Title

Quantification of HER2-low and ultra-low expression in breast cancer specimens by quantitative IHC and artificial intelligence

Authors

Frederik Aidt[1*], Elad Arbel[2], Itay Remer[2], Oded Ben-David[2], Amir Ben-Dor[2], Daniela Rabkin[2], Kirsten Hoff[1], Karin Salomon[1], Sarit Aviel-Ronen[3, **], Gitte Nielsen[4], Jens Mollerup[1], Lars Jacobsen[1], Anya Tsalenko[5]

Author affiliations

[1]Agilent Technologies, Glostrup, Denmark

[2]Agilent Research Laboratories, Tel Aviv, Israel

[3]Adelson School of Medicine, Ariel University, Ariel, Israel

[4]Agilent Technologies, Carpinteria, CA, USA,

[5]Agilent Research Laboratories, Santa Clara, CA, US

*Corresponding author: Frederik Aidt, Agilent Technologies Denmark ApS, Produktionsvej 42, 2600 Glostrup, Denmark. Phone: +45 8830 5061; E-mail: frederik.aidt@agilent.com

**Contract for Agilent Research Laboratories, Tel Aviv, Israel

Abstract

Recent advancements in antibody drug conjugate therapies have significantly broadened treatment options for a substantial subgroup of breast cancer patients. However, sensitive, accurate and quantitative evaluation of HER2 expression based on current immunohistochemistry (IHC) assays remain challenging, especially in low and ultra-low HER2 expression ranges.

We developed a novel methodology for quantifying HER2 protein expression, targeting breast cancer cases in the HER2 IHC 0 and 1+ categories. We measured HER2 expression using quantitative IHC (qIHC) that enables precise and tunable HER2 detection across different expression levels as demonstrated in formalin-fixed paraffin-embedded cell lines**.** Additionally, we developed an AI-based interpretation of HercepTest™ mAb pharmDx (Dako Omnis) (HercepTest™ mAb) using qIHC measurements as the ground truth. Both methodologies allowed spatial resolution and visualization of low and ultra-low levels of HER2 expression across entire tissue sections to demonstrate and enable quantification of heterogeneity of HER2 expression.

Serial sections of 82 formalin-fixed paraffin-embedded tissue blocks of invasive breast carcinoma with HER2 IHC scores 0 or 1+ were stained with H&E, HercepTest™ (mAb), qIHC and p63, then scanned and digitally aligned. Tumor areas were manually selected and reviewed by expert pathologists. HER2 expression was quantitatively evaluated based on the qIHC assay in each 128x128µm^2 area within tumor regions. We observed statistically significant differences in HER2 expression between IHC 0, 0<IHC<1+, and IHC 1+ groups, and high level of spatial heterogeneity of the HER2 expression levels within the same tissue, up to five-fold in some cases. We demonstrated high slide-level tumor region agreement of estimates of HER2 expression between the AI-based interpretation of HercepTest™ mAb and the qIHC ground truth with a Pearson correlation of 0.94, and R^2 of 0.87.

The developed methodologies can be used to stratify HER2 low-expression patient groups, potentially improving the interpretation of IHC assays and maximizing therapeutic benefits. This method can be implemented in histology labs without requiring a specialized workflow.

Introduction

Alongside FDA-approval of the anti-HER2 antibody drug trastuzumab (Herceptin®) for treatment of HER2 positive breast cancer in 1998, HercepTest™ was approved as an aid in the assessment of patients where trastuzumab treatment was considered [1]. The HercepTest™ approval included a scoring system of HER2 expression in breast cancer specimens with negative IHC scores defined as IHC 0 and 1+, weakly positive as IHC 2+ and strongly positive as IHC 3+ [1]. Due to the more recent introduction of potent antibody drug conjugates (ADCs), such as trastuzumab deruxtecan (T-DXd; Enhertu®), a clinically relevant HER2-low group has been suggested as HER2 IHC 1+ or IHC 2+ without *HER2* gene amplification and a HER2-ultra-low group representing a subset of IHC 0 cases with incomplete and faint staining in ≤10% of tumor cells has also been suggested [2](Figure 1).

Recent data from the DESTINY-Breast04 trial (NCT03734029) showed that metastatic breast cancer cases with low levels of HER2 expression were clinically actionable. Subsequently treatment with T-DXd was approved in the US in August 2022 and in the EU in January 2023 for HER2-low breast cancer patients with unresectable or metastatic disease [3]. In a recent press release from the ongoing DESTINY-Breast06 trial (NCT04494425), a significant and clinically meaningful effect from T-DXd was also reported for metastatic breast cancer patients with ultra-low HER2 expression [4]. Thus, the distinction between low levels of HER2 expression that was previously not clinically meaningful [3], has become critical today.

Identification of patients exhibiting low or ultra-low HER2 expression levels in tumor cell membranes represents a nontrivial challenge in the pathology laboratory. Currently approved semiquantitative HER2 IHC assays were designed to differentiate between cases characterized as IHC 1+ and 2+, because this aids in the identification of patients with HER2 positive staining that could benefit from treatment with trastuzumab. The complexity in detecting low levels of HER2 expression relevant for T-DXd treatment is further challenged by high interobserver variability [5-8], although it has been reported that relatively short pathologist training sessions may lead to acceptable accuracy in identifying HER2 negative (IHC 0) and HER2-low samples [9].

It has been suggested that quantitative HER2 expression information could be retrieved from orthogonal methods such as those evaluating tissue RNA expression [10, 11], HER2 protein by mass spectrometry or quantitative immunofluorescence [12]. Recently, HER2 protein detected by immunofluorescence, and HER2 RNA correlated well to HER2 IHC scores from two FDA-approved HER2 IHC assays [10]. However, this study did not show a difference between HER2 IHC classes characterized as 0, ultralow or 1+ with respect to HER2 protein and HER2 RNA levels, and the lack of discrimination of HER2 expression in the IHC 0, ultralow, and 1+ groups were attributed to suboptimal sensitivity of the HER2 IHC assays in this range.

Using immunofluorescence-based detection, relative HER2 score levels were obtained from breast cancer specimens of a tissue microarray and was also successfully used to provide prognostic information of HER2 cases with low HER2 expression [13]. In a more recent study, HER2 protein was quantified by liquid chromatography mass spectrometry (LC-MS) in cell lines and correlated to HER2 immunofluorescence of the same cell lines enabling conversion of the immunofluorescence signal in a breast cancer tissue microarray to actual HER2 protein levels [12]. However, despite their potential to provide quantitative HER2 expression information, these techniques have not been widely adopted as standard practice in pathology and may require additional equipment. In addition, immunofluorescence signals are relative and can be influenced by factors such as excitation intensity and acquisition settings, affecting consistency and reproducibility in clinical applications.

A different, sensitive and quantitative IHC (qIHC) methodology that determines HER2 expression in FFPE tissue was presented by Jensen et al. [14]. This method differs from others as it generated countable chromogenic dots based on improved catalyzed reporter deposition (iCARD) [15] to reflect HER2 protein levels in tissue sections or cell lines with less influence from IHC staining variation. This assay was demonstrated to enable sensitive, reproducible, and quantitative HER2 determination with a lower limit of detection than observed in traditional IHC staining. qIHC measurement of the biomarker alkylpurine-DNA-N-glucosylase has previously been shown to demonstrate prognostic value in glioblastoma multiforme. Patients with high expression of this biomarker detected by qIHC were shown to have improved overall survival [16].

Whole slide image (WSI) analysis supported by artificial intelligence (AI) can provide relevant aid in interpretation of stained breast cancer specimens and reduce inter-observer variation when interpreting HER2 levels [17, 18] and in particular with heterogenous samples or challenging HER2-low cases [19, 20]. It has also been demonstrated that whole slide deep learning based image analysis enables patient stratification with significant improvement in progression-free survival and a higher prevalence of patient selection compared to the manual scoring method based on ASCO/CAP guidelines [21]. However, in studies involving deep learning algorithms, a consensus method among several specialist pathologists is typically employed to provide the ground truth for algorithm development [19, 20]. This method remains especially challenging in the HER2-low and ultra-low signal ranges as the staining may be faint and barely perceptible and/or represent reactivity in less than 10% of tumor cells.

In this work, we developed a novel methodology for quantifying HER2 protein expression targeting breast cancer cases in the HER2 IHC 0 and 1+ categories. We estimated HER2 expression based on the analysis of qIHC – stained tissue slides using a modified version of the qIHC method by Jensen et al. [14]. As first suggested by Jensen et al. [14] we now demonstrate that qIHC can be used as a tunable system, where the qIHC dot count is directly proportional to the concentration to the dot-generating component. We used the qIHC based measurements to obtain quantitative ground truth information of available HER2 epitopes in 82 formalin-fixed paraffin embedded (FFPE) breast cancer specimens and used it to develop AI-based interpretation of WSI of HercepTest™ mAb pharmDx (Dako Omnis) (HercepTest™ mAb) stained slides. Unlike other studies that rely on ground truth HER2 expression from domain experts that could face challenges due to suboptimal interobserver agreement and scalability, the use of qIHC for ground truth HER2 expression allowed us to use objective HER2 expression determinations from the HER2-null, HER2-ultra-low and HER2-low groups to overcome these challenges during AI-model training. We observed statistically significant HER2 expression by qIHC between these groups and we demonstrated the capability of final algorithms to infer heterogeneous HER2 expression spatially across the entire breast cancer tissue section in graphical overlays of stained tissues, and to visualize low levels of HER2 expression that are challenging to detect by the human eye.

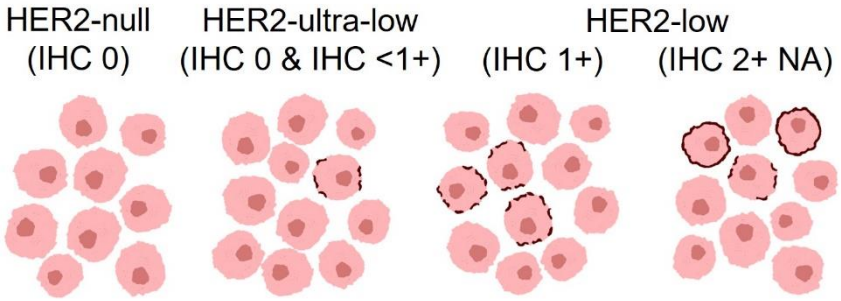

Figure 1. Definitions of HER2 negative tumors used in this study: HER2-null with no membrane staining; HER2-ultra-low with weak membrane staining in less than 10% of tumor cells; and HER2-low with very weak to moderate membrane staining in more than 10% of tumor cells and non-amplified (NA) *HER2* ISH status. These definitions align with the European Society for Medical Oncology consensus statements presented by [22].

Material and Methods

Tissue and Cell Lines

A total of 151 formalin-fixed, paraffin-embedded (FFPE) breast cancer specimens, encompassing both resections and biopsies were screened for this study. From these, 82 specimens having solid tumor areas and characterized as HER2 IHC 0 or 1+ by ASCO/CAP guidelines were selected for algorithm development. One HER2 IHC 0 sample had small 2+ areas and was excluded from group comparison in Figure 4B. According to the criteria outlined in Figure 1A, these breast cancer specimens represent HER2-null (n=15), HER2-ultra-low (n=26), and HER2-low (n=41). Cases without invasive cancer or with very sparse tumor representation were not included. All specimens in the study were residual, de-identified, originated from different individuals, and were procured from commercial sources or local hospitals with no possibility for patient identification. The study was performed in agreement with the World Medical Association Declaration of Helsinki (2013) and no study protocol was submitted for Ethics Committee approval as this type of analytical study is exempt from Ethics Committee approval in Denmark. FFPE blocks containing cell lines (CCLs) MDA-MB-468,

MDA-MB-231, MDA-MB-175, MDA-MB-453, and SK-BR-3 used in this study were originally described by Jensen et al. [14].

Sectioning

Sections were cut at 4 µm on a Leica microtome (RM2255, Leica Biosystems, Wetzlar, Germany), mounted on K8020 FLEX IHC Microscope Slides (Agilent Technologies, Glostrup, Denmark) and baked for 1 hour at 60 °C. Six sequential and numbered sections from the breast cancer specimens were cut to ensure spatial proximity between the different staining results (see Figure S1).

Staining of Tissue Sections

Staining with H&E (CS700 and CS701, Agilent Technologies, Glostrup, Denmark) was performed according to instructions by the manufacturer on a Dako Coverstainer (Agilent Technologies, Glostrup, Denmark). Slides were stained with HercepTest™ mAb pharmDx (Dako Omnis) (GE001, Agilent Technologies, Glostrup, Denmark) or monoclonal mouse-anti human p63 Protein Clone DAK-p63 (Dako Omnis) (GA662, Agilent Technologies, Glostrup, Denmark), according to the manufacturer's instructions. Following staining, slides were dehydrated in ethanol, transferred to xylene, and then coverslipped on a Sakura TissueTek coverfilm coverslipper (Sakura, Osaka, Japan).

The principle of HER2 qIHC staining that generates countable dots based on the iCARD technology is schematically illustrated in Figure 2A. The staining was essentially performed as previously described by Jensen et al. [14,] Lohse et al. [15] on an Autostainer Link 48 (Agilent Technologies, Glostrup, Denmark) with a few modifications. The highly sensitive monoclonal rabbit anti-HER2 antibody (DG44) from HercepTest™ mAb was used as the primary antibody in Step 1 at a concentration of 1.5 µg/mL. A polymer consisting of polyclonal goat anti-rabbit antibodies conjugated to a dextran backbone with multiple horseradish peroxidase (HRP) enzymes and a fixed concentration of 3 nM unlabeled goat anti-rabbit antibodies was used in Step 2. A fluorescein containing substrate reporter [15] was precipitated as shown in Step 3, and an HRP-conjugated F(ab') antibody recognizing the fluorescein substrate in Step 4 was used to generate DAB-based qIHC dots in Step 5. In this study we also varied the concentration of the HRP-conjugated dextran polymer in Step 2 to control the qIHC dot density. For negative control staining, the primary antibody was omitted from the staining process. Slides were then dehydrated and coverslipped as described for IHC staining.

Whole Slide Image Scanning

Scanning was performed on a brightfield Ultra Fast Scanner (Philips, Eindhoven, The Netherlands) at a resolution of 0.25 µm/pixel. WSI images were exported from the Philips IMS in Philips TIFF format (quality 80) for further processing. Additionally, qIHC slides for dot detection model training were scanned on an Axioscan Z.1 with Zen 3.4 acquisition software (Carl Zeiss, Oberkochen, Germany) at a resolution of 0.22 µm/pixel. For each slide, a z-stack of 5 planes was captured with the middle plane being the focus depth and with 1 µm intervals between planes to cover the entire tissue thickness. Extended depth of focus (EDF) using the contrast method in the acquisition software was used to flatten focus in the 5 planes to a single plane.

Alignment of Whole Slide Image Scans

All WSI scans (Figure S1) were aligned to the HercepTest™ mAb WSI scan of the same specimen. Alignment was performed using an in house developed algorithm based on rigid transformations, incorporating only lateral image translation and rotation, to maintain the physical dimensions of tissue sections.

Tumor Region Annotations on WSI Scans

H&E-stained slides were utilized to identify tumors showing solid growth pattern (solid tumor regions), non-tumor areas, carcinoma *in situ* (CIS), and necrotic areas (Figure S2A). Solid tumor regions were included when an estimated ≥ 80% of the cells were invasive tumor. Large areas populated with lymphocytes, stroma, fat cells, normal ducts, necrotic areas, or areas out of focus within the solid tumor regions were excluded. CIS exclusion was guided by p63 immunostained sections. Tumor regions were manually annotated on HercepTest™ mAb stained breast carcinoma sections (Figure S2B) with the definitions HER2-null, HER2-ultra-low and HER2-low as described in Figure 1, and a HER2 score was assigned based on the ASCO/CAP guideline for the entire specimen [23, 24]. Tissue folds that could affect qIHC dot detection were annotated on the qIHC WSI scan for exclusion (Figure S2C). aided in identifying normal ducts and CIS (Figures S2D).

Detection of qIHC Dots

An established UNET AI-model used for nuclei detection and segmentation [25, 26] was adapted to detect qIHC dots by identifying their centroid proximity. The model was iteratively trained using selected fields of view from HER2 qIHC stained WSIs scanned on the Ultra Fast Philips scanner with manually marked dots. To enhance accuracy of qIHC dot detection, we used EDF WSIs of the same fields of views scanned on the Zeiss Axioscan Z.1 to improve manual annotation of faint and out-of-focus qIHC dots (Supplementary Figure S3). In addition, qIHC negative controls, i.e., slides stained without the HER2 antibody were used as additional training samples to improve the capability of distinguishing noisy patterns and avoid false-positive dots

qIHC Patch Count Analysis

Patches for training and evaluation of the AI-model were defined around every single nucleus in annotated tumor regions [25, 26]. Nuclei were used as patch centers to ensure preferential sampling of patches that contained cells. A patch size of 128 µm × 128 µm, equal to 512 × 512 pixels with a resolution of 0.25 pixel/µm, was used to balance spatial resolution and statistical stability of dot densities. Slide qIHC count for qIHC stained sections was calculated by averaging the patch qIHC dot counts from all patches in annotated tumor areas. For heatmap visualization in Figure 4, patch qIHC counts were computed for patches centered around each pixel in the visualization. For heatmap visualizations in Figure 5, patches were selected on a 32 µm × 32 µm grid, leading to patch overlap of 75%

Statistical Analysis

Statistical analyses were performed using the Pingouin version 0.5.4 Python statistical package. The Kruskal-Wallis H-test, a non-parametric one-way ANOVA, was used to compare medians of slide qIHC counts between slides with different overall IHC scores. Where the Kruskal-Walli's test indicated significance, the Wilcoxon rank-sum post hoc test, with Holm P value adjustment for multiple testing was used to test for significant differences between each pair of overall IHC score groups. Comparison of average slide qIHC counts of HER2-null slides stained with anti-HER2 antibody and negative controls was done using a paired, two-sided t-test with Welch correction for unequal variances (Figure 4B). A single specimen was removed from the qIHC count analysis due to a discrepant HER2 IHC score.

AI-Model Development for Predicting HER2 Expression

The method for developing AI-models from the standard ResNet50 architecture [27] included training to quantitatively predict the patch qIHC dot count in the HercepTest™ mAb stained image patches as shown in Figure 5A. The ground truth HER2 expression for model training was determined by the HER2 qIHC dot count detection on sequentially stained sections and aligned to the HercepTest™ mAb stained section. To investigate performance of the AI-model development method, k-fold cross-

validation analysis was applied splitting the complete specimen dataset into eight non-overlapping subsets, each containing 72 (or 71) training specimens and 10 (or 11) test specimens (see Figure S5). Following development of a ResNet50 model with training specimens of one subset, this AI-model was then evaluated with the test specimens not used for training of this model. Thereby k-fold cross-validation ensured that variation in all sections were used for testing and proved the overall applicability of our AI-model development method. Within each training subset, tumor regions in each specimen were further divided into non-overlapping training and validation regions creating approximately 80%/20% train-validation non-overlapping image patch samples. The patches sampled were balanced to reflect patch qIHC dot count distribution and specimen representation during training. Model training optimization was performed using the Adam optimizer and guided by mean-squared-error (MSE) loss, with a reduce-on-plateau learning rate scheduler and early stopping criterion, both informed by validation MSE loss.

Results

qIHC quality compared with traditional HER2 IHC

To illustrate the difference of a HER2 qIHC staining versus a traditional HER2 IHC staining, Figure 2B has one example of an FFPE breast cancer cell from the MDA-MB-453 cell line stained by IHC using HercepTest™ mAb and one cell stained by qIHC. In this example the HercepTest™ mAb staining generates a strong DAB staining on the cell membrane, whereas HER2 qIHC staining results in two dots located on the cell membrane and cytosol. The dots generated by qIHC are often in the vicinity of the cell membrane, but due to the size of dots and the nature of the dot generating process, they may also locate to other cellular parts as seen here.

qIHC Cell Line Results

Different concentrations of the HRP-conjugated polymer to detect the primary anti-HER2 antibody and investigate the number of HER2 qIHC dots generated with different HER2 expression levels in five breast cancer cell lines (Figure 2C and Figure S4). These cell lines exhibit HER2 expression levels from 1.5×10^3 to 5.6×10^5 HER2 receptors/cell and have HER2 IHC scores from 0 to 3+ [14]. Figure 2C illustrates the correlation between number of qIHC dots per cell and the concentration of the HRP-conjugated polymer. It is observed that doubling of the polymer concentration leads to doubling of the number of qIHC dots. For the three cell lines MDA-MB-468, MDA-MB-231, and MDA-MB-175 with less HER2 expression, a linear relationship across the tested polymer concentration range, from 0.125 pM to 2 pM and number of qIHC dots was seen. At low polymer concentrations the two cell lines MDA-MB-453 and SK-BR-3 with higher HER2 expression, also had a linear relationship between polymer concentration and qIHC dots, but at concentrations of 1 pM and 2 pM this linearity was lost. This is due to a saturating dot density as dots begin to overlap and the dot detection algorithm cannot estimate actual number of dots. The unspecific dots from the qIHC staining protocol at five different polymer concentrations is also illustrated in Figure 2C, where red dots represent qIHC staining in which the HER2 antibody has been replaced by the negative control reagent at increasing concentrations of the polymer. The mean number of dots per cell is similar across all tested cell lines and increases with increasing polymer concentration. The image patches in Figure 2D illustrate how the dots in the SK-BR-3 cell line begin to overlap at high polymer concentrations preventing detection of single dots. Overall, the qIHC results presented in Figure 2 demonstrates the scalability of the qIHC method and how an improved signal-to-noise level can be achieved by trimming the concentration of the HRP-conjugated polymer.

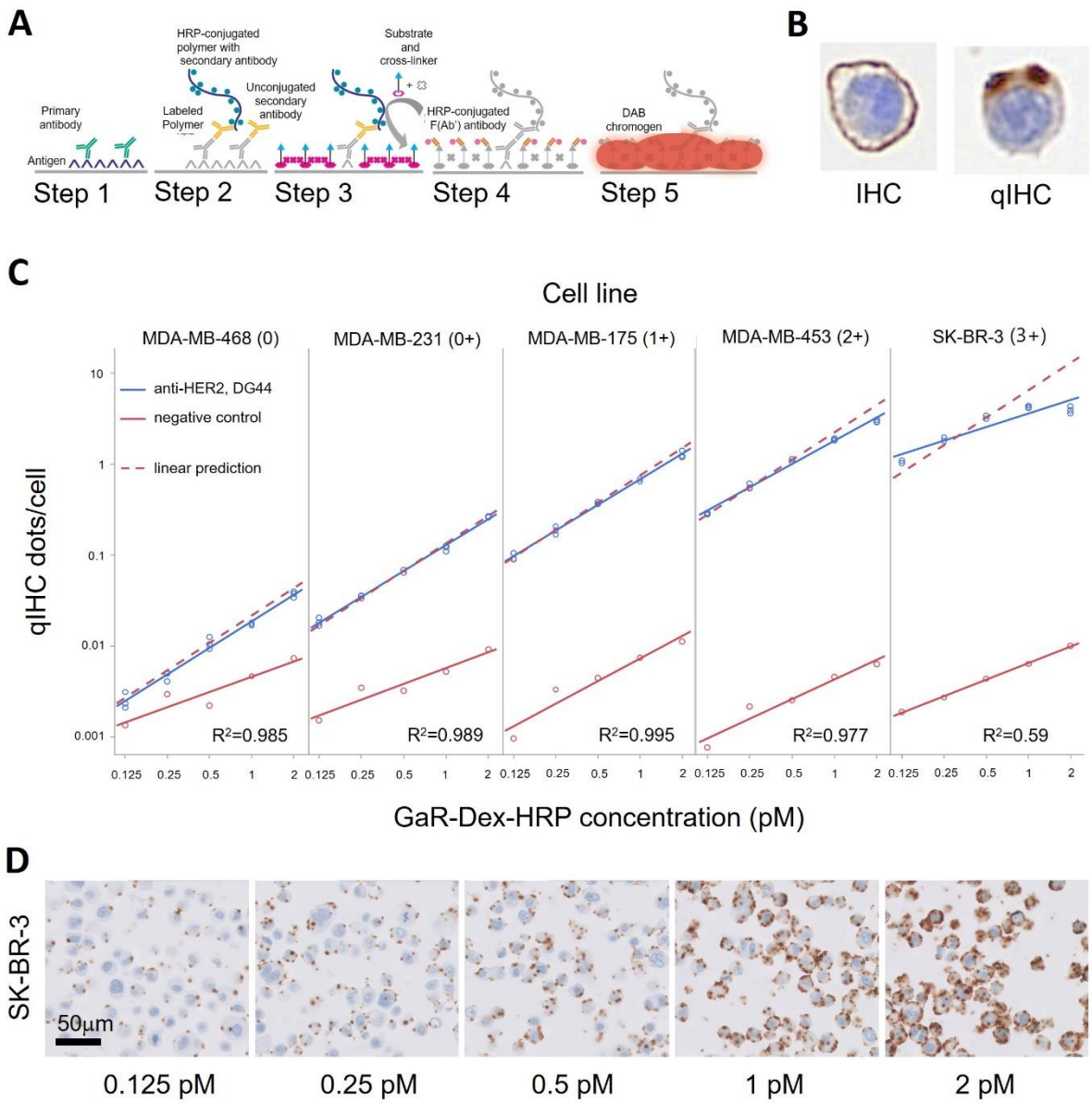

Figure 2. Principle of qIHC. **A**. Schematic illustration of the qIHC staining procedure used in this study, illustration modified from Jensen et al. [14]. **B.** Example of one MDA-MB-453 FFPE cell stained with standard IHC (HercepTest™ mAb), and one cell stained by HER2 qIHC. **C**. HER2 qIHC dots/cell as a function of the concentration of HRP-conjugated goat anti-rabbit dextran polymer (GaR-Dex-HRP) in five different breast cancer cell lines using the monoclonal rabbit anti-HER2 antibody (clone DG44) tested in triplicates (blue circles, blue linear fit with R^2 coefficients given) and negative control reagent (antibody diluent without HER2 antibody) in singlets (red circles, red linear fit). The dotted line represents predicted HER2 qIHC dots/cell if doubling or halving the GaR-Dex-HRP concentration calculated from the mean HER2 qIHC dots/cell at 0.5 pM GaR-Dex-HRP. **D.** HER2 qIHC staining at 40X magnification in the SK-BR-3 cell line at different GaR-Dex-HRP concentrations.

qIHC on Breast cancer Specimens

We then applied the qIHC staining method using negative control reagent or HRP-conjugated polymer on 82 FFPE breast cancer specimens with sequential sections stained with H&E,

HercepTest™ mAb, and p63 as described in the supplementary Figure S1. Figure 3 shows representative staining results from HER2-low, HER2-ultra-low and HER2-null breast cancer specimens stained with HercepTest™ mAb (left panel) and HER2 qIHC (right panel). The middle panel in Figure 3 represents a digital enhancement to visually clarify the faint/barely perceptible HER2 staining, and in the right panel the number of qIHC dots detected in these image patches is given in the lower right corner. As anticipated, the qIHC dot number increases from the HER2-null to the HER2-ultra-low and HER2-low patches.

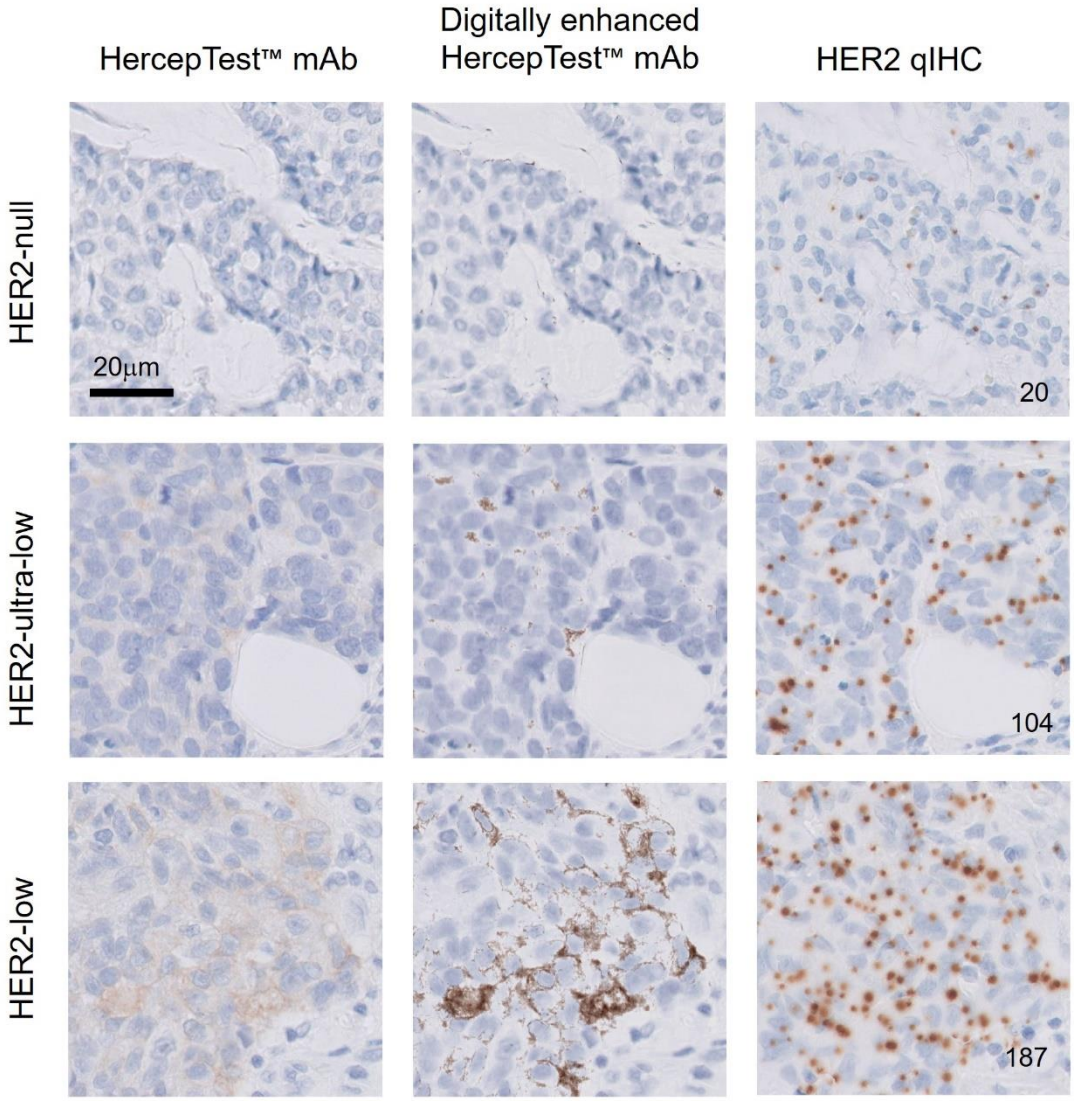

Figure 3. Comparison between IHC, digitally enhanced IHC and qIHC. Top, middle and bottom rows with representative field of views (FOVs) of HER2-null, HER2-ultra-low and HER2-low breast cancer tissues, respectively. Left and right columns have FOVs of aligned sections with HercepTest™ mAb and HER2 qIHC, respectively. The middle column represents the HercepTest™ mAb stained section with a digital enhancement to emphasize faint/barely perceptible membrane staining. The number of HER2 qIHC dots is indicated in the bottom right corner of each FOV.

qIHC Analysis and Spatial Distribution of HER2 Expression

AI-based automated detection of qIHC dots is illustrated in Figure 4A as squared boxes added around recognized dots. This AI-based qIHC dot detection model achieved high accuracy with an F1-score of 0.96 in comparison with human annotated ground-truth on the evaluation FOVs. We calculated the slide qIHC count for all specimens and illustrated those in the box plot in Figure 4B grouped by HER2 category. The median slide qIHC count is significantly different between the three HER2 categories, and the median number of qIHC dots per image patch is also significantly different when comparing the 15 HER2-null specimens stained with negative control reagent versus staining with the primary anti-HER2 antibody (Figure 4B). For these 15 specimens characterized as HER2-null the low number of qIHC dots per patch following staining with the negative control reagent illustrates the background level of this system and the high signal-to-noise observed in qIHC. We used a sliding window approach to count dots, allowing us to quantify HER2 expression across the entire tissue sections. In Figure 4C and 4D, we demonstrate such quantitative spatial distribution of HER2 expression from the qIHC dots detected. Figure 4C illustrates a HercepTest™ mAb stained tissue WSI in the HER2 IHC 0 category with manual annotations of solid tumor regions as described in the material and method section and in the supplementary Figure S2. Two small areas with higher magnification are also shown below panel 4C. In Figure 4D a qIHC dots-count heatmap is overlaid on this tissue to reveal the spatial heterogeneity of the quantitative HER2 expression across this WSI, and the same subregions as under Figure 4C is seen under Figure 4D to illustrate this heatmap at higher magnification.

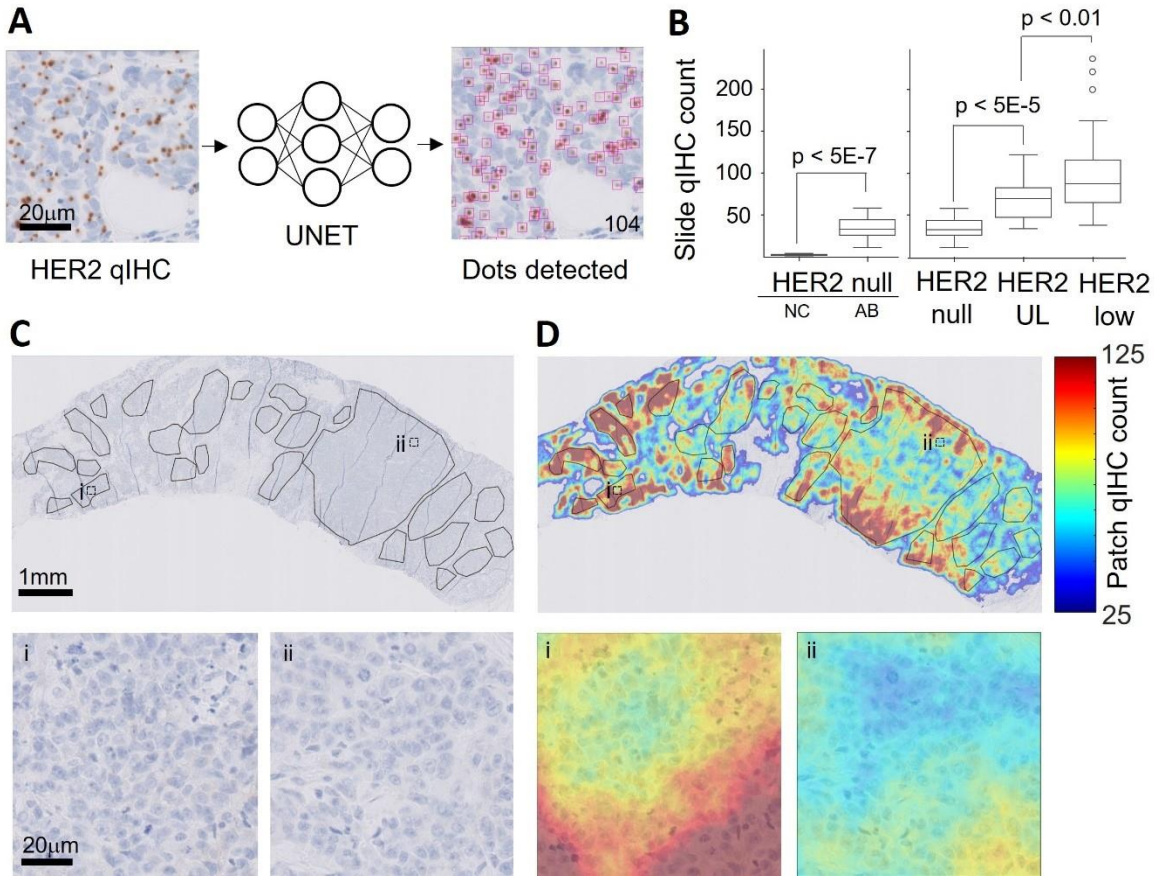

Figure 4. qIHC dot detection and visualization. **A.** Principle of HER2 qIHC dot detection by a UNET deep-neural-network showing detected dots with magenta boxes, and the total number of detected

dots in the bottom right corner. **B**. Left panel has the average qIHC dots count per patch for HER2-null cases stained with negative control reagents (NC) and HER2 antibody (AB). The p-value is given for a paired, two-sided t-test with Welch correction for unequal variances between the NC and AB groups. Right panel has average qIHC dots count per patch in the three categories HER2-null, HER2-ultra-low (UL), and HER2-low (low) and p-values are from median comparison using the Kruskal-Wallis H-test. **C**. Top row has breast cancer tissue WSI stained with HercepTest™ mAb and solid tumor area annotations in black markings. Bottom row has enlarged areas i and ii from top row. **D**. Same specimen as in **C** with the observed patch qIHC count overlaid using a heat map. Blue to red colors indicate low and high patch qIHC count, respectively.

AI Model Prediction of HER2 Expression Using patch qIHC Count Ground Truth

We then trained AI-models to predict HER2 expression on the HercepTest™ mAb WSI scans using the ground truth HER2 expression from qIHC dot count as described in the Materials and Methods section (Figure 5A). The resulting AI-models could quantitatively predict the number of qIHC dots per patch in WSI of HercepTest™ mAb stained specimens. Figure 5B illustrates the correlation of predicted qIHC dots from such AI-models as a function of the actual number of observed qIHC dots. A Pearson correlation coefficient at 0.94, equal to R^2 at 0.87 was found. This correlation was found by splitting the entire dataset of 82 specimens in several training and testing subsets as described in the Materials and Methods section, and for each training subset a new AI-model was developed and evaluated on the test subset (Figure S5). Using this procedure, we ensured that none of the tested AI-models were trained on the same tissue WSI used for evaluation tests. Figure 5C shows examples of two HER2 expression heatmaps, in which the left one represents the actual HER2 qIHC dots detected, and the right one represents the predicted quantitative HER2 expression output from an AI-model that received the HercepTest™ mAb WSI scan as input. For this specific breast cancer WSI, the Pearson correlation coefficient describing how well the predicted HER2 expression correlated with actual HER2 expression was found at 0.82, with R^2 at 0.67.

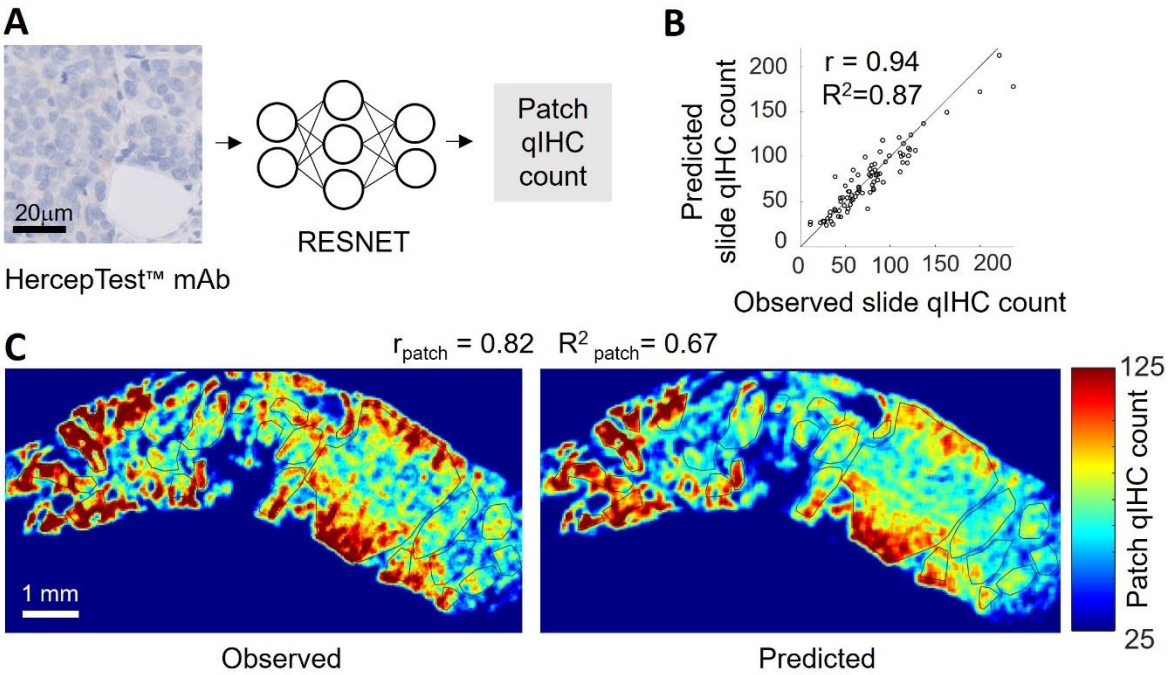

Figure 5. qIHC patch count prediction and visualization. **A.** Principle of patch HER2 qIHC dot prediction by the trained ResNet50 deep neural network. The HercepTest™ mAb WSI scan is processed by the trained ResNet50 AI-model that outputs the predicted numbers of HER2 qIHC dots per patch. **B**. AI predicted HER2 slide qIHC counts as a function of the actual number of HER2 slide qIHC counts. The correlations given are based on slide comparisons across 82 tissues **C**. Left panel with heatmap from one breast cancer specimen to illustrate patch qIHC dot count of observed HER2 expression, and right panel the AI-predicted HER2 qIHC dots count per patch. Blue to red colors indicate low and high patch qIHC count, respectively. Solid tumor annotations are shown with black markings. The correlations given are based on patch comparisons.

Discussion

In this paper, we introduced a new methodology for quantifying HER2 expression, specifically targeting breast cancer cases in the HER2 IHC 0 and 1+ categories, using an analysis of qIHC-stained tissue slides. Additionally, we developed an AI-based interpretation of HercepTest™ mAb using qIHC analysis as ground truth.

Recent successes of ADC based therapies, as demonstrated in Destiny Breast-04 and Destiny-Breast-06 clinical trials, along with the FDA approval of T-DXd for the treatment of metastatic HER2-positive and HER2-low breast cancer, have expanded the treatment options for the significant subgroup of breast cancer patients. Results from the Destiny Breast-06 clinical trial showed that T-DXd provides statistically significant and clinically meaningful benefits to HER2-ultra-low breast cancer patients as well.

However, sensitive, accurate and quantitative evaluation of HER2 expression based on currently approved IHC assays remains challenging, especially in low and ultra-low ranges of HER2 [3]. The ability to correlate and differentiate clinical outcomes with HER2 expression depends not only on the efficacy of the tested drugs but also on the performance of diagnostic assays employed to assess HER2 expression. Advanced computational approaches could improve the interpretation of such IHC assays and could be of high benefit for identifying the best treatment options for current and future HER2 targeted therapies. More importantly, it could provide a basis for objectively assessing the lower level of HER2 protein expression that maximizes therapeutic benefits, while minimizing drug exposure risks for patients.

As first suggested by Jensen et al. [14] we now demonstrate that qIHC can be used as a tunable system, where the qIHC dot count is directly proportional to the concentration to the dot-generating component. This allows for quantitative HER2 detection with a high signal/noise ratio at different protein expression levels, as demonstrated by our reported reliable and consistent quantitative expression information in FFPE cell lines that have HER2 expression ranges from IHC 0 to 3+. Hence, even though the FFPE tissues used in this study are exclusively HER2 IHC 0 and IHC 1+, the method can be adapted to cover the full HER2 expression range. Additionally, the method can be extended to other cancer types depending on HER2 pattern of expression as well as intra-tumoral heterogeneity.

We also demonstrate that qIHC slides can be automatically and accurately processed using an AI detection model to identify qIHC dots, enabling scalable processing of qIHC staining. By integrating the detected qIHC dots over patches and averaging them across tumor regions, qIHC can provide continuous quantification of HER2 expression, allowing statistically significant differentiation between various IHC HER2 low categories above the background level of qIHC staining, as shown in Figure 4B. Alternatively, different thresholds could be applied, particularly at ultra-low levels, to provide varying degrees of stratification in future CDx analyses or studies.

In addition, patch-based quantification of HER2 expression using qIHC staining can be spatially mapped across the entire breast cancer tissue, providing valuable insights into HER2 expression heterogeneity. This information can be graphically overlaid on HER2 IHC stained WSI for both visualization and quantification of the heterogeneity, as demonstrated in Figure 4C-D. Furthermore, using qIHC staining, we were able to visualize low levels of HER2 expression that are challenging to detect with the human eye using traditional IHC (subpanels i and ii in Figure 4C-D), or even expression levels below the limit of detection of currently available HER2 IHC assays. While patch-based qIHC scores are currently summarized into global scores, they could conceivably be adapted to offer more detailed, and spatially resolved measurements of HER2 expression in low and ultra-low regimes. Additionally, new metrics leveraging qIHC-based spatial analysis of HER2 expression could be developed, which may be particularly useful for studying mechanism and therapeutic effects of Enhertu® or other ADC drugs, such as the bystander effect.

In addition, we successfully demonstrated a method for using our qIHC staining and analysis as ground truth for training AI models to extract information from the HercepTest™ mAb pharmDx stain. This method unlocks the ability to infer continuous quantitative expression levels from traditional HER2 IHC. Two key factors contributed to the success of our method. First, the quantitative, high sensitivity, tunability and scalability of qIHC staining and analysis for HER2 expression make it an ideal candidate for providing reliable and consistent ground truth for AI modeling. Second, the relatively high sensitivity of HercepTest™ mAb [28] generates minute membrane staining signal even in low and ultra-low HER2 expression tissues, although it is difficult to detect by the human eye. This weak signal, which an AI model can learn to detect, is made apparent when the DAB signal is digitally enhanced, as shown in Figure 3. Using our AI model training method, we achieved excellent agreement between model inference and measured qIHC expression at the slide level. This agreement extended to low levels of HER2 expression (Figure 5B). Additionally, strong consistency was observed in the spatial quantification of HER2 expression across the entire breast cancer tissue section, providing valuable insights into HER2 heterogeneity (Figure 5C).

While the results of this study are promising, current qIHC analysis faces limitations in differentiating between cytoplasmic and membranal HER2 expression. However, future developments, such as multiplexing membranal markers with qIHC staining using different chromogens and applying algorithmic approaches, could assist in distinguishing between these expression patterns. It should be emphasized that the data provided from this work is purely investigational, and the observations brought forward requires additional clinical validation before being employed in the diagnostic space.

Several reports use orthogonal methods as opposed to IHC to examine whether alternative quantitative methods can help improve predictive capabilities while maintaining the important spatial resolution of specimens intact [10, 29], but require specialized workflows and has so far shown limited resolution in distinguishing HER2 IHC 0, HER2 IHC ultra-low and HER2 IHC low groups. As we demonstrate in this paper, a qIHC based workflow or AI-based interpretation of HercepTest™ mAb could be used to measure the HER2 expression in FFPE specimens without requiring a specialized workflow not currently in place in histology labs, in contrast to other quantitative methods for measuring HER2. Due to the large dynamic range, the technology can be used to separate even low-expression patient groups. qIHC staining is not limited to HER2 but can be extended to any IHC biomarker at any expression level.

Looking ahead, more detailed, automatic, spatial, and continuous quantification of HER2 expression across WSI may provide insights into the impact of HER2 heterogeneity on T-DXd response and help

optimize patient selection for treatment decisions. Moreover, the methods presented here could be expanded to cover the full range of HER2 expression and applied to other tumors beyond breast cancer. Additionally, these techniques hold the potential for measuring multiple biomarkers through multiplexing and could even be extended to sparse tumors in the future.

In conclusion, this work illustrates a potential for significant advancement in the field of pathology by utilizing the power AI based interpretation of IHC measurements to better quantify and spatially resolve protein expression in cancer tissue, ultimately providing better tools for studying response to therapy.

References

1. Jørgensen JT, Winther H, Askaa J, Andresen L, Olsen D, Mollerup J. A Companion Diagnostic With Significant Clinical Impact in Treatment of Breast and Gastric Cancer. *Front Oncol.* 2021;11:676939. https://doi.org/10.3389/fonc.2021.676939.
2. Venetis K, Crimini E, Sajjadi E, et al. HER2 Low, Ultra-low, and Novel Complementary Biomarkers: Expanding the Spectrum of HER2 Positivity in Breast Cancer. *Front Mol Biosci.* 2022;9:834651. https://doi.org/10.3389/fmolb.2022.834651.
3. Curigliano G, Dent R, Earle H, et al. Open questions, current challenges, and future perspectives in targeting human epidermal growth factor receptor 2-low breast cancer. *ESMO Open.* 2024;9(4):102989. https://doi.org/10.1016/j.esmoop.2024.102989.
4. Zouein J, Karam E, Strickler JH, Kourie HR. Trastuzumab-Deruxtecan: Redefining HER2 as a Tumor Agnostic Biomarker. *Target Oncol.* 2024;19(5):705-710. https://doi.org/10.1007/s11523-024-01079-4.
5. Fernandez AI, Liu M, Bellizzi A, et al. Examination of Low ERBB2 Protein Expression in Breast Cancer Tissue. *JAMA Oncol.* 2022;8(4):1-4. https://doi.org/10.1001/jamaoncol.2021.7239.
6. Karakas C, Tyburski H, Turner BM, et al. Interobserver and Interantibody Reproducibility of HER2 Immunohistochemical Scoring in an Enriched HER2-Low-Expressing Breast Cancer Cohort. *Am J Clin Pathol.* 2023;159(5):484-491. https://doi.org/10.1093/ajcp/aqac184.
7. Zaakouk M, Quinn C, Provenzano E, et al. Concordance of HER2-low scoring in breast carcinoma among expert pathologists in the United Kingdom and the republic of Ireland -on behalf of the UK national coordinating committee for breast pathology. *Breast.* 2023;70:82-91. https://doi.org/10.1016/j.breast.2023.06.005.
8. Turashvili G, Gao Y, Ai DA, et al. Low interobserver agreement among subspecialised breast pathologists in evaluating HER2-low breast cancer. *J Clin Pathol.* 2023. https://doi.org/10.1136/jcp-2023-209055.
9. Rüschoff J, Penner A, Ellis IO, et al. Abstract HER2-13: HER2-13 Proficiency assessment of HER2-low breast cancer scoring with the Ventana PATHWAY 4B5 and Dako HercepTest HER2 assays and the impact of pathologist training. *Cancer Res.* 2023;83(HER2-13). https://doi.org/doi.org/10.1158/1538-7445.SABCS22-HER2-13.
10. Li X, Lee JH, Gao Y, et al. Correlation of HER2 Protein Level With mRNA Level Quantified by RNAscope in Breast Cancer. *Mod Pathol.* 2024;37(2):100408. https://doi.org/10.1016/j.modpat.2023.100408.
11. Tyburski H, Karakas C, Finkelman BS, Turner BM, Zhang H, Hicks DG. In ER-Positive, HER2-Negative Breast Cancers, HER2 mRNA Levels Correlate Better with Clinicopathologic Features and Oncotype DX Recurrence Score than HER2 Immunohistochemistry. *Lab Invest.* 2024;104(3):100309. https://doi.org/10.1016/j.labinv.2023.100309.
12. Moutafi M, Robbins CJ, Yaghoobi V, et al. Quantitative measurement of HER2 expression to subclassify ERBB2 unamplified breast cancer. *Lab Invest.* 2022;102(10):1101-1108. https://doi.org/10.1038/s41374-022-00804-9.

13. Harigopal M, Barlow WE, Tedeschi G, et al. Multiplexed assessment of the Southwest Oncology Group-directed Intergroup Breast Cancer Trial S9313 by AQUA shows that both high and low levels of HER2 are associated with poor outcome. *Am J Pathol.* 2010;176(4):1639-1647. https://doi.org/10.2353/ajpath.2010.090711.
14. Jensen K, Krusenstjerna-Hafstrøm R, Lohse J, Petersen KH, Derand H. A novel quantitative immunohistochemistry method for precise protein measurements directly in formalin-fixed, paraffin-embedded specimens: analytical performance measuring HER2. *Mod Pathol.* 2017;30(2):180-193. https://doi.org/10.1038/modpathol.2016.176.
15. Lohse J, Petersen KH, Woller NC, Pedersen HC, Skladtchikova G, Jørgensen RM. Improved catalyzed reporter deposition, iCARD. *Bioconjug Chem.* 2014;25(6):1036-1042. https://doi.org/10.1021/bc400311g.
16. Fosmark S, Hellwege S, Dahlrot RH, et al. APNG as a prognostic marker in patients with glioblastoma. *PLoS One.* 2017;12(6):e0178693. https://doi.org/10.1371/journal.pone.0178693.
17. Jung M, Song SG, Cho SI, et al. Artificial intelligence-powered human epidermal growth factor receptor 2 (HER2) analyzer in breast cancer as an assistance tool for pathologists to reduce interobserver variation. *Journal of Clinical Oncology.* 2022;40:e12543. https://doi.org/10.1200/JCO.2022.40.16_suppl.e12543.
18. Palm C, Connolly CE, Masser R, et al. Determining HER2 Status by Artificial Intelligence: An Investigation of Primary, Metastatic, and HER2 Low Breast Tumors. *Diagnostics (Basel).* 2023;13(1). https://doi.org/10.3390/diagnostics13010168.
19. Wu S, Yue M, Zhang J, et al. The Role of Artificial Intelligence in Accurate Interpretation of HER2 Immunohistochemical Scores 0 and 1+ in Breast Cancer. *Mod Pathol.* 2023;36(3):100054. https://doi.org/10.1016/j.modpat.2022.100054.
20. Bannier PA, Broeckx G, Herpin L, et al. Development of a deep-learning model tailored for HER2 detection in breast cancer to aid pathologists in interpreting HER2-low cases. *Histopathology.* 2024;85(3):478-488. https://doi.org/10.1111/his.15274.
21. Kapil A, Spitzmuller A, Brieu N, et al. HER2 quantitative continuous scoring for accurate patient selection in HER2 negative trastuzumab deruxtecan treated breast cancer. *Sci Rep.* 2024;14(1):12129. https://doi.org/10.1038/s41598-024-61957-9.
22. Tarantino P, Viale G, Press MF, et al. ESMO expert consensus statements (ECS) on the definition, diagnosis, and management of HER2-low breast cancer. *Ann Oncol.* 2023;34(8):645-659. https://doi.org/10.1016/j.annonc.2023.05.008.
23. Wolff AC, Hammond MEH, Allison KH, Harvey BE, McShane LM, Dowsett M. HER2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update Summary. *J Oncol Pract.* 2018;14(7):437-441. https://doi.org/10.1200/JOP.18.00206.
24. Wolff AC, Somerfield MR, Dowsett M, et al. Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: ASCO-College of American Pathologists Guideline Update. *J Clin Oncol.* 2023;41(22):3867-3872. https://doi.org/10.1200/JCO.22.02864.
25. Arbel E, Remer I, Ben-Dor A. Deep Learning Based Partial Annotation Framework for Instance Segmentation in Histopathology Images. *Medical Imaging with Deep Learning 2019.* 2019.
26. Kumar N, Verma R, Anand D, et al. A Multi-Organ Nucleus Segmentation Challenge. *IEEE Trans Med Imaging.* 2020;39(5):1380-1391. 8880654. https://doi.org/10.1109/TMI.2019.2947628.
27. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 2016:770-778.
28. Rüschoff J, Friedrich M, Nagelmeier I, et al. Comparison of HercepTest mAb pharmDx (Dako Omnis, GE001) with Ventana PATHWAY anti-HER-2/neu (4B5) in breast cancer: correlation with HER2 amplification and HER2 low status. *Virchows Arch.* 2022;481(5):685-694. https://doi.org/10.1007/s00428-022-03378-5.

29. Hicks DG, Buscaglia B, Goda H, et al. A novel detection methodology for HER2 protein quantitation in formalin-fixed, paraffin embedded clinical samples using fluorescent nanoparticles: an analytical and clinical validation study. *BMC Cancer.* 2018;18(1):1266. https://doi.org/10.1186/s12885-018-5172-1.

Supplementary Information

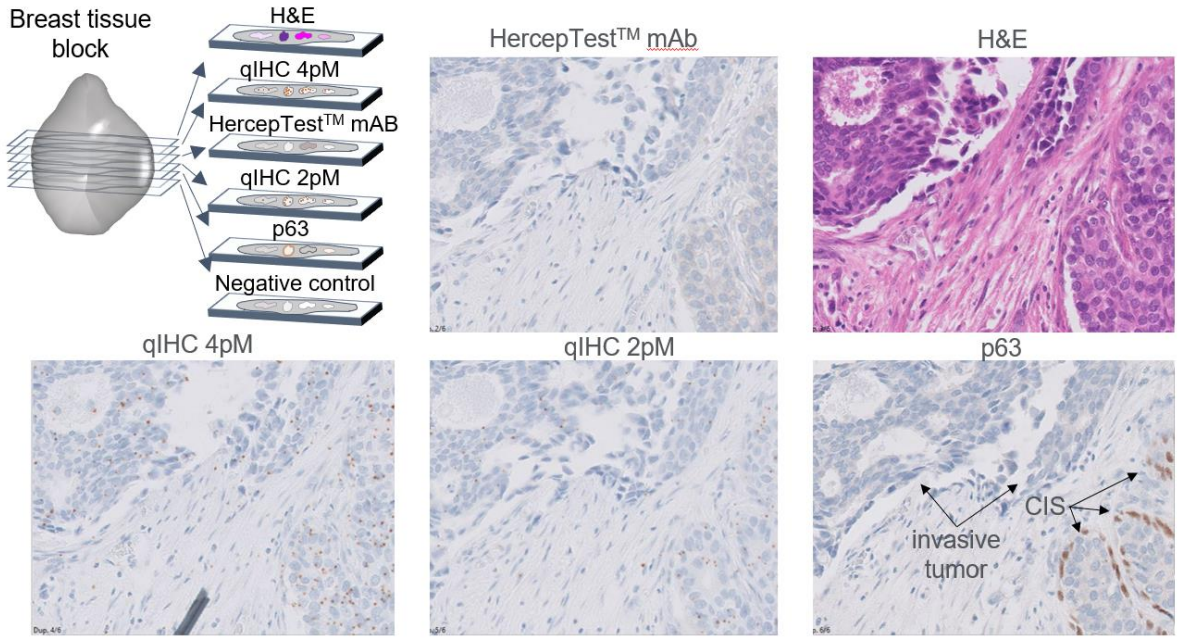

Figure S1. Overview of tissue sections used. Upper, left panel illustrates the six sequential sections cut from each breast cancer specimen and how these sections were stained. Remaining panels show a representative, aligned field of view for five stained sections with invasive tumor and carcinoma in situ (CIS) indicated in the p63 stained section.

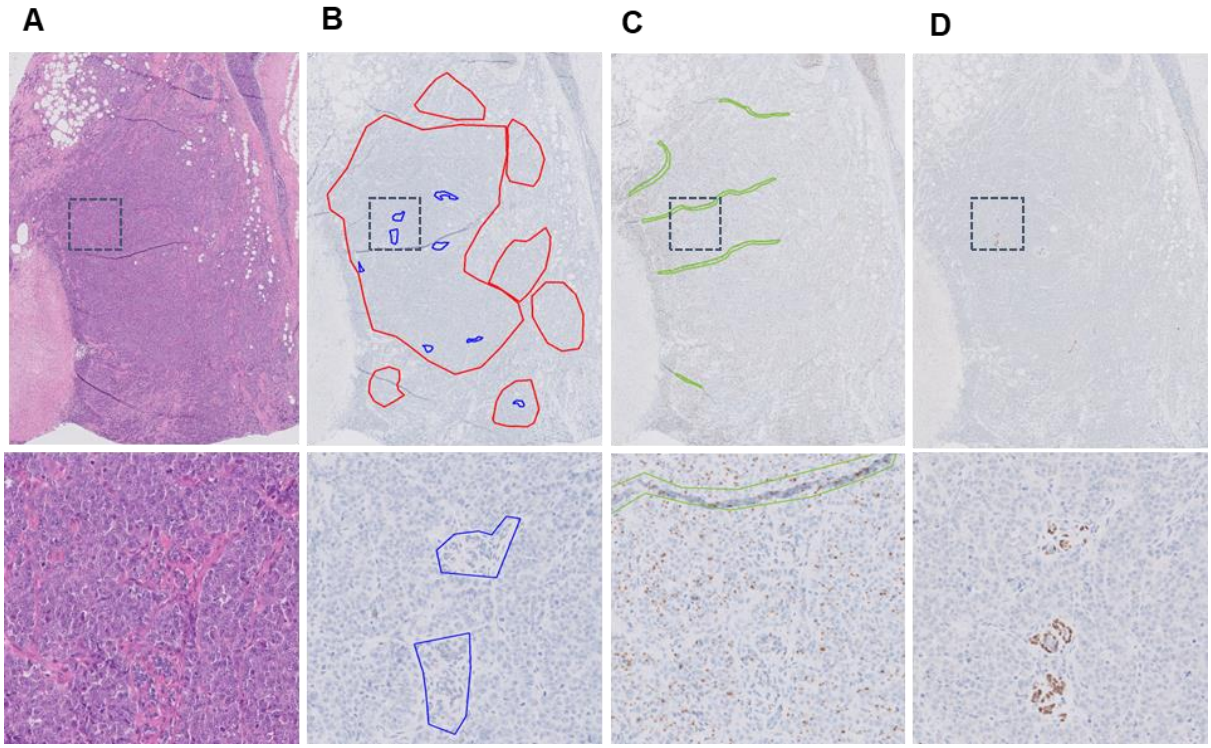

Figure S2. Illustration of tumor annotations in different sections of a breast cancer specimen. Images in the top row represent an overview, and the bottom row a zoomed-in area as indicated by the dashed box. **A**. H&E counterstain to support identification of solid tumor regions and non-tumor areas. **B**. HER2 IHC stain by HercepTest™ mAb where solid tumor regions were encircled red, while large non-tumor areas representing stroma and normal ducts were encircled blue for exclusion. **C**. HER2 qIHC staining for dot detection in which know artefacts such as tissue folds were marked green for exclusion. **D**. The p63 IHC stain supported identification of normal ducts and CIS structures to be excluded.

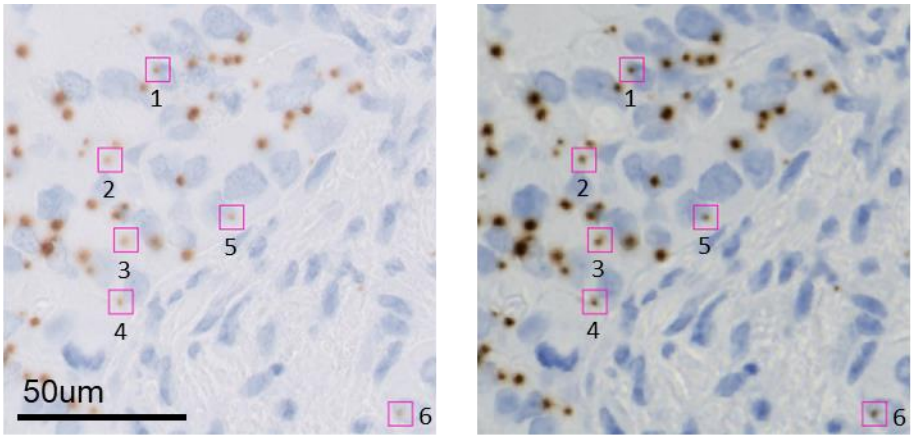

Figure S3. Comparison between single plane and extended depth of focus (EDF) for dot detection. Left image represents a single plane field of view from a HER2 qIHC stained breast cancer section scanned on the Philips Ultra Fast scanner. The magenta boxes highlight six faint qIHC dots that might be consider as staining artifacts instead of real dots. Right image represents the same slide area scanned using the extended depth focus option on the Zeiss Axioscan.Z1 with software-based flattening of five focus planes, dots are clearly observed.

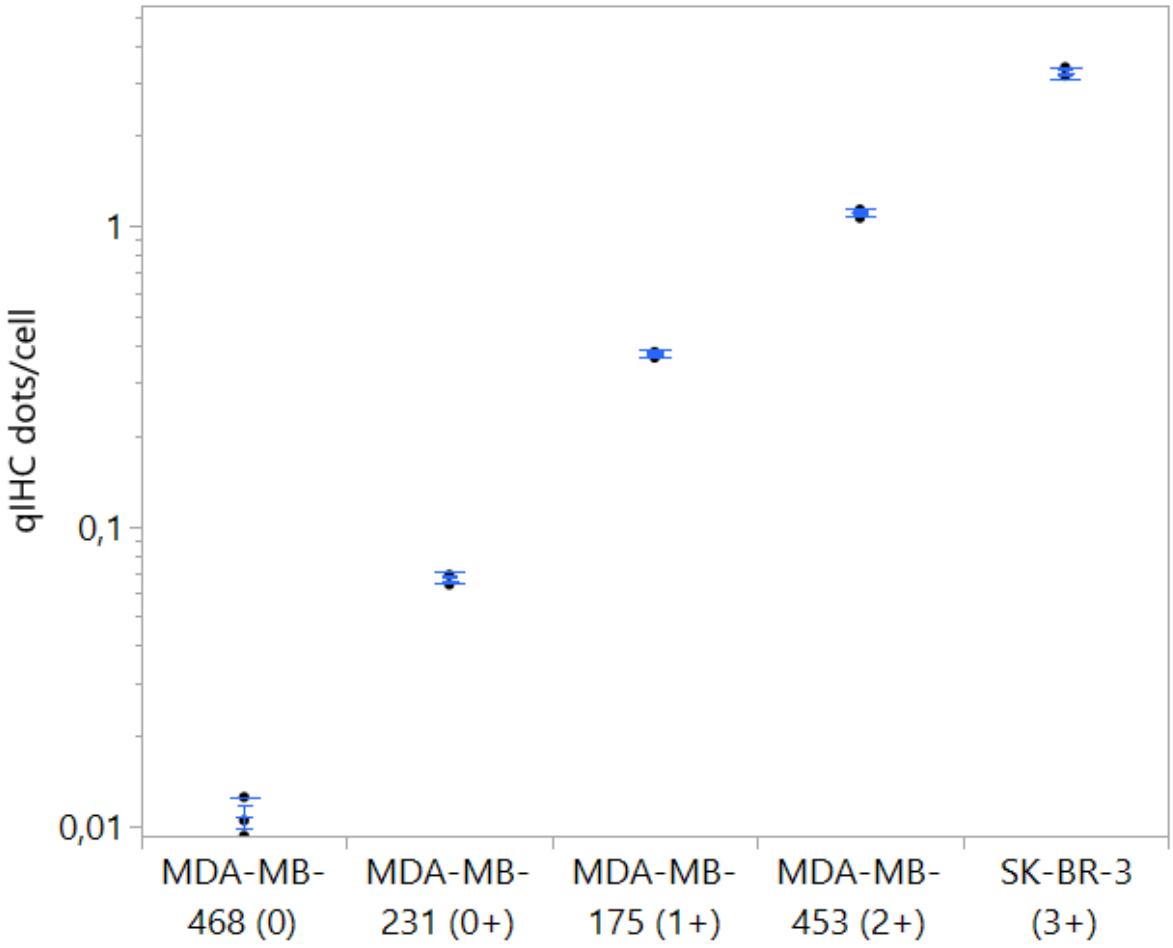

Figure S4: Sensitivity and variability of qIHC in different HER2 expressing cell lines. HER2 qIHC dots/cell in five different breast cancer cell lines using the monoclonal rabbit anti-HER2 antibody (clone DG44) tested in triplicates at a concentration of 1 pM of HRP-conjugated goat anti-rabbit dextran polymer. Bars show the 95% confidence limits of the mean for each cell line.

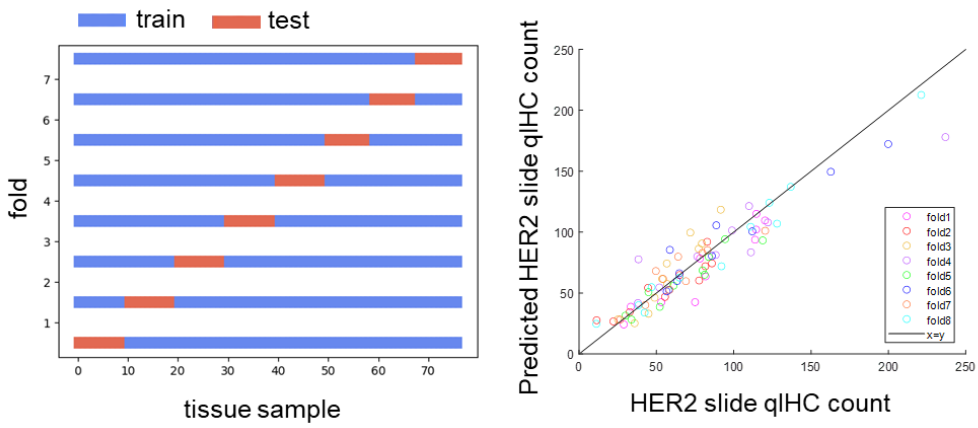

Figure S5. Principle of k-fold validation. Left panel illustrates the principle of k-fold cross validation where 82 breast cancer specimens were subdivided into eight non-overlapping subsets, each containing 72 (or 71) specimens for AI-training, and 10 (or 11) for testing the resulting AI-model. The different colors in the right panel illustrate the predicted HER2 qIHC dots per patch versus the actual HER2 qIHC dots per patch from the eight separate train and test events. The line has been drawn to illustrate the perfect correlation where y=x.

Acknowledgements

The authors express their sincere appreciation to Steve Laderman, Lou Welebob, Darlene Solomon, Sam Raha, and Lars Von Gersdorff for their unwavering support and encouragement. Special thanks to Claudia Gottstein, Ghislain Noumsi, Doug Clark, and Emin Orudjev for their invaluable scientific discussions and to Josh Littrell for statistical assistance. We appreciate the histology assistance provided by Charlotte Aagerup, Anne Falkenskjold, Anne Maarbjerg, and Marianne Marcussen.

Author Contributions

FA, EA, IR, OB, DR, AB, KH, KS, LJ, JM, and AT contributed to study concept and design. FA optimized IHC assays and generated the data. EA, IR, OB, and DR trained AI models and analyzed the data. AB developed supporting software tools. KH, KS, and SA reviewed and annotated IHC stained tissues. JM, IR, EA, OB, FA, and AT wrote the manuscript, GN edited the manuscript. LJ and AT supervised the research.

Funding

The authors did not receive external funding for this study.

Declaration of competing interests

FA, EA, IR, OB, DR, AB, KH, KS, JM, LJ, and AT are employees and shareholders of Agilent Technologies. SA performed this work as a consultant for Agilent Technologies.

Data Availability Statement

Aggregated data is available on reasonable request.

Ethics statement and patient consent

All specimens in the study were residual, de-identified, originated from different individuals, and were procured from commercial sources or local hospitals with no possibility for patient identification. The study was performed in agreement with the World Medical Association Declaration of Helsinki (2013). No study protocol was submitted for Ethics Committee approval as this type of analytical study is exempt from Ethics Committee approval in Denmark.