

# Further Evidence to Support Bimodality of Oestrogen Receptor Expression in Breast Cancer

Abir A. Muftah<sup>1,2</sup>, Mohammed Aleskandarany<sup>1</sup>, Sultan N. Sonbul<sup>1</sup>, Christopher C. Nolan<sup>1</sup>, Maria Diez Rodriguez<sup>1</sup>, Carlos Caldas<sup>3</sup>, Ian O. Ellis<sup>1</sup>, Andrew R. Green<sup>1</sup>, Emad A. Rakha<sup>1</sup>.

<sup>1</sup> Division of Cancer and Stem Cells, School of Medicine, The University of Nottingham and Nottingham University Hospitals NHS Trust, Nottingham City Hospital, Nottingham, UK, NG5 1PB.

<sup>2</sup> Department of Pathology, Faculty of Medicine, University of Benghazi, Libya.

<sup>3</sup>Cancer Research UK Cambridge Institute, Li Ka Shing Centre, University of Cambridge, Robinson Way, Cambridge CB2 0RE, UK

## Correspondence:

Emad Rakha, FRCPath

Professor of Breast pathology. Division of Cancer and Stem Cells, School of Medicine, The University of Nottingham and Nottingham University Hospitals NHS Trust, Nottingham City Hospital, Nottingham, UK, NG5 1PB. Email: emad.rakha@nottingham.ac.uk

**Keyword:** breast cancer; oestrogen receptor; expression pattern; bimodality; ESR,

**Running title:** ER expression pattern in breast cancer

## ABSTRACT

**Background and Aim:** Although oestrogen receptor (ER)-negative breast cancers (BC) do not respond to hormone therapy, the response of ER-positive BC is reported to be variable which may suggest a dose dependent effect. This study aimed to assess the pattern of ER-expression in BC at the protein (IHC) and transcriptome (microarray-based gene expression) levels. **Materials and Methods:** ER IHC expression was assessed in a large series of BC including 3649 core biopsies and 1892 cases prepared as TMA stained using specific antibodies. *ESR1* mRNA expression was assessed in the METABRIC study (1980 cases) using Linear Models for Microarray Data (LIMMA) and results were compared with protein levels. **Results:** IHC data confirmed bimodality of ER expression with 92.2% and 89.2% of the cases showed completely negative (<1%) or highly positive ( $\geq 70\%$ ) expression on the cores and TMA respectively. Weak positive cases (1-10%) and intermediate-expression (11%-69%) cases were infrequent (2.7% & 5.1% and 1.6% & 9.2% in cores and TMA respectively) and did not show any survival difference to ER negative tumours. When full-face sections of the corresponding excision specimens were immunostained, 47% of the ER low/intermediate group were deemed to ER negative. Transcriptomic data not only showed significant correlation between *ESR1* mRNA and protein expression levels but also confirmed the bimodality of ER at the mRNA levels. **Conclusion:** Our study provides further evidence that ER expression is bimodal and that it is observed at both mRNA and protein levels. The reported poor survival of BC patients with low ER-expression in the early clinical trials may be related to the inclusion of ER-negative cases.

## INTRODUCTION

Global gene expression studies of breast cancer (BC) have demonstrated that oestrogen receptor (ER) is the main determinant of BC molecular profiles and that ER positive and negative disease BCs are different diseases <sup>1, 2</sup>. Although several randomised clinical trials have demonstrated that ER negative BC does not respond to endocrine therapy and the effect of endocrine treatments is restricted to ER positive disease, the response of ER positive tumours is variable. Only two thirds of ER positive BC patients treated with endocrine therapy respond <sup>3</sup>. Several studies have reported correlation between the level of ER expression and response to endocrine therapy and clinical trials indicate that only 50% of tumours with an Allred score of 4–6 respond to endocrine treatments compared to 75%–80% in tumours with a score of 7 or 8 <sup>3-6</sup>.

Before immunohistochemistry (IHC), ER expression was measured by radiolabelled ligand binding assays (LBA) using fresh-frozen tumour samples which showed a continuum of values. For more than two decades, ER expression has been assessed using IHC and is used to predict response to endocrine therapy<sup>4, 7</sup>. In addition to being a prognostic marker, some authors reported that the linear distribution of ER expression reported with LBA was observed with IHC <sup>4, 5</sup>. However, other authors have challenged the concept of ER linear expression and they provided evidence that ER expression is essentially bimodal <sup>8, 9</sup>.

Identifying genes with bimodal expression patterns from large-scale gene expression profiling data has provided new insights into the distribution of expression of key

genes. At the transcriptomic level, there are two different classes of genes. The first class is consistent of those with a Gaussian or continuous distribution: two small groups of tumours have very high expression and very low expression respectively, with the rest; which represent the majority, fall somewhere in between. The second class is composed of genes with a bimodal distribution of expression. This class has the majority of tumours with either high levels of gene expression or no expression, and only a relatively few tumours fall in between <sup>10</sup>. Previous studies have reported significant correlations between *ESR1* expression and the clinical ER status <sup>11</sup>. *ESR1* has a high Bimodality Index score and it can be used to classify samples into two distinct expression states <sup>12</sup>.

This study aimed to assess the pattern of ER expression, both at the protein (IHC) and transcriptome (microarray-based gene expression) levels, to comprehensively understand the clinical and biological value of tumours expressing low/intermediate levels of ER. To achieve this, we reviewed three cohorts of primary BC: 1) 3649 ER IHC stained core biopsies performed at Nottingham City Hospital 2) 1892 cases prepared as tissue microarray (TMA) and IHC stained for ER and 3) *ESR1* mRNA level of 1980 BC cases that were included in the METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) study <sup>13</sup>.

## **MATERIALS AND METHODS**

### **Patients Cohorts**

This study was based on 3 patient's cohorts and was approved by the Nottingham Research Ethics Committee 2 under the title 'Development of a molecular genetic classification of breast cancer'. ER was assessed using IHC in the first and second cohorts and mRNA in the third cohort. Regarding the level of ER expression, weak positive cases are defined as those with ER expression between 1% and 10% while intermediate expression cases are defined as those scored between 11% and 69%. The negative cases are defined according to the ASCO/CAP guidelines<sup>14</sup> with a score of <1% while the highly positive cases are defined as those expressing ER in 70% or more of the tumour cells.

1) The first cohort comprised a consecutive series of symptomatic and screen detected BC patients who had ER status assessed on a preoperative core needle biopsy (CNB) at Nottingham City Hospital in routine practice between March 2008 and November 2014 (n=3649 cases). CNB were fixed, processed and stained according to standardised protocol as previously published <sup>15</sup>. The primary antibodies used were 1D5 (Dako; UK), diluted 1/100 and a pre-diluted SP1 clone (Roche; UK).

2) The second cohort comprised 1892 patients diagnosed between 1988 and 1998 whose BC tissues were prepared as tissue microarray (TMA) as part of the previously published Nottingham Tenovus Primary Breast Carcinoma Series <sup>16</sup>. This is a consecutive well-characterised series of early stage primary operable BC patients aged 70 years or less. In this cohort, patients' outcome, including regional and distant events, survival and time to the event, was recorded and annually updated. This included BC specific survival (BCSS). Moreover, the clinical details of the patients including age and menopausal status as well as the tumour details including tumour size, grade, stage, lymphovascular invasion (LVI) and lymph node

status were available. TMA sections of this series were stained with the ER antibody SP1 clone (Dako, UK), using a dilution 1:150 for 30 minutes incubation. Levels of immunohistochemical expression were assessed by microscopic eyeballing as the percentage of ER positively stained invasive tumour cells.

3) The third cohort comprised 1980 BC cases that were included in the METABRIC cohort <sup>13</sup>. In this study, the extracted and purified DNA probes were hybridised to Affymetrix SNP 6.0 arrays (Affymetrix, Santa Clara, CA) at the quality control criteria established by AROS Applied Biotechnology (Aarhus, Denmark). The Illumina Total prep RNA amplification kit (Ambion, Warrington, UK) was used for total extraction and purification of total RNA. The generated biotinylated cRNA was hybridised with Illumina Human HT-12 v3 Expression Beadchips from the same manufacturer <sup>17</sup>.

The resulted gene expression data from the microarray experiments was statistically analysed by applying the Linear Models for Microarray Data (LIMMA) inclusive software package that is compatible with the Affymetrix data. LIMMA software package comprised Student's t- test, base 2 logarithms of fold changes between the normal and tested samples, log<sub>2</sub>FC, average expression measurements, t-test values, p-values, adjusted p-values, and log-odds of differential expression, B values as its primary statistical outputs. In this cohort, the survival data and the ER immunostaining was available for 262 cases from Nottingham.

### **Immunohistochemistry**

Surgical excision specimens of 55 cases from the first cohort that showed a low / intermediate ER expression on CNB were immunostained to confirm the level of ER expression. These 55 cases comprised 39 from the low ER expression group (ER

scored between 1% and 10%) and 16 from the intermediate ER expression group (ER scored between 11% and 69%). The immunohistochemistry technique was applied to full-face sections that are considered as the “gold standard” for ER assessment using the standard protocol. In brief, antigen retrieval was carried out using citrate at pH 6.0 for 20 min at the microwave. Manual immunohistochemistry staining was used by Novolink Max Polymer Detection Kit, Ref: RE7280-K. Peroxidase blocking was done by applying peroxidase block for 5 min. Optimised primary antibody, EP1 anti ER Rabbit monoclonal antibody (Dako, Ref- M3643) was applied and incubated for 30 minute at room temperature followed by Novolink polymer for 30 min and enzyme substrate for 5 min. Then, Novolink haematoxylin was added for 6 min to each slide. Finally, the slides were dehydrated and mounted by coverslip using DPX (BDH, Leica Microsystems, Newcastle UK). Nuclear ER staining was then scored using percent and H-score system by two pathologists (AAM and EAR).

### **Statistical analysis**

For IHC expression, in the first and second cohort, the statistical tests were performed using IBM SPSS software version 22 (SPSS Inc., Chicago, IL, USA). The data was categorised into three groups, the first group included those cases with ER expression scored as <1%. The second group included cases between 1-69%. The third group contained cases with ER positive  $\geq 70\%$  which is similar to the cut-off point used by Collins et al.<sup>8</sup>.

In the second cohort, additionally, survival curves were analysed using the Kaplan–Meier method with significance determined by the Log Rank (LR) test. The association between subcategories and the clinicopathological variables as well as

PgR and HER2 as basic prognostic and predictive markers were evaluated using the Chi square test. For all statistical tests, p-value was taken into consideration as a significant value when it was  $<0.05$ .

For transcriptomic data of the third cohort, 1980 BC cases were used to demonstrate the *ESR1* gene expression distribution pattern. However, patient outcome and immunostaining of ER were only available for 262 cases. These cases were subdivided into three subgroups according to the changes in the distribution of the curve using the SPSS software. The first cut-off point was 6.5 and the second cut-off point was 8.6. The high expression group (group 3) included those cases with *ESR1* expression level of  $>8.6$ . Gene expression was compared with protein levels and patient outcome. Survival curves were analysed by Kaplan–Meier method with significance determined by the LR test and a p-value  $<0.05$  considered significant.

## RESULTS

Analysis of the 3649 cases assessed on CNBs showed that the majority of cases (92.2%) were either strongly positive ( $\geq 70\%$ ) or negative ( $< 1\%$ ) for ER while weak ER positive (1-10%) cases (2.7%) and the intermediate-expression (11-69%) group (5.1%) were infrequent. Figure 1 shows the bimodal distribution of ER expression where those with  $< 1\%$  expression represented 22.4 % of cases and strong ER positive cases represented 69.8 % of the total cases. To further assess the existence and frequency of ER low/intermediate expressing tumours, 55 cases were immunostained using full-face sections of excision specimens. Out of those 55 cases, 26 (47.3%) resulted in negative ER expression on excision specimens while 29 cases remained in the low/intermediate group; reducing the frequency of low/intermediate group further to approximately 3.9%.. Regarding the low ER expression cases assessed on CNBs (39 cases), 16 cases remained positive for ER while 23 cases changed into negative ER expression. For the intermediate ER expression tumours assessed on CNBs (16 cases), 13 cases remained positive while 3 cases changed to the negative ER expression subgroup. Figure 2 illustrated negative, low, intermediate and highly positive ER expression cases.

The frequency distribution of ER staining, based on the percentage of ER positive tumour cells, in the TMA series of 1892 cases is shown in Figure 3. Similar to those in CNBs, this series showed bimodality of expression with the completely negative and strongly positive cases representing 89.2% of the cases. Weak and intermediate expression of ER was infrequent representing 1.6% and 9.2% of cases, respectively.

Due to the small number of cases in the weak positive and intermediate expression, they were combined and data was analysed accordingly; negative group (<1%), intermediate group (1-69%) and highly positive group ( $\geq 70\%$ ). The associations of these subgroups and other clinicopathological variables, progesterone receptor (PR) and HER2 status is shown in Table 1. The association of the ER subgroup with HER2 was highly significant where 25% and 22% of the negative and intermediate ER groups were HER2 positive compared to 6% of the strong ER positive groups that showed HER2 positivity. Progesterone receptor is an ER dependent protein <sup>18</sup>. In this study, we assessed the frequency of expression of PR in the TMA series using the previously published IHC stained PR <sup>16</sup> and this showed a bimodal distribution similar to ER (Figure 4).

Outcome analysis showed significant association between ER expression groups and patients' outcome (Figure 5;  $p < 0.001$ , LR=23.5). However, further analysis showed no significant survival difference between the ER negative group and the intermediate ER expression group ( $p = 0.324$  and LR=0.97), while there was a difference between the intermediate group and the highly positive group ( $p = 0.022$  and LR=5.21). Expectedly, the difference in patient outcome between the negative and highly positive group was highly significant ( $p < 0.001$  and LR=22.05).

Consistent with IHC results, analysis of the METABRIC cases ( $n = 1980$ ) showed a bimodal distribution of *ESR1* mRNA (Figure 6) with 18.2 % of cases in group one (representing the ER negative group) and 72 % in group three (representing the ER highly positive group) whilst group two (the ER intermediate group) represented only 9.8% of cases. There was a significant positive correlation between the *ESR1* mRNA

and protein expression levels (p-value <0.001). The majority (79%) of ER IHC negative cases were in the negative *ESR1* mRNA group. In addition, 97.1% of the strongly ER IHC positive group belonged to the mRNA highly positive group. Interestingly 77.8% of the IHC intermediate group (1-69%) correlated to the negative and strongly positive mRNA group (Table 2). Outcome analysis showed significant differences between the groups (Figure 7; p=0.001 and LR= 13.28). No significant difference between the negative group and the intermediate group was identified (p=0.74 and LR=0.10). However, the difference between the highly positive group and the intermediate group regarding the patient outcome was significant (p=0.01 and LR= 6.52).

## DISCUSSION

The concept of bimodality and linear distribution of ER expression in breast cancer remains a subject of debate. The current study provides evidence that the distribution of ER expression is bimodal at both protein and gene expression levels using IHC and gene expression microarrays technology respectively. At the protein level our data are consistent with the results reported by *Collins* et al. in which 99% of 817 cases were completely negative or positive <sup>8</sup> and with *Nadji* et al. who showed that ER IHC in BC is an all-or-none phenomenon <sup>9</sup>. In support of our findings, several other authors have demonstrated the bimodal distribution of ER immunostaining including *Bogina* et al. (362 cases) <sup>19</sup>, *Badve* et al. (776 cases) <sup>20</sup>, *Pinto* et al. (360 cases) <sup>21</sup> and *Khoshnoud* et al. (683 cases) <sup>22</sup>.

The *Schnitt* group have commented on the reproducibility of the continuity of expression seen using the LBA when a large population-based study was tested using IHC <sup>9</sup>. They described the relationship between the real quantity of ER protein in the malignant cell nuclei and the apparent amount of ER demonstrated by IHC as a highly complex process and related this to the effect of the pre-analytic factors on the IHC results <sup>23</sup>. The pre-analytical factors and the overall sensitivity of IHC were considered as the main explanation for the bimodality of the ER distribution. It has been demonstrated that IHC can vary depending on a pre-analytical factors as for instance tissue fixation<sup>24, 25</sup> and antigen retrieval. *Umemura* et al demonstrated a positive correlation regarding a linear relationship between the LBA and a low sensitivity IHC assays. This correlation was reduced using a highly sensitive IHC test with nonlinear correlation and the biochemical assays resulted in a shift of the low ER cases toward the higher end of ER positivity <sup>26</sup>. This reason was used to explain

the low frequency of the low ER cases. Interestingly although the techniques and the testing level used to evaluate ER distribution is totally different, in our study when the 1980 cases of METABRIC BC data set were evaluated, the similarity between the transcriptomic and proteomic levels were significantly high. Both IHC and mRNA levels using large data sets demonstrated the bimodal distribution. Therefore, the explanation that the bimodal distribution is a result of the pre-analytical factors or the highly sensitive IHC test has to be re-considered. The bimodal gene expression for *ESR1* using 123 patients was presented previously by Wang et al.<sup>12</sup>. Their method was also applied to the MDA133 breast cancer microarray dataset<sup>27</sup>.

This study demonstrated strong correlation between ER protein levels (IHC) and *ESR1* mRNA levels as assessed using gene expression microarrays. Interestingly a large proportion of the IHC intermediate group related to the strong *ESR* expresser group. However, this can be explained by either using different cut-offs in each cohort or by the differential translation of ER in some cases. The use of 70% as a cut-off to define the intermediate group in this study may have resulted in the underestimation of the level of ER positivity defined using *ESR* mRNA. This study also did not assess the role of variable ER expression in the breast cancer associated stroma on the strength of the association between ER IHC and *ESR* mRNA levels

At the proteomic as well as the transcriptomic level, our results showed no significant difference between the ER negative and the intermediate expression groups and outcome. Therefore, our findings raise the clinical question whether patients with low positive ER expression are actually a positive cases or the misclassification of ER-negative cancers as borderline weak ER positive. In this study, approximately 50% of the low/intermediate IHC ER positive group changed to ER negative when full-face

sections of excision specimens were immunostained. Although this may reflect a proportion of cases with false positive ER expression on core biopsies <sup>28</sup>, further research is needed to evaluate this important group of cases. In this study HER2 overexpression was more frequent in the low and intermediate ER expresser groups compared to the strong ER expresser group. The high frequency of HER2 overexpression in these tumours may support their aggressive behaviour and that their response to therapy may be similar to ER-negative group than to the strong ER positive tumours. Some of these tumours would be classified in the luminal B or HER2-enriched classes however; further study is warrant to investigate this point.

In conclusion, our study provides further evidence to support the concept that ER expression in BC is essentially bimodal with the vast majority of cases being either ER negative or strongly ER positive. The biological and clinical significance of the intermediate expression group particularly the low ER expression subgroup needs further investigation.

**Conflict of Interest:** The authors have no conflicts of interest.

**Author Contributions:**

**Abir A. Muftah:** Writing the manuscript, IHC staining and scoring, data collection, analysis and interpretation.

**Mohammed Aleskandarany:** Data analysis and interpretation, contribution in writing and reviewing the manuscript.

**Sultan N. Sonbul:** Data analysis and interpretation.

**Christopher C. Nolan and Maria Diez Rodriguez:** Helping in lab work.

**Carlos Caldas:** Providing data.

**Andy R Green:** Data collection, contribution in writing and reviewing the manuscript.

**Ian O. Ellis:** Data analysis and interpretation.

**Emad A. Rakha:** Generated the hypothesis and design of the study, data interpretation, contribution in writing, reviewing and approving the manuscript.

All authors approved the final version.

## REFERENCES

1. Colombo PE, Milanezi F, Weigelt B, et al. Microarrays in the 2010s: the contribution of microarray-based gene expression profiling to breast cancer classification, prognostication and prediction. *Breast Cancer Res.* 2011;13:212.
2. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature.* 2000;406:747-52.
3. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet.* 2005;365:1687-717.
4. Harvey JM, Clark GM, Osborne CK, et al. Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. *J Clin Oncol.* 1999;17:1474-81.
5. Elledge RM, Green S, Pugh R, et al. Estrogen receptor (ER) and progesterone receptor (PgR), by ligand-binding assay compared with ER, PgR and pS2, by immunohistochemistry in predicting response to tamoxifen in metastatic breast cancer: a Southwest Oncology Group Study. *Int J Cancer.* 2000;89:111-7.
6. Early Breast Cancer Trialists' Collaborative G, Davies C, Godwin J, et al. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet.* 2011;378:771-84.
7. Barnes DM, Harris WH, Smith P, et al. Immunohistochemical determination of oestrogen receptor: comparison of different methods of assessment of staining and correlation with clinical outcome of breast cancer patients. *Br J Cancer.* 1996;74:1445-51.
8. Collins LC, Botero ML, Schnitt SJ. Bimodal frequency distribution of estrogen receptor immunohistochemical staining results in breast cancer: an analysis of 825 cases. *Am J Clin Pathol.* 2005;123:16-20.
9. Nadji M, Gomez-Fernandez C, Ganjei-Azar P, et al. Immunohistochemistry of estrogen and progesterone receptors reconsidered: experience with 5,993 breast cancers. *Am J Clin Pathol.* 2005;123:21-7.
10. Kernagis DN, Hall AH, Datto MB. Genes with bimodal expression are robust diagnostic targets that define distinct subtypes of epithelial ovarian cancer with different overall survival. *J Mol Diagn.* 2012;14:214-22.
11. Ji Y, Coombes K, Zhang J, et al. RefSeq refinements of UniGene-based gene matching improve the correlation of expression measurements between two microarray platforms. *Appl Bioinformatics.* 2006;5:89-98.
12. Wang J, Wen S, Symmans WF, et al. The bimodality index: a criterion for discovering and ranking bimodal signatures from cancer gene expression profiling data. *Cancer Inform.* 2009;7:199-216.
13. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature.* 2012;486:346-52.
14. 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. *J Clin Oncol.* 2010;28:2784-95.
15. Hodi Z, Chakrabarti J, Lee AH, et al. The reliability of assessment of oestrogen receptor expression on needle core biopsy specimens of invasive carcinomas of the breast. *J Clin Pathol.* 2007;60:299-302.
16. Rakha EA, El-Sayed ME, Green AR, et al. Biologic and clinical characteristics of breast cancer with single hormone receptor positive phenotype. *J Clin Oncol.* 2007;25:4772-8.

17. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 486. England 2012. p. 346-52.
18. Kastner P, Krust A, Turcotte B, et al. Two distinct estrogen-regulated promoters generate transcripts encoding the two functionally different human progesterone receptor forms A and B. *EMBO J*. 1990;9:1603-14.
19. Bogina G, Zamboni G, Sapino A, et al. Comparison of anti-estrogen receptor antibodies SP1, 6F11, and 1D5 in breast cancer: lower 1D5 sensitivity but questionable clinical implications. *Am J Clin Pathol*. 2012;138:697-702.
20. Badve SS, Baehner FL, Gray RP, et al. Estrogen- and progesterone-receptor status in ECOG 2197: comparison of immunohistochemistry by local and central laboratories and quantitative reverse transcription polymerase chain reaction by central laboratory. *J Clin Oncol*. 2008;26:2473-81.
21. Pinto AE, Areia F, Pereira T, et al. Clinical relevance of the reappraisal of negative hormone receptor expression in breast cancer. *Springerplus*. 2013;2:375.
22. Khoshnoud MR, Lofdahl B, Fohlin H, et al. Immunohistochemistry compared to cytosol assays for determination of estrogen receptor and prediction of the long-term effect of adjuvant tamoxifen. *Breast Cancer Res Treat*. 2011;126:421-30.
23. Schnitt SJ. Estrogen receptor testing of breast cancer in current clinical practice: what's the question? *J Clin Oncol*. 2006;24:1797-9.
24. Werner M, Chott A, Fabiano A, et al. Effect of formalin tissue fixation and processing on immunohistochemistry. *Am J Surg Pathol*. 2000;24:1016-9.
25. Goldstein NS, Ferkowicz M, Odish E, et al. Minimum formalin fixation time for consistent estrogen receptor immunohistochemical staining of invasive breast carcinoma. *Am J Clin Pathol*. 2003;120:86-92.
26. Umemura S, Itoh J, Itoh H, et al. Immunohistochemical evaluation of hormone receptors in breast cancer: which scoring system is suitable for highly sensitive procedures? *Appl Immunohistochem Mol Morphol*. 2004;12:8-13.
27. Hess KR, Anderson K, Symmans WF, et al. Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *J Clin Oncol*. 2006;24:4236-44.
28. Rakha EA, Lee AH, Roberts J, et al. Low-estrogen receptor-positive breast cancer: the impact of tissue sampling, choice of antibody, and molecular subtyping. *J Clin Oncol*. 2012;30:2929-30; author reply 31.

## Figure legends:

**Figure 1:** The distribution of oestrogen receptor (ER) IHC expression as assessed on core needle biopsies (3649 cases) demonstrating the bimodal distribution.

**Figure 2:** Illustrated different levels of ER expression, A) negative ER expression. B) Low ER expression. C) Intermediate level of ER expression and D) Highly positive ER expression.

**Figure 3:** The frequency distribution of ER IHC staining scores in the second cohort (1892 cases) assessed on tissue microarrays (TMA) showing similar bimodal distribution as figure 1.

**Figure 4:** The distribution of progesterone receptor (PR) IHC expression as assessed on the second cohort demonstrating the bimodal distribution similar to ER expression.

**Figure 5:** Kaplan Meier plots illustrating BCSS and IHC subgroups (negative, intermediate and highly positive group).

**Figure 6:** Histogram representing a bimodal transcript level distribution for the ESR1 gene.

**Figure 7:** Kaplan Meier plots illustrating of breast cancer specific survival (BCSS) and transcriptomic subgroups (group one = negative, group two = intermediate, and group three = highly positive group)