



Research article

Development of machine learning-based malignant pericardial effusion-related model in breast cancer: Implications for clinical significance, tumor immune and drug-therapy

Wendi Zhan^{a,b,1}, Haihong Hu^{a,b,1}, Bo Hao^b, Hongxia Zhu^{a,b}, Ting Yan^d,
Jingdi Zhang^{a,b}, Siyu Wang^e, Saiyang Liu^f, Taolan Zhang^{b,c,*}

^a School of Pharmacy, Hengyang Medical College, University of South China, 28 Western Changsheng Road, Hengyang, Hunan, 421001, China

^b Department of Pharmacy, The First Affiliated Hospital, Hengyang Medical School, University of South China, Hengyang, Hunan, 421001, China

^c Phase I Clinical Trial Center, The First Affiliated Hospital, Hengyang Medical School, University of South China, Hengyang, Hunan, 421001, China

^d Department of Breast and Thyroid Surgery, The First Affiliated Hospital, Hengyang Medical School, University of South China, Hengyang, Hunan, 421001, China

^e Department of Medical Oncology, The First Affiliated Hospital, Hengyang Medical School, University of South China, Hengyang, Hunan, 421001, China

^f Shandong University of Traditional Chinese Medicine, Jinan, Shandong, 250355, China

ARTICLE INFO

Keywords:

Breast cancer
Malignant pericardial effusion
Subtypes
Immune infiltration
Prognostic model
Web-based tool

ABSTRACT

Background: Malignant pericardial effusion (MPE) is a common complication of advanced breast cancer (BRCA) and plays an important role in BRCA. This study aims to construct a prognostic model based on MPE-related genes for predicting the prognosis of breast cancer.

Methods: The BRCA samples are analyzed based on the expression of MPE-related genes by using an unsupervised cluster analysis method. This study processes the data by least absolute shrinkage and selection operator and multivariate Cox analysis, and uses machine learning algorithms to construct BRCA prognostic model and develop web tool.

Results: BRCA patients are classified into three clusters and a BRCA prognostic model is constructed containing 9 MPE-related genes. There are significant differences in signature pathways, immune infiltration, immunotherapy response and drug sensitivity testing between the high and low-risk groups. Of note, a web-based tool (<http://wys.helyly.top/cox.html>) is developed to predict overall survival as well as drug-therapy response of BRCA patients quickly and conveniently, which can provide a basis for clinicians to formulate individualized treatment plans.

Conclusion: The MPE-related prognostic model developed in this study can be used as an effective tool for predicting the prognosis of BRCA and provides new insights for the diagnosis and treatment of BRCA patients.

* Corresponding author. The First Affiliated Hospital, Department of Pharmacy, Hengyang Medical School, 69 Chuanshan Road, Hengyang, Hunan, 421000, China.

E-mail address: TaolZhan@usc.edu.cn (T. Zhang).

¹ Wendi Zhan and Haihong Hu are equally contributed to this work.

<https://doi.org/10.1016/j.heliyon.2024.e27507>

Received 25 September 2023; Received in revised form 30 January 2024; Accepted 29 February 2024

Available online 2 March 2024

2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

1. Introduction

Breast cancer (BRCA) is a major threat to women’s health worldwide, with an estimated 2 million new cases diagnosed each year [1]. Despite advances in diagnosis and treatment, BRCA remains a major public health challenge [2]. The biology of BRCA is complex with multiple factors that contribute to its development and progression, including genetic mutations, hormonal factors, lifestyle, and environmental factors, among others [3,4]. One of the major challenges in managing BRCA is predicting the course of the disease and making an accurate prognostic assessment [5]. Accurate prognostic diagnosis of BRCA is expected to enable efficient and personalized

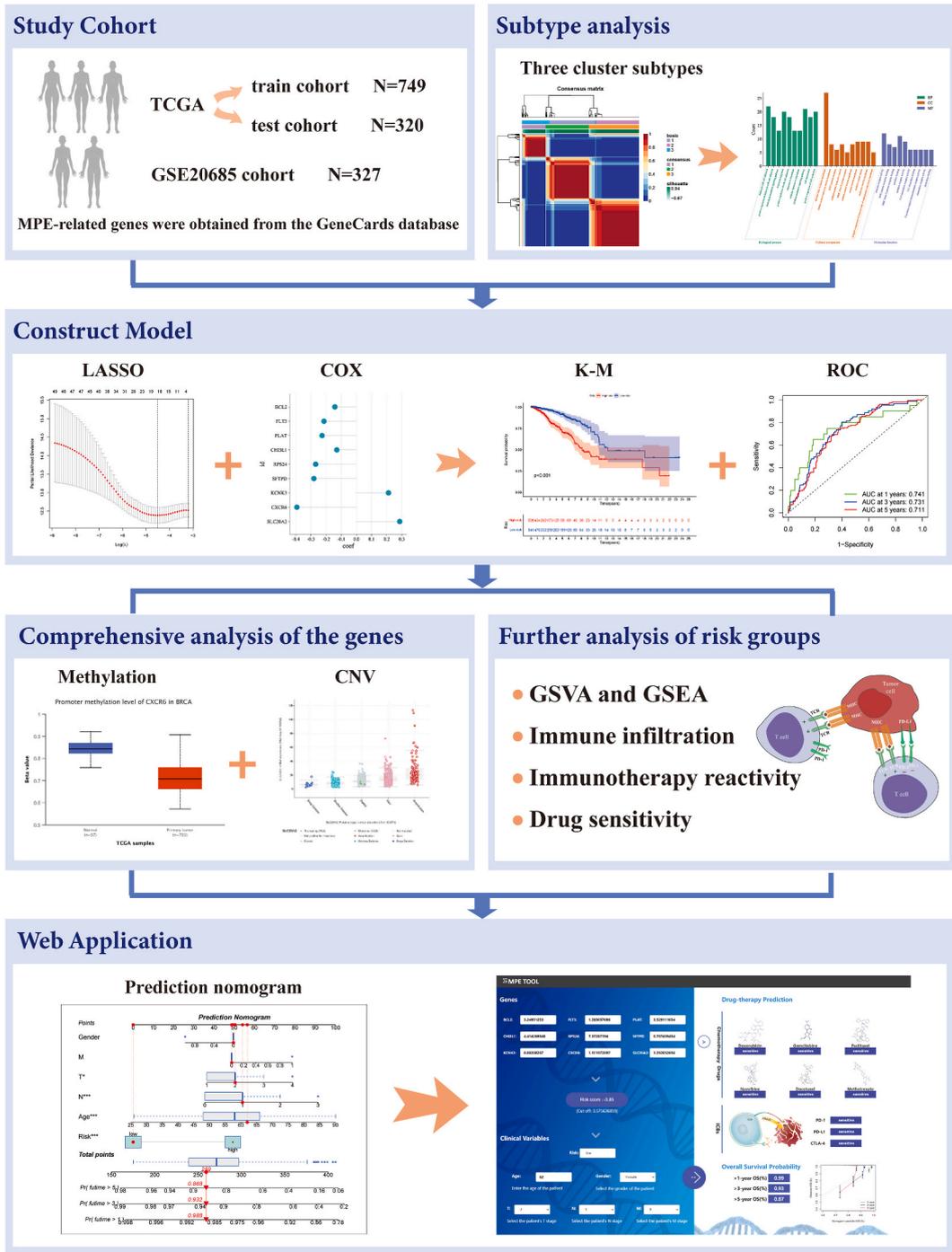


Fig. 1. Workflow of the study.

treatment to improve the survival rate of patients [6,7]. Many different clinical and biological factors are currently used to assess prognosis, including tumor size, lymph node status, histological grading, and hormone receptor status, but uncertainty remains [8–11]. The heterogeneity of BRCA and individual differences in treatment response and disease progression highlight the need for more accurate and personalized prognostic tools.

Malignant pericardial effusion (MPE) refers to excessive accumulation of fluid in the pericardial cavity caused by malignant tumors [12]. The accumulation of fluid can cause compression of the heart, leading to symptoms such as chest pain, shortness of breath, and cardiac tamponade [13–15]. MPE is usually a complication of advanced-stage cancer, with lung cancer, breast cancer, and lymphoma being the most common malignancies that metastasize to the pericardium [16,17]. The incidence of MPE in postmortem patients with malignant tumors is 10–15% and up to 21% [18]. The presence of MPE has been shown to be associated with poor prognosis, increased morbidity, and decreased survival in cancer patients [19]. The median survival time for BRCA patients with MPE ranges from a few weeks to six months, with most patients surviving less than one year [20,21]. Therefore, the identification of molecular markers and pathways associated with MPE may provide insights into BRCA biology and contribute to the diagnosis and treatment of BRCA. However, there is still no reliable MPE-related biomarkers to predict the prognosis of BRCA, which limits the application of MPE in the diagnosis and treatment of BRCA.

In this study, we classified BRCA patients into three clustered subtypes based on the expression of MPE-related genes and developed a risk model for BRCA prognosis prediction. The DNA methylation and copy number variation of the core genes of the model were analyzed, and potential therapeutic targets of BRCA were explored. Based on the prognostic model, we next performed analyses of immune infiltration, immunotherapy response, and drug sensitivity in risk subgroups. Finally, we developed a web-based tool combining risk score and other clinical variables to predict OS and drug-therapy response in BRCA patients. Our study provides new insights into the diagnosis and treatment of BRCA patients.

2. Methods

The workflow of this research was shown in Fig. 1.

2.1. Data collection and preprocessing

In this study, gene expression data and clinical data from BRCA patients were downloaded from GDC TCGA BRCA cohorts in UCSC Xena browser (<https://xenabrowser.net/>) [22]. A total of 1069 breast cancer patients with complete survival information were obtained after excluding normal samples and samples from the same patient. Subsequently, the transcriptome data with FPKM format were processed to normalize by `normalizeBetweenArrays` function of `limma` package [23]. Then, the 1069 BRCA patients were randomly divided into TCGA training cohort containing 749 patients and TCGA testing cohort including 320 patients with 7:3 ratio. Additionally, 327 breast cancer patients of GSE20685 which downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>) was utilized as an external validation cohort [24]. Detailed description of all cohorts can be found in Table 1. MPE-related genes were obtained from the GeneCards database (<https://www.genecards.org/>) and those with relevance score

Table 1
The clinical characteristics of BRCA in TCGA cohort and GSE20685.

Variables	TCGA train cohort (N=749)	TCGA test cohort (N=320)	GSE cohort (N=327)
Incomplete	N = 124	N = 52	0
Age			
>=65 years	167	80	22
<65 years	458	188	305
Sex			
Female	617	265	327
Male	8	3	0
M classification			
M0	617	259	244
M1	8	9	83
N classification			
N0	314	124	122
N1	205	94	102
N2	67	36	63
N3	39	14	40
T classification			
T1	173	62	101
T2	367	162	188
T3	62	36	30
T4	23	8	8
Stage classification			
Stage I	121	38	NA
Stage II	357	164	NA
Stage III	139	57	NA
Stage IV	8	9	NA

greater than 1 were screened for subsequent analysis.

2.2. Construction of breast cancer subtypes of malignant pericardial effusion by unsupervised clustering analysis

After screening prognostic genes associated with MPE using univariate Cox analysis, unsupervised cluster analysis was performed based on the expression of MPE -related genes using the non-negative matrix factorization (NMF) algorithm through “NMF” function to classify patients into different clustered subtypes [25]. Then, to verify the results of the classification, we used principal component analysis (PCA) to analyze the distribution differences of the cluster subtypes. Kaplan-Meier (KM) analysis was used to compare the overall survival (OS) of different clusters by survival package. To explore the effect of MPE on the BRCA mutation profile, the R package “maftools” was used to depict a waterfall plot of the mutation landscape in different clusters [26,27].

2.3. Differential analysis of distinct clusters

Since cluster 2 had the worst OS, we used the “limma” R package to screen cluster 2 against cluster 1 and cluster 3 for differentially expressed genes (DEGs) and plotted the differential gene heatmap [28]. The DEGs between Cluster 1 and Cluster 2, and the DEGs between Cluster 3 and Cluster 2 were merged as MPE -related differentially expressed genes. Through “ggplot2”, “clusterProfiler”, “DOSE” and “enrichplot” R package [29], Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway were performed for the differential genes to explore the biological functions and pathways of related genes.

2.4. Construction and validation of prognostic model

Prognostic features were initially constructed using the training set and subsequently validated in the validation set. To meticulously identify genes associated with BRCA prognosis, a comprehensive methodology was employed. In the univariate Cox regression analysis, survival outcomes of BRCA patients were individually assessed against the expression levels of each gene. We explicitly calculated hazard ratios (HR) and corresponding p-values for each gene, with a stringent criterion for statistical significance ($p < 0.05$). After univariate Cox analysis, the least absolute shrinkage and selection operator (LASSO) Cox proportional risk regression was implemented to further screen for characteristic genes related to breast cancer prognosis using glmnet package [30]. This method applied a penalty term to the regression coefficients to prevent overfitting. Rigorous parameter settings, including optimization for the regularization parameter, were employed to identify a subset of genes with the most significant impact on BRCA prognosis. Subsequently, a multivariate Cox regression analysis was meticulously applied to identify the optimal set of genes for constructing robust risk models predicting the prognosis of BRCA patients. This involved assessing the combined effect of the selected genes from Lasso regression on survival outcomes. The coefficients derived from the multivariate Cox model were used to assign weights to each gene, capturing their respective contributions to the overall risk score. In this model, the prognostic risk score was calculated using the following formula: $\text{risk score} = \sum_{i=1}^n (\text{expression} * \text{coef})$. The Expression represents the expression level of each selected gene, and Coef represents its corresponding coefficient from the multivariate Cox model. The summation across all selected genes yields an individualized risk score for each patient, indicating their relative risk of adverse outcomes associated with BRCA. To assess the predictive performance of the model, BRCA patients were divided into high-risk and low-risk groups based on the median risk score. The K-M survival curves were plotted using the R packages “Survival” and “Survminer” to compare the survival of the low and high-risk groups, and a p-value of < 0.05 was considered statistically significant [31]. In addition, receiver operating characteristic (ROC) curves over time were constructed using The R package “survival ROC” [32]. The performance of the risk model was evaluated according to the area under the area under curve (AUC) of the ROC curve.

2.5. Comparison of prediction accuracy with existing breast cancer prognostic models

We validated our proposed breast cancer prognostic model through a comparative analysis with three established models from the literature. The first model which focused on hypoxia- and lactate metabolism-related signatures by Li et al., featuring ESRP1, MAFF, SLC2A1, DARS2, and TH [33]. Wang et al. presented the second model associated with ferroptosis, including ALOX15, CISD1, CS, GCLC, GPX4, SLC7A11, EMC2, G6PD, and ACSF2 [34]. The third model was an immune-related prognostic model conducted by Yao et al., encompassing SOCS3, TCF7L2, TSLP, NPR3, ANO6, and HMGB3 [35]. To ensure consistency with the literature and reduce data dimensionality, we extracted gene expression levels for each model. Subsequently, multivariate Cox regression analysis yielded regression coefficients for each gene. After calculating risk scores for individual samples in TCGA cohort, we evaluated the predictive power and clinical utility using metrics such as the receiver operating characteristic (ROC) curves, concordance index (C-index) and decision curve analysis (DCA) using the timeROC and ggDCA packages.

2.6. Analysis of DNA promoter methylation and copy number variation

The analysis of DNA promoter methylation was conducted by the University of Alabama at Birmingham CANcer data analysis Portal (<http://ualcan.path.uab.edu/>) (UALCAN) [36]. UALCAN was an interactive portal for in-depth analysis of TCGA gene expression data, which contained a large number of cancer-related data including promoter methylation data of patients in TCGA. These data could be used directly to analyze differences in methylation status of selected genes in tumor samples and normal samples.

After selecting TCGA breast cancer patients as study subjects in the database, gene names could be entered to obtain visualization results of methylation degree, and p-values of statistical analysis between the two groups. Copy number variation (CNV) is a significant type of genetic variation in humans, and it is closely associated with tumor initiation and progression [37]. The analysis of Copy number variation (CNV) for breast cancer patients in the Cancer Genome Atlas (TCGA) was conducted by cBioPortal (<https://www.cbioportal.org>). Patient data were obtained from TCGA through cBioPortal, and specific genes of interest were selected for CNV analysis. The cBioPortal interface facilitated the visualization of copy number alterations (CNAs) with tools such as oncoprints and segmented plots. Amplifications and deletions were identified based on segmented copy number data, with thresholds for defining alterations. By utilizing this tool, we examined the relationship between CNV status of the model core genes and their corresponding gene expression levels [38].

2.7. Gene set variation analysis and gene set enrichment analysis

In order to identify potential marker pathways, we used the “limma”, “GSEABase” and “GSVA” package to conduct gene set variation analysis (GSVA) for high-risk and low-risk patients in the TCGA all dataset [39]. The R package “limma”, “DOSE”, “clusterProfiler” and “enrichplot” was used to perform gene set enrichment analysis (GSEA) of marker gene sets and visualized them using the R package “enrichplot” [40,41].

2.8. Immune landscape of high- and low-risk group

The ESTIMATE Score, Immune Score, Stromal Score and Tumor purity of BRCA samples were calculated according to the ESTIMATE algorithm of the ESTIMATE package [42]. The ESTIMATE Score and Tumor Purity provides an estimation of the overall tumor purity, while the immune score and stromal score represent the infiltration levels of immune cells and stromal cells, respectively. We employed the CIBERSORT algorithm for transcriptome data analysis, enabling the extraction of expression levels for 22 distinct immune cell types within each sample. In addition to the differences in immune score and immune infiltration among various risk groups, the correlation between immune cells and risk score was also analyzed using the R package “limma”, “reshape2”, “ggpubr” and “ggExtra” [43].

2.9. Prediction of immunotherapy response

To assess the responsiveness of patients in the high- and low-risk group to immunotherapy, we investigated the differential expression of immune checkpoints between the two groups. These genes play a crucial role in regulating immune responses. Furthermore, an immunophenoscore (IPS) was obtained from the Cancer Immunome Atlas (TCIA) (<https://tcia.at/>) [44], with higher scores indicating higher response to immune checkpoint blockade (ICB). Based on the expression status of CTLA-4 and PD-1, we analyzed the IPS in different risk groups and visualized the Differential distribution with violin plots.

2.10. Drug sensitivity

The therapeutic value of the prognostic model was further determined by predicting the chemotherapy response of patients with “PRRophetic” package [45]. We calculated the half maximal inhibitory concentration (IC50) which often used to assess sensitivity to drug therapy of six chemotherapeutic drugs including methotrexate, doxorubicin, gemcitabine, gefitinib, paclitaxel, and vinorelbine commonly used in the clinical treatment of BRCA. The box plot was used to visualize the differential IC50 between the high-risk and low-risk groups.

2.11. The establishment of a prediction nomogram

The independent prognostic effects of clinical variables (age, gender and TNM stage) and risk score were explored using univariate and multivariate Cox regression analyses and visualized by drawing forest plots. Subsequently, a nomogram was constructed to predict the probability of 1-year, 3-year and 5-year OS of BRCA by combining clinical variables and risk score through the R package “rms”, and a calibration curve was drawn to assess the predictive capability of the nomogram [46]. Next, predictive nomogram performance and clinical utility were evaluated using the 1-year, 3-year and 5-year area under the ROC curve (AUC) and decision curve analysis (DCA) plots, implemented by the R packages “timeROC” and “ggDCA”, respectively [47].

2.12. Statistical analysis

Analysis in this study employed R software (versions 4.0.3 and 4.1.3) along with pertinent R packages obtained from Bioconductor and CRAN. Distinctions between two groups were evaluated using the Wilcoxon test, while comparisons involving more than two groups were conducted using the Kruskal–Wallis test. Pearson’s test was employed for correlation analyses. Median values served as the basis for all truncation values associated with grouping, and statistical significance was established at a P value < 0.05.

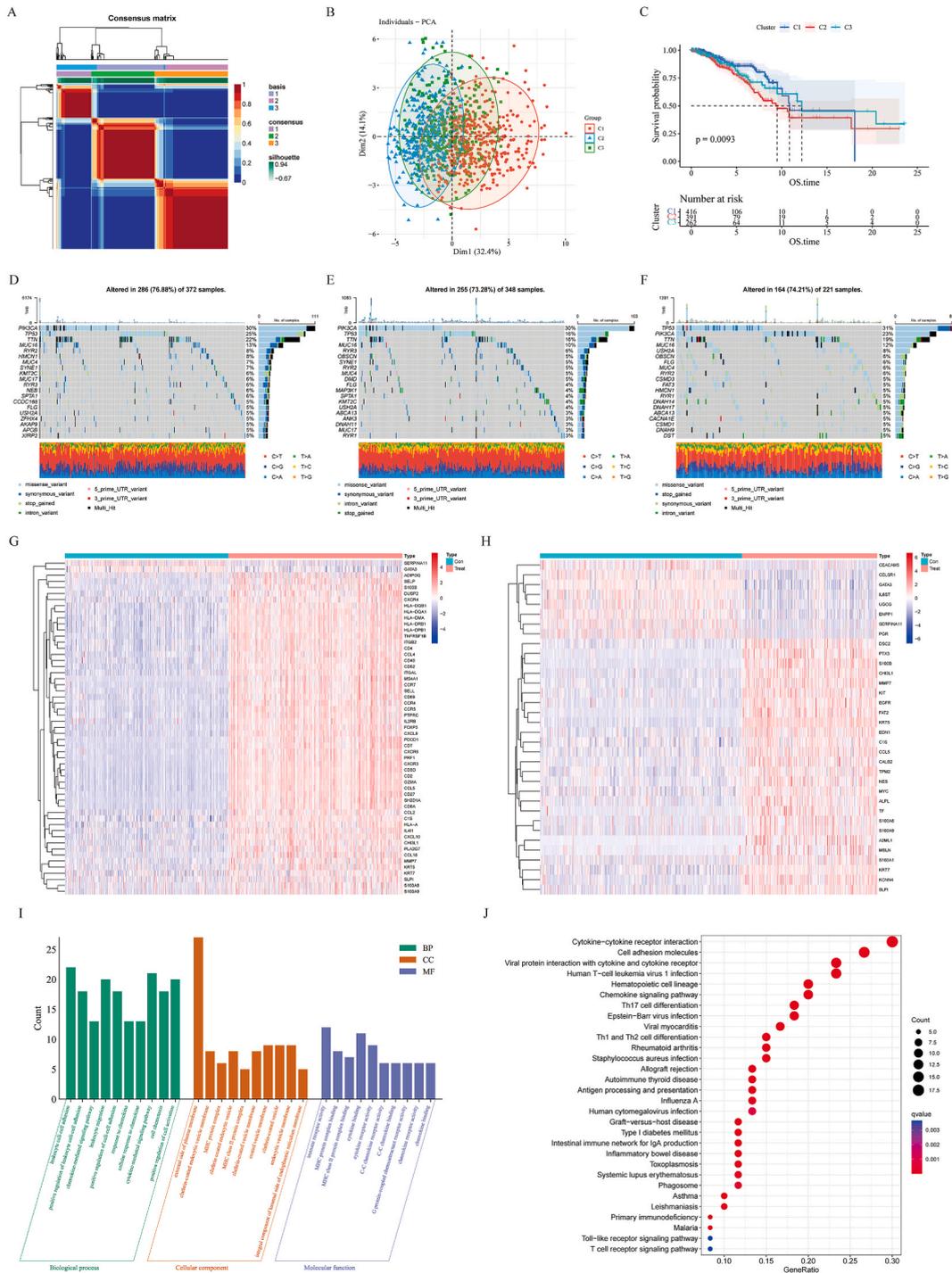


Fig. 2. Subtypes of BRCA and their characteristics. (A) Based on the expression of MPE, the BRCA subtypes was constructed by NMF algorithm. (B) Data were visualized using PCA. (C) K-M survival analysis of OS for the three subtypes. (D–F) The mutation waterfall plots of three subtypes. (G) Differential expression profiling between cluster1 and cluster2. (H) Differential expression profiling between cluster3 and cluster2. (I) GO enrichment analysis of DEGs. (J) KEGG pathway analysis of DEGs.

3. Results

3.1. BRCA cluster subtypes based on MPE

Based on the expression of MPE, we used the TCGA-BRCA dataset to perform unsupervised cluster analysis to classify BRCA patients into three different cluster subtypes (Fig. 2A). The PCA showed that the three clusters differed significantly (Fig. 2B) and that BRCA patients in cluster 2 had worse survival than those in cluster 1 and 3 (Fig. 2C). We then investigated the gene mutation distribution of the three different subtypes to explore the effect of MPE on the mutation profile of BRCA. The mutation waterfall plots showed that PIK3CA, TP53, TTN and MUC16 were the most common mutant genes in the three clusters (>10% mutation rate), and the mutation rates of TP53 and USH2A were more significant in cluster 3 (Fig. 2D-F). The results of subtype difference analysis showed that there were 55 DEGs in cluster 1 and cluster 2, 34 DEGs in cluster 3 and cluster 2, and 74 total DEGs (Fig. 2G and H). To further investigate the

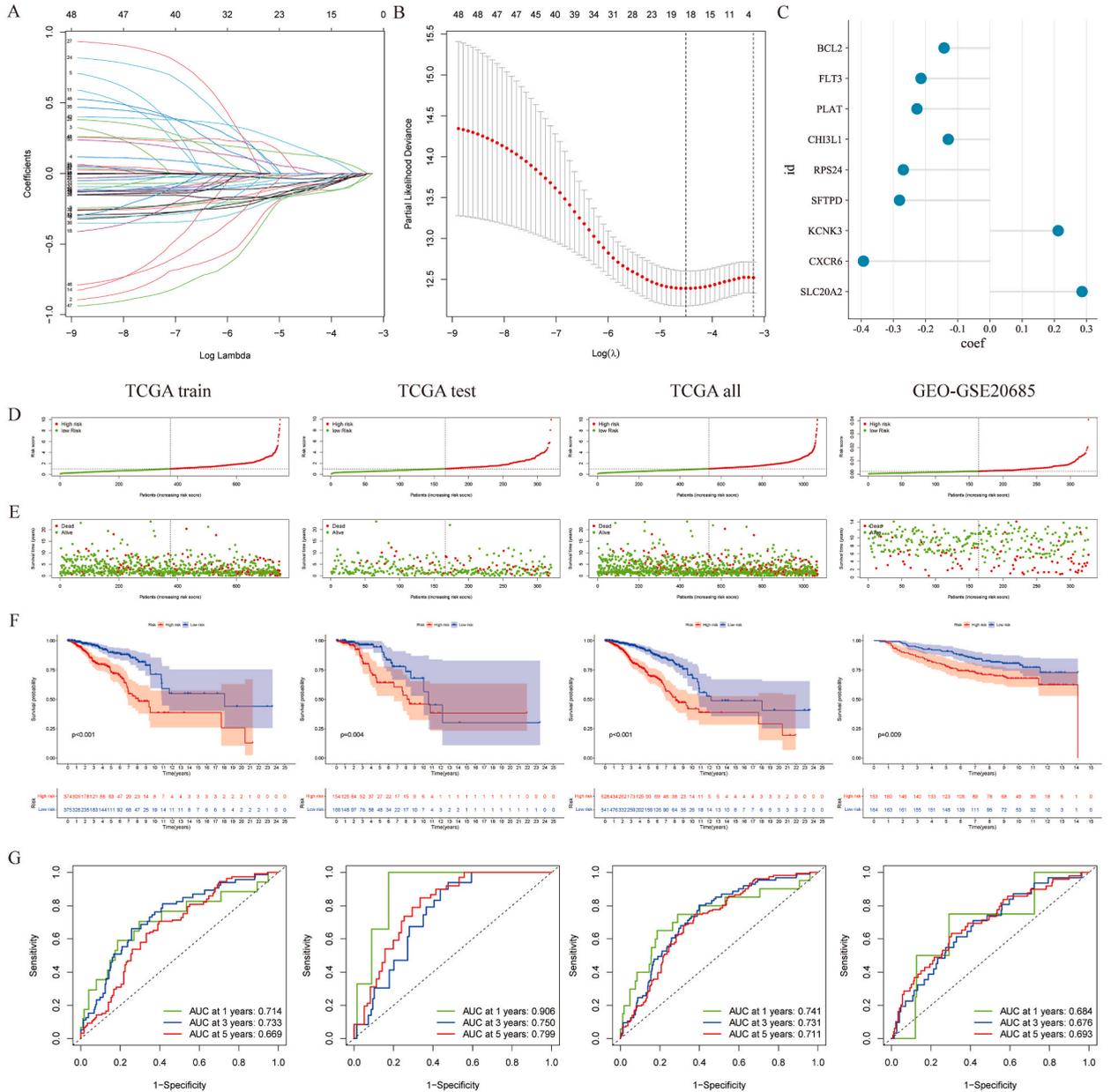


Fig. 3. Construction and validation of prognostic models. (A, B) Lasso regression further screened the genes of the prognostic model. (C) Genes screened by multivariate COX regression analysis that included in the prognostic model. (D) The risk score distribution diagram of four cohorts. (E) The survival status diagram of four cohorts. (F) K-M survival analysis. (G) Time-related ROC curve analysis.

biological function of 74 DEGs, GO enrichment analysis and KEGG pathway analysis were performed (Fig. 2I and J). In terms of biological processes, DEGs are mainly involved in leukocyte cell-cell adhesion and cell chemotaxis. In terms of cellular composition, DEGs are mainly enriched in the external side of plasma membrane. In terms of molecular functions, DEGs are mainly enriched in immune receptor activity and cytokine binding. In terms of action pathways, these genes are mainly involved in the Cytokine-cytokine receptor interaction pathway.

3.2. Construction and validation of prognostic models

Univariate Cox regression analysis was used to identify 50 MPE-related genes associated with the prognosis of BRCA. Based on these genes, further screening was performed using Lasso regression and multivariate Cox regression analysis (Fig. 3A–C). An MPE-related prognostic model containing these 9 related genes (BCL2, FLT3, PLAT, CHI3L1, RPS24, SFTPD, KCNK3, CXCR6, SLC20A2) was finally constructed (Fig. 3C). The calculation formula was as follows: Risk Score = $BCL2 * (-0.142438368) + FLT3 * (-0.214186716) + PLAT * (-0.226695372) + CHI3L1 * (-0.129528376) + RPS24 * (-0.269126914) + SFTPD * (-0.280900134) + KCNK3 * (0.211854921) + CXCR6 * (-0.39308609) + SLC20A2 * (0.286056043)$. Based on the formula, we calculated the risk score for each patient and consequently divided them into low-risk groups and high-risk groups. As shown in the risk score distribution diagram (Fig. 3D) and the survival status diagram (Fig. 3E), in all four cohorts, patients with high-risk index had less time to survival and higher mortality than those with low-risk index. The K-M analysis was performed to determine the value of risk scores in predicting patient prognosis, which resulted in significantly reduction in OS time in the high-risk group compared to the low-risk group (Fig. 3F). Time-related ROC curve analysis (Fig. 3G) showed AUCs of 0.714, 0.733, and 0.669 at 1, 3, and 5 years in the TCGA test cohort, 0.906, 0.750, and 0.799 in the TCGA train cohort, and 0.741, 0.731, and 0.711 in the TCGA all cohort, respectively. The AUCs in the GSE20685 cohort were 0.684, 0.676, and 0.693. It indicates that the prediction performance of this model is good. To underscore the merits of the MPE-related prognostic model developed in this study, we conducted a comparative analysis with three other breast cancer signatures. Fitting parameters for genes in the various models were obtained in the TCGA cohort, enabling the calculation of corresponding risk

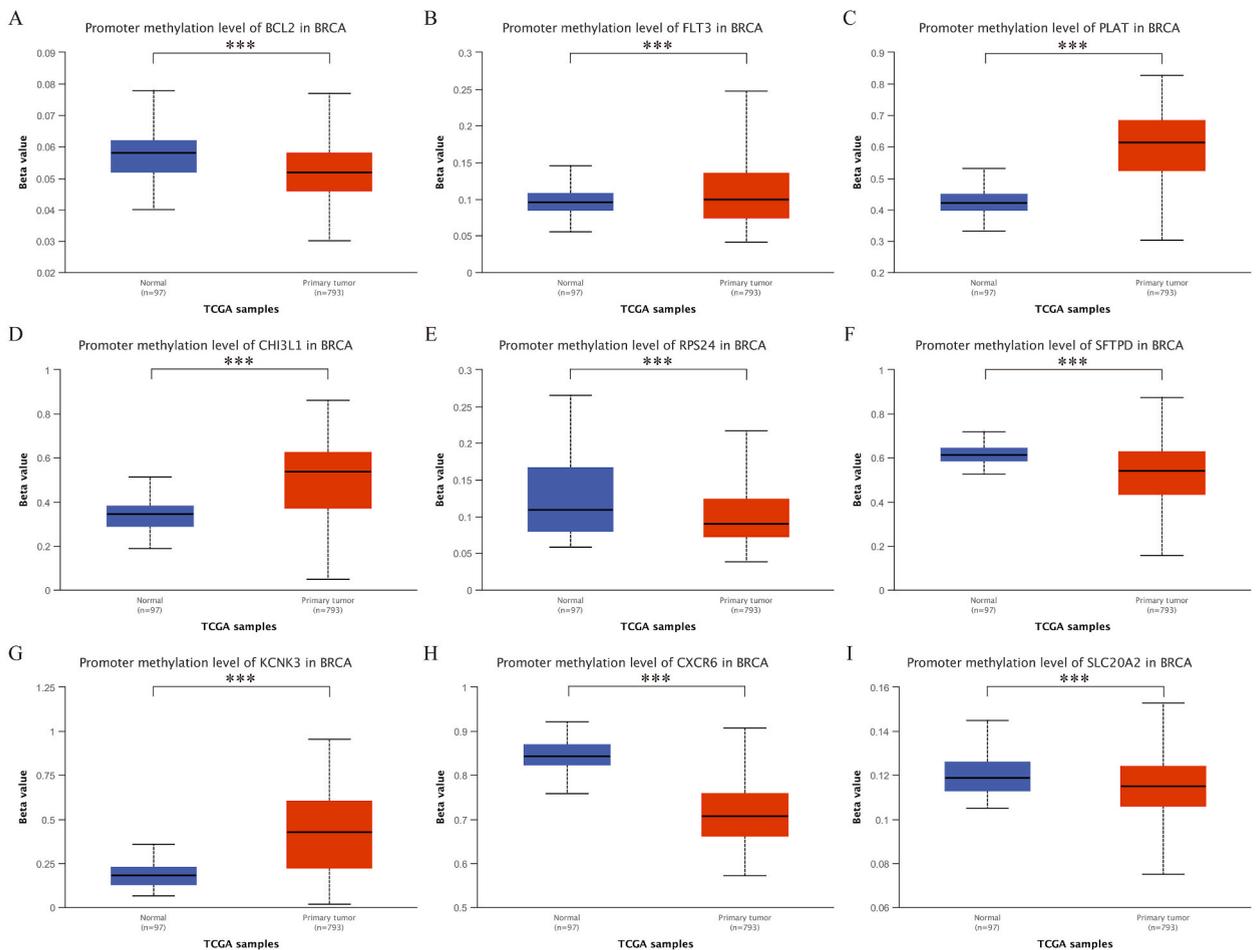


Fig. 4. Promoter methylation of core genes. Boxplots visualized the methylation levels of BCL2 (A), FLT3 (B), PLAT (C), CHI3L1 (D), RPS24 (E), SFTPD (F), KCNK3 (G), CXCR6 (H) and SLC20A2 (I) in normal and BRCA tissues.

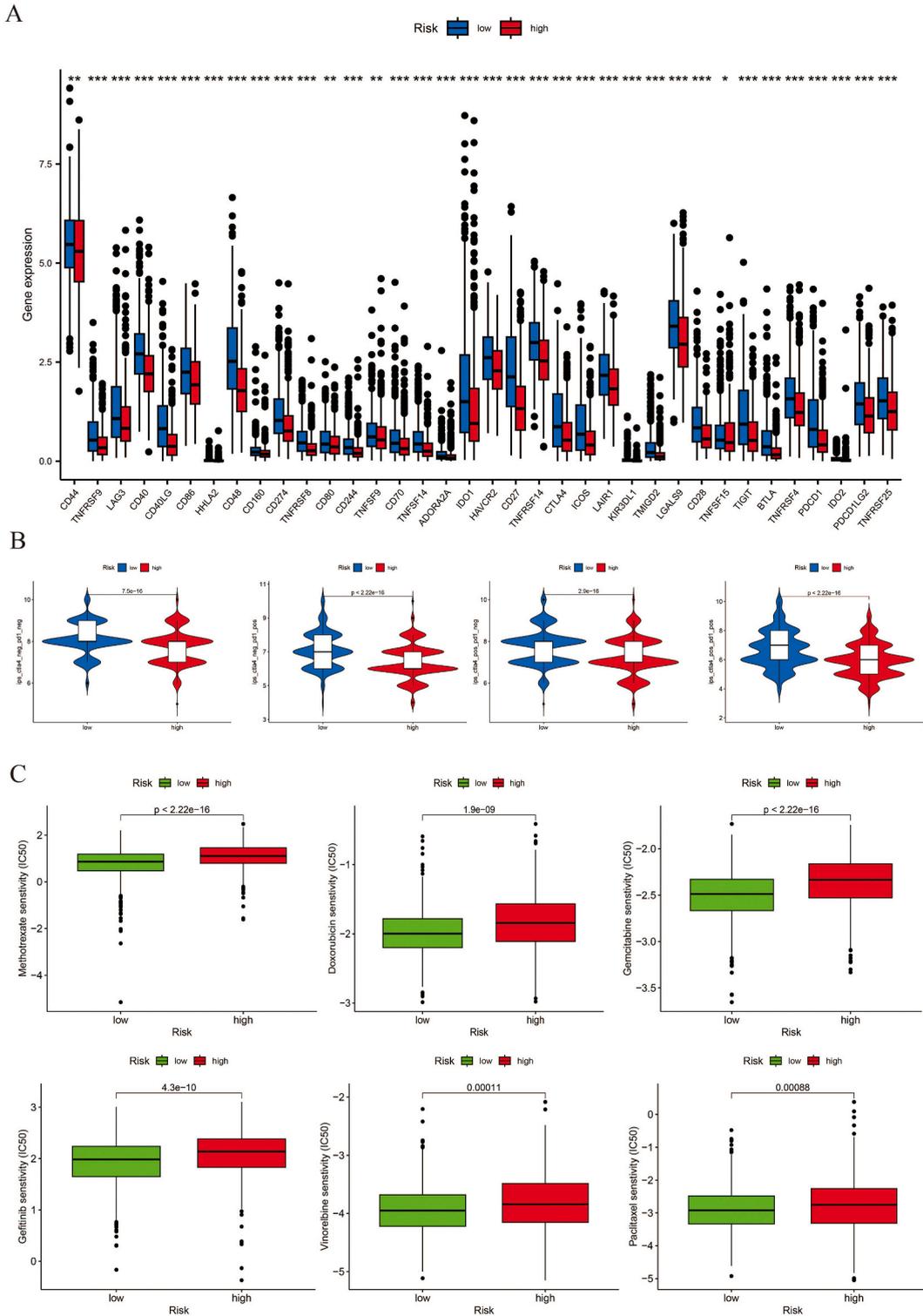


Fig. 6. Drug-therapy Prediction. (A) Immune checkpoint gene expression. (B) Immunophenotyping scores assess the potential clinical efficacy of immune checkpoint inhibitors in different risk groups. (C) Test of susceptibility to antineoplastic drugs of Methotrexate, Doxorubicin, Gemcitabine, Gefitinib, Vinorelbine and Paclitaxel.

scores for each patient. The patients were then stratified into high and low-risk groups using the established grouping method. Survival curves consistently revealed higher survival rates in the low-risk groups across all models (Supplementary Figs. 1A–D). Additionally, we assessed the prediction accuracy of these models. With the exception of the Yao et al. signature (AUC = 0.548, 0.609, 0.626), the

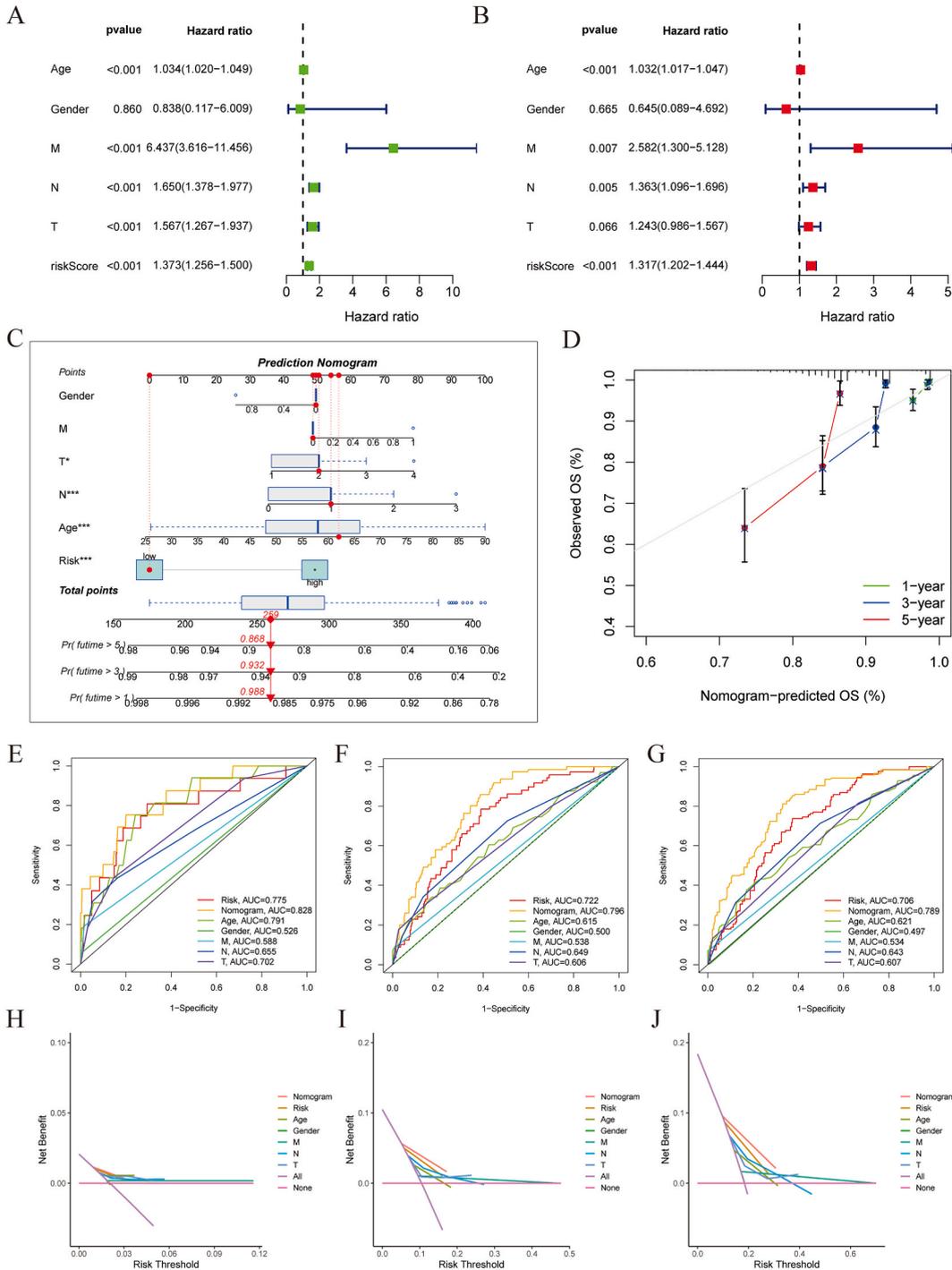


Fig. 7. Nomogram for predicting OS. (A) Univariate Cox regression analysis was used to analyze the predictive ability of age, TNM stage and risks core. (B) Multivariate Cox regression analysis was used to analyze the predictive ability of age, TNM stage and risks core. (C) The predictive nomogram combining clinicopathological features and risk score. When the total point was 259, the 1-year, 3-year and 5-year survival rates were 0.988, 0.932 and 0.868, respectively. (D) Calibration curve of the predictive nomogram. (E–G) ROC curves for 1, 3 and 5 years of the nomogram. (H–J) DCA evaluated the performance of the nomogram.

remaining three signatures demonstrated notable potential in predicting breast cancer survival over 1-, 3-, and 5-year intervals, as evidenced by the area under the receiver operating characteristics curve. Notably, the MPE-related prognostic model developed in this study exhibited superior accuracy, boasting AUCs of 0.741, 0.731, and 0.711 at 1-year, 3-year, and 5-year intervals (Supplementary Figs. 1E–H). Further validation through C-index, RMS, and DCA analyses consistently affirmed the heightened accuracy of MPE-related prognostic model in predicting breast cancer survival (Supplementary Figs. 1I–K).

3.3. Potential therapeutic target for BRCA

Promoter DNA methylation affects transcriptional inhibition and tumorigenesis [48], so we explored the methylation values of core genes in the model in normal tissues and BRCA tissues. The results showed that the promoter methylation levels of BCL2, RPS24, SFTPD, CXCR6 and SLC20A2 were significantly decreased in tumor tissues, while the promoter methylation levels of FLT3, PLAT, CHI3L1 and KCNK3 were significantly increased in tumor tissues (Fig. 4). DNA copy number is also a major determinant of gene expression, so we performed the correlation analysis between CNV and mRNA expression of core genes in the model. PLAT, CHI3L1, and SLC20A2 had a large proportion of CNV and were concentrated in the occurrence of amplification (Supplementary Fig. 2A). The mRNA expression of BCL2, PLAT, CHI3L1, RPS24 and SLC20A2 were significantly affected by the CNV, and the deletion or amplification might be responsible for affecting the expression of these genes, while the expression of FLT3, SFTPD, KCNK3 and CXCR6 had little correlation with the CNV, which might be due to other factors (Supplementary Fig. 2B). These can provide a reference for the search of potential therapeutic targets for BRCA.

3.4. Immune characteristics in high- and low-risk groups

To explore biological behavioral differences between different risk groups in the model, we performed GSVA and GSEA using the TCGA all data set. GSVA results showed that 11 signature pathways, including pentose phosphate pathway, terpenoid backbone biosynthesis and DNA replication, were significantly enriched in high-risk group compared with low-risk group (Supplementary Fig. 3A). GSEA confirmed that citrate cycle TCA cycle, DNA replication, homologous recombination, mismatch repair, and oocyte meiosis pathways were up-regulated in high-risk groups, while chemokine signaling pathway, cytokine-cytokine receptor interaction, hematopoietic cell lineage, primary immunodeficiency and T cell receptor signaling pathway were down-regulated in the low-risk group (Supplementary Fig. 3B). The results of the ESTIMATE algorithm indicated that BRCA patients in the low-risk group had considerably higher Immune, Stromal and ESTIMATE scores and significantly lower Tumor purity compared to the high-risk group (Fig. 5A and B). Furthermore, we also investigated the relationship between risk groups and immune infiltration. The infiltration of immune cells including naive B cells, Plasma cells, CD8 T cells, CD4 memory activated T cells, gamma delta T cells, resting NK cells, resting Dendritic cells was significantly higher in the low-risk group. However, the infiltrating level of M0 Macrophages and M2 Macrophages is completely opposite (Fig. 5C). In addition to the significant difference between high and low risk groups, the levels of immune cell infiltration were also correlated with risk scores (Fig. 5D).

3.5. The model may be a potential predictor of drug-therapy sensitivity in BRCA patients

Immune checkpoint gene expression levels are strongly correlated with therapeutic response to immune checkpoint inhibitors. To assess the potential of prognostic models to predict immunotherapy response in BRCA patients, we analyzed the differential expression of immune checkpoint genes between the high-risk and low-risk groups. As shown in Fig. 6A, the expression of the immune checkpoint gene was significantly higher in BRCA patients in the low-risk group than in those in the high-risk group. We then evaluated the potential clinical efficacy of the immune checkpoint inhibitor in different risk groups using IPS, which estimate immunogenicity to predict a patient's response to an immune checkpoint inhibitor. The violin plot showed that the IPS score was higher in the low-risk group than in the high-risk (Fig. 6B). These results suggest to us that patient in the low-risk group responded better to immunotherapy than those in the high-risk group. We also examined the relationship between the prediction model and general chemotherapy response. The IC50 of these conventional chemotherapy drugs in the low-risk group was significantly lower than that in the high-risk group, suggesting that patients in the low-risk group may be more responsive to receiving these chemotherapy drugs (Fig. 6C).

3.6. Establishment and validation of predictive nomogram for individualized evaluation

To explore the independent predictive ability of clinicopathological variables and risk score, univariate and multivariate Cox regression analyses were performed for age, gender, TNM stage, and risk score. The results showed that age, TNM stage and risks core all had good independent prediction ability, while gender showed poor performance due to the prevalence of BRCA in female (Fig. 7A and B). Next, we developed a nomogram combining clinicopathological features and risk score to predict survival in patients with BRCA. As shown in Fig. 7C, when the total point was 259, the 1-year, 3-year and 5-year survival rates were 0.988, 0.932 and 0.868, respectively. The calibration curves for 1-year, 3-year, and 5-year OS prediction showed that the predictive nomogram performed well (Fig. 7D). ROC curve analysis showed that the 1-year, 3-year, and 5-year AUC of the nomogram were 0.828, 0.796 and 0.789, respectively, which were significantly higher than the other parameters (Fig. 7E–G). In addition, DCA was added to evaluate the predictive nomogram, which had the highest net benefit and a wider range of threshold probabilities compared with other parameters such as risks core or TNM stage alone (Fig. 7H–J). These results suggested that the nomogram exhibited good predictive performance and is more suitable for predicting the prognosis of patients with BRCA in clinical practice.

3.7. Webpage deployment tool

Finally, we implemented the prognostic model into a Web application with custom algorithms to help predict OS and drug-therapy response in BRCA patients with personalized characteristics (wys.helyly.top/cox.html) (Fig. 8). First, the expression of MPE-related genes were entered into the application to provide a risk prediction for BRCA patient, which was then combined with the clinical variables of the patient's age, gender and TNM stage to predict the patient's drug-therapy response and OS of 1, 3 and 5 years, that presented on the web page. An example is a low-risk 62-year-old female with stage T2, N1, M0. This patient is expected to be sensitive to chemotherapy drugs of Docetaxel, Doxorubicin, Gemcitabine, Methotrexate, Paclitaxel, Vinorelbine and immunotherapy with PD-1, PD-L1, and CTLA-4. And the OS of 1 -, 3 - and 5-year of this patient are 99%, 93% and 87%, respectively. The tool is user-friendly and convenient to operate, which is promises to provide help for individualized prediction of prognosis in BRCA patients.

4. Discussion

The morbidity and mortality of BRCA continues to show an upward trend, posing a serious threat to the health of women worldwide [49]. An accurate prognosis is essential to make informed treatment decisions and to determine the appropriate management strategy for each patient [50]. At present, the prognostic indicators used in clinical practice, such as lymph node status and tumor factors, are subject to uncertainty [51–53]. MPE is a life-threatening complication of BRCA, and the presence of MPE indicates that cancer cells have spread to the pericardium (the sac surrounding the heart) and may interfere with its function [54–56]. Thus, we constructed a prognostic model based on MPE-related genes that may have better predictive value for BRCA patients.

In this study, the high-throughput expression profile data of TCGA was used to construct the prognosis model of BRCA, and the data of TCGA and GEO were used to verify the model. We revealed three BRCA subtypes associated with patient survival based on MPE expression. PIK3CA, TP53, TTN and MUC16 were the most frequently mutated genes in these three cluster subtypes. While the mutation rates of TP53 and USH2A were more significant in cluster 3. Somatic mutations in the PIK3CA gene occur in up to 40% of primary BRCA [57–59]. TP53 is a tumor suppressor gene that is responsible for regulating the proliferation of tumor cells, and tumors with mutations in TP53 tend to be more aggressive and have a poorer prognosis [60]. Patients with tumor treatment-induced cardiomyopathy have an excess of rare mutations in the dilated cardiomyopathy gene, with TTN being the most common [61]. MUC16 mutations are the third most common cancer mutated gene, especially in ovarian cancer [58,62]. Previous studies have found that missense mutations in USH2A are associated with better therapeutic efficacy and survival outcomes in cancer [63]. Based on the three

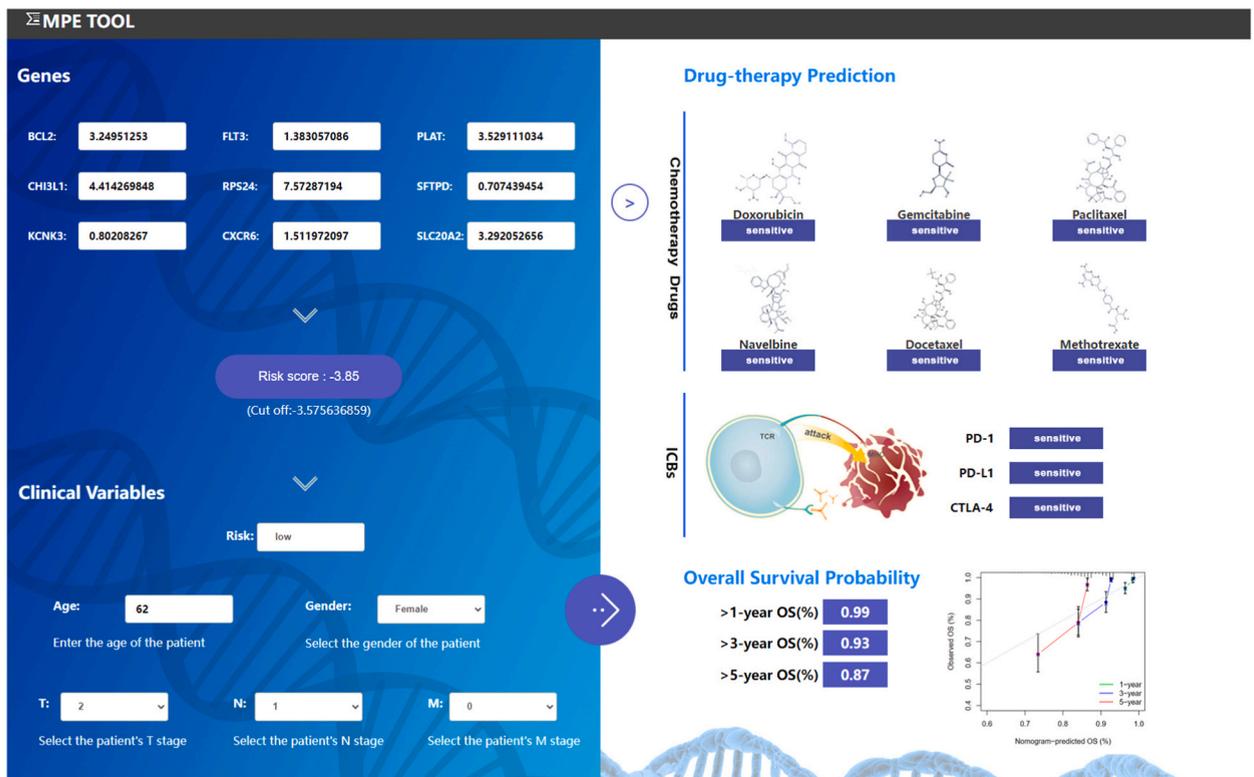


Fig. 8. Webpage interface of the prediction tool. Users can input the patient's MPE-related gene expression and clinical variables of the patient's age, gender and TNM stage on the left panel, and the response to chemotherapy drugs and immunotherapy and OS of 1 -, 3 -, 5-year of this patient will be output on the right panel.

subtypes, 74 differentially expressed MPE -related genes were identified. GO and KEGG analysis showed that these genes were mainly enriched in leukocyte-cell adhesion regulation, cytokine-mediated signaling pathway and cell chemotaxis, mainly through metabolic pathway, tumor transcriptional disorder pathway and Cytokine-cytokine receptor interaction pathway affect the prognosis of BRCA. These results are consistent with previous findings. Patients with pericardial effusion usually have elevated serum leukocyte counts and cytokine-induced killer cells can treat MPE [64,65], suggesting that it is associated with leukocyte cell-cell adhesion, cytokine-mediated signaling pathway, and Cytokine-cytokine receptor interaction pathway.

Univariate Cox regression analysis identified 50 MPE -related genes that were closely associated with survival of BRCA patients, and nine core genes (BCL2, FLT3, PLAT, CHI3L1, RPS24, SFTPD, KCNK3, CXCR6, SLC20A2) were finally selected for model construction after LASSO regression and multivariate Cox regression analysis. Among these, FLT3 (FMS-Like Tyrosine kinase-3) is a type 3 receptor tyrosine kinase with important roles in the proliferation, differentiation, and survival of hematopoietic stem cells, precursor B cells, etc. [66,67]. It is one of the most commonly mutated genes in acute myeloid leukemia (AML) [68]. In recent years, more and more studies have found that FLT3, as an important receptor tyrosine kinase in cell signaling, can lead to abnormal cell proliferation and induce tumorigenesis [69,70]. We calculated the risk score for each BRCA patient and classified them into high-risk group or low-risk group by median risk score. Risk scores, survival status analysis and survival curves showed that the high-risk group had less time to survive, higher mortality and significantly lower survival than the low-risk group. ROC curves also showed that the model performed well in predicting survival in BRCA patients.

We constructed a predictive nomogram combining age, gender, TNM stage and risk score to predict survival of BRCA patients. Calibration curves, ROC curves and DCA analysis showed that this nomogram had good predictive performance. For the convenience of clinical application, we have developed a web-based tool to quickly predict the OS as well as drug-therapy response of BRCA patients, which can provide a basis for clinicians to formulate individualized treatment plans and provide more effective treatment strategies. Inevitably, there are still some limitations in this study. Although our results have been validated in clinical samples, the sample size needs to be expanded. In addition, the potential molecular mechanisms of the genes we used to model lack further in vivo or in vitro functional experiments.

5. Conclusion

In summary, we developed an MPE-related model and constructed a web-based tool to predict the prognosis and therapy response of BRCA patients, which has been well validated from multiple aspects. This model has excellent reliability and accuracy. Our study contributes to the individualized treatment and prognosis assessment of BRCA patients and provides a new perspective for the discovery of immunotherapy drugs, which has certain clinical significance.

Funding

This study received funding from the National Natural Science Foundation of China, China (grant number 82003802), the Science and Technology Program of Hunan Health Commission, China (grant number 20201978), the China Scholarship Council, China (grant number 201808430085) and Clinical Research 4310 Program of the First Affiliated Hospital of the University of South China, China (grant number 20224310NHYCG04), Science and technology innovation Program of Hengyang City, China (grant number 202250045223), the Natural Science Foundation of Hunan Province, China (grant number 2019JJ50542 and 2023JJ50156).

Data availability

The data associated with this study have been deposited into a publicly available repository. The raw data utilized in this study are available in the UCSC xenabrowser (<https://xenabrowser.net/>), GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) under the accession number GSE20685 and GeneCards database (<https://www.genecards.org/>).

Ethical approval

There are no ethical considerations applicable to our article.

CRedit authorship contribution statement

Wendi Zhan: Writing – original draft, Visualization. **Haihong Hu:** Visualization, Conceptualization. **Bo Hao:** Visualization, Software. **Hongxia Zhu:** Visualization. **Ting Yan:** Writing – review & editing. **Jingdi Zhang:** Writing – review & editing. **Siyu Wang:** Writing – review & editing. **Saiyang Liu:** Writing – review & editing. **Taolan Zhang:** Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to express our heartfelt gratitude to the TCGA database, GEO database and GeneCards database for providing valuable resources for data analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e27507>.

References

- [1] Q. Zhu, et al., SEC14L3 plays a tumor-suppressive role in breast cancer through a Wnt/beta-catenin-related way, *Exp. Cell Res.* 417 (1) (2022) 113161.
- [2] J.A. de la Mare, et al., Breast cancer: current developments in molecular approaches to diagnosis and treatment, *Recent Pat. Anti-Cancer Drug Discov.* 9 (2) (2014) 153–175.
- [3] J.C. Keen, N.E. Davidson, The biology of breast carcinoma, *Cancer* 97 (3 Suppl) (2003) 825–833.
- [4] K. Tan, M.J. Naylor, The influence of Modifiable factors on breast and prostate cancer risk and disease progression, *Front. Physiol.* 13 (2022) 840826.
- [5] M. Fahad Ullah, Breast cancer: current perspectives on the disease status, *Adv. Exp. Med. Biol.* 1152 (2019) 51–64.
- [6] M. Zubair, S. Wang, N. Ali, Advanced approaches to breast cancer classification and diagnosis, *Front. Pharmacol.* 11 (2020) 632079.
- [7] H. Xu, et al., Prognostic role of vitamin D receptor in breast cancer: a systematic review and meta-analysis, *BMC Cancer* 20 (1) (2020) 1051.
- [8] M.V. Dieci, et al., Integration of tumour infiltrating lymphocytes, programmed cell-death ligand-1, CD8 and FOXP3 in prognostic models for triple-negative breast cancer: analysis of 244 stage I-III patients treated with standard therapy, *Eur. J. Cancer* 136 (2020) 7–15.
- [9] E.P. Olthof, et al., The prognostic value of the number of positive lymph nodes and the lymph node ratio in early-stage cervical cancer, *Acta Obstet. Gynecol. Scand.* 101 (5) (2022) 550–557.
- [10] K. Rajendran, M. Sudalaimuthu, S. Ganapathy, Cytological grading of breast carcinomas and its prognostic implications, *Cureus* 14 (9) (2022) e29385.
- [11] W.L. Donegan, Prognostic factors. Stage and receptor status in breast cancer, *Cancer* 70 (6 Suppl) (1992) 1755–1764.
- [12] S. Albugami, et al., Etiology of pericardial effusion and outcomes post pericardiocentesis in the western region of Saudi Arabia: a single-center experience, *Cureus* 12 (1) (2020) e6627.
- [13] S.E. Janus, B.D. Hoit, Effusive-constrictive Pericarditis in the Spectrum of Pericardial Compressive Syndromes, *Heart*, 2021.
- [14] A. Hein, Y.H. Wai, Subacute cardiac tamponade with massive pericardial effusion, *Cureus* 14 (12) (2022) e32250.
- [15] C.S. Restrepo, et al., Imaging findings in cardiac tamponade with emphasis on CT, *Radiographics* 27 (6) (2007) 1595–1610.
- [16] A.S. Barroso, et al., Pericardial mesothelioma presenting as a suspected ST-elevation myocardial infarction, *Rev. Port. Cardiol.* 36 (4) (2017) 307 e1–e307 e5.
- [17] F.A. Shepherd, et al., Tetracycline sclerosis in the management of malignant pericardial effusion, *J. Clin. Oncol.* 3 (12) (1985) 1678–1682.
- [18] S.R. Kim, et al., Effect of anti-inflammatory drugs on clinical outcomes in patients with malignant pericardial effusion, *J. Am. Coll. Cardiol.* 76 (13) (2020) 1551–1561.
- [19] H.L. Gornik, M. Gerhard-Herman, J.A. Beckman, Abnormal cytology predicts poor prognosis in cancer patients with pericardial effusion, *J. Clin. Oncol.* 23 (22) (2005) 5211–5216.
- [20] D. El Haddad, et al., Outcomes of cancer patients undergoing percutaneous pericardiocentesis for pericardial effusion, *J. Am. Coll. Cardiol.* 66 (10) (2015) 1119–1128.
- [21] L. Jia, et al., Differential expression of vascular endothelial growth factor-A, -C and -D for the diagnosis and prognosis of cancer patients with malignant effusions, *Oncol. Lett.* 10 (2) (2015) 667–674.
- [22] Y. Tang, et al., Identification and validation of a prognostic model based on three MVI-related genes in hepatocellular carcinoma, *Int. J. Biol. Sci.* 18 (1) (2022) 261–275.
- [23] W.Y. Cai, et al., Identification of a tumor microenvironment-relevant gene set-based prognostic signature and related therapy targets in gastric cancer, *Theranostics* 10 (19) (2020) 8633–8647.
- [24] K.J. Kao, et al., Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization, *BMC Cancer* 11 (2011) 143.
- [25] J. Gong, et al., HCC subtypes based on the activity changes of immunologic and hallmark gene sets in tumor and nontumor tissues, *Briefings Bioinf.* 22 (5) (2021).
- [26] A. Mayakonda, et al., Maftools: efficient and comprehensive analysis of somatic variants in cancer, *Genome Res.* 28 (11) (2018) 1747–1756.
- [27] V.F. Grabinski, O.W. Brawley, Disparities in breast cancer, *Obstet. Gynecol. Clin. N. Am.* 49 (1) (2022) 149–165.
- [28] M.E. Ritchie, et al., Limma powers differential expression analyses for RNA-sequencing and microarray studies, *Nucleic Acids Res.* 43 (7) (2015) e47.
- [29] G. Yu, et al., clusterProfiler: an R package for comparing biological themes among gene clusters, *OMICS* 16 (5) (2012) 284–287.
- [30] M. Zhang, et al., An immune-related signature predicts survival in patients with lung adenocarcinoma, *Front. Oncol.* 9 (2019) 1314.
- [31] T.T. Liu, et al., Identification of CDK2-related immune forecast model and ceRNA in lung adenocarcinoma, a pan-cancer analysis, *Front. Cell Dev. Biol.* 9 (2021) 682002.
- [32] C. Yan, et al., System analysis based on the cuproptosis-related genes identifies LIPT1 as a novel therapy target for liver hepatocellular carcinoma, *J. Transl. Med.* 20 (1) (2022) 452.
- [33] J. Li, et al., A novel hypoxia- and lactate metabolism-related signature to predict prognosis and immunotherapy responses for breast cancer by integrating machine learning and bioinformatic analyses, *Front. Immunol.* 13 (2022) 998140.
- [34] D. Wang, et al., Identification of the prognostic value of ferroptosis-related gene signature in breast cancer patients, *BMC Cancer* 21 (1) (2021) 645.
- [35] Y. Yao, et al., Development of a novel immune-related gene prognostic index for breast cancer, *Front. Immunol.* 13 (2022) 845093.
- [36] D.S. Chandrashekar, et al., UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses, *Neoplasia* 19 (8) (2017) 649–658.
- [37] Z.N. Lye, M.D. Purugganan, Copy number variation in domestication, *Trends Plant Sci.* 24 (4) (2019) 352–365.
- [38] E. Cerami, et al., The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data, *Cancer Discov.* 2 (5) (2012) 401–404.
- [39] S. Hanzelmann, R. Castelo, J. Guinney, GSEA: gene set variation analysis for microarray and RNA-seq data, *BMC Bioinf.* 14 (2013) 7.
- [40] A. Liberzon, et al., The Molecular Signatures Database (MSigDB) hallmark gene set collection, *Cell Syst* 1 (6) (2015) 417–425.
- [41] A. Subramanian, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Natl. Acad. Sci. U. S. A.* 102 (43) (2005) 15545–15550.
- [42] K. Yoshihara, et al., Inferring tumour purity and stromal and immune cell admixture from expression data, *Nat. Commun.* 4 (2013) 2612.
- [43] B. Chen, et al., Profiling tumor infiltrating immune cells with CIBERSORT, *Methods Mol. Biol.* 1711 (2018) 243–259.
- [44] P. Charoentong, et al., Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade, *Cell Rep.* 18 (1) (2017) 248–262.

- [45] P. Geeleher, N. Cox, R.S. Huang, pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels, *PLoS One* 9 (9) (2014) e107468.
- [46] L. Li, et al., A nomogram model to predict prognosis of patients with genitourinary sarcoma, *Front. Oncol.* 11 (2021) 656325.
- [47] Q.W. Wang, W.W. Lin, Y.J. Zhu, Comprehensive analysis of a TNF family based-signature in diffuse gliomas with regard to prognosis and immune significance, *Cell Commun. Signal.* 20 (1) (2022) 6.
- [48] J. Smith, et al., Promoter DNA hypermethylation and paradoxical gene activation, *Trends Cancer* 6 (5) (2020) 392–406.
- [49] A. Montero, et al., Postmastectomy radiation therapy in early breast cancer: utility or futility? *Crit. Rev. Oncol. Hematol.* 147 (2020) 102887.
- [50] F. Zhao, et al., Survival and prognostic factors of adult intracranial ependymoma: a single-institutional analysis of 236 patients, *Am. J. Surg. Pathol.* 45 (7) (2021) 979–987.
- [51] K. Zhang, et al., A novel systematic oxidative stress score predicts the prognosis of patients with operable breast cancer, *Oxid. Med. Cell. Longev.* 2021 (2021) 9441896.
- [52] C. Tang, et al., Lymph node status have a prognostic impact in breast cancer patients with distant metastasis, *PLoS One* 12 (8) (2017) e0182953.
- [53] S. O'Grady, M.P. Morgan, Microcalcifications in breast cancer: from pathophysiology to diagnosis and prognosis, *Biochim. Biophys. Acta Rev. Canc* 1869 (2) (2018) 310–320.
- [54] A.C. Bibby, et al., ERS/EACTS statement on the management of malignant pleural effusions, *Eur. Respir. J.* 52 (1) (2018).
- [55] D. He, et al., Novel therapies for malignant pleural effusion: anti-angiogenic therapy and immunotherapy, *Int. J. Oncol.* 58 (3) (2021) 359–370 (Review).
- [56] S.I. Bashour, B.J. Mankidy, D.R. Lazarus, Update on the diagnosis and management of malignant pleural effusions, *Respir. Med.* 196 (2022) 106802.
- [57] F. Mosele, et al., Outcome and molecular landscape of patients with PIK3CA-mutated metastatic breast cancer, *Ann. Oncol.* 31 (3) (2020) 377–386.
- [58] Y. Samuels, V.E. Velculescu, Oncogenic mutations of PIK3CA in human cancers, *Cell Cycle* 3 (10) (2004) 1221–1224.
- [59] N. Cancer Genome Atlas, Comprehensive molecular portraits of human breast tumours, *Nature* 490 (7418) (2012) 61–70.
- [60] K. Sabapathy, D.P. Lane, Therapeutic targeting of p53: all mutants are equal, but some mutants are more equal than others, *Nat. Rev. Clin. Oncol.* 15 (1) (2018) 13–30.
- [61] P. Garcia-Pavia, et al., Genetic variants associated with cancer therapy-induced cardiomyopathy, *Circulation* 140 (1) (2019) 31–41.
- [62] D. Thomas, et al., Isoforms of MUC16 activate oncogenic signaling through EGF receptors to enhance the progression of pancreatic cancer, *Mol. Ther.* 29 (4) (2021) 1557–1571.
- [63] D. Yang, et al., USH2A mutation and specific driver mutation subtypes are associated with clinical efficacy of immune checkpoint inhibitors in lung cancer, *J. Zhejiang Univ. - Sci. B* 24 (2) (2023) 143–156.
- [64] J.E. Roberts, et al., Association between parapneumonic effusion and pericardial effusion in a pediatric cohort, *Pediatrics* 122 (6) (2008) e1231-e1235.
- [65] H. Wang, et al., Curative effects of dendritic cells combined with cytokine-induced killer cells in patients with malignant pericardial effusion, *Med. Sci. Mon. Int. Med. J. Exp. Clin. Res.* 22 (2016) 4159–4163.
- [66] N.J. Short, et al., Emerging treatment paradigms with FLT3 inhibitors in acute myeloid leukemia, *Ther Adv Hematol* 10 (2019) 2040620719827310.
- [67] A. Nogami, et al., FLT3-ITD confers resistance to the PI3K/Akt pathway inhibitors by protecting the mTOR/4EBP1/Mcl-1 pathway through STAT5 activation in acute myeloid leukemia, *Oncotarget* 6 (11) (2015) 9189–9205.
- [68] A.J. Ambinder, M. Levis, Potential targeting of FLT3 acute myeloid leukemia, *Haematologica* 106 (3) (2021) 671–681.
- [69] M. Katoh, M. Katoh, Precision medicine for human cancers with Notch signaling dysregulation, *Int. J. Mol. Med.* 45 (2) (2020) 279–297 (Review).
- [70] Y.L. Lai, et al., Novel FLT3/AURK multikinase inhibitor is efficacious against sorafenib-refractory and sorafenib-resistant hepatocellular carcinoma, *J. Biomed. Sci.* 29 (1) (2022) 5.