Research Article

Development and Verification of a Prognostic Stemness-Related Gene Signature in Triple-Negative Breast Cancer

Xueqi Ou,[1] **Yeru Tan** [iD],[2] **Guanfeng Gao,**[1] **Song Wu,**[1] **Jinhui Zhang,**[1] **Hailin Tang** [iD],[1] **Hongbo Zhu** [iD],[2] **and Anli Yang** [iD][1]

[1]*Sun Yat-Sen University Cancer Center, State Key Laboratory of Oncology in South China,*
 Collaborative Innovation Center for Cancer Medicine, Guangzhou, China
[2]*Department of Medical Oncology, The First Affiliated Hospital, Hengyang Medical School, University of South China,*
 Hengyang, China

Correspondence should be addressed to Hongbo Zhu; wonderfulzhu1985@163.com and Anli Yang; yangal@sysucc.org.cn

Background. It is well known that cancer stem cells can induce cancer metastasis, which causes the majority of cancer-related death, especially in triple-negative breast cancer (TNBC). TNBC features a high metastatic rate and low metastasis-free survival and is regarded as the most malignant subtype of breast cancer. The purpose of this study is to explore prognostic biomarkers that can predict metastasis of triple-negative breast cancer. *Methods.* The human triple-negative breast cancerSUM149PT cells were used for the study. The cancer stem cell spheres (sum149-Stem) and paired adherent cancer cells (sum149-Tumor) were collected to extract total RNAs. RNA-seq was used to analysis the mRNA expression of cancer stem cells and paired adherent cancer cells. Two different gene expression omnibus datasets (https://www.ncbi.nlm.nih.gov/gds), GSE58812 and GSE33926, were used to explore the mechanism of different expression genes between stem cells and adherent cancer cells. Seven genes showed prognostic function in all datasets. The STITCH database (https://www.stitchdata.com/) was used to explore the possible metastasis-inhibiting drugs that can target the seven genes. Each gene expression was compared by Pearson analysis. The receiver operating characteristic curve (ROC) and Kaplan–Meier survival curve were performed to assess the metastasis prognostic ability of the seven-gene modeling two different GEO datasets. *Results.* A subset of 7 stemness-related genes (SRGs) containing UCN, ST3GAL5, FDPS, HK2, MALL, LMTK3, and CRHR2 were identified in three independent cohorts. Univariate Cox analysis showed that ST3GAL5 plays an antitumor role in TNBC metastasis, and the other 6 genes promote the metastatic progression of TNBC. The ability of the 7-SRGs gene Cox model to predict TNBC metastasis was constructed with the GSE58812 dataset. Most of the genes showed significant expression in patients with different risk levels. Additionally, the model showed predictive value in another GEO dataset of TNBC patients. ROC curves indicated that the seven-gene model has a significant predictive value of TNBC metastasis. *Conclusions.* Expression analysis of the 7-SRGs signature model at diagnosis has predictive value for metastasis in TNBC patients.

1. Introduction

Breast cancer is the most commonly diagnosed cancer and the leading cause of cancer-related death in the female population worldwide [1]. Breast cancer can be classified into five subtypes: luminal A, luminal B, HER2 (human epidermal growth factor receptor 2) overexpressing, triple-negative breast cancer (TNBC), and unclassified [2, 3]. TNBC is characterized by lacking of estrogen receptor (ER), progesterone receptor (PR), and HER2 by immunohisto-chemistry (IHC) (also defined by FISH). TNBC accounts for about 15–20% of all breast cancers [4]. Compared to other subtypes of breast cancer, women suffering from TNBC have a poor outcome and low survival rate, with great potential for metastasis and relapse [5]. However, the therapeutic option for this subtype of cancer is monotonous. Due to negative PR, ER, and HER-2, hormonal therapy (target on ER and PR) and treatment with trastuzumab (target on

HER-2) are ruled out [6, 7]. Therefore, chemotherapy with Anthracyclines or taxanes remains the standard of treatment. Although standard chemotherapy can be effective at early-stage TNBC, the median overall survival (OS) of those who develop the metastatic disease is only 4–12 months, per a recent retrospective multicenter analysis reported [8, 9]. Hence, new effective therapies for patients with metastatic disease are needed.

Advanced genomic profiling of TNBC has reported that TNBC harbors abundant cancer stem cells (CSCs), especially tumor cells with enriched ALDH1 and CD44+/CD24– expression signatures [10, 11]. Multiple studies have shown that CSCs serve as determining factors that contribute to the self-renewal capacity and heterogeneity of tumor cells [12]. CSCs are responsible for chemotherapy resistance as a result of several mechanisms, including high expression of multidrug resistance (MDR) transporters such as ATP-binding cassette (ABC) efflux transporters of P-glycoprotein (P-gp/ABCB1), multidrug resistance-associated protein 1 (MRP1/ABCC1), and breast cancer resistance protein (BCRP/ABCG2) [13]. As a result, current chemotherapeutic drugs can not totally eradicate CSCs, and there is an urgent need for novel compounds that target CSCs by regulating the expression of associated genes and their signaling pathways.

Cancer stem cells are a group of tumor cells with strong carcinogenic ability. These cells can self-renew to maintain the growth and proliferation of tumors and lead to the formation of metastasis. Studies have shown that cancer cells expressing stem cell markers have been detected in the blood of breast cancer patients, and when they are inoculated into immunodeficient mice, these cells will metastasis to bone, liver, lung, and other organs [14]. High expression of stem cell markers in cancer is associated with poor prognosis and metastatic recurrence [15]. The acquisition of stem cell characteristics is related to epithelial-mesenchymal transition (EMT). EMT is a crucial process of embryogenesis [16]. Neoplastic cells undergoing EMT could lose their intercellular adhesion and apical polarity and gain a migratory behavior that surpasses these barriers [17]. Additionally, the overexpression of EMT transcription factors promote tumor metastasis in breast cancer and pancreatic cancer cells [18, 19]. By transplanting EMT into immunodeficient mice, the increased tumor spread potential has been associated with EMT and tumor cell stemness. Promoting the expression of EMT, such as transcription factors, TWIST1 or SNAIL1 in breast epithelial cells increases the ability of secondary tumors caused by transplantation [20, 21]. Compared with traditional two-dimensional (2D) cell culture systems, three-dimensional (3D) analysis can better stimulate cell conditions in the body and affect the formation of subpopulations of cancer cells with stem cell-like properties. Compared with the 2D model, the 3D cancer stem cell model can better simulate the tumor microenvironment, promote the formation of ECM, and stimulate the increase in the expression of stemness-related genes. In vitro-grown mammospheres, subsequently implanted orthotopically in mice, showed that uPAR signaling can induce CSC-like behavior in breast cells [22]. Signatures of genes in CSCs are proven to predict tumor metastasis, which provides

evidence that CSCs may be metastatic precursors [23], for instance, a review released in 2017 revealed that several microRNAs (miRNAs) may also participate in the activation of CSC-related metastasis [24].

In our study, we extracted total RNA from cancer stem cell spheres (sum149-Stem) and paired adherent cancer cells (sum149-Tumor) and then analyzed the mRNA expression by RNA-seq. Afterwards, bioinformatics analysis was utilized to determine potential metastasis-correlated mRNA, and their prognostic ability on metastasis was assessed. Besides, we explored possible metastasis-inhibiting drugs. Consequently, a subset of 7 stemness-related genes (SRGs) was identified. All SRGs are significantly differentially expressed in the sum149-Stem sequence, and 5 of these genes have been documented to support cancer cell stemness. After knocking down LMKT3, we found that the formation of stem cell spheres decreased significantly and the invasion of triple-negative breast cancer cells was inhibited. Based on the correlation between cancer stem cells and metastasis, we created a gene signature with SRGs expression. The SRGs signature was proved to be of predictive value for metastasis in TNBC, which may provide a valuable therapeutic target for the treatment of TNBC with the metastatic disease through eliminating cancer stem cells.

2. Materials and Methods

2.1. Cancer Cell Culture. Human triple-negative breast cancer cell lines SUM149PT and BT549 were purchased from American Type Culture Collection (ATCC) and cultured according to the supplier's instructions.

2.2. Cancer Cell Sphere Formation. The adherent cancer cells were collected by trypsin digestion. The culture medium was discarded by centrifugation at 1000 rpm (157rcf) for 3 mins. The cells were washed with PBS 3 times and immediately suspended in a sphere formation medium. The sphere formation medium was made by supplementing RPMI1640 with 20 ng/ml EGF, 20 ng/ml bFGF, 5 ml Penicillin-Streptomycin solution, and 50 ml B27. About 100 cancer cells were added into the ultralow attachment dishes (Corning Inc., Corning, NY, USA) and incubated for 7–14 days in the condition of 37°C, 5% CO_2 for the formation of spheres. The culture medium was changed every 4 days.

2.3. RNA-Sequence Analysis of Cancer Cells. The total RNA of triple negative breast cancer cells was extracted by Trizol (Ambion) reagents. RNA samples were quantified by NanoDrop ND-2000 (Thermo Scientific) and OD 260/280 > 1.8, and total RNA concentration >100 ng was considered adequate for RNA-sequence analysis. The eukaryotic mRNA was enriched on mRNA Enriching Beads Oligo dT. The mRNA was disrupted into short fragments after adding the fragmentation buffer. Using the mRNA as a template, a strand was synthesized with a six-base random primer (random hexamer) and then the buffer, dNTPs, DNA polymerase I, and RNase H to synthesize double-stranded

DNA. AMPure XP bead was used to purify double-stranded cDNA. The purified double-stranded cDNA was first repaired, a tail was added, and the sequencing connector was connected, and then the fragment size was selected with AMPure XP beads. Finally, the PCR amplification was carried out and AMPure XP beads were used to purify PCR products to obtain the final library. The Qubit 2.0 is used for preliminary quantification, the library is diluted, and then the insert fragment size of the library is detected by Agilent 2100. After the insert fragment meets the expectations, the effective concentration of the library is obtained by using the real-time PCR(qPCR) method to ensure the quality of the library. The high-throughput sequencing platform was used to pool different libraries to flow cell BOT clusters according to the requirement of effective concentration and data volume (HiSeq/MiSeq) sequencing.

2.4. qRT-PCR and siRNA Transfection. Total RNA was extracted with TRIzol reagent (Invitrogen). qRT-PCR was performed with SYBR Premix Ex Taq (Takara). Primer information is listed in Supplementary Table 1. The siRNA of LMTK3 was designed through the DSIR website tool and purchased from GeneCopoeia (USA). The sense sequence of si-LMTK3 is 5′-GCAACUACAAGGAGGACUACU-3′, and the antisense sequence is 5′-UAGUCCUCCUUGUAGUUG CUG-3′. Transfection was conducted with Lipofectamine 3000 (Invitrogen).

2.5. Transwell Assay. Matrigel was melted and added into transwell chambers (BD Biosciences). 2×10^4 cells were suspended in serum-free culture medium and added into the upper chamber. Culture medium with 10% fetal bovine serum was added to the lower chambers. The cells were incubated at 37°C for 24 h. Methanol was added to fix the cells in the lower chamber. The invasive cells were stained using 0.01% crystal violet and counted under the microscope.

2.6. The Acquisition and Normalization Gene Expression Data. The workflow of this study is shown in Figure 1. All sample data were acquired from the gene expression omnibus. The inclusion criteria are as follows: molecular subtype diagnosed with triple-negative breast cancer; gene expression profile and valid metastasis data. Patients were excluded from the study if missing records of follow-up. Totally, 107 patients from GSE58812 and 48 patients from GSE33926 were included [25]. GEO data were normalized using quantile normalization. All gene expressions were transformed into the FPKM. The metastatic occurrence was used as the analysis outcome. The patients were divided into a high-expression group and a low-expression group according to the average gene expression. The univariate Cox regression analysis was performed if the low expression of a gene is a risk factor for patients.

2.7. The Analysis of TNBC Patient Data. The heatmap was applied to visualize the gene expression characteristics of different cell types with the Heatmap package in R. The common genes among the three gene sets were depicted with a Venn Diagram package in R software. DAVID database [26] was used to predict the biological functions and relative pathways of DEGs. Predictive correlation between genes and target drugs came from the STITCH database [27]. Forest plot was performed to visualize the hazard ratio and significance of the prognostic genes to tumor metastatic events. The correlation between the seven genes was analysis and visualized by the corrplot package with Pearson analysis in the R platform. The MFS Kaplan–Meier survival curves of each gene were depicted with the ggsurvplot package. Log-rank *t* test was used to calculate the significance of high or low-expression groups. The Cox proportional hazard model was employed to assess the prognostic value of the identified signature. A patient was classified into the high- or low-risk group by median risk scores. A two-tailed *P* value <0.05 was considered significant. AUC was calculated with the time ROC package in R(19). The signature model was established with multivariate Cox hazard ratio analysis. The risk score was calculated through the following formula:

$$
\begin{aligned}
\text{Riskscore} = & (0.139 * \text{expression level of CRHR2}) + (-0.800 * \text{expression level of ST3GAL5}) \\
& + (0.438 * \text{expression level of LMTK3}) + (0.027 * \text{expression level of MALL}) \\
& + (0.107 * \text{expression level of HK2}) + (0.460 * \text{expression level of FDPS}) + (0.209 * \text{expression level of UCN}).
\end{aligned}
$$

$$(1)$$

3. Results

3.1. Triple Negative Breast Cancer Data Acquisition. The whole study was conducted following the design in Figure 1. We compared the different expression profiles between TNBC adherent cancer cells and stem cells. Two GEO TNBC data sets were applied to explore the function of different expression genes (DEGs). The triple-negative breast cancer data set and clinical data were acquired from GEO (https://www.ncbi.nlm.nih.gov/geo/). 107 patients from GSE58812

and 48 patients from GSE33926 were included based on the inclusion criteria.

3.2. The Analysis of Different Expression Genes (DEGs). Genes withFDR (false discovery rate) < 0.05 and | log2FC|(log2FoldChange) > 1 were considered expressed significantly between different groups. A total of 2877 genes were found differently expressed in adherent cancer cells and cancer stem cells. The distinct expression profile of each gene
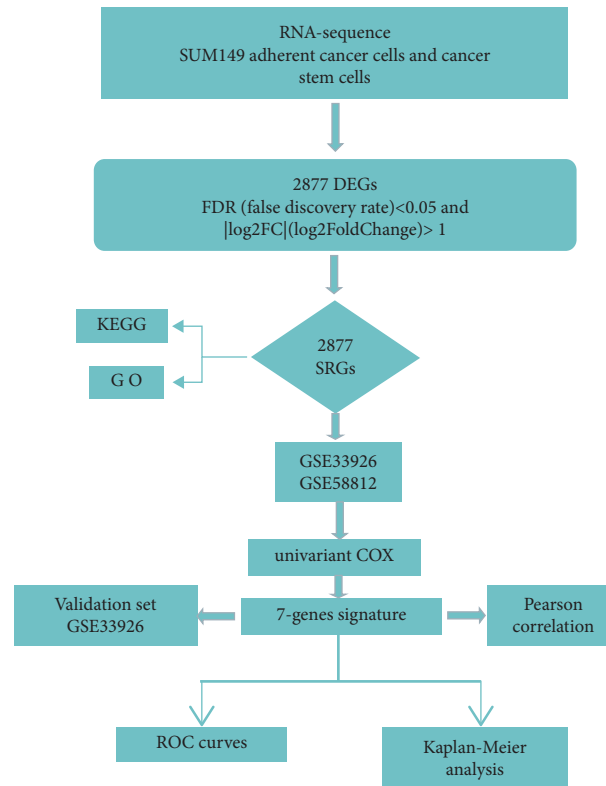
FIGURE 1: The flow chart of the research strategy.

in adherent cancer cells (T149) and cancer stem cells (S149) were shown in Figure 2(a), and 1827 genes were identified as upregulated in cancer stem cells, while 1049 genes down-regulated. GO (gene ontology) and KEGG analysis were conducted to explore the biological function of different expression genes (DEGs). GO analysis results identified DEGs closely related to development, procession, and cell differentiation (Figure 2(c)). The top 20 pathway terms enriched were shown in this study (Figure 2(d)). In cancer stem cells, KEGG analysis revealed that the DEGs were enriched in metabolism pathways such as glycine, serine and threonine, phenylalanine, steroid biosynthesis metabolism, and immunology pathways (complement and coagulation cascades).

3.3. Stemness-Related Genes(SRGs) Were Highly Associated with TNBC Metastasis.

To validate the relationship between SRGs and triple-negative breast cancer metastasis, the study selected two independent TNBC gene expression omnibus (GEO) datasets. All GEO datasets contain the information on cancer metastasis and gene expression. Univariate Cox analysis was applied to explore the prognostic value of SRGs in two GEO datasets. In the GSE58812 dataset, 888 genes were found to significantly affecting tumor metastasis. Additionally, 1297 genes were found to significantly affecting tumor metastasis in the GSE33926 dataset. 7 genes were found significantly affected TNBC metastasis, including CRHR2, MALL, LMTK3, UCN, FDPS, HK2, and

ST3GAL5 (Figure 3(a)). Setting clinical metastasis as the outcome, Kaplan–Meier curve was used to measure the predictive value of each of the seven genes on tumor metastasis. As the results showed, the high-expression of six genes, including CRHR2, MALL, LMTK3, UCN, FDPS, and HK2, are related to a high rate of metastasis (Figures 3(b)–3(g)). Combined with Figure 2(b), the six genes were upregulated in TNBC stem cells when compared with TNBC adherent cells, which can support the result of those Kaplan–Meier curves that the six genes are correlated with TNBC metastasis, respectively. Instead, high ST3GAL5 expression levels were significantly associated with a high metastatic ratio (Figures 3(h)), and 2(b) shown when comparing with TNBC adherent cells, the expression of ST3GAL5 was downregulated in TNBC stem cells, and summarize those evidence, ST3GAL5 may acts as an anti-tumor factor. To explore the potential drugs for inhibiting TNBC metastasis, the online gene-drug database STITCH was used to investigate the interactions between SRGs and potential target molecules. The 7 assumed target proteins were predicted to interact with 6 drugs in the STITCH database (Figures 3(i)–3(j)). Most of these drugs were confirmed having the antitumor ability.

3.4. The Construction of a Gene Signature with Seven SRGs for Predicting Metastasis of TNBC.

The forest plot showed the prognostic role of the seven SRGs for TNBC metastasis with an overall corrected hazard ratio (HR). Consistent with the
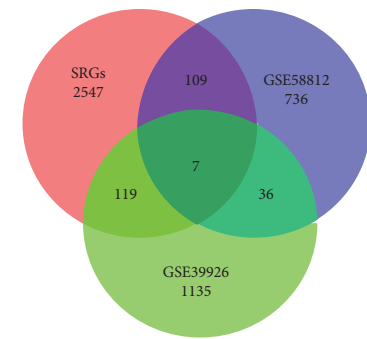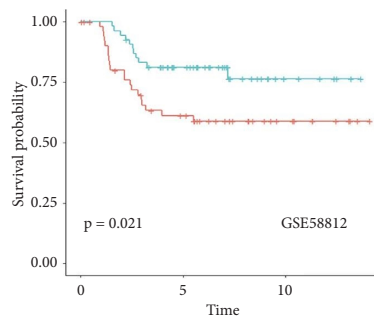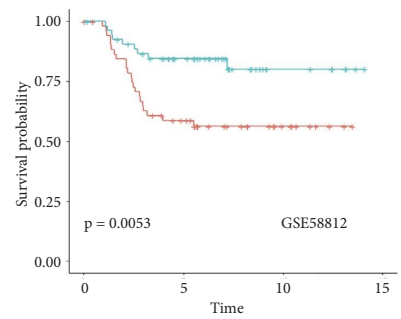
(a)

(b)
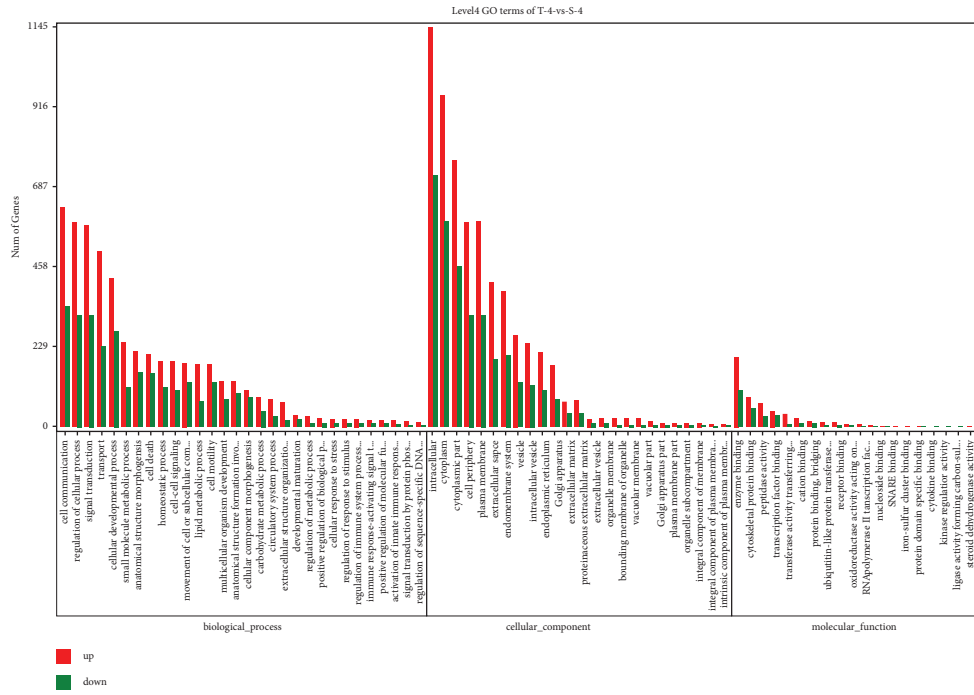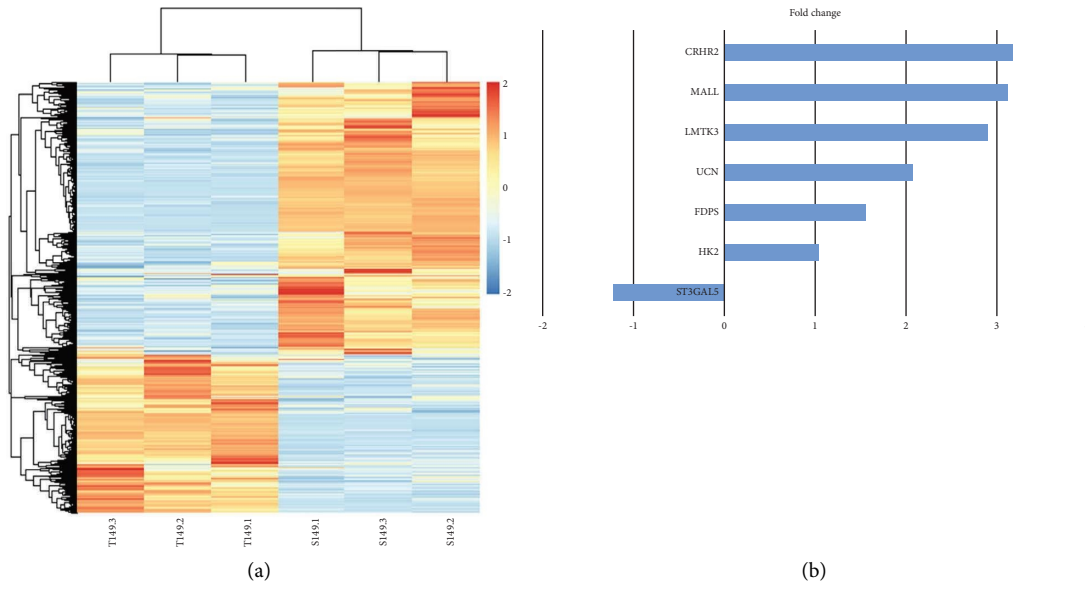
(c)

Figure 2: Continued.
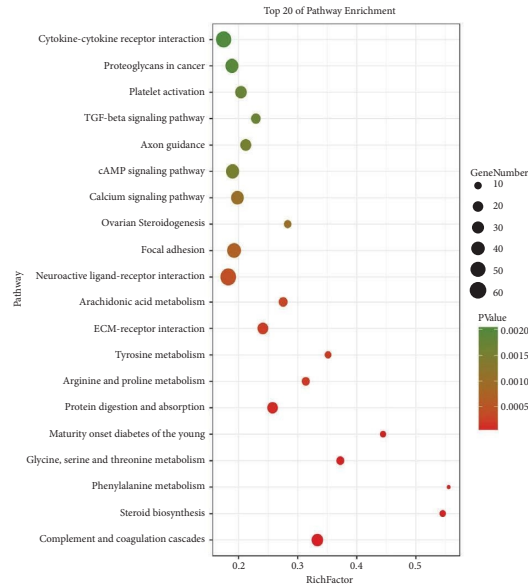
(d)

Figure 2: The differently expressed genes between TNBC stem cells and adherent cancer cells: (a) heat map showing a distinct expression profile between adherent cancer cells (T149) and cancer stem cells (S149), (b) the log2 fold change (logFC) indicates the mean expression level for each gene, (c) GO function of the SRGs, and (d) top 20 relative KEGG pathways of SRGs.
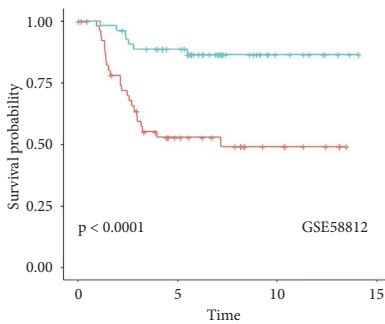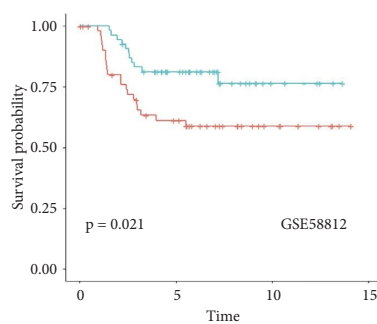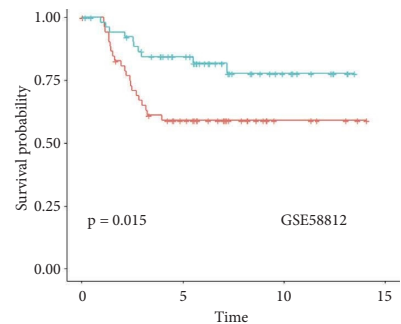

(a)

CRHR2
— group=high
— group=low

(b)

MALL
— group=high
— group=low

(c)

LMTK3
— group=high
— group=low

(d)

UCN
— group=high
— group=low

(e)

FDPS
— group=high
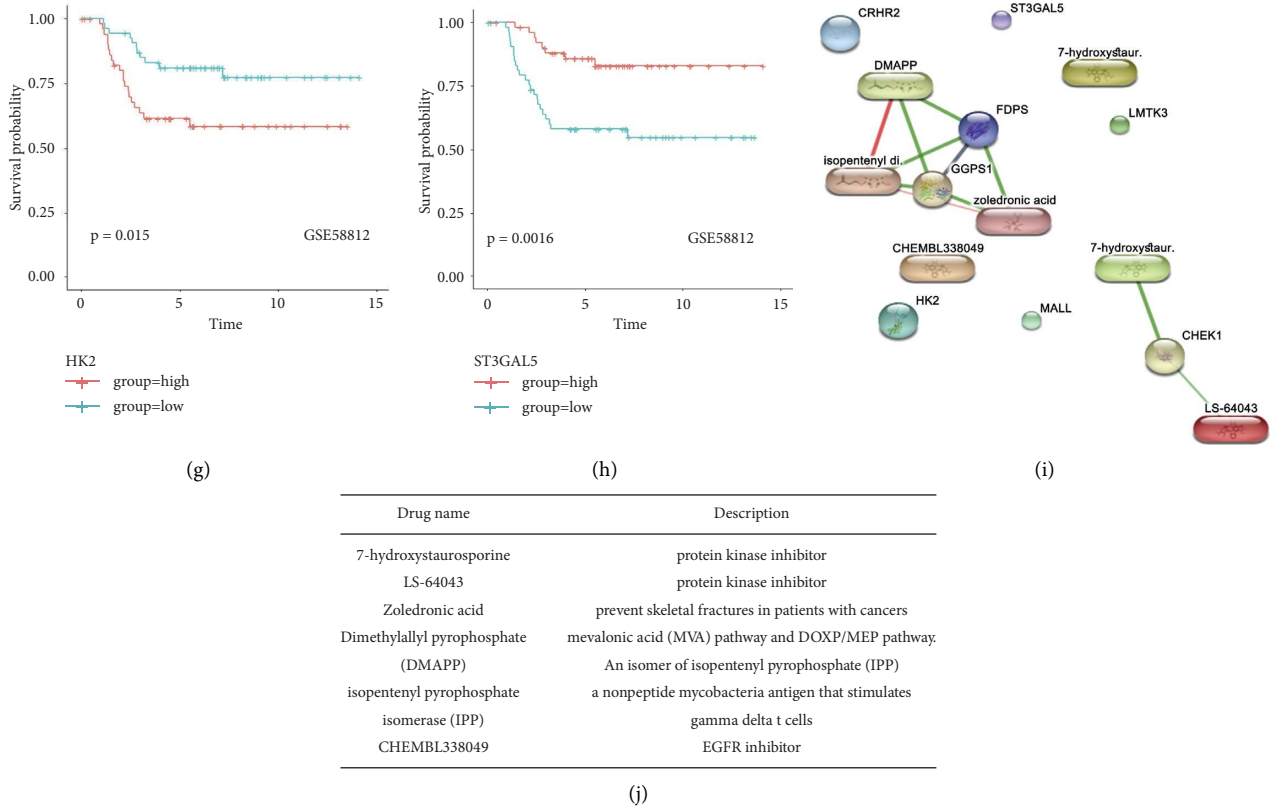— group=low

(f)

Figure 3: Continued.

Figure 3: Seven SRGs significantly affect the metastasis of TNBC. (a) The common prognostic genes among three cohorts. (b–h) The Kaplan–Meier survival curves of each of the seven SRGs. (i) The target drugs of the seven SRGs. (j) The description of the target drugs.

results of Kaplan–Meier curve, low-expression of ST3GAL5 is a risk factor for metastasis. Instead, the low expression of the other 6 genes can prevent metastasis (Figure 4(a)). Pearson correlation analysis revealed that the correlation between the expression levels of each gene within the 7 genes set (Figure 4(b)). Based on the previous result, we tried to constructed a multivariate Cox model to study whether the combination of the seven SRGs could act as a predicting index for metastasis of TNBC. Among these 107 patients from database GSE58812, 31 patients were diagnosed with clinical metastasis, and the remaining 76 patients were metastasis-free till the end of follow-up. The Cox regression coefficients were used in the risk prognostic model. The 107 patients were divided into high- and low-risk groups according to the risk score. The high-risk group is consisted of patients with prognostic scores higher or equal to the median of the total risk score, while the low-risk group is composed of patients with prognostic scores lower than the median. The distribution of the prognostic score and the status of metastasis of each patient are shown in Figure 4(c). Each of the seven genes expression between two different risk groups was shown in Figure 4(d). A time-dependent ROC curve with an AUC (area under curve) at 0.769, which indicated the capability of the 7-SRGs signature to predict metastasis at 18 months (Figure 4(e)).

3.5. Knockdown of LMKT3 Inhibit the Cell Sphere Formation and Invasion of Triple-Negative Breast Cancer Cells.

The expression level of the 7-SRGs were explored in triple-negative breast cancer cells. The results demonstrated that CRHR2, MALL, LMTK3, UCN, FDPS, and HK2 mRNA levels were upregulated in TNBC stem cells. While ST3GAL5 was downregulated (Figure 5(a)). After knocking down LMKT3, we found that the formation of stem cell spheres decreased significantly, and the invasion of triple-negative breast cancer cells was inhibited (Figures 5(b) and 5(c)).

3.6. The Seven SRGs Signature Precisely Predicts Metastasis of TNBC.

Notably, patients with high prognostic scores ($n = 20$) appeared prone to metastasis ($p = 0.02$) (Figure 6(a)) and poor survival outcome (Figure 6(b)). These results suggest that the 7-SRGs gene signature at diagnosis may be useful for the prediction of metastasis in TNBC patients. To further validate the significance of the 7-SRGs signature in predicting TNBC metastasis, we applied the Cox proportional hazard model to another independent GEO dataset (GSE33926). Consisting of the observations made with the discovery cohort, the the7-SRGs signature model divided the patients into the high- and low-risk groups based on the median risk score and the patients with high
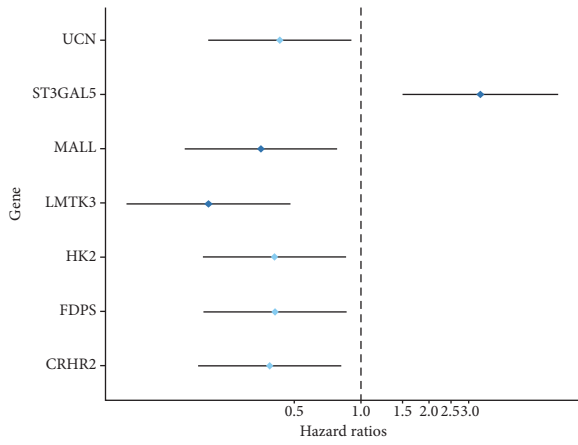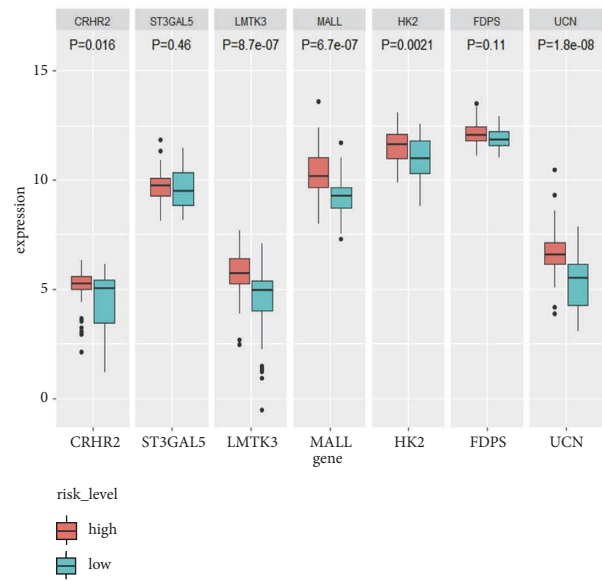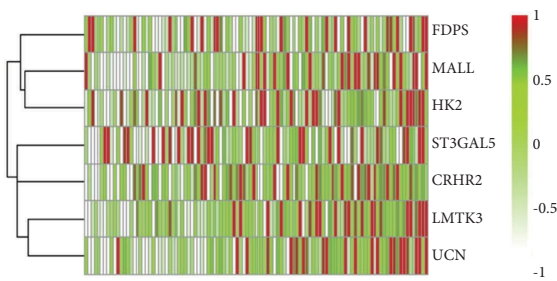
(a)
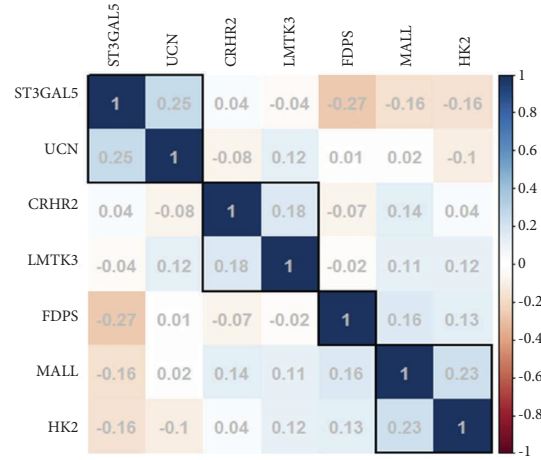
(b)
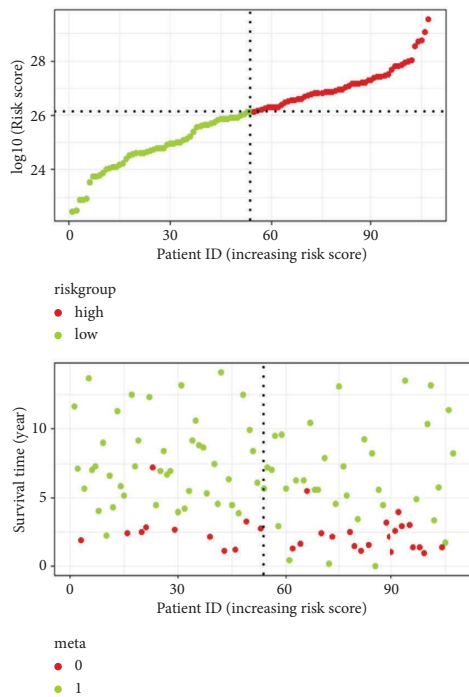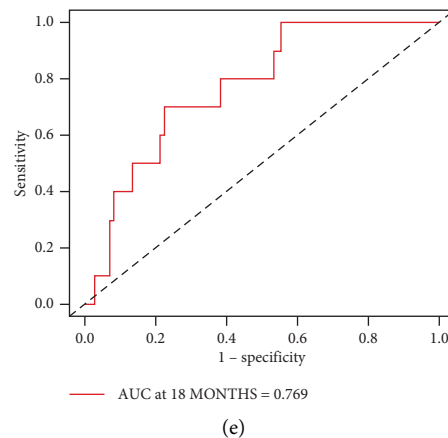

(c)

(d)

FIGURE 4: Continued.

(e)

FIGURE 4: The 7 SRGs gene signature model: (a) hazard ratio of each gene in the 7 SRGs gene signature model, (b) the correlation of the 7 SRGs genes, (c) risk score distribution and cut-off point distribution of breast cancer patient survival outcomes. Heat map showing the expression levels of the seven genes in breast cancer patients, (d) the expression level of each gene in the 7 SRGs gene signature between high-risk and low-risk groups, and (e) the ROC curves of the 7 SRGs gene signature.

prognostic scores ($n = 24$) appeared prone to metastasis ($p = 0.018$) (Figure 6(c)). When time-dependent ROC curves created at 1-, 3-, and 5-years were used to evaluate the power of the 7-gene signature to predict metastasis, the area under the curve (AUC) of the classification model reached 0.962 at 1-year, 0.785 at 3-years, and 0.837 at 5-years (Figure 6(d)). These results suggest that the 7-SRGs gene signature may also be useful as a predictive marker in TNBC patients.

4. Discussion

Triple-negative breast cancer is the subtype of breast cancer with the poorest prognosis. Due to the lack of hormone receptors, there are few effective target drugs for the treatment of TNBC. TNBC patients showed higher metastasis and re-currence rate compared to other breast cancer subtypes. It is well known that cancer stem cells are closely associated with cancer metastasis [28]. In this study, we compared the ex-pression profile between adherent and suspended cancer cells and validated a set of DEGs that might play roles in main-taining cancer cell stemness. As cancer stem cells are im-portant to the invasion and metastasis of neoplastic cells, two metastasis GEO datasets were used to explore the value of stemness-related genes in the progression of metastasis. After selecting the common genes in two GEO datasets, we found 7 genes are critical in the metastasis of TNBC patients.

In these 7 genes, the expression of ST3GAL5 is inversely correlated with the metastasis of TNBC. Other 6 genes promote the occurrence of TNBC metastasis. It is interesting that all seven proteins are upregulated in the high-risk group, indicating that the model can well divide the risk factors and is good for predicting the clinical outcomes. In the seven SRGs, urocortin (UCN) encodes a member of the corticotropin-releasing factor I family and is highly related to glioma stemness. UCN is an endogenous ligand for both corticotropin-releasing factor receptor 1 and corticotropin-releasing factor receptor 2(CRHR2), that belong to G-pro-tein-coupled receptors (GPCRs) and play different roles in the central and periphery nervous systems [29]. It was shown

that CRH was also related to TNBC metastasis. CRH family peptides were found in a series of endocrine-related-cancer tissues, such as breast cancer and prostate cancer. It is re-ported that CRH-UCN in cancer cells regulates cellular proliferation, apoptosis, and migration. CRH inhibited TGFβ1-induced EMT of breast cancer cells via CRHR1 and R2, where Snail1 and Twist-repressed E-cadherin expression is involved. CRH-UCN signaling pathway might play a vital role in triple-negative breast cancer stem cells. ST3 beta-galactoside alpha-2,3-sialyltransferase 5(ST3GAL5) is known to participate in the induction of cell differentiation, proliferation, and integrin-mediated cell adhesion [30]. It is reported that ST3GAL5 can prevent muscle invasion and the malignant progression of bladder cancer [31]. ST3GAL5 expression knockdown promotes the chemoresistance of AML cells. The expression of ST3GAL5 increased signifi-cantly in MCF-7 stem cells. However, this study showed that ST3GAL5 is downregulated in sum149 stem cells and cor-related with a good prognosis. The reason for this phe-nomenon might be due to the heterogeneity in different breast cancer subtypes. Lemur tyrosine kinase 3(LMTK3) is an oncogenic receptor tyrosine kinase (RTK) implicated in various types of cancers, including breast, gastric, and co-lorectal cancer and melanoma [32–35]. KRAB-associated protein 1 (KAP1) binding to LMTK3. PP1α stabilized LMTK3/KAP1 complex by specifically suppressing KAP1 phosphorylation at LMTK3-associated chromatin regions, inducing chromatin condensation and resulting in LMTK3-bound tumor suppressor-like gene silencing, and H3K9me3 modification [36]. Mal-T cell differentiation protein-like (MALL), is reported to be upregulated in ovarian carci-noma [37]. MALL overexpression was associated with high-grade serous ovarian carcinoma stem cells. Patients with MALL upregulation had poor clinical outcomes [38]. Hexokinase 2(HK2) phosphorylates glucose to produce glucose-6-phosphate, the first step in most glucose meta-bolism pathways. HK2 is expressed at a high level in cancer cells. HK2 ablation inhibits the neoplastic phenotype of human gallbladder, lung, and breast cancer cells both *in vitro*
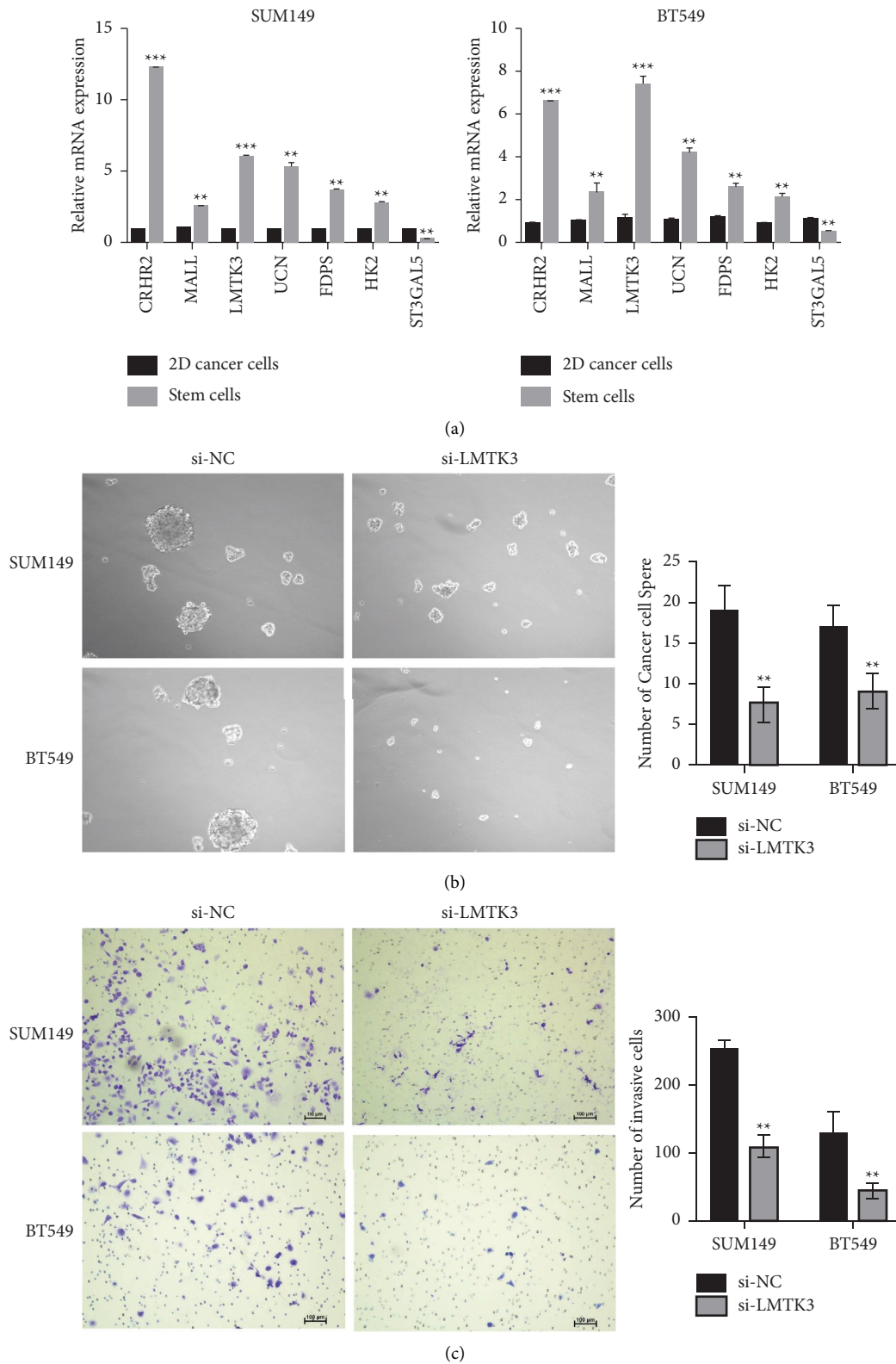
(a)

(b)

(c)

FIGURE 5: Knockdown of LMKT3 inhibits the cell sphere formation and invasion of triple-negative breast cancer cells: (a) the relative expression level of the 7 SRGs in SUM149 and BT549, (b) the rate of stem cell sphere formation decreased significantly in si-LMTK3 groups, and (c) transwell assay was performed to measure the invasion abilities in SUM149 and BT549 cells.
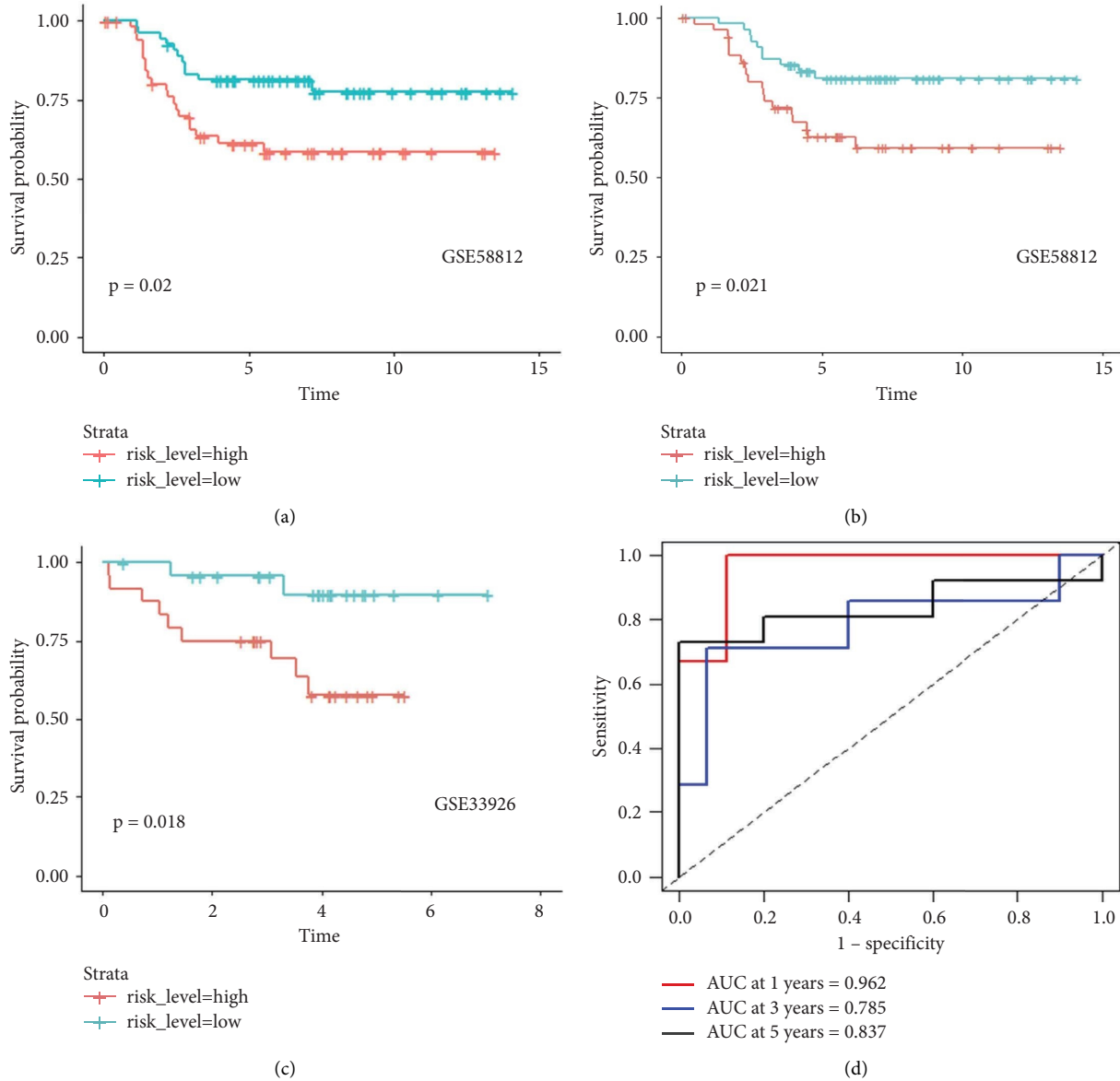
(a)

(b)

(c)

(d)

Figure 6: Prognostic value of the 7 SRGs signature model: (a) the Kaplan–Meier metastasis-free survival curves of the 7-SRGs signature in GSE58812, (b) the Kaplan–Meier overall survival curves of the 7-SRGs signature, (c) the Kaplan–Meier metastasis -free survival curves of the 7-SRGs signature in GSE33926, and (d) the ROC curves of the 7 SRGs signature in GSE33926.

and *in vivo* [39–41]. HK2 deletion in lung cancer cells suppressed glucose-derived ribonucleotides and impaired glutamine-derived carbon utilization in anaplerosis, thus inhibiting the proliferation of cancers that rely on rapid glucose metabolism [41]. HK2 can promote prostate cancer stem cell self-renewal by enhancing glucose metabolism. Farnesyl diphosphate synthase (FDPS), a mevalonate pathway enzyme that catalyzes the production of geranyl diphosphate and farnesyl diphosphate (FDP) from isopentenyl pyrophosphate and dimethyl-allyl pyrophosphate [42]. FDPS plays a critical role in cholesterol biosynthesis (CBS) and protein prenylation [43]. FDPS was found highly expressed in several cancers. Upregulated FDPS expression promotes PTEN-deficient prostate cancer proliferation by

activating AKT and ERK signaling by prenylation Rho A, Rho G, and CDC42 small GTPases [44]. The FDPS inhibitor significantly reduces the formation of glioma spheres. Most of the seven SRGs have close interactions. They play important roles in cancer progression. The model established by combing the expression of 7 SRGs might be a good prognostic marker of cancer metastasis.

In summary, we select the stemness-related genes that are significantly related to tumor metastasis in two independent cohorts. Based on the multivariant COX model, we established a model combining the expression of SRGs and the risk coefficient of the genes. Our results suggest that these seven proteins contribute to the metastasis of triple-negative breast cancer and have predictive value for clinical outcomes.

Data Availability

The data used to support the findings of this study are presented in the article, including GEO datasets (https://www.ncbi.nlm.nih.gov/gds) and the STITCH database (https://www.stitchdata.com/).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Xueqi Ou and Yeru Tan contributed equally to this work.

Acknowledgments

Supplementary Materials

Supplementary Table 1: Primer sequence of 7 stemness-related genes (UCN, ST3GAL5, FDPS, HK2, MALL, LMTK3, and CRHR2) are listed in the table. (*Supplementary Materials*)

References

[1] H. Sung, J. Ferlay, R. L. Siegel et al., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

[2] A. Prat, E. Pineda, B. Adamo et al., "Clinical implications of the intrinsic molecular subtypes of breast cancer," *The Breast*, vol. 24, no. 2, pp. S26–S35, 2015.

[3] J. Xie, S. Zheng, Y. Zou et al., "Turning up a new pattern: identification of cancer-associated fibroblast-related clusters in TNBC," *Frontiers in Immunology*, vol. 13, Article ID 1022147, 2022.

[4] A. C. Garrido-Castro, N. U. Lin, and K. Polyak, "Insights into molecular classifications of triple-negative breast cancer: improving patient selection for treatment," *Cancer Discovery*, vol. 9, no. 2, pp. 176–198, 2019.

[5] G. Bianchini, J. M. Balko, I. A. Mayer, M. E. Sanders, and L. Gianni, "Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease," *Nature Reviews Clinical Oncology*, vol. 13, no. 11, pp. 674–690, 2016.

[6] D. Ma, S. Y. Chen, J. X. Ren et al., "Molecular features and functional implications of germline variants in triple-negative breast cancer," *Journal of the National Cancer Institute*, vol. 113, no. 7, pp. 884–892, 2021.

[7] Y. Zou, S. Zheng, X. Xie et al., "N6-methyladenosine regulated FGFR4 attenuates ferroptotic cell death in recalcitrant HER2-positive breast cancer," *Nature Communications*, vol. 13, no. 1, p. 2672, 2022.

[8] L. A. Carey, E. C. Dees, L. Sawyer et al., "The triple negative paradox: primary tumor chemosensitivity of breast cancer

subtypes," *Clinical Cancer Research*, vol. 13, no. 8, pp. 2329–2334, 2007.

[9] Y. Zou, J. Xie, S. Zheng et al., "Leveraging diverse cell-death patterns to predict the prognosis and drug sensitivity of triple-negative breast cancer patients after surgery," *International Journal of Surgery*, vol. 107, Article ID 106936, 2022.

[10] G. Honeth, P. O. Bendahl, M. Ringner et al., "The CD44+/CD24- phenotype is enriched in basal-like breast tumors," *Breast Cancer Research*, vol. 10, no. 3, p. R53, 2008.

[11] E. Resetkova, J. S. Reis-Filho, R. K. Jain et al., "Prognostic impact of ALDH1 in breast cancer: a story of stem cells and tumor microenvironment," *Breast Cancer Research and Treatment*, vol. 123, no. 1, pp. 97–108, 2010.

[12] K. G. K. Deepak, R. Vempati, G. P. Nagaraju, V. R. Dasari, D. N. Rao, and R. R. Malla, "Tumor microenvironment: challenges and opportunities in targeting metastasis of triple negative breast cancer," *Pharmacological Research*, vol. 153, Article ID 104683, 2020.

[13] L. Zinzi, M. Contino, M. Cantore, E. Capparelli, M. Leopoldo, and N. A. Colabufo, "ABC transporters in CSCs membranes as a novel target for treating tumor relapse," *Frontiers in Pharmacology*, vol. 5, p. 163, 2014.

[14] S. Q. Geng, A. T. Alexandrou, and J. J. Li, "Breast cancer stem cells: multiple capacities in tumor metastasis," *Cancer Letters*, vol. 349, no. 1, pp. 1–7, 2014.

[15] L. Walcher, A. K. Kistenmacher, H. Suo et al., "Cancer stem cells-origins and biomarkers: perspectives for targeted personalized therapies," *Frontiers in Immunology*, vol. 11, p. 1280, 2020.

[16] S. Lamouille, J. Xu, and R. Derynck, "Molecular mechanisms of epithelial-mesenchymal transition," *Nature Reviews Molecular Cell Biology*, vol. 15, no. 3, pp. 178–196, 2014.

[17] P. Liu, Z. Wang, X. Ou et al., "The FUS/circEZH2/KLF5/feedback loop contributes to CXCR4-induced liver metastasis of breast cancer by enhancing epithelial-mesenchymal transition," *Molecular Cancer*, vol. 21, no. 1, p. 198, 2022.

[18] K. Zeng, B. He, B. B. Yang et al., "The pro-metastasis effect of circANKS1B in breast cancer," *Molecular Cancer*, vol. 17, no. 1, p. 160, 2018.

[19] M. V. Recouvreux, M. R. Moldenhauer, K. M. O. Galenkamp et al., "Glutamine depletion regulates Slug to promote EMT and metastasis in pancreatic cancer," *Journal of Experimental Medicine*, vol. 217, no. 9, Article ID e20200388, 2020.

[20] E. J. Tan, A. K. Olsson, and A. Moustakas, "Reprogramming during epithelial to mesenchymal transition under the control of TGFβ," *Cell Adhesion & Migration*, vol. 9, no. 3, pp. 233–246, 2015.

[21] M. S. Kim, H. S. Lee, Y. J. Kim, D. Y. Lee, S. G. Kang, and W. Jin, "MEST induces Twist-1-mediated EMT through STAT3 activation in breast cancers," *Cell Death & Differentiation*, vol. 26, no. 12, pp. 2594–2606, 2019.

[22] M. Jo, B. M. Eastman, D. L. Webb, K. Stoletov, R. Klemke, and S. L. Gonias, "Cell signaling by urokinase-type plasminogen activator receptor induces stem cell-like properties in breast cancer cells," *Cancer Research*, vol. 70, no. 21, pp. 8948–8958, 2010.

[23] A. Agliano, A. Calvo, and C. Box, "The challenge of targeting cancer stem cells to halt metastasis," *Seminars in Cancer Biology*, vol. 44, pp. 25–42, 2017.

[24] T. B. Steinbichler, D. Savic, J. Dudas et al., "Cancer stem cells and their unique role in metastatic spread," *Seminars in Cancer Biology*, vol. 60, pp. 148–156, 2020.

[25] W. H. Kuo, Y. Y. Chang, L. C. Lai et al., "Molecular characteristics and metastasis predictor genes of triple-negative

breast cancer: a clinical study of triple-negative breast carcinomas," *PLoS One*, vol. 7, no. 9, Article ID e45831, 2012.

[26] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nature Protocols*, vol. 4, no. 1, pp. 44–57, 2009.

[27] D. Szklarczyk, A. Santos, C. von Mering, L. J. Jensen, P. Bork, and M. Kuhn, "Stitch 5: augmenting protein-chemical interaction networks with tissue and affinity data," *Nucleic Acids Research*, vol. 44, no. 1, pp. D380–D384, 2016.

[28] E. Vagia, D. Mahalingam, and M. Cristofanilli, "The landscape of targeted therapies in TNBC," *Cancers*, vol. 12, no. 4, p. 916, 2020.

[29] L. Jin, C. Zhu, X. Wang et al., "Urocortin attenuates tgfβ1-induced Snail1 and slug expressions: inhibitory role of Smad7 in smad2/3 signaling in breast cancer cells," *Journal of Cellular Biochemistry*, vol. 116, no. 11, pp. 2494–2503, 2015.

[30] J. Y. Kan, S. H. Moi, W. C. Hung et al., "Comprehensive transcriptomic analysis identifies ST8SIA1 as a survival-related sialyltransferase gene in breast cancer," *Genes*, vol. 11, no. 12, p. 1436, 2020.

[31] S. Ouyang, J. H. Liu, Z. Ni, G. F. Ding, and Q. Z. Wang, "Downregulation of ST3GAL5 is associated with muscle invasion, high grade and a poor prognosis in patients with bladder cancer," *Oncology Letters*, vol. 20, no. 1, pp. 828–840, 2020.

[32] J. Stebbing, K. Shah, L. C. Lit et al., "LMTK3 confers chemoresistance in breast cancer," *Oncogene*, vol. 37, no. 23, pp. 3113–3130, 2018.

[33] C. Wang, M. Yang, X. Gu, and Y. Gu, "Lemur tyrosine kinase-3 (LMTK3) induces chemoresistance to cetuximab in colorectal cancer via the ERK/MAPK pathway," *Bioengineered*, vol. 12, no. 1, pp. 6594–6605, 2021.

[34] L. R. Klug, A. E. Bannon, N. Javidi-Sharifi et al., "LMTK3 is essential for oncogenic KIT expression in KIT-mutant GIST and melanoma," *Oncogene*, vol. 38, no. 8, pp. 1200–1210, 2019.

[35] A. Ditsiou, T. Gagliano, M. Samuels, V. Vella, C. Tolias, and G. Giamas, "The multifaceted role of lemur tyrosine kinase 3 in health and disease," *Open Biol*, vol. 11, no. 9, Article ID 210218, 2021.

[36] Y. Xu, H. Zhang, V. T. Nguyen et al., "LMTK3 represses tumor suppressor-like genes through chromatin remodeling in breast cancer," *Cell Reports*, vol. 12, no. 5, pp. 837–849, 2015.

[37] L. Zanotti, C. Romani, L. Tassone et al., "MAL gene overexpression as a marker of high-grade serous ovarian carcinoma stem-like cells that predicts chemoresistance and poor prognosis," *BMC Cancer*, vol. 17, no. 1, p. 366, 2017.

[38] G. Duns, E. Vigano, D. Ennishi et al., "Characterization of DLBCL with a PMBL gene expression signature," *Blood*, vol. 138, no. 2, pp. 136–148, 2021.

[39] J. Chen, Y. Yu, H. Li et al., "Long non-coding RNA PVT1 promotes tumor progression by regulating the miR-143/HK2 axis in gallbladder cancer," *Molecular Cancer*, vol. 18, no. 1, p. 33, 2019.

[40] L. Jiao, H. L. Zhang, D. D. Li et al., "Regulation of glycolytic metabolism by autophagy in liver cancer involves selective autophagic degradation of HK2 (hexokinase 2)," *Autophagy*, vol. 14, no. 4, pp. 671–684, 2018.

[41] K. C. Patra, Q. Wang, P. T. Bhaskar et al., "Hexokinase 2 is required for tumor initiation and maintenance and its systemic deletion is therapeutic in mouse models of cancer," *Cancer Cell*, vol. 24, no. 3, pp. 399–428, 2013.

[42] N. T. Ha and C. H. Lee, "Roles of farnesyl-diphosphate farnesyltransferase 1 in tumour and tumour microenvironments," *Cells*, vol. 9, no. 11, p. 2352, 2020.

[43] X. Wang, X. Zhang, Y. Chen et al., "Cardiac-specific deletion of FDPS induces cardiac remodeling and dysfunction by enhancing the activity of small GTP-binding proteins," *The Journal of Pathology*, vol. 255, no. 4, pp. 438–450, 2021.

[44] P. Seshacharyulu, S. Rachagani, S. Muniyan et al., "FDPS cooperates with PTEN loss to promote prostate cancer progression through modulation of small GTPases/AKT axis," *Oncogene*, vol. 38, no. 26, pp. 5265–5280, 2019.