Journal of Pathology Informatics 13 (2022) 100116

Contents lists available at ScienceDirect

Journal of Pathology Informatics

journal homepage: www.elsevier.com/locate/jpi

Original Research Article

Computer-aided scoring of *erb-b2 receptor tyrosine kinase 2* (*HER2*) gene amplification status in breast cancer

Alyson Yoder, Landon J. Inge *, Chen-Chun Chen, Vijay R. Marati, Trung Kien Nguyen, Karel Zuiderveld, Jim Martin, Sarah Gladden, Mohammad Saleh Miri, Raghavan Venugopal, Bryan Lopez, Jim Ranger-Moore, Christoph Guetter

Ventana Medical Systems, Inc., Tucson, AZ, USA

ARTICLE INFO

Keywords:
Image analysis
Algorithm
HER2
Breast cancer
Bright-field ISH

ABSTRACT

Background: Identification of HER2 protein overexpression and/or amplification of the *HER2* gene are required to qualify breast cancer patients for HER2 targeted therapies. *In situ* hybridization (ISH) assays that identify *HER2* gene amplification function as a stand-alone test for determination of HER2 status and rely on the manual quantification of the number of *HER2* genes and copies of chromosome 17 to determine *HER2* amplification.

Methods: To assist pathologists, we have developed the uPath HER2 Dual ISH Image Analysis for Breast (uPath HER2 DISH IA) algorithm, as an adjunctive aid in the determination of *HER2* gene status in breast cancer specimens. The objective of this study was to compare uPath HER2 DISH image analysis vs manual read scoring of VENTANA HER2 DISH-stained breast carcinoma specimens with ground truth (GT) gene status as the reference. Three reader pathologists reviewed 220, formalin-fixed, paraffin-embedded (FFPE) breast cancer cases by both manual and uPath HER2 DISH IA methods. Scoring results from manual read (MR) and computer-assisted scores (image analysis, IA) were compared against the GT gene status generated by consensus of a panel of pathologists. The differences in agreement rates of *HER2* gene status between manual, computer-assisted, and GT gene status were determined.

Results: The positive percent agreement (PPA) and negative percent agreement (NPA) rates for image analysis (IA) vs GT were 97.2% (95% confidence interval [CI]: 95.0, 99.3) and 94.3% (95% CI: 90.8, 97.3) respectively. Comparison of agreement rates showed that the lower bounds of the 95% CIs for the difference of PPA and NPA for IA vs MR were –0.9% and –6.2%, respectively. Further, inter- and intra-reader agreement rates in the IA method were observed with point estimates of at least 96.7%.

Conclusions: Overall, our data show that the uPath HER2 DISH IA is non-inferior to manual scoring and supports its use as an aid for pathologists in routine diagnosis of breast cancer.

Background

Breast cancer is the most common cancer among women and the leading cause of global cancer mortality with more than 680 000 deaths among women in 2020.[1] Human erb-b2 receptor tyrosine kinase 2 (*ERBB2 or HER2*), a member of the epidermal growth factor receptor family, is overexpressed or amplified in 15%–20% of breast cancers[2,3] and overexpression/amplification is associated with a higher rate of recurrence and death.[4,5] Routine clinical use of trastuzumab and other HER2 targeted therapies have dramatically changed outcomes for patients with HER2 overexpression, providing improved survival, and quality of health.[6,7] Determination of HER2 positivity, by either overexpression or gene amplification, is fundamental to the current treatment paradigm, ensuring that patients receiving therapy can derive the most benefit from HER2-targeted therapies. In current practice, HER2 status is determined through immunohistochemical and/or *in situ* hybridization (dual color fluorescence [FISH] or dual bright-field chromogenic *in situ* hybridization [DISH]) assays.[8] For *in situ* hybridization, the necessity of manual quantification of *HER2* gene amplification, while reproducible, remains a time-consuming and expensive process.[9] Furthermore, with pathologists facing ever-expanding work responsibilities, combined with a shrinking number of new pathologists entering the field, there is a growing unmet need for tools to assist pathologists in managing their daily work.[10–12] Over the past several years, the increasing availability of slide scanners that can digitize a glass slide and produce a whole slide image (WSI) at high-resolution, combined with advances in computers and data storage has brought the promise of digital pathology into reality. In addition to full digitization of the pathologist workflow, the ability to leverage WSI and

* Corresponding author at: 1910 E. Innovation Park Dr., Tucson, AZ 85755, USA.
E-mail addresses: alyson.yoder@contractors.roche.com (A. Yoder), landon.inge@roche.com (L.J. Inge), chen_chun.chen@roche.com (C.-C. Chen), nguyenk8@gene.com (T.K. Nguyen), karel.zuiderveld@roche.com (K. Zuiderveld), jim.martin@contractors.roche.com (J. Martin), sarah.gladden@roche.com (S. Gladden), saleh.miri@roche.com (M.S. Miri), raghavan.venugopal@roche.com (R. Venugopal), bryan.lopez@roche.com (B. Lopez), jim.ranger-moore@roche.com (J. Ranger-Moore), christoph.guetter@roche.com (C. Guetter).

http://dx.doi.org/10.1016/j.jpi.2022.100116
Received 10 January 2022; Received in revised form 5 June 2022; Accepted 24 June 2022
Available online 05 July 2022

computer image analysis algorithms to aid pathologists with quantitative and qualitative tasks are also now possible. Indeed, the need for manual quantification of *HER2* gene amplification in both FISH and DISH assays represents an ideal use case for image analysis algorithms to function as an aid for pathologists. As an example, automated image analysis has been used for FISH analysis and was found to be concordant with manual analysis,[13–15] despite increased variability at higher HER2 counts.[16]

Here we have developed the uPath HER2 Dual ISH image analysis algorithm for breast (uPath HER2 DISH IA) as a clinical adjunctive aid for pathologists to score breast cancers stained with the VENTANA HER2 Dual ISH DNA Probe Cocktail assay. In this study, we demonstrate that the uPath HER2 DISH IA algorithm displays high inter- and intra-reader agreement between pathologists, and is concordant to manual assessment of the VENTANA HER2 Dual ISH DNA Probe Cocktail assay. Collectively our data show that the uPath HER2 DISH IA algorithm is robust, reproducible and non-inferior to manual pathologist scoring.

Methods

Tissue specimens, tissue processing, and ISH staining

Invasive breast cancer specimens used for validation and development of the uPath HER2 DISH IA algorithm were sourced from commercial vendors as formalin-fixed paraffin-embedded (FFPE) tissues and absent of identifying data, except for the histological diagnosis. All samples were reviewed by a Roche pathologist to confirm histological diagnosis and ensure the presence of viable tumor before use. FFPE breast cancers were sectioned at 4 μM, mounted onto positively charged glass slides and stored at room temperature until stained. Tissue sections were de-paraffinized, exposed to an antigen retrieval reagent, and stained with VENTANA HER2 Dual ISH DNA Probe Cocktail and counterstained with hematoxylin and bluing reagent on the BenchMark ULTRA platform (Ventana Medical Systems, Inc.) according to the recommended manufacturer staining procedure. The 220 de-identified, FFPE breast carcinoma tissue specimens used for the studies described below were sourced for an unrelated institutional review board (IRB) approved study assessing reproducibility of the VENTANA HER2 Dual ISH DNA Probe Cocktail assay as part of the regulatory approval for the assay.

uPath HER2 dual ISH image analysis algorithm development

The VENTANA HER2 Dual ISH DNA Probe Cocktail assay is an FDA approved/CE marked[1] diagnostic device used for determination of *HER2* gene status in breast cancer. Regulatory approval of the VENTANA HER2 Dual ISH DNA Probe Cocktail assay is based upon manual counting *HER2* copies via silver *in situ* hybridization (SISH) and chromosome 17 (Chr17) copies via chromogenic red *in situ* hybridization (Re ISH) on a single slide as described in the VENTANA HER2 Dual ISH DNA Probe Cocktail interpretation guide for breast tissue specimens.[17] The uPath HER2 Dual ISH IA is intended to assist in interpretation of VENTANA HER2 Dual ISH staining by automating cell selection, counting of *HER2* gene and Chromosome 17 copies, and quantification of the HER2/Chr17 ratio (Fig. 2). The uPath HER2 Dual ISH IA is a ranking algorithm that utilizes a feature set, e.g. cell shape (roundness metrics, cell segmentation), to identify tumor nuclei and *HER2* gene and Chr17 copies. Although the algorithm itself is not a deep learning algorithm, the feature set used for the ranking and selection of tumor nuclei was derived using an advanced machine learning approach during development. The uPath HER2 Dual ISH IA algorithm is composed of several individual image processing steps that are tightly integrated within the user workflow in the uPath Enterprise software. They include: (1) unmixing color images into blue, red, and black channels, (2) detecting "dots" (e.g. assay signals of HER2 loci and Chr17 copies) in the red and black channels, (3) segmenting cells into scorable cells, and (4) scoring the segmentation based on quality and ranking (Fig. 3). Dots are detected based upon measurement of the signal amplification and the center of the signal center. Cell segmentation is defined as "scorable" if they include both black and red signals and if the segmentation quality is found to be high. Scorable cells are ranked by segmentation quality (i.e. confidence), the number of detected signals per cell, and the likelihood of it being a tumor cell (as defined by cell area). Cell segmentation is performed via an adaptive voting-based algorithm that generates a quality score for each segmented cell. The algorithm ranks the individual cells throughout the region of interest (ROI) using the criteria 1–4 as described above. Twenty cells with the highest ranked scores are then selected by the algorithm for HER2/Chr17 assessment. More in-depth details of the development of the uPath HER2 Dual ISH IA algorithm are proprietary and are not publicly available at this time.

The early feasibility studies utilized pathologist review (n = 3) to establish the accuracy and precision of the cell ranking by the uPath HER2 Dual ISH IA algorithm. The algorithm's accuracy was assessed at the cell level using a data set created from pathologist (n = 1) selection of 50 non-overlapping ROIs from 50 images (1 ROI per image). For each ROI, the algorithm selected 20 cells that were within the cell segmentation boundary and had both HER2 and Chr17 signals (Fig. 3). A total of 1000 cells, generated from the 50 ROIs, were independently evaluated by a panel of 3 pathologists to ensure accuracy of cell segmentation, cell selection, and dot count accuracy.

To verify the algorithm's ability to segment cells, pathologists visually assessed cells (n = 20) selected by the uPath HER2 Dual ISH IA algorithm to determine if segmentation of each cell was acceptable or unacceptable. Segmentation was considered acceptable if the algorithm correctly incorporated the countable dots of a single cell in the segment and non-acceptable if algorithm did not incorporate the countable dots of a cell, or incorporated extra dots from an adjacent cell and/or included miscellaneous signals. To verify the algorithm's ability to select cells, pathologists reviewed the algorithm's selection of the 20 selected cells (out of all cells in the ROI) as either acceptable or unacceptable. Acceptable cells were tumor cells with the highest number of distinct black and red signals representative of that tumor area in a field of view (FOV). Unacceptable cells were lymphocytes or stroma cells that did not have distinct black and red signals, or were not one of the cells in a FOV with the highest number of both black and red signals. A minimum of 2 pathologists had to agree for a cell to be considered acceptable. In order to evaluate the acceptability of the algorithm's selection of HER2 and Chr17 signals, pathologists reviewed the HER2 and Chr17 results generated by the algorithm in selected cells (n = 20). Pathologists (n = 3) reviewed the dot detection results and made corrections by either adding missing or removing incorrectly detected signals, i.e. algorithm dots. The total of HER2 and Chr17 signals after correction were considered the pathologist manual count for HER2 and Chr17 signals, respectively, for that ROI. The median number (generated from 3 different reads by 3 different pathologists) was considered the consensus count for either HER2 or Chr17. The R-squared (R^2) values (Coefficient of determination) between the algorithm count (before correction) and Consensus (defined as a panel of 3 pathologists who provided the consensus scores) pathologist counts (after correction) for HER2 and Chr17 dots were compared. R^2 values greater or equal to 0.9 were considered acceptable.

Pathologist training

A total of 9 pathologists, who were previously qualified to manually evaluate VENTANA HER2 Dual ISH-stained breast cancer specimens for *HER2* gene status, participated in the study. Three groups, Screening, Consensus, and Reader pathologists performed specific study activities (Fig. 1) and underwent training specific for the study protocol prior to performing study activities. Reader (pathologists who participated in digital and manual scoring) Pathologists (n = 3) received training regarding the use the uPath HER2 Dual ISH IA algorithm. Training for the uPath HER2 Dual ISH IA algorithm was conducted by a Roche Pathologist involved in the development of algorithm. The training took place in multiple parts,

[1] uPath HER2 Dual ISH Image Analysis for Breast (uPath HER2 DISH IA) algorithm is *CE-IVD marked. In the US, For Research Use Only. Not for use in diagnostic procedures.*

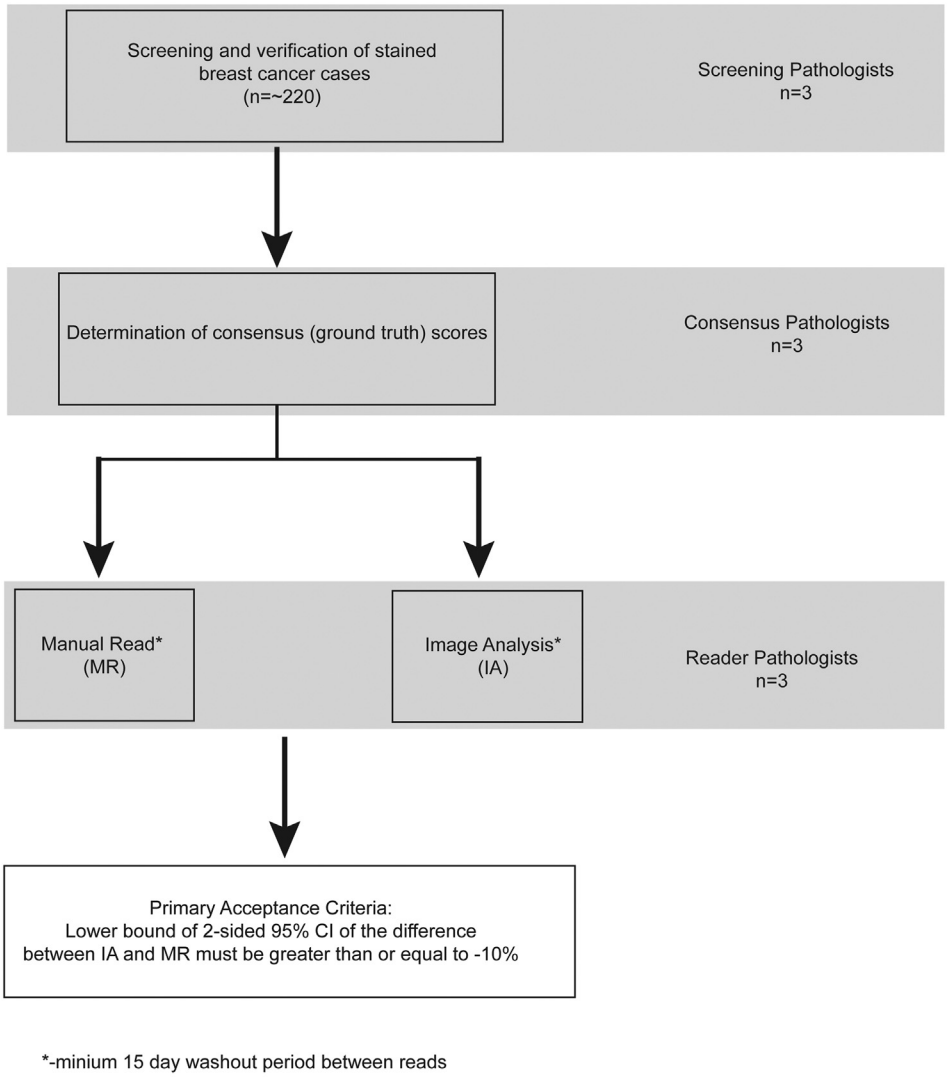

Fig. 1. Study workflow. Flowchart presenting the workflow of the validation studies.

including a group session, self-study (6 cases with consensus scores), mini test (6 cases with consensus scores), and a final proficiency test (12 cases with consensus scores). Reader Pathologists were required to complete a final proficiency test and achieve 90% overall percent agreement to consensus to ensure that they were comfortable and proficient in using the HER2 Dual ISH IA algorithm on Roche uPath enterprise software.

ISH staining and case selection for validation studies

A total of 220 de-identified, FFPE breast carcinoma tissue specimens (§1.2.1) stained with the VENTANA HER2 Dual ISH DNA Probe Cocktail on BenchMark ULTRA instruments at external laboratories under an unrelated study protocol were retrieved from that study's archives and used for the validation studies. VENTANA HER2 Dual ISH- and H&E-stained slides for all 220 selected cases were reviewed by the Screening Pathologists to ensure staining and adequate tumor tissue (Fig. 1). Slides that met the criteria and had a valid VENTANA HER2 Dual ISH staining run according to the package insert were included in the study. To establish the GT of each case, a panel of Consensus Pathologists (n = 3) evaluated the HER2/Chr17 ratio for the selected cases and scored cases as shown in Table 1. GT status was established by two Consensus Pathologists independently scoring each case by manually counting cells on glass slides using a microscope according to the VENTANA HER2 Dual ISH Interpretation Guide[17] (P/N 1018859, Cat.No.760-6072). Due to the inability to ensure that pathologists count the same exact cells and regions, consensus for cases was determined by agreement on amplification status (defined as non-borderline amplified, borderline amplified, borderline non-amplified, and non-borderline non-amplified) instead of exact HER2/Chr17 ratios. If the amplification status assigned by the 2 Consensus Pathologists agreed with each other for a given case, then the agreed status was assigned as the GT status. In the event that the Consensus Pathologists disagreed on the *HER2* gene status for a given case, the third Consensus Pathologist independently scored the case in the same manner as the first two Consensus Pathologists. The majority amplification status among the 3 Consensus Pathologists was the GT status. However, if all 3 Consensus Pathologists disagreed (e.g., amplified, non-amplified, and unevaluable), then for those cases the 3 Consensus Pathologists met together under a multi-headed scope and achieved a GT status by consensus. Selected glass

Table 1
Summary of case qualification and enrollment.

Case bins	Number of cases enrolled
Non-borderline amplified cases (Ratio>2.6)	99
Borderline amplified cases (Ratio ≥2.0 and ≤2.5)	11
Borderline non-amplified cases (Ratio ≥1.5 and ≤1.9)	11
Non-borderline non-amplified cases (Ratio<1.4)	99
Total, no.	220

slides were scanned on the VENTANA DP 200 slide scanners[2] at 40x magnification. Resulting digital slide images were checked for quality before transfer to the centralized server for analysis using the Roche uPath enterprise software and the uPath HER2 Dual ISH IA algorithm. The number of cases distributed to each score bin (non-borderline amplified, borderline amplified, borderline non-amplified, and non-borderline non-amplified) were determined based upon the reported clinical prevalence.[8,18–22]

Method comparison study

The primary endpoint of this method comparison study was to determine whether the uPath HER2 Dual ISH IA algorithm scoring of breast cancer slides was non-inferior to manual pathologist scores. To accomplish this, HER2 scoring results from three Reader Pathologists' IA and Manual Read (MR) modalities were compared against Consensus Pathologists' GT *HER2* gene status for the 220 selected cases. All pathologists who participated in the study were blinded to all previous *HER2* gene status information of the cases. MR were conducted by three Reader Pathologists who independently scored each of the 220 cases using the microscope at a magnification of 600x per the VENTANA HER2 Dual ISH DNA Probe Cocktail assay interpretation guide (Fig. 2A).[17] After completing a minimum 14 day washout following MR evaluation, the same three Reader Pathologists performed IA using the uPath HER2 Dual ISH IA algorithm reads according to the following steps: (1) selection of region of interest (ROI) containing 20 tumor nuclei with the help of the Heat Map (Fig. 2B), (2) the HER2/Chr17 ratio was recorded from the HER2 Dual ISH algorithm, (3) if deemed necessary by the pathologist, individual cells were edited within the ROI and replaced. The number of cells added or deleted was recorded. For cases where the HER2 score for the case was within 1.8–2.2 inclusive, the Reader Pathologist selected a second, non-overlapping ROI and applied the HER2 Dual ISH algorithm to another 20 nuclei and recorded the score from the 40 selected nuclei, in accordance with the FDA approved, CE marked VENTANA HER2 Dual ISH DNA Probe Cocktail assay.[23] Override scores at both the cell and overall score level were also captured and recorded. *HER2* gene status was defined as either amplified (HER2/Chr17 ratio $\geq$ 2.0) or non-amplified (HER2/Chr17 ratio < 2.0) in accordance with *HER2* scoring described in the VENTANA HER2 Dual ISH assay interpretation guide. *HER2* gene status was used as a reference to determine the difference in agreement rates between MR and IA modalities. The difference in positive, negative, and overall agreement rates of *HER2* gene status between IA vs GT and MR vs GT were calculated.

Inter-reader agreement for image analysis study

Inter-reader agreement was performed in addition to the primary analysis. The most common *HER2* gene status given by the 3 Reader Pathologists (e.g. if a case is determined to be positive in 2 observations and negative in 1 observation from 3 IA reads, the status mode for that case would be positive) was calculated as the modal *HER2* gene status. If the modal status for a case could not be determined, the *HER2* gene status determined during case qualification was used as the reference. Inter-reader agreement (OPA, PPA, NPA) rates for each reader was compared to the modal status.

Intra-reader reproducibility and scanner precision studies

In addition to our primary objective to assess the non-inferiority of HER2 Dual ISH IA algorithm to manual scoring, we also assessed intra-reader reproducibility and inter-scanner precision in parallel studies to evaluate the robustness of the HER2 Dual ISH IA algorithm. To assess intra-reader reproducibility, a subset of primary breast cancer cases (18 Amplified; 2 Borderline amplified; 2 Borderline-non-amplified; 18 Non-amplified) were randomly selected from the 220 case set. A single pathologist assessed the digital images for each case in 3 separate reading rounds using the HER2 Dual ISH IA. A minimum of a 15-day washout period was required between each round. Next, the pathologist assigned HER2 status for each case as either amplified (HER2/Chr17 ratio $\geq$ 2.0) or non-amplified (HER2/Chr17 ratio < 2.0) in accordance with the HER2 scoring described in the VENTANA HER2 Dual ISH assay interpretation guide. Scanner precision was determined using glass slides from the intra-reader assessment. Slides were scanned 3 times on 3 different VENTANA DP 200 slide scanners to produce digital images. Each resulting digital image was analyzed by the HER2 Dual ISH IA to assign a HER2 score (HER2/Chr17 ratio) to the case. Agreement rates between IA evaluations of multiple scans of the same HER2 DISH-stained primary breast specimens from different scanners were compared against the case modal status.

Statistical analysis

All analyses were performed using SAS (version 9.4 or older) software. Ventana biostatisticians produced all analyses and summaries using SAS version 9.4. By-observation data listings were prepared for all data collected on data forms. The planned analyses and study results at times used 'positive' and 'negative' to denote 'amplified' and 'non-amplified', respectively. The aggregated scores from all 3 Reader Pathologists were used to determine the difference of PPA and NPA point estimates for the overall difference of agreement in *HER2* gene status. The 95% CIs for these analyses were calculated using the percentile bootstrap method with cases stratified by the 4 screening bins. In order to demonstrate non-inferiority of IA to MR, the lower bounds of the difference of PPA and NPA 95% CIs had to be greater than or equal to -10%. The difference of PPA, NPA, and OPA point estimates and the difference of OPA 95% CI were calculated for descriptive purposes only. In order to demonstrate adequate inter-reader agreement, the point estimates for PPA and NPA had to be both $\geq$85%. During data analysis, a biostatistician assigned each case a *HER2* gene status (amplified or non-amplified) based on its assigned HER2 score. Scoring used the HER2 scoring categories of amplified (HER2/Chr17 ratio $\geq$ 2.0) and non-amplified (HER2/Chr17 ratio < 2.0). For analysis purposes, an amplified gene status greater than or equal to 2.0 was positive and a non-amplified gene status less than 2.0 was negative.

Randomization and bias minimization

All study cases included in the study were randomized before the pathologist's assessment began. Sample identification on glass slides were obscured and replaced by study case IDs. Several measures were employed to mitigate the risk of bias, including selection bias and recall bias. These measures included random selection of cases from tissue cohort (as specified in the study randomization plans), randomization in pathologist's reading order, a 15-day washout period for intra-reader reading sessions, a 14-day washout period between IA and MR methods, and blinding of Reader Pathologists to the previous HER2 IHC/ISH results.

Results

uPath HER2 dual ISH image analysis algorithm development

The uPath HER2 Dual ISH IA algorithm is a ranking algorithm that uses a feature set of image criteria developed through a learning approach to rank cells based on dot detection, cell segmentation, and tumor cell probability. A description of the algorithm calculation of HER2 amplification status is described in the Methods section. To evaluate the algorithm's capability to segment cells, HER2 DISH stained images were analyzed by the algorithm and reviewed by pathologists (n = 3) to determine the correct segmentation of single cells. On average, there were 17.86 cells having acceptable segmentation by 2 out of 3 pathologists, which translated to 89% consensus accuracy. Pathologists (n = 3) also evaluated the algorithm's cell selection accuracy by reviewing the algorithm performance in selecting the 20 highest ranked tumor cells, defined as tumor cells with distinct black

[2] *The VENTANA DP 200 slide scanner is CE-IVD marked. In the US, it is classified as For Research Use Only. Not for use in diagnostic procedures.*

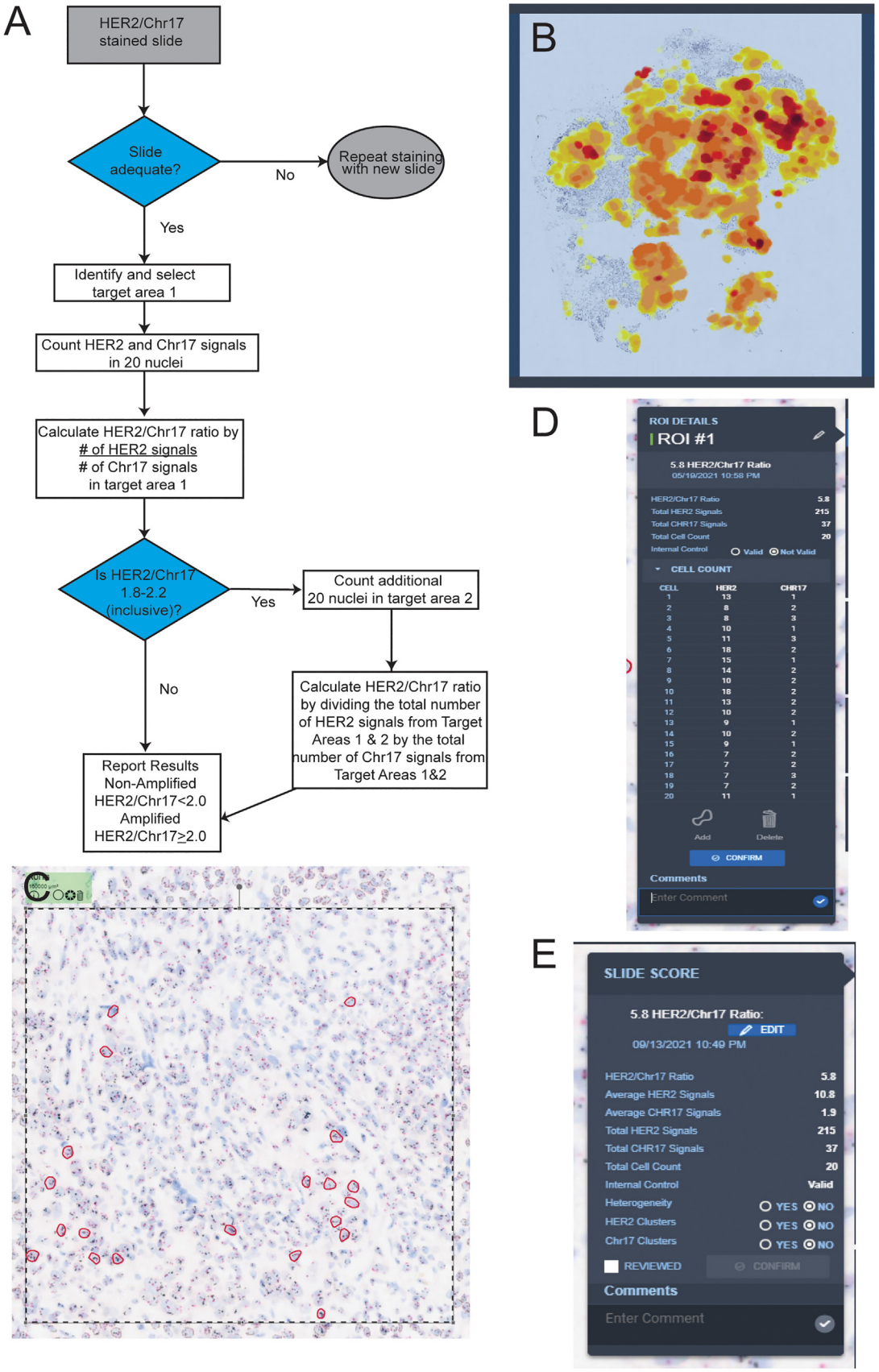

Fig. 2. Overview of the uPath HER2 dual ISH image analysis algorithm workflow. (A) Manual scoring workflow of the VENTANA HER2 Dual ISH DNA Probe Cocktail Assay. (B–E) The uPath HER2 DISH IA algorithm utilizes a heat map (B) in the slide viewer to orient the pathologist and highlight potential areas of HER2:Chr17 amplification. (C) An overlay of segmented tumor cells is provided to the pathologist to allow review of tumor cells and for placement of the region of interest (ROI) for analysis. Placement of the ROI can be performed at a magnification of 20x. Once placed, the uPath HER DISH IA algorithm will select 20 cells within the ROI and calculate the raw HER2 and Chr17 counts and the HER2:Chr17 ratio. (D) After analysis, HER2 *and* Chr17 counts for each individual cell and the calculated HER2:Chr17 ratio are displayed. The pathologist is able to review the selected cells and either remove and add new cells or confirm the analysis. (E) The result is then displayed and additional diagnostic information (e.g. heterogeneity, clusters) can be included for the final report.

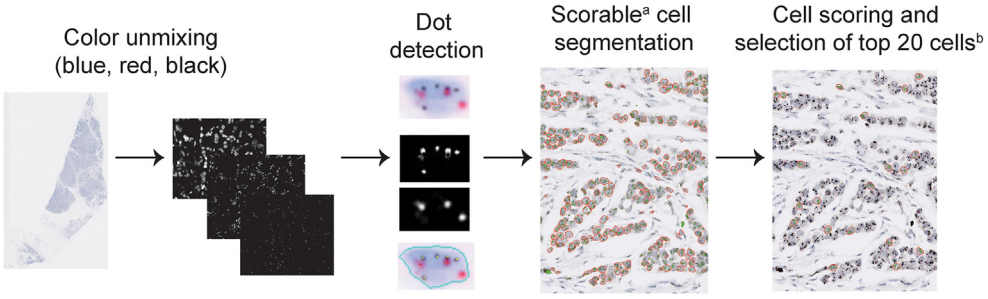

Fig. 3. Algorithm processing feature set. The algorithm's feature set is composed of several individual criteria which include: unmixing color images into blue, red, and black channels, detecting "dots" (e.g. HER2 loci and Chr17 copies) in the red and black channels, segmenting cells into scorable cells, scoring the segmentation based on quality and ranking and selecting the top 20 cells. [a]A scorable cell is defined as a cell with both red and black dots with a high segmentation confidence. [b]The top 20 cells are selected based on a high segmentation confidence, a high number of dots, and the likelihood of the cell being a tumor cell.

and red signals and containing the highest number of both black and red signals representative of that tumor area. On average, there were 16.29 cells called acceptable by 2 out of 3 pathologists, which translated to an 81% consensus accuracy. To assess the algorithm's performance in detection of individual HER2 and Chr17 signals (or dots), we compared the Coefficient of determination or R-squared (R^2) between the algorithm count (before correction) and Consensus Pathologist counts (after correction) for HER2 and Chr17 signals. On average, the R^2 for red signal detection was 0.926 and R^2 for black signal detection was 0.99, indicative of a high level of correlation. Based upon these criteria, the uPath HER2 Dual ISH IA algorithm met the pre-defined criteria for feasibility testing. The code for the algorithm was locked and deployed for the method comparison studies.

Comparison of image analysis to manual read method and ground truth

Concordance between the uPath HER2 Dual ISH IA algorithm and MR methods was evaluated by comparing each method's HER2 amplification status to GT HER2 amplification status established by three Consensus Pathologists (Fig. 1). A total of 220, FFPE breast carcinoma cases stratified into 4 case bins based on HER2 amplification status were used for these studies (Table 1). The distribution of HER2 status reflects the prevalence of amplified, borderline, and non-amplified breast carcinoma specimens in routine clinical practice.[8,18–22] The agreement rates for each Reader Pathologist were calculated in the direction of IA vs GT minus MR vs GT (Supplementary Tables 1–3). The aggregated HER2 amplification status from all 3 Reader Pathologists were used to determine the difference between MR vs GT scoring and IA vs GT scoring. (Table 2) The differences between IA and MR (IA - MR) agreement rates were 0.3% (1/321) (2-sided 95% CI:-0.9, 1.8) for PPA, -3.3% (-11/331) (2-sided 95% CI: -6.2, -1.2) for NPA, and -1.5% (-10/652) (2-sided 95% CI: -3.1, -0.2) for OPA, meeting the pre-determined acceptance criteria for the study. Overall, in terms of GT, there were 2 false negatives, where the HER2 Dual ISH IA read was negative and the MR was positive, vs 14 false positives, where HER2 Dual ISH IA was positive and the MR was negative. The net difference of false positives vs false negatives, (14 - 2 = 12) represents 12/652 = 1.8% of all observations.

Inter-reader agreement for image analysis

Inter-reader agreement for the IA reading method was assessed by comparing each reader's *HER2* gene status result for a case to the respective modal *HER2* gene status. The aggregated reads from the Reader Pathologists (n = 3) are presented in Table 3 and showed a high level of agreement for the IA reading method for readers to the modal *HER2* gene status. Similarly, the agreement for the MR *HER2* gene status compared to the inter-reader case-level was also very high with a PPA rate of 98.2% (319/325) (95% CI: 97.0, 99.4) and a NPA rate of 98.8% (328/332) (95% CI: 97.6, 99.7) (Supplementary Table 4).

Direct comparison of image analysis and manual reads agreement rates

Agreement between IA and MR was also compared directly between Reader Pathologists and to the overall (aggregated) scores (Table 4). Consistent with our other comparisons, the IA reading method displayed strong agreement to manual scoring, both as an overall aggregate score (OPA 97.5%, [95% CI: 96.0, 98.8]) and between individual readers (Reader 1-OPA 99.1% [95% CI: 96.7, 99.7], Reader 2-OPA 98.1% [95% CI: 95.3, 99.3], and Reader 3-OPA 95.4%, [95% CI: 91.7, 97.5]).

Table 2
Difference in agreement rates between manual reads vs. ground truth, and image analysis vs. ground truth.

		MR status		
GT status	IA status	Amplified	Non-amplified	Total
Amplified	Amplified	309	3	312
	Non-amplified	2	7	9
	Total	311	10	321
	MR PPA n/N (%) (95% CI)	311/321(96.9)	(94.2, 99.0)	
	IA PPA n/N (%) (95% CI)	312/321(97.2)	(95.0, 99.3)	
	Difference of PPA (IA - MR) n/N (%) (95% CI)	1/321 (0.3)	(-0.9, 1.8)	
Non-amplified	Amplified	8	11	19
	Non-amplified	0	312	312
	Total	8	323	331
	MR NPA n/N (%) (95% CI)	323/331 (97.6)	(95.8, 99.4)	
	IA NPA n/N (%) (95% CI)	312/331 (94.3)	(90.8, 97.3)	
	Difference of NPA (IA - MR) n/N (%) (95% CI)	-11/331 (-3.3)	(-6.2, -1.2)	
Overall	Amplified	317	14	331
	Non-amplified	2	319	321
	Total	319	333	652
	MR OPA n/N (%) (95% CI)	634/652 (97.2)	(95.7, 98.6)	
	IA OPA n/N (%) (95% CI)	624/652 (95.7)	(93.8, 97.5)	
	Difference of OPA (IA - MR) n/N (%) (95% CI)	-10/652 (-1.5)	(-3.1, -0.2)	

GT ground truth, *IA* image analysis, *MR* manual read; *PPA* positive percent agreement, *NPA* negative percent agreement, *OPA* overall percent agreement, *CI* confidence interval.
Note: 2-sided 95% confidence intervals were calculated using the percentile bootstrap method with cases stratified by the 4 screening bins.
Note: This analysis includes all 3 study Reader Pathologists' results.

Table 3
Agreement of IA *HER2* gene status in comparison to the inter-reader case-level.

		Modal result			Agreement		
Reader	Reader result	Amplified	Non-amplified	Total	Measurement	% (n/N)	95% CI[a]
Overall	Amplified	322	11	333	PPA	98.8 (322/326)	(97.6, 99.7)
	Non-amplified	4	318	322	NPA	96.7 (318/329)	(94.8, 98.4)
	Total	326	329	655	OPA	97.7 (640/655)	(96.6, 98.6)
Reader 1	Amplified	109	5	114	PPA	100.0 (109/109)	(96.6, 100.0)
	Non-amplified	0	105	105	NPA	95.5 (105/110)	(89.8, 98.0)
	Total	109	110	219	OPA	97.7 (214/219)	(94.8, 99.0)
Reader 2	Amplified	105	0	105	PPA	96.3 (105/109)	(90.9, 98.6)
	Non-amplified	4	109	113	NPA	100.0 (109/109)	(96.6, 100.0)
	Total	109	109	218	OPA	98.2 (214/218)	(95.4, 99.3)
Reader 3	Amplified	108	6	114	PPA	100.0 (108/108)	(96.6, 100.0)
	Non-amplified	0	104	104	NPA	94.5 (104/110)	(88.6, 97.5)
	Total	108	110	218	OPA	97.2 (212/218)	(94.1, 98.7)

IA image analysis, *PPA* positive percent agreement, *NPA* negative percent agreement, *OPA* overall percent agreement, *CI* confidence interval.
[a] For reader overall result, the confidence intervals were calculated using the percentile bootstrap method. For each reader result, the confidence intervals were calculated using the Wilson Score method. Note: Only observations with non-missing evaluation results were included in this analysis.

Table 4
Agreement between manual read and image analysis algorithm.

		Manual read result			Agreement		
Reader	IA result	Positive	Negative	Total	Measurement	% (n/N)	95% CI[a]
Overall	Amplified	317	14	331	PPA	99.4 (317/319)	(98.4, 100.0)**
	Non-amplified	2	319	321	NPA	95.8 (319/333)	(92.8, 98.2)**
	Total	319	333	652	OPA	97.5 (636/652)	(96.0, 98.8)**
Reader 1	Amplified	112	2	114	PPA	100.0 (112/112)	(96.7, 100.0)*
	Non-amplified	0	105	105	NPA	98.1 (105/107)	(93.4, 99.5)*
	Total	112	107	219	OPA	99.1 (217/219)	(96.7, 99.7)*
Reader 2	Amplified	101	3	104	PPA	99.0 (101/102)	(94.7, 99.8)*
	Non-amplified	1	111	112	NPA	97.4 (111/114)	(92.5, 99.1)*
	Total	102	114	216	OPA	98.1 (212/216)	(95.3, 99.3)*
Reader 3	Amplified	104	9	113	PPA	99.0 (104/105)	(94.8, 99.8)*
	Non-amplified	1	103	104	NPA	92.0 (103/112)	(85.4, 95.7)*
	Total	105	112	217	OPA	95.4 (207/217)	(91.7, 97.5)*

Note: Only observations with non-missing evaluation results were included in this analysis.
IA image analysis, *PPA* positive percent agreement, *NPA* negative percent agreement, *OPA* overall percent agreement, *CI* confidence interval.
[a] The confidence intervals were calculated using the percentile bootstrap method (**) or the Wilson Score method (*).

Image analysis slide level overrides

As part of the IA workflow, Reader Pathologists are allowed to override the HER2/Chr17 ratio calculated by the uPath HER2 DISH IA and enter a manual HER2/Chr17 ratio in instances where they disagree with the algorithm generated HER2/Chr17 ratio (slide-level override) as a quality check on the performance of the algorithm. To understand the disagreement between the pathologists and the uPath HER2 DISH IA, the frequency of IA slide-level overrides was calculated for each Reader Pathologist as the percentage of total cases evaluated by that Reader Pathologist using IA for which a slide-level override occurred (Table 5). A slide-level override would have occurred if the Reader Pathologist decided to ignore the IA HER2/Chr17 ratio completely and manually provided a HER2/Chr17 ratio significant enough to change the final amplification status after the HER2/Chr17 ratio was overridden. There was a single instance where Reader Pathologist 1 chose to override the IA HER2/Chr17 ratio due to disagreement with the IA results. No disagreement or slide overrides occurred for Reader Pathologists 2 and 3. In addition to the slide-level override, the uPath HER2 DISH IA allows the pathologist to delete and replace individual cells (within the ROI) selected by the algorithm (cell selection override) before algorithm computation of the HER2/Chr17 ratio. For an IA cell selection override to occur, at least one of the cells (within a single ROI) selected by IA had to be deleted and a new cell added by the Reader Pathologist. Out of a total of 693 ROIs, 472 cells (68.1%) were accepted by the Reader Pathologists (n = 3) with no cells removed or added. In cases where the Reader Pathologists did perform a cell selection override, a total of 185 ROIs (26.7%) had 5 cells or less removed and replaced. Reader Pathologist 3 had the most IA cell selection overrides followed by Reader Pathologists 2 and 1. The frequency of IA cell selection overrides were summarized for each reader per ROI and provided in Supplementary Table 5.

Table 5
IA score slide level overrides.

Description	Reader 1	Reader 2	Reader 3	Overall
	n (%)	n (%)	n (%)	n (%)
Final clinical status obtained[a]	219	218	218	655
Confirm	218 (99.5)	218 (100.0)	218 (100.0)	654 (99.8)
Override[b]	1 (0.5)	0 (0.0)	0 (0.0)	1 (0.2)
Final clinical status not obtained	1	2	2	5

IA image analysis.
[a] The final clinical score from image analysis was obtained with the reader confirmation or override.
[b] A slide-level override occurred if a reader decided to ignore the IA score completely and manually provided a slide score.

Intra-reader agreement for IA and scanner precision studies

As a secondary assessment of the performance of the uPath HER2 DISHIA, individual pathologist reads from the 3 reading rounds were used to evaluate intra-reader agreement of the HER2 Dual ISH IA (Table 6) and showed an overall OPA of 100% (120/120), (95% CI: 96.9,

Table 6
Agreement of IA *HER2* gene status in comparison to intra-reader case-level.

		Modal result			Agreement		
Round	Reader Result	Amplified	Non-amplified	Total	Measurement	% (n/N)	95% CI[a]
Overall	Amplified	60	0	60	PPA	100.0 (60/60)	(94.0, 100.0)
	Non-amplified	0	60	60	NPA	100.0 (60/60)	(94.0, 100.0)
	Total	60	60	120	OPA	100.0 (120/120)	(96.9, 100.0)
Round 1	Amplified	20	0	20	PPA	100.0 (20/20)	(83.9, 100.0)
	Non-amplified	0	20	20	NPA	100.0 (20/20)	(83.9, 100.0)
	Total	20	20	40	OPA	100.0 (40/40)	(91.2, 100.0)
Round 2	Amplified	20	0	20	PPA	100.0 (20/20)	(83.9, 100.0)
	Non-amplified	0	20	20	NPA	100.0 (20/20)	(83.9, 100.0)
	Total	20	20	40	OPA	100.0 (40/40)	(91.2, 100.0)
Round 3	Amplified	20	0	20	PPA	100.0 (20/20)	(83.9, 100.0)
	Non-amplified	0	20	20	NPA	100.0 (20/20)	(83.9, 100.0)
	Total	20	20	40	OPA	100.0 (40/40)	(91.2, 100.0)

IA image analysis, *PPA* positive percent agreement, *NPA* negative percent agreement, *OPA* overall percent agreement, *CI* confidence interval.
[a] All the confidence intervals were calculated using the Wilson Score method.

100.0). We also assessed the performance of the uPath HER2 DISH IA between 3 different scanners, which displayed considerable agreement with a PPA and NPA rate of 98.9% (95% CI: 97.8, 100.0) and 100% (95% CI: 97.8, 100.0) respectively.

Discussion

In the current study, we describe the development and validation of the uPath HER2 DISH IA as a tool for the computer-aided evaluation of the VENTANA HER2 Dual ISH DNA Probe Cocktail assay in determining HER2 amplification status in breast cancer. Using image analysis and a feature ranking technique, the uPath HER2 DISH IA automatically identifies and segments tumor and non-tumor cells, selects individual tumor cells for analysis, identifies *HER2* gene and Chr17 copies (dots), quantifies HER2 and Chr17 copies and calculates the HER2/Chr17 ratio. Critically, the algorithm relies on pathologist selection of a region of interest within the invasive breast tumor to perform these functions. Thus, accurate performance of the HER2 Dual ISH IA is inherently dependent upon the expertise of the pathologist. Using multiple pathologists and a large clinically relevant case cohort, we found that the uPath HER2 DISH IA is non-inferior to manual scoring and displays high inter- and intra-reader reproducibility. Collectively, these data suggest that the uPath HER2 DISH IA can be used in a clinical setting to support pathologists in determining *HER2* gene status.

The uPath HER2 DISH IA algorithm is intended to function as part of a complete solution for determining HER2 amplification status, working in concert with the VENTANA HER2 Dual ISH DNA Probe Cocktail assay and the BenchMark staining platform, eliminating the need for internal validation of the algorithm and staining procedure. As such, the design of the uPath HER2 DISH IA algorithm's workflow is in accordance with the CE marked/FDA approved VENTANA HER2 Dual ISH DNA Probe Cocktail assay manual interpretation. Because this study was designed to demonstrate the non-inferiority of the uPath HER2 DISH IA algorithm to manual evaluation of the VENTANA HER2 Dual ISH DNA Probe Cocktail assay, we did not perform any studies that compare against other assays (e.g. FISH) currently being used for determining *HER2* gene status. However, understanding the performance of the uPath HER2 DISH IA algorithm in relation to other approaches for determine *HER2* gene status is an important aspect of future work.

Application of image analysis algorithms as tools to help pathologists in routine time-consuming tasks, such as screening for invasive cancer, grading of prostate cancer biopsies and ISH analysis for HER2 quantification is an area of active work. A recently published study of a laboratory developed (LD) IA algorithm for HER2 DISH quantification showed that a LD IA algorithm displayed a 0.98 correlation coefficient for HER2 quantification compared with pathologist MR using a cohort of 22 breast cancers.[24] In a study of IA for FISH, authors reported moderate agreement between automated (tile image capture) vs manual FISH as measured by Cohen's Kappa (0.48 [95% CI: 0.23, 0.72]).[16] For comparison, we observed a high overall percent agreement (OPA) rate (95.7%) of image analysis vs GT, and more importantly displayed a narrow 95% confidence interval (95% CI: 93.8, 97.5). The image analysis-manual read difference in OPA was –1.5% (95% CI: –3.1, –0.2) and the inter-reader agreement rates for the uPath HER2 Dual ISH IA had point estimates of at least 96.7%. Finally, we observed that the frequency of slide level overrides significant enough to change the HER2 amplification status scored by the IA algorithm was extremely low with only 1 slide override out of 655 cases scored by the 3 pathologists. Notably our results were achieved using a large cohort of 220 FFPE samples and 3 pathologists. Overall, these data, performed with a large cohort of samples, suggest that the uPath HER2 Dual ISH IA is robust and could be used within routine clinical practice.

Increased work efficiencies is one of the commonly used reasons proposed for implementation of IA algorithms into routine clinical practice. Although our study was not designed to investigate potential gains in work efficiency, there are reports indicating that IA may help provide efficiencies in assessment of *HER2* gene status. Indeed, a recent study found that application of IA in real-world clinical practice reduced diagnostic time by 65%.[25] Likewise, Hossain and colleagues reported that the application of the LD IA for HER2 DISH also reduced diagnostic time, from 3 to 20 min per manual score compared with 1 min for the algorithm to quantify the HER2/Chr17 ratio after pathologist annotation.[24] Thus, it will be critical for future studies to address the value of the uPath HER2 DISH IA in increasing workflow efficiencies, as demonstrating reduction of diagnostic interpretation time has substantial value for both the patient and oncologist.

Future directions

Although early guidelines recommended evaluation of a minimum of 60 cells[26] to accurately ascertain *HER2* gene amplification using ISH, more recent updates have revised that number to a minimum of 20 cells (2018 and 2013 guidelines).[8,27] Despite widespread adoption of these guidelines, there is evidence that inter-reader agreement improves when larger (60 cells) numbers of cells are evaluated and could help reduce errors in diagnostic decision-making for borderline HER2 cases.[28,29] However, the need to assess ever larger numbers of cells would require a correlative increase in work for pathologists, which, when balanced with an ever-increasing workload, may not be possible. Again, the application of an IA algorithm and automation of the many tasks required for HER2 ISH assessment would directly address these workload demands, while enabling robust and accurate assessment of larger cohorts of cells. Moreover, implementation of IA to evaluate large cohorts of tumor cells may have added benefits towards improving patient diagnosis. We hope that the availability of the uPath HER2 Dual ISH IA algorithm may enable future studies into this topic.

Although evaluation of HER2 protein expression and gene amplification are routinely used to identify patients for HER2 targeted therapy, efforts have been made to utilize data regarding HER2 amplification and

expression to provide additional diagnostic insights. Data suggest that increased HER2 amplification may predict pathologic complete response in the neoadjuvant setting,[30–32] as well as predict progression-free survival in metastatic patients.[33] Similarly, characterization of HER2 intratumoral heterogeneity may act as a predictor for therapeutic response.[34–36] However, the reliance on manual interpretation for these studies has complicated their ability to be implemented clinically. The ability of image analysis algorithms to compute large amounts of data may provide an avenue for quantitative assessment of HER2 amplification and *HER2* gene heterogeneity.[37–39] The application of IA algorithms might also provide more consistent quantitative measures for determining pathologic complete response and/or intratumoral heterogeneity. As the uPath HER2 DISH IA algorithm is implemented into clinical practice, we hope that future studies will address the role of IA in the clinical workflow, as well as the ability of IA to provide deeper insights that could improve overall HER2 diagnosis.

Sources of support

All the authors are employees of Ventana Medical Systems.

Competing Interests Statement

All the authors are employees of Ventana Medical Systems. LI, JRM, TKN, CG, KZ, RV, SG and SM are shareholders in Roche Diagnostics International, Ltd.

Authors' Contributions

CG, JRM, SG, JM, KZ, TKN, SM, RV and CC performed study concept and design, analysis and generation of data; LI and AY performed interpretation of data and prepared original manuscript, CC performed data analysis. VM, CC, and CG assisted in drafting of the manuscript. All authors have read and approved the manuscript and have contributed sufficiently to the project to be included as an author.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We have no acknowledgements at this time.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpi.2022.100116.

References

1. Global Cancer Observatory. International Agency for Research on Cancer, World Health Organization. Breast Fact Sheet. Accessed June 8, 2021. Available from: https://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheet.pdf 2021.
2. Stenehjem DD, Yoo M, Unni SK, et al. Assessment of HER2 testing patterns, HER2+ disease, and the utilization of HER2-directed therapy in early breast cancer. Breast Cancer (Dove Med Press) 2014;6:169–177. https://doi.org/10.2147/BCTT.S69416.
3. Varga Z, Noske A, Ramach C, Padberg B, Moch H. Assessment of HER2 status in breast cancer: overall positivity rate and accuracy by fluorescence in situ hybridization and immunohistochemistry in a single institution over 12 years: a quality control study. BMC Cancer 2013;13:615. https://doi.org/10.1186/1471-2407-13-615.
4. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science 1987;235:177–182. https://doi.org/10.1126/science.3798106.
5. Tandon AK, Clark GM, Chamness GC, Ullrich A, McGuire WL. HER-2/neu oncogene protein and prognosis in breast cancer. J Clin Oncol 1989;7:1120–1128. https://doi.org/10.1200/JCO.1989.7.8.1120.
6. Cameron D, Piccart-Gebhart MJ, Gelber RD, et al. Herceptin Adjuvant Trial Study T. 11 years' follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive early breast cancer: final analysis of the HERceptin Adjuvant (HERA) trial. Lancet 2017;389:1195–1205. https://doi.org/10.1016/S0140-6736(16)32616-2.
7. Rugo H, Brammer M, Zhang F, Lalla D. Effect of trastuzumab on health-related quality of life in patients with HER2-positive metastatic breast cancer: data from three clinical trials. Clin Breast Cancer 2010;10:288–293. https://doi.org/10.3816/CBC.2010.n.037.
8. Wolff A, Hammond E, Allison K, Harvey B, Mangu P, Bartlett J, 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. J Clin Oncol 2018;36:2105–2122. https://doi.org/10.1200/JCO10.1200/JCO.2018.
9. Yaziji H, Gown AM. Accuracy and precision in HER2/neu testing in breast cancer: are we there yet? Hum Pathol 2004;35:143–146. https://doi.org/10.1016/j.humpath.2004.01.002.
10. Metter DM, Colgan TJ, Leung ST, Timmons CF, Park JY. Trends in the US and Canadian pathologist workforces from 2007 to 2017. JAMA Netw Open 2019;2, e194337. https://doi.org/10.1001/jamanetworkopen.2019.4337.
11. Park S, Parwani A, Aller R, et al. The history of pathology informatics: a global perspective. J Pathol Inform 2013;4:7. https://doi.org/10.4103/2153-3539.112689.
12. Wilson E. Digital Pathology Gives Rise to Computational Pathology. Accessed 5/5/2021. Available from:: https://digitalpathologyassociation.org/_data/files/2014_pathology_visions/pv14_presentations/19c_in-sourcing_workshop_cornish.pdf 2021.
13. Furrer D, Jacob S, Caron C, Sanschagrin F, Provencher L, Diorio C. Validation of a new classifier for the automated analysis of the human epidermal growth factor receptor 2 (HER2) gene amplification in breast cancer specimens. Diagn Pathol 2013;8:17. https://doi.org/10.1186/1746-1596-8-17.
14. Konsti J, Lundin J, Jumppanen M, Lundin M, Viitanen A, Isola J. A public-domain image processing tool for automated quantification of fluorescence in situ hybridisation signals. J Clin Pathol 2008;61:278–282. https://doi.org/10.1136/jcp.2007.048991.
15. van der Logt EM, Kuperus DA, van Setten JW, et al. Fully automated fluorescent in situ hybridization (FISH) staining and digital analysis of HER2 in breast cancer: a validation study. PLoS One 2015;10, e0123201. https://doi.org/10.1371/journal.pone.0123201.
16. Thakral G, Wey A, Rahman M, Fang R, Lum C. Agreement of different methods for tissue based detection of HER2 signal in invasive breast cancer. Pathol Oncol Res 2017;23:79–84. https://doi.org/10.1007/s12253-016-0091-4.
17. Ventana Medical Systems. *VENTANA HER2 Dual ISH DNA Probe Cocktail Interpretation Guide for Breast Carcinoma.* 2019.
18. Ballard M, Jalikis F, Krings G, et al. "Non-classical" HER2 FISH results in breast cancer: a multi-institutional study. Mod Pathol 2017;30:227–235. https://doi.org/10.1038/modpathol.2016.175.
19. Press MF, Sauter G, Buyse M, et al. HER2 gene amplification testing by Fluorescent In Situ Hybridization (FISH): comparison of the ASCO-College of American Pathologists Guidelines with FISH scores used for enrollment in Breast Cancer International Research Group clinical trials. J Clin Oncol 2016;34:3518–3528. https://doi.org/10.1200/JCO.2016.66.6693.
20. Press MF, Villalobos I, Santiago A, et al. Assessing the New American Society of Clinical Oncology/College of American Pathologists guidelines for HER2 testing by fluorescence in situ hybridization: experience of an academic consultation practice. Arch Pathol Lab Med 2016;140:1250–1258. https://doi.org/10.5858/arpa.2016-0009-OA.
21. Shah MV, Wiktor AE, Meyer RG, et al. Change in pattern of HER2 fluorescent in situ hybridization (FISH) results in breast cancers submitted for FISH testing: experience of a reference laboratory using US Food and Drug Administration criteria and American Society of Clinical Oncology and College of American Pathologists guidelines. J Clin Oncol 2016;34:3502–3510. https://doi.org/10.1200/JCO.2015.61.8983.
22. Stoss OC, Scheel A, Nagelmeier I, et al. Impact of updated HER2 testing guidelines in breast cancer–re-evaluation of HERA trial fluorescence in situ hybridization data. Mod Pathol 2015;28:1528–1534. https://doi.org/10.1038/modpathol.2015.112.
23. Center for Devices and Radiological Heath of the Food and Drug Administration. *VENTANA HER2 Dual ISH DNA Probe Cocktail PMA P190031 approval letter, July 28.* 2020.
24. Hossain MS, Hanna MG, Uraoka N, et al. Automatic quantification of HER2 gene amplification in invasive breast cancer from chromogenic in situ hybridization whole slide images. J Med Imaging (Bellingham) 2019;6, 047501. https://doi.org/10.1117/1.JMI.6.4.047501.
25. da Silva LM, Pereira EM, Salles PG, et al. Independent real-world application of a clinical-grade automated prostate cancer detection system. J Pathol 2021;254:147–158. https://doi.org/10.1002/path.5662.
26. Wolff AC, Hammond MEH, Schwartz JN, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. J Clin Oncol 2007;25:118–145. https://doi.org/10.1200/jco.2006.09.2775.
27. Wolff AC, Hammond MEH, Hicks DG, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. J Clin Oncol 2013;31:3997–4013. https://doi.org/10.1200/jco.2013.50.9984.
28. Polónia A, Caramelo A. HER2 in situ hybridization test in breast cancer: quantifying margins of error and genetic heterogeneity. Mod Pathol 2021. https://doi.org/10.1038/s41379-021-00813-x.
29. Polónia A, Eloy C, Pinto J, Braga AC, Oliveira G, Schmitt F. Counting invasive breast cancer cells in the HER2 silver in-situ hybridization test: how many cells are enough? Histopathology 2017;71:247–257. https://doi.org/10.1111/his.13208.

30. Arnould L, Arveux P, Couturier J, et al. Pathologic complete response to trastuzumab-based neoadjuvant therapy is related to the level of HER-2 amplification. Clin Cancer Res 2007;13:6404–6409. https://doi.org/10.1158/1078-0432.CCR-06-3022.
31. Guiu S, Gauthier M, Coudert B, et al. Pathological complete response and survival according to the level of HER-2 amplification after trastuzumab-based neoadjuvant therapy for breast cancer. Br J Cancer 2010;103:1335–1342. https://doi.org/10.1038/sj.bjc.6605939.
32. Schneeweiss A, Chia S, Hegg R, et al. Evaluating the predictive value of biomarkers for efficacy outcomes in response to pertuzumab- and trastuzumab-based therapy: an exploratory analysis of the TRYPHAENA study. Breast Cancer Res 2014;16:R73. https://doi.org/10.1186/bcr3690.
33. Giuliani R, Durbecq V, Di Leo A, et al. Phosphorylated HER-2 tyrosine kinase and Her-2/neu gene amplification as predictive factors of response to trastuzumab in patients with HER-2 overexpressing metastatic breast cancer (MBC). Eur J Cancer 2007;43:725–735. https://doi.org/10.1016/j.ejca.2006.11.019.
34. Filho OM, Viale G, Trippa L, et al. HER2 heterogeneity as a predictor of response to neoadjuvant T-DM1 plus pertuzumab: results from a prospective clinical trial. J Clin Oncol 2019;37:502. https://doi.org/10.1200/JCO.2019.37.15_suppl.502.
35. Ohlschlegel C, Zahel K, Kradolfer D, Hell M, Jochum W. HER2 genetic heterogeneity in breast carcinoma. J Clin Pathol 2011;64:1112–1116. https://doi.org/10.1136/jclinpath-2011-200265.
36. Seol H, Lee HJ, Choi Y, et al. Intratumoral heterogeneity of HER2 gene amplification in breast cancer: its clinicopathological significance. Mod Pathol 2012;25:938–948. https://doi.org/10.1038/modpathol.2012.36.
37. Bartlett AI, Starczynski J, Robson T, et al. Heterogeneous HER2 gene amplification: impact on patient outcome and a clinically relevant definition. Am J Clin Pathol 2011;136:266–274. https://doi.org/10.1309/AJCP0EN6AQMWETZZ.
38. Chang MC, Malowany JI, Mazurkiewicz J, Wood M. "Genetic heterogeneity" in HER2/neu testing by fluorescence in situ hybridization: a study of 2,522 cases. Mod Pathol 2012;25:683–688. https://doi.org/10.1038/modpathol.2011.206.
39. Hanna WM, Ruschoff J, Bilous M, et al. HER2 in situ hybridization in breast cancer: clinical implications of polysomy 17 and genetic heterogeneity. Mod Pathol 2014;27:4-18. https://doi.org/10.1038/modpathol.2013.103.