Artificial intelligence–powered immune phenotyping of advanced or metastatic urothelial carcinoma clinical trial samples from hematoxylin and eosin–stained whole slide images

J. Conway,[1] L. Yu,[1] Y. Belhe,[1]* S. Jain,[1]* H. J. Grote,[2] M. Lin,[1] B. Trotter,[1] D. Sanghavi,[1] A. Gerhold-Ay,[2] X. Wang,[3] S. Brutus,[1] T. Mrowiec[2]

[1]PathAI, Boston, MA, USA; [2]The healthcare business of Merck KGaA, Darmstadt, Germany; [3]EMD Serono, Billerica, MA, USA
*Affiliation at the time the research was conducted.

SCOPE

- **We report an artificial intelligence (AI)–powered approach to determine slide–level immune phenotypes directly from digitized hematoxylin and eosin (H&E)–stained whole slide images (WSI) of clinical advanced or metastatic urothelial carcinoma (UC) samples, with potential use as an alternative to CD8 immunohistochemistry (IHC)–based approaches to identify patients with hot and cold tumors**

CONCLUSIONS

- **An association between inflamed immune phenotypes and higher CD8 scores at the tumor core supports the potential use of this AI method as an alternative to CD8 IHC-based approaches to identify patients with hot and cold tumors in advanced or metastatic UC**

Correspondence: **Jake Conway, jake.conway@pathai.com; Sergine Brutus, sergine.brutus@pathai.com; Thomas Mrowiec, thomas.mrowiec@emdgroup.com**

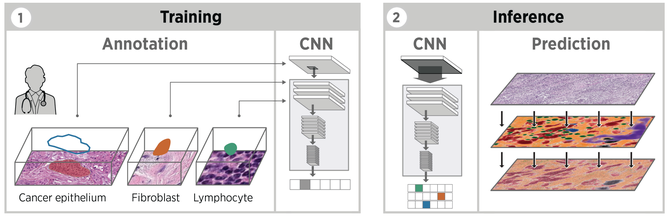
BACKGROUND

- First-line (1L) maintenance with avelumab, an anti–PD-L1 antibody, combined with best supportive care (BSC) significantly prolonged overall survival vs BSC alone in the JAVELIN Bladder 100 clinical trial enrolling patients with advanced or metastatic UC that had not progressed with 1L platinum-containing chemotherapy[1,2]
- Results from this trial led to the approval of avelumab 1L maintenance in various countries worldwide, and the JAVELIN Bladder regimen is an established standard of care in international treatment guidelines based on level 1 evidence[3-8]
- PathAI machine learning (ML) models have previously been deployed to characterize the tumor microenvironment in a range of cancer types[9]

METHODS

- To assess the tumor microenvironment in advanced or metastatic UC, we developed ML–based models to identify cell types and tissue regions in digitized H&E–stained WSI from the JAVELIN Bladder 100 trial. ML–quantified features were used for subsequent slide–level immune phenotyping of these clinical trial samples
- Models previously trained using samples from The Cancer Genome Atlas were refined using extensive tissue regions and cell type annotations on 703 formalin-fixed, paraffin–embedded, H&E–stained WSI scanned on MIRAX (40×) to identify artifacts, tissue regions (cancer, stroma, and necrosis), and cell types (cancer epithelial cells, lymphocytes, macrophages, fibroblasts, and granulocytes) (**Figure 1**). Precision, recall, and F1 scores were calculated to evaluate model performance
- In accordance with pathologist guidelines at Merck KGaA, Darmstadt, Germany, H&E slide–level digital immune phenotypes (excluded, inflamed, and desert) were determined using ML–derived features
- The distribution of H&E slide–level digital immune phenotypes was calculated across available 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

Figure 1. Model Development



Schematic illustrating CNN model development approach for pixel-level feature predictions.
CNN, convolutional neural network.

RESULTS

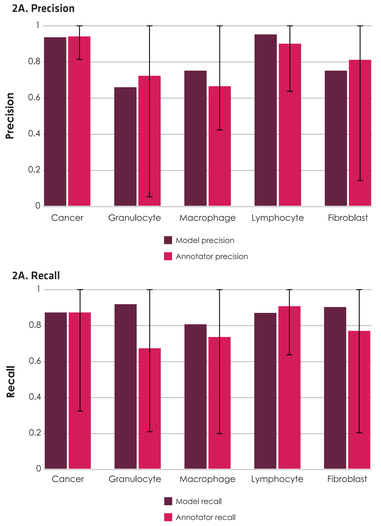
- Model performance evaluation in representative frames of tissue regions that were exhaustively annotated showed high correlation of model predictions and pathologist consensus ("model vs consensus"), comparable to agreement among 5 pathologists ("annotator vs consensus") (**Table 1**)
- Precision, 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 higher than that of an average human annotator (Cohen κ, 0.816 vs 0.680) (**Figure 2A** and **2B**)
- Calculation of the distribution of H&E slide–level digital immune phenotypes across trial arms showed that most samples in each arm were classified by the model as excluded (approximately 84% of all samples), followed by inflamed (approximately 15%). The model identified very few samples as desert (approximately 1%) (**Figure 3**)
- Association between immune phenotype and the gold standard CD8 IHC score at the tumor core showed that samples with the inflamed phenotype had higher CD8 scores at the tumor core than samples with the excluded phenotype (Mann-Whitney p<0.026) (**Figure 4**)
- Inflamed vs excluded slides show the difference in relative abundance of lymphocytes in tumor compartments (**Figure 5**)

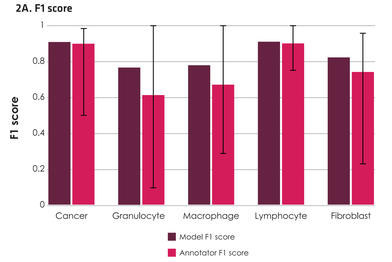
Table 1. Frames-based analysis of model performance

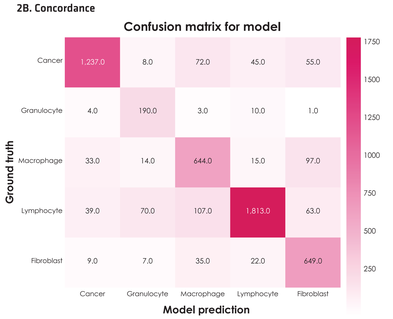
Cell model class	Interpathologist Pearson estimate (95% CI)*	PathAI algorithm Pearson estimate (95% CI)†
Cancer cell	0.89 (0.84-0.93)	0.82 (0.77-0.87)
Lymphocyte	0.89 (0.86-0.92)	0.90 (0.87-0.93)
Fibroblast	0.71 (0.63-0.78)	0.74 (0.66-0.80)
Granulocyte	0.65 (0.53-0.74)	0.73 (0.66-0.79)
Macrophage	0.44 (0.34-0.52)	0.53 (0.42-0.63)

Model performance vs consensus was evaluated on frames of small tissue regions that were exhaustively annotated.
*Annotator vs consensus. †Model vs consensus.

Figure 2. Model Performance



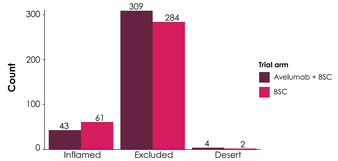




Confusion matrices	Model	Aggregated annotator
Accurate predictions, %	0.8647	0.8220
Innaccurate predictions, %	0.1353	0.1780
Concordance with consensus	**Model**	**Average annotator**
Cohen κ	0.816	0.680
Kramer V	0.800	0.761

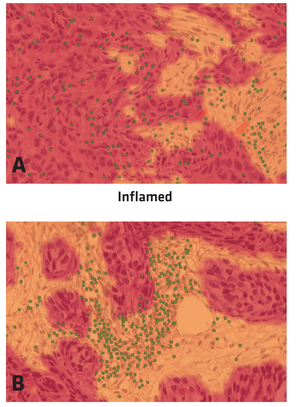
Average annotator matrix is computed by averaging per-class metrics for each annotator, for which each annotator has equal weight. Aggregated annotator confusion matrix is generated using all an2222notations, for which each annotation has equal weight.

Figure 3. Distribution of H&E slide–level immune phenotypes across trial arms



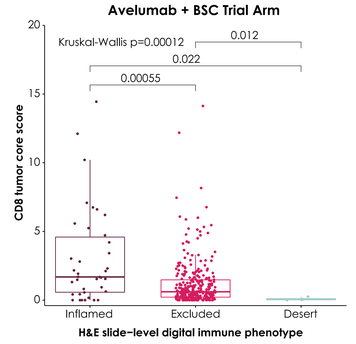
BSC, best supportive care; **H&E**, hematoxylin and eosin.

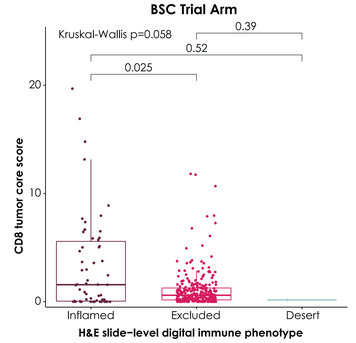
Figure 5. Machine learning model–generated tissue and cell overlays of WSI with inflamed vs excluded immune phenotypes



PathAI model–generated tissue- and cell-level overlays. (A) Overlays on WSI with inflamed phenotype. (B) Overlays on WS with excluded phenotype.
Tissue-level overlays show model predictions for regions of cancer epithelium (red) and cancer stroma (orange). Cell-level overlays show model predictions for lymphocytes (green).

Figure 4. Association between digital immune phenotype and CD8 IHC score





BSC, best supportive care; **H&E**, hematoxylin and eosin.

REFERENCES 1. Powles T, et al. N Engl J Med. 2020;383(13):1218-30. **2.** Powles T, et al. J Clin Oncol. 2022;40(Suppl 6). Abstract 487. **3.** Bavencio (avelumab). Prescribing information. EMD Serono: 2022. **4.** Bavencio (avelumab). Summary of product characteristics. Merck Europe BV, Amsterdam, the Netherlands, an affiliate of Merck KGaA, Darmstadt, Germany; 2022. **5.** NCCN Clinical Practice Guidelines. Bladder Cancer. V1.2022. **6.** Powles T, et al. Ann Oncol. 2022;33(3):244-58. **7.** Cathomas R, et al. Eur Urol. 2022;81(1):95-103. **8.** Japanese Urological Association. Accessed September 15, 2022. https://www.urol.or.jp/lib/files/other/guideline/39_bladder_cancer_2019_rev2021_info.pdf. **9.** Diao J, et al. Nat Comms. 2021;12:1613.

DISCLOSURES JC is an employee and shareholder of PathAI, Inc, Boston, MA, USA. **LY** is an employee and shareholder of PathAI, Inc, Boston, MA, USA. **YB** and **SJ** were employees of PathAI, Inc, Boston, MA, USA at the time research was conducted. **HJG** is employee of the healthcare business of Merck KGaA, Darmstadt, Germany and a shareholder in Merck KGaA, Darmstadt, Germany. **ML**, **BT**, and **DS** are employees and shareholders of PathAI, Inc, Boston, MA, USA. **AGA** is an employee of the healthcare business of Merck KGaA, Darmstadt, Germany. **XW** is an employee of EMD Serono and a shareholder in Merck KGaA, Darmstadt, Germany. **SB** is an employee and shareholder of PathAI, Inc, Boston, MA, USA. **TM** is an employee of the healthcare business of Merck KGaA, Darmstadt, Germany and a shareholder in Merck KGaA, Darmstadt, Germany.
ACKNOWLEDGMENTS This research was sponsored 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. Editorial support was provided by ClinicalThinking and funded by Merck and Pfizer.

Abstract No. 554. Presented at the Society for Immunotherapy of Cancer (SITC) 37th Annual Meeting November 8-12, 2022; Boston, MA, USA.