

An AI-driven computational biomarker model from H&E slides recovers cases with low levels of HER2 from immunohistochemically HER2-negative breast cancers

Marra A¹, Goldfinger MH², Millar EKA³, Hanna M², Rothrock B, Lee MCH¹, Wang YK¹, van Eck A¹, Trlifaj Tydlitova L. ¹, Sue J ¹, Wen YH², Stefanelli J ¹, Retamero J¹,Hamilton P¹, Fuchs TJ¹, Klimstra D¹, Chandarlapaty S⁴, Reis-Filho JS²

1. Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY 10021, USA 2. Paige AI, 11 Times Sq, New York, NY 10036, USA; 3. Department of Anatomical Pathology, NSW Health Pathology, St George Hospital, Sydney, Australia; 4. Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10021, USA

Background:

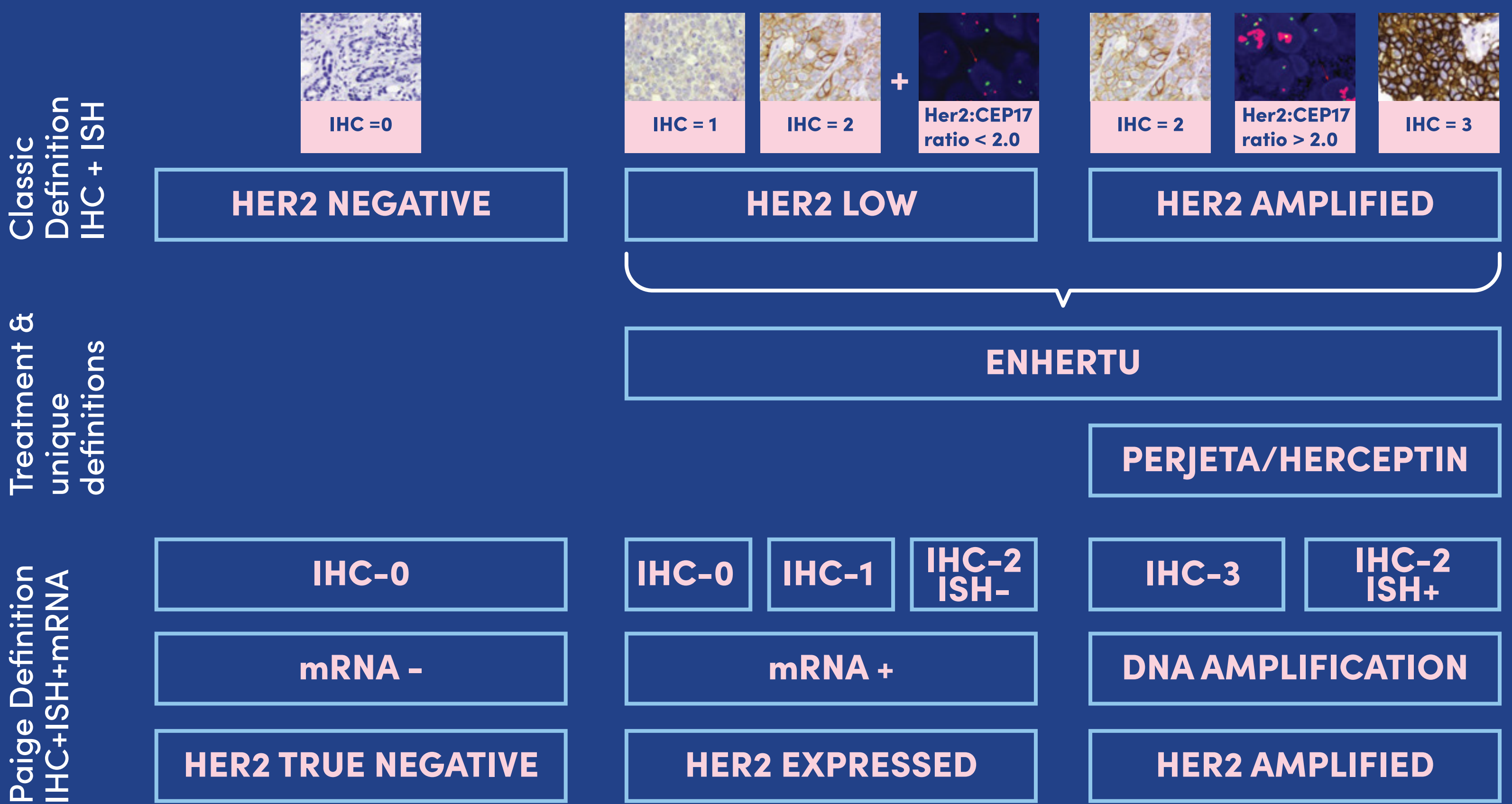
Novel anti-HER2 antibody drug conjugates (ADCs) have shown efficacy in breast cancers (BCs) expressing low levels of HER2 (i.e., IHC 1+/2+). A subset of BCs classified as HER2 0 by current IHC methods may express HER2 protein. We sought to define whether BCs expressing HER2 could be accurately detected by deep learning (DL) methods applied to H&E whole slide images (WSIs), using a combination of IHC and HER2 mRNA expression as ‘gold standard’.

Methods:

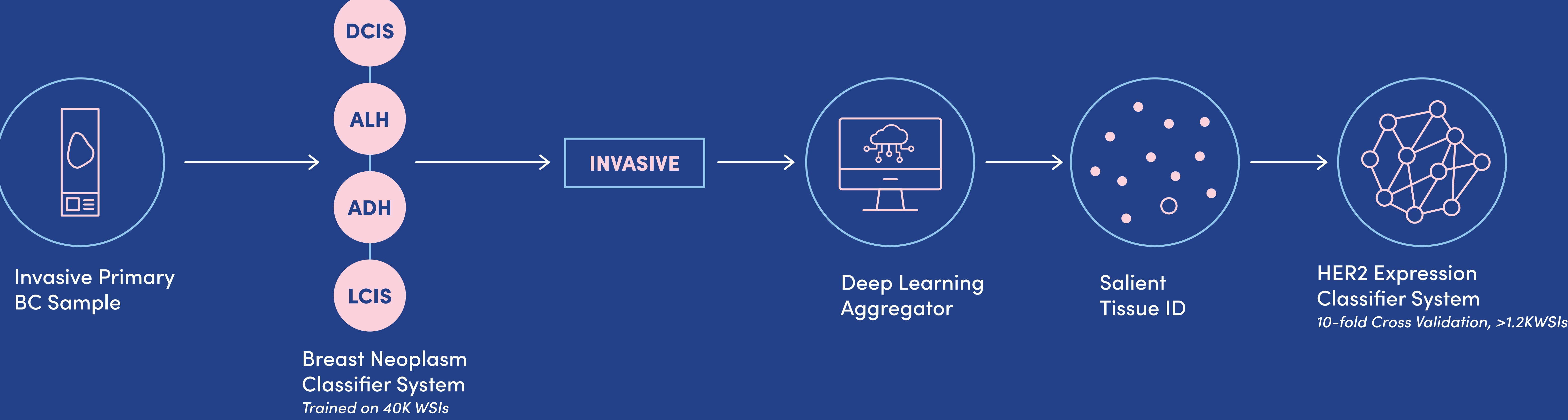
1479 H&E-stained WSIs from 417 primary BCs were categorized according to HER2 IHC, FISH and HER2 copy number amplification. All HER2 0 and HER2-low (i.e., 1+ and 2+) samples were also tested for HER2 mRNA expression. For the purpose of defining true negatives we selected cases that were both negative for HER2 expression via IHC (IHC-0) and contained no ERBB2 mRNA (less than 7.6 standardized expression score). For cases selected as HER2-low, IHC-1+ and IHC-2+/ISH- were selected for training. HER2 amplified were defined as having IHC-2+/ISH+, IHC-3+ and DNA amplification as identified via MSK-IMPACT. A SE-ResNet-50 CNN and aggregator were trained from WSIs of H&E sections at 20x. Slide-level predictions were evaluated with 8-fold cross-validation.

Establishing a New Standard to Assess HER2 Expression

IHC, FISH & Transcriptomics: A New Ground Truth



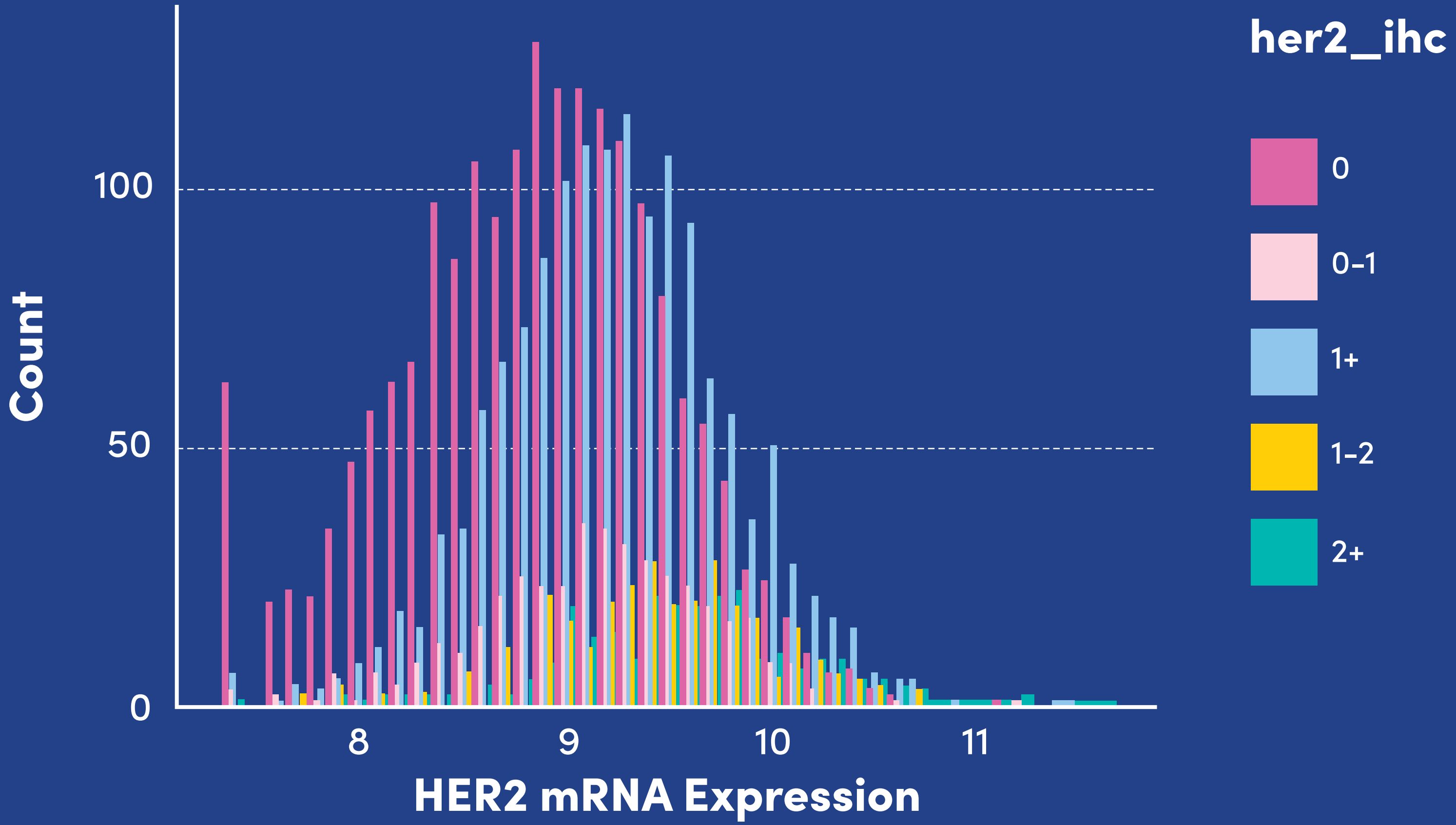
HER2 Model Development



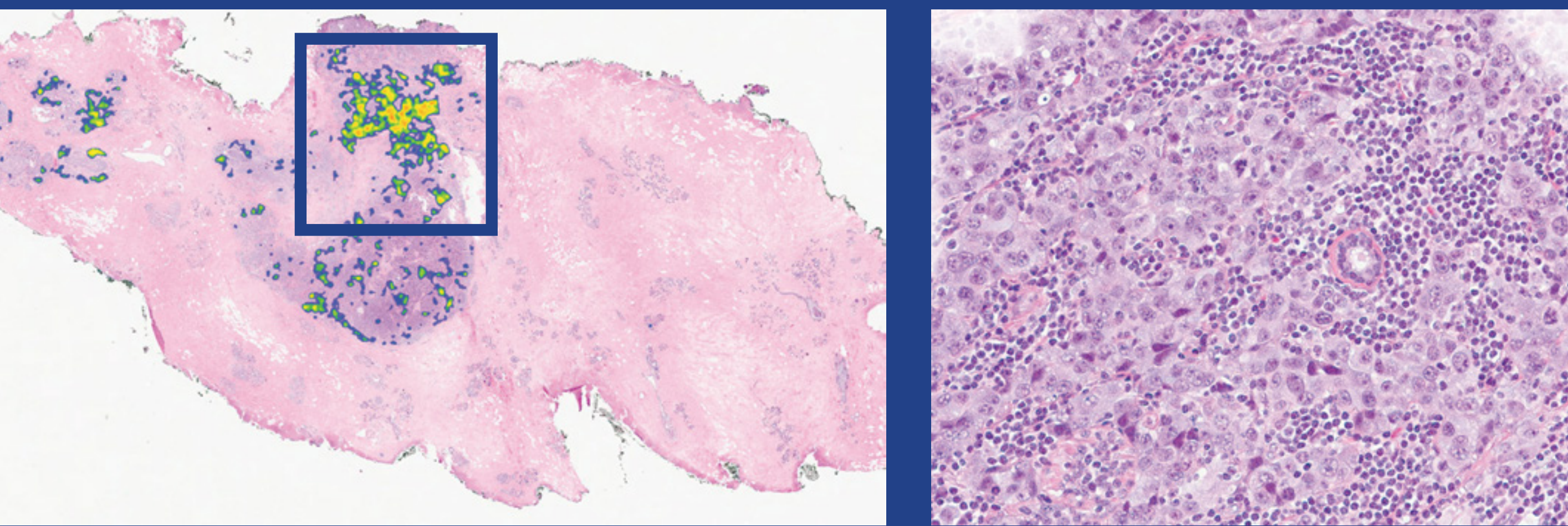
Results:

A total of 1098, 820, 122, and 148 WSIs were categorized as HER2 0, 1+, 2+ and 3+ by IHC staining, respectively. mRNA expression data revealed a range of <7.6 to 11.6 (SD=0.63) for HER2 mRNA expression level. When stratified by HER2 mRNA expression, a cut-off of <7.6 mRNA was selected to represent HER2-null, which included IHC 0 (n=32), IHC 1+ (n=3) and IHC 2+ (n=1). HER2 IHC 0 to 2+ with HER2 mRNA expression >9 and HER2 FISH not amplified were considered as being HER2-low (IHC 0, n=494; IHC 1+, n=562; IHC 2+, n=103). Cases with HER2 IHC 3+ and/or FISH amplification were considered HER2-positive. Model development was based on 417 cases (1479 WSIs) including 32 HER2-null, 292 HER2-low and 93 HER2-positive cases. When distinguishing HER2-low and amplified cases from HER2-null, the model's performance had an AUC 0.78, a sensitivity of 73%, a specificity of 78%, PPV of 21%, NPV of 97%, and F1=0.85. The model identified 21/25 (84%) HER2 IHC 0 and mRNA<7.6, classified as HER2-null, and 136/167 (81%) HER2 IHC 1+/2+ and mRNA>9, classified as HER2-low.

HER2 Expression Distribution Per HER2 IHC Category



The algorithm detects phenotypic patterns within a WSI containing a breast excision, where the sample is found to be completely negative for HER2 expression (heatmap). A higher powered image of the area identified as HER2 true negative is strongly associated with invasive cancer.

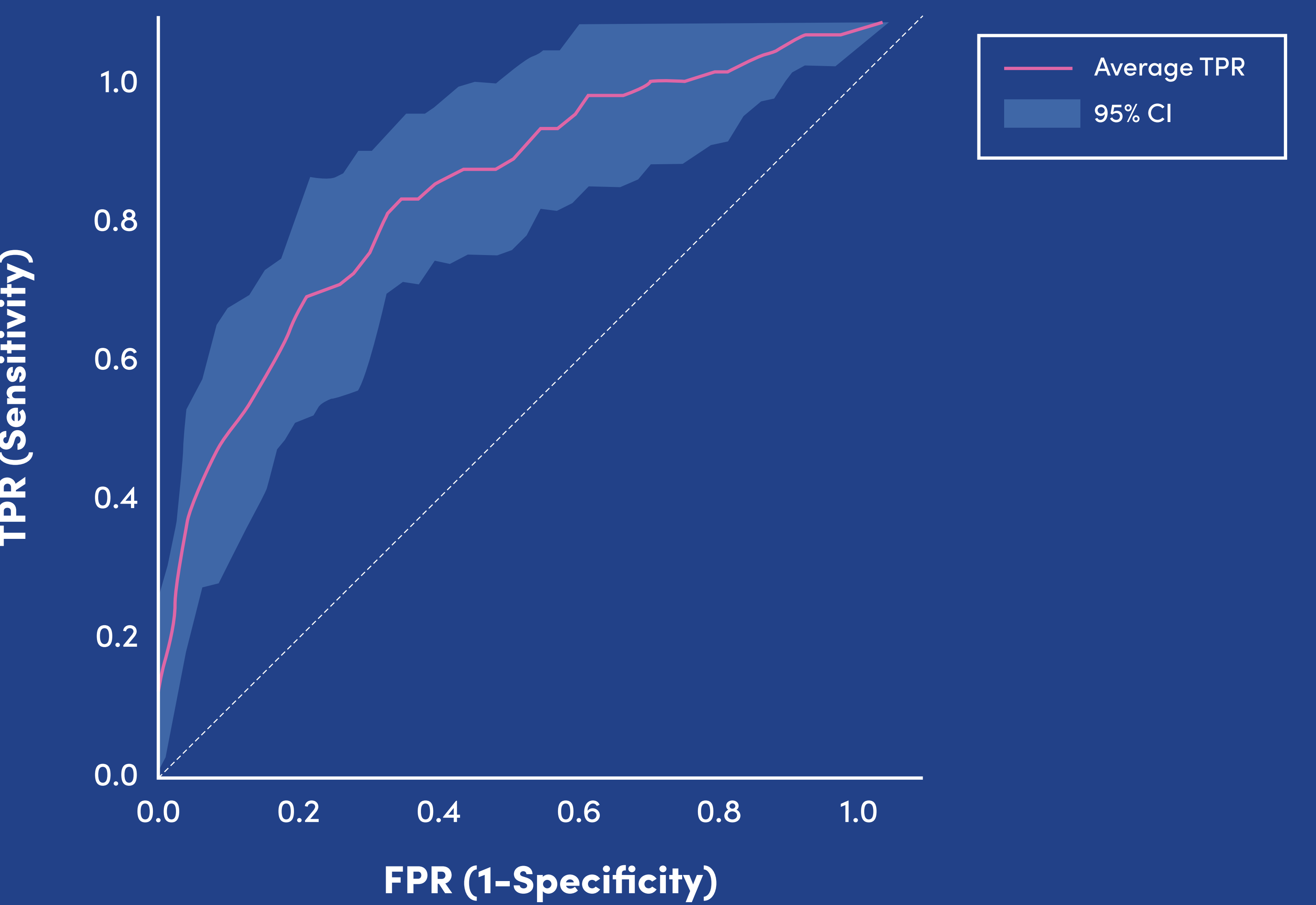


Conclusion:

Our AI system applied to H&E-stained WSIs can distinguish between BCs lacking any HER2 protein and mRNA (HER2 true negative) and HER2-low tumors, warranting further validation in cohorts of patients treated by new HER2 ADCs to support its use in trials and future clinical decision-making.

Slide Level Classification

AUC: 0.78 ±0.08
14.75 ± 4.2 +ve samples per fold. 172.12 ± 16.6 -ve



HER2 Model Performance

Metrics	Case Level	Slide Level
Sensitivity*	0.78	0.73
Specificity*	0.78	0.77
PPV*	0.23	0.21
NPV*	0.98	0.97

*Calculated on test set

Sensitivity: # true positive / (# true positive + # false negative)
Specificity: # true negative / (# true negative + # false positive)
PPV: # true positive / (# true positive + # false positive)
NPV: # true negative / (# true negative + # false negative)