Development and Validation of a Prognostic Classifier Based on Lipid Metabolism-Related Genes for Breast Cancer

Nan Wang 1, Yuanting Gu 1, Lin Li 1, Jiangrui Chi 1, Xinwei Liu 1, Youyi Xiong 1, Chaochao Zhong 2

1Department of Breast Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, People’s Republic of China; 2Department of Plastic Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, People’s Republic of China

Correspondence: Nan Wang, Email fccwangn2@zzu.edu.cn

Background: The changes of lipid metabolism have been implicated in the development of many tumors, but its role in breast invasive carcinoma (BRCA) remains to be fully established. Here, we attempted to ascertain the prognostic value of lipid metabolism-related genes in BRCA.

Methods: We obtained RNA expression data and clinical information for BRCA and normal samples from public databases and downloaded a lipid metabolism-related gene set. Ingenuity Pathway Analysis (IPA) was applied to identify the potential pathways and functions of Differentially Expressed Genes (DEGs) related to lipid metabolism. Subsequently, univariate and multivariate Cox regression analyses were utilized to construct the prognostic gene signature. Functional enrichment analysis of prognostic genes was achieved by the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). Kaplan-Meier analysis, Receiver Operating Characteristic (ROC) curves, clinical follow-up results were employed to assess the prognostic potency. Potential compounds targeting prognostic genes were screened by Connectivity Map (CMap) database and a prognostic gene-drug interaction network was constructed using Comparative Toxicogenomics Database (CTD). Furthermore, we separately validated the selected marker genes in BRCA samples and human breast cancer cell lines (MCF-7, MDA-MB-231).

Results: IPA and functional enrichment analysis demonstrated that the 162 lipid metabolism-related DEGs we obtained were involved in many lipid metabolism and BRCA pathological signatures. The prognostic classifier we constructed comprising SDC1 and SORBS1 can serve as an independent prognostic marker for BRCA. CMap filtered 37 potential compounds against prognostic genes, of which 16 compounds could target both two prognostic genes were identified by CTD. The functions of the two prognostic genes in breast cancer cells were verified by cell function experiments.

Conclusion: Within this study, we identified a novel prognostic classifier based on two lipid metabolism-related genes: SDC1 and SORBS1. This result highlighted a new perspective on the metabolic exploration of BRCA.

Keywords: lipid metabolism, breast invasive carcinoma, BRCA, prognostic classifier, SDC1, SORBS1

Introduction

According to 2020 GLOBOCAN statistics of A Cancer Journal for Clinicians, breast invasive cancer (BRCA) is the most common type of cancer worldwide, comprising approximately 11.7% of new cancer cases. Despite advances in screening, research, and treatment, BRCA incidence continues to increase. BRCA continues to be a major public health problem globally. The risk factors for BRCA are complex and numerous, such as postpartum childbirth, obesity, metabolic syndrome, life stress, and lack of exercise. However, the correlation between these risk factors and the biological behavior of cancer cells has not yet been studied in depth, resulting in low value for the accurate prediction of the prognosis. Among these risk factors, the role of lipid metabolism disorders in the occurrence and development of BRCA has started to be appreciated. Establishing a more accurate model for predicting the prognosis of BRCA patients based on biomarkers related to BRCA lipid-metabolism is expected to become a valuable new research direction.
Changes in lipid metabolism and signal transduction have been recognized as one of the signs of abnormal cell growth and cancer progression. In addition to being a component of the cell membrane and a source of energy storage, lipids can also activate cell proliferation, differentiation, and migration, as well as induce cell death. Abnormal lipid metabolism has been reported to be related to the occurrence and development of a variety of malignant tumors, such as prostate cancer, ovarian cancer, pancreatic cancer, liver cancer, and BRCA. Studies have found that changes in lipid metabolism exist during the early stages of BRCA development, and a lipid-rich environment promotes the proliferation and migration of BRCA cells. Some bioinformatics studies have found that tumor gene sets and inflammation-related gene sets are highly expressed in adipocytes adjacent to BRCA, while histocytology studies have found that the lipid metabolism of residual cancer cells in patients undergoing neoadjuvant therapy for BRCA differs from that before treatment and in normal tissues. These findings suggest that genes related to lipid metabolism may become targets for BRCA treatment. Changes in lipid metabolism may also affect the target organ tropism of BRCA metastasis. It was found that the BRCA cells that are more prone to brain metastasis have significant changes in lipid metabolism. Although current studies have determined that abnormal lipid metabolism is related to BRCA progression and resistance to treatment, there is currently no specific diagnosis and treatment plan incorporating this information due to the complexity of the effects of lipid metabolism-related genes on BRCA.

The use of bioinformatics to analyze the targets of disease diagnosis and treatment has been widely used. In this study, we analyzed and obtained the biological processes of differential lipid metabolism genes with prognostic value in BRCA based on The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Molecular Signatures Database (MSigDB) databases. We identified two genes, SDC1 and SORBS1, which we screened, analyzed, and verified as biomarker genes that may play important roles in BRCA. In addition, we not only verified the expression changes of these two genes in BRCA samples and breast cancer cell lines (MCF7, MDA-MB-231), but also verified the functions of these two genes in breast cancer cells by cell function experiments. Finally, we constructed a prognostic classifier and preliminarily verified it with datasets in the GEO database and clinical follow-up results we collected. The validation results showed that our prognostic classifier has a clinical value and can provide a new reference for the diagnosis and treatment of BRCA patients. In addition, we used Connectivity Map (CMAP) and the Comparative Toxicogenomics Database (CTD) to investigate compounds targeting prognostic genes, and to provide new insights for the synthesis of new drugs.

**Materials and Methods**

**Cell Lines**

MDA-MB-231 and MCF-7 cells were obtained from ATCC (Shanghai, China) and cultured in DMEM (Invitrogen, Carlsbad, CA, USA) at 37°C under 5% CO2.

**Clinical Sample Acquisition**

Tumor tissue and normal breast tissue adjacent to the tumor were collected from 30 BRCA patients. All the patients were introduced to surgery at the First Affiliated Hospital of Zhengzhou University, Henan, China, from December 2020 to January 2021 and did not receive any anti-cancer treatment before surgery. All patients were between 18 and 80 years old and agreed to accept the postoperative follow-up plan. Tissue specimens were collected within 30 min after surgery and quickly frozen in liquid nitrogen. Postoperative monitoring and treatment continued in accordance with relevant consensus guidelines. The degree of tumor differentiation was graded according to the WHO grading system. Overall survival (OS) was defined as the time between surgery and death. The study was approved by the Ethics Committee of the First Affiliated Hospital of Zhengzhou University and conducted according to the principles expressed in the Declaration of Helsinki. Written informed consent was obtained from each patient prior to collection of any samples. Detailed information of each patient, including serial number, sampling time, pathology type, ER positivity rate, HER2 status, Ki-67 expression, lymph node stage, tumor size, age, and side of affected breast, is available in Supplementary Table 1.

In order to verify the clinical applicability of our prognostic model, we collected the clinical data and pathological sections of 50 patients with BRCA who underwent surgery in our hospital from January 2015 to March 2015. The plan
for clinical patient follow-up and collection of patient clinical data to verify the prognosis model was approved by the Ethics Committee of the First Affiliated Hospital of Zhengzhou University. We followed up by phone or in-person to collect the survival information of the 50 patients. All patients were between 18 and 80 years old and agreed to accept the post-operative follow-up plan. The personal information of each patient is strictly confidential, and as such, written informed consent was obtained from each patient or its entrusted. The clinical conditions and follow-up results of the patients are shown in Supplementary Table 2.

Data Sources
We acquired RNA expression data from 1171 samples (1072 breast cancer samples and 99 normal samples) from TCGA database. Of the 1072 BRCA samples, only 1069 BRCA samples recorded complete survival information, so candidate prognosis-related lipid metabolism differentially expressed genes (DEGs) were identified based on these samples. In addition, only 881 of the 1069 BRCA samples had complete clinical information, which was needed for the sample to be used in the construction of the prognostic classifier. The patients’ characteristics were presented in Supplementary Table 3.

The GSE109169 dataset28 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE109169) was downloaded from the GEO database (25 BRCA samples, 25 normal samples) and was utilized to legitimize the expression pattern of candidate prognosis-related lipid metabolism DEGs. The patients’ characteristics were presented in Supplementary Table 4. Furthermore, we downloaded the GSE2068529 dataset from the GEO database, which was utilized as independent external validation set to authenticate the prognostic efficacy of the multigene signature. The GSE20685 dataset contains 327 BRCA samples with complete survival information and matched transcriptomic data (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20685). The characteristics of the BRCA samples in the GSE20685 dataset were available in Supplementary Table 5. Moreover, 1499 genes related to lipid metabolism were obtained by de-duplication from 13 gene sets related to lipid metabolism, which were downloaded from the Gene Ontology (GO) and the Molecular Signatures Database (MSigDB) websites. These 13 gene sets were GO_GALACTOLIPID_METABOLIC_PROCESS, GO_GLYCEROLIPID_METABOLIC_PROCESS, GO_GLYCEROPHOSPHOLIPID_METABOLIC_PROCESS, GO_GLYCOSPHINGOLIPID_METABOLIC_PROCESS, GO_LIPID_METABOLIC_PROCESS, GO_LIPROTEIN_METABOLIC_PROCESS, GO_MEMBRANE_LIPID_METABOLIC_PROCESS, GO_NEUTRAL_LIPID_METABOLIC_PROCESS, GO_REGULATION_OF_LIPID_METABOLIC_PROCESS, GO_REGULATION_OF_LIPOPROTEIN_METABOLIC_PROCESS, GO_REGULATION_OF_PHOSPHOLIPID_METABOLIC_PROCESS, GO_SPHINGOLIPID_METABOLIC_PROCESS. All lipid metabolism-related genes are integrated into Supplementary Table 6.

BRCA-FPKM data were obtained from the TCGA database and log2 transformed for the gene expression files before use; the expression profiles acquired from the GEO database were log2 transformed. Moreover, we used the normalize Between Arrays function of the limma package to normalize the gene expression profiles of GSE109169 and GSE20685 datasets.

Translation level validation of identified prognostic genes was performed using the Human Protein Atlas30 database (http://www.proteinatlas.org/), an online database containing immunohistochemical expression data for approximately 20 of the most common cancers.

Differential Analysis
Transcriptome RNA-seq data of 1171 cases (normal samples, 99 cases; BRCA samples, 1072 cases) were downloaded from TCGA database. R package limma31 was used to perform differentiation analysis of the gene expression, and DEGs were generated by the comparison between the BRCA samples vs the normal samples (Supplementary Table 7). DEGs with |log2 fold change | > 1 and adjusted $P < 0.05$ were considered significant.

Ingenuity Pathway Analysis (IPA)
Overlapping genes (Supplementary Table 8) for DEGs and lipid metabolism-related genes were analyzed by the IPA software (version 1–19-00). RNA sequencing results were imported into IPA, and the enrichment status of canonical pathways and diseases and functions were assessed by $P$-values32,33. Z scores were indicated by orange or blue to denote activation or inhibition of pathways. In IPA, terms with $P < 0.05$ were considered significantly enriched; furthermore, in
IPA-Canonical pathway analysis, the activation and inhibition of the enriched pathway were assessed by Z-score, specifically, Z-score > 0 and \( P < 0.05 \) indicated that the pathway was active, while Z-score < 0 and \( P < 0.05 \), then the pathway was considered to be in the inhibited state.

**Identification of Candidate Prognosis-Related Lipid Metabolism DEGs**

Lipid metabolism DEGs were included in the univariate Cox regression analysis, and genes with a \( P < 0.05 \) were selected to be used in the multivariate Cox regression analysis. Similarly, genes with a \( P < 0.05 \) were considered as candidate prognosis-related lipid metabolism DEGs. Subsequently, GO and KEGG analyses were carried out based on the above genes using the R software clusterProfiler package. GO analysis consisted of three components: biological process (BP), molecular function (MF) and cellular component (CC). Entries with an adjusted (adj.) \( P < 0.05 \) were considered significantly enriched.

**Knock-Down and Overexpression Experiments of SDC1 and SORBS1**

The empty plasmids used for constructing controls of SDC1 gene overexpression and SORBS1 gene overexpression (oe-SDC1 control and oe-SORBS1 control), the non-targeting plasmid used for constructing controls of SDC1 gene knockdown and SORBS1 gene knockdown (si-SDC1 control and si-SORBS1 control), SDC1 knockdown plasmids, SORBS1 knockdown plasmids, SDC1 overexpression plasmids and SORBS1 overexpression plasmids were designed and constructed by Genechem (Shanghai, China). The sequences of the plasmids used in this study are provided in Supplementary Table 9. According to the manufacturer’s instruction, cells of different groups were transfected with these plasmids using Lipofectamine 2000. 72 hours after transfections, the effects of knockdown or overexpression were examined by qRT-PCR and Western blot.

**RNA Isolation, Reverse Transcription, and Real-Time PCR**

Total RNA was isolated from 30 paired tissues and 8 groups of experimental cells using TriQuick Reagent (Solarbio, Shanghai, China) according to the manufacturer’s instructions. Then the concentration and purity of the RNA solution was quantified using a NanoDrop 2000 nucleic acid protein quantifier (Thermo Fisher Scientific, Waltham, MA, USA). qRT-PCR was performed as described previously.\(^{34}\) Briefly, the extracted RNA was reverse-transcribed to cDNA using the FastQuant RT Kit with gDNA eraser (TIANGEN, Beijing, China) prior to qRT-PCR. The reverse transcription reaction consisted of 4 µL of 5x reaction buffer, 1 µL of oligo (dt) primer, 1 µL of random hexamer primer, 1 µL of Servicebio RT Enzyme Mix, and 0.1 ng–5 µg total RNA, and finally the reaction was brought to 20 µL using RNase-free water. The qRT-PCR reaction consisted of 2 µL of reverse transcription product, 10 µL of 2X SYBRGreen qPCR Master Mix (High ROX, Servicebio, Wuhan, China), 0.4 µL each of forward and reverse primer, and 7.2-µL nuclease-free water. PCR was performed in a MiniAmp Thermal Cycler (A37834, Thermo Fisher Scientific, Waltham, MA, USA) under the following conditions: 95°C for 3 s, followed by 40 cycles of 95°C for 15s and 60°C for 30s. The GAPDH gene served as an internal control. 2-ΔΔTct method was used to calculate the RNA levels of tumor samples, paired adjacent samples and 8 groups of experimental cells. Primer sequences used for qRT-PCR are shown in Table 1.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer</th>
<th>Sequence (5’→3’)</th>
<th>Length</th>
<th>Tm</th>
<th>GC %</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDC1</td>
<td>Forward</td>
<td>ACAGGCTATTCCACGTCTC</td>
<td>19</td>
<td>54.58</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>ATGGGCGAGACTCACGCTC</td>
<td>20</td>
<td>56.54</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Forward</td>
<td>CCAATCGAGAAGACGAGAAACG</td>
<td>23</td>
<td>56.05</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>ACCCGGCCCTACTGTCACTCTT</td>
<td>21</td>
<td>56.95</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>Forward</td>
<td>CTGGGCTACACTTGAAGGCA</td>
<td>19</td>
<td>55.46</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>AAGTGGTAGTCCCTT</td>
<td>21</td>
<td>57.03</td>
<td>52</td>
</tr>
<tr>
<td>SORBS1</td>
<td>Forward</td>
<td>ACGGCTATTCCACGTCTC</td>
<td>19</td>
<td>54.58</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>ATGGGCGAGACTCACGCTC</td>
<td>20</td>
<td>56.54</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Forward</td>
<td>CCAATCGAGAAGACGAGAAACG</td>
<td>23</td>
<td>56.05</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>ACCCGGCCCTACTGTCACTCTT</td>
<td>21</td>
<td>56.95</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>Forward</td>
<td>CTGGGCTACACTTGAAGGCA</td>
<td>19</td>
<td>55.46</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>AAGTGGTAGTCCCTT</td>
<td>21</td>
<td>57.03</td>
<td>52</td>
</tr>
</tbody>
</table>

Table 1 Primer Sequences Used for qRT-PCR
Western Blot Analysis

Western blot analysis was conducted using an SDS-PAGE electrophoresis system (Bio-Rad Laboratories, Hercules, CA, USA). Briefly, the total protein content was extracted from tissue or 8 groups of experimental cells using a RIPA buffer (P0013C, Beyotime, Shanghai, China). Protein samples were separated by SDS-PAGE and transferred onto nitrocellulose membranes (88,520, Thermo Fisher Scientific, Waltham, MA, USA), which were subsequently blocked 12 h at 4°C with 5% skimmed milk containing TBST (Tris-buffered salt solution, containing 50 mmol/L Tris-HCl, 150 mmol/L NaCl, 0.1% v/v Tween-20, pH 7.4) solution. Antibodies against SDC1 (ab128936; Abcam, Cambridge, MA, USA), SORBS1 (ab224129; Abcam, Cambridge, MA, USA), and GAPDH (T0004; Affinity Biologicals, Cincinnati, OH, USA) were used as primary antibodies. The samples were incubated with horseradish peroxidase–conjugated secondary antibodies at 37°C for 1 h. The membrane was imaged using an Amersham Imager 600 (GE Healthcare UK Limited, Little Chalfont, UK).

Immunohistochemistry

In the present study, we performed immunohistochemical analysis on 30 BRCA tissue samples and their paired cancer-adjacent normal breast tissue samples. The tissues were collected and fixed in 4% paraformaldehyde, embedded in paraffin, and sectioned at 6-μm intervals. The sections were placed in EDTA (pH 9.0) for antigen repair, washed with PBS (pH 7.4), treated with 3% H2O2, blocked with goat serum, and then incubated 12 h at 4°C with SORB1 (1:100; ab224129, Abcam, Cambridge, MA, USA) and SDC1 (1:1600; 10593-1-AP, Proteintech, Rosemont, IL, USA). Subsequently, the secondary antibody goat anti-rabbit IgG (1:200; GB23303, Servicebio) was incubated for 1 h at 37°C, and the positive sites were labelled with a DAB (diaminobenzidine) color development solution (G1211, Servicebio, Wuhan, China). Finally, hematoxylin staining (G1004, Servicebio, Wuhan, China) was performed to visualize the nuclei. Images (200×) were captured with a microscope (XSP-C204, COIC, Chongqing, China), and three different visual fields were analyzed. Image-Pro Plus 6 processing software (Media Cybernetics, Inc., Rockville, MD, USA) was used to analyzed results and to calculate average optical density, \[ AOD = \frac{\text{integrated OD}}{\text{measurement area}}. \] The higher the AOD, the higher the expression of the protein.37,38 The person imaging the slides was uninformed during the experiment and analysis.

Invasion Assays

Cell invasion assay was done performed according to the manufacturer’s instructions. Briefly, the upper surface of a Transwell (8-μm pore size; Corning, Corning, NY, USA) was coated with 20μL diluted Matrigel (BD Biosciences, Franklin Lakes, NJ, USA). Equal numbers of the indicated cells were seeded in the upper chamber of the Transwell in serum-free medium, with FBS added to the lower chamber. After incubation for 48 hours, non-invading cells were removed with a cotton swab, and the invaded cells were stained with crystal violet. Average invaded cell number per field of view was obtained from 5 random fields.

Scratch Assays

Cells were plated in 24-well cell culture plates (1 × 10^5 cells/well) followed by overnight incubation. When culture was confluent (24 h post transfection), a p-200 pipette tip was used to score two vertical lines and one horizontal line (average width: 700–900 μm) simulating a “wound” by scratching the culture. All the wells were washed with sterile PBS twice and reference markings were drawn near the scratch area from the bottom side of the plate with a fine tip marker. Scratch images were captured within the marked area using an inverted microscope (Laxco, Mill Creek, WA, USA) with a 10X magnification objective piece.

Prediction Model

We randomly divided TCGA database of 881 BRCA samples containing complete clinical and survival information into a training set (n = 617) and a testing set (n = 264) based on a 7:3 ratio. The characteristics between the BRCA samples in the TCGA-training and -testing sets were not significantly different (Supplementary Table 10), indicating that they could be used for subsequent analysis. In the training set, univariate and multivariate Cox regression analyses were applied to extract the
optimal prognosis-related lipid metabolism DEGs based on selected candidate prognosis-related lipid metabolism DEGs. Ultimately, genes with a $P < 0.05$ were used to construct a prognostic gene signature. Prognosis-related lipid metabolism DEGs were visualized on the forest plots. Risk scores were simultaneously calculated for each patient in the training set, testing set, and external validation set of the following equations: risk score = (expression of Gene$_1$ × $\beta_1$Gene$_1$) + (expression of Gene$_2$ × $\beta_2$Gene$_2$) + (… expression of Gene$_n$ × $\beta_n$Gene$_n$). $\beta$ was represented by the regression coefficient, which was generated by the step multivariate Cox proportional hazards regression model. Based on this score, patients were classified into high- and low-risk score groups on the basis of the median classification method.

**Prognostic Analysis**
Risk curves were plotted with the pheatmap package. The R survival package was employed to assess the relationship between lipid metabolism-related gene signature and OS ($P < 0.05$) and to plot Kaplan-Meier (K-M) curves. Subsequently, the survival ROC package was executed to create ROC curves and calculate AUC in order to estimate the predictive accuracy of the prognostic gene signature. The above analyses were performed simultaneously in the training set, testing set, and the external validation set (GSE20685 dataset). Univariate and multivariate independent prognostic analyses were developed by survival software packages to confirm whether risk scores could be applied as the independent prognostic indicator. The rms and survival packages of R were performed to construct a nomogram. Subsequently, calibration curves were plotted to evaluate the agreement between actual and predicted survival. Decision curve analysis (DCA) was carried out to calculate the net clinical benefit for each model. The optimal model was the one with the greatest net benefit calculated.

**Connectivity Map (CMap) Analysis**
CMap is designed to reveal the relationships among genes, compounds, and biological conditions, which is a systematic, data-driven process. We resorted to CMap02 in order to screen potential compounds that might target prognostic genes. Compounds with $|\text{enrichment score}| \geq 0.5$ and a $P < 0.05$ were selected as potential therapeutic drugs for BRCA. The compounds were then further filtered by CMap’s CLUE tool (https://clue.io/) for related Mechanism of Actions (MoA) and their inhibitors in order to investigate their joint intrinsic mechanism of action.

**Predicting Prognostic Gene-Chemotherapy Drug Interaction Network by CTD**
A network of interactions between prognostic genes and chemotherapeutic agents was constructed using the Comparative Toxicogenomics Database (CTD) to obtain chemotherapeutic agents that could reduce or increase prognostic gene expression levels. Briefly, chemotherapeutic drugs for SORBS1 and SDC1 were searched for in the CTD, and the prognostic gene-drug interaction networks were visualized by using Cytoscape 3.8.2 (https://cytoscape.org/).

**Statistical Analysis**
The Jvenn website (http://jvenn.toulouse.inra.fr/app/example.html) was used to produce a Venn diagram for intersection analysis. A K–M curve with Log rank test was used to assess the OS differences between different groups. An ANOVA test was performed to detect the association of the risk score with clinical characteristics, and also, to reveal the differences in the levels of risk scores within different subtypes of clinical characteristics. The statistical analyses were conducted using the R software. A $P < 0.05$ indicated statistically significant differences.

**Model Validation by Detailed Data from 2015.01 to 2015.03**
After collecting clinicopathological and follow-up data and conducting immunohistochemistry staining, correlations between SDC1 and SORBS1 expression and survival prognosis were analyzed.

**Results**
**Identification of Differentially Expressed Lipid Metabolism-Related Genes**
RNA sequencing data of 1072 BRCA samples and 99 normal samples were extracted from TCGA database. A differential analysis based on the R package limma was performed with normal samples as controls in order to screen
for genes that were aberrantly expressed in BRCA, and genes satisfying $|\log_2 \text{FC}| > 1$ and adjusted $P < 0.05$ were identified as DEGs. We identified a total of 1732 BRCA-related DEGs, of which 644 genes were upregulated and 1088 were downregulated in the BRCA group compared with the normal group (Figure 1A; Supplementary Table 7). Subsequently, to recognize lipid metabolism-related DEGs, we performed intersection analysis based on a list of 1732 DEGs and 1499 lipid metabolism-related genes. The results were shown in Figure 1B, and a total of 162 overlapping genes were identified, which were defined as lipid metabolism-related DEGs (Supplementary Table 8). Further, the heatmap demonstrated the expression pattern of these lipid metabolism-related DEGs between 1072 BRCA and 99 normal samples in the TCGA database (Figure 1C), including 34 up-regulated genes and 128 down-regulated genes (BRCA vs normal). Canonical pathway analysis of these lipid metabolism-related DEGs in IPA revealed that several pathways related to lipid metabolism (eg, fatty acid β-oxidation I, fatty acid α-oxidation, and fatty acid activation), carcinogenesis (eg, AMPK signaling, JAK/STAT signaling, and ERK/MAPK signaling), and immune response (eg, MIF regulation of innate immunity, IL-7 signaling pathway, and glioblastoma multiforme signaling) were inhibited. Furthermore, we demonstrated that these genes were correlated with the negative regulation of estrogen-dependent BRCA signaling, HER2 signaling in BRCA, and BRCA regulation by stathmin1 (Supplementary Figure 1A; Supplementary Table 11). The diseases and functions analysis confirmed that lipid metabolism and small molecule biochemistry were the most abundant pathways (Supplementary Figure 1B; Supplementary Table 12).

Functional Enrichment Analysis of Candidate Prognosis-Related Lipid Metabolism DEGs

In the present analysis, we included 1069 BRCA samples that contained complete survival information. Univariate Cox analysis screened 76 genes with $P < 0.05$ from the identified 162 lipid metabolism-related DEGs (Table 2). Thirteen candidate prognosis-related lipid metabolism genes were obtained from the subsequent multivariate Cox analysis. Through examination with a forest plot, genes with HR > 1 were found to include LEPR, FGF2, GPAM, SDC1, CCDC3, PCK2, PLTP and HSD17B13, which may be risk factors for BRCA prognosis; while ABCD2, FABP4, SORBS1, FAM126A and GP1HBP1 with HR < 1 could be termed as protective factors for BRCA prognosis (Figure 2A). We then validated the expression patterns of the 13 genes mentioned above.
using the GSE109169 dataset. The results showed that the expression trends of these genes were consistent with those in the TCGA database. Except for SDC1, the expression of the other 12 genes was downregulated in the BRCA samples (Figure 2B and C). The underlying features of these genes were ascertained by GO and KEGG pathway analysis (Supplementary Figure 2). Undoubtedly, several GO terms related to lipid metabolism were found in biological processes, such as “positive regulation of lipid metabolic process,” “regulation of lipid biosynthetic process,” “regulation of lipid metabolic process,” “positive regulation of lipid biosynthetic process,” “glycerolipid metabolic process,” and “triglyceride metabolic process” (Supplementary Table 13).

### Table 2: Univariate COX Analysis of Different Lipid Metabolism Genes (P < 0.05)

<table>
<thead>
<tr>
<th>ID</th>
<th>HR</th>
<th>HR.95L</th>
<th>HR.95H</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEPR</td>
<td>1.08005</td>
<td>1.0446</td>
<td>1.11663</td>
<td>6.04E-06</td>
</tr>
<tr>
<td>ADM</td>
<td>1.02309</td>
<td>1.012557</td>
<td>1.033609</td>
<td>1.44E-05</td>
</tr>
<tr>
<td>ADH1C</td>
<td>1.05337</td>
<td>1.025811</td>
<td>1.081671</td>
<td>0.000121</td>
</tr>
<tr>
<td>CCDC3</td>
<td>1.01049</td>
<td>1.005017</td>
<td>1.01601</td>
<td>0.000168</td>
</tr>
<tr>
<td>PLTP</td>
<td>1.00573</td>
<td>1.002641</td>
<td>1.008835</td>
<td>0.000274</td>
</tr>
<tr>
<td>PKK1</td>
<td>1.02409</td>
<td>1.010228</td>
<td>1.038143</td>
<td>0.000783</td>
</tr>
<tr>
<td>FGF2</td>
<td>1.03733</td>
<td>1.015572</td>
<td>1.059562</td>
<td>0.000943</td>
</tr>
<tr>
<td>GOS2</td>
<td>1.00482</td>
<td>1.000617</td>
<td>1.002347</td>
<td>0.001225</td>
</tr>
<tr>
<td>MGLL</td>
<td>1.01302</td>
<td>1.005283</td>
<td>1.020816</td>
<td>0.001225</td>
</tr>
<tr>
<td>HCAR2</td>
<td>1.04276</td>
<td>1.016626</td>
<td>1.069574</td>
<td>0.001225</td>
</tr>
<tr>
<td>ACACB</td>
<td>1.01582</td>
<td>1.00608</td>
<td>1.025671</td>
<td>0.001412</td>
</tr>
<tr>
<td>CD36</td>
<td>1.00284</td>
<td>1.001089</td>
<td>1.004609</td>
<td>0.001494</td>
</tr>
<tr>
<td>KLF4</td>
<td>1.00852</td>
<td>1.003137</td>
<td>1.013405</td>
<td>0.001546</td>
</tr>
<tr>
<td>ADH1B</td>
<td>1.00186</td>
<td>1.000613</td>
<td>1.002761</td>
<td>0.00207</td>
</tr>
<tr>
<td>DGAT2</td>
<td>1.00737</td>
<td>1.00267</td>
<td>1.012108</td>
<td>0.00209</td>
</tr>
<tr>
<td>PTGIS</td>
<td>1.02536</td>
<td>1.009068</td>
<td>1.041933</td>
<td>0.002183</td>
</tr>
<tr>
<td>LGALS12</td>
<td>1.01164</td>
<td>1.004028</td>
<td>1.019325</td>
<td>0.00268</td>
</tr>
<tr>
<td>CAV1</td>
<td>1.00249</td>
<td>1.000862</td>
<td>1.004126</td>
<td>0.002729</td>
</tr>
<tr>
<td>LRPI</td>
<td>1.00787</td>
<td>1.002627</td>
<td>1.01315</td>
<td>0.00321</td>
</tr>
<tr>
<td>TEK</td>
<td>1.05472</td>
<td>1.017956</td>
<td>1.092815</td>
<td>0.00325</td>
</tr>
<tr>
<td>ENPP2</td>
<td>1.01509</td>
<td>1.00498</td>
<td>1.025306</td>
<td>0.00363</td>
</tr>
<tr>
<td>RETSAT</td>
<td>1.00534</td>
<td>1.001763</td>
<td>1.00894</td>
<td>0.003423</td>
</tr>
<tr>
<td>ANXA1</td>
<td>1.00369</td>
<td>1.001071</td>
<td>1.005673</td>
<td>0.004047</td>
</tr>
<tr>
<td>GAM</td>
<td>1.00501</td>
<td>1.001526</td>
<td>1.008489</td>
<td>0.004761</td>
</tr>
<tr>
<td>PLIN1</td>
<td>1.00121</td>
<td>1.00037</td>
<td>1.002057</td>
<td>0.004782</td>
</tr>
<tr>
<td>HSDL17B13</td>
<td>1.09067</td>
<td>1.026844</td>
<td>1.159093</td>
<td>0.004846</td>
</tr>
<tr>
<td>ADIPOQ</td>
<td>1.00146</td>
<td>1.000426</td>
<td>1.002506</td>
<td>0.00571</td>
</tr>
<tr>
<td>ACSL1</td>
<td>1.00275</td>
<td>1.000788</td>
<td>1.004715</td>
<td>0.005984</td>
</tr>
<tr>
<td>BMX</td>
<td>1.16895</td>
<td>1.045744</td>
<td>1.30666</td>
<td>0.006012</td>
</tr>
<tr>
<td>CYP2U1</td>
<td>1.15477</td>
<td>1.042081</td>
<td>1.279664</td>
<td>0.00602</td>
</tr>
<tr>
<td>LIPE</td>
<td>1.00391</td>
<td>1.001119</td>
<td>1.006724</td>
<td>0.006048</td>
</tr>
<tr>
<td>AADAC</td>
<td>1.07426</td>
<td>1.020586</td>
<td>1.130726</td>
<td>0.00615</td>
</tr>
<tr>
<td>LPL</td>
<td>1.00133</td>
<td>1.000377</td>
<td>1.002294</td>
<td>0.006318</td>
</tr>
<tr>
<td>VAV3</td>
<td>0.98945</td>
<td>0.981893</td>
<td>0.997073</td>
<td>0.006749</td>
</tr>
<tr>
<td>FABP4</td>
<td>1.00029</td>
<td>1.000082</td>
<td>1.000517</td>
<td>0.007006</td>
</tr>
<tr>
<td>ANGPTL4</td>
<td>1.01123</td>
<td>1.003002</td>
<td>1.019543</td>
<td>0.007398</td>
</tr>
<tr>
<td>PDE3B</td>
<td>1.03854</td>
<td>1.010131</td>
<td>1.067755</td>
<td>0.007336</td>
</tr>
<tr>
<td>MLX1PL</td>
<td>1.04972</td>
<td>1.012711</td>
<td>1.088168</td>
<td>0.008075</td>
</tr>
<tr>
<td>CIDEA</td>
<td>1.00631</td>
<td>1.001635</td>
<td>1.011008</td>
<td>0.008109</td>
</tr>
<tr>
<td>EGR1</td>
<td>1.00075</td>
<td>1.00019</td>
<td>1.001317</td>
<td>0.00872</td>
</tr>
<tr>
<td>QKI</td>
<td>1.05652</td>
<td>1.013967</td>
<td>1.100873</td>
<td>0.008763</td>
</tr>
<tr>
<td>PDGFRA</td>
<td>1.01727</td>
<td>1.004373</td>
<td>1.031259</td>
<td>0.00912</td>
</tr>
<tr>
<td>CEBPA</td>
<td>1.01291</td>
<td>1.002986</td>
<td>1.02148</td>
<td>0.009333</td>
</tr>
</tbody>
</table>

Abbreviations: COX analysis, Cox proportional hazards model; HR, hazard ratio.
Furthermore, “lipid droplet” and “external side of plasma membrane” were the most remarkably enriched GO items of CC (Supplementary Figure 2A). As implied by the KEGG pathway analysis (Supplementary Figure 2B), candidate prognosis-related lipid metabolism genes were primarily enriched along two pathways—The proliferator-activated receptor (PPAR) signaling pathway and adipocytokine signaling pathway (Supplementary Table 14).

Filtering Candidate Prognosis-Related Lipid Metabolism DEGs for Prognostic Prediction

Two prognosis-related genes, **SDC1** (HR = 1.2638, \( P = 0.03076 \)) and **SORBS1** (HR = 0.7628, \( P = 0.03576 \)), were retained based on the univariate Cox regression analysis in the training set (Figure 3A). Subsequently, we proceeded to risk model gene selection utilizing multivariate Cox regression (Figure 3B). The results indicated that **SDC1** (\( P = 0.037 \)) and **SORBS1** (\( P = 0.043 \)) were the optimal variables for the construction of prognostic signature. According to the hazard ratio (HR), **SDC1** might be a promoter of BRCA (HR = 1.244, 95% CI: 1.014–1.523) and **SORBS1** was a promising tumor suppressor gene (HR = 0.768, 95% CI: 0.595–0.991). Subsequently, based on the TCGA-training set, we revealed the association of the two identified prognostic genes with OS in BRCA patients by K-M curves. The relatively short OS of high expression of **SDC1**-BRCA compared to low expression of **SDC1**-BRCA (Figure 3C; \( P = 0.048 \)); and the expression of **SORBS1** was positively associated with OS in BRCA patients, with high expression of **SORBS1** implying longer OS (Figure 3D; \( P = 0.0096 \)), which evidence corroborated the inferences made based on HR values. Next, the expression values of prognostic genes and multivariate Cox regression coefficients (Supplementary Table 15) were utilized to generate risk scores for individual samples in the training set. Moreover, IPA exposed a sophisticated network of interactions involving the two prognostic genes (Supplementary Figures 1C–E).

The 2-gene based prognostic signature was assessed in the training set. Depending on the training set-median risk score, the 617 samples were divided into high (n = 308) and low (n = 309) risk groups, and the risk score curves and survival status of all samples were presented in Figure 3E. K-M survival analysis showed that the risk score based on the 2-gene prognostic signature could significantly discriminate the clinical outcomes of BRCA patients, with a low-risk score implying a better prognosis (\( P = 0.003 \); Figure 3F). Subsequently, ROC curve analysis showed that the risk score predicted the AUC of 1, 3, and 5-year OS for training set-BRCA patients to be 0.694, 0.681, and 0.618, respectively (Figure 3G). Further, we evaluated the general applicability of the risk score system in the testing set and external
The performance of risk score in the TCGA-testing set (n = 264) and the independent external validation set (GSE20685 dataset, n = 327) was similar to that in the TCGA-training set. High-risk scores were significantly associated with poor prognosis (all P < 0.05). The AUC of the ROC curve at 1, 3, and 5 years was 0.635, 0.628, and 0.611 in the testing set, and the AUCs in the GSE20685 dataset were 0.819, 0.672, and 0.650, respectively. This evidence suggested that the SDC1 and SORBS1-based prognostic signature possessed robust prognostic predictive power. Additionally, heat maps were constructed to visualize the expression of prognostic genes in each cohort (Supplementary Figure 3A–C), with SDC1 being relatively highly expressed in the high-risk group and SORBS1 being relatively overexpressed in the low-risk group. Moreover, immunohistochemical results, pathological results, and follow-up results of 50 patients (Supplementary Table 16) collected from January 2015 to March 2015 were divided into high-risk and low-risk groups according to this prognostic model, and the difference in the survival curves between the two groups was found to be statistically significant by survival analysis (P = 2.516e-02) (Supplementary Figure 3D).
The Obtained Two-Gene Signature Was Linked to Pathological Features of BRCA

We further investigated whether the genetic signature was implicated in the pathological features of BRCA. Specifically, in the AJCC pathologic T-stage subgroup (Figure 4A), risk score levels were significantly higher in the T2/T4 subtype than in the T1 subtype; interestingly, risk score levels were markedly lower in the T3 subtype compared with the T2 subtype; compared with the T3 subtype, risk score levels were remarkably higher in the T4 subtype; however, risk score levels were comparable between the T1 and T3 and T2 and T4 subtypes. In the AJCC pathologic N-stage subgroup (Figure 4B), risk score levels were increased considerably in the N1 and N2 subtypes relative to the N0 subtype; however, risk score levels were not statistically different between the N0 and N3, N1 and N2/N3, and N2 and N3 subtypes. In the AJCC pathologic stage subgroup (Figure 4C), the risk score levels were proportional to the stage level, with the higher the stage level, the higher the risk score level; however, there was no significant difference in the risk score levels among stage II, stage III, and stage IV. We also assessed the relationship between different types of BRCA and risk scores (Figure 4D). Of the 4 types of BRCA, HER2-enrich BRCA had the highest risk score, followed by luminal B, basal-like, and luminal A BRCA. Such results suggested a substantial interaction between lipid metabolism genes and clinical molecular features.

Figure 4 Scatter dot plot shows the association between risk score of TCGA breast cancer samples and clinical characteristics. (A) Risk scores in the different pathologic T stages of BC. (B) Risk scores in the different pathologic N stages of BC. (C) Risk scores in the different pathologic stages of BC. (D) Risk scores of different BC subtypes. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

Abbreviations: BC: breast cancer; TCGA: The Cancer Genome Atlas.
Furthermore, we performed stratified survival analysis in the TCGA-training set, TCGA-test set, and independent external validation set, respectively, to explore whether the prognostic value of the risk score was applicable to other clinical factors. All patients were divided into designated subgroups based on clinical characteristics in the corresponding datasets. BRCA samples in the TCGA-training set and test set were analyzed stratified according to age, gender, pathologic T-stage, pathologic N-stage, pathologic M-stage, and pathologic tumor stage. In the TCGA-training set, the 2-gene signature was useful in >60 years, female, T2, T3, N2, N3, M0, and stage III subgroups, with clinically and statistically significant prognostic value (Supplementary Figure 4). In the TCGA-test set, the prognostic model was able to significantly differentiate the clinical outcomes of patients in the <60 years, female, N3, M1, and stage IV subgroups (Supplementary Figure 5). BRCA samples in the independent external validation set were analyzed stratified according to age, gender, metastatic events, regional recurrence or not, adjuvant chemotherapy, pathologic N stage, pathologic M stage, pathologic tumor stage, and subtype. In the independent external validation set, the 2-gene signature had clinically and statistically significant prognostic value in the <60, female, no regional recurrence, N0, and type I subgroups (Supplementary Figure 6).

The Independent Utility of the Prognostic Classifier
First, univariate Cox regression analysis indicated that age, tumor stage, AJCC pathologic stage, AJCC pathologic TNM stage, and risk score were correlated with OS (Figure 5A). Subsequent multivariate Cox regression analysis suggested that age, AJCC pathologic stage, and risk score could serve as independent prognostic factors (Figure 5B). A nomogram was derived based on the Cox regression analysis results that employed four independent prognostic factors to predict the prognosis of patients with BRCA in TCGA database (Figure 5C). The concordance index of the nomogram was 0.739. The calibration curves suggested that the nomogram (combined model) was possibly under- or overestimating the mortality (Figure 5D). Decision curve analysis (DCA) evidenced that the combined model exhibited the best net returns for the 5-year OS as opposed to 3-year OS (Figures 5E and F). Regrettably, we were unable to capture the net returns of the combined model on 1-year OS, possibly owing to the inadequacy of the 1-year OS data. Collectively, these results indicated that the nomogram constructed using a combined model may be the optimal nomogram in predicting long-term survival (5 years) for patients with BRCA as compared to the nomogram obtained using individual prognostic factors, which may assist in clinical management.

Validation of Prognostic Genes Based on Clinical Samples
We performed qRT-PCR analysis to assess the expression levels of the two genes used to construct our prognostic model. Consistent with the results of the biochemical analysis, SORBS1 was significantly downregulated in the tumor samples as compared to the adjacent healthy samples (Figure 6A), while SDC1 was markedly overexpressed in the tumor samples (Figure 6B). Furthermore, we performed Western blot analysis and found that protein expression of SORBS1 was downregulated in BRCA (Figures 6C), and the protein expression level of SDC1 was upregulated in BRCA (Figure 6D). The results of immunohistochemistry showed positive staining for SORBS1 and SDC1 proteins in a brown or tan color, which revealed the localization of SORBS1 and SDC1 to be in the cytoplasm/membrane of BRCA tissue and adjacent normal breast tissue, respectively (Figure 6E and F). The AOD of SORBS1 was higher in the adjacent healthy group as compared to the BRCA group (Figure 6G; \( P < 0.05 \)). The AOD of SDC1 protein was significantly higher in the BRCA group than in the adjacent healthy group (Figure 6H; \( P < 0.05 \)). Besides, we explored the protein expression levels of these prognostic genes in BRCA using the HPA database. Consistently, protein levels of SDC1 were not expressed in normal breast tissue, while moderate expression levels of this gene were observed in breast cancer tissue (Supplementary Figure 7A); meanwhile, moderate protein expression levels of SORBS1 were observed in normal breast tissue, while low protein expression levels of these genes were observed in breast cancer tissue (Supplementary Figure 7B). In conclusion, the present results suggested that the transcriptional and translational expression levels of SDC1 were overexpressed in BRCA patients while SORBS1 were overexpressed in normal tissues.
CMap Analysis Identified Novel Candidate Compounds Targeting the Prognostic Genes

To identify potential compounds capable of targeting prognostic lipid metabolism-related genes (SDC1 and SORBS1), we screened a total of 76 compounds that met the enrichment score > 0.5, P < 0.05 by the CMap02 database (Supplementary Table 17). Unfortunately, only 37 of the 76 compounds could be traced in the CLUE tool. The 37 compounds that were able to...
Figure 6 Expression levels of SORBS1 and SDC1 proteins in clinical samples. (A) Expression levels of SORBS1 measured by qRT-PCR analysis. (B) Expression levels of SDC1 measured by qRT-PCR analysis. (C) Representative image of Western blot analysis of SORBS1 protein. N, Adjacent healthy samples; T, BRCA samples. (D) Representative image of Western blot analysis of SDC1 protein. (E) Immunohistochemical staining (SORBS1 protein) of BRCA and adjacent healthy groups; magnification, ×200. (F) Immunohistochemical staining (SDC1 protein) of BRCA and adjacent healthy groups; magnification, ×200. (G and H) The AOD of SORBS1 and SDC1 protein in BRCA and adjacent healthy groups. Mean ± SEM; n = 30; ***P < 0.001 vs adjacent healthy. ****P < 0.0001 vs adjacent healthy.

Abbreviations: SORBS1, sorbin and SH3 domain containing 1; SDC1, Syndecan-1; AOD, average optical density.
inhibit prognostic gene expression as described above are illustrated in Supplementary Figure 8A. Subsequent CMap MoA analysis revealed 44 mechanisms of action common to the above compounds. Amitriptyline was involved in four MoAs, including norepinephrine inhibitor, norepinephrine reuptake inhibitor, serotonin receptor agonist, and serotonin reuptake inhibitor, while the MoAs involving acetylcholine release stimulant, acetylcholinesterase inhibitor, butyrylcholinesterase inhibitor, and potassium channel antagonist were all modulated by the compound tacrine. Furthermore, a total of 11 compounds shared the following six mechanisms: 1) adrenergic receptor agonist, 2) bacterial cell wall synthesis inhibitor, 3) glucocorticoid receptor agonist, 4) histone deacetylase (HDAC) inhibitor, 5) progesterone receptor agonist, and 6) sterol demethylase inhibitor.

**Prognostic Gene-Drug Interaction Network Analysis**

Interactions between prognostic genes and drugs used for cancer treatment were probed through the CTD database and visualized via Cytoscape. As depicted in Supplementary Figure 8B, there were 50 and 42 drugs that affected the expression of SDC1 and SORBS1, respectively, of which 16 drugs regulated the expression of both prognostic genes. For instance, abrine reduced the SDC1 expression levels, while upregulating the SORBS1 expression (Supplementary Table 18).

**SDC1 Could Promote the Migration and Invasion of Breast Cancer Cells**

To verify transduction efficacy, RT-PCR and Western blot were used to detect the expression of SDC1 following 24 h of transduction (Figure 7A and B). To explore the effect of SDC1 on breast cell invasion and migration, transwell invasion assays and wound healing assays were performed. Transwell invasion assays demonstrated that si-SDC1 could reduce the invasion ability of MCF-7 and MDA-MB-231 cells and oe-SDC1 could enhance the invasion ability of MCF-7 and MDA-MB-231 cells in vitro (Figure 7C and D). The wound healing assays demonstrated that the migratory abilities of MCF-7 and MDA-MB-231 cells were significantly suppressed following transfection of si-SDC1 (Figure 7E and F). Similarly, compared with the corresponding negative control groups, the speed of wound closure was significantly faster in the MCF-7 and MDA-MB-231 cells transfected with SDC1 plasmid (Figure 7E and F). These results suggest that SDC1 can increase the invasiveness of breast cancer cells.

**SORBS1 Could Inhibit the Migration and Invasion of Breast Cancer Cells**

To verify transduction efficacy, qRT-PCR and Western blot were used to detect the expression of SORBS1 following 24h of transduction (Figure 8A and B). In the transwell invasion assays, the invasion ability of MCF-7 and MDA-MB-231 cells in the si-SORBS1 group was significantly increased compared with si-SORBS1 control group, and the invasion ability of MCF-7 and MDA-MB-231 cells in the oe-SORBS1 control groups was significantly increased compared with oe-SORBS1 groups (Figure 8C and D). The wound healing assays demonstrated that the migratory abilities of MCF-7 and MDA-MB-231 cells were significantly promoted following transfection of si-SORBS1 (Figure 8E and F). Similarly, the speed of wound closure was significantly slower in the MCF-7 and MDA-MB-231 cells transfected with SORBS1 plasmid compared with the negative control groups (Figure 8E and F). Taken together, SORBS1 inhibits the migration and invasion of breast cancer cells, and SORBS1 depletion conversely promotes the migration and invasion of breast cancer cells.

**Discussion**

The high annual incidence rate of BRCA has a serious impact on human health and the social economy. Detection of BRCA at an early stage is a critical step for successful treatment and improvement of the prognosis. The transformation of metabolic pattern is one of the important characteristics of tumor cells. A number of studies have found that tumor cells will undergo lipid metabolism reprogramming to meet the needs of rapid proliferation. Because the tumor microenvironment is rich in adipokines, abnormal lipid metabolism plays a more important role in breast cancer than other malignant tumors. In recent years, great achievements have been made in abnormal lipid metabolism in tumors. Lipid molecules with the diagnostic potential for BRCA are constantly being discovered. Studies have found that the abnormal lipid metabolism of BRCA cells is closely related to their resistance to HER2 inhibitors and CDK4/6 inhibitors. In addition, some studies have reported changes of phospholipids in the plasma, serum, and urine.
of BRCA patients. However, lipid metabolism involves both exogenous and endogenous processes, and its impact on BRCA risk and prognosis needs to be further determined. Thus, lipid metabolism-related genes in BRCA may be a breakthrough point that can further improve the prognosis of breast cancer patients.

Figure 7 Effects of SDC1 knockdown or overexpression on BC cells' invasion and migration. (A) Quantitative PCR analysis of SDC1 expression levels in si-SDC1 control, si-SDC1, oe-SDC1 control and oe-SDC1 breast cancer cell lines (MCF-7 and MDA-MB-231). (B) Western blot of SDC1 expression levels in si-SDC1 control, si-SDC1, oe-SDC1 control and oe-SDC1 breast cancer cell lines (MCF-7 and MDA-MB-231). (C) Effects of SDC1 knockdown or overexpressed on invasion of MDA-MB-231 breast cancer cells. (D) Effects of SDC1 knockdown or overexpressed on invasion of MCF-7 breast cancer cells. (E) Effects of SDC1 knockdown or overexpressed on migration of MDA-MB-231 breast cancer cells showed by scratch assays. (F) Effects of SDC1 knockdown or overexpressed on migration of MCF-7 breast cancer cells showed by scratch assays. **P < 0.01.
In this study, a risk assessment classifier closely related to the prognosis of BRCA was constructed through the analysis of differentially expressed lipid metabolism-related genes with prognostic value in BRCA and revealed the correlation between BRCA and lipid metabolism. We not only used the GEO database, but also the cell function

Figure 8 Effects of SORBS1 knockdown or overexpression on BC cells' invasion and migration. (A) Quantitative PCR analysis of SORBS1 expression levels in si-SORBS1 control, si-SORBS1, oe-SORBS1 control and oe-SORBS1 breast cancer cell lines (MCF-7 and MDA-MB-231). (B) Western blot of SORBS1 expression levels in si-SORBS1 control, si-SORBS1, oe-SORBS1 control and oe-SORBS1 breast cancer cell lines (MCF-7 and MDA-MB-231). (C) Effects of SORBS1 knockdown or overexpressed on invasion of MDA-MB-231 breast cancer cells. (D) Effects of SORBS1 knockdown or overexpressed on invasion of MCF-7 breast cancer cells. (E) Effects of SORBS1 knockdown or overexpressed on migration of MDA-MB-231 breast cancer cells showed by scratch assays. (F) Effects of SORBS1 knockdown or overexpressed on migration of MCF-7 breast cancer cells showed by scratch assays. **: $P < 0.01$; ***: $P < 0.001$.

In this study, a risk assessment classifier closely related to the prognosis of BRCA was constructed through the analysis of differentially expressed lipid metabolism-related genes with prognostic value in BRCA and revealed the correlation between BRCA and lipid metabolism. We not only used the GEO database, but also the cell function
experiments and the clinical data of tissue samples to verify that this prognostic classifier can offer good clinical application value.

In addition, we analyzed the correlation between the prognostic classifier and pathological characteristics and clinical characteristics and found that the risk of TNBC was significantly higher. TNBC is histologically defined as lacking estrogen and progesterone receptors (ER/PR) and human epidermal growth factor receptor 2 (HER2) amplification, it is a subtype of BRCA with the worst prognosis and highest risk because of the lack of effective therapeutic targets. 58,59 However, no targeted drugs for TNBC have been discovered so far. 60,61 At present, the main systemic treatment for TNBC is still chemotherapy. 62 Antibody drug conjugates (ADCs), immune-oncology (I/O) therapies, and polyadenosine diphosphate-ribose polymerase (PARP) inhibitors have joined the armamentarium against specific types of TNBC but have limited durable efficacy in TNBC patients. 63 Our findings may provide information for the development of drugs that target lipid metabolism genes, thus, making it possible to find new pathways in the treatment of TNBC.

In order to further explore the mechanism of the lipid metabolism DEGs related to BRCA survival, we performed network analysis using IPA and KEGG and GO enrichment analysis on the 13 genes selected. We found that PPAR signaling pathways and adipokine signaling pathways had the highest correlation with these genes, and IPA uncovered a significant enrichment of lipid metabolism pathways and estrogen-dependent BRCA signaling. Moreover, the effect of PPAR signaling pathway and estrogen-dependent signaling in BRCA was previously confirmed and the related drugs are in clinical application, 64-67 so these results verified our research’s directions and methods are scientific from the other side. The top GO terms involved in lipid metabolism were “lipid droplet” and “external side of the plasma membrane.” Adipokine signaling pathways are the main pathways of lipid metabolism, and their relevance to tumors has been confirmed in many studies. 68-71 They play an independent and combined role in the activation of intracellular signaling networks related to the malignant phenotype of BRCA cells. 72 The GO term “external side of the plasma membrane” has also been reported to be associated with testicular cancer. 73 Lipid droplets 74 are the natural immune center that integrates cell metabolism and host defense and are gradually being recognized as a prominent feature of many cancers. 75,76 Studies have reported that lipid droplet and the PPAR signaling pathway are closely related to the metastasis of TNBC. 77-79 Furthermore, lipid droplets have been found to have increased expression in drug-resistant TNBC cells. 71 In addition, based on the current progress of immunotherapy for BRCA and the increasing recommendation of immunotherapy in the treatment regimen of TNBC, 80 although our functional enrichment analysis did not indicate that the lipid metabolism-related DEGs were related to immune cells, we still calculated the correlation between SDC1 and SORBS1 and immune cells based on TIMER database, but no positive result was obtained. This means that SDC1 and SORBS1 do not affect the prognosis of BRCA through immune function pathway, which also suggests that it may be a new direction for us to explore the diagnosis and treatment of BRCA from the factors related to lipid metabolism. These findings have a certain correlation with the poor prognosis of TNBC in the prognostic classifier we established in this study and support the development of therapies for TNBC targeting lipid metabolism.

The genes we used for prognostic classifier constructing were SORBS1 and SDC1. SORBS1 encodes a CasitasB.1 lineage lymphoma (CBL)–associated protein that functions in the signaling and stimulation of insulin, and it also been confirmed to be related to COPD. 81-83 Studies have found that SORBS1 can participate in immune-related gene signal transduction, and its increased expression is closely related to poor prognosis in cervical cancer, 84 gastric cancer, 85 prostate cancer, 86 and colorectal cancer. 87 However, the exact role and diagnostic significance of SORBS1 in BRCA are not yet clear. Syndecan-1 (SDC1, also known as CD138) is a key regulator of fatty acid synthesis and catalyzes the formation of mono-unsaturated fatty acids (MUFA) from saturated fatty acids (SFA). 88-91 SDC1 has been implicated in various cancers, including BRCA. 92-95 The inactivation of SDC1 can lead to the consumption of unsaturated fatty acids, thereby inhibiting cancer cells. 91,96,97 However, the imbalance in the ratio of saturated to unsaturated fat can lead to a cellular stress response and death in both normal and transformed cells. 91,98 Therefore, although SDC1 has potential as a biomarker, its complex relationship with the body limits its clinical application. 99-101 In this study, we constructed a prognostic classifier based on SORBS1, which is currently less studied, and SDC1, which has a complex effect on the body. In addition, we examined the correlation between invasiveness and expression of SDC1 and SORBS1 in MCF-7 and MDA-MB-231 breast cancer cell lines by cell function experiments. This is the first prognostic classifier that combines the two genes in order to jointly assess their prognostic risk in BRCA and this classifier can change the limited application status of SDC1 and SORBS1.
Through the CMap02 database, we screened potential drugs that target genes related to lipid metabolism as a means to control the progression of BRCA. We also used the CTD database to explore the interaction between selected prognostic genes and cancer treatment drugs. We found that vorinostat, abrine, and NSC 689534 are very promising drugs. Vorinostat is a simple molecule histone deacetylase (HDAC) inhibitor which has both oral and vena formulation. The main indication currently approved by the FDA is the treatment of skin T-cell lymphoma and its combination with pembrolizumab (a monoclonal anti-PD1 antibody) can increases the response rate of head and neck malignancies to immunotherapy. Whereas, pembrolizumab combined with standard chemotherapy plays an important role in the treatment of advanced TNBC, and this may mean that we can try adding vorinostat to this treatment to increase the tumor response rate to pembrolizumab. In addition, the main metabolic pathway of vorinostat is via hydroxamic acid glucuronidation and oxidative cleavage of the aliphatic methylene chain. This confirms our findings that this drug may be effective in treating BRCA patients carrying the prognosis-related lipid metabolism genes we screened. Abrine is an alkaloid, and the main component of the traditional Chinese medicine cochinchinensis. In addition, abrine has a structure that can form hydrogen bonds with peroxisome PPARs, thereby affecting the PPAR pathway. Our analysis found that abrine reduces the expression level of SDC1 while increasing the expression of SORBS1, which indicated that it may have a dual protective effect on BRCA patients. The inhibitory effect of abrine on BRCA cells has been confirmed in several in vitro experiments. It is our hope that our research results can prompt experts in the field of chemistry and pharmacy to pay attention to this compound with therapeutic potential. NSC 689534 (2-pyridinecarbaldehyde N, N-bis (2-pyridinylmethyl) thiosemicarbazone) is a member of the heterocyclic thiosemicarbazone family of compounds, and its biological activity is significantly different between the unbound state and the metal-bound state. In recent years, there have been reports about the effects of these drugs on human cancer, but most of them are pharmacological studies. It is found that the chelating ability of thiosemicarbazone NSC 689534 can induce oxidative/ER stress and inhibit tumor growth in vitro and in vivo. Our findings suggest that NSC 689534 is a potential drug for the treatment of BRCA which is worthy of further preclinical investigation.

Our analysis provides a reference for the prognosis of BRCA patients, a basis for the diagnosis and treatment of patients, and a new research direction for BRCA research. This study still has some certain limitations, these results are preliminary that mainly based on the secondary mining and analysis of previously published datasets. Although PCR experiments, Western blot experiments, immunohistochemistry experiments and cell function experiments were performed in this study to initially validate the association between the two genes and breast cancer, we still need to conduct prospective studies in order to further verify the clinical applicability of this prognostic classifier.

Conclusions

In summary, we have constructed a prognostic classifier of BRCA based on two lipid metabolism-related genes: SDC1 and SORBS1. This classifier has a value in predicting the disease-free survival rate of BRCA patients and identifying high-risk patients. Our results strengthen the underestimated role of abnormal tumor lipid metabolism in the prognosis of BRCA. The translational application of this classifier will guide clinicians to make a more informed decision regarding adjuvant systemic treatments and choosing follow-up plan.

Abbreviations

qRT-PCR, Real-Time Quantitative Reverse Transcription PCR; SDC1, Syndecan-1; TNBC, triple negative breast cancer; AMPK, Adenosine 5’-monophosphate (AMP)-activated protein kinase; AUC, area Under Curve; COPD, chronic obstructive pulmonary disease; EDTA, Ethylene Diamine Tetraacetic Acid; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; MIF, Macrophage Migration Inhibitory Factor; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; AJCC, American Joint Committee on Cancer; BRCA, breast cancer; CMap, Connectivity Map; CTD, Comparative Toxicogenomics Database; DEGs, differentially expressed genes; GEO, Gene Expression Omnibus; GO, Gene Ontology; IPA, Ingenuity Pathway Analysis; PPARs, proliferator-activated receptors; TCGA, The Cancer Genome Atlas.

Data Sharing Statement

Some or all data, models, or code generated or used during the study are available from the corresponding author by request.
Ethics Statement
Written informed consent was obtained from the patients/participants for publication of this article. The studies involving human participants were reviewed and approved by the Ethics Committee of the First Affiliated Hospital of Zhengzhou University.

Acknowledgments
We thank the experimental platform of Translational Medicine Center of the First Affiliated Hospital of Zhengzhou University for supporting this study. And we thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

Author Contributions
All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation. All authors gave final approval of the version to be published and agreed to be accountable for all aspects of the work.

Funding
Health Commission of Henan Province, Youth Project of Medical Science and Technology Project of Henan Province, Award Number: SBGJ202103070; Recipient: Nan Wang, M.D. Science and Technology Department of Henan Province; Award Number: LHGJ20200356; Recipient: Nan Wang, M.D.

Disclosure
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.

References


77. Wright HJ, Hou J, Xu B, et al. CDCP1 drives triple-negative breast cancer metastasis through reduction of lipid-droplet abundance and stimulation of fatty acid oxidation.


