Gastric Cancer (2022) 25:1131–1132
https://doi.org/10.1007/s10120-022-01336-3

LETTER TO THE EDITOR

PD-L1 IHC testing: issues with interchangeability evidence based on laboratory-developed assays of unknown analytical performance

Emina Emilia Torlakovic[1]

Received: 31 July 2022 / Accepted: 18 August 2022 / Published online: 26 September 2022

Dear Editor,

I read with interest your recently published article in Gastric Cancer entitled "Choice of PD-L1 immunohistochemistry assay influences clinical eligibility for gastric cancer immunotherapy" [1]. This, like many other similar papers, is the results of what authors refer to as "unmet clinical and logistical need for harmonization" for PD-L1 testing. This study explores so-called "interchangeability" between the Dako 22C3, Dako 28-8 and Ventana SP142 assays as predictive biomarkers in gastric cancer.

Consequences of this and similar studies are such that they may significantly impact clinical practice as well as be used as evidence that regulatory bodies will use to approve or deny approval for certain biomarkers for certain purposes (e.g., FDA, Health Canada, or other). Therefore, it is critical that this type of evidence is transparent for what it is and how it applies to clinical practice and other published studies on interchangeability of PD-L1 biomarker assays.

I have reviewed this paper carefully and found that it unfortunately has several issues in the study design and interpretation of the study results, as follows:

- All published clinical trials used regulatory agency-approved PD-L1 biomarker assays, specifically DAKO PD-L1 IHC 22C3 pharmDx, DAKO PD-L1 IHC 28-8, and VENTANA PD-L1 (SP142) Assay, or VENTANA PD-L1 (SP263) Assay. These IHC assays have established analytical sensitivity, which is generally stable, and the results are generally reproducible, and the scoring schemes are directly linked to the respective FDA-approved assay and clinical trial(s), which are using specific immune checkpoint inhibitor and are also designated as companion diagnostics (CDx). Primary antibody used by these CDx is not an "assay", but it is just a primary antibody. Since the authors have used these primary antibodies with different, laboratory-developed tests (LDTs), which were not IHC, but multiplex assay mIHC/IF with Opal Multiplex fIHC kit, the results of this study are not applicable to clinical practice. Although the authors used the same primary Ab clones, it is not clear why they assume that their LDTs will have the same analytical performance as the original FDA-approved CDx. There is also no attempt by the authors to compare their 22C3 LDT with DAKO PD-L1 IHC 22C3 pharmDx, their 28-8 LDT with DAKO PD-L1 28-8 pharmDx, or their SP142 LDT with VENTANA PD-L1 (SP142) Assay, first separately and also when multiplexed. It is a common knowledge and historical experience of many proficiency testing programs that the conditions of IHC protocols beyond the primary Ab are critically important for IHC results. Some LDTs may be good, others may be insufficient for the purpose for which they are developed. Therefore, fit-for-purpose diagnostic and technical validation of LDTs is essential before they are to be considered to be similar to clinically validated CDx. The authors cited their previous publication where they stated that they have compared what they refer to as "conventional IHC" to their multiplex LDT with three different clones, which did not include 28-8 clone [2]. Not only that the 28-8 clone was not included, but also there is no clear statement whether "conventional IHC" assays were performed according to CDx specifications, or the pre-diluted antibodies were used in their own LDT for each IHC assay. Even if they have used CDx assays as per specification protocols, the purpose of the study was not to validate the multiplex LDT against CDx assays, but their results were compared and concordance rates

This comment refers to the article available online at https://doi.org/10.1007/s10120-022-01301-0.

✉ Emina Emilia Torlakovic
emina.torlakovic@usask.ca

[1] Department of Pathology and Laboratory Medicine, Royal University Hospital, University of Saskatchewan, Saskatoon, SK, Canada

show, where only 2 out of 9 comparisons for concordance were above 90%, with lowest concordance being 67%. With these results, we can be reasonably assured that the mIHC/IF was not validated for diagnostic equivalence against respective CDx assays.

- Sample degradation is an important consideration in studies of PD-L1 expression as it has been shown that for at least some clones (e.g., 22C3) paraffin blocks older that 3 years may already show degradation of the PD-L1 epitope recognized by this clone [3, 4]. This may or may not be the same for different PD-L1 clones and it may cause background noise where two different assays using two different clones are compared. However, even the "new cohort" is very old with the newest samples being from 2013. Therefore, all samples are much older than 3 years. It is uncertain if this is possibly causing low(er) sensitivity with 22C3 clone.
- The authors use the same scoring schemes for the readout of their LDT as they are used by CDXs for each primary Ab. It is a serious mistake to apply the same scoring system to assays of unknown and presumed different analytical sensitivity. If authors developed LDTs with higher analytical sensitivity, using the same scoring would lead to higher number of positive cases and, the other way around, with an LDT of lower analytical sensitivity than that of the relevant CDX, the number of positive cases would be lower. Since the analytical sensitivity of their LDT is unknown, the results could go both ways. The need to align analytical sensitivity with a scoring scheme was recently emphasized for ROS1 IHC assays [5].
- The authors used correlation (Spearman's correlation) for the analysis of the results. This is a common, but serious mistake. Correlation should never be used to compare two methods that are "measuring" (or assessing) the same variable/parameter. While it makes sense to assess the correlation between the height and weight of a person, it does not make sense to assess the correlation between two methods that measure the height of a person. There will be a correlation. What really matters in comparing different predictive qualitative assays (such are PD-L1 IHC assays) is their accuracy, which is assessed by diagnostic sensitivity and specificity, calculated from the number of true-positive, false-positive, true-negative, and false-negative results where a candidate assay is compared to designated reference method (or comparator assay) [6]. This is elaborated in detail in meta-analysis of PD-L1 interchangeability studies (7). Similarly, mean scores are completely irrelevant because they do not tell us anything about diagnostic errors (e.g., false-negative or false-positive results).

In summary, the use of multiplex immunohistochemistry/immunofluorescence (mIHC/IF) LDT for the simultaneous assessment of PD-L1 expression is a very interesting and promising methodology. However, based on the design of the study, methods applied, and terminological confusion, the results of this study are not applicable to clinical practice. The conclusions about the performance of FDA-approved assays and their interchangeability in gastric cancer cannot be presumed based on the results of this LDT; the authors provided no evidence that this multiplex LDT has the same test performance characteristic and is equivalent to corresponding FDA-approved assays based solely on the fact that they use the same primary antibodies.

Editor-in-Chief
Yasuhiro Kodera, Nagoya.

References

1. Yeong J, Lum HYJ, Teo CB, et al. Choice of PD-L1 immunohistochemistry assay influences clinical eligibility for gastric cancer immunotherapy. Gastric Cancer. 2022;25:741–50.
2. Yeong J, Tan T, Chow ZL, Cheng Q, Lee B, Seet A, Lim JX, Lim JCT, Ong CCH, Thike AA, Saraf S, Tan BYC, Poh YC, Yee S, Liu J, Lim E, Iqbal J, Dent R, Tan PH. Multiplex immunohistochemistry/immunofluorescence (mIHC/IF) for PD-L1 testing in triple-negative breast cancer: a translational assay compared with conventional IHC. J Clin Pathol. 2020;73:557–62.
3. Takeda M, Kasai T, Naito M, Tamiya A, Taniguchi Y, Saijo N, Naoki Y, Okishio K, Shimizu S, Kojima K, Nagoya A, Sakamoto T, Utsumi T, Yoon HE, Matsumura A, Atagi S. programmed death-ligand 1 expression with clone 22c3 in non-small cell lung cancer: a single institution experience. Clin Med Insights Oncol. 2019;9(13):1179554918821314.
4. Gagné A, Wang E, Bastien N, Orain M, Desmeules P, Pagé S, Trahan S, Couture C, Joubert D, Joubert P. Impact of specimen characteristics on pd-l1 testing in non-small cell lung cancer: validation of the IASLC PD-L1 testing Recommendations. J Thorac Oncol. 2019;12:2062–70.
5. Cheung CC, Smith AC, Albadine R, Bigras G, Bojarski A, Couture C, Cutz JC, Huang WY, Ionescu D, Itani D, Izevbaye I, Karsan A, Kelly MM, Knoll J, Kwan K, Nasr MR, Qing G, Rashid-Kolvear F, Sekhon HS, Spatz A, Stockley T, Tran-Thanh D, Tucker T, Waghray R, Wang H, Xu Z, Yatabe Y, Torlakovic EE, Tsao MS. Canadian ROS proto-oncogene 1 study (CROS) for multi-institutional implementation of ROS1 testing in non-small cell lung cancer. Lung Cancer. 2021;160:127–35.
6. Garrett PE, Lasky FD, Meier KL, et al. 2008 User protocol for evaluation of qualitative test performance: approved guideline. Wayne, Pa.: Clinical and Laboratory Standards Institute.
7. Torlakovic E, Lim HJ, Adam J, Barnes P, Bigras G, Chan AWH, Cheung CC, Chung JH, Couture C, Fiset PO, Fujimoto D, Han G, Hirsch FR, Ilie M, Ionescu D, Li C, Munari E, Okuda K, Ratcliffe MJ, Rimm DL, Ross C, Røge R, Scheel AH, Soo RA, Swanson PE, Tretiakova M, To KF, Vainer GW, Wang H, Xu Z, Zielinski D, Tsao MS. "Interchangeability" of PD-L1 immunohistochemistry assays: a meta-analysis of diagnostic accuracy. Mod Pathol. 2020;33(1):4–17.