Enrichment of CD44 in basal-type breast cancer correlates with EMT, cancer stem cell gene profile, and prognosis

Abstract: Cluster of differentiation 44 (CD44) is a transmembrane glycoprotein that serves as the receptor for the extracellular matrix component hyaluronic acid. CD44 has been reported to play key roles in cell proliferation, motility, and survival, but its role in breast cancer remains controversial. In this study, we conducted a meta-analysis. A total of 23 published Gene Expression Omnibus databases were included to evaluate the association between CD44 mRNA expression and clinicopathological characteristics or prognosis of the patients with breast cancer. Our analysis revealed that CD44 expression was associated with clinicopathological features, including the histological grade, estrogen receptor status, progesterone receptor status, and human epidermal growth factor receptor-2 status. Higher levels of CD44 expression were observed in the basal subtype of breast cancer both at the mRNA and protein levels (odds ratio [OR] = 2.08, 95% confidence interval [CI]: 1.72–2.52; OR = 2.11, 95% CI: 1.67–2.68). Patients with CD44 overexpression exhibited significantly worse overall survival (hazard ratio = 1.27; 95% CI: 1.04–1.55). Whole gene profile analysis revealed that CD44 expression was enriched in basal-type breast cancer and correlated with epithelial–mesenchymal transition and cancer stem cell gene profiles. In summary, our analyses indicated that CD44 potentially might be a prognostic marker for breast cancer and thus can serve as a therapeutic target for basal-type breast cancer.

Keywords: breast cancer, CD44, survival prediction, meta-analysis, biomarker

Introduction

Breast cancer is one of the most common female cancers, accounting for approximately 28% of all female cancers and the second leading cause of cancer-related deaths in women. Progress has been made to the earlier diagnosis and better treatment of breast cancer during the past few decades, leading to the 5-year survival rates of breast cancer patients at approximately 85%. However, distant metastasis and recurrence still occur and result in poor prognosis. Therefore, there is an urgent need for identifying novel biomarkers that can be used to screen high-risk patients and help predict the progression and prognosis of breast cancer.

Cluster of differentiation 44 (CD44) is a complex transmembrane glycoprotein that is encoded by the CD44 gene on chromosome 11p13. CD44 consists of seven extracellular domains, a transmembrane domain, and a cytoplasmic domain. CD44 has several isoforms, including CD44s and CD44v. Functionally, CD44 was initially identified as the receptor for the extracellular matrix component, hyaluronic acid (HA), and was involved in multiple physiological and pathological processes, such as angiogenesis, cell adhesion, inflammation, and cancer development. In addition, CD44 has...
been reported to play important roles in cell proliferation, motility, and survival. A recent study indicated that CD44 expression was elevated in tumor-initiating cells in many kinds of cancer. Thus, CD44 is thought to be a biomarker for cancer stem cells (CSCs). Subsequent functional studies have shown that CD44 is involved in tumorigenesis and metastasis in many cancer types such as colon, bladder, gastric, and breast cancers. Studies on CD44 expression have suggested a correlation between it and clinical outcome in patients with breast cancer. It has been shown that the overexpression of CD44 has a bad impact on survival of breast cancer patients, but different results were also reported. Currently, the role of CD44 in breast cancer has not been clearly defined. To investigate the role of CD44 in breast cancer, a meta-analysis was performed. Our analysis indicated that CD44 expression was elevated in basal-type breast cancer. Currently, there are no effectively targeted therapies for patients with this subtype of breast cancer and prognosis is poor compared with other subtypes. Since CD44 expression is associated with mesenchymal and CSC signature and predicts poor prognosis, our study indicates that CD44 may represent a potential therapeutic target for basal-type breast cancer.

Materials and methods

Database and literature search

We performed a comprehensive search of relevant Gene Expression Omnibus (GEO) databases for CD44 mRNA expression and literature of CD44 protein level. First, we searched the ArrayExpress for uploaded databases within the topic of interest, using the search terms “breast cancer” by filtering Homo sapiens, RNA array, array assay, and all arrays. We also searched Oncomine for databases of breast cancer with mRNA information of CD44. Second, PubMed was reviewed to identify potentially relevant literatures using the search terms associated with CD44 (“CD44 antigen”, “hyaluronan-binding protein”, “receptors”, “hyaluronan”) and breast cancer (“breast neoplasm”, “breast tumor”, “breast carcinoma”, “mammary cancer”). The references were also searched to discover additional relevant publications.

Inclusion and exclusion criteria

This meta-analysis collected data aimed at evaluating the role of CD44 expression in breast cancer at both mRNA and protein levels. Databases that met the following criteria were included: 1) the datasets were about breast cancer; 2) CD44 expression was measured in these databases; 3) the sample capacity was >50; and 4) clinical information of patients was showed in these databases. The exclusion criteria were as follows: 1) the datasets were about animals such as mice and rabbits and 2) the datasets were about DNA, rather than RNA. When several databases shared the same patient population, only the latest and most complete datasets were included. Literature that met the following criteria were included: 1) patients recruited in the study were pathologically diagnosed with breast cancer; 2) CD44 expression was measured in breast cancer tissues; and 3) the hazard ratio (HR)/odds ratio (OR) and corresponding 95% confidence interval (CI) were reported or could be statistically extracted from the study. The exclusion criteria were as follows: 1) reviews, case reports, comments, letters, and conference abstract and 2) ineligible samples or those where the required data were not available. When several articles were from the same patient population, the latest or most complete article was included.

Data extraction

Data were abstracted in a standardized collection form, with information recorded as follows: last name of first author, publication year, country, duration, tumor–node–metastasis (TNM) stage, quality score, detection, and cutoff values for CD44. We reviewed ArrayExpress and Oncomine and identified 23 independent human breast cancer microarray datasets with CD44 mRNA expression and clinical data. Overall survival (OS), recurrence-free survival (RFS), and metastasis-free survival (MFS) were evaluated by Cox proportional HRs and 95% CIs using these numerical data. If HRs were not given in an article, we used the methods described by Tierney et al to calculate the statistical variables from published survival curves. The quality of observational studies was evaluated according to the Newcastle–Ottawa Quality Assessment Scale. This scale reflects patient selection, study comparability, and outcomes and is based on the identification of eight sources of potential study bias. Two reviewers performed the literature search, study selection, and data abstraction independently, and disagreements between the reviewers were solved by discussion.

Statistical analysis

Statistical analysis was performed based on the requirements of the meta-analysis of observational studies. The STATA software package (Version 12.0; StataCorp LP, College Station, TX, USA) was utilized to perform the meta-analysis. The random-effect model was employed when heterogeneity was present, and the fixed-effect model was used when homogeneity was demonstrated. The heterogeneity of publication was evaluated by means of the chi-square-based Q statistic and inconsistency index ($I^2$) statistic. Begg’s and Egger’s tests were employed to assess the publication bias. HRs
were employed to assess the survival outcome of patients with breast cancer who had high CD44 expression, and HR >1 indicated that high expression of CD44 predicted worse survival of patients. The OR and 95% CI were used to evaluate the association between CD44 expression and clinicopathological parameters.

**Results**

**Search results**

The flow diagram for the identification of relevant studies is shown in Figure 1. A total of 1,472 datasets and 1,147 literatures were initially identified by our search approach. For GEO databases, after the sample capacity and clinical information were checked, 23 datasets21,27–48 met the criteria for this analysis. For 1,147 literatures, after title/abstract scanning and full-text reading, 12 eligible articles22,49–59 were included. Table 1 shows the features of these 23 studies. Four Gene Expression Omnibus series (GSE) datasets were analyzed for finding the difference in CD44 mRNA expression between breast tumors and normal breast tissues. For finding the association between CD44 mRNA expression and TNM stage, tumor grade, estrogen receptor (ER) status, progesterone receptor (PR) status, human epidermal growth factor receptor-2 (HER2) status, and basal-like breast cancer, four, 13, eleven, four, six, and seven GSE datasets, respectively, were analyzed. To estimate the prognostic role of CD44 mRNA expression in OS, RFS, and MFS, eleven, ten, and nine, respectively, GSE datasets were adopted. Three GSE datasets were analyzed for the association between CD44 mRNA expression and the RFS in basal-like breast cancer. Table 2 shows the characteristics of 12 studies. A total of nine, eight, seven, and five articles were assessed for the correlation between CD44 protein abundance and ER status, PR status, HER2 status, and basal-like breast cancer, respectively. Clinical stages I and II were grouped as early-stage disease, whereas stages III and IV were grouped as late-stage disease. Clinical T stages 1 and 2 were identified as early T stage, and 3 and 4 were identified as late T stage. Clinical N stages 1 and 2 were classified into early N stage, and 3 and 4 were classified into late N stage. Histological grades I and II were pooled as low-grade disease, and III and IV were pooled as high-grade disease.

**CD44 expression correlates with clinicopathological features of breast cancer**

Eighteen studies assessed the association between CD44 mRNA expression and tumor clinicopathological features. Our meta-analysis indicated that CD44 expression in breast cancer tissues was increased when compared with that in normal breast tissues (pooled OR =1.15, 95% CI: 1.02–1.31, Cochran’s Q test P=0.070, and I²=57.5%; Figure 2A). However, there was no statistically significant correlation between CD44 expression and tumor TNM stage (pooled OR =1.10, 95% CI: 0.94–1.29, Cochran’s Q test P=0.039, and I²=64.1%; Figure 2B), T stage (pooled OR =1.00, 95% CI: 0.84–1.19, Cochran’s Q test P=0.137,
Table 1 Characteristics of the included studies by CD44 mRNA expression in the meta-analysis

<table>
<thead>
<tr>
<th>References</th>
<th>Year</th>
<th>Country or area</th>
<th>Duration (months)</th>
<th>Stage</th>
<th>Quality score</th>
<th>Detection</th>
<th>Cutoff values</th>
<th>Patients with CD44 overexpression (total number)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desmedt et al</td>
<td>2007</td>
<td>Canada</td>
<td>NA</td>
<td>NA</td>
<td>8</td>
<td>Microarray</td>
<td>Median expression: 9.34</td>
<td>99 (198)</td>
</tr>
<tr>
<td>Heikkinen et al</td>
<td>2011</td>
<td>Finland</td>
<td>36</td>
<td>NA</td>
<td>8</td>
<td>Microarray</td>
<td>Median expression: 11.05</td>
<td>92 (183)</td>
</tr>
<tr>
<td>Terunuma et al</td>
<td>2014</td>
<td>USA</td>
<td>NA</td>
<td>I–III</td>
<td>9</td>
<td>Microarray</td>
<td>Median expression: 11.12</td>
<td>31 (61)</td>
</tr>
<tr>
<td>Sircoulomb et al</td>
<td>2010</td>
<td>France</td>
<td>240</td>
<td>NA</td>
<td>7</td>
<td>Microarray</td>
<td>Median expression: 7.60</td>
<td>26 (51)</td>
</tr>
<tr>
<td>Clarke et al</td>
<td>2013</td>
<td>Canada</td>
<td>180</td>
<td>NA</td>
<td>8</td>
<td>Microarray</td>
<td>Median expression: NA</td>
<td>39 (77)</td>
</tr>
<tr>
<td>Ma et al</td>
<td>2009</td>
<td>USA</td>
<td>36</td>
<td>NA</td>
<td>7</td>
<td>Microarray</td>
<td>Median expression: 2.33</td>
<td>19 (38)</td>
</tr>
<tr>
<td>Hatzis et al</td>
<td>2011</td>
<td>USA</td>
<td>120</td>
<td>I–IV</td>
<td>9</td>
<td>Microarray</td>
<td>Median expression: 9.63</td>
<td>254 (508)</td>
</tr>
<tr>
<td>Tofigh et al</td>
<td>2014</td>
<td>Canada</td>
<td>NA</td>
<td>I–IV</td>
<td>8</td>
<td>Microarray</td>
<td>Median expression: 11.63</td>
<td>161 (321)</td>
</tr>
<tr>
<td>Kao et al</td>
<td>2011</td>
<td>Taiwan</td>
<td>156</td>
<td>I–IV</td>
<td>9</td>
<td>Microarray</td>
<td>Median expression: 8.06</td>
<td>164 (327)</td>
</tr>
<tr>
<td>Minn et al</td>
<td>2005</td>
<td>USA</td>
<td>NA</td>
<td>NA</td>
<td>8</td>
<td>Microarray</td>
<td>Median expression: 1,674.60</td>
<td>50 (99)</td>
</tr>
<tr>
<td>Desmedt et al</td>
<td>2011</td>
<td>Canada</td>
<td>NA</td>
<td>NA</td>
<td>8</td>
<td>Microarray</td>
<td>Median expression: 2.94</td>
<td>61 (120)</td>
</tr>
<tr>
<td>Schmidt et al</td>
<td>2008</td>
<td>Germany</td>
<td>120</td>
<td>NA</td>
<td>9</td>
<td>Microarray</td>
<td>Median expression: 1,595</td>
<td>100 (200)</td>
</tr>
<tr>
<td>Nagalla et al</td>
<td>2013</td>
<td>USA</td>
<td>NA</td>
<td>NA</td>
<td>8</td>
<td>Microarray</td>
<td>Median expression: 10.33</td>
<td>70 (139)</td>
</tr>
<tr>
<td>Loi et al</td>
<td>2007</td>
<td>Canada</td>
<td>NA</td>
<td>NA</td>
<td>9</td>
<td>Microarray</td>
<td>Median expression: 6.52</td>
<td>164 (327)</td>
</tr>
<tr>
<td>Dedeurwaerder et al</td>
<td>2011</td>
<td>Canada</td>
<td>NA</td>
<td>NA</td>
<td>8</td>
<td>Microarray</td>
<td>Median expression: 7.18</td>
<td>45 (90)</td>
</tr>
<tr>
<td>Miller et al</td>
<td>2005</td>
<td>Singapore</td>
<td>NA</td>
<td>NA</td>
<td>9</td>
<td>Microarray</td>
<td>Median expression: 8.04</td>
<td>126 (251)</td>
</tr>
<tr>
<td>Symmans et al</td>
<td>2010</td>
<td>USA</td>
<td>NA</td>
<td>NA</td>
<td>9</td>
<td>Microarray</td>
<td>Median expression: 10.78</td>
<td>149 (298)</td>
</tr>
<tr>
<td>Pawitan et al</td>
<td>2005</td>
<td>Sweden</td>
<td>24</td>
<td>NA</td>
<td>8</td>
<td>Microarray</td>
<td>Median expression: 7.04</td>
<td>80 (159)</td>
</tr>
<tr>
<td>Wang et al</td>
<td>2005</td>
<td>USA</td>
<td>180</td>
<td>NA</td>
<td>9</td>
<td>Microarray</td>
<td>Median expression: 1,285.15</td>
<td>143 (286)</td>
</tr>
<tr>
<td>Hu et al</td>
<td>2009</td>
<td>USA</td>
<td>NA</td>
<td>NA</td>
<td>7</td>
<td>Microarray</td>
<td>Median expression: NA</td>
<td>40 (80)</td>
</tr>
<tr>
<td>Hennessy et al</td>
<td>2009</td>
<td>USA</td>
<td>NA</td>
<td>NA</td>
<td>9</td>
<td>Microarray</td>
<td>Median expression: NA</td>
<td>47 (94)</td>
</tr>
<tr>
<td>Bild et al</td>
<td>2006</td>
<td>USA</td>
<td>NA</td>
<td>NA</td>
<td>9</td>
<td>Microarray</td>
<td>Median expression: 62.00</td>
<td>80 (158)</td>
</tr>
<tr>
<td>Minn et al</td>
<td>2007</td>
<td>USA</td>
<td>NA</td>
<td>NA</td>
<td>7</td>
<td>Microarray</td>
<td>Median expression: 1,636.50</td>
<td>29 (58)</td>
</tr>
</tbody>
</table>

Abbreviations: CD44, cluster of differentiation 44; NA, not available.
and \( F=33.9\% \); Figure 2C), and N status (pooled \( OR=0.98, 95\% CI: 0.91–1.06, \) Cochran’s \( Q \) test \( P=0.006, \) and \( F=57.8\% \); Figure 2D). Patients with breast cancer with higher histological grade were likely to have a higher content of CD44 at both mRNA (pooled \( OR=1.15, 95\% CI: 1.06–1.25, \) Cochran’s \( Q \) test \( P=0.582, \) and \( F=0\% \); Figure 2E) and protein levels (pooled \( OR=1.11, 95\% CI: 1.02–1.20, \) Cochran’s \( Q \) test \( P=0.055, \) and \( F=45.8\% \); Figure 2F).

**CD44 expression correlates with molecular subtypes of breast cancer**

The association of CD44 expression with ER, PR, HER2 status, and basal-like breast cancer was also analyzed. At the mRNA level, CD44 was inversely correlated with ER status (pooled \( OR=1.93, 95\% CI: 1.69–2.20, \) Cochran’s \( Q \) test \( P=0.002, \) and \( F=63.9\% \); Figure 3A), PR status (pooled \( OR=1.31, 95\% CI: 1.14–1.51, \) Cochran’s \( Q \) test \( P=0.173, \) and \( F=39.8\% \); Figure 3B), and HER2 status (pooled \( OR=1.05, 95\% CI: 1.00–1.10, \) Cochran’s \( Q \) test \( P=0.000, \) and \( F=82.4\% \); Figure 3C). Interestingly, CD44 mRNA expression was higher in basal-like tumors than in the luminal subtype of breast cancer (pooled \( OR=2.08, 95\% CI: 1.72–2.52, \) Cochran’s \( Q \) test \( P=0.001, \) and \( F=72.1\% \); Figure 3D). At the protein level, CD44 expression was conversely linked to ER status (pooled \( OR=1.31, 95\% CI: 1.15–1.48, \) Cochran’s \( Q \) test \( P=0.329, \) and \( F=12.7\% \); Figure 3E). However, there is no statistical significance in terms of an association between CD44 expression and PR status (pooled \( OR=0.99, 95\% CI: 0.90–1.08, \) Cochran’s \( Q \) test \( P=0.816, \) and \( F=0\% \); Figure 3F) or HER2 status (pooled \( OR=1.03, 95\% CI: 0.98–1.08, \) Cochran’s \( Q \) test \( P=0.008, \) and \( F=65.5\% \); Figure 3G) at protein level. Moreover, CD44 protein abundance in basal-like tumors was much higher than in the luminal subtype of breast cancer (pooled \( OR=2.11, 95\% CI: 1.67–2.68, \) Cochran’s \( Q \) test \( P=0.017, \) and \( F=66.9\% \); Figure 3H).

**CD44 mRNA expression correlates with breast cancer survival**

The association between CD44 expression level and breast cancer patient survival was analyzed. Our analysis indicated that there was a significant correlation between CD44 overexpression and the poor OS rate (pooled \( OR=1.27, 95\% CI: 1.04–1.55, \) Cochran’s \( Q \) test \( P=0.505, \) and \( F=0\% \); Figure 4A). However, CD44 expression was not statistically significant in terms of an association between the RFS rate (pooled \( OR=1.04, 95\% CI: 0.89–1.23, \) Cochran’s \( Q \) test \( P=0.417, \) and \( F=2.4\% \); Figure 4B) and the MFS rate (pooled \( OR=1.30, 95\% CI: 0.89–1.90, \) Cochran’s \( Q \) test \( P=0.010, \) and \( F=60.2\% \);
Figure 4C). Subcategory analyses according to the molecular classification of breast cancer were also performed. We found that higher CD44 mRNA expression correlated with worse RFS in patients with basal-like breast cancer (pooled OR = 1.84, 95% CI: 1.17–2.87, Cochran’s Q test P = 0.074, and F = 0.0%; Figure 4D). However, there was no statistically significant correlation between CD44 mRNA expression and the survival performance of patients with luminal subtype of breast cancer. The latter included the OS rate (pooled OR = 1.14, 95% CI: 0.73–1.79, Cochran’s Q test P = 0.296, and F = 0.296%), the RFS rate (pooled OR = 0.99, 95% CI: 0.75–1.31, Cochran’s Q test P = 0.258, and F = 0.258%), and the MFS rate (pooled OR = 1.25, 95% CI: 0.65–2.38, Cochran’s Q test P = 0.010, and F = 69.7%). Kaplan–Meier survival analysis of GSE3143
Figure 3 Association between CD44 expression and molecular subtype.

Notes: Association between CD44 mRNA with ER status (A), PR (B), HER2 (C), and basal–luminal (D). Association between CD44 protein with ER status (E), PR (F), HER2 (G), and basal–luminal (H).

Abbreviations: CD44, cluster of differentiation 44; CI, confidence interval; ER, estrogen receptor; HER2, human epidermal growth factor receptor-2; OR, odds ratio; PR, progesterone receptor.
demonstrated that there was a significant effect of CD44 on OS (P=0.016; Figure 5A). Kaplan–Meier survival analysis of GSE6532 showed that there were no significant effects of CD44 on the RFS in all population of breast cancer (P=0.743; Figure 5B), but it was inversely associated with the MFS rate (P=0.007; Figure 5C). Kaplan–Meier survival analysis of GSE25066 demonstrated that there was a significant effect of CD44 on the RFS in basal-like breast cancer (P=0.023; Figure 5F) but no significant association between CD44 mRNA expression and the RFS in all molecular subtypes (P=0.136; Figure 5D) or in luminal breast cancer (P=0.215; Figure 5E).

In all, the results from the CD44 mRNA profile indicated that higher CD44 expression predicted a poorer prognosis in patients with breast cancer subtype.

**CD44 correlates with epithelial–mesenchymal transition and CSC markers**

The association between CD44 and epithelial–mesenchymal transition (EMT) or CSC-related genes was also assessed. The results indicated that there was a positive relation between CD44 and SNAI1 (R=0.87, P<0.001; Figure 6A), SLUG (R=0.66, P<0.001; Figure 6B), ZEB1 (R=0.82, P<0.001; Figure 6C), CDH2 (R=0.83, P<0.001; Figure 6D), TWIST (R=0.40, P<0.001; Figure 6E), and VIM (R=0.72, P<0.001; Figure 6F). The association between CD44 and CSC markers was also evaluated. It was shown that CD44 was positively associated with ALDH1 (R=0.53, P<0.001; Figure 6G), SOX2 (R=0.86, P<0.001; Figure 6H), NANOG (R=0.78, P<0.001; Figure 6I), KLF4 (R=0.74, P<0.001; Figure 6J), MYC (R=0.68, P<0.001; Figure 6K), and OCT4 (R=0.87, P<0.001; Figure 6L).

**Publication bias**

Publication bias statistics were obtained using Begg’s and Egger’s tests, and did not indicate any significant publication bias; CD44 mRNA expression: breast cancer: Begg’s test P=0.734, Egger’s test P=0.905; TNM stage: Begg’s test P=1, Egger’s test P=0.796; tumor size: Begg’s test P=0.466,
Egger’s test \( P = 0.362 \); lymph node metastasis: Begg’s test \( P = 0.945 \), Egger’s test \( P = 0.097 \); histological grade: Begg’s test \( P = 0.246 \), Egger’s test \( P = 0.948 \); expression of ER: Begg’s test \( P = 0.640 \), Egger’s test \( P = 0.313 \); expression of PR: Begg’s test \( P = 1 \), Egger’s test \( P = 0.809 \); expression of Her2: Begg’s test \( P = 0.260 \), Egger’s test \( P = 0.494 \); basal-like breast cancer: Begg’s test \( P = 1 \), Egger’s test \( P = 0.77 \); OS: Begg’s test \( P = 0.436 \), Egger’s test \( P = 0.436 \); RFS: Begg’s test \( P = 0.592 \), Egger’s test \( P = 0.612 \); MFS: Begg’s test \( P = 0.251 \), Egger’s test \( P = 0.146 \); OS of luminal breast cancer: Begg’s test \( P = 1 \), Egger’s test \( P = 0.642 \); RFS of luminal breast cancer: Begg’s test \( P = 0.260 \), Egger’s test \( P = 0.436 \); MFS of luminal breast cancer: Begg’s test \( P = 0.806 \), Egger’s test \( P = 0.528 \); RFS of basal-like breast cancer: Begg’s test \( P = 1 \), Egger’s test \( P = 0.698 \). Protein level: histological grade: Begg’s test \( P = 0.917 \), Egger’s test \( P = 0.911 \); expression of ER: Begg’s test \( P = 0.917 \), Egger’s test \( P = 0.099 \); expression of PR: Begg’s test \( P = 0.266 \), Egger’s test \( P = 0.743 \); expression of HER2: Begg’s test \( P = 1 \), Egger’s test \( P = 0.434 \); and basal-like breast cancer: Begg’s test \( P = 0.462 \), Egger’s test \( P = 0.065 \).

**Discussion**

Molecular characterization contributes to the discovery of biomarkers and potential targets for anticancer therapy, which is the basis of precise medicine.\(^6\) Accumulating evidence suggests that CD44 is a marker of tumor-initiating cells, plays a role in tumorigenesis, and linked to the progression of breast cancer.\(^{15,61-63}\) CD44 was also reported to have an impact on the prognosis of breast cancer including recurrence\(^6\) and chemoresistance.\(^6\) Uchino et al found that the upregulation of CD44 represented an aggressive subtype in noninvasive breast cancer cell.\(^6\) The blockade of CD44 intracellular domain (CD44ICD) cleavage and nuclear translocation have been shown in cancer cells. The activation of CD44 by HA promoted the chemoresistance in breast cancer cells.\(^6\) CD44/cellular prion protein interaction has an effect on the responses to neoadjuvant chemotherapy in patients with breast cancer and exhibits aggressive behaviors of breast cancer cells.\(^6\) CD44–STAT3 interaction plays an important role in breast cancer invasion.\(^6\) Moreover, Cox regression analysis showed that ezrin and CD44 co-expression were independent prognostic factors of breast cancer.\(^6\)
In our meta-analysis, the role of CD44 in breast cancer, at both mRNA and protein levels, was investigated. We found that the mRNA level of CD44 was higher in breast tumor tissues than in normal breast tissues, indicating that CD44 might participate in the tumorigenesis of specific subtypes of breast cancer. Moreover, our meta-analysis suggests a positive association between histological grade and the CD44 levels. This would indicate that patients with high expression of CD44 mRNA might have poor prognosis, because high-grade tumor tends to be more aggressive and tends toward early recurrence. It has been shown that CD44 was activated in breast cancer cells but inactivated in normal cells in vitro and in vivo. However, the association between CD44 mRNA expression and TNM stage, T stage, and N stage was not statistically significant.

Based on the status of ER, PR, and HER2, breast cancer could be divided into five molecular subtypes, including normal-like, luminal A, luminal B, HER2-overexpressing, and
basal-like breast cancer. Each subtype exhibits distinctive expression patterns of specific molecules, clinical outcomes, and responses to adjuvant chemotherapy. Some studies indicated that CD44 expression was negatively associated with the status of ER, PR, and HER2. At the mRNA level, our meta-analysis showed that CD44 expression was significantly inversely associated with the status of ER, PR, and HER2. Consistently, significant correlation between CD44 expression and ER status was found at its protein level. Among the five molecular subtypes, basal-like breast cancer tends to be more aggressive and there is a lack of effective therapy, resulting in poorer outcomes. Jang et al. showed that CD44(+)CD24(−) subpopulation was much higher in basal-like breast cancer than that in non-basal-like cancer, and that CD44(+)CD24(−) cells had a high capacity of proliferation, migration, invasion, and tumorigenesis. By providing a highly hydrated environment favoring cellular invasion, HA–CD44 interaction contributed to the progression of basal-like breast cancer. Consistently, our results showed that CD44 expression was higher in basal-like breast cancers than in luminal breast cancer or all other subtypes.

CD44 is critical for regulating EMT. CD44 activation can lead to the expression of epithelial growth factor receptor and the activation of phosphoinositide-3 kinase/Akt. CD44 also upregulates N-cadherin, α-actin, vimentin, fibronectin, and other EMT markers. The latter is involved in cell invasion and migration. By knocking down CD44 expression in human hepatoma cell line HLE, the levels of snail and vimentin were decreased, which was correlated with a less-mesenchymal-like phenotype. Consistently, our analysis indicated that CD44 expression was significantly associated with mesenchymal gene SNAI1, SLUG, ZEB1, CDH2, and TWIST.

CD44 is a well-known breast CSC marker that plays a role in promoting tumorigenesis of breast cancer through interaction with its intracellular domain and stemness factors such as NANOG, OCT4, and SOX2. Analysis of gene expression profiles revealed that CD44 is closely associated with key stem cell genes ALDH1, SOX2, NANOG, KLF4, OCT4, and MYC. Since CSC is thought to be a major cause for cancer progression and therapeutic resistance, the role of CD44 in breast cancer might be attributable to those stem cell factors.

Studies identified several genes that might have prognostic values for breast cancer, including urokinase plasminogen activator and its inhibitor and the genes in the DACH–EYA–SIX pathway. Interestingly, insulin-like growth factor 1 receptor expression showed different prognostic values for patients with different subtypes of breast cancer. Ubiquitin protein D and KLF4 have been reported to predict the response to chemotherapy. Accumulating evidence indicates that CD44 could be a prognostic biomarker for breast cancer. Our meta-analysis suggested that CD44 high expression could be a prognostic marker for OS. Although there was no association between CD44 expression and RFS in the whole population of breast cancer, a significant association between CD44 mRNA expression and RFS in patients with basal-like breast cancer was identified. This agrees with a previous study showing that patients with CSC markers CD44(+)/CD24(−) had a lower survival rate, while patients without this subpopulation had a higher survival rate in basal-like breast cancer. Some studies showed that CD44 expression was positively correlated with the metastasis of breast carcinoma, but others reported opposite results. Martin and Jiang found that CD44 was markedly reduced in patients with ductal breast cancer with metastasis. Our meta-analysis showed that CD44 expression has no significant effect on the MFS (Figure 4C), but some GSE data did demonstrate that CD44 was correlated with the MFS (Figure 5C). Breast cancer metastasis is a complicated process which is involved in the alteration of a number of proteins, including epithelial growth factor receptor and transforming growth factor-β. Considering the complex regulation of the metastasis process of breast cancer, the effects of CD44 on the MFS might be covered by other factors.

Heterogeneity tests are essential to a meta-analysis. In this study, the evidence of minor heterogeneities was observed with respect to TNM stage, ER status, molecular subtypes, and the MFS. However, there was substantial heterogeneity with respect to HER2 status. This result might be due to the following aspects: 1) The sample size is limited, indicating that multicenter prospective studies are needed. 2) The variations in assessing CD44 mRNA expression might also contribute to heterogeneity. The cutoff value was estimated in 23 studies using the median CD44 level measured by gene microarray. 3) Publication bias is worth considering in meta-analyses. This study was a meta-analysis based on GEO datasets and published studies. Thus, our analysis has the following limitations: 1) we cannot exclude the publication bias; 2) the relevant papers were limited; and 3) methods and cutoff values used to assess CD44 expression were different.

**Conclusion**

In conclusion, our meta-analysis suggests that CD44 might be a prognostic factor for patients with breast cancer, particularly for the basal-like breast cancer. Since CD44 expression was elevated in basal-type breast cancer and its expression levels were correlated with EMT and CSC signatures, these
considerations might partially explain why patients with basal-type breast cancer have a high risk of metastasis and relapse. Moreover, our meta-analysis might help identify subpopulation of patients with breast cancer for CD44-based therapy in the future.

Acknowledgment
This work was supported by the National Natural Science Foundation of China (Nos 81572608, 81172422, and 81072169).

Disclosure
The authors report no conflicts of interest in this work.

References
Enrichment of CD44 in basal-type breast cancer


