

Prognostic Analysis of Patients with Breast Cancer Based on Tumor Mutational Burden, DNA Damage Repair Genes, and Immune Infiltration

Tianshu Yang

Capital Medical University

Jiaxiang Liu

National Cancer Center, Chinese Academy of Medical Sciences and Peking Union Medical College

Xin Wang

National Cancer Center, Chinese Academy of Medical Sciences and Peking Union Medical College

Tianyu Ma

Capital Medical University

Xin Yin

Capital Medical University

Yunkai Yang

National Cancer Center, National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Baowen Yuan

National Cancer Center, National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Xu Teng

Capital Medical University

Hefen Yu

Capital Medical University

Shuang Wang

Shandong Provincial ENT Hospital affiliated to Shandong University

Wei Huang (✉ wei Huang@ccmu.edu.cn)

Capital Medical University

Research Article

Keywords: Breast cancer, TMB, Cox regression analysis, Cox-LASSO regression analysis, Risk prognostic model, Tumor-infiltrating immune cells.

Posted Date: October 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-970339/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Breast cancer has a high tumor-specific death rate and poor prognosis. In this study, we aimed to provide a basis for the prognostic risk in patients with breast cancer using significant gene sets selected by analyzing tumor mutational burden (TMB) and DNA damage repair (DDR).

Methods

Breast cancer genomic and transcriptomic data were obtained from The Cancer Genome Atlas, and the breast tumor samples were grouped into high- and low-TMB groups based on the calculated TMB. The differentially expressed DDR genes between the two groups were used to construct the breast cancer prognosis model; they were also further analyzed, and 10 key prognostic genes were screened. Using least absolute shrinkage and selection operator (LASSO)-Cox regression analysis, seven genes were selected to construct a linear risk assessment model related to survival, and the patients were divided into high- and low-risk groups.

Results

The ability of the model to predict the 5-year and 10-year survival was assessed by time-dependent receiver operating characteristic (ROC) curves. The GSE26085 dataset was used to verify the prognostic risk model. Finally, immune cell abundance and infiltration between the high- and low-risk groups were compared.

Conclusions

We established a risk prognostic model based on the MDC1, PARP3, PSMB1, PSMB9, PSMD2, PSMD7, and PSMD14 genes, providing a basis for further exploration of a population-based prediction of prognosis and immunotherapy response in patients with breast cancer.

Introduction

Breast cancer is one of the most common malignant tumors occurring in women. In recent years, the incidence of breast cancer has increased annually and has gradually become a veritable health risk for women. In 2018, new cases of breast cancer accounted for 24.2% of the total number of cancer cases in women, representing approximately 2.08 million people, and deaths due to breast cancer accounted for 15% of the total number of cancer-related deaths, representing approximately 630,000 people in the world. Because of the high tumor-specific death rate of breast cancer, the prognosis of patients with breast cancer should be investigated. Currently, traditional clinical and pathological staging cannot

effusively mirror tumor heterogeneity and predict the prognosis. With the development of cDNA microarray and high-throughput sequencing technology, prediction models based on the combination of gene sequencing data and clinical data have gained considerable attention for the diagnosis and treatment of breast cancer.

Tumor mutational burden (TMB) is defined as the total number of somatic gene coding errors, base substitutions, and gene insertion or deletion errors detected per million bases. Mutations are recognized by T cells, subsequently activating the immune response. Thus, TMB can reflect the curative effect of therapy to a certain extent. TMB, particularly microsatellite instability, is related to programmed death (PD) ligand 1 (PD-L1) levels in cancer cells. Moreover, the accumulation of mutations in the tumor genome can result in the translation of abnormal proteins through mutated mRNAs, leading to the production of new antigens and the presentation of new human leukocyte antigen (HLA) complexes in tumor cells (1). Therefore, the TMB index has been permitted by the Food and Drug Administration (FDA) for use in predicting the efficacy of pan-tumor immunotherapy. TMB is also closely associated with the expression of PD-1 and PD-L1, affecting the response to immunotherapy. Thus, the higher the TMB, the more likely the tumor cells would be discerned by the immune system and the higher the probability of immunotherapy efficacy (2, 3). Therefore, optimizing the algorithm for distinguishing high- and low-risk groups using TMB and improving the differentiation of patients with respect to adaptive immunotherapy response must be the focus of future research.

The DNA damage repair (DDR) pathway is vital in ensuring the accurate transmission of genetic material. Changes in the DDR pathway play a predictive and prognostic role in anticancer therapy (4). The occurrence and development of tumors are associated to abnormalities in the DDR pathway, such as the mutations in the homologous recombination repair (HRR) gene *BRCA1/2* in breast cancer (5). Approximately 10% of breast cancer cases occur in patients with germline pathogenic variants of *BRCA1*, *BRCA2*, and other DDR genes, which are correlated with an increased risk of breast, and other cancers. Studies have shown that DDR-related gene mutations are significantly correlated with TMB and that these genes can improve immunotherapy efficacy, which is associated with favorable outcomes (6). Hence, the establishment of a DDR gene panel based on TMB is crucial in optimizing the benefits and improving the therapeutic effects of immunologic agents, such as immune checkpoint inhibitors (ICIs), in patients with breast cancer.

Previous studies have used RNA sequencing to directly screen all differential genes and obtain the gene set related to breast cancer prognosis or to analyze the TMB of each tumor and predict the biomarkers associated with immunotherapy and TMB. In this study, we used breast cancer genomic and transcriptomic data and calculated TMB and to construct a breast cancer prognostic model based on differentially expressed DDR genes between high- and low-TMB groups and to determine potential biomarkers related to breast cancer prognosis, ultimately providing a theoretical reference value for the prognostic risk of patients with breast cancer. By analyzing the breast cancer genomic and transcriptomic data from The Cancer Genome Atlas (TCGA), we preliminarily identified seven DDR genes associated with a high TMB, constructed a risk score model based on TMB and the DDR genes, and verified the model

using the GSE26085 dataset from the Gene Expression Omnibus. We expect that our findings will provide a comprehensive basis for further exploration of a population-based prediction of prognosis and immunotherapy response in patients with breast cancer.

Materials And Methods

Genome and transcriptome data

Breast cancer genomic and transcriptomic data were obtained from the University of California, Santa Cruz (UCSC) Xena database. Samples without survival status and survival time were filtered out, while samples with both genomic and transcriptomic data were retained. The GSE20685 dataset served as the validation cohort. The clinical information for each sample is presented in Supplementary Table 1. The workflow is illustrated in Figure 8.

Tumor mutational burden

TMB was calculated for each sample using the TMB formula (27). The samples were divided into high- and low-TMB groups according to the quartile TMB value (upper quartile vs. residual; lower quartile vs. residual; mean/medium as cutoff).

DDR gene sets and mutational landscape

DDR gene sets were collated as previously mentioned (7). Gene information is shown in Supplementary Table 2. The “maftools” R package was used to depict the DDR mutational landscape of the high- and low-TMB groups.

Analysis of differential DDR gene expression profile

Differentially expressed DDR genes in the high- and low-TMB groups were screened according to $|\log_2(\text{FC})| > 0.26$ (i.e., 1.2-fold differential expression) and $P\text{-value} < 0.05$ (28, 29). DDR genes were intersected to obtain differentially expressed DDR genes using the “edgeR” R package.

Enrichment analyses and signaling pathway analysis of DDR genes

KEGG analysis and GO functional enrichment analysis of the differentially expressed DDR genes was performed using the “clusterProfiler” R package. Signaling pathway analysis of the differentially expressed DDR genes was conducted using the GSEA software.

Construction and assessment of risk prognostic model

For the differentially expressed DDR genes, univariate Cox regression analysis was used to select genes associated with prognosis using $P\text{-value} < 0.05$ as a filtering condition. The LASSO-Cox regression model was also used to select genes related to prognosis, and the correlation coefficients of these genes were obtained to construct a risk prognostic model. The risk score for each patient was determined according

to the model, and the median of the risk scores served as the cutoff value for dividing the patients into high-and low-risk groups. The time-dependent ROC curves were used to evaluate the ability of the model to predict the 5-year and 10-year survival rates. The survival curves of the high-and low-risk groups were also analyzed. The GSE20685 dataset was used to validate the risk prognostic model.

Analysis of immune cell abundance

The “CIBERSORT” R package was used to assess the abundance of 22 subtypes of immune cells in the high- and low-risk groups.

Differential analysis of immune cell infiltration by gene copy number

The Tumor Immune Estimation Resource (TIMER) database (<https://cistrome.shinyapps.io/timer/>) was used to analyze the infiltration of different immune cell subtypes based on the gene copy number.

Cell culture

MCF-7, MDA-MB-231, MCF-10A and T-47D cell lines were obtained from the Chinese Academy of Medical Sciences. MCF-7, MDA-MB-231, MCF-10A and T-47D cell lines were cultured with Dulbecco’s modified Eagle medium (DMEM) or Roswell Park Memorial Institute (RPMI) 1640 medium, were supplemented with 10% fetal bovine serum (FBS), 100 units/mL penicillin, and 100 mg/mL streptomycin (Thermo Fisher Scientific, Inc., Waltham, MA, USA). The cells were cultured in a humidified incubator equilibrated with 5% CO₂ at 37°C.

Real-time quantitative polymerase chain reaction analysis

To quantify the mRNA of target genes by RT-qPCR, total RNA was extracted from MCF-10A, T-47D, MCF-7, and MDA-MB-231 cells using TRIzol reagent. Reverse transcription was performed using the RevertAid First Strand cDNA Synthesis Kit (Roche, Basel, Switzerland) according to the manufacturer’s instructions. β-Actin was used as the internal reference. The relative expression of target genes was calculated using the 2^{-ΔΔCt} method.

EdU incorporation assay

MDA-MB-231 cells with depletion MDC1 and control cells were seeded into 6-well plates at a density of 0.8 × 10⁵ cells/ml to adhere overnight. After that, DNA proliferation was detected using an EdU assay kit according to the manufacturer’s instructions (RiboBio, Guangzhou, China).

Cell invasion assay

Transwell chamber filters (BD) were coated with Matrigel diluted (1:10) in serum-free medium. Depletion MDC1 MDA-MB-231 cells and control cells were seeded into the upper chamber of the transwell chambers, and the chambers were transferred into 24-well plates containing 500 μl culture medium per well. After 20 h of incubation, cells in the upper chamber were fixed with 4% formaldehyde, washed with

PBS, and stained with crystal violet for half an hour. Images of invasive cells were captured using a light microscope.

Western blot

Total protein was assessed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis using 10% gels, followed by immunoblotting with the indicated antibodies: anti-Fibronectin, anti-Vimentin and anti- β -actin (SigmaAldrich), anti-E-cadherin, anti- α -catenin, anti- γ -catenin and anti-N-cadherin (BD Bioscience). The β -actin content was analyzed as the loading control.

Statistical analyses

All statistical analyses were performed using the RStudio software 3.6.3 (RStudio, Boston, MA, USA) and SPSS Statistics software (SPSS, Inc., Chicago, IL, USA). All data were visualized using GraphPad Prism 8 (GraphPad Software, Inc, San Diego, CA, USA). Kaplan-Meier curve analyses were performed using the “survminer” R package. For all statistical tests, two-tailed $P < 0.05$ denoted statistical significance, which is indicated by * ($P < 0.05$), ** ($P < 0.01$), *** ($P < 0.001$), or **** ($P < 0.0001$).

Results

TMB calculation and correlation with clinical parameters

Breast cancer genomic and transcriptomic data were obtained from a total of 960 tumor samples. The TMB value of each sample was calculated to evaluate the correlation between TMB and the clinical parameters. The survival rate of the high-TMB group was significantly lower than that of the low-TMB group ($P = 6.734e-04$) (Figure 1A). As shown in Figures 1B and 1C, TMB was significantly higher in older patients than in younger ones ($P = 0.0033$); however, TMB was not significantly different in terms of each clinical stage between older and younger patients ($P > 0.05$). TMB was significantly different between the T1 and T2 groups, the T3 and T4 groups ($P < 0.05$) (Figure 1D), the N0 and N1 groups, and the N1 and N2 groups ($P < 0.05$) (Figure 1E). However, TMB was not significantly different between the M0 and M1 groups ($P = 0.12$) (Figure 1F).

DDR gene set and mutational landscape

DDR gene sets were sorted out from 10 DDR-related signaling pathways, and a total of 193 DDR gene sets were obtained (7). Gene information is shown in Supplementary Table 3. The landscape map of the DDR gene mutations in the high- and low-TMB group was visualized using the “maftools” R package (Figure 2). The top five high-frequency mutant genes in the high-TMB group were *TP53* (51%), *PRKDC* (4%), *BRCA2* (4%), *BRCA1* (4%), and *ATM* (3%). The top five high-frequency mutant genes in the low-TMB group were *TP53* (18%), *ATM* (1%), *PRKDC* (1%), *BRCA2* (1%), and *CDKN1B* (1%). It is important to note that 51% of the samples in the high-TMB group had a *TP53* mutant, while only 18% of the samples in the low-TMB group had a *TP53* mutant (Figures 2A, B). Overall, the high-TMB group had a higher frequency of gene mutations than the low-TMB group. In the high- and low-TMB groups, the median number of

mutations was 51.5% and 19%, respectively. In addition, 90% of the mutations were point mutations (Figures 2C, E). Through the co-occurrence and exclusive analyses of these mutant genes, a total of 72 mutant gene pairs were obtained in the high-TMB group, and only one significant mutant gene pair was obtained in the low-TMB group (Figures 2D, F).

Screening of differentially expressed DDR genes in the high- and low-TMB groups.

A volcano plot of 6,424 differentially expressed genes was obtained between the two TMB groups (Figure 3A). A total of 67 differentially expressed genes were obtained by intersecting the DDR genes. A series of oncogenes, including *EXO1*, *CCNE1*, *CCNE2*, *POLR2F*, *TP53BP1*, and *PSMA8*, was shown to be activated or inactivated (Figure 3B). Information on the 67 differentially expressed genes is listed in Supplementary Table 2. To further understand the molecular function of the differentially expressed genes in breast cancer, we performed Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis and Gene Ontology (GO) enrichment analysis. KEGG analysis revealed that the differentially expressed DDR genes were mainly enriched in terms of the proteasome, mismatch repair, and homologous recombination (Figure 3C). The GO enrichment results revealed that the top five enriched signaling pathways of the 67 differentially expressed genes were the Nuclear factor kappa B (NF- κ B) inducing kinase (NIK)/NF- κ B signaling pathway, DNA repair pathway, Wnt signaling pathway, proteasome complex pathway, and damaged DNA binding pathway (Figure 3D). Furthermore, Gene Set Enrichment Analysis (GSEA) revealed that in the high-TMB group, the differentially expressed genes were mainly enriched in terms of the cell cycle, DNA replication, proteasome, and oocyte meiosis (Figure 3E).

Screening of prognostic factors by univariate Cox regression analysis

To screen the genes related to prognosis, we performed univariate Cox regression analysis using the 67 differentially expressed DDR genes. Ten prognosis-related genes were selected: *MDC1*, *PARP3*, *POLR2K*, *PSMB1*, *PSMB9*, *PSMD2*, *PSMD7*, *PSMD14*, *RFC3*, and *UBE2T*. Information on these 10 genes is listed in Supplementary Table 4. We divided the 960 breast cancer samples into high- and low-expression groups according to the median expression of the prognosis-related genes in the samples. We also performed Kaplan-Meier survival analysis using the 10 DDR genes in the high- and low-expression groups. The survival rates of the two groups were analyzed using the log-rank test. There was a significant difference in terms of the overall survival rate between the high- and low-expression groups. The high expression of *PARP3* and the low expression of *POLR2K*, *PSMB1*, *PSMD2*, and *PSMD14* significantly prolonged the survival time of patients, improving outcomes and reducing recurrence rates ($P < 0.05$) (Figures 4A). In addition, we explored the expression and related pathways of the 10 genes using the Gene Set Cancer Analysis (GSCALite) database. As shown in Figure 4B, the expression levels of the genes were partially enhanced in breast cancer and lung adenocarcinoma. *MDC1*, *POLR2K*, *PSMB1*, *PSMB9*, *PSMD2*, *PSMD7*, *PSMD14*, *RFC3*, and *UBE2T* were highly activated in the apoptotic and cell cycle pathways and reserved in the Ras/mitogen-activated protein kinase (MAPK) pathway (Figure 4C). To validate the expression of *MDC1*, *PARP3*, *POLR2K*, *PSMB1*, *PSMB9*, *PSMD2*, *PSMD7*, *PSMD14*, *RFC3*, and *UBE2T* in breast cancer cells, we achieved real-time quantitative polymerase chain reaction (RT-qPCR) to identify the mRNA levels

of these 10 genes in T-47D, MCF-7, and MDA-MB-231 human breast cancer cells and MCF-10A mammary epithelial cells (normal cells) (Figure 4D). The expression levels of *MDC1*, *POLR2K*, *PSMB1*, *PSMD2*, *PSMD7*, *PSMD14*, *RFC3*, and *UBE2T* were higher in breast cancer cells than in normal cells. In contrast, the expression levels of *PARP3* and *PSMB9* were lower in breast cancer cells than in normal cells. These results are consistent with those of the survival analysis, indicating the suitability of the prediction model based on the prognosis-related genes.

Analysis and evaluation of the risk diagnostic model for breast cancer

We implemented univariate Cox regression analysis to screen 10 prognosis-related genes (Figure 5A). Then, we used the least absolute shrinkage and selection operator (LASSO)-Cox regression model to reduce the dimensionality. The parameter lambda.min was selected as the critical point for the linear risk assessment model composed of seven genes (*MDC1*, *PARP3*, *PSMB1*, *PSMB9*, *PSMD2*, *PSMD7*, and *PSMD14*) (Figures 5B, C). The gene descriptions and biological processes are shown in Supplementary Table 5. The median risk score of all patients was used as the cutoff value. The patients were divided into the high-risk (n = 480) and low-risk (n = 480) groups using the risk diagnostic model to calculate the risk score of each patient in the training cohort. By analyzing the time-dependent receiver operating characteristic (ROC) curves, we discovered that the model exhibited a clinical significance in terms of the 5-year and 10-year survival rates in patients with breast cancer (area under the curve [AUC] = 0.632 and 0.645, respectively), indicating the good predictive ability of the model in breast cancer (Figure 5D). Additionally, Kaplan-Meier analysis displayed that the overall survival rate of patients in the high-risk group was significantly lower than that of patients in the low-risk group (P = 1.708e-04), indicating the suitability of the risk diagnostic model for predicting the prognosis of patients with breast cancer (Figure 5F). After successfully establishing the risk diagnostic model, we used the GSE26085 dataset as the validation cohort to analyze the overall survival rates and ROC curves. We found that the overall survival rate of patients in the high-risk group was significantly lower than that of patients in the low-risk group (P = 1.694e-02). We also found that the model exhibited a clinical significance in terms of the 5-year and 10-year survival rates (area under the curve [AUC] = 0.641 and 0.647, respectively) in the validation cohort (Figures 5E, G). Subsequently, we explored the protein levels of *MDC1*, *PARP3*, *PSMB1*, *PSMB9*, *PSMD2*, *PSMD7*, and *PSMD14* in normal breast tissues and breast cancer tissues. As shown in Figure 5H, the protein levels of *MDC1*, *PSMB1*, *PSMB9*, *PSMD2*, *PSMD7*, and *PSMD14* were increased, while that of *PARP3* was decreased in breast cancer tissues compared with normal tissues.

Tumor-infiltrating immune cells in the risk diagnostic model

Then, we investigated 22 tumor-infiltrating immune cell subtypes in the high- and low-risk groups in the breast cancer training cohort. Of the subtypes, 11 varied significantly between the high- and low-risk groups. Furthermore, the CD8+ T cells, activated NK cells, M0 macrophages, M2 macrophages, resting dendritic cells, and resting mast cells showed significant differences in terms of expression between the high- and low-risk groups (P < 0.0001) (Figure 6A). Furthermore, we investigated the effects of the seven DDR genes on immune cell infiltration in the training cohort using the TIMER database

(<https://cistrome.shinyapps.io/timer/>). Different types of somatic copy number alterations (SCNAs), including those with two copies missing, one copy missing, normal copy number, or multiple copy numbers, in the seven genes were shown to significantly regulate immune cell infiltration in the breast cancer microenvironment (Figure 6B-H). Only the abundance of CD8+ T cells showed significant differences among all seven genes, indicating that CD8+ T cells may be a potential biomarker for distinguishing patients with favorable responses to immunotherapy based on our risk diagnostic model.

Functional identification of risk diagnostic model in breast cancer

We first collected 6 paired breast cancer tissues and adjacent normal tissues to verify the mRNA expression of these 7 genes in vivo. Consistent with the IHC results (Figure 5H), the expression of MDC1, PSMB1, PSMD2, PSMD7 and PSMD14 were significant augmented in breast cancer patients ($P < 0.05$). In contrast, the expression of PARP3 were significant declined in breast cancer patients ($P < 0.05$) (Figure 7A). After that, to test the function of risk diagnostic model in breast cancer, we chose MDC1 as candidate gene to detect cell proliferation, cell growth, cell invasion in breast cancer cells. We achieved a growth assay in MDA-MB-231 harboring loss of function of MDC1. Figure 7B indicated that MDC1 depletion significant decreased the proliferation rate ($P < 0.05$). The results of EdU assays further shown that MDC1 depletion could inhibit cell proliferation (Figure 7C). We next assessed the invasive capabilities by the transwell invasion assays and found that MDC1 knockdown led to significant decline in the invasive capacity of MDA-MB-231 cells ($P < 0.05$) (Figure 7D). To further investigate the inhibition of invasive capacity of MDA-MB-231 cells with MDC1 depletion, we detected the epithelial markers and mesenchymal markers expression in breast cancer cells. Results showed that MDC1 depletion augmented E-cadherin, α -catenin and γ -catenin at both the mRNA and protein levels, whereas the mRNA and protein levels for mesenchymal markers, including N-cadherin, vimentin and fibronectin were decreased (Figure 7E, F). These results suggested that MDC1 in risk diagnostic model could promote the proliferation, EMT, and invasion potential of breast cancer cells.

Discussion

TMB can be used to forecast the efficacy of the immune checkpoint blockade (ICB) therapy and has become a useful biomarker for recognizing patients with cancer who will benefit from immunotherapy. However, research on the use of DDR genes from tumors with a high TMB to construct a risk model for cancer treatment and prognosis is currently limited. In this study, we employed Cox regression analysis and LASSO-Cox regression analysis to construct a linear risk assessment model using differentially expressed DDR genes between the high- and low-TMB groups based on the genomic and transcriptomic data and clinical information of the verification cohort. With such stratification strategies, clinicians would be able to conveniently personalize medical treatment and health management for each patient with breast cancer.

Several studies have investigated the use of TMB in identifying patients who may respond to immunotherapy (8, 9). In recent years, such studies have mapped and characterized changes in TMB in pathological mechanisms. One study evaluated the distribution of TMB in 100,000 cancer cases and found that a group of patients exhibited a high TMB, which was associated with microsatellite instability. A group of somatic mutations in the *PMS2* gene promoter was also identified to 10% of skin cancers and was significantly correlated with an increase in TMB (10). Furthermore, another study reported that the tumor types with the highest percentage of mutations were thyroid cancer, breast cancer, and melanoma (11). ICB therapy produces a lasting anti-tumor effect in a variety of cancers, but not all patients respond to this therapy. The evaluation of more than 300 patient samples of 22 different cancer types in four major clinical trials showed that TMB and T cell inflammatory gene expression profiles play a joint predictive role in distinguishing responders and non-responders to PD-1 antibody therapy (12). Clinical studies that investigated TMB discovered that TMB was significantly associated with the wild-type epidermal growth factor receptor (*EGFR*) gene and a *TP53* mutation-positive status in 92 patients with lung cancer who endured surgery between 2013 and 2016 (13). Therefore, analyzing specific genetic changes, such as in *TP53*, may be a useful alternative in predicting TMB.

In this study, we mapped the landscape of DDR gene mutations in the high- and low-TMB groups and found that the top five high-frequency mutant genes in the high-TMB group were *TP53*, *PRKDC*, *BRCA2*, *BRCA1*, and *ATM*. Moreover, we found that the frequency of gene mutations in the high-TMB group was higher than that in the low-TMB group. *TP53* is known to have a major role in the regulation and repair of genomic damage. *TP53* gene mutations are one of the most common mutations in several cancers. There is considerable evidence proving that *TP53* affects TMB. Patients with *TP53* mutations represent a different molecular cohort that exhibits a poor prognosis. It was discovered that the expression of PD-L1 was enhanced in the hematopoietic stem cells of patients with *TP53* mutations, which were related to the upregulation of MYC and downregulation of the p53 transcriptional target miR-34a. It is worth noting that patients with *TP53* mutations showed a significant decrease in the number of bone marrow infiltrating T cells, leading to a decrease in ICOS⁺ and 4-1BB⁺ natural killer (NK) cells (14). Moreover, the microenvironment of *TP53*-mutant myelodysplastic syndromes (MDS) have been shown to possess immune-dominant and immune-evasive phenotypes, which may provide better therapeutic effects for patients with such *TP53* mutations. *BRCA1/2* alterations are caused by somatic or germline mutations or homologous recombination (HR)-related defects caused by other factors (15). For instance, *BRCA1* promoter methylation or other potential mutations in DDR genes can lead to *BRCA1/2* deficiency in patients with breast cancer (16). The combination therapy of cisplatin and PD-1 significantly enhanced the anti-tumor immunity of *BRCA1*-deficient mice, resulting in a strong systemic and intratumoral immune response (17). Furthermore, triple-negative breast cancer (TNBC) with *BRCA1* mutations treated with ICB therapy was reported to improved clinical outcomes. Ultimately, *TP53*, *PRKDC*, *BRCA2*, *BRCA1*, and *ATM* mutations may be potential biomarkers for predicting the clinical response of patients to immunotherapy to improve breast cancer survival.

We analyzed a risk diagnostic model based on seven DDR genes, *PSMD2*, *PSMD7*, *PSMD14*, *PARP3*, *MDC1*, *PSMB1*, and *PSMB9*, reflecting an enhanced level of predicting the survival and prognosis of

patients with breast cancer. The 26S proteasome non-ATPase regulatory subunit (PSMD) 2 (PSMD2), PSMD7, and PSMD14 proteins participate in the ubiquitin-proteasome system, which plays a potential role in the proliferation and progression of tumor cells. In hepatocellular carcinoma cells, *PSMD2* knockout reduced the formation of lipid droplets and modulated the expression of genes associated with lipid synthesis through the p38-JNK and AKT signaling pathways (18). PSMD7 has similar functions in breast cancer. The level of PSMD7 was significantly elevated in breast cancer and was closely related to tumor subtype, tumor size, lymph node invasion, and tumor-node-metastasis (TNM) stage. PSMD14 participates in the regulation of cancer occurrence and progression through a variety of molecular mechanisms. In our study, the *PARP3*, *POLR2K*, *PSMB1*, and *PSMD2* genes were significantly associated with the overall survival of patients with breast cancer, and the high expression levels of *POLR2K*, *PSMB1*, and *PSMD2* were related to low survival rates. Proteasome β -subunit 1 (PSMB1), a member of the proteasome β -subunit family, was found to interact with inhibitor of $\text{I}\kappa\text{B}$ kinase ϵ (IKK- ϵ) and promote the degradation of IKK- ϵ . The binding of PSMB1 to the *BCL-3* oncogene is necessary for proteasome degradation. As such, cells with a *PSMB1* deletion were found to exhibit defects in the polyubiquitin degradation of the BCL-3 protein (19). In addition, PSMB1 was shown to affect the response of follicular lymphoma to bortezomib and that the presence of the *PSMB1* rs12717 minor allele predicted the ineffectiveness of bortezomib in patients with myeloma (20). Based on our results, combined with the above evidence, our risk diagnostic model has the prospective to predict the survival and prognosis of patients with breast cancer. Moreover, three DDR genes, *POLR2K*, *PSMB1*, and *PSMD2*, which are closely linked to tumorigenesis, may be used as potential biomarkers for predicting the prognosis of patients with breast cancer.

During cancer treatment, an effective immune response is required to damage the function of tumor cells and ultimately destroy them. However, tumor cells have evolved a variety of mechanisms to escape immune surveillance, resulting in the inhibition of immune cell function and loss of the anti-tumor immune response. Therefore, a new type of monoclonal antibody, ICIs, has become one of the most critical immunotherapeutic methods in cancer treatment. Therefore, in immunotherapy research, analysis of immune cell infiltration in cancer is necessary. So, we discovered that there were significant differences in terms of immune cell abundance, particularly in terms of CD8+ T cells, activated NK cells, and M0, M1, and M2 macrophages, between the high- and low-TMB groups. CD8+ T cells can produce and express T cell receptors in the thymus. T cells induce immune responses in cancer, autoimmunity, and infection, and Th cells and CD8+ T cells play an essential role in tumor progression. In a study involving mice with TNBC, memory CD8+ T cells were found to be improved in the peripheral blood. In another study, the effect of *Plasmodium* infection on mouse breast cancer cells was determined to be linked to the initiation of an anti-tumor immune response regulated by CD8+ T cells (21). In addition, an analysis of clinical samples of metastatic melanoma revealed that the coexistence of CD20+ B cells and CD8+ T cells in tumors was related to the improvement of survival of patients with metastatic melanoma. Thus, CD8+ T cells likely play a key role in the immune microenvironment of melanoma, improving clinical outcomes, and may predict the prognosis of patients subjected to ICB therapy (22). NK cells, which also play a main role in immunity, are not only involved in immunoregulation and anti-tumor and anti-viral infection

responses, but also participate in hypersensitivity and autoimmunity on certain occasions. In a TNBC xenotransplantation model, the distribution and aggregation patterns of NK cells in the tumor site were found to differ across the distinctive stages of tumor progression (23).

There are many events of gene copy number variation in the progression of breast cancer. In this study, we analyzed SCNAs and immune cell infiltration to assess the function of our risk diagnostic model based on seven genes. The SCNAs of *PARP3*, *PSMB1*, *PSMD7*, and *PSMD1* showed significant differences in terms of immune infiltration, particularly in terms of B cell, CD8+ T cell, CD4+ T cell, macrophage, and dendritic cell infiltration. Poly (ADP-ribose) polymerase (PARP) 3 (PARP3) exhibits high homology with PARP1 and PARP2. PARP3 plays a role in DNA single- and double-strand break repair and humoral immunity. PARP3 was reported to be associated with the progression of gliomas and breast cancer. Moreover, the inhibition of PARP3 in lung cancer cells and osteosarcoma cells was found to increase telomerase activity, promote telomere maintenance, and lessen gene instability (24). Similarly, the absence of PARP3 was reported to enhance NADPH oxidase 4 (NOX4)-induced oxidative stress and reduce mechanistic target of rapamycin complex 2 (mTORC2) activation, resulting in an inefficient differentiation of neural stem cells or progenitor cells into astrocytes after birth (25). PARP3 was also found to interact with glycogen synthase kinase 3 beta (GSK3 β), a positive regulator of ubiquitin and rapamycin-insensitive companion of mammalian target of rapamycin (RICTOR) degradation, producing adenosine diphosphate. Knockout or inhibition of the *PARP3* gene aggravated centrosome amplification and genomic instability, reducing the proliferation, survival, and tumorigenicity of *BRCA1*-deficient TNBC cells (26). These results suggest that targeting the catalytic activity of PARP3 is a suitable approach for the treatment of *BRCA1*-related tumors through the RICTOR/mTORC2 signaling pathway. Currently, research on PARP3 and the tumor immune microenvironment is still limited. Our results suggest that focusing on PARP3, PSMB1, PSMD7, and PSMD14 and their roles in immunotherapy is a reasonable strategy for breast cancer treatment.

Conclusions

In summary, we established and validated a risk diagnostic model based on TMB characteristics and DDR genes and showed that the model has potential applications in predicting the clinical benefits of ICB therapy and the prognosis of patients with breast cancer. This model can also be used to determine patients with breast cancer who would respond favorably to immunotherapy. The limitation of this study is that the molecular mechanisms of the DDR genes were not fully explored. Therefore, prospective studies are needed to verify our risk diagnostic model and to further elucidate the detailed molecular mechanisms of the seven DDR genes as clinical biomarkers.

Declarations

Ethics approval and consent to participate

The study was authorized by the Ethics Committee of Capital Medical University, and the written informed consent was signed by each patient. All the experiment protocol for involving human data was in accordance with the guidelines of national/international/institutional or Declaration of Helsinki in the manuscript.

Consent for publication

All authors give consent for the publication of the manuscript in BMC Cancer.

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request. The datasets for this study can be found in the GEO database <https://www.ncbi.nlm.nih.gov/geo/>, KM-plotter <http://kmplot.com/analysis/index.php?p=service&cancer=breast> and The Human Protein Atlas <https://www.proteinatlas.org/>.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Natural Science Foundation of Beijing (7204241) and the National Natural Science Foundation of China (81902960). The State Key Laboratory of Molecular Oncology (SKLMO-KF2021-21).

Authors' contributions

TYM and WH contributed conception and designed the research study. XY, YKY and HFY collected and organized the database. TSY, BWY and XT performed data analysis. TSY and XY performed the validation experiments. TSY and WH wrote the manuscript. TYM and SW revised the manuscript. All authors read and finally approved the manuscript.

Acknowledgements

We would like to thank Editage (www.editage.cn) for English language editing.

References

1. Kakoti S, Sato H, Laskar S, Yasuhara T, Shibata A. DNA Repair and Signaling in Immune-Related Cancer Therapy. *Front Mol Biosci.* 2020;7:205.
2. Tray N, Weber JS, Adams S. Predictive Biomarkers for Checkpoint Immunotherapy: Current Status and Challenges for Clinical Application. *Cancer Immunol Res.* 2018;6(10):1122-8.

3. Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev Cancer*. 2019;19(3):133-50.
4. Chatterjee N, Walker GC. Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen*. 2017;58(5):235-63.
5. Rennert G, Bisland-Naggan S, Barnett-Griness O, Bar-Joseph N, Zhang S, Rennert HS, et al. Clinical outcomes of breast cancer in carriers of BRCA1 and BRCA2 mutations. *N Engl J Med*. 2007;357(2):115-23.
6. Shao Y, Hu X, Yang Z, Lia T, Yang W, Wu K, et al. Prognostic factors of non-muscle invasive bladder cancer: a study based on next-generation sequencing. *Cancer Cell Int*. 2021;21(1):23.
7. Das S, Camphausen K, Shankavaram U. Pan-Cancer Analysis of Potential Synthetic Lethal Drug Targets Specific to Alterations in DNA Damage Response. *Front Oncol*. 2019;9:1136.
8. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500(7463):415-21.
9. Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. *Science*. 2015;348(6230):69-74.
10. Chalmers ZR, Connelly CF, Fabrizio D, Gay L, Ali SM, Ennis R, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med*. 2017;9(1):34.
11. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med*. 2017;23(6):703-13.
12. Cristescu R, Mogg R, Ayers M, Albright A, Murphy E, Yearley J, et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science*. 2018;362(6411).
13. Ozaki Y, Muto S, Takagi H, Watanabe M, Inoue T, Fukuhara M, et al. Tumor mutation burden and immunological, genomic, and clinicopathological factors as biomarkers for checkpoint inhibitor treatment of patients with non-small-cell lung cancer. *Cancer Immunol Immunother*. 2020;69(1):127-34.
14. Sallman DA, McLemore AF, Aldrich AL, Komrokji RS, McGraw KL, Dhawan A, et al. TP53 mutations in myelodysplastic syndromes and secondary AML confer an immunosuppressive phenotype. *Blood*. 2020;136(24):2812-23.
15. Knijnenburg TA, Wang L, Zimmermann MT, Chambwe N, Gao GF, Cherniack AD, et al. Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. *Cell Rep*. 2018;23(1):239-54 e6.
16. Nik-Zainal S, Morganella S. Mutational Signatures in Breast Cancer: The Problem at the DNA Level. *Clin Cancer Res*. 2017;23(11):2617-29.
17. Nolan E, Savas P, Policheni AN, Darcy PK, Vaillant F, Mintoff CP, et al. Combined immune checkpoint blockade as a therapeutic strategy for BRCA1-mutated breast cancer. *Sci Transl Med*. 2017;9(393).

18. Tan Y, Jin Y, Wu X, Ren Z. PSMD1 and PSMD2 regulate HepG2 cell proliferation and apoptosis via modulating cellular lipid droplet metabolism. *BMC Mol Biol.* 2019;20(1):24.
19. Keutgens A, Zhang X, Shostak K, Robert I, Olivier S, Vanderplasschen A, et al. BCL-3 degradation involves its polyubiquitination through a FBW7-independent pathway and its binding to the proteasome subunit PSMB1. *J Biol Chem.* 2010;285(33):25831-40.
20. Varga G, Mikala G, Kiss KP, Kosoczki E, Szabo E, Meggyesi N, et al. Proteasome Subunit Beta Type 1 P11A Polymorphism Is a New Prognostic Marker in Multiple Myeloma. *Clin Lymphoma Myeloma Leuk.* 2017;17(11):734-42.
21. Pan J, Ma M, Qin L, Kang Z, Adah D, Tao Z, et al. Plasmodium infection inhibits triple negative 4T1 breast cancer potentially through induction of CD8(+) T cell-mediated antitumor responses in mice. *Biomed Pharmacother.* 2021;138:111406.
22. Cabrita R, Lauss M, Sanna A, Donia M, Skaarup Larsen M, Mitra S, et al. Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature.* 2020;577(7791):561-5.
23. Uong TNT, Lee KH, Ahn SJ, Kim KW, Min JJ, Hyun H, et al. Real-Time Tracking of Ex Vivo-Expanded Natural Killer Cells Toward Human Triple-Negative Breast Cancers. *Front Immunol.* 2018;9:825.
24. Fernandez-Marcelo T, Frias C, Pascua I, de Juan C, Head J, Gomez A, et al. Poly (ADP-ribose) polymerase 3 (PARP3), a potential repressor of telomerase activity. *J Exp Clin Cancer Res.* 2014;33:19.
25. Rodriguez-Vargas JM, Martin-Hernandez K, Wang W, Kunath N, Suganthan R, Ame JC, et al. Parp3 promotes astrocytic differentiation through a tight regulation of Nox4-induced ROS and mTorc2 activation. *Cell Death Dis.* 2020;11(11):954.
26. Beck C, Rodriguez-Vargas JM, Boehler C, Robert I, Heyer V, Hanini N, et al. PARP3, a new therapeutic target to alter Rictor/mTORC2 signaling and tumor progression in BRCA1-associated cancers. *Cell Death Differ.* 2019;26(9):1615-30.
27. Paul MR, Pan TC, Pant DK, Shih NN, Chen Y, Harvey KL, et al. Genomic landscape of metastatic breast cancer identifies preferentially dysregulated pathways and targets. *J Clin Invest.* 2020;130(8):4252-65.
28. Xu H, Stamova B, Ander BP, Waldau B, Jickling GC, Sharp FR, et al. mRNA Expression Profiles from Whole Blood Associated with Vasospasm in Patients with Subarachnoid Hemorrhage. *Neurocrit Care.* 2020;33(1):82-9.
29. Ha TJ, Swanson DJ, Kirova R, Yeung J, Choi K, Tong Y, et al. Genome-wide microarray comparison reveals downstream genes of Pax6 in the developing mouse cerebellum. *Eur J Neurosci.* 2012;36(7):2888-98.

Figures

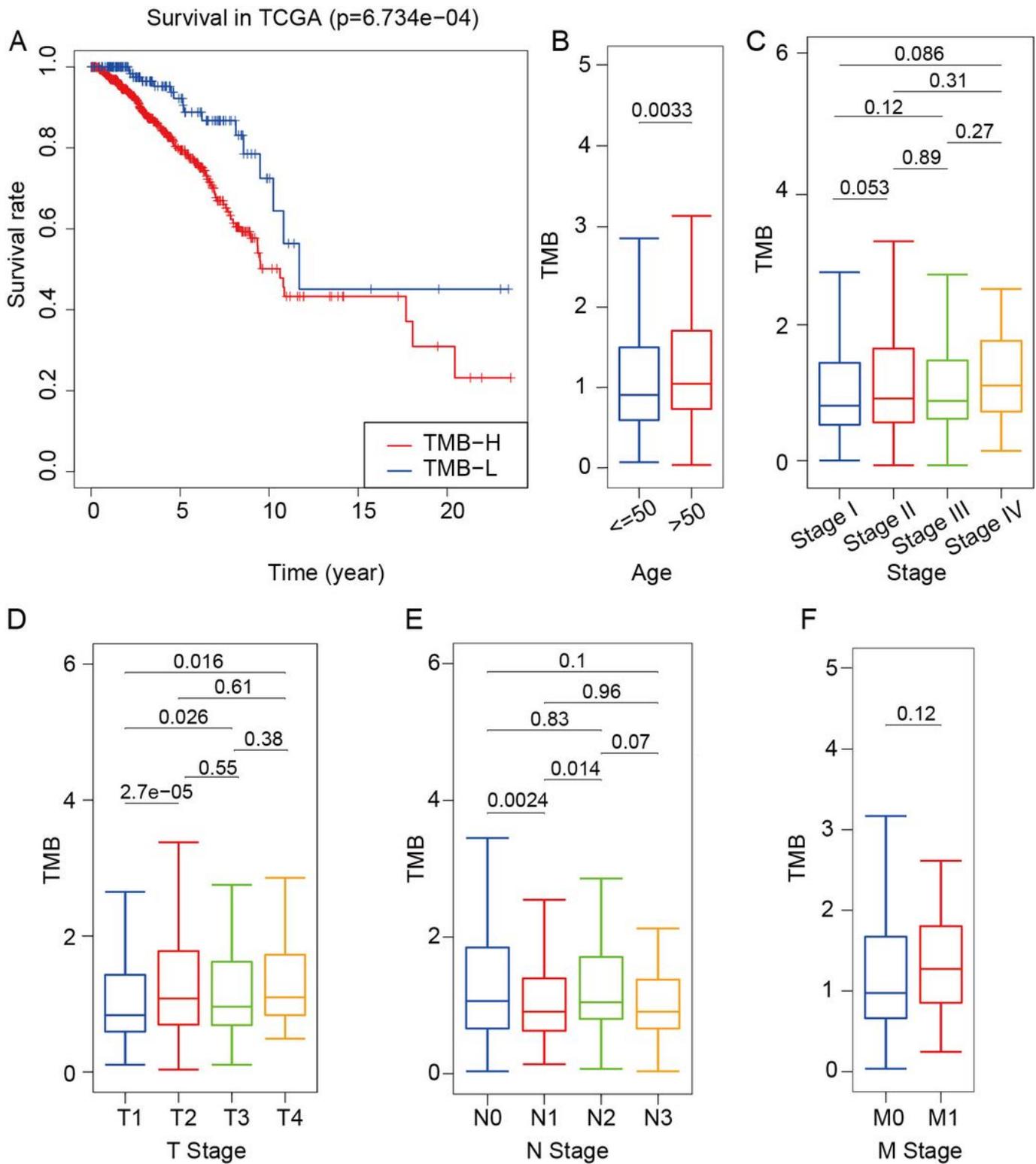


Figure 1

Difference in survival between the high- and low-TMB groups and correlation between TMB and clinical indexes. (A) Survival curve of training cohort based on TMB. (B–F) Correlational analyses between TMB and patient age, clinical stage, T stage, N stage and M stage.

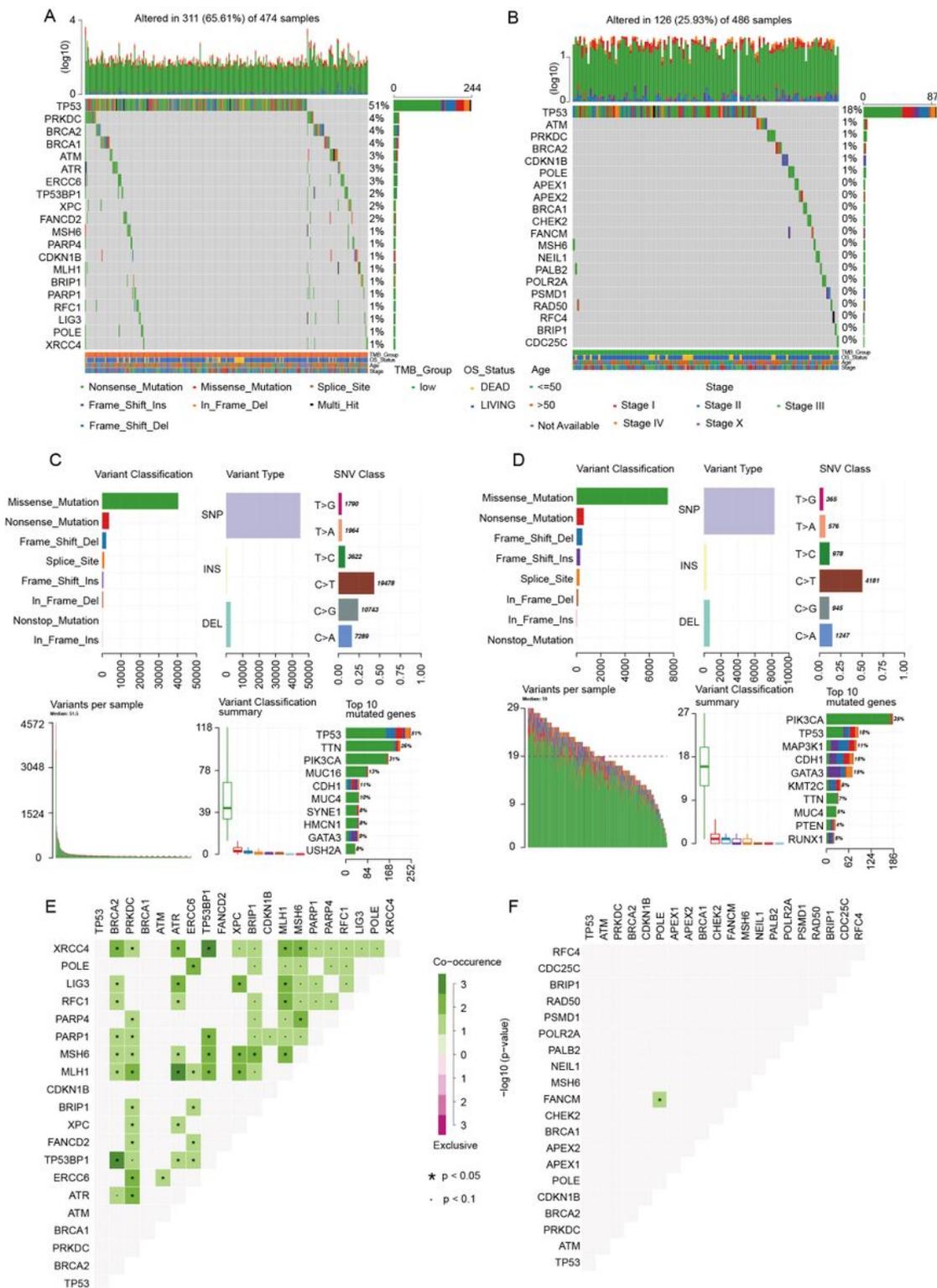


Figure 2

Mutational landscape of breast cancer. (A) Landscape of gene mutations of each high-TMB breast cancer sample in a waterfall plot. (B) Landscape of gene mutations of each low-TMB breast cancer sample in a waterfall plot. (C) Summary of statistical calculations for the frequency of mutation types in the high-TMB group. Missense mutation, single nucleotide polymorphism (SNP), and C > T mutation are the most frequent. Top 10 mutant genes in the high TMB group include TP53, TTN, PIK3CA, MUC16,

CDH1, MUC4, SYNE1, HMCN1, GATA3, and USH2A. (D) Summary of statistical calculations for the frequency of mutation types in the low-TMB group. Missense mutation, SNP, and C > T mutation are the most frequent. Top 10 mutant genes in the low-TMB group include PIK3CA, TP53, MAP3K1, CDH1, GATA3, KMT2C, TTN, MUC4, PTEN, and RUNX1. (E) Co-occurrence and exclusive analyses in the high-TMB group. (F) Co-occurrence and exclusive analyses in the low-TMB group.

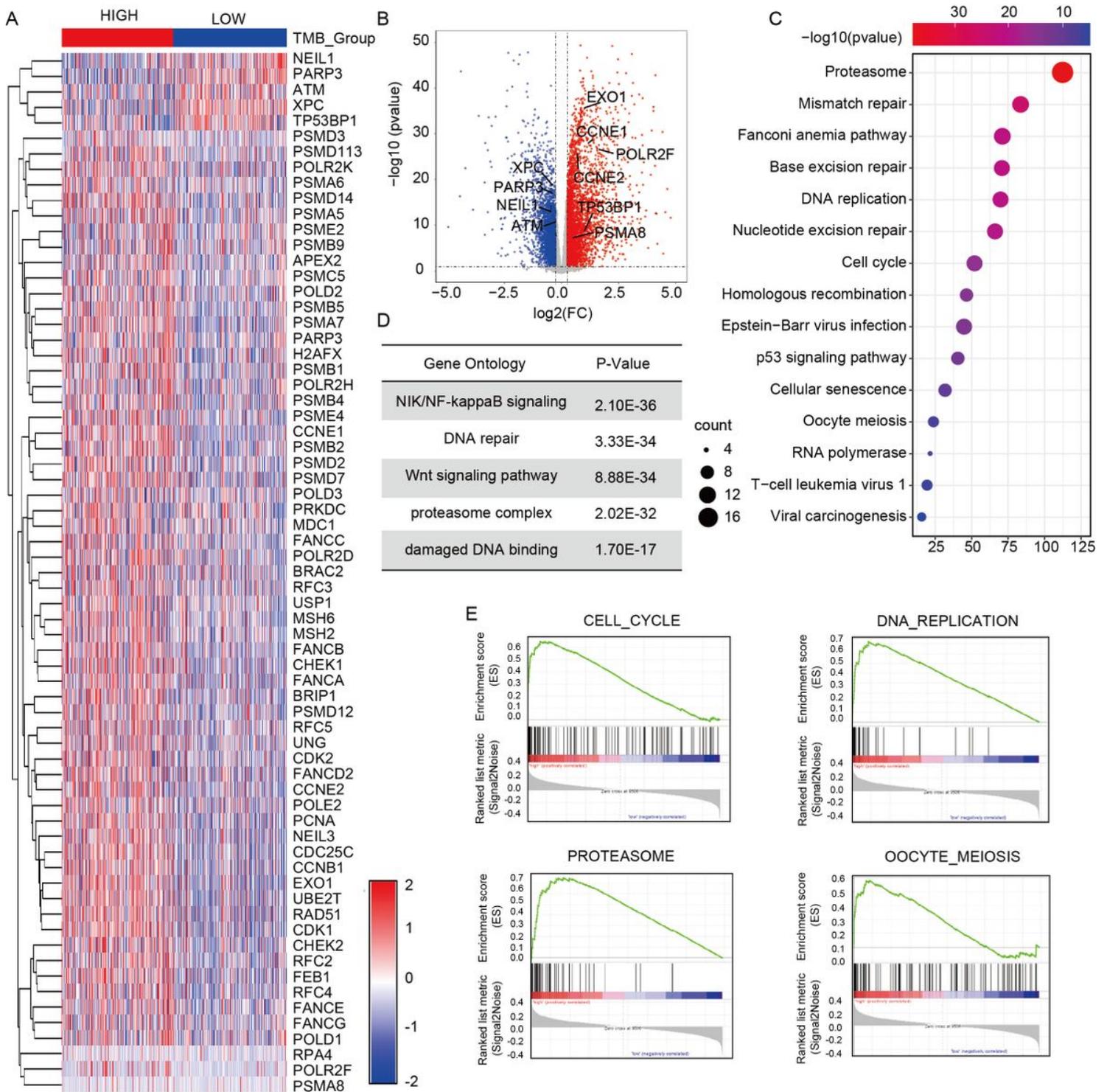


Figure 3

Screening of differentially expressed DDR genes of in the two TMB groups. (A) Heat map of 67 differentially expressed DDR genes in the high- and low-TMB groups. (B) Volcano plot of 6,424 differentially expressed genes in the high- and low-TMB groups. (C) KEGG analysis of the differential genes in breast cancer. (D) GO enrichment analysis of the differences in gene expression. (E) GSEA in the high- and low-TMB groups. Top four pathways enriched in the high-TMB group are cell cycle, DNA replication, proteasome, and oocyte meiosis.

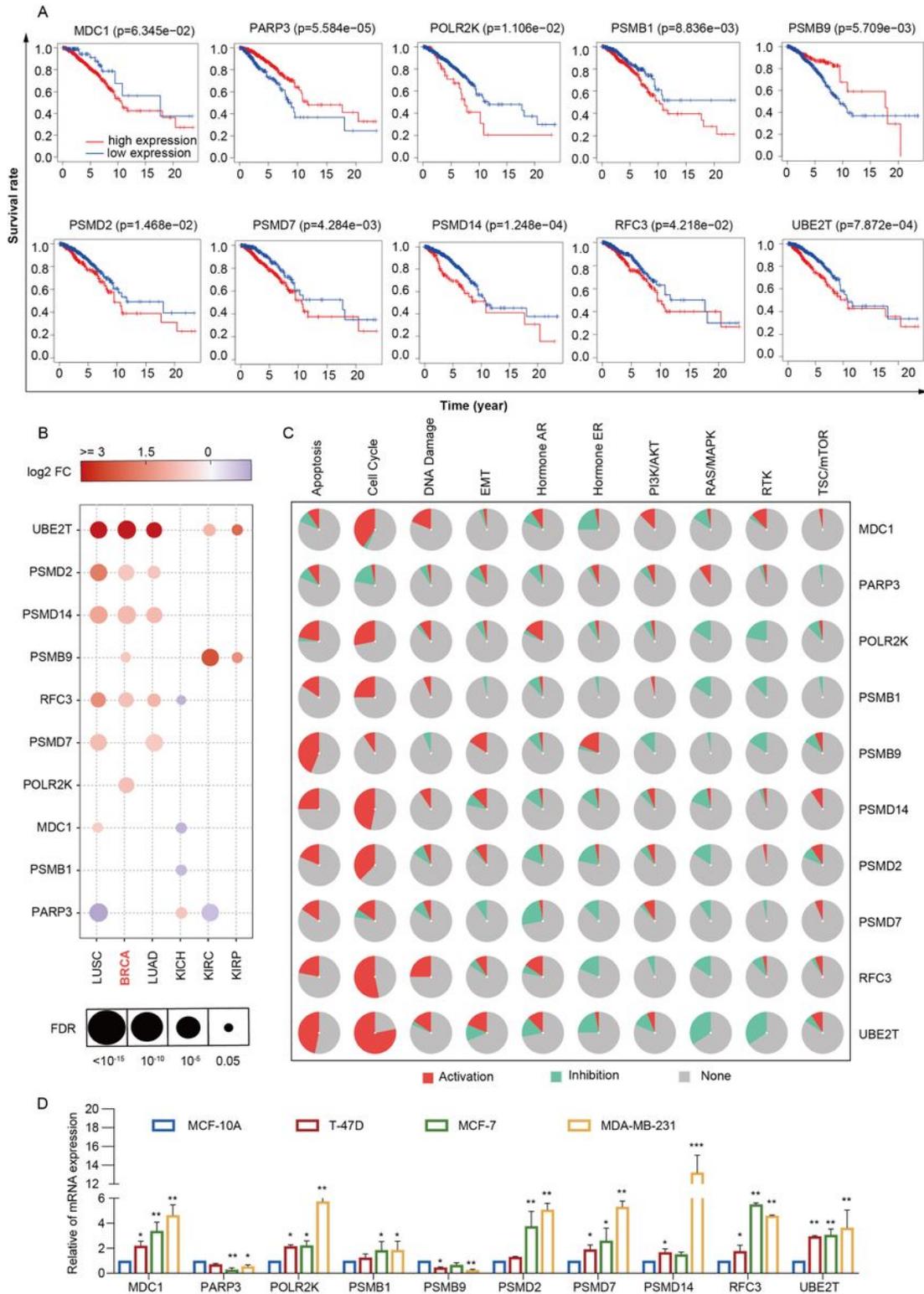


Figure 4

Comprehensive analysis of 10 DDR genes. (A) Kaplan–Meier analysis of MDC1, PARP3, POLR2K, PSMB1, PSMB9, PSMD2, PSMD7, PSMD14, RFC3 and UBE2T. (B) Expression of MDC1, PARP3, POLR2K, PSMB1, PSMB9, PSMD2, PSMD7, PSMD14, RFC3 and UBE2T in lung squamous cell carcinoma, breast cancer, lung adenocarcinoma, kidney chromophobe, kidney renal clear cell carcinoma, and kidney renal papillary carcinoma from the GSCALite database. (C) Different pathways correlated with MDC1, PARP3, POLR2K, PSMB1, PSMB9, PSMD2, PSMD7, PSMD14, RFC3 and UBE2T from the GSCALite database. (D) RT-qPCR analyses of MDC1, PARP3, POLR2K, PSMB1, PSMB9, PSMD2, PSMD7, PSMD14, RFC3 and UBE2T expressions in MCF-10A, T-47D, MCF-7, and MDA-MB-231 cells. Bars represent the mean \pm SD of triplicate cell cultures (*P < 0.05, **P < 0.01, ***P < 0.001).

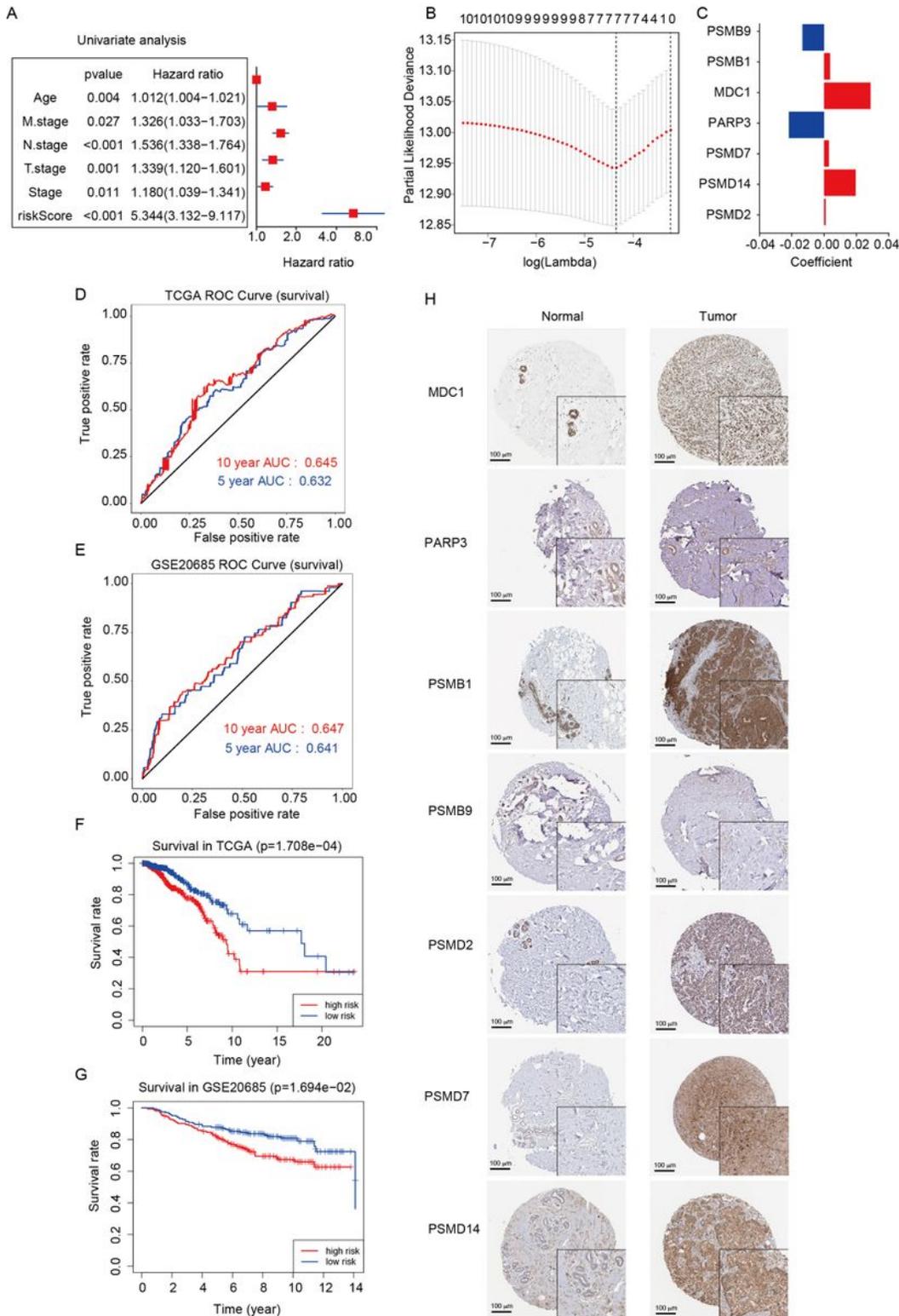


Figure 5

Analysis and evaluation of the risk diagnostic model for breast cancer. (A) Forest map of univariate Cox regression analysis for the selection of 10 genes related to prognosis. (B) Determination of the Lambda coefficient of LASSO-Cox regression analysis. (C) The regression coefficients (coincident values) of the seven genes included in the model. (D, E) AUC of ROC curves displaying the predictive accuracy of the risk scores in the training and validation cohorts. (F, G) Kaplan-Meier analysis of the risk diagnostic model in

the training and validation cohorts. (H) Immunohistochemistry of the MDC1, PARP3, PSMB1, PSMB9, PSMD2, PSMD7 and PSMD14 proteins in breast cancer and normal breast tissues from the Human Protein Atlas.

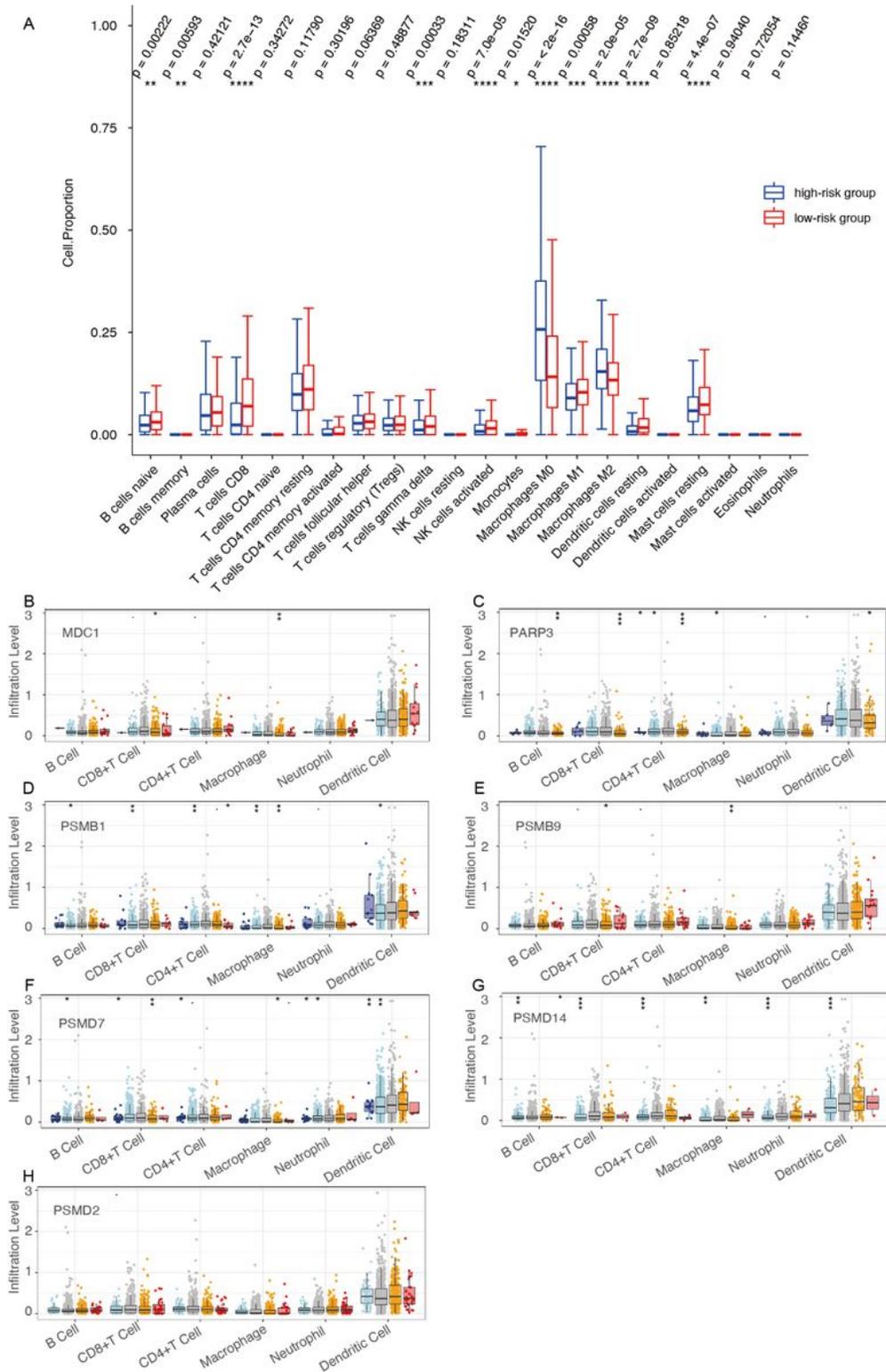


Figure 6

Tumor-infiltrating immune cells in the risk diagnostic model. (A) Analysis of tumor-infiltrating immune cells in the high- and low-risk groups. (B) Analysis of the differences in immune infiltration and SCNAs

among seven DDR genes (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$). I, RT-qPCR analyses of MDC1, PARP3, PSMB1, PSMB9, PSMD2, PSMD7 and PSMD14 expression in paired human breast (n = 6) cancer tissues and adjacent normal tissues.

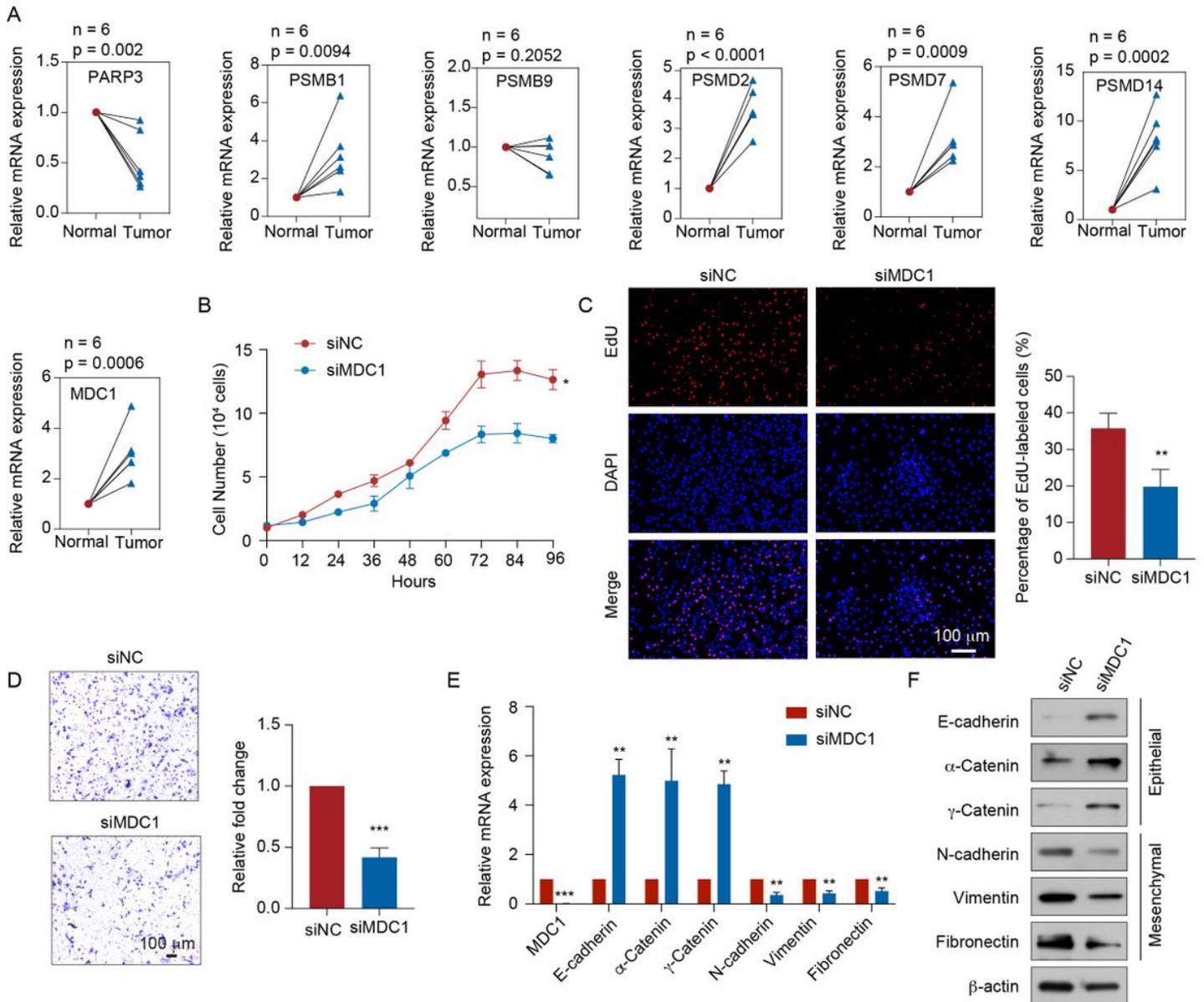


Figure 7

Functional identification of risk diagnostic model in breast cancer. (A) RT-qPCR analyses of MDC1, PARP3, POLR2K, PSMB1, PSMB9, PSMD2, PSMD7, PSMD14, RFC3 and UBE2T expressions in 6 paired breast cancer tissues and adjacent normal tissues. Bars represent the mean \pm SD for three independent experiments (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). (B) Cell proliferation curves of the MDC1 depleted (with siRNA) in MDA-MB-231 cells. Error bars represent means \pm SD for three independent experiments (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). (C) EdU incorporation assays were performed on MDC1 depleted MDA-MB-

231 cells. Representative images are shown on the left, and statistical analysis is shown on the right (*P < 0.05, **P < 0.01, ***P < 0.001). (D) Cell invasion assays were performed using the matrigel transwell filters in MDC1 depleted MDA-MB-231 cells. Invading cells were stained and counted. Representative images are shown on the left, and statistical analysis is shown on the right (*P < 0.05, **P < 0.01, ***P < 0.001). (E) RT-qPCR analyses of EMT markers expressions in MDC1 depleted MDA-MB-231 cells. Bars represent the mean ± SD for three independent experiments (*P < 0.05, **P < 0.01, ***P < 0.001). (F) Western blot using the indicated antibodies were performed on total protein extracted from MDC1 depleted MDA-MB-231 cells.

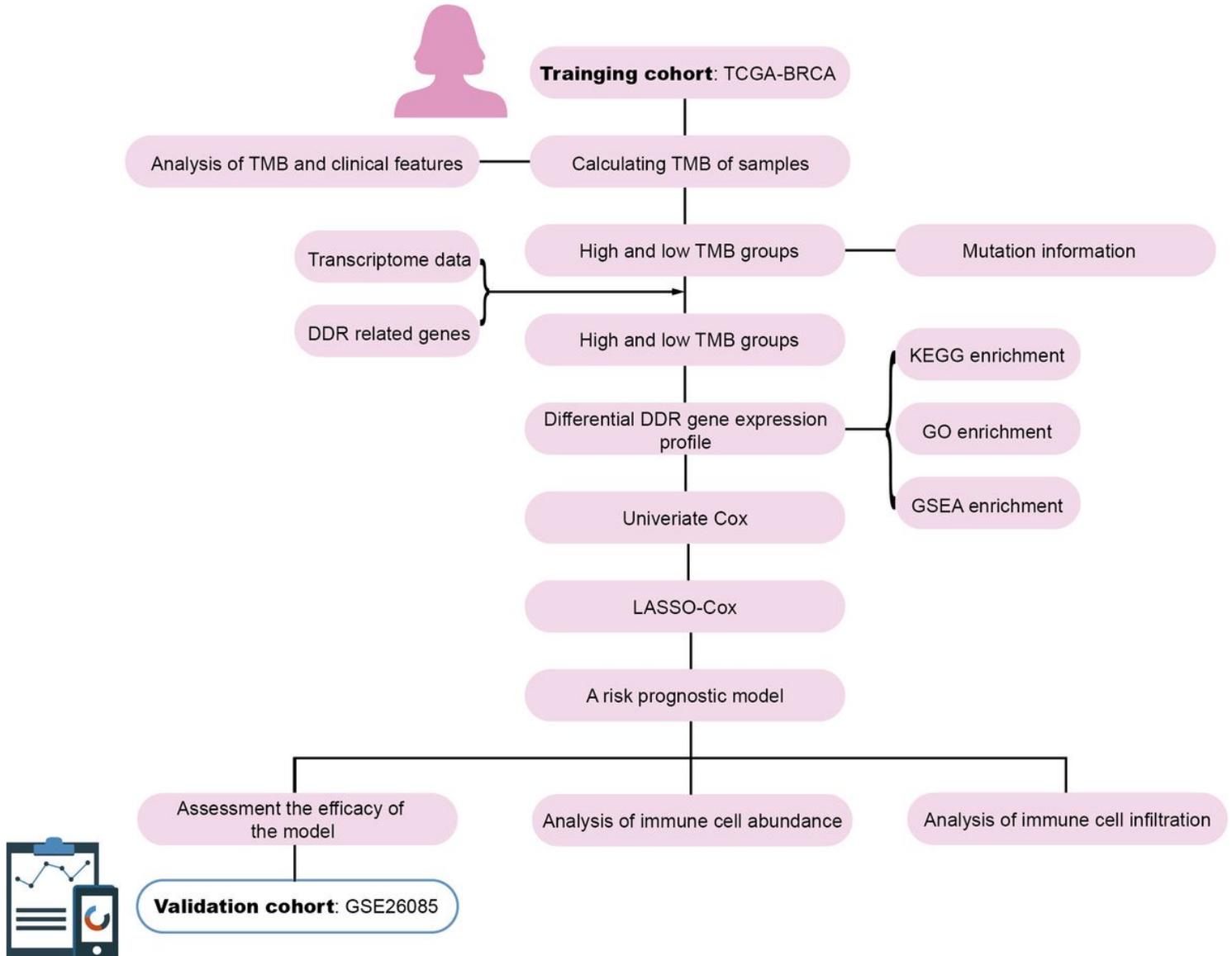


Figure 8

Workflow of the establishment of the risk prognostic model for breast cancer.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1.xls](#)
- [SupplementaryTable2.xls](#)
- [SupplementaryTable3.xls](#)
- [SupplementaryTable4.xls](#)
- [SupplementaryTable5.xls](#)