Supplementary Information

Cytidine deaminase axis modulated by miR-484 differentially regulates cell proliferation and

chemoresistance in breast cancer

Note: Supplementary Methods, 6 Supplementary figures and 5 Supplementary tables.

Supplementary Fig. S1 qPCR analysis of selected genes was performed in MDA-231-Gem and parental MDA-231 cells. Genes were selected based on the GO analysis

Supplementary Fig. S2 qPCR analysis of 7 genes in the gemcitabine metabolism pathway was performed in MDA-231-Gem and parental MDA-231 cells.

Supplementary Fig. S3 The expression of miR-339-5p, miR-484 and miR-345-5p was assessed by miRNA qPCR in MDA-231-Gem cells compared to MDA-231 cells. U6 served as an internal control.

Supplementary Fig. S4 Constructing an mRNA-miRNA network associated with the gemcitabine metabolism pathway.

Supplementary Fig. S5 Patient cohorts for IHC analysis.

Supplementary Fig. S6 Patient cohort for mRNA and miRNA expression analyses

Supplementary Table S1 Final concentrations of gemcitabine in the culture medium corresponding to each cycle.

Supplementary Table S2 Nucletide sequences of PCR primers used for amplification of ALDH1A3, CDA, DST, FMO3, MAOA, PTHLH, XDH, UGT1A1, hENT1, NT5C, DCK, DCTD, CMPK1, NME1.

Supplementary Table S3 Selected Genes Related to Drug Metabolism/Resistance according to the GO analysis.

Supplementary Table S4 Clinicopathological variables and CDA expression in the studied cases.

Supplementary Table S5 Univariate and multivariate analysis of DFS.

Supplementary Methods

Patients and specimens

Patient cohorts for immunohistochemical (IHC) analysis:

In this study, we retrospectively obtained 193 pathologically proven primary breast cancer samples and 36 non-cancerous mammary controls to examine the CDA expression level via IHC analysis. This study was approved by the institutional review board (IRB) of Fudan University Shanghai Cancer Center (FDSCC). The breast cancer patients in this cohort fulfilled the following inclusion criteria: (i) female patients diagnosed with stage I to III primary breast cancer; (ii) patients with unilateral invasive ductal carcinoma (IDC); ductal carcinomas *in situ* were excluded; (iii) patients without any evidence of metastasis at diagnosis; (iv) patients underwent a mastectomy and axillary lymph node dissection or breast conservation surgery followed by adjuvant non-gemcitabine-based chemotherapy; the patients with any prior chemotherapy (neoadjuvant chemotherapy) were excluded in our cohorts. We also excluded patients if they had no FFPE tumor sample available and presence of any tumor type other than breast cancer (no co-morbidities). As shown in **Supplementary Fig. S5**, a total of 1,709 samples that met the eligibility criteria were collected from 2,514 patients who were diagnosed as breast cancer at the Department of Breast Surgery in FDSCC between August 2001 and March 2006. All samples were obtained with informed consent.

Next, these breast cancer cases were characterized as luminal-like, human epidermal growth factor receptor 2 (HER2)-enrichment, and triple-negative subtypes according to the expression statuses of estrogen receptor (ER), progesterone receptor (PR), and HER2. In this study, ER, PR, and HER2 expression statuses were examined by IHC staining at the Department of Pathology in FDSCC. Additionally, the patients with HER2 expression status (IHC, score=2) were subjected to florescence *in situ*

hybridization (FISH) screening for HER2 gene amplification. The HER2 overexpression subgroup was defined as FISH positive or an IHC staining score=3. Based on the above criteria, a total of 1,650 breast cancer cases with available ER, PR and HER2 statuses were characterized as luminal-like (ER+ and/or PR+; n=1027), human epidermal growth factor receptor 2 (HER2)-enriched (HER2+, ER- and PR-; n=262), and triple-negative (ER-, PR-, and HER2-; n=361) subtypes. Subsequently, we used the complete random sampling method to collect 100 luminal-like subtype cases, 50 HER2-enriched subtype cases and 50 triple-negative subtype cases to construct the tissue microarrays (TMAs).

In addition, non-cancerous mammary tissues were also collected as controls for these TMAs. During the period from January 2013 to February 2013, 109 non-cancerous mammary controls with pathologically confirmed benign mammary diseases were collected from women who had come to the Outpatient Department at FDSCC for breast cancer screening. The detailed inclusion and exclusion criteria for these controls are illustrated in **Supplementary Fig. S5**. A total of 36 non-cancerous mammary tissues were randomly collected for the TMAs construction. All these samples were obtained with informed consent.

Finally, a total of 200 IDC breast cancer cases and 36 non-cancerous mammary tissues were included for the TMAs. Of these cases, 7 breast cancer cases experienced duplicate tissue core loss after IHC staining. Thus, the remaining 193 cancerous and 36 non-cancerous mammary samples were included in the subsequent analysis. As shown in **Supplementary Table S4**, the correlations between clinicopathological parameters and the expression level of CDA were evaluated in this cohort of 193 cancerous cases using contingency tables and Pearson's χ^2 or Fisher's exact tests.

In this cohort, the breast cancer patients were regularly followed, and the clinical outcomes of 177 cases were obtained, with the last update occurring in September 2013. The follow-up period was defined as the time from surgery to the last observation for censored cases or recurrence/death for complete

observations. Disease-free survival (DFS) was defined as the time from the date of primary surgery to the date of recurrence/breast cancer-specific death or September 2013. The categories analyzed for DFS were first recurrence of disease at a local, regional, or distant site and breast cancer-specific death. Patients with study end date and loss of follow-up were considered censored. Thus, these 177 cancerous cases were analyzed in the subsequent Kaplan-Meier analysis and Cox proportional hazard analysis.

In this cohort, all therapeutic regimen decisions were based on the Chinese Anti-Cancer Association guidelines for the treatment of breast cancer. According to the guidelines, gemcitabine has not been recommended for adjuvant chemotherapy of primary breast cancer, thus most patients would not receive gemcitabine during adjuvant chemotherapy in our cohorts. Therefore, enrolled patients underwent a mastectomy and axillary lymph node dissection or breast conservation surgery followed by non-gemcitabine-based adjuvant chemotherapy. Moreover, disease-free survival (DFS) is defined as the time from the date of primary surgery to the date of recurrence/breast cancer-specific death, making sure that these patients did not receive gemcitabine before our study endpoint (DFS) in this study.

At present, gemcitabine is recommended for first-line treatment of recurrent or metastatic breast cancer, according to Chinese Anti-Cancer Association guidelines or National Comprehensive Cancer Network guidelines. However, the patients who received gemcitabine treatment in our hospital were confirmed to have recurrent or metastatic breast cancer for which operative treatment was not suitable. Due to this limitation, we could not obtain sufficient breast cancer samples to evaluate the CDA expression in patients who were treated with gemcitabine in this study. An additional cohort of patients receiving gemcitabine-based chemotherapy may be needed to form a general understanding of the role of CDA in breast cancer.

Patient cohort for mRNA and miRNA expression analyses:

In this study, we also retrospectively collected paired primary breast cancer and corresponding adjacent non-cancer tissues to examine the mRNAs and microRNAs expression profiles by qPCR analysis. As we could not achieve the frozen samples for the RNA extraction in our first cohort of 2001 through 2006, we used paired frozen samples from 30 breast cancer patients that met the eligibility criteria between March 2013 and May 2013. These recruited patients had surgically confirmed stage I to III primary invasive ductal carcinoma and received a mastectomy and axillary lymph node dissection at the FDSCC. Similarly, gemcitabine was not utilized for patients in this cohort. The consort-type diagram illustrated the detailed inclusion and exclusion criteria of this cohort (**Supplementary Fig. S6**). This study was approved by the institutional review board (IRB) of FDSCC, and all participants provided informed consent to participate in this research.

Patient cohort for Kaplan-Meier Plotter analysis:

We used a large public clinical database (Kaplan-Meier Plotter, short for KM plotter) of breast cancer to evaluate the association between the CDA expression and clinical outcome, with the following restricted condition: 1) 120 months of follow-up time, 2): including all the clinical subtypes of breast cancer, 3) systemically treated cases, 4) auto select best cutoff. Primary purpose of the tool is a meta-analysis based *in silico* biomarker assessment. In the revised manuscript, we evaluated the effects of CDA expression on disease-free survivals (DFS) of 3,455 patients with the latest version of this database (2014 version; http://www.kmplot.com/analysis/index.php?p=service).

Tissue microarray (TMA)

TMAs were constructed using above 200 paraffin-embedded blocks of breast tumors and 36 blocks of non-cancerous mammary controls. The hematoxylin and eosin (HE)-stained slides from tumors were evaluated to identify representative tumor regions from which two 1.0-mm tissue cores were retrieved and transferred into recipient array blocks using a tissue micro arrayer (UNITMA Instruments, Seoul, Korea). TMAs were composed of duplicate cores from different areas of the same tumor to compare staining patterns. TMA sections were subsequently dewaxed in xylene and rehydrated in ethanol for IHC staining.

Immunohistochemical (IHC) staining

Immunohistochemistry for CDA were conducted using a two-step protocol (GTVision[TM]III). Briefly, TMA sections were washed with phosphate-buffered saline (PBS) after rehydration and then treated with 3% hydrogen peroxide for 10 min to block endogenous peroxidase activity. The antigens were retrieved by boiling the TMAs in citrate buffer (pH 6.0) at 100°C for 5 min. The TMAs were blocked with 10% normal goat serum for 1 h at room temperature (RT) and incubated in a humid chamber at 4°C overnight with polyclonal rabbit anti-human CDA antibody (Cell Signaling Technology, CA). Following washes with PBS, all of the TMAs were incubated for 30 min with secondary antibody (GTVision[TM]III Detection System/Mo&Rb) at RT. The sections were counterstained with Gill hematoxylin and mounted after clearing with xylene.

IHC variable evaluation

TMAs representing duplicate samples from each case were stained and scored semi-quantitatively. Staining was graded based on the staining intensity (1, weak; 2, moderate; 3, strong) and the percentage

of cells stained (1, 0 to <10%; 2, 10 to <50%; 3, 50 to 100%). Scoring was conducted according to a sum index (SI) of the intensity and percentage of CDA-positive cells as follows: SI, 2, scored as 0; SI, 3, scored as 1; SI, 4, scored as 2; SI, 5 or 6, scored as 3. If the score was equal to or greater than two, the tumor was considered to have high CDA expression; otherwise, low CDA expression was classified. Scoring was reviewed in parallel by two experienced breast disease pathologists who were blinded to all clinical data.

miRNA microarray and mRNA microarray

Total RNA was extracted using a mirVana™ miRNA Isolation Kit (Ambion, Austin, TX) according to the manufacturer's instructions. Human miRNA V18.0 microarray chips (Agilent Technology, Austin, TX) and analyzed using the GeneSpring 11.0 software (Agilent Technologies, Santa Clara, CA). For the mRNA microarray, human U133 Plus 2.0 chips (Affymetrix, Santa Clara, CA) were utilized to identify differences between the samples. The miRNA and mRNA microarrays were performed by the Shanghai Bio Corporation (Shanghai, China). These microarray data have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus database under the accession number GSE63140. The candidate genes related to gemcitabine resistance were generated by combining the gene list from the Gene Set Enrichment Analysis (GSEA) with "drug" as keywords and the discrepant genes from our mRNA microarray data. Gene ontology (GO) analysis was performed for the intersected genes to identify the relevant gemcitabine resistance pathway.

Cytotoxicity and cell proliferation assays

Cells in the logarithmic growth phase were plated into a 96-well plate. After allowing the cells to

adhere overnight, complete medium was replaced with medium containing serially diluted

chemotherapeutic reagent. The Cell Counting Kit-8 (CCK-8) assay (Dojindo Laboratories, Kumamoto,

Japan) was used to construct a Hdose-response curveH to obtain half the maximal inhibitory

concentration (IC50) of each compound. The IC50 was calculated using the XLFit curve fitting

software (Microsoft Inc., CA). Cell proliferation was also quantified using the CCK-8 assay according

to the manufacturer's instructions. All experiments were performed three times in triplicate.

Colony formation assay

Cells were seeded at a low density and treated with different concentrations of a chemotherapeutic

reagent and cultured for 14 days to promote colony formation. The colonies were stained with 2%

methylene blue/50% ethanol, and colonies containing 50 or more cells were counted.

Cell cycle distribution analysis

Cells of each cell line were harvested and fixed in ice-cold 70% ethanol overnight. The fixed cells

were washed with PBS and stained with a solution containing 25 μg/mL propidium iodide (Shenggong,

Shanghai, China), 10 μg/mL RNase A, 0.2% Triton X-100 and 0.05 mM EDTA in PBS for 30 minutes

in the dark. For each sample, at least 20,000 cells were analyzed by flow cytometry (FACStation, BD

Biosciences) and FlowJo (TreeStar Inc., Ashland, OR).

Luciferase reporter assay targeting CDA 3′-UTR

The 3'-UTR of CDA (NM__001785) was amplified from human genomic DNA using appropriate

primers and subcloned into the region directly downstream of the Renilla gene stop codon in the

psiCHECK2 vector (Promega, Madison, WI) to generate psiCHECK2-CDA-3'-UTR constructs. With appropriate primers, PCR amplification of the 3'UTR sequence of CDA generated a series of mutant psiCHECK2-CDA-3'-UTR reporter vectors. The miR-339-5p, miR-345-5p and miR-484 mimic duplexes were synthesized by Genepharma (Shanghai, China). HEK 293T cells were transfected with a mixture of reporter constructs and miR duplexes using Lipofectamine 2000 (Invitrogen, Carlsbad, CA). After 48 hours of incubation, firefly and Renillaluciferase activities were measured using the dual-luciferase reporter assay system (Promega, Madison, WI) from the cell lysates.