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

Nan Wang (✉ fccwangn2@zzu.edu.cn)

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

Lin Li

Zhengzhou University First Affiliated Hospital

Jiangrui Chi

The First Affiliated Hospital of Zhengzhou University

Xinwei Liu

The First Affiliated Hospital of Zhengzhou University

Youyi Xiong

The First Affiliated Hospital of Zhengzhou University

Chaochao Zhong

The First Affiliated Hospital of Zhengzhou University

Research Article

Keywords: lipid metabolism, Breast invasive carcinoma (BRCA), prognostic classifier, SDC1, SORBS1

Posted Date: December 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1105292/v1>

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

[Read Full License](#)

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

Nan Wang^{1*}, Lin Li¹, Jiangrui Chi¹, Xinwei Liu¹, Youyi Xiong¹, Chaochao Zhong²

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

fccwangn2@zzu.edu.cn

Number of words: 6022

Number of figures and tables: 10

Keywords: lipid metabolism, Breast invasive carcinoma (BRCA), prognostic classifier, *SDCI*, *SORBS1*

Abstract

Background: Alterations in lipid metabolism have been implicated in the development of many tumors. However, the contribution of different lipid metabolism pathways to 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 to harvest lipid metabolism-related genes. IPA was applied to identify the potential pathways and functions of DEGs related to lipid metabolism. Subsequently, univariate and multivariate Cox regression analyses were utilized to construct the prognostic gene signature and independent prognostic analyses. Thereafter, the differential expression of the selected marker genes *SDCI* and *SORBS1* in clinical tissue samples was verified by qRT-PCR, western blotting, and immunohistochemical experiments. Functional enrichment analysis of prognostic genes was achieved by the GO and KEGG databases. Moreover, Kaplan-Meier analysis, ROC curves, clinical immunohistochemistry conditions and follow-up results were employed to assess the prognostic potency. Potential compounds targeting prognostic genes were then screened by CMap database and a prognostic gene-drug interaction network was constructed using Comparative Toxicogenomics Database.

Results: IPA demonstrated that the 162 lipid metabolism-related DEGs we obtained were involved in a variety of lipid metabolism and BRCA pathological signatures. Subsequent functional enrichment analysis of candidate prognostic lipid metabolism DEGs also revealed a similar outcome. The prognostic classifier we constructed comprising *SDCI* and *SORBS1* has a strong prognostic potency that was verified by the clinical conditions and follow-up results, it also can serve as an independent prognostic marker for BRCA. CMap filtered 37 potential compounds against prognostic genes. CTD indicated that the two prognostic genes had 16 drugs in common.

Conclusion: Within this study, we identified a novel prognostic classifier based on two lipid metabolism-related genes: *SDCI* and *SORBS1*. This classifier had accurately predicted the prognosis of our follow-up BRCA patients and this result highlighted a new perspective on the metabolic exploration of BRCA. In addition, *SDCI* and *SORBS1* could serve as a possible new target for the synthesis of BRCA drugs.

1 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 cases¹. Despite advances in screening, research, and treatment, BRCA incidence continues to increase^{2,3}. 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 exercise⁴. 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 tumor prognosis^{5,6}. Among these risk factors, the role of lipid metabolism disorders in the occurrence and development of BRCA has started to be appreciated⁷⁻⁹. 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 progression¹⁰. 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^{11,12}. Abnormal lipid metabolism has been reported to be related to the occurrence and development of a variety of malignant tumors, such as prostate cancer¹³, ovarian cancer¹⁴, pancreatic cancer¹⁵, liver cancer¹⁶, and BRCA¹⁷. 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^{18,19}. 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^{20,21}. 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 metabolism²². 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 BRCA²³⁻²⁵.

The use of bioinformatics to analyze the targets of disease diagnosis and treatment has been widely used^{26,27}. In this study, we analyzed and obtained the biological processes of differential lipid metabolism genes with prognostic value in BRCA based on the 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, *SDCI* and *SORBS1*, which we screened, analyzed, and verified as biomarker genes that may play important roles in BRCA. Finally, we constructed risk assessment models and preliminarily verified them with datasets in the GEO database and clinical follow-up results we collected. The validation results showed that our risk assessment models have 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

82 Toxicogenomics Database (CTD) to investigate compounds targeting prognostic genes, and to
83 provide new insights for the synthesis of new drugs.

84 **2 Materials and Methods**

85 **2.1 Clinical Sample Acquisition**

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

100 In order to verify the clinical applicability of our prognostic model, we collected the clinical data and
101 pathological sections of 50 patients with BRCA who underwent surgery in our hospital from January
102 2015 to March 2015. The plan for clinical patient follow-up and collection of patient clinical data to
103 verify the prognosis model was approved by the Ethics Committee of the First Affiliated Hospital of
104 Zhengzhou University. We followed up by phone or in-person to collect the survival information of
105 the 50 patients. All patients were between 18 and 80 years old and agreed to accept the post-operative
106 follow-up plan. The personal information of each patient is strictly confidential, and as such, written
107 informed consent was obtained from each patient or its entrustment. The clinical conditions and
108 follow-up results of the patients are shown in Supplementary Table 2.

109 **2.2 Data Sources**

110 We acquired RNA expression data from 1,171 samples (1,072 breast cancer samples and 99 normal
111 samples) from TCGA database. Of the 1,072 BRCA samples, only 1,069 BRCA samples recorded
112 complete survival information, so candidate prognosis - related lipid metabolism differentially
113 expressed genes (DEGs) were identified based on these samples. In addition, only 881 of the 1,069
114 BRCA samples had complete clinical information, which was needed for the sample to be used in the
115 construction of the prognostic classifier.

116 The GSE109169 dataset was downloaded from the GEO database (25 BRCA samples, 25 normal
117 samples) and was utilized to legitimize the expression pattern of candidate prognosis-related lipid
118 metabolism DEGs. Furthermore, the GSE20685 dataset, containing 327 BRCA samples, served as an
119 external validation set to verify the prognostic validity of the prognostic signature. Moreover, 1,499
120 genes related to lipid metabolism were obtained by de-duplication from 13 gene sets related to lipid
121 metabolism, which were downloaded from the Gene Ontology (GO) and the Molecular Signatures
122 Database (MSigDB) websites. These 13 gene sets were
123 GO_GALACTOLIPID_METABOLIC_PROCESS,

124 GO_GLYCEROLIPID_METABOLIC_PROCESS,
 125 GO_GLYCEROPHOSPHOLIPID_METABOLIC_PROCESS,
 126 GO_GLYCOSPHINGOLIPID_METABOLIC_PROCESS, GO_LIPID_METABOLIC_PROCESS,
 127 GO_LIPOPROTEIN_METABOLIC_PROCESS,
 128 GO_MEMBRANE_LIPID_METABOLIC_PROCESS,
 129 GO_NEUTRAL_LIPID_METABOLIC_PROCESS,
 130 GO_PHOSPHOLIPID_METABOLIC_PROCESS,
 131 GO_REGULATION_OF_LIPID_METABOLIC_PROCESS,
 132 GO_REGULATION_OF_LIPOPROTEIN_METABOLIC_PROCESS,
 133 GO_REGULATION_OF_PHOSPHOLIPID_METABOLIC_PROCESS,
 134 GO_SPHINGOLIPID_METABOLIC_PROCESS. All lipid metabolism-related genes are integrated
 135 into Supplementary Table 3.

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

140 Translation level validation of identified prognostic genes was performed using the Human Protein
 141 Atlas²⁸ database (<http://www.proteinatlas.org/>), an online database containing immunohistochemical
 142 expression data for approximately 20 of the most common cancers.

143 2.3 Differential Analysis

144 Transcriptome RNA-seq data of 1171 cases (normal samples, 99 cases; BRCA samples, 1072 cases)
 145 were downloaded from TCGA database. R package limma²⁹ was used to perform differentiation
 146 analysis of the gene expression, and DEGs were generated by the comparison between the BRCA
 147 samples vs. the normal samples (Supplementary Table 4). DEGs with $|\log_2 \text{fold change}|^{30} > 1$ and
 148 adjusted $P < 0.05$ were considered significant.

149 2.4 Ingenuity Pathway Analysis (IPA)

150 Overlapping genes (Supplementary Table 5) for DEGs and lipid metabolism-related genes were
 151 analyzed by the IPA software (version 1-19-00). RNA sequencing results were imported into IPA, and
 152 the enrichment status of canonical pathways and diseases and functions were assessed by P -values
 153 ^{31,32}. Z scores were indicated by orange or blue to denote activation or inhibition of pathways.

154 2.5 Identification of Candidate Prognosis-Related Lipid Metabolism DEGs

155 Lipid metabolism DEGs were included in the univariate Cox regression analysis, and genes with a P
 156 < 0.05 were selected to be used in the multivariate Cox regression analysis. Similarly, genes with a P
 157 < 0.05 were considered as candidate prognosis-related lipid metabolism DEGs. Subsequently, GO
 158 and KEGG analyses were carried out based on the above genes using the R software clusterProfiler
 159 package. GO analysis consisted of three components: biological process (BP), molecular function
 160 (MF) and cellular component (CC). Entries with a $P < 0.05$ were considered significantly enriched.

161 2.6 RNA Isolation, Reverse Transcription, and Real-Time PCR

162 Total RNA was isolated from 30 paired tissues using TriQuick Reagent (Solarbio, Shanghai, China)
 163 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 previously³³. 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 15 s and 60°C for 30 s. The *GAPDH* gene served as an internal control. RNA levels were calculated for tumor samples and paired adjacent samples using the 2-ΔCt method. Primer sequences used for qRT-PCR are shown in Table 1.

2.7 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 using a RIPA buffer (P0013C, Beyotime, Shanghai, China). Protein samples were separated by SDS-PAGE and transferred onto nitrocellulose membranes (88520, 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).

2.8 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 SORBS1 (1:500; 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 density³⁴, where AOD=integrated OD/measurement area³⁵. The higher the AOD, the higher the expression of the protein^{36; 37}. The person imaging the slides was uninformed during the experiment and analysis.

2.9 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. 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: $\text{risk score} = (\text{expression of Gene}_1 \times \beta_1 \text{Gene}_1) + (\text{expression of Gene}_2 \times \beta_2 \text{Gene}_2) + (\dots \text{expression of Gene}_n \times \beta_n \text{Gene}_n)^{38,39}$. β 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.

2.10 Prognostic Analysis

Risk curves were plotted with the pheatmap package. The R survival package⁴⁰ 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^{41,42} 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 external validation set. 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 rms⁴³ and survival packages of R were performed to construct a nomogram. Subsequently, calibration curves were plotted to evaluate the agreement between actual and predicted survival⁴⁴. 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 calculated⁴⁵.

2.11 Connectivity Map (CMap) Analysis

CMap is designed to reveal the relationships among genes, compounds, and biological conditions, which is a systematic, data-driven process⁴⁶. 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.

2.12 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/>).

2.13 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 conducted to detect the pathologic

247 differences between high- and low-risk patients. The statistical analyses were conducted using the R
248 software. A $P < 0.05$ indicated statistically significant differences.

249 2.14 Model Validation by Detailed Data from 2015.01 to 2015.03

250 After collecting clinicopathological and follow-up data and conducting immunohistochemistry
251 staining, correlations between *SDCI* and *SORBS1* expression and survival prognosis were analyzed.

252 3 Results

253 3.1 Identification of Differentially Expressed Lipid Metabolism-Related Genes

254 RNA sequencing data of 1072 BRCA samples and 99 normal samples were extracted from TCGA
255 database. A differential analysis based on the R package limma was performed with normal samples
256 as controls in order to screen for genes that were aberrantly expressed in BRCA, and genes satisfying
257 $|\log_2 FC| > 1$ and adjusted $P < 0.05$ were identified as DEGs. We identified a total of 1,732 DEGs
258 relevant to BRCA, 162 of which were lipid metabolism-related genes. A total of 644 DEGs were
259 upregulated while 1,088 were downregulated (Figures 1A,B). Thirty-four genes associated with lipid
260 metabolism were upregulated while 128 were downregulated (Figure 1C). Canonical pathway
261 analysis of these lipid metabolism - related DEGs in IPA revealed that several pathways related to
262 lipid metabolism (e.g., fatty acid β -oxidation I, fatty acid α -oxidation, and fatty acid activation),
263 carcinogenesis (e.g., AMPK signaling, JAK/STAT signaling, and ERK/MAPK signaling), and
264 immune response (e.g., MIF regulation of innate immunity, IL-7 signaling pathway, and glioblastoma
265 multiforme signaling) were inhibited. Furthermore, we demonstrated that these genes were correlated
266 with the negative regulation of estrogen-dependent BRCA signaling, HER2 signaling in BRCA, and
267 BRCA regulation by stathmin1 (Supplementary Figure 1A; Supplementary Table 6). The diseases
268 and functions analysis confirmed that lipid metabolism and small molecule biochemistry were the
269 most abundant pathways (Supplementary Figure 1B; Supplementary Table 7).

270 3.2 Functional Enrichment Analysis of Candidate Prognosis-Related Lipid Metabolism DEGs

271 In the present analysis, we included 1,069 BRCA samples that contained complete survival
272 information. Univariate Cox analysis screened 76 genes with $P < 0.05$ from the identified 162 lipid
273 metabolism-related DEGs (Table 2). Thirteen candidate prognosis - related lipid metabolism genes
274 were obtained from the subsequent multivariate Cox analysis. Through examination with a forest
275 plot, genes with $HR > 1$ were found to include *LEPR*, *FGF2*, *GPAM*, *SDCI*, *CCDC3*, *PCK2*, *PLTP*
276 and *HSD17B13*, which may be risk factors for BRCA prognosis; while *ABCD2*, *FABP4*, *SORBS1*,
277 *FAM126A* and *GPIHBP1* with $HR < 1$ could be termed as protective factors for BRCA prognosis
278 (Figure 2A). We then validated the expression patterns of the 13 genes mentioned above using the
279 GSE109169 dataset. The results showed that the expression trends of these genes were consistent
280 with those in the TCGA database. Except for *SDCI*, the expression of the other 12 genes was
281 downregulated in the BRCA samples (Figure 2B,C). The underlying features of these genes were
282 ascertained by GO and KEGG pathway analysis (Figures 2D,E). Undoubtedly, several GO terms
283 related to lipid metabolism were found in biological processes, such as “positive regulation of lipid
284 metabolic process,” “regulation of lipid biosynthetic process,” “regulation of lipid metabolic
285 process,” “positive regulation of lipid biosynthetic process,” “glycerolipid metabolic process,” and
286 “triglyceride metabolic process” (Supplementary Table 8). Furthermore, “lipid droplet” and “external
287 side of plasma membrane” were the most remarkably enriched GO items of CC. As implied by the
288 KEGG pathway analysis, 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 9).

3.3 Filtering Candidate Prognosis-Related Lipid Metabolism DEGs for Prognostic Prediction

Two prognosis - related genes, *SDCI* (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). Therefore, a two-gene signature was gained as a classifier, and a risk score for individual patients was generated with the expressed values and multivariate Cox regression coefficients (Supplementary Table 10). Where *SDCI* was deemed to be a pro-oncogene (Figure 3I; hazard ratio (HR) = 1.2443), *SORBS1* was found to be a protective gene (Figure 3J; hazard ratio (HR) = 0.7682), as determined by the hazard ratio (HR) of each prognostic gene. Moreover, IPA exposed a sophisticated network of interactions involving the two prognostic genes (Supplementary Figures 1C–E).

Patients were categorized into high-risk or low-risk groups based on the median risk score. As patient risk scores escalated, the mortality increased progressively (Figures 3C,D). K–M analysis indicated that the OS of high-risk patients was noticeably shorter than that of low-risk patients in both the training and testing sets ($P < 0.02$; Figures 3F,G). The AUC of the ROC curves demonstrated that the two-gene signature had a tolerable prognostic predictive potency (Figures 3K,L). In order to validate the gene signature, we also calculated the risk scores for patients in the external validation set (GSE20685 dataset) relying on the same methodology. Unexpectedly, we obtained a consensus result (Figures 3E,H,M). We constructed a heatmap to display the expression of prognostic genes in each cohort (Supplementary Figure 2).

3.4 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. As illustrated in Figure 4, higher risk scores were generally distributed among patients with higher American Joint Committee on Cancer (AJCC) pathologic T stage, AJCC pathologic N stage, and tumor stage. We also assessed the relationship between different types of BRCA and risk scores. Of the 4 types of BRCA, triple-negative BRCA had the highest risk score, followed by luminal B BRCA. Such results suggested a substantial interaction between lipid metabolism genes and clinical molecular features.

3.5 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 0.744. 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,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. Additionally, immunohistochemical results, pathological results, and follow-up results of 50 patients 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 ($P = 2.516e-02$) (Figure 5G).

3.6 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 3A); 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 3B). 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.

3.7 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 $|\text{enrichment score}| > 0.5$, $P < 0.05$ by the CMap02 database (Supplementary Table 11). Unfortunately, only 37 of the 76 compounds could be traced in the CLUE tool. The 37 compounds that were able to inhibit prognostic gene expression as described above are illustrated in Figure 7. 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.

3.8 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 Figure 8, there were 50 and 42 drugs that

affected the expression of *SDCI* and *SORBS1*, respectively, of which 16 drugs regulated the expression of both prognostic genes. For instance, abrine reduced the *SDCI* expression levels, while upregulating the *SORBS1* expression (Supplementary Table 12).

4 Discussion

The high annual incidence rate of BRCA has a serious impact on human health and the social economy⁴⁷. Detection of BRCA at an early stage is a critical step for successful treatment and improvement of the prognosis⁴⁸. 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^{49,50}. 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^{12,19}. Lipid molecules with the diagnostic potential for BRCA are constantly being discovered^{12,51}. Studies have found that the abnormal lipid metabolism of BRCA cells is closely related to their resistance to HER2 inhibitors and CDK4/6 inhibitors⁵². In addition, some studies have reported changes of phospholipids in the plasma, serum^{53,54}, and urine^{55,56} 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.

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 clinical data and immunohistochemical results of tissue samples to verify that this prognostic classifier I offered good clinical application value in assessing the prognosis of patients. Compared with other published features, our prognostic classifier has better predictive ability.

In addition, we analyzed the correlation between the prognostic classifier and pathological conditions 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^{57,58}. However, no targeted drugs for TNBC have been discovered so far^{59,60}. At present, the main systemic treatment for TNBC is still chemotherapy⁶¹. 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 patients⁶². 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 application⁶³⁻⁶⁶, 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 studies⁶⁷⁻⁷⁰. They play an independent and combined role in the activation of intracellular signaling networks related to the malignant phenotype of BRCA cells⁷¹. The GO term “external side of the plasma membrane” has also been reported to be associated with testicular cancer⁷². Lipid droplets⁷³ are the natural immune center that integrates cell metabolism and host defense and are gradually being recognized as a prominent feature of many cancers⁷⁴⁻⁷⁵. Studies have reported that lipid droplet and the PPAR signaling pathway are closely related to the metastasis of TNBC⁷⁶⁻⁷⁸. Furthermore, lipid droplets have been found to have increased expression in drug-resistant TNBC cells²¹. In addition, based on the current progress of immunotherapy for BRCA and the increasing recommendation of immunotherapy in the treatment regimen of TNBC⁷⁹, 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 *SDCI* and *SORBS1* and immune cells based on TIMER database, but no positive result was obtained. This means that *SDCI* 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 *SDCI*. *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 COPD⁸⁰⁻⁸². 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 cancer⁸³, gastric cancer⁸⁴, prostate cancer⁸⁵, and colorectal cancer⁸⁶. However, the exact role and diagnostic significance of *SORBS1* in BRCA are not yet clear. Syndecan-1 (*SDCI*, 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)⁸⁷⁻⁹⁰. *SDCI* has been implicated in various cancers, including BRCA⁹¹⁻⁹⁴. The inactivation of *SDCI* can lead to the consumption of unsaturated fatty acids, thereby inhibiting cancer cells^{90,95,96}. 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^{90,97}. Therefore, although *SDCI* has potential as a biomarker, its complex relationship with the body limits its clinical application⁹⁸⁻¹⁰⁰. In this study, we constructed a prognostic classifier based on *SORBS1*, which is currently less studied, and *SDCI*, which has a complex effect on the body. This is the first prognostic classifier that combines the two in order to jointly assess their prognostic risk in BRCA. The successful construction of this classifier has given *SDCI* clinical application value and changed its limited application status.

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, abiraterone, 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 increase the response rate of head and neck malignancies to immunotherapy¹⁰¹. 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 chain⁶⁶. 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*¹⁰². In addition, abrine has a structure that can form hydrogen bonds with peroxisome PPARs, thereby affecting the PPAR pathway¹⁰³. Our analysis found that abrine reduces the expression level of *SDCI* 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^{104,105}. 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^{106,107}. 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 vivo¹⁰⁷. 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. Although we have constructed an independent prognostic classifier of BRCA with high clinical value, this study still has certain limitations. The results of this study are preliminary and are mainly based on the secondary mining and analysis of previously published datasets. Although we have performed PCR experiments, western blot experiments, immunohistochemistry experiments and validation by clinical information and follow-up data, we still need to conduct prospective studies in order to further verify the clinical applicability of this discovery.

5 Conclusions

In summary, we have successfully constructed a prognostic classifier of BRCA based on two lipid metabolism-related genes: *SDCI* and *SORBS1*. This classifier was validated using clinical tissue samples and confirmed to have value in predicting the disease-free survival rate of BRCA patients and identifying high-risk patients. This classifier has the characteristics of high discrimination accuracy, strong risk stratification ability, good correction performance, and wide application range and has a highly correlated drug action network. 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.

6 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.

7 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.

8 Conflict of Interest

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.

9 Author Contributions

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

10 Funding

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

11 Data Availability Statement

Some or all data, models, or code generated or used during the study are available from the corresponding author by request.

12 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.

13 References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: a cancer journal for*

- 538 *clinicians*. 2021;71(3):209-249.
- 539 2. Britt KL, Cuzick J, Phillips KA. Key steps for effective breast cancer prevention. *Nature reviews*
- 540 *Cancer*. 2020;20(8):417-436.
- 541 3. Thorat MA, Balasubramanian R. Breast cancer prevention in high-risk women. *Best practice &*
- 542 *research Clinical obstetrics & gynaecology*. 2020;65:18-31.
- 543 4. Paluch-Shimon S, Cardoso F, Partridge AH, et al. ESO-ESMO 4th International Consensus
- 544 Guidelines for Breast Cancer in Young Women (BCY4). *Annals of oncology : official journal of*
- 545 *the European Society for Medical Oncology*. 2020;31(6):674-696.
- 546 5. Jin G, Lv J, Yang M, et al. Genetic risk, incident gastric cancer, and healthy lifestyle: a meta-
- 547 analysis of genome-wide association studies and prospective cohort study. *The Lancet Oncology*.
- 548 2020;21(10):1378-1386.
- 549 6. Carr PR, Weigl K, Edelmann D, et al. Estimation of Absolute Risk of Colorectal Cancer Based on
- 550 Healthy Lifestyle, Genetic Risk, and Colonoscopy Status in a Population-Based Study.
- 551 *Gastroenterology*. 2020;159(1):129-138.e129.
- 552 7. Johnson KE, Siewert KM, Klarin D, et al. The relationship between circulating lipids and breast
- 553 cancer risk: A Mendelian randomization study. *PLoS medicine*. 2020;17(9):e1003302.
- 554 8. McLaughlin SA, Brunelle CL, Taghian A. Breast Cancer-Related Lymphedema: Risk Factors,
- 555 Screening, Management, and the Impact of Locoregional Treatment. *Journal of clinical*
- 556 *oncology : official journal of the American Society of Clinical Oncology*. 2020;38(20):2341-2350.
- 557 9. Caffa I, Spagnolo V, Vernieri C, et al. Fasting-mimicking diet and hormone therapy induce breast
- 558 cancer regression. *Nature*. 2020;583(7817):620-624.
- 559 10. Corn KC, Windham MA, Rafat M. Lipids in the tumor microenvironment: From cancer
- 560 progression to treatment. *Progress in lipid research*. 2020;80:101055.
- 561 11. Blevé A, Durante B, Sica A, Consonni FM. Lipid metabolism and cancer immunotherapy:
- 562 immunosuppressive myeloid cells at the crossroad. *International Journal of Molecular Sciences*.
- 563 2020;21(16):5845.
- 564 12. Guo R, Chen Y, Borgard H, et al. The function and mechanism of lipid molecules and their roles
- 565 in the diagnosis and prognosis of breast cancer. *Molecules*. 2020;25(20):4864.
- 566 13. Pang B, Zhu Y, Ni J, et al. Extracellular vesicles: the next generation of biomarkers for liquid
- 567 biopsy-based prostate cancer diagnosis. *Theranostics*. 2020;10(5):2309-2326.
- 568 14. Mukherjee A, Chiang CY, Daifotis HA, et al. Adipocyte-Induced FABP4 Expression in Ovarian
- 569 Cancer Cells Promotes Metastasis and Mediates Carboplatin Resistance. *Cancer research*.
- 570 2020;80(8):1748-1761.
- 571 15. Auciello FR, Bulusu V, Oon C, et al. A Stromal Lysolipid-Autotaxin Signaling Axis Promotes
- 572 Pancreatic Tumor Progression. *Cancer discovery*. 2019;9(5):617-627.
- 573 16. Wu L, Zhang X, Zheng L, et al. RIPK3 Orchestrates Fatty Acid Metabolism in Tumor-Associated
- 574 Macrophages and Hepatocarcinogenesis. *Cancer immunology research*. 2020;8(5):710-721.
- 575 17. Wang J, Shidfar A, Ivancic D, et al. Overexpression of lipid metabolism genes and PBX1 in the
- 576 contralateral breasts of women with estrogen receptor-negative breast cancer. *International*
- 577 *journal of cancer*. 2017;140(11):2484-2497.
- 578 18. Balaban S, Shearer RF, Lee LS, et al. Adipocyte lipolysis links obesity to breast cancer growth:
- 579 adipocyte-derived fatty acids drive breast cancer cell proliferation and migration. *Cancer &*
- 580 *metabolism*. 2017;5:1.
- 581 19. Marino N, German R, Rao X, et al. Upregulation of lipid metabolism genes in the breast prior to
- 582 cancer diagnosis. *NPJ breast cancer*. 2020;6:50.
- 583 20. Havas KM, Milchevskaya V, Radic K, et al. Metabolic shifts in residual breast cancer drive tumor
- 584 recurrence. *J Clin Invest*. 2017;127(6):2091-2105.
- 585 21. Sirois I, Aguilar-Mahecha A, Lafleur J, et al. A Unique Morphological Phenotype in
- 586 Chemoresistant Triple-Negative Breast Cancer Reveals Metabolic Reprogramming and PLIN4

- Expression as a Molecular Vulnerability. *Molecular cancer research : MCR*. 2019;17(12):2492-2507.
22. Jin X, Demere Z, Nair K, et al. A metastasis map of human cancer cell lines. *Nature*. 2020;588(7837):331-336.
 23. Heer E, Harper A, Escandor N, Sung H, McCormack V, Fidler-Benaoudia MM. Global burden and trends in premenopausal and postmenopausal breast cancer: a population-based study. *The Lancet Global health*. 2020;8(8):e1027-e1037.
 24. Demark-Wahnefried W, Rogers LQ, Gibson JT, et al. Randomized trial of weight loss in primary breast cancer: Impact on body composition, circulating biomarkers and tumor characteristics. *International journal of cancer*. 2020;146(10):2784-2796.
 25. Hultsch S, Kankainen M, Paavolainen L, et al. Association of tamoxifen resistance and lipid reprogramming in breast cancer. *BMC cancer*. 2018;18(1):850.
 26. Chen Q, Meng X, Liao Q, Chen M. Versatile interactions and bioinformatics analysis of noncoding RNAs. *Briefings in bioinformatics*. 2019;20(5):1781-1794.
 27. Yan X, He B, Hu L, Gao J, Chen S, Jiang G. Insight into the endocrine disrupting effect and cell response to butyltin compounds in H295R cell: Evaluated with proteomics and bioinformatics analysis. *The Science of the total environment*. 2018;628-629:1489-1496.
 28. Shahid M, Mubeen A, Tse J, et al. Branched chain in situ hybridization for albumin as a marker of hepatocellular differentiation: Evaluation of manual and automated in situ hybridization platforms. *American Journal of Surgical Pathology*. 2015;39(1):25–34.
 29. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
 30. Cancer Genome Atlas Research Network. Electronic address aadhe, Cancer Genome Atlas Research N. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell*. 2017;32(2):185-203 e113.
 31. Hasegawa M, Takahashi H, Rajabi H, et al. Functional interactions of the cystine/glutamate antiporter, CD44v and MUC1-C oncoprotein in triple-negative breast cancer cells. *Oncotarget*. 2016;7(11):11756–11769.
 32. Yi S, Zhang H, Gong L, et al. Deep Sequencing and Bioinformatic Analysis of Lesioned Sciatic Nerves after Crush Injury. *PLoS One*. 2015;10(12):e0143491.
 33. Cai B, Ma W, Ding F, et al. The long noncoding RNA CAREL controls cardiac regeneration. *Journal of the American College of Cardiology*. 2018;72(5):534-550.
 34. Jin L, Qin C, Qi X, Hong T, Yang X, Zhu X. 2020.
 35. Hu J, Song K, Zhang J, Zhang Y, Tan BZ. Effects of menstrual blood-derived stem cells on endometrial injury repair. *Molecular medicine reports*. 2019;19(2):813-820.
 36. Chen F, Pan SX, Feng HL. [Distribution and content of transforming growth factor- β 1 and vascular endothelial growth factor in each layer of concentrated growth factors]. *Beijing da xue xue bao Yi xue ban = Journal of Peking University Health sciences*. 2016;48(5):860-865.
 37. Andrade Wde C, Silva LF, Coelho MC, Tannuri AC, Alves VA, Tannuri U. Effects of the administration of pentoxifylline and prednisolone on the evolution of portal fibrogenesis secondary to biliary obstruction in growing animals: immunohistochemical analysis of the expression of TGF- β and VEGF. *Clinics (Sao Paulo, Brazil)*. 2012;67(12):1455-1461.
 38. Liao X, Zhu G, Huang R, et al. Identification of potential prognostic microRNA biomarkers for predicting survival in patients with hepatocellular carcinoma. *Cancer Manag Res*. 2018;10:787-803.
 39. Liao X, Liu X, Yang C, et al. Distinct Diagnostic and Prognostic Values of Minichromosome Maintenance Gene Expression in Patients with Hepatocellular Carcinoma. *Journal of Cancer*. 2018;9(13):2357-2373.
 40. Zhang C, Zheng Y, Li X, Hu X, Qi F, Luo J. Genome-wide mutation profiling and related risk

- signature for prognosis of papillary renal cell carcinoma. *Ann Transl Med.* 2019;7(18):427.
41. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics.* 2005;61(1):92-105.
 42. Lin P, He RQ, Ma FC, et al. Systematic Analysis of Survival-Associated Alternative Splicing Signatures in Gastrointestinal Pan-Adenocarcinomas. *EBioMedicine.* 2018;34:46-60.
 43. Sun D, Tian L, Zhu Y, et al. Subunits of ARID1 serve as novel biomarkers for the sensitivity to immune checkpoint inhibitors and prognosis of advanced non-small cell lung cancer. *Molecular medicine (Cambridge, Mass).* 2020;26(1):78.
 44. Liu Y, Wu L, Ao H, et al. Prognostic implications of autophagy-associated gene signatures in non-small cell lung cancer. *Aging.* 2019;11(23):11440-11462.
 45. Liu GM, Zeng HD, Zhang CY, Xu JW. Identification of a six-gene signature predicting overall survival for hepatocellular carcinoma. *Cancer cell international.* 2019;19:138.
 46. He Z, Wang C, Xue H, Zhao R, Li G. Identification of a Metabolism-Related Risk Signature Associated With Clinical Prognosis in Glioblastoma Using Integrated Bioinformatic Analysis. *Front Oncol.* 2020;10:1631.
 47. Wang N, Gu Y, Li L, et al. Circular RNA circMYO9B facilitates breast cancer cell proliferation and invasiveness via upregulating FOXP4 expression by sponging miR-4316. *Archives of biochemistry and biophysics.* 2018;653:63-70.
 48. Jørgensen KJ, Paludan-Müller AS. Breast Cancer Screening and Diagnosis. *Annals of internal medicine.* 2020;172(12):838-839.
 49. Fhu CW, Ali A. Fatty Acid Synthase: An Emerging Target in Cancer. *Molecules (Basel, Switzerland).* 2020;25(17).
 50. Kim JA. Peroxisome Metabolism in Cancer. *Cells.* 2020;9(7).
 51. Zeleznik OA, Clish CB, Kraft P, Avila-Pacheco J, Eliassen AH, Tworoger SS. Circulating Lysophosphatidylcholines, Phosphatidylcholines, Ceramides, and Sphingomyelins and Ovarian Cancer Risk: A 23-Year Prospective Study. *Journal of the National Cancer Institute.* 2020;112(6):628-636.
 52. Feng WW, Kurokawa M. Lipid metabolic reprogramming as an emerging mechanism of resistance to kinase inhibitors in breast cancer. *Cancer Drug Resistance.* 2020;3(1):10.20517/cdr.22019.20100.
 53. Yang L, Cui X, Zhang N, et al. Comprehensive lipid profiling of plasma in patients with benign breast tumor and breast cancer reveals novel biomarkers. *Analytical and bioanalytical chemistry.* 2015;407(17):5065-5077.
 54. Hammad LA, Wu G, Saleh MM, et al. Elevated levels of hydroxylated phosphocholine lipids in the blood serum of breast cancer patients. *Rapid communications in mass spectrometry : RCM.* 2009;23(6):863-876.
 55. Kim H, Min HK, Kong G, Moon MH. Quantitative analysis of phosphatidylcholines and phosphatidylethanolamines in urine of patients with breast cancer by nanoflow liquid chromatography/tandem mass spectrometry. *Analytical and bioanalytical chemistry.* 2009;393(6-7):1649-1656.
 56. Min HK, Kong G, Moon MH. Quantitative analysis of urinary phospholipids found in patients with breast cancer by nanoflow liquid chromatography-tandem mass spectrometry: II. Negative ion mode analysis of four phospholipid classes. *Analytical and bioanalytical chemistry.* 2010;396(3):1273-1280.
 57. Vagia E, Mahalingam D, Cristofanilli M. The landscape of targeted therapies in TNBC. *Cancers.* 2020;12(4):916.
 58. Keenan TE, Tolaney SM. Role of Immunotherapy in Triple-Negative Breast Cancer. *Journal of the National Comprehensive Cancer Network : JNCCN.* 2020;18(4):479-489.
 59. Park SY, Choi JH, Nam JS. Targeting cancer stem cells in triple-negative breast cancer. *Cancers.*

- 2019;11(7):965.
60. Dees S, Ganesan R, Singh S, Grewal IS. Bispecific Antibodies for Triple Negative Breast Cancer. *Trends in cancer*. 2021;7(2):162-173.
 61. Garrido-Castro AC, Lin NU, Polyak K. Insights into Molecular Classifications of Triple-Negative Breast Cancer: Improving Patient Selection for Treatment. *Cancer discovery*. 2019;9(2):176-198.
 62. Huynh MM, Pambid MR, Jayanthan A, Dorr A, Los G, Dunn SE. The dawn of targeted therapies for triple negative breast cancer (TNBC): a snapshot of investigational drugs in phase I and II trials. *Expert opinion on investigational drugs*. 2020;29(11):1199-1208.
 63. Tacconi EM, Badie S, De Gregoriis G, et al. Chlorambucil targets BRCA1/2-deficient tumours and counteracts PARP inhibitor resistance. *EMBO molecular medicine*. 2019;11(7):e9982.
 64. Keung MY, Wu Y, Badar F, Vadgama JV. Response of Breast Cancer Cells to PARP Inhibitors Is Independent of BRCA Status. *Journal of clinical medicine*. 2020;9(4).
 65. Arun B, Akar U, Gutierrez-Barrera AM, Hortobagyi GN, Ozpolat B. The PARP inhibitor AZD2281 (Olaparib) induces autophagy/mitophagy in BRCA1 and BRCA2 mutant breast cancer cells. *International journal of oncology*. 2015;47(1):262-268.
 66. Chen XQ, Zhang F, Su QC, Zeng C, Xiao FH, Peng Y. Methylome and transcriptome analyses reveal insights into the epigenetic basis for the good survival of hypomethylated ER-positive breast cancer subtype. *Clinical epigenetics*. 2020;12(1):16.
 67. Kisková T, Kassayová M. Resveratrol action on lipid metabolism in cancer. *International Journal of Molecular Sciences*. 2019;20(11):2704.
 68. Xu H, Zhou S, Tang Q, Xia H, Bi F. Cholesterol metabolism: New functions and therapeutic approaches in cancer. *Biochimica et biophysica acta Reviews on cancer*. 2020;1874(1):188394.
 69. Ondruskova N, Cechova A, Hansikova H, Honzik T, Jaeken J. Congenital disorders of glycosylation: Still "hot" in 2020. *Biochimica et biophysica acta General subjects*. 2021;1865(1):129751.
 70. Guillaumond F, Bidaut G, Ouaisi M, et al. Cholesterol uptake disruption, in association with chemotherapy, is a promising combined metabolic therapy for pancreatic adenocarcinoma. *Proc Natl Acad Sci U S A*. 2015;112(8):2473-2478.
 71. Christodoulatos GS, Spyrou N, Kadillari J, Psallida S, Dalamaga M. The Role of Adipokines in Breast Cancer: Current Evidence and Perspectives. *Current obesity reports*. 2019;8(4):413-433.
 72. Ke ZB, Wu YP, Huang P, et al. Identification of novel genes in testicular cancer microenvironment based on ESTIMATE algorithm-derived immune scores. *J Cell Physiol*. 2021;236(1):706-713.
 73. Graney PL, Tavakol DN, Chramiec A, Ronaldson-Bouchard K, Vunjak-Novakovic G. Engineered models of tumor metastasis with immune cell contributions. *iScience*. 2021;24(3):102179.
 74. Bosch M, Sánchez-Álvarez M, Fajardo A, et al. Mammalian lipid droplets are innate immune hubs integrating cell metabolism and host defense. *Science*. 2020;370(6514):eaay8085.
 75. Li Z, Liu H, Luo X. Lipid droplet and its implication in cancer progression. *American journal of cancer research*. 2020;10(12):4112-4122.
 76. 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. *Proc Natl Acad Sci U S A*. 2017;114(32):E6556-e6565.
 77. Kwong SC, Jamil AHA, Rhodes A, Taib NA, Chung I. Metabolic role of fatty acid binding protein 7 in mediating triple-negative breast cancer cell death via PPAR- α signaling. *Journal of lipid research*. 2019;60(11):1807-1817.
 78. Apaya MK, Hsiao PW, Yang YC, Shyur LF. Deregulating the CYP2C19/epoxy-eicosatrienoic acid-associated FABP4/FABP5 signaling network as a therapeutic approach for metastatic triple-negative breast cancer. *Cancers*. 2020;12(1):199.
 79. Kao TJ, Wu CC, Phan NN, et al. Prognoses and genomic analyses of proteasome 26S subunit,

- ATPase (PSMC) family genes in clinical breast cancer. *Aging*. 2021;13(14):17970.
80. Zhou N, Chen X, Xi J, et al. Novel genomic targets of valosin-containing protein in protecting pathological cardiac hypertrophy. *Scientific reports*. 2020;10(1):18098.
81. Chang TJ, Wang WC, Hsiung CA, et al. Genetic variation of SORBS1 gene is associated with glucose homeostasis and age at onset of diabetes: A SAPHIRE Cohort Study. *Scientific reports*. 2018;8(1):10574.
82. Brandsma CA, Guryev V, Timens W, et al. Integrated proteogenomic approach identifying a protein signature of COPD and a new splice variant of SORBS1. *Thorax*. 2020;75(2):180-183.
83. Mei J, Xing Y, Lv J, et al. Construction of an immune-related gene signature for prediction of prognosis in patients with cervical cancer. *International immunopharmacology*. 2020;88:106882.
84. Gong C, Hu Y, Zhou M, et al. Identification of specific modules and hub genes associated with the progression of gastric cancer. *Carcinogenesis*. 2019;40(10):1269-1277.
85. Aakula A, Kohonen P, Leivonen SK, et al. Systematic Identification of MicroRNAs That Impact on Proliferation of Prostate Cancer Cells and Display Changed Expression in Tumor Tissue. *European urology*. 2016;69(6):1120-1128.
86. Cho WC, Jang JE, Kim KH, Yoo BC, Ku JL. SORBS1 serves a metastatic role via suppression of AHNK in colorectal cancer cell lines. *International journal of oncology*. 2020;56(5):1140-1151.
87. Zhu S, Xiang X, Xu X, Gao S, Mai K, Ai Q. TIR Domain-Containing Adaptor-Inducing Interferon- β (TRIF) Participates in Antiviral Immune Responses and Hepatic Lipogenesis of Large Yellow Croaker (*Larimichthys Crocea*). *Front Immunol*. 2019;10:2506.
88. Yokoyama S, Cai Y, Murata M, et al. A novel pathway of LPS uptake through syndecan-1 leading to pyroptotic cell death. *eLife*. 2018;7.
89. Igal RA. Stearoyl-CoA desaturase-1: a novel key player in the mechanisms of cell proliferation, programmed cell death and transformation to cancer. *Carcinogenesis*. 2010;31(9):1509-1515.
90. Gharbaran R. Advances in the molecular functions of syndecan-1 (SDC1/CD138) in the pathogenesis of malignancies. *Critical reviews in oncology/hematology*. 2015;94(1):1-17.
91. Spinler K, Bajaj J, Ito T, et al. A stem cell reporter based platform to identify and target drug resistant stem cells in myeloid leukemia. *Nat Commun*. 2020;11(1):5998.
92. Stow JL, Hung Y, Wall AA. Macropinocytosis: Insights from immunology and cancer. *Current opinion in cell biology*. 2020;65:131-140.
93. Betriu N, Bertran-Mas J, Andreeva A, Semino CE. Syndecans and Pancreatic Ductal Adenocarcinoma. *Biomolecules*. 2021;11(3).
94. Hollósi P, Váncza L, Karázi K, et al. Syndecan-1 Promotes Hepatocyte-Like Differentiation of Hepatoma Cells Targeting Ets-1 and AP-1. *Biomolecules*. 2020;10(10).
95. Reszegi A, Karázi K, Tóth G, et al. Overexpression of Human Syndecan-1 Protects against the Diethylnitrosamine-Induced Hepatocarcinogenesis in Mice. *Cancers (Basel)*. 2021;13(7).
96. Chen J, Tang J, Chen W, et al. Effects of syndecan-1 on the expression of syntenin and the migration of U251 glioma cells. *Oncol Lett*. 2017;14(6):7217-7224.
97. Sundberg EL, Deng Y, Burd CG. Syndecan-1 Mediates Sorting of Soluble Lipoprotein Lipase with Sphingomyelin-Rich Membrane in the Golgi Apparatus. *Developmental cell*. 2019;51(3):387-398.e384.
98. Gharbaran R, Goy A, Tanaka T, et al. Fibroblast growth factor-2 (FGF2) and syndecan-1 (SDC1) are potential biomarkers for putative circulating CD15+/CD30+ cells in poor outcome Hodgkin lymphoma patients. *Journal of hematology & oncology*. 2013;6:62.
99. Lu W, Li N, Liao F. Identification of Key Genes and Pathways in Pancreatic Cancer Gene Expression Profile by Integrative Analysis. *Genes*. 2019;10(8).
100. Bowlt Blacklock K, Birand Z, Biasoli D, et al. Identification of molecular genetic contributors to canine cutaneous mast cell tumour metastasis by global gene expression analysis. *PloS one*. 2018;13(12):e0208026.

101. Rodriguez CP, Wu QV, Voutsinas J, et al. A Phase II Trial of Pembrolizumab and Vorinostat in Recurrent Metastatic Head and Neck Squamous Cell Carcinomas and Salivary Gland Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2020;26(4):837-845.
102. Yang M, Al Zaharna M, Chen YS, Li L, Cheung HY. In vitro antioxidant activities and anti-proliferative properties of the functional herb *Abrus cantoniensis* and its main alkaloid abrine. *Food & function*. 2014;5(9):2268-2277.
103. Chen KC, Chen CY. In Silico Identification of Potent PPAR- γ Agonists from Traditional Chinese Medicine: A Bioactivity Prediction, Virtual Screening, and Molecular Dynamics Study. *Evidence-based complementary and alternative medicine : eCAM*. 2014;2014:192452.
104. Laskar S, Espino O, Bandyopadhyay D. Isolation, Solid-state Structure Determination, In Silico and In Vitro Anticancer Evaluation of an Indole Amino Acid Alkaloid L-Abrine. *Current cancer drug targets*. 2019;19(9):707-715.
105. Lee WY, Chen KC, Chen HY, Chen CY. Potential mitochondrial isocitrate dehydrogenase R140Q mutant inhibitor from traditional Chinese medicine against cancers. *BioMed research international*. 2014;2014:364625.
106. García-Tojal J, García-Orad A, Díaz AA, et al. Biological activity of complexes derived from pyridine-2-carbaldehyde thiosemicarbazone. Structure of. *Journal of inorganic biochemistry*. 2001;84(3-4):271-278.
107. Hancock CN, Stockwin LH, Han B, et al. A copper chelate of thiosemicarbazone NSC 689534 induces oxidative/ER stress and inhibits tumor growth in vitro and in vivo. *Free radical biology & medicine*. 2011;50(1):110-121.

Figure 1. Identification of metabolism-related genes and construction of a prognostic classifier. **(A)** A volcano plot of all DEGs is shown combined with $|\log_2FC|$ and an adjusted p-value. Red represents 644 upregulated DEGs. Green represents 1,088 downregulated DEGs. **(B)** Extraction of metabolism-related genes from the DEGs. **(C)** Heatmap of lipid metabolism-related DEGs between tumor and matched adjacent normal tissue. Different colors represent the expression trend of lipid metabolism-related DEGs in the two groups, red represents high expression of genes and blue represents low expression of genes. DEGs, differentially expressed genes; FC, fold change.

Figure 2. Functional enrichment analysis of candidate prognosis-related lipid metabolism DEGs **(A)** Forest map of lipid metabolism-related DEGs associated with prognosis. **(B)** Expression of prognostic lipid metabolism DEGs in TCGA, where the horizontal axis represents different gene, the vertical axis represents the gene expression distribution, where different colors represent different groups. Asterisks represent levels of significance *** $p < 0.001$, **** $p < 0.0001$. **(C)** Expression of prognostic lipid metabolism DEGs in GSE109169, where the horizontal axis represents different gene, the vertical axis represents the gene expression distribution, where different colors represent different groups. Asterisks represent levels of significance *** $p < 0.001$, **** $p < 0.0001$. **(D)** TOP15 GO term enrichment analysis of lipid metabolism-related DEGs. The abscissa indicates gene ratio and the enriched pathways were presented in the ordinate. **(E)** KEGG pathway enrichment analysis of lipid metabolism-related DEGs. The abscissa indicates gene ratio and the enriched pathways were presented in the ordinate. DEGs, differentially expressed genes; TCGA, The Cancer Genome Atlas; GSE, GEO Series; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Figure 3. Risk score independent of the prognostic analysis. **(A)** Univariate regression analysis of the training set. There are only two genes, *SDCI* and *SORBS1*, had p values less than 0.05. **(B)** Multiple regression analysis of the training set showed that the p values of the two genes were still less than 0.05. **(C)** The risk score of the training set. The horizontal axis represent risk score and sample

survival respectively. Vertical reference lines represent sample locations corresponding to median risk score values and sample survival time respectively. **(D)** The risk curve of the test set. **(E)** Validation of the risk curve using the GSE20685 dataset. **(F)** K–M survival curve of the high- and low-risk groups of the training set, there was significant difference in survival time between the two groups. **(G)** K–M survival curve of the high- and low-risk groups in the test set. **(H)** Validation of the K–M survival curve using the GSE20685 dataset. **(I)** Survival analysis of the signature in patients stratified by *SDC1* status. **(J)** Survival analysis of the signature in patients stratified by *SORBS1* status. **(K)** Survival ROC of the training set, AUC (Area under Curve) values of 1, 3 and 5 years were all greater than 0.6. **(L)** Survival ROC of the test set. **(M)** Validation of the survival ROC using the GSE20685 dataset. ROC, receiver operating characteristic curve; K-M, Kaplan-Meier; *SDC1*, *Syndecan-1*; *SORBS1*, *sorbin and SH3 domain containing 1*.

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. BC, breast cancer; TCGA, The Cancer Genome Atlas.

Figure 5. Nomogram for predicting survival probability in the TCGA dataset. **(A)** Independent prognosis-univariate COX analysis, the results showed that the p values of other factors except gender were all less than 0.05. **(B)** Independent prognosis-multivariate COX analysis, the results showed that age, pathological stage and riskScore were independent prognostic factors. **(C)** Nomogram to predict survival probability at 1, 3, and 5 years. **(D)** The calibration curves suggested that the nomogram was possibly under- or over-estimating the mortality. **(E)** The 3-year decision curve for the nomogram and other clinical traits. **(F)** The 5-year decision curve for the nomogram and other clinical traits. **(G)** Based on this model, 50 clinical cases were divided into low-risk and high-risk groups, with the high-risk group having a worse prognosis. The survival difference between the two groups was statistically significant. COX analysis, Cox proportional hazards model.

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. **(D)** Representative image of western blot analysis of *SDC1* protein. **(E)** Immunohistochemical staining (*SORBS1* protein) of BRCA and adjacent healthy groups; magnification, $\times 200$. **(F)** Immunohistochemical staining (*SDC1* protein) of BRCA and adjacent healthy groups; magnification, $\times 200$. **(G and H)** The AOD of *SORBS1* and *SDC1* protein in BRCA and adjacent healthy groups. Mean $\pm$ SEM; $n = 30$; *** $P < 0.001$ vs. adjacent healthy; **** $P < 0.0001$ vs. adjacent healthy; *SORBS1*, sorbin and SH3 domain containing 1; *SDC1*, *Syndecan-1*; AOD, average optical density.

Figure 7. MoA analysis results via CMap database. A heatmap showing each compound (perturbagen) from the CMap that shares a MoA (rows). The heatmap is sorted by descending number of compounds with a shared MoA. The compounds in this heatmap have an absolute value of enrichment score ≥ 0.5 and might be capable of targeting the metabolism-related signature. Two drugs (fenoterol and guanabenz) have the MoA of Adrenergic receptor agonist; Two drugs (cefaclor and cetergoline) share MoA of Bacterial cell wall synthesis inhibitor; Two drugs (fludrocortisone and fludroxycortide) share the MoA of Glucocorticoid receptor agonist; Two drugs (scriptaid and vorinostat) shared the MoA of HDAC inhibitor. Two drugs (danazol and Dydrogesterone) share the MoA of Progesterone agonist; Two drugs (econazole and slucanazole) shared the MoA of Sterol demethylase inhibitor. CMap, Connectivity Map; MoA, mode of action.

873 **Figure 8.** Interactions between prognostic genes and drugs used for cancer treatment. Red indicates
 874 that the drug increases expression of the biomarkers, while green indicates that the drugs decreases
 875 expression of the biomarkers.

876 **Table 1.** Primer sequences used for qRT-PCR.

Gene	Primer	Sequence (5'→3')	Length	Tm	GC %
<i>SDCI</i>	Forward Primer	ACGGCTATTCCCACGTCTC	19	54.58	57
	Reverse Primer	TCTGGCAGGACTACAGCCTC	20	56.54	60
<i>SORBS1</i>	Forward Primer	CACAATCGAGAACAGCAAAAACG	23	56.05	43
	Reverse Primer	ACCCGCCTACTGTCATCCTTT	21	56.95	52
<i>GAPDH</i>	Forward Primer	CTGGGCTACACTGAGCACC	19	55.46	63
	Reverse Primer	AAGTGGTCGTTGAGGGCAATG	21	57.03	52

877 **Table 2.** Univariate COX analysis of different lipid metabolism genes ($P < 0.05$).

id	HR	HR.95L	HR.95H	p-value
<i>LEPR</i>	1.080015	1.0446	1.11663	6.04E-06
<i>ADM</i>	1.023029	1.012557	1.033609	1.44E-05
<i>ADHIC</i>	1.05337	1.025811	1.081671	0.000121
<i>CCDC3</i>	1.010499	1.005017	1.01601	0.000168
<i>PLTP</i>	1.005733	1.002641	1.008835	0.000274
<i>PCK1</i>	1.02409	1.010228	1.038143	0.000619
<i>FGF2</i>	1.037334	1.015572	1.059562	0.000703
<i>G0S2</i>	1.001482	1.000617	1.002347	0.000783
<i>MGLL</i>	1.01302	1.005283	1.020816	0.000943
<i>HCAR2</i>	1.042764	1.016626	1.069574	0.001225
<i>ACACB</i>	1.015828	1.00608	1.025671	0.001412
<i>CD36</i>	1.002847	1.001089	1.004609	0.001494
<i>KLF4</i>	1.008258	1.003137	1.013405	0.001546
<i>ADH1B</i>	1.001686	1.000613	1.002761	0.00207
<i>DGAT2</i>	1.007378	1.00267	1.012108	0.002099
<i>PTGIS</i>	1.025369	1.009068	1.041933	0.002183
<i>LGALS12</i>	1.011648	1.004028	1.019325	0.00268
<i>CAVI</i>	1.002493	1.000862	1.004126	0.002729
<i>LRP1</i>	1.007875	1.002627	1.01315	0.003231

<i>TEK</i>	1.054722	1.017956	1.092815	0.00325
<i>ENPP2</i>	1.015092	1.00498	1.025306	0.003363
<i>RETSAT</i>	1.005345	1.001763	1.00894	0.003423
<i>ANXA1</i>	1.003369	1.001071	1.005673	0.004047
<i>GPAM</i>	1.005001	1.001526	1.008489	0.004761
<i>PLIN1</i>	1.001213	1.00037	1.002057	0.004782
<i>HSD17B13</i>	1.090967	1.026844	1.159093	0.004846
<i>ADIPOQ</i>	1.001466	1.000426	1.002506	0.005711
<i>ACSL1</i>	1.00275	1.000788	1.004715	0.005984
<i>BMX</i>	1.168945	1.045744	1.30666	0.006012
<i>CYP2U1</i>	1.154779	1.042081	1.279664	0.00602
<i>LIPE</i>	1.003918	1.001119	1.006724	0.006048
<i>AADAC</i>	1.074246	1.020586	1.130726	0.006155
<i>LPL</i>	1.001335	1.000377	1.002294	0.006318
<i>VAV3</i>	0.989454	0.981893	0.997073	0.006749
<i>FABP4</i>	1.000299	1.000082	1.000517	0.007006
<i>ANGPTL4</i>	1.011239	1.003002	1.019543	0.007398
<i>PDE3B</i>	1.038544	1.010131	1.067755	0.007536
<i>MLXIPL</i>	1.049762	1.012711	1.088168	0.008075
<i>CIDEA</i>	1.006311	1.001635	1.011008	0.008109
<i>EGR1</i>	1.000754	1.00019	1.001317	0.00872

<i>QKI</i>	1.056527	1.013967	1.100873	0.008763
<i>PDGFRA</i>	1.017727	1.004373	1.031259	0.00912
<i>CEBPA</i>	1.012191	1.002986	1.02148	0.009333

878 *COX analysis, Cox proportional hazards model; HR, hazard ratio.*

Figures

Figure 1

Identification of metabolism-related genes and construction of a prognostic classifier. (A) A volcano plot of all DEGs is shown combined with $|\log_{2} \text{FC}|$ and an adjusted p-value. Red represents 644 upregulated DEGs. Green represents 1,088 downregulated DEGs. (B) Extraction of metabolism-related genes from the DEGs. (C) Heatmap of lipid metabolism-related DEGs between tumor and matched adjacent normal tissue. Different colors represent the expression trend of lipid metabolism-related DEGs in the two groups, red represents high expression of genes and blue represents low expression of genes. DEGs, differentially expressed genes; FC, fold change.

Figure 2

Functional enrichment analysis of candidate prognosis-related lipid metabolism DEGs (A) Forest map of lipid metabolism-related DEGs associated with prognosis. (B) Expression of prognostic lipid metabolism DEGs in TCGA, where the horizontal axis represents different gene, the vertical axis represents the gene expression distribution, where different colors represent different groups. Asterisks represent levels of significance *** $p < 0.001$, **** $p < 0.0001$. (C) Expression of prognostic lipid metabolism DEGs in GSE109169, where the horizontal axis represents different gene, the vertical axis represents the gene expression distribution, where different colors represent different groups. Asterisks represent levels of significance *** $p < 0.001$, **** $p < 0.0001$. (D) TOP15 GO term enrichment analysis of lipid metabolism-related DEGs. The abscissa indicates gene ratio and the enriched pathways were presented in the ordinate. (E) KEGG pathway enrichment analysis of lipid metabolism-related DEGs. The abscissa indicates gene ratio and the enriched pathways were presented in the ordinate. DEGs, differentially expressed genes; TCGA, The Cancer Genome Atlas; GSE, GEO Series; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Figure 3

Risk score independent of the prognostic analysis. (A) Univariate regression analysis of the training set. There are only two genes, *SDC1* and *SORBS1*, had p values less than 0.05. (B) Multiple regression analysis of the training set showed that the p values of the two genes were still less than 0.05. (C) The risk score of the training set. The horizontal axis represent risk score and sample survival respectively. Vertical reference lines represent sample locations corresponding to median risk score values and sample

survival time respectively. (D) The risk curve of the test set. (E) Validation of the risk curve using the GSE20685 dataset. (F) K–M survival curve of the high- and low-risk groups of the training set, there was significant difference in survival time between the two groups. (G) K–M survival curve of the high- and low-risk groups in the test set. (H) Validation of the K–M survival curve using the GSE20685 dataset. (I) Survival analysis of the signature in patients stratified by *SDC1* status. (J) Survival analysis of the signature in patients stratified by *SORBS1* status. (K) Survival ROC of the training set, AUC (Area under Curve) values of 1, 3 and 5 years were all greater than 0.6. (L) Survival ROC of the test set. (M) Validation of the survival ROC using the GSE20685 dataset. ROC, receiver operating characteristic curve; K-M, Kaplan-Meier; *SDC1*, *Syndecan-1*; *SORBS1*, *sorbin* and *SH3 domain containing 1*.

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. BC, breast cancer; TCGA, The Cancer Genome Atlas.

Figure 5

Nomogram for predicting survival probability in the TCGA dataset. (A) Independent prognosis-univariate COX analysis, the results showed that the p values of other factors except gender were all less than 0.05. (B) Independent prognosis-multivariate COX analysis, the results showed that age, pathological stage and riskScore were independent prognostic factors. (C) Nomogram to predict survival probability at 1, 3, and 5 years. (D) The calibration curves suggested that the nomogram was possibly under- or over-estimating the mortality. (E) The 3-year decision curve for the nomogram and other clinical traits. (F) The 5-year decision curve for the nomogram and other clinical traits. (G) Based on this model, 50 clinical cases were divided into low-risk and high-risk groups, with the high-risk group having a worse prognosis. The survival difference between the two groups was statistically significant. COX analysis, Cox proportional hazards model.

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. (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; SORBS1, sorbin and SH3 domain containing 1; SDC1, Syndecan-1; AOD, average optical density.

Figure 7

MoA analysis results via CMap database.A heatmap showing each compound (perturbagen) from the CMap that shares a MoA (rows). The heatmap is sorted by descending number of compounds with a shared MoA. The compounds in this heatmap have an absolute value of enrichment score ≥ 0.5 and might be capable of targeting the metabolism-related signature. Two drugs(fenoterol and guanabenz) have the MoA of Adrenergic receptor agonist;Two drugs(cefaclor and cetergoline)share MoA of Bacterial cell wall synthesis inhibitor;Two drugs (fludrocortisone and fludroxycortide)share the MoA of

Glucocorticoid receptor agonist;Two drugs (scriptaid and 869 vorinostat)shared the MoA of HDAC inhibitor. Two drugs (danazol and Dydrogesterone) share the MoA of Progesterone agonist; Two drugs (econazole and slucanazole) shared the MoA of Sterol demethylase inhibitor. CMap, Connectivity Map; MoA, mode of action.

Figure 8

Interactions between prognostic genes and drugs used for cancer treatment. Red indicates that the drug increases expression of the biomarkers, while green indicates that the drugs decreases expression of the biomarkers.

Supplementary Files

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

- [SupplementaryFigure1.jpg](#)
- [SupplementaryFigure2.jpg](#)
- [SupplementaryFigure3.tif](#)
- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable3.xlsx](#)
- [SupplementaryTable4.xlsx](#)
- [SupplementaryTable5.xlsx](#)
- [SupplementaryTable6.xlsx](#)
- [SupplementaryTable7.xlsx](#)
- [SupplementaryTable8.xlsx](#)
- [SupplementaryTable9.xlsx](#)
- [SupplementaryTable10.xlsx](#)
- [SupplementaryTable11.xlsx](#)
- [SupplementaryTable12.xlsx](#)