ORIGINAL RESEARCH

ADAMTS8 Expression is a Potential Prognostic Biomarker for Postoperative Metastasis in Lymph Node-Negative Early-Stage Invasive Breast Carcinoma Patients

Yao Li[1,]*
Xin Yang[1,]*
Jie Sun [2]
Yangyang Zhao[3]
Qi Zhou [3]
Bin Hua[1]

[1]Breast Center, Department of Thyroid-Breast-Hernia Surgery, Department of General Surgery, Beijing Hospital, National Center of Gerontology, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing, People's Republic of China; [2]Graduate School of Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, People's Republic of China; [3]The Key Laboratory of Geriatrics, Beijing Hospital, National Center of Gerontology, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing, People's Republic of China

*These authors contributed equally to this work

Correspondence: Bin Hua
Breast Center, Department of Thyroid-breast-hernia Surgery, Department of General Surgery, Beijing Hospital, National Center of Gerontology, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing, People's Republic of China
Tel +86-10-85136127
Fax +86-10-65132969
Email huabinbjh@126.com

Purpose: Invasive breast carcinoma (BC) is the most common malignant breast tumor. Most lymph node-negative (LN−) early-stage BC patients usually have a good prognosis, but 7% of patients still develop metastasis after surgery. It is not yet clear how to screen candidates with poorer prognosis in LN− early-stage patients, so that they can receive intensive therapy. Hence, we expect to identify a prognostic biomarker to assess postoperative metastasis in LN− early-stage BC patients.

Patients and Methods: Screening and verifying of candidate genes by gene expression profiling of LN− early-stage BC samples (n = 640) from 3 independent public datasets. Univariable and multivariable Cox regression analyses showed the relation between the candidate genes and postoperative metastasis. Distant metastasis-free survival (DMFS) analysis was performed to examine the prognostic significance. Quantitative real-time polymerase chain reaction (qRT-PCR) assays were performed to examine *ADAMTS8* expression and prognostic association in our clinical samples (n = 25).

Results: In the discovery cohort (TCGA and GSE20685 datasets), we found that *ADAMTS8* tend to be low expression in LN− early-stage BC, and low *ADAMTS8* expression was associated with postoperative metastasis and shortened DMFS. Moreover, the above finding was confirmed in the validation cohort (GSE6538 dataset). Lower *ADAMTS8* expression was related to poorer prognostic clinical stage and PAM50 subtypes and shorter DMFS. Gene enrichment analysis indicated that *ADAMTS8* may be correlated with BC metastasis. qRT-PCR assays of our clinical tumor sample showed that patients with low *ADAMTS8* expression seem to be prone to developing metastasis and have a shorter DMFS time.

Conclusion: Our research shows that low *ADAMTS8* expression is associated with postoperative metastasis and shortened DMFS in LN− early-stage BC patients, which suggests that *ADAMTS8* may be a potential prognostic marker for postoperative metastasis in LN− early-stage BC patients.

Keywords: *ADAMTS8*, lymph node-negative early-stage invasive breast carcinoma, distant metastasis-free survival, biomarker

Introduction

Female breast cancer is the most commonly diagnosed cancer in the world, and more than 2.3 million women are diagnosed in 2020 and are still the leading cause of cancer death.[1] Invasive breast carcinoma (BC) accounts for 80–90% of breast cancer and is the most common type of pathology and is a heterogeneous disease at the molecular level.[2–5]

In the last decade, systemic therapy, including endocrine therapy, anti-HER2 therapy, immunotherapy and chemotherapy, has achieved success in improving the clinical outcomes of BC patients after surgery, including early-stage BC.[6–9] Although most lymph node-negative (LN−) early-stage BC patients have a good prognosis,[10–12] postoperative metastasis still occurs. The risk of metastasis may be underestimated in this group of BC patients. Whether to intensify treatment for LN− early-stage BC patients remains controversial.[7] Hence, it is necessary to define prognostic markers to assess the risk of postoperative metastasis in LN− early-stage BC patients.

Traditional clinical and pathological characteristics, such as patient age, tumor size, histological features, lymph node status, and molecular subtypes, can predict the outcomes of breast cancer but are not accurate, which will cause over- or undertreatment.[13,14] With the development of bioinformatics, all kinds of gene signatures, including intrinsic molecular subtypes, 21-gene recurrence scores (RS) and MammaPrint 70 genetic testing, were used to predict the risk of postoperative metastasis in patients with early-stage breast cancer.[15–18] We found that most studies on early-stage breast cancer outcomes involve all stage I–II BC patients, according to the definition of early-stage BC,[5] but few studies have focused only on LN− early-stage BC patients.[9,19–21]

In this study, we compared the gene expression profiles of patients (n = 640) between postoperative metastatic and non-metastatic LN− early-stage BC patients from The Cancer Genome Atlas (TCGA provisional dataset, 2021) and Gene Expression Omnibus (GEO) databases.[22,23] After a series of rigorous bioinformatics screenings and analyses, we determined that low *ADAMTS8* expression correlated with shortened distant metastasis-free survival (DMFS) in these patients. Furthermore, qRT–PCR assays were used to test the relationship between *ADAMTS8* expression levels and DMFS in our clinical samples (n = 25) with a matched case study. Finally, we identified *ADAMTS8* as a potential prognostic marker for postoperative metastasis in LN− early-stage BC, although there were several limitations in our study; for example, our clinical sample size was small.

Materials and Methods
Data Extraction
Transcriptome and clinical data from BC patients were downloaded from TCGA (https://portal.gdc.cancer.gov/), and the microarray gene expression profiles and clinical data were retrieved from the GEO (GSE20685, GSE6532) (https://www.ncbi.nlm.nih.gov/gds) database. Patients who

met the following criteria were included: (1) female patients were diagnosed with BC between the ages of 20–85 years; (2) LN− status at diagnosis; (3) the size of the primary tumors was T1-T3 stage; (4) no distant metastasis; (5) patients accepted standard systemic therapy; and (6) metastatic status was recorded at the end of follow-up. Patients with only local or regional recurrence were excluded from this study. The discovery cohorts TCGA (n = 390) and GSE20685 (n = 131, Affymetrix Human Genome U133 Plus 2.0 Array (GPL570)) datasets were used as cohorts to identify common differentially expressed genes (DEGs) and detect the correlation between common DEGs and postoperative metastatic and DMFS time. The testing cohort (GSE6532 dataset; n = 119; Affymetrix Human Genome U133 (GPL96)) was used to verify the differences in expression and prognostic value of the DEGs. The detailed clinical information of the three datasets is included in Supplementary Table S1.

Data Processing
The raw counts of RNA-sequencing data from the TCGA dataset were corrected and normalized using the R package "limma". Preprocessing procedures were used to process raw data of the Affymetrix microarray in the GSE20685 and GSE6532 datasets, including RMA background correction, and the "affy" R package was also applied to complete $\log_2$ transformation, quantile normalization and median polishing algorithm summarization. Probes were annotated by the Affymetrix annotation files. The PAM50 gene expression data in TCGA and GES20685 were extracted and applied against reference data of median gene expression to bin each sample into the intrinsic molecular subtypes of BC patients using the "genefu" R package. The detailed candidate gene screening process is shown in Figure 1. All data analyses and screening procedures referred to the RECIST criteria.[24,25]

Analysis of Differentially Expressed Genes
In the discovery cohort, the common DEGs were screened between postoperative metastatic patients and non-metastatic patients. The differences in expression of the final candidate genes were validated in the validation cohort. The filtering condition for all DEGs was $p<0.05$ and $|\log_2$fold change (FC)$| > 1$.
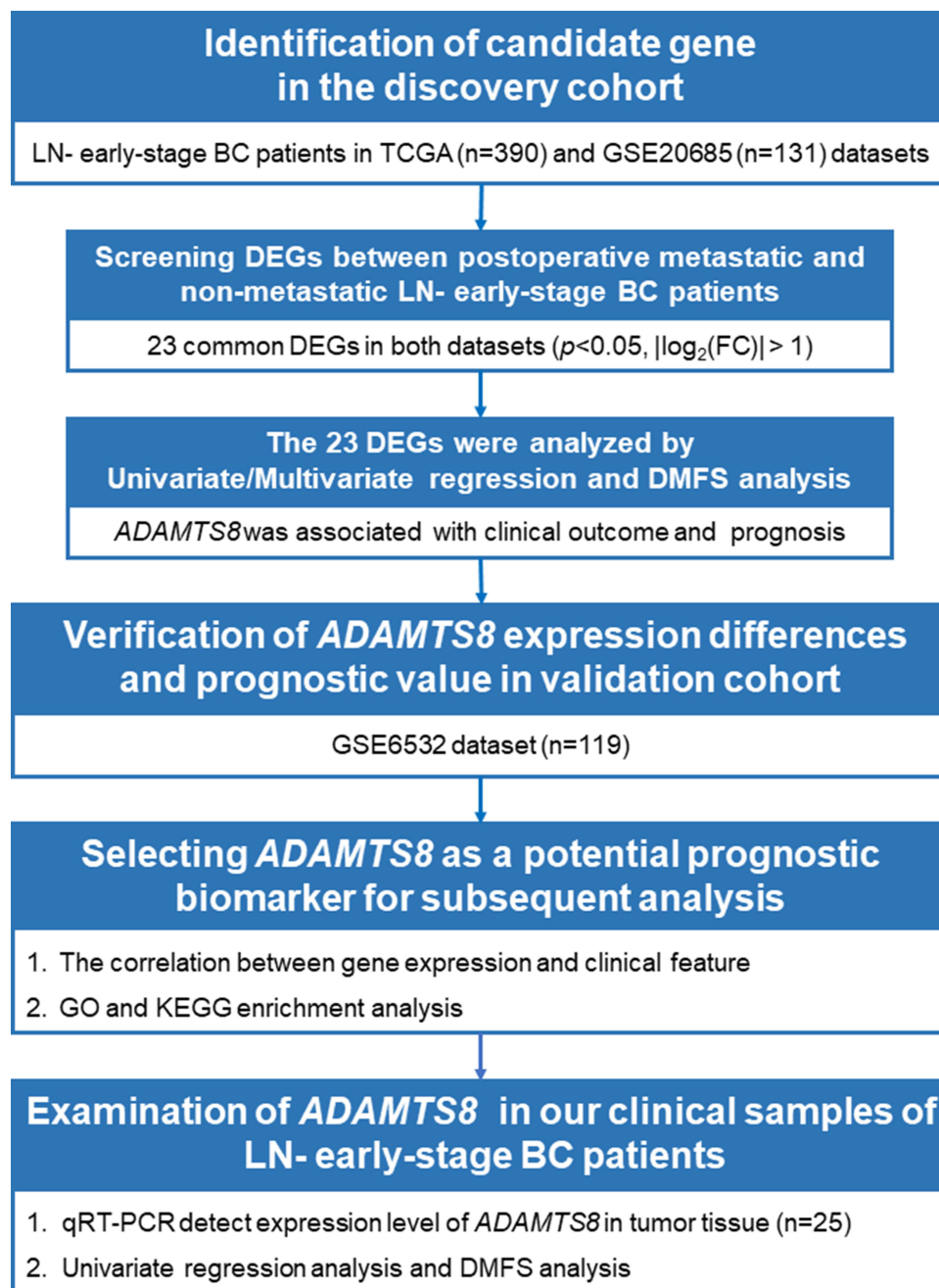
Figure 1 Study workflow.
Abbreviations: LNN, lymph node-negative; BC, invasive breast carcinoma; DEGs, differentially expressed genes; DMFS, distant metastasis-free survival; log$_2$(FC), log$_2$(fold change); qRT–PCR, quantitative real-time polymerase chain reaction.

Univariate and Multivariate Cox Regression Analysis

The common DEGs with significant differences in univariate analysis were entered into multivariate analysis. Multivariate Cox regression with forward stepwise regression was performed to investigate the impact of independent factors on postoperative metastasis, and the hazard ratios (HRs) with 95% confidence intervals (CIs) and p values are reported.

Distant Metastasis-Free Survival Analysis

The data of postoperative metastasis in LN− early-stage patients were used to calculate DMFS. DMFS analysis was conducted to test the ability of the DEGs to predict the prognosis of LN− early-stage BC patients in the discovery cohort, in the testing cohort, and in our LN− early-stage BC tumor specimen. Kaplan–Meier survival curves were used to compare

the high and low expression groups according to the hazard ratio (HR) and log-rank *p* value.

Gene Ontology and Pathway Enrichment Analysis

To explore the potential molecular biological functions of *ADAMTS8*, we used the online bioinformatic tool DAVID (https://david.ncifcrf.gov/)[26] to analyze the enrichment of Gene Ontology (GO)[27] and pathway (KEGG)[28] between the high and low *ADAMTS8* expression groups using the TCGA and GSE20685 datasets (*p*<0.05).

Examination in Our Clinical Tumor Specimens of LN− Early-Stage BC Patients

We examined the candidate gene expression levels in clinical tumor specimens obtained from breast tumor tissue removed during surgery and stored in liquid nitrogen in the Breast Centre of Beijing Hospital. The sample screening criteria were consistent with the patient screening criteria in the Data Extraction section. The follow-up time was from surgical resection of the tumor to May 2021.

We found 8 LN− early-stage patients who developed postoperative metastasis in our sample library. We matched each metastatic patient according to their molecular subtype, age, and date of surgery to approximately 2 non-metastatic patients as controls (n = 17). A total of 25 patient information and matching methods were showed in Supplementary Table S3.

We extracted total RNA of tumor specimens using RNAiso Plus (TaKaRa). qRT–PCR assays were performed to detect *GAPDH* and *ADAMTS8* mRNA using the PrimeScript RT reagent Kit and SYBR Premix Ex Taq (TaKaRa). The primers for qPCR were as follows: *GAPDH*: forward 5′-GGAGCGAGATCCCTCCAAAAT-3′, reverse 5′-GGCTGTTGTCATACTTCTCATGG-3′; *ADAMTS8*: forward 5′-GTGGCAGCCCGAATCYACAAGCA-3′, reverse 5′-AGTGTAAGCCCCCCATTGTCGGA-3′. *GAPDH* was used as control. The relative expression levels of *ADAMTS8* mRNAs were calculated and quantified using the $2^{-\Delta\Delta ct}$ method.[29]

Statistical Analysis

DEG analysis was conducted on the TCGA datasets using the "edgeR" R package and on the GEO datasets using the "limma" R package. We used the "survival" and "survminer" R packages to draw Kaplan–Meier

survival curves and compared the high and low expression groups according to the log-rank *p* value. Univariate and multivariate Cox regression analyses with forward stepwise regression based on the likelihood ratio test (forward LR model) were performed to investigate the impact of independent factors on DMFS, and the hazard ratios (HRs) with 95% confidence intervals (CIs) and p values are reported.

SPSS version 25 (IBMCorp, Armonk, N.K. USA) and R software version 4.0.5 (R core Team, Vienna, Austria) were used to statistically analyze the data. The *chi-squared test* or *Fisher's exact probability test* was used to process the categorical variables, and *t*-test, *ANOVA test* or the *Mann–Whitney U-test* was used to process the continuous variables. Linear regression models were used to detect linear trends. A value of *p* < 0.05 was considered statistically significant.

Results
Data Collection

A total of 640 LN− early-stage BC patients were collected in public databases. In the TCGA dataset (n = 390), 23 patients developed distant metastasis, and 367 patients were included as a negative control. In the GES20685 dataset (n = 131), 12 patients developed distant metastasis, and 119 patients were included as a negative control. In the GSE6532 dataset (n = 119), 10 patients developed distant metastasis, and 109 patients were included as a negative control. The clinical information is summarized in Supplementary Table S1.

Identification of Candidate Gene in the Discovery Cohort

In the discovery cohort, a total of 893 and 314 DEGs were extracted from the TCGA and GES20685 datasets, respectively. Nine common DEGs were upregulated, and 14 common DEGs were downregulated in both datasets (Figure 2A and B). Twenty-three common DEGs in the GSE20685 dataset are presented as a heatmap in Figure 2C.

Then, Kaplan–Meier curves of DMFS analysis showed that only 3 of the 23 common DEGs were significantly associated with the prognosis, including *ADAMTS8* (TCGA: log-rank *p* = 0.000; GSE20685: log-rank *p* = 0.002), *ZBTB16* (TCGA: log-rank *p* = 0.034; GSE20685: log-rank *p* = 0.022), and *LEP* (TCGA: log-rank *p* = 0.028; GSE20685: log-rank *p* = 0.018), and their low expression suggested shorter DMFS times in both datasets (Figure 2D–I).
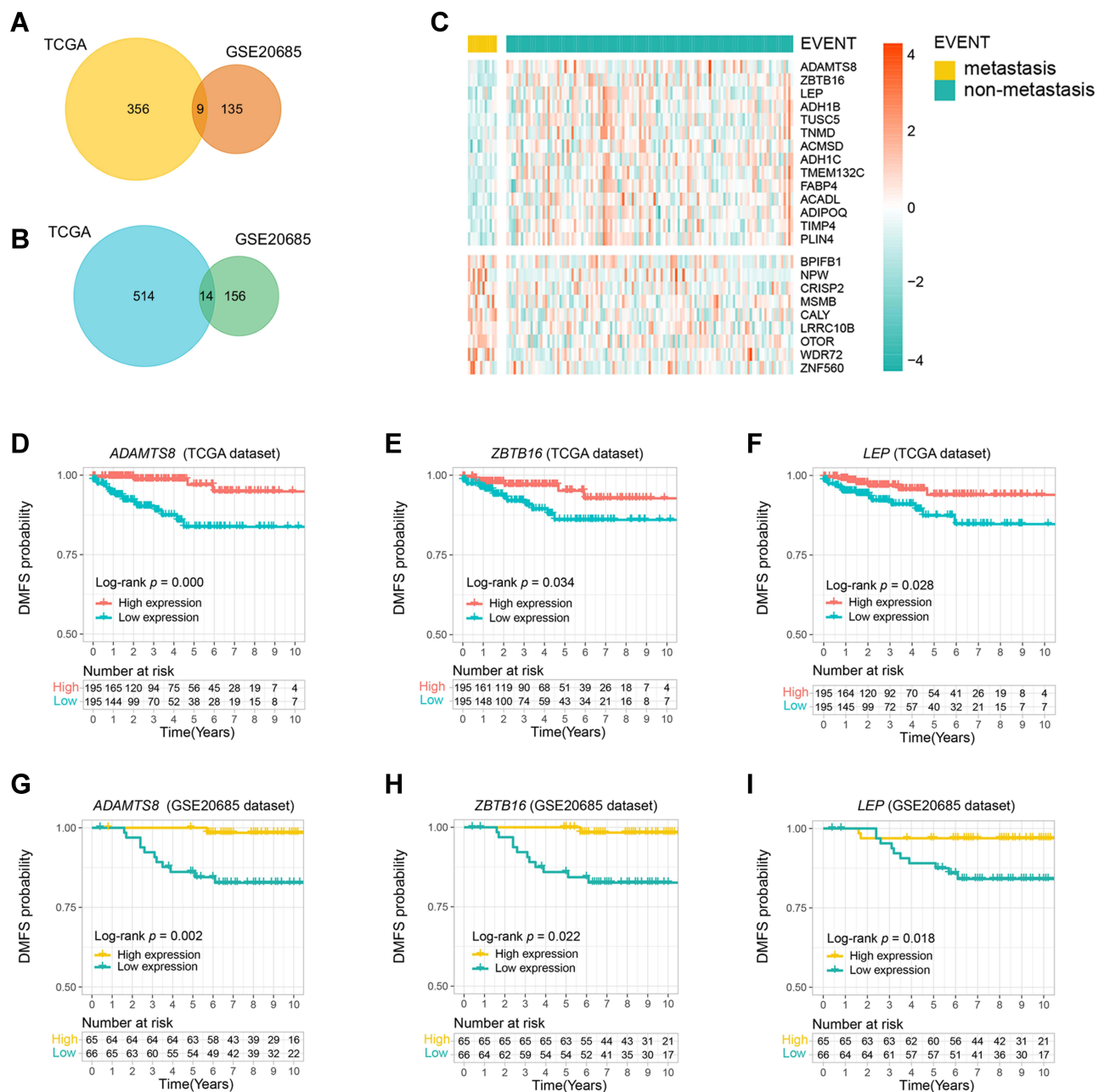
Figure 2 Low *ADAMTS8* expression was correlated with a reduced postoperative distant metastasis-free survival (DMFS) time in LNN early-stage BC patients. (**A** and **B**) A total of 23 genes were simultaneously differentially expressed in both discovery cohorts (TCGA and GSE20685), 9 DEGs were upregulated in postoperative metastasis patients, and 14 DEGs were downregulated in postoperative metastasis patients. (**C**) The heatmap shows significant differences of the 23 DEGs in the GSE20685 dataset. Kaplan–Meier curves of DMFS between the high and low expression groups in the discovery cohort stratified by 3 genes: (**D** and **G**) stratified by *ADAMTS8*, (**E** and **H**) stratified by *ZBTB16*, (**F** and **I**) stratified by *LEP*.

In addition, the 23 common DEGs were analyzed by univariate and multivariate Cox regression (Table 1). We found that low *ADAMTS8* expression was a high risk factor for postoperative metastasis in both discovery datasets (TCGA: HR = 5.637 [1.915–16.595], $p = 0.002$; GSE20685: HR = 14.550 [1.853–114.230], $p = 0.011$).

Furthermore, we combined the expression level of *ADAMTS8* and standard clinical prognostic variables

(age, PAM50 and stage), regardless of their statistical significance, into the multivariate Cox regression analysis in the TCGA and GSE20685 datasets (Table 2). The results showed that the expression level of *ADAMTS8* was an independent prognostic factor correlated with DMFS in both datasets (TCGA: HR = 0.136 [0.038–0.493], $p = 0.002$; GSE20685: HR = 0.119 [0.014–0.977], $p = 0.047$).

Table 1 Univariate and Multivariate Cox Regression (Forward LR Model) Analysis the 23 Common DEGs to Detect the Correlation Between DEGs Expression Levels and Clinical Outcome of LN− Early-Stage BC Patients

TCGA Dataset						GSE20685 Dataset						
DEG (Low vs High)	**Univariate Analysis**			**Multivariate Analysis**			**DEG (Low vs High)**	**Univariate Analysis**			**Multivariate Analysis**		
	p	**HR**	**95.0% CI**	*p*	**HR**	**95.0% CI**		*p*	**HR**	**95.0% CI**	*p*	**HR**	**95.0% CI**
ACADL	0.676	1.191	0.525–2.699				**ACADL**	**0.019**	11.690	1.509–90.590			
ACMSD	0.462	1.363	0.597–3.108				ACMSD	0.108	2.921	0.791–10.793			
ADAMTS8	**0.002**	5.637	1.915–4.984	**0.002**	5.637	1.915–16.595	**ADAMTS8**	**0.017**	12.008	1.550–93.053	**0.011**	14.550	1.853–114.230
ADH1B	0.088	2.111	0.894–16.595				**ADH1B**	**0.040**	4.922	1.078–22.468			
ADH1C	0.243	1.636	0.715–3.742				**ADH1C**	**0.020**	11.362	1.467–88.019			
ADIPOQ	0.352	1.480	0.648–3.381				**ADIPOQ**	**0.020**	11.362	1.467–88.019			
BPIFB1	0.262	0.619	0.268–1.431				BPIFB1	0.226	0.476	0.143–1.582			
CALY	0.156	0.537	0.228–1.268				CALY	0.092	0.325	0.088–1.200			
CRISP2	0.157	0.538	0.228–1.269				CRISP2	0.204	0.459	0.138–1.525			
FABP4	0.366	1.463	0.641–3.340				**FABP4**	**0.033**	5.207	1.141–23.770			
LEP	**0.034**	2.610	1.073–6.350				**LEP**	**0.034**	5.189	1.137–23.687			
LRRC10B	0.117	0.490	0.201–1.194				**LRRC10B**	**0.026**	0.177	0.039–0.809	**0.011**	14.550	1.853–114.230
MSMB	0.886	0.942	0.416–2.136				MSMB	0.544	1.426	0.453–4.494			
NPW	0.061	0.428	0.176–1.040				NPW	0.499	0.673	0.214–2.120			
OTOR	0.491	0.748	0.328–1.708				**OTOR**	**0.029**	0.184	0.040–0.838	**0.035**	0.192	0.041
PLIN4	0.437	1.387	0.608–3.165				**PLIN4**	**0.035**	5.141	1.126–23.466			
TIMP4	0.251	1.634	0.706–3.781				**TIMP4**	**0.033**	5.239	1.148–23.917			
TMEM132C	0.079	2.161	0.915–5.105				**TMEM132C**	**0.017**	11.979	1.546–92.828			
TNMD	0.334	1.503	0.658–3.432				TNMD	0.086	3.137	0.849–11.589			
TUSC5	0.166	1.811	0.7824.194				**TUSC5**	**0.017**	12.064	1.557–93.453	**0.031**	9.690	1.238
WDR72	0.414	0.709	0.310–1.619				WDR72	0.192	0.450	0.1351.495			
ZBTB16	**0.041**	2.525	1.038–6.140				**ZBTB16**	**0.017**	12.096	1.561–93.704			
ZNF560	0.310	0.648	0.281–1.498				ZNF560	0.095	0.329	0.089–1.214			

Note: Bold figure note: this variable is statistically significant.
Abbreviations: HR, hazard ratio; CI, confidence interval.

Table 2 Multivariable Cox Regression Analyses of *ADAMTS8* in TCGA and GSE20685 Datasets

Variables	TCGA			GSE20685		
	p value	HR	95.0% CI	*p* value	HR	95.0% CI
ADAMTS8 (high vs low)	**0.002**	0.136	0.038–0.493	**0.047**	0.119	0.014–0.977
Age (40–59 vs <40)	0.560	0.674	0.180–2.533	0.191	0.442	0.130–1.502
Age (≥60 vs <40)	0.277	0.479	0.127–1.804	0.981	0.000	0.000
Pam50 (Her2 vs Basal)	0.431	1.802	0.416–7.812	0.926	1.124	0.096–13.189
Pam50 (LumA vs Basal)	0.757	1.268	0.282–5.708	0.982	1.024	0.129–8.097
Pam50 (LumB vs Basal)	0.601	1.353	0.436–4.204	0.057	4.769	0.956–23.788
Pam50 (Normal vs Basal)	0.007	17.085	2.161–135.098	0.994	0.000	0.000
Stage (stage II vs stage I)	0.593	0.783	0.319–1.922	0.346	0.575	0.182–1.818

Note: Bold figure note: this variable is statistically significant.
Abbreviations: HR, hazard ratio; CI, confid.

In conclusion, we found that only *ADAMTS8* appeared to be reduced in LN− early-stage BC patients with metastasis, and its low expression was associated with postoperative metastasis and shortened DMFS time. Thus, we focused attention on *ADAMTS8* in this study.

Verification of Prognostic Value of ADAMTS8 in the Validation Cohort

In the validation cohort (GSE6532), we found that the expression level of *ADAMTS8* was lower than that in metastatic patients (Figure 3A, $\log_2$(metastasis/non-metastasis) = −1.621, *p* = 0.000). Patients with low *ADAMTS8* expression had an increased risk of metastasis (GSE6532: HR=9.416 [1.193–74.328], *p* = 0.033). In addition, the lower the *ADAMTS8* expression was, the shorter the DMFS time (Figure 3B, log-rank *p* = 0.009).

The results in the discovery and validation cohorts imply the potential ability of *ADAMTS8* to predict the outcomes of LN− early-stage BC patients.
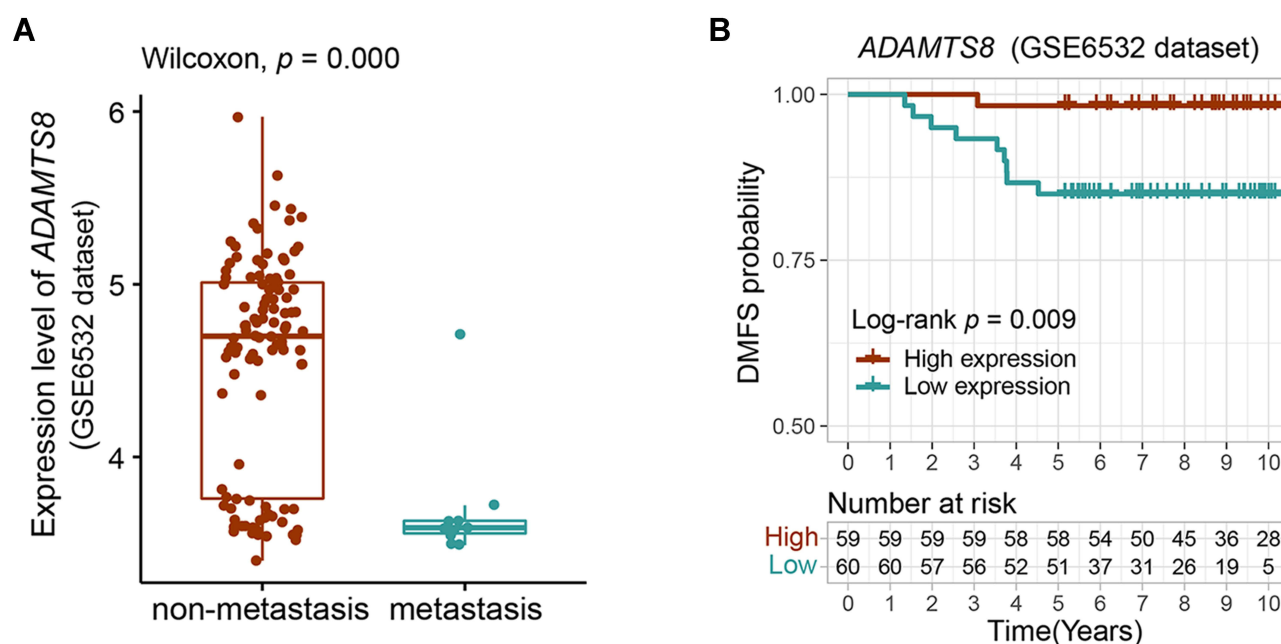

Figure 3 Verification of difference in expression and prognostic value of *ADAMTS8* in the validation cohort. (**A**) The expression level of *ADAMTS8* in metastatic and non-metastatic LN− early-stage BC patients. (**B**) Kaplan–Meier curves of DMFS between the *ADAMTS8* high and *ADAMTS8* low expression groups.

Analysis of the Correlation Between ADAMTS8 Gene Expression and Clinical Features

First, we studied the difference in the distribution of clinical features, including age, tumor size, clinical stage, molecular subtypes and DMFS, between *ADAMTS8* high and low expression in TCGA and GSE20685 datasets. As there were no molecular subtypes of BC patients in the datasets,

we used the PAM50 gene set to define the intrinsic molecular subtypes of BC patients as luminal A (LumA), luminal B (LumB), HER2-enriched (Her2), basal-like (basal), and normal-like (normal).[17] The detailed clinical data are presented in Supplementary Table S1, and the PAM50 hierarchical clustering of TCGA and GSE20685 datasets is presented by heatmaps in Figure 4A and Supplementary Figure S1A, respectively. There were obvious differences
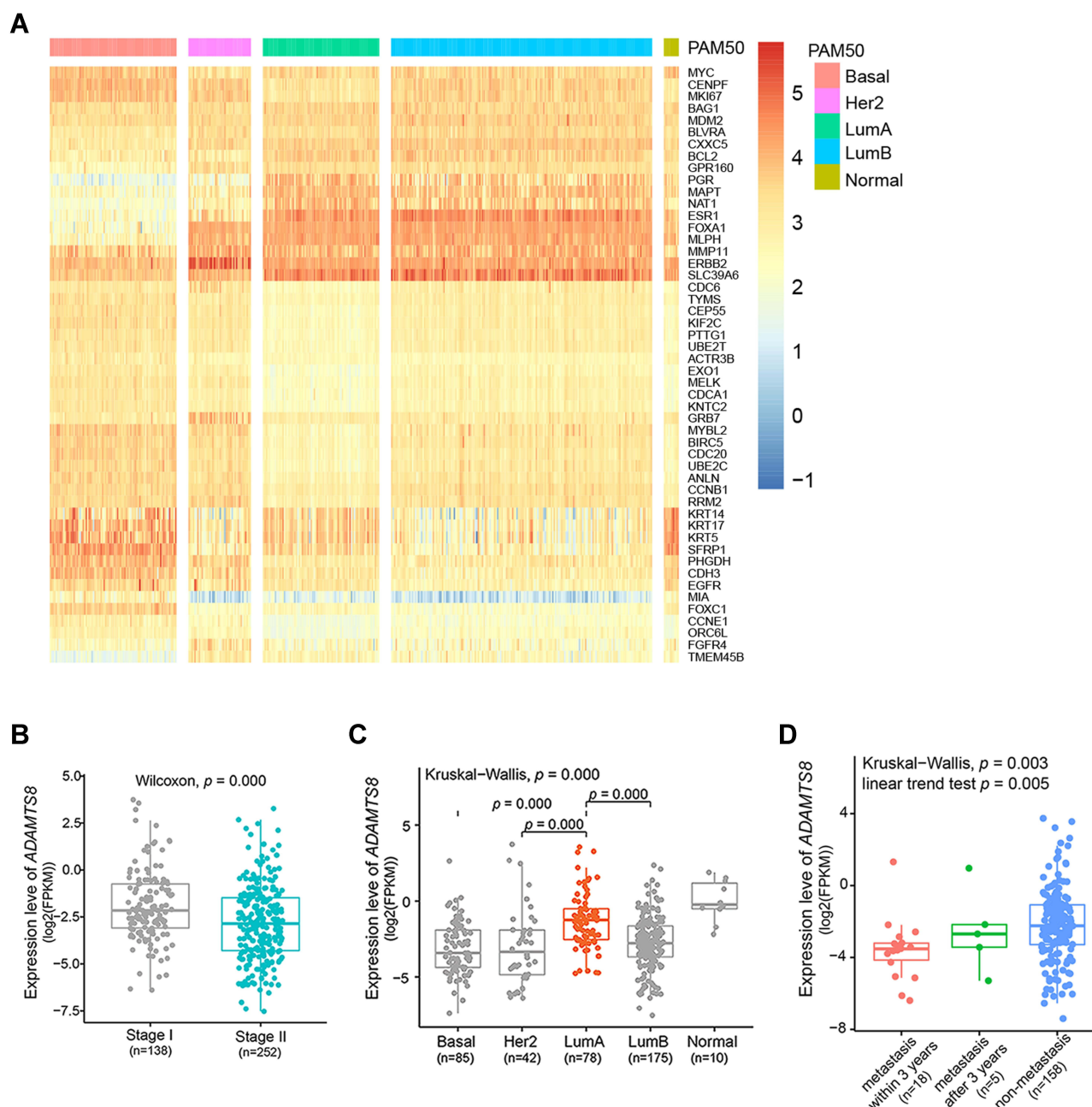

Figure 4 Lower *ADAMTS8* expression was related to poorer prognostic clinical stage and PAM50 subtypes and shorter DMFS time in the TCGA dataset. (**A**) PAM50 gene expression hierarchical in a heatmap. (**B–D**) Boxplots showing the distribution of *ADAMTS8* expression in patients stratified by clinical stage, PAM50 subtype and DMFS time performance status.

in the distribution of clinical stage, PAM50 subtypes and DMFS performance status in both datasets, but there were no common significant differences between *ADAMTS8* expression and age and tumor size in both datasets at the same time. Among all patients, the expression trend of *ADAMTS8* was lower in stage II patients than in stage I patients (Figure 4B; Supplementary Figure S1B), and more than half of the stage II patients (TCGA: 56.3%; GSE20685: 59.1%) had low *ADAMTS8* expression (Table 3). The expression trend of *ADAMTS8* was lower in the non-LumA subtype than in the LumA subtype (Figure 4C; Supplementary Figure S1C), and less than half of the LumA patients (TCGA: 24.4%; GSE20685: 41.2%) had low *ADAMTS8* expression (Table 3). Next, we grouped the TCGA and GSE20685 datasets by DMFS. We identified *ADAMTS8* expression in the subgroups of patients with metastasis within 3 years, metastasis after 3 years and non-metastasis. Patients with follow-up times less than 3 years were excluded. The detailed clinical data are presented in Supplementary Table S2. We found that the expression trend of *ADAMTS8* between the three subgroups was significantly different, and the linear trend test $p<0.05$ (Figure 4D, Kruskal–Wallis $p = 0.003$, linear trend test $p = 0.005$; Supplementary Figure S1D, Kruskal–Wallis $p = 0.043$, linear trend test $p = 0.034$). The patients in the metastasis within 3 years subgroup exhibited the lowest

ADAMTS8 expression trend, which implied that the lower the *ADAMTS8* expression was, the shorter the DMFS time.

GO and KEGG Analysis Reveals the Cell Signaling Pathways of ADAMTS8

All DEGs between the high *ADAMTS8* expression groups and the low *ADAMTS8* expression groups were analyzed by DAVID software and the GO and KEGG results ($p < 0.05$). As shown in Figure 5A–D, the DEGs were significantly enriched in GO terms and KEGG pathways, including ECM-receptor interaction, focal adhesion, angiogenesis, PI3K-AKT signaling pathway, cell proliferation and cell cycle, which showed that *ADAMTS8* may participate in the development and metastasis of BC through these pathways.

Examining ADAMTS8 Expression Level and Its Association with Prognosis in Our Clinical Specimens of LN− Early-Stage BC

The expression levels of *ADAMTS8* were detected by qRT–PCR in 25 LN− early-stage BC tumor tissue samples. The median age of the enrolled patients was 59 years, the oldest 73 years and the youngest 36 years. The median distant metastasis-free survival time for 8 patients who developed metastases was 3.41 years (range: 1.17–4.92 years), and for 17 non-metastatic patients, it was 6.17 years (range: 4.92–7.83 years). All patients had undergone

Table 3 Clinical Features of the Patients in TCGA and GSE20685 Datasets and Correlation of *ADAMTS8* Expression in LN− Early-Stage BC Patients

Feature	Subgroup	TCGA			GSE20685		
		Low (n=195)	High (n=195)	p^{a} value	Low (n=66)	High (n=65)	p^{a} value
Age N(%)	<40	13(61.9)	8(38.1)	0.081	17(68.0)	8(32.0)	0.116
	40–59	78(44.1)	99(55.9)		42(47.7)	46(52.3)	
	≥60	104(54.2)	88(45.8)		7(38.9)	11(61.1)	
Tumor size N(%)	T1	53(38.4)	85(61.6)	0.001	27(41.5)	38(58.5)	0.065
	T2	131(58.5)	93(41.5)		39(60.0)	26(40.0)	
	T3	11(39.3)	1760.7)		0(0.00)	1(100.0)	
Stage N(%)	I	53(38.4)	85(56.3)	**0.001**	27(41.5)	38(58.5)	**0.045**
	II	142(56.3)	110(43.7)		39(59.1)	27(40.9)	
Pam50 N(%)	Basal	53(62.4)	32(37.6)	**0.000**	18(72.0)	7(28.0)	**0.004**
	Her2	25(59.5)	17(40.5)		8(38.1)	13(61.9)	
	LumA	19(24.4)	59(75.6)		21(41.2)	30(58.8)	
	LumB	98(56.0)	77(44.0)		19(65.5)	10(34.5)	
	Normal	0(0.00)	10(100.0)		0(0.00)	5(100.0)	

Note: ^{a}p values were derived from chi-square test. Bold figure note: this variable is statistically significant.

A TCGA: KEGG enrichment analysis

B GSE20685: KEGG enrichment analysis

C TCGA: GO enrichment analysis
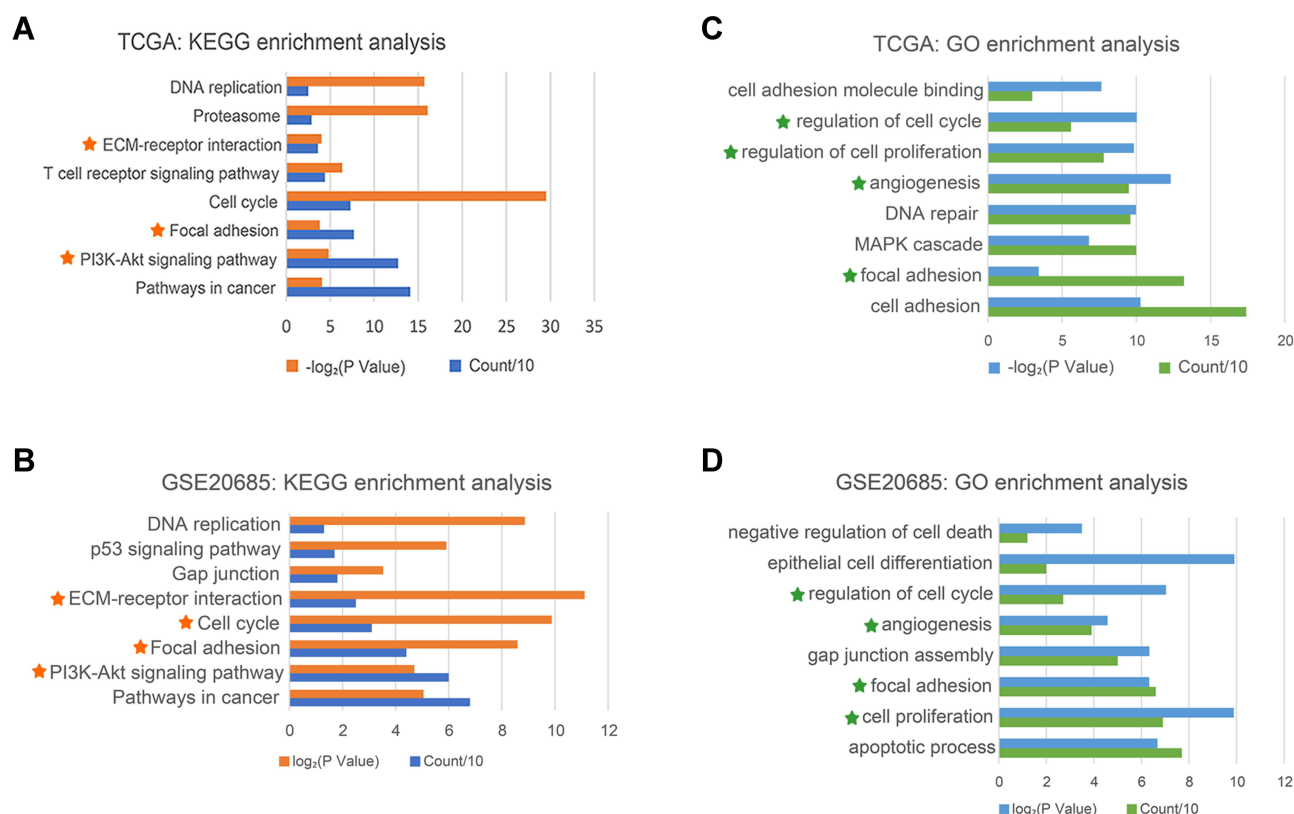
D GSE20685: GO enrichment analysis

Figure 5 Gene ontology and pathway enrichment analysis showed that *ADAMTS8* may participate in the development and metastasis of IDC. GO and KEGG analyses revealed the biological function of *ADAMTS8* in the TCGA dataset (**A** and **C**) and in the GSE20685 dataset (**B** and **D**). Pentagram marks indicate simultaneous enrichment of two datasets.

surgery and accepted systemic therapy. The detailed clinical information of 25 patients is included in Supplementary Table S3.

The results showed that the mRNA expression levels of *ADAMTS8* were lower in patients who developed metastases (Figure 6A, WilCoxon $p = 0.009$). Univariate Cox regression showed that low *ADAMTS8* expression was related to metastasis developed (HR = 8.639 [1.059–70.503], $p = 0.044$). Low expression of *ADAMTS8* was related to shortened DMFS compared with high *ADAMTS8* expression (Figure 6B, log-rank $p = 0.015$). There was no correlation between *ADAMTS8* expression and clinical features except DMFS, and the metastatic patients within 3 years were all concentrated in low *ADAMTS8* expression (Supplementary Table S4).

Discussion

In three independent datasets, we compared the gene expression profiles of LN− early-stage BC patients between the metastatic group and the non-metastatic group to identify DEGs and found for the first time that low *ADAMTS8*

expression is correlated with postoperative metastasis and shortened DMFS in this group of patients. In our clinical tumor specimens, we detected *ADAMTS8* expression levels in LN− early-stage BC patients by qRT–PCR. The results showed that *ADAMTS8* was significantly lower in patients who developed postoperative metastases. Downregulation of *ADAMTS8* was related to shorter DMFS.

ADAMTS8 is a member of the ADAMTS family known to have antiangiogenic properties. It has been shown to specifically inhibit endothelial growth factor VEGF-mediated angiogenesis in endothelial cells in vitro.[30] Some studies have demonstrated that *ADAMTS8* displays antitumor properties by antagonizing EGFR-MEK-ERK signaling.[31] Increasing evidence has indicated that *ADAMTS8* is an antioncogene in some cancers,[32–35] including breast cancer. Some studies have demonstrated that *ADAMTS8* inhibits cell proliferation and invasion and induces apoptosis in breast cancer (BC).[36] Another study demonstrated that *ADAMTS8* combined with *ADAMTS15* or 16 other immune genes was associated with overall survival in BC.[37,38] However, to the best of our
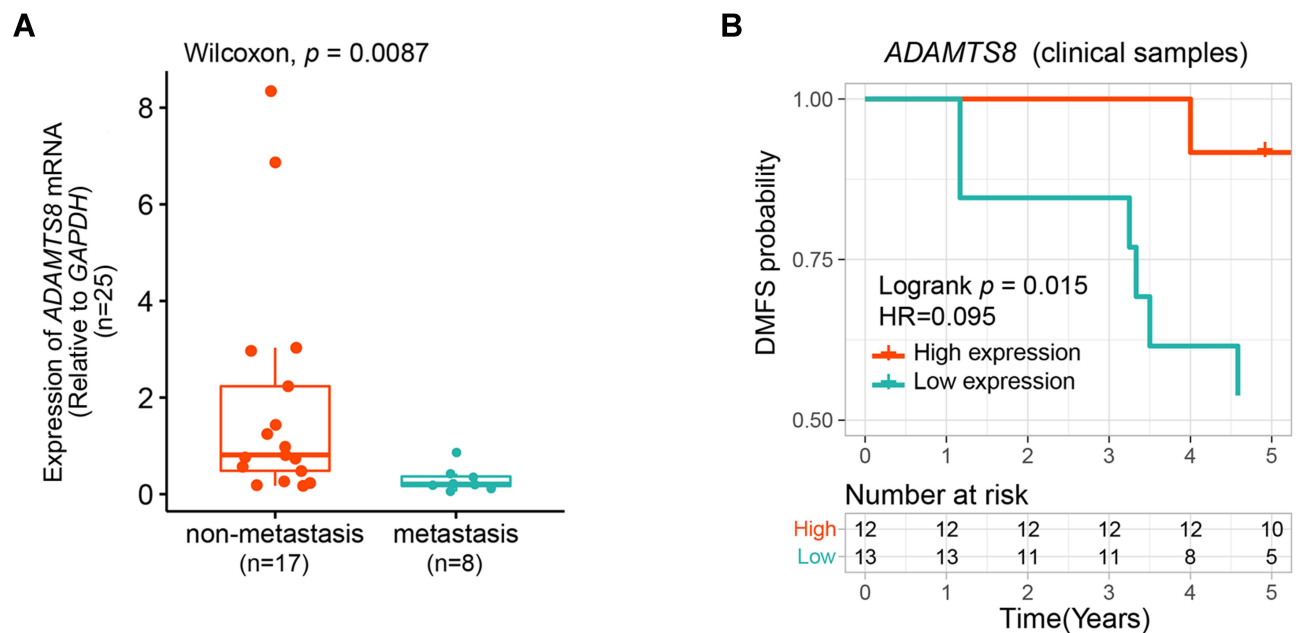
A

B

Wilcoxon, *p* = 0.0087

ADAMTS8 (clinical samples)

Logrank *p* = 0.015
HR=0.095

Figure 6 *ADAMTS8* is a potential prognostic biomarker for DMFS in real clinical samples of LNN early-stage BC patients. (**A**) qRT–PCR assays of the expression level of *ADAMTS8* in BC tumor samples. (**B**) Kaplan–Meier curves of DMFS between the *ADAMTS8* high and *ADAMTS8* low expression groups.

knowledge, no studies have reported the prognostic value of the *ADAMTS8* expression level for postoperative metastasis in LN− early-stage BC patients.

In TCGA and GSE20685 datasets, correlation analysis between *ADAMTS8* expression and clinical features indicated that the prediction results of *ADAMTS8* on patient prognosis were consistent with traditional prognostic indicators, including clinical stage, PAM50 subtypes and DMFS. In general, stage II patients are considered to have a worse prognosis than stage I patients. LN− LumA patients, which have the lowest proliferative ability of tumor cells and are sensitive to endocrine therapy, are considered to have the best prognosis.[39–41] The shorter DMFS, the poorer prognosis. In our study, lower *ADAMTS8* expression was related to poorer prognostic clinical stage and PAM50 subtypes and shorter DMFS. In our clinical samples, DMFS stratification of *ADAMTS8* expression showed similar results, and the metastatic patients within 3 years all had low *ADAMTS8* expression. No other clinical features were significantly different between high and low *ADAMTS8* expression, which may be due to our small clinical sample size.

To better understand the function of *ADAMTS8*, we performed GO and KEGG enrichment analysis, which showed that the DEGs between *ADAMTS8* low expression *ADAMTS8* and high expression were significantly enriched in ECM-receptor interaction, PI3K-AKT signaling pathway,[42–44] focal adhesion, and angiogenesis.[45–48] The PI3K-AKT pathway is

a downstream effector of the ECM-receptor pathway, and the results implied that breast tumor cell metastasis in patients with low *ADAMTS8* expression could be associated with ECM-receptor interactions and PI3K-AKT pathway activation, which has never been reported. However, further experiments are needed to directly test this hypothesis.

There were several limitations await to be addressed in our study. First, the sample size was insufficient for multivariate Cox regression analysis and in the stratified analysis of clinical features in our clinical samples. Second, no further experiments explored the role of *ADAMTS8* in breast cancer cell movement and metastasis.

Conclusion

In conclusion, we first found that the downregulated expression of *ADAMTS8* was associated with LN− early-stage BC patients prone to developing postoperative metastasis. We validated that *ADAMTS8* could be a potential biomarker for postoperative metastasis in LN− early-stage BC patients, and it is helpful to screen candidates with poorer prognosis in LN− early-stage patients so that they can receive intensive therapy.

Ethics Statement

The clinical tumor specimens from LN− early-stage BC patients were collected from the Breast Cancer Biobank of the Breast Centre at Beijing Hospital. The study protocol was approved by the Ethics Committee of Beijing Hospital

on the basis of the Declaration of Helsinki (IRB Number in Ethical approval: 2017BJYYEC-086-02), and written informed consent was obtained from the patients.

Funding

Beijing Hospital Clinical Research 121 Project(Project Number: BJ-2019-191); Beijing Hospital Project (Project Number: BJ-2021-212).

Disclosure

The authors declare there are no competing interest.

References

1. Lei S, Zheng R, Zhang S, et al. Global patterns of breast cancer incidence and mortality: a population-based cancer registry data analysis from 2000 to 2020. *Cancer commun*. 2021;41 (11):1183–1194. doi:10.1002/cac2.12207
2. Koboldt DC, Fulton R, McLellan M, et al.Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61–70. doi:10.1038/nature11412
3. Waks AG, Winer EP. Breast cancer treatment: a review. *JAMA*. 2019;321(3):288–300. doi:10.1001/jama.2018.19323
4. Cserni G. Histological type and typing of breast carcinomas and the WHO classification changes over time. *Pathologica*. 2020;112 (1):25–41. doi:10.32074/1591-951X-1-20
5. Harbeck N, Penault-Llorca F, Cortes J, et al. Breast cancer. *Nat Rev Dis Primers*. 2019;5(1):66. doi:10.1038/s41572-019-0111-2
6. Harbeck N, Gnant M. Breast cancer. *Lancet*. 2017;389 (10074):1134–1150. doi:10.1016/S0140-6736(16)31891-8
7. Shah C, Smith BD, Royce TJ. Balancing treatment deintensification strategies in early stage breast cancer. *Int J Radiat Oncol Biol Phys*. 2020;107(5):959–963. doi:10.1016/j.ijrobp.2020.04.037
8. Korde LA, Somerfield MR, Carey LA, et al. Neoadjuvant chemotherapy, endocrine therapy, and targeted therapy for breast cancer: ASCO guideline. *Int J Cancer*. 2021;39(13):1485–1505. doi:10.1200/ JCO.20.03399
9. Agostinetto E, Rediti M, Fimereli D, et al. HER2-low breast cancer: molecular characteristics and prognosis. *Cancers*. 2021;13(11):2824. doi:10.3390/cancers13112824
10. Pan H, Gray R, Braybrooke J, et al. 20-year risks of breast-cancer recurrence after stopping endocrine therapy at 5 years. *N Engl J Med*. 2017;377(19):1836–1846. doi:10.1056/NEJMoa1701830
11. Chu QD, Hsieh MC, Lyons JM, Wu XC. 10-year survival after breast-conserving surgery compared with mastectomy in Louisiana women with early-stage breast cancer: a population-based study. *J Am Coll Surg*. 2021;232(4):607–621. doi:10.1016/j.jamcollsurg.2020.11.011
12. Choi HY, Park M, Seo M, Song E, Shin SY, Sohn YM. Preoperative axillary lymph node evaluation in breast cancer: current issues and literature review. *Ultrasound Q*. 2017;33(1):6–14. doi:10.1097/ RUQ.0000000000000277
13. Curigliano G, Burstein HJ, Winer EP, et al. De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen International expert consensus conference on the primary therapy of early breast cancer 2017. *Ann Oncol*. 2017;28(8):1700–1712. doi:10.1093/annonc/mdx308
14. Harris LN, Ismaila N, McShane LM, et al. Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: American society of clinical oncology clinical practice guideline. *Int J Cancer*. 2016;34(10):1134–1150. doi:10.1200/ JCO.2015.65.2289

15. Cardoso F, Van't Veer LJ, Bogaerts J, et al. 70-gene signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med*. 2016;375(8):717–729. doi:10.1056/NEJMoa1602253
16. Albain KS, Barlow WE, Shak S, et al. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. *Lancet Oncol*. 2010;11(1):55–65. doi:10.1016/S1470-2045(09) 70314-6
17. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *Int J Cancer*. 2009;27 (8):1160–1167. doi:10.1200/JCO.2008.18.1370
18. Krop I, Ismaila N, Andre F, et al. Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: American society of clinical oncology clinical practice guideline focused update. *Int J Cancer*. 2017;35 (24):2838–2847. doi:10.1200/JCO.2017.74.0472
19. Kwa M, Makris A, Esteva FJ. Clinical utility of gene-expression signatures in early stage breast cancer. *Nat Rev Clin Oncol*. 2017;14(10):595–610. doi:10.1038/nrclinonc.2017.74
20. Zhu Y, Wang T, Tong Y, Chen X, Shen K. 21-gene recurrence assay associated with favorable metabolic profiles in HR-positive, HER2-negative early-stage breast cancer patients. *Front Endocrinol*. 2021;12:725161. doi:10.3389/fendo.2021.725161
21. Kok M, Koornstra RH, Mook S, et al. Additional value of the 70-gene signature and levels of ER and PR for the prediction of outcome in tamoxifen-treated ER-positive breast cancer. *Breast*. 2012;21(6):769–778. doi:10.1016/j.breast.2012.04.010
22. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res*. 2013;41 (D1):D991–D995. doi:10.1093/nar/gks1193
23. Blum A, Wang P, Zenklusen JC. SnapShot: TCGA-analyzed tumors. *Cell*. 2018;173(2):530. doi:10.1016/j.cell.2018.03.059
24. Kern SE. Why your new cancer biomarker may never work: recurrent patterns and remarkable diversity in biomarker failures. *Cancer Res*. 2012;72(23):6097–6101. doi:10.1158/0008-5472.CAN-12-3232
25. Sauerbrei W, Taube SE, McShane LM, Cavenagh MM, Altman DG. Reporting recommendations for tumor marker prognostic studies (REMARK): an Abridged explanation and elaboration. *J Natl Cancer Inst*. 2018;110(8):803–811. doi:10.1093/jnci/djy088
26. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44–57. doi:10.1038/nprot.2008.211
27. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat Genet*. 2000;25(1):25–29. doi:10.1038/75556
28. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000;28(1):27–30. doi:10.1093/nar/ 28.1.27
29. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods*. 2001;25(4):402–408. doi:10.1006/ meth.2001.1262
30. Vázquez F, Hastings G, Ortega MA, et al. METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new family of proteins with angio-inhibitory activity. *J Biol Chem*. 1999;274 (33):23349–23357. doi:10.1074/jbc.274.33.23349
31. Choi GC, Li J, Wang Y, et al. The metalloprotease *ADAMTS8* displays antitumor properties through antagonizing EGFR-MEK-ERK signaling and is silenced in carcinomas by CpG methylation. *Mol Cancer Res*. 2014;12(2):228–238. doi:10.1158/1541-7786.MCR-13-0195
32. Zhang X, Li D, Jia C, Cai H, Lv Z, Wu B. METTL14 promotes tumorigenesis by regulating lncRNA OIP5-AS1/miR-98/*ADAMTS8* signaling in papillary thyroid cancer. *Cell Death Dis*. 2021;12(6):617. doi:10.1038/s41419-021-03891-6

33. Wu Z, Shi Y, Ren S, Ju Y, Hu Y, Wu J. *ADAMTS8* inhibits progression of esophageal squamous cell carcinoma. *DNA Cell Biol.* 2020;39 (12):2300–2307. doi:10.1089/dna.2020.6053

34. Li L, Yuan S, Zhao X, Luo T. *ADAMTS8* is frequently down-regulated in colorectal cancer and functions as a tumor suppressor. *Biochem Biophys Res Commun.* 2020;524(3):663–671. doi:10.1016/j.bbrc.2020.01.020

35. Dunn JR, Reed JE, Du Plessis DG, et al. Expression of ADAMTS-8, a secreted protease with antiangiogenic properties, is downregulated in brain tumours. *Br J Cancer.* 2006;94(8):1186–1193.

36. Zhao X, Yang C, Wu J, Nan Y. *ADAMTS8* targets ERK to suppress cell proliferation, invasion, and metastasis of hepatocellular carcinoma. *Onco Targets Ther.* 2018;11:7569–7578. doi:10.2147/OTT.S173360

37. Zhang Z, Li J, He T, Ding J. Bioinformatics identified 17 immune genes as prognostic biomarkers for breast cancer: application study based on artificial intelligence algorithms. *Front Oncol.* 2020;10:330. doi:10.3389/fonc.2020.00330

38. Porter S, Span PN, Sweep FC, et al. *ADAMTS8* and ADAMTS15 expression predicts survival in human breast carcinoma. *Int J Cancer.* 2006;118(5):1241–1247. doi:10.1002/ijc.21476

39. Zhang Y, Schnabel CA, Schroeder BE, et al. Breast cancer index identifies early-stage estrogen receptor-positive breast cancer patients at risk for early- and late-distant recurrence. *Clin Cancer Res.* 2013;19(15):4196–4205. doi:10.1158/1078-0432.CCR-13-0804

40. Pu M, Messer K, Davies SR, et al. Research-based PAM50 signature and long-term breast cancer survival. *Breast Cancer Res Treat.* 2020;179(1):197–206. doi:10.1007/s10549-019-05446-y

41. Almstedt K, Mendoza S, Otto M, et al. EndoPredict(®) in early hormone receptor-positive, HER2-negative breast cancer. *Breast Cancer Res Treat.* 2020;182(1):137–146. doi:10.1007/s10549-020-05688-1

42. Gilkes DM, Semenza GL, Wirtz D. Hypoxia and the extracellular matrix: drivers of tumour metastasis. *Nat Rev Cancer.* 2014;14 (6):430–439. doi:10.1038/nrc3726

43. Gkretsi V, Stylianopoulos T. Cell adhesion and matrix stiffness: coordinating cancer cell invasion and metastasis. *Front Oncol.* 2018;8:145. doi:10.3389/fonc.2018.00145

44. Aoki M, Fujishita T. Oncogenic roles of the PI3K/AKT/mTOR axis. *Curr Top Microbiol Immunol.* 2017;407:153–189. doi:10.1007/82_2017_6

45. Carragher NO, Frame MC. Focal adhesion and actin dynamics: a place where kinases and proteases meet to promote invasion. *Trends Cell Biol.* 2004;14(5):241–249. doi:10.1016/j.tcb.2004.03.011

46. Burridge K. Focal adhesions: a personal perspective on a half century of progress. *FEBS J.* 2017;284(20):3355–3361. doi:10.1111/febs.14195

47. Lamalice L, Le Boeuf F, Huot J. Endothelial cell migration during angiogenesis. *Circ Res.* 2007;100(6):782–794. doi:10.1161/01.RES.0000259593.07661.1e

48. Ma Q, Reiter RJ, Chen Y. Role of melatonin in controlling angiogenesis under physiological and pathological conditions. *Angiogenesis.* 2020;23(2):91–104. doi:10.1007/s10456-019-09689-7