Evaluating Image Analysis Approaches Toward "Harmonization" of PD-L1 Assays

Allison S. Harney, Staci J. Kearney, Carsten Schnatwinkel, Famke Aeffner, Luke Pratte, Zach Pollack, Jenifer Caldara, Karen Ryall, Joseph Krueger, Daniel Rudmann, Roberto Gianani

Flagship Biosciences

FlagshipBio.com | info@flagshipbio.com

Abstract

The commercial diagnostic landscape for PD-L1 immunohistochemistry (IHC) assays is highly complex. Multiple different companion or complementary diagnostic tests exist for therapeutics targeting the PD-1/PD-L1 pathway, each using a different interpretation to inform therapeutic decision-making. Flagship Biosciences envisions the utilization of Computational Tissue Analysis (cTA™) to develop an approach that could harmonize the interpretation of individual PD-L1 diagnostic tests. Specifically, when a single, continuous cTA-based scoring system is applied across each assay, the assays can be mathematically normalized, harmonizing PD-L1 assay scoring.

In a proof-of-concept study, non–small cell lung cancer (NSCLC) patient samples were stained with the FDA-approved Dako 28-8 and Dako 22C3 tests, as well as the in-house SP142 and E1L3N assays. The cTA platform was used to identify tissue and cell-specific Biofeatures™ and then generate digital scores for PD-L1 test comparison. The performance of the cTA platform in scoring a PD-L1 IHC assay was first examined by comparing the digitally generated PD-L1 scores for the 28-8 assay with (1) manual PD-L1 scores generated by multiple pathologists and (2) an orthogonal reference method (ie, NanoString™). The comparison of manual and digital scores (using cTA) demonstrated that the cTA approach significantly reduced variability in PD-L1 scoring. Additionally, the digitally generated PD-L1 scores showed better correlation to the reference method than did the manual PD-L1 scores.

Following evaluation of the cTA platform performance in scoring the 28-8 PD-L1 assay, the digitally generated scores for each of the 4 PD-L1 assays were compared. The FDA-AACR-ASCO "PD-L1 Blueprint" working group has previously identified similarities and differences between these 4 commercialized assays. Similarly, digital quantification of membrane staining intensity in the tumor compartment using the cTA platform showed that the average intensities of the 22C3 and 28-8 assays were similar, while the SP142 intensity was lower and the E1L3N intensity was higher. The percentage of PD-L1–positive cells identified in each assay was highly correlated across the reference range of PD-L1 expression for each assay. Based on the proof of concept demonstrated in this study, a cTA approach is a method that could potentially enable harmonization of the PD-L1 tests through use of a digital pathology platform.

Whole-Slide Scoring of PD-L1 Using cTA™

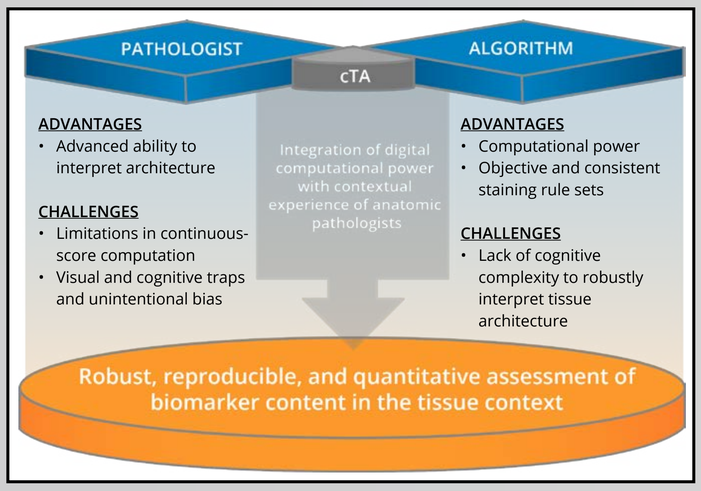
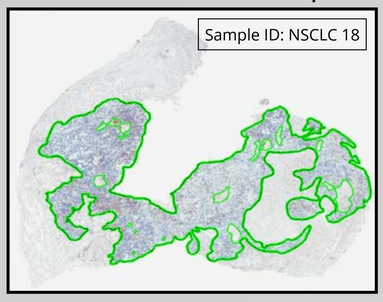


Image analysis tools overcome some of the challenges in conventional anatomic pathology practice, particularly for analyzing complex tissue architecture and heterogenous biomarker expression. In a computer-aided workflow, a digital image of the stained tissue is created with digital pathology components. Algorithms analyze the tissue captured in the high-resolution image and provide a digitally derived score. The use of cTA™-aided scoring allows for more accurate and direct cell counting and scoring across the whole slidethan can be achieved with manual scoring.

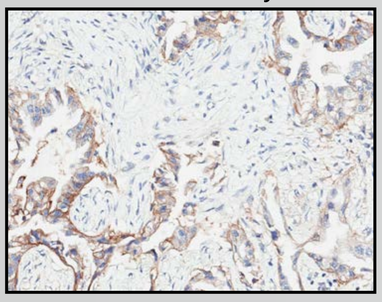
The cTA markup is a visual representation of the data generated by the algorithm.

cTA Markup: PD-L1⁻ Tumor Cell PD-L1⁺ Tumor Cell

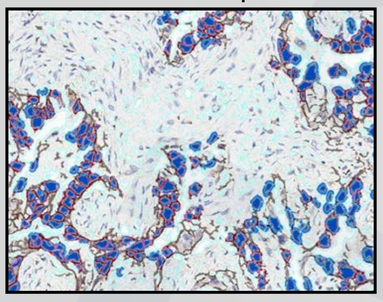
Whole-Slide cTA Markup



PD-L1 IHC Assay



cTA Markup



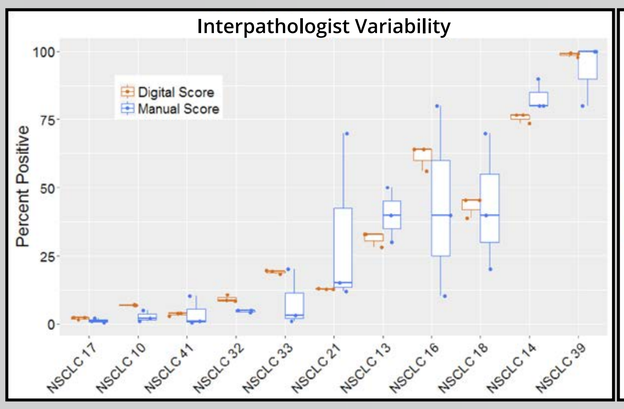
Interpathologist and Intrapathologist Scoring Variability

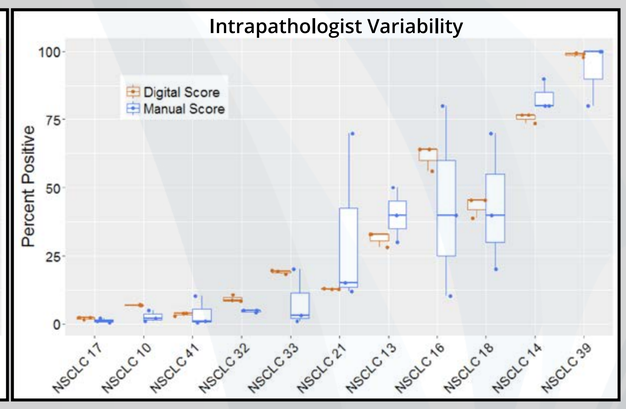
To examine the performance of the cTA™-based solution as a pathologist aid in comparison to manual pathology scoring, interpathologist and intrapathologist scores for PD-L1 were evaluated with whole-slide manual scoring and cTA-aided scoring. Interpathologist assessments for both manual scores and cTA-aided scores were from 3 different pathologists. For intrapathologist assessments, the same pathologist completed manual and cTA-aided scoring on 3 separate days with a 2-week washout period between scoring.

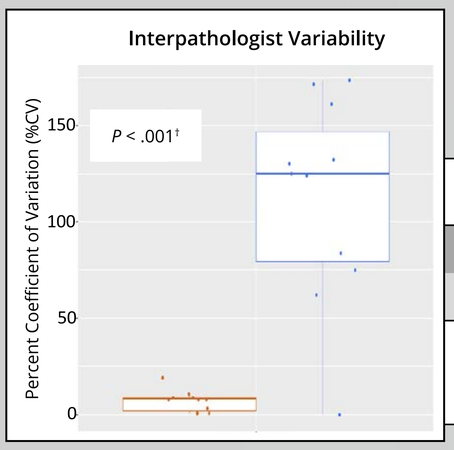
Pathologist assessments included manual pathology scoring and review of algorithm performance for cTA-aided scoring, including adjustment of certain algorithm parameters to increase accuracy of staining or cellular detection if appropriate.

For both interpathologist and intrapathologist variability assessments, the within-sample standard deviation of the cTA-aided score was less than that of the manual score in 10 out of 11 samples

The cTA platform significantly reduces variability in PD-L1 scoring.





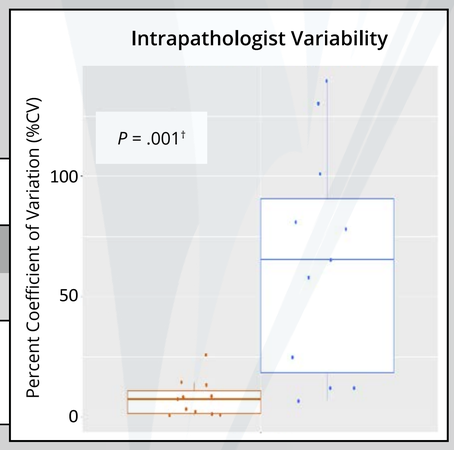


Digital scoring
Manual scoring

Median %CV	Manual	Digital
Inter-	124.9%	7.8%
Intra-	65.4%	7.6%

Each graph represents scoring of 11 PD-L1–stained NSCLC samples.

† Wilcoxon signed-rank test.

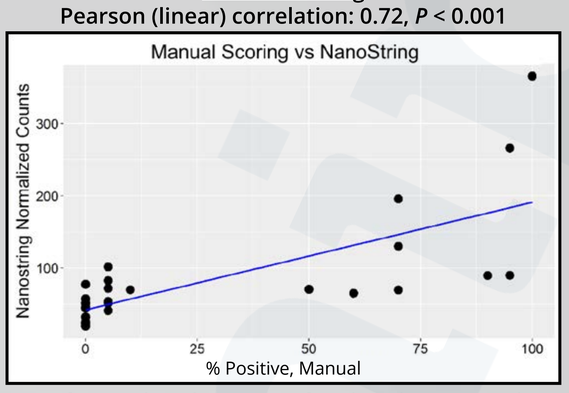


The intrasample %CV for both the interpathologist and intrapathologist assessments were significantly reduced for cTA-aided pathologist scoring (digital scoring) as compared to manual scoring by a pathologist.

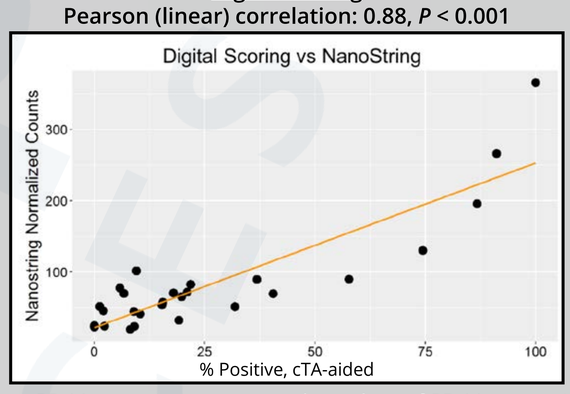
Comparison of PD-L1 Scoring to a Reference Standard

cTA™-aided scoring more accurately correlates with PD-L1 gene expression.

Manual Scoring
Pearson (linear) correlation: 0.72, *P* < 0.001



Digital Scoring
Pearson (linear) correlation: 0.88, *P* < 0.001



To confirm analytical accuracy of analyte detection in the IHC assays, PD-L1 scores were compared to values of PD-L1 expression derived from an orthogonal method, namely NanoString. cTA-aided digital PD-L1 scores and manual pathology scores were both compared to values of PD-L1 gene expression determined by NanoString using formalin-fixed, paraffin-embedded sections from the tissue blocks.

cTA-aided PD-L1 scoring improved performance across the diagnostic spectrum.

Cut-point Agreement: Manual vs Digital Scoring

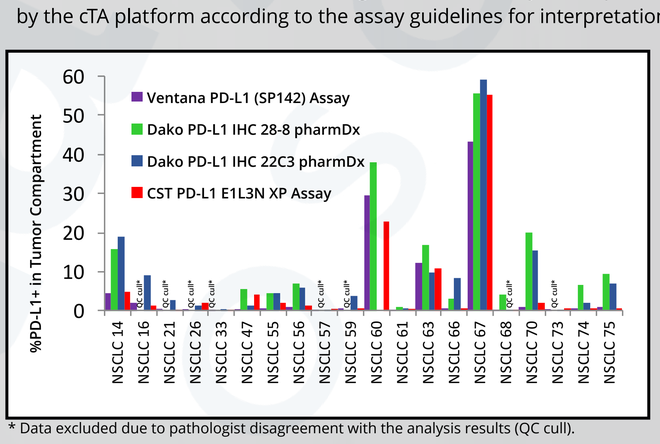
Cut-Point	1%	5%	10%	25%	50%
Agreement	66%	70%	70%	89%	85%
Positive Agreement	100%	94%	90%	78%	56%
Negative Agreement	18%	36%	59%	94%	100%

Assessment of agreement of manual and digital scoring data demonstrated poor negative agreement at lower percent-positive cut-points, indicating that potentially beneficial (positive) patient samples may be excluded based on manual pathology scoring. Overall, cTA-aided scoring has a higher number of samples that are found to be PD-L1 positive.

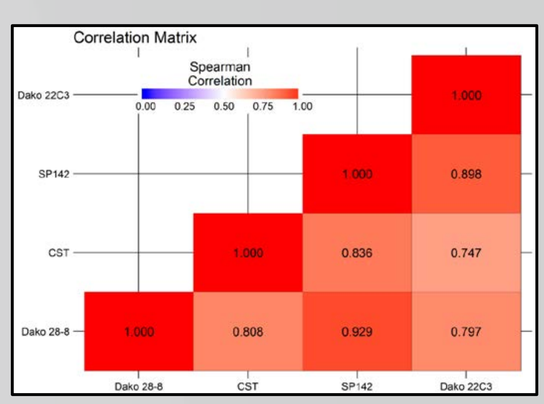
Comparison of PD-L1 IHC Assays Using cTA™-Aided Scoring

cTA™-aided PD-L1 scoring identified differences and similarities in IHC assay scoring.

Since cTA-based digital scoring of PD-L1 IHC assays provided a better diagnostic continuum, we used the cTA platform to investigate the similarities and differences among 4 PD-L1 IHC assays. The percentage of PD-L1-positive cells for each IHC assay was quantified by the cTA platform according to the assay guidelines for interpretation.



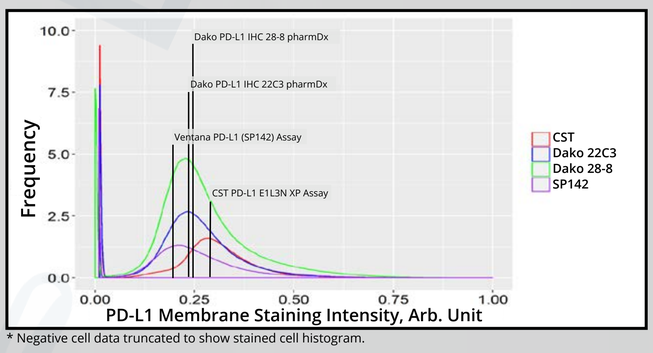
* Data excluded due to pathologist disagreement with the analysis results (QC cull).



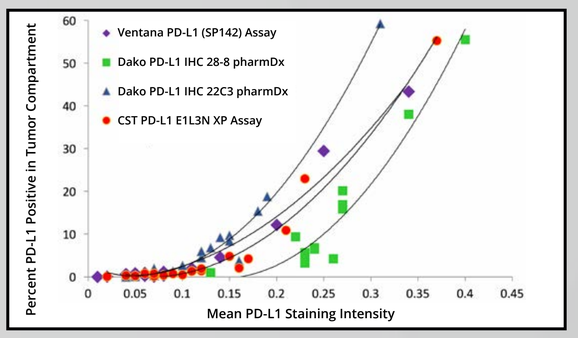
The comparison of the percentage of PD-L1-positive cells in each sample for the 4 IHC assays demonstrates that there are differences in cellular identification. To examine how similar the 4 assays were in identifying PD-L1-positive cells, samples were classified as PD-L1-positive or -negative at the 0% end point. The correlation matrix demonstrates the similarities in each of the PD-L1 IHC assays.

Assessment of PD-L1 IHC Scoring Harmonization Using Staining-Intensity Data

The cTA™ platform demonstrates that the relationship between PD-L1 scoring and IHC staining intensity is differential and inconsistent between and within assays.



* Negative cell data truncated to show stained cell histogram.



In understanding the differences among the 4 PD-L1 IHC assays, IHC staining intensity was investigated as a measurement that could be combined with the percentage of PD-L1-positive cells as a means of harmonizing scoring and interpretation of PD-L1 across all 4 assays.

While all 4 PD-L1 IHC assays had a strong correlation between mean membrane staining intensity and the percentage of PD-L1-positive cells, when a cTA-aided scoring method was used to determine the PD-L1 membrane staining intensity on a continuous scale, the CST PD-L1 E1L3N XP Assay demonstrated the highest PD-L1 membrane staining intensity overall, while the Ventana PD-L1 (SP142) Assay demonstrated lower PD-L1 membrane staining intensity than the other assays but had a similar staining profile to the Dako PD-L1 IHC 22C3 and Dako PD-L1 IHC 28-8 pharmDx assays.

Conclusions

As compared with a manual scoring approach, cTA™-aided scoring

- **improves precision in the scoring of a challenging biomarker stain such as PD-L1.**
- **demonstrates higher accuracy as determined by the correlation of a reference method (ie, mRNA expression) with IHC scoring.**
- **better captures the full diagnostic spectrum of PD-L1 scoring and better defines positivity for PD-L1 samples that have a low IHC staining intensity.**
- **can be used to understand PD-L1 scoring and staining intensity in multiple PD-L1 IHC assays to develop a robust method for harmonizing assay interpretation.**