Prospective Clinical Trials

Breast cancer tissue (purchased from Avaden Biosciences or anonymized) were stained with HER2 Immunohistochemistry (IHC; Ventana HER2 [4B5] Assay) and digitized into whole slide images (WSI) using Aperio ScanScope XT slidescanners for the model training, pathologist label collection, and validation phase. N=689 breast cancer tissue samples were stained with HER2 IHC stain (Ventana HER2 [4B5] Assay) and digitized into whole slide images (WSI) across 5 laboratories in the US. Breast cancer tissue (Harvested from Avidex Biosciences or anonymized from the AstraZeneca biobank) was from primary and metastatic tumors, core needle biopsy and surgical resections, lobular and ductal carcinoma across tumor grades and HER2 expression levels and cancer types. WSI were stratified into training (n=407), validation (n=110), and test sets (n=173). Multiple convolutional neural network (CNN)-based ML models were trained and validated an automated ML-based model as a quality control tool for HER2 testing and monitoring in clinical trials. The ML model was trained using whole slide images (WSI) from multiple sources to quantify HER2 expression, and measure stain intensity, antigen content, tumor area, and DCIS (ductal carcinoma in-situ) across a diversity of breast cancer phenotypes. Model quantified HER2 scores were consistent with pathologist consensus scores across breast cancer tissue types. These results support incorporation of ML-based algorithms into clinical trial workflows to monitor HER2 testing quality including scoring, tissue quality, and assay performance.

**Cell-Level Scores**

ML cells were validated against a consensus (median) of manual counts from 5 independent pathologists in 320 representative frames (Figure 1B,C). In pre-test data, there was strong agreement between ML-model and pathologist consensus scores for all cell types except for fairly positive HER2 cells where ML-based quantification identified more cells on average (Table 1). Table 1. ML Model Quantified and Pathologist HER2 Cell Level Scores Pearson correlation values for consensus cell count correlation with ML Model in evaluated frames, 95% CI.

**Slide-Level Scores**

HER2 slide-level scores were generated by automatically applying the rules derived from 2018 ASCO/CAP guidelines and compared with consensus scores from 3 independent pathologists in the test set (Figure 3). In the test set, automatically generated ML-ASCO/CAP scores showed substantial consistency with the pathologists’ consensus scores across the IHC categories (ICC 0.88 [95% CI 0.82-0.92]) and Figure 4, Precision Score. Agreement improved further when models were trained to agree with pathologists by adjusting the cut-offs by moving ML predicted HER2 weak to moderate partial positive tumor cells to the 2+ score (ICC 0.91 [95% CI 0.89-0.93]) and Figure 4, Adjusted Score.

**Validation**

Figure 4. Confusion Matrices of Test Set (n=173)

Comparison of Pathologist and PathAI Precision Algorithm Scoring Precision scores (left) and Adjusted Score (right). The 2+ HER2 tumor score was generated by automatically applying the 2018 ASCO/CAP guidelines. Table 1. ML Model Quantified and Pathologist HER2 Cell Level Scores Pearson correlation values for consensus cell count correlation with ML Model in evaluated frames, 95% CI.

**Future Directions**

Deployment of the HER2 QC Tool for Use in Clinical Trials

Validated ML models were incorporated into a clinical trial monitoring tool that supports the uploading of WSI to the PathAI cloud-based platform, deployment of ML models, and reporting of case-level (Figure 5), and trial level results (Figure 6). The HER2 QC can be used at clinical laboratories to reproducibly and rapidly monitor sample adequacy and HER2 assays in active clinical trials. Figure 5. Sample Individual HER2 Case Report from the PathAI Clinical Trial Support Platform

Report shows ML model Adjusted and Precision scores, tumor area, and tumor cell count (positive and negative HER2 tumor cells). Additional readouts include artifact area, DCIS area and turn-around time.

**Figure 6. Sample HER2 Trial Report from the PathAI Clinical Trial Support Platform Generated Using Simulated Data**

Trial report contains comparisons of ML-model predicted and pathologist scores using the Precision (left) and Adjusted (right) algorithms.

**Deployment of HER2 QC Tool for Use in Clinical Trials**

Validation of HER2 scores

Figure 1A. HER2 Negative Cancer Cell.

Figure 1B. HER2 Weak To Moderate Complete Membranous Positive Cancer Cell.

Figure 1C. HER2 Complete Membranous Positive Cancer Cell.

Figure 1B,C. In test set data, there was strong agreement between ML-model and pathologist consensus scores for all cell types except for fairly positive HER2 cells where ML-based quantification identified more cells on average (Table 1). Table 1. ML Model Quantified and Pathologist HER2 Cell Level Scores Pearson correlation values for consensus cell count correlation with ML Model in evaluated frames, 95% CI.

**Authors**

Benjamin Glass1, Michel Emilio Vandenbergh2, Surya Teja Chavali, Syed Ashar Jose1, Martin Fehlen4, Shantini Sridharan, Hunter Elliott1, Sootha Rao1, Michael Montalvo, Murray Resnick1, Ilan Wapinski3, Andrew Beck1, Craig Barker2, 1PathAI, Boston, MA; 2AstraZeneca R&D, Cambridge, United Kingdom; 3AstraZeneca, Gaithersburg, MD

**Acknowledgments**

This study was funded by AstraZeneca Research and Development. This poster template was developed by SciStories LLC. https://scistories.com/