**Joseph Krueger**, Chief Scientific Officer, Flagship Biosciences, Inc.

# Using Artificial Intelligence To Determine Patient Response To Immunotherapy

## Key points:

- *While the current diagnostic strategies for PD-L1 checkpoint inhibitors identify some responders well, the majority of patients still need diagnostic approaches that better predict response.*
- *PD-L1 IHC remains a valuable tool for predicting patient response, but can be limited by the ability to correctly use the test to interpret the complex underlying biology and response profile.*
- *New IHC interpretation methods are capturing this biology and defining better patient selection tests, but the interpretations are complicated and hard for pathologists to implement.*

## Measuring PD-L1 Differently for Better Patient Response Profiles

The greatest potential impact for immuno-oncology patient selection is by applying new technology to existing methods of PD-L1 IHC testing. There is a growing understanding that a simplistic approach to measuring programmed death ligand 1 (PD-L1) expression no longer matches the understanding of the complex profile needed to identify a responsive patient. As a result, PD-L1 immunohistochemistry (IHC) interpretations (ie, scoring paradigms) are becoming more complex. For example, the "first-generation" PD-L1 tests only examined the amount of PD-L1 expressed by the tumor to predict response (eg, the tumor proportion score), but newer tests are including the important information about PD-L1 expression on the immune cells that interact with the tumor cells.

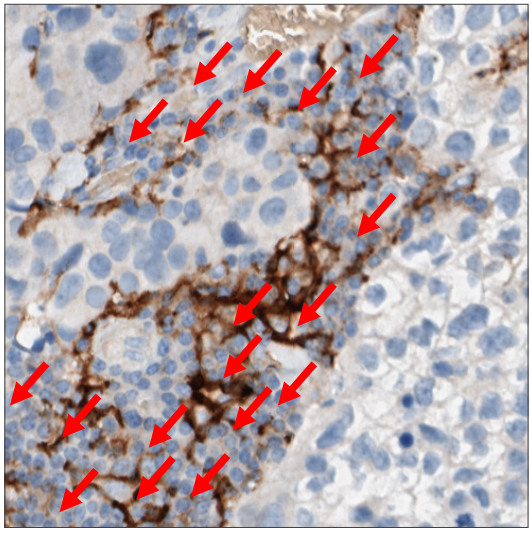
Among the various PD-L1 in vitro diagnostics (IVDs), PD-L1 expression is measured as 3 distinct phenotypes (Figure 1):

1. Presence of PD-L1 in tumor cells
2. Presence of PD-L1 in the immune cells in the tumor microenvironment (TME)
3. Presence of PD-L1 in the tumor-infiltrating immune cells

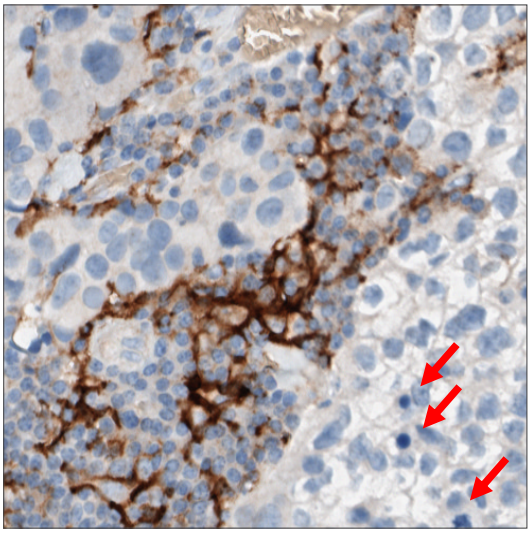
The presence or absence of PD-L1 on each of these cell types has different biological significance. Thus, there is a matrix of biological cell types whose descriptions are dependent on their spatial locations and spatial relationships to the TME and tumor nests. The more recently approved PD-L1 CDx are incorporating these tumor and TME profiles into the scoring and

**Figure 1:** Description of biologically distinct immune cells dependent on their location and PD-L1 expression.

**Immune cells in TME**



**Immune cells in tumor nests**



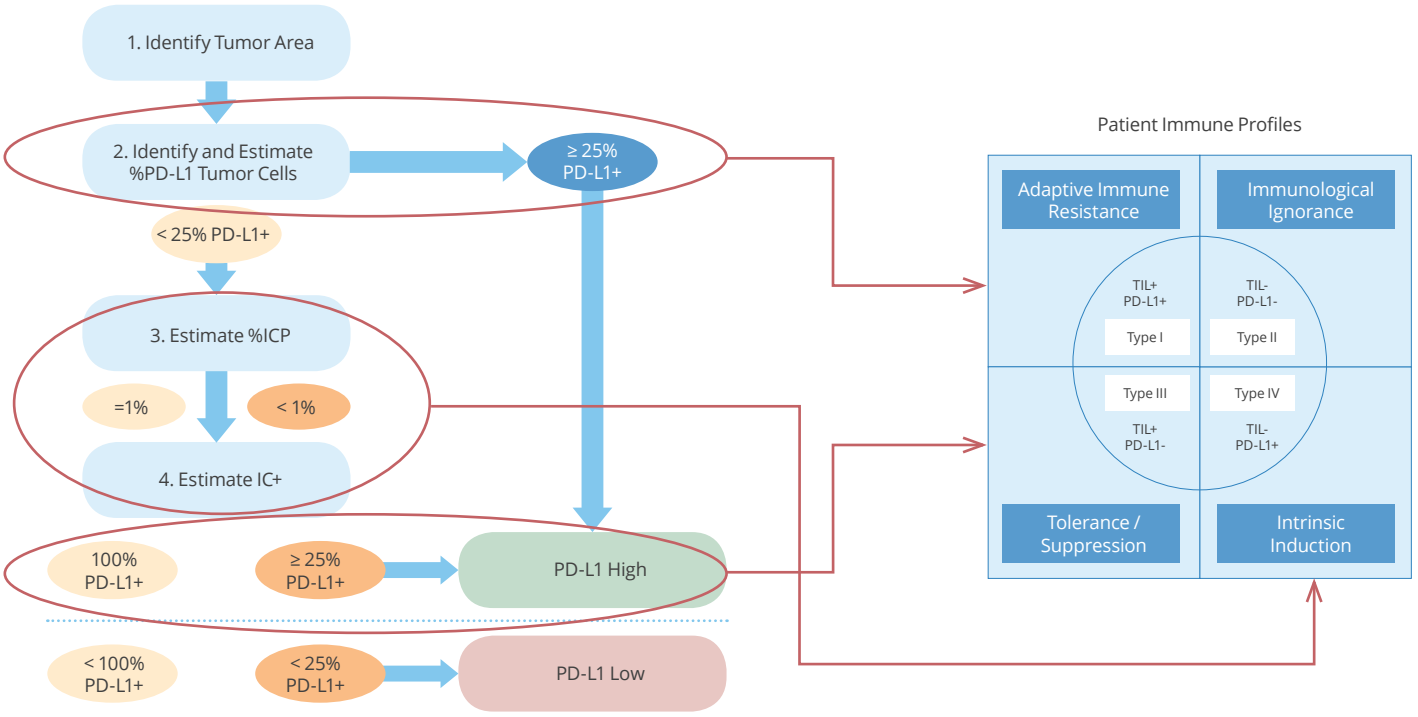
interpretation paradigms. For example, the PD-L1 SP263 companion diagnostic test for TECENTRIQ® in urothelial carcinoma requires measurement of PD-L1 expression in the different tumor and immune cell types through a multifaceted scoring paradigm. To deliver a diagnostic score for the test, the pathologist must evaluate (1) PD-L1 expression in tumor cells, (2) the presence and expression of PD-L1 in immune cells in the TME, and (3) the presence and expression of PD-L1 in immune cells in the tumor nests.

The presence or absence of PD-L1 on tumor and immune cells and the cells' locations and spatial relationships can be used to describe aspects of the 4 different tissue phenotypes of patient response (Figure 2):

- **Adaptive Immune Resistance (type I),** which is characterized by PD-L1 expression on the tumor and immune cells with infiltration into the tumor nests
- **Immunological Ignorance (type II)**, which is characterized by the lack of immune infiltration and PD-L1 expression
- **Intrinsic Induction (type III)**, which is characterized by PD-L1 expression on tumor cells but not immune cells
- **Tolerance/Suppression (type IV),** which is characterized by PD-L1 expression on the immune cells but not the tumor cells, with no infiltration of immune cells into the tumor nests

While each of these descriptions are general and simplified, they are meant to capture a particular facet of PD-L1 biology that determines a patient's response to programmed cell death 1 (PD-1)/PD-L1 inhibitor therapy. Although thse types of scoring paradigms can present challenges for the pathologist to execute, it represents progress toward evaluating the different facets of PD-L1 biology essential for predicting patient response.

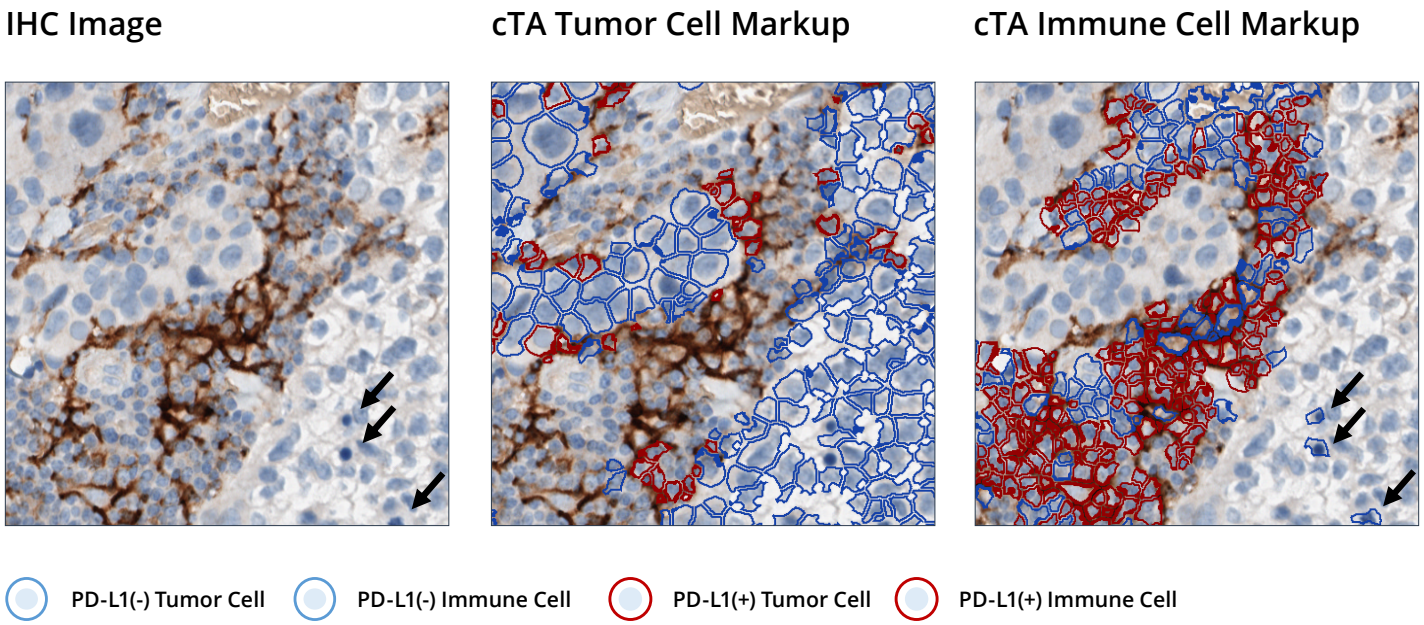
**Figure 2:** Descriptions of 4 different types of biological responses to PD-1/PD-L1 inhibitor therapy based on PD-L1 expression in tumor or immune cells measured by IHC. Abbreviations: IC+, immune cell in the tumor sycytium; ICP PD-L1 postive immune cell in the tumor microenvironment; PD-L1, programmed cell death ligand 1; TIL, tumor-infiltrating lymphocyte.




# Improving PD-L1 IHC Using Artificial Intelligence

Apart from the confusion introduced by multiple tests and uses of these tests in the marketplace (ie, the "PD-L1 harmonization" challenge), the newer, more complex scoring methods can introduce new challenges for the objectivity and reproducibility of the pathologists performing these tests. Computational methods have been seen as a long-standing approach for assisting with the challenges pathologists face with complex tissue scoring algorithms. In fact, more than 15 years ago, pathologists began to apply computer-driven image analysis to whole slide images of tissue to assist with the complexities of IHC testing. A key advantage of this approach is that it leverages the use of standard IHC workflows performed in clinical laboratories. In these cases, the evaluation of a slide under a microscope is replaced by a computer algorithm evaluating a digital image of a slide under the guidance of a pathologist. To date, however, the application of this method as an FDA-cleared/-approved IVD has been limited to reinforcement of pathologist-derived end points, such as for human epidermal growth factor receptor 2 (Her2), estrogen receptor (ER), progesterone receptor (PR), and Ki-67 in breast cancer, by computer-assisted analysis. The widespread adoption of high-throughput digital slide scanners and more recent advances in computational methods has now allowed more sophisticated approaches than ever before imagined to be applied to the clinical setting, which are capable of meeting the current challenges of PD-L1 IHC testing.

**Figure 3:** Example of Flagship's cTA AI applied to the PD-L1 SP263 IHC companion diagnostic.




Recent advances in machine learning algorithms and data science allow for the creation of data-rich tissue profiles that capture the key tissue context information about PD-L1. These advances facilitate novel artificial intelligence (AI) based capabilities for IHC testing. A fundamental application of AI could be to increase the objectivity and reproducibility of scoring in IHC interpretation. Beyond this, existing FDA-approved IHC tests can also be re-evaluated by AI to create more predictable, powerful tests than those currently applied.

A real-world example of this type of AI application is Flagship's cTA platform, our AI-based computational tissue analysis technology, which can reproduce the SP263 scoring method objectively (Figure 3). A similar approach could be applied to the equally challenging PD-L1 IHC 22C3 PharmDx CDx test for gastric (GEJ) cancer. With the more recently developed SP263 and 22C3 scoring methods, the pathologist is required to make simultaneous and discrete measurements of the tumor and immune cell interactions. Flagship's AI technology analytically separates and reports PD-L1 positivity in the tumor, immune cells in the TME, and immune cells within the tumor nest to fulfill this complex scoring method (Figure 3). In this manner, Flagship's AI can be used to increase the accuracy and precision of an existing PD-L1 IHC approach in the clinical setting without changing the IHC wet chemistry or causing significant disruption in the normal procedures performed in pathology laboratories.

## Novel PD-L1 IHC-Based Predictive Tests Using Flagship's AI

While Flagship's AI can improve the current IHC assessments for better general reproducibility, it is also likely to result in more patient response sensitivity, specificity, and accuracy due to the breadth of data and ability to establish correlations in the data for summary decision-making. Potentially the most valuable contribution of AI to evaluating PD-L1 IHC is its ability to detect discrete and non-obvious patterns in PD-L1 IHC that pathologists cannot detect; enabling the creation of

even more useful patient signatures from existing PD-L1 IHC tests. Flagship's cTA can record multiple data points for each cell in the tissue, including staining, morphological, organizational, and spatial aspects, which we describe as Biofeatures. In a resection biopsy sample with over a million cells, this results in potentially billions of Biofeature data points.

This creates the ability to improve patient profiling by allowing the use of more sophisticated diagnostic cut points based on more compelling profile of the tissue biopsy sample than a pathologist can perform. Flagship's cTA AI methods have been developed over the course of 8 years and over 500 client projects. Critically, Flagship has created a technology and clinical testing environment that meets the real-world requirements of drug developers, pathologists, and regulatory agencies. Since Flagship's AI allows thousands of different attributes of tissue to be examined simultaneously, the information must be digested by machine learning to summarize it into a single score or output for decision-making.

Importantly, the use of AI to distill data into a single score does not result in a reduction in any of the complexity in the data as part of the scoring algorithm. The entire dataset is recorded and used by the machine learning processes to deduce a conclusion that best fits the desired purpose. Notably, for developing a novel pathology AI approach, the test or end point does not have to be defined in advance of the initial design. Flagship's cTA process simply requires knowledge of the input (the IHC slide image) and a specific clinical outcome that the diagnostic information will be associated with.

The patient outcome could include, but is not limited to:

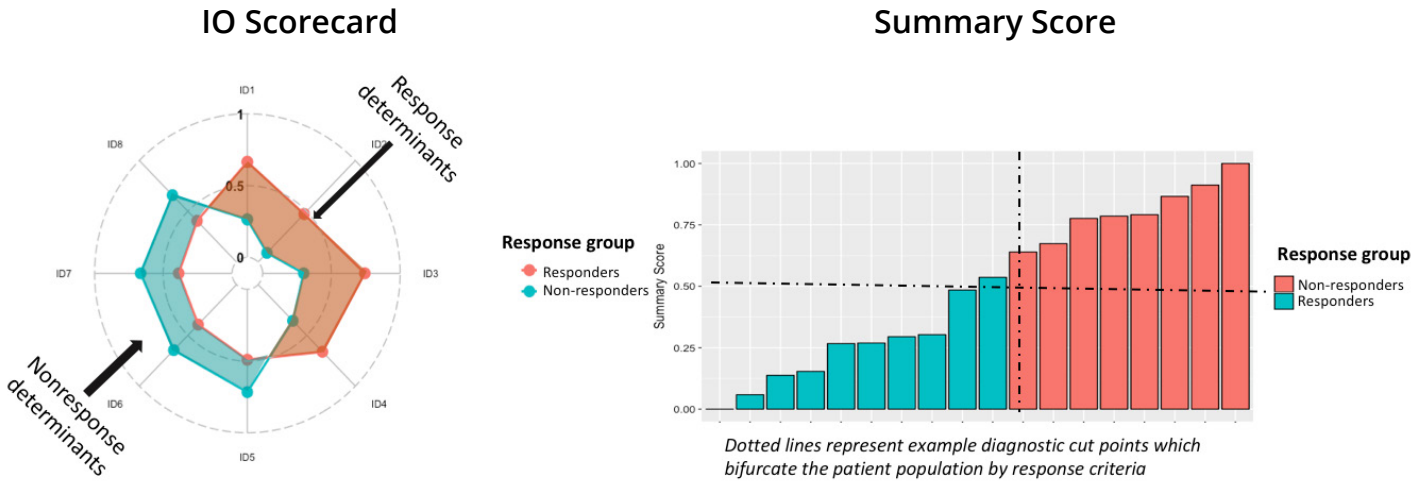
- A cohort of responsive or nonresponsive patients
- Patients with a high versus low tumor mutational burden or microsatellite instability
- Patients with a specific genomic or proteomic profile or drug target
- Predose and postdose biopsies for pharmacodynamic evaluations

**Figure 4:** Inputs for Flagship's cTA AI approach.

| WHAT | HOW | EXAMPLES |
|---|---|---|
| **Morphology** | **FFPE solid tumor biopsy** | biomarker-independent cell features (density of tumor nest, TME, etc) |
| ↓ **Biomarker(s)** | **IHC or IF, monoplex or multiplex** | Flagship IO panel or client-chosen custom assay |
| ↓ **Spatial relationships** | **Flagship Biosciences' cTA** | tumor margin, nearest neighbor, etc. |
| + **Novel and/or study-specific end points** | **Client-Specific** | drug targets/activities (Agonists, immune activators, etc) |

One way of integrating the complex patient profile that can be deduced from a single PD-L1 IHC slide is demonstrated through Flagship's Immuno-oncology (IO) Scorecard™ method. Using Flagship's cTA approach, the IO Scorecard is developed by allowing the AI to determine which Biofeatures are most predictive of a patient outcome in a training set of PD-L1 IHC slides from patients with known responses. The AI determines which Biofeatures are favored in responsive patients and which Biofeatures are favored in nonresponsive patients (Figure 5). The Biofeatures are weighted and organized by the AI to deliver a summary score, which can accurately classify responders and nonresponders through a single cut point (Figure 5, right). This information is also displayed in the IO Scorecard format (Figure 5, left), a graphical representation of the patient's response dataset for PD-L1 IHC testing, demonstrating the severity of the Biofeatures, which create the response profile. This deduction of complex data through AI is now captured in a simple, digestible format for pathologists and oncologists to evaluate as part of the PD-L1 IHC component of patient testing.

**Figure 5:** The basis for the IO Scorecard approach using Flagship's cTA AI.




## Bringing the Value of Flagship's AI to Patients Through Clinical Laboratories

Along with a WSI analysis markup (Figure 3), the pathologist and oncologist can now mutually examine the same dataset in a simple but informative reporting format, such as shown in Figure 5, with sufficient understanding and communication. The summary score reports the clinically validated binary diagnostic for the oncologist's decision-making, the IO Scorecard allows the pathologist to describe the specific nuances of a patient sample important for the oncologist's interpretation, and the WSI analysis markup allows the pathologist to evaluate the performance of the AI and understand the AI's decision process to validate its interpretation as correct. In this way, AI not only translates the deep, intrinsic information in a PD-L1 slide into a digestible summary, but it also aids in the communication between the pathologist and oncologist.

Importantly, this approach does not require that the standard clinical diagnostic laboratory workflow be changed significantly. Each laboratory does not have to buy new equipment, learn new technology, or hire highly trained technical staff to support this enabling technology. Existing PD-L1 IHC tests can be used without changes to the IHC protocol, and any slide scanner that meets the minimum specifications can be used to create a slide image. Slide images are simply uploadeded to Flagship's database though a devoted virtual portal. Flagship then runs the analysis remotely and returns it to the clinical laboratory for viewing. Critically, this is all performed under Flagship's College of American Pathologists/Clinical Laboratory Improvement Amendments laboratory that applies robust quality management system standards. The data analysis process is typically completed within a few hours, and the pathologist simply logs in to view the analysis markup for accuracy, review the test results for acceptance, and sign the report (Figure 6). The data output of the analysis and IO scorecard is captured in a summary report (Figure 6). During execution of the test, a pathologist provides key inputs and quality assurance evaluations to the image analysis process. The resulting image markup and primary results (such as percent positive, etc) are reported and used by the pathologist to evaluate analytical performance of the cTA process and report

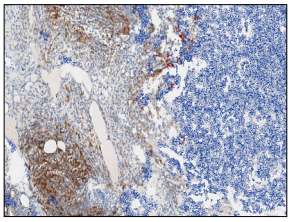
**Figure 6:** Hypothetical IO scorecard report format the SP263 PD-L1 IHC assay. Abbreviations: AI, artificial intelligence; cTA, computational tissue analysis; IC, PD-L1 positive immune cell in tumor syncytium; ICP, immune cell in tumor microenvironment; ID, identification; IHC, immunohistochemistry; IO, immuno-oncology; PD-1, programmed cell death 1; PD-L1, programmed death ligand 1; TIL, tumor-infiltrating lymphocyte; TME, tumor microenvironment.

# IO Scorecard Report for VENTANA PD-L1 (SP263) IHC Assay for Urothelial Carcinoma

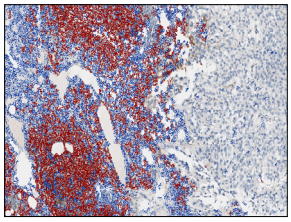
## Patient ID: ABC123

### Pathologist cTA Report

**cTa Tumor Cell Markup**



**cTA Immune Cell Markup**



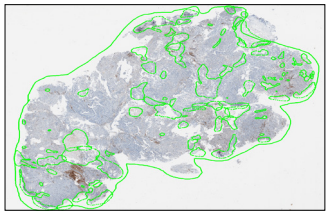
Negative Cell  Positive Cell

| % TUMOR | ICP | IC+ | SCORE |
|---|---|---|---|
| 11.12% | 13.41% | 30.98% | High |

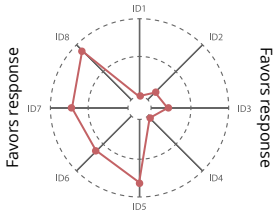
**Pathologist Notes:** Tumor shows moderate PD-L1 staining but with high infiltration in TME and high PD-L1 expression in TILs and moderate infiltrating lymphocytes. Focal PD-L1 expression with high heterogeneity across the tissue.

### IO Scorecard Report

**Whole Tissue Image**



**IO Scorecard**




| | Response Favor | Responder mean | Non-resp. mean | Patient Score |
|---|---|---|---|---|
| ID1 | No | 26 | 67 | 3 |
| ID2 | No | 9 | 43 | 18 |
| ID3 | No | 29 | 74 | 31 |
| ID4 | No | 32 | 66 | 11 |
| ID5 | Yes | 72 | 49 | 86 |
| ID6 | Yes | 65 | 39 | 69 |
| ID7 | Yes | 67 | 39 | 72 |
| ID8 | Yes | 66 | 41 | 88 |

**Patient overall score of 75** (range 0-100) - Indicates patient is highly likely to respond to therapy. Cutpoint for response is patient score >50

**Diagnosis: PD-L1 POSITIVE**
Patient has a significantly increased chance of response to PD-1/PD-L1 therapy

results. While the pathologist has the opportunity to evaluate the whole slide image virtually, example images which represent the tissue are captured in the report alongside the appropriate scoring paradigm (Figure 6 left, SP263 PD-L1 IHC scoring format used as an example).

The results of the AI process, which create the resultant scorecard output and a summary diagnostic score (positive/negative), are also captured in the report (Figure 6, right). A low magnification image of the whole tissue is provided, alongside the IO scorecard plot which shows how the individual patient profile favors or antagonizes response. Importantly, the primary values of the key features in the plot are reported, with reference values for typical responders and non-responders for comparison. This information is provided for the oncologist to evaluate how the patient summary score is characterized. The overall patient score (as a result of the weighted score) is reported (score is normalized to a range of 0-100) and its distance from the diagnostic cutpoint value.

The oncologist can use this information to weigh the medical decision directed by the diagnostic score (for example, the oncologist may choose to give a patient with most of the positive features of a responder, who falls just below the diagnostic cutpoint, anti PD-1/PD-L1 therapy based on their specific clinical situation). If the IO scorecard also is clinically validated to predict survival outcome, a more detailed survival prediction and/or odds ratio can be provided.

The pathologist can then forward the summary report to the oncologist and/or bring a completed report to the tumor board for discussion. Even if the testing approach is novel to both the pathologist and oncologist, this pathologist-enabling report ensures that each professional is allowed to thoroughly examine the data created to execute his or her discretionary judgement for patient care.

In this way, Flagship's AI embodies the ultimate goal of digital and computational pathology: to support pathologists in making better decisions and communicating results effectively to oncologists. This pathologist-enabling technology also solves the human capital problems encountered with the use of PD-L1 IHC testing in the clinic. Namely, only the most skilled personnel (pathologists) are able to interpret these complex IHC scoring algorithms, which comes at a great expense since these personnel need training and time to deliver results for individual tests.

Faced with demands to deliver a widening base of testing methods to support immuno-oncology, with more and more complexity in each application, while in a resource-deficient environment, the only viable solution to meeting the performance and scale demands of medical practitioners is to reduce the time and effort needed by pathologists to make critical decisions. Flagship's cTA combines the beneficial application of AI technology with the practical requirements of the clinical laboratory to deliver on the promise of PD-1/PD-L1 therapy using existing PD-L1 IHC tests and to meet the needs of patient care today.

Flagship BIOSCIENCES

7575 West 103rd Avenue, Suite 102 Westminster, CO 80021

303.325.5894 • info@flagshipbio.com • @FlagshipBio