The Molecular Mechanisms Underlying Reduced E-cadherin Expression in Invasive Ductal Carcinoma of the Breast: High Throughput Analysis of Large Cohorts

Mansour Alsaleem[1,2], Michael S Toss[1], Chitra Joseph[1], Mohammed Aleskandarany[1,7], Sasagu Kurozumi[1], Ibrahim Alshankyty[5], Angela Ogden[8], Padmashree C G Rida[8], Ian O Ellis[1], Ritu Aneja[8,] Andrew R. Green[1], Nigel P Mongan[3,4] and Emad A. Rakha[1,7]

[1] Nottingham Breast Cancer Research Centre, Division of Cancer and Stem Cells, School of Medicine, University of Nottingham, Nottingham, UK.
[2] Faculty of Applied Medical Sciences, Onizah Community College, Qassim University, Qassim, Saudi Arabia.
[3] Cancer Biology and Translational Research, Faculty of Medicine and Health Sciences, University of Nottingham, Nottingham, UK.
[4] Department of Pharmacology, Weill Cornell Medicine, New York, USA.
[5] Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah, Saudi Arabia.
[7] Faculty of Medicine, Menoufyia University, Shebin El Kom, Egypt.
[8] Department of Biology, Georgia State University, Atlanta, GA, USA

Corresponding author:

Professor Emad Rakha

Department of Histopathology, Division of Cancer and Stem Cells, School of Medicine, The University of Nottingham and Nottingham University Hospitals NHS Trust, Nottingham City Hospital, Nottingham, NG5 1PB, UK
Email: Emad.Rakha@nottingham.ac.uk

Key Words: E-cadherin, mechanisms, invasive ductal carcinoma, copy number, Next generation sequence

Running title: Molecular mechanisms of E-cadherin loss in invasive ductal carcinoma of the breast

ABSTRACT

E-cadherin is a tumor suppressor gene in invasive lobular breast cancer. However, a proportion of high-grade ductal carcinoma shows reduced/loss of E-cadherin. In this study, we assessed the underlying mechanisms and molecular implications of E-cadherin loss in invasive ductal carcinoma. This study utilized large, well-characterized cohorts of early stage breast cancer evaluated E-cadherin expression via various platforms including immunohistochemistry, microarray analysis using Illumina HT-12V3, copy number analysis using Affymetrix SNP 6.0 arrays and next generation sequencing for differential gene expression. Our results showed 27% of high-grade invasive ductal carcinoma showed reduced/loss of E-cadherin membranous expression. *CDH1* copy number loss was in 21% of invasive ductal carcinoma which also showed low *CDH1* mRNA expression (p=0.003). *CDH1* copy number was associated with copy number loss of *TP53, ATM, BRCA1* and *BRCA2* (p<0.001). 79% of invasive ductal carcinoma with reduced *CDH1* mRNA expression showed elevated expression of E-cadherin transcription suppressors *TWIST2, ZEB2, NFKB1, LLGL2, CTNNB1* (p<0.01). Reduced/loss E-cadherin expression was associated with differential expression of 2143 genes including those regulating Wnt (*FZD2, GNG5, HLTF, WNT2, and CER1*), and PIK3-AKT (*FGFR2, GNF5, GNGT1, IFNA17, and IGF1*) signaling pathways. Interestingly, key genes differentially expressed between invasive lobular carcinoma and invasive ductal tumors did not show association with E-cadherin loss in invasive ductal carcinoma. We conclude that E-cadherin loss in invasive ductal carcinoma is likely a consequence of genomic instability occurring during carcinogenesis. Potential novel regulators controlling E-cadherin expression in invasive ductal carcinoma warrant further investigation.

INTRODUCTION

Worldwide, breast cancer accounts for 23% of total diagnosed cancer cases, and is the second leading cause of cancer-related death amongst women [1]. In solid tumors, cell-cell decohesion is a recognized phenomenon allowing tumor cells to grow invasively into surrounding tissues [2]. E-cadherin, a calcium-dependent adhesion molecule encoded by the *CDH1* gene located on chromosome 16q22.1 [3], plays an important role in gland formation, cell differentiation, polarity and maintaining the integrity of epithelial cells [4]. Subsequently, decreased expression of E-cadherin, which is frequently seen in breast cancer, may lead to cellular de-differentiation and invasiveness [5-6].

Reduced/loss of E-cadherin expression in the vast majority of invasive lobular carcinomas and lobular carcinoma in situ together with loss of CDH1 gene copy number [7-8-9] or CDH1 gene mutation [10] in a large proportion of cases suggests a plausible role for E-cadherin as a tumor suppressor gene [11-9]. However, there is limited evidence to support a role for E-cadherin as a tumor suppressor gene in invasive ductal carcinoma [12]. In fact, ductal carcinoma *in situ* and low grade invasive ductal carcinoma generally show stronger E-cadherin membrane staining than that seen in the normal breast epithelial cells, denoting increased expression, rather than a loss of expression [13]. Although some studies indicated that a proportion of invasive ductal carcinoma shows loss/reduced E-cadherin protein expression, these tumors were typically high-grade aggressive tumors. Of note, accumulating evidence suggests that high grade invasive ductal carcinoma are characterized by genomic instability with loss of increasing number of tumor suppressor genes during the carcinogenesis process that contributes to their aggressive behavior [12]. In addition, reduced/loss of E-cadherin expression is frequently associated with loss of estrogen expression, larger tumor size, and with the development of metastasis and recurrence [14-15-16-17]. These findings suggest that E-cadherin loss occurs as a late event in the process of carcinogenesis arising in association with or as a part of genomic instability, rather than as an early neoplastic event as seen in invasive lobular carcinoma [13-18-19]. However, the reasons for dysregulation of E-cadherin protein expression remain ill-defined [20].

We therefore aimed to study the mechanisms of reduced/loss E-cadherin expression in high grade invasive ductal carcinoma compared to invasive lobular carcinoma and its potential molecular implications.

MATERIALS AND METHODS

Study cohort

This study was conducted on multiple well-characterized cohorts of high grade invasive ductal carcinoma using different molecular techniques (Supplementary Table 1). First, a well-characterized cohort of primary grade 3 invasive ductal carcinoma from patients presenting to Nottingham City Hospital between 1989 and 1998 (n=813), and for whom detailed clinicopathologic data were available was used to determine E-cadherin expression using immunohistochemistry [12]. The mean patient age was 52 years (range 18-71), and tumor size ranged in diameter from 0.1 cm to 5 cm at time of presentation, with a mean tumor size of 2 cm (Supplementary Table 2). To understand the molecular biology of E-cadherin expression, high grade invasive ductal carcinoma (n=883) cases in the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) cohort[21] were used to investigate copy number alterations and *CDH1* mRNA expression. The mean patient age was 59 (range 26-96), and mean tumor size at time of presentation was 3 cm (range from 1 cm to 18 cm). In the METABRIC-invasive ductal carcinoma series, DNA/RNA was isolated from fresh frozen samples and transcriptional profiling was obtained using the Affymetrix SNP 6.0 Illumina Total Prep RNA Amplification Kit and Illumina Human HT-12 v3 Expression Bead Chips (Ambion, Warrington, UK). Copy number alteration was considered at the gene level by segments and the Šidák correction [22], whereas gene expression data were pre-processed and normalized as described previously [21]. In this cohort, patients with estrogen positive tumor and /or lymph node negative at time of diagnosis did not receive adjuvant chemotherapy, whereas those with estrogen negative tumors and lymph node positive status received adjuvant treatment. Next generation RNA sequencing was conducted on an additional triple negative breast cancer cohort (n=106) to investigate E-cadherin reduced/loss expression in this subtype of breast cancer. The mean patient age was 48 (range 27-69) and tumors size ranged in diameter from 1 to 6 cm at time of presentation, with a mean tumor size of 2 cm (Supplementary Table 2).

Immunohistochemistry staining and scoring

Mouse monoclonal anti-E-cadherin antibody [Cl;4A2C7, Ref#180223, LOT 954621A, Invitrogen, UK] was used to assess protein expression on immunohistochemically-stained tissue sections after prior

validation of the antibody by Western blot using MDA-MB-231 and MDA-MB-157 breast cancer cell lysates (obtained from American Type Culture Collection, Rockville, MD, USA). Immunohistochemistry staining procedure was performed using Novocastra Novolink TM Polymer Detection Systems kit (Code: RE7280-K, Leica, Biosystems, UK) on 4 μm tissue microarray sections [20]. Sections were incubated for 24 hours with the anti-CDH1 antibody diluted to a concentration of 1:25. Scoring of membranous protein expression was performed using the modified histo-score [23]. We used the lower quartile from the modified histochemical score value (i.e. 85) to stratify the cohort into high and reduced/loss E-cadherin expression groups. Cases in the METABRIC cohort were stratified using a similar approach for total CDH1 mRNA expression. Copy number alteration and CDH1 mRNA expression were correlated with E-cadherin protein expression in the same cases where available (n=131).

RNA sequencing

RNA sequencing was performed on representative formalin-fixed paraffin-embedded blocks of triple negative breast cancers (n=106) which had also been assessed histopathologically for tumor burden. Invasive tumor cells were micro-dissected from unstained tissue sections where tissue burden was at least 50% of the tissue section area. Micro-dissected tissues were deparaffinized, rehydrated, and centrifuged to remove excess ethanol. RNA was extracted using the Omega Mag-Bind XP formalin fixed paraffin embedded RNA isolation kit (Omega, M2595-01) and Kingfisher Flex magnetic particle separator (ThermoFisher) as per manufacturer's instructions. RNA was measured with a Nanodrop 2000c spectrophotometer (Thermo Scientific). First strand cDNA synthesis was performed on approximately 100 ng RNA at 25°C for 10 min, 42°C for 15 min, and 70°C for 15 minutes using random hexamers and ProtoScript II Reverse Transcriptase (New England BioLabs, Ipswich, MA). Second strand synthesis and RNA sequencing libraries were prepared using the Illumina TruSeq RNA access library kit (Illumina, RS-301-2002) and sequenced on an Illumina HiSeq 2500 using PE75 run chemistry. The targeted read count was 60M total reads per sample. Sequencing was performed at the Emory Integrated Genomics Core Facility, Emory University, Atlanta, USA. Raw FastQ sequence reads files were quality assessed and adapter processed using the trim galore wrapper for Fastqc and Cutadapt with reads with phred scores >30 retained. The resultant quality trimmed reads were aligned

to the hg38 (GRCh38.83) build of the human genome using the STAR aligner. Transcript abundance quantification were performed using HTSEQ [34]. Only one sample per patient was included in downstream analyses by random selection. Differential gene expression was assessed using Robina implementation of Edge-R [24].

Pathway analysis

The online public available web-based gene set analysis tool, Webgestalt, (http://www.webgestalt.org/option.php) was used to identify differentially regulated canonical pathways. This pathway analysis was based on transcripts differentially expressed at the $p<0.05$ level and generated by Robina analysis, including only unbiased hits with significant z-scores based on network-adjusted p-values <0.05 using KEGG pathway database [25].

Statistical analysis

IBM SPSS 24.0 (Chicago, IL, USA) software was used for statistical analysis. The chi-squared test was used to assess the effect of copy number alteration on reduced/loss of *CDH1* mRNA expression. Furthermore, we evaluated copy number alteration of established tumor suppressor genes in cases that exhibited reduced/loss of *CDH1* mRNA expression and copy number loss via copy number alteration in the METABRIC cohort, to infer genetic instability as the likely driver of the reduced/loss of *CDH1* mRNA expression using chi-squared test. Mann-Whitney test was used to compare the expression of *CDH1* mRNA expression with expression of well-established transcription factors affecting E-cadherin expression[26]. Furthermore, we evaluated expression of a set of genes previously demonstrated to have 93% predictive accuracy in distinguishing invasive lobular carcinoma from invasive ductal carcinoma via the prediction analysis for microarrays test [27]. Expression of proteins related to DNA repair and proliferation were compared to expression of the E-cadherin protein using the Mann-Whitney test. Furthermore, the association of E-cadherin protein expression with that of transcription factors mRNA expression (assessed using next generation sequencing -HTSEQ values) was evaluated using the Mann-Whitney test. Two tailed *p*-value <0.05 was considered as statistically significant. RNA-Seq values were expressed as standard error of means in GRAPH PAD PRISM v.7 for data presentation.

RESULTS

Evaluation of E-cadherin protein expression in the high grade invasive ductal breast carcinoma cohort (n=813)

The specificity of E-cadherin antibody was validated by western blot that showed a single specific band at the expected molecular weight (~100 kDa). A total of 217/813 (27%) of high-grade invasive ductal carcinoma and 46/106 (43%) of triple negative breast cancer showed reduced/loss membrane expression of E-cadherin. Within the METABRIC cohort, reduced/loss *CDH1* mRNA expression was observed in 208/883 (23%) cases. Furthermore, triple negative breast cancer showed reduced/loss of *CDH1* mRNA expression in 90/235 (38%) cases. Reduced/loss *CDH1* mRNA expression cases were observed in 104 cases of the basal (37%), 18 cases of the HER2 enriched (11%), 40 cases of luminal A (27%), 29 cases of luminal B (12%), and 17 cases of the Normal like (29%) molecular subtypes (Supplementary Table 3 & Supplementary Figure 1). In the subset of cases that were included in the METABRIC dataset (n=131), there was a positive linear correlation between *CDH1* mRNA and the dichotomized E-cadherin protein expression (r=0.27, p=0.002).

Reduced/loss E-cadherin protein expression was associated with *GammaH2AX (*p<0.0001) and PTEN (p=0.003) protein expression (Table 1).

E-cadherin copy number alteration in ductal breast cancer

To investigate whether reduced/loss of E-cadherin expression in the invasive ductal carcinoma cases is due to copy number alteration, we examined copy number alteration and *CDH1* mRNA levels. We observed that 44/208 (21%) of cases showed significant association between loss of CDH1 copy number and reduced/loss *CDH1* mRNA expression (p=0.003) (Supplementary Table 4). Only 1 case with copy number loss did not show any association with the transcription factors investigated while the remaining cases showed upregulation of one or more transcription factors (Supplementary Table 5). Interestingly, 77% of tumors presenting with reduced/loss *CDH1* mRNA expression did not show *CDH1* copy number loss, indicating that other mechanisms are implicated. Subsequently, investigating the triple negative tumors, only 7/90 (8%) of cases showed copy number to be associated with reduced/loss *CDH1* mRNA expression. However, there was no statistical association

between copy number loss and reduced/loss of *CDH1* mRNA expression (p=0.10) (Supplementary Table 6). More importantly, amongst those cases, only 1 (copy number loss) case did not show any association with any transcription factors while the rest of the 6/90 (7%) cases (copy number loss) showed upregulation of one or more transcription factors (Supplementary Table 7). Moreover, 83/90 (92%) of triple negative tumors with reduced/loss *CDH1* mRNA expression showed neutral/amplified *CDH1* copy number expression.

In addition, reduced/loss *CDH1* mRNA expression in invasive ductal carcinoma showed copy number loss of multiple well-established breast cancer tumor suppressor genes located at different chromosome loci; *TP53, ATM, BRCA1, and BRCA2* (p<0.001) (Supplementary Table 8).

Expression of E-cadherin suppressor transcription factors

In cases with reduced/loss E-cadherin expression (n=208) from the METABRIC cohort, upregulated mRNA expression was observed with *ZEB2* (56%), *TWIST2* (54%), *NFKB1* (54%), *ZEB1* (53%), *TWIST1* (52%), *SLUG* (51%), *SNAIL* (50%), *GSK3BETA* (49%), *TGFB1* (47%), *LLGL2 (38%),* and *CRUMBS3* (34%). Only 4% of the cases were affected by 9 or more upregulated transcription factors (Supplementary Table 9 & Supplementary Figure 2). Upregulated expression of *TWIST2*, *ZEB2*, *NFKB1, LLGL2* and *CRUMBS3* were significantly associated with reduced/loss of *CDH1* mRNA expression (Table 2). In triple negative breast cancer with reduce/loss E-cadherin expression, upregulated mRNA expression was observed with *ZEB2* (63%), *SLUG* (62%), *TWIST2* (59%), *TWIST1* (57%), *ZEB1* (54%), *SNAIL* (52%), *TGFB1* (51%), *GSK3BETA* (50%), *NFKB1* (46%), *LLGL2 (24%),* and *CRUMBS3* (24%) (Supplementary Table 10 & Supplementary Figure 3). Only 3% of the cases harbored 9 or more upregulated transcription factors (Supplementary Table 7). Upregulated expression of *TWIST2*, *TWIST1*, *ZEB2*, *ZEB1*, *SLUG*, *LLGL2* and *CRUMBS3* were significantly associated with reduced/loss of *CDH1* mRNA expression (Table 3).

Proteins associated with E-cadherin expression in invasive triple negative ductal breast carcinoma

There was no significant statistical correlation between reduced/ loss of E-cadherin expression with transcription factors, DNA repair family, nor other markers such as ki67, ATM and PTEN on the protein level in triple negative breast cancer (Table 4 A, B & Supplementary Figure 4A)

E-cadherin loss and expression of genes deferentially expressed between invasive lobular carcinoma and invasive ductal carcinoma within the triple negative breast cancer cohort.

There was no significant association between reduced/ loss of E-cadherin expression in the high grade triple negative ductal cancer and those genes differentially expressed between invasive lobular and ductal carcinoma (*Cathepsin B, TPI1, SPRY1, SCYA14, TFAP2B, thrombospondin 4, Osteopontin, HLA-G, CHC1*) [27] (Table 5 & Supplementary Figure 4B).

Genomic study and pathway analysis

Next generation sequencing identified 2143 differentially expressed genes (Benjamin-Hochberg; $p < 0.05$, differentially expressed by >two-fold, false discovery rate <0.05). Triple negative invasive ductal carcinoma with reduced/loss E-cadherin expression (n=46) showed 849 significantly overexpressed and 1294 downregulated genes. It is noteworthy that dysregulation of genes regulating Wnt signaling pathway, the top predicted master regulator of E-cadherin expression, based on *p*-value, whose activity could explain protein expression differences were (*FZD2, GNG5, HLTF, WNT2, and CER1*), PIK3-AKT signaling pathway top predicted master regulator controlling E-cadherin expression were (*FGFR2*, *GNF5, GNGT1, IFNA17, and IGF1*) (Table 6). Importantly, key genes differentially expressed between invasive lobular carcinoma and invasive ductal tumors [27] did not show association with E-cadherin reduced/loss of expression in the invasive triple negative ductal carcinoma (Table 5).

DISCUSSION

Reduced/loss of E-cadherin expression is recognized as part of the main molecular events driving loss of cell-cell adhesion and thus facilitating cancer invasion and metastasis [28]. Some authors have suggested that E-cadherin can serve as a phenotypic marker to distinguish between invasive lobular carcinoma and non-invasive lobular tumors [27]. Mechanisms seeding reduced/loss of E-cadherin expression comprise *CDH1* gene mutation [10], truncating mutation [29], promoter hypermethylation [30] and transcriptional inactivation [31]. Reduced/loss of E-cadherin expression is observed in 84% of invasive lobular carcinomas [9]. Several studies have shown that approximately 38% of high-grade invasive ductal tumors show reduced/ loss of E-cadherin expression, and this phenomenon has been linked to aggressive tumor behavior. Interestingly, *CDH1* gene mutations were not identified in this subgroup [11-19-32].

One of the recognized mechanisms leading to reduced/loss of E-cadherin expression is loss of heterozygosity at chromosome 16q22.1, where the *CHD1* gene is located [33]. Studies investigating the mechanism underlying reduced/loss of E-cadherin protein expression in invasive lobular carcinoma cases uncovered loss of wild-type allele due to loss of heterozygosity at 16q22.1 occurring in more than 70% of cases [7-8]. Furthermore, *CDH1* gene mutation and promoter hypermethylation were observed in 20% and 56% of invasive lobular carcinomas, respectively [7]. Interestingly, co-occurrence of these mechanisms rarely occurs in invasive lobular tumors [34]. Remarkably, mutational inactivation of *CDH1* gene mostly coexists with loss of the wild type allele in invasive lobular carcinoma [35]. Since reduced/loss of E-cadherin expression in invasive lobular tumors is predominantly caused by loss of heterozygosity, it has been suggested that copy number loss of the *CDH1* gene can be utilized to discriminate between invasive ductal carcinoma and invasive lobular tumors when it is difficult to differentiate them based on histological evaluation [36]. Our investigation revealed that copy number loss occurred in only 21% of invasive ductal carcinomas displaying reduced/loss of E-cadherin expression. Therefore, other mechanisms must underlie the downregulation of E-cadherin in the majority of cases. Other mechanisms of E-cadherin reduced/loss of expression without copy number loss include DNA hypermethylation, a mechanism that may induce the *CDH1* reduced/loss of mRNA expression detected in 60% of metastatic invasive ductal carcinoma [37].

Loss of *CDH1* gene at 16q22.1 in invasive lobular carcinoma is one of the main genet events and is observed early in the process of carcinogenesis in lobular carcinomas. We hypothesized that reduced/loss of E-cadherin expression in a subset of invasive ductal tumors might be the result of genomic instability and occurs as a late event during the process of cancer progression. Our results demonstrate that loss of *CDH1* copy number is associated with copy number aberrations of multiple well-established breast cancer tumor suppressor genes located at different chromosomes; copy number loss of *ATM* (11q22.3), *PTEN (*10q23.31), *RB1* (13q14.2), *TP53* (17p13.1), *BRCA1* (17q21.31), and *BRCA2* (13q13.1) tumor suppressor genes. Moreover, DNA damage response pathways which are crucial for detecting DNA lesions and arresting the cell cycle until the DNA is repaired or inducing cell death if cells sustain irreparable DNA damage [38], play key roles in preventing genetic instability and tumorigenesis [39]. Investigation of correlations between reduced/loss of E-cadherin expression and expression of biomarkers related to DNA damage response pathways in breast cancer revealed negative correlation between reduced/loss of E-cadherin protein expression and GammaH2AX and PTEN expression suggesting that reduced/loss of E-cadherin expression is associated with impaired DNA damage response and likely, genomic instability. Taken together, these results support our hypothesis that reduced/loss of E-cadherin expression in invasive ductal carcinomas is associated with genomic instability.

Reduced/loss of E-cadherin expression can also be caused by overexpression of its associated transcription factors [40-41-42-26]. Our results showed a negative correlation between reduced/loss of *CDH1* mRNA expression and the mRNA expression of transcription factors known to suppress E-cadherin expression and cause disruption of cell-cell adhesion[43-44]; in fact, 76% of cases harboring reduced/loss of *CDH1* mRNA show upregulation of one or more of these transcriptional repressors.

Remarkably, other key players in epithelial-mesenchymal transition such as TGFBeta1, SNAIL and SLUG did not show any correlation with E-cadherin reduced/loss of mRNA expression. These observations suggest that reduced/loss of E-cadherin expression is not merely a surrogate for epithelial-mesenchymal transition but represents a readout of other pathways controlling E-cadherin expression at membranes level.

Of note, reduced/loss of E-cadherin protein expression occurs in up to 50% of triple negative invasive ductal carcinoma, which may contribute to increased lymph node metastasis, and poor patient outcomes [45]. We observed a negative correlation between reduced/loss of *CDH1* mRNA expression and the mRNA expression of multiple transcription factors known to suppress E-cadherin expression in our triple negative breast cancer cohort. On the contrary, when we investigated the same genotype within the cohort tested by next generation sequencing, none of these transcription factors showed statistically significant associations with E-cadherin expression. It is possible that different molecular mechanisms regulate E-cadherin expression, although we cannot exclude the possibility that our cohort is small to such associations.

More importantly, genes differentially expressed between invasive ductal and invasive lobular breast tumors as identified by Waldman et al [27] and could represent the effect of E-cadherin loss in lobular carcinoma compared to ductal tumors showed no statistically significant difference, when tested on mRNA level, in breast cancer cases showing reduced/loss of E-cadherin expression compared to these tumour with normal expression. This may indicate not only that more complex molecular mechanisms are responsible for E-cadherin reduced/loss of protein expression in these cases but also E-cadherin loss in ductal carcinoma does not produce the same effects in lobular tumors. This may also be supported by the lack of morphological features and metastatic behavior characteristic of lobular carcinomas in ductal tumors lacking E-cadherin expression.

In this study, differential gene expression using next generation sequencing investigating differences between cases with reduced/loss of E-cadherin expression and cases with normal/high expression showed dysregulation of genes regulating *PIK3-AKT* signaling pathway. Our analysis exposed a negative correlation between the genes regulating this pathway and reduced/loss of E-cadherin protein expression suggesting that overexpression of those indicators may promote signaling via the *PIK3-AKT* pathway and thus negatively regulate E-cadherin expression. Receptors such as insulin like growth factor receptor 1, can induce the activity of *Akt* pathway [46]. Our results are in agreement with reports indicating activation of *PIK3-AKT* represses E-cadherin expression and stimulates cell migration [47]. Nonetheless, dysregulation of genes regulating Wnt signaling pathway was also present

in our results. Mutation or deregulation of gene expression of the canonical Wnt pathway is implicated in cancer [48-49-50].

Our study limitation relates to comparing gene expression data obtained from microarrays, as used in the METABRIC cohort comprising different molecular subtypes of invasive ductal tumors, and the RNA sequencing dataset available for our triple negative breast cancers only. We have chosen triple negative breast cancer to study E-cadherin protein expression in invasive ductal carcinoma cases as up to 50% of this molecular subtype show reduced/loss of E-cadherin protein expression [45- 51]. On the contrary, studies have shown that reduced/loss of E-cadherin expression occurs in 23% and 27% of luminal and HER2 enriched subtypes, respectively [51]. RNA-Seq approaches cover multiple aspects of the transcriptome without any a *priori* knowledge, allowing to identify novel transcripts, splice junctions and noncoding RNAs [52]. We acknowledge that comparison between these two different approaches may or may not provide the same results due to intrinsic differences in assay design [53]. For instance, next generation sequencing may have different lower limits of detection or may encompass different genomic regions [52]. More importantly, invasive ductal carcinoma cases used in the METABRIC cohort comprise different molecular subtypes, while the RNA sequencing data was acquired for a triple negative breast cancer cohort, which also may play a role in our study. Therefore, further validation of our findings is warranted.

Conclusion: Reduced/loss E-cadherin expression in invasive ductal carcinoma is a complex biological phenomenon, which, according to the findings of this study, appears to be a part of the genomic instability process occurring late in the process of carcinogenesis rather than an initial neoplastic event and results in different effects to those produced in invasive lobular carcinomas. Using the high throughput next generation sequencing, we have unraveled potential novel regulators controlling different signaling pathways that regulate E-cadherin protein expression in invasive ductal carcinoma. These regulators warrant further investigation and validation using different platforms.

Acknowledgements

We thank the Nottingham Health Science Biobank and Breast Cancer Now Tissue Bank for the provision of tissue samples.

Conflict of interest

There were no conflicts of interest for any of the other authors.

References

1 Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394–424.

2 Hart IR, Saini A. Biology of tumour metastasis. Lancet (London, England) 1992;339:1453–1457.

3 Berx G, Staes K, van Hengel J, et al. Cloning and characterization of the human invasion suppressor gene E-cadherin (CDH1). Genomics 1995;26:281–289.

4 Paredes J, Figueiredo J, Albergaria A, et al. Epithelial E- and P-cadherins: Role and clinical significance in cancer. Biochim Biophys Acta - Rev Cancer 2012;1826:297–311.

5 Behrens J, Mareel MM, Van Roy FM, et al. Dissecting tumor cell invasion: epithelial cells acquire invasive properties after the loss of uvomorulin-mediated cell-cell adhesion. J Cell Biol 1989;108:2435–2447.

6 Frixen UH, Behrens J, Sachs M, et al. E-cadherin-mediated cell-cell adhesion prevents invasiveness of human carcinoma cells. J Cell Biol 1991;113:173–185.

7 Sarrió D, Moreno-Bueno G, Hardisson D, et al. Epigenetic and genetic alterations of APC and CDH1 genes in lobular breast cancer: relationships with abnormal E-cadherin and catenin expression and microsatellite instability. Int J Cancer 2003;106:208–215.

8 Huiping C, Sigurgeirsdottir JR, Jonasson JG, et al. Chromosome alterations and E-cadherin gene mutations in human lobular breast cancer. Br J Cancer 1999;81:1103–1110.

9 de Leeuw WJF, Berx G, Vos CBJ, et al. Simultaneous loss of E-cadherin and catenins in invasive lobular breast cancer and lobular carcinomain situ. J Pathol 1997;183:404–411.

10 Sgeirsson KSA, Joâ Nasson JG, Tryggvadoâ Ttir L, et al. Altered expression of E-cadherin in breast cancer: patterns, mechanisms and clinical signi®cance. [cited 12 July 2018]. Available from: https://skemman.is/bitstream/1946/21610/1/Asgeirsson2000.pdf.

11 Acs G, Lawton TJ, Rebbeck TR, et al. Differential Expression of E-Cadherin in Lobular and Ductal Neoplasms of the Breast and Its Biologic and Diagnostic Implications. Am J Clin Pathol 2001;115:85–98.

12 Rakha EA, Teoh TK, Lee AHS, et al. Further evidence that E-cadherin is not a tumour suppressor gene in invasive ductal carcinoma of the breast: an immunohistochemical study. Histopathology 2013;62:695–701.

13 Jeschke U, Mylonas I, Kuhn C, et al. Expression of E-cadherin in human ductal breast cancer carcinoma in situ, invasive carcinomas, their lymph node metastases, their distant metastases, carcinomas with recurrence and in recurrence. Anticancer Res;27:1969–1974.

14 da Silva BB, dos Santos AR, Pires CG, et al. E-cadherin expression in estrogen receptor-positive and negative breast carcinomas of postmenopausal women. Eur J Gynaecol Oncol 2010;31:90–93.

15 Siitonen SM, Kononen JT, Helin HJ, et al. Reduced E-cadherin expression is associated with invasiveness and unfavorable prognosis in breast cancer. Am J Clin Pathol 1996;105:394–402.

16 Mahler-Araujo B, Savage K, Parry S, et al. Reduction of E-cadherin expression is associated with non-lobular breast carcinomas of basal-like and triple negative phenotype. J Clin Pathol 2008;61:615–620.

17 Lipponen P, Saarelainen E, Ji H, et al. Expression of E-cadherin (E-CD) as related to other prognostic factors and survival in breast cancer. J Pathol 1994;174:101–109.

18 McCart Reed AE, Kutasovic JR, Lakhani SR, et al. Invasive lobular carcinoma of the breast: morphology, biomarkers and 'omics. Breast Cancer Res 2015;17:12.

19 Gillett CE, Miles DW, Ryder K, et al. Retention of the expression of E-cadherin and catenins is associated with shorter survival in grade III ductal carcinoma of the breast. J Pathol 2001;193:433–441.

20 Rakha EA, Abd El Rehim D, Pinder SE, et al. E-cadherin expression in invasive non-lobular

carcinoma of the breast and its prognostic significance. Histopathology 2005;46:685–693.

21 Curtis C, Shah SP, Chin S-F, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012;486:346–352.

22 Šidák Z. Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. J Am Stat Assoc 1967;62:626–633.

23 McCarty KS, Miller LS, Cox EB, et al. Estrogen receptor analyses. Correlation of biochemical and immunohistochemical methods using monoclonal antireceptor antibodies. Arch Pathol Lab Med 1985;109:716–721.

24 Lohse M, Bolger AM, Nagel A, et al. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Res 2012;40:W622–W627.

25 Wang J, Vasaikar S, Shi Z, et al. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. Nucleic Acids Res 2017;45:W130–W137.

26 Baranwal S, Alahari SK. Molecular mechanisms controlling E-cadherin expression in breast cancer. Biochem Biophys Res Commun 2009;384:6–11.

27 Korkola JE, DeVries S, Fridlyand J, et al. Differentiation of lobular versus ductal breast carcinomas by expression microarray analysis. Cancer Res 2003;63:7167–7175.

28 Debies MT, Welch DR. Genetic basis of human breast cancer metastasis. J Mammary Gland Biol Neoplasia 2001;6:441–451.

29 Strathdee G. Epigenetic versus genetic alterations in the inactivation of E-cadherin. Semin Cancer Biol 2002;12:373–379.

30 Graff JR, Herman JG, Lapidus RG, et al. E-cadherin expression is silenced by DNA hypermethylation in human breast and prostate carcinomas. Cancer Res 1995;55:5195–5199.

31 Batlle E, Sancho E, Francí C, et al. The transcription factor Snail is a repressor of E-cadherin gene expression in epithelial tumour cells. Nat Cell Biol 2000;2:84–89.

32 Oka H, Shiozaki H, Kobayashi K, et al. Expression of E-cadherin cell adhesion molecules in

human breast cancer tissues and its relationship to metastasis. Cancer Res 1993;53:1696–1701.

33 Cleton-Jansen AM, Callen DF, Seshadri R, et al. Loss of heterozygosity mapping at chromosome arm 16q in 712 breast tumors reveals factors that influence delineation of candidate regions. Cancer Res 2001;61:1171–1177.

34 Cheng C-W, Wu P-E, Yu J-C, et al. Mechanisms of inactivation of E-cadherin in breast carcinoma: modification of the two-hit hypothesis of tumor suppressor gene. Oncogene 2001;20:3814–3823.

35 Berx G, Cleton-Jansen AM, Strumane K, et al. E-cadherin is inactivated in a majority of invasive human lobular breast cancers by truncation mutations throughout its extracellular domain. Oncogene 1996;13:1919–1925.

36 Li J, Wang K, Li S, et al. DNA copy number aberrations in breast cancer by array comparative genomic hybridization. Genomics Proteomics Bioinformatics 2009;7:13–24.

37 Caldeira JRF, Prando ÉC, Quevedo FC, et al. CDH1promoter hypermethylation and E-cadherin protein expression in infiltrating breast cancer. BMC Cancer 2006;6:48.

38 Sengodan SK, K.H. S, Nadhan R, et al. Regulation of epithelial to mesenchymal transition by BRCA1 in breast cancer. Crit Rev Oncol Hematol 2018;123:74–82.

39 Yasmeen A, Liu W, Dekhil H, et al. BRCA1 mutations contribute to cell motility and invasion by affecting its main regulators. Cell Cycle 2008;7:3781–3783.

40 Becker KF, Atkinson MJ, Reich U, et al. E-cadherin gene mutations provide clues to diffuse type gastric carcinomas. Cancer Res 1994;54:3845–3852.

41 Berx G, Cleton-Jansen AM, Nollet F, et al. E-cadherin is a tumour/invasion suppressor gene mutated in human lobular breast cancers. EMBO J 1995;14:6107–6115.

42 Cleton-Jansen A-M. E-cadherin and loss of heterozygosity at chromosome 16 in breast carcinogenesis: different genetic pathways in ductal and lobular breast cancer? Breast Cancer Res 2002;4:5–8.

43 Spaderna S, Schmalhofer O, Wahlbuhl M, et al. The Transcriptional Repressor ZEB1 Promotes Metastasis and Loss of Cell Polarity in Cancer. Cancer Res 2008;68:537–544.

44 Aigner K, Dampier B, Descovich L, et al. The transcription factor ZEB1 (δEF1) promotes tumour cell dedifferentiation by repressing master regulators of epithelial polarity. Oncogene 2007;26:6979–6988.

45 Kashiwagi S, Yashiro M, Takashima T, et al. Significance of E-cadherin expression in triple-negative breast cancer. Br J Cancer 2010;103:249–255.

46 Foulstone E, Prince S, Zaccheo O, et al. Insulin-like growth factor ligands, receptors, and binding proteins in cancer. J Pathol 2005;205:145–153.

47 Irie HY, Pearline R V, Grueneberg D, et al. Distinct roles of Akt1 and Akt2 in regulating cell migration and epithelial-mesenchymal transition. J Cell Biol 2005;171:1023–1034.

48 Bienz M, Clevers H. Linking colorectal cancer to Wnt signaling. Cell 2000;103:311–320.

49 Prasad CP, Gupta SD, Rath G, et al. Wnt signaling pathway in invasive ductal carcinoma of the breast: relationship between beta-catenin, dishevelled and cyclin D1 expression. Oncology 2007;73:112–117.

50 Prasad CP, Mirza S, Sharma G, et al. Epigenetic alterations of CDH1 and APC genes: Relationship with activation of Wnt/β-catenin Pathway in invasive ductal carcinoma of breast. Life Sci 2008;83:318–325.

51 Liu J-B, Feng C-Y, Deng M, et al. E-cadherin expression phenotypes associated with molecular subtypes in invasive non-lobular breast cancer: evidence from a retrospective study and meta-analysis. World J Surg Oncol 2017;15:139.

52 Hurd PJ, Nelson CJ. Advantages of next-generation sequencing versus the microarray in epigenetic research. Briefings Funct Genomics Proteomics 2009;8:174–183.

53 Wolff A, Bayerlová M, Gaedcke J, et al. A comparative study of RNA-Seq and microarray data analysis on the two examples of rectal-cancer patients and Burkitt Lymphoma cells. PLoS One 2018;13:e0197162.

Table 1: Correlation between level of proteins associated with altered E-cadherin expression in high grade invasive breast cancer cohort (n=813)

Identifier		Protein of interest	E-cadherin expression frequency		Mean RANK		Z-score	*p*-value
			Low (%)	High (%)	Low	High		
Protein Level	Transcription Factor	TGFBeta1	143 (27)	371 (73)	257	258	-0.04	0.96
		TWIST2	130 (25)	376 (75)	259	251	-0.60	0.540
		ZEB	126 (26)	352 (74)	240	239	-0.19	0.840
	DNA repair family	BRCA1	178 (26)	491 (74)	336	335	-0.05	0.950
		BRCA2	132 (26)	375 (74)	248	256	-1.19	0.230
		GammaH2ax	120 (23)	387 (77)	221	264	-3.48	**<0.001**
		RAD51	110 (25)	320 (75)	199	221	-1.96	**0.050**
	Proliferation and other markers family	Ki67	174 (25)	501 (75)	329	341	-1.14	0.250
		PTEN	104 (28)	260 (72)	164	189	-2.93	**0.003**
		ATM	143 (28)	358 (72)	245	253	-0.70	0.480
		TP53	214 (26)	590 (74)	408	400	-0.47	0.630

Significant *p* values are in bold

Table 2: Correlation between mRNA levels of the genes associated with altered E-cadherin expression in breast cancer in the METABRIC cohort

Identifier	Genome of interest	E-cadherin expression frequency		Mean RANK		Z- score	*p*-value
		Low (%)	High (%)	Low	High		
Transcription Factor mRNA Level	*TGFB1*	208 (23)	675 (77)	443	442	-0.07	0.940
	TWIST2			481	430	-2.51	**0.012**
	TWIST1			466	434	-1.58	0.110
	ZEB2			489	427	-3.05	**0.002**
	ZEB1			445	441	-0.21	0.830
	SLUG			463	436	-1.33	0.180
	SNAIL			457	437	-0.94	0.340
	NFKB1			489	427	-3.06	**0.002**
	LLGL2			349	470	-6.00	**<0.001**
	GSK3B			439	443	-0.150	0.880
	CRUMBS			335	475	-6.93	**<0.001**

Significant *p* values are in bold

Table 3: Correlation between mRNA level of the genes associated with E-cadherin expression in triple negative high grade invasive ductal carcinoma in the METABRIC cohort

Identifier	Gene of interest	E-cadherin expression frequency		Mean RANK		Z-score	*p*-value
		Low (%)	H high (%)	Low	High		
Transcription Factor mRNA Level	*TGFB1*	90 (38)	145 (62)	114	120	- 0.74	0.460
	TWIST2			130	110	- 2.16	**0.030**
	TWIST1			132	109	- 2.51	**0.010**
	ZEB2			138	106	- 3.49	**<0.001**
	ZEB1			134	108	- 2.825	**0.005**
	SLUG			129	111	- 2.048	**0.040**
	SNAIL			113	121	- 0.94	0.340
	NFKB1			115	119	- 0.47	0.630
	LLGL2			89	135	- 5.00	**<0.001**
	GSK3B			108	124	-1.71	0.080
	CRUMBS			91	135	- 4.80	**<0.001**

***Significant *p* values are in bold**

***Dichotomization of mRNA level was based on median of the total expression of each gene**

Table 4A: Correlation between level of proteins known to control E-cadherin expression using triple negative invasive breast carcinoma cohort (n=106)

Identifier		Protein of interest	E-cadherin expression frequency		Mean RANK		Z-score	*p*-value
			Low (%)	High (%)	Low	High		
Protein Level	Transcription Factor	TGFBeta1	27 (47)	30 (53)	27	31	-0.87	0.380
		TWIST2	26 (48)	28 (52)	27	31	-0.23	0.810
		ZEB	25 (49)	26 (51)	27	25	-0.79	0.430
	DNA repair family	BRCA1	33 (44)	41(56)	37	38	-0.26	0.790
		BRCA2	26 (45)	31 (55)	26	32	-1.90	0.060
		GammaH2ax	27 (43)	35 (57)	28	34	-1.47	0.140
		RAD51	23 (51)	22 (49)	21	24	-0.85	0.390
	Proliferation and other markers family	Ki67	35 (47)	40 (53)	42	34	-1.61	0.100
		PTEN	21 (47)	19 (53)	22	19	-0.96	0.390
		ATM	28 (49)	29 (51)	29	29	-0.18	0.850
		TP53	39 (41)	45 (59)	43	42	-0.32	0.750

Table *4B:* Correlation between level mRNA expression of other genes known to control E-cadherin expression using triple negative invasive breast carcinoma cohort (n=106)

Identifier		Gene of interest	E-cadherin expression frequency		Mean RANK		*Z*- score	*p*-value
			Low (%)	High (%)	Low	High		
mRNA Level	Transcription Factor family	*TGFB1*	46 (43)	60 (57)	56	51	-0.88	0.370
		TWIST2			56	51	-1.28	0.200
		TWIST1			53	53	-0.01	0.990
		ZEB2			56	52	-0.74	0.450
		ZEB1			57	51	-0.94	0.340
		SLUG			51	55	-0.69	0.480
		SNAIL			55	53	-0.43	0.660
		NFKB1			57	51	-0.95	0.340
		LLGL2			58	50	-1.33	0.180
		GSK3B			54	53	-0.29	0.760
		CRUMBS			53	54	-0.07	0.930

Table 5: Genes differentially expressed between lobular versus ductal breast carcinomas in triple negative breast cancer cohort (n=106)

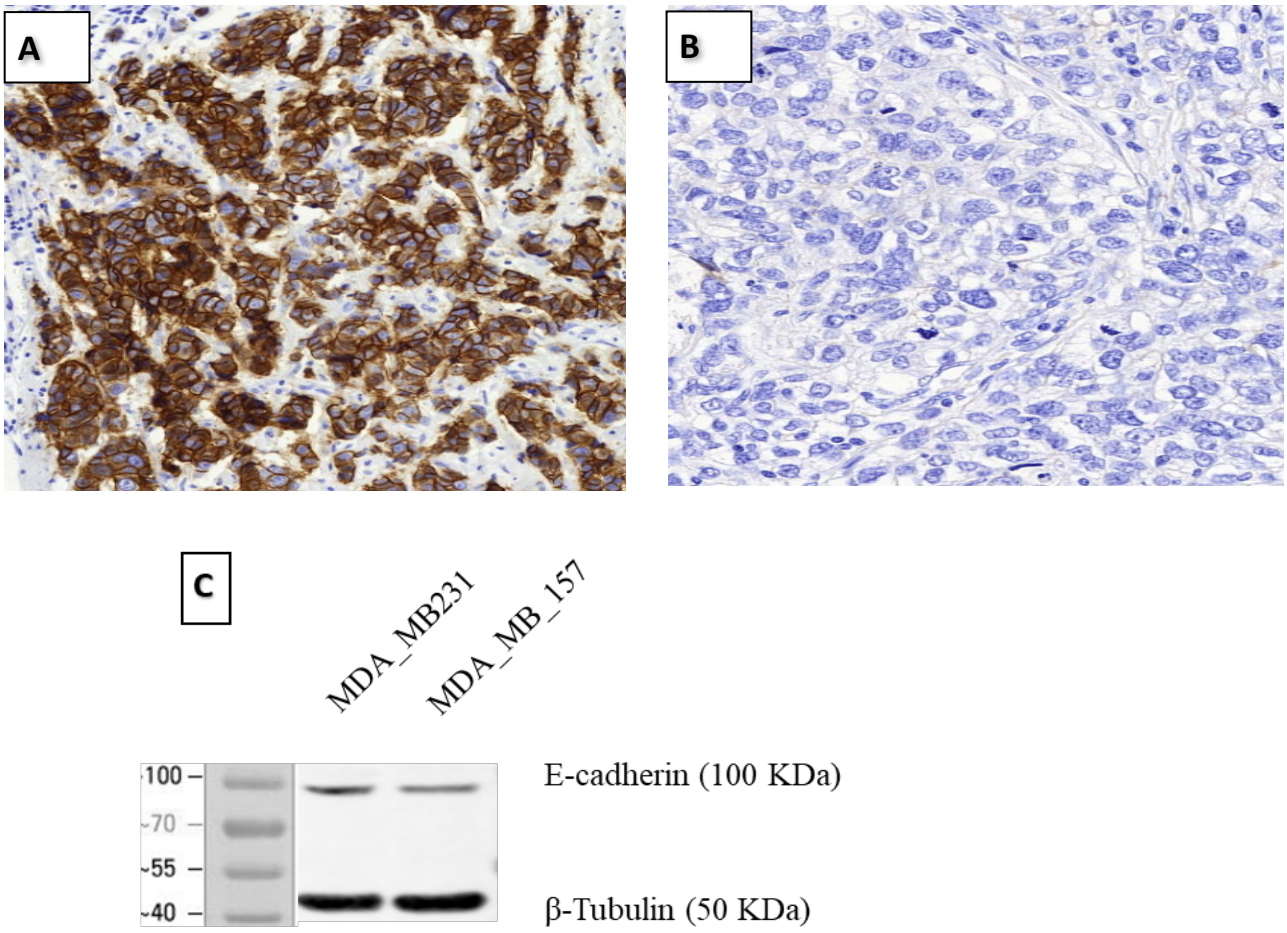
Identifier	Gene of Interest	E-cadherin expression frequency		Mean RANK		*Z*-score	*p*-value
		Low (%)	High (%)	Low	High		
mRNA Level	*Cathepsin B*	46 (43)	60 (57)	51	55	-0.57	0.560
	TPI1			56	51	-0.87	0.380
	SPRY1			54	53	-0.39	0.690
	SCYA14			54	53	-0.20	0.830
	TFAP2B			52	55	-0.46	0.640
	Thrombospondin 4			53	54	-0.09	0.920
	Osteopontin			54	53	-0.26	0.790
	HLA-G			52	55	-0.51	0.600
	CHC1			53	54	-0.13	0.890

Table 6: Pathway analysis results using Web gestalt to identify differentially regulated canonical pathways in the triple negative breast cancer cohort

Master regulator	Gene symbol	Gene name	score	FDR score from GSEA	*p* value from GSEA
PI3K-Akt signalling pathway	*FGFR2*	Fibroblast growth factor receptor 2	0.04	0.0069	**<0.001**
	GNG5	G protein subunit gamma 5	0.03		
	GNGT1	G protein subunit gamma transducin1	0.02		
	IFNA17	Interferon alpha 17	0.03		
	IGF1	Insulin like growth factor	0.03		
Wnt signalling pathway	*FZD2*	frizzled class receptor 2	0.01	0.0024	**<0.001**
	GNG5	G protein subunit gamma 5	0.03		
	HLTF	helicase like transcription factor	0.01		
	WNT2	Wnt family member 2	0.03		
	CER1	Cerberus 1, DAN family BMP antagonist	0.01		

Significant P values are in bold

Supplementary Figure 1: Western blotting & Immunohistochemistry

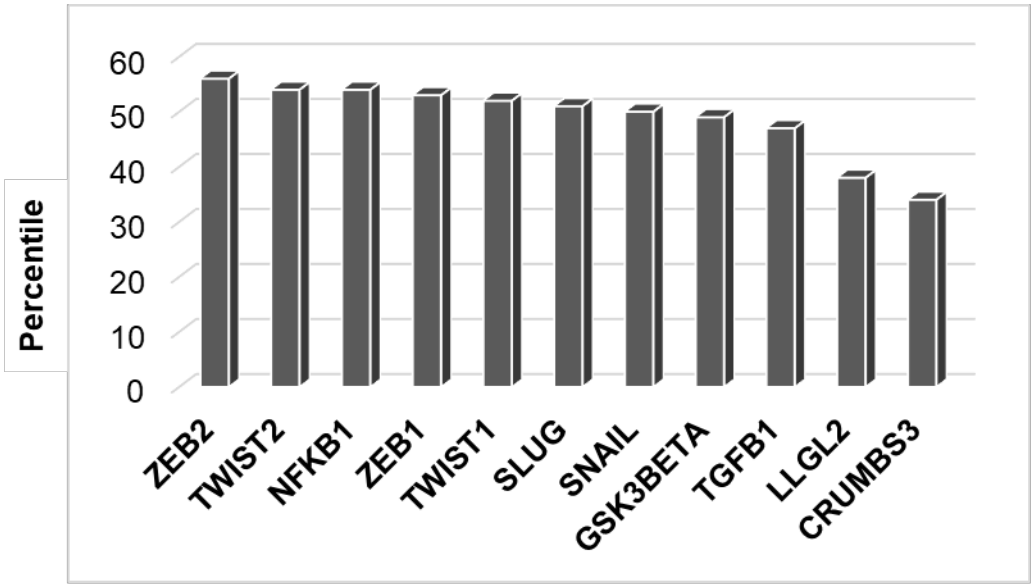


A. Immunohistochemistry for E-cadherin expression in Invasive ductal carcinoma of the breast showing strong membranous staining.

B. Immunohistochemistry for E-cadherin expression in Invasive ductal carcinoma of the breast showing complete loss of membranous staining.

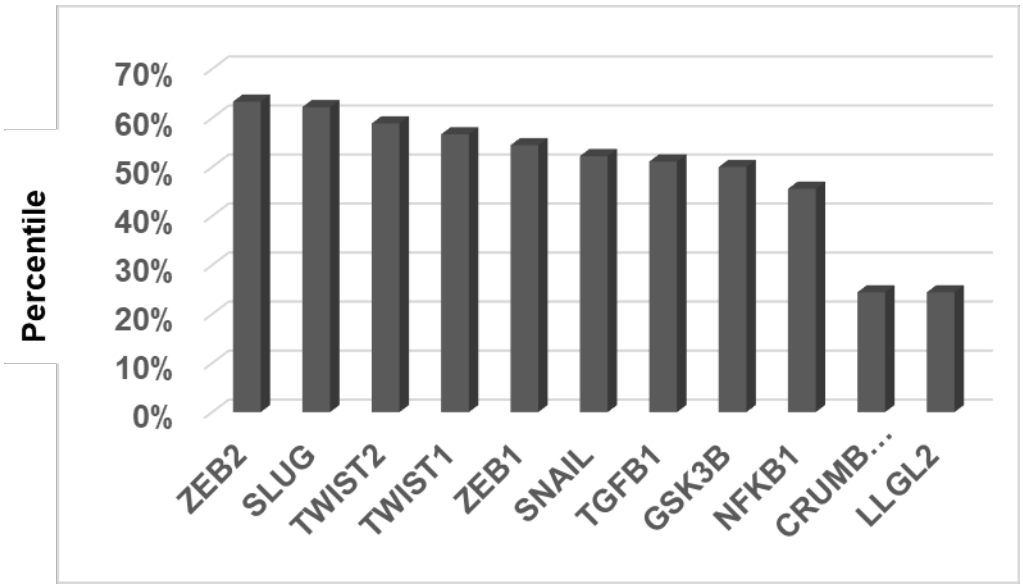
C. Specificity of the E-cadherin antibody (clone 4A2C7, Ref#180223, LOT 954621A, Invitrogen, UK) was confirmed by western blotting using MDA-MB231 and MDA-MB157 cell lines (The American Type Culture Collection; Rockville, MD, USA), which indicated a specific band at approximately 100 kDa.

Supplementary Figure2: Frequency of genes associated with reduced E-cadherin expression in invasive ductal carcinoma in the METABRIC cohort



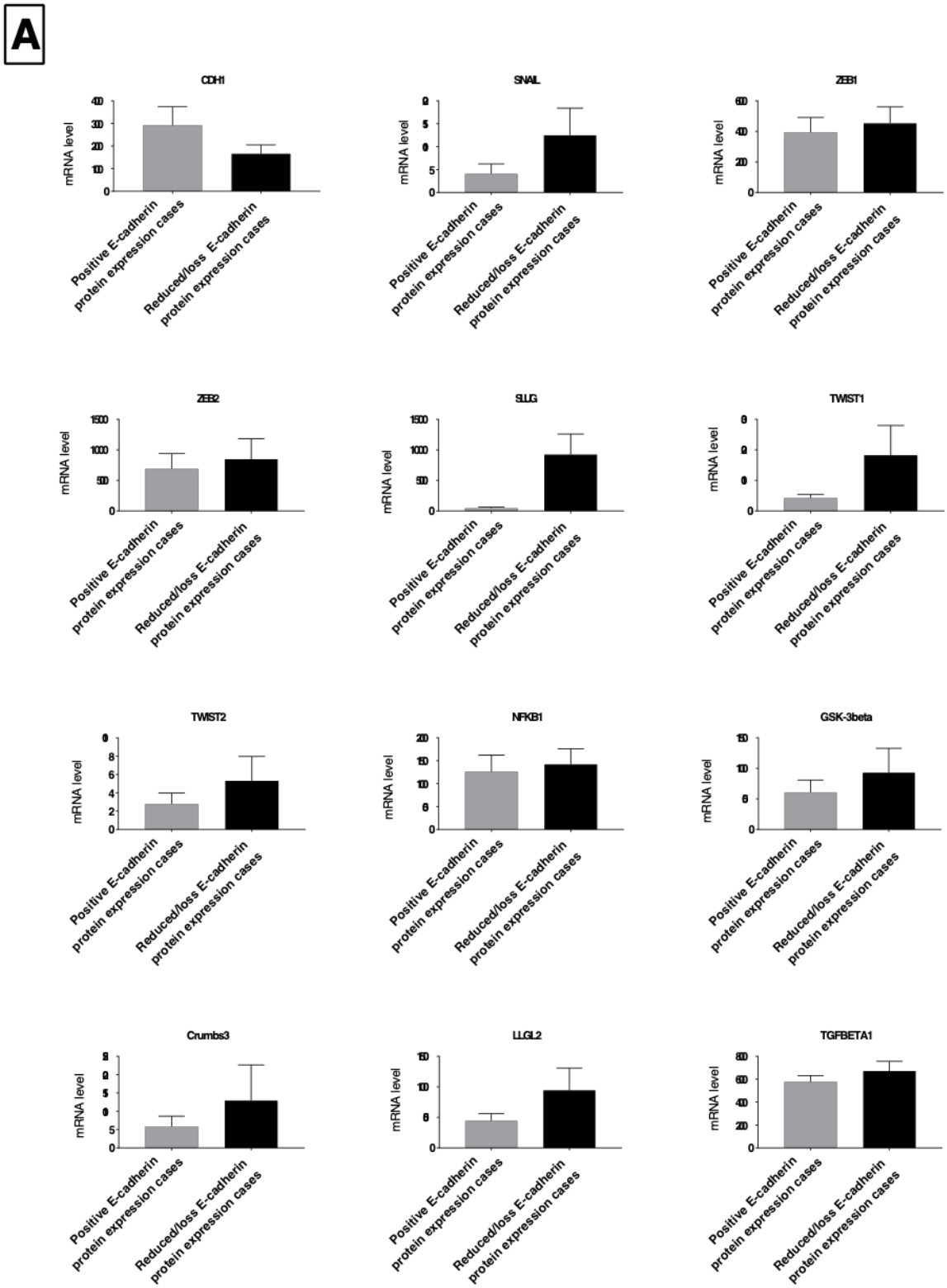
* Dichotomisation of *mRNA level was based on Median of the total expression of each gene*.

Supplementary Figure 3: Frequency of genes associated with reduced E-cadherin expression in the triple negative invasive ductal carcinoma in the METABRIC cohort



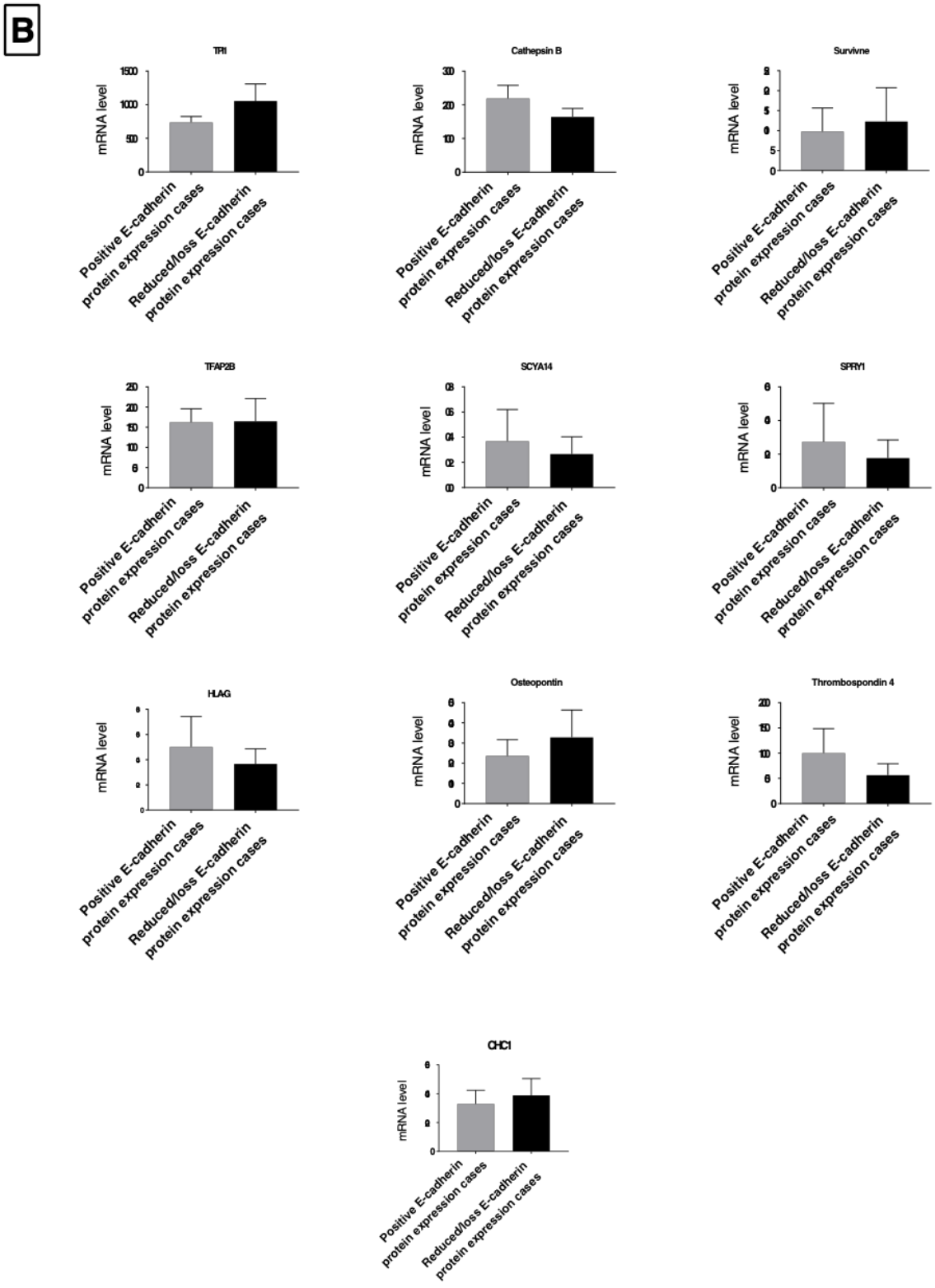
Dichotomization of mRNA level was based on median of the total expression of each gene

Supplementary Figure 4A: Frequency of Transcription factors suppressing E-cadherin expression in Cases showing reduced/loss CDH1 mRNA expression in the triple negative breast cancer cohort



- **Dichotomization of mRNA was based on frequency of E-cadherin cases**

Supplementary Figure 4B: Differentially expressed genes between lobular versus ductal breast carcinomas in the triple negative breast cancer cohort



- **Dichotomization of mRNA was based on frequency of E-cadherin cases**

Cohort	Immunohistochemistry	Microarray data	Next Generation sequencing
High grade invasive ductal carcinoma (n=813)	Yes	No	Yes*
High grade invasive ductal carcinoma within METABRIC series (n=883)	No	Yes	No

Supplementary Table 1: Summary of the study cohorts

METABRIC: Molecular Taxonomy of Breast Cancer International Consortium
*Next generation sequencing was carried out on a sub cohort of triple negative invasive breast cancer (n=106)

Supplementary Table 2: Clinicopathological data of patient's cohorts used in the study

Clinicopathological Parameter	Whole Invasive ductal carcinoma cohort (n=813) N (%)	High grade Invasive ductal carcinoma in METABRIC cohort (n=883) N (%)	Triple negative -Invasive ductal carcinoma cohort (n=106) N (%)
Age (years)			
< 50	326 (40)	266 (30)	59 (56)
≥ 50	487 (60)	617 (70)	47 (44)
Tumour size (cm)			
≤ 2cm	319 (39)	325 (37)	29 (27)
>2cm	494 (61)	551 (63)	77 (73)
Stage			
1	455 (56)	186 (27)	64 (60)
2	258 (32)	420 (62)	29 (27)
3	100 (12)	73 (11)	13 (13)
Estrogen receptor status			
Positive	452 (56)	372 (42)	0 (0)
Negative	361 (44)	511 (58)	106 (100)
Nottingham prognostic index			
Good prognosis	1 (0.1)	0 (0)	2 (2)
Moderate prognosis	567 (69.9)	705 (80)	72 (68)
Poor prognosis	245 (30)	178 (20)	32 (30)

Supplementary Table 3: Frequency of CDH1 mRNA level cases in invasive ductal carcinoma of the breast (all METABRIC cohort) based on PAM50 classification

PAM50 Classification	*Cases with reduced/ loss CDH1* mRNA *expression*	Percentage of *cases with reduced/ loss CDH1* mRNA *expression*	*Cases with normal/ above threshold CDH1* mRNA *Expression*	Percentage of c*ases with normal/above threshold CDH1* mRNA *Expression*	**TOTAL**
Basal Type	104	37%	174	63%	278
HER2 Enriched Type	18	11%	141	89%	159
Luminal A Type	40	27%	110	73%	150
Luminal B Type	29	12%	208	88%	237
Normal Like Type	17	29%	42	71%	59
Total	208		675		883

Dichotomization of CDH1 mRNA level was based on the lower quartile of the total expression of the gene.

Supplementary Table 4: *CDH1* copy number alteration (CNA) in invasive ductal carcinoma of the breast (all METABRIC cohort)

CNA region for *CDH1*	CNA	*Cases with reduced/loss CDH1* mRNA *expression*	*Cases with normal/above threshold CDH1* mRNA *Expression*	Total	*p*-value
Chromosome 16q22.1	Loss (%)	44 (35%)	81 (65%)	125 (100%)	**0.003**
	Gain (%)	4 (15%)	23 (85%)	27 (100%)	
	Neutral (%)	160 (22%)	571 (78%)	731 (100%)	
	Total (%)	208 (23%)	675 (77%)	883 (100%)	

Significant *p* values are in bold

Supplementary Table 5: Frequency of genes associated with reduced E-cadherin expression in invasive ductal carcinoma in the METABRIC cohort

Frequency of Cumulative Number of Upregulation of Transcription Factor suppressor of *CDH1* mRNA expression in cases with reduced/loss of *CDH1* mRNA expression	Frequency	Percentage
11	1	1%
10	1	1%
9	5	3%
8	19	12%
7	35	21%
6	42	26%
5	38	23%
4	32	20%
3	21	13%
2	8	5%
1	5	3%

Dichotomization of mRNA level was based on median of the total expression of each gene.

Supplementary Table 6: *CDH1* copy number alteration (CNA) in the high-grade triple negative invasive ductal carcinoma in the METABRIC cohort

CNA region for *CDH1*	CNA	*Cases with reduced/loss CDH1* mRNA *expression*	*Cases with normal/above threshold CDH1* mRNA *Expression*	Total	Asymptotic Significance (2-sided) P-value
Chromosome 16q22.1	Loss (%)	7 (70%)	3 (30%)	10 (100%)	0.10
	Gain (%)	4 (36%)	7 (64%)	11 (100%)	
	Neutral (%)	79 (37%)	135 (63%)	214 (100%)	
	Total (%)	90 (38%)	145 (62%)	235 (100%)	

Supplementary Table 7: Frequency of genes associated with reduced E-cadherin expression in invasive ductal carcinoma in triple negative breast cancer within the METABRIC cohort

Frequency of cumulative number of upregulation of transcription factors' suppressor of *CDH1* mRNA expression in cases with reduced/loss of *CDH1* mRNA expression	Frequency of cases	Percentage
10	1	1%
9	2	2%
8	10	11%
7	20	24%
6	15	17%
5	14	16%
4	11	12%
3	9	10%
2	3	3%
1	4	4%

Dichotomization of mRNA level was based on median of the total expression of each gene.

Supplementary Table 8: Copy number alteration (CNA) within other chromosomes in invasive ductal carcinomas with reduced/loss CDH1 mRNA expression in METABRIC cohort

Gene	*Cases with reduced/loss CDH1* mRNA *expression*			*p*-value
	Loss	Gain	Neutral	
ATM	26 (12%)	2 (1%)	180 (87%)	**<0.001**
PTEN	17 (8%)	8 (4%)	178 (88%)	0.53
RB1	36 (17%)	24 (12%)	145 (71%)	0.46
TP53	25 (12%)	0 (0%)	183 (88%)	**<0.001**
BRCA1	13 (6%)	8 (4%)	187 (90%)	**<0.001**
BRCA2	12 (6%)	15 (7%)	180 (87%)	**<0.001**

Significant *p* values are in bold

Supplementary Table 9: Genes associated with reduced E-cadherin expression in invasive ductal carcinoma in METABRIC cohort

Upregulation of transcription factors' suppressor of *CDH1* mRNA expression in cases with reduced/ loss of *CDH1* mRNA expression	Percentage
ZEB2	56%
TWIST2	54%
NFKB1	54%
ZEB1	53%
TWIST1	52%
SLUG	51%
SNAIL	50%
GSK3BETA	49%
TGFB1	47%
LLGL2	38%
CRUMBS3	34%

Dichotomization of mRNA level was based on median of the total expression of each gene.

Supplementary Table 10: Genes associated with reduced E-cadherin expression in the triple negative invasive ductal carcinoma in the METABRIC cohort

Upregulation of transcription factor suppressor of *CDH1* mRNA expression in cases with reduced/ loss of *CDH1* mRNA expression	Percentage
ZEB2	63%
SLUG	62%
TWIST2	59%
TWIST1	57%
ZEB1	54%
SNAIL	52%
TGFB1	51%
GSK3BETA	50%
NFKB1	46%
LLGL2	24%
CRUMBS3	24%

Dichotomization of mRNA level was based on median of the total expression of each gene

Supplementary Table 11: Genomic study and pathway analysis based on data generated from next generation RNA sequencing in triple negative breast cancer cohort

Identifier	Log Fold Change	False Discovery Rate	*p*-value
FZD2	3.60	0.01	**<0.001**
GNG5	-2.73	0.03	**<0.001**
HLTF	3.39	0.01	**<0.001**
WNT2	-2.47	0.03	**<0.001**
CER1	4.17	0.01	**<0.001**
FGFR2	-3.26	0.04	**<0.001**
GNGT1	-4.49	0.02	**<0.001**
IFNA17	-6.90	0.03	**<0.001**
IGF1	-2.99	0.03	**<0.001**

- **Negative (-) correlation with CDH1 genomic and protein expression (downregulated when CDH1 is low/negative cases)**
- **Positive (+) correlation with CDH1 genomic and protein expression (upregulated when CDH1 is low/negative cases)**
- **Significant *p* values are in bold**