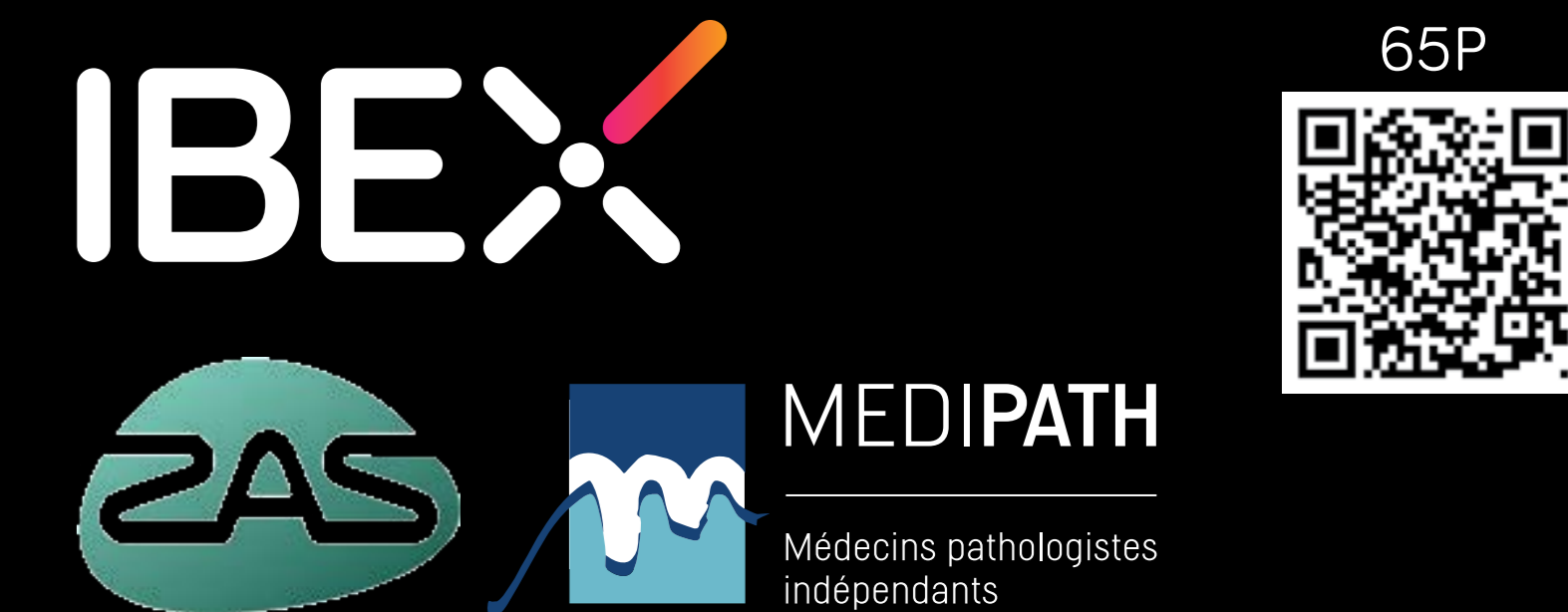


Improved Accuracy and Consistency in Breast Cancer HER2 Scoring Using an AI-Based Diagnostic Solution: IBE²



Frederik Deman^{1,2}, Michel Vandenberghe³, Maya Grinwald⁴, Tsuyoshi Matsuo³, Marios Gavrielides⁵, Cheng Cui³, Isabell Loftin⁵, Jack Henry Rawson³, Jimena Trillo Tinoco⁵, JD Alvarez⁵, Charo Garrido⁵, Sabine Declercq¹, Glenn Broeckx¹, Jeanne Thomassin⁶, Rinat Kiperman⁴, Raz Ziv⁴, Shahar Aizenbud Sagal⁴, Giuseppe Malle⁴, Judith Sandbank⁴, Manuela Vecsler⁴
¹Department of Pathology, Ziekenhuis aan de Stroom (ZAS), Antwerp, Belgium, ²Department of Diagnostic Sciences, Faculty of Medicine and Health Sciences, University of Ghent, Belgium; ³AstraZeneca, Cambridge, UK; ⁴Ibex Medical Analytics, ⁵Daichi Sankyo, Basking Ridge, NJ, USA; ⁶MediPath, Frejus, France

BACKGROUND

HER2 constitutes a significant prognostic and predictive biomarker in breast carcinoma (BC), which undergoes routine assessment in all patients diagnosed with invasive BC. The recently demonstrated effectiveness of HER2-targeted therapy in HER-low (IHC 1+ or IHC 2+/ISH-) expressing tumors emphasizes the necessity for a precise and standardized HER2 scoring process¹².

OBJECTIVE

This retrospective validation study aims to determine whether an artificial intelligence (AI) solution contributes to pathologists' consistency and accuracy in interpreting HER2 scores, according to 2023 ASCO/CAP guidelines, in BC core needle biopsies and excisions. This clinical accuracy study supported CE-IVDR certification.

METHODS

Study design Two-arm multi-site multi-reader study:

- 370 invasive breast carcinoma slides from 2 sites scanned at 40X
- 4 reader pathologists performed HER2 scoring with and without AI assistance thus each slide was reviewed by 2 pathologists blindly
- Ground truth (GT) established independently by a majority of 2 out of 3 expert breast pathologists (a total of 6 breast experts participated thus each slide was reviewed by 3 experts blindly.)
- HER2 was scored according to the 2023 ASCO/CAP guidelines⁴ without additional clinical considerations
- Both study arms reports were compared to GT
- Ibex Breast HER2³ solution was developed based on fully supervised, multilayered convolutional neural networks (CNNs) designed to detect the invasive tumor area, classify tumor cells based on their staining pattern and derive a slide-level HER2 score by applying ASCO/CAP guidelines⁴
- Average agreement with GT i.e. accuracy, and inter-observer agreement was measured for both arms.

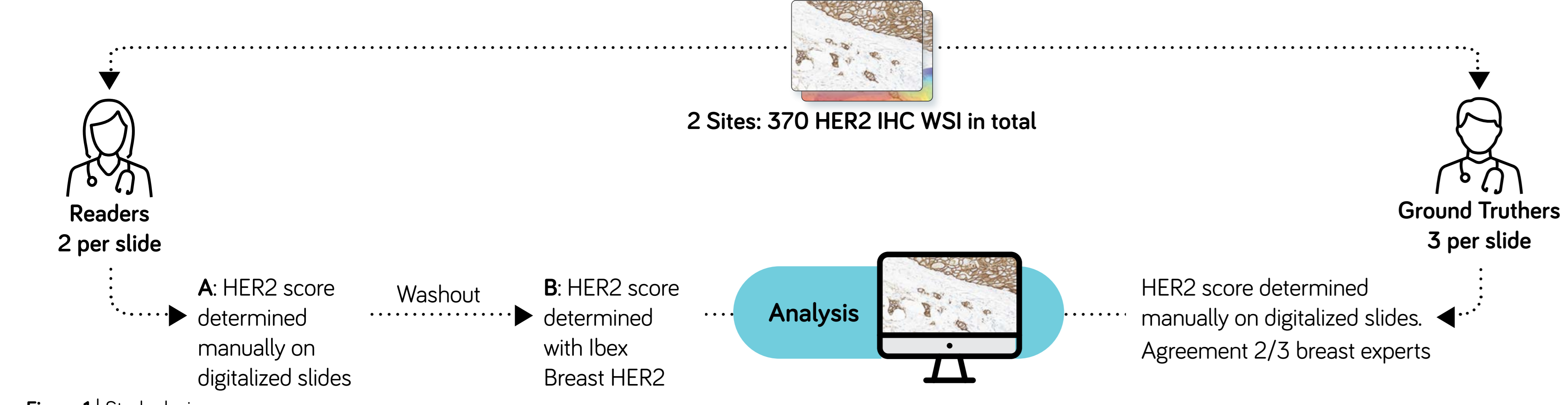


Figure 1 | Study design

Study Cohort

The cohort included 370 biopsies and excisions with different BC subtypes from 2 different European centres. HER2 slides were stained with anti-HER2 antibody (4B5, VENTANA) and scanned with different scanners (Leica Aperio, Philips UFS).

		Site 1	Site 2	Total
Scanners		Philips UFS	Leica Aperio	
# Slides		176	194	370
Age (Mean, years)		68	64	66
Gender	Female	176	191	367
	Male	0	3	3
Procedure	Biopsy	175	91	266
	Excision	1	103	104
Invasive Carcinoma type	IDC	133	162	295
	ILC	32	29	61
	Other	11	3	14
HER2 score by original report	0	66	59	125
	+1	63	69	132
	+2	34	39	73
	+3	13	27	40

Table 1 | Study Cohort

RESULTS

HER2 AI standalone performance

- The AI solution demonstrated high accuracy in HER2 scoring: 92.90% for 0 vs. 1+/2+/3+ and 85.52% overall
- The AI showed similar levels of accuracy across all labs, scanners and invasive carcinoma types (not shown here).

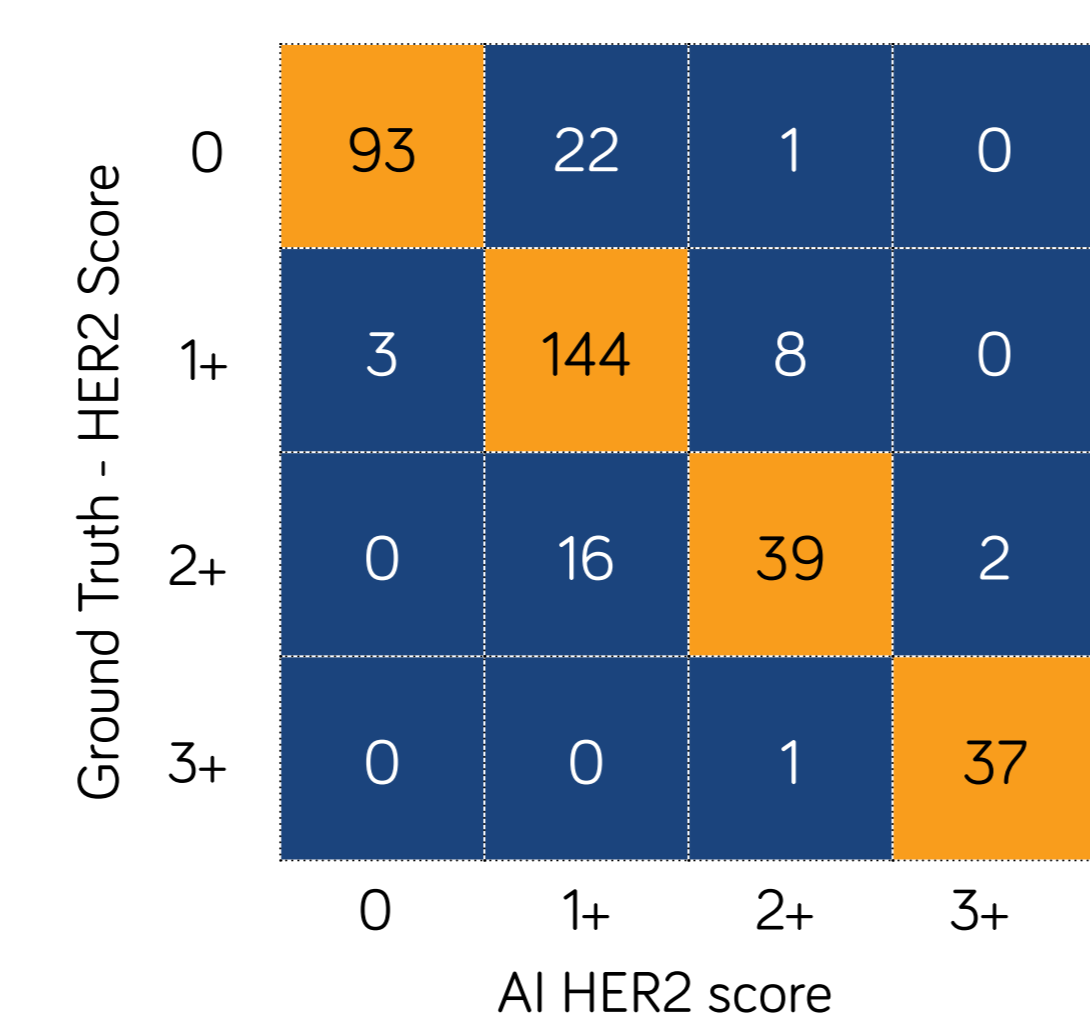


Figure 2 | AI performance | The confusion matrix was calculated for AI score vs. GT for all four HER2 scores.

Pathologists' performance - Accuracy

- Readers' overall agreement with GT was significantly higher when assisted by AI (86.6%) than without AI (77.5%), p<0.05.

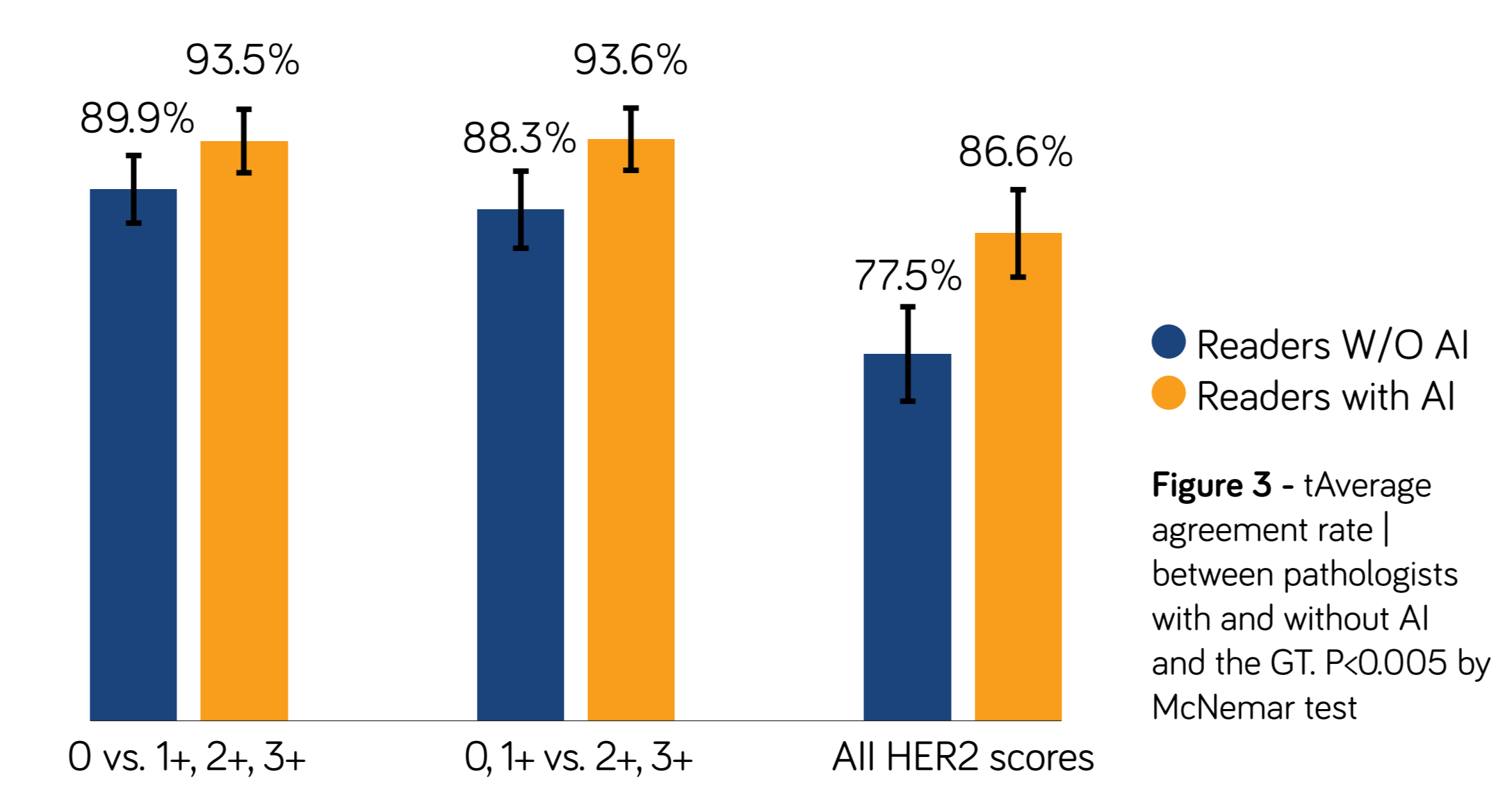


Figure 3 - Average agreement rate | between pathologists with and without AI and the GT. P<0.0005 by McNemar test

Inter-observer agreement - Consistency

- Readers' overall inter-observer agreement was significantly higher when assisted by AI.

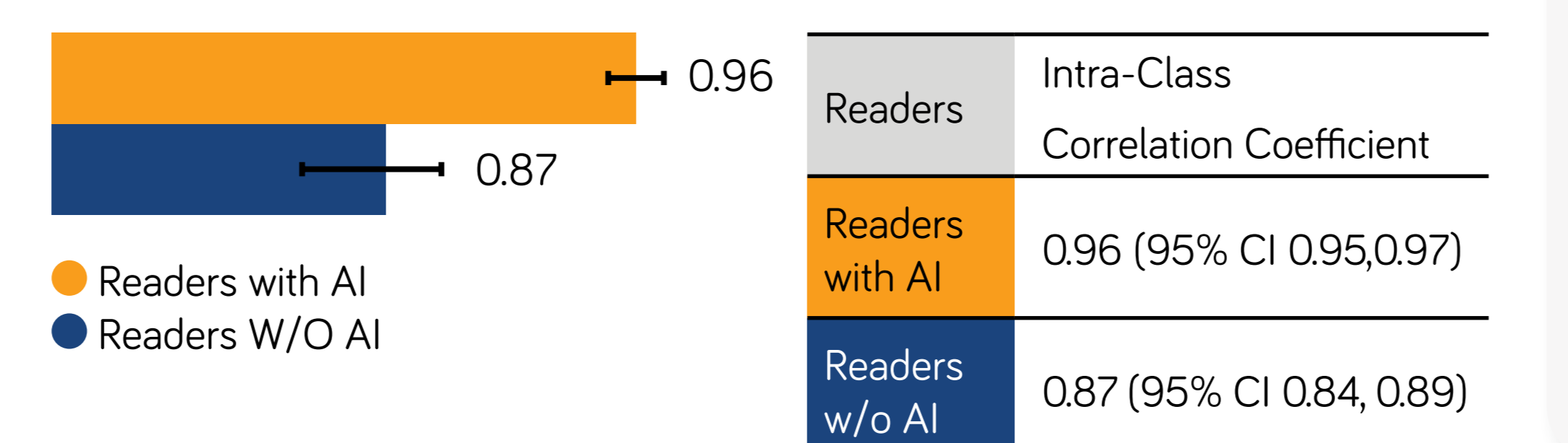


Figure 4 | Inter-observer agreement

Readers	Intra-Class Correlation Coefficient
Readers with AI	0.96 (95% CI 0.95,0.97)
Readers w/o AI	0.87 (95% CI 0.84, 0.89)

Table 2 | Inter-observer agreement. *95% CI do not overlap

Examples of pathologists HER2 scoring with and without AI

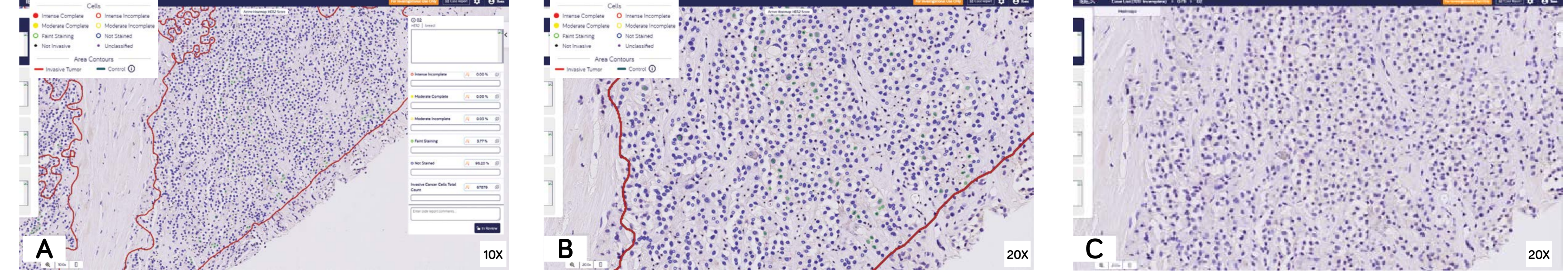


Figure 5 | HER2-ultralow (IHC 0 with membrane staining) slide; A. 10X with AI, B. 20X with AI, C. 20X without AI. AI could assist in identifying HER2-ultralow slides when the % of stained cells is below the 10% cutoff and hard to assess. ILC Pleomorphic carcinoma biopsy, GT HER2 scores = 0,0,0, Readers W/O AI scores = 1+,0, Readers with AI scores = 0,0, AI = 3.77% Faint complete stained cells.

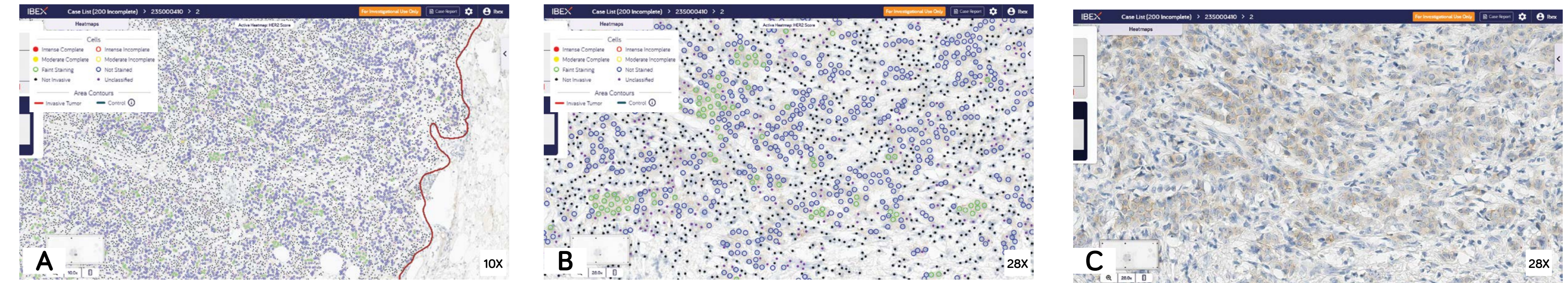


Figure 6 | HER-low (IHC 1+ or IHC 2+/ISH-); A. 10X with AI, B. 28X with AI, C. 28X without AI. AI could assist in identifying HER2-low slides when the % of stained cells is slightly above the 10% cutoff. IDC carcinoma excision, GT HER2 scores = 0,1+,1+, Readers W/O AI scores = 1+,0, Readers with AI scores = 1+,1+, AI = 14.49% Faint complete stained cells.

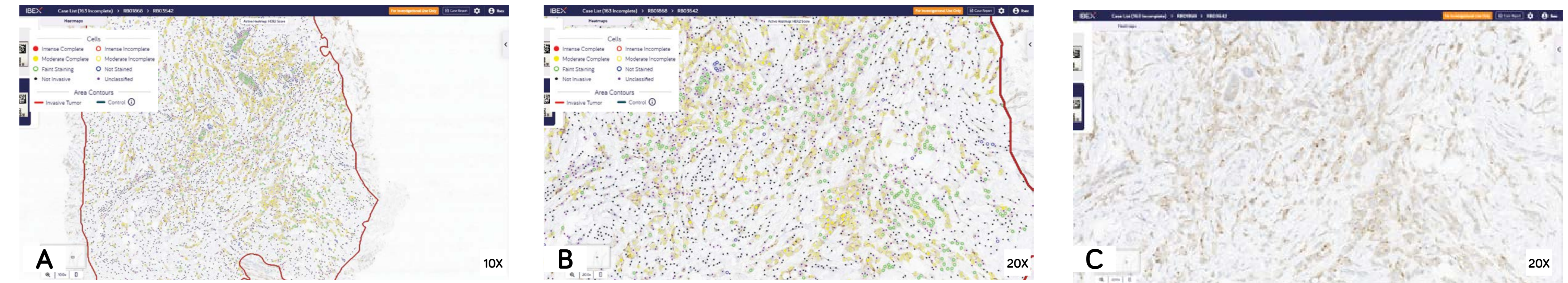


Figure 7 | HER-low slide; A. 10X with AI, B. 20X with AI, C. 20X without AI. AI could assist in distinguishing between 1+ and 2+ slides. ILC carcinoma biopsy, GT HER2 scores = 1+,1+,2+, Readers W/O AI scores = 2+,2+, Readers with AI scores = 1+,1+, AI = 28.78% Faint incomplete, 46.9% Moderate incomplete and 8.96% Moderate complete stained cells.

User Survey 100% of pathologists report they feel

- Consistent** in their HER2 scoring when assisted by the AI
- Confident** about their scoring accuracy when assisted by the AI application
- Satisfied** with the application. It was useful and assisted them in scoring borderline cases

Figure 8 | User feedback survey completed by the 4 reader pathologists

CONCLUSIONS

- Pathologists supported by AI showed significant improvements in HER2 scoring overall accuracy and consistency for all analyzed clinical cut-offs.
- AI solutions could be used as decision-support tools for pathologists, enhancing the reproducibility and consistency of HER2 scoring and may support optimal treatment pathways and better patient outcomes.

Disclosures | Dr. Deman confirms there is no conflict of interest to declare. M. Grinwald, R. Kiperman, R. Ziv, S. Aizenbud Sagal, G. Malle, J. Sandbank, M. Vecsler are employees of Ibex Medical Analytics and have an ownership interest in stock options. M. Vandenberghe, T. Matsuo, C. Cui, J. H. Rawson are employees of AstraZeneca and have an ownership interest in stock/stock options. M. Gavrielides, I. Loftin, J. Trillo Tinoco, J. D. Alvarez, C. Garrido are employees of Daichi Sankyo and have an ownership interest in stock/stock options.
 Acknowledgments | This study is sponsored by Ibex Medical Analytics and AstraZeneca Pharmaceuticals in collaboration with Daichi Sankyo, Inc.
 Copies of this poster obtained through QR, AR and/or text key codes are for personal use only and may not be reproduced without written permission of the authors

References: 1. Modi et al, N Engl J Med, 2022; 387:9; 2. Mosele et al, Nat Med 2023; 29, 2110–2120; 3. Globerson et al, Cancer Res 2023; 83:P6-04-05; 4. Wolff et al, Arch Pathol Lab Med. 2018, 142(11):1364-1382