Analysis of PD-L1 IHC tests using NIST SRM 1934–traceable reference materials: A new paradigm for development of predictive IHC biomarkers

Short Title:
New paradigm for IHC biomarkers

Abstract:
BACKGROUND. The challenges in accurate patient stratification for immune checkpoint inhibitors have been compounded by the fact that the FDA–cleared PD-L1 IHC tests are analytic ‘black boxes’. Relatively basic analytic parameters such as lower limit of detection and analytic dynamic range are unknown to both developers of assays as well as end users. The recent development of standardized PD-L1 immunohistochemistry (IHC) reference materials enables quantitative test characterizations that were not previously possible.

METHODS. We surveyed 41 PD-L1 testing laboratories in North America and Europe, quantitatively defining each of the PD-L1 tests’ analytic performance in terms of lower limit of detection and dynamic range. All four commercial PD-L1 kits were assessed by multiple laboratories. A variety of laboratory–developed tests (LDTs) were also assessed. The reference materials incorporated defined concentrations of PD-L1 peptide (intracellular domain) or recombinant extracellular domain protein, traceable to NIST Standard Reference Material 1934. Each laboratory received a slide with 10 separate PD-L1 calibrator concentrations: 2,200 – 600,000 molecules of PD-L1 extracellular domain or 34,000 – 2,200,000 molecules of PD-L1 intracellular domain. The calibrator concentrations ranged from those that are below the lower limit of detection to others that yield maximal staining.

RESULTS. The data obtained with the four PD-L1 kits (VENTANA PD-L1 (SP263) Assay, VENTANA PD-L1 (SP142) Assay, DAKO PD-L1 IHC 28–8 pharmDx and DAKO PD-L1 IHC 22C3 pharmDx assays) revealed that the lower limits of detection (PD-L1 molecules per cell equivalent) are approximately: 50,000 – 180,000 (SP263), 800,000 – 1,200,000 (SP142), 220,000 – 360,000 (28–8), and 200,000 – 400,000 (22C3). The dynamic ranges for all of these tests are generally narrow, spanning less than a log concentration of PD-L1. The SP263 and SP142 assays showed no overlap of their analytic response curves. This means that a maximal stain intensity with SP263 kit can be associated with zero staining with the SP142 kit. Consequently, it is not possible to compensate for the variability in analytic sensitivity between these two tests by adjusting the percent positive cell cutoff. The 28–8 was more sensitive than, but statistically indistinguishable from the 22C3 assay. Laboratory–developed tests (LDTs) using these and other primary antibodies have their own unique analytic performance characteristics.

CONCLUSIONS. The PD-L1 reference materials enable precise definitions of analytic test performance and linking them with clinical management thresholds. Therefore, this tool finds its most important implementation at the stage of development of new IHC predictive biomarkers in clinical trials as well as at the stage of methodology transfer to clinical IHC laboratories. Furthermore, our results also help define more precisely the possibility for assay interchangeability and to what degree the assays may be harmonized.