Abstract #471

AIM PD-L1-NSCLC: Artificial intelligence-powered PD-L1 quantification for accurate prediction of tumor proportion score in diverse, multi-stain clinical tissue samples

BACKGROUND

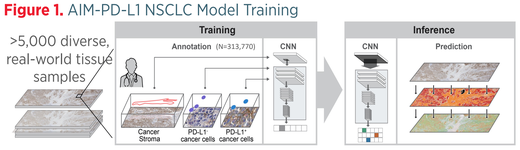
Patients with NSCLC can benefit from treatment with PD-L1-targeting immunotherapy, and current guidelines recommend quantification of the PD-L1 biomarker for patient tissue samples.[1] Currently, PD-L1 expression is assessed by pathologist using an approved PD-L1 immunohistochemistry assay.[2] However, manual assessment is challenging because the four different FDA approved different PD-L1 immunohistochemical assays have different scoring criteria.[2] Additionally, pathologist inter- and intra-variability can affect scoring.[3]

Here, we report the development and validation of machine learning (ML) models for the quantification of PD-L1 in non-small cell lung cancer (NSCLC) that are clone agnostic and can be incorporated into clinical trials with varied workflows.

METHODS

- AIM-PD-L1 NSCLC ML models, based on convolutional neural networks (CNNs), were trained using a large, diverse, real-world dataset of >5,000 clinical biopsies and resections of primary and metastatic NSCLC adenocarcinoma and squamous cell carcinoma (**Figure 1**).
- Samples were stained for PD-L1 expression using all four FDA approved PD-L1 clones **SP263** (N=1,320), **SP142** (N=1,829) (both Ventana Medical Systems Inc., Tucson AZ), **28-8** (N=1,331), or **22C3** (N=843) (both Agilent Technologies, Santa Clara CA) and digitized (Aperio AT2, Leica Biosystems Imaging, Vista, CA)
- Whole slide images (WSIs) were annotated with 313,770 annotations by board certified pathologists to label tissue regions (cancer epithelium, cancer stroma, and necrosis), and cell types (PD-L1+ and PD-L1- cancer cells, and PD-L1+ and PD-L1- immune cells, including lymphocytes, and macrophages).
- For each WSI, Human Interpretable Features (HIFs) representing the number of cancer cells were automatically extracted from the model and a slide-level Tumor Proportion Score (TPS) was calculated as the proportion of PD-L1+ cancer cells divided by the total cancer cells in a tumor region.

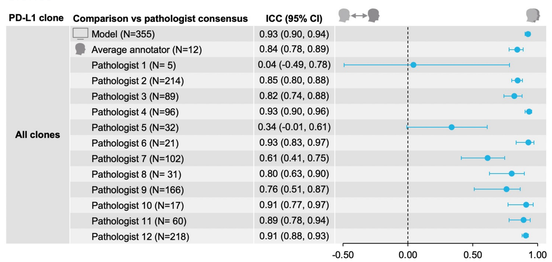
Figure 1. AIM-PD-L1 NSCLC Model Training



- Models were applied to 355 WSI not used for training (22C3 N=68; 28-8 N=137; SP142 N=72; SP263 N=78), and generated slide-level TPS scores.
- Slide-level TPS scores were also provided by 12 expert pathologists on WSIs
- Model performance was assessed by comparing model-predicted slide-level TPS with the consensus TPS of pathologists using intraclass correlation coefficient (ICC) statistics across all PD-L1 clones together, and individually for each clone.
- Pathologist performance across all PD-L1 clones together and individually for each clone was assessed using ICC statistics by:
  - Comparing the average TPS score for all scoring pathologists with the pathologist consensus TPS score
  - Comparing the performance of each individual pathologist with the consensus TPS score of all scoring pathologists

RESULTS

Figure 2. Comparison of Model-predicted and Pathologist TPS Across All PD-L1 Clones



AIM-PD-L1 TPS predictions were consistently highly concordant with the consensus TPS of all scoring pathologists across all four PD-L1 clones together and for each clone separately (Model vs pathologist consensus **Figures 2-6**). In each case, N refers to the number of datapoints or slides scored. Specifically, ICCs were calculated as 0.93 (95% CI 0.90-0.94) across all clones; 0.93 (95% CI 0.89-0.96) for 22C3; 0.90 (95% CI 0.85-0.93) for 28-8; 0.96 (95% CI 0.93-0.97) for SP263; 0.88 (95% CI 0.79-0.93) for SP142.

Figure 3. Comparison of Model-predicted and Pathologist TPS for 22-C3

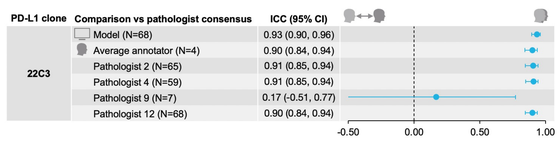
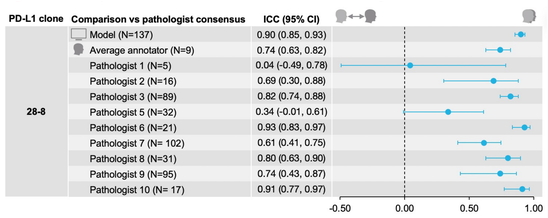


Figure 4. Comparison of Model-predicted and Pathologist TPS for 28-8



RESULTS

Average pathologist performance was assessed across all four clones (**Figure 2**) and for each clone separately (**Figures 3-6**). In all cases, there was high concordance between the average pathologist score and the pathologist consensus score (Average annotator vs pathologist consensus, **Figures 2-6**). Here, N refers to the number of annotators in the average.

We also investigated individual pathologist performance across across all four clones (**Figure 2**) and for each clone separately (**Figures 3-6**). Not all pathologists scored all slides, for each comparison the scoring pathologists are listed in the figure, and N refers to the number of datapoints or slides scored. Overall, there was high concordance between individual pathologist scores and the consensus scores, however, for each comparison broad variation in pathologist scores is observed. For three clones, there are pathologists that perform slightly better than the model (**Figures 4-6**) suggesting that the model may be outperformed by an individual expert pathologist but will consistently perform better than the average pathologist.

CONCLUSIONS

AIM-PD-L1 NSCLC predicted TPS scores with strong concordance with a pathologist consensus score across all four approved PD-L1 clones in a diverse dataset of real-world tissue samples.

Figure 5. Comparison of Model-predicted and Pathologist TPS for SP263

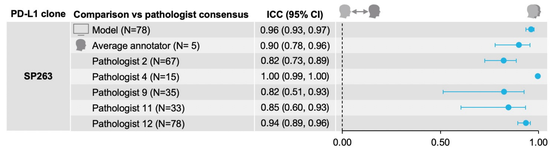
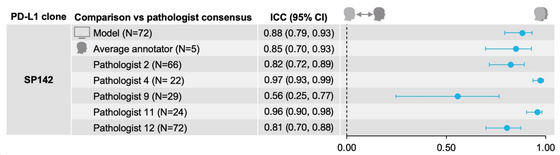


Figure 6. Comparison of Model-predicted and Pathologist TPS for SP142

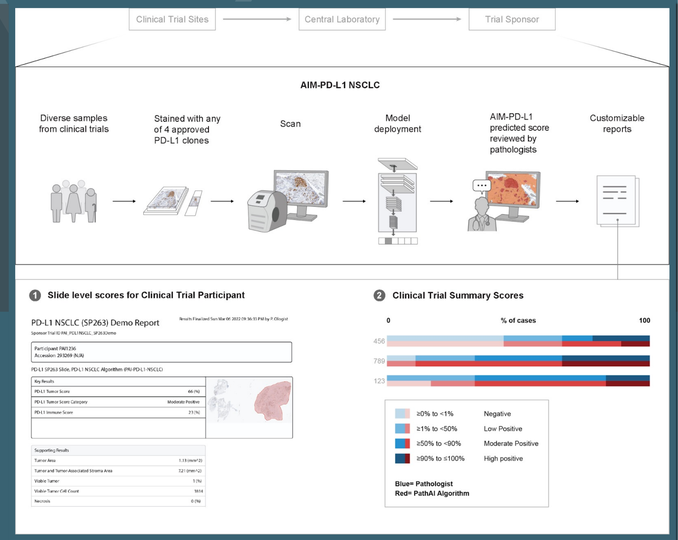


FUTURE DIRECTIONS

Figure 7 illustrates how the validated AIM-PD-L1 NSCLC models can be incorporated into a clinical trial digital pathology workflow facilitated by the PathAI Clinical Trial Services Platform. This platform supports multiple steps of digital pathology clinical trial workflows, enabling the automated ingestion of WSIs, deployment of ML models, pathologist review facilitation as specified by trial design, and reporting of case-level and trial-level results.

AIM-PD-L1 NSCLC can be used in clinical trials to reproducibly and rapidly quantify PD-L1 for patient enrollment and stratification.

Figure 7. AIM-PD-L1 NSCLC Incorporation Into Clinical Trial Workflows



AUTHORS

Michael Griffin, Mevlana Gemici, Ashar Javed, Nishant Agrawal, Murray Resnick, Limin Yu, Sara Hoffman, Victoria Mountain, Jamie Harisiades, Megan Rothney, Benjamin Glass, Ilan Wapinski, Andrew Beck, Eric Walk*

PathAI, Boston, MA

*contact Eric Walk : eric.walk@pathai.com

REFERENCES

1. Bironzo P, Di Maio M. A review of guidelines for lung cancer. J Thorac Dis. 2018;10(Suppl 13):S1556-S1563. doi:10.21037/jtd.2018.03.54
2. Kim H, Chung JH. PD-L1 Testing in Non-small Cell Lung Cancer: Past, Present, and Future [published correction appears in J Pathol Transl Med. 2020 Mar;54(2):196]. J Pathol Transl Med. 2019;53(4):199-206. doi:10.4132/jptm.2019.04.24
3. Rimm DL, Han G, Taube JM, Yi ES, Bridge JA, Flieder DB, Homer R, West WW, Wu H, Roden AC, Fujimoto J, Yu H, Anders R, Kowalewski A, Rivard C, Rehman J, Batenchuk C, Burns V, Hirsch FR, Wistuba II. A Prospective, Multi-institutional, Pathologist-Based Assessment of 4 Immunohistochemistry Assays for PD-L1 Expression in Non-Small Cell Lung Cancer. JAMA Oncol. 2017 Aug 1;3(8):1051-1058

ACKNOWLEDGMENTS

We thank Biosciences Communications for their expert assistance with data visualization. https://www.bioscicom.net/
This poster template was developed by SciStories LLC. https://scistories.com/

PathAI