Poster 287P

COMPARATIVE EVALUATION OF 7 ARTIFICIAL INTELLIGENCE (AI) TOOLS FOR HER2-LOW AND ULTRALOW SCORING IN BREAST CANCER

Jean-Sébastien Frenel[1], Véronique Verriele-Beurrier[1], Tanguy Perennec[1], Antoine Gbessemehlan[2], Olivier Kerdraon[1], Mario Campone[1], Loig Vaugier[1,9], Irène Spiridon[1], Leslie Tessier[1], Khouloud Ben Lazreg[1], François Bocquet[1], Michel E Vandenberghe[3], Ross J Hill[3], John W Longshore[4], Olivier Bourgeois[5], Aurélie Cottard[5], Thomas Di Maio[6], Frédérique Penault-Llorca[7], Camille Franchet[8]

1. Institut de Cancerologie de L'Ouest, Saint-Herblain, France , 2. Digital Medical Hub, 75004 Paris, France, 3. AstraZeneca, Cambridge, United Kingdom, 4. AstraZeneca, Durham, NC, USA, 5. AstraZeneca, La Defense, France, 6. AstraZeneca, Baar, Switzerland, 7. Department of Pathology, Centre Jean Perrin, Clermont-Ferrand, France, 8. Department of Pathology, Univ Toulouse, Oncopole Claudius Regaud, IUCT-Oncopole, Toulouse, France, 9. Groupe Hospitalier Bretagne Sud, Lorient, France

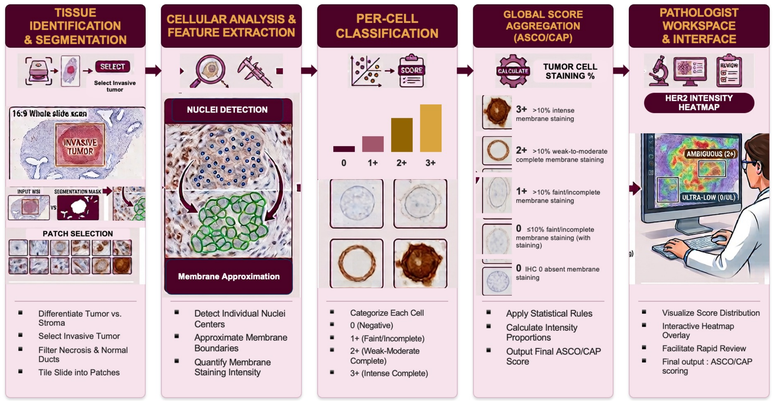
Objective: Assess the AI Tools for HER2 scoring available on the market

- Comparative evaluation of the diagnostic performance of 7 CE-marked (or CE marking in-process) AI tools from different continents on the same digitised slides.

Conclusions

- AI tools may improve the accuracy of HER2-low and HER2-ultralow assessment by pathologists, potentially informing treatment decision-making.
- Persistent differences in staining intensity, pattern and percentage, including between expert references, support the use of AI as an interpretive aid rather than a substitute for pathologists.
- Variability in performance of AI tools indicates that selection and validation of AI tools are important for successful implementation of AI support for HER2-low and HER2-ultralow assessment.

ARTIFICIAL INTELLIGENCE STEPWISE APPROACH TO SCORE HER2 IN BREAST CANCER



This study is sponsored by Institut de Cancérologie de l'Ouest (ICO) and AstraZeneca. In March 2019, AstraZeneca entered into a global development and commercialization collaboration agreement with Daiichi Sankyo for trastuzumab deruxtecan (T-DXd; DS-8201).

Introduction

- T-DXd has expanded clinically relevant HER2 categories beyond the historic positive/negative dichotomy to HER2-low (IHC 1+ or 2+/ISH−), HER2-ultralow (IHC 0 with membrane staining) and HER2 null (IHC 0 absent membrane staining).
- Due to potential treatment eligibility, discrimination at the lowest HER2 expression levels has become clinically meaningful.
- Manual scoring of HER2 expression is subject to inter-observer variability, potentially affecting treatment selection for patients with breast cancer
- AI-assisted scoring may enhance harmonization in HER2 status assessment, while maintaining its consistency with independent expert consensus.

Methods

ICO (Institut de Cancérologie de l'Ouest) Cohort

- n = 200
- 120 biopsies (60%), 80 surgical specimens (40%)
- 159 primary BC (79.5%), 41 metastatic BC (20.5%)
- Staining with PATHWAY HER2 (4B5) Assay (Roche Diagnostic Solutions, Tucson, AZ, USA)
- Leica Aperio GT 450, definition 40×, 0.26 µm/pixel
- 2 References: Consensus of 3 expert pathologists from the GEFPICS (France) & LabCorp (Switzerland)

IUCT (Institut Universitaire du Cancer de Toulouse) -Oncopole Cohort

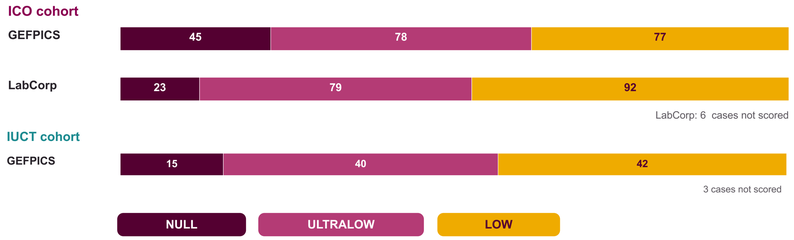
- n = 100
- 60 biopsies (60%), 40 surgical specimens (40%)
- 74 primary (74%), 26 metastases (26%)
- Staining with PATHWAY HER2 (4B5) Assay (Roche Diagnostic Solutions, Tucson, AZ, USA)
- 3DHISTECH Pannoramic 250, definition 20×, 0.24 µm/pixel
- Reference : Consensus of 2 expert pathologists from the GEFPICS (France)

Endpoints

- Primary endpoint: agreement (weighted κ) of each tool with (1) GEFPICS expert consensus and (2) LabCorp.
- Secondary endpoint : Ranking: Top two tools based on a composite score comprising 75% main-cohort performance, 20% external validation, and 5% interpretability.
- AI tool manufacturers remained anonymous throughout the project.

Results

Reference score distribution



- K ranged from 0.59 to 0.79 in the main cohort and reached 0.78 in the external validation cohort according to GEFPICS. Tools C and E had the highest composite scores.
- Agreement in the external cohort (κ = 0.78 and 0.74 for the best tools) exceeded the agreement observed between the two expert references themselves.
- HER2-low vs non-low was well discriminated, whereas HER2-ultralow vs other categories was more challenging.

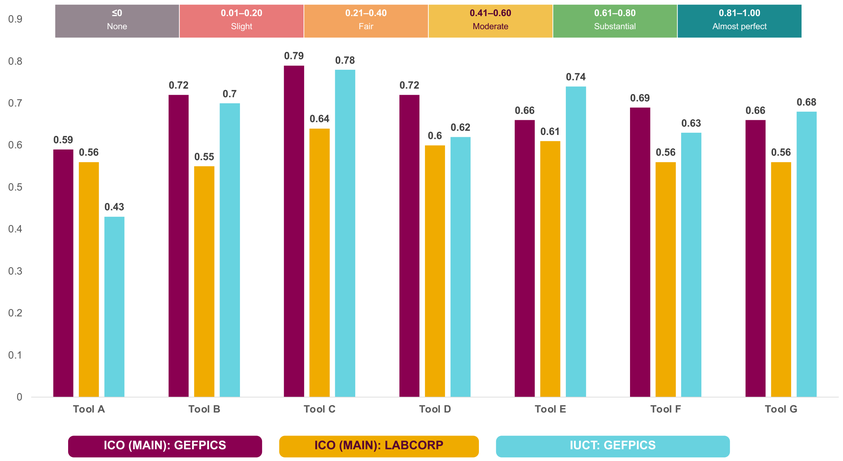
Top composite scores

Tool C: 0.66 **Tool E: 0.59**

Both exceeded the expert-consensus vs LabCorp benchmark.

Tool C and Tool E delivered the best overall performance, with demonstrated robustness across external cohort.

Global agreement - Weighted κ by AI tool and reference



Best performing AI tool vs GEFPICS (Confusion matrix, ICO cohort)

	GEFPICS NULL	GEFPICS ULTRALOW	GEFPICS LOW
Tool C NULL	29 (64.4%)	3 (3.8%)	0 (0.0%)
Tool C ULTRALOW	14 (31.1%)	64 (82.1%)	9 (11.7%)
Tool C LOW	2 (4.4%)	11 (14.1%)	68 (88.3%)

κ = 0.79 • SUBSTANTIAL AGREEMENT

- Strongest diagonal agreement: ultralow 64/78 and low 68/77.
- Most residual discordance lies at adjacent category borders, especially null ↔ ultralow.
- The matrix shows where κ is gained—and where clinically difficult boundaries remain.

Abbreviations

AI, artificial intelligence; IHC, immunohistochemistry; ISH, in situ hybridization; κ, kappa; T-DXd, trastuzumab deruxtecan. GEFPICS : Groupe d'Etude des Facteurs Pronostiques et Immunohistochimiques dans le Cancer du Sein

Acknowledgments

We thank the AI tool providers for their participation in the study; Mathieu Grajoszex and Rebecca Serruya, for their contributions to discussions and reflections on the study design and for their project support; Lisa Gauquelin, for her involvement in data management and statistical analysis; Fatima-Zahra El Barche, for her support with the study report and protocol writing; Morgane Menard and Antoine Guyonvarch, for their involvement in project management, logistics, and data collection

Disclosures

OB, JWL, AC, MEV, RHJ, and TdM are full-time employee of AstraZeneca and hold equity/shares in AstraZeneca. JWL is the current chair of the Precision Cancer Consortium. LT declares honoraria from Primaa for and advisory role. LV declares honoraria from Bayer, Ipsen, Astra Zeneca, BMS, Johnson and Johnson, Novartis, MSD, Recordati and Accord Healthcare for public speaking. LV declares direct research funding from Research Labs for radiotherapy delineation. LV declares financial support from Astra Zeneca for lung cancer project. LV is merbmer of ESTRO working group for radiation induced heart toxicity. MC declares honoraria from Novartis, BeOne Pfizer and Lilly fr Advisory Role. MC is member of Scientific Advisory Board Of National Cancer Institute. TdM is Board Member of GMT Science and Shareholder of Kestra Medical. TdM was a speaker at Sophia Genetics Events. VV and CF declares honoraria from AstraZeneca. VV declares honoraria from MSD.