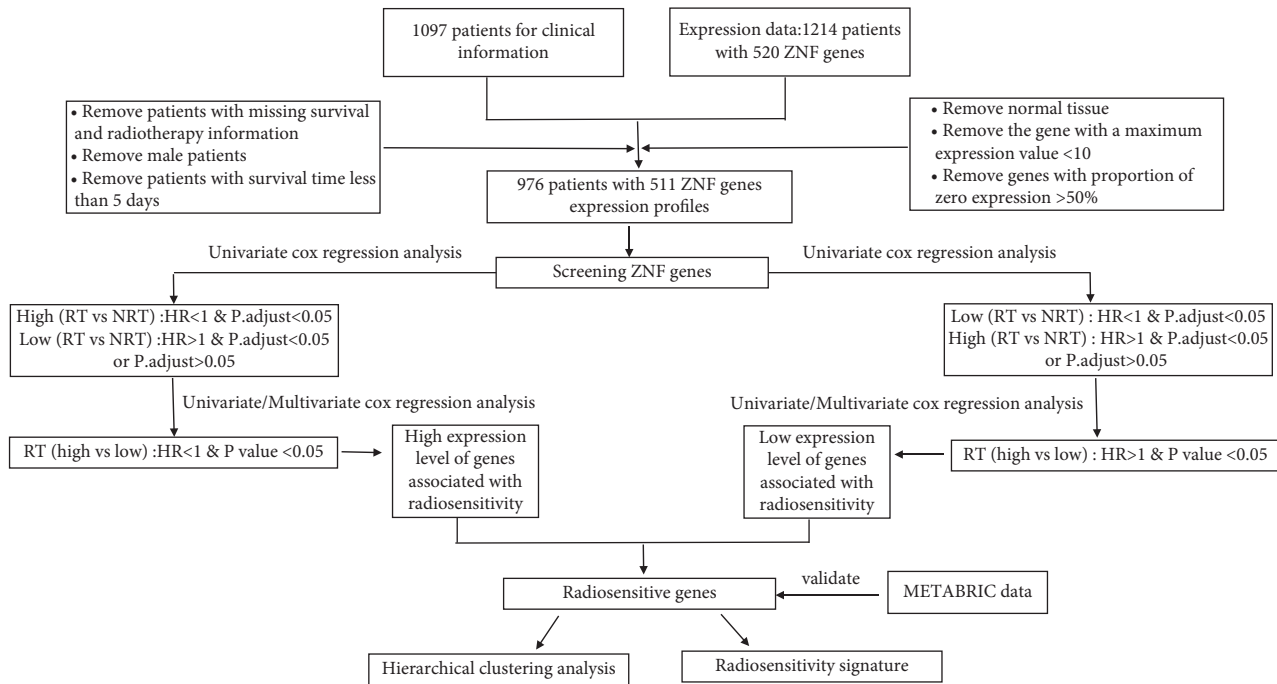
FIGURE 1: The workflow of data cleaning and analysis steps.

groups; thus, it is concluded that high expression of this gene was associated with radiosensitivity. The selected genes were determined based on the corresponding risk ratio and adjusted P value and those with significant interaction effects were selected for further analysis.

The distribution of expressed ZNF genes was skewed (Figure S1). According to the median expression value of each ZNF gene, the patients were divided into high and low expression subgroups. The Kaplan–Meier method was used to estimate survival time for the two subgroups, and the two-sided log-rank test was applied to determine statistical significance. Univariate and multivariate proportional hazard regression models were used to analyze the relationships between gene expression levels and radiosensitivity. Statistical analysis was performed in R version 3.6.3 using the packages "survival" and "rms." The Benjamini–Hochberg method was used to perform multiple corrections. $P < 0.05$ was considered statistically significant. Missing values were multiply imputed using the R package "mice." Hierarchical cluster analysis was performed by using R the package "heatmap." The CoxBoost algorithm used the R package CoxBoost (https://github.com/binderh/CoxBoost) [37]. Random survival forests (RSF) algorithm used the R package "randomForestSRC." The data analysis process is also summarized in Figure 1.

3. Results

3.1. Clinical Information and Survival Analysis.
Clinical information on the 976 breast cancer patients included in the TCGA database is summarized in Table S1. Of these patients, 557 received radiotherapy. Multivariate analysis showed that receiving radiotherapy did not result in a statistically significant overall survival (OS) benefit when the whole patient

cohort was analyzed ($P = 0.068$). The only clinical factors associated with the overall survival of these breast cancer patients were age ($P < 0.001$), pathological stage ($P < 0.001$), and having received chemotherapy ($P = 0.025$). A total of 1798 breast cancer patients is included in the METABRIC database, with 1094 patients receiving radiotherapy. After univariate and multivariate Cox regression analysis, radiotherapy was still related to the overall survival of breast cancer patients (Table S2). In the GSE31863 dataset, a total of 121 patients (66 patients receiving radiotherapy) were included in this study (Table S3). Univariate Cox regression analysis showed that there was no significant correlation between radiotherapy and the OS of breast cancer patients ($P = 0.127$), but in multivariate Cox regression analysis, radiotherapy showed a survival benefit ($P = 0.03$).

3.2. Radiosensitive ZNF Gene Development.
Based on the detailed procedure shown in Figure 1, we analyzed the effects of radiotherapy on the overall survival of breast cancer patients with high and low levels of ZNF gene expression. According to the corresponding risk ratio and adjusted P value, the high expression of 253 ZNF genes and the low expression of a further 106 ZNF genes were informative for the outcome of radiotherapy. Of these, the expression levels of 51 genes were related to the overall survival of breast cancer patients receiving radiotherapy (Tables S4 and S5). There were 40 overlapping ZNF genes in the METABRIC database for validation. Univariate Cox regression analysis showed that the expression levels of 7 genes were related to radiosensitivity (Table S6). Multivariate Cox regression analysis of the two databases showed that high expression of ZNF644 and low expression of ZNF341, ZNF541, and ZNF653 were associated with radiosensitivity of breast cancer.

Figure 2 depicts a comparison of the Kaplan–Meier curves between radiotherapy and gene expression levels. When ZNF644 was expressed at a high level, patients who received radiotherapy had a longer survival time than those who did not receive radiotherapy. However, in patients with low ZNF644 expression, there were no significant survival differences between those receiving or not receiving radiotherapy. Among patients who did receive radiotherapy, those with high ZNF644 expression had significantly longer survival than patients with low level expression. For the other 3 genes, the results were the opposite; i.e., high expression resulted in no significant survival differences between patients receiving or not receiving radiotherapy but when these genes were expressed at low levels, patients receiving radiotherapy had a survival advantage. Thus, of patients receiving radiotherapy, those with low expression of these genes had significantly longer survival than patients with high expression. Figure 2 shows the results of univariate analysis of the TCGA data, and Figure 3 shows the results of univariate analysis of the METABRIC data which are consistent with these. Thus, similar conclusions can be drawn from the analyses in Figures 2 and 3.

Multivariate analysis was also performed adjusting for age, pathological stage, ER, PR, HER, histological type, and chemotherapy (Figure 4). The four identified ZNF genes with the potential to select patients who are most likely to benefit from radiotherapy were shown to be independent of clinical variables.

3.3. Correlation of ZNF Radiosensitive Genes.
The results of the correlation analysis are depicted in Figures 5(a) and 6(a) showing an interesting correlation pattern. ZNF644 expression was negatively correlated with the other 3 genes (ZNF341, ZNF541, and ZNF653), which were positively correlated with each other.

3.4. Cluster and Survival Analysis.
We further extracted expression patterns of these 4 genes for hierarchical cluster analysis (Figure 5(b)). The patients were classified into two groups according to this hierarchical cluster analysis. The red bar represents clustered radiosensitive (RS) patients ($n = 623$), and the black bar represents clustered radioresistant (RR) patients ($n = 353$). Patients in the RS group had significantly longer survival after receiving radiotherapy than patients who did not receive radiotherapy ($P < 0.001$) (Figure 5(c)). In contrast, radiotherapy did not improve the overall survival in the RR group ($P = 0.68$) (Figure 5(d)). Among the patients receiving radiotherapy, the survival of patients in the RS group was significantly longer than that in the RR group ($P = 0.003$) (Figure 5(e)). The results of the hierarchical cluster and univariate Cox regression analysis of the validation data are shown in Figure 6. Subsequently, multivariate analysis was further performed to assess the effect of radiotherapy on overall survival for the clusters of RS and RR patients. Table 1 summarizes the raw and adjusted HR and P values by univariate and multivariate Cox regression analysis. Similarly, the univariate and multivariate Cox regression analysis results of the METABRIC data

were consistent with the above outcome. Patients in the GSE31683 dataset were separated into clustered RS and RR groups (Figure S3). In the cluster RS group, patients who received radiotherapy tended to have better recurrence-free survival than those who did not ($P = 0.045$). By contrast, radiotherapy did not show an RFS benefit in the clustered RR group ($P = 0.6$). However, there was no significant improvement in RFS after radiotherapy in the RS group ($P = 0.32$).

3.5. Cox Prediction Model.
Among patients with TCGA data who received radiotherapy, a 4-mRNA ZNF gene signature comprising ZNF644, ZNF341, ZNF541, and ZNF653 was established. Thus, the risk score = $(-0.61414^*$ expression level of ZNF644) + $(0.21667^*$ expression level of ZNF341) + $(-0.14907^*$ expression level of ZNF541) + $(-0.06819^*$ expression level of ZNF653). The coefficients were assessed by multivariate regression analysis based on these 4 genes (Table S6). Patients could be divided into low- and high-risk groups based on the median risk score. Hence, the low-risk group was classified as the RS group, which had a higher survival rate after radiotherapy, while the high-risk group was classified as the RR group, which had a poorer outcome after radiotherapy (Figure 7(a)). In the RS group, patients who received radiation therapy had better overall survival than those who did not (Figure 7(b)). In contrast, radiotherapy did not improve overall survival in the RR group (Figure 7(c)). METABRIC data were used to validate this signature, with the result shown in Figures 7(d)–7(f). Multivariate Cox regression analysis showed that this signature was a predictive factor independent of other clinical factors (Figure 8). According to Cox regression analysis, the significant interaction between this 4-gene-based signature and radiotherapy indicates that it may have predictive value for identifying those patients who will benefit from radiotherapy. Patients in the GSE31863 dataset were separated into RS and RR groups using the cut-off established above (Figure S4). In the RS group, the RFS of patients who received radiotherapy did not improve significantly compared with patients who did not receive radiotherapy ($P = 0.082$). Besides, among radiotherapy patients, the RFS of RS patients was not significantly higher than that of RR patients ($P = 0.73$). KM plot showed that this gene signature did not have significant interaction with radiotherapy for predicting RFS.

3.6. CoxBoost Prediction Model.
CoxBoost is an algorithm that allows the implementation of boosting in conjunction with the Cox model [38]. We used the "CoxBoost" package to fit the Cox model (tune by CoxBoost (): step. no = 100; penalty = 100). The coefficients of the selected multivariate Cox model were set by the last step of CoxBoost (Figure S5). Similarly, based on a combination of the gene coefficients and gene expression count, the risk score of each patient was calculated. Patients were divided into low- and high-risk groups based on the median risk score. Among radiotherapy patients, the prognosis of RS patients was significantly better than that of RR patients (Figures S6(a) and S6(d)). In the
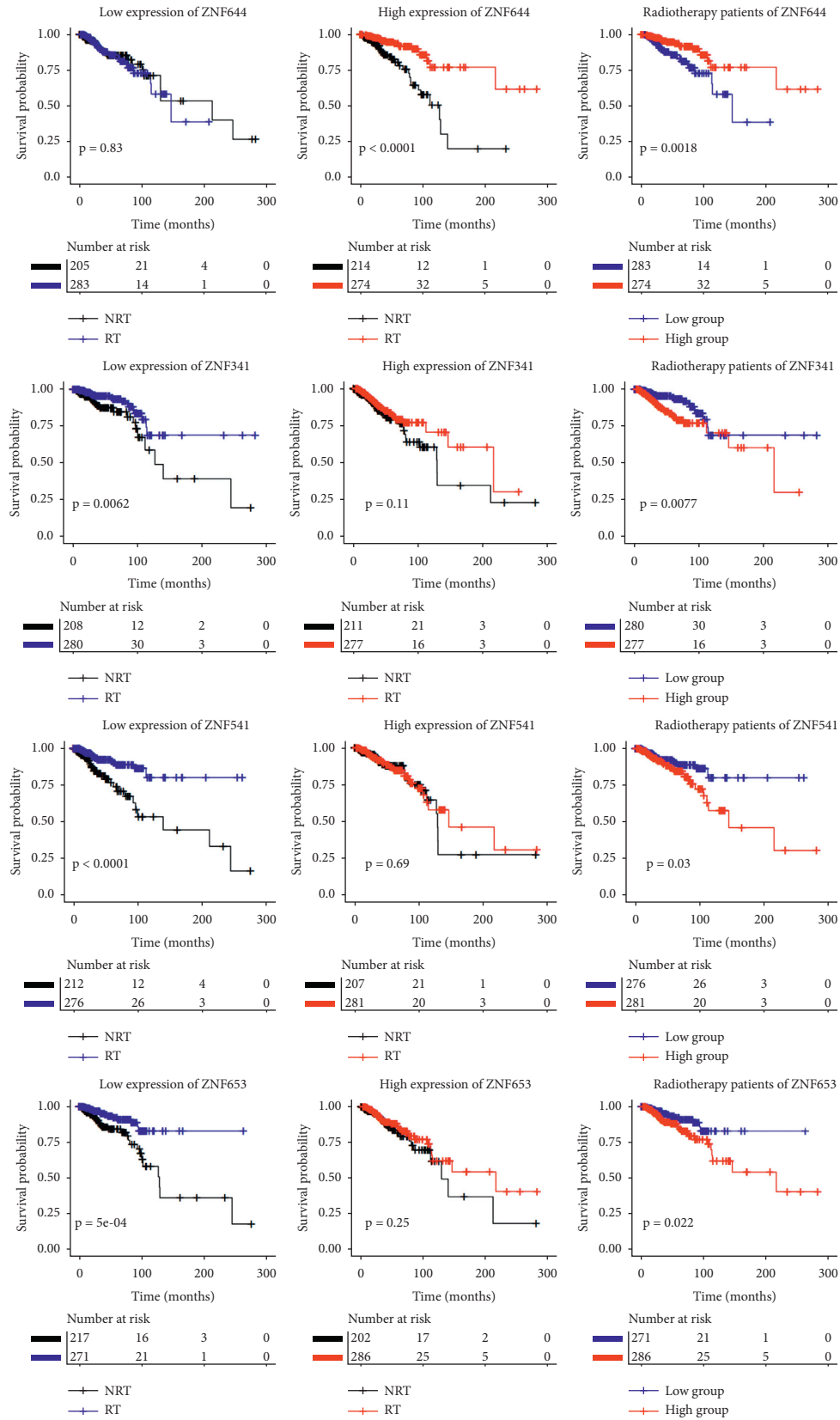
Figure 2: The survival curves comparison of radiotherapy and ZNF gene expression levels in TCGA data. RT, radiotherapy; NRT, nonradiotherapy.
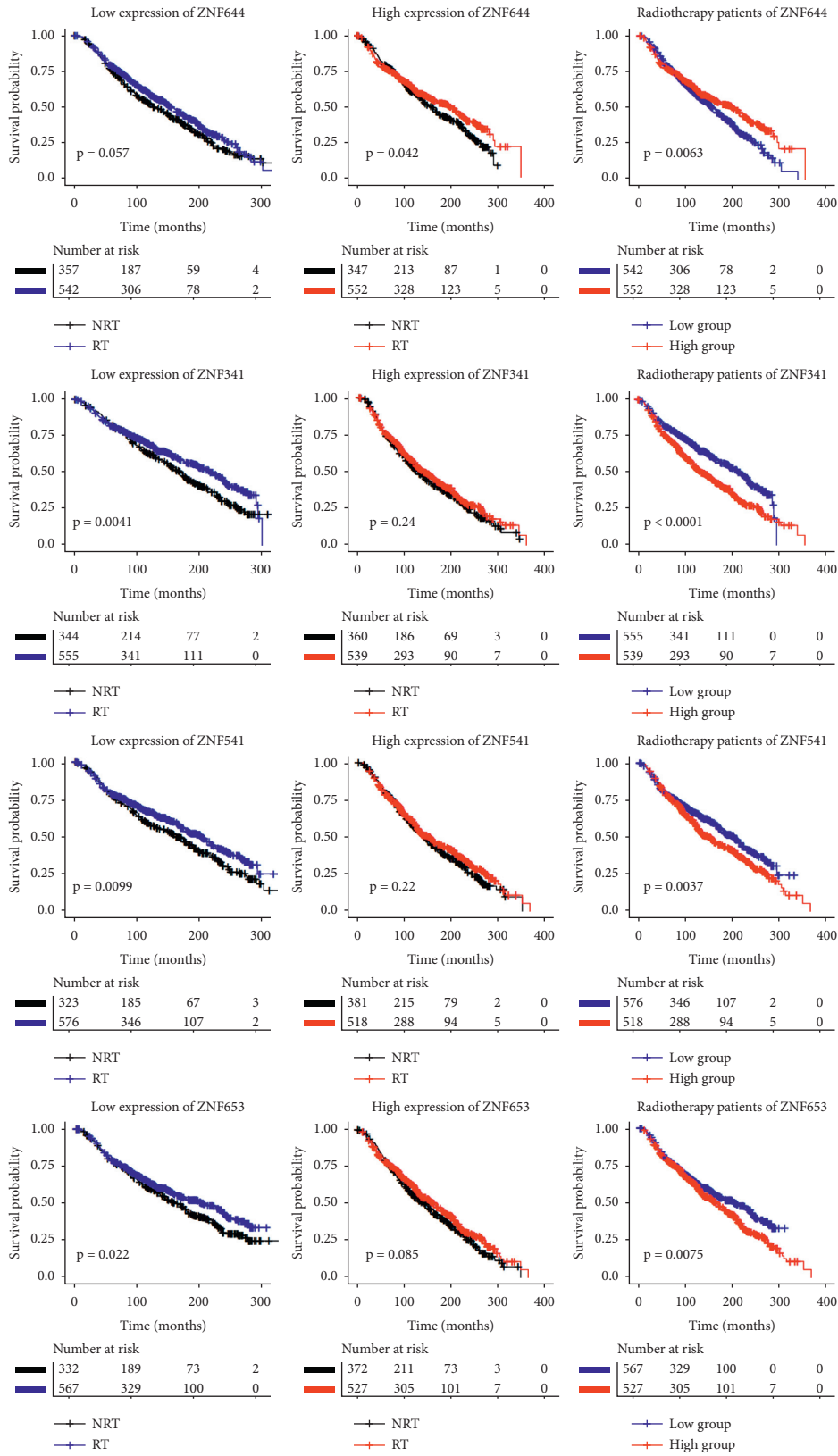
FIGURE 3: The survival curves comparison of radiotherapy and ZNF gene expression levels in METABRIC data. RT, radiotherapy; NRT, nonradiotherapy.

Gene	HR (95%CI)	P-value
ZNF644 RT (High vs Low)	0.467 (0.253-0.860)	0.015
ZNF341 RT (High vs Low)	2.046 (1.126-3.717)	0.019
ZNF54 RT (High vs Low)1	1.984 (1.081-3.640)	0.027
ZNF653 RT (High vs Low)	1.91 (1.042-3.501)	0.036

(a)

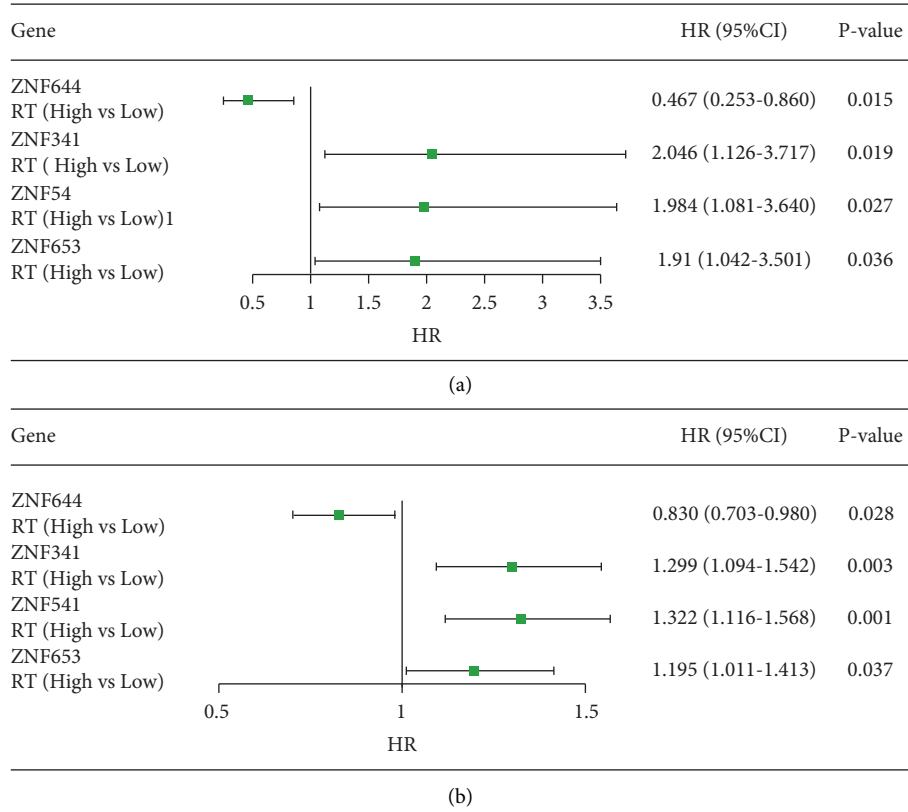| Gene	HR (95%CI)	P-value
ZNF644 RT (High vs Low)	0.830 (0.703-0.980)	0.028
ZNF341 RT (High vs Low)	1.299 (1.094-1.542)	0.003
ZNF541 RT (High vs Low)	1.322 (1.116-1.568)	0.001
ZNF653 RT (High vs Low)	1.195 (1.011-1.413)	0.037

(b)

FIGURE 4: The HR estimation for high expression (High) vs. low expression (Low) of ZNF genes in the multivariable Cox regression models of radiotherapy population. (a) TCGA data, the adjusted factors are age, pathological stage, histological type, ER, PR, HER, and chemotherapy. (b) METABRIC data, the adjusted factors are age, grade, histological type, ER, PR, HER, and chemotherapy.
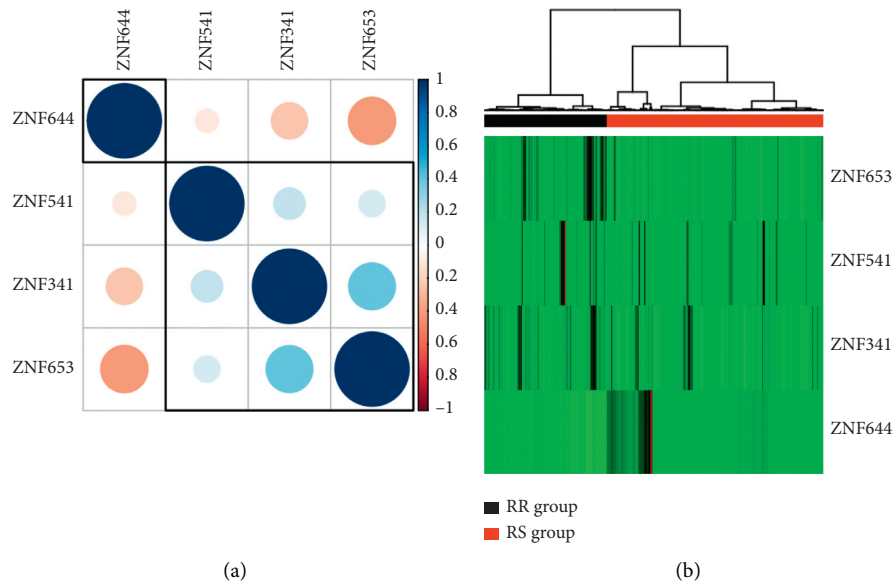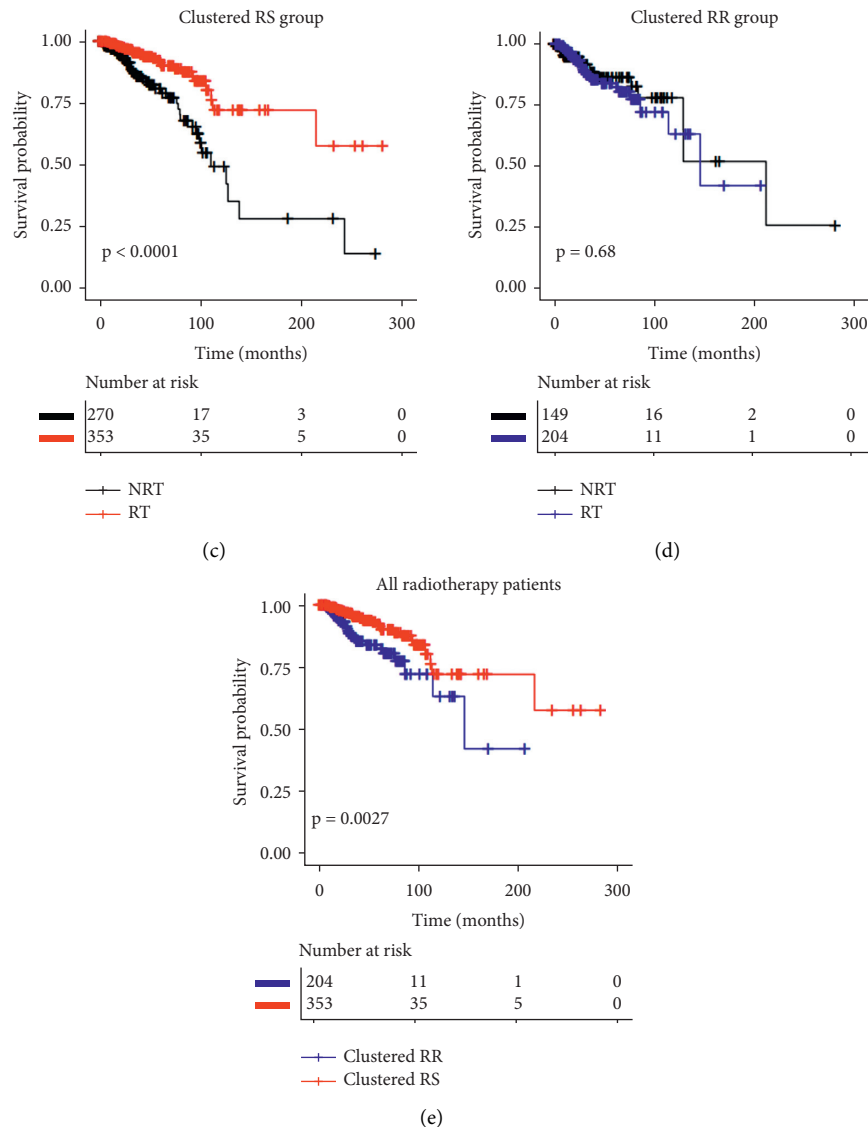

(a)

(b)

FIGURE 5: Continued.

(c)

(d)

(e)

FIGURE 5: TCGA data. (a) Correlation plot based on Pearson's correlation test results to show the correlation of gene expression among the 4 ZNF family members. (b) Hierarchical clustering analysis. The top red and black bands denote the clustered radiosensitive (RS) and radioresistant (RR) patients, respectively. (c–e) The survival curves under radiotherapy and nonradiotherapy for both clustered radiosensitive (RS) and radioresistant (RR) patients.

classed RS group, patients who received radiotherapy had a longer OS rate than those who did not receive radiotherapy (Figures S6(b) and S6(e)), but no significant difference was observed in the classed RR group (Figures S6(c) and S6(f)). Multivariate Cox regression analysis showed that this signature was a predictive factor independent of other clinical factors (Figure S7). The results of the CoxBoost model were similar to the results of the Cox model. KM plot showed that this gene signature did not have significant interaction with radiotherapy for predicting RFS (Figure S8).

3.7. Random Forests Prediction Model. To find out how useful these genes actually are for the prediction of survival after radiotherapy, we constructed a random survival forest model based on these four genes to predict the survival of

patients after radiotherapy. As shown in Figure S9, the model was symbiosis into 500 binary survival trees, and each survival tree had 20 terminal nodes on average. The out-of-bag error to verify and predict the survival outcome of the model was 33.24%. The model was used to predict the survival outcome of the METABRIC dataset, and the error rate is 50.24%. This model was used to predict the recurrence of patients in the GSE31863 dataset after radiotherapy, with an error rate of 54.54%.

3.8. Comparison with Previously Published Gene Signatures. We compared two existing gene signatures for predicting radiosensitivity, a 31-gene cluster and the radiation sensitivity index (RSI), in the TCGA and METABRIC datasets. Neither 31-gene signature nor RSI showed a significant
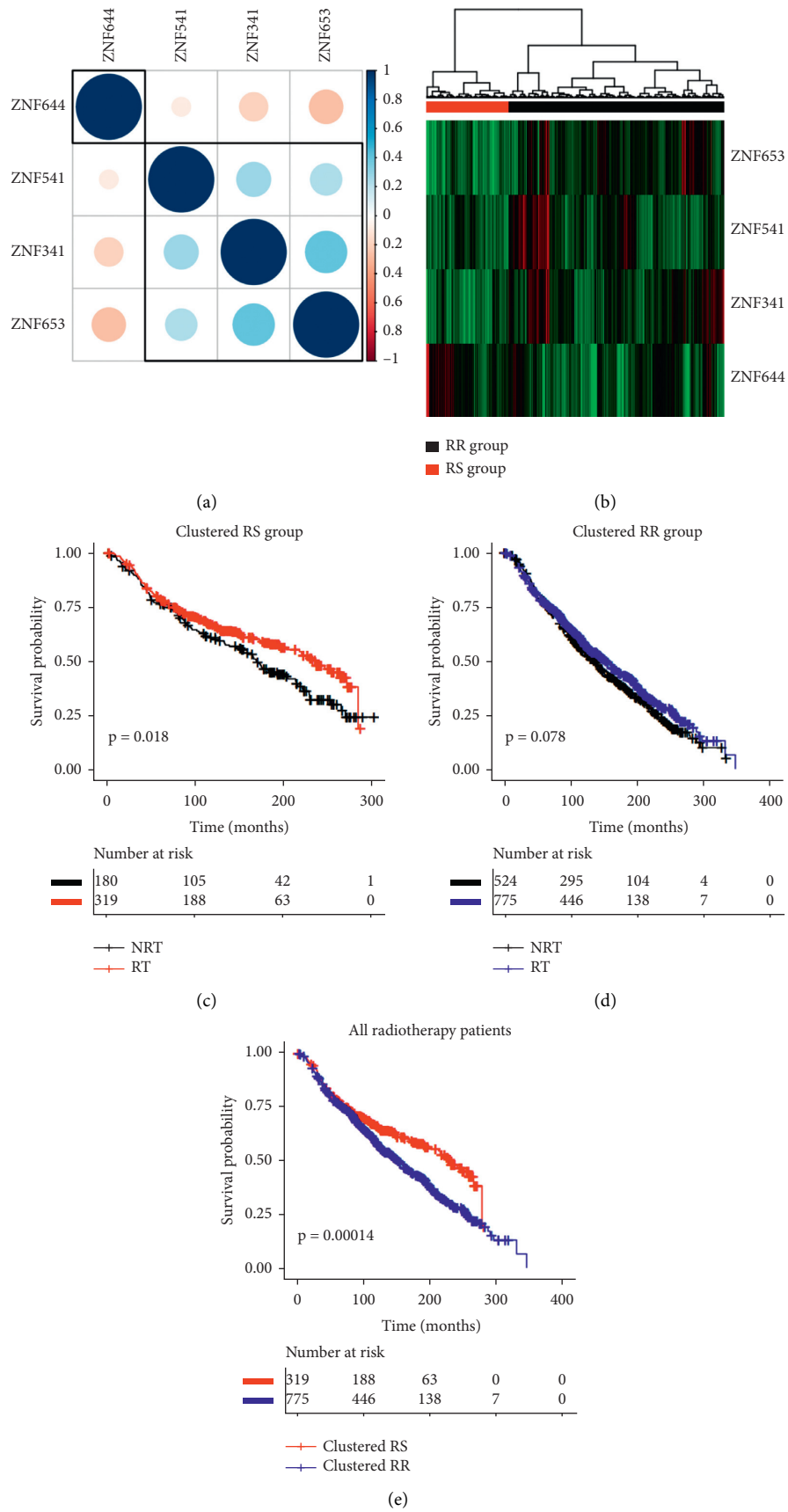
(a)

■ RR group
■ RS group

(b)

(c)

(d)

(e)

FIGURE 6: METABRIC data. (a) Correlation plot based on Pearson's correlation test results to show the correlation of gene expression among the 4 ZNF family members. (b) Hierarchical clustering analysis. The top red and black bands denote the clustered radiosensitive (RS) and radioresistant (RR) patients, respectively. (c–e) The survival curves under radiotherapy and nonradiotherapy for both clustered radiosensitive (RS) and radioresistant (RR) patients.

TABLE 1: Hierarchical clustering analyses.

	Raw		Adjusted	
	HR (95% CI)	P value	HR (95% CI)	P value
TCGA data				
Clustered RS (RT/NRT)	0.346 (0.210–0.570)	<0.001	0.253 (0.147–0.438)	<0.001
Clustered RR (RT/NRT)	1.143 (0.612–2.135)	0.676	1.411 (0.682–2.918)	0.353
RT patients (RS/RR)	0.436 (0.250–0.762)	0.004	0.437 (0.242–0.788)	0.006
METABRIC data				
Clustered RS (RT/NRT)	0.734 (0.567–0.950)	0.019	0.650 (0.491–0.862)	0.003
Clustered RR (RT/NRT)	0.882 (0.767–1.014)	0.078	0.868 (0.751–1.004)	0.056
RT patients (RS/RR)	0.687 (0.565–0.835)	<0.001	0.741 (0.606–0.907)	0.004

The adjusted factors included age, pathological stage (grade), histological type, ER, PR, HER, and chemotherapy. HR, hazard ratio; 95% CI, 95% confidence interval; RT, radiotherapy; NRT, nonradiotherapy; RS, radiosensitive; RR, radioresistant.
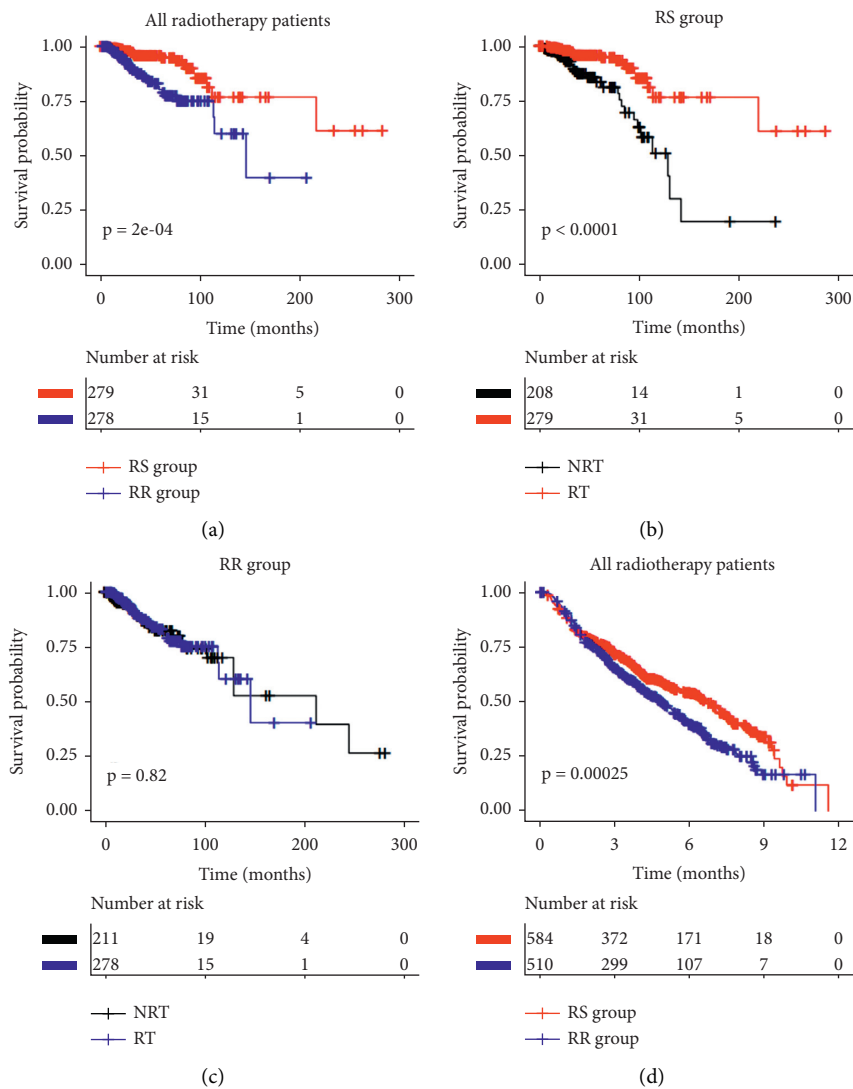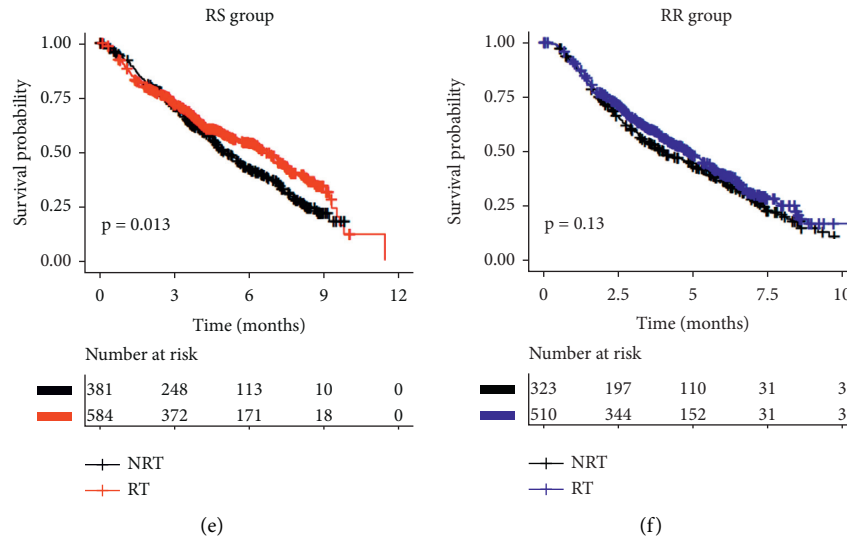

FIGURE 7: Continued.

(e)   (f)

FIGURE 7: Overall survival stratified by Cox prediction model. The survival curves comparison of radiotherapy and 4-gene-based signature in (a–f) TCGA data and (d–f) METABRIC data. RT, radiotherapy; NRT, nonradiotherapy; RS, radiosensitive; RR, radioresistant.
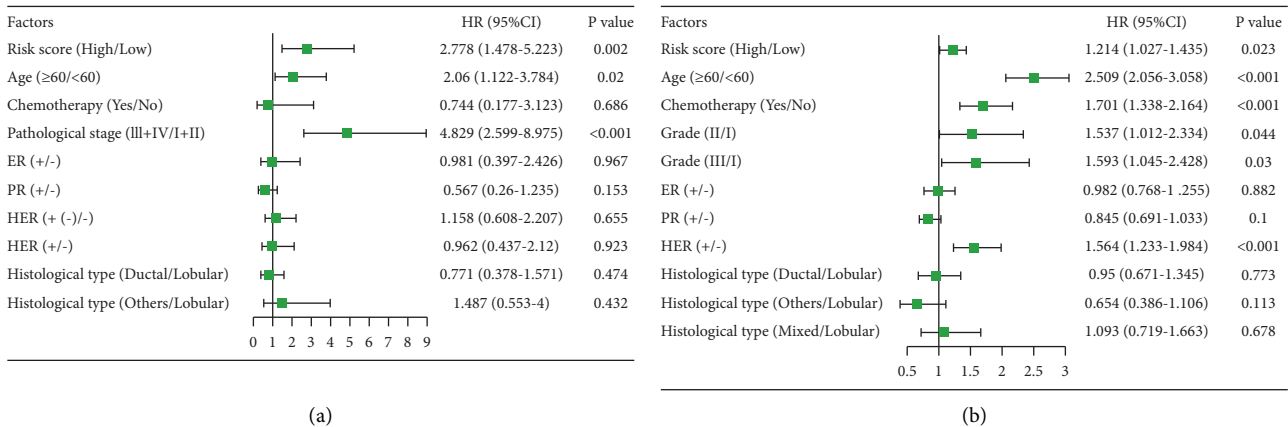

(a)   (b)

FIGURE 8: Forest plot of hazard ratios. Multivariable Cox regression analyses of the risk score, age, pathological stage (grade), histological type, ER, PR, HER, and chemotherapy in TCGA data (a) and METABRIC data (b). The blue squares represent the hazard ratio (HR). 95% CI, 95% confidence interval.

interaction with radiotherapy in the two datasets. Using hierarchical clustering based on the gene expression profile of the 31-gene signature, we divided the total patients into two clusters (Figures S10 and S11). Patients in the red bar were assigned to the RS group, and the survival rate of patients after radiotherapy was higher than that of patients without radiotherapy. Conversely, in the RR groups, there was no significant difference in survival rate between the radiotherapy and nonradiotherapy groups. However, among radiotherapy patients, the survival rate of clustered RS patients was not significantly better than that of clustered RR patients.

For the RSI, the 25th percentile of RSI in patients receiving radiotherapy was used as the cut-off value for dividing patients into RS and RR groups, as in previous studies [21]. RSI did not show significant results in predicting radiotherapy benefits in either the TCGA or METABRIC datasets (Figure S12).

4. Discussion

Radiotherapy is one of the main methods of breast cancer treatment, which has made substantial progress over the years with the development of improved treatment plans and implementation methods. However, through multivariate Cox regression analysis, we found that radiotherapy was not an important clinical factor related to the overall survival of the whole cohort of breast cancer patients in the TCGA dataset. However, it was significantly associated with overall survival in the METABRIC validation data. And, radiotherapy was significantly associated with recurrence-free survival in the GSE31863 dataset. Therefore, we concluded that not all patients can benefit from radiotherapy, nor is the absolute benefit equal across risk groups. Due to differences in tumor radiosensitivity, the response of cancer patients to radiotherapy varies greatly, while serious adverse

consequences are universal. Therefore, the most important issue in this respect is to be able to accurately predict which patients will benefit from radiotherapy before beginning treatment. In the era of precision medicine, exploring the radiosensitivity of patients at the genetic level has attracted widespread attention. Cancer risk and radiosensitivity are commonly associated with deficiencies in DNA-damage recognition, DNA repair, cell cycle checkpoint activation, stress responses and signal transduction, premature senescence pathways, and/or apoptosis [39, 40]. It is known that hundreds of gene products are involved in determining the radiosensitivity of cells, and this number is still increasing as data are generated [40]. The zinc finger protein gene family is one of the largest human gene families and plays an important role in transcriptional regulation. In this study, we explored whether and which ZNF genes may be useful as radiosensitivity signatures and for personalized medical decisions.

In the present study, we analyzed female breast cancer patient clinical and gene expression information from the TCGA database. A total of 511 ZNF genes and 976 patients was involved in our initial analysis. First, ZNF genes were screened by using a two-step Cox regression analysis. The main idea of this step was to determine whether patients with lower or higher expression of ZNF genes were differently sensitive to radiotherapy. Next, univariate and multivariate Cox regression analysis was used to analyze the relationship between ZNF gene expression and overall survival in the patient cohort receiving radiotherapy. The results of TCGA and METABRIC data analyses indicated that ZNF gene signatures had statistically significant interactions with radiotherapy independent of clinical variables. These selected ZNF genes revealed that the effect of radiotherapy on patients is different according to their different levels of expression. It was assumed that some of the patients express several of these sensitivity marker genes, and those patients would be expected to have a relatively high probability of survival. They are referred to here as radiosensitive patients, and accordingly, the sensitivity genes were designated as providing a radiosensitive gene signature. We have identified a total of 4 ZNF genes related to the radiosensitivity of breast cancer. High expression of ZNF644 might be associated with the radiosensitivity of breast cancer patients, and reciprocally, low expression of the other 3 genes (ZNF341, ZNF544, and ZNF653) marked patients in the radiosensitive group. We found crossing survival curves with significant P values, e.g., survival curves for the gene ZNF341 (Figure 2). Landmark analyses were performed according to a landmark point at 112 months, with the KM survival analysis for events that occurred up to 112 months and events that occurred between 112 months and the end of the follow-up period. Figure S2 shows the results of the landmark analyses of the overall survival and its components. Among patients receiving radiotherapy, patients with low expression of ZNF341 had significantly longer survival than patients with a high expression before the landmark point of 112 months. After the landmark point of 112 months, the survival rate of the low expression group was higher than that of the high expression group, but the

difference was not statistically significant. The predictive value of longer-term radiosensitivity according to ZNF341 expression needs further study.

To explore the potential relationship between the radiosensitivity-associated genes, correlation coefficients between the 4 ZNF genes were calculated. These results indicated that the 3 genes may share some common features or functions, and the 4 ZNF genes may have a synergistic prognostic effect. Due to the limited information available, it has not been proven possible to establish transcriptional regulatory networks for these four genes. The result of hierarchical cluster analysis showed that all breast cancer patients could be divided into two groups according to their expression of these 4 genes. Patients clustered as radiosensitive had significantly longer survival when compared with nonradiosensitive patients. Thus, the 4-mRNA ZNF signature comprising ZNF644, ZNF341, ZNF541, and ZNF653 was established. Cox regression and CoxBoost analysis showed that there was a significant interaction between this signature and the overall survival of radiotherapy. The random survival forest model also showed that these four genes could predict the survival status of patients after radiotherapy. We used clustering, Cox modeling, CoxBoost modeling, and random survival forest methods to demonstrate that these four genes have the value of predicting radiosensitivity in breast cancer patients. These radiosensitivity marker genes were designed to be predictive because they were identified in patients who received radiation therapy, using an appropriate control group of patients who did not receive radiotherapy. Furthermore, in the GSE31863 dataset, the RS patients did not show significantly longer recurrence-free survival than the RR patients. However, in the cluster RS group, the radiotherapy patients had significantly longer recurrence-free survival than the nonradiotherapy patients. In the RS group, recurrence-free survival was longer in radiotherapy patients than in nonradiotherapy patients, although no significant difference was achieved. However, this still provides evidence for the radiosensitivity study of breast cancer recurrence-free survival. Due to the limited sample size of GSE31863, the extrapolation of this gene signature to predict other outcomes of radiosensitivity requires further research. In addition, by comparing with the previously developed radiosensitivity gene signatures, we found that this ZNF signature has obvious advantages. In both clustering and modeling, the overall survival of RS patients after radiotherapy was significantly higher than that of RR patients after radiotherapy. There was a significant interaction between the 4-gene signature and radiotherapy according to the Cox regression analysis, suggesting that the ZNF signature could have predictive value. This ZNF signature has the potential to select patients who are most likely to benefit from radiotherapy.

The zinc finger protein (KRAB-ZFP) family accounts for at least one-third of mammalian ZNF proteins. KRAB-ZFP contains an effector motif denoted the Krüppel-associated box or KRAB, which seems to be unique to vertebrates [23, 41, 42]. The KRAB-ZNF family is very complex, containing a large number of homologous genes, gene isoforms,

and pseudogenes [26]. There are more than 800 ZNF genes in the human genome, including 8000 ZNF domains. Interestingly, approximately 266 of these genes are found on chromosome 19, of which 202 are KRAB-ZNF genes [43]. ZNF541 and ZNF653 are both located on chromosome 19. Although all chromosomes are unique, chromosome 19 can be considered unusual in that it possesses the highest gene density of all human chromosomes, more than twice the average for the whole genome. Studies have found that gene mutations on chromosome 19 are associated with the occurrence of malignant tumors [44]. There are few reports on the biological functions of ZNF541 and ZNF653. Some studies have pointed out that the aberrant expression of ZNF644 may be a potential pathogenic factor for rheumatoid arthritis and strong myopia [45, 46].

ZNF341, a transcription factor with 12 Cys2His2 zinc fingers, regulates STAT3 (signal transducer and activator of transcription 3) and many other genes, including STAT1, because it can bind to the STAT3 promoter and activate the transcription of the STAT3 reporter plasmid [47]. STAT3 is a central regulator of tumor immune tolerance. Importantly, STAT3 can act as a negative feedback regulator, and targeted inhibition of STAT3 can enhance the efficacy of radiotherapy for head and neck cancer [48], lung cancer [49], and breast cancer [50]. Studies have shown that transcription factor ZNF341 is a positive regulator of STAT3 expression [51]. It is known that ZNF341 transcriptionally regulates the expression of STAT3 which seems to be necessary for the efficient repair of damaged DNA. Therefore, based on Comet assay results, it was hypothesized that increased radiosensitivity was due to the lack of efficient DNA repair [52]. Further studies are needed to better understand how ZNF341 regulates radiotherapy responses.

The main limitation of our study is the use of retrospective cohorts. We also need some related experiments to verify these gene signatures. Ideally, predictive biomarkers should be tested in prospective randomized controlled trials. However, we have verified these sensitivity signatures in two large retrospective databases, which provide directions for follow-up clinical studies. There are no reports on associations between ZNF genes and radiosensitivity of breast cancer patients to date. Therefore, the mechanisms of ZNF genes and radiosensitivity of breast cancer need further study.

5. Conclusion

Associations between cancer radiosensitivity and ZNF family gene expression have not been reported before. In the present study, we explored this association by using the TCGA and METABRIC data. According to our results, patients with high expression of ZNF644 were sensitive to radiotherapy, while low expression of ZNF314, ZNF544, and ZNF653 marked those who were radiosensitive. Although the biological mechanism responsible for these findings is not clear, and further work is necessary, at least the present study suggests that these 4 ZNF genes could act as radiosensitivity signatures for breast cancer patients. Our study provides important clues for determining the radiosensitivity of breast cancer and also offers a new avenue for clinical improvement in breast cancer treatment.

Data Availability

The datasets used in the present study are available from The Cancer Genome Atlas database (http://cancergenome.nih.gov/), Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) database (http://www.cbioportal.org/), and GEO database (https://www.ncbi.nlm.nih.gov/geo/).

Disclosure

The funding body did not play any roles in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Derui Yan, Mingjing Shen, Ping Zeng, and Zaixiang Tang contributed to study conception and design; Derui Yan, Mingjing Shen, and Zixuan Du were responsible for data collection and cleaning; Derui Yan, Mingjing Shen, and Zaixiang Tang were involved in real data analysis and interpretation; Derui Yan and Mingjing Shen drafted the manuscript, with assistance from Jianping Cao and Ye Tian. All authors reviewed the manuscript. Derui Yan and Mingjing Shen contributed equally to this work.

Acknowledgments

Supplementary Materials

Figure S1: expression distributions of ZNF genes of patients with breast cancer. Figure S2: the landmark analysis of ZNF341 in radiotherapy patients with TCGA data. Figure S3: GSE31863 data. (a) Hierarchical clustering analysis. The top red and black bands denote the clustered radiosensitive (RS) and radioresistant (RR) patients, respectively. (b–d) The survival curves under radiotherapy and nonradiotherapy for both clustered radiosensitive (RS) and radioresistant (RR) patients. Figure S4: recurrence-free survival stratified by Cox

model in GSE31863. (a) RS group. (b) RR group. (c) All radiotherapy patients. RT, radiotherapy; NRT, non-radiotherapy; RS, radiosensitive; RR, radioresistant. Figure S5: coefficient paths for the TCGA data. Figure S6: overall survival stratified by CoxBoost prediction model. The survival curves comparison of radiotherapy and 4-gene-based signature in (a–c) TCGA data and (d–f) METABRIC data. RT, radiotherapy; NRT, nonradiotherapy; RS, radiosensitive; RR, radioresistant. Figure S7: the HR estimation for high expression (High) vs. low expression (Low) of ZNF genes in the CoxBoost prediction model of radiotherapy patients. (a) TCGA data, the adjusted factors are age, pathological stage, histological type, ER, PR, HER, and chemotherapy. (b) METABRIC data, the adjusted factors are age, grade, histological type, ER, PR, HER, and chemotherapy. Figure S8: recurrence-free survival stratified by CoxBoost prediction model in GSE31863. (a) All radiotherapy patients. (b) RS group. (c) RR group. RT, radiotherapy; NRT, non-radiotherapy; RS, radiosensitive; RR, radioresistant. Figure S9: a random forests prediction model. (a) Model prediction error rate for different numbers of survival trees. The prediction results of the random survival forest model on the (b) TCGA dataset, (c) METABRIC dataset, and (d) GSE31863. Figure S10: TCGA data. (a) Hierarchical clustering analysis based on 31 genes. The top red and black bands denote the clustered radiosensitive (RS) and radioresistant (RR) patients, respectively. (b–d) The survival curves under radiotherapy and nonradiotherapy for both clustered radiosensitive (RS) and radioresistant (RR) patients. Figure S11: METABRIC data. (a) Hierarchical clustering analysis based on 31 genes. The top red and black bands denote the clustered radiosensitive (RS) and radioresistant (RR) patients, respectively. (b–d) The survival curves under radiotherapy and non-radiotherapy for both clustered radiosensitive (RS) and radioresistant (RR) patients. Figure S12: overall survival stratified by RSI. The survival curves comparison of radiotherapy and 4-gene-based signature in (a–c) TCGA data and (d–f) METABRIC data. RT, radiotherapy; NRT, non-radiotherapy; RS, radiosensitive; RR, radioresistant. Table S1: univariate and multivariate Cox regression analysis of breast cancer clinical indicators and total survival in the TCGA data. Table S2: univariate and multivariate Cox regression analysis of breast cancer clinical indicators and total survival in the METABRIC data. Table S3: univariate Cox regression analysis of radiotherapy patients and nonradiotherapy patients in ZNF genes high/low expression subgroups in the TCGA data. Table S4: univariate Cox regression analysis of ZNF genes expression and overall survival in the radiotherapy patients in the TCGA data. Table S5: univariate Cox regression analysis of radiotherapy patients and nonradiotherapy patients in ZNF genes high/low expression subgroups in the METABRIC data. Table S6: multivariate Cox regression analysis of ZNF644, ZNF341, ZNF541, and ZNF653 genes in radiotherapy patients with TCGA data. (*Supplementary Materials*)

References

[1] J. Ferlay, M. Colombet, I. Soerjomataram et al., "Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods," *International Journal of Cancer*, vol. 144, no. 8, pp. 1941–1953, 2019.

[2] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[3] S. A. Castaneda and J. Strasser, "Updates in the treatment of breast cancer with radiotherapy," *Surgical Oncology Clinics of North America*, vol. 26, no. 3, pp. 371–382, 2017.

[4] J. Poleszczuk, K. Luddy, L. Chen et al., "Neoadjuvant radiotherapy of early-stage breast cancer and long-term disease-free survival," *Breast Cancer Research*, vol. 19, no. 1, p. 75, 2017.

[5] Ebctcg, "Effect of radiotherapy after mastectomy and axillary surgery on 10-year recurrence and 20-year breast cancer mortality: meta-analysis of individual patient data for 8135 women in 22 randomised trials," *The Lancet*, vol. 383, no. 9935, pp. 2127–2135, 2014.

[6] C. Speers and L. J. Pierce, "Postoperative radiotherapy after breast-conserving surgery for early-stage breast cancer," *JAMA Oncology*, vol. 2, no. 8, pp. 1075–1082, 2016.

[7] S. Darby, P. Mcgale, and C. Correa, "Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10 801 women in 17 randomised trials," *The Lancet*, vol. 378, no. 9804, pp. 1707–1716, 2011.

[8] E. Lee, C. Takita, J. L. Wright et al., "Characterization of risk factors for adjuvant radiotherapy-associated pain in a tri-racial/ethnic breast cancer population," *Pain*, vol. 157, no. 5, pp. 1122–1131, 2016.

[9] J. J. Hu, J. J. Urbanic, L. D. Case et al., "Association between inflammatory biomarker C-reactive protein and radiotherapy-induced early adverse skin reactions in a multiracial/ethnic breast cancer population," *Journal of Clinical Oncology*, vol. 36, no. 24, pp. 2473–2482, 2018.

[10] Z. Tian, J. Tang, X. Liao, Q. Yang, Y. Wu, and G. Wu, "Identification of a 9-gene prognostic signature for breast cancer," *Cancer Medicine*, vol. 9, no. 24, pp. 9471–9484, 2020.

[11] S. Huang, L. Murphy, and W. Xu, "Genes and functions from breast cancer signatures," *BMC Cancer*, vol. 18, no. 1, p. 473, 2018.

[12] J. Sun, X. Chen, Z. Wang et al., "A potential prognostic long non-coding RNA signature to predict metastasis-free survival of breast cancer patients," *Scientific Reports*, vol. 5, no. 1, Article ID 16553, 2015.

[13] S. Bao, H. Zhao, J. Yuan et al., "Computational identification of mutator-derived lncRNA signatures of genome instability for improving the clinical outcome of cancers: a case study in breast cancer," *Briefings in Bioinformatics*, vol. 21, no. 5, pp. 1742–1755, 2020.

[14] S. Bao, T. Hu, J. Liu et al., "Genomic instability-derived plasma extracellular vesicle-microRNA signature as a minimally invasive predictor of risk and unfavorable prognosis in breast cancer," *Journal of Nanobiotechnology*, vol. 19, no. 1, p. 22, 2021.

[15] K. V. Ballman, "Biomarker: predictive or prognostic?" *Journal of Clinical Oncology*, vol. 33, no. 33, pp. 3968–3971, 2015.

[16] H. Sun, J. Gu, Z. Li et al., "Low expression of GLIS2 gene might associate with radiosensitivity of gastric cancer," *Journal of Oncology*, vol. 2019, Article ID 2934925, 10 pages, 2019.

[17] B.-S. Jang and I. A. Kim, "A radiosensitivity gene signature and PD-L1 status predict clinical outcome of patients with

invasive breast carcinoma in the Cancer Genome Atlas (TCGA) dataset," *Radiotherapy and Oncology*, vol. 124, no. 3, pp. 403–410, 2017.

[18] J. Meng, P. Li, Q. Zhang, Z. Yang, and S. Fu, "A radiosensitivity gene signature in predicting glioma prognostic via EMT pathway," *Oncotarget*, vol. 5, no. 13, pp. 4683–4693, 2014.

[19] H. S. Kim, S. C. Kim, S. J. Kim et al., "Identification of a radiosensitivity signature using integrative metaanalysis of published microarray data for NCI-60 cancer cells," *BMC Genomics*, vol. 13, no. 1, p. 348, 2012.

[20] S. Eschrich, H. Zhang, H. Zhao et al., "Systems biology modeling of the radiation sensitivity network: a biomarker discovery platform," *International Journal of Radiation Oncology, Biology, Physics*, vol. 75, no. 2, pp. 497–505, 2009.

[21] S. A. Eschrich, W. J. Fulp, Y. Pawitan et al., "Validation of a radiosensitivity molecular signature in breast cancer," *Clinical Cancer Research*, vol. 18, no. 18, pp. 5134–5143, 2012.

[22] S. K. Rhie, L. Yao, Z. Luo et al., "ZFX acts as a transcriptional activator in multiple types of human tumors by binding downstream from transcription start sites at the majority of CpG island promoters," *Genome Research*, vol. 28, no. 3, pp. 310–320, 2018.

[23] T. Collins, J. R. Stone, and A. J. Williams, "All in the family: the BTB/POZ, KRAB, and SCAN domains," *Molecular and Cellular Biology*, vol. 21, no. 11, pp. 3609–3615, 2001.

[24] J. H. Laity, B. M. Lee, and P. E. Wright, "Zinc finger proteins: new insights into structural and functional diversity," *Current Opinion in Structural Biology*, vol. 11, no. 1, pp. 39–46, 2001.

[25] H. Yang, Y. Lu, Y. Zheng et al., "shRNA-mediated silencing of ZFX attenuated the proliferation of breast cancer cells," *Cancer Chemotherapy and Pharmacology*, vol. 73, no. 3, pp. 569–576, 2014.

[26] M. Machnik, R. Cylwa, K. Kiełczewski et al., "The expression signature of cancer-associated KRAB - ZNF factors identified in TCGA pan-cancer transcriptomic data," *Molecular Oncology*, vol. 13, no. 4, pp. 701–724, 2019.

[27] Y. Xiao, T. Xiang, X. Luo et al., "Zinc-finger protein 545 inhibits cell proliferation as a tumor suppressor through inducing apoptosis and is disrupted by promoter methylation in breast cancer," *PLoS One*, vol. 9, no. 10, Article ID e110990, 2014.

[28] H. Yang, L. Pan, C. Xu et al., "Overexpression of tumor suppressor gene ZNF750 inhibits oral squamous cell carcinoma metastasis," *Oncology Letters*, vol. 14, no. 5, pp. 5591–5596, 2017.

[29] A. Luo, X. Zhang, L. Fu, Z. Zhu, and J.-T. Dong, "Zinc finger factor ZNF121 is a MYC-interacting protein functionally affecting MYC and cell proliferation in epithelial cells," *Journal of Genetics and Genomics*, vol. 43, no. 12, pp. 677–685, 2016.

[30] B. Liu, X. Xing, X. Li, Q. Guo, T. Xu, and K. Xu, "ZNF259 promotes breast cancer cells invasion and migration via ERK/GSK3β/snail signaling," *Cancer Management and Research*, vol. 10, pp. 3159–3168, 2018.

[31] J. A. Vendrell, A. Thollet, N. T. Nguyen et al., "ZNF217 is a marker of poor prognosis in breast cancer that drives epithelial-mesenchymal transition and invasion," *Cancer Research*, vol. 72, no. 14, pp. 3593–3606, 2012.

[32] R. Otsuka, Y. Akutsu, H. Sakata et al., "ZNF750 expression is a potential prognostic biomarker in esophageal squamous cell carcinoma," *Oncology*, vol. 94, no. 3, pp. 142–148, 2018.

[33] B. Schnabl, D. Valletta, G. Kirovski, and C. Hellerbrand, "Zinc finger protein 267 is up-regulated in hepatocellular carcinoma

and promotes tumor cell proliferation and migration," *Experimental and Molecular Pathology*, vol. 91, no. 3, pp. 695–701, 2011.

[34] L. Hu, W. Wang, J. Cai et al., "Aberrant expression of ZNF268 alters the growth and migration of ovarian cancer cells," *Oncology Letters*, vol. 6, no. 1, pp. 49–54, 2013.

[35] Y. Zhu, P. Qiu, and Y. Ji, "TCGA-assembler: open-source software for retrieving and processing TCGA data," *Nature Methods*, vol. 11, no. 6, pp. 599-600, 2014.

[36] Z. Tang, Q. Zeng, Y. Li et al., "Predicting radiotherapy response for patients with soft tissue sarcoma by developing a molecular signature," *Oncology Reports*, vol. 38, no. 5, pp. 2814–2824, 2017.

[37] V. Weyer and H. Binder, "A weighting approach for judging the effect of patient strata on high-dimensional risk prediction signatures," *BMC Bioinformatics*, vol. 16, no. 1, p. 294, 2015.

[38] C. Zemmour, F. Bertucci, P. Finetti et al., "Prediction of early breast cancer metastasis from DNA microarray data using high-dimensional cox regression models," *Cancer Informatics*, vol. 14s2, no. Suppl 2, pp. CIN.S17284–38, 2015.

[39] S. Hornhardt, U. Rößler, W. Sauter et al., "Genetic factors in individual radiation sensitivity," *DNA Repair*, vol. 16, pp. 54–65, 2014.

[40] M. B. Parliament and D. Murray, "Single nucleotide polymorphisms of DNA repair genes as predictors of radioresponse," *Seminars in Radiation Oncology*, vol. 20, no. 4, pp. 232–240, 2010.

[41] S. Huntley, D. M. Baggott, A. T. Hamilton et al., "A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors," *Genome Research*, vol. 16, no. 5, pp. 669–677, 2006.

[42] W. Wang, J. Cai, Y. Lin et al., "Zinc fingers function cooperatively with KRAB domain for nuclear localization of KRAB-containing zinc finger proteins," *PLoS One*, vol. 9, no. 3, Article ID e92155, 2014.

[43] J. Grimwood, L. A. Gordon, A Olsen et al., "The DNA sequence and biology of human chromosome 19," *Nature*, vol. 428, no. 6982, pp. 529–535, 2004.

[44] A. M. Rose, "Cancer and the junkyard chromosome: how repeat DNA sequence on chromosome 19 influences risk of malignant disease," *Oncotarget*, vol. 9, no. 62, pp. 31942–31944, 2018.

[45] H. Ye, J. Zhang, J. Wang et al., "CD4 T-cell transcriptome analysis reveals aberrant regulation of STAT3 and Wnt signaling pathways in rheumatoid arthritis: evidence from a case-control study," *Arthritis Research & Therapy*, vol. 17, no. 1, p. 76, 2015.

[46] F. Liu, J. Wang, Y. Xing, and T. Li, "Mutation screening of 17 candidate genes in a cohort of 67 probands with early-onset high myopia," *Ophthalmic and Physiological Optics*, vol. 40, no. 3, pp. 271–280, 2020.

[47] A. August, "Who regulates whom: ZNF341 is an additional player in the STAT3/T H 17 song," *Science Immunology*, vol. 3, no. 24, 2018.

[48] D. Moreira, S. Sampath, H. Won et al., "Myeloid cell-targeted STAT3 inhibition sensitizes head and neck cancers to radiotherapy and T cell-mediated immunity," *Journal of Clinical Investigation*, vol. 131, no. 2, p. 131, 2021.

[49] J.-Y. Kim, H.-J. Kim, C.-W. Jung, T. S. Lee, E. H. Kim, and M.-J. Park, "CXCR4 uses STAT3-mediated slug expression to maintain radioresistance of non-small cell lung cancer cells: Emerges as a potential prognostic biomarker for lung cancer," *Cell Death & Disease*, vol. 12, no. 1, 2021.

[50] B. Johari, M. Rahmati, L. Nasehi, Y. Mortazavi, M. H. Faghfoori, and H. Rezaeejam, "Evaluation of STAT3 decoy oligodeoxynucleotides' synergistic effects on radiation and/or chemotherapy in metastatic breast cancer cell line," *Cell Biology International*, vol. 44, no. 12, pp. 2499–2511, 2020.

[51] S. Frey-Jakobs, J. M. Hartberger, M. Fliegauf et al., "ZNF341 controls STAT3 expression and thereby immunocompetence," *Science Immunology*, vol. 3, no. 24, 2018.

[52] S. Cekic, J. M. Hartberger, S. Frey-Jakobs et al., "Cancer tendency in a patient with ZNF341 deficiency," *Journal of Clinical Immunology*, vol. 40, no. 3, pp. 534–538, 2020.