Screening of DNA Damage Repair Genes Involved in the Prognosis of Triple-Negative Breast Cancer Patients Based on Bioinformatics

Nan Wang

The First Affiliated Hospital of Zhengzhou University <https://orcid.org/0000-0002-0983-3389>

Yuanting Gu

The First Affiliated Hospital of Zhengzhou University

Jiangrui Chi

The First Affiliated Hospital of Zhengzhou University

Xinwei Liu

The First Affiliated Hospital of Zhengzhou University

Youyi Xiong

The First Affiliated Hospital of Zhengzhou University

Chaochao Zhong

The First Affiliated Hospital of Zhengzhou University

Fang Wang

The First Affiliated Hospital of Zhengzhou University

Xinxing Wang

The First Affiliated Hospital of Zhengzhou University <https://orcid.org/0000-0003-2894-783X>

Lin Li (✉ fcclil6@zzu.edu.cn)

The First Affiliated Hospital of Zhengzhou University <https://orcid.org/0000-0001-6484-7866>

Primary research

Keywords: DNA damage response, triple-negative breast cancer, prognostic model

Posted Date: June 11th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-594124/v1>

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

[Read Full License](#)

Abstract

Background: Triple-negative breast cancer (TNBC) is a special subtype of breast cancer with poor prognosis. DNA damage response (DDR) is one of the hallmarks of this cancer. However, the association of DDR genes with the prognosis of TNBC is still unclear.

Methods: We identified differentially expressed genes (DEGs) between normal and TNBC samples from The Cancer Genome Atlas (TCGA). DDR genes were obtained from the Molecular Signatures Database (MSigDB) through six DDR gene sets. We then overlapped the DEGs with DDR genes. Based on univariate and LASSO Cox regression analyses, a prognostic model was constructed to predict overall survival (OS). Kaplan–Meier (K–M) analysis and receiver operating characteristic (ROC) curve were used to assess the performance of the prognostic model. Cox regression analysis was applied to identify independent prognostic factors in TNBC. The prognostic model was validated using an independent dataset. Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were performed by using gene set enrichment analysis (GSEA). Single-sample gene set enrichment analysis (ssGSEA) was employed to estimate immune cells related to this prognostic model. Finally, we constructed a transcriptional factor (TF) network and a competing endogenous RNA (ceRNA) regulatory network.

Results: 23 differentially expressed DDR genes were detected between TNBC and normal samples. The six-gene prognostic model we developed was shown to be related to OS in TNBC using univariate and LASSO Cox regression analyses. By drawing ROC curve and KM curve, we determined the effectiveness of the risk model. The prognostic value of the six-gene prognostic model was further validated using the GSE58812 dataset. The GSEA analysis indicated that the genes in the high-risk group were mainly correlated with leukocyte migration, cytokine interaction with cytokine receptors, oxidative phosphorylation, autoimmune diseases, and coagulation cascade. The mutation data revealed that the mutation frequency of the two groups was the same, while the mutated genes were different. The gene-TF regulatory network showed that Replication Factor C subunit 4 (RFC4) occupied the dominant position.

Conclusion: We identified six gene markers related to DDR, which can predict prognosis and serve as an independent biomarker for TNBC patients.

Introduction

Breast cancer is the most common malignancy in women. The International Agency for Research on Cancer (IARC) of the World Health Organization reported that the number of new cases of breast cancer reached 2.26 million in 2020, and breast cancer has become the most prevalent malignant tumor in the world [1]. Breast cancer is a biologically and clinically heterogeneous disease with several recognized tissue and molecular subtypes with different etiologies, risk factors, treatment responses, and prognoses [2–4]. Triple-negative breast cancer (TNBC) is defined as a type of breast cancer with negative expression of estrogen (ER), progesterone (PR), and human epidermal growth factor receptor-2 (HER2) [5]. Compared

with other breast cancer subtypes, TNBC is highly aggressive and has a high rate of early recurrence. Patients with TNBC usually relapse within 5 years after surgery, and the overall prognosis is very poor [6]. Due to the lack of ER, PR, and HER2 expression in TNBC tumors, few therapies targeting specific molecular targets have been able to significantly improve the prognosis of patients with TNBC disease, and chemotherapy remains the standard of treatment of TNBC [5, 7]. Although many patients with early-stage TNBC disease are cured by chemotherapy, the overall median survival with the current treatment regimen is 13–18 months among those who develop metastatic disease [8]. Therefore, in order to improve the prognosis and curative effect of TNBC patients, it is urgent to obtain new and effective biomarkers.

DNA damage response (DDR) pathways are an important mechanism to correct and repair DNA damage, which can inhibit cell aging, apoptosis, and carcinogenesis in time and ensure normal life activities [9]. DDR consists of eight pathways: 1) mismatch repair (MMR), 2) base excision repair (BER), 3) nucleotide excision repair (NER), 4) homologous recombination repair (HRR), 5) non-homologous end ligation (NHEJ), 6) checkpoint factor (CPF), 7) Fanconi's anemia (FA), and 8) variable DNA synthesis (TLS). The interaction of these pathways can repair DNA damage accurately and timely, prevent gene distortion, and ensure the integrity of the genome [10]. Recent studies have shown that increasing DNA damage and reducing the DNA repair ability of cancer cells lead to genome distortion of cancer cells, but that distinguishing these cells from normal cells can improve the effectiveness of cancer treatment [11]. DDR genes can broaden therapy options for cancer patients by cancer-driving effects and significance in clinical and translational medicine [12]. For example, DDR alterations are independently associated with the therapeutic response to PD-1/PD-L1 inhibitors and are positively correlated with a higher tumor mutation burden (TMB) [13, 14]. Poly-ADP-ribose polymerase inhibitor (PAPRI) therapy had a better effect on cancer patients with BRCA1/2 mutations [15, 16]. Moreover, many studies have demonstrated that tumors with deleterious DDR mutations are more sensitive to platinum-based therapy [17]. Therefore, DDR genes are very important to the prognosis of patients, but there has yet to be a systematic study of DDR genes in TNBC.

In this study, we downloaded the sequencing data and DDR datasets of TNBC patients from The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO) databases and performed bioinformatics analysis on them to comprehensively evaluate whether the expression level of DDR-related genes can predict the prognosis of TNBC patients. The aim of our study was to identify new potential prognostic markers for TNBC and establish new prognostic models to assist in the formulation of diagnosis and treatment strategies. In addition, we stratified the risk of TNBC patients by establishing a prognostic model, and then performed more specific treatments on the patients according to the results of the different risk assessments, so as to avoid unnecessary active treatments for the patients.

Materials And Methods

Data collection and differential expression analysis

The gene expression data and clinical information of TNBC patients were acquired from the TCGA database (<https://portal.gdc.cancer.gov/cart>) and GEO database (<https://www.ncbi.nlm.nih.gov/gds>). We collected 99 normal and 113 TNBC samples from the TCGA database. DEGs were selected using the “limma” package ($|\log_2(\text{fold change})| > 1, P < 0.05$). We obtained six DDR gene sets that encompassed several DDRs pathways, including mismatch repair, base excision repair, checkpoint factors, nucleotide excision repair, homologous recombination repair, Fanconi’s anemia, and nonhomologous end-joining, from MSigDB, (<http://www.gsea-msigdb.org/gsea/msigdb/>), and 57 duplicate DDRs were removed. Finally, we collected 154 DDR genes and intersected them with DEGs (DE-DDRs).

Survival analysis

From a total of 113 TNBC samples, after excluding a TNBC sample lacking survival data, we used 112 TNBC samples to construct a prognostic model. Univariate and LASSO Cox analyses were used to identify prognostic genes that were significantly associated with OS. Univariate Cox analysis was used to initially identify the potential prognostic genes ($P < 0.02$). Next, we used the R package “glmnet” for the LASSO Cox regression analysis to construct the prognostic model for TNBC patients. The risk score was calculated according to the standardized expression level of each gene and its corresponding regression coefficient. The following formula was used: Risk score = $(\text{Coefficient}_{\text{mRNA1}} \times \text{Expression}_{\text{mRNA1}}) + (\text{Coefficient}_{\text{mRNA2}} \times \text{Expression}_{\text{mRNA2}}) + \dots + (\text{Coefficient}_{\text{mRNA}_n} \times \text{Expression}_{\text{mRNA}_n})$. Setting the median risk score of the TCGA database as the cut-off value, 112 TNBC samples were divided into low- or high-risk groups. The R package “survival” was performed to generate the K-M survival curve, and the R package “survivalROC” was used to generate time-dependent ROC curves to evaluate the predictive power of the prognostic model. The GSE58812 dataset, which contained 107 TNBC samples, was used for validation of the above results. Univariate and multivariate Cox regression analyses were performed to analyze the independent prognosis of the six-gene prognostic model. All independent prognostic factors were used to construct a nomogram to predict the survival of TNBC patients at 3 and 5 years.

Gene set enrichment analysis (GSEA)

To better understand the functional pathways of the high- and low-risk groups, we used GSEA to perform GO and KEGG enrichment analyses. GSEA was performed by using clusterProfiler. $P < 0.05$ was considered statistically significant.

Mutation analysis

The somatic mutation data of the 112 TNBC samples were obtained from the TCGA database. We used the “maftools” tool to comprehensively analyze mutation status in TNBC. The “somaticInteractions” function in the R package “maftools” was used to perform a Fisher test on the mutated genes in order to obtain the interaction relationships between them.

Immune analysis

The enrichment levels of 28 immune signatures in each TNBC sample were quantified by single-sample gene set enrichment analysis (ssGSEA) in the R package “GSVA.” Heat maps and violin plots were drawn

to observe the difference in the level of various immune cell infiltration between the high- and low-risk groups. Finally, the correlation between 6 genes and 28 immune signatures was calculated by the Spearman method.

Transcription factors (TFs) and ceRNA network construction

The Network analyst database (<http://www.networkanalyst.ca/faces/home.xhtml>) is an online visual analysis platform for gene expression analysis and meta-analysis. In this study, the Network analyst database was used to search the TFs related to the hub genes, which refers to the genes used to construct the prognostic model.

The mRNA-miRNA and miRNA-lncRNA interactions were predicted by using the miRanda database (<http://www.microrna.org/microrna/home.do>). To improve the reliability of the ceRNA network, we used a binding score > 500 and minimal folding free energy (MFE) < -50 for the predicted mRNA-target miRNA interaction. The screening criteria of miRNA-lncRNA were as follows: binding score > 4000 and MFE < -400.

Statistical analysis

All analyses were performed using the R software. Univariate and LASSO Cox regression analyses were used to assess the relationship between the prognostic model and OS. The Kaplan–Meier (K–M) method was used to assess survival analysis with a log-rank test. The ROC curves were used to detect the sensitivity and specificity of the prognostic model. $P < 0.05$ was considered statistically as significant.

Results

Identification of DE-DDRs

Following analysis of the TCGA database using limma, a total of 2,178 DEGs were identified in 113 TNBC and 99 normal samples ($|\log_2(\text{fold change})| > 1$, $P < 0.05$). Figure 1A illustrates the 946 upregulated and 1,232 downregulated genes using a volcano plot. As shown by the Venn diagrams in Fig. 1B, we selected 23 significant DE-DDRs (the intersection of 154 DDRs and 2,178 DEGs) for subsequent analysis. GO function annotation of the DE-DDRs was performed by the R package. These genes were mainly enriched in DNA replication, DNA recombination, and DNA-dependent DNA replication (Fig. 1C).

Construction of prognostic model in the TCGA database

In order to establish a prognostic model, univariate Cox regression analysis was performed on 23 genes, of which 6 genes were significantly associated with the OS of TCGA-TNBC (Fig. 2A). The regression coefficients of these six genes were calculated via LASSO COX regression analysis (Fig. 2B). The prognostic model achieved the best performance when the six genes were used (Fig. 2C). The formula of the model was: risk score = $(-0.18330185 \times \text{expression level of PARP1}) + (0.25938239 \times \text{expression level of ...})$

of BRIP1) + $(-0.71002582 \times \text{expression level of RMI2}) + (-0.05379813 \times \text{expression level of RFC4}) + (-1.01590214 \times \text{expression level of EXO1}) + (1.46313437 \times \text{expression level of RAD51})$. According to the median risk score, 56 of the 112 TNBC samples were classified into the high-risk group ($n = 56$), and the remaining 56 samples were classified into the low-risk group ($n = 56$) (Fig. 2D). Survival analysis indicated that the OS was lower in the high-risk group than the low-risk group ($P < 0.05$) (Fig. 2E). The time-dependent ROC curves were used to evaluate the prediction effect of the risk score, and the AUC was 0.821 at 3 years and 0.745 at 5 years (Fig. 2F). The relationships between risk score and clinicopathological features (age, sex, pathological stage and TNM stage) are shown in Fig. 2G.

Validation of the six-gene prognostic model in the GEO database

In order to verify the robustness of the prognostic model, we applied the model to the GEO cohort for external validation. Patients in the GSE58812 dataset ($n = 107$) were divided into a high-risk group ($n = 53$) and low-risk group ($n = 54$) using the formula obtained from TCGA-TNBC cohort (Fig. 3A). Consistent with the TCGA cohort, the survival probability of high-risk patients was significantly lower than that of low-risk patients (Fig. 3B). As shown in Fig. 3C, the AUC of the ROCs was 0.574 for 3 years and 0.663 for 5 years. Since there were only three patients with data regarding the 1-year follow-up, we did not plot the ROC curve of the 1-year follow-up for the TCGA and GSE58812 datasets. Collectively, these results indicated that the six-gene prognostic model was robust for survival prediction.

Independent prognostic value of the six-gene prognostic model

Next, we used univariate and multivariate Cox regression analyses to evaluate whether the six-gene prognostic model could serve as an independent predictor for TNBC patients. Univariate Cox regression analysis showed the variables of tumor stage, TNM stage, and risk score were significantly associated with OS ($P < 0.05$) (Fig. 4A). Multivariate Cox regression analysis indicated that N stage, T stage, and risk score were independent risk factors correlated with OS ($P < 0.05$) (Fig. 4B). Moreover, a nomogram was constructed to predict the possibility of 3-year and 5-year OS in TNBC patients by integrating the six-gene prognostic model with other clinicopathological characteristics (T and N stage). As shown in Fig. 4C, D, the nomogram and calibration curve demonstrated that the six-gene prognostic model was a valuable indicator for prognostic prediction.

Gene set enrichment analyses

We performed GSEA to identify 672 GO terms and 30 KEGG pathways associated with the high- and low-risk groups in the TCGA cohort ($P < 0.05$). As shown in Fig. 5A, B and Additional file 1: Table S1, the genes in the high-risk group were mainly enriched in leukocyte migration, cytokine interaction with cytokine receptors, oxidative phosphorylation, autoimmune diseases, and coagulation cascade. The genes in the low-risk group were mainly enriched in ATPase activity, chromatin remodeling, DNA replication,

methylation, and cell cycle (Fig. 5C, D and Additional file 2: Table S2). In summary, the enrichment analysis revealed potential pathways that could serve as targets in TNBC treatment.

Landscape of mutation profiles in low-and high-risk groups

Since the DDR is closely related to somatic mutations [18], we further explored the mutation status of the high- and low-risk groups. After analyzing the mutation data, we found missense mutations accounted for the most mutations in the high- and low-risk groups. The main variant type was single-nucleotide polymorphism (SNP), with the most common single nucleotide variants being C > T (Fig. 6A, B). Figure 6C, D show the top 20 most frequently mutated genes in the high- and low-risk groups with ranked percentages. The mutation frequency of the two groups was the same (94% vs. 93.75%), while the mutated genes were different. Additionally, the associations across the top 20 mutated genes are shown in Additional file 3: Figure S1A, B, where green represents co-expression and red represents mutually exclusive relationships.

Immune status between low- and high-risk groups

To further explore the relationship between the six-gene prognostic model and the immune system, ssGSEA was used to evaluate the expression profiles of 28 immune signature genomes in the high- and low-risk groups. The heat map in Fig. 7A shows that in the TCGA database, 28 types of immune cells in the low- and high-risk groups showed a certain degree of heterogeneity. The violin plot of the 28 immune cells showed that, as compared with the low-risk group, the content of memory B cells and T follicular helper cells increased, while the content of myeloid-derived suppressor cells (MDSCs) decreased (Fig. 7B). We also showed the correlation analysis of 6 genes and 28 immune cells (Additional file 4: Fig. 2A, B).

Construction of TF and ceRNA regulatory network

A gene-TF regulatory network was constructed using the Network Analyst database for the six-gene prognostic model we developed in this study. As shown in Fig. 8A, we were unable to search for TF related to *BRIP1* and *PARP1*. The constructed transcriptional regulatory network included 105 interaction pairs among 4 genes and 87 TFs, of which *RFC4* regulated most of the TF and occupied the dominant position. Finally, we predicted the target miRNAs of the six-gene prognostic model and lncRNAs that may be related to miRNAs via the miRanda database. A total of 271 lncRNA-miRNA-mRNA pairs were obtained, including 94 lncRNAs, 25 miRNAs, and 3 mRNAs (Fig. 8B). It is worth noting that three mRNAs in this ceRNA network including *BRIP1* and *PARP1*, were not found in the Network Analyst database.

Discussion

DNA is the source of genetic information, and maintaining its integrity is vital to sustaining life. Therefore, cells have evolved specialized DDR mechanisms to maintain the integrity of the genome [19, 20]. DDR plays an important role in maintaining homeostasis within the cell [21]. Cancer cells are characterized by genomic instability, which is conducive to the accumulation of driver mutations and the expansion of tumor heterogeneity [12, 22]. DDR mechanisms can repair mutated genes during the early stage of cancer

and hinder the development of tumors [23]. However, with the development of cancer, DDRs may cause tumor cells to develop resistance to cytotoxic drugs [24, 25]. The occurrence and development of cancer are often accompanied by the inactivation of one or more DDR pathways; however, cancer cells are therefore more dependent on the remaining DDR pathways than normal cells [25]. This phenomenon suggests that there are potential weaknesses in tumors that can be targeted by innovative therapies that follow the concept of synthetic lethality. In the concept of synthetic lethality, two pathway defects (independently non-toxic) become fatal when combined [26–29]. In this study, we screened out six prognostic genes from DDR genes, constructed a risk model based on the bio-analysis of DDR genes, and conducted immune-related analysis based on the model. Our research provides a theoretical basis and reference for the diagnosis and treatment of TNBC.

From the TNBC patient data in the TCGA database, we obtained 2,178 differential genes and intersected them with DDRs genes to obtain 23 DE-DDRs. We then constructed a risk model based on six prognostic genes (*PARP1*, *RAD51*, *EXO1*, *BRIP1*, *RMI2*, and *RFC4*) using univariate analysis and Lasso analysis and determined the effectiveness of the risk model by drawing an ROC curve and a K–M curve. In addition, independent prognostic analysis of the risk model and verification of the model based on the GEO dataset confirmed that the risk model constructed in this study can effectively predict the prognosis of TNBC. In order to investigate the reasons why the model could effectively predict the prognosis of TNBC patients, we conducted enrichment analysis, mutation analysis, and immunocorrelation analysis (immunoinfiltration and immunocorrelation score) on the high- and low-risk groups defined by the model. We found that there were certain differences in immunity between the high- and low-risk groups. Finally, we constructed a TF regulatory network and ceRNA network based on model gene prediction and demonstrated the regulatory function of these key genes.

Poly(ADP-ribose) Polymerase-1 (PARP1) is a member of the PARP family, which has 17 members total and plays a role in various biological functions, including synthetic lethality, DNA repair, apoptosis, necrosis, and histone binding. PARP1, a chromatin-bound nuclear enzyme that is activated by DNA damage, is a validated therapeutic target for cancers and other human diseases [30, 31]. PARP1 can inhibit the expression of PD-L1 on the surface of TNBC cells by interacting with the nucleic acid-binding domain of nucleophosmin (NPM1), thus playing a key role in the tumor-related immune escape of TNBC [32]. In our study, we found that the *PARP1* gene correlated positively with eosinophils.

RAD51 is a strand transferase that aggregates into nucleoprotein filaments on single strands of DNA and promotes the exchange of DNA strands with undamaged homologous chromatin [33]. RAD51 is a component of the cellular DDR, and as such, inhibition of RAD51 sensitizes cancer cells to DNA-damaging drugs [34, 35]. Studies have found that RAD51 can mediate breast cancer stem cells (CSCs) to develop resistance to PARP inhibitors in TNBC [36]. In our study, we found a negative correlation between the *RAD51* gene and immune dense cells.

BRIP1, which belongs to the Fanconi anemia (FA) gene family, was first identified via tandem mass spectrometry through its physical interaction with BRCA1 [37]. BRIP1 is essential to the stability of the

genome, and its normal active expression is necessary for the repair of DNA interstrand cross-links (ICL) [38]. Although pathogenic mutations and a large number of missense mutations in *BRIP1* have been discovered through genetic testing, the impact of these mutations on the molecular function and subsequent role of *BRIP1* in cancer risk is uncertain [39–41]. Studies have found that *BRIP1* can promote the invasion of breast cancer (BC) cells by regulating the expression of multiple downstream target genes, such as *MGAT5*, *EPCAM*, and *CXCL 12*, especially in the triple-negative phenotype MDA-MB-231 cell line [42]. In our study, we found a positive correlation between the *BRIP1* gene and monocytes.

Exonuclease 1 (EXO1) is associated with increased levels of genomic instability in the telomere region, and this widespread genomic instability can promote cancer progression [43]. EXO1 is a therapeutic target of TNBC that serves an important role in the DDR by inhibiting the activity of PARP [44, 45]. In our study, we found that the *EXO1* gene has a positive correlation with eosinophils.

RecQ-mediated genome instability protein 2 (RMI2) plays a vital role when the spindle assembly point (SAC) is activated during mitosis [46]. RMI2 is widely considered to play a crucial role in DNA damage repair. High expression of RMI2 was confirmed to be associated with the worse prognosis in pancreatic cancer [47] and lung cancer [48]. RMI2 was also reported to act as a tumor promoter by mediating MYCN/PARP DDR signaling pathway in neuroendocrine prostate cancer [49].

Human Replication Factor C (RFC) is a polyprotein composed of five distinct subunits that are highly conserved through evolution and plays an important role in DNA repair after DNA damage [50, 51]. Human replication factor C subunit 4 (RFC4) is a member of the RFC family that is often overexpressed in cancer, promoting tumor progression and resulting in worse survival outcomes by regulating tumor cell proliferation and cell cycle. RFC4 has been reported to be overexpressed in a variety of malignancies, including prostate cancer, cervical cancer, colorectal cancer, and head and neck squamous cell carcinoma [52–55]. It can promote tumor progression and lead to worse survival outcomes by regulating cell proliferation and cell cycle arrest in tumors [56]. In our study, we found that the *RFC4* gene had a positive correlation with type 2 T helper cells and a negative correlation with mast cells.

Our results demonstrate that further elucidating the functions of these six DDR-related genes in TNBC may improve our understanding of the biological basis of breast cancer and provide new therapeutic targets. The poor prognosis of TNBC seems to depend on the multi-layered interaction between DNA repair gene mutations, cell proliferation, and the immune response. By including prediction-related biological features, such as immune cells, our six-gene model displayed better predictive value than previously published immune features.

In this study, the correlation between our six gene markers related to the DDR and immune characteristics has been characterized to a certain extent. This model outperforms the prognostic performance of individual clinicopathological prognostic factors and published markers of disease-free survival (DFS) gene expression, further reinforcing the fact that the immune response is an important component of TNBC. Analyzing the function of the six genetic signatures not only helped us to understand the biological basis of the risk association, but also aided us in making treatment decisions. The main

limitations of this study are the retrospective nature of the study and the absence of functional validation of the genes included in the signature, which will be further verified in future experiments and prospective studies.

Conclusion

We screened and identified 6 DE-DDRs (*PARP1*, *RAD51*, *EXO1*, *BRIP1*, *RMI2*, and *RFC4*) as prognostic genes through comprehensive bioinformatics analysis and constructed a risk model that can effectively predict the prognosis of TNBC. In addition, we found that the high- and low risk TNBC groups, as defined by the model, exhibited differences in immune-related analysis (immune infiltration, immune-related scores). The above analysis provides a theoretical basis and reference for the research and treatment of TNBC.

Abbreviations

AUC

under the ROC curve; BC:breast cancer; BER:base excision repair; ceRNAs:competing endogenous RNAs; CPF:checkpoint factor; DDR:DNA damage response; DEGs:differentially expressed genes; DFS:disease-free survival; EXO1:Exonuclease 1; FA:Fanconi's anemia; GEO:Gene Expression Omnibus; GO:Gene Ontology; GSEA:gene set enrichment analysis; HER2:human epidermal growth factor receptor-2; HRR:homologous recombination repair; IARC:International Agency for Research on Cancer; KEGG:Kyoto Encyclopedia of Genes and Genomes; MDSCs:myeloid-derived suppressor cells; MFE:minimal folding free energy; MMR:Mismatch repair; MSigDB:Molecular Signatures Database; NER:nucleotide excision repair; NHEJ:non-homologous end ligation; NPM1:nucleic acid-binding domain of nucleophosmin; OS:overall survival; PAPRI:poly-ADP-ribose polymerase inhibitor; PR:progesterone; RFC:Replication Factor C; RMI2:RecQ-mediated genome instability protein 2; ROC:receiver operating characteristic; SAC:spindle assembly point; SNP:single-nucleotide polymorphism; ssGSEA:single-sample gene set enrichment analysis; TCGA:the Cancer Genome Atlas; TF:transcriptional factor; TLS:Translesion Synthesis; TMB:tumor mutation burden; TNBC:triple-negative breast cancer.

Declarations

Ethics approval and consent to participate

The studies involving human participants were reviewed and approved by the Ethics Committee of the First Affiliated Hospital of Zhengzhou University.

Consent for publication

All authors approved the submitted version.

Availability of data and materials

Publicly available datasets were analyzed in this study. This data can be found here: TCGA database (<https://portal.gdc.cancer.gov/cart>), MSigDB (<http://www.gsea-msigdb.org/gsea/msigdb/>), miRanda database (<http://www.microrna.org/microrna/home.do>) and GEO database (<https://www.ncbi.nlm.nih.gov/gds>).

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

Science and Technology Department of Henan Province; Award Number: LHGJ20200356; Recipient: Nan Wang, M.D

Authors' contributions

Nan Wang and Youyi Xiong performed the data analysis and wrote the manuscript. Jiangrui Chi, Xinwei Liu and Chaochao Zhong contributed to the data analysis and manuscript revision. Fang Wang and Xinxing Wang contributed to literature search and data extraction. Lin Li and Yuanting Gu proofread the manuscript. Nan Wang and Lin Li conceived and designed the study. All authors contributed to the article and approved the submitted version.

Acknowledgements

We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209–49.
2. Pashayan N, Antoniou AC, Ivanus U, Esserman LJ, Easton DF, French D, et al. Personalized early detection and prevention of breast cancer: ENVISION consensus statement. *Nat Rev Clin Oncol*. 2020;17(11):687–705.
3. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am J Hum Genet*. 2019;104(1):21–34.

4. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, et al. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet*. 2018;391(10125):1023–75.
5. Bianchini G, Balko JM, Mayer IA, Sanders ME, Gianni L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nat Rev Clin Oncol*. 2016;13(11):674–90.
6. Emens LA. Breast cancer immunotherapy: facts and hopes. *Clin Cancer Res*. 2018;24(3):511–20.
7. Lee JS, Yost SE, Yuan Y. Neoadjuvant treatment for triple negative breast cancer: recent progresses and challenges. *Cancers (Basel)*. 2020;12(6).
8. André F, Zielinski CC. Optimal strategies for the treatment of metastatic triple-negative breast cancer with currently approved agents. *Ann Oncol*. 2012;23(Suppl 6):vi46–51.
9. Roos WP, Thomas AD, Kaina B. DNA damage and the balance between survival and death in cancer biology. *Nat Rev Cancer*. 2016;16(1):20–33.
10. Scarbrough PM, Weber RP, Iversen ES, Brhane Y, Amos CI, Kraft P, et al. A cross-cancer genetic association analysis of the DNA repair and DNA damage signaling pathways for lung, ovary, prostate, breast, and colorectal cancer. *Cancer Epidemiol Biomarkers Prev*. 2016;25(1):193–200.
11. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*. 2014;505(7484):495–501.
12. Hu Y, Guo M. Synthetic lethality strategies: Beyond BRCA1/2 mutations in pancreatic cancer. *Cancer Sci*. 2020;111(9):3111–21.
13. Turner N, Tutt A, Ashworth A. Hallmarks of 'BRCAness' in sporadic cancers. *Nat Rev Cancer*. 2004;4(10):814–9.
14. Cerrato A, Morra F, Celetti A. Use of poly ADP-ribose polymerase [PARP] inhibitors in cancer cells bearing DDR defects: the rationale for their inclusion in the clinic. *J Exp Clin Cancer Res*. 2016;35(1):179.
15. Teo MY, Seier K, Ostrovnaya I, Regazzi AM, Kania BE, Moran MM, et al. Alterations in DNA damage response and repair genes as potential marker of clinical benefit from PD-1/PD-L1 blockade in advanced urothelial cancers. *J Clin Oncol*. 2018;36(17):1685–94.
16. Faraoni I, Graziani G. Role of BRCA mutations in cancer treatment with Poly(ADP-ribose) polymerase (PARP) inhibitors. *Cancers (Basel)*. 2018;10(12):487.
17. Tutt A, Tovey H, Cheang MCU, Kernaghan S, Kilburn L, Gazinska P, et al. Carboplatin in BRCA1/2-mutated and triple-negative breast cancer BRCAness subgroups: the TNT Trial. *Nat Med*. 2018;24(5):628–37.
18. Tian W, Shan B, Zhang Y, Ren Y, Liang S, Zhao J, et al. Association between DNA damage repair gene somatic mutations and immune-related gene expression in ovarian cancer. *Cancer Med*. 2020;9(6):2190–200.

19. Wengner AM, Scholz A, Haendler B. Targeting DNA damage response in prostate and breast cancer. *Int J Mol Sci.* 2020;21(21):8273.
20. Malaquin N, Carrier-Leclerc A, Dessureault M, Rodier F. DDR-mediated crosstalk between DNA-damaged cells and their microenvironment. *Front Genet.* 2015;6:94.
21. Goldstein M, Kastan MB. The DNA damage response: implications for tumor responses to radiation and chemotherapy. *Annu Rev Med.* 2015;66:129–43.
22. Lin J, Shi J, Guo H, Yang X, Jiang Y, Long J, et al. Alterations in DNA damage repair genes in primary liver cancer. *Clin Cancer Res.* 2019;25(15):4701–11.
23. Ali R, Rakha EA, Madhusudan S, Bryant HE. DNA damage repair in breast cancer and its therapeutic implications. *Pathology.* 2017;49(2):156–65.
24. Brandsma I, Fleuren EDG, Williamson CT, Lord CJ. Directing the use of DDR kinase inhibitors in cancer treatment. *Expert Opin Investig Drugs.* 2017;26(12):1341–55.
25. Carusillo A, Mussolino C. DNA Damage: From Threat to Treatment. *Cells.* 2020;9(7).
26. Blackford AN, Jackson SP. ATM, ATR, and DNA-PK: the trinity at the heart of the DNA damage response. *Mol Cell.* 2017;66(6):801–17.
27. Pilié PG, Tang C, Mills GB, Yap TA. State-of-the-art strategies for targeting the DNA damage response in cancer. *Nat Rev Clin Oncol.* 2019;16(2):81–104.
28. Yap TA, Plummer R, Azad NS, Helleday T. The DNA damaging revolution: PARP inhibitors and beyond. *Am Soc Clin Oncol Educ Book.* 2019;39:185–95.
29. Nickoloff JA, Jones D, Lee SH, Williamson EA, Hromas R. Drugging the cancers addicted to DNA repair. *J Natl Cancer Inst.* 2017;109(11).
30. Jain PG, Patel BD. Medicinal chemistry approaches of poly ADP-Ribose polymerase 1 (PARP1) inhibitors as anticancer agents—A recent update. *Eur J Med Chem.* 2019;165:198–215.
31. Cao C, Yang J, Chen Y, Zhou P, Wang Y, Du W, et al. Discovery of SK-575 as a highly potent and efficacious proteolysis-targeting chimera degrader of PARP1 for treating cancers. *J Med Chem.* 2020;63(19):11012–33.
32. Qin G, Wang X, Ye S, Li Y, Chen M, Wang S, et al. NPM1 upregulates the transcription of PD-L1 and suppresses T cell activity in triple-negative breast cancer. *Nat Commun.* 2020;11(1):1669.
33. San Filippo J, Sung P, Klein H. Mechanism of eukaryotic homologous recombination. *Annu Rev Biochem.* 2008;77:229–57.
34. Tsai MS, Kuo YH, Chiu YF, Su YC, Lin YW. Down-regulation of Rad51 expression overcomes drug resistance to gemcitabine in human non-small-cell lung cancer cells. *J Pharmacol Exp Ther.* 2010;335(3):830–40.
35. Quiros S, Roos WP, Kaina B. Rad51 and BRCA2—New molecular targets for sensitizing glioma cells to alkylating anticancer drugs. *PLoS One.* 2011;6(11):e27183.
36. Liu Y, Burness ML, Martin-Trevino R, Guy J, Bai S, Harouaka R, et al. RAD51 mediates resistance of cancer stem cells to PARP inhibition in triple-negative breast cancer. *Clin Cancer Res.*

2017;23(2):514–22.

37. Rutter JL, Smith AM, Dávila MR, Sigurdson AJ, Giusti RM, Pineda MA, et al. Mutational analysis of the BRCA1-interacting genes ZNF350/ZBRK1 and BRIP1/BACH1 among BRCA1 and BRCA2-negative probands from breast-ovarian cancer families and among early-onset breast cancer cases and reference individuals. *Hum Mutat.* 2003;22(2):121–8.
38. Moyer CL, Ivanovich J, Gillespie JL, Doberstein R, Radke MR, Richardson ME, et al. Rare BRIP1 missense alleles confer risk for ovarian and breast cancer. *Cancer Res.* 2020;80(4):857–67.
39. Weber-Lassalle N, Hauke J, Ramser J, Richters L, Groß E, Blümcke B, et al. BRIP1 loss-of-function mutations confer high risk for familial ovarian cancer, but not familial breast cancer. *Breast Cancer Res.* 2018;20(1):7.
40. Easton DF, Lesueur F, Decker B, Michailidou K, Li J, Allen J, et al. No evidence that protein truncating variants in BRIP1 are associated with breast cancer risk: implications for gene panel testing. *J Med Genet.* 2016;53(5):298–309.
41. Lu C, Xie M, Wendl MC, Wang J, McLellan MD, Leiserson MD, et al. Patterns and functional implications of rare germline variants across 12 cancer types. *Nat Commun.* 2015;6:10086.
42. Rizeq B, Sif S, Nasrallah GK, Ouhtit A. Novel role of BRCA1 interacting C-terminal helicase 1 (BRIP1) in breast tumour cell invasion. *J Cell Mol Med.* 2020;24(19):11477–88.
43. Maciejowski J, de Lange T. Telomeres in cancer: tumour suppression and genome instability. *Nat Rev Mol Cell Biol.* 2017;18(3):175–86.
44. Quist J, Mirza H, Cheang MCU, Telli ML, O'Shaughnessy JA, Lord CJ, et al. A Four-gene decision tree signature classification of triple-negative breast cancer: Implications for targeted therapeutics. *Mol Cancer Ther.* 2019;18(1):204–12.
45. Li Y, Wang Y, Zhang W, Wang X, Chen L, Wang S. BKM120 sensitizes BRCA-proficient triple negative breast cancer cells to olaparib through regulating FOXM1 and Exo1 expression. *Sci Rep.* 2021;11(1):4774.
46. Pradhan A, Singh TR, Ali AM, Wahengbam K, Meetei AR. Monopolar spindle 1 (MPS1) protein-dependent phosphorylation of RecQ-mediated genome instability protein 2 (RMI2) at serine 112 is essential for BLM-Topo III α -RMI1-RMI2 (BTR) protein complex function upon spindle assembly checkpoint (SAC) activation during mitosis. *J Biol Chem.* 2013;288(47):33500–8.
47. Xu MD, Liu SL, Zheng BB, Wu J, Wu MY, Zhang Y, et al. The radiotherapy-sensitization effect of cantharidin: Mechanisms involving cell cycle regulation, enhanced DNA damage, and inhibited DNA damage repair. *Pancreatology.* 2018;18(7):822–32.
48. Zhan W, Liu Y, Gao Y, Gong R, Wang W, Zhang R, et al. RMI2 plays crucial roles in growth and metastasis of lung cancer. *Signal Transduct Target Ther.* 2020;5(1):188.
49. Zhang W, Liu B, Wu W, Li L, Broom BM, Basourakos SP, et al. Targeting the MYCN-PARP-DNA damage response pathway in neuroendocrine prostate cancer. *Clin Cancer Res.* 2018;24(3):696–707.
50. Kim HS, Brill SJ. Rfc4 interacts with Rpa1 and is required for both DNA replication and DNA damage checkpoints in *Saccharomyces cerevisiae*. *Mol Cell Biol.* 2001;21(11):3725–37.

51. Krause SA, Loupart ML, Vass S, Schoenfelder S, Harrison S, Heck MM. Loss of cell cycle checkpoint control in *Drosophila* Rfc4 mutants. *Mol Cell Biol.* 2001;21(15):5156–68.

52. Erdogan E, Klee EW, Thompson EA, Fields AP. Meta-analysis of oncogenic protein kinase Ciota signaling in lung adenocarcinoma. *Clin Cancer Res.* 2009;15(5):1527–33.

53. Kang BY, You H, Bandyopadhyay S, Agrawal N, Melchert RB, Basnakian AG, et al. Cervical cancer isolate PT3, super-permissive for adeno-associated virus replication, over-expresses DNA polymerase delta, PCNA, RFC and RPA. *BMC Microbiol.* 2009;9:79.

54. Narayan G, Bourdon V, Chaganti S, Arias-Pulido H, Nandula SV, Rao PH, et al. Gene dosage alterations revealed by cDNA microarray analysis in cervical cancer: identification of candidate amplified and overexpressed genes. *Genes Chromosomes Cancer.* 2007;46(4):373–84.

55. Slebos RJ, Yi Y, Ely K, Carter J, Evjen A, Zhang X, et al. Gene expression differences associated with human papillomavirus status in head and neck squamous cell carcinoma. *Clin Cancer Res.* 2006;12(3 Pt 1):701–9.

56. Yao NY, O'Donnell M. The RFC clamp loader: structure and function. *Subcell Biochem.* 2012;62:259–79.

Figures

Figure 1

Bioinformatics analysis of the expression of DE-DDRs. (A) A Volcano plot illustrating the comparison of differential gene expression in tumors vs. normal tissue. The red dots indicate that the gene expression level is upregulated (tumor samples relative to normal samples), the blue dots indicate that the gene expression level is downregulated (tumor samples relative to normal samples), and the gray dots indicate that there is no significant difference between these genes. (B) A Venn diagram representation showing the intersection of DDRs and DEGs. (C) The results of the GO annotation of DE-DDRs

Figure 2

Establishment and evaluation of the prognostic model. (A) The forest plots illustrate univariate Cox analysis of the 6 genes significantly associated with OS. (B, C) LASSO coefficient profiles of 6 genes significantly associated with OS. (D) The Risk curve of the risk model. The risk value of patients increases from left to right. According to the median value, the samples were divided into high- and low-risk groups. (E) K–M survival curve of the Risk score. In the figure, the ordinate indicates the survival rate, and the abscissa indicates the total survival time. The red curve represents the high-risk group, and the blue curve represents the low-risk group. The difference between high- and low-risk groups was 0.0048, indicating a significant difference. (F) The ROC curve used to evaluate the effectiveness of the risk model. (G) The top of the heat map shows different clinical characteristics, in which the first line denotes the high-low risk grouping, orange represents the low-risk group samples, and green represents the high-risk group samples. The tree on the left shows the clustering analysis results of different genes from different samples

Figure 3

Validation of the six-gene prognostic model using the GEO database. (A) The Risk curve of the risk model. The risk value of patients increases from left to right. According to the median value, the samples were divided into high- and low-risk groups. (B) K–M survival curve of verification set–Risk score. The ordinate indicates survival rate, and the abscissa indicates total survival time. The red curve represents the high-risk group, and the blue curve represents the low-risk group. The difference between high- and low-risk groups was significant (0.0012). (C) Validation Set–ROC Curve to evaluate the effectiveness of the risk model

Figure 4

Independent prognostic value of the six-gene prognostic model. (A) Independent prognostic factors as determined by the Univariate Cox regression analysis. The left side represents the gene, corresponding P value, and HR value. The red square on the right indicates that the HR value is greater than 1, the green square indicates that the HR value is less than 1, and the line segments on both sides of the square are 95% confidence intervals of the HR value. (B) Independent prognostic factors as determined by the Multiple Cox regression analysis. (C) The nomogram to predict overall survival was created based on three independent prognostic factors. Each factor corresponds to a score, and the sum of the total scores of each factor corresponds to the total score. The 1-year, 3-year, and 5-year survival rate is predicted according to the total score. (D) The correction curve based on the above prediction model. The c-index of the model is 0.902371, and the corrected c-index is 0.887527

Figure 5

Gene set enrichment analysis (GSEA). (A) The top 10 enriched GO pathways in the high-risk group. (B) The top 10 enriched GO pathways in the low-risk group. (C) The top 10 enriched KEGG pathways in the high-risk group. (D) The top 10 enriched KEGG pathways in the low-risk group

Figure 6

Landscape of mutation profiles in the low- and high-risk groups. (A) Overview of mutation types in the high-risk group. (B) Overview of mutation types in the low-risk group. (C) Waterfall Plot of the top 20 genes with the most mutations in the high-risk group. (D) Waterfall Plot of the top 20 genes with the most mutations in the low-risk group

Figure 7

Analysis of immune infiltration patterns in breast cancer samples from TCGA dataset via an ssGSEA score-based method. “ns” represents that there is no significant difference in the infiltration of immune cells between the two samples; “*” represents $P < 0.05$. (A) Heatmap of ssGSEA scores (red = positive, blue = negative). (B) Boxplot of ssGSEA scores of the 28 representative gene sets

Figure 8

Regulatory network of the risk model genes. (A) Transcription factor regulation network diagram. Only four of the six model genes (RAD51, EXO1, RMI2, and RFC4) predicted the corresponding transcription factors. In the picture, red is the model gene and blue is TF. (B) Transcription factor regulation network diagram. In the picture, red is the model gene, green is miRNA, and blue is lncRNA. Three model genes, 25 miRNAs, and 94 lncRNAs were screened out

Supplementary Files

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

- [renameda6be7.xlsx](#)
- [renamedfb436.xlsx](#)
- [supplementfig1.png](#)
- [SUPPLEMENTFIG2.png](#)