Deep learning-based assessment of HER2-low expression on breast cancer H&E digital whole slide images

Oakley, J.*, Reis-Filho Jf, Klimstra D, Goldfinger MH, Wang YK, Marra A*.

*Paige AI, 47 Times Sq, New York, NY 10036, USA. 1Department of Pathology, Memorial Sloan Kettering Cancer Center, 125 York Ave, New York NY 10065

Background:
Antibody drug conjugates (ADCs) against HER2 have shown meaningful clinical activity in HER2-low breast cancer, defined as an ERBB2 expression by immunohistochemistry (IHC) with no gene amplification by in situ hybridization (ISH) techniques. However, these methods were originally developed for accurate detection of HER2 gene amplification (ISH), not IHC, their sensitivity and robustness to the detection of low levels of ERBB2 transcription and translation are not well understood. We have recently described a deep learning algorithm that can detect signatures of HER2 expression based on tissue slides scanned H&E whole slide images (WSI) of breast cancers for which IHC and mRNA expression levels of HER2 are available. Here we describe the application of our algorithm to two independent breast cancer cohorts.

Methods:
A model was developed based on recognition of invasive breast cancer is HER2 IHC, and then trained via a computational neural network with multiple instances using the heavy public dataset of genes on HER2-negative and HER2-expressed (low) AI. For training, low negatives were defined as having HER2 IHC 0 and mRNA below 15% HCC in tissue slides, while high negatives were defined as IHC 0 and mRNA 0% (Oncomine HER2 gene expression). IHC 0 cases with mRNA 0% were removed from the training cohort. The resulting model (HER2Complete) was used to distinguish HER2-negative from HER2-low cases with an AUC of 0.91 (95% CI: 0.88-0.93). HER2Complete was then made public for assessment against HER2 IHC and mRNA expression of 20,603 breast cancers in TCGA and TCGA-Ovarian, respectively. For the TCGA cohort, immunohistochemistry transcriptional data (RNAs) as a reference for HER2 mRNA expression were determined and HER2 expression was defined as IHC 0 or 1+ and the 99th percentile of the geometric mean of expression of three reference genes not expressing. In About TCGA. Values less than this reference cut-off in the TCGA cohort were considered HER2 "not expressed.""