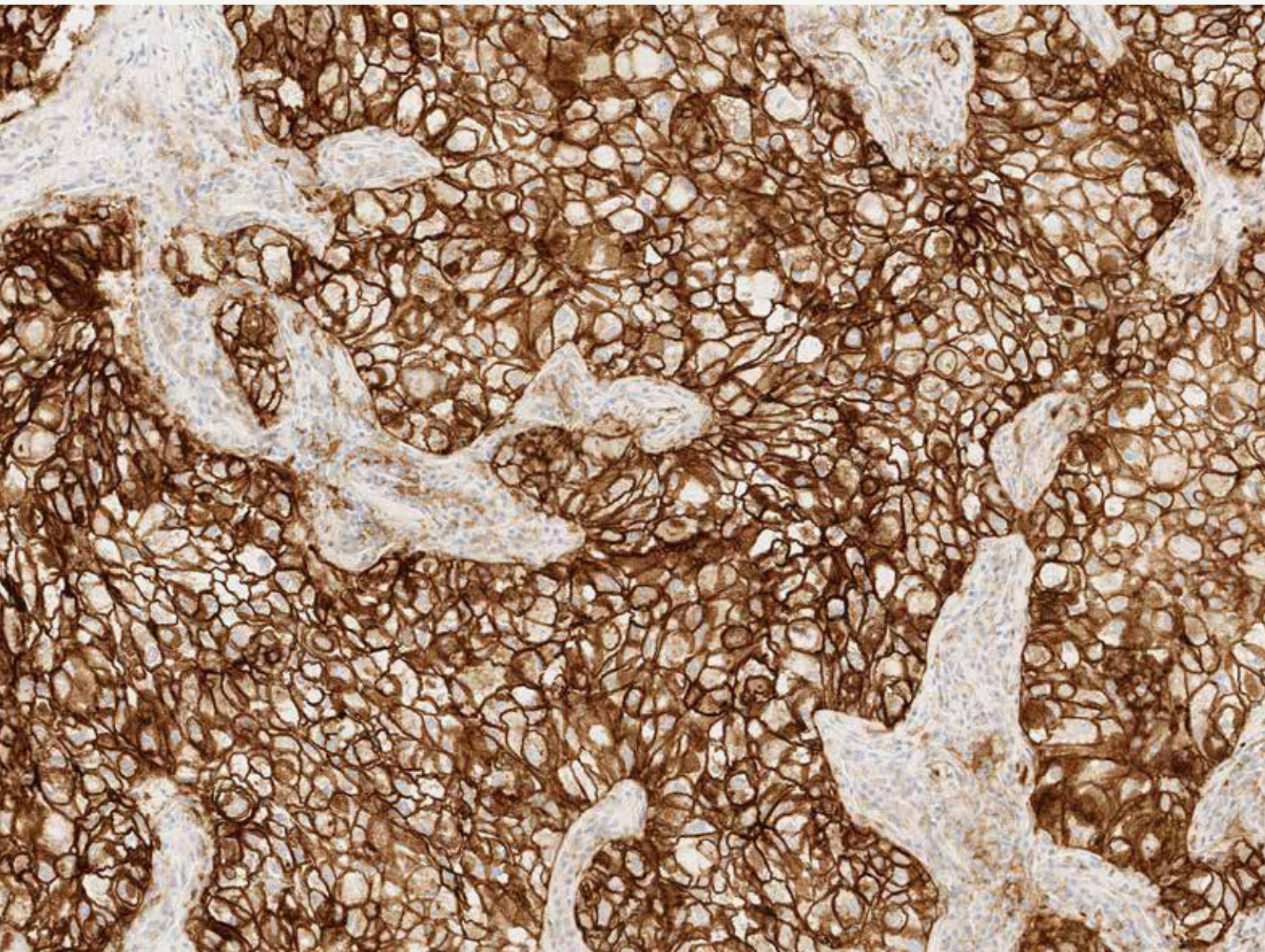
VENTANA®

Roche

WHITE PAPER

Precision and Repeatability of the VENTANA PD-L1 (SP263) Assay Across Seven Different Tumor Types



WHITE PAPER

Precision and Repeatability of the VENTANA PD-L1 (SP263) Assay Across Seven Different Tumor Types

Lupe Manriquez, M.D., Dorothy Hayden, M.D., Fangru Lian, M.D., Chunyan Liu, M.D., Janine Feng, M.D., Alma Nielsen, MS, Pallavi Patil, PhD, Peiyi Wang, MS, Pengfei Gu, MS, Karina Schnittker PhD, Bryan Roland, PhD, Michelle Quiroz, PhD, and Amy Hanlon Newell, PhD

Abstract

Checkpoint inhibitor immunotherapy is a critical type of cancer treatment that relies on the immunohistochemical (IHC) detection of checkpoint biomarkers, such as programmed death ligand-1 (PD-L1), to determine patient eligibility. However, the wide number of available PD-L1 assays complicates diagnosis by forcing pathologists to manage and use a high number of diagnostic assays for each drug therapy and indication. As such, the pathology community seeks to identify a more universal PD-L1 IHC assay that maintains excellent tissue assessment reproducibility across a variety of cancer types, thereby increasing user familiarity and cross-correlation while reducing hospital burden. To this end, Roche's VENTANA PD-L1 (SP263) Assay (SP263 Assay) reproducibility and pathological precision were measured across seven different disease indications: non-small cell lung cancer (NSCLC), urothelial cancer (UC), squamous cell carcinoma of head & neck (SCCHN), melanoma, renal cell carcinoma (RCC), gastric or gastroesophageal junction (GEJ) adenocarcinoma, and hepatocellular carcinoma (HCC). Inter-pathologist overall percent agreement (OPA) for UC and NSCLC was 93% and 93.5%, respectively. OPA was even higher for other cancers, reaching 100% for RCC and HCC. Furthermore, the OPA remained high (>88%) even when day of assessment, instrument, reagent lot, and performing laboratory were varied. Together, the precision studies demonstrate the SP263 Assay is reproducible for PD-L1 detection across a variety of cancers. The adoption of the SP263 Assay may provide consolidation of PD-L1 IHC diagnostics.

Key Findings

The SP263 Assay precision studies indicate:

- Highly concordant tissue PD-L1 evaluation when assessed by different pathologists
- Reproducible PD-L1 tissue assessment regardless of day performed
- Reliable tissue staining and scoring independent of reagent lot or instrumentation
- Consistent assay performance when executed at different laboratories
- Precise PD-L1 expression for diagnostic evaluation across seven different cancer tissues.

Background

While the global research and medical communities continue to make remendous strides in the fight against cancer, it remains the second leading cause of mortality worldwide and a persistent challenge to human health.[1, 2]
With the declining risk of heart disease-associated death, cancer is expected to become the leading cause of mortality in the U.S. in the near future.[3]

Given the wide variety of cancer types and patient variability, drug development efforts have focused on more nuanced and versatile cancer treatments which are able to treat multiple types. In particular, research efforts have uncovered a number of molecular mechanisms that cancerous cells use to avoid immune recognition and cytotoxic damage by T-cells.[4] Often, cancer cells overexpress immunosuppressive molecular signals, such as PD-L1, which help regulate the immune system against over activity.[5] As a result of hijacking these immune checkpoints, T-cells fail to recognize cancer tissue as immunogenic.[5, 6]

This tumor "immunoediting"[6] is both a challenge and an opportunity. Immune checkpoints represent an important target for broadly effective medical intervention and drug development. By disrupting these interactions, immune cells can be re-activated and mount an effective response against tumor cells.[4] There has been significant research focused on identifying immune checkpoints, understanding how to inhibit them, and detecting biomarkers that help determine effective treatment. In 2018, the significance of this work was recognized with a Nobel Prize in Physiology/Medicine.[7]

Immunotherapy, where antibodies disrupt critical mechanisms required for tumor growth and survival, has grown to particular prominence in the past decade. From 2011 to 2020, the FDA approved seven checkpoint inhibitor immunotherapies.[4] Of these, most target the PD-1 (programmed death-1) /PD-L1 checkpoint, indicating the particular importance of this pathway in immune regulation and cancer treatment.[4, 8] PD-L1 is a transmembrane protein that suppresses the adaptive immune system by binding to receptors PD-1 and B7-1 (CD80).[8] Aberrant expression of PD-L1 is found in a multitude of tumor types.5 Importantly, detection of PD-L1 overexpression in tumor samples through IHC correlates with improved clinical outcomes with checkpoint inhibitors targeting PD-1/PD-L1.[5, 9, 10, 11, 12] As a result, PD-L1 represents a key biomarker for determining checkpoint immunotherapy eligibility for patients with seven different tumor types using IHC.5 While re-activating T-cell anti-tumor immunity by interrupting the PD-1/PD-L1 pathway has become part of the standard of care in oncology,[12] it is not applicable to all patients since inhibiting checkpoint functions can cause autoimmune side effects.[5, 13]

There are a number of commercially available IHC assays and tissue evaluation scoring algorithms intended as companion diagnostic tests to identify patients more likely to benefit from PD-L1 checkpoint therapies and reduce risky or ineffective use. However, the heterogeneity of multiple platforms, antibodies, scoring systems, and licensing indications has complicated pathological analysis.[10, 11, 12] The use of many different tests places significant training burdens on pathologists and creates greater uncertainty. Together, these challenges can strain hospital and laboratory resources.[11, 12]

There is an unmet need in the pathology community to use a more universal PD-L1 IHC assay with high reproducibility and excellent sensitivity to tumor and immune cell expression across a variety of cancer types. Finding such an IHC assay would increase user familiarity and enable improved cross-laboratory and crosspathologist comparison, all while reducing hospital burden.

Studies were conducted to assess the performance of Roche's SP263 Assay across a number of different cancer types and report on the possible utility of the SP263 Assay as a universal PD-L1 assay. Furthermore, previous reports have indicated that the antibody clone SP263 used in the SP263 Assay has strong clinical comparability to other IHC tests used in the clinic to assess NSCLC tumors.[11, 12] One such report showed that the SP263 Assay achieved the highest cross-assay concordance rate across all platforms and assays used.[10] Adding to these results, the BenchMark IHC/ISH platforms are found in a high proportion of pathology laboratories, making it easy for many labs to seamlessly adopt the test, without needing to switch platforms.

However, it remains undetermined whether its analytical reliability and reproducibility is more broadly applicable to additional PD-L1-expressing cancer types. The technical repeatability and reproducibility of the SP263 Assay was examined across seven different disease indications: UC, NSCLC, SCCHN, melanoma, RCC, gastric or GEJ adenocarcinoma, and HCC.

Methods

Tumor Specimens	Formalin-fixed and paraffin-embedded (FFPE) specimens from NSCLC, UC, SCCHN, melanoma, RCC, gastric or GEJ adenocarcinoma, and HCC were obtained from commercial sources.
Staining	IHC was performed using the same protocol on each of the seven different tissue types using the BenchMark ULTRA IHC/ISH fully automated instrument, with the staining protocol summarized in **Table 1.**
Evaluation	NSCLC and melanoma were evaluated by pathologists determining the percentage of tumor cell staining, whereas UC and SCCHN were evaluated for both percent tumor cell and percentage of immune cell staining (Fig. 1). Gastric or GEJ adenocarcinoma and HCC were evaluated for combined percentage tumor and immune cell score. RCC was evaluated for percentage immune cell staining over tumor area.

Experimental Design

Inter/intra-pathologist precision	50–114 unique indication tissue cases representing a range of PD-L1 expression levels were independently evaluated by three pathologists. Each pathologist read the same set of samples twice, with a two-week minimum wash out period between reads.
Inter/intra-day precision	14–24 unique indication tissue cases representing a range of PD-L1 expression levels were stained across three non-consecutive days as well as on the same day.
Inter-instrument and inter-lot precision	19–24 unique indication tissue cases representing a range of PD-L1 expression levels were stained with three lots of the SP263 Assay, three lots of OptiView DAB IHC Detection Kit, and across three BenchMark ULTRA IHC/ISH instruments.
Inter-laboratory precision	28–35 unique indication tissue cases representing a range of PD-L1 expression levels were stained in three external laboratories across five non-consecutive days over a 20-day period and then evaluated by two pathologists at each site. The samples were stratified as either positive or negative according to designated scoring cut-offs for each cancer type.[14] For each comparison, the modal result was determined, and the result from each test sample was compared to its respective case-level modal result. From there, the result was deemed either concordant or discordant. Results were aggregated across cases and the overall percent agreement (OPA) was calculated for each study.

Table 1: Staining Procedure for the SP263 Assay Across All Indications

Procedure Parameter	Selection
Deparaffinization	Selected
Baking	Optional 60°C, 12 minutes
Cell Conditioning	Cell Conditioning 1, 64 minutes
Pre-primary Antibody Peroxidase	Selected
Antibody (Primary) or Negative Reagent Control	16 minutes, 36°C
OptiView HQ Linker	8 minutes (default)
OptiView HQ Multimer	8 minutes (default)
Counterstain	Hematoxylin II, 4 minutes
Post Counterstain	Bluing Reagent, 4 minutes

Results

The SP263 Assay is well-characterized for assessing PD-L1 expression in UC and NSCLC. Staining outcomes and OPA in pathological assessment were benchmarked to results from these tissue types. Staining of both lung and urothelial tissue samples clearly permitted the visualization of both PD-L1(+) tumor and immune cells for use in assessing the sample **(Fig. 1)**.

Pathologists Comparisons

To assess the reliability and reproducibility of the SP263 Assay, FFPE tissue samples were stained and assessed independently to understand their inter-pathologist precision of the SP263 Assay. OPA for UC and NSCLC samples were 93% and 93.5%, respectively **(Table 2).** OPA for other indications was even higher (>93.5%), even reaching 100% for RCC and HCC samples.

Sample assessments for all cancer types were also compared to a replicate sample staining and assessment performed at least two weeks later by the same pathologist to determine intra-pathologist precision. OPA for intrapathologist comparisons ranged between 92.4% for UC and 99% for RCC and gastric or GEJ adenocarcinoma samples **(Table 2).**

Day-to-Day Comparisons

To further explore the reproducibility of sample assessment using the SP263 Assay, the same tissue sample was stained and assessed on three non-consecutive days (i.e., inter-day precision assessment) for all cancer tissues **(Fig. 2).** Impressively, the comparison of three separate tissue assessments resulted in an inter-day OPA ranging from 93.8% to 100%.

In addition, sample staining and assessment within the same day (i.e., intra-day precision assessment) also led to high concordance **(Table 2)**. Similar to inter-day OPA, most cancer types intra-day comparisons resulted in 94.2% to 100% OPA.

Figure 1: Tumor Cell and Immune Cell Staining Detected by the SP263 Assay.

Representative cases from UC (left) and NSCLC (right) demonstrated both tumor cell (black arrow) and immune cell (red arrow) staining by the SP263 Assay.

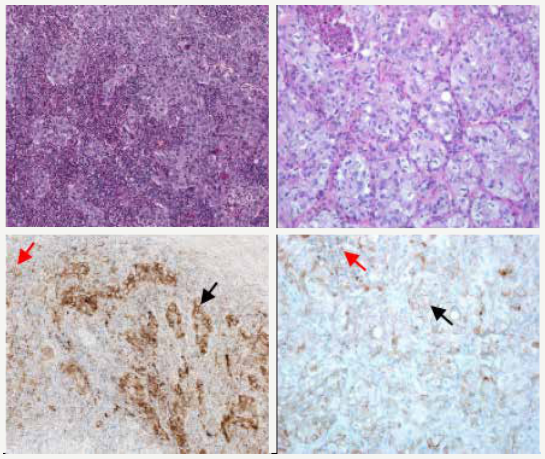
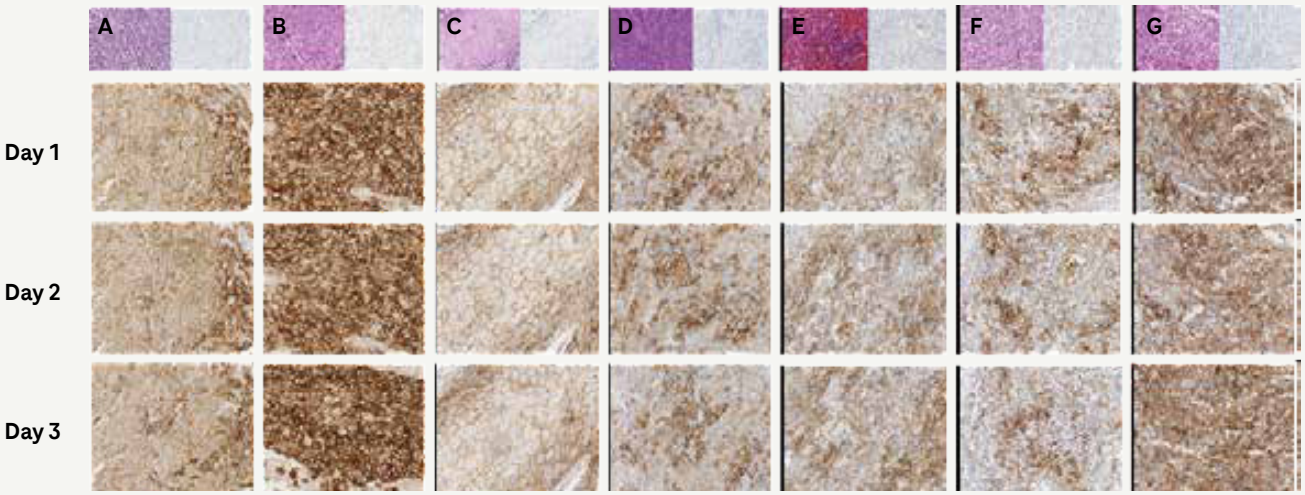


Figure 2: Inter-Day Reproducibility Staining.

Tissues represented: A) UC, B) NSCLC, C) SCCHN, D) melanoma, E) RCC, F) gastric or GEJ adenocarcinoma, G) HCC. Small insets display H&E and negative reagent control staining for each tissue. Larger IHC images display representative staining with the SP263 Assay across three different days on the same tissue sample.



Instrument, Reagent Lot, and Laboratory Comparisons

Reproducibility of any diagnostic assay is also bound by the variations possible between instruments, antibody reagent production lots, and discrete laboratories. To test the possibility that these factors may impact assay reliability and repeatability, the SP263 Assay testing was performed using multiple instruments and lots, as well as in different laboratories.

Three separate BenchMark Ultra IHC/ISH instruments and three different lots of the SP263 Assay and OptiView DAB IHC Detection Kits were used to assess cancer tissues and across instrument and lots. OPAs were 100% for four cancer types (NSCLC, RCC, gastric/GEJ, and HCC), 99.2% for SCCHN, 99.0% for UC, and 98.4% for melanoma **(Table 2).**

Furthermore, inter-laboratory comparisons were also performed to account for differences that may occur between pathologists working in different lab environments. In this case, tissue samples were stained and reviewed in three separate laboratories on five non-consecutive days over a 20-day period. At that point, each cancer (except melanoma and HCC) were evaluated by two different pathologists from each site. While the lowest OPA's was slightly lower here compared with other OPA groupings, agreement was still high, ranging from 88.3% for NSCLC to 98.6% for SCCHN.

Table 2: Consistency and Repeatability of the SP263 Assay in Seven Indications

	Overall Percent Agreement (OPA)						
Disease Indications	Interpathologist (%)	Intrapathologist (%)	Inter-day (%)	Intra-day (%)	Interinstrument (3 ULTRAs) (%)	Inter-lot (3 Ab lots & 3 detection lots) (%)	Inter-lab (%)
UC	93.0	92.4	100.0	99.2	99.0	99.0	92.6
NSCLC	93.5	96.2	100.0	100.0	100.0	100.0	88.3
SCCHN	98.0	98.7	93.8	94.2	99.2	99.2	98.6
Melanoma	96.0	96.0	100.0	100.0	98.4	98.4	N/A
RCC	100.0	99.0	100.0	100.0	100.0	100.0	93.2
Gastric or GEJ adenocarcinoma	99.3	99.0	100.0	100.0	100.0	100.0	96.3
HCC	100.0	98.7	100.0	100.0	100.0	100.0	N/A

Conclusions

In order to assess the reliability and reproducibility of the SP263 Assay, precision studies were performed using seven different cancer tissues: NSCLC, UC, SCCHN, Melanoma, RCC, Gastric/GEJ, and HCC. Through these studies, inter-/intra-pathologist and inter-/intra-day consistency was measured through percent concordance. In addition, variability due to differing instruments, assay reagent lots, and laboratories was also investigated. Importantly, all studies showed an OPA of 88% or greater for PD-L1 concordance to the designated cut-offs for each indication **(Table 2).** Given that an OPA of greater than 85% is considered acceptable concordance for IHC diagnostic assays, the OPA values from this study point to excellent assessment agreement for all cancer tissues investigated with the SP263 Assay.[15, 16, 17] Furthermore, these OPA's compare favorably to OPA's reported previously using a different PD-L1 IHC diagnostic assay on NSCLC where, 1% cut-off inter-pathologist and intrapathologist OPA were 84.2% and 89.7%, respectively.[18]

Collectively, the precision studies point to high consistency and repeatability in PD-L1 detection by the SP263 Assay across a variety of cancer types. These studies agree with previous findings that demonstrated the SP263 Assay's particular ability to generate consistent results across centers and platforms as well its potential to assist PD-L1 laboratory developed test (LDT) harmonization.10 Dependable and consistent IHC assessments remain critical as they can guide patient treatments and impact outcomes. Sample staining and assessment by independent pathologists, resulted in excellent agreement (inter-pathologist, **Table 2)**. These evaluations were also internally consistent when reproduced by the same pathologist following a two-week gap (intra-pathologist, **Table 2**). Pathologist OPA was independent of run-to-run staining variability and scoring were performed. There was often 100% agreement when replicates were performed on the same day (intra-day) or different days (inter-day, **Fig. 2 & Table 2**). Furthermore, differing instruments and product lots had no negative effect on OPA, demonstrating no detectable batch-to-batch or instrument-associated variability (inter-instrument/lot).

Since assay outcome confusion is most likely to occur when pathologists adopt and train on an assay independentlyin separate geographic locations and lab environments, cross-laboratory comparisons were also performed to further test the potential of interpretation variability.[11] Maintaining consistent assay performance and results while performing the same test at different sites represents a distinct challenge for diagnostic testing, especially for IHC. Despite this, OPAs for each of the five cancer types investigated in multiple laboratories remained high (88-98.6%). The lowest OPA of this particular study was for NSCLC, which is a CE-IVD marked indication and the most highly analyzed in clinical literature for the SP263 Assay.[19] "With earlier reports exhibiting the SP263 Assay's consistency[10,11,12] and high diagnostic sensitivity compared with other FDA-approved assays (including 22C3 pharmDx and 28-8 pharmDx)19 for NSCLC, higher OPAs for other cancers points to the assay's potential to provide meaningful PD-L1 detection for additional clinical indications."[19]

The reproducibility of the SP263 Assay with differing pathologists, timelines, materials, and laboratories points to the highly trainable nature of the assay as well as its consistency across tumor types. Since the FDA has recently approved PD-1/PD-L1 checkpoint immunotherapies in NSCLC, melanoma, SCCHN, UC, gastric, cervical, esophageal, and triple-negative breast cancer,4 further studies are ongoing to assess additional clinical utility for the SP263 Assay in a wider variety of tissue and cancer types.

As further PD-1/PD-L1 targeted therapies are approved for an even broader range of tumor types, the SP263 Assay should function as a robust and reproducible tool for assessing and quantifying PD-L1 expression. The data discussed in this report illustrates that laboratory adoption of the SP263 Assay permits reliable data collection in several different tissue types, potentially eliminating the need to use multiple assays. The ability to consolidate testing into one assay can help reduce hospital costs and increase pathologist familiarity for the benefit of patients and healthcare professionals alike.

1 Murphy SL, Xu J, Kochanek KD, Arias E. Mortality in the United States, 2017. NCHS Data Brief 2018;1-8.
2 WHO Cancer Fact Sheet, https://www.who.int/news-room/fact-sheets/detail/cancer, (accessed April 30, 2020)
3 Weir HK, Anderson RN, Coleman King SM, Soman A, et al. Heart disease and cancer deaths - trends and projections in the United States, 1969–2020. Prev Chronic Dis 2016;13: E157.
4 Vaddepally RK, Kharel P, Pandey R, Garje R, Chandra AB. Review of indications of FDA-approved immune checkpoint inhibitors per NCCN guidelines with the level of evidence. Cancers 2020;12:738.
5 Patel SP, Kurzrock R. PD-L1 Expression as a predictive biomarker in cancer immunotherapy. Mol Cancer Ther 2015;14(4):847-56.
6 Dunn GP, Old LJ, Schreiber RD. The three Es of cancer immunoediting. Annu Rev Immunol 2004;22:329–60.
7 The Nobel Prize in Physiology or Medicine 2018. Nobelprize.org: The Nobel Assembly at Karolinska Institutet 2018;1-5.
8 Keir ME, Butte MJ, Freeman GJ, Sharpe AH. PD-1 and its ligands in tolerance and immunity. Annu Rev Immunol 2008;26:677-704.
9 Hsu J-M, Li C-W, Lai Y-J, and Hung M-C. Posttranslational modifications of PD-L1 and their applications in cancer therapy. Cancer Res 2018;78(22):6349-53.
10 Adam J, Stang NL, Rouquette I, Cazes A, et al. Multicenter harmonization study for PD-L1 IHC testing in non-small cell lung cancer. Ann Oncol 2018;29:953-8.
11 Udall M, Rizzo M, Kenny J, Doherty J, et al. PD-L1 diagnostic tests: a systemic literature review of scoring algorithms and test-validations metrics. Diagn Pathol 2018;13:12.
12 Tsao MS, Kerr KM, Kockx M, Beasley M-B, et al. PD-L1 Immunohistochemistry comparability study in real-life clinical samples: results of Blueprint phase 2 project. J Thoracic Oncol 2018;13(9):1302-11.
13 Hamid O, Robert C, Daud A, Hodi FS, Hwu W-J, Kefford R, et al. Safety and tumor responses with lambrolizumab (anti–PD-1) in melanoma. N Engl J Med 2013;369:134–44.
14 For NSCLC and UC, CE Marked or approved cut-offs, respectively, for VENTANA PD-L1 (SP263) Assay were utilized. For the remainder of indications, cut-offs are under investigational studies and cannot be published.
15 Fitzgibbons PL, Bradley LA, Fatheree LA, Alsabeh R, et al. Principles of analytic validation of immunohistochemical assays. Arch Pathol Lab Med 2014;138:1432-43.
16 Lin F, Chen Z. Standardization of Diagnostic Immunohistochemistry. Arch Pathol Lab Med 2014;138:1564-77.
17 McHugh ML. Interrater reliability: the kappa statistic. Biochemia Medica 2012;22(3):276-82.
18 Cooper WA, Russell PA, Cherian M, Duhig EE, et al. Intra- and interobserver reproducibility assessment of PD-L1 biomarker in non-small cell lung cancer. Clinical Cancer Res 2017;23(16):4569-77.
19 Torlakovic, E, et al., "Interchangeability" of PD-L1 immunohistochemistry assays: a meta-analysis of diagnostic accuracy. Modern Pathology 2020;;33(1):4-17.

Roche Diagnostics
9115 Hague Road
Indianapolis, IN 46256

diagnostics.roche.com