Evaluation of Cell Lines in Combination with Artificial Intelligence as Quality Tools to Monitor HER2 IHC Test Reproducibility

Birgit Truumees[1], Jeppe Thagaard[2], Stine Amtoft Nielsen[2], Mateusz Tylicki[2], Lise Emanuelsen[1], Søren Nielsen[1]

Affiliation: [1]NordiQC, Department of Pathology, Aalborg University Hospital, Denmark, [2]Visiopharm A/S, Denmark

Disclosures: Birgit Truumees: None; Jeppe Thagaard: Employee, Visiopharm A/S; Stine Amtoft Nielsen: Employee, Visiopharm A/S; Mateusz Tylicki: Employee, Visiopharm A/S; Lise Emanuelsen: None; Søren Nielsen: None.

Background

Human epidermal growth factor receptor 2 (HER2) is an important immunohistochemical (IHC) biomarker for breast cancer (BC) to determine the eligibility of HER2-targeted therapies both for classical HER2 overexpression and HER2-low status. For correct treatment decision, accurate and precise HER2 IHC testing is fundamental. In this study, we evaluated if cell lines with relevant and critical expression levels of HER2 in combination with artificial intelligence (AI) could be used to identify inaccurate HER2 IHC assays.

Objectives

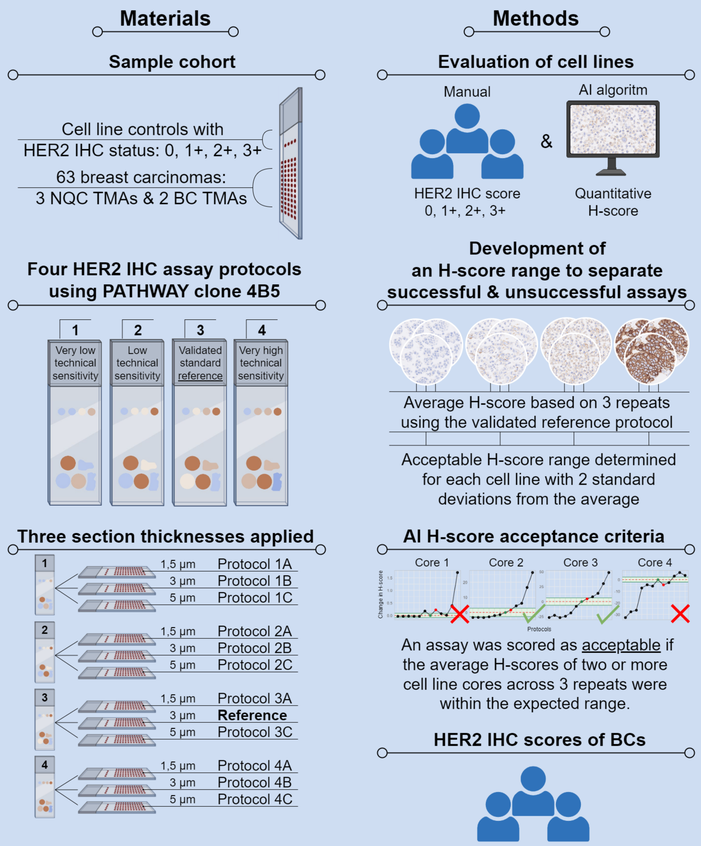
- Evaluation of cell lines to predict HER2 IHC assay accuracy.
- Impact on HER2 IHC scores in BC using IHC assays with different technical sensitivities.
- Investigation of HER2 IHC scores in relation to slide thickness of cell lines and BCs.
- Comparison of manual and artificial intelligence read-out as cell line based quality control.

References

- Røge, R., Nielsen, S., Riber-Hansen, R., Vyberg, M. (2021). Image analyses assessed cell lines as potential performance controls of Ki-67 immunostained slides. *Applied Immunohistochemistry & Molecular Morphology*, 29(2), 95-98.
- Aung, T. N., et al. (2021). A new tool for technical standardization of the Ki67 immunohistochemical assay. *Modern Pathology*, 34(7), 1261-1270.

Design

The manual and AI read-outs of passed versus failed cell lines were compared to HER2 scores in BCs for the ability to separate accurate and inaccurate HER2 IHC assays. A reference protocol and protocols with different “forced errors” were applied on all samples. Manually evaluated cell lines were passed if the expected HER2 level in all 4 cores was obtained and failed if one or more of the cores showed a change in the expected HER2 IHC score. Cell lines were scored manually by 3 reviewers and by AI using Qualitopix (Visiopharm), whereas BCs were only scored manually.



Results

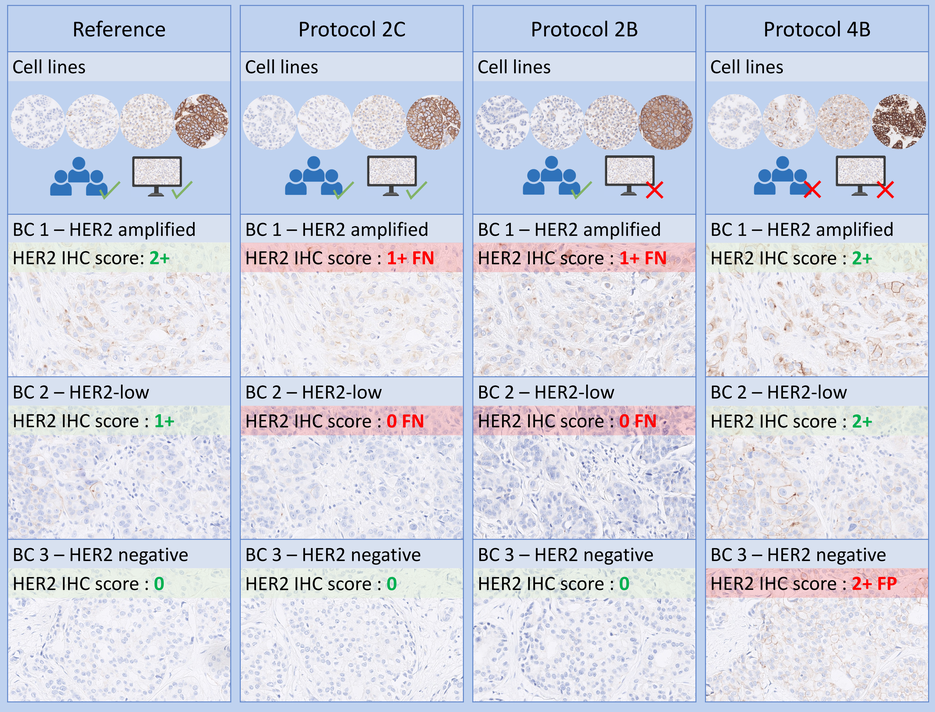
Changes in the expected HER2 expression levels in cell lines was correlated to corresponding changes of HER2 IHC scores in BCs (see Table 1). A range of concordance and disconcordance between manual and AI supported read-out of cell lines was seen to identify accurate versus inaccurate HER2 IHC assays (see Figure 1). IHC assays with a reduced technical sensitivity provided a lower proportion of HER2-low and HER2 classical positive BCs. IHC assays with increased technical sensitivity especially increased the number of HER2-low BCs (see Graph 1). The sensitivity using AI to separate accurate versus inaccurate HER2 results in cell lines was superior to manual read-out for both HER2 classical and HER2-low (see Table 2).

Table 1. Correlation of HER2 IHC scores in cell lines and BCs

		Prot. 1A	Prot. 1B	Prot. 1C	Prot. 2A	Prot. 2B	Prot. 2C	Prot. 3A	Ref.	Prot. 3C	Prot. 4A	Prot. 4B	Prot. 4C
Scoring of cell lines	Manual	✗	✗	✗	✗	✓	✓	✓	✓	✓	✗	✗	✗
	AI algoritm	✗	✗	✗	✗	✗	✓	✗	✓	✗	✗	✗	✗
Clinical impact on BCs	HER2 classic	FN	FN	FN	FN	FN	FN	FN	NO	NO	NO	NO	NO
	HER2-low	FN	FN	FN	FN	FN	FN	FN	NO	FP	FP	FP	FP

✓ - pass; ✗ - fail; FN – false negative; FP – false positive; NO – no clinical impact

Figure 1. Examples of passed and failed cell lines and corresponding HER2 IHC scores in BCs



Graph 1. Proportion of HER2 scores across protocols

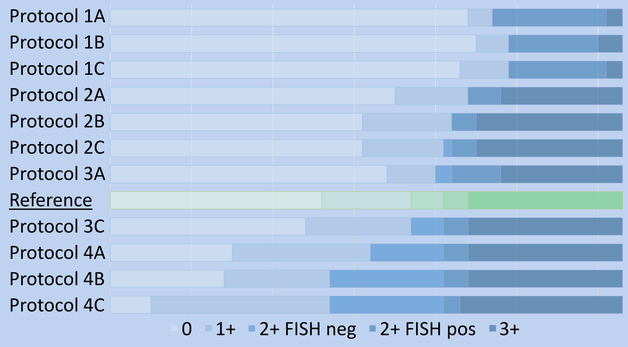


Table 2. Sensitivity of cell lines to predict HER2 assay accuracy by manual and AI supported read-out

Number of BC samples with clinically critical change of HER2 IHC score (FN/FP)				
	HER2 classical		HER2 low	
	Manual	AI	Manual	AI
Cell lines passed ✓	3	1	20	5
Cell lines failed ✗	17	19	106	121
Sensitivity	**85%**	**95%**	**84%**	**96%**

Discussion

Based on this initial feasibility study, the combination of cell lines together with AI was found to be a potential tool to evaluate the accuracy of HER2 IHC assays. More studies with enriched numbers of BCs at the critical thresholds for both HER2 overexpression and HER2-low must be performed. In addition to the HER2 IHC assays with “forced errors” included in this study, the most commonly applied real-life IHC protocol settings must be incorporated. Also, the ability, precision and robustness of this quality tool to help secure HER2 IHC reproducibility in each IHC assay conducted in a diagnostic setting must be further evaluated and validated.

Download the poster

USCAP 113th Annual Meeting, 2024, Baltimore

NordiQC.org / Visiopharm.com