

# Reproducibility of PD-L1 assessment in non-small cell lung cancer – know your limits but never stop trying to exceed them

Marius Ilie<sup>1,2,3</sup>, Paul Hofman<sup>1,2,3</sup>

<sup>1</sup>Université Côte d'Azur, University Hospital Federation OncoAge, Laboratory of Clinical and Experimental Pathology, Pasteur Hospital, Nice, France; <sup>2</sup>Université Côte d'Azur, Institute for Research on Cancer and Ageing, Nice (IRCAN), Inserm U1081 and UMR CNRS 7284, Nice, France;

<sup>3</sup>Université Côte d'Azur, University Hospital Federation OncoAge, Hospital-Related Biobank (BB-0033-00025), Pasteur Hospital, Nice, France

*Correspondence to:* Paul Hofman. Laboratory of Clinical and Experimental Pathology, Pasteur Hospital, 30 Avenue de la voie Romaine, 06002 Nice, France. Email: hofman.p@chu-nice.fr.

*Provenance:* This is an invited Editorial commissioned by Section Editor Dr. Hongbing Liu, MD, PhD (Department of Respiratory Medicine, Jinling Hospital, Nanjing University School of Medicine, Nanjing, China).

*Comment on:* Cooper WA, Russell PA, Cherian M, *et al.* Intra- and Interobserver Reproducibility Assessment of PD-L1 Biomarker in Non-Small Cell Lung Cancer. *Clin Cancer Res* 2017;23:4569-77.

Submitted Oct 11, 2017. Accepted for publication Oct 17, 2017.

doi: 10.21037/tlcr.2017.10.13

**View this article at:** <http://dx.doi.org/10.21037/tlcr.2017.10.13>

Immunotherapy targeting the programmed death 1 (PD-1)/programmed death-ligand 1 (PD-L1) pathway has demonstrated strong and durable anti-tumoral immune responses with significantly improved overall survival in patients with locally advanced or metastatic non-small cell lung cancer (NSCLC) (1).

PD-1 or CD279 is a type 1 transmembrane protein expressed on the surface of activated immune cells, including T cells, B cells, monocytes, natural killer cells, regulatory T cells, dendritic cells, and macrophages (2). The binding of PD-1 to its major ligand PD-L1, B7-H1 or CD274 decreases the ability of activated T cells to produce an effective immune response and prevents the host immune system from destroying tumor cells. PD-L1 is widely expressed in hematopoietic cells, including macrophages, dendritic cells, mast cells, T cells, and B cells, as well as in non-hematopoietic cells, including epithelial, endothelial, and tumor cells (2).

Monoclonal antibodies blocking the interaction between PD-1 and PD-L1 have confirmed the clinical activity in various solid tumors, including NSCLC (3,4). Three drugs were approved by the FDA in a second-line setting for NSCLC patients. The anti-PD-1 inhibitor pembrolizumab was approved for use in NSCLC patients with  $\geq 1\%$  PD-L1 expression on tumor cells, whereas nivolumab, an anti-PD-1 inhibitor, and atezolizumab, an anti-PD-L1

inhibitor, demonstrated clinical benefit over chemotherapy regardless of the PD-L1 expression (3,4). Still, treatment with atezolizumab showed significant improvement of overall survival in patients showing  $\geq 50\%$  PD-L1 expression in tumor cells or  $\geq 10\%$  in immune cells (5). In contrast, pembrolizumab is the only agent recently holding a first-line indication for single-agent treatment in NSCLC patients with  $\geq 50\%$  PD-L1 expression on tumor cells (6,7). Based on the results of clinical trials, these drugs were also approved with their own companion or complementary PD-L1 immunohistochemistry (IHC) assay. Currently, there is one PD-L1 FDA-approved companion diagnostic, the 22C3 pharmDx assay (Agilent Technologies, Santa Clara, CA, USA) for use with pembrolizumab (for NSCLC), and two PD-L1 complementary diagnostics, namely, the SP142 assay (Ventana, Tucson, AZ, USA) for use with atezolizumab (for NSCLC and bladder indications) and the 28-8 assay (Agilent) for use with nivolumab (for NSCLC and melanoma indications) (8). In addition, the SP263 assay (Ventana) is CE marked, but not yet FDA approved, for selecting NSCLC patients for treatment with pembrolizumab or nivolumab (8,9).

Substantial effort to harmonize the different PD-L1 assays has recently been made and now demonstrates relative analytical comparability. Overall, several comparative studies performed independently by different

investigators demonstrated that the tumor proportion score (TPS) of PD-L1 was comparable when the 22C3, 28-8, and SP263 assays were used, whereas the SP142 assay exhibited fewer stained tumor cells (9-11).

Currently, there are two cut points generally accepted for all the PD-L1 IHC assays in NSCLC (i.e., 1% of tumor cells in a second-setting, and 50% of tumor cells in a first-line setting) (6,7).

Although tremendous efforts have been made to adjust several methodological and biological aspects that may influence the outcome of the PD-L1 assay (i.e., tumor heterogeneity, pre-analytical parameters, assay harmonization, validation; IASLC Atlas of PD-L1 Testing in Lung Cancer), the post-analytical phase, and in particular the interpretation and scoring of PD-L1 expression, may hamper the robustness of a PD-L1 assay for surgical pathologists in their daily practice (12,13).

The IHC technique can hold the intrinsic disadvantage of a quite subjective interpretation, in particular when PD-L1 is expressed on both tumor and tumor-infiltrating immune cells. There is limited data assessing the reproducibility of interpretation and scoring of PD-L1 expression in NSCLC tissue samples and most studies only involve a few pathologists or small numbers of tumors, making it easy to achieve concordance (12-16). Whereas some studies assessed PD-L1 interpretation reproducibility on resection specimens, only a few studies used smaller sized specimens such as tissue micro-arrays, but none on bronchial or transthoracic biopsies (12-16).

Cooper *et al.*, assessed the intra- and inter-observer reproducibility as well as the impact of training of the pathologists in scoring of PD-L1 in NSCLC using the FDA-approved companion diagnostic PD-L1 22C3 pharmDx assay (Agilent) considering two cut points for positivity, 1% and 50% PD-L1 stained tumor cells (12). A representative sample, in terms of demography (i.e., age, type of practice and experience, median 15 years), of ten surgical pathologists were randomly and equally assigned to two subgroups to evaluate a total of 108 NSCLC samples. Subgroup one analyzed all samples on 2 consecutive days. Subgroup two performed the same assessments except they received a 1-hour training session prior to the second assessment. This study reports results on the reproducibility of the pathologists' assessments with a large number of both observers and samples, ensuring good reliability in terms of precision of the calculated values and robustness of the study results, which are more likely to reflect real-life practice (12).

The gold standard TPS was established by two lead investigators trained and certified to assess PD-L1 22C3 pharmDx staining after a 2-day training course provided by Agilent prior to the study. It is interesting to note that among the 789 samples assessed to establish the gold standard PD-L1 TPS, there still was a lack of consensus for 8 (1%) of cases. Moreover, the authors report a lower prevalence of PD-L1 positivity (10.3% with PD-L1 TPS  $\geq$ 50%) compared to previously published data (i.e., approximately 23%) (7). The nature of the samples used in the study, including the use of tissue micro-arrays, early stage rather than late stage tumors, a high proportion of well differentiated tumors and potentially the use of archived samples, could explain this difference (12).

This is the first study to report on intra-observer reproducibility. The overall percent agreement (OPA) of intraobserver reproducibility was of 89.7% for the 1% cut point sample set, and of 91.3% for the 50% cut point sample set. Overall, there was a mean of 9.5% of irreproducible intraobserver assessments. Interestingly, for the 1% cut point sample set there were 4% of irreproducible negative (1<sup>st</sup> day) to positive (2<sup>nd</sup> day) cases, and 6.3% irreproducible positive (1<sup>st</sup> day) to negative (2<sup>nd</sup> day) cases. Likewise, for the 50% cut point sample set there were 4.3% of irreproducible negative (1<sup>st</sup> day) to positive (2<sup>nd</sup> day) cases, and 4.3% irreproducible positive (1<sup>st</sup> day) to negative (2<sup>nd</sup> day) cases. These results are not insignificant for the patients, when compared to the prevalence of a TPS of  $\geq$ 50% of 30.2% in the screened population in the open-label, phase III KEYNOTE-024 trial, which allowed approval by the FDA of pembrolizumab in the first-line setting in NSCLC patients (6). It is possible that intraobserver bias may lead to a shift towards the choice of the PD-1/PD-L1 inhibitor, whether it requires or not PD-L1 testing, and thus places extra emphasis on reducing inter-pathologist variability (17).

In addition, the OPA of interobserver reproducibility was of 84.2% for the 1% cut point sample set, and 81.9% for the 50% cut point sample set. Overall, a mean of 17% of assessments were irreproducible between observers, with a rate of 8% of irreproducible negative to positive cases and 7.8% of positive to negative cases for the 1% cut point sample set, while for the 50% cut point sample set there were as high as 14.1% of irreproducible negative to positive cases, and 4% of irreproducible positive to negative cases.

Importantly, the training had no impact on the interobserver reproducibility at 1% (i.e., OPA 82% versus 82.3%). For the 50% cut point sample set, the OPAs were similar for the first and second assessments with only a

slight improvement for the second assessment (i.e., 78.3 versus 81.7), suggesting that the training had little impact on the interobserver reproducibility at 50%. However, the training lasted only one hour, and in addition to giving strategies to optimally assess PD-L1 expression, it also covered elements maybe less important for interpretation, such as the biology of PD-L1 and development of the assay.

Variability in pathologists' assessment was high for samples with PD-L1 TPS, between 30% and 80% when compared with the gold standard. This is of particular concern especially around the cut point  $\geq 50\%$  TPS established for first-line treatment with pembrolizumab in locally advanced or metastatic NSCLC. Pathologists had a tendency to underestimate the PD-L1 TPS when weak membranous or incomplete staining or concomitant cytoplasmic staining was observed, although several pathologists also overestimated PD-L1 TPS, most likely related to PD-L1 expressed in the tumor-infiltrating immune cells. While normal cells and tumor-associated immune cells were excluded from the scoring, no information was presented on how such an exclusion analysis was performed.

A variety of pitfalls and artifacts, such as non-specific background, edge artifacts, crush artifacts, necrosis, or poor fixation, may be encountered when evaluating PD-L1 staining (IASLC Atlas of PD-L1 Testing in Lung Cancer). Potential false-positive results can be assigned to intra-alveolar or tumor-infiltrating macrophages that exhibit strong membranous staining or to stromal elements (inflammatory or endothelial cells) that can show different intensities of staining. In addition, lung samples may contain anthracotic or other pigments in the cytoplasm, which may confound IHC interpretation. Comparison with the hematoxylin and eosin staining morphology and histochemical stainings may be useful to exclude such non-tumoral staining, particularly for small biopsy samples (IASLC Atlas of PD-L1 Testing in Lung Cancer).

Similarly, a recent study showed that up to 20% of the analyzed cases were differently classified as positive or negative by any pathologist compared with the gold standard using the cut point  $\geq 1\%$  TPS, whereas there was better agreement between pathologists using the cut point  $\geq 50\%$  TPS (0–5% of cases) (13).

Thus, there is some evidence that the intraobserver and interobserver reproducibility presents an intrinsic source of error and bottlenecks for PD-L1 staining assessment. In this context, some solutions for optimization should be adopted to urgently improve the assessment of PD-L1 expression in routine clinical practice.

Education and repeated training of pathologists may improve consistency in PD-L1 assessment, but only up to a point beyond which improvement in methodology is needed (12). In particular, training should be conducted with classic and difficult examples, strategies for tricky cases, and on-line educational material, to optimize pathologists' assessment of PD-L1 staining. Several external quality assessment schemes have been developed, while the rate of adherence for pathologists remains low (i.e., European Society of Pathology, NordiQC, AFAQAP). In addition, intra-laboratory quality assessment of the PD-L1 interpretation should be performed continuously in a pathology laboratory based on a weekly interpathologists' evaluation grid. Obtaining a second opinion for difficult cases on an on-line platform may be a strategy to improve diagnostic accuracy.

Finally, the variability among pathologists coupled with the inherent tumor heterogeneity signifies that a more objective and veritably more quantitative strategy is needed. In the study by Cooper *et al.*, the gold standard for PD-L1 assessment was not fully objective but consisted of subjective assessment of PD-L1 expression by two trained pathologists (12). Interestingly, digital computer-assisted methods may improve IHC quantification. However, the availability of these approaches is limited and still needs standardization (18). In cases where manual scoring is severely hampered, an alternative digital method may be considered in quantifying PD-L1 expression.

It would be beneficial to evaluate PD-L1 expression combined with multicolor IHC assays to better characterize the immune tumor microenvironment by staining cells such as CD8 T cells, macrophages, myeloid-derived suppressor cells, natural-killers or regulatory T cells (19). These new approaches may contribute to optimized biomarker assessment of clinical samples, as well as improvement of the predictive value of PD-L1 expression on both tumor cells and immune cells for immunotherapy.

In conclusion, several approaches have the potential to improve the assessment by pathologists of PD-L1 staining in clinical practice. Personalized cancer immunotherapy should integrate in the future the evaluation of PD-L1 expression along with specific mechanisms through which cancers escape the anti-tumor immune response (20).

## Acknowledgements

The authors would like to thank the “Conseil Départemental des Alpes Maritimes 06”, the “Comité

Départementale 06 de la Ligue contre le Cancer”, and the “Cancéropôle PACA” for their financial support.

## Footnote

*Conflicts of Interest:* Paul Hofman is a member of different industrial scientific advisory boards (Roche, MSD, AstraZeneca, Novartis, Bristol-Myers Squibb, Pfizer, Qiagen, Janssen, Biocartis) for which he receives honorarium. Marius Ilié has no conflicts of interest to declare.

## References

- Ilie M, Hofman V, Dietel M, et al. Assessment of the PD-L1 status by immunohistochemistry: challenges and perspectives for therapeutic strategies in lung cancer patients. *Virchows Arch* 2016;468:511-25.
- Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer* 2012;12:252-64.
- Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. *N Engl J Med* 2015;373:1627-39.
- Herbst RS, Baas P, Kim DW, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet* 2016;387:1540-50.
- Rittmeyer A, Barlesi F, Waterkamp D, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *Lancet* 2017;389:255-65.
- Reck M, Rodriguez-Abreu D, Robinson AG, et al. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. *N Engl J Med* 2016;375:1823-33.
- Garon EB, Rizvi NA, Hui R, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. *N Engl J Med* 2015;372:2018-28.
- Scheerens H, Malong A, Bassett K, et al. Current Status of Companion and Complementary Diagnostics: Strategic Considerations for Development and Launch. *Clin Transl Sci* 2017;10:84-92.
- Marchetti A, Barberis M, Franco R, et al. Multicenter Comparison of 22C3 PharmDx (Agilent) and SP263 (Ventana) Assays to Test PD-L1 Expression for NSCLC Patients to Be Treated with Immune Checkpoint Inhibitors. *J Thorac Oncol* 2017;12:1654-63.
- Scheel AH, Baenfer G, Baretton G, et al. Interlaboratory-concordance of PD-L1 immunohistochemistry for non-small cell lung cancer. *Histopathology* 2017. [Epub ahead of print].
- Hirsch FR, McElhinny A, Stanforth D, et al. PD-L1 Immunohistochemistry Assays for Lung Cancer: Results from Phase 1 of the Blueprint PD-L1 IHC Assay Comparison Project. *J Thorac Oncol* 2017;12:208-22.
- Cooper WA, Russell PA, Cherian M, et al. Intra- and Interobserver Reproducibility Assessment of PD-L1 Biomarker in Non-Small Cell Lung Cancer. *Clin Cancer Res* 2017;23:4569-77.
- Brunnstrom H, Johansson A, Westbom-Fremer S, et al. PD-L1 immunohistochemistry in clinical diagnostics of lung cancer: inter-pathologist variability is higher than assay variability. *Mod Pathol* 2017;30:1411-21.
- Scheel AH, Dietel M, Heukamp LC, et al. Harmonized PD-L1 immunohistochemistry for pulmonary squamous-cell and adenocarcinomas. *Mod Pathol* 2016;29:1165-72.
- Phillips T, Simmons P, Inzunza HD, et al. Development of an automated PD-L1 immunohistochemistry (IHC) assay for non-small cell lung cancer. *Appl Immunohistochem Mol Morphol* 2015;23:541-9.
- Cooper WA, Tran T, Vilain RE, et al. PD-L1 expression is a favorable prognostic factor in early stage non-small cell carcinoma. *Lung Cancer* 2015;89:181-8.
- Chow JC, Cheung KM, Cho WC. Atezolizumab in non-small cell lung cancer: the era of precision immunoncology. *Ann Transl Med* 2017;5:265.
- Aeffner F, Wilson K, Martin NT, et al. The Gold Standard Paradox in Digital Image Analysis: Manual Versus Automated Scoring as Ground Truth. *Arch Pathol Lab Med* 2017;141:1267-75.
- Hofman P, Beaulande M, Hadj SB, et al. Automated brightfield multiplex immunohistochemistry to quantify biomarkers related to immune senescence: Relationships with survival in non-small cell lung cancer patients. *J Clin Oncol* 2017;35:e20500.
- Ribas A. Adaptive Immune Resistance: How Cancer Protects from Immune Attack. *Cancer Discov* 2015;5:915-9.

**Cite this article as:** Ilie M, Hofman P. Reproducibility of PD-L1 assessment in non-small cell lung cancer—know your limits but never stop trying to exceed them. *Transl Lung Cancer Res* 2017;6(Suppl 1):S51-S54. doi: 10.21037/tlcr.2017.10.13