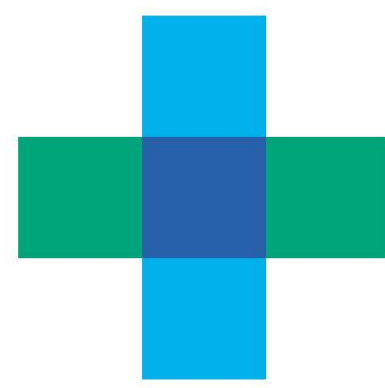


#516: AI-based HER2-low IHC scoring in breast cancer
across multiple sites, clones, and scanners.Patrick Frey¹, A. Mamilos², E. Minin³, R. Banisch¹, S. Günther¹, C. Schmidt⁴, N. Abele⁴, A. Hartmann⁴, P.A. Fasching⁵, M. Ruebner⁵, H. Huebner⁵, T. Lang¹, Ramona Erber⁴

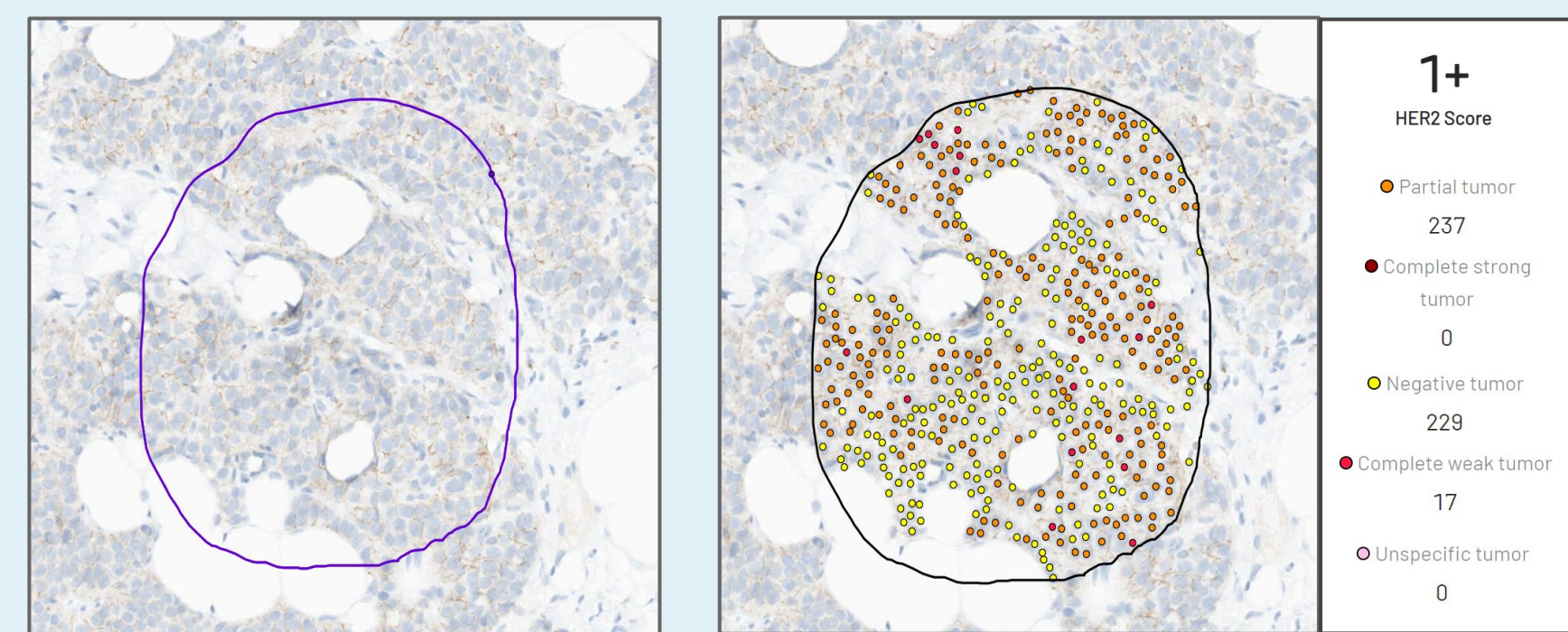
1 Mindpeak GmbH, Hamburg, Germany; 2 Institute of Pathology, University of Regensburg, Regensburg, Germany; 3 Institute of Pathology, Clinical Center Osnabrueck, Osnabrueck, Germany; 4 Institute of Pathology, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Comprehensive Cancer Center Erlangen-EMN (CCC ER-EMN), Erlangen, Germany; 5 Department of Gynaecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Comprehensive Cancer Center Erlangen-EMN (CCC ER-EMN), Erlangen

Background: Assessment of immunohistochemical (IHC) HER2 expression plays a pivotal role in breast cancer diagnostics. In the era of HER2-low and HER2-targeted antibody drug conjugates, accurate discrimination of defined HER2 IHC scores is essential. At the same time, HER2 IHC scoring suffers from poor interobserver concordance. Artificial intelligence (AI) may optimize this scoring in regard to standardization, accuracy and efficiency, but previous approaches fail to show the required consistency across samples from different sites, clones and scanning hardware.

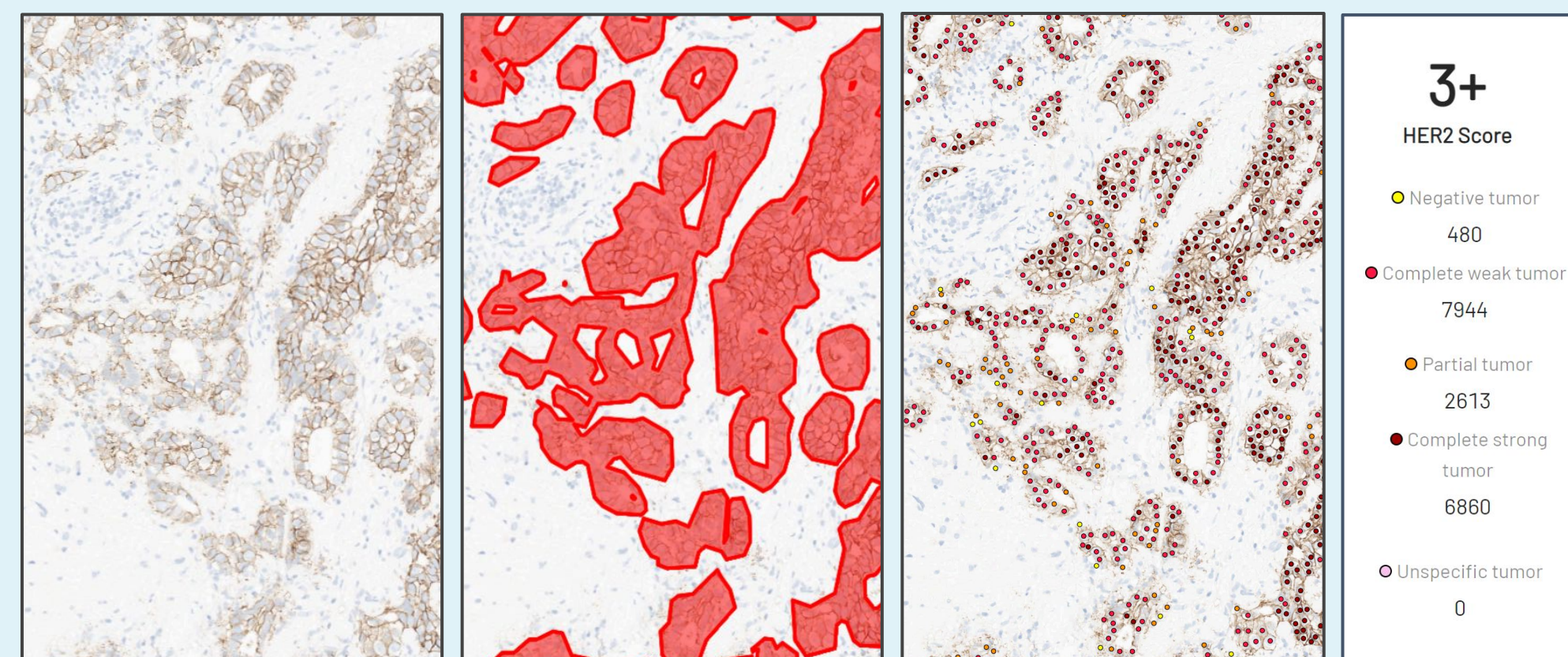
Methods: We investigated the use of AI-based HER2 IHC quantifier [*Mindpeak Breast HER2 RoI* ^{CE} and *Breast HER2 (WSI)*] software to support pathologists in standardized HER2 IHC assessment in breast cancer.

Mindpeak Breast HER2 RoI ^{CE}

1. Pathologist draws ROI
2. AI identifies cells and scores HER2



Mindpeak Breast HER2 (WSI) ^{RuO}: Fully automatic analysis of entire WSI: tumor detection, cell identification, HER2-scoring

**Performance evaluation study:**

To assess performance of the AI-assistance tool, we conducted a validation study measuring interrater-agreement. As a reference: **HER2 interobserver variability in humans is 68.2±4.7% according to literature** ^{[1],[2],[3]}.

Cohort 1 (ROI & WSI scoring):

N=150 WSIs from 3 institutions

4 different scanners (3D Histech, Leica, Hamamatsu, Philips)

2 Antibodies (Ventana, CellMarque)

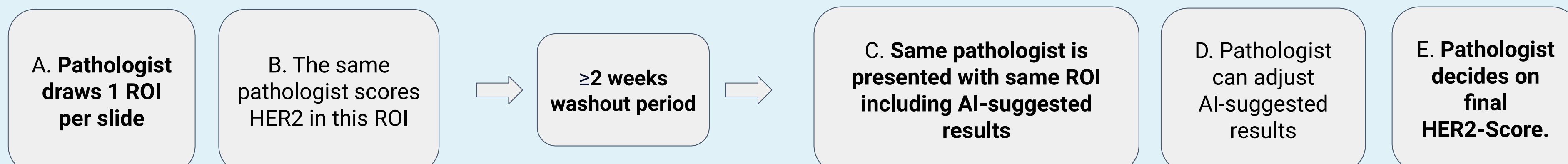
Cohort 2 (WSI scoring):

N=94 WSIs

single institute, scanner and clone

Study setup:

1. ROI scoring (performed by 3 pathologists with multi-year experience in HER2 scoring)

**2. WSI scoring**

AI scores HER2 on entire WSIs fully automatically (no human intervention). Comparison to scores from clinical workflow.

ROI scoring**AI-assisted vs. Human**

Agreement: 91.3%

(Specificity: 94.4%, Sensitivity: 90.9%)

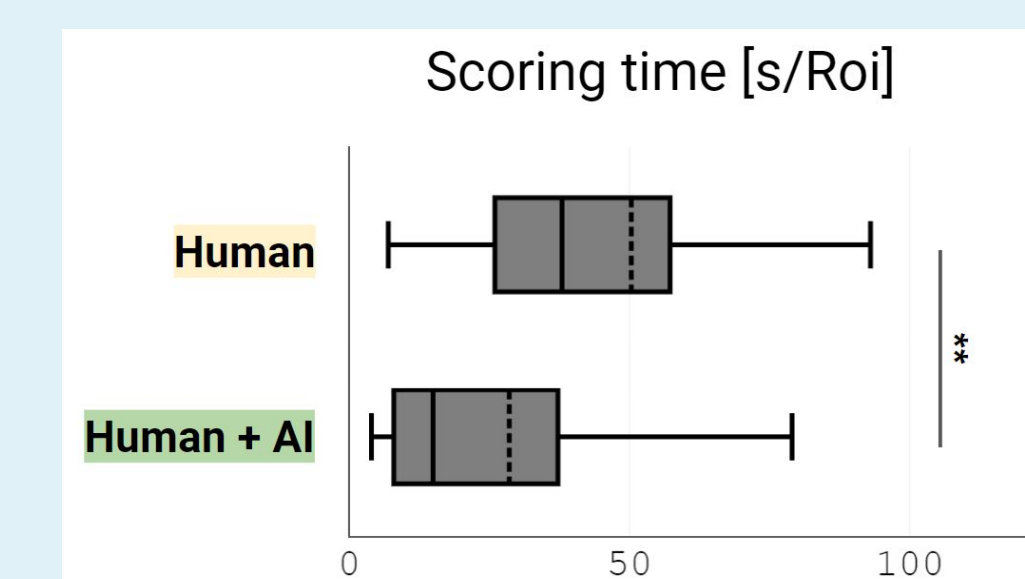
	HER2-Score	Human	
		0	1+/2+/3+
Human + AI	0	17	12
	1+/2+/3+	1	120

AI-only vs. Human

Agreement: 86.7%

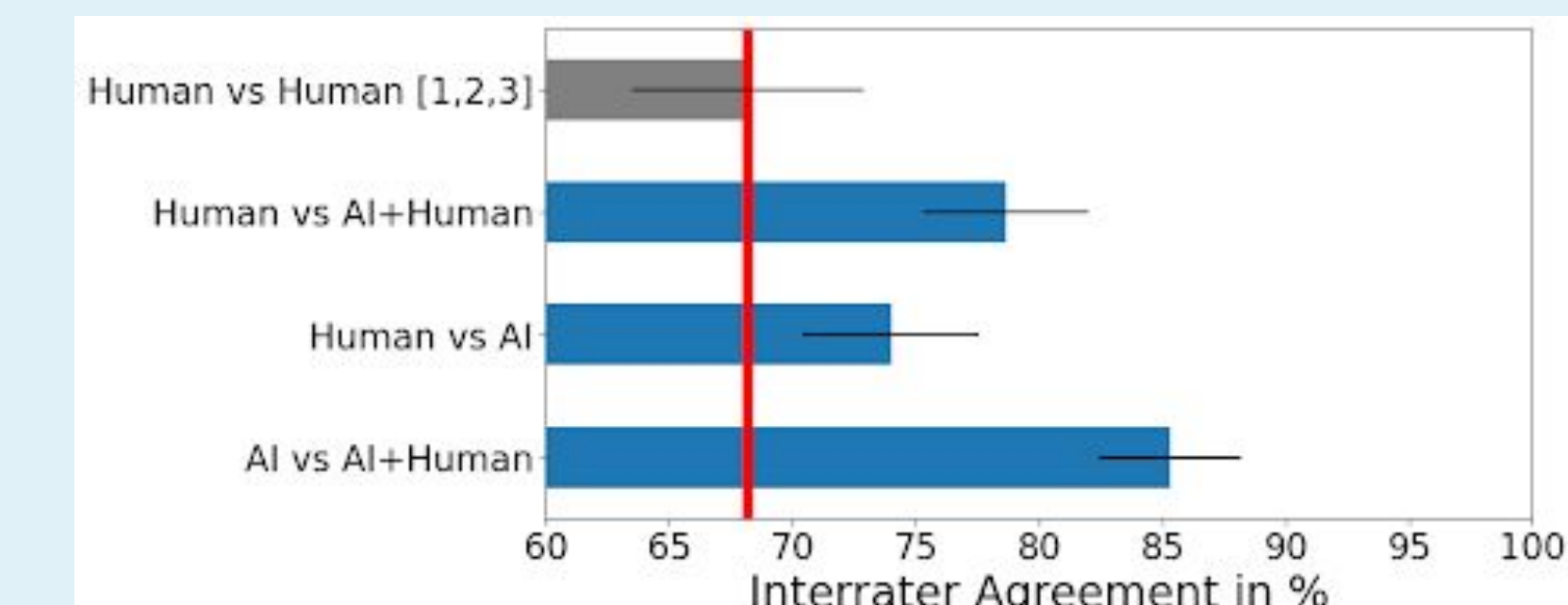
(Specificity: 100%, Sensitivity: 84.8%)

	HER2-Score	Human	
		0	1+/2+/3+
AI-only	0	18	20
	1+/2+/3+	0	112

AI-Assistance reduces scoring time.

(Mean: 50s vs. 29 s - Median: 38s vs. 15s)

Agreement rates across all classes (0 vs. 1+ vs. 2+ vs. 3+) with use of the AI tool are higher than literature-reported agreement between humans ^{[1],[2],[3]}.

**Fully automated WSI scoring****HER2neg. vs. other****Cohort 1**

Agreement: 89.1%

	HER2-Score	Clinical report score	
		0	1+/2+/3+
AI result	0	20	8
	1+/2+/3+	7	103

Cohort 2

Agreement: 86.2%

	HER2-Score	Clinical report score	
		0	1+/2+/3+
AI result	0	13	5
	1+/2+/3+	8	68

Clinical HER2-score categories

HER2 negative (0) vs. low (1+/2+^{ISH-}) vs. positive (2+^{ISH+}/3+)

Cohort 1: Agreement: 84.1%

Cohort 2: Agreement: 84.0%

Ability to correctly predict ISH results**Cohort 1**

	Clinical Score 2+	ISH-	AI result			
			0	1+	2+	3+
Clinical Score 2+	ISH-	0	10	25	2	
		ISH+	0	2	0	3

31% (13/42) of ISH tests could have been avoided with AI.
(AI result 1+ has 83% Specificity)

Cohort 2

	Clinical Score 2+	ISH-	AI result			
			0	1+	2+	3+
Clinical Score 2+	ISH-	0	23	12	0	
		ISH+	0	1	1	0

62% (23/37) of ISH tests could have been avoided with AI.
(AI result 1+ has 96% Specificity)

Conclusions: Across challenging validation data from four institutions and four scanners, scoring with the support of an AI HER2 IHC quantifier software showed very high agreement when discriminating HER2-neg from HER2-low/pos cases and high accuracy for general HER2 scoring. When using AI-assistance, scoring time was reduced by around 50%. Altogether, these results demonstrate the potential of AI solutions to **increase consistency and efficiency of HER2 scoring** and ultimately to improve patient outcome.

References:

- [1]: Layfield LJ, Frazier S, Esebua M, Schmidt RL. Interobserver reproducibility for HER2/neu immunohistochemistry: A comparison of reproducibility for the HercepTest™ and the 4B5 antibody clone. *Pathol Res Pract*. 2016 Mar;212(3):190-5. doi: 10.1016/j.prp.2015.11.016. Epub 2015 Dec 31. PMID: 26839158.
- [2]: Casterá Redal, Carlos, and Laia Bernet Vegué. "HER2 immunohistochemistry inter-observer reproducibility in 205 cases of invasive breast carcinoma additionally tested by ISH". *Annals of diagnostic pathology* vol. 48 (2020): 151566. doi:10.1016/j.anndiagpath.2020.151566
- [3]: Mamoon, Nadira et al. "Interobserver variability in HER-2/neu reporting on immunohistochemistry in breast carcinomas." *JPMa. The Journal of the Pakistan Medical Association* vol. 64,2 (2014): 151-4.

Correspondence: Dr. Patrick Frey; patrick.frey@mindpeak.ai, +49 40 3567 6797