HER2-13

Proficiency Assessment of HER2-Low Breast Cancer Scoring With the Ventana PATHWAY 4B5 and Dako HercepTest HER2 Assays and the Impact of Pathologist Training

Josef Rüschoff,[1] Alexander Penner,[1] Ian O. Ellis,[2] Elizabeth Hale Hammond,[3] Annette Lebeau,[4] Robert Y. Osamura,[5] Fréderique Penault-Llorca,[6] Federico Rojo,[7] Neil Atkey,[8] Andreas H. Scheel,[9] Corrado D´Arrigo,[10] Hans-Ulrich Schildhaus,[1] Akira Moh,[11] Chirag Desai,[11] Giuseppe Viale[12]

[1]Discovery Life Sciences Biomarker Services, Kassel, Germany; [2]University of Nottingham, Nottingham, United Kingdom; [3]Intermountain Healthcare and University of Utah School of Medicine, Salt Lake City, UT, USA; [4]University Medical Center Hamburg-Eppendorf, Hamburg, Germany; [5]Nippon Koukan Hospital, Kawasaki and Keio University School of Medicine, Tokyo, Japan; [6]Centre Jean Perrin, University Clermont Auvergne, Clermont-Ferrand, France; [7]Hospital Universitario Fundacion Jimenez Diaz, Madrid, Spain; [8]Diaceutics PLC, Belfast, United Kingdom; [9]Institute of Pathology, University of Cologne, Cologne, Germany; [10]Poundbury Cancer Institute for Personalised Medicine, Dorchester, United Kingdom; [11]Daiichi Sankyo, Basking Ridge, NJ, USA; [12]European Institute of Oncology IRCCS, and University of Milan, Milan, Italy

Background

- Prior to DESTINY-Breast04, breast cancers expressing low levels of human epidermal growth factor receptor 2 (HER2-low), defined as either HER2 immunohistochemistry (IHC) 2+ without gene amplification as determined by in situ hybridization (ISH–) or IHC 1+, were categorized as HER2-negative when identifying treatment approaches for patients with breast cancer[1-4]
- Although there is an emerging need to distinguish lower ranges of HER2 IHC expression based on the results of DESTINY-Breast04,[1] a recent College of American Pathologists (CAP) survey data set showed a low concordance among pathologists in distinguishing HER2 IHC 0 from HER2 IHC 1+[5]
- DESTINY-Breast04 demonstrated that trastuzumab deruxtecan (T-DXd), a HER2-targeting antibody-drug conjugate, significantly prolonged progression-free survival and overall survival of patients with HER2-low (IHC 2+/ISH– or IHC 1+) metastatic BC (mBC) compared with those patients treated with physician's choice of chemotherapy[1]
- As a result of the efficacy demonstrated in DESTINY-Breast04, the US Food and Drug Administration (FDA) approved T-DXd for the treatment of HER2-low mBC and the Ventana PATHWAY 4B5 companion diagnostic (4B5) to determine patients having HER2-low mBC[1,6,7]
- Here, we report on current real-world proficiency for interpreting HER2-low disease and the impact of training for participating pathologists in HER2-low scoring
- The objective of this study was to provide comprehensive concordance data of HER2-low scoring in breast cancer, including:
  - The assessment of baseline proficiency of pathologists of HER2-low scoring in breast cancer for the 4B5 and Dako HercepTest (HcT) assays, separately
  - The assessment of the effect of training on pathologist proficiency and to provide comprehensive concordance data of HER2-low scoring in breast cancer for the 4B5 and HcT assays, separately, after specific reader training

Conclusions

- Results from this real-world global study demonstrate that overall score concordance with a new category of HER2-low was above the 80% overall rater agreement (ORA) benchmark for both 4B5 and HcT and is higher than previously reported[5]
- With an improvement in positive percent agreement (PPA) for HER2 IHC 0 scoring and an improvement in negative percent agreement (NPA) in HER2-low scoring after training, training significantly improved the ability of pathologists to identify HER2 IHC 0 and HER2-low cases
- These data demonstrate pathologists' ability to achieve an acceptable level of accuracy for identifying HER2-low patients even after short-term training; however, additional training techniques and experience are needed to further improve accuracy

Poster presented at the San Antonio Breast Cancer Symposium (SABCS) Annual Meeting; December 6-10, 2022; San Antonio, TX, USA

This study was sponsored by Daiichi Sankyo, in collaboration with AstraZeneca. In 2019, AstraZeneca entered into a global development and commercialization collaboration agreement with Daiichi Sankyo for T-DXd (DS-8201).

Methods

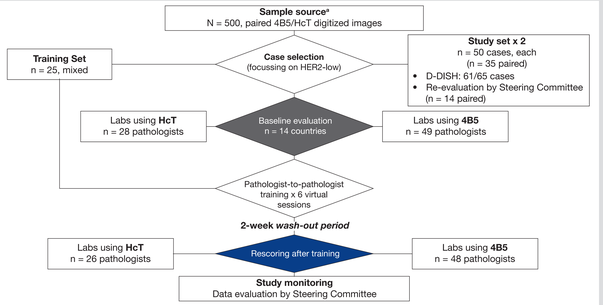
Case Selection and Study Design

- Pathologists (N = 80 [planned]) from laboratories across the United States, European Union, Japan, Australia, and Brazil used a digital pathology platform (Pathotrainer) to interpret HER2 digital images using ASCO/CAP 2018 scoring criteria (**Figure 1** and **Figure 2**) with some practical considerations such as increased time and high power magnification
- The study of scoring proficiency consisted of 3 steps (**Figure 1**):
  1. Initial baseline evaluation
  2. Half-day (4 hour), virtual pathologist-to-pathologist training
  3. Rescoring after a 2-week washout period
- 2 whole-slide imaging samples of 50 representative study cases each (35 paired samples), were compiled for 4B5 or HcT stained tumor and another sample set (n = 25) was developed for the virtual training session
- A steering committee (SC) of 8 pathology experts was formed to guide the study
  - Cases considered challenging due to difficult-to-interpret staining patterns were reevaluated by the SC members
- The primary endpoint was real-world concordance and ORA
  - The secondary endpoints were post-training concordance and ORA and correct identification of HER2 zero (IHC 0) and HER2-low (IHC 2+/ISH– or IHC 1+)

Statistical Analysis

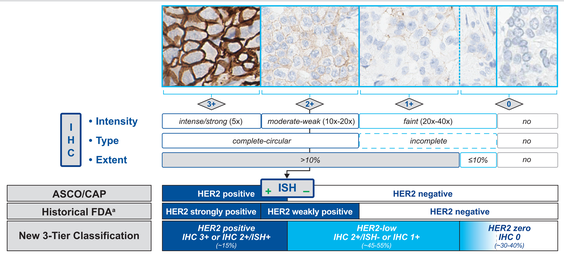
- Concordance and efficacy of training were measured by Cohen's weighted kappa (κ) coefficient and ORA and receiver operating characteristic (ROC) curve statistics
  - Concordance as assessed by Cohen's κ (range 0.–1.0) is regarded as perfect/optimal at >0.80 and substantial at >0.6-0.8[8]
  - Acceptable ORA percentage for test evaluation in general is ≥80%[9], with the ideal ORA percentage for HER2 tests ≥95%[10]
  - ROC results are considered excellent for area under the curve (AUC) values between 0.9-1.0, good for 0.8-0.9, and fair between 0.7-0.8. Lower values indicate a poor (0.6-0.7) or failed test (0.5-0.6)
  - Significance level was set to $P < 0.05$ and calculated by chi-square test and samples considered not evaluable (NOE) by a participant were excluded from calculations

Figure 1. Sample Selection and Study Design



[a]Astra-Zeneca R&D, Cambridge, United Kingdom.

Figure 2. Classification Scheme of HER2 Staining According to FDA[11,12] and ASCO/CAP 2018[2] in Comparison to New 3-Tier Classification (HER2-Positive, HER2-Low and HER2 IHC 0)



[a]The interpretation of score 2+ has changed from "weak positive" to "equivocal" after ISH had been introduced. The "historical FDA classification" is "positive" (IHC3+/2+) versus "negative" (IHC 0/1+). IHC stained tissue sections (upper row) are evaluated stepwise first by determining intensity of membrane staining using magnification rule, followed by the assessment of circularity and finally the percentage of stained tumor cells. Classification of diagnostic groups is based on these 3 criteria in combination with HER2 ISH data in IHC 2+ cases.

Results

Characteristics of Study Cases and Participating Pathologists

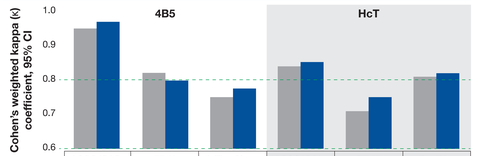
- A total of 91 (N = 46 4B5 and N = 45 HcT) stained study cases fulfilled the eligibility criteria for proficiency analysis; several cases were excluded by the SC because they were not evaluable or the consensus on score could not be reached
  - Both study sets were comparable with respect to type of specimen (resections and biopsies), histological tumor type (lobular and not otherwise specified), and tumor grade
  - Distribution between the 3 categories of HER2 positive, -low, and IHC 0 were 13%, 52%, and 35% for the 4B5 assay and 27%, 33%, and 40% for the HcT assay, respectively
    - » Differences in distribution of HER2 positive, -low, and IHC 0 between 4B5 and HcT occurred because ISH data could be obtained only after cases had been assigned to the 4B5 or HcT immunostains
- Pretraining baseline or real-world scores were obtained for 77 pathologists in 14 countries (n = 49 for 4B5, n = 28 for HcT) and 74 pathologists completed post-training scores (n = 48 for 4B5, n = 26 for HcT)

Inter-rater Concordance for HER2 Scoring

- HER2 scoring proficiency of pathologists was high for both assays when assessed on ASCO/CAP binary HER2 negative and positive status, irrespective of training (4B5: κ = 0.96 vs 0.97, ORA = 98.9% vs 99.4% at baseline and after training, respectively; HcT: κ = 0.84 vs 0.85, ORA = 94.3% vs 94.7% at baseline and after training, respectively) (**Figure 3** and **Figure 4**)
  - Inter-rater concordance for the new 3-tier classification (HER2 IHC 0 vs HER2-low [IHC 2+/ISH– or IHC 1+] vs HER2 positive [IHC 3+ or IHC 2+/ISH+]) was within the acceptable range for both assays at baseline (4B5: κ = 0.75, ORA = 82.8%; HcT: κ = 0.81, ORA = 84.1%) and after training (4B5: κ = 0.79, ORA = 84.9%; HcT: κ = 0.82, ORA 85.3%) (**Figure 3** and **Figure 4**)

Results (continued)

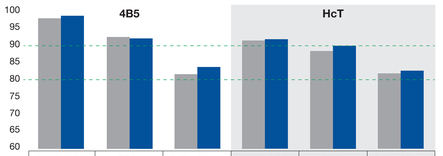
Figure 3. Kappa Analysis for the Concordance Between Reference Group and Participating Pathologists for HER2 Binary (ASCO/CAP and FDA) and New 3-Tier Classification (including HER2-low)



	4B5			HcT		
	ASCO/CAP[a]	FDA[b]	3-Tier Class[c]	ASCO/CAP[a]	FDA[b]	3-Tier Class[c]
Baseline	0.96 (0.90-1.0)	0.82 (0.54-0.95)	0.75 (0.61-0.92)	0.84 (0.51-0.97)	0.72 (0.42-0.90)	0.81 (0.70-0.95)
After training	0.97 (0.94-1.0) $P = 0.26$	0.81 (0.52-0.95) $P = 0.45$	0.79 (0.64-0.93) $P = 0.06$	0.85 (0.57-0.97) $P = 0.67$	0.75 (0.45-0.92) $P = 0.197$	0.82 (0.74-0.95) $P = 0.47$

Concordance as assessed by Cohen's κ (range 0.-1.0) is regarded as perfect/optimal at >0.80 and substantial at >0.6-0.8
[a]Current ASCO/CAP 2018 guideline classification; HER2 positive (IHC 3+ or IHC 2+/ISH+) vs HER2 negative (IHC 2+/ISH–, IHC 1+, IHC 0).
[b]Historical FDA classification; HER2-positive (IHC 3+ or IHC 2+ vs HER2 negative (IHC 1+ or IHC 0).
[c]New 3-tier classification with a category for HER2-low; HER2 positive (IHC 3+ or IHC 2+/ISH+) vs HER2-low (IHC 2+/ISH– or IHC 1+) vs HER2 IHC 0.

Figure 4. Overall Rater Agreement Between Reference Group and Participating Pathologists for HER2 Binary (ASCO/CAP and FDA) and New 3-Tier Classification (including HER2-low)



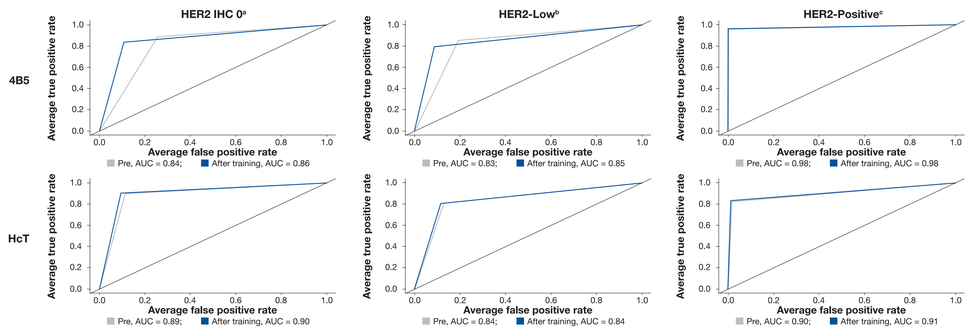
	4B5			HcT		
	ASCO/CAP[a]	FDA[b]	3-Tier Class[c]	ASCO/CAP[a]	FDA[b]	3-Tier Class[c]
Baseline	98.9%	92.4%	82.8%	94.3%	88.3%	84.1%
After training	99.4% ($P = 0.18$)	92.2% ($P = 0.79$)	84.9% ($P = 0.15$)	94.7% ($P = 0.69$)	89.8% ($P = 0.19$)	85.3% ($P = 0.54$)

[a]Current ASCO/CAP 2018 guideline classification; HER2 positive (IHC 3+ or IHC 2+/ISH+) vs HER2 negative (IHC 2+/ISH–, IHC 1+, IHC 0).
[b]Historical FDA classification; HER2 positive (IHC 3+ or IHC 2+ vs HER2 negative (IHC 1+ or IHC 0).
[c]New 3-tier classification with a category for HER2-low; HER2 positive (IHC 3+ or IHC 2+/ISH+) vs HER2-low (IHC 2+/ISH– or IHC 1+) vs HER2 IHC 0.

Subgroup Analysis for HER2 IHC 0 and HER2-Low

- AUC of ROC showed excellent performance values for HER2 positive diagnostics with both tests (≥0.9) (**Figure 5**)
  - Identification of HER2 IHC 0 and HER2-low was good (AUC 0.8-0.9) with both tests and could be improved by short-term training, especially in 4B5 users
  - For the 4B5 assay, ROC improved for HER2 IHC 0 and HER2-low, particularly for 4B5 assay readers
- According to agreement analyses of both assays (**Table 1**), there was general improvement for HER2 IHC 0 and HER2-low scoring after training.
  - 74.6% of HER2 IHC 0 tumors were correctly diagnosed before training and 89.2% after training ($P < 0.0001$), an improvement in PPA, indicating a reduction of false positive scoring
  - HER2-low scoring demonstrated a significant increase in NPA after training for the 4B5 test from 80.6% to 91.1% ($P < 0.0001$)

Figure 5. Receiver Operating Characteristics for 4B5 and HcT Users



[a]IHC 0, UL+) = 0; (IHC 1+, IHC 2+ [ISH+/-], 3+) = 1.
[b]IHC 0, UL+; IHC 2+ [ISH+], IHC 3+) = 0; (IHC 1+, IHC 2+ [ISH–]) = 1.
[c]IHC 0, UL+; IHC 1+, IHC 2+ [ISH–]) = 0; (IHC 2+ [ISH+], 3+) = 1

Table 1. Agreement Analyses for HER2 Zero and HER2-Low Scoring

	4B5						HcT					
	PPA		NPA		ORA		PPA		NPA		ORA	
HER2 Score	Baseline	After training	Baseline	After training	Baseline	After training	Baseline	After training	Baseline	After training	Baseline	After training
Zero	74.6%	89.2%	88.8%	83.5%	83.9%	85.5%	88.8%	90.9%	89.8%	89.9%	89.4%	90.3%
P-value[a]	< 0.00001		≤ 0.00005		NA		Not significant at < 0.05		Not significant at < 0.05		NA	
Low	85.0%	79.3%	80.6%	91.1%	82.9%	85.0%	80.5%	80.4%	87.0%	88.4%	84.8%	85.7%
P-value[a]	< 0.0004		< 0.00001		NA		Not significant at < 0.05		Not significant at < 0.05		NA	
Positive	96.1%	95.8%	99.4%	99.9%	98.9%	99.4%	82.1%	83.0%	98.5%	98.7%	94.4%	94.7%
P-value[a]	NA		NA		NA		NA		NA		NA	

[a]According to chi-square statistics and chi-square statistics with Yates correction, significance at $P < 0.05$

Abbreviations

4B5, Ventana PATHWAY 4B5 Assay; ASCO/CAP, American Society of Clinical Oncology/College of American Pathologists; AUC, area under the curve; D-DISH, dual-color dual-hapten in situ hybridization; FDA, US Food and Drug Administration; HcT, Dako HercepTest; HER2, human epidermal growth factor receptor 2; IHC, immunohistochemistry; ISH, in situ hybridization; mBC, metastatic breast cancer; NA, not analyzed; NPA, negative percent agreement; PPA, positive percent agreement; ORA, overall rater agreement; ROC, receiver operating characteristic; T-DXd, trastuzumab deruxtecan; UL, ultra low.

References

1. Modi S et al. *N Engl J Med*. 2022;387:9-20.
2. Wolff AC et al. *J Clin Oncol*. 2018;36:2105-2122.
3. Tarantino P et al. *J Clin Oncol*. 2020;38:1951-1962.
4. Eiger D et al. *Cancers (Basel)*. 2021;13:1015.
5. Fernandez et al. *JAMA Oncol*. 2022;2022;8:1-4.
6. Enhertu. Package insert. Daiichi Sankyo, Inc.; 2022.
7. Food and Drug Administration. *List of Cleared or Approved Companion Diagnostic Devices*. Accessed October 11, 2022. https://www.fda.gov/medical-devices/in-vitro-diagnostics/list-cleared-or-approved-companion-diagnostic-devices-in-vitro-and-imaging-tools
8. Landis JR, Koch GG. *Biometrics*. 1977;33:159-174.
9. College of American Pathologists. Proficiency Testing Manual 2019. Available at https://documents.cap.org/documents/2019-ProficiencyTesting-Manual_FINAL.pdf. Accessed 4 November 2022.
10. College of American Pathologists, Anatomic Pathology Checklist 2021. Available at https://pathology.jhu.edu/campus/CAP-Checklist/Anatomic-Pathology-2021.pdf. Accessed 4 November 2022.
11. HercepTest Code K5204 [package insert] Dako Cytomation, Colostrup, Denmark. 2020.
12. Jorgensen JT et al. *Front Oncol*. 2021;11:676939.

Acknowledgments

Judy Yu, a former AstraZeneca employee, provided expertise and technical insights to support the study. Amy Hanlon Newell, Andrew Livingston, Anne-Marie Boothman, Maria Haughton, and Victoria de Giorgio-Miller provided insightful comments. Sharon Allen oversaw the project management. Ouzna Morsli provided general support. This study was sponsored by Daiichi Sankyo, in collaboration with AstraZeneca. In 2019, AstraZeneca entered into a global development and commercialization collaboration agreement with Daiichi Sankyo for T-DXd (DS-8201). Under the guidance of authors, assistance in medical writing and editorial support was provided by ApotheCom, and was funded by Daiichi Sankyo.

Disclosures

Dr. Josef Rüschoff is an employee of Targos – a Discovery Life Science company. He has advisory board memberships with Daiichi Sankyo, AstraZeneca, BMS, GSK, MSD, Merck and Roche, with payments made to employer.