

ARTIFICIAL INTELLIGENCE (AI)-POWERED IMMUNE PHENOTYPING OF ADVANCED OR METASTATIC UROTHELIAL CARCINOMA (AUC) CLINICAL TRIAL SAMPLES FROM HEMATOXYLIN AND EOSIN (H&E)-STAINED WHOLE SLIDE IMAGES (WSI)

¹Jake Conway*, ¹Limin Yu, ¹Yash Belhe, ¹Suyog Jain, ²Hans Grote, ¹Mary Lin, ¹Benjamin Trotter, ¹Darpan Sanghavi, ²Aslihan Gerhold-Ay, ³XiaoZhe Wang, ¹Sergine Brutus, ²Thomas Mrowiec. ¹PathAI, Boston, MA, United States; ²The Healthcare Business of Merck KGaA, Darmstadt, Germany; ³EMD Serono, Billerica, MA, United States

Background CD8 immune phenotype status is associated with response to anti-PD-L1 therapy in aUC. To assess the tumor microenvironment in aUC, we developed machine learning (ML)-based models to identify cell types and tissue regions in digitized H&E-stained WSI from the JAVELIN Bladder 100 trial, which showed that avelumab (anti-PD-L1) first-line maintenance plus best supportive care (BSC) significantly prolonged overall survival vs BSC alone in patients with aUC. ML-quantified features were used for subsequent slide-level immune phenotyping of these clinical trial samples.

Methods Models previously trained using samples from The Cancer Genome Atlas were refined using 25,926 tissue regions annotations and 183,259 cell type annotations on 704 formalin-fixed, paraffin-embedded, H&E-stained WSI scanned on MIRAX (40x) to identify artifacts, tissue regions (cancer, stroma, and necrosis), and cell types (cancer epithelial cells, lymphocytes, macrophages, fibroblasts, and granulocytes). Precision, recall, and F1 scores were calculated to evaluate model performance. Following pathologist guidelines, H&E slide-level digital immune phenotypes (excluded, inflamed, and desert) were determined using thresholds of lymphocyte area within stroma and proportion of lymphocytes vs all cells in cancer epithelium. The distribution of H&E slide-level digital immune phenotypes was calculated across samples in both trial arms, and the association between immune phenotype and CD8 score at the tumor core was determined using the Kruskal-Wallis and Mann-Whitney U tests.

Results Precision, recall, and F1 scores for model predictions were comparable to those of an average annotator across cell types. The model's concordance with consensus was also higher than that of an average human annotator (Cohen's kappa, 0.816 vs 0.680). Calculation of the distribution of H&E slide-level digital immune phenotypes across trial arms showed that most samples in each arm had an excluded immune phenotype, and a slightly higher proportion of samples in the BSC-only arm had an inflamed phenotype. Association between immune phenotype and the gold-standard CD8 immunohistochemistry (IHC) score at the tumor core showed that samples with the inflamed immune phenotype had higher CD8 scores at the tumor core than samples with the excluded immune phenotype (Mann-Whitney, $p < 0.026$).

Conclusions We show an AI-powered approach to determine slide-level immune phenotypes directly from digitized H&E-stained WSI of clinical aUC samples. An association between inflamed immune phenotypes and higher CD8 scores at the tumor core supports the potential use of this method as an alternative to CD8 IHC-based approaches to identify patients who may derive optimal benefit from anti-PD-L1 treatment.

Acknowledgements This research was supported by the healthcare business of Merck KGaA, Darmstadt, Germany (CrossRef Funder ID: 10.13039/100009945) as part of an alliance between the healthcare business of Merck KGaA, Darmstadt, Germany and Pfizer.

Trial Registration NCT02603432 (ClinicalTrials.gov)

Ethics Approval The trial protocol was approved by the independent ethics committee or institutional review board at each participating center.

<http://dx.doi.org/10.1136/jitc-2022-SITC2022.0554>