Refining the definition of HER2 low class in invasive breast cancer

Nehal M Atallah[1,2,3], Michael S Toss[3,4], Andrew R Green[3], Nigel P Mongan[5], Graham Ball[6], Emad A Rakha [1,2]

[1]Department of Histopathology, School of Medicine, the University of Nottingham and Nottingham University Hospitals NHS Trust, Nottingham, UK, [2]Department of Pathology, Faculty of Medicine, Menoufia University, Egypt, [3]Division of Cancer and Stem Cells, Biodiscovery Institute, School of Medicine, University of Nottingham, Nottingham, UK, [4]Histopathology Department, Royal Hallamshire Hospital, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK, [5]School of Veterinary Medicine and Sciences, University of Nottingham, Sutton Bonington, UK, [6]Division of Life Sciences, Nottingham Trent University, Nottingham, UK

Corresponding Author:

Professor Emad A Rakha

Department of Histopathology, Nottingham University Hospital NHS Trust,

City Hospital Campus, Hucknall Road, Nottingham, NG5 1PB, UK

Tel: (44) 0115-9691169, Fax: (44) 0115-9627768

Email: emad.rakha@nottingham.ac.uk, Emad.rakha@nuh.nhs.uk

ORCID: 0000-0002-5009-5525

Running title: Refine HER2 low definition

Keywords: Breast cancer, refining, HER2 low, *HER2* mRNA, ANN

Abstract

Background: Emerging evidence indicates that breast cancer (BC) patients whose tumours express HER2 protein without *HER2* gene amplification (HER2-low), can benefit from antibody-drug conjugates (ADC). However, the current definition of HER2-low BC remains incomplete with low rates of concordance. This study aims to refine HER2-low definition with emphasis on distinguishing HER2 score 0 from score 1+ to identify patients who are eligible for ADC.

Methods: BC cohort (n=363) with HER2 IHC scores 0, 1+ and 2+ (without *HER2* gene amplification) and available *HER2* mRNA was included. HER2 staining intensity, pattern, and subcellular localisation were reassessed. Artificial neural network analysis was applied to cluster the cohort and to distinguish HER2 score 0 from 1+. Reproducibility and reliability of the refined criteria were tested.

Results: HER2 IHC score 1+ was refined as membranous staining in invasive cells as either: 1) faint intensity in ≥20% of cells regardless the circumferential completeness, 2) weak complete staining in ≤10%, 3) weak incomplete staining in >10%, 4) moderate incomplete staining in ≤10%. Based on this, 63% of the HER2 negative cases were reclassified as positive (HER2-low). The refined score showed perfect observer agreement compared to the moderate agreement in the original clinical scores. Similar results were generated when the refined score was applied on the independent BC cohorts. A proposal to refine the definition of other HER2 classes is presented.

Conclusion: This study refined the definition of HER2-low BC based on correlation with *HER2* mRNA and distinguished between HER2 IHC score 1+ and score 0 tumours.

INTRODUCTION

Human epidermal growth factor receptor 2 (HER2) gene is amplified in about 15% of invasive breast cancer (BC) leading to HER2 protein overexpression [1-4]**.** HER2 testing in routine practice is performed using immunohistochemistry (IHC) to assess the level of protein expression which is reported using a range of 0 to 3+ score [5, 6]. HER2 positive BC is defined as IHC score 3+ or score 2+ with evidence of *HER2* gene amplification using *in situ hybridization* (ISH) technique. HER2 positive BC patients are eligible for therapies that target the HER2 pathways [7-9]. BC with HER2 IHC score 2+ that lacks evidence of *HER*2 gene amplification is currently classified as HER2 negative similar to cases showing IHC score of 0 or 1+ [5, 6] and do not benefit from anti-HER2 therapy. However, recent data have demonstrated that some of the HER2 directed antibody-drug conjugates (ADC) such as trastuzumab-emtansine (T-DM1) and trastuzumab-deruxtecan (T-DXD) can improve the outcome of patients with BC that express HER2 protein without evidence of *HER2* gene amplification [10]. These cases included BC with HER2 IHC score 1+ or score 2+ without *HER*2 gene amplification, which are defined as the HER2-low class [11-14].

ADCs are molecules consisting of a recombinant monoclonal antibody covalently bound to a cytotoxic drug via a linker. After antibody binding to the specific antigen on the targeted cell surface, the cytotoxic drug gets internalized, and is released intracellularly where it can exert its effect. ADC effect relies on the presence an extracellular protein receptor which acts as a carrier for the cytotoxic agent to achieve targeted effect with no or minimal levels of cytotoxicity to the normal cells, rather than on the oncogenic effect of the protein. Patients' recruitment to the ongoing HER2 low positive clinical trials which are testing the effect of ADCs in BC are based on the

existing definition of HER2 categories as described in the American Society of Clinical Oncology and College of American Pathologists (ASCO/CAP) [6].

Although the ASCO/CAP guideline recommendations provided comprehensive definition of HER2 staining pattern and the categorization of cases into 4 IHC scores (0 - 3+), the distinction between IHC score 0 and 1+ is not sufficiently detailed and lack relevant evidence and some scenarios of HER2 expression patterns are missing [5, 6]. This could participate in the high discordance rates in HER2 status assessments reported in some studies [15-17].

Although clinical response can provide the best tool to define the lower limit of the HER2-low class, the number of recruited patients in such randomised clinical trials, particularly those close to the threshold of positivity, is typically too limited to develop a robust definition. In this study, we have used a large cohort of BC that express low levels of HER2 protein without evidence of *HER2* gene amplification and applied an artificial neural network (ANN) model to refine the definition of HER2-low class of BC with an emphasis on distinguishing HER2 score 1+ and 0 categories. We have used the *HER2* mRNA levels as a ground truth to reflect the level of *HER2* gene expression. ANNs can learn and model non-linear and complex relationships [18-22]**.** We have also tried to refine the existing definitions of HER2 IHC categories by completing the missing scenarios utilising the existing data and our experience.

MATERIALS AND METHODS

This study was conducted on a primary invasive BC cohort (n=363) from patients presenting at Nottingham University Hospitals NHS Trust with a HER2 IHC score 0, 1+ and 2+ without gene amplification. Transcriptomic data on *HER2* mRNA expression was available for this cohort within the recorded Oncotype DX report [23] which was

carried out as part of the patients' clinical care for management**.** Briefly, mRNA levels were obtained from tumour samples extracted from formalin fixed paraffin embedded tissue using high-throughput, real-time, reverse transcription-polymerase chain reaction. Normalized expression measurements were calculated as the mean cycle threshold (CT) for the 5 reference genes minus the mean CT of triplicate measurements for each individual gene. *HER2* mRNA level ranged between 5.0-10.8 units with a mean of 9 units.

The clinicopathological data including age at diagnosis, tumour size, histological grade, histologic tumour type, axillary lymph node status, lympho-vascular invasion (LVI) and Nottingham prognostic index (NPI) were available (**Supplementary Table 1**). The patients mean age at diagnosis was 59 years while the mean invasive tumour size was 2.2 cm (range 0.1-11.5 cm). All cases were oestrogen receptor (ER) positive and HER2 negative. ER (and progesterone receptor (PR)) positivity were assessed according to ASCO/CAP guidelines if ≥1% of the invasive tumour cell nuclei are immunoreactive[24]. HER2 staining was completed on the Ventana Benchmark ULTRA immunohistochemistry automated staining system using the Ventana PATHWAY anti-HER-2/*neu*, Rabbit Monoclonal ready to use primary antibody in combination with Ventana detection kits. No antigen retrieval was required according to the protocol.

Appropriate positive and negative controls were included for each staining run as per the published guidelines [5, 6]. Protein expression assessment was carried out in routine clinical practice using light microscopy on the diagnostic core needle biopsies. The reported HER2 scoring categories in the clinical setting were retrieved from the patient records.

Detailed re-assessment of HER2 IHC protein expression

HER2 expression within the invasive tumour cells only of each case, was reassessed and presented in detail. This included: 1) Cellular localisation of protein expression (membranous, cytoplasmic or both), 2) Intensity of staining divided into 5 grades (negative, faint, weak, moderate, and strong). In addition to the comparison with the positive and negative controls, the magnification rule was used to guarantee high inter-observer agreement. Strong HER2 staining was assessed as those cases displaying unequivocal membranous staining seen easily at low power magnification (2x or 4x), while unequivocal membranous staining (moderate to weak) was only assigned at medium magnification (10x to 20x, respectively). Faint staining can only be appreciated at 40x magnification whereas weak staining can be appreciated at 20x magnification.[25]. Cases were assessed using NIKON NI-U Microscope, Nikon UK, Branch of Nikon Europe B.V. UNITED KINGDOM. Different intensities within the same tumour were assessed to reflect the heterogeneity. 3) The percentage of each intensity. 4) Distribution/completeness of membranous staining as either complete circumferential membranous or incomplete lateral or basolateral staining. 5) H score was calculated as follows: % of weak intensity X 1 + % of moderate Intensity X 2+ % of strong intensity X 3). In addition, the % of faint intensity was assessed and multiplied by 0.5 to produce a total score of 350. Each incomplete membranous staining is multiplied by 0.5, while complete membranous staining is multiplied by 1.

HER2 staining on full face sections of resection specimen

HER2 IHC staining, and scoring was performed on core biopsies while *HER2* mRNA level was assessed on resection specimens. Thus, for the cases that showed discrepancy between HER2 IHC score and mRNA level (n=30) i.e., high mRNA level

and HER2 score 0, or HER2 score 2+ with low mRNA level, repeating HER2 IHC staining on the full-face sections using the same tissue block tested for oncotype DX was performed. Whenever possible, the same tissue block that was used to run the Oncotype DX test, was stained with HER2. The staining protocol was similar to the core biopsy staining as described above.

Defining cut-off for HER2 score 1+ versus score 0

Two steps were followed to define HER2 score 1+ **(Figure 1)**

Step 1: K-means clustering

The K-means technique aims to partition the data into k-groups such that the sum of squares from points to the assigned cluster centres is minimized. *HER2* mRNA values were classified, using k-means, into 2 clusters based on their similarity of expression across multiple HER2 scoring parameters. Those cases which had a score of 1+ or 0, were clustered into two groups based on *HER2* mRNA level and the detailed IHC scoring performed. Cluster 1 was defined as HER2 negative (0) while cluster 2 represented HER2 positive (1+). HER2 2+ were excluded from the clustering to avoid data bias.

Step 2: Artificial neural network model (ANN)

ANN model (NeuroSolution version 7.0, NeuroDimension, Gainesville, FL with a range of hidden nodes in three layers, with a Levenburg Marquardt algorithm and a TanH activation function) was used to set cut off point for defining HER2 1+ based on the k-means clusters defined in step 1. A Monte Carlo Cross Validation approach was used to train a population of models and early stopping was undertaken using a randomly extracted unseen cross validation set with subsequent validation on a test set (n=38)

which was kept completely blind to training process. Weight regularization was conducted during training.

The model was trained with the detailed HER2 scoring parameters including various intensities (faint, weak, moderate) and distribution of each intensity if present either complete and incomplete in addition to total percentage of positive cells and cytoplasmic staining as input and *HER2* mRNA-based clusters as an output variable. The ANN model determined which of the input parameters predicted HER2 score 1+ with high level of accuracy. Sensitivity and specificity with the produced response curves was applied to set cut-off for the most participating parameter.

Predictions of trained models were examined to decide predicted probability of K-means cluster membership. These were examined to determine a probabilistic cut point for HER2 score 1+. Model performance was further assessed by finding the area under the curve (AUC) of a constructed receiver operator characteristic (ROC) curve. AUC of 1 were seen across the three cross validation cohorts as well as 100% classification rates.

After setting the cut points, a new refined score for HER2 was developed and applied. To detect the accuracy of our refined score against the clinical score, we used the same neural network to build a discriminating model of both HER2 scores using the clinicopathologic parameters as input units and the HER2 score as an output. The differentiating performance of the ANN models was evaluated with AUC as well as the true and false positivity rates.

To test the reliability of using *HER2* mRNA as a dichotomizing variable, we assessed the correlation between *HER2* mRNA, protein level and gene amplification levels in a large independent cohort of primary BCs obtained from two publicly available datasets;

The Cancer Genome Atlas (TCGA) (n=614) and the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) (n=288).

Reliability and reproducibility of the refined HER2 IHC score

The efficiency and reproducibility of the refined HER2 scoring method against the current guidelines were tested. HER2 was scored twice according to the existing ASCO/CAP; once by the clinical team at time of diagnosis and a second score was carried out by experienced pathologists (NA and MT) who have more than 5-years' experience in histopathology and supervised by an experienced breast pathology consultant (ER) with more than 20 years' experience in the field of breast pathology. The agreement between two scores was assessed. Moreover, the inter-observer agreement of the refined score was assessed between both observers and the intra-observer agreement was examined through rescoring the cases again after 3 months washout period.

Correlation between refined HER2 score with the clinicopathological variables

The correlation between the clinicopathologic parameters, including *HER2* mRNA level, HER2 scores including the refined and the original clinical scores was carried out. In addition, *HER2* mRNA k-means clusters were correlated with the other clinicopathologic parameters.

Statistical Analysis

SPSS v.24 was used to carry out the statistical analysis. Correlations were analysed using chi-square (χ^2) test, Fisher's exact test, Kruskal–Wallis and Wilcoxon rank sum test with continuity correction, where appropriate. The concordance analysis was

performed using the Cohen's Kappa test. All differences were considered significant at $p < 0.05$.

RESULTS

Patterns of HER2 protein expression

Of the cases in the study cohort, 81% showed a degree of HER2 expression regardless of the pattern and/or the percentage of positive cells. It was observed that each case had a mixture of expression patterns, intensities, and cellular localisation. The most frequent pattern observed was incomplete faint staining which presented in 78% of the cohort followed by complete faint expression which in 58% of cases with or without other patterns of expression. Moderate incomplete staining had the lowest proportion among all patterns of expression (4%). Detailed description of HER2 expression in terms of staining intensities, patterns and percentages are summarized in **(Supplementary Table 2 and Supplementary Figure 1).**

Correlation between HER2 IHC score and mRNA level

Discrepancy between HER2 IHC protein expression score on core biopsy and mRNA level was seen in 30/363 (8%) of cases. IHC score 2+ with low mRNA level, as defined based on k mean clustering analysis, was presented in (2/30) cases while the remaining 28 cases had IHC score 0 with high mRNA level. Upon staining those 30 cases on full face tissue sections, all cases that were score 0 on core biopsy were completely negative (score 0) in the invasive tumour cells with weak to moderate staining within the *in situ* component. While the two cases that were score 2+ in core biopsy turned to be completely negative (score 0) on full face sections.

Clustering of *HER2* mRNA

A total of 308 cases with complete data on different HER2 expression patterns and *HER2* mRNA level were available for k-mean clustering analysis. The dataset was divided into 2 clusters; Cluster 1 (n=109) and Cluster 2 (n=199), based on mRNA level at cut off 8.7units. Based on the new cut off, the mean values ± SD of *HER2* mRNA in HER2 in HER2 IHC 0, 1+ and 2+ was 8.66 ± 0.69, 9.1± 0.66 and 9.4 ± 0.78 (Supplementary Figure 2A).

ANN model sensitivity

The parameters that predicted HER2 cluster 2 (equivalent to score 1+) were, faint complete staining at ≥ 20% of invasive tumour cells and/or faint incomplete staining in ≥ 20%, weak complete staining in ≤10%, weak incomplete staining in >10%, moderate incomplete membranous staining in ≤10% of invasive tumour cells. Cytoplasmic staining and total percentage of positive cells did not define the cluster **(Table 1, Supplementary Figure 3). Figure 2 (a, b)** shows a schematic illustration of different scenarios of HER2 expression in BC and the corresponding score based on the refined criteria compared to the existing guidelines [5, 6, 26, 27].

Based on the new defined cut points for low HER2 IHC scoring, 136/363 (37%) cases were scored 0, 140/363 cases (39%) with score 1+ and 87/363 cases (24%) were score 2+ compared to 126/363 (35%), 156/363 (43%) and 81/87 (22%), for the original scores 0, 1+ and 2+, respectively (Supplementary Figure 2B).

Faint staining intensity was the most predominant pattern in HER2 score 1+ (41/140), followed by weak incomplete staining in >10% of invasive tumour cells and then weak complete staining less than or equal 10% of cells. Exclusive moderate expression

either complete or incomplete <10% was not found in HER2 1+ as a unique pattern but was expressed in combination of other patterns (**Figure 3 I, II).**

The AUC for the refined score was 0.92 with true positive rate of 92% and false positive rate of 13%. The AUC for the original clinical score was 0.71 with 69% and 34% as true positive rates and false positive rates, respectively **(Supplementary Figure 4)**

Reproducibility of the refined HER2 IHC score

The degree of concordance between the score given on the original clinical settings and the re-defined score upon applying the existing HER2 scoring criteria was substantial (kappa=0.6). Exact score agreement was 79%, while number of discordant cases was 75/363 cases (20%), 47 of them were between IHC score 1+ vs. 0 and the remaining 28 discordant cases were between 1+ vs. 2+. None of the cases showed score 2+ vs. 0 discrepancy. Regarding the refined score, the intra-observer concordance showed perfect agreement (kappa=0.8) with 87% exact score agreement. Furthermore, the interobserver agreement was perfect (kappa=0.9) with 89% exact score agreement. Overall, there were 36/363 cases (10%) with discordance. **Tables 2** details the agreement levels. There was strong association between H score and HER2 IHC scores ($p<0.001$).

The association between various clinicopathological parameters and the refined HER2 score in comparison with the original score

Both HER2 scores showed statistically significant correlation with lymph node status and mRNA level **(Table 3**). The refined score, but not the original score, showed statistically significant correlation with Oncotype DX recurrence score ($p = 0.02$) where score 0 was associated with higher risk Oncotype DX groups. Moreover, there was statistically significant difference between HER2 IHC 1+ and 2+ using the refined score

regarding lymph node metastasis, where IHC 1+ was associated with lymph node metastasis. When compared to IHC score 0 cases, there was statistically significant correlation between HER2 low cluster and low tumour grade ($p<0.001$), lower pleomorphism score ($p=0.001$) and low mitotic count ($p<0.001$), less DCIS within tumour ($p<0.001$) and more lymph node metastasis ($p=0.03$) **(Supplementary Table 3).**

Within the external cohorts used, there was a significant correlation between *HER2* mRNA level and different HER2 IHC scores (from 0-3 and HER2 low only) and *HER2* gene copy number in TCGA and METABRIC cohorts with $p <0.001$ (**Supplementary Figures and 6).**

DISCUSSION

Accurate assessment of HER2 status is integral to the care of patients with BC. Recognizing this the ASCO/CAP HER2 working group released their guideline recommendations on HER2 testing in 2007, which were updated thereafter to provide a clearer guidance for HER2 testing and assessment.

At least 16 scenarios of HER2 expression patterns exist when considering the combination of staining intensity (faint, weak, moderate, and strong), membrane completeness (complete versus incomplete) and the cut-off (e.g.,10%) used to classify the percentage of HER2 in the invasive tumour cells into 2 main categories. However, not all the scenarios have been defined (see below) which in turn led to a degree of subjectivity and discordance in HER2 scoring. Some studies indicated that the concordance rates among pathologists remains low [15, 28-30], raising a concern regarding the need to refine the scoring criteria. Moreover, the distinction between HER2 IHC score 0 from score 1+, was not clinically relevant, and for practical purposes

these two groups have often been combined and/or used alternatively in routine practice. Fernandez et al. demonstrated that the current standard assays utilised in the clinical setting do not efficiently differentiate IHC scores 0 or 1+ and only 26% of these cases had 90% concordance agreement [16]. Also, Schettini and colleagues showed that multi-rater overall kappa score was 0.7, equivalent to substantial agreement, and almost half of the discordant cases were between IHC score 0 versus 1+ [31].

All previous attempts for the definition aimed at separating HER2 positive from HER2 negative BC for therapeutic and prognostic purposes [5, 6, 27, 32, 33] as patients with tumours that show a low or moderate levels of HER2 protein expression without confirmed gene amplification are currently not candidates to anti-HER2 agents [9]. This category, which accounts for 45-55% of BC, is known as HER2-low class of BC which include IHC score 1+ or 2+ with non-amplified *HER2* gene by ISH [34, 35]. With the promising response rate of ADC in HER2-low BC patients [12, 36-39], we hypothesized that refining the definition of HER2-low positive class with precise scoring criteria for this group will lead to better scoring concordance levels and better personalization of ADC therapy.

Borderline HER2-low BC can be demarcated from HER2 positive cases through gene amplification assays, but the lower limit of protein expression beyond which the tumour is considered HER2 negative is not fully identified. In this study, we aimed to refine the definition of different HER2 scoring categories through providing a clearer, easier, and applicable interpretation approach for different HER2 expression scenarios. We also sought to provide a definition for HER2-low positive BC through distinguishing HER2 IHC score 1+ from score 0 by using the mRNA expression as ground truth. The rationale behind using mRNA level to dichotomize our cases instead of the patient

outcome, is that at this low level of protein expression, *HER2* is not the driver oncogene and the clinical behaviour of the tumour and outcome is typically not dependent on the activation of the HER2 pathways [31, 40]. This was supported by Denkert et al. who demonstrated that there was no difference between HER2-low and HER2 negative tumours in the triple negative BC cohort [41].

Multiple studies show that rates of concordance for HER2 between core biopsy and excision specimens of 98% to 99% are achievable [42-44]. We have demonstrated that *HER2* mRNA was reliable in reflecting HER2 protein level both on core biopsy and full-face sections. Our results revealed that *HER2* mRNA is statistically significant in differentiating not only HER2 positive from negative BC, but also in HER2-low class where it can separate them into two distinct groups, and which are correlated with IHC protein level and gene amplification. Our study also showed that HER2 mRNA significantly correlates with HER2 protein and gene amplification levels supported by data from TCGA and METABRIC cohorts. This was supported in other studies that showed high concordance threshold between *HER2* mRNA and IHC and gene amplification [45-49]. The discrepancy between mRNA level and IHC score that was observed in few cases could be explained by intratumoural heterogeneity, and ratio of malignant to non-malignant cells within tumours, which can dilute the influence of the tumour cells on the result, leading to false-low mRNA level [39, 50]. While false high mRNA level in HER2 score 0 cases was mainly due to the presence of HER2 expression within the *in-situ*.

We had described ten possibilities for the HER2 expression patterns in BC tumour cells related to the staining intensity, localisation, and the circumferential staining completeness. Using a trained ANN model, we identified which pattern has the highest weight to differentiate HER2 score 1+ from score 0 based on the ground truth

represented by the mRNA level. We found that at faint intensity, the percentage of expression was more effective than membranous pattern of expression whether complete or incomplete. Based on our data, any faint HER2 protein expression in 20% or more can be considered as IHC score 1+. For weak staining, our results were consistent with the 2007 ASCO/CAP HER2 guidelines [27] and updated UK guidelines [5] in the definition of HER2 1+ (weak complete staining less than 10% and weak incomplete more than 10%, respectively).

The established algorithm for HER2 scoring according to ASCO/CAP guidelines, encompasses 10 out of the 16 possible scenarios for HER2 expression patterns. In this study, we tried to complete the missing HER2 expression possibilities based on the current study results, data from the various published HER2 scoring guideline recommendations, and our personal experience. Although most of these undefined scenarios are infrequent, like strong incomplete expression and moderate complete less than 10%, providing more objective criteria and adding more guidance to their scoring would improve the concordance rate among pathologists and consequently better HER2 categorization and management decision making.

To guarantee high inter-observer agreement, the magnification rule was also used to define faint staining which is areas showing barely visible expression defined as membranous staining confirmed only at x40, corresponding to faint intensity. This rule is applied and efficient in the assessment of HER2 in gastric carcinoma [25].

Although H score showed significant association with HJER2 scores, we did not include as a parameter to refine HER2 low definition. Histo score (H score) has been used for assessment of HER2 expression in previous studies, yet it is not approved for routine clinical [51-53]. The limitation of using H score is the non-linearity of the score which is due to the heavier weighting of higher intensity staining over lower intensity

staining to calculate the score. One more fallacy of using H score in assessment of HER2 expression is that it cannot address faint intensity and completeness of membranous staining. Thus, the H-score, which was designed as a standard scoring scheme to provide continuous scores, is not well suited for the scoring HER2 in BC [51] and would provide more unclarity to pathologists and clinicians.

Based on the refined score, the proportion of HER2 score 1+ cases decreased by 5% in comparison with the original ASCO/CAP definition that was used in the original scoring in the clinical setting. This could be explained by increasing the cut-off from the 10% used in clinical practice to 10% to 20% in the faint category. From this we can assume that there could be false increase in score 1+ category in the recent guidelines which may have affected the response rate for ADC in HER2 score 1+ BC patients. The refined score was more efficient in predicting HER2 score 1 + than the current applied score.

The inter-observer agreement between HER2 scores based on existing guidelines showed substantial concordance. This magnitude of concordance is in line with others reproducibility studies [16, 31]. Schettini and colleagues showed that multi-rater agreement was substantial and almost half of the discordant cases were between IHC score 0 versus IHC score 1+ [31]. Moreover, in the Phase 1b trastuzumab deruxtecan study, the concordance between local and central pathology was 70% for HER2 IHC score 1+ [12].

The inter- and intra-observer agreement for the two scoring sessions according to our refined criteria was near perfect with reduction of discordant cases between HER2 scores 1+ vs 0 by more than 70%. These results support the fact that current scoring criteria for HER2 scores 1+ and 0 are subjective and less reproducible among pathologists. Guidelines should be updated or refined to distinguish between HER2

scores 0 and 1+ especially in the upcoming era of ADC therapy. Recent studies revealed that 40% of patients with HER2-low BC achieved partial response to T-DXD [12, 54].

The refined score showed stronger association with the clinicopathological parameters than the current applied score. Also, it showed statistical significance with Oncotype DX scores. Our results agreed with both Schettini *et al.,* and Tan *et al.,* who declared that HER2-low BC is apparently more associated with axillary lymph node involvement compared to HER2 score 0 tumours [31, 55]**.** Overall, HER2 protein expression and mRNA level in IHC 1+ category was associated with low tumour grade, low mitotic count, special histological types of BC and low risk of recurrence based on Oncotype DX as described in other studies [31, 41, 55].

This study has some limitations including that the mRNA levels were measured on full face sections, whereas the IHC score was assessed on core biopsy. To overcome this issue, we selected cases with conflicting *HER2* mRNA expression and IHC scores and re-stained them on resection specimen blocks. The cohort had low number of outcome events in term of BC related deaths or disease recurrence, so outcome analysis and therapy effects were not feasible in this cohort. Therefore, we have used the mRNA level as our ground truth in classifying patients. . Due to the study design, the cohort did not include ER negative BC. However, this study aimed at refining the scoring of HER2 protein expression rather than assessing its oncogenic effect or its interaction with other proteins, thus we believe that the refined scoring criteria can be generalised and applied to ER negative tumours.

CONCLUSION

This is the first study to discuss refining the HER2-low positive BC focusing on the distinction between IHC score 1+ IHC score 0 to provide a more reproducible and non-arbitrary scoring criteria compared with the current definition which is more subjective. *HER2* mRNA level is strongly correlated with HER2 protein expression. Further investigations and clinical trials using ADC in HER2-low class BC using the refined criteria is warranted.

Acknowledgements

The authors are part of the PathLAKE digital pathology consortium. PathLAKE is one of the UK Government's 5 AI Centres of Excellence and has received in kind industry investment from Philips for digital pathology equipment, software, and other services. These new centres are supported by a £50m investment from the Data to Early Diagnosis and Precision Medicine strand of the government's Industrial Strategy Challenge Fund, managed and delivered by UK Research and Innovation (UKRI). N.A. is supported and funded by the Egyptian Ministry of Higher Education and Scientific Research.

Conflict of interest: The authors declare no conflict of interest

Ethics Approval and Consent to Participate: Ethical Approval was obtained for this study and approved by REC (REF. no. 19/SC/0363) under the title "PathLAKE". All cases included in the study were fully anonymized.

Author Contributions:

Concept and Design: N.A., E.R., Collection of Data: N.A., M.T., Identifying and reviewing the images: N.A., M.T., Analysis of Data: N.A., M.T., E.R., Interpretation of Results: All Authors, Writing and Reviewing Manuscript: All Authors

All authors reviewed and agreed on the final revision of the manuscript.

Funding: N.A. is funded by the Egyptian Ministry of Higher Education and Scientific Research

Data Availability: All data used in this study are available and can be accessed upon reasonable request. The following publicly available datasets were used on

https://identifiers.org/cbioportal:brca_tcga;

https://identifiers.org/cbioportal:brca_metabric

REFERENCES

1. Marchiò C, Annaratone L, Marques A, Casorzo L, Berrino E, Sapino A, editors. Evolving concepts in HER2 evaluation in breast cancer: Heterogeneity, HER2-low carcinomas and beyond. Seminars in cancer biology; 2021: Elsevier.

2. Harbeck N, Gnant M. Interpretation of the evidence for the efficacy and safety of statin therapy. Lancet. 2017;389(10074):1134-50.

3. Appert-Collin A, Hubert P, Crémel G, Bennasroune A. Role of ErbB receptors in cancer cell migration and invasion. Frontiers in pharmacology. 2015;6:283.

4. Ali R, Wendt MK. The paradoxical functions of EGFR during breast cancer progression. Signal transduction and targeted therapy. 2017;2(1):1-7.

5. Rakha EA, Pinder SE, Bartlett JM, Ibrahim M, Starczynski J, Carder PJ, et al. Updated UK Recommendations for HER2 assessment in breast cancer. Journal of clinical pathology. 2015;68(2):93-9.

6. Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JM, et al. Human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline focused update. Archives of pathology & laboratory medicine. 2018;142(11):1364-82.

7. Pagliarini R, Shao W, Sellers WR. Oncogene addiction: pathways of therapeutic response, resistance, and road maps toward a cure. EMBO reports. 2015;16(3):280-96.

8. Loibl S, Gianni L. HER2-positive breast cancer. The Lancet. 2017;389(10087):2415-29.

9. Fehrenbacher L, Cecchini RS, Geyer Jr CE, Rastogi P, Costantino JP, Atkins JN, et al. NSABP B-47/NRG oncology phase III randomized trial comparing adjuvant chemotherapy with or without trastuzumab in high-risk invasive breast cancer negative for HER2 by FISH and with IHC 1+ or 2+. Journal of Clinical Oncology. 2020;38(5):444.

10. Modi S, Jacot W, Yamashita T, Sohn J, Vidal M, Tokunaga E, et al. Trastuzumab Deruxtecan in Previously Treated HER2-Low Advanced Breast Cancer. N Engl J Med. 2022;387(1):9-20.

Formatted: English (UK)

11. Rinnerthaler G, Gampenrieder SP, Greil R. HER2 directed antibody-drug-conjugates beyond T-DM1 in breast cancer. International journal of molecular sciences. 2019;20(5):1115.

12. Modi S, Park H, Murthy RK, Iwata H, Tamura K, Tsurutani J, et al. Antitumor activity and safety of trastuzumab deruxtecan in patients with HER2-low–expressing advanced breast cancer: results from a phase Ib study. Journal of Clinical Oncology. 2020;38(17):1887.

13. Zhang X, Huang AC, Chen F, Chen H, Li L, Kong N, et al. Novel development strategies and challenges for anti-Her2 antibody-drug conjugates. Antibody Therapeutics. 2022;5(1):18-29.

14. Saura C, Thistlethwaite F, Banerji U, Lord S, Moreno V, MacPherson I, et al. A phase I expansion cohorts study of SYD985 in heavily pretreated patients with HER2-positive or HER2-low metastatic breast cancer. Journal of Clinical Oncology. 2018;36(15_suppl):1014-.

15. Perez EA, Suman VJ, Davidson NE, Martino S, Kaufman PA, Lingle WL, et al. HER2 testing by local, central, and reference laboratories in specimens from the North Central Cancer Treatment Group N9831 intergroup adjuvant trial. Journal of Clinical Oncology. 2006;24(19):3032-8.

16. Fernandez A, Liu M, Bellizzi A. Examination of low HER2 expression in breast cancer. Presented at: 2021 San Antonio Breast Cancer Symposium; December 7-10, 2021; San Antonio, TX. Abstract P1-02-02; 2021.

17. Grimm EE, Schmidt RA, Swanson PE, Dintzis SM, Allison KH. Achieving 95% cross-methodological concordance in HER2 testing: causes and implications of discordant cases. American journal of clinical pathology. 2010;134(2):284-92.

18. Jamarani SM, Behnam H, Rezairad G, editors. Multiwavelet based neural network for breast cancer diagnosis. GVIP 05 Conference; 2005.

19. Chae H-D, Park CM, Park SJ, Lee SM, Kim KG, Goo JM. Computerized texture analysis of persistent part-solid ground-glass nodules: differentiation of preinvasive lesions from invasive pulmonary adenocarcinomas. Radiology. 2014;273(1):285-93.

20. Swathi S, Kumar PS, Sarma P. Approach of Jordan Elman Neural Network to Diagnose Breast Cancer on Three Different Data Sets. International Journal of Advanced Research in Computer and Communication Engineering. 2015;4(11).

21. Puri M, Lloyd M, Bui M. Role of an Artificial Neural Network Classifier in Nuclear Pleomorphic Feature Analysis of Histopathological Images of Breast Cancer. Artificial Neural Network for Drug Design, Delivery and Disposition: Elsevier; 2016. p. 377-91.

22. Saranya G, Pravin A. A comprehensive study on disease risk predictions in machine learning. International Journal of Electrical and Computer Engineering. 2020;10(4):4217.

23. Baehner FL. The analytical validation of the Oncotype DX Recurrence Score assay. ecancermedicalscience. 2016;10.

24. Allison KH, Hammond MEH, Dowsett M, McKernin SE, Carey LA, Fitzgibbons PL, et al. Estrogen and Progesterone Receptor Testing in Breast Cancer: ASCO/CAP Guideline Update. J Clin Oncol. 2020;38(12):1346-66.

25. Rüschoff J, Hanna W, Bilous M, Hofmann M, Osamura RY, Penault-Llorca F, et al. HER2 testing in gastric cancer: a practical approach. Modern Pathology. 2012;25(5):637-50.

26. Wolff AC, Hammond MEH, Hicks DG, Dowsett M, McShane LM, Allison KH, 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. Archives of Pathology and Laboratory Medicine. 2014;138(2):241-56.

27. Wolff AC, Hammond MEH, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. Archives of pathology & laboratory medicine. 2007;131(1):18-43.

28. Dybdal N, Leiberman G, Anderson S, McCune B, Bajamonde A, Cohen RL, et al. Determination of HER2 gene amplification by fluorescence in situ hybridization and concordance with the clinical trials immunohistochemical assay in women with

metastatic breast cancer evaluated for treatment with trastuzumab. Breast cancer research and treatment. 2005;93(1):3-11.

29. Paik S, Bryant J, Tan-Chiu E, Romond E, Hiller W, Park K, et al. Real-world performance of HER2 testing—national surgical adjuvant breast and bowel project experience. Journal of the National Cancer Institute. 2002;94(11):852-4.

30. Roche PC, Suman VJ, Jenkins RB, Davidson NE, Martino S, Kaufman PA, et al. Concordance between local and central laboratory HER2 testing in the breast intergroup trial N9831. Journal of the National Cancer Institute. 2002;94(11):855-7.

31. Schettini F, Chic N, Brasó-Maristany F, Paré L, Pascual T, Conte B, et al. Clinical, pathological, and PAM50 gene expression features of HER2-low breast cancer. NPJ Breast Cancer. 2021;7(1):1-13.

32. Carlson RW, Moench SJ, Hammond MEH, Perez EA, Burstein HJ, Allred DC, et al. HER2 testing in breast cancer: NCCN Task Force report and recommendations. Journal of the National Comprehensive Cancer Network. 2006;4(S3):S-1-S-22.

33. Ellis I, Bartlett J, Dowsett M, Humphreys S, Jasani B, Miller K, et al. Best Practice No 176: Updated recommendations for HER2 testing in the UK. Journal of clinical pathology. 2004;57(3):233-7.

34. Tarantino P, Hamilton E, Tolaney SM, Cortes J, Morganti S, Ferraro E, et al. HER2-low breast cancer: pathological and clinical landscape. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2020;38(17):1951-62.

35. Agostinetto E, Rediti M, Fimereli D, Debien V, Piccart M, Aftimos P, et al. HER2-Low Breast Cancer: Molecular Characteristics and Prognosis. Cancers. 2021;13(11):2824.

36. Banerji U, van Herpen CM, Saura C, Thistlethwaite F, Lord S, Moreno V, et al. Trastuzumab duocarmazine in locally advanced and metastatic solid tumours and HER2-expressing breast cancer: a phase 1 dose-escalation and dose-expansion study. The Lancet Oncology. 2019;20(8):1124-35.

37. Clifton GT, Hale D, Vreeland TJ, Hickerson AT, Litton JK, Alatrash G, et al. Results of a randomized phase IIb trial of nelipepimut-S+ trastuzumab versus

trastuzumab to prevent recurrences in patients with high-risk HER2 low-expressing breast cancer. Clinical Cancer Research. 2020;26(11):2515-23.

38. Nakada T, Sugihara K, Jikoh T, Abe Y, Agatsuma T. The latest research and development into the antibody–drug conjugate,[fam-] trastuzumab deruxtecan (DS-8201a), for HER2 cancer therapy. Chemical and Pharmaceutical Bulletin. 2019;67(3):173-85.

39. Pernas S, Tolaney SM. HER2-positive breast cancer: new therapeutic frontiers and overcoming resistance. Therapeutic advances in medical oncology. 2019;11:1758835919833519.

40. Krishnamurti U, Silverman JF. HER2 in breast cancer: a review and update. Advances in anatomic pathology. 2014;21(2):100-7.

41. Denkert C, Seither F, Schneeweiss A, Link T, Blohmer J-U, Just M, et al. Clinical and molecular characteristics of HER2-low-positive breast cancer: pooled analysis of individual patient data from four prospective, neoadjuvant clinical trials. The Lancet Oncology. 2021;22(8):1151-61.

42. Lebeau A, Turzynski A, Braun S, Behrhof W, Fleige B, Schmitt WD, et al. Reliability of human epidermal growth factor receptor 2 immunohistochemistry in breast core needle biopsies. J Clin Oncol. 2010;28(20):3264-70.

43. Lee AH, Key HP, Bell JA, Hodi Z, Ellis IO. Concordance of HER2 status assessed on needle core biopsy and surgical specimens of invasive carcinoma of the breast. Histopathology. 2012;60(6):880-4.

44. Rakha EA, Pigera M, Shaaban A, Shin SJ, D'Alfonso T, Ellis IO, et al. National guidelines and level of evidence: comments on some of the new recommendations in the American Society of Clinical Oncology and the College of American Pathologists human epidermal growth factor receptor 2 guidelines for breast cancer. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2015;33(11):1301-2.

45. De P, Smith BR, Leyland-Jones B. Human epidermal growth factor receptor 2 testing: where are we? : American Society of Clinical Oncology; 2010. p. 4289-92.

46. Noske A, Loibl S, Darb-Esfahani S, Roller M, Kronenwett R, Müller B, et al. Comparison of different approaches for assessment of HER2 expression on protein

and mRNA level: prediction of chemotherapy response in the neoadjuvant GeparTrio trial (NCT00544765). Breast cancer research and treatment. 2011;126(1):109-17.

47. Baehner FL, Achacoso N, Maddala T, Shak S, Quesenberry Jr CP, Goldstein LC, et al. Human epidermal growth factor receptor 2 assessment in a case-control study: comparison of fluorescence in situ hybridization and quantitative reverse transcription polymerase chain reaction performed by central laboratories. Journal of Clinical Oncology. 2010;28(28):4300-6.

48. Susini T, Bussani C, Marini G, Nori J, Olivieri S, Molino C, et al. Preoperative assessment of HER-2/neu status in breast carcinoma: the role of quantitative real-time PCR on core-biopsy specimens. Gynecologic oncology. 2010;116(2):234-9.

49. Dabbs DJ, Klein ME, Mohsin SK, Tubbs RR, Shuai Y, Bhargava R. High false-negative rate of HER2 quantitative reverse transcription polymerase chain reaction of the Oncotype DX test: an independent quality assurance study. J Clin Oncol. 2011;29(32):4279-85.

50. Shafi H, Astvatsaturyan K, Chung F, Mirocha J, Schmidt M, Bose S. Clinicopathological significance of HER2/neu genetic heterogeneity in HER2/neu non-amplified invasive breast carcinomas and its concurrent axillary metastasis. Journal of Clinical Pathology. 2013;66(8):649-54.

51. Potts SJ, Krueger JS, Landis ND, Eberhard DA, Young GD, Schmechel SC, et al. Evaluating tumor heterogeneity in immunohistochemistry-stained breast cancer tissue. Laboratory Investigation. 2012;92(9):1342-57.

52. Saura C, Matito J, Oliveira M, Wildiers H, Brufksy AM, Waters SH, et al. Biomarker Analysis of the Phase III NALA Study of Neratinib+ Capecitabine versus Lapatinib+ Capecitabine in Patients with Previously Treated Metastatic Breast Cancer. Clinical Cancer Research. 2021;27(21):5818-27.

53. Kim EK, Kim KA, Lee CY, Shim HS. The frequency and clinical impact of HER2 alterations in lung adenocarcinoma. PLoS One. 2017;12(2):e0171280.

54. Modi S, Saura C, Yamashita T, Park YH, Kim S-B, Tamura K, et al. Trastuzumab deruxtecan in previously treated HER2-positive breast cancer. New England Journal of Medicine. 2020;382(7):610-21.

55. Tan RSYC, Ong WS, Lee K-H, Lim AH, Park S, Park YH, et al. HER2 expression, copy number variation and survival outcomes in HER2-low non-metastatic breast cancer: an international multicentre cohort study and TCGA-METABRIC analysis. BMC Medicine. 2022;20(1):105.

Figure Legends

Figure1: Flowchart summarizing cohort selection and different steps carried out.

Figure 2 A: Schematic illustration of different scenarios of HER2 expression in breast cancer and the corresponding score from different existing guidelines**. Figure 2 B:** Recommended HER2 scoring algorithm based on immunohistochemistry (IHC) stained slides.

Figure 3 I: An illustrated diagram and photomicrographs showing different HER2 membranous staining patterns and intensities; Faint complete **(1, A),** Faint incomplete **(2, B),** Weak complete **(3, C)**, Weak incomplete **(4, D),** Moderate complete **(5, E),** Moderate incomplete **(6, F). Figure 3 II:** Graphical description highlighting degree of Intratumoural heterogeneity within HER2 low category. A: HER2 IHC stained slide showing different staining intensities within the same tumour, B: Pie chart showing that 60% of HER2 1+ cases were scored based on heterogenous mixed expression patterns, while only 40% were scored based on single homogenous staining. Moderate staining was present in addition to other staining patterns and not alone in HER2 1+ category. Sparkline graphs show multiple combinations of heterogenous patterns in example of HER2 score 1+ (C), and another example of HER2 score 2+ (D).