P730

An Empirical Framework for Validating Artificial Intelligence–Derived PD-L1 Positivity Predictions Applied to Urothelial Carcinoma

Andy Beck,[1] Benjamin Glass,[1] Hunter Elliott,[1] Jennifer K. Kerner,[1] Aditya Khosla,[1] Abhik Lahiri,[1] Harsha Pokkalla,[1] Dayong Wang,[1] Ilan Wapinski,[1] George Lee,[2] Vipul Baxi,[2] Cyrus Hedvat,[2]* Dimple Pandya,[2] Michael Montalto[2]

[1]PathAI, Boston, MA, USA; [2]Bristol-Myers Squibb, Princeton, NJ, USA

*At the time the analysis was conducted

Background

- Assessing programmed death ligand 1 (PD-L1) immunohistochemistry (IHC) expression plays an important role in identifying patients likely to benefit from anti–programmed death-1/PD-L1 therapies in advanced cancer, including urothelial carcinoma (UC)
- Studies have shown moderate-to-strong interobserver agreement for pathologist assessment of PD-L1 expression on tumor cells, with moderate-to-poor concordance for immune cell scoring[1–3]
- Thus, conventional pathologist estimation of whole-slide image scores is a suboptimal approach to obtain reference data for the evaluation of the performance of image-analysis algorithms, especially for immune cell scoring

Study Objectives

- Develop a platform to collect exhaustive annotations of PD-L1 positivity from a crowdsourced network of pathologists for analytic validation of artificial intelligence (AI)-based algorithms
- Evaluate the performance of an AI-based predictor of PD-L1 expression on tumor and/or immune cells in the tumor microenvironment using the validated platform

Methods

- PD-L1 expression was assessed by IHC using the PD-L1 IHC 28-8 pharmDx assay (Dako, Agilent Technologies Co)
- The training set consisted of 293 pretreatment samples from commercially obtained sources and from patients with platinum-resistant metastatic UC who were enrolled in clinical trials of nivolumab (CheckMate 032 [NCT01928394] and CheckMate 275 [NCT02387996])
  - From these, we obtained 105,514 annotations of tumor and immune cells from 43 pathologists
- To establish a reference dataset for manual vs digital concordance using our platform, we generated a subset of 80 images ("frames") sampled across different cell densities within a validation dataset comprising a subset of 100 samples from CheckMate 032 and CheckMate 275
  - We then collected exhaustive annotations from 5 different pathologists for each frame to produce quantitative estimates of PD-L1 positivity on tumor and immune cells in each frame of the validation dataset
- Altogether, 66,187 annotations (**Table 1**) were collected and used to compute pathologist consensus scores for each frame
- In the validation step these scores were then correlated with each individual pathologist (inter-reader agreement) and with the PathAI-derived automated scores for evaluation of manual vs digital agreement

Table 1. Summary of the number of annotations collected from 35 distinct pathologists used in generating a ground-truth dataset for concordance studies across 80 frames selected per cell type

Cell type	PD-L1 positive	PD-L1 negative	Total
Cancer cells	3728	14,328	18,056
Immune cells (macrophages + lymphocytes)	13,430	34,701	48,131
Total	17,158	49,029	66,187

Results

- The application of the frames-based validation framework is shown in **Figure 1**
- The PathAI platform (not intended for diagnostic use) showed significantly stronger correlation with reference median consensus scores compared with scores generated by individual pathologists for quantifying PD-L1 positivity of lymphocytes (r = 0.744 vs 0.598) and macrophages (r = 0.68 vs 0.287) (**Figure 2**)
- There was no significant difference in correlation with consensus between PathAI-derived and individual pathologist–derived assessment of positivity on tumor cells (r = 0.837 vs 0.857) (**Figure 2**)

Figure 1. Frames-based validation framework. Left panel: Process for recruiting and training pathologists. Right panel: Process for generating image frames, obtaining pathologist annotations, and comparing the pathologist consensus to the AI-derived scores

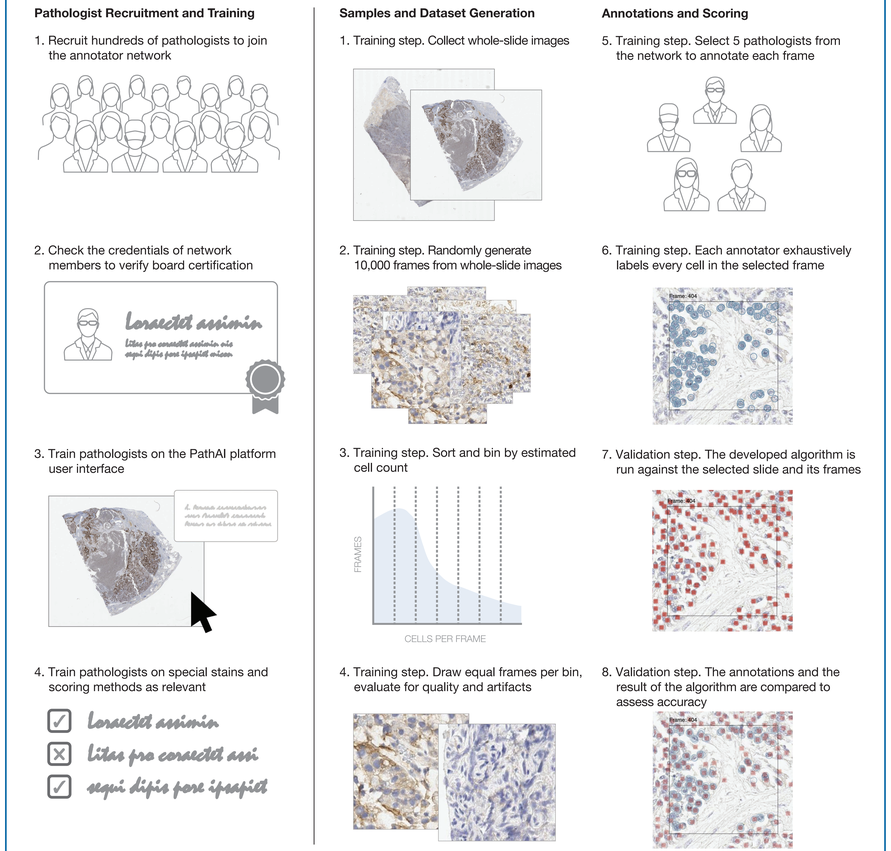
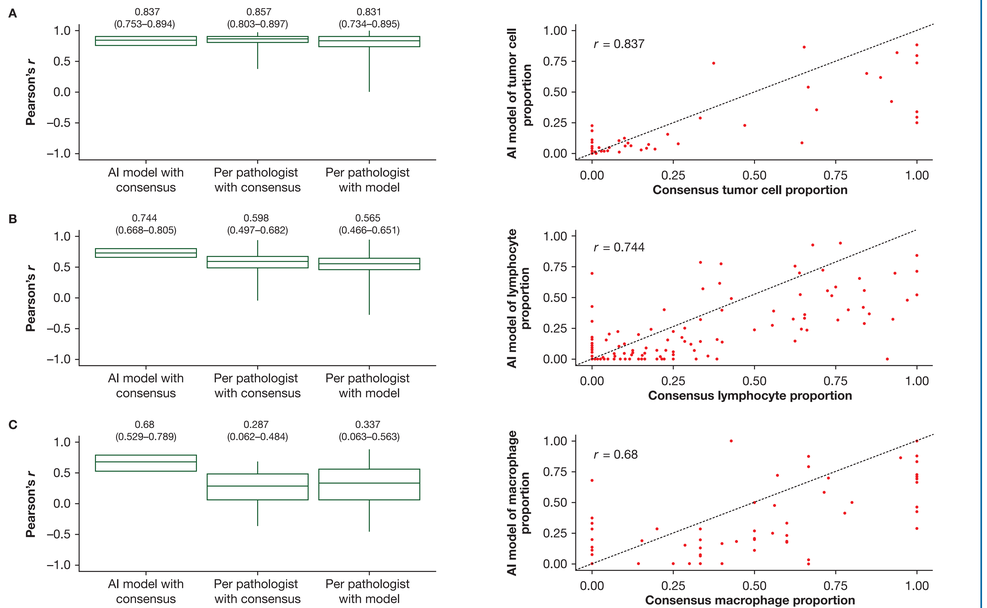


Figure 2. Correlation box plots of (A) tumor cell proportion, (B) lymphocyte proportion, and (C) macrophage proportion



The proportion score refers to the percentage of tumor, lymphocyte or macrophage cells that are PD-L1–positive. Values shown above the plots are median (95% CI) for AI model with consensus, and adjusted mean (95% CI) for per pathologist with consensus and per pathologist with model. Boxes indicate median or adjusted mean values and 95% CIs. Error bars represent the minimum and maximum values. CI, confidence interval.

Conclusions

- We validated performance of the PathAI platform for automated assessment of PD-L1 expression on tumor and/or immune cells and demonstrated that the AI-based predictors perform similar to or better than pathologist-based scoring in all cell types tested
- These results from a retrospective analysis support the concordance of the PathAI platform for PD-L1 quantification on tumor and/or immune cells in UC and suggest its applicability in other tumor types, including non-small cell lung cancer[4]
- These data demonstrate that AI-powered assessment represents a reproducible and potentially generalizable approach to interpretation of IHC assays

References

1. Zajac M, et al. *Diagn Pathol* 2019;14(1). doi: 10.1186/s13000-019-0873-6.
2. Rimm DL, et al. *JAMA Oncol* 2017;3:1051–1058.
3. Tsao MS, et al. *J Thorac Oncol* 2018;13:1302–1311.
4. Baxi V, et al. Oral presentation at the Society for Immunotherapy of Cancer (SITC) 34th Annual Meeting; November 6–10, 2019; National Harbor, MD, USA. Abstract O65.

Acknowledgments

- The patients and families who made the trials possible
- The clinical study teams who participated in these trials
- Bristol-Myers Squibb (Princeton, NJ) and ONO Pharmaceutical Company Ltd. (Osaka, Japan)
- Dako, an Agilent Technologies Inc company, for the anti-CD8 monoclonal antibody C8/144B and for collaborative development of the PD-L1 IHC 28-8 pharmDx assay
- All authors contributed to and approved the presentation; editorial assistance was provided by John Copier, PhD, and Jay Rathi, MA, of Spark Medica Inc, funded by Bristol-Myers Squibb

Email: Andy.Beck@pathai.com