



## PD-L1 Expression Scoring: Noninterchangeable, Noninterpretable, Neither, or Both

Niki Gavrielatou , MD, Saba Shafi, MD, Patricia Gaule, PhD, David L. Rimm , MD, PhD\*

Department of Pathology, Yale University School of Medicine, New Haven, CT, USA

\*Correspondence to: David L. Rimm, MD, PhD, Department of Pathology, Yale University School of Medicine, 310 Cedar Street, BML 116, PO Box 208023, New Haven, CT 06520-8023, USA (e-mail: david.rimm@yale.edu).

The IMpassion 130 trial led to the approval of anti-PD-L1 agent atezolizumab in combination with nab-paclitaxel as first-line treatment for unresectable locally advanced or metastatic triple-negative breast cancer (1). Schmid and colleagues showed improved progression-free median survival (PFS) by 2 months (hazard ratio [HR] = 0.63, 95% confidence interval [CI] = 0.50 to 0.80) and, most importantly, prolonged median overall survival (OS) by 7 months (HR = 0.71, 95% CI = 0.54 to 0.94) among patients with immune cell (IC) PD-L1 expression of 1% or more (ie, IC  $\geq$  1%). This led to US Food and Drug Administration approval and widespread use of this therapy. The atezolizumab label specifies the use of the Ventana PD-L1 SP142 immunohistochemical (IHC) assay (SP142) where only patients with "IC > 1%" are qualified for the drug. This biomarker requirement has opened Pandora's box because: 1) it is the stipulated IHC assay but has been proven to be less sensitive than other PD-L1 detection assays, 2) it uses a different system for pathologist-based analysis, and 3) it is not equivalent to other assays that have subsequently been approved in the same class (2), it requires the pathologist to know which treatment will be chosen prior to performing the assay (or to perform 2 unstandardized assays for the same biomarker).

To begin to address this problem, in this issue of the Journal, Rugo et al. (3) performed a post hoc exploratory analysis on the biomarker-evaluable population of IMpassion 130 (68.1% of the intent-to-treat population) to investigate the analytical concordance and outcome differences between the SP142 and 2 other PD-L1 IHC assays: Ventana SP263 (SP263) and the Agilent/Dako 22C3 (22c3). The SP142 assay detected the lowest number of PD-L1 positive cases stratified by the IC of more than 1% cutoff (46.4%) in comparison with SP263 (74.9%), 22C3 (73.1%), and 22C3 using combined positive score (CPS > 1; 80.9%). The overall percentage agreement between the assays comparing the IC of more than 1% cutoff for all 3 assays in addition to the CPS of more than 1 for 22C3 was less than 70%, suggesting analytical discordance among the assay readouts. Though all assays demonstrated similar results for clinical outcome with the IC of more than 1% cutoff (PFS HR = 0.60-0.68; OS HR = 0.74-0.79), the

investigators emphasize that SP142 showed the greatest difference in median PFS and OS values. Exploring the differences among subgroups with combined positivity, double-positive cases showed the highest clinical activity for PFS and OS. However, SP263-positive cases also showed improved PFS and OS (HR = 0.64, 95% CI = 0.53 to 0.79; HR = 0.75, 95% CI = 0.59 to 0.96). To attempt to define assay equivalence among assays with differential sensitivity, a mathematical model was employed using the optimal combinations for the overall percentage agreement, positive percentage agreement, and negative percentage agreement, and cutoffs were determined as an IC of 4% or higher for SP263 and CPS 10 for 22C3. However, assay concordance remained poor, and almost one-quarter of SP142 PD-L1-positive cases were undetected by the new cutoffs. Nonetheless, SP263 IC of 4% or higher demonstrated similar hazard ratios to SP142 IC of 1% or higher subgroup in predicting improved PFS and OS. Rugo et al. (3) conclude that the best assay for selecting patients for atezolizumab is the SP142 assay, and even with adjustment of cutoffs for differential assay sensitivities, the other assays do not perform as well, therefore concluding that the assays are noninterchangeable.

Different scoring systems and thresholds are used to determine PD-L1 positivity by immunohistochemistry, as summarized in Table 1 of Rugo et al. (3). The tumor proportion score (TPS) is defined as the percent of viable tumor cells showing partial or complete membrane staining, regardless of intensity. This assay has been consistently validated as accurate and reproducible (4,5). The IC assay is less well defined and has 2 scoring systems and multiple organ system-specific cut-points. This assay has been shown to be not reproducible in 2 large multi-institutional studies, one of which included pathologist training (5). The combined positive score (CPS) is the number of PD-L1 staining cells (tumor cells, lymphocytes, macrophages) divided by the total number of viable tumor cells, multiplied by 100. CPS offers the advantage of eliminating the need to choose between tumor and immune cell PD-L1 expression as a predictive biomarker (6). The CPS score has not been validated in a

Received: May 20, 2021; Accepted: May 26, 2021

© The Author(s) 2021. Published by Oxford University Press. All rights reserved. For permissions, please email: journals.permissions@oup.com

large, prospective, multi-institutional study, although that work is in progress.

Although Rugo et al. (3) provide critical information about assay noninterchangeability, they do not address the important question of assay interpretation. Although these methods are US Food and Drug Administration approved, it does not mean they are exempt from science related to assay interpretation. In fact, the specific system used for the assays in question is interpretation of IC. Three prospective, statistically powered, multi-institutional studies have been conducted to assess IC, and all 3 suggest pathologist agreement was poor, summarized by interclass correlation coefficients of less than 0.3 (clinically unacceptable) or overall percentage agreement less than 50% (4,5,7). In Rugo et al. (3), interpretation and scoring were carried out by 8 separate pathologists who participated in specific training programs for each assay. The authors have provided detailed information about the training schema (Supplementary Figure 2 and Supplementary Table 1, available online). Only a single pathologist received training in reading all 3 assays, 5 pathologists received training in 2 assays, and 2 pathologists received single assay training. Concordance between the pathologists could not be assessed as authors state that “each immunostained slide was read by a single trained pathologist per scoring algorithm.” The authors make the case that this is comparable to real-world situations. The authors thus base all calculations on a score that could have less than a 50% agreement with scoring by another pathologist. It is a limitation of this work that the interpretation, although perhaps the current standard of care, may be inaccurate or nonreproducible. These results also raise questions about the role of the SP142 diagnostic test, perhaps failing to enrich the responding population in the highly similar IMpassion131 trial, which failed to meet its primary endpoints.

Finally, even assuming the assays were accurately read, the data in Table 3 of Rugo et al. (3) raise an interesting question of interchangeability related to outcome, not concordance. Although SP142 is the only approved assay for clinical use, interchangeability with the SP263 assay might be considered to maximize patient benefit. The authors focus on the longer median OS difference (9.4 vs 3.3 months) for SP142, however, median survival is less informative than hazard ratio, which better accounts for the whole population not just the behavior of the median in each population. If the hazard ratio is considered for OS, the SP142 assay is essentially equivalent (HR = 0.74) compared with SP263 (HR = 0.75). In fact, because of the number of patients at risk, the SP142 assay is not statistically significant (95% CI exceeds 1.0), whereas the SP263 is statistically significant. The hazard ratio means that biomarker-positive patients are about 0.75 as likely to have an event (death) compared with biomarker-negative patients. However, because the SP263 assay has 460 positives compared with SP142 having only 285 positives, it appears that more patients would benefit if the SP263 assay was used than would benefit with SP142. Rugo et al. (3) show strong data about noninterchangeability with respect to

assay concordance, however, an argument can be made that the assays are, in fact, interchangeable with respect to outcome.

## Funding

Dr Rimm is funded by the Breast Cancer Research Foundation #20-198.

## Notes

**Role of the funder:** The funder had no role in the writing of this editorial or the decision to submit it for publication.

**Disclosures:** In the last 3 years, DLR has served as an advisor for Astra Zeneca, Agendia, Amgen, BMS, Cell Signaling Technology, Cepheid, Danaher, Daiichi Sankyo, Genoptix/Novartis, GSK, Konica Minolta, Merck, NanoString, PAIGE.AI, Perkin Elmer, Roche, Sanofi, Ventana, and Ultivue. Amgen, Cepheid, NavigateBP, NextCure, and Konica Minolta fund research in DLR's lab. The other authors have nothing to disclose.

**Author contributions:** NG: Writing—Original Draft, Writing—Review and Editing. SS: Writing—Original Draft, Writing—Review and Editing. PG: Conceptualization, Writing—Original Draft, Writing—Review and Editing. DR: Conceptualization, Methodology, Writing—Original Draft, Writing—Review and Editing, Supervision.

## Data Availability

Not applicable.

## References

- Schmid P, Adams S, Rugo HS, et al. Atezolizumab and nab-paclitaxel in advanced triple-negative breast cancer. *N Engl J Med*. 2018;379(22):2108–2121.
- Cortes J, Gescon DW, Rugo HS, et al. Pembrolizumab plus chemotherapy versus placebo plus chemotherapy for previously untreated locally recurrent inoperable or metastatic triple-negative breast cancer (KEYNOTE-355): a randomised, placebo-controlled, double-blind, phase 3 clinical trial. *Lancet*. 2020;396(10265):1817–1828.
- Rugo H, Loi S. PD-L1 immunohistochemistry assay comparison in atezolizumab plus nab-paclitaxel-treated advanced triple-negative breast cancer. *J Natl Cancer Inst*. 2021.
- Rimm DL, Han G, Taube JM, et al. A prospective, multi-institutional, pathologist-based assessment of 4 immunohistochemistry assays for PD-L1 expression in non-small cell lung cancer. *JAMA Oncol*. 2017;3(8):1051–1058.
- Tsao MS, Kerr KM, Kockx M, et al. PD-L1 immunohistochemistry comparability study in real-life clinical samples: results of blueprint phase 2 project. *J Thorac Oncol*. 2018;13(9):1302–1311.
- Kulangara K, Zhang N, Corigliano E, et al. Clinical utility of the combined positive score for programmed death ligand-1 expression and the approval of pembrolizumab for treatment of gastric cancer. *Arch Pathol Lab Med*. 2019;143(3):330–337.
- Reisenbichler ES, Han G, Bellizzi A, et al. Prospective multi-institutional evaluation of pathologist assessment of PD-L1 assays for patient selection in triple negative breast cancer. *Mod Pathol*. 2020;33(9):1746–1752.