Activation of the YY1-UGT2B7 axis promotes mammary estrogen homeostasis dysregulation and exacerbates breast tumor metastasis

Jiahao Xu\textsuperscript{1,2#}, Ying Zhou\textsuperscript{1#}, Shiqing He\textsuperscript{1}, Yinghao Wang\textsuperscript{1}, Jiachen Ma\textsuperscript{1}, Changwen Li\textsuperscript{4}, Zhao Liu\textsuperscript{3*}, Xueyan Zhou\textsuperscript{1*}

1. Jiangsu Key Laboratory of New Drug Research and Clinical Pharmacy, Xuzhou Medical University, Xuzhou, China. 221004.

2. The First People’s Hospital of Changzhou, Changzhou, China. 213003.

3. Department of Thyroid and Breast Surgery, the Affiliated Hospital of Xuzhou Medical University, Xuzhou, China. 221002.

4. Department of Breast Surgery, Xuzhou Central Hospital Xuzhou Clinical School of Xuzhou Medical University, Xuzhou, China. 221009.

\# These authors contributed equally to this work.
Running title:

Role and mechanism of YY1-UGT2B7 axis in breast cancer metastasis

*Corresponding Author:

Prof. Xueyan Zhou, Jiangsu Key Laboratory of New Drug Research and Clinical Pharmacy, College of Pharmacy, Xuzhou Medical University, 209 Tongshan Road, Xuzhou 221004, China. Phone: +86 516 83262630 E-mail: zxy851107@xzhmu.edu.cn

Prof. Zhao Liu, Department of Thyroid and Breast Surgery, the Affiliated Hospital of Xuzhou Medical University, 99 West Huaihai Road, Xuzhou, 221002, China, Tel: +0086-516-13852222788. Email: xylzhao9999@163.com

The number of text pages: 48

The number of tables: 3

The number of figures: 12

The number of references: 59

The number of words in the Abstract: 243

The number of words in the Introduction: 994

The number of words in the Discussion: 2007
Abbreviations:

YY1: Yin-yang 1; UGTs: UDP-glucuronosyltransferases; ER: estrogen receptor; UGT2B7: UDP-glucuronosyltransferase 2B7; E1: estrone; E2: 17β-estradiol; E3: estriol; 2-OHE2/1: 2-hydroxyestradiol/estrone; 4-OHE2/1: 4-hydroxyestradiol/estrone; 2-MeOE2/1: 2-methoxyestradiol/estrone; 4-MeOE2/1: 4-methoxyestradiol/estrone; 16-epiE3: 16-epiestriol; 17-epiE3: 17-epiestriol; 16α-OHE1: 16α-hydroxyestrone; CYP: cytochrome P450; UDPGA: glucuronide; BRCA1: breast cancer type 1 susceptibility protein; TMT: tandem mass tag; IHC: immunohistochemistry; real-time PCR: Quantitative reverse transcriptase PCR; Co-IP: Coimmunoprecipitation; ChIP: Chromatin immunoprecipitation assays; RFS: recurrence-free survival; OPLS-DA: Orthogonal Projections to Latent Structures-Discriminant Analysis;
Abstract

Metastasis is the most common pathway of cancer death. The lack of effective predictors of breast cancer metastasis is a pressing issue in clinical practice. Therefore, exploring the mechanism of breast cancer metastasis to uncover reliable predictors is very important for the clinical treatment of breast cancer patients. In this study, TMT quantitative proteomics technology was used to detect protein content in primary breast tumor tissue samples from patients with metastatic and non-metastatic breast cancer at diagnosis. We found that the high expression of YY1 is strongly associated with poor prognosis in high-grade breast cancer. YY1 expression was detected in both clinical tumor tissue samples and tumor tissue samples from MMTV-PyMT mice. We demonstrated that up-regulation of YY1 expression was closely associated with breast cancer metastasis and that high YY1 expression could promote the migratory invasive ability of breast cancer cells. Mechanistically, YY1 directly binds to the UGT2B7 mRNA initiation sequence ATTCAT, thereby transcriptionally regulating the inhibition of UGT2B7 expression, UGT2B7 can regulate the development of breast cancer by regulating estrogen homeostasis in the breast, and the abnormal accumulation of estrogen, especially 4-OHE2, promote the migration and invasion of breast cancer cells, ultimately causing the development of breast cancer metastasis. In conclusion, YY1 can regulate the UGT2B7-estrogen metabolic axis and induce disturbances in estrogen metabolism in breast tumors, ultimately
leading to breast cancer metastasis. Disturbances in estrogen metabolism in the breast tissue may be an important risk factor for breast tumor progression and metastasis.

**Keywords** Drug-metabolism-enzymes; Breast Cancer; Yin-Yang 1 (YY1); UDP-glucuronosyltransferase 2B7 (UGT2B7); Estrogen; Metastasis

**Significance Statement**

In this study, we propose for the first time a regulatory relationship between YY1 and UGT2B7/estrogen metabolism axis and explore the molecular mechanism. Our study shows that the YY1/UGT2B7/estrogen axis plays an important role in the development and metastasis of breast cancer. This study further elucidates the potential mechanisms of YY1-mediated breast cancer metastasis and the possibility and promise of YY1 as a predictor of cancer metastasis.

**Introduction**

Breast cancer is the most common cancer diagnosed worldwide and is one of the major causes of cancer deaths (Sung et al., 2021). Studies have reported that over 400,000 people die from breast cancer each year and that 90% of deaths are due to the development of metastases and related complications (Cancer Genome Atlas, 2012; Siegel, Miller, Wagle, & Jemal, 2023). In light of significant advances in diagnosis, surgery and the development of anticancer drugs, the survival rate for primary breast cancer is now close to 100%. However, once distant metastases occur, the survival rate drops to just 25% (Torre, Siegel, Ward, & Jemal, 2016). The lack of effective therapeutic targets and predictors is one of the key reasons for the high lethality of metastatic breast cancer.
Estrogen homeostasis is vital for various physiological processes such as energy metabolism regulation and sexual development. Emerging research indicates that the disruption of estrogen balance and subsequent abnormal accumulation of estrogen are critical factors in breast cancer development (Parida & Sharma, 2019; Zhuang et al., 2022). Estrogens have been implicated in the development of breast cancer due to their ability to stimulate cell growth and proliferation via receptor-mediated processes and the presence of toxic metabolites. The parent estrogens, estrone (E1) and estradiol (E2), are hypothesized to directly promote tumorigenesis by activating the estrogen receptor (ER) and initiating downstream pro-mitogenic transcriptional programs (Santen, Yue, & Wang, 2015).

The endogenous conversion of estrogen into genotoxic metabolites has been identified as a potential mechanism, independent of estrogen receptor (ER) signaling, contributing to estrogen-dependent breast tumorigenesis (Santen et al., 2015). After hydroxylation, estrogen forms catechol estrogens, specifically 2-hydroxyestrogens (2-OHE1/E2) and 4-hydroxyestrogens (4-OHE1/E2). These catechol estrogens can be converted into methoxyestrogens (2-MeOE1/E2, 4-MeOE1/E2) through methylation by COMT enzyme or further oxidized by cytochrome P450 (CYP) enzymes to form estrogen quinones. Estrogen quinones can chemically react with the guanine and adenine bases in DNA, leading to the formation of depurinating estrogen-DNA adducts (Dallal et al., 2014; Rajan et al., 2021). Studies have reported that concentrations of 4-OHE2 in breast cancer biopsies can be up to three times higher than those observed in normal breast tissue. Our previous studies have also found that
estrogen homeostasis is significantly dysregulated during the development of breast cancer (Zhou et al., 2017). In particular, concentrations of 4-OHE2/1 and 2-OHE2/1 were significantly elevated in the mammary tissue of breast cancer rats (Zhou et al., 2018). Therefore, it is reasonable to suggest that restoring estrogen homeostasis could serve as a valuable therapeutic approach for treating and potentially preventing breast cancer.

In addition to methylation, endogenous estrogens and catechol estrogens can undergo glucuronidation in the liver through the action of UDP-glucuronosyltransferases (UGTs), resulting in their conjugation with glucuronic acid. UGTs are phase II enzymes that catalyze the covalent addition of glucuronide (UDPGA) to facilitate the elimination and metabolism of estrogen (Lu, Lu, Wang, Wang, & Zhang, 2018). In contrast to other metabolic pathways of estrogen, the UGT-mediated process leads to the formation of glucuronides, which lack biological activity and are easily excreted from tissues into the circulation (Mitra, Basu, & Owens, 2009). In previous studies, we have established that UDP-glucuronosyltransferases (UGTs) play a significant role in the metabolism and elimination of estrogen. The glucuronidation capacity of UGTs has been shown to impact the regulation of estrogen signaling pathways and the pathogenesis of breast cancer (Hao et al., 2022). It has been reported that UGT1A1, UGT1A8 and UGT1A9 of the UGT1A family and UGT2B7 of the UGT2B family are the main metabolic enzymes that metabolize estrogen and related substances, with UGT2B7 having high expression specific to breast tissue (Zhao et al., 2020; Zhou et al., 2018). The literature reports that UGT2B7 is expressed at low levels or even
absent in breast tumor tissues, leading to the accumulation of estrogen in breast tissue, while supra-physiological concentrations of estrogen can mediate the overexpression of various growth factors, and promote cell growth and tumorigenesis (Guillemette, Belanger, & Lepine, 2004; Zhou et al., 2017). In our previous study, we verified that UGT enzymes contributes to estrogen elimination. The glucuronidation capacity of UGT enzymes influences the estrogen signaling pathway and the pathogenesis of breast cancer. Among them, UGT2B7 may play a key role in the elimination of estrogen (Hao et al., 2022). Thus, it is clear that UGT2B7 efficiently metabolizes estrogen and its toxic metabolites in the mammary tissue of rats with breast cancer and that its functional activity will directly affect the in vivo exposure, homeostatic balance and physiopathological activity of estrogen.

Yin-Yang 1 (YY1) is a zinc finger protein that was first identified as a member of the YY family (Shi, Seto, Chang, & Shenk, 1991). As a transcription factor, its downstream target genes are involved in a range of cellular processes involved in tumor progression, including cell proliferation, invasion, metastasis and angiogenesis (Khachigian, 2018). Structurally, YY1 has an activation domain in the N-terminus along with a repression domain in the C-terminus (Khachigian, 2018). Therefore, the transcriptional regulatory role of YY1 can either activate or suppress the transcription of its downstream genes, depending on the interaction between YY1 and its associated cofactors and the binding status of these target gene promoters (Thomas & Seto, 1999). There is growing evidence that YY1 can promote the development and progression of many cancers (Sarvagalla, Kolapalli, &
Vallabhapurapu, 2019). However, its functional role in breast cancer progression is controversial (Wottrich et al., 2017). YY1 can promote the invasion of Erb-B2 receptor tyrosine kinase 2 subtype breast cancers by upregulating the expression of ERBB2 and its transcriptional coactivator protein 2 (AP-2)(Hickish et al., 2022). Furthermore, YY1 has the ability to suppress cell proliferation by upregulating the expression of breast cancer type 1 susceptibility protein (BRCA1). Moreover, YY1 binds to the BRCA1 promoter, exerting positive regulatory effects (Lee et al., 2012).

However, the exact role of YY1 in high-grade breast cancer, especially in metastatic breast cancer, needs to be further elucidated. Therefore, this study will investigate whether YY1 can directly transcribe and regulate the expression of UGT2B7, thereby inducing abnormal accumulation of estrogen in the tumor site and ultimately promoting breast cancer metastasis.

Methods

Clinical sample collection

The clinical trial in this study was a Case-control study conducted at the Affiliated Hospital of Xuzhou Medical University from October 2017 to March 2023. The tumor tissues and serum samples of the breast cancer patients came from the affiliated hospital of Xuzhou Medical University, Xuzhou, China. During surgery, Venous blood samples were collected during the follicular phase, which was determined by a questionnaire survey. Additionally, premenopausal female patients with malignant breast cancer were also recruited to collect samples of cancerous tissues and para-
cancerous tissues. We collected a series of fresh-frozen tissues and clinical data from breast cancer patients. Tumor tissues were obtained from patients who underwent surgery and promptly stored at -80°C. None of the female participants included in our study had undergone systemic hormone therapy. The inclusion and exclusion criteria of this study were consistent with those previously reported. (Sampson et al., 2017). The experiment was approved by the Ethics Committee of the Affiliated Hospital of Xuzhou Medical University (No. XYFY2017-KL008-01) and this study was registered in the Chinese Clinical Trial Register on May 13, 2017 (No. ChiCTR-DOD-17011393) and was performed in accordance with the Declaration of Helsinki. Each patient was provided with written informed consent.

Animals and treatments

All the experiments that involved animals were approved and conducted under the oversight of the Animal Ethics Committee of Xuzhou Medical University (NO. 202208S040). Four-week-old female MMTV-PyMT transgenic mice were obtained from the Shanghai Southern Model Animal Centre (Shanghai, China), and nontransgenic FVB/n female siblings served as negative controls. Mice in each group were randomly assigned. This paper follows the growth characteristics of this model, which includes a proliferative phase at 4-5 weeks, an early carcinogenic phase at 9-10 weeks, and a late carcinogenic phase at 12-13 weeks. Therefore, MMTV-PyMT mice were divided into three groups, namely, the hyperplasia group, the early cancer group and the late cancer group, with six mice in each group.

For the ectopic implant trial, four-week-old female BALB/C mice were injected with
2×10⁶ cancer cells that were transduced with shCtrl or shYY1 lentiviral vector mixed with Matrigel (1:1) via subcutaneous injection. The humane endpoints were when the largest tumor size was >15 mm in diameter. None of the mice reached the endpoints of the present study. The laboratory mice were housed under controlled conditions, including a temperature range of 22–24°C, humidity levels between 40 and 60%, a 12-h light/dark cycle, and continuous access to food and water to ensure their well-being during research. The animal health and behaviour were monitored for animals every 3 days during the trial. Tumor volumes were evaluated every 3 days after injection. The tumor volume was calculated using the following formula: \( V = (L \times W^2) \times 0.5 \) (V, volume of tumor; L, length of tumor; W, width of tumor). After two months, mice were humanely sacrificed under 5% isoflurane for 10-20 min. Mice were placed into a chamber filled with vapor of the anesthetic isoflurane until respiration ceased and continued to be exposed to isoflurane until 2 min after respiratory arrest. The tumors were weighed, and volumes were measured for tumorigenesis evaluation.

For the colonization assay, BALB/C mice were injected with 2 x 10⁵ 4T1-GFP/LUC cells that were transduced with shCtrl or shYY1 lentiviral vector mixed with Matrigel. The humane endpoints were when the mice showed hind limb weakness or paralysis. None of the mice reached the endpoints of the present study. An IVIS kinetics imaging system (Caliper LifeSciences) was applied to monitor tumor metastasis via tail vein injection. To assess pulmonary metastasis in a nude mouse model and it was monitored every 3 days by the IVIS kinetics imaging system. After four weeks, the mice were humanely sacrificed under overdosed isoflurane and placed into a chamber.
filled with vapor of the anesthetic isoflurane until respiration ceased. The lungs were collected, fixed and sectioned. The number of metastatic lung nodules was counted using H&E staining.

**Bioinformatic analysis**

Cluster analysis: Hierarchical clustering analysis was performed using Cluster 3.0 and Java Treeview software. Definitions for technical terms were provided upon their initial mention. Similarities were evaluated using the Euclidean distance algorithm, and the average linkage clustering algorithm—leveraging observation centroids—was employed for clustering. Alongside the dendrogram, a heatmap was frequently included as a visual tool. Furthermore, standard academic formatting was adhered to, ensuring the text was devoid of grammatical or punctuation inaccuracies.

Subcellular localization: CELLO (http://cello.life.nctu.edu.tw/) which is a multi-class SVM classification system, was used to predict protein subcellular localization.

Domain annotation: Protein sequences are analyzed using InterProScan to identify specific protein domain signatures contained within the Pfam member database of InterPro.

GO annotation: The protein sequences of the selected differentially expressed proteins underwent a localized exploration through NCBI BLAST+ client software (ncbi-blast-2.2.28+-win32.exe) and InterProScan to locate homologous sequences. Following this, the discovered sequences were annotated, and gene ontology (GO) terms were associated using Blast2GO software. The results of the GO annotations were presented visually utilizing R scripts.
Enrichment analysis: Enrichment analysis was carried out utilizing Fisher's exact test, with the entire set of quantified proteins employed as the reference dataset. The Benjamini-Hochberg correction method was then utilized to rectify the obtained p-values for multiple comparisons. Only functional categories and pathways displaying p-values lower than 0.05 were considered statistically significant.

**Cell culture and cell treatment**

The breast cancer cell line MCF-7 was provided by Dr. Yanyan Yu. The breast cancer cell line MDA-MB-231, the breast cancer cell line 4T1 and HEK 293T cells were provided by Dr. Zhao Liu. MCF-7 cells and MDA-MB-231 cells were maintained in DMEM (KGM12800-500, Jiangsu Kaiji Biotechnology Co., Ltd.) supplemented with 10% FBS (FB25015, Clark) at 37°C in a 5% CO2 atmosphere. 4T1 cells were maintained in DMEM/F12 (C11875500BT, Gibco) supplemented with 10% FBS (FB25015, Clark). All the cell lines tested negative for mycoplasma.

**Transient transfection and lentiviral transduction**

A YY1 overexpression plasmid and negative control vector, YY1-specific shRNA and non-coding shRNA were synthesized by Genechem Co., Ltd. (Shanghai, China). The vectors used for shRNA and overexpression are GV344 and GV492. The sequence of the shRNA targeting YY1 was as follows: 5′- GCTTCGAGGATCAGATTCTCA -3′. The sequence of the control shRNA was as follows: 5′- TTCTCCGAACGTGTCACGT -3′. For transient transfection, cells were seeded in six-well plates at a density of 4×10^5 cells/well. After reaching 70% confluence, the cells were transfected with the shRNAs or plasmids. shRNA (80 pmol) or plasmids (2 pmol) were used for transient transfection.
were transfected into the cells using Lipofectamine 2000 reagent (GenePharma Co., Ltd.) at 37°C for 8 h according to the manufacturer's protocol. Eight hours after transfection, the cell culture medium was discarded and fresh DMEM containing 10% FBS was added to each well. After 48 h, transfection efficiency was further assessed by Western blot analysis and the following experiments were performed. For lentiviral transduction, the 2nd generation system was used in the lentivirus transduction experiment. MOI values and optimal infection conditions were determined by pre-testing. Cells were seeded in six-well plates at a density of 3×10^5 cells/well. After reaching 30% confluence, the cells were transfected with the shRNAs. The shRNAs (MOI, 40) were transduced into cells at 37°C for 24 h according to the manufacturer's protocol. The cell culture medium was then discarded and fresh DMEM containing 10% FBS was added to each well. After 72 h, transduction efficiency was observed using a fluorescence microscope. Stable YY1 knockdown or overexpression cell lines were selected using 2 µg/ml puromycin for 7 days. Western blot analysis was performed to detect the knockdown or overexpression efficiency of YY1.

**LC–MS/MS**

Estrone (E1), 17β-estradiol (E2), estriol (E3), 16-epiestriol (16-epiE3), 17-epiestriol (17-epiE3), 2-hydroxyestradiol/estrone (2-OHE2/1), 4-hydroxyestradiol/estrone (4-OHE2/1), 16α-hydroxyestrone (16α-OHE1), 2-methoxyestradiol/estrone (2-MeOE2/1), 4-methoxyestradiol/estrone (4-MeOE2/1), d5-E2, dansylchloride and vitamin C were purchased from Sigma Aldrich.

The analysis was performed with the HPLC series system (SCIEX ExionLC AD,
Singapore) coupled with a linear ion trap mass spectrometer as the detector (SCIEX TRIPLEQUAD 5500, AB Sciex Instruments, Singapore) under electrospray ionization (ESI) conditions in positive mode. Chromatographic separation was carried out using an Agilent ZORBAX Extend-C18 column (2.1 mm inner diameter x 100 mm length with a particle size of 1.8 µm) maintained at a temperature of 40°C. The separation process involved the use of mobile phases A (consisting of 100% H2O with 0.1% formic acid) and B (comprising 100% acetonitrile with 0.1% formic acid). The gradient elution was programmed in the following manner: initiating with 70% B for 2 minutes, followed by a linear increase from 30% B to 90% B over 2.5 minutes, a wash step with 90% B for 2 minutes, a gradual rise from 90% B to 95% B in 1.5 minutes, a wash with 95% B for 4 minutes, returning to 70% B over 0.5 minutes, and re-equilibration for 1.5 minutes, resulting in a total runtime of 13 minutes. The flow rate was set at 0.3 mL/min. MRM scan mode was used for positive ion detection; detection of ion pairs, residence time, fragmentation voltage and collision energy are listed in Table SI. The method of LC-MS/MS was carried out according to the method previously reported(Xu, Keefer, Ziegler, & Veenstra, 2007).

Quantitative reverse transcriptase PCR (real-time PCR)

RNA was extracted with TRIzol® reagent and isolated according to the manufacturer's protocol, and cDNA was generated using reverse transcription (PrimeScript RT kit). mRNA expression was measured by real-time PCR analysis using a LightCycler® 480 II. The primers were designed by Sangon Biotech and their specificity was checked by melting curve analysis. The sequences of the primers are
listed in Table SII.

**Western blotting**

After total protein extraction, the proteins were separated by SDS–PAGE, and their content was quantified by immunoblotting with specific antibodies. Primary antibodies included antibodies against YY1 (D595Z, Cell Signaling Technology), UGT2B7 (DF12140, Affinity), E-cadherin (BS1098, Bioworld), N-cadherin (22018-1-AP, Proteintech), Vimentin (BS4483, Bioworld), MMP2 (66366-Ig, Proteintech), MMP9 (10375-2-AP, Proteintech) and GAPDH (AP0063, Bioworld). The intensity of the bands was quantified using Odyssey®Sa (LICOR, USA). The expression of GAPDH was used as a control to quantify the signal intensity of the target protein.

**Wound healing assay**

The scratch assay was performed with $10^5$ cells in a 6-well plate. After the cell density reached 80%-90% confluence, a scratch was created under sterile conditions using a 200 μl sterile pipette tip, and the migration of cells toward the notch was observed. Then, the cells were washed twice with PBS to remove dead cells and cultured in FBS-free medium. Three replicate wells were performed in 6-well plates for each experimental condition.

**Coimmunoprecipitation (Co-IP)**

The interaction between YY1 and UGT2B7 was analyzed by Co-IP using YY1 and UGT2B7 antibodies. For CoIP assays, a total of approximately 500 μg of protein extract was incubated with 1 mg of normal IgG (Beyotime Biotech Inc., Jiangsu, China) and 10 μl of fully mixed Protein A/G Agarose (Beyotime Biotech Inc.,...
Jiangsu, China) before being gently shaken at 4°C for two hours to prevent non-specific binding. The supernatant was then collected and used for IP following centrifugation at 3000 g for 5 minutes. Next, 2 μg of either YY1 or UGT2B7 primary antibody was added to the supernatant and gently shaken overnight at 4°C. IgG was employed as a negative control. The following day, 20 μL of protein A/G Agarose in a completely suspended state was introduced into the reaction mixture and left at 4°C for four hours. The beads containing bound protein were then precipitated by centrifugation and thoroughly washed five times with Co-IP buffer. They were subsequently denatured by boiling in SDS-sample buffer before the resulting immunoprecipitate was subjected to immunoblot analysis. The method of Co-IP was carried out according to the method previously reported.

**Chromatin immunoprecipitation assays (ChIP)**

ChIP assays were used to measure the binding of YY1 to the CpG-rich region of the UGT2B7 promoter, which we detected using the ChIP kit (Epigentek, p-2001). The method of Co-IP was carried out according to the method previously reported. Quantitative PCR was performed to detect protein-associated promoter regions using the primers listed in Table SII.

**Promoter luciferase assay**

For 293T cells, 1x10⁵ cells were seeded in a 12-well plate and cotransfected with wild-type or mutant YY1 and Renilla. Luc (for normalization), and UGT2B7. The method of Co-IP was carried out according to the method previously reported.

**Immunohistochemistry (IHC)**
Formalin-fixed paraffin-embedded sections of primary and xenograft tumor tissue were used for IHC. IHC assays were performed as previously described (Rodriguez et al., 2020). The tissue slides were subjected to deparaffinization, rehydration, and epitope retrieval and were incubated at 4°C overnight with primary antibodies against YY1 (66281-1-Ig, Proteintech) and UGT2B7 (DF12140, Affinity). Slides were examined with a digital section scanning system (Olympus VS120 microscope). Linearity measurements were performed with Image-Pro Plus (Media Cybernetics, Silver Spring, MD).

**Statistical analysis**

Statistical analyses were performed using Prism8 software. Kaplan–Meier analysis was used to calculate the survival differences between divided groups, and a log-rank test was used to compare differences. All differences between groups were examined for statistical significance using a two-tailed Student's t test or one-way ANOVA to compare multiple groups; the results are presented as the mean ± SEM. A P value of < 0.05 was considered statistically significant for all the data sets. All in vitro and in vivo experiments were repeated at least three times.

**Results**

**TMT quantitative proteomics screening for differentially expressed proteins**

To identify key factors that promote breast cancer metastasis, we performed TMT (tandem mass tag) quantitative proteomics analysis of six metastatic breast cancer tumor tissues and six nonmetastatic breast cancer tumor tissues. The results showed
that the proteomics analysis yielded a total of 849433 secondary spectra (total number of database matched spectra 49449), 23092 peptides identified (total number of unique peptides 21098) and 4320 proteins identified (4313 quantifiable proteins) (Fig.1A). The proteins in the comparison groups were plotted on a volcano plot (Fig.1B) using the two factors of fold change and \( P \) value (T test) as criteria, with fold change (FC) > 1.2-fold (upregulation > 1.2-fold or downregulation < 0.83-fold) and \( P \) value < 0.05 (T test or other) as the criteria; 767 upregulated proteins and 798 downregulated proteins were identified between the comparison groups. Then, we performed subcellular localization analysis of all differentially expressed proteins using the subcellular structure prediction software CELLO, and the results showed that most of the differentially expressed proteins were localized in the nucleus (Fig.1C). The structural domain prediction software InterProScan was used to predict the structural domains of the differentially expressed proteins, and the results showed that the nucleic acid recognition region was a significantly enriched structural domain for the differentially expressed proteins (Fig.1D). GO functional annotation of all differentially expressed proteins was performed using Blast2Go (https://www.blast2go.com/) software, and the number of differentially expressed proteins was counted at the GO secondary functional annotation level (Fig.1E). To comprehensively figure out the functions, localization, and biological pathways related to differential proteins, protein annotation is performed through the GO analysis. The top twenty items in cellular components, biological processes, and molecular functions were listed in Fig. 1F-H. When referring to biological process
category, nuclear-transcribed mRNA catabolic process was significantly altered, and YY1 was involved in this process as a nuclear transcription factor (Fig. 1F). For the molecular function, structural molecule activity was changed significantly, and YY1 was involved in this function as it has a regulatory function in the activity of DNA-binding transcription factors (Fig. 1G). Thus, it is well reasonable to believe that this protein is relevant to the enriched GO terms.

Finally, we comprehensively screened 207 differentially expressed target proteins according to the screening principles (including t test, difference ploidy greater than 2, expression abundance analysis, etc.), including 106 proteins whose functions and structures were included in the NCBI and UniProt protein databases. We selected the top ten proteins with upregulated and downregulated ploidy for analysis and found that the most differentially expressed protein in the proteomics assay was YY1 (Fig.1I, Table SIII). YY1 was selected as the candidate protein that is located in the nucleus, has a regulatory function in DNA binding transcription factor activity and has the highest differential ploidy.

**High expression of YY1 in breast cancer is associated with metastasis**

A plethora of evidence suggests that YY1 plays a crucial role in cancer cell proliferation and tumor growth. However, the involvement of YY1 in cancer metastasis remains unclear. To address this issue, we initially employed immunohistochemistry (IHC) technique to examine the expression of YY1 in human breast cancer samples. As shown in Fig.2A, YY1 protein levels were dramatically increased in metastasized breast cancer samples. In addition, the protein levels of
MMP9, which is a common indicator protein in cancer metastasis, were detected in clinical samples. IHC analyses showed that MMP9 protein levels were significantly increased in metastasized breast cancer samples when compared with the primary breast cancer samples and the adjacent tissues (Fig.2B). Notably, the expression of the YY1 and MMP9 proteins exhibited a clear correlation ($R^2=0.2568$ and $P=0.003$, Fig.S1A). Furthermore, patients with breast cancer with either high YY1 and MMP9 mRNA levels had decreased recurrence-free survival (RFS) (Fig.2C-D). In addition, YY1 protein expression was significantly increased in a higher degree of breast cancer specimens (Fig.2E). These results indicated that elevated levels of YY1 expression are closely associated with the development and metastasis of breast cancer.

In an effort to characterize the dynamic changes in estrogen during breast cancer metastasis, we analyzed the changes in estrogen content at different stages, utilizing the mammary-specific polyomavirus middle T antigen overexpression mouse model (MMTV-PyMT) of breast cancer. MMTV-PyMT mice is characterized by multistep carcinogenesis, including hyperplasia at 4–5 weeks, carcinoma at 9–10 weeks, and advanced carcinoma at 12–13 weeks(Ershaid et al., 2019).

To investigate the changes in YY1 throughout breast cancer progression, we analyzed lung nodules in MMTV-PyMT mice at different times. As shown in Fig.S1B-C, MMTV-PyMT mice with advanced carcinoma developed distinct lung nodules, indicating that distant metastases had developed by this time. We examined tumor tissues from MMTV-PyMT mice at distinct tumourigenic stages including
hyperplasia group, carcinoma group, and advanced carcinoma group and found that YY1 expression increased significantly as breast cancer progression advanced, with the highest levels occurring when distal metastases occurred (Fig.2F). Moreover, as shown in Fig.2G, IHC staining also confirmed that YY1 expression increased significantly when breast cancer had metastasized. Taken together, these results suggest that elevated expression of YY1 is closely linked to breast cancer metastasis.

**Alteration of YY1 expression impacts cancer cell invasion and migration and the expression of EMT genes**

We found that YY1 knockdown in MCF-7 cells and MDA-MB-231 cells (Fig.3A, Fig.S2A-B) significantly inhibited migration and invasion ability (Fig.3C,E). In addition, we also found that YY1 overexpressed in MCF-7 and MDA-MB-231 cells (Fig.3B, Fig.S2C-D) significantly promoted breast cancer cell migration and invasion ability (Fig.3D,F).

The initiation of cancer metastasis, including invasion, migration and intravascular metastasis, is heavily dependent on tumor cell spread. Consistent with previous observations, reduced expression of YY1 resulted in changes in EMT genes, including E-cadherin, N-cadherin, Vimentin, MMP2 and MMP9, all of which are essential for the mobility and dissemination of cancer cells (Luo et al., 2018; Qin et al., 2018) (Fig.3G). Conversely, overexpression of YY1 upregulated these genes (Fig.3H). Taken together, these results suggest that upregulation of YY1 expression promotes the migration and invasion of breast cancer cells, ultimately leading to breast cancer metastasis.
YY1 regulates the UGT-mediated estrogen metabolic axis to promote breast cancer metastasis

First of all, this study has confirmed that alterations in YY1 expression can affect breast cancer metastasis. Abnormal estrogen metabolism plays a critical role in breast cancer metastasis, but it is unknown whether YY1 can regulate estrogen metabolism to affect breast cancer metastasis. Since UGT2B7 is a critical mediator of estrogen accumulation, we determined the effect of YY1 on UGT2B7 expression. As shown in Fig. 4A, Fig.S3A, we found that UGT2B7 expression was increased after knockdown of YY1 in MCF-7 and MDA-MB-231 cells. Not surprisingly, UGT2B7 expression was significantly reduced when YY1 was overexpressed (Fig. 4B, Fig.S3B). Other enzymes responsible for the estrogen metabolism, including COMT, CYP1A1 and CYP1B1 were also detected (Fig.S3C-D). These findings suggest that YY1 may regulate estrogen metabolism by modulating UGT2B7 expression.

To further elucidate whether YY1 directly suppressed UGT2B7 transcription, we used promoter luciferase assay to identify putative YY1 binding sites on the UGT2B7 promoter and found that mutation of one of these sites (-1396~1391) abolished YY1 inhibition of UGT2B7 expression. We identified the putative YY1 binding site on the UGT2B7 promoter by a dual luciferase reporter gene and found that mutation of one of the sites (TCCATT) eliminated the inhibitory effect of YY1 on UGT2B7 (Fig.4C). To understand how YY1 carried out this transcriptional suppression, we performed coimmunoprecipitation (Co-IP) assays and found that YY1 interacted directly with
UGT2B7 in breast cancer cells (Fig. 4D). To further investigate whether YY1 binds to the UGT2B7 promoter and undergoes transcriptional repression, we performed Chromatin immunoprecipitation (ChIP) assays using cells with or without YY1 knockdown. The results showed that YY1 bound to UGT2B7 at the initiation sequence and significantly suppressed UGT2B7 expression. We carried out the same experiments on cells with or without YY1 overexpression, and the conclusions were consistent (Fig. 4E).

Furthermore, animal experiments in vivo were conducted to verify that YY1 has a direct effect on UGT2B7. We found that the expression of UGT2B7 decreased significantly with cancer progression, especially when metastasis occurred and UGT2B7 was barely expressed. As shown in Fig. 4F, the expression of UGT2B7 in the tumor tissues of MMTV mice with advanced carcinoma was much lower than that of mice in other stages. In addition, we further validated this result by IHC experiments (Fig. 4G). We observed that UGT2B7 expression was lower in tumor tissues from metastatic breast cancer patients than in those from the primary group by IHC of tumor tissues from breast cancer patients; furthermore, UGT2B7 expression was highest in non-tumor tissues, and the results were highly consistent with previous results (Fig. 4H). Taken together, these results indicate that YY1 is a direct transcription factor of UGT2B7 that suppresses UGT2B7 expression through transcriptional regulation.

Abnormal accumulation of estrogen during breast cancer metastasis

Our group is working on exploring the detection of estrogen levels in serum and
tissues and has utilized a well-established method for the detection of estrogen based on LC–MS/MS (Hao et al., 2022). A total of 146 premenopausal patients with metastatic breast cancer (metastatic group), 132 premenopausal patients with primary breast cancer (primary group), and 143 matched control subjects (normal group) were included in the study. LC/MS-MS analytical method was employed to measure the concentrations of E1, E2, E3, and 11 other estrogen metabolites in serum samples. We found an increase in total estrogen accumulation as the breast cancer metastasis, and multivariate analysis (Orthogonal Projections to Latent Structures-Discriminant Analysis, OPLS-DA), which is an unsupervised statistical method, suggested that with the occurrence of breast cancer metastasis, there are changes in estrogen metabolism (Fig. 5A, Fig. S4A). Similarly, six cases of tumor tissue from patients diagnosed with metastatic breast cancer, six cases of tumor tissues from nonmetastatic breast cancer patients and six cases of paracancerous tissues were tested for hormones; analysis revealed estrogen accumulation in patients with metastatic breast cancer (Fig. 5B, Fig. S4B). To better substantiate our results, we performed the same hormone assays and analyses on serum and tumor tissue from MMTV-PyMT mice at different times, and the results once again demonstrated that breast cancer metastasis is closely associated with abnormal estrogen accumulation (Fig. 5C-D, Fig. S4C-D).

To further identify the key estrogens that promote breast cancer metastasis, we compared 14 estrogens. We found that the levels of 4-OHE2 are strongly associated with breast cancer progression and metastasis and those in metastatic breast cancer patients were highest levels of expression (Fig. S4E). The results of measuring the
levels of 14 estrogens in tumor tissue from breast cancer patients showed significantly increased concentrations of 4-OHE2 in metastatic breast cancer tissues compared to tissues from the another groups as well (Fig.5E). These results suggest that the estrogen balance is disrupted in breast cancer patients and that estrogen and its metabolites accumulate abnormally in breast cancer patients. Notably, we observed a more pronounced accumulation of estrogen in patients with metastatic breast cancer, confirming the close association between disturbances in estrogen metabolism and metastasis in breast cancer. Through analysis of serum and tumor tissue samples from MMTV-PyMT mice at different times, we found that the levels of 4-OHE2 in advanced carcinoma mice was significantly higher compared with the levels of estrogens in MMTV-PyMT serum at different stages (Fig.S4F), indicating that estrogens significantly accumulate in the serum of mice with metastatic breast cancer. As with clinical trials, we also measured the levels of estrogens in tumor tissues at different stages (Fig.5F). The results of the experiment were consistent with the previous results. All these experimental results proved that abnormal accumulation of estrogens is closely related to breast cancer metastasis.

Furthermore, whether the accumulation of 4-OHE2 contributes to the progression and metastasis of breast cancer was further confirmed, 1 μM 4-OHE2 stimulation was performed on MCF-7 cells and MDA-MB-231 cells (4-OHE2 group) and a blank control group (Control group) without any treatment, respectively. Scratch assays on MCF-7 cells and MDA-MB-231 cells in the Control and 4-OHE2 groups, respectively, showed that the migration ability of breast cancer cells was significantly
enhanced after the administration of 4-OHE2 stimulation (Fig.5G-H). This suggests that the accumulation of 4-OHE2 promotes the migratory ability of breast cancer cells, thus promoting breast cancer metastasis. Further, Transwell assays were performed on Control and 4-OHE2 groups of MCF-7 and MDA-MB-231 cells respectively, and the results showed that the invasive ability of breast cancer cells was significantly enhanced after administration of 4-OHE2 stimulation (Fig.5I-J). This suggests that the accumulation of 4-OHE2 can promote the invasive ability of breast cancer cells, thus promoting the metastasis of breast cancer.

**YY1 regulates the UGT2B7-estrogen metabolic axis to promote breast cancer metastasis**

These YY1-knockdown cells were then used to determine the role of YY1 in development of breast cancer by ectopic xenograft assays where LUC-expressing MCF-7 cells, without (shCtrl) or with (shYY1) YY1 depletion, were subcutaneously injected into immunocompromised BALB/C-NU/NU (nude) mice (Fig.6A). Eight weeks after subcutaneous injection, tumor size and tumor weight were evaluated. Mice bearing shCtrl cell-derived, YY1-expressing tumors had larger sizes and weights than the mice bearing shYY1 cell-derived, YY1-depleted tumors (Fig.6B, Fig.S5A-C). Moreover, in those experiments, 5 out of 6 YY1-shCON (83.3%) mice developed mammary tumors, and 3 out of 8 (37.5%) YY1-shRNA mice developed tumors (Fig.S5D), thus suggesting that suppression of YY1 expression reduces the development of breast cancer. Simultaneous xenograft experiments were conducted in which 4T1 cells without (shCtrl) or with (shYY1) YY1 depletion were
subcutaneously injected into Balb/c mice, following the same procedure as before (Fig.S5E). Interestingly, we also observed higher mRNA expression of YY1 and lower mRNA expression of UGT2B7 in the YY1-expressing tumors than in the YY1-depleted tumors (Fig.6C,Fig.S5F-G). It is reported that mouse UGT2B1 is equivalent to the functionality and drug selectivity to human UGT2B7, UGT2B1 was selected for this part instead of UGT2B7 expression assay (Jarrar, Jarrar, Abu-Shalhoob, Abed, & Sha'ban, 2019). We observed that UGT2B7 expression was lower in tumor tissues from the YY1-shCON group than in the YY1-shRNA group by IHC (Fig.6D-E). We further examined the changes in estrogen levels in serum and tissues in tumors and found that estrogen levels were significantly higher in the shCON group than in the shYY1 group (Fig.6F). As shown in Fig.S5H-I, there was no significant change in the mRNA expression of CYP1A1, CYP1B1, and COMT in the YY1-shRNA groups. Thus, we confirmed that YY1 can promote tumor growth at the primary site by downregulating UGTB7 to cause abnormal estrogen metabolism.

As metastasis is a crucial step to determine whether cancer cells can survive in distant organs, we injected 4T1 cells, with or without YY1 knockdown into mouse tail vein to examine whether YY1 contributes to metastasis ability. (Fig.6G). We further continued to observe the in vivo dynamics of tumour cells through small animal live imaging for 4 weeks after the initial tail vein injection showed that mice injected with YY1-depleted 4T1 cells had significantly less distal metastases than mice injected with control cells (Fig.6H). Therefore, YY1 can promote the distal metastasis and survival of breast cancer cells in vivo.
Taken together, this study suggests that YY1 inhibits the expression of UGT2B7 through transcriptional regulation, which causes abnormal accumulation of estrogen, especially 4-OHE2, and ultimately promotes the development of breast cancer progression and metastasis (Fig.7).

**Discussion**

The main unresolved challenge in breast cancer is the availability of effective treatment after metastasis. With advances in medical care, most breast cancer patients can be almost cured through early screening and diagnosis using different treatment strategies, such as targeted therapy, hormonal therapy, radiation therapy, surgery and chemotherapy (Bo et al., 2021; Fahad Ullah, 2019). However, for patients with distal metastases, there is a lack of effective clinical treatment, and the goal of treatment is usually only to improve quality of life and survival (Park et al., 2022). Thus, a deeper understanding of the molecular events that drive the progression of breast cancer metastasis and the development of more effective management strategies is imperative. In this study, we found a strong correlation between high expression of YY1 and poor prognosis in advanced breast cancer. Furthermore, our findings confirmed that YY1 exerts its effects on breast cancer metastasis by inhibiting the transcription of UGT2B7, leading to an imbalance in estrogen metabolism.

Substantial epidemiological, laboratory and clinical evidence suggests that increased endogenous estrogen is associated with an increased risk of breast cancer (Kulkoyluoglu-Cotul, Arca, & Madak-Erdogan, 2019; Patel, Homaei, Raju, &
Meher, 2018). It has been reported that estrogen promotes breast cancer progression through activation of the ER and that specific estrogen metabolites, mainly CE-3 and 4-Q, may contribute to cancer progression by mediating DNA damage (E. Cavalieri, Rogan, & Chakravarti, 2004). In this study, we showed that YY1 expression is significantly increased in breast tissue as breast cancer progresses, especially when distal metastases occur. Furthermore, we unexpectedly found that elevated YY1 could regulate the estrogen metabolic axis, leading to a significant accumulation of estrogens, especially estradiol and its hydroxylated metabolites, in the breast tissue. Mechanistically, YY1 regulates the estrogen metabolic axis by repressing the transcription of UGT2B7, leading to the progression and metastasis of breast tumors. There is growing evidence that YY1 can promote the development and progression of many cancers (Hua et al., 2018; Wang et al., 2020; Yang et al., 2020). However, its functional role in breast cancer progression has been controversial. One is that YY1 can promote the invasion of Erb-B2 receptor tyrosine kinase 2 subtype breast cancers by upregulating the expression of ERBB2 and its transcriptional coactivator AP-2 (Begon, Delacroix, Vernimmen, Jackers, & Winkler, 2005; Harbeck et al., 2016). Another is that YY1 can bind to and positively regulate the BRCA1 promoter and inhibit cell proliferation through BRCA1 expression in breast cancer (Khachigian, 2018).

Interestingly, we performed a comparison of tumor tissues collected from patients with metastatic breast cancer and primary breast cancer by proteomic analysis, which is a biotechnology widely used in oncology for molecular target mining (Ding, Wang,
Ji, & Chen, 2022). We found that YY1 was the most significantly differentially expressed protein and was significantly overexpressed in metastatic breast cancer tissues (Fig.2). Therefore, we speculated that elevated expression of YY1 is closely associated with the development of breast cancer. To further substantiate our conjecture, we chose MMTV-PyMT spontaneous breast cancer mice to mimic the dynamics of mammary tumor development in humans. The mouse model (MMTV-PyMT) is the most commonly used GEMM for cancer research (Lin et al., 2003). Specifically, in the context of breast cancer, the molecular and histological progression observed in mammary lesions of MMTV-PyMT mice closely resembles that of human mammary gland disease. Therefore, this model serves as a valuable tool for cancer researchers aiming to enhance their understanding of tumor biology (Maglione et al., 2001; Rennhack et al., 2019). We found that YY1 was significantly elevated in tumor tissue as breast cancer progressed, particularly with the development of distal metastases (Fig.2F-G). This is highly consistent with our conjecture. Next, we performed extensive in vitro studies on breast cancer cells and found that upregulation of YY1 expression promoted the migration and invasive ability of breast cancer cells, and conversely, downregulation of YY1 expression significantly inhibited the migration and invasive ability of breast cancer cells (Fig.3, Fig.S2). Although a limitation of the study is that the knockdown efficiency for YY1 is not very strong, even the modest knockdown achieved caused significant differences in subsequent assays. As YY1 is expressed in all tissues and organs of the body, it is essential for growth and development, and it has been previously reported
that the survival time of mice subjected to systemic YY1 knockout is very short, which is not conducive to experiments. Therefore, instead of breeding MMTV-PyMT mice with a transgenic model of YY1 knockout, we used in situ implantation of BALB/C mice and tail vein injection of breast cancer cells, two conventional tumor research tools that allow us to more clearly observe the effect of altered YY1 expression on breast cancer progression and metastasis. We found that YY1 can regulate the progression and metastasis of breast cancer (Fig.6, Fig.S5).

Previous studies have shown that as breast cancer progression occurs, estrogens, particularly estradiol and its hydroxylated metabolites, accumulate significantly in breast tissue and that tumor metastasis can be triggered by estrogens through activation of the immune microenvironment and production of inflammatory vesicles (Frontiers Production, 2019; Segovia-Mendoza & Morales-Montor, 2019). In this study, LC–MS/MS was used for quantification of the abundance of estrogens across multiple tissues and serum samples as previously described. We demonstrated the correlation between estrogen and breast cancer progression again in this study. We recruited 146 premenopausal female patients 132 premenopausal female patients and 143 premenopausal healthy female women (Sampson et al., 2017). To fully illustrate the role of dysregulated estrogen balance and breast cancer progression, we tested both serum and tumor tissue from different subgroups of breast cancer patients. As shown in Fig.5, the concentrations of multiple estrogens were significantly increased in breast cancer patients, particularly in patients with malignant metastatic breast cancer. In addition, we found that estrogen hydroxylated metabolites (4-OHE2)
accumulated significantly in breast tumor tissue compared to estrogen levels in paracancerous tissue. It is reported that the estrogens including E2, 4-OHE2 and 16-OHE1 could display a proliferative effect on MCF-7 cells which is accompanied by a down-regulation of apoptosis (Seeger, Wallwiener, Kraemer, & Mueck, 2006). To further verify whether 4-OHE2 has a facilitative effect on breast cancer migration invasion, this study was performed on two different breast cancer cell models and the results reconfirmed this. In addition to this, we also found that 4-OHE2 has a promoting effect on ER-negative breast cancer cells. It has been shown that 4-OHE2 can also promote breast cancer development by causing DNA double-strand breaks in estrogen receptor-negative breast cells and thus increasing the associated DNA damage (Savage et al., 2014). The results are consistent with previous studies, suggesting that disturbances in estrogen metabolism are closely associated with breast cancer progression and metastasis and may be an important breakthrough in exploring the mechanisms of breast cancer metastasis. We further found that when YY1 expression was downregulated, the abnormal accumulation of estrogen in serum, especially in breast tissue, was significantly reduced. These results suggested that altered expression of YY1 can affect the metabolism of estrogen and thus promote the development and metastasis of breast cancer. E1 and E2 undergo competitive pathways of irreversible hydroxylations catalyzed by NADPH-dependent CYP enzymes, including CYP1A1 and CYP1B1 (Cribb et al., 2006). Hydroxylation occurs at positions C2, C4, and C16, resulting in the formation of catechol estrogens: 2-OHE1/E2, 4-OHE1/E2, and 16α-OHE1. E3 is produced through the hydroxylation of
E2 or 16α-OHE1 (E. L. Cavalieri, Rogan, & Zahid, 2017). Catechol estrogens are further metabolized through methylation by the enzyme COMT (Guldberg & Marsden, 1975), leading to the formation of 2-MeOE1/E2, 4-MeOE1/E2. As shown in Fig.S3 and Fig. S5, there was no significant changes of the COMT, CYP1A1, and CYP1B1 expression compared with the control group. However, the expression of UGT2B7 mRNA was significantly increased in YY1-knockdown cell/animal models (Fig.4A-B and Fig.S5D). These findings provided conclusive evidence that YY1 regulates estrogen metabolism mainly by modulating UGT2B7.

UGTs are phase II enzymes that catalyze the covalent addition of UDPGA to facilitate the elimination of estrogen and its metabolites (Cheng et al., 1998). Our research group has previously worked on the correlation between UGTs and estrogen metabolism and demonstrated that UGTs contribute to the elimination of estrogen and can influence estrogen signaling pathways and UGT2B7 may play a key role in the elimination of estrogen (Hao et al., 2022). UGT2B7, a member of the UGTs family, exhibits high expression specificity in breast tissue and plays a critical role in estrogen metabolism homeostasis. Reduced expression or weakened functional activity of UGT2B7 directly influences the homeostasis of estrogen in vivo. Moreover, supraphysiological concentrations of estrogen can mediate the overexpression of various growth factors, thereby promoting cell growth and tumorigenesis (Hao et al., 2022; Zhou et al., 2018; Zhou et al., 2017). Additionally, in our previous study, we found that inactivation of UGT2B7 leads to a decrease in the expression of Glu-4-OHE2.
(glucuronide conjugate of 4-OH-estradiol) and inhibits the intracellular metabolic elimination of 4-OHE2 (Zhou et al., 2017). Coincidentally, the results of this study found that YY1 binds directly to UGT2B7, which is highly expressed in the mammary gland of UGTs, and can inhibit the transcription of UGT2B7, thereby downregulating UGT2B7 expression. We have found through extensive experiments in vitro and in vivo that when YY1 expression is upregulated, it can inhibit UGT2B7 expression, leading to abnormal estrogen accumulation; conversely, when YY1 expression is downregulated, it can alleviate the inhibition of UGT2B7, thus reducing or eliminating abnormal estrogen accumulation (Fig. 6, Fig. S5). These results provide strong evidence that YY1 can influence estrogen metabolism by regulating the expression of UGT2B7, which ultimately leads to breast cancer progression and metastasis. In short, YY1 can regulate the UGT2B7-estrogen metabolic axis to influence breast cancer progression and metastasis. However, this study did not further investigate the regulatory role and mechanism of UGT2B7 on 4-OHE2, which is a limitation of this study. These areas require further exploration in future research endeavors.

YY1 is a zinc finger transcription factor whose specific function at a particular promoter is determined by its interaction with several key regulatory proteins (Thomas & Seto, 1999). Interestingly, YY1 has been found to be overexpressed in several types of cancer, including colon (Chinnappan et al., 2009), breast (Thomassen, Tan, & Kruse, 2008) and prostate cancers (Seligson et al., 2005). It has been shown that YY1 can indirectly promote EMT through the NF-κB/Snail/YY1/RKIP loop, and its
transcriptional upregulation of Snail, a zinc finger protein, has been identified as an inducer of epithelial-mesenchymal transition (EMT) (Palmer et al., 2009). This upregulation of Snail induces EMT by repressing epithelial cell markers, such as E-cadherin. Markers such as E-cadherin, which promotes upregulation of Snail expression, also lead to increased expression of the mesenchymal markers N-cadherin and vimentin (Cho & Bonavida, 2017). Increasingly, studies have shown that YY1 plays an important role in cancer metastasis. For example, Yuan, Peng et al. demonstrated that the KRAS/NF-κB/YY1/miR-489 signaling axis controls metastasis in pancreatic cancer (Yuan et al., 2017), and Wang, Jing et al. showed that the YY1-regulated CCDC43-ADRM1 axis promotes proliferation and metastasis in gastric cancer (Wang et al., 2020), among others. These various activities mediated by YY1 clearly indicate that YY1 may be a desirable target and that its inhibition may be a novel strategy for treating cancer.

Our clinical data and animal models unequivocally demonstrate that YY1 can serve as a prognostic biomarker for breast cancer. Interestingly, the expression of YY1 is also associated with unfavorable prognosis in various other cancers, including gastric cancer, prostate cancer, pancreatic cancer, colorectal cancer, and lung cancer. This suggests a broad role for YY1 in malignancy. YY1 is aberrantly highly expressed in tumor cells and contributes to tumorigenesis and progression by promoting tumor cell cycle progression and cell proliferation, angiogenesis, and reprogramming of glucose and lipid metabolism; targeting YY1 has been considered a promising antitumor treatment strategy, but no targeted YY1 inhibitors have been successfully developed.
for clinical oncology treatment. In this study, we propose for the first time a regulatory relationship between YY1 and the estrogen metabolic axis and explore the molecular mechanism. Our study shows that the YY1/UGT2B7/estrogen axis plays an important role in the development and metastasis of breast cancer. This study further elucidates the potential mechanisms of YY1-mediated breast cancer metastasis and the possibility and promise of YY1 as a predictor of cancer metastasis.

**Conclusion**

In summary, we investigated the close association between disturbances in estrogen metabolism and breast cancer progression and found that YY1 regulates estrogen metabolism, thereby promoting breast cancer metastasis. We further found that YY1 transcriptionally regulates UGT2B7 expression and thus estrogen metabolism, ultimately leading to the development of breast cancer metastasis. The results of this study suggest that the YY1-UGT2B7-estrogen metabolism axis, as an innovative pathway associated with breast cancer, may be a potential target for the prevention of cancer metastasis.

**Ethical approval and consent to participate**

This study was registered in the Chinese Clinical Trial Register on May 13, 2017 (No. ChiCTR-DOD-17011393), and was performed in accordance with the Declaration of Helsinki. All the experiments that involved animals were approved and conducted under the oversight of the Animal Ethics Committee of Xuzhou Medical University.
Consent for publication

Not applicable.

Availability of data and materials

All data described in the manuscript are contained within the manuscript and the supplemental data. All raw mass spectrometry data have been uploaded to the ProteomeXchange Consortium via the iProX repository with the dataset identifier PXD047637. The dataset can be accessed using the following URL: https://www.iprox.cn/page/PSV023.html?url=1702018010554dXFZ; Password: IZDK.

Competing interests

The authors declare that they have no conflicts of interest.

Funding

This study was supported by the Natural Science Foundation of China (No. 82173883, China); the Science and Technology Foundation of Xuzhou (No. KC21010, China); the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 18KA350002, China); the Provincial Commission of Health and Family Planning in Jiangsu Province (No. H2017079, China) and the Science and Technology Planning Project of Jiangsu Province (No. BE2019636, China).

Author contributions

Conception and design: XYZ, ZL, CWL. Collected and analyzed the data: JHX, YZ, SQH. Performed the experiments: JHX, YZ, SQH, YHW, JCM. Writing the
manuscript: XYZ, JHX, YZ. All authors read and approved the final manuscript.

Acknowledgements

We thank all volunteers in this study for their cooperation and thank the Department of Thyroid and Breast Surgery, Affiliated Hospital of Xuzhou Medical University for assisting in the collection of clinical samples from patients with breast cancer. We also thank the core facilities of the Key Laboratory of Immunology, Xuzhou Medical University for their technical support in flow cytometry.

References


Figure legends

Figure.1 Screening of the most differential protein YY1 by proteomic analysis

(A) Statistical histogram of identification and quantitative results. (B) Differential protein volcano plot. The horizontal coordinate is the difference in fold, the vertical coordinate is the significance of the difference in P-value. (C) Pie chart showing the number and proportion of proteins in each subcellular organelle. (D) Bar chart showing the number of proteins in Domain (top 20). (E) The GO annotation statistics of differentially expressed proteins (level 2) indicate the secondary function annotation information. (F) GO functional enrichment bubble diagram under biological process classification. (G) Bubble diagram of GO functional enrichment under cell component classification. (H) Molecular functional classification under GO functional enrichment bubble map. (I) Heat map for the top 10 up-regulated proteins and top 10 down-regulated proteins.

Figure.2 Confirmation of the correlation between YY1 expression and breast cancer progression and metastasis
(A) Expression levels of YY1 during breast disease progression. The data are presented as means ± SEM and significant differences detected using T test. (B) Expression levels of MMP9 during breast disease progression. The data are presented as means ± SEM and significant differences detected using T test. (C) Overall survival curve. K-M Plotter was used to analyse the role of YY1 in breast cancer prognosis. (D) Overall survival curve. K-M Plotter was used to analyse the role of MMP9 in breast cancer prognosis. (E) YY1 expression in tumours of breast cancer samples at different stages was analysed by IHC. The data are presented as means ± SEM and significant differences detected using T test. (F) The protein expression levels of YY1 in MMTV-PyMT mice at different stages was analysed by western blot. The data are presented as means ± SEM and significant differences detected using T test, n=6. (G) YY1 expression in tumours of MMTV-PyMT mice at different stages was analysed by IHC, n=6. The data are presented as means ± SEM. ***P < 0.001, **P < 0.01, *P < 0.05.

**Figure 3** YY1 expression can promote breast cancer metastasis

(A-B) Changes in the protein expression levels of YY1 in MCF-7 cells and MDA-MB-231 cells after virus infection. Three replicate experiments were performed, and the data are presented as means ± SEM and significant differences detected using T test. (C-F) MCF-7 cells stably expressing YY1-shCON, YY1-shRNA, MDA-MB-231 cells stably expressing YY1-shCON, YY1-shRNA, MCF-7 cells stably expressing YY1-OENC, YY1-OE and MDA-MB-231 cells stably expressing YY1-OENC, YY1-OE were subjected to wound healing assay for cell migration and transwell assays for...
cell invasion. Three replicate experiments were performed, and the data are presented as means ± SEM and significant differences detected using T test. (G-H) The protein expression levels of E-cadherin, N-cadherin, Vimentin, MMP2 and MMP9 in YY1-shRNA and YY1-OE cell models of MCF-7 cells and MDA-MB-231 cells. The data are presented as means ± SEM. n=3, ***P < 0.001, **P < 0.01, *P < 0.05.

**Figure 4** YY1 can repress expression by regulating UGT2B7 transcription

(A-B) The protein expression levels of UGT2B7 in YY1-shRNA and YY1-OE cell models of MCF-7 cells and MDA-MB-231 cells. Three replicate experiments were performed, and the data are presented as means ± SEM and significant differences detected using T test. (C) Diagram shows three putative YY1 binding sites on the UGT2B7 promoter predicted using ISMAR and the mutated promoter sequences used in luciferase reporter assays. Luciferase reporter assays were conducted using 293T cells cotransfected with pcDNA. YY1 and the wild-type (Wt) or mutant UGT2B7 promoter constructs. Three replicate experiments were performed, and the data are presented as means ± SEM and significant differences detected using T test. (D) Co-IP of UGT2B7 and YY1 in MCF-7 cells and MDA-MB-231 cells. IgG was used as a control. The experiments were repeated three times. Co-IP of YY1 and UGT2B7 in MCF-7 cells and MDA-MB-231 cells. IgG was used as a control. The experiments were repeated three times. (E) ChIP-qPCR analysis of UGT2B7 occupancy on the YY1 promoter region containing putative UGT2B7 binding site in YY1-shRNA and YY1-OE cell models of MCF-7 cells and MDA-MB-231 cells. Three replicate experiments were performed, and the data are presented as means ± SEM and
significant differences detected using T test. (F-G) The protein expression levels of UGT2B7 in MMTV-PyMT mice at different stages was analysed by western blot and IHC. The data are presented as means ± SEM and significant differences detected using T test, n=6. (H) The protein expression levels of UGT2B7 in breast cancer samples at different stages was analysed by IHC. The data are presented as means ± SEM and significant differences detected using T test, n=6. ***P < 0.001, **P < 0.01, *P < 0.05.

Figure 5 Abnormal accumulation of oestrogen is closely associated with breast cancer metastasis

(A) The total estrogen level in serum samples from 146 premenopausal patients with malignant metastatic breast cancer (Metastasis group), 132 premenopausal patients with primary breast cancer (Primary group) and 143 matched controls (Normal group) were measured by LC/MS-MS analysis. The results are analysed by Student’s t-test and presented as the means ± SEMs. (B) The total estrogen level in cancer tissues samples from patients diagnosed with metastatic breast cancer (n=6), nonmetastatic breast cancer patients (n=6) and paracancerous tissues (n=6). The results are analysed by Student’s t-test and presented as the means ± SEMs. (C-D) The total estrogen level in serum and cancer tissues from MMTV-PyMT mice at different times (n=6). The results are analysed by Student’s t-test and presented as the means ± SEMs. (E) The concentrations of 14 estrogens in cancer tissues from human patients. (F) The concentrations of 14 estrogens in cancer tissues from MMTV-PyMT mice at different times. The data are presented as means ± SEM and significant differences detected...
using T test. (G-H) MCF-7 cells and MDA-MB-231 cells in different treatment groups were subjected to wound healing assay for cell migration for cell invasion. (I-J) MCF-7 cells and MDA-MB-231 cells in different treatment groups were subjected to transwell assays for cell invasion. Three replicate experiments were performed, and the data are presented as means ± SEM and significant differences detected using T test, n=3. ***P < 0.001, **P < 0.01, *P < 0.05.

**Figure.6 YY1 contributes to the development of breast cancer metastasis by regulating the UGT2B7-estrogen metabolic axis**

(A) Schematic illustration of the establishment of xenograft tumours in BALB/C-NU/NU (nude) mice. (B) Bioluminescence imaging of cancer tissues between YY1-shRNA and YY1-shCON groups. Sive mice were used for each group. (C) The protein expression levels of YY1 and UGT2B7 in YY1-shRNA and YY1-shCON groups (n=6) were analysed by western blot. The data are presented as means ± SEM and significant differences detected using T test. (D-E) YY1 and UGT2B7 expression in tumours of breast cancer samples from YY1-shRNA and YY1-shCON groups was analysed by IHC. The data are presented as means ± SEM and significant differences detected using T test. (F) The concentrations of 14 estrogens in serum samples from YY1-shRNA and YY1-shCON groups. The data are presented as means ± SEM and significant differences detected using T test. (G) Schematic illustration of the establishment of xenograft tumours in BALB/C-NU/NU (nude) mice. (H) Bioluminescence imaging of cancer tissues between YY1-shRNA and YY1-shCON groups. n=6. ***P < 0.001, **P < 0.01, *P < 0.05.
Figure 7 YY1 inhibits the expression of UGT2B7 through transcriptional regulation, which causes abnormal accumulation of estrogen, especially 4-OHE2, and ultimately promotes the development of breast cancer progression and metastasis.