Identification of molecular characteristics associated with disulfidptosis and ferroptosis-related genes in breast cancer, along with immune cell infiltration Analysis and Development of a prognostic risk model

Xiao-Qing Pei
Shandong Provincial Hospital Affiliated to Shandong First Medical University

Wen-Hao Wang
Shandong Provincial Hospital Affiliated to Shandong First Medical University

Zhen-dan Zhao (.zzdslyy@126.com)
Shandong Provincial Hospital Affiliated to Shandong First Medical University

Hua-wei Zhang
Shandong Provincial Hospital Affiliated to Shandong First Medical University

Research Article

Keywords:

Posted Date: October 5th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3393444/v1

License: ☕️️️ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background

Breast cancer (BRCA) is a common and highly heterogeneous malignancy with a relative poor prognosis. Disulfidptosis is a newly discovered mode of cell death induced by disulfide stress. Like ferroptosis, disulfidptosis is a novel form of programmed cell death. The goal of this research was to explore new biomarkers associated with disulfidptosis and ferroptosis that can guide the treatment of BRCA.

Methods

We collected expression and clinical information about BRCA patients from public database. We comprehensively characterized the relationship between disulfidptosis and ferroptosis-related genes and the molecular characteristics as well as immune cell infiltration of tumor tissue. Next, a risk model was identified based on COX regression model with least absolute shrinkage and selection operator regression (LASSO) algorithm. Besides, we evaluated the prognostic value and treatment sensitivity of various subtypes.

Results

Specific molecular characteristics and model possessed a promising predictive potential. Immune infiltration and treatment sensitivity analysis all showed differences between groups.

Conclusions

Our molecular subtypes and risk model, with strong immune prediction and prognostic prediction capabilities, are committed to guiding BRCA treatment.

Introduction

Pyroptosis, necroptosis, and ferroptosis represent, so far, are the best-characterized non-apoptotic forms of regulated cell death [1]. Ferroptosis is an iron-mediated death characterized by excessive accumulation of lipid peroxides and reactive oxygen species (ROS), ultimately leading to cell membrane damage [2]. Ferroptosis is a significant regulatory mechanism involved in the growth and development of various malignancies [3]. Currently, ferroptosis is regarded as a potential therapeutic target for breast cancer. The proliferation of breast cancer cells was reportedly inhibited by activation of ferroptosis [4]. Recently it has been found that PARP1 degrader could promote ferroptosis by downregulating the SLC7A11 pathway in p53-positive breast cancer cells, which provides new insights for treating breast cancer patients [5].
More recently, Liu et al. demonstrated a new type of cell death pathway named disulfidptosis [6]. In highly SLC7A11 expressed cells, glucose starvation can cause abnormal accumulation of cystine and other disulfide, which induced disulfide stress and increase of disulfide bond in actin cytoskeleton and result in action filament contraction, destruction of cytoskeleton structure, and ultimately cell death [6, 7].

Both ferroptosis and disulfidptosis play crucial role in breast cancer [8, 9]. However, it remains unclear whether ferroptosis combined with disulfidptosis-related genes can be used as a prognosis marker, drug sensitivity, or immunotherapy response in breast cancer. For predicting prognosis and selecting treatment regimens, it is extremely important to construct molecular subtypes and characterize the corresponding immune microenvironment. Hence, this study combined genes associated with ferroptosis and disulfidptosis to investigate their expression and roles in breast cancer. We identified specific molecular subtypes for prognosis prediction and drug response assessment. We hope our study will guide the precise treatment of breast cancer.

**Method**

**Data collection**

Ferroptosis and disulfidptosis-related gene expression, somatic mutation data, CNV files and clinical information of breast cancer patients were obtained from The Cancer Genome Altas (TCGA) database (https://portal.gdc.cancer.gov//, accessed on 10 March 2023). Besides, GSE159956 dataset from GEO database (http://www.ncbi.nlm.nih.gov/geo/) was utilized to acquire clinical parameters and normalized gene expression data. After removing samples without clinical follow-up information, survival time, and survival status, 113 normal control samples from TCGA database, 1096 breast cancer samples from TCGA database and 295 breast cancer samples from GSE159956 dataset were obtained. 236 ferroptosis genes were obtained from the world's first database of ferroptosis regulators and markers and ferroptosis-disease associations (http://www.zhounan.org/ferrdb/legacy/operations/download.html). Disulfidptosis-related gene lists were acquired from recently published literature [6]. The protein-protein interaction (PPI) information was obtained by the STRING database (https://string-db.org/). Moreover, we analyzed tumor mutation burden (TMB) using the package “maftools”. The flowchart of the study design was illustrated in Fig. 1.

**Consensus cluster analysis**

We used K-means clustering algorithms and the “ConsensusClusterPlus” R package to identify different prognosis-related patterns. Overall survival (OS) curve was generated by the Kaplan-Meier (KM) method. Principle component analysis (PCA) was applied to identify to compare the composition of different clusters. Gene set variation analysis (GSVA) was performed to identify differences in biological function of prognosis-related genes using the Kyoto Encyclopedia of Genes and Genomes (KEGG) gene set (c2.cp.kegg.v7.2.gmt) and Gene Ontology (GO) gene set (c5.go.Hs.symbols.gmt).
Relationship among molecular patterns, clinical characters, and TME

To determine the clinical significance of the cluster, we investigated associations between molecular patterns and clinical characteristics including N-stage and M-stage, as well as survival outcomes. Immune scores and stromal scores of breast cancer patients were calculated by the ESTIMATE algorithm. Simultaneously, single-sample gene set enrichment analysis (ssGSEA) and CIBERSORT algorithms were used to assess the extent of immune cell infiltration.

Identification of DEGs between molecular patterns

We used the “limma” R package to distinguish differentially expressed genes (DEGs), and genes with P < 0.05 and |log2FoldChange(FC)|>0.585 were screened as significant DEG. The R package “clusterProfile” was used to perform GO and KEGG pathway enrichment analysis.

Risk model construction

Univariate COX regression analysis was used to screen for prognosis-related DEGs. For a more comprehensive assessment, the samples were divided into specific gene clusters using the “biomanager” and “consensus cluster plus” R packages according to prognosis-related DEGs. Ferroptosis and disulfidptosis related-risk model was constructed to provide a better assessment of the prognosis of breast cancer patients. The samples of the TCGA database were designed as the training set and the samples of the GEO database were treated as the test set to construct the risk model based on the training set. Subsequently, a multivariate COX regression model after least absolute shrinkage and selection operator (LASSO) regression was used to construct best prognostic signature by selecting genes from prognostic DEGs. As a result, 6 hub genes and their correlative coefficients were obtained to develop the risk signature. Ultimately, we generated risk score with the following formula: risk score= Σ (gene expression × coefficient). The “survminer” package in R was used to determine the optimal threshold for risk scores to classify patients into high- and low-risk groups.

Risk signature validation

KM survival analysis was applied to compare the OS between the high-risk and low-risk groups. PCA was performed to reduce the dimensionality and visualize difference between low-risk and high-risk groups. Univariate and multivariable COX regression analyses were respectively carried out in order to determine whether the risk score had prognostic significance independently. A nomogram was established to predict the 1, 3, and 5-year OS in breast cancer patients, through integration with the clinicopathological characteristics and risk score, using the “RMS” package of the R. Time-dependent receiver operating characteristic curve (ROC) analysis of 1-year, 3-year and 5-year survival was analyzed using the “timeROC” R package. Gene set enrichment analysis (GSEA) was conducted using “clusterProfiler” R package to further explore the potential molecular mechanism among different risk groups.
Exploration of TME and Drug sensitivity between high- and low-risk groups

With above methods, we assessed immune cell infiltration and immune function in high-risk and low-risk groups. Moreover, The CIBERSORT, CIBERSORT-ABS, QUANTISEQ, MCP-COUNTER, XCELL, TIMER and EPIC algorithms were applied for immune infiltration estimations. To predict the clinical outcome of immune checkpoint inhibitor blockade treatment, we also explored the expression of immune checkpoints such as CDC80, LAG3, NRP1, TNFSF9, TNFRSF9 and TNFSF15. To assess the therapeutic effects of chemotherapy drugs in high-risk and low-risk groups, we evaluated the half-maximal inhibitory concentration (IC50) values of chemotherapy drugs commonly used to treat breast cancer patients using the “pRRophetic” package.

Real time PCR

Cancer tissue samples and precancerous tissue samples were collected from Provincial Hospital Affiliated to Shandong First Medical University. The RNA-easy Isolation Reagent (Vazyme Nanjing, China) was applied to extract RNA, while cDNA was synthesized using the Reverse Transcription Kit (SparkJade, Shandong, China). Quantitative real-time PCR (RT-PCR) was carried out using the SPARKscript PCR Kit ((SparkJade, Shandong, China). The relative mRNA expression level was measured by 2-ΔΔCT method. The primer sequence were EIF4EBP1-F: 5’-ACCAGCCCTTCCAGTGATGAG-3’, EIF4EBP1-R: 5’-CGCCGCCCCTATCTTCTTC-3’, SIAH2-F: 5’-CCGCAAGCAAGCGAGAC-3’, SIAH2-R: 5’-CGCACCCACACCTCATG-3’, APOD-F: 5’-GCCACCGACTATGAGAC-3’, APOD-R: 5’-ACTGTTTCTGAGGAGATAGGG-3’, LTF-F: 5’-CCGCGCTGGACAGGACTG-3’, LTF-R: 5’-CGCCATACAGACAGACAGAG-3’, SERPINA1-F: 5’-AAGAGCGTCGGTGCTGGA-3’, SERPINA1-R: 5’-CGATGGCTCAGCAGCCTTATG-3’, CD24-F: 5’-CTGCTGGCGACTGCTTACC-3’, CD24-R: 5’-CCTTGGTGGTGATGTTT-3’, GAPDH-F: 5’-GTATGACACAGCCTCAAGT-3’, GAPDH-R: 5’-GTCCTCCACGATACCAAAG-3’.

Statistical analysis

Differences between the two groups were determined by t-tests. A log-rank test was used to determine differences between KM curves. All statistical significances were denoted at P < 0.05 unless otherwise noted.

Results

Identification of prognostic ferroptosis and disulfidptosis-related genes in breast cancer samples.

We identified the expression levels of the 236 genes related to ferroptosis and 23 genes related to disulfidptosis in tumor and normal samples based on the TCGA-BRCA dataset. As shown in Fig. 2A, 108 DEGs were obtained using a criteria of P < 0.05 and |logFC| >0.585. The top 50 genes were shown in heatmap (Fig. 2B). To evaluate the influence of above DEGs on prognosis, survival analysis was
performed using univariate COX and KM method. Finally, we found that 20 genes were closely related to the prognosis of breast cancer. According to supplementary (Additional File1: Fig. S1), ALOX15B, ZFP36, TP36, JUN, SLC1A4, SLC2A1, CD01, MYB and MUC1 high expression indicated a better prognosis of OS, whereas high expression of NGB, SLC7A11, RPM2, ASNS, TFRC, FH, ELAVL1, HSF1, CA9, SLC3A2 and PRDX1 suggested a poor prognosis (Additional File2: Table S1).

Genetic mutation landscape of prognosis-related genes (PRGs) in breast cancer

The network of protein-protein interaction (PPI) was first constructed using STRING (https://string-db.org/) to analyze interaction of prognostic-related genes (PRGs). As shown in Fig. 3A, JUN, SLC2A1, RRM2, TFRC, SLC7A11, SLC3A2 were hub genes. Following this lead, we summarized the incidence of copy number variations (CNV) and somatic mutations of PRGs in breast cancer. 55 of the 991 samples had gene mutations, and the results showed that MYB was the gene with highest mutation rate, followed by CA9 and ALOX15B (Fig. 3B). We next analyzed the CNV frequency in breast cancer. The locations of CNV alteration for 20 PRGs on chromosomes are shown in Fig. 3C. Among them, the CNV of MUC1, FH, HSF1, TFRC, TP63, SLC3A2, MYB, ZFP36, SLC2A1, JUN, PRDX1, CA9 and SLC1A4 increased, while ELAVL1, RPM2, SLC7A11, ASNS, CD01, NGB and ALOX15B exhibited an extensive decreased in CNV (Fig. 3D).

Generation of ferroptosis and disulfidptosis genes subtypes in breast cancer

To investigate the relationship between PRGs and tumorigenesis, 1109 BRCA patients from TCGA-BRCA and 295 patients from GSE159956 were included in this study. A network of PRGs interactions, regulatory relationships and their prognostic significance in BRCA patients is shown in Fig. 4A. To further explore the associations between the expression of the 20 PRGs and BRCA molecular subtypes, consensus clustering analysis was conducted. Our results suggested that the clustering variable \( k = 2 \) seemed to be the optimum choice for classifying the whole cohort in cluster A (\( n = 1381 \)) and cluster B (\( n = 1386 \)) (Fig. 4B). According to PCA analysis, there was a difference between two groups (Fig. 4C). The KM curves showed a significance difference in the OS between the two groups. Individuals in subtype B had shorter OS (Fig. 4D). The heatmap implied the expression discrepancy of 20 PRGs in two clusters and their links to clinically relevant information (Fig. 4E).

Features of TME cell infiltration in different clusters

The GO enrichment and KEGG pathway analyses were performed to explore the biological processes and pathways associated with the two clusters using GSVA (Additional File3: Table S2). Enriched biological processes in cluster A included later sprouting from an epithelium, mammary gland development and mammary gland epithelium development, while major biological processes in cluster B include meiotic
cell cycle, cell cycle and replication, DNA replication initiation (Fig. 5A). The pathways of complement and coagulation cascades, valine leucine and isoleucine degradation were enriched in cluster A. In contrast, those of cell cycle, homologous recombination, DNA replication were enriched in cluster B (Fig. 5B). Thereafter, we examined the infiltration levels of 23 human immune cells utilizing the ssGSEA algorithm (Additional File 4: Table S3). As shown in Fig. 5C, immune infiltration of various immune cells differed significantly between the two clusters. We found that activated B cell, activated CD4 T cell, activated dendritic cell, CD56bright natural killer cell and CD56dim natural killer cell were lowly expressed in cluster A. Subsequently, based on 366 DEGs investigated in both clusters by using the “limma” R package (Additional File 5: Table S4), we performed GO and KEGG enrichment analysis. The results revealed that extracellular matrix organization, collagen-containing extracellular matrix and extracellular matrix structural constituent were significantly enriched in biological processes, cell components and molecular functions (Fig. 5D, Additional File 6: Table S5). KEGG enrichment analysis revealed the these DEGs were mainly related to cell cycle, PI3K-AKT pathways (Fig. 5E, Additional File 7: Table S6).

Secondary clustering: Identification of gene subgroups based on DEGs.

Given that two clusters had a remarkable survival difference, we then applied the univariate COX regression analysis to identify the survival significance of 366 DEGs among cluster A and B. As a result, 219 genes were ultimately determined (Additional File 8: Table S7). To better understand the potential mechanism for the differential prognosis, a secondary consensus clustering was performed based on 219 prognostic genes. All samples were classified into three clusters (cluster A-C) (Fig. 6A). The result of PCA analysis indicated that there were significant differences among three clusters (Fig. 6B). The KM analysis is revealed that patients in cluster A had a significantly best overall survival (Fig. 6C). Moreover, the correlation among clinicopathological variables, PRG clusters and gene clusters was shown in Fig. 6D. Through the ssGSEA algorithm, we determined the degree of infiltration of the different immune cell population in three clusters. As shown in Fig. 6E, cluster A display significant higher eosinophil, mast cell and type17 T helper cell infiltration. As expected, there were remarkable difference among PRGs expression of three gene clusters (Fig. 6F, Additional File 9: Table S8).

Construction and validation of prognostic risk model

To construct a risk signature, LASSO regression analysis was performed for 219 DEGs. Next, genes obtained by LASSO regression were analyzed by multivariate COX regression analysis to further screen core genes. Eventually, 6 genes were enrolled to construct the risk signature. Risk score was calculated by the following formula: Risk score = (0.132165×EIF4EBP1 expression) + (-0.169712277×SIAH2 expression) + (-0.070124216×APOD expression) + (-0.050397905×LTF expression) + (-0.143307142×SERPINA1 expression) + (0.152959868×CD24 expression). Furthermore, we incorporate the TCGA cohort in training group and the GEO cohort in test group to improve the accuracy of the prognostic model. Patients from TCGA cohort were split into two groups, with low- and high-risk scores, according to the best cut-off value. The KM survival curves demonstrated that the OS time was longer in the low-risk group compared with that in high-risk group (Fig. 7A). The relative expression of these genes for every
patient were shown in Fig. 7B. Consistently, the results showed that patients with a high-risk score had lower survival probability and were more likely to die than those with a low-risk score (Fig. 7C, D). The classification ability of the prognosis signature was assessed by PCA (Fig. 7E). Moreover, predictability and accuracy of the model were investigated by ROC analysis with AUC values for 1-, 3-, and 5-year OS of 0.742, 0.645 and 0.638 respectively (Fig. 7F). To validate the observations described above, we applied the same methodology to GSE159956 cohort. In general, results were relatively consistent between the training and validation set (Fig. 7G-K). The time-dependent ROC curve showed that the AUC of 1-, 3-, and 5-years was 0.674, 0.717 and 0.721 respectively (Fig. 7L). Figure 7M demonstrated the connection between PRG clusters, gene clusters, risk layers, and survival status. Moreover, we found a significance difference between two PRG clusters and three gene clusters in risk score. Patients with PRG cluster B and gene cluster C had higher risk scores (Fig. 7M, O).

The risk model performs a substantial predictive prognostic value

To determine whether the prognostic signature for OS was an independent prognostic factor for BRCA patients, we performed the univariate and multivariate COX regression analysis. According to the univariate regression assessment, the hazard (HR) of risk score and the 95% confidence interval were found to be 2.0107 and 1.5077–2.6815 (Fig. 8A). After statistical adjustment for other variables with multivariate COX analysis, the risk score was still independently prognostic of OS (HR = 2.2323, 95%CI = 1.5939–2.8499) (Fig. 8B). The nomogram was constructed in accordance with the independent factors including N stage, M stage and risk score to better predict survival for BRCA patients at 1-, 3-, and 5-years (Fig. 8C). The accuracy and robustness of the prognostic nomogram were demonstrated by the calibration curve (Fig. 8D). In order to explore the potential mechanism leading to the difference between the two groups, we conducted GSEA. As a result, pathways such as amyotrophic lateral sclerosis, Rap1 signaling pathway, and transcriptional misregulation in cancer were significantly enriched in high-risk group, while thyroid hormone synthesis and Fanconi anemia pathway were mainly enriched in low-risk group (Fig. 8E, Additional File10: Table S9).

Evaluation of immune infiltration and drug sensitivity between high- and low- groups

The association between the signature and immune infiltration was investigated using the TIMER, CIBERSORT, CIBERSORT-ABS, XCELL, QUANTISEQ, EPIC, and MCP-counter algorithms, which showed in Fig. 9A. Considering the importance of infiltrating immune cells in TME, the abundance of infiltrating immune cells in different risk groups was evaluated by CIBERSORT. As a result, large differences in the distribution of immune cells were observed between high- risk group and low- risk group (Fig. 9B, C, Additional File11: Table S10). As a result, high- risk group was remarkably rich in immune-suppressive cell infiltration, including T cell regulatory (Tregs) and macrophages (M0). Similarly, spearman correlation analysis revealed a positive association between the level of risk score and the level of regulatory T cell
as assessed by ssGSEA, which indicated that part of the reason for the poor prognosis of patients in the high-risk subgroup may be the immunosuppressive microenvironment (Fig. 9D, Additional File12: Table S11). Given the essentiality that immune checkpoint inhibitors attribute to immunotherapies, we further focused on the difference in the expression of immune checkpoints between the two groups. As we have seen, the expression of CD80, LAG3, NRP1, TNFRSF9, TNFSF9 and TNFSF15 were significantly overexpressed in the high-risk population, indicating a highly immunosuppressed state in the high-risk group (Fig. 9E).

To improve the therapeutic effect of patients with BRCA, we investigated the sensitivity difference of common chemotherapy drugs among two groups. The results of the GDSC database analysis showed that IC50 values of drugs including Bexarotene, Imatinib, Metformin, Nilotinib, Pazopanib, Sorafenib and Tipifarnib were higher in patients of the high-risk group than those of the low-risk group, which indicated that patients in low-risk category may be more responsive to those anticancer drugs (Additional File13: Fig. S2A). Inversely, the IC50 levels of Camptothecin, Cisplatin, Cyclopamine, Docetaxel, Doxorubicin, Etoposide and Paclitaxel in low-risk group were higher than that in high-risk group (Additional File13: Fig. S2B).

**Multidimensional validation**

Through searching the public database, we found that mRNA of the 6 core genes differentially expressed in normal tissues and BRCA tissues. As shown in Fig. 10A-B, EIF4EBP1 and CD24 were highly expressed in cancer tissues, while the mRNA of SIAH2, APOD, LTF, SERPINA1 were elevated in normal tissues (Fig. 10C-F). To validate these results, RT-PCR was performed in 24 breast cancer tissues and 24 normal tissues. Result of RT-PCR showed consistency with those of the public databases (Fig. 10G-L). To investigate the functions of 6 hub genes in BRCA patients, we first analyzed the relationship between their expression and BRCA patients’ survival prognosis, using an online database Kaplan-Meier (KM) Plotter Online Tool (http://www.kmplot.com). As a result, high expression of EIF4EBP1 and CD24 predicts poor prognosis in BRCA while high expression of SIAH2, APOD, LTF, SERPINA1 were associated with a prolonged overall survival of patients with breast cancer (Fig. 10M-R). Next, the datasets GSE25066, GSE124647, GSE11121 form GEO databases further validated these results obtained from KM-plotter online tool (Fig. 10S-X).

**Discussion**

Disulfidptosis was recently defined as a distinct type of regulated cell death which is related to metabolism [6]. It is reported that the depletion of NADPH triggered by glucose starvation, glucose transporter inhibitors, or by providing excess cystine caused the production of disulfide stress in the cell, leading to the formation of disulfide bonds between actin cytoskeleton proteins and the collapse of the F-actin network, eventually initiated disulfidptosis [10]. Additionally, the inhibitors of ferroptosis, apoptosis, necroptosis and autophagy could not reverse this particular cell death [6]. Currently, disulfidptosis has been studied in BRCA through the use of bioinformatics approaches [9, 11, 12]. However, this study innovatively gathers disulfidptosis and ferroptosis genes to comprehensively describe how disulfidptosis
and ferroptosis related-genes patterns are clinically significant and how they may be related to TME features. In addition, the disulfidptosis and ferroptosis related-risk model was constructed to promote the understanding of TME and assist doctors to develop more effective immunotherapeutic strategies.

In current study, the expression of disulfidptosis-related genes and ferroptosis-related genes were identified in TCGA cohort and a total of 20 PRGs were screened from 108 DEGs. Next, GSE159956 cohort from GEO database were include for a comprehensive analysis. Then, BRCA patients were divided into two clusters (cluster A and B) with different prognosis, immune infiltrations and functions by using unsupervised clustering analysis on 20 PRGs. Subsequent GSVA enrichment analysis revealed that cluster B was mainly enriched in DNA-related cellular processes such as cell cycle and replication, and DNA replication initiation while cluster B was significantly related to mammary gland development biological processes including mammary gland epithelium development and mammary gland duct morphogenesis. Furthermore, GO and KEGG functional enrichment analyses have manifested that DEGs among two clusters were highly associated with biological processes and pathways such as cell cycle, PI3K-AKT signaling, and P53 signaling. In further continuation, three clusters (cluster A, B and C) with different clinical characteristics, immune functions and environmental influence were identified for BRCA based on the DEGs related to subgroups signature. After LASSO regression and multivariate COX regression analysis, six genes- EIF4EBP1, SIAH2, APOD, LTF, SERPINA1, and CD24 were selected to build the risk score model. Interestingly, patients with lower risk score had a better prognosis, which supported the prognostic value of the risk model. Furthermore, according to the univariate and multivariate COX regression analysis, the risk signature was proved to be an independent prognosis factor. Subsequently, ROC analysis validated the predictive robustness of risk model over 1-, 3-, and 5- years of OS.

To further understand the underlying cause for the difference in prognosis in patients in the high- and low- risk groups, the following aspects were systematically analyzed. In the first place, the unique molecular mechanism might be contributed to this discrepancy. GSEA pathway enrichment analysis revealed that molecular function in high risk-group mainly enriched in cancer related pathway such as Rap1 signaling pathway and transcriptional misregulation in cancer. On the other hand, the genes in the low-risk group were mainly enriched in “thyroid hormone synthesis” and “Fanconi anemia pathway”. Altogether, the various molecular pathway may account for one of the possible reason for the markedly different prognosis of patients in high- and low- risk groups.

Immune characteristics may be another reason contributing to the above phenomenon. CIBERSORT algorithm demonstrated that the infiltration level of T cells regulatory (Tregs) were elevated in high-risk group, while mast cells resting were more abundant in the low-risk group. Increasing evidence has shown that Treg cells played a negative role in anti-tumor immunity and mast cells predicted a better prognosis in breast cancer [13, 14]. In addition, ssGSEA analysis revealed that the infiltration of activated CD4 T cells with role of tumor-promoting immunity was proportional to the risk score while the infiltration of NK cells with function of anti-tumor immunity inversely correlated with risk score [15, 16]. Therefore, the worse prognosis in the high-risk group may be due to intense immunosuppression.
To determine whether the risk score could predict the drug response of individual patients, we analyzed the relationship between the IC50 values and risk score. The high risk score was correlated with lower IC50 of Camptothecin, Cisplatin, Cyclopamine, Docetaxel, Doxorubicin, Etoposide and Paclitaxel, suggesting higher sensitivity. For several years, immunotherapy has shown to be promising therapy for many cancers, there is no exception for breast cancer. Studies, however, have shown that not all therapies used effective on all patients. Further experience and research are required to differentiate between “responder” and “non-responder” patients. A large number of studies have demonstrated that the efficacy of immunotherapy relied on the adequate expression of immune checkpoints [17, 18]. In this study, the vast immunological checkpoint genes were up-regulated in BRCA tissues in the high-risk group, suggesting that patients in the high-risk group are more likely to benefit from immunotherapy.

Next, we verified the expression of six hub genes (EIF4EBP1, SIAH2, APOD, LTF, SERPINA1, CD24) in BRCA tissues and para-cancerous tissues using the TCGA database, GEO database, and clinical specimens collected by ourselves. It has been reported that EIF4EBP1 functioned as oncogene and knockdown of EIF4EBP1 inhibited the proliferation of breast cancer cell [19]. Multiple studies have proven that CD24 was overexpressed in breast cancer [20]. A meta-analysis revealed that overexpression of CD24 was associated with higher histological grade, shorter overall survival and disease-free survival in patients with BRCA [21]. Our analysis of the database indicated that the expression of EIF4EBP1 and CD24 were prominently up-regulated in BRCA compared with normal tissues, which was also verified in our breast cancer clinical specimens. In our risk model, the high expression of EIF4EBP1 and CD24 was positively correlated with risk score. SERPINA1, a member of the seine protease inhibitor family, is downregulated in several types of tumors, but its role in breast cancer has not been yet reported [22]. We found that SERPINA1 was downregulated in high-risk patients, which may predict a poor prognosis. In this study, the database and clinical sample verification results showed that SERPINA1 expression substantially lower in BRCA tissues than that in normal tissues. Moreover, the results from database indicated that elevated expression of SERPINA1 was an indicative of a longer survival in BRCA patients. Our study provided first insight into the expression and potential prognostic value of SERPINA1 in patients with BRCA, which suggested that SERPINA1 might be an effective target for BRCA therapies.

In summary, we have systematically discussed the molecular changes, intracellular pathways, and immune cell infiltration features related to disulfidptosis and ferroptosis in BRCA in our study. After using bioinformatic techniques, we selected six genes (EIF4EBP1, SIAH2, APOD, LTF, SERPINA1, CD24), verified their different expression in tumor and para-cancerous tissue samples at tissue level, and then analyzed the relationship between the expression of SERPINA1 protein and prognosis of BRCA patients. The herein-described risk model is able to reliably predict BRCA patient prognosis.

Admittedly, this study has certain limitations. Firstly, most of the data used for the analysis were derived from public database. As the data was analyzed retrospective, it was prone to several biases, which affected the accuracy of the results. Secondly, disulfidptosis is a newly discovered kind of programmed cell death, and there is little information about the role of disulfidptosis in oncogenesis. Last but not least,
this study only carried out the superficial experiments, lack of biological functional verification experiments of six genes in this model.

**Declarations**

**Ethics approval and consent to participate**

The protocol was approved by Ethics Committee for Biomedical Research Involving Human Beings of Shandong Provincial Hospital (SZRJJ: NO.2022-071).

**Consent for publication**

All the authors provided consent for publication.

**Availability of data and materials**

The datasets analyzed during the current study are available in TCGA database (https://portal.gdc.cancer.gov/, accessed on 10 March 2023), GEO database (http://www.ncbi.nlm.nih.gov/geo/) including GSE159956 GSE27562 GSE11121 GSE124647 GSE250066 GSE124647, and Kaplan-Meier (KM) Plotter Online Tool (http://www.kmplot.com). 236 ferroptosis genes were obtained from the world's first database of ferroptosis regulators and markers and ferroptosis-disease associations (http://www.zhounan.org/ferrdb/legacy/operations/download.html). The protein-protein interaction (PPI) information was obtained by the STRING database (https://string-db.org/).

**Competing interests**

The authors declare no competing interests.

**Funding**

This study was supported by The 2022 Research Incubation Fund Project of Shandong Provincial Hospital Affiliated to Shandong First Medical University (2022FY121).

**Authors' contributions**

PXQ: Performed experiments and Data analysis; ZZD: Manuscript writing and study design; ZHW: Revised the manuscript; WWH: Data interpretation.

**Acknowledgements**

Not Applicable.

**References**


Figures
Figure 1

The flowchart of this study.
Figure 2

Identification of ferroptosis and disulfidptosis-related genes in breast cancer. A: Volcano plot showed differentially expressed genes between normal and cancer tissues of BRCA in TCGA datasets. B: Heatmap showed differentially expressed ferroptosis and disulfidptosis genes in TCGA datasets.
Analysis of expression and mutation of PRGs in BRCA. A: The PPI network acquired from the STRING database among the PRGs. B: Genetic alteration on a query of PRGs. C: Circus plots of chromosome distributions of PRGs. D: Frequencies of CNV gain, loss, and non-CNV among PRGs.
Figure 4

Differences in OS and distribution of PRGs between two clusters. A: The network of correlation with PRGs in BRCA. B: The heatmap of consensus clustering (k=2) in BRCA. (C, D): Analysis of PCA (C) and Kaplan-Meier curve (D) among two clusters. E: Two clusters were identified by the PAGs with a heatmap and clinicopathological profiles.
Figure 5

Biological characteristics and TME cell infiltration in different subtypes. A, B: GO (A) and KEGG (B) enrichment via GSVA between two clusters in TCGA and GSE159956. C: The infiltration level of 23 immune cell types in two clusters. D, E: GO (D) and KEGG (E) analyses for biological functions and pathways. *P < 0.05, **P < 0.01, ***P < 0.001.
Figure 6

Identification of gene subgroups. A: 3 matrices from consensus clustering. B, C: Analysis of PCA (B) and Kaplan-Meier curve (C) among three clusters. D: Heatmap and clinicopathological features between three gene subgroups. E: The infiltration level of 23 immune cell types in three clusters. F: The expression level of PAGs in three gene subgroups. *P < 0.05, **P < 0.01, ***P < 0.001
Figure 7

PCA between the two groups of testing set. L: ROC curve for the survival prediction model in testing set. M: Alluvial diagram displaying the changes in PRG clusters, gene clusters, risk scores and survival outcomes. M, N: The difference in risk score among PAG cluster (M) and gene cluster (N).

Figure 8
The predictive value of the risk model. A, B: Multivariate (A) and univariate (B) COX analyses based on the entire set, respectively. C: The nomogram. D: Calibration curve for evaluating nomogram performance. E: The exploration of biological pathways based on GSEA in the high- and low-risk groups.

Figure 9
Characterization of immune cells infiltration in TME. A: Heatmap for immune responses based on the TIMER, CIBERSORT, CIBERSORT-ABS, QUANTISEQ, MCP-COUNTER, XCELL and EPIC algorithms in high-risk and low-risk groups. B, C: Relative percent (B) and fraction (C) of expression features of 22 immune cells in different groups using the CIBERSORT algorithm. D: The correlation between immune cell infiltration and risk score based on ssGSEA results. E: The relationship between risk model and immune checkpoints. ***P < 0.001, ****P < 0.0001.

Figure 10

Differential expression and prognosis of 6 hub genes. A–E: The different expression of 6 core genes in normal tissues and breast cancer tissues from public databases. G–L: RT-PCR validation of differences of 6 hub genes in normal tissues and breast cancer tissues. M–R: KM plotter was used to analyze the correlation of 6 core genes’ expression and overall survival of BRCA patients. S–X: Influence of 6 hub genes’ expression on overall survival of BRCA patients from GEO databases. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.
• AdditionalFile1Fig.S1.tif
• AdditionalFile2TableS1.xlsx
• AdditionalFile3TableS2.xlsx
• AdditionalFile4TableS3.xlsx
• AdditionalFile5TableS4.xlsx
• AdditionalFile6TableS5.xlsx
• AdditionalFile7TableS6.xlsx
• AdditionalFile8TableS7.xlsx
• AdditionalFile9TableS8.xlsx
• AdditionalFile10TableS9.xlsx
• AdditionalFile11TableS10.xlsx
• AdditionalFile12TableS11.xlsx
• AdditionalFile13Fig.S2.tif
• SupplementaryinformationLegends.docx