Histopathology

Histopathology 2020, **77**, 437–447. DOI: 10.1111/his.14129

The prognostic significance of ALDH1A1 expression in early invasive breast cancer

Maryam Althobiti,[1,2,3] Rokaya El Ansari,[1,2] Mohammed Aleskandarany,[1,2] Chitra Joseph,[1,2] Michael S Toss,[1,2] Andrew R Green[1,2] & Emad A Rakha[1,2]

[1]*Nottingham Breast Cancer Research Centre, Division of Cancer and Stem Cells, School of Medicine, University of Nottingham Biodiscovery Institute, Nottingham, UK,* [2]*The University of Nottingham and Nottingham University Hospitals NHS Trust, Nottingham City Hospital, Nottingham, UK, and* [3]*Department of Clinical Laboratory Science, College of Applied Medical Science, Shaqra, Saudi Arabia*

Date of submission 24 January 2020
Accepted for publication 28 April 2020
Published online *Article Accepted* 5 May 2020

Althobiti M, El Ansari R, Aleskandarany M, Joseph C, Toss M S, Green A R & Rakha E A (2020) *Histopathology* **77**, 437–447. https://doi.org/10.1111/his.14129

The prognostic significance of ALDH1A1 expression in early invasive breast cancer

Aims: Aldehyde dehydrogenase family 1 member A1 (ALDH1A1) is reportedly a key ALDH isozyme linked to the cancer stem cells (CSC) of many solid tumours, where it is involved in self-renewal, differentiation and self-protection. In this study, the prognostic significance of ALDH1A1 expression in early invasive breast cancer (BC) and its role as a BC stem cell (BCSC) were evaluated.

Methods and results: ALDH1A1 expression was assessed, using immunohistochemistry and tissue microarrays, in a large well-characterised BC cohort. *ALDH1A1* mRNA expression was also assessed at transcriptomic levels, utilising data from the Molecular Taxonomy of Breast Cancer International Consortium. The associations of ALDH1A1 with clinicopathological parameters, other stem cell markers and patient outcomes were determined. ALDH1A1 was expressed in 71% of BC cases at both the protein and mRNA levels. High ALDH1A1 expression was associated with poor prognostic features, including high grade, poor Nottingham Prognostic Index (NPI), lymph node metastasis and highly proliferative ER^+ (luminal B) and triple-negative (TNBC) subtypes. ALDH1A1 expression was positively correlated with the expression of CD44, CD24, TWIST, SOX9, EPCAM and CD133. The high immunoexpression of ALDH1A1 was significantly associated with poor BC-specific survival ($P < 0.001$), and specifically in the luminal B and TNBC subtypes ($P = 0.042$ and $P = 0.003$, respectively). The immunoexpression of ALDH1A1 was an independent predictor of poor prognosis ($P = 0.015$).

Conclusions: ALDH1A1, as assessed using immunohistochemistry, seems to act as a BCSC marker associated not only with other BCSC markers but also with poor prognostic characteristics and poor outcomes, particularly in the luminal B and TNBC subtypes.

Keywords: ALDH1A1, breast cancer, immunohistochemistry, patient outcome, stem cell markers

Introduction

Breast cancer (BC) remains a life-threatening disease despite the advanced achievements in BC therapy.[1] Worldwide, it is the second most common cause of cancer-related mortality among women.[2] Cancer stem cells (CSC) are a subpopulation of cells within the primary tumour mass, which possess self-renewal, differentiation and potential tumorigenic properties.[3] The identification of CSC-specific biomarkers has been validated in BC models and has been shown to play a key role in drug resistance and poor BC outcomes.[4,5] Some molecular biomarkers have been studied

Address for correspondence: E A Rakha, Department of Histopathology, Nottingham University Hospital NHS Trust, City Hospital Campus, Hucknall Road, Nottingham NG5 1PB, UK. e-mail: emad.rakha@nottingham.ac.uk and emad.rakha@nuh.nhs.uk

previously in cancer and identified as BCSCs, including aldehyde dehydrogenase family 1 member A1 (ALDH1A1), CD44, CD24 and CD133.[6,7] ALDH1 is the most common BCSC marker, and has been studied in both *in-vitro* and *in-vivo* BC models.[8] ALDH1 is an enzyme that catalyses aldehydes to carboxylic acids and is also involved in the metabolism of several aliphatic and aromatic aldehydes.[9] ALDH1A1 isozyme oxidises retinaldehyde to retinoic acid, which regulates the expression of the genes involved in tumour-initiating stem-like cells, thereby initiating tumour growth and resistance to drugs.[10,11] Much emphasis has been placed on ALDH1A1 as a CSC marker. High expression of ALDH1A1 has been reported as a poor prognostic marker in several tumour types and is associated with poor patient outcomes.[12,13] However, the clinical significance of ALDH1A1 expression in human BC using immunohistochemistry (IHC) needs to be defined. This study, therefore, assessed the mRNA and protein expression of ALDH1A1 in BC with an emphasis on their correlation with other BCSCs markers, as well as with clinicopathological features and patient outcomes.

Materials and methods

TUMOUR CHARACTERISTICS OF THE STUDY COHORT (IHC)

The study cohort for the evaluation of ALDH1A1 protein expression comprised 930 early-stage (I–III) invasive BCs without distant metastases at the time of diagnosis (M0) derived from a retrospective series of primary breast carcinoma patients presenting at Nottingham City Hospital between 1988 and 1998.[14] The patients' clinical and pathological data included age at diagnosis, histological tumour type, tumour size, nodal stage (based on the Nottingham system for nodal stage: stage 1 for node-negative, stage 2 for one to three positive nodes and stage 3 for four or more positive nodes), Nottingham Prognostic Index (NPI) and lymphovascular invasion (LVI) status. Survival data were available and prospectively maintained, including (i) BC-specific survival (BCSS), defined as the duration (in months) from the date of the primary surgical treatment to the time of death from BC, and (ii) distant metastasis-free survival (DMFS), defined as the duration (in months) from surgery to the first event of distant metastasis.[15,16] The mean follow-up period of the study cohort was 130 months, median was 120 months and the range was 239 months. Patients were uniformly treated based on tumour features, NPI and hormone receptor status. Endocrine therapy was given to patients who had ER^+ tumours with high NPI scores (>3.4), whereas no adjuvant therapy was given to patients with excellent NPI scores (≤3.4). Premenopausal patients with moderate and poor NPI scores were candidates for chemotherapy, while postmenopausal patients with moderate or poor NPI scores were given hormonal therapy only. None of the patients in the current study cohort received neoadjuvant therapy. Data for the expression of oestrogen receptor (ER), progesterone receptor (PgR) and human epidermal growth factor receptor 2 (HER2), as well as Ki67, were also available.[17,18] Hormone receptor and HER2 status were assessed according to the recent ASCO guidelines. ER and PgR positivity were defined as ≥1%. HER2 positivity was defined as ≥10% of tumour cells showing strong membranous staining as score +3, where chromogenic *in-situ* hybridisation technique (CISH) was used to assess the gene amplification status in borderline cases (+2). Ki67 was considered positive/high if ≥10% of invasive BC cells showed nuclear positivity. BC molecular subtypes including luminal A ($ER^+/HER2^-$; Ki67 < 10%), luminal B ($ER^+/HER2^-$; Ki67 ≥ 10%), HER2-positive class ($HER2^+$ regardless of ER status) and TN (ER^-, PgR^- and $HER2^-$) were defined based on IHC profile.[19,20]

Immunohistochemical expression of BC stem cell-like markers, including CD133, CD24, CD44, TWIST (unpublished data), EPCAM and SOX9 (unpublished data), was previously performed and scored.[21–23] The number of informative cases scored per each marker was variable based on the available tissue microarray (TMA) cores.

The clinicopathological parameters for this study cohort are summarised in Table 1.

ALDH1A1 ANTIBODY VALIDATION

The specificity of the anti-ALDH1A1 antibody (rabbit monoclonal, HPA002123; Sigma-Aldrich, Poole, UK) was validated using Western blotting in MCF7 and HeLa human cell-line lysates, obtained from the American Type Culture Collection (Rockville, MD, USA). The ALDH1A1 primary antibody dilution used was 1:250, and the mouse monoclonal primary antibody beta-actin (diluted to 1:5000; Sigma-Aldrich) was used as a loading control. IRDye 800CW donkey antirabbit fluorescent secondary antibody and IRDye 800CW donkey antimouse fluorescent secondary antibody (LI-COR Biosciences, Lincoln, NE, USA) were used at a dilution of 1:15000. Odyssey Fc with Image Studio version 4.0 was used to visualise the protein

Table 1. The association of ALDH1A1 expression and clinicopathological parameters in invasive breast cancer Nottingham cohort for IHC protein expression

Parameters	ALDH1A1 expression (IHC)			
	N (%)	Low, *N* (%)	High, *N* (%)	χ^2 *P*-value
Patient age (years)				
<50	594 (64)	178 (30)	416 (70)	1.99
≥50	332 (36)	85 (26)	247 (74)	0.158
Tumour size (cm)				
≤2	467 57)	126 (27)	341 (73)	0.730
>2	355 (43)	134 (30)	320 (70)	0.90
Tumour grade				
Grade 1	148 (19)	50 (38)	82 (62)	9.91
Grade 2	263 (34)	87 (30)	202 (70)	**0.007**
Grade 3	373 (47)	123 (25)	378 (75)	
Tubule formation				
1	34 (4)	12 (35)	22 (65)	1.14
2	301 (34)	83 (27)	228 (73)	0.564
3	553 (62)	151 (27)	402 (73)	
Mitotic count scores				
1	264 (29)	91 (35)	173 (65)	11.29
2	188 (21)	53 (28)	135 (72)	**0.004**
3	446 (50)	102 (23)	344 (77)	
Nuclear pleomorphism				
1	18 (2)	9 (50)	9 (50)	13.76
2	311 (35)	103 (33)	208 (67)	**0.001**
3	567 (63)	134 (24)	433 (76)	
Axillary nodal stage*				
Stage 1	527 (65)	173 (30)	399 (70)	12.81
Stage 2	213 (27)	57 (21)	216 (79)	**0.002**
Stage 3	66 (8)	30 (39)	47 (61)	
Nottingham Prognostic Index				
Good prognostic group	206 (26)	90 (36)	116 (64)	9.58
Moderate prognostic group	408 (50)	26 (26)	382 (74)	**0.008**
Poor prognostic group	199 (24)	36 (24)	163 (76)	
Vascular invasion status				
Negative	490 (67)	183 (30)	418 (70)	4.094
Positive	244 (33)	77 (24)	242 (77)	**0.043**
IHC/molecular subtypes				
Luminal A	198 (25)	67 (34)	131 (66)	
Luminal B	301 (37)	66 (22)	235 (78)	9.11
HER2	130 (16)	32 (25)	95 (75)	**0.028**
Triple-negative	178 (22)	44 (25)	134 (75)	
Ki67				
Negative (<10%)	230 (37)	71 (30)	159 (70)	4.030
Positive (≥10%)	395 (63)	93 (24)	302 (76)	0.045

The significant *P*-value in bold.
ALDH1A1, Aldehyde dehydrogenase family 1 member A1; IHC, Immunohistochemistry; HER2, Human epidermal growth factor receptor 2.
*Ndal status is based on the Nottingham system nodal stage (1 for node-negative, 2 for 1–3 positive nodes and 3 for ≥4 positive nodes).

bands (LI-COR Biosciences). A single band for ALDH1A1 was observed at 56 kDa (the predicted size), which confirmed the specificity of the antibody (Figure S1A).

TISSUE MICROARRAY AND IHC

Formalin-fixed paraffin-embedded (FFPE) BC tissue samples were arrayed as described previously,[24] and constructed using a GrandMaster TMA arrayer Machine (3DHISTECH, Budapest, Hungary), where a 0.6-mm core was removed from the donor blocks to the recipient blocks. Each case was represented by a single core taken from the invasive edge of the tumour.

IHC staining was performed on 4-μm TMA sections using a Novolink polymer detection system (Leica, Newcastle, UK). In brief, the slides with the TMA sections were dewaxed and rehydrated. Antigen retrieval was performed in a citrate buffer (pH 6.0) in a

microwave (Whirlpool JT359 Jet Chef 1000W) for 20 min. The optimal dilution of ALDH1A1 antibody in IHC was 1:100, and this was incubated for 1 h at room temperature. Liver tissue was used as a positive control, while the negative control was obtained by omitting the primary antibody application step of the staining protocol. In addition, prior to immunostaining of the TMAs, full-face tissue sections from 20 randomly selected BC FFPE were stained for ALDH1A1 and assessed for the distribution of staining to determine the appropriateness of using the TMAs.

ASSESSMENT OF ALDH1A1 PROTEIN EXPRESSION

The TMA-stained slides were scanned at ×20 magnification into high-resolution digital images using a NanoZoomer scanner (NanoZoomer; Hamamatsu Photonics, Welwyn Garden City, UK). The images were viewed using Xplore viewing software viewer (Philips, Guildford, UK). ALDH1A1 staining was evaluated based on semiquantitative scoring using the modified histochemical score (H-score), which estimated both the intensity and the percentage of stained cells. The intensity was assessed as 0 = negative, 1 = weak, 2 = moderate and 3 = strong, and the percentage (0–100%) of positive stained tumour cells was evaluated. The final H-score was calculated by multiplying the percentage of positive cells (0–100) by the level of intensity (0–3), generating a total range of 0–300.[25] Furthermore, to test for the interobserver scoring reproducibility, 10% of the cases were randomly selected and double-scored. All cores were scored independently, blinded to histopathological data and patient outcomes.

TRANSCRIPTOMIC ANALYSIS OF *ALDH1A1*

The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) cohort of 1980 BC patients was evaluated for *ALDH1A1* mRNA expression and gene copy number (CN) aberration. In this cohort, DNA/RNA were isolated from fresh frozen samples and transcription profiling was achieved using the Illumina HT-12V3 platform. Details of the experimental and analytical methods used have been published.[26,27] In the cohort used for the current study, patients with ER-positive and/or lymph node-negative tumours did not receive adjuvant chemotherapy, while those with ER-negative and/or lymph node-positive tumours received this treatment. The relationships between *ALDH1A1* mRNA expression, CN aberration and clinicopathological parameters were investigated. The association between ALDH1A1 expression at both protein and mRNA levels was investigated in overlapping cases for both the METABRIC and Nottingham cohorts (n = 184 cases).

The clinicopathological parameters for this study cohort are summarised in Table S2.

STATISTICAL ANALYSIS

IBM SPSS 24.0 (Chicago, IL, USA) software was used for statistical analysis. Interobserver agreement was determined using kappa statistics. Furthermore, ALDH1A1 expression and other BCSCs were categorised at the proteomic and transcriptomic levels using X-tile software based on prediction of patient survival (X-tile Bioinformatics Software, Yale University, version 3.6.1). Associations between the categorical groups of ALDH1A1 (including protein and mRNA levels and CN aberration) and clinicopathological parameters and other BCSCs were analysed using the χ^2 test. Associations with patient outcome were assessed using Kaplan–Meier curves and the log-rank test. Cox proportional hazard regression models were built for multivariate survival analyses to estimate the hazard ratios (HR) of ALDH1A1 adjusted by other well-known prognostic factors. A P-value of less than 0.05 (two-tailed) was considered significant in all statistical tests. This study followed the criteria for the reporting recommendations for tumour marker prognostic studies (REMARK).[28]

RESEARCH INVOLVING HUMAN PARTICIPANTS AND/OR ANIMALS

This study was approved by the Nottingham Research Ethics Committee 2 under the title 'Development of a molecular genetic classification of breast cancer' and the North West–Greater Manchester Central Research Ethics Committee under the title 'Nottingham Health Science Biobank (NHSB)', reference number 15/NW/0685. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committees and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Release of data was also pseudoanonymised as per the UK Human Tissue Act regulations. This article does not contain any studies with animals performed by any of the authors.

Results

ALDH1A1 EXPRESSION IN BC

We observed a high level of reproducibility of ALDH1A1 scoring between the two observers as

assessed by the kappa statistic of almost perfect agreement (kappa = 0.865). ALDH1A1 immunostaining showed homogeneous staining distribution throughout the stained full-face tissue sections, which validated the use of TMAs, as shown in Figure S1B,C. ALDH1A1 expression was localised in the cytoplasm of the tumour cells with varying intensities, from absent to strong (Figure 1A,B). However, ALDH1A1 was further expressed in immune infiltrates and the stromal fibroblasts cells (Figure S1D,E). Moreover, strong intensity of ALDH1A1 expression was observed in the minority of tumour cells, while the majority showed weak to moderate intensity.

Positive ALDH1A1 expression at the X-tile-generated cut-off point of H-score 5 (>5 H-score) was observed in 663 of 930 (71%) cases. When ALDH1A1 protein expression was examined in BC molecular subtypes, ALDH1A1 was significantly highly expressed in the highly proliferative tumours, including luminal B ($P = 0.028$) (Table 1) although, at the mRNA level, *ALDH1A1* was also highly expressed in 1405 of 1980 (71%) cases at cut-off point (6.9 log fold-change) using X-tile. ALDH1A1 protein expression was not significantly associated with *ALDH1A1* mRNA expression ($P > 0.05$). *ALDH1A1* CN gain was observed in 35 of 1980 (1.7%) cases in the study cohort, and 32 of 1980 (1.6%) cases showed CN loss.

ALDH1A1 AND CLINICOPATHOLOGICAL PARAMETERS

ALDH1A1 protein expression was associated with aggressive prognostic features of BC including high grade ($P = 0.007$), high mitotic count ($P = 0.004$), increased nuclear pleomorphism ($P = 0.001$), poor NPI ($P = 0.008$), advanced nodal stage (four or more positive nodes) ($P = 0.002$) and LVI ($P = 0.043$), as shown in Table 1. High expression of *ALDH1A1* mRNA was positively associated with positive axillary nodes ($P = 0.002$), poor NPI ($P = 0.029$), small tumour size ($P < 0.0001$) and normal-like subtype ($P < 0.0001$), Table 2. There was no significant association between ALDH1A1 CN gain/loss and clinicopathological parameters ($P > 0.05$).

ASSOCIATION WITH OTHER BC STEM CELL MARKERS

The correlation of *ALDH1A1* mRNA and/or ALDH1A1 protein expression with BCSC-like markers was assessed using METABRIC (mRNA expression) and the available markers in the cohort (IHC expression). The biomarkers selected for this analysis were based on previous investigations reporting the marker as being either a BCSC marker or associated with the biological functions of ALDH1A1. ALDH1A1 was positively associated with CD24 ($n = 578$) ($P = 0.018$), CD44 ($n = 519$) ($P = 0.022$), CD133 ($n = 522$) ($P = 0.039$), TWIST ($n = 631$) ($P < 0.0001$), EPCAM ($n = 930$) ($P < 0.0001$) and SOX9 ($n = 929$) ($P = 0.034$), as shown in Table 3.

ALDH1A1 mRNA was positively associated with *CD44* ($n = 1156$) ($P = 0.023$), *TWIST* ($n = 1159$) ($P < 0.0001$), *SOX9* ($n = 1529$) ($P = 0.037$) and *EpCAM* ($n = 1343$) ($P = 0.012$), as shown in Table 4. There was no significant association between ALDH1A1 CN gain/loss and any of the BCSC markers investigated ($P > 0.05$).

OUTCOME ANALYSIS

Univariate analysis showed that a high expression of ALDH1A1 expression was associated with poor BCSS ($P < 0.0001$). When the analysis was limited to the

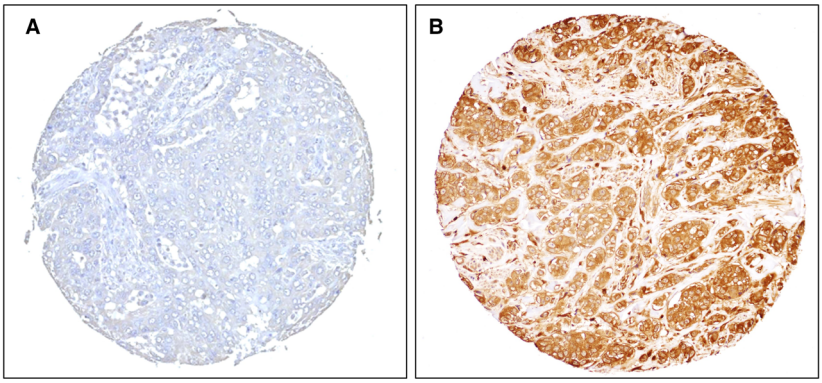

Figure 1. Immunohistochemical (IHC) expression of aldehyde dehydrogenase family 1 member A1 (ALDH1) in invasive breast cancer (BC). **A**, Negative expression of ALDH1A1. **B**, Positive IHC of ALDH1A1 expression. [Colour figure can be viewed at wileyonlinelibrary.com]

Table 2. The association of ALDH1A1 expression and clinicopathological parameters in invasive breast cancer METABRIC cohort for mRNA expression

	ALDH1A1 expression (mRNA)			
Parameters	*N* (%)	Low, *n* (%)	High, *n* (%)	χ^2 *P*-value
Patient age (years)				
<50	424 (21)	92 (25)	271(75)	2.12
≥50	1426 (79)	435 (29)	1055(71)	0.145
Tumour size (cm)				
≤2	623 (32)	124 (21)	467 (79)	220.80
>2	1337 (68)	404 (32)	871 (68)	<0.0001
Tumour grade				
Grade 1	148 (19)	39 (24)	123 (76)	1.94
Grade 2	263 (34)	203 (28)	534 (72)	0.379
Grade 3	373 (47)	265 (29)	643 (71)	
Axillary nodal stage*,*				
Stage 1	1035 (52)	308 (31)	681 (69)	12.17
Stage 2	623 (31)	162 (28)	428 (73)	**0.002**
Stage 3	315 (16)	64 (21)	241 (79)	
Nottingham Prognostic Index				
Good prognostic group	649 (34)	192 (30)	457 (70)	7.05
Moderate prognostic group	1049 (55)	305 (29)	744 (71)	**0.029**
Poor prognostic group	193 (10)	39 (20)	154 (80)	
PAM50 subtypes				
Luminal A	689 (37)	184 (27)	505 (73)	51.35
Luminal B	459 (24)	154 (34)	305 (66)	**<0.0001**
HER2**	308 (16)	107 (35)	201 (65)	
Basal	233 (12)	72 (31)	161 (69)	
Normal-like	196 (11)	17 (9)	179 (91)	

The significant *P*-value in bold.
ALDH1A1, Aldehyde dehydrogenase family 1 member A1; METABRIC, Molecular Taxonomy of Breast Cancer International Consortium.
*Nodal status is based on the Nottingham system nodal stage (1 for node-negative, 2 for 1–3 positive nodes and 3 for ≥4 positive nodes).
**Human epidermal growth factor receptor 2.

Table 3. Correlation of ALDH1A1 expression and other breast cancer stem cell markers at protein expression level

	ALDH1A1 status (IHC)			
Biomarkers	*N* (%)	Low, *n* (%)	High, *n* (%)	χ^2 *P*-value
CD44				5.25
Negative	291 (56)	90 (31)	201 (69)	**0.022**
Positive	228 (44)	50 (22)	178 (78)	
CD24				5.57
Negative	491 (85)	145 (29)	346 (71)	**0.018**
Positive	87 (15)	15 (17)	72 (83)	
CD133				4.274
Negative	442 (85)	212 (34)	230 (66)	**0.039**
Positive	80 (15)	18 (22)	62 (78)	
SOX9				4.48
Negative	377 (41)	121 (32)	256 (68)	**0.034**
Positive	552 (59)	142 (26)	410 (74)	
EPCAM				18.40
Negative	640 (69)	209 (33)	431 (67)	**<0.0001**
Positive	290 (31)	55 (19)	235 (81)	
TWIST				17.38
Negative	266 (42)	96 (36)	170 (64)	**<0.0001**
Positive	365 (58)	77 (21)	288 (79)	

The significant *P*-value in bold.
ALDH1A1, Aldehyde dehydrogenase family 1 member A1; IHC, Immunohistochemistry; HER2, Human epidermal growth factor receptor 2.

molecular subtypes, high ALDH1A1 expression, at the protein level, was associated with poor outcomes in the TNBC ($P = 0.003$) and luminal B subtypes ($P = 0.042$), but not in luminal A or $HER2^+$ tumours ($P > 0.05$) (Figure 2A–C). Cox proportional-hazards models incorporating the standard prognostic parameters (tumour size, tumour grade and nodal stage) showed that ALDH1A1 was an independent predictor of poor prognosis [$P = 0.001$, HR = 1524, 95% confidence interval (CI) = 1.180–1.968], as shown in Table 5. However, ALDH1A1 lost its significant association with outcome when analysis was limited to the luminal B and TNBC molecular subtypes, as shown in Table S1.

Table 4. Correlation of ALDH1A1 expression and other breast cancer stem cell markers at mRNA expression level

	ALDH1A1 status (mRNA)			
Biomarkers	*N* (%)	Low, *n* (%)	High, *n* (%)	χ^2 *P*-value
CD44				
Negative	907 (78)	260 (29)	647 (71)	5.17
Positive	249 (22)	90 (36)	159 (64)	**0.023**
CD24				
Negative	697 (41)	199 (29)	498 (71)	0.49
Positive	1009 (59)	304 (30)	705 (70)	0.482
CD133				
Negative	1027 (63)	313 (31)	711 (69)	1.87
Positive	596 (37)	163 (27)	433 (73)	0.170
SOX9				
Negative	732 (48)	420 (57)	312 (43)	4.35
Positive	797 (52)	499 (63)	298 (37)	**0.037**
EPCAM				
Negative	917 (68)	521 (57)	396 (43)	6.359
Positive	426 (32)	273 (64)	153 (36)	**0.012**
TWIST				
Negative	948 (82)	398 (42)	550 (58)	47.56
Positive	211 (18)	35 (17)	176 (83)	**<0.0001**

The significant *P*-value in bold.
ALDH1A1, Aldehyde dehydrogenase family 1 member A1.

ALDH1A1 mRNA expression level showed that high expression was associated with a favourable patient outcome ($P < 0.0001$). However, no significant association of *ALDH1A1* mRNA expression was identified with outcome in the BC molecular subtypes ($P > 0.05$). Multivariant analysis showed that *ALDH1A1* mRNA expression was an independent predictor of good prognosis ($P < 0.0001$, HR = 0.640, 95% CI = 0.527–0.777), as shown in Table 5. There was no significant association between ALDH1A1 CN gain/loss and patient outcome ($P > 0.05$).

Discussion

ALDH1A1 plays a major role in the progression of several solid tumours.[6,29–31] Ginestier *et al.*[4] have reported that ALDH1A1 is a specific marker for isolated BCSC. Furthermore, ALDH1A1 has been reported as being a marker for CSC in solid cancers such as lung[13] and colorectal cancer.[12]. However, Chang *et al.*[12] showed that ALDH1A1 is associated with improved patient outcome in ovarian cancer.

In this study, ALDH1A1 expression was assessed at the genomic, transcriptomic and proteomic levels in two large cohorts (METABRIC for both genomic, transcriptomic and Nottingham series for proteomic expression) of invasive BC, in order to understand its prognostic significance and utility as a BCSC marker. High protein and mRNA expression of ALDH1A1 was noted in 71% of the BC tumours, which was much higher than theoretically predicted. CSCs represent a small percentage (0.05–1%) of tumour cell populations.[32,33] Studies have reported that the expression of ALDH1A1 is much lower (~50%) in other study cohorts compared to the results of the current study.[34–36] However, the results of this study are in agreement with those of another study.[37] Such variability may have resulted from the use of different cut-offs and the adopted definition of ALDH1A1 positivity. Investigating the prognostic significance of ALDH1A1 in breast cancer using a large cohort, we have derived a cut-off value based on the prediction of patient survival, which might explain some of the differences seen with other studies. Additionally, the lack of concordance with other studies could partly be explained by the difference in the scoring methods. The H-scoring was used in this study, which is a widely accepted system in both clinical and research settings, and previous studies have used other scoring systems.

In addition, it could be speculated that the diffuse immunostaining of ALDH1A1 expression as an individual tissue marker may not predict the BCSCs population; rather, a combination of expression of more than one BCSC marker may more precisely determine CSCs using IHC. For example, Neumeister *et al.*[38] showed that both ALDH1A1 and CD44 expression in combination, using multiplex immunofluorescent in BC tissues, can predict patient survival rather than the isolated assessment of either. Furthermore, in the current study we observed strong immunoreactivity of ALDH1A1 in a minority of malignant cell populations, which could indicate that only this subpopulation may represent BCSCs. A previous study has also demonstrated low levels of strong ALDH1 expression in BC with only focal positivity, which may be consistent with the concept of CSCs that are represented by only a subpopulation of tumour cells.[39]

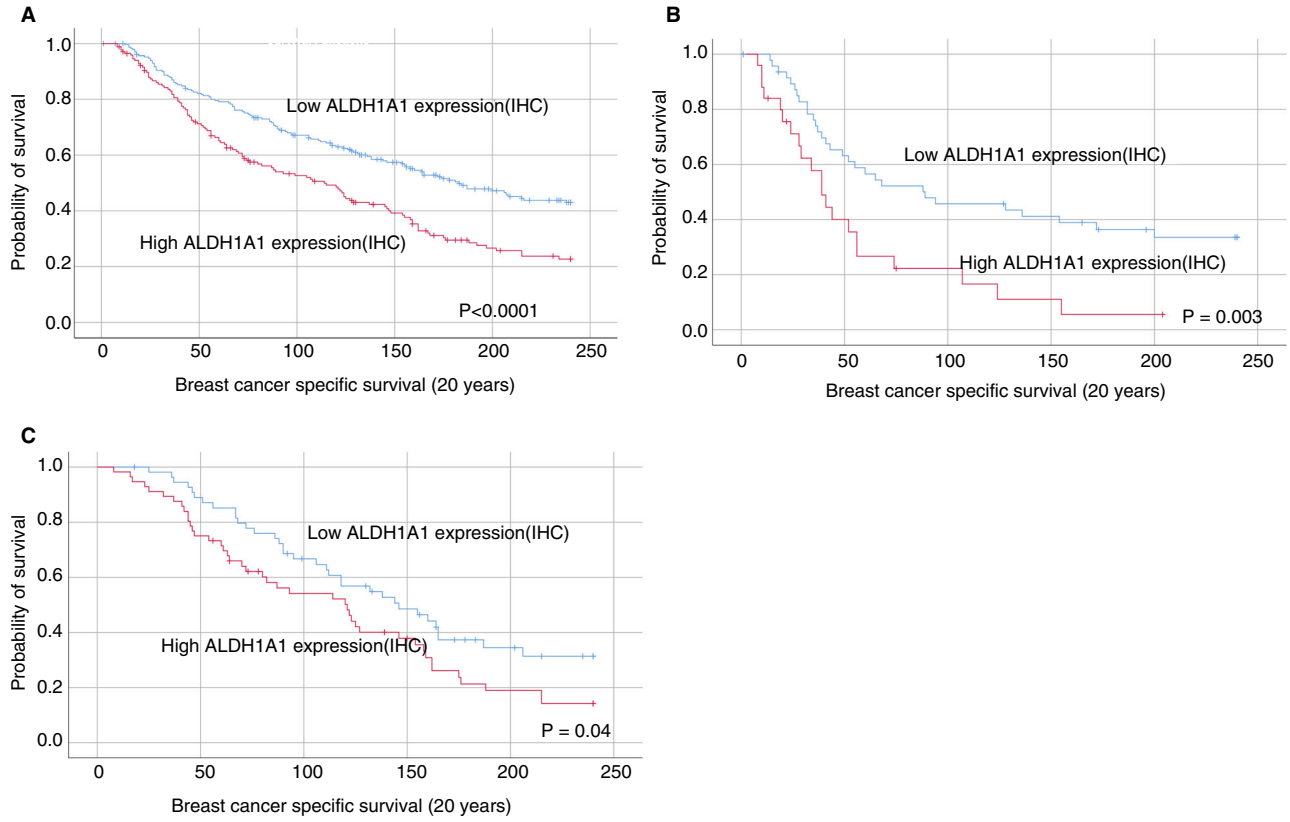

Figure 2. Kaplan–Meier survival plots for immunohistochemical expression of aldehyde dehydrogenase family 1 member A1 (ALDH1). **A–C**, BCSS in the whole breast cancer (BC) cohort, in triple-negative (TN) BC cases and in luminal B subtype, respectively. [Colour figure can be viewed at wileyonlinelibrary.com]

Table 5. Multivariate Cox regression hazard analysis including BC standard prognostic clinicopathological parameters and ALDH1A1 IHC and mRNA expression in whole cohort

	IHC expression			mRNA expression		
Parameters	*P*-value	Hazard ratio	95% CI	*P*-value	Hazard ratio	95% CI
ALDH1A1	**0.001**	1.524	1.180–1.968	<0.0001	0.640	0.527–0.777
Tumour size	0.008	1.395	1.089–1.785	<0.0001	1.516	1.203–1.910
Tumour grade	<0.0001	1.846	1.529–2.273	<0.0001	1.364	1.163–1.599
Axillary nodal stage	<0.0001	1.790	1.531–2.132	<0.0001	1.92	1.703–2.174

The significant *P*-value in bold.
ALDH1A1, Aldehyde dehydrogenase family 1 member A1; BC, Breast cancer; IHC, Immunohistochemistry; CI, Confidence interval.

Identification of CSCs is commonly accomplished using several biomarkers. There is no universal marker, nor is there a consensus for CSCs markers for all cancer types, and some molecules are frequently shared across entities. Al-Hajj *et al.*[40] identified a BCSC with respect to cell surface markers (CD44/CD24) shown to drive tumour growth. Furthermore, Ahmed *et al.*[22] showed that expression of CD44 and CD24 in BC tissue samples was significantly associated with more aggressive behaviour and worse patient outcomes. The findings of the current study indicate a significant correlation between the ALDH1A1 and $CD44^{+}/CD24^{+}$ phenotypes. Furthermore, ALDH1A1 also showed a significant positive

association with other BCSC markers, including CD133, TWIST1, SOX9 and EPCAM, at protein expression level. CD133/prominin 1 is a promising BCSC candidate marker, and maintains the characteristics of self-renewal, high proliferation and drug resistance.[21] Recently, an assessment of CD133 expression in BC concluded that its overexpression is likely a stem cell feature and prognostic marker in BC.[41] EPCAM and TWIST1 are well-established BCSCs and epithelial–mesenchymal transition (EMT) markers that can maintain self-renewal, the pluripotent phenotype and EMT in cancer progression.[42,43] Domenic *et al.*[44] reported that SOX9 expression may maintain human breast luminal progenitor cells, and may be associated with more aggressive tumours in BC.[45] The results of this study indicate that ALDH1A1 may be used not only as a CSC marker, but also as a prognostic and predictive marker, in BC. The protein expression of ALDH1A1 was more frequent in the highly proliferative (luminal B and TNBC) tumours, which suggests that ALDH1A1 may contribute to cell proliferation and tumorigenesis of BC.[46] Furthermore, in agreement with another study,[47] the current study showed that ALDH1A1 is positively associated with the proliferation marker Ki67. ALDH1A1 mainly catalyses retinaldehyde to retinoic acid, which subsequently binds and activates the retinoic acid or the retinoid X receptors in the nucleus of the cell, and thus promotes target gene expression. The downstream genes of the retinoic acid pathway are also involved in the differentiation and proliferation of tumour cells.[45] The study results demonstrate that high expression of ALDH1A1 at both protein and mRNA levels has a positive association with poor prognostic features. Fei *et al.*[39] concluded that ALDH1A1 predicts patient outcome in TNBC. Another study[48] reported that high expression of ALDH1A1 is associated with larger tumour size, higher grade and advanced tumour stage in BC, which is in agreement with the current study. The results of the current study showed high expression of ALDH1A1 in the luminal B and TN subtypes, which could further reflect the heterogeneity of BC with respect to stemness marker expression. Importantly, ALDH1A1 was an independent prognostic marker in BC, which may have clinical relevance in terms of the potential to improve survival rate prediction, especially in these specific subtypes.[48]

Contradictory findings between protein levels and mRNA levels were observed in the outcome analysis for this study, but they were similar to other study findings.[49] *ALDH1A1* mRNA expression showed an association with improved BC patients' survival. Another study[49] indicated that ALDH1A1 expression was associated with good outcomes in TNBC. Sjöström *et al.*[50] reported that stromal ALDH1A1 was associated with better distant disease-free survival in BC. A possible explanation of the survival discrepancy in mRNA expression is that the stromal expression of ALDH1A1 was not completely removed during tissue processing in the process of RNA extraction. This observation suggests that ALDH1A1 expression in stromal cells might be in cytotoxic T, B or dendritic cells, which are known to suppress tumour progression.[49] For IHC expression in this study, we only considered ALDH1A1 protein expression in the invasive tumour cells. Moreover, there was no significant correlation between ALDH1A1 protein and mRNA expression in a subset of 184 cases, which could be attributed to the same reason.

Conclusions

This study has demonstrated and confirmed that ALDH1A1 expression is associated with poor prognostic characteristics and survival outcomes in BC. Expression in BC was also associated with stem cell markers. The combination of the expression of ALDH1A1 with other BCSC markers in BC tissue samples may, therefore, serve as a tool to accurately identify the BCSC subpopulation.

Acknowledgement

We thank the University of Nottingham (Nottingham Life Cycle 6) for funding.

Conflicts of interest

The authors declare that they have no conflict of interests.

Author's contribution

Conception and design: MA, ER and AG. Staining and scoring analysis: MA and MT. Data analysis and interpretation: MA and RE. Manuscript writing and Reviewing: MA, RE, MT, CJ, MA, AG and ER.

References

1. Butti R, Gunasekaran VP, Kumar TVS, Banerjee P, Kundu GC. Breast cancer stem cells: biology and therapeutic implications. *Int. J. Biochem. Cell Biol.* 2019; **107**; 38–52.

2. Ferlay J, Colombet M, Soerjomataram I *et al*. Estimating the global cancer incidence and mortality in 2018: Globocan sources and methods. *Int. J. Cancer* 2019; **144**; 1941–1953.
3. Pires BR, De Amorim IS, Souza LD, Rodrigues JA, Mencalha AL. Targeting cellular signaling pathways in breast cancer stem cells and its implication for cancer treatment. *Anticancer Res.* 2016; **36**; 5681–5691.
4. Ginestier C, Hur MH, Charafe-Jauffret E *et al*. Aldh1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell* 2007; **1**; 555–567.
5. Ablett MP, O'Brien CS, Sims AH, Farnie G, Clarke RB. A differential role for cxcr4 in the regulation of normal versus malignant breast stem cell activity. *Oncotarget* 2014; **5**; 599–612.
6. Kong Y, Lyu N, Wu J *et al*. Breast cancer stem cell markers cd44 and aldh1a1 in serum: distribution and prognostic value in patients with primary breast cancer. *J. Cancer* 2018; **9**; 3728–3735.
7. Xia P. Cd133 mrna may be a suitable prognostic marker for human breast cancer. *Stem Cell Invest.* 2017; **4**; 87.
8. Martinez-Cruzado L, Tornin J, Santos L *et al*. Aldh1 expression and activity increase during tumor evolution in sarcoma cancer stem cell populations. *Sci. Rep.* 2016; **6**; 27878.
9. Moreb JS, Ucar D, Han S *et al*. The enzymatic activity of human aldehyde dehydrogenases 1a2 and 2 (aldh1a2 and aldh2) is detected by aldefluor, inhibited by diethylaminobenzaldehyde and has significant effects on cell proliferation and drug resistance. *Chem. Biol. Interact.* 2012; **195**; 52–60.
10. Vassalli G. Aldehyde dehydrogenases: not just markers, but functional regulators of stem cells. *Stem Cells Int.* 2019; **2019**; 3904645.
11. Ciccone V, Terzuoli E, Donnini S, Giachetti A, Morbidelli L, Ziche M. Correction to: Stemness marker aldh1a1 promotes tumor angiogenesis via retinoic acid/hif-1alpha/vegf signalling in mcf-7 breast cancer cells. *J. Exp. Clin. Cancer Res.* 2019; **38**; 45.
12. Yang W, Wang Y, Wang W, Chen Z, Bai G. Expression of aldehyde dehydrogenase 1a1 (aldh1a1) as a prognostic biomarker in colorectal cancer using immunohistochemistry. *Med. Sci. Monit.* 2018; **24**; 2864–2872.
13. Alamgeer M, Ganju V, Szczepny A *et al*. The prognostic significance of aldehyde dehydrogenase 1a1 (aldh1a1) and cd133 expression in early stage non-small cell lung cancer. *Thorax* 2013; **68**; 1095–1104.
14. Alshareeda AT, Negm OH, Albarakati N *et al*. Clinicopathological significance of ku70/ku80, a key DNA damage repair protein in breast cancer. *Breast Cancer Res. Treat.* 2013; **139**; 301–310.
15. Althobiti M, Aleskandarany MA, Joseph C *et al*. Heterogeneity of tumour-infiltrating lymphocytes in breast cancer and its prognostic significance. *Histopathology* 2018; **73**; 887–896.
16. Imrich S, Hachmeister M, Gires O. Epcam and its potential role in tumor-initiating cells. *Cell Adh. Migr.* 2012; **6**; 30–38.
17. Muftah AA, Aleskandarany MA, Al-Kaabi MM *et al*. Ki67 expression in invasive breast cancer: the use of tissue microarrays compared with whole tissue sections. *Breast Cancer Res. Treat.* 2017; **164**; 341–348.
18. Muftah AA, Aleskandarany M, Sonbul SN *et al*. Further evidence to support bimodality of oestrogen receptor expression in breast cancer. *Histopathology* 2017; **70**; 456–465.
19. Hammond ME, Hayes DF, Dowsett M *et al*. American society of clinical oncology/college of American pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). *Arch. Pathol. Lab. Med.* 2010; **134**; e48–e72.
20. Wolff AC, Hammond MEH, Hicks DG *et al*. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American society of clinical oncology/college of American pathologists clinical practice guideline update. *Arch. Pathol. Lab. Med.* 2014; **138**; 241–256.
21. Joseph C, Arshad M, Kurozomi S *et al*. Overexpression of the cancer stem cell marker CD133 confers a poor prognosis in invasive breast cancer. *Breast Cancer Res. Treat.* 2018; **174**; 387–399.
22. Ahmed MA, Aleskandarany MA, Rakha EA *et al*. A cd44(−)/cd24(+) phenotype is a poor prognostic marker in early invasive breast cancer. *Breast Cancer Res. Treat.* 2012; **133**; 979–995.
23. Agboola AJ, Paish EC, Rakha EA *et al*. Epcam expression is an indicator of recurrence in basal-like breast cancer. *Breast Cancer Res. Treat.* 2012; **133**; 575–582.
24. Abd El-Rehim DM, Ball G, Pinder SE *et al*. High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cdna expression analyses. *Int. J. Cancer* 2005; **116**; 340–350.
25. McCarty KS Jr, McCarty KS Sr. Histochemical approaches to steroid receptor analyses. *Semin. Diagn. Pathol.* 1984; **1**; 297–308.
26. Curtis C, Shah SP, Chin SF *et al*. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012; **486**; 346–352.
27. Pereira B, Chin SF, Rueda OM *et al*. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat. Commun.* 2016; **7**; 11479.
28. McShane LM, Altman DG, Sauerbrei W *et al*. Reporting recommendations for tumour marker prognostic studies (remark). *Br. J. Cancer* 2005; **93**; 387–391.
29. Qian X, Wagner S, Ma C *et al*. Prognostic significance of aldh1a1-positive cancer stem cells in patients with locally advanced, metastasized head and neck squamous cell carcinoma. *J. Cancer Res. Clin. Oncol.* 2014; **140**; 1151–1158.
30. van der Waals LM, Borel Rinkes IHM, Kranenburg O. Aldh1a1 expression is associated with poor differentiation, 'right-sidedness' and poor survival in human colorectal cancer. *PLoS ONE* 2018; **13**; e0205536.
31. Li XS, Xu Q, Fu XY, Luo WS. Aldh1a1 overexpression is associated with the progression and prognosis in gastric cancer. *BMC Cancer* 2014; **14**; 705.
32. Chang B, Liu G, Xue F *et al*. Aldh1 expression correlates with favorable prognosis in ovarian cancers. *Mod. Pathol.* 2009; **22**; 817–823.
33. Li Y, Kong D, Ahmad A, Bao B, Sarkar FH. Pancreatic cancer stem cells: emerging target for designing novel therapy. *Cancer Lett.* 2013; **338**; 94–100.
34. Toll AD, Dasgupta A, Potoczek M *et al*. Implications of enhancer of zeste homologue 2 expression in pancreatic ductal adenocarcinoma. *Hum. Pathol.* 2010; **41**; 1205–1209.
35. Kida K, Ishikawa T, Yamada A *et al*. Effect of ALDH1 on prognosis and chemoresistance by breast cancer subtype. *Breast Cancer Res. Treat.* 2016; **156**; 261–269.
36. Nalwoga H, Arnes JB, Wabinga H, Akslen LA. Expression of aldehyde dehydrogenase 1 (aldh1) is associated with basal-like markers and features of aggressive tumours in african breast cancer. *Br. J. Cancer* 2009; **102**; 369.

37. Neumeister V, Agarwal S, Bordeaux J, Camp RL, Rimm DL. In situ identification of putative cancer stem cells by multiplexing aldh1, cd44, and cytokeratin identifies breast cancer patients with poor prognosis. *Am. J. Pathol.* 2010; **176**; 2131–2138.
38. Ma F, Li H, Li Y *et al*. Aldehyde dehydrogenase 1 (aldh1) expression is an independent prognostic factor in triple negative breast cancer (tnbc). *Medicine* 2017; **96**; e6561.
39. Waterworth A. Introducing the concept of breast cancer stem cells. *Breast Cancer Res.* 2003; **6**; 53.
40. Horst D, Kriegl L, Engel J, Kirchner T, Jung A. Cd133 expression is an independent prognostic marker for low survival in colorectal cancer. *Br J Cancer* 2008; **99**; 1285–1289.
41. Gonzalez B, Denzel S, Mack B, Conrad M, Gires O. Epcam is involved in maintenance of the murine embryonic stem cell phenotype. *Stem Cells* 2009; **27**; 1782–1791.
42. Grzegrzolka J, Biala M, Wojtyra P *et al*. Expression of emt markers slug and twist in breast cancer. *Anticancer Res.* 2015; **35**; 3961–3968.
43. Domenici G, Aurrekoetxea-Rodriguez I, Simoes BM *et al*. A Sox2-Sox9 signalling axis maintains human breast luminal progenitor and breast cancer stem cells. *Oncogene* 2019; **38**; 3151–3169.
44. Chakravarty G, Moroz K, Makridakis NM *et al*. Prognostic significance of cytoplasmic sox9 in invasive ductal carcinoma and metastatic breast cancer. *Exp. Biol. Med.* 2011; **236**; 145–155.
45. Yao J, Jin Q, Wang X-D, Zhu H-J, Ni Q-C. Aldehyde dehydrogenase 1 expression is correlated with poor prognosis in breast cancer. *Medicine* 2017; **96**; e7171.
46. Inwald EC, Klinkhammer-Schalke M, Hofstädter F *et al*. Ki-67 is a prognostic parameter in breast cancer patients: results of a large population-based cohort of a cancer registry. *Breast Cancer Res. Treat.* 2013; **139**; 539–552.
47. Marchitti SA, Brocker C, Stagos D, Vasiliou V. Non-p450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. *Expert Opin. Drug Metab. Toxicol.* 2008; **4**; 697–720.
48. Liu Y, Baglia M, Zheng Y *et al*. Aldh1a1 mrna expression in association with prognosis of triple-negative breast cancer. *Oncotarget* 2015; **6**; 41360–41369.
49. Sjostrom M, Hartman L, Honeth G *et al*. Stem cell biomarker ALDH1A1 in breast cancer shows an association with prognosis and clinicopathological variables that is highly cut-off dependent. *J. Clin. Pathol.* 2015; **68**; 1012–1019.
50. Bednarz-Knoll N, Nastaly P, Zaczek A *et al*. Stromal expression of aldh1 in human breast carcinomas indicates reduced tumor progression. *Oncotarget* 2015; **6**; 26789–26803.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. (A) Western blot of rabbit monoclonal anti-ALDH1A1 antibody shows a single specific band (left green band) at the expected molecular weight (56 kDa) in HeLa cell lysates. The red bands represent the beta-actin (positive control) at 42 kDa molecular weight. (B, C) Homogenous expression of ALDH1A1 in invasive breast cancer. (D, E) ALDH1A1 expression in stromal cells.

Table S1. Multivariate Cox regression hazard model including other prognostic clinicopathological parameters and ALDH1A1 (IHC) in luminal B and TNBC.

