Artificial Intelligence (AI) Assisted Detection of Microsatellite Instability / Mismatch Repair Deficiency (MSI-H / dMMR) in Multiple Tumor Types from Whole Slide Images (WSI) of H&E Sections

Joe Oakley[1]; David S Klimstra[1]; Yikan Wang[1]; Jeremy D Kunz[1]; Jillian Sue[1]; Matthew C. H. Lee[1]; Ran A Godrich[1]; Jan Bernhard[1]; Lida Tydlitatova[1]; Chad Vanderbilt[2]

1. Paige.AI Inc., 11 Times Square, 37 Floor, New York, USA 2. Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY 10021, USA

Authors JO, DK, YW, JK, JS, ML, RG, JB, and LT report they are employees of Paige, with salary and stock compensation, and Paige sponsored this work.
Author CV has Equity; Intellectual Property Rights; Professional Services and Activities; interests with Paige.AI, Inc.: (Uncompensated).

Background:

Universal screening for MSI-H / dMMR in CRC for Lynch syndrome is a formal recommendation of the National Comprehensive Cancer Network (NCCN).[1] Screening gastric carcinoma (GC) and CRC for MSI-H/dMMR to predict immunotherapy response is recommended by the College of American Pathologists (CAP).[2] Currently, this screening is performed by tissue and cost intensive methods such as immunohistochemistry (IHC), polymerase chain reaction (PCR) and/or next generation sequencing (NGS). We sought to create AI assisted screening of WSI of H&E sections from CRC and GC for MSI-H/dMMR as a tissue sparing, less expensive screen to triage cases as either likely or unlikely to harbor MSI-H/dMMR.

Design:

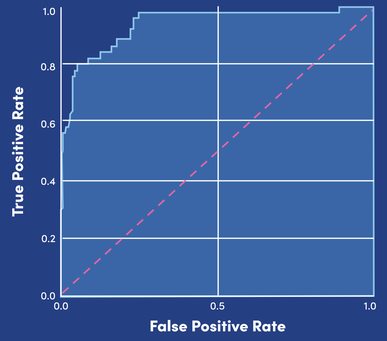
1,366 WSIs of H&E-stained CRC sections with known MSI-H status were used to train an AI-algorithm which is trained in phases. Phase 1 uses features from a colon cancer detection model to train an MSI-H/dMMR attention-based classifier. The model from phase 1 is used to attend regions of high activation to collect patch-based pseudo-labels to train a fully supervised CNN in phase 2. Finally, an attention-based classifier is trained on the MSI specific features from phase 2. We tested against 900 CRC cases of known MSI-H status from internal, The Cancer Genome Atlas (TCGA), and Pathology AI Platform (PAIP) datasets, noting that we do not train on TCGA or PAIP data.

The phase 2 model is again used to collect pseudo-labeled patches on a combined CRC and GC dataset which includes 587 additional WSIs of H&E-stained sections with known MSI-H status. This allows us to train another fully supervised MSI-H specific feature extractor for CRC and GC combined. A final attention-based classifier was trained on these features. This model was tested against 480 GC cases of known MSI-H status from internal and TCGA datasets.

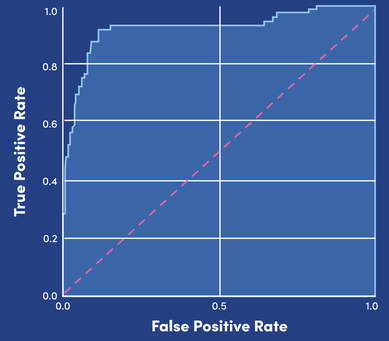
Results:

An area under the receiver operating characteristic (AUC) curve of 0.939 in CRC and 0.905 in GC were obtained with the models on internal hold out data. MSI-H was identified in TCGA with sensitivity / specificity / AUC of 0.917 / 0.762 / 0.929 in CRC and 0.918 / 0.510 / 0.842 in GC. MSI-H in the PAIP dataset was identified with sensitivity / specificity / AUC of 0.917 / 0.8 / 0.945. Sensitivity of 0.917 with 0.765 specificity was obtained for combined CRC in TCGA and PAIP.

Internal Validation Dataset (AUC=0.9340)



TCGA Colorectal MSI Dataset (AUC=0.9226)



PAIP MSI Dataset (AUC=0.9571)

Conclusion:

We trained an AI-assisted digital pathology assay for MSI-H/dMMR from WSIs of H&E-stained sections in CRC (AUC 0.939) and GC (AUC 0.905), and validated on external TCGA and PAIP datasets.[3]

Our assay shows superior performance characteristics to those presented for previous digital assays for MSI-H in CRC.

We show the ability to detect MSI-H accurately in multiple tissue types. The CAP recommends MSI-H/dMMR testing in solid tumors where immunotherapy is being considered. However, this is rarely done and the best assay to use is not always certain, as low prevalence has discouraged widespread screening outside of CRC, endometrial cancer and GC.[1]

Our success in training a multi-histology MSI-H model, coupled with future work using a digital pathology Foundation Model, may result in a digital pathology assay that can efficiently screen for MSI-H in many tumor types, and helping identify patients who may benefit from immunotherapy.[4]

We also plan to conduct real world studies of our assay to confirm its performance in clinical settings, and prove the medical economic benefit of triaging MSI-H/dMMR testing with digital pathology.

Citation:

1. NCCN Guidelines Version 2.2023 Genetic/Familial High-Risk Assessment: Colorectal.
2. Bartley AN, Mills AM, Konnick E, et al. Mismatch repair and microsatellite instability testing for immune checkpoint inhibitor therapy: Guideline from the College of American Pathologists in Collaboration with the Association for Molecular Pathology and Fight Colorectal Cancer. Arch Pathol Lab Med. 2022;146(10):1194-1210. doi:10.5858/arpa.2021-0632-CP
3. Svrcek M, Saillard C, Dubois R, et al. Blind validation of MSIntuit, an AI-based pre-screening tool for MSI detection from colorectal cancer H&E slides. Poster presented at: European Society for Medical Oncology (ESMO); May 9th - 13th 2022; Paris France.
4. Vorontsov E, Bozkurt A, Casson AI et al. Virchow: A Million-Slide Digital Pathology Foundation Model. ArXiv. Updated preprint posted online January 18, 2024. arXiv:2309.07778v5

Paige®