H&E-Based MSI Predictor in Prostate Cancer Generalizes to External Site Stain and Scanner Characteristics

Poster 218

Geoffrey Schau[1], Kunal Nagpal[1], Rohan Joshi[1], Rachel Baits[1], Sebastian Pretzer[1], Irvin Ho[!], Adam Cole[2,3], Roberto Nussenzveig[3], Abigail Gordhamer[2,3], Martin C Stumpe[1], Nike Beaubier[1], and Matt Leavitt[2,3,4]

[1]Tempus Labs, Inc., Chicago, IL // [2]PathNet Labs, Little Rock, AR // [3]DDx Foundation, Lehi, UT // [4]Lumea, Lehi, UT // **Correspondence:** geoffrey.schau@tempus.com

Disclosure: Tempus Labs, PathNet Labs, and Lumea are a for-profit companies

INTRODUCTION

Microsatellite instability (MSI) is associated with patient response to immunotherapy in several cancer types. Previous studies have shown that artificial intelligence-based imaging assays can infer MSI status from H&E whole slide images (WSIs) but assessment of external site generalizability remains a key challenge. In this study, we develop and evaluate a model trained to predict MSI status from WSIs in prostate cancer. Further, we directly evaluate stain and scanner generalizability by assessing our model on an internal test set and an external test set that contains a serial section of each slide in the internal test set but stained at a different site and scanned using a different scanner model.

METHODS

This study assessed a real-world data cohort composed of 2,253 patient samples of primary prostate cancer with digitized H&E images and next-generation sequencing (NGS) confirmed MSI status (69 MSI-High [MSI-H], 2183 microsatellite stable [MSS]) from the development site. Of these, 114 cases (MSI-H= 29, MSS=85) were assigned to an internal test set and held-out from training. All internal test set slides were scanned on a Leica GT450 and a serial section for each of these samples was stained at an external site with a Leica AT2 (Table 1). All remaining data were assigned to a 4-fold cross-validation development dataset (MSI-H=40, MSS=2098), which was used to train four ensemble of attention-based multiple instance learning (MIL) models to predict MSI-H with a ResNet-18 backbone. Prediction scores from the four models were averaged to produce a final MSI-H prediction score for each of the test set slides.

Acknowledgements: We thank the Scientific Communications for visualization and poster review.

References:

1. Alam MR, Abdul-Ghafar J, Yim K, Thakur N, Lee SH, Jang H-J, Jung CK, Chong Y. Recent Applications of Artificial Intelligence from Histopathologic Image-Based Prediction of Microsatellite Instability in Solid Cancers: A Systematic Review. Cancers. 2022

2. Hildebrand LA, Pierce CJ, Dennis M, Paracha M, Maoz A. Artificial Intelligence for Histology-Based Detection of Microsatellite Instability and Prediction of Response to Immunotherapy in Colorectal Cancer. Cancers. 2021; 13(3):391

SUMMARY

- We developed a paired dataset of adjacent slides stained and scanned at **different clinical sites** to directly measure the **generalizability** of an imaging-based deep learning model
- We used this data to illustrate that a model trained to **predict MSI status from H&E images** is robust to differences across institutions in predictive performance, but may need calibration prior to setting prediction thresholds.

RESULTS

Paired dataset generation for generalization evaluation

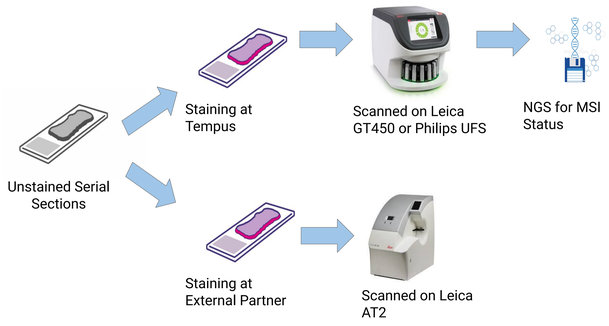

Figure 1. A paired generalization dataset is composed of adjacent serial sections stained and scanned at two different sites

	Train Cohort		Test Cohort	
	MSS	MSI-H	MSS	MSI-H
Gleason Score				
Unknown	806	15	26	8
6	11	0	0	0
7	269	1	4	0
8	233	5	4	2
9	690	14	26	9
10	89	5	25	10
Procedure Type				
Unknown	186	1	4	0
Biopsy	1135	28	51	17
Resection	777	11	30	12
Total	***2098***	***40***	***85***	***29***

Table 1. Patient cohort and clinical description

Characterization of model performance on both internally and externally stained test sets

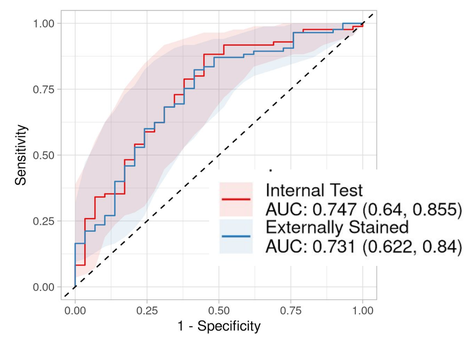

Figure 2. Receiver operating characteristic curves for both development and deployment sites are similar, suggesting robust generalizability of the MSI predictive model (area under the curve[AUC] with 95% CIs).

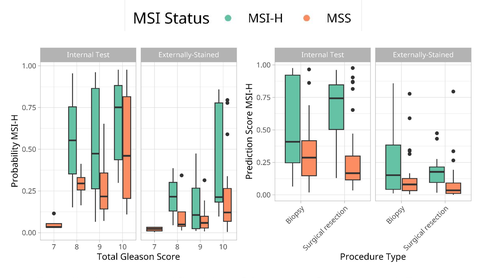

Figure 3. Predictions stratified by Gleason Score and Procedure Type. In both the internal and externally-stained test sets, MSI prediction scores were greater in GS>7 and in surgical resections.

High-attention tiles illustrate a focus on dense tumor tissue

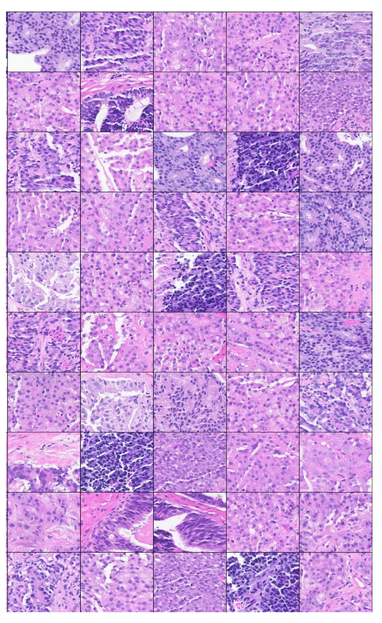

Figure 4. The 50 highest-attention tiles across all MSI-H slide images in the test set suggest the model tends to focus on high-grade, dense tumor tissue when predicting high microsatellite instability.