AI-Based HER2 Scoring Requires IHC Assay-Specific Training

K.B. Anderson[1,2,3], D. Wang[4], A. Saxena[4], R. Decamillis[4], A. Papine[5], F. Soussaline[6], M. Cossutta[5], M. Korsdal Stensballe[1], S. Huus Pedersen[1], N.M. Villena Salinas[1], S. Nielsen[2]

[1]Aalborg University Hospital, Department of Pathology, Aalborg, Denmark; [2]NordiQC, Aalborg, Denmark;

[3]Aalborg University, Department of Clinical Medicine, Faculty of Medicine, Aalborg, Denmark;

[4]Sakura Finetek USA, Sakura Bioscience, Department of Research and Development, Torrance, United States;

[5]Diomedes, Paris, France;

[6]FMS MedTech, Paris, France

Background and objective

Targeted treatment in breast carcinoma (BC) now includes HER2-low and HER2-ultralow categories. Artificial intelligence (AI) and image analysis may improve scoring reproducibility, but performance and accuracy may depend on the HER2 IHC assay used. We evaluated the cross-assay generalizability of a HER2 image analysis algorithm and the effects of assay-specific parameterization and threshold recalibration.

PROJECT OVERVIEW

98 breast carcinoma cases

HER2 IHC
Assay A: PATHWAY (Roche)
Assay B: EP3 (Sakura Finetek)

HER2 reference standard score
4 reviewers · 2025 ASCO/CAP criteria

Existing HER2 image analysis algorithm parameterized on Assay A for HER2-low

PART 1: CROSS-ASSAY EVALUATION

Initial evaluation
Image analysis algorithm adapted for Assay A

Model tested on
Assay A and Assay B

Binary clinical grouping
Not treatment-eligible: 0, 0+
Treatment-eligible: 1+, 2+, 3+

Compare image analysis results with HER2 reference score

PART 2: ASSAY B-SPECIFIC PARAMETERIZATION

Parameterizationfor Assay B staining characteristics

Model recalibrated
to include HER2-ultralow

Evaluation of a cohort of
92 BC cases

Updated binary clinical grouping
Not treatment-eligible: 0
Treatment-eligible: 0+, 1+, 2+, 3+

Compare image analysis results with HER2 reference score

Results of the project

Download the poster!

Materials and Methods

Ninety-eight BC cases were stained with two HER2 IHC assays according to the manufacturers' recommendations: Assay A (PATHWAY (4B5), Roche) and Assay B (EP3, Sakura Finetek, USA). Four reviewers scored the slides according to the 2025 ASCO/CAP criteria to establish a HER2 status in each BC to serve as reference score. An existing HER2 image-analysis algorithm[1,2], parameterized for Assay A for HER2-low classification was evaluated on both assays using two clinical categories: not treatment-eligible (0 and 0+) and treatment-eligible (1+, 2+ and 3+).

The model was subsequently adapted to Assay B by human parameterization, and the intensity thresholds were then recalibrated to include HER2-ultralow cases. The adapted algorithm was then evaluated in 92 BC cases. For this analysis, the decision threshold was changed to distinguish not treatment-eligible (0) from treatment-eligible (0+, 1+, 2+ and 3+), thereby including ultralow (0+) in the treatment-eligible group.

Scan the QR code for detailed documentation of the digital image-analysis workflow and assay-specific parameterization.

Results

Initial image analysis algorithm performance

For HER2-low classification, agreement between the image-analysis algorithm and the consolidated reference standard was higher for Assay A than for Assay B (94% compared to 81%).

Effect of assay B specific parameterization and threshold recalibration to HER2-ultralow

Assay B specific parameterization increased agreement from 81% to 91%.

Recalibrating the intensity threshold to include HER2-ultralow cases resulted in 92% agreement for Assay B. This reduced false-negative cases from 7 to 1 but increased false-positive cases from 1 to 6.

When the Assay B adapted and recalibrated algorithm was applied to Assay A, agreement decreased from 94% to 88%.

RESULTS OVERVIEW

Existing HER2 image analysis algorithm parameterized on Assay A for HER2-low

PART 1: CROSS-ASSAY EVALUATION

Assay A: 94% agreement
Assay B: 81% agreement

PART 2: ASSAY B-SPECIFIC PARAMETERIZATION

Assay B-specific parameterization

Assay A: 92% agreement
Assay B: 91% agreement

Threshold recalibration to include HER2 ultralow

Assay A: 88% agreement
Assay B: 92% agreement

Table 1. Agreement changed after assay-specific image-processing parameterization and ultralow-extended threshold selection, with improved agreement for Assay B but reduced agreement for Assay A when the adapted workflow was cross-assay evaluated.

Workflow / comparison	Classification threshold	Assay A PATHWAY	Assay B EP3
Initial evaluation original algorithm adapted on Assay A	HER2-low 0/0+ vs 1+/2+/3+	94%	81%
After Assay B-specific image-processing parameterization	HER2-low 0/0+ vs 1+/2+/3+	92%	91%
After Ultralow-extended threshold selection with Assay B-adapted workflow	HER2-ultralow included 0 vs 0+/1+/2+/3+	88%	92%

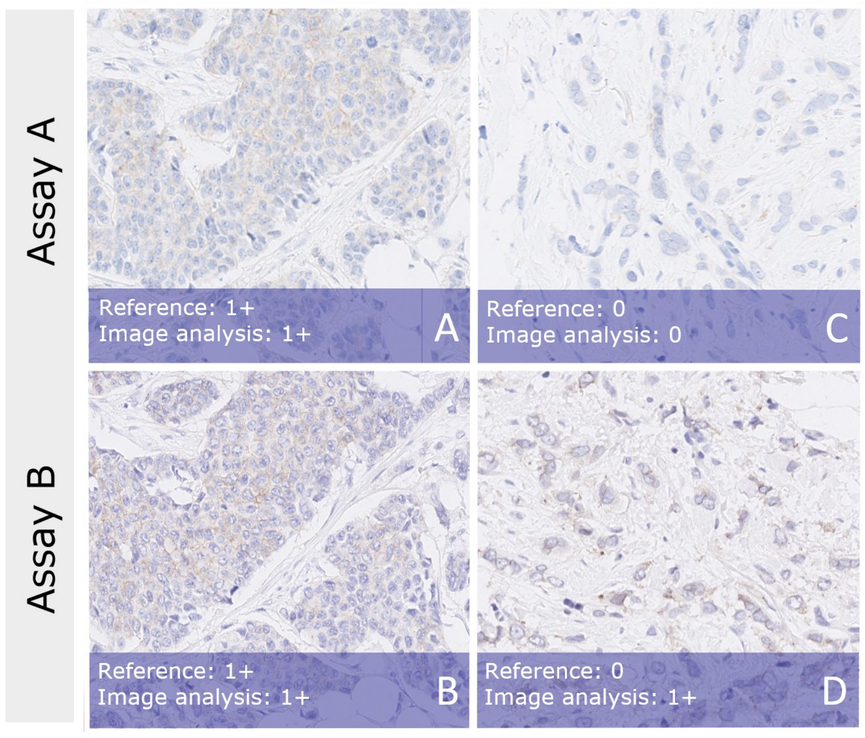

Figure 1. Representative assay-dependent image analysis performance before Assay B specific parameterization.
A and B show the same breast carcinoma stained with Assay A and Assay B, respectively. The carcinoma was scored as HER2-low 1+ both by reference and image analysis scoring. C and D show another carcinoma:
With Assay A the carcinoma was scored HER2 0 by both reference and image analysis scoring, while Assay B was scored as HER2 0 by reference scoring but HER2-low 1+ by image analysis, representing a false-positive result. This example illustrates the need for assay-specific model adaptation.

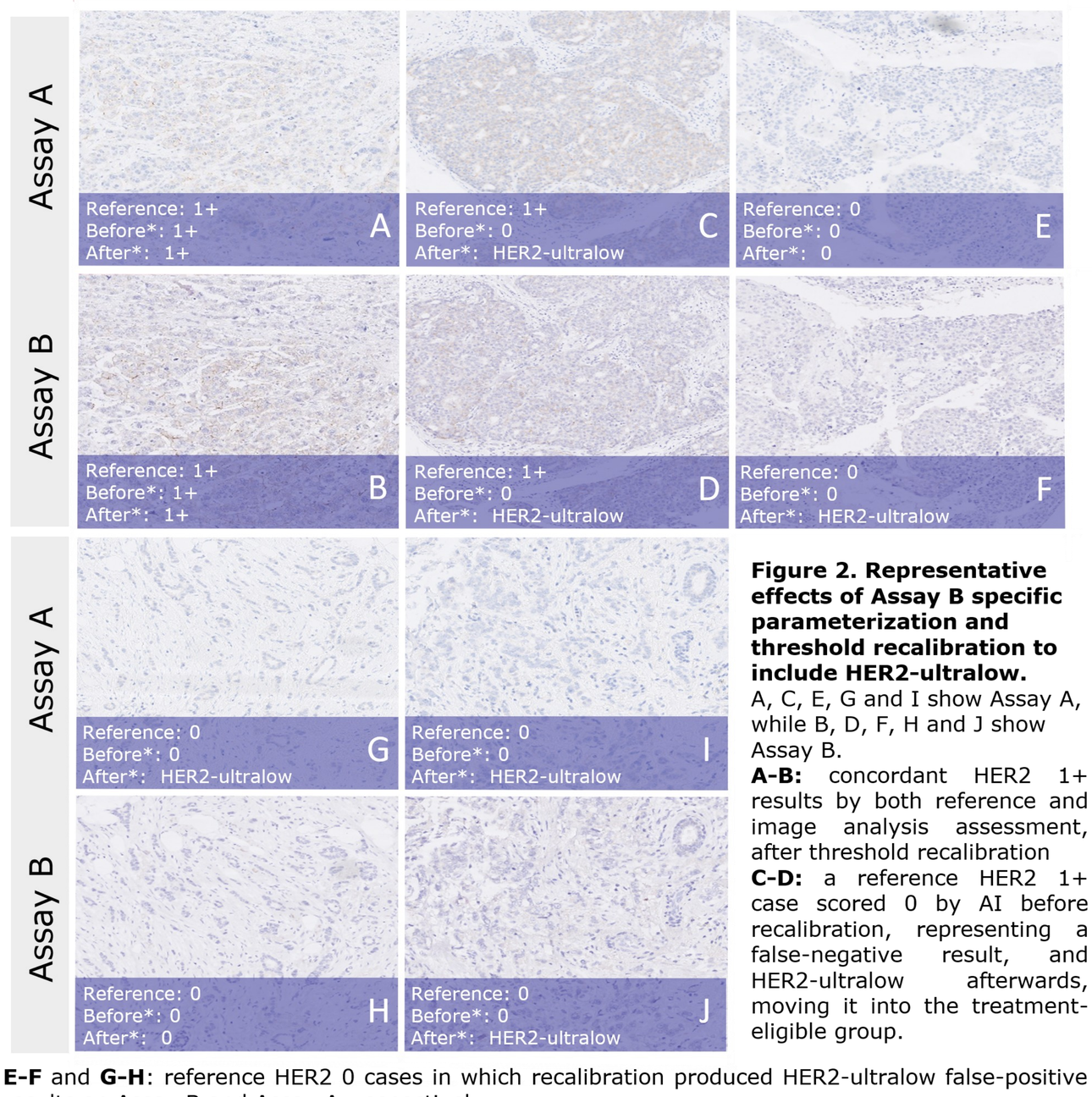

Figure 2. Representative effects of Assay B specific parameterization and threshold recalibration to include HER2-ultralow.
A, C, E, G and I show Assay A, while B, D, F, H and J show Assay B.
A-B: concordant HER2 1+ results by both reference and image analysis assessment, after threshold recalibration
C-D: a reference HER2 1+ case scored 0 by AI before recalibration, representing a false-negative result, and HER2-ultralow afterwards, moving it into the treatment-eligible group.
E-F and **G-H**: reference HER2 0 cases in which recalibration produced HER2-ultralow false-positive results on Assay B and Assay A, respectively.
I-J: reference HER2 0 cases classified as HER2-ultralow by image analysis on both assays after recalibration. Overall, recalibration for HER2-ultralow reduced false-negative results but increased false-positive results.

* Image analysis results before and after threshold recalibration to include HER2-ultralow

CONCLUSIONS

HER2 image analysis algorithm performance was assay dependent.

Initial HER2 image analysis showed highest agreement for Assay A being used to train the algorithm

Assay B-specific parameterisation was associated with higher agreement with Assay B and gave reduced agreement for Assay A.

The threshold recalibration changed the balance between false-negative and false-positive results.

When the Assay B-adapted and recalibrated model was applied to Assay A, that had shown the best initial performance, agreement decreased from 94% to 88%.

These findings highlight the need for assay-specific algorithm setup, independent validation and verification at the intended clinical decision threshold before deployment.

1. Cossutta M, Truumees B, Papine A, Kristoffersen HL, Tatarinova E, Emanuelsen L, Homsy C, Soussaline F, Nielsen S. Computer-Aided Quantitative Image Analysis of Multiple Human Epidermal Growth Factor Receptor 2 Immunohistochemistry Assays in Breast Carcinoma. Arch Pathol Lab Med. 2026 Apr 14;150(8):610-619. doi: 10.5858/arpa.2025-0316-OA. PMID: 41980744.
2. Dabbs DJ., Torlakovic E., Nielsen S., Parry SC., Yu J., Stoos C., Clark B., Høeg H., Thagaard J., Sompuram SR., Naber SP., Yagi Y., Sayre J., Vani K., Cossutta M., Soussaline F., Papine A., t'Hart NA., Szabolcs MJ., Jasani B., Kinloch M., Chiriboga L., Miller K., Bogen S. New standards in HER2 low testing: the CASI-01 comparative methods study. EBioMedicine 2025;120:105919

Co-authors affiliated with Sakura Finetek, USA, Diomedes and FMS MedTech contributed technical and methodological expertise. They did not participate in or influence HER2 interpretation, establishment of the reference standard or data analysis. The authors declare no additional conflicts of interest.