European Review for Medical and Pharmacological Sciences 2023; 27: 256-274

Integrative analyses of radiation-related genes and biomarkers associated with breast cancer

W.-C. DAN[1], X.-Y. GUO[2], G.-Z. ZHANG[1], S.-L. WANG[2], M. DENG[2], J.-L. LIU[3]

[1]Department of Dermatology, Beijing Hospital of Traditional Chinese Medicine, Capital Medical University
[2]Department of Radiation Oncology State Key and Laboratory of Molecular Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
[3]Guang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing, China

Abstract. – OBJECTIVE: In addition to significantly reducing breast cancer recurrence risk, radiotherapy also prolongs patients' lives. However, radiotherapy-related genes and biomarkers still remain poorly understood. The present study aimed to identify radiation-associated genes in breast cancer.

MATERIALS AND METHODS: Breast cancer data were downloaded from Gene Expression Omnibus (GEO) and UCSC Xena database. The gene ontology (GO) enrichment and gene set enrichment analysis (GSEA) were performed for annotation and integrated discovery. Protein-protein interaction (PPI) network was constructed by STRING database and hub genes were identified. Then, immunohistochemistry and tissue expression of key genes was analyzed by using the Human Protein Atlas (HPA) and GEPIA database. Genes associated with prognosis were identified by performing univariate cox analysis.

RESULTS: We identified 341 differentially expressed genes related to radiotherapy in breast cancer patients. PPI analysis revealed a total of 129 nodes and 516 interactions and identified five hub genes (EGFR, FOS, ESR1, JUN, and IL6). In addition, 11 SDEGs THBS1, SERPINA11, NFIL3, METTL7A, KCTD12, HSPA6, EGR1, DDIT4, CCDC3, C11orf96, and BCL2A1 candidate genes can be used as potential diagnostic markers. The calibration curve and ROC indicate good probability consistencies of 3-years and 5-year survival rates of patients between estimation and observation.

CONCLUSIONS: Our findings provide novel insight into the functional characteristics of breast cancer through integrative analysis of GEO data and suggest potential biomarkers and therapeutic targets for breast cancer.

Key Words:
Bioinformatics, Radiation therapy, Biomarkers, Breast cancer.

Introduction

Breast cancer is the most common cancer type among women worldwide, and it is also the leading cause of tumor-related female death[1-3]. According to the latest data from the International Agency for Research on Cancer (IARC) of the World Health Organization, more than 2.26 million breast cancer cases occurred in 2020, the first time surpassing lung cancer and becoming the most prominent cancer in the world[4-6]. In 2020, about 420,000 new cases of breast cancer and nearly 120,000 deaths occurred in China. Chemotherapy, radiotherapy, surgery, targeted therapy, and immunotherapy treatment strategies have greatly improved the prognosis of breast cancer patients[7]. Radiotherapy is an essential treatment for breast cancer. It can reduce the recurrence and prolong the survival time of breast cancer patients and is a necessary palliative treatment for patients with inoperable locally advanced and metastatic breast cancer[8-10]. However, there is increasing evidence that the response of breast cancer patients to radiation therapy is heterogeneous, which significantly impacts clinical effectiveness and quality of life. Hence, this study aimed to identify potential diagnostic biomarkers and biological functions related to breast cancer from the Gene Expression Omnibus. Furthermore, cross-validation investigated radiation-related differentially expressed genes (DEGs) to distinguish patients with breast cancer from healthy controls. Moreover, the biological processes (BPs) involved were analyzed using gene ontology (GO) enrichment and gene set enrichment analysis (GSEA) pathways for the SDEGs. In addition, overlapping SDEGs screened

Corresponding Author: Guo Xinyuan, MD; e-mail: 458967607@qq.com

via protein-protein interaction (PPI) network was selected for their functional similarity, and their diagnostic value was assessed. Our study provides insights into breast cancer's molecular mechanisms based on its pathophysiology.

Breast cancer can be divided into different types according to its molecular features, histopathological manifestations, and clinical results[11,12]. However, different types can not fully describe the clinical significance of breast cancer. Thus, exploring the difference in the effectiveness of radiotherapy for breast cancer and preventing overcrowding and undertreatment remains an urgent challenge. In addition, more and more evidence shows that early diagnosis and treatment will lead to a good prognosis, early detection and improvement of treatment will significantly reduce the mortality of breast cancer patients, and post-RT treatment can effectively improve the clinical effect[13]. Moreover, cancer survival related to RT-induced symptoms is essential, which may affect the quality of life (QOL)[14]. There is a lack of reliable and detailed clinical biomarkers for predicting breast cancer after radiotherapy, and the pathways and genes related to breast cancer radiotherapy are still unclear. Therefore, it is urgent to look for new diagnostic markers with high sensitivity and specificity to distinguish breast cancer from benign breast diseases and normal samples.

As an interdisciplinary discipline, Bioinformatics has made a breakthrough in medical research[15-19]. Many bioinformatics studies[20-23] have been used to predict the mechanism of drug resistance and detect molecular biomarkers in radiotherapy. However, tumors may gradually adapt to changes in physical and chemical characteristics in the body's microenvironment during radiotherapy and gain resistance to radiotherapy. The poor prognosis of breast cancer is related to radiotherapy resistance, and the poor pathology may aggravate radiotherapy resistance. More personalized RT therapy tailored to individual risk and tumor biology will help improve patients' prognoses. Currently, several radiation-related genes are reported, while no applicated biomarkers are in the clinic. There are few studies on the changes in breast cancer gene expression during radiotherapy[24,25]. Eschrich et al[24] reported 10 hub genes (*AR*, *cJun*, *STAT1*, *PKC*, *RelA*, *cABL*, *SUMO1*, *CDK1*, *HDAC1*, and *IRF1*) associated with intrinsic radiosensitivity, relating different pathways, such as cell cycle, DNA damage response, histone deacetylation, proliferation and apoptosis. However, the low statistical sample makes it difficult to assess the prognosis of breast cancer patients with RT treated in 2 independent BC patients. Another study[25] reported that radiation response was associated with p53. None of these hub genes was validated in clinical practice. Therefore, identifying differentially expressed genes may overcome the molecular mechanism of radiotherapy resistance to breast cancer.

In this study, we aimed to combine the machine learning algorithm (SVM) with various bioinformatics methods to determine the potential diagnostic markers of breast cancer based on the gene expression dataset from the GEO database. Univariate cox analysis was used to identify prognosis-related vital genes. These genes expression of various tissues were analyzed in the Gene Expression Profiling Interactive Analysis (GEPIA) and cancer genome atlas (TCGA). Furthermore, CIBERSORT was initially used to estimate the difference in immune infiltration between normal and breast cancer tissue in 22 immune cells. This study investigated early breast cancer's possible molecular immune mechanism and the association between infiltrating immune cells and diagnostic markers.

Materials and Methods

Data Acquisition

TCGA transcriptome data, and clinical data for breast cancer, including 120 normal and 1,097 cancer samples, were downloaded from the UCSC Xena database (http://xena.ucsc.edu/)[26]. Reliable breast cancer radiotherapy profiles of GSE59733 and gene expression profiles of GSE71053 were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/)[27]. The GSE59733 dataset is based on the Affymetrix Human Almac Xcel Array GPL18990 provided by Horton JK. It contains 9 tumor samples before radiotherapy and 10 tumor samples after radiotherapy. The GSE71053 dataset was uploaded by Pedersen IS, including 6 normal samples and 12 tumor samples[28].

Differentially Expressed Genes Between Pre- and Post-Radiotherapy

The ComBat function of SVA package was applied to remove batch effects on the GSE59733 dataset[29]. Then, the PCA analysis was carried out on the batch-corrected data by princomp function to check whether breast cancer samples could be clearly distinguished before and after radiotherapy. The limma package was performed to screen Radiation differentially expressed genes (RDEGs) between pre-radiation tumor samples and post-ra-

diation tumor samples from the GSE59733 dataset[30]. The volcano diagram was generated by ggplot2 package[31]. The 3D PCA plot was generated by the scatterplot3d R package. The *p*-value of the RDEGs was calculated by using the *t*-test method. $|\log_2 \text{foldchange}| > 1$ and $p\text{-value} < 0.05$ were the cut-off criteria for RDEGs. Before and after the pre- and post-radiation, Hub genes with significant differential expression were selected. The pROC package was applied to analyze *FDCSP* between pre- and post-radiotherapy[32]. Boxplot was generated by the ggplot2 package.

Functional and Pathway Enrichment Analysis

The "median" expression of *FDCSP* was used as the group cut-off to assign the higher and lower expression of FDCSP as the high and low groups of the GSE59733 dataset. Limma package in R was performed to screen Single gene differentially expressed genes[29]. In order to identify SDEGs, we used a *p*-value of 0.05 and a $|\log_2 FC|$ of greater than one as cut-off c-value criteria. The clusterProfiler package[33] was used to analyze Gene Ontology (GO) functions[34] and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways[35] of SDEGs. GO annotation analysis is a ubiquitous method for large-scale functional enrichment of genes, including the cellular component (CC), molecular function (MF), and biological process (BP). KEGG is a widely used database for analyzing information about genomes, biological pathways, diseases and drugs. Adjusted $p\text{-value} < 0.05$ was considered statistically significant. According to the phenotypic correlation, the trend of gene distribution in a set of predefined genes classified in the gene table can be better described.

Gene set enrichment analysis (GSEA) was carried out between the high-risk and low-risk groups according to the gene expression data by cluster profile package. We selected "c2.cp.kegg.v7.0.symbols.gmt" as the internal reference gene set, false discovery rate (FDR) < 0.25 and p value < 0.05 were considered as significant enrichment. The GSVA package*36* was used to perform gene set variation analysis with the high and low gene expression groups, the Hallmark, KEGG, GO-BP, GO-CC, and GO-MF were selected as reference gene sets.

Construction of PPI Network Analysis Underlying FDSCP

The interaction relationship of the SDEGs was predicted from the STRING (https://www.string-db.org/) database[37], and the protein-protein interaction (PPI) network was constructed to visualize using Cytoscape software[38]. The MCODE[39] and cytoHubba[40] plug-ins identified Hub genes in the PPI network. Meanwhile, the mRNA-miRNA interaction relationship of *FDCSP* was constructed based on miRWalk database[41] (http://mirwalk.umm.uni-heidelberg.de/) and visualized by Cytoscape software.

Estimation of the Association Between Immune Infiltration of Hub Genes and Its Related Diagnostic Markers

By default, CIBERSORT deconvolutions the transcriptome expression matrix based on the linear support vector regression (SVG) to evaluate the composition and frequency of immune cells in mixed cells[42]. The immune cell infiltration was obtained by using genes in the signature matrix with the CIBERSORT method. The pheatmap package in R (https: //CRAN. R-project. Org/package = pheatmap) plot the distribution of 22 immune cells in each sample. The ggplot2 package plotted the distribution of immune cells in samples. The correlation of 22 kinds of immune cell infiltration was visualized by correlation heat map generated with the corrplot package. The violin graph plotted by the ggplot2 software package was used to visualize the differences in 22 kinds of immune cell infiltration in *FDCSP* expression groups. The correlation between FDCSP, hub gene, and immune cell infiltration was analyzed and then visualized with the ggplot2 package.

The Expression Analysis of Hub Genes

HPA (https://www.proteinatlas.org/) is a biological research platform based on the TCGA database, which uses various combinatorial techniques to characterize the expression of proteins in tissues and cells, including the localization and distribution of thousands of proteins in various cancer tissues. GEPIA (http://gepia.cancer-pku.cn/) is an online analysis website based on transcriptome sequencing data from 9,736 tumor samples and 8,587 normal samples from TCGA and GTEx databases, which can be used to evaluate the correlation between the two genes in cancer. We analyzed the expression and distribution of *FDCSP* in the HPA database (https://www.proteinatlas.org/)[43] and GEPIA database (http://gepia.cancer-pku.cn/)[44].

Construction of Predicted Hub Genes-Based Prognostic Model

Univariate cox analysis was conducted by using survival package that screened the prognosis-re-

lated SDEGs. Two machine learning algorithms, SVM and Lasso Rogers regression were further used to screen the diagnostic markers. The Lasso Rogers regression model analysis was carried out with the glmnet package[45], the SVM model analysis was carried out with the randomForest package[46]. The final result is the intersection of the diagnostic markers obtained from the two algorithms. The TCGA-BRCA dataset was divided into a 1:1 training and validation sets. In the training set, multivariate cox analysis was performed to construct the prediction model. Kaplan-Meier survival and ROC curve were performed on the validation and all sets to verify the predicted prognostic model.

The Risk Score and Clinical Nomogram Construction

To further obtain clinical prognostic factors, univariate and multivariate cox analysis was performed on a training dataset with clinicopathological features (risk score, age, tumor stage, N stage, T stage and M stage). These clinical prognostic factors constructed two nomograms. ROC and calibration curves are used to evaluate the prediction efficiency of the nomogram chart.

Statistical Analysis

All analyses in this study were performed with R software (Version 4.0.2), and p-value < 0.05 was considered statistically significant. For the comparison of continuous variables between two groups, the statistical significance of normally distributed variables was estimated by the independent student t-test, and the Mann Whitney U test analyzed differences between non normally distributed variables. The Chi-square or Fisher's exact test was used to compare and analyze the statistical significance between two groups of categorical variables. The survival package of R was used to perform survival analysis, Kaplan Meier survival curves were used to show survival differences, and the log-rank test was used to assess the significance of differences in survival times between the two groups. Univariate and multivariate Cox analyses were based on the survival R package, and lasso analysis was based on the glmnet R package. All statistical p-values were two-sided; a p-value < 0.05 was considered statistically significant.

Results

The Expression of Genes Differs Between Pre- and Post-Radiotherapy

Data normalization is necessary to obtain valid results in gene expression analysis before downstream analyses. After normalization, the median expression of all samples in the GSE59733 dataset at the same level after the batch effect is removed (Figure 1). PCA results revealed that samples could be clearly distinguished between pre- and post-radiotherapy (Figure 2C). A total of 341 RDEGs were discovered, including 183 up-regulated RDEGs and 158 down-regulated RDEGs.

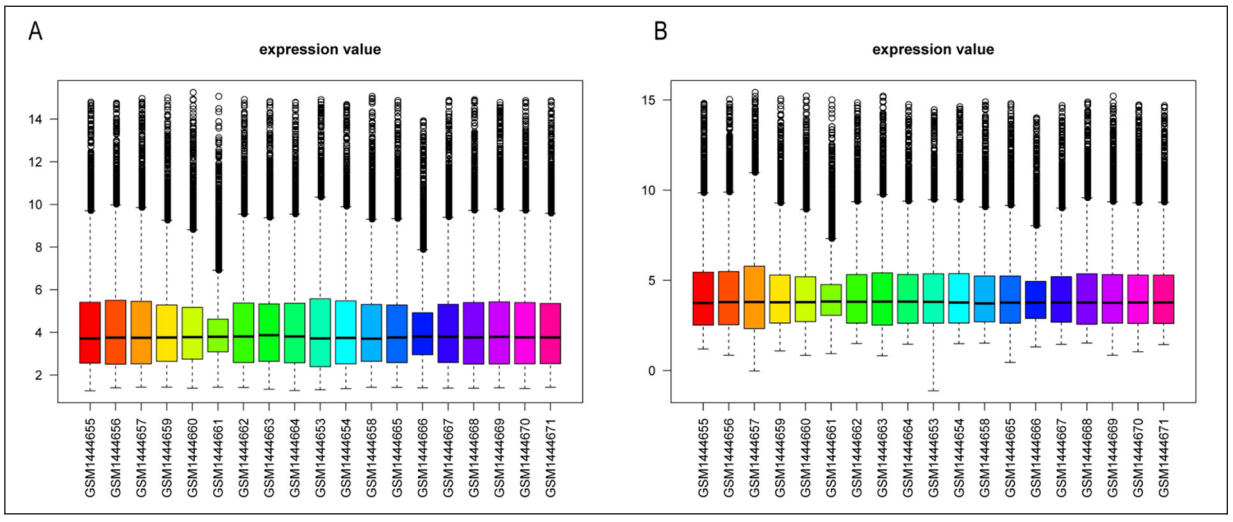

Figure 1. The boxplots for gene expression of each sample in the GSE59733 dataset. **A**, Boxplot of gene expression of each sample prior to normalization. **B**, Boxplot of gene expression of each sample after normalization.

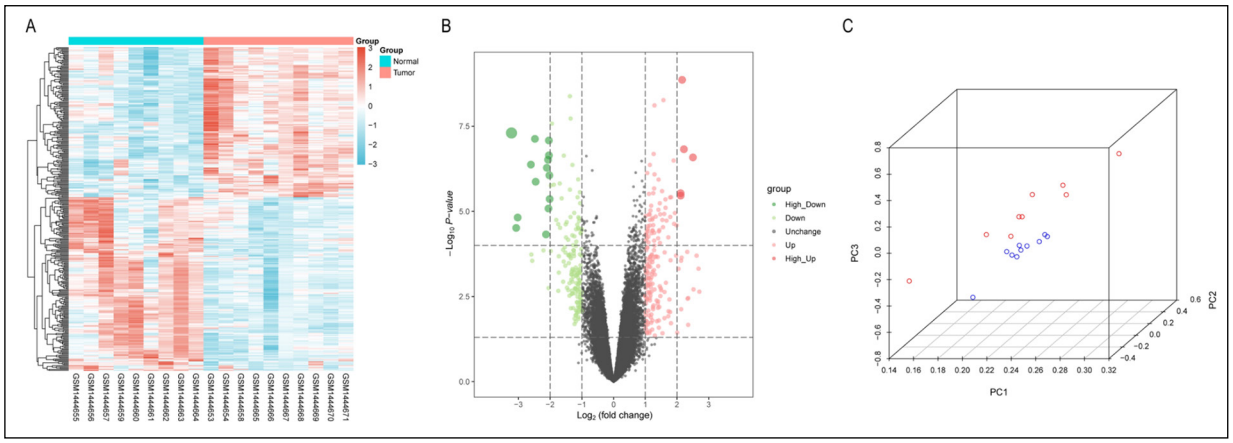

Figure 2. Differential expression profiles and PCA analysis on GSE59733 dataset of pre- and post-radiotherapy. **A**, Heat-map overview of the differentially expressed genes. Red and blue color represent high and low expression level of genes, respectively. **B**, Volcano plot of DEGs in GSE59733 dataset. Red and green color indicate relatively higher and lower gene expression levels, respectively. **C**, Principal component analysis on gene expression profiles of GSE59733 dataset. Red and blue color represent pre and post-radiotherapy, respectively.

Differential Expression Analysis

The *follicular dendritic cell secreted peptide* (*FDSCP*) gene that has a significantly different expression before and after radiotherapy and is reported in breast cancer was selected as the research object. These genes profiles were conducted to analyze expression differences before and after radiotherapy. Figure 3 showed that the expression of the *FDCSP* gene changed significantly before and after radiotherapy ($p < 0.05$), and ROC analysis showed that the AUC value of the *FDCSP* gene was 0.922, which had a relatively high prediction accuracy.

Analysis of Pathway and Functional Enrichment

One hundred eighty-nine candidate genes (64 up-regulated genes and 125 down-regulated genes) were obtained from the hub gene differential expression analysis of FDCSP genes. The overlapping SDEGs were subjected to GO analysis and KEGG pathway enrichment analysis. In GO enrichment analysis, the terms such as "response to steroid hormone", "response to reactive oxygen species" and "response to glucocorticoid" were significantly enriched. In KEGG pathway analysis, several significantly pathway involved, such as "IL-7 signal pathway", "rheumatoid arthritis", and "MAPK signal pathway" were found (**Supplementary Figure 1**). A summary of the total number of GO terms is given in Table I. Detailed information on KEGG enriched pathway is shown in Table II. Then, GSEA and GSVA (hallmark gene set) were used to investigate the relationship between enriched pathways and tumor characteristics based on the ffgiold change of *FDSCP* on GSE59733 dataset. Gene set variation analysis (GSVA) was used to explore further the differences between subtypes based on biological process (BP), cellular component (CC), and molecular function (MF) in terms of GO and KEGG pathways, the top four pathways as shown in Figure 4. Furthermore, the high-expression group showed a significantly enriched hub gene with blue bars. In contrast, the green bars indicate the pathways significantly enriched in the low-expression group of hub genes. A summary of GSEA analysis result are shown in Table III and **Supplementary Figure 2**.

Developing the PPI Network Analysis Underlying the FDSCP

The PPI network constructed with SDEGs and STRING database is shown in Figure 5 A. The results showed that the PPI network contains 129 nodes and 516 interactions. The color of each node represents the node degree for each gene. Color red and green indicate a high and low degree of node, respectively. Five hub genes were screened with MCODE (Figure 5 B) and cytoHubba (Figure 5 C) plug-in in Cytoscape software. Additionally, *FDSCP* gene -miRNA interaction network is shown in Figure 5 D, color red indicates *FDSCP* gene, and the color blue represents miRNA interacting with *FDSCP* gene (752 nodes in total).

Table I. GO enrichment analysis results.

GO-BP			
ID	**Description**	**Count in gene set**	***p*-value**
GO:0048545	response to steroid hormone	18	7.79E-09
GO:0000302	response to reactive oxygen species	14	1.66E-08
GO:0051384	response to glucocorticoid	11	6.35E-08
GO:0022612	gland morphogenesis	10	9.93E-08
GO:1901654	response to ketone	12	1.32E-07
GO:0031099	regeneration	12	1.74E-07
GO:0031960	response to corticosteroid	11	1.84E-07
GO:0097193	intrinsic apoptotic signaling pathway	14	2.54E-07
GO:0032355	response to estradiol	10	2.81E-07
GO:0048660	regulation of smooth muscle cell proliferation	11	2.82E-07
GO-CC			
ID	**Description**	**Count in gene set**	***p*-value**
GO:0062023	collagen-containing extracellular matrix	19	1.46E-09
GO:0031983	vesicle lumen	13	5.80E-06
GO:0060205	cytoplasmic vesicle lumen	12	2.87E-05
GO:0034774	secretory granule lumen	11	8.53E-05
GO:0030055	cell-substrate junction	12	0.000188
GO:0005925	focal adhesion	11	0.000626
GO:0005924	cell-substrate adherens junction	11	0.000665
GO:1904724	tertiary granule lumen	4	0.001157
GO:0005604	basement membrane	5	0.00122
GO:0045111	Intermediate filament cytoskeleton	8	0.001273
GO-MF			
ID	**Description**	**Count in gene set**	***p*-value**
GO:0001085	RNA polymerase II transcription factor binding	8	5.31E-05
GO:0005201	extracellular matrix structural constituent	8	7.58E-05
GO:0008083	growth factor activity	8	7.58E-05
GO:0008201	heparin binding	8	9.76E-05
GO:0005539	glycosaminoglycan binding	9	0.000149
GO:0019838	growth factor binding	7	0.000168
GO:0004714	transmembrane receptor protein tyrosine kinase activity	5	0.000179
GO:0043394	proteoglycan binding	4	0.000237
GO:0048018	receptor ligand activity	13	0.000238
GO:0042379	chemokine receptor binding	5	0.000241

Evaluation of Hub Genes and Associated Diagnostic Markers and Immune Infiltration

The composition and abundance of 22 kinds of immune cells with CIBERSORT method are visualized in a heatmap and histogram plot. There were statistically significant differences in the infiltration rate of immune cells between the group with high expression of the *FDCSP* gene and the group with low expression (Figure 6 A-D). Memory-activated $CD4^+$ T cells show the strongest negative correlation with Gamma-delta (γδ) T cells and macrophages M0, whereas it was positively

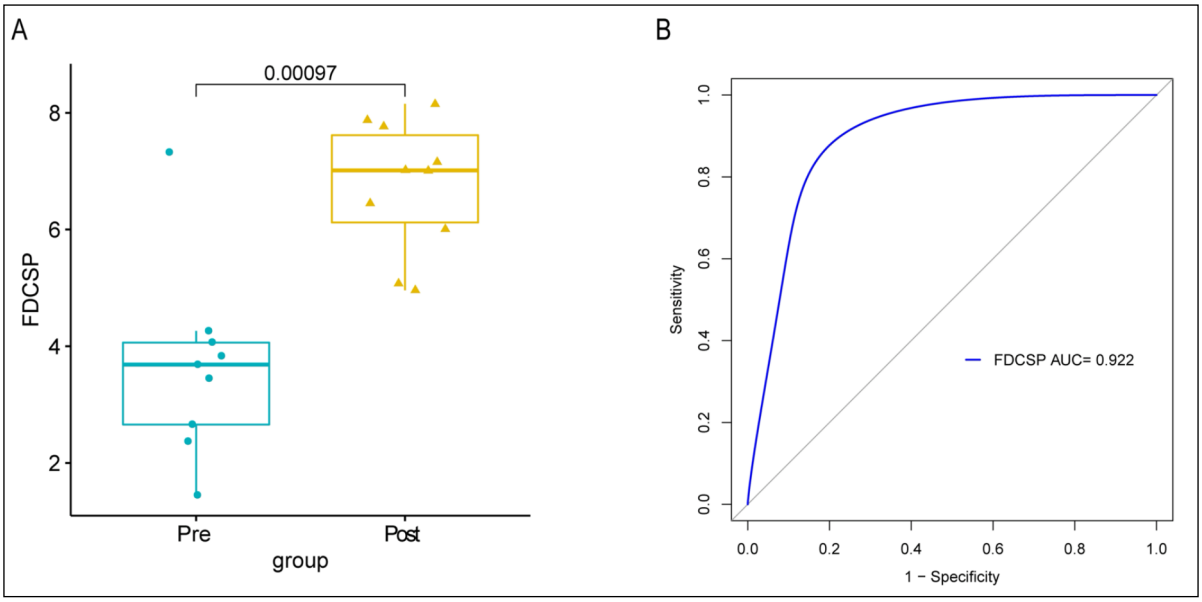

Figure 3. *FDCSP* gene expression profiles before and after radiotherapy and ROC curve. **A**, Changes in *FDSCP* gene expression before and after radiotherapy. **B**, The ROC curve analysis of *FDSCP* gene.

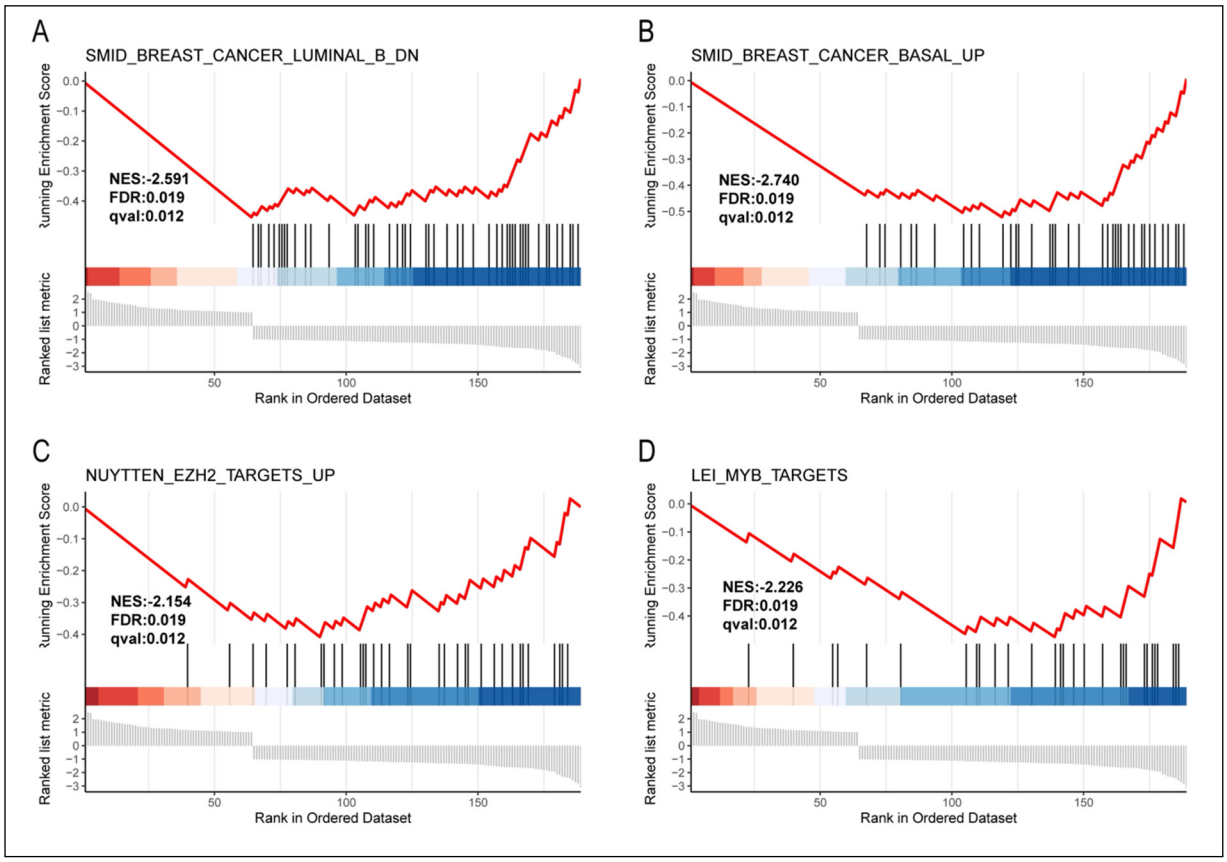

Figure 4. Results of gene set enrichment analysis (GSEA). **A**, Gene set enrichment analysis indicates SMID_ BREAST_ CANCER_ LUMINAL_ B_ DN signaling pathways are enriched in breast cancer. **B**, Gene set enrichment analysis indicates SMID_ BREAST_ CANCER_ BASAL_ The up pathway is enriched in breast cancer. **C**, Gene set enrichment analysis indicates that Nuytten_ EZH2_ TARGETS_ The up pathway is enriched in breast cancer. **D**, Gene set enrichment analysis indicated that Lei_ MYB_ The targets pathway is enriched in breast cancer.

Table II. KEGG enrichment analysis

ID	Description	Count in gene set	*p*-value
hsa04657	IL-17 signaling pathway	8	1.38E-05
hsa05323	Rheumatoid arthritis	7	0.000108
hsa04010	MAPK signaling pathway	12	0.000168
hsa05417	Lipid and atherosclerosis	10	0.000214
hsa04915	Estrogen signaling pathway	8	0.000216
hsa05166	Human T-cell leukemia virus 1 infection	10	0.000248
hsa04668	TNF signaling pathway	7	0.000344
hsa04151	PI3K-Akt signaling pathway	12	0.000904
hsa05150	Staphylococcus aureus infection	6	0.000924
hsa01522	Endocrine resistance	6	0.001029
hsa04657	IL-17 signaling pathway	8	1.38E-05
hsa05323	Rheumatoid arthritis	7	0.000108

Table III. Results of gene set enrichment analysis (GSEA)

ID	NES	p.adjust	qvalues
SMID_BREAST_CANCER_LUMINAL_B_DN	-2.591	0.0193	0.0124
SMID_BREAST_CANCER_BASAL_UP	-2.740	0.0193	0.0124
NUYTTEN_EZH2_TARGETS_UP	-2.154	0.0193	0.0124
LEI_MYB_TARGETS	-2.226	0.0193	0.0124

Table IV. Clinicopathological characteristics of high and low FDCSP gene groups in TCGA breast cancer patients.

Characteristic	Level	High-FDCSP Group (n=456)	Low-FDCSP Group (n=456)	Overall (N=913)
Age				
	Mean (SD)	56.3 (12.4)	59.0 (13.3)	57.7 (12.9)
	Median [Min, Max]	56.0 [26.0, 90.0]	59.0 [26.0, 90.0]	58.0 [26.0, 90.0]
T stage				
	T1	130 (28.5%)	106 (23.2%)	236 (25.8%)
	T2	261 (57.2%)	281 (61.5%)	542 (59.4%)
	T3	57 (12.5%)	45 (9.8%)	102 (11.2%)
	T4	8 (1.8%)	25 (5.5%)	33 (3.6%)
N stage				
	N0	230 (50.4%)	224 (49.0%)	454 (49.7%)
	N1	150 (32.9%)	151 (33.0%)	301 (33.0%)
	N2	49 (10.7%)	54 (11.8%)	103 (11.3%)
	N3	27 (5.9%)	28 (6.1%)	55 (6.0%)
M stage				
	M0	451 (98.9%)	445 (97.4%)	896 (98.1%)
	M1	5 (1.1%)	12 (2.6%)	17 (1.9%)
Gender				
	female	454 (99.6%)	448 (98.0%)	902 (98.8%)
	male	2 (0.4%)	9 (2.0%)	11 (1.2%)
Stage				
	stage i	86 (18.9%)	74 (16.2%)	160 (17.5%)
	stage ii	265 (58.1%)	270 (59.1%)	535 (58.6%)
	stage iii	100 (21.9%)	101 (22.1%)	201 (22.0%)
	stage iv	5 (1.1%)	12 (2.6%)	17 (1.9%)

correlated with resting dendritic cells and macrophages M1.

The Violin plot exhibited the immune infiltration subpopulations correlated profiles. Native B cell, T cell CD8$^+$, memory-activated T cell CD4$^+$, T follicular helper cell, Gamma-delta (γδ) T cell, resting NK cell, Macrophage M0, Macrophage M1, activated myeloid dendritic cell, and Neutrophil shared higher proportion in high *FDCSP* expression group. On the contrary, the proportion of memory B cells, plasma B cells, naive CD4$^+$ T cells, activated Mast cells, activated NK cells, and regulatory T cells (Tregs) are lower. Furthermore, we analyzed the correlation between hub genes (*EGFR*, *ESR1*, *FOS*, *IL6*, and *JUN*) based on the GSE59733 dataset and immune cell infiltration in BRCA tissues. Significant differences among various subtype immune cells were found between the group with high expression of the *FDCSP* gene and the group with low expression (**Supplementary Figure 3**).

FDCSP Immunohistochemistry and Tissue Expression Analysis

The expression of *FDCSP* gene was analyzed based on HPA and GEPIA databases. Gene expression of various organs and immunohistochemistry of normal tissues and breast cancer tissues can be seen in HPA database (Figure 7 A), and gene expression of various tissues can be seen in GEPIA database (Figure 7 B). Additionally, the ROC results of the *FDCSP* gene in the GSE71053 dataset revealed that the *FDCSP* gene effectively distinguished breast cancer tumor tissues from normal tissues (AUC = 0.704) (Figure 7 C).

Construction of a Hub Genes-Based Prognostic Model

The baseline data sheets of the patients are shown in Table IV. Eleven associated prognosis SDEGs were identified by performing a univariate cox analysis with all SDEGs (Figure 8 A).

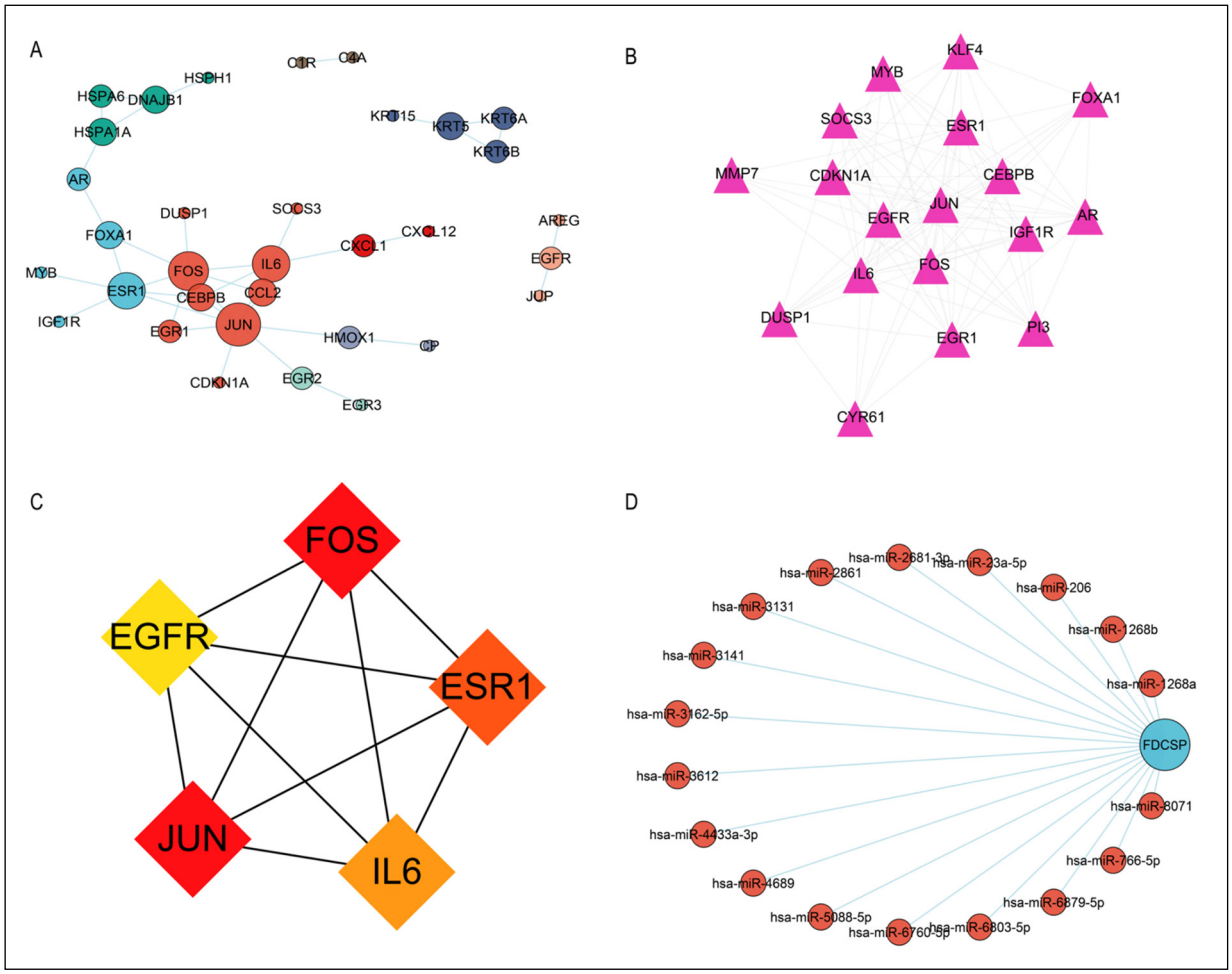

Figure 5. PPI networks of SDEGs and FDSCP gene-miRNA . **A**, PPI networks of the SDEGs; (**B**) Hub genes screened by MCODE plug-in; (**C**) Identification of the hub genes by cytoHubba plug-in. **D**, *FDCSP* gene-miRNA interaction network.

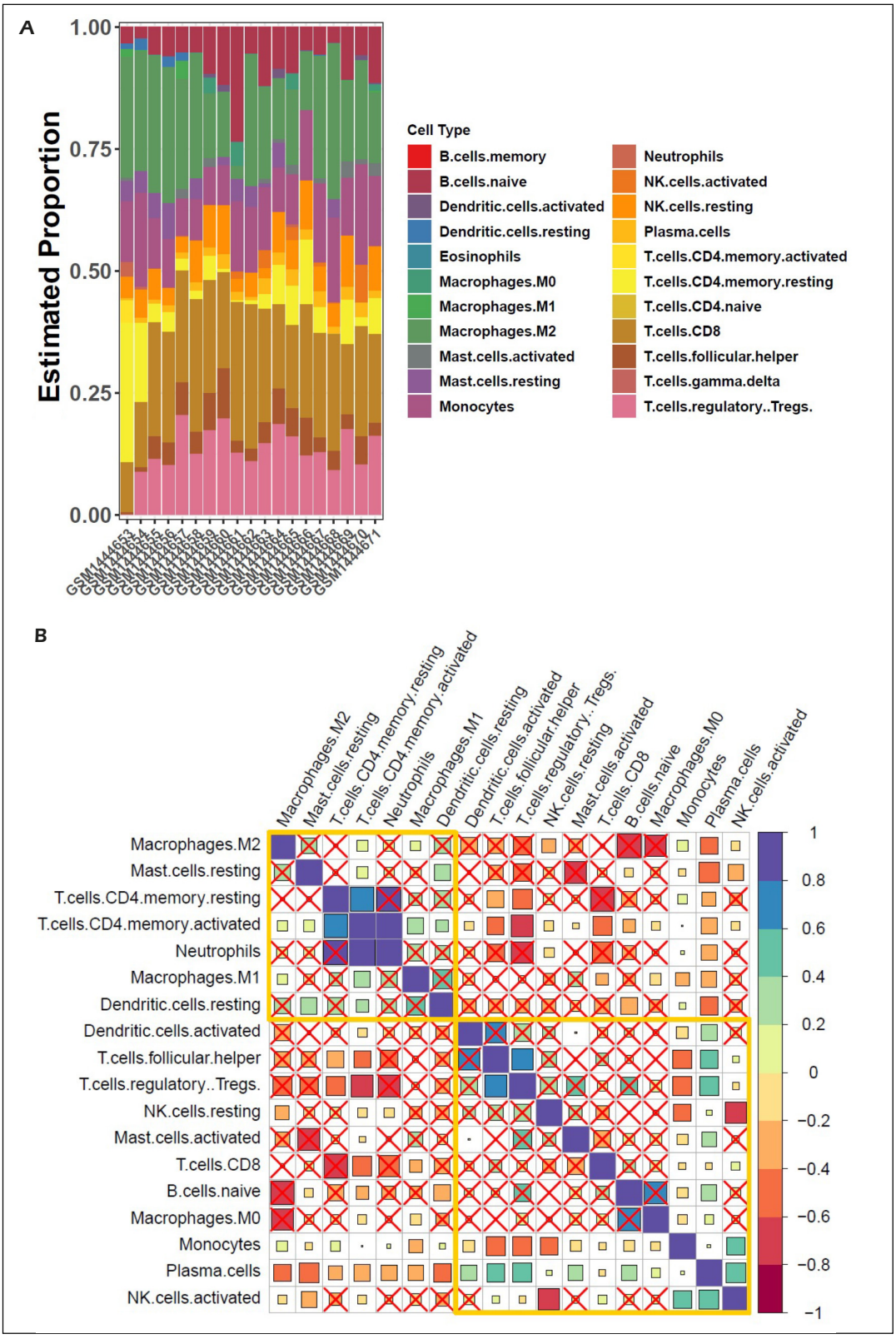

Figure 6. Immune cell infiltration evaluation and visualization of GSE59733 dataset. **A**, Proportion of the 22 immune cell types in BRCA tissues. **B**, Correlation matrix between the 22 immune cell types. Green represents positive correlation, and brown represents negative correlation.

Figure continued

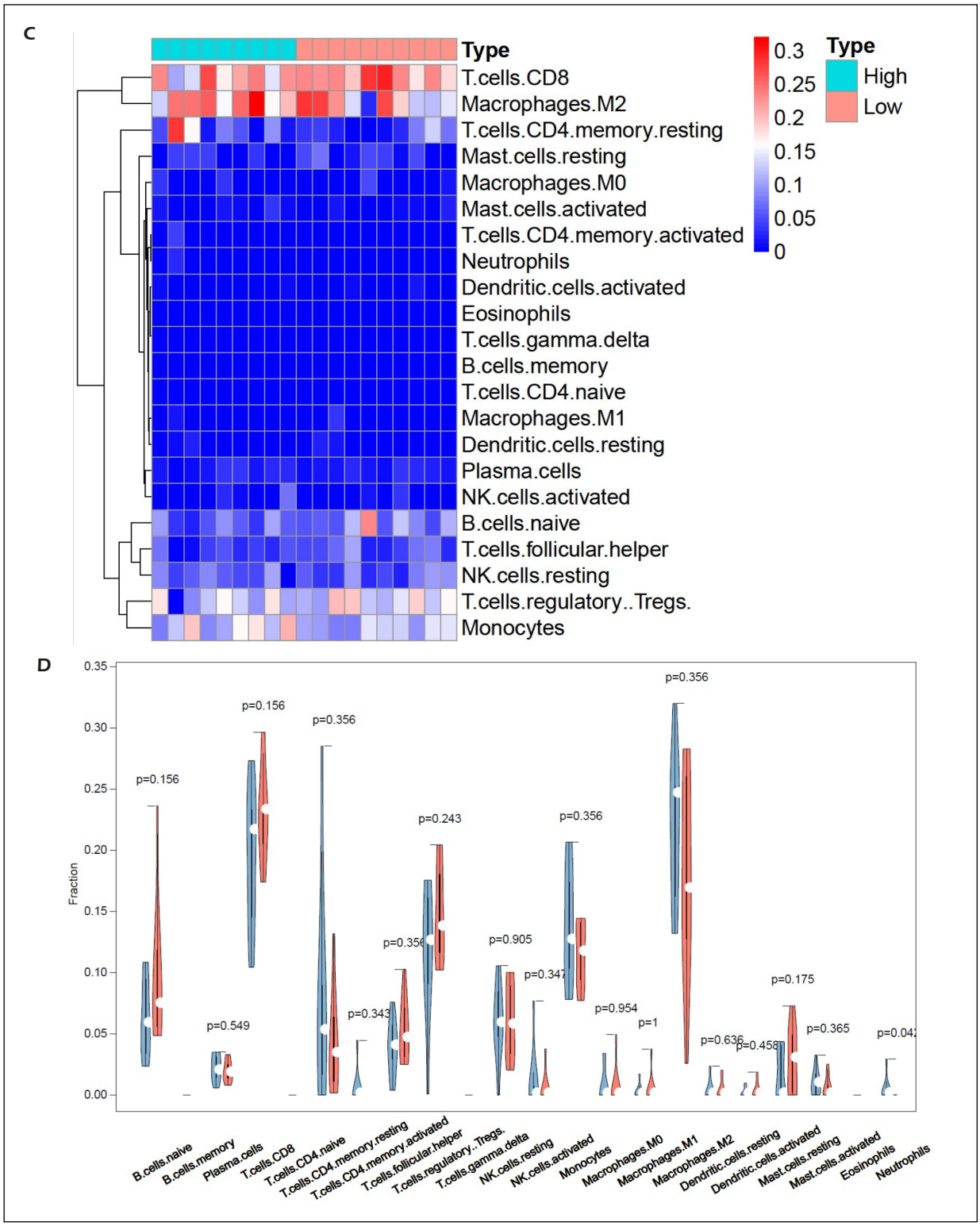

Figure 6 ***(Continued)***. **C**, Heatmatablep of immune cell infiltration between high and low *FDCSP* gene expression groups. **D**, Violin map of immune cell proportion in high *FDCSP* expression group (blue) and low *FDCSP* expression group (red).

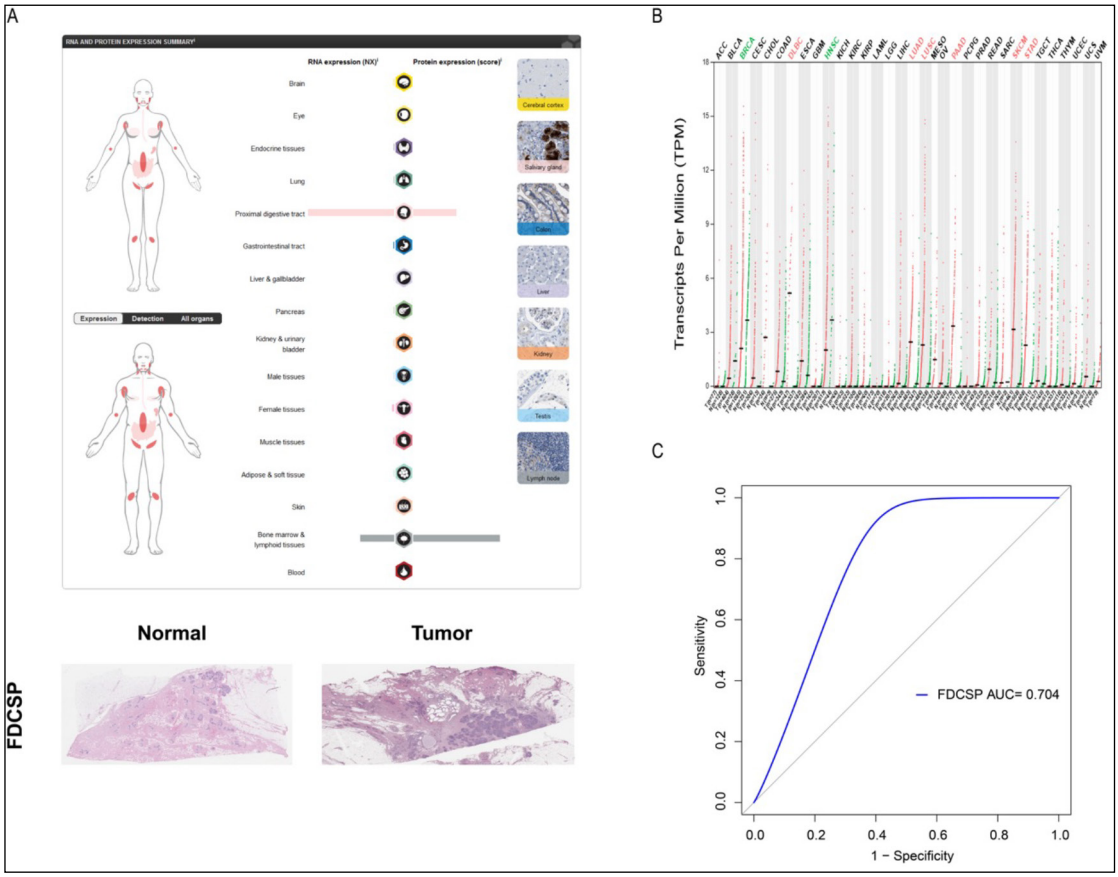

Figure 7. The expression analysis of *FDCSP* gene in HPA and GEPIA database. **A**, *FDCSP* gene expression in HPA and its immunohistochemical map in breast cancer normal tissues and tumor tissues. **B**, *FDSCP* gene expression profile in GEPIA. **C**, ROC curve of breast cancer from normal tissue and tumor tissue predicted by *FDCSP* gene in GSE71053 dataset.

The lasso regression algorithm (Figure 8 B-C) and SVM algorithm (Figure 8 D-E) were used to identify potential diagnostic markers. 11 genes were screened from SDEGs as diagnostic markers using the lasso regression algorithm. The SVM algorithm was used to identify 11 genes as diagnostic markers. These candidate diagnostic markers identified by the two methods completely overlapped, and 11 associated diagnostic genes were obtained. To further test the effectiveness of the diagnosis, 7 genes were identified (Figure 8 G), and multivariate cox analysis on the TCGA-BRCA training dataset was to construct a predicted prognostic model. We further validated the predicted prognostic model on the validation dataset and all datasets (**Supplementary Figure 4**). Table V shows the results of univariate and multivariate cox analysis.

Nomogram

Univariate (Figure 9 A) and multivariate (Figure 9 B) cox analysis was used to analysis on clinical features (risk score, age, tumor stage, N stage, T stage and M stage). The results revealed that risk score, age and M stage were associated with the prognosis of patients in TCGA. Therefore, based on the TCGA-BRCA training dataset, we developed a nomogram to predict overall patient survival at 3 and 5 years. The calibration curve and ROC curve results show that patients' estimated survival probabilities at 3 and 5 years are consistent with the actual probabilities (Figure 9 C-E).

Discussion

The radiotherapy process of breast cancer patients is heterogeneous[47,48], which significantly impacts clinical efficacy and quality of life. Therefore, there is an urgent need to find radiation-associated biomarkers for breast cancer's early diagnosis and prognosis. In addition, radiotherapy-related biomarkers enable BC patients to

receive more personalized targeted immunotherapy. Similar improvements in personalized radiotherapy have not been applied clinically. Hence, researchers are increasingly looking for new diagnostic biomarkers and studying the components of breast cancer immune cell infiltration that may positively impact the clinical outcome of breast cancer patients. In this study, we performed an integrated analysis of TCGA transcriptome data and clinical data of breast cancer to identify effective diagnostic biomarkers for breast cancer.

In the present study, two BC datasets were downloaded from GEO, and a total of 341 DEGs (183 up- and 158 down-regulated genes) were identified using cross-validation. The GO and KEGG enrichment analyses showed that the DEGs were associated with various cancer-related functions and pathways, such as steroid hormone, reactive oxygen species and glucocorticoids. Steroid receptors (SRs) are subjected to many post-translational modifications by the reversible addition of various molecular parts, including phosphorylation, acetylation, methylation, glycosylation and ubiquitination. Active oxygen is generated in mitochondria, peroxisome and the endoplasmic reticulum. ROS are generated after radiotherapy in breast cancer. Glucocorticoids play an essential role in embryonic development and tissue homeostasis and possess important anti-inflammatory and immunosuppressive properties. These functional abnormalities will lead to abnormal physiological functional pathways, including those related to genetics/genomics, oxidative stress, neuroplasticity and inflammation.

Furthermore, the top 20 hub genes associated with breast cancer, identified in the PPI network based on STRING database, showed high functional similarity and diagnostic values for breast cancer. PPI network analysis using the STRING database showed several effective central genes. Biological process, cell component and molecular function also evaluated the enrichment pathway. Survival analysis further supported the robustness of the above results.

With the development of next-generation sequencing technologies and the prognosis effects of radiotherapy, transcriptome research on radiation-related breast cancer have been carried out. For instance, in a study based on the original breast cancer dataset GSE59733 from GEO, 82 DEGs were identified, and *FOS*, *CCL2*, and *CXCL12* were strongly proposed as hub genes[49]. Another bioinformatics analysis of breast cancer gene expression profiles in GSE1561 datasets and

Table V. Results of Univariate Cox analysis and Multivariate Cox analysis.

Univariate Cox analysis					
Gene	**HR**	**z**	***p*-value**	**lower**	**upper**
CCDC3	1.26	4.66	0.00	1.14	1.39
METTL7A	1.32	3.82	0.00	1.14	1.52
DDIT4	0.85	-3.11	0.00	0.77	0.94
BCL2A1	0.86	-3.01	0.00	0.78	0.95
NFIL3	1.18	2.72	0.01	1.05	1.32
KCTD12	1.17	2.58	0.01	1.04	1.32
C11orf96	1.14	2.55	0.01	1.03	1.26
EGR1	1.09	2.17	0.03	1.01	1.17
HSPA6	0.89	-2.10	0.04	0.80	0.99
SERPINA11	0.95	-2.09	0.04	0.90	1.00
THBS1	1.12	2.01	0.04	1.00	1.25
Multivariate Cox analysis					
Id	**coef**	**HR**	**HR.95L**	**HR.95H**	***p*-value**
EGR1	-0.09	0.91	0.81	1.02	0.11
THBS1	0.10	1.11	0.97	1.27	0.13
BCL2A1	-0.24	0.79	0.69	0.90	0.00
NFIL3	0.30	1.35	1.12	1.62	0.00
DDIT4	-0.19	0.83	0.73	0.94	0.00
SERPINA11	-0.09	0.91	0.86	0.97	0.00
C11orf96	0.18	1.20	1.04	1.38	0.01

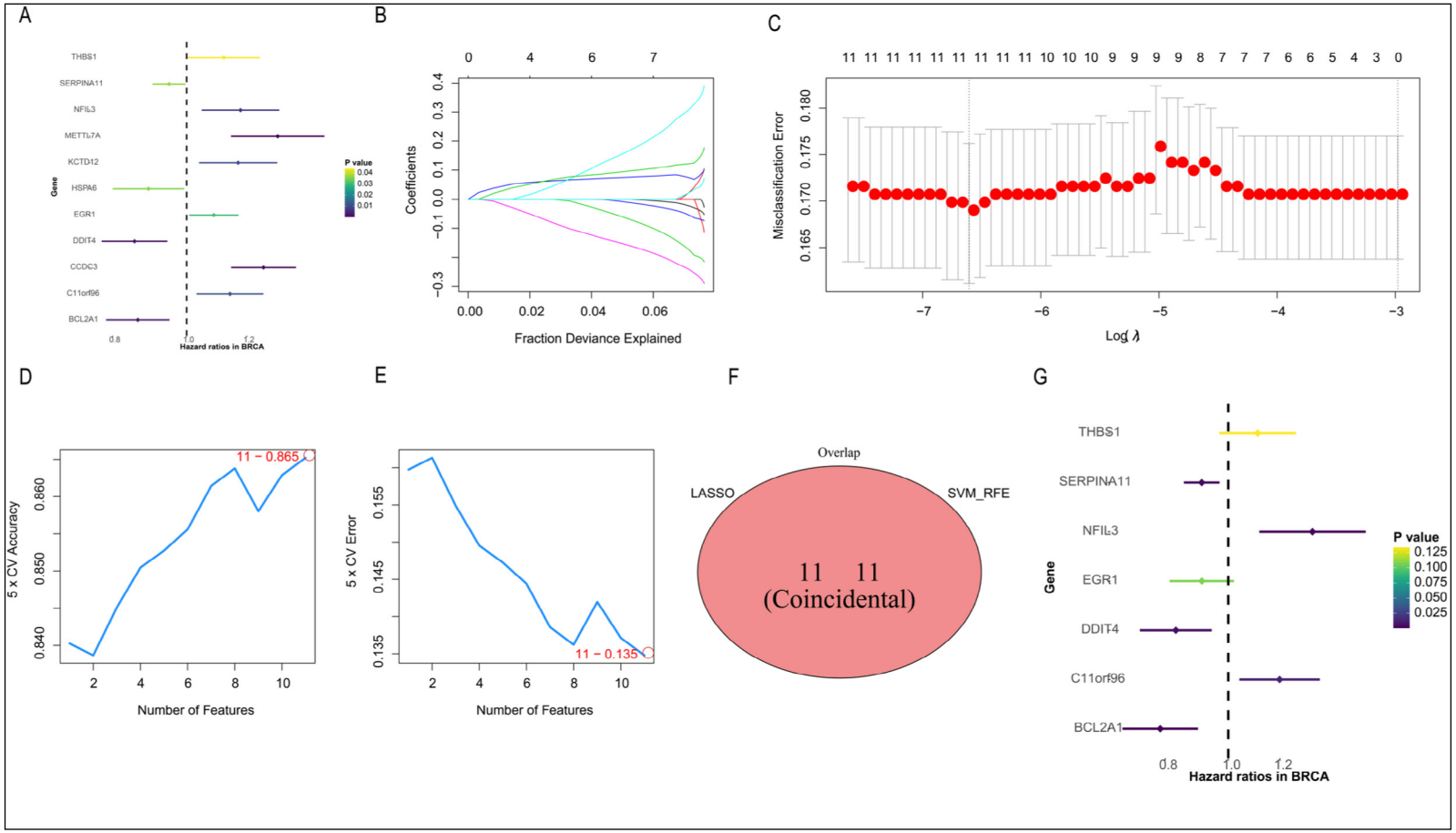

Figure 8. LASSO, SVM, Univariate and multivariate cox regression algorithms for identifying independent prognostic factors. **A**, Univariate cos regression of the 11 screened prognosis-associated genes. **B-C**, LASSO coefficient profiles of the 11 hub genes, in which the lowest cross-validation error rate is the best predictors of the model. **D**, SVM performance accuracy. **E**, SVM performance error. **F**, Venn plot showing hub genes identified in common by the lasso and SVM algorithms. **G**, Forest map of multivariate cox regression for 11 diagnostic genes, 7 genes were obtained for establishing prediction prognostic model.

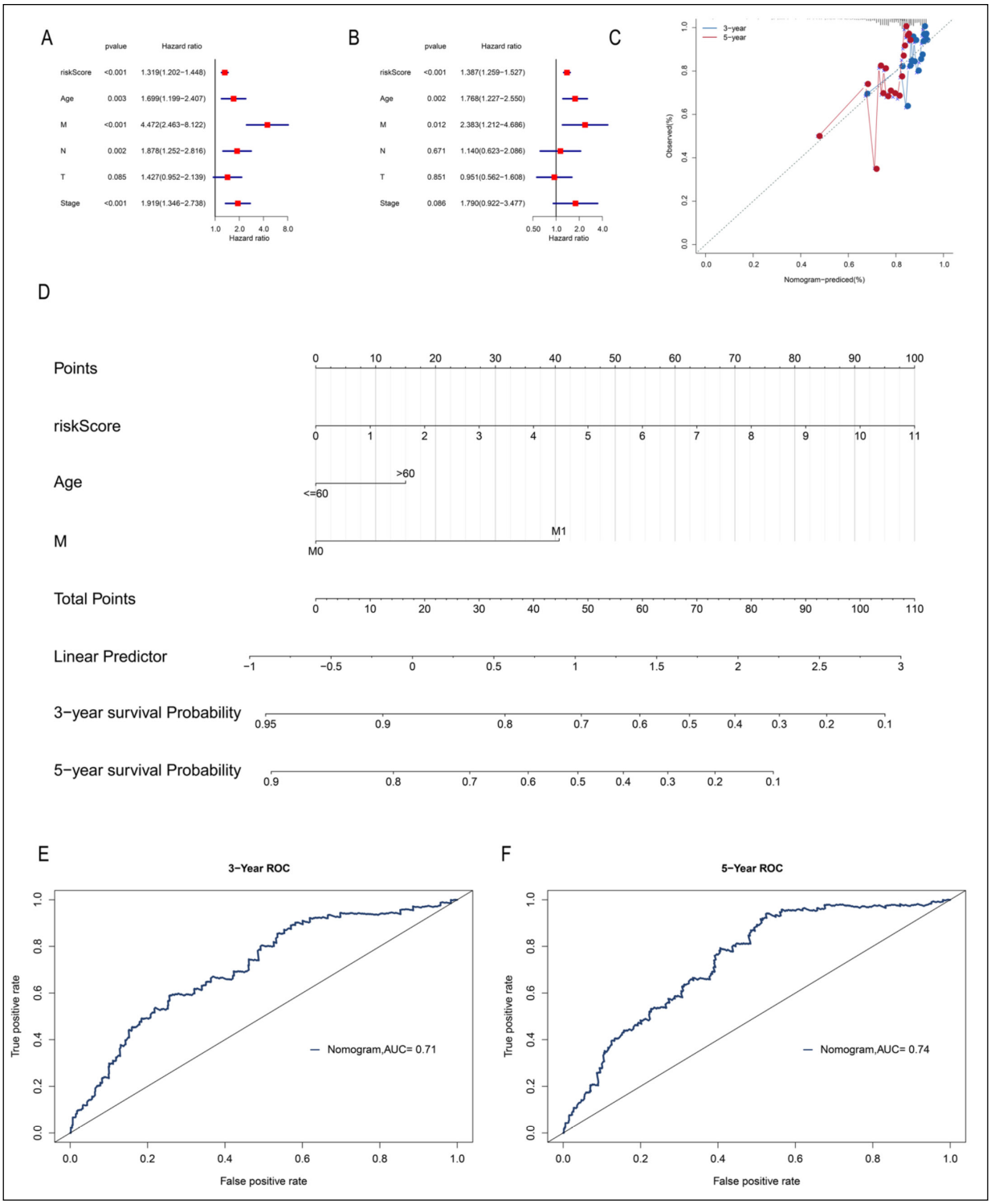

Figure 9. Nomogram construction and verification. **A**, Univariate Cox regression analysis of clinical characteristics for overall survival in the breast cancer. **B**, Multivariate Cox regression analysis of clinical characteristics for overall survival in the breast cancer. **C**, Calibration plot of the nomogram. **D**, The nomogram consists of risk score, age, and stage. **E-F**, Nomogram validation of 3-year and 5-year ROC curve.

five hub genes (*CCNB2*, *FBXO5*, *KIF4A*, *MCM10*, and *TPX2*) were identified as potential biomarkers associated with breast cancer prognosis[50]. Compared with these published similar studies, the present study has some advantages: First, we employed the limma package to screen radiation differentially expressed genes (*RDEGs*) between pre-radiation tumor samples and post-radiation tumor samples from the GSE59733 dataset. To our knowledge, radiation-related transcriptomic expression has not been analyzed for breast cancer. Secondly, Two machine learning algorithms, SVM and Lasso Rogers regression were further used to screen the diagnostic markers. Two machine learning algorithms could build a more precise classification model with feature variables. These features ensured the credibility of our results.

We screened five hub genes (*EGFR*, *ESR1*, *FOS*, *IL6*, and *JUN*) from the PPI network in the present study. According to existing studies, the hub genes play key roles in various cancer-related biological processes. FOS is one of five hub genes. FOS plays a critical role in the development and progression of breast cancer by mediating the transcription of AP-1 target genes[51]. Nonetheless, few studies reveal the response of FOS in breast cancer to radiation[49]. In our analysis, FOS was one of the hub proteins with the highest interaction in the PPI network. Thus, whether altering its gene expression may affect the response of breast cancer to radiation is still an urgent problem to be studied in-depth. Another hub gene ESR1 was identified as potential prognostic biomarkers for breast cancer and is expressed by several types of cancer[52]. It is reported that mutations of the ESR 1 gene is a prognostic factor related to low survival rate. Previous studies have reported that mutation of ESR 1 could affect hormone resistance and reduce the therapeutic response[53,54]. The most important diagnostic markers with the highest degree of connection among the central genes were selected. Kaplan-Meier plotter with p-value < 0.05 was used to verify the recurrence-free survival rate of hub genes. Furthermore, we also analyzed the correlation between hub genes (*EGFR*, *ESR1*, *FOS*, *IL6*, and *JUN*) and the infiltration of immune cells in BRCA tissue using the GSE59733 dataset. Significant differences exist among the immune cells of each subtype of the *FDCSP* gene expression group.

In addition, the Lasso-Rogers regression and support vector machines (SVG) were further used to screen and identify potential diagnostic markers. 11 genes were obtained from SDEGs using the lasso regression algorithm as diagnostic markers. Finally, according to the identified gene biomarkers, the available evidence is obtained from the reported experimental literature on breast cancer. In addition, our analysis shows that the nomogram has a good calibration efficiency. The ROC results of the *FDCSP* gene in the GSE71053 dataset revealed that the *FDCSP* gene could effectively distinguish breast cancer tissue from normal tissue. The candidate genes *THBS1*, *SERPINA11*, *NFIL3*, *METTL7A*, *KCTD12*, *HSPA6*, *EGR1*, *DDIT4*, *CCDC3*, *C11orf96*, and *BCL2A1* in breast cancer cells after radiotherapy were verified. These genes can be used as potential prognostic biomarkers and therapeutic targets for breast cancer, but more evidence is needed to support the basis of computational analysis biology.

Limitations

The limitations of our research should also be acknowledged. First of all, in analyzing the DEGs, it is difficult to consider some important factors, for example, different ages, races, regions, tumor stages and patient classification, because of the complexity of the data set in our study. Secondly, according to the results, the five hub genes were up-regulated in breast cancer, but the mechanism of up-regulation was unclear. Third, the sample size of breast cancer radiotherapy is quite low and more evidence is needed to understand the biological basis. Finally, this study mainly focuses on analyzing the expression levels and OS of five hub genes. Whether these hub genes could be used as biomarkers or improve breast cancer's diagnostic accuracy and specificity needs further study. We will collect more relevant samples in future clinical trials and design prospective trials to improve the statistical power and achieve more meaningful results.

Conclusions

The present study has identified several key genes (*EGFR*, *ESR1*, *FOS*, *IL6*, and *JUN*) that might be considered novel and potential breast cancer biomarkers. These results may provide a novel understanding of the prognosis effects of radiotherapy on breast cancer, identifying several potential biomarkers for its diagnosis and treatment.

Conflict of Interest
The Authors declare that they have no conflict of interests.

Data Availability
The datasets generated during and/or analyzed during the current study are available in the repository GEO database (https://www.ncbi.nlm.nih.gov/geo/).

Author Contributions
X.-Y. Guo and W.-C. Dan designed the study. X.-Y. Guo and W.-C. Dan ran the search strategy. W.-C. Dan and S.-L. Wang collected data, and X.-Y. Guo and LJL re-checked the data. W.-C. Dan and X.-Y. Guo performed the analysis. S.-L. Wang and G.-Z. Zhang re-checked the data. W.-C. Dan and X.-Y. Guo wrote the manuscript. All the listed authors have reviewed and revised the manuscript.

Funding
This study was supported by the Non-profit Central Research Institute Fund of the Chinese Academy of Medical Sciences to M.D (2021-RC310-013), Beijing Hope Run Special Fund of Cancer Foundation of China to M.D (LC2021R02) and the CAMS Innovation Fund for Medical Sciences to M.D (2021-I2M-1-067).

Acknowledgments
We thank Lao Xinyuan for the suggestions on this article.

Ethics Approval
Not applicable.

References

1) Ataollahi MR, Sharifi J, Paknahad MR and Paknahad A. Breast cancer and associated factors: a review. J Med Life 2015; 8: 6-11.
2) Momenimovahed Z and Salehiniya H. Epidemiological characteristics of and risk factors for breast cancer in the world. Breast Cancer (Dove Med Press) 2019; 11: 151-164.
3) Feng Y, Spezia M, Huang S, Yuan C, Zeng Z, Zhang L, Ji X, Liu W, Huang B, Luo W, Liu B, Lei Y, Du S, Vuppalapati A, Luu HH, Haydon RC, He TC and Ren G. Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. Genes Dis 2018; 5: 77-106.
4) Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A and Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021; 71: 209-249.
5) Lei S, Zheng R, Zhang S, Chen R, Wang S, Sun K, Zeng H, Wei W and He J. Breast cancer incidence and mortality in women in China: temporal trends and projections to 2030. Cancer Biol Med 2021; 18: 900-909.
6) Goodarzi E, Beiranvand R, Naemi H, Rahimi S and Khazaei Z. Geographical distribution incidence and mortality of breast cancer and its relationship with the Human Development Index (HDI): an ecology study in 2018. WCRJ 2020; 7: e1468.
7) Waks AG and Winer EP. Breast Cancer Treatment: A Review. Jama 2019; 321: 288-300.
8) Vaidya JS, Bulsara M, Baum M, Alvarado M, Bernstein M, Massarut S, Saunders C, Sperk E, Wenz F and Tobias JS. Intraoperative radiotherapy for breast cancer: powerful evidence to change practice. Nat Rev Clin Oncol 2021; 18: 187-188.
9) Cheng YJ, Nie XY, Ji CC, Lin XX, Liu LJ, Chen XM, Yao H and Wu SH. Long-Term Cardiovascular Risk After Radiotherapy in Women With Breast Cancer. J Am Heart Assoc 2017; 6: 290-333.
10) He MY, Rancoule C, Rehailia-Blanchard A, Espenel S, Trone JC, Bernichon E, Guillaume E, Vallard A and Magné N. Radiotherapy in triple-negative breast cancer: Current situation and upcoming strategies. Crit Rev Oncol Hematol 2018; 131: 96-101.
11) Tsang JYS and Tse GM. Molecular Classification of Breast Cancer. Adv Anat Pathol 2020; 27: 27-35.
12) Russnes HG, Lingjærde OC, Børresen-Dale AL and Caldas C. Breast Cancer Molecular Stratification: From Intrinsic Subtypes to Integrative Clusters. Am J Pathol 2017; 187: 2152-2162.
13) Lee K, Kruper L, Dieli-Conwright CM and Mortimer JE. The Impact of Obesity on Breast Cancer Diagnosis and Treatment. Curr Oncol Rep 2019; 21: 41.
14) Poleszczuk J, Luddy K, Chen L, Lee JK, Harrison LB, Czerniecki BJ, Soliman H and Enderling H. Neoadjuvant radiotherapy of early-stage breast cancer and long-term disease-free survival. Breast Cancer Res 2017; 19: 75.
15) Tang D, Zhao X, Zhang L, Wang Z and Wang C. Identification of hub genes to regulate breast cancer metastasis to brain by bioinformatics analyses. J Cell Biochem 2019; 120: 9522-9531.
16) Yang K, Gao J and Luo M. Identification of key pathways and hub genes in basal-like breast cancer using bioinformatics analysis. Onco Targets Ther 2019; 12: 1319-1331.
17) Dong P, Yu B, Pan L, Tian X and Liu F. Identification of Key Genes and Pathways in Triple-Negative Breast Cancer by Integrated Bioinformatics Analysis. Biomed Res Int 2018; 2018: 2760918.
18) Zeng X, Shi G, He Q and Zhu P. Screening and predicted value of potential biomarkers for breast cancer using bioinformatics analysis. Sci Rep 2021; 11: 20799.

19) Parvizpour S, Razmara J and Omidi Y. Breast cancer vaccination comes to age: impacts of bioinformatics. Bioimpacts 2018; 8: 223-235.

20) Allmer J. Computational and bioinformatics methods for microRNA gene prediction. Methods Mol Biol 2014; 1107: 157-175.

21) Peng Q, Zhu J, Shen P, Yao W, Lei Y, Zou L, Xu Y, Shen Y and Zhu Y. Screening candidate microRNA-mRNA regulatory pairs for predicting the response to chemoradiotherapy in rectal cancer by a bioinformatics approach. Sci Rep 2017; 7: 11312.

22) Liu D, Li B, Shi X, Zhang J, Chen AM, Xu J, Wang W, Huang K, Gao J, Zheng Z, Liu D, Wang H, Shi W, Chen L and Xu J. Cross-platform genomic identification and clinical validation of breast cancer diagnostic biomarkers. Aging (Albany NY) 2021; 13: 4258-4273.

23) Tabl AA, Alkhateeb A, ElMaraghy W, Rueda L and Ngom A. A Machine Learning Approach for Identifying Gene Biomarkers Guiding the Treatment of Breast Cancer. Front Genet 2019; 10: 256.

24) Eschrich SA, Pramana J, Zhang H, Zhao H, Boulware D, Lee JH, Bloom G, Rocha-Lima C, Kelley S, Calvin DP, Yeatman TJ, Begg AC and Torres-Roca JF. A gene expression model of intrinsic tumor radiosensitivity: prediction of response and prognosis after chemoradiation. Int J Radiat Oncol Biol Phys 2009; 75: 489-496.

25) Amundson SA, Do KT, Vinikoor LC, Lee RA, Koch-Paiz CA, Ahn J, Reimers M, Chen Y, Scudiero DA, Weinstein JN, Trent JM, Bittner ML, Meltzer PS and Fornace AJ, Jr. Integrating global gene expression and radiation survival parameters across the 60 cell lines of the National Cancer Institute Anticancer Drug Screen. Cancer Res 2008; 68: 415-424.

26) Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, Banerjee A, Luo Y, Rogers D, Brooks AN, Zhu J and Haussler D. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol 2020; 38: 675-678.

27) Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S and Soboleva A. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res 2013; 41: D991-995.

28) Pedersen IS, Thomassen M, Tan Q, Kruse T, Thorlacius-Ussing O, Garne JP and Krarup HB. Differential effect of surgical manipulation on gene expression in normal breast tissue and breast tumor tissue. Mol Med 2018; 24: 57.

29) Leek JT, Johnson WE, Parker HS, Jaffe AE and Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 2012; 28: 882-883.

30) Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015; 43: e47.

31) Wilkinson L. ggplot2: Elegant Graphics for Data Analysis by WICKHAM, H. Biometrics 2011; 67: 671-679.

32) Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC and Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011; 12: 77.

33) Yu G, Wang LG, Han Y and He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics 2012; 16: 284-287.

34) Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM and Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000; 25: 25-29.

35) Kanehisa M, Furumichi M, Tanabe M, Sato Y and Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 2017; 45: D353-d361.

36) Hänzelmann S, Castelo R and Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 2013; 14: 7.

37) von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P and Snel B. STRING: a database of predicted functional associations between proteins. Nucleic Acids Res 2003; 31: 258-261.

38) Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003; 13: 2498-2504.

39) Bandettini WP, Kellman P, Mancini C, Booker OJ, Vasu S, Leung SW, Wilson JR, Shanbhag SM, Chen MY and Arai AE. MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study. J Cardiovasc Magn Reson 2012; 14: 83.

40) Chin CH, Chen SH, Wu HH, Ho CW, Ko MT and Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol 2014; 8 Suppl 4: S11.

41) Dweep H, Gretz N and Sticht C. miRWalk database for miRNA-target interactions. Methods Mol Biol 2014; 1182: 289-305.

42) Chen B, Khodadoust MS, Liu CL, Newman AM and Alizadeh AA. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. Methods Mol Biol 2018; 1711: 243-259.

43) Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J and Pontén

F. Proteomics. Tissue-based map of the human proteome. Science 2015; 347: 1260419.

44) Tang Z, Li C, Kang B, Gao G, Li C and Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res 2017; 45: W98-W102.

45) Engebretsen S and Bohlin J. Statistical predictions with glmnet. Clin Epigenetics 2019; 11: 123.

46) Alderden J, Pepper GA, Wilson A, Whitney JD, Richardson S, Butcher R, Jo Y and Cummins MR. Predicting Pressure Injury in Critical Care Patients: A Machine-Learning Model. Am J Crit Care 2018; 27: 461-468.

47) Dimitrov G. AM, Popova Y., Vasileva K., Milusheva Y., Troianova P. Molecular and genetic subtyping of breast cancer: the era of precision oncology. WCRJ 2022; 9: e2367.

48) Caputo R, Cianniello D, Giordano A, Piezzo M, Riemma M, Trovò M, Berretta M and De Laurentiis M. Gene Expression Assay in the Management of Early Breast Cancer. Curr Med Chem 2020; 27: 2826-2839.

49) Zhu C, Ge C, He J, Zhang X, Feng G and Fan S. Identification of Key Genes and Pathways Associated With Irradiation in Breast Cancer Tissue and Breast Cancer Cell Lines. Dose Response 2020; 18: 1559325820931252.

50) Saunus JM, De Luca XM, Northwood K, Raghavendra A, Hasson A, McCart Reed AE, Lim M, Lal S, Vargas AC, Kutasovic JR, Dalley AJ, Miranda M, Kalaw E, Kalita-de Croft P, Gresshoff I, Al-Ejeh F, Gee JMW, Ormandy C, Khanna KK, Beesley J, Chenevix-Trench G, Green AR, Rakha EA, Ellis IO, Nicolau DV, Jr., Simpson PT and Lakhani SR. Epigenome erosion and SOX10 drive neural crest phenotypic mimicry in triple-negative breast cancer. NPJ Breast Cancer 2022; 8: 57.

51) Fu J, Cheng L, Wang Y, Yuan P, Xu X, Ding L, Zhang H, Jiang K, Song H, Chen Z and Ye Q. The RNA-binding protein RBPMS1 represses AP-1 signaling and regulates breast cancer cell proliferation and migration. Biochim Biophys Acta 2015; 1853: 1-13.

52) Qin X and Song Y. Bioinformatics Analysis Identifies the Estrogen Receptor 1 (ESR1) Gene and hsa-miR-26a-5p as Potential Prognostic Biomarkers in Patients with Intrahepatic Cholangiocarcinoma. Med Sci Monit 2020; 26: e921815.

53) Dustin D, Gu G and Fuqua SAW. ESR1 mutations in breast cancer. Cancer 2019; 125: 3714-3728.

54) Hishida M, Nomoto S, Inokawa Y, Hayashi M, Kanda M, Okamura Y, Nishikawa Y, Tanaka C, Kobayashi D, Yamada S, Nakayama G, Fujii T, Sugimoto H, Koike M, Fujiwara M, Takeda S and Kodera Y. Estrogen receptor 1 gene as a tumor suppressor gene in hepatocellular carcinoma detected by triple-combination array analysis. Int J Oncol 2013; 43: 88-94.