Annals of Diagnostic Pathology 48 (2020) 151602

Contents lists available at ScienceDirect

Annals of Diagnostic Pathology

journal homepage: www.elsevier.com/locate/anndiagpath

# Utility of PD-L1 testing on non-small cell lung cancer cytology specimens: An institutional experience with interobserver variability analysis

Oleksandr Kravtsov[a,*], Christopher P. Hartley[b], Yuri Sheinin[a], Bryan C. Hunt[a], Juan C. Felix[a], Tamara Giorgadze[a]

[a] Medical College of Wisconsin, Department of Pathology, 9200 W. Wisconsin Avenue, Milwaukee, WI 53226, United States of America
[b] Mayo Clinic, Department of Laboratory Medicine and Pathology, 200 First Street SW, Rochester, MN 55905, United States of America

## 1. Introduction

Programmed death ligand 1 (PD-L1) is a membrane protein that serves as a ligand of programmed cell death 1 (PD-1) inhibitory receptor, that is present on immune system cells such as T-lymphocytes. PD-L1 normally prevents activation of cellular autoimmune response via its inhibitory mechanism, however the same mechanism is one of the ways cancer cells of various types can evade the immune response [1]. Immune checkpoint blockade is emerging as an effective type of non-small cell lung cancer (NSCLC) immunotherapy. In 2015 United States Food and Drug Administration (FDA) approved pembrolizumab for treatment of metastatic NSCLC in cases with positive PD-L1 expression in at least 1% tumor cells and failed chemotherapy with other agents [2]. Pembrolizumab use was extended to first-line therapy in advanced NSCLC cases with no EGFR and/or ALK mutation and PD-L1 expression of 50% or more [3]. PD-L1 IHC clone 22C3 pharmDx qualitative assay (Agilent, Santa Clara, CA) is approved by FDA as the companion diagnostic test to determine eligibility of advanced NSCLC patients for pembrolizumab treatment [4]. Tumor Proportion Score (TPS) is used to evaluate PD-L1 status in NSCLC patients, which is the percentage of viable tumor cells showing partial or complete membranous staining of any intensity. Other PD-L1 antibody clones are being developed such as Ventana SP263, however studies showed conflicting results in comparison with 22C3 clone with discrepancies at clinically relevant cutoffs [5]. Minimally invasive approaches such as endobronchial ultrasound-guided (EBUS) fine-needle aspiration are increasingly used for tissue diagnosis of lung cancer. Although cytology preparations provide valuable material for PD-L1 assessment in lung cancer, the literature on this topic is limited. Our aim was to review PD-L1 performed on cytology specimens at our institution to evaluate adequacy and interobserver reproducibility of PD-L1 reporting.

## 2. Materials and methods

A retrospective database search was performed for all cases of NSCLC that had PD-L1 immunohistochemical stain performed on cytological preparations from January 2015 to August 2018. The PD-L1 IHC 22C3 pharmDx kit (Agilent, Santa Clara, CA) with 22C3 clone of anti-PD-L1 antibody was used on Dako Link 48 platform by our laboratory according to the manufacturer's instructions throughout the entire testing period. Cytology aspirates were collected into methanol-based fixative (CytoLyt; Hologic Corp., Marlborough, MA). Cell blocks were prepared via plasma-thrombin method with consecutive formalin fixation, paraffin embedding and microtomy using standard laboratory procedures. Fifty cytology cases were randomly selected and re-evaluated by 3 cytopathologists with experience in the evaluation and reporting of PD-L1 stains to study interobserver reproducibility. All 3 graders were blinded to the previously reported PD-L1 status. Partial or complete membranous staining of any intensity was considered for positive for the purpose of Tumor Proportion Score (TPS) calculation. The cases with < 100 tumor cells were considered insufficient for PD-L1 evaluation. < 1%, 1–50%, and > 50% TPS cutoffs were used corresponding to negative, low positive, and high positive results respectively. Interobserver variation was compared across TPS categories via the Fleiss kappa statistic. Interobserver agreement was analyzed for low positive, high positive, negative and unsatisfactory categories and separately analyzed for positive (combined low positive and high positive into a single category), negative and unsatisfactory categories. Kappa values were interpreted as follows: 0 indicates less than a chance agreement, 0.01–0.20 indicates slight agreement, 0.21–0.40 indicates fair agreement, 0.41–0.60 indicates moderate agreement, 0.61–0.80 indicates substantial agreement, and 0.81–0.99 indicates almost perfect agreement.

## 3. Results

### 3.1. Clinicopathological data and adequacy assessment

NSCLC cytology cases (#75), including 47 adenocarcinomas and 28 squamous cell carcinomas were identified (see Table 1). PD-L1 stains

* Corresponding author at: 350 W 11th St., Indianapolis, IN 46202, United States of America.
*E-mail addresses:* olkrav@iu.edu (O. Kravtsov), Hartley.Christopher@mayo.edu (C.P. Hartley), ysheinin@mcw.edu (Y. Sheinin), bhunt@mcw.edu (B.C. Hunt), jcfelix@mcw.edu (J.C. Felix), tgiorgadze@mcw.edu (T. Giorgadze).

https://doi.org/10.1016/j.anndiagpath.2020.151602

**Table 1**
PD-L1 results on cytology specimens.

| Diagnosis | Negative, n (%) | Low positive n (%) | High positive n (%) | Insufficient n (%) |
|---|---|---|---|---|
| Adenocarcinoma | 12 (16) | 7 (9) | 19 (25) | 9 (12) |
| Squamous cell carcinoma | 6 (8) | 8 (11) | 11 (15) | 3 (4) |
| Total | 18 (24) | 15 (20) | 30 (40) | 12 (16) |

**Table 2**
Interobserver variability across all TPS categories (50 cases, 3 independent graders).

| Result category | TPS cutoff | Kappa value | *P*-value | Variability interpretation |
|---|---|---|---|---|
| High positive | ≥50% | 0.71 | < 0.01 | Substantial agreement |
| Low positive | ≥1%–49% | 0.45 | < 0.01 | Moderate agreement |
| Negative | < 1% | 0.71 | < 0.01 | Substantial agreement |
| Unsatisfactory | N/A | 0.83 | < 0.01 | Almost perfect agreement |
| All categories | N/A | 0.66 | < 0.01 | Substantial agreement |

TPS — Tumor Proportion Score.

performed on cytology preparations were adequate for evaluation (> 100 tumor cells present) in 84% of all cases. Overall, 60% of NSCLC cases showed positive PD-L1 results (20% low positive or 1%–49% TPS and 40% high positive or ≥50% TPS) and 24% were negative or < 1% TPS. Upon review of clinical data 27 patients were found to be deceased at the time of database search. The distribution of PD-L1 results in deceased patients showed no significant differences with the overall distribution (Fisher's exact test, $p = 0.63$).

### 3.2. Interobserver variability

Interobserver variability analysis with the Fleiss kappa methodology across all TPS categories (see Table 2) showed overall substantial agreement among the three independent observers (Kappa value 0.66, $p < 0.01$). 33 of 50 samples (66%) had perfect concordance among all three observers. Almost perfect agreement was achieved within unsatisfactory category (Kappa value 0.83, $p < 0.01$). The lowest kappa value was obtained within low positive category (Kappa value 0.45, moderate agreement, $p < 0.01$). Substantial agreement was achieved within high positive and negative categories (Kappa values 0.71 in both, $p < 0.01$ in both). See Fig. 1 for paired comparisons of PD-L1 results by observer and TPS category.

**Table 3**
Interobserver variability with combined low positive and high positive results into the single positive category (50 cases, 3 independent graders).

| Result category | TPS cutoff | Kappa value | *p*-Value | Variability interpretation |
|---|---|---|---|---|
| Positive | ≥1% | 0.69 | < 0.01 | Substantial agreement |
| Negative | < 1% | 0.71 | < 0.01 | Substantial agreement |
| Unsatisfactory | N/A | 0.83 | < 0.01 | Almost perfect agreement |
| All categories | N/A | 0.72 | < 0.01 | Substantial agreement |

TPS — Tumor Proportion Score.

Overall, slightly better interobserver agreement is seen if low positive and high positive results are combined into a single positive category (see Table 3) using ≥1% TPS cutoff (Kappa value 0.72, substantial agreement, $p < 0.01$). 39 of 50 samples (78%) had perfect concordance among all three observers. See Fig. 2 for paired comparisons of PD-L1 results by observer and category.

Different TPS category was assigned in 17 of 50 (34%) cases by at least 1 observer (see Table 4). 3 of 17 cases (6% from total 50 cases) had different TPS category assigned by all 3 independent graders. Diagnosis of squamous cell carcinoma was made in 6 of 17 tumors with discrepant PD-L1 results and 11 of 17 cases were diagnosed as adenocarcinoma.

11 of 50 (22%) cases with interobserver discrepancies of PD-L1 interpretation are identified if low positive and high positive results are combined into a single positive category using ≥1% TPS cutoff.

8 of 50 (16%) cases had discrepancies in negative versus positive PD-L1 results. 3 of 50 (6%) had discrepancies in unsatisfactory versus positive PD-L1 results (2 cases with unsatisfactory versus low positive results and 1 case with unsatisfactory versus high-positive result).

## 4. Discussion

Validation of feasibility of cytological preparations such as cell blocks for accurate and reliable assessment of PD-L1 is of great clinical importance for NSCLC patients as fine-needle aspiration cytology is being increasingly used as the primary source for tissue diagnosis. There are countries where most of the patients have only cytological material available for diagnosis, however, none of the major clinical trials have validated the use of cytological specimens for PD-L1 status assessment [6,7].

Interpretation of PD-L1 immunohistochemistry on cytological

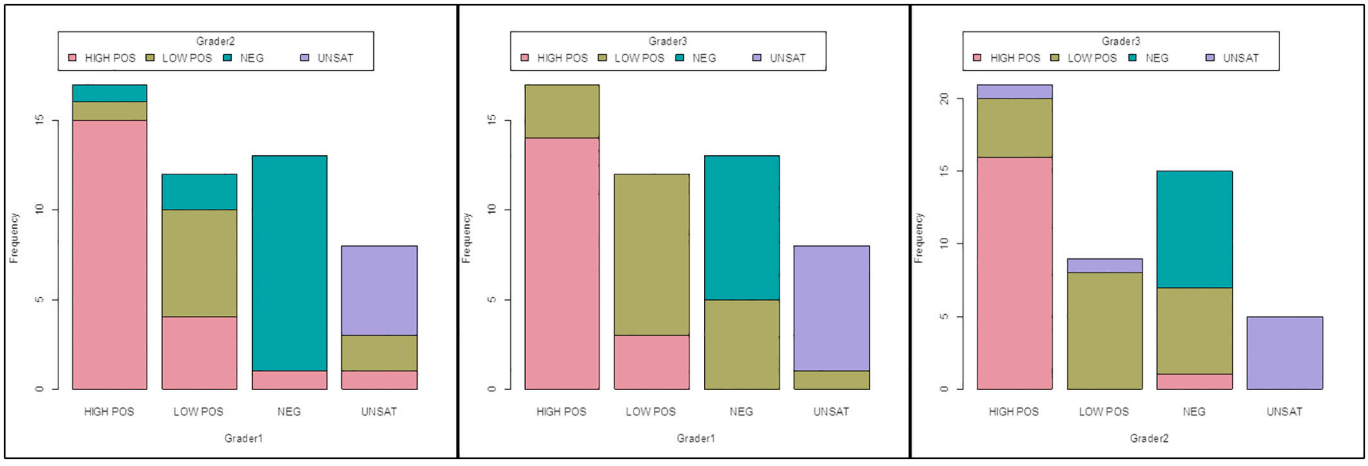


**Fig. 1.** Paired comparisons of PD-L1 results by observer and TPS category (1% and 50% cutoffs, all TPS categories).

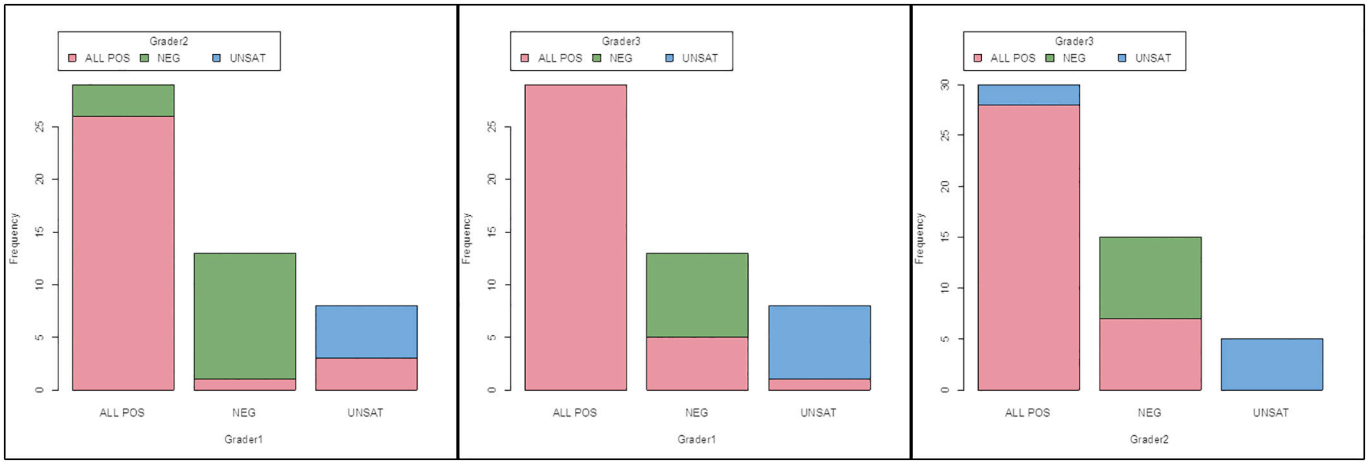


**Fig. 2.** Paired comparisons of PD-L1 results by observer and category (1% cutoff only, low positive and high positive results are combined into the single positive category).

**Table 4**
PD-L1 results with discrepant interobserver interpretation by case and by grader.

| Case | PD-L1 results | | | Diagnosis |
|---|---|---|---|---|
| | Grader 1 | Grader 2 | Grader 3 | |
| 1 | LOW POS | **NEG** | LOW POS | Adenocarcinoma |
| 2 | NEG | HIGH POS | LOW POS | SCC |
| 3 | LOW POS | **HIGH POS** | LOW POS | SCC |
| 4 | HIGH POS | NEG | LOW POS | Adenocarcinoma |
| 5 | **LOW POS** | HIGH POS | HIGH POS | SCC |
| 6 | NEG | NEG | **LOW POS** | Adenocarcinoma |
| 7 | UNSAT | **LOW POS** | UNSAT | Adenocarcinoma |
| 8 | LOW POS | **HIGH POS** | LOW POS | SCC |
| 9 | NEG | NEG | **LOW POS** | SCC |
| 10 | **HIGH POS** | LOW POS | LOW POS | SCC |
| 11 | NEG | NEG | **LOW POS** | Adenocarcinoma |
| 12 | NEG | NEG | **LOW POS** | Adenocarcinoma |
| 13 | **UNSAT** | LOW POS | LOW POS | Adenocarcinoma |
| 14 | **LOW POS** | HIGH POS | HIGH POS | Adenocarcinoma |
| 15 | UNSAT | **HIGH POS** | UNSAT | Adenocarcinoma |
| 16 | LOW POS | NEG | HIGH POS | Adenocarcinoma |
| 17 | HIGH POS | HIGH POS | **LOW POS** | Adenocarcinoma |

Bolded are outlying results in cases with concordant results of 2 graders, NEG — negative, LOW POS — low positive, HIGH POS — high positive, UNSAT — unsatisfactory, SCC — squamous cell carcinoma.

preparation is by its nature more challenging than on biopsies or surgical resection specimens. The main concern is having limited amounts of tumor tissue available for both establishing the diagnosis and assessment of PD-L1 status. Rapid on-site evaluations (ROSE) are of utmost importance for obtaining adequate FNA samples for the diagnosis and all ancillary studies. Pathologists, cytotechnologists and cytopathology fellows in our practice that are performing ROSE have been instructed to request additional FNA passes for cell block preparations since we started PD-L1 testing. Our data showed that cell block preparations had more than 100 tumor cells in most cases (84%) and, therefore, were feasible for assessment of PD-L1 status. This conclusion is in keeping with prior published research, including the study by Heymann et al., that showed that PD-L1 assessment is feasible in cytological preparations, and the results are concordant with biopsy and surgical resection specimens [8].

The question of accuracy and reliability of PD-L1 immunohistochemistry interpretation in cell block preparations is a complex issue as it can be influenced by many factors including, but not limited to, interpreter background and experience, preanalytical issues such as type of fixative used in cytology preparation, heterogeneity of PD-L1 expression in tumor and problems with reliable differentiation of tumor cells from benign epithelial (see Fig. 3A) and inflammatory cells on the PD-L1 stained cell block preparations, especially in cases with dispersed tumor cells not forming easily identifiable and discreet clusters. PD-L1 expression in macrophages (see Fig. 3B) is another important pitfall that is especially relevant in PD-L1 assessment of well-differentiated pulmonary adenocarcinomas, as macrophages and adenocarcinoma cells share some morphologic features such as large cellular size and sometimes require correlation with tumor morphology on H&E-stained biopsy sampling and immunohistochemical assessment of CD68 and TTF1 expression for definitive distinction. Only PD-L1 expression in tumor cells should be counted towards TPS calculation, unlike Combined Positive Score (CPS), that includes positive staining in tumor cells, lymphocytes and macrophages and is used for PD-L1 assessment in esophageal squamous cell carcinoma, gastric or gastroesophageal adenocarcinoma, cervical carcinoma and other cancer sites [9].

It is unclear how the usage of different fixative solutions affects the concordance of staining results as PD-L1 IHC kits were initially validated only for formalin-fixed paraffin-embedded tissue samples. Our group has performed clinical and technical validation of PD-L1 status assessment using cell block preparation fixed by methanol-based solution (CytoLyt). Several studies were performed comparing the concordance of PD-L1 results on cytological preparation using different types of fixatives and showed conflicting results [10,11]. Canadian Association of Pathologists has issued comprehensive guidelines for PD-L1 biomarker testing that in one of its points recommends immediate fixation of cytology specimens in 10% neutral buffered formalin with subsequent spinning into a cell pellet for the purpose of the result compatibility with validated biopsy samples [12]. PD-L1 immunohistochemistry analysis of direct cytology smears have been suggested by Lozano et al. [13]

We investigated interobserver variability among 3 board certified cytopathologists in our institution, that routinely report PD-L1 status on cytological preparations and showed overall substantial agreement

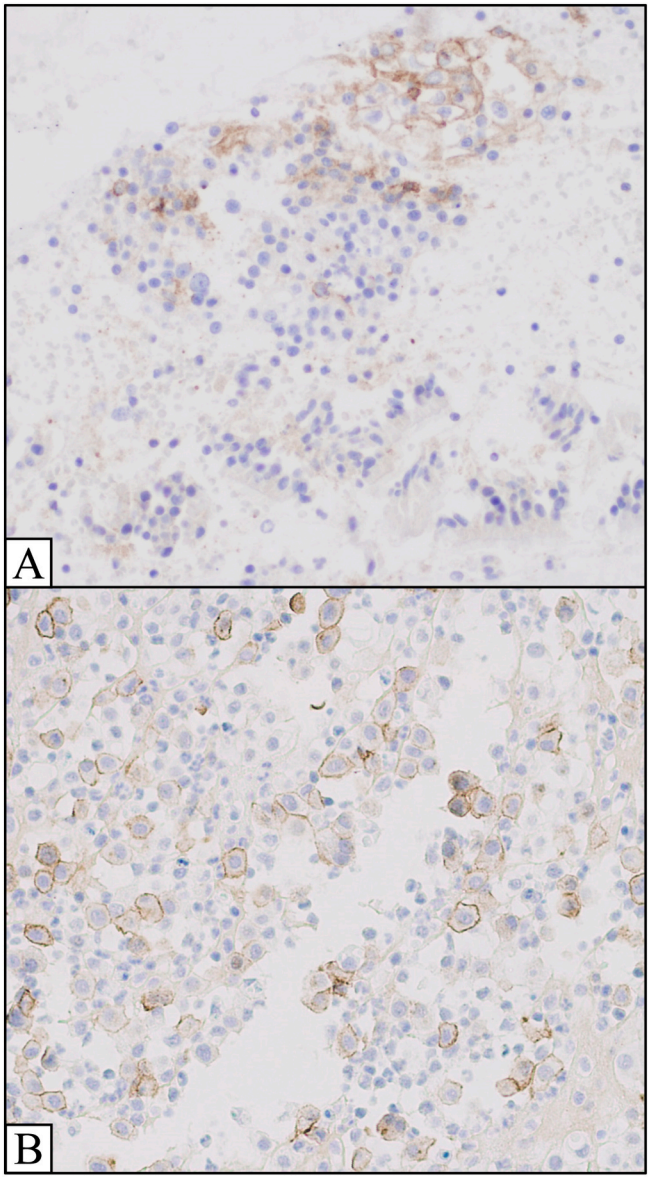


**Fig. 3.** Common pitfalls in PD-L1 status assessment on cell block preparations. A. Low positive PD-L1 expression in squamous cell carcinoma of the lung (upper right) and negative expression in benign bronchial epithelium (bottom), cell block, 200× magnification. B. PD-L1 expression in macrophages, cell block, 200× magnification. Adenocarcinoma cells with negative PD-L1 expression not shown in this photograph.

(Kappa value: 0.66) across all TPS categories using 1% and 50% cutoffs. Those results are comparable with previously published studies on interobserver variability of PD-L1 assessment on cytology specimens [14-16]. We did not include any recent cases in our evaluation to avoid the possibility of recall bias. PD-L1 status assessment was the most reliable in highly cellular specimens with PD-L1 expression close to 100%, highly cellular specimens with PD-L1 expression of 0% and acellular specimens. Upon review of cases with interobserver discrepancies (see Table 4) we found that majority of such cases had PD-L1 expression close to 1 and 50% cutoffs, prevalent expression of PD-L1 in macrophages or borderline tumor cellularity (see Fig. 4). Most of the discrepant cases were interpreted as low positive by at least one of the independent observers. With this data and the fact that in our interobserver variability study low positive TPS category had the lowest kappa value (0.45) across all TPS categories we suggest that results in low positive TPS category are the least reliable PD-L1 status assessment results on cell block preparations. Adenocarcinoma diagnosis was more prevalent than SCC among cases with interobserver discrepancies. This can be explained by the fact that adenocarcinoma cells tend to be more dispersed in the cytological preparations and, therefore can be harder to recognize on immunohistochemically stained slides. Comparing the morphology and distribution of PD-L1 positive cells with that of tumor cells identified on the cell block preparations with H&E or immunohistochemical stains (TTF-1, p40, p63, cytokeratins) may be helpful in these cases. The routine practice at our institution is to review equivocal or borderline PD-L1 interpretations at a cytopathology consensus conference.

Although interobserver agreement is statistically good, the clinical impact of result discrepancy can be very significant as the patient may or may not get immune checkpoint inhibitor therapy based on the result. Therefore, we recommend achieving a group consensus among multiple pathologists with experience in PD-L1 assessment in cases with borderline tumor cellularity or cases with PD-L1 expression close to 1% and 50% cutoffs and/or recommend repeat tissue sampling if achieving consensus is problematic.

## Disclosure

The submission builds on an abstract which was presented at USCAP 2019.

## Declaration of competing interest

All authors were active participants of this study, and none has connection to any companies or products mentioned in this manuscript.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

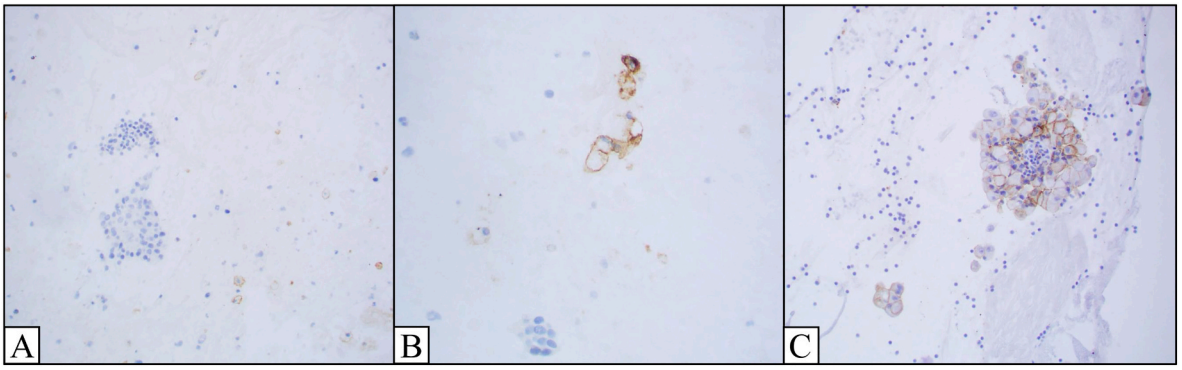


**Fig. 4.** Selected clinical cases with interobserver discrepancies of PD-L1 status assessment. A. Case 2 (consensus result: negative). Squamous cell carcinoma cluster with negative PD-L1 expression (left) and macrophages staining with PD-L1 (right), cell block, 100× magnification. B. Case 2 (consensus result: negative). Small cluster of squamous cell carcinoma with positive PD-L1 expression, representing overall less than 1% of tumor cells, cell block, 200× magnification. C. Case 15 (consensus result: unsatisfactory). Adenocarcinoma with borderline tumor cellularity (~90 cells), cell block, 200× magnification.

## References


[1] Keir ME, Butte MJ, Freeman GJ, Sharpe AH. PD-1 and its ligands in tolerance and immunity. Annu Rev Immunol 2008;26:677–704. https://doi.org/10.1146/annurev.immunol.26.021607.090331.

[2] Herbst RS, Baas P, Kim D-W, Felip E, Pérez-Gracia JL, Han J-Y, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet Lond Engl 2016;387:1540–50. https://doi.org/10.1016/S0140-6736(15)01281-7.

[3] Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. N Engl J Med 2016;375:1823–33. https://doi.org/10.1056/NEJMoa1606774.

[4] Kintsler S, Cassataro MA, Drosch M, Holenya P, Knuechel R, Braunschweig T. Expression of programmed death ligand (PD-L1) in different tumors. Comparison of several current available antibody clones and antibody profiling. Ann Diagn Pathol 2019;41:24–37. https://doi.org/10.1016/j.anndiagpath.2019.05.005.

[5] Munari E, Rossi G, Zamboni G, Lunardi G, Marconi M, Sommaggio M, et al. PD-L1 assays 22C3 and SP263 are not interchangeable in non–small cell lung cancer when considering clinically relevant cutoffs: an Interclone evaluation by differently trained pathologists. Am J Surg Pathol 2018;42:1384–9. https://doi.org/10.1097/PAS.0000000000001105.

[6] Yatabe Y, Kerr KM, Utomo A, Rajadurai P, Tran VK, Du X, et al. EGFR mutation testing practices within the Asia Pacific region: results of a multicenter diagnostic survey. J Thorac Oncol Off Publ Int Assoc Study Lung Cancer 2015;10:438–45. https://doi.org/10.1097/JTO.0000000000000422.

[7] Tsao MS, Kerr KM, Kockx M, Beasley M-B, Borczuk AC, Botling J, et al. PD-L1 immunohistochemistry comparability study in real-life clinical samples: results of blueprint phase 2 project. J Thorac Oncol Off Publ Int Assoc Study Lung Cancer 2018;13:1302–11. https://doi.org/10.1016/j.jtho.2018.05.013.

[8] Heymann JJ, Bulman WA, Swinarski D, Pagan CA, Crapanzano JP, Haghighi M, et al. PD-L1 expression in non-small cell lung carcinoma: comparison among cytology, small biopsy, and surgical resection specimens. Cancer Cytopathol 2017;125:896–907. https://doi.org/10.1002/cncy.21937.

[9] Davis AA, Patel VG. The role of PD-L1 expression as a predictive biomarker: an analysis of all US Food and Drug Administration (FDA) approvals of immune checkpoint inhibitors. J Immunother Cancer 2019;7:278. https://doi.org/10.1186/s40425-019-0768-9.

[10] Lloyd IE, Zhou W, Witt BL, Chadwick BE. Characterization of PD-L1 immunohistochemical expression in cell blocks with different specimen fixation and processing methods. Appl Immunohistochem Mol Morphol 2019;27:107–13. https://doi.org/10.1097/PAI.0000000000000572.

[11] Wang H, Agulnik J, Kasymjanova G, Wang A, Jiménez P, Cohen V, et al. Cytology cell blocks are suitable for immunohistochemical testing for PD-L1 in lung cancer. Ann Oncol 2018;29:1417–22. https://doi.org/10.1093/annonc/mdy126.

[12] Cheung CC, Barnes P, Bigras G, Boerner S, Butany J, Calabrese F, et al. Fit-for-purpose PD-L1 biomarker testing for patient selection in Immuno-oncology: guidelines for clinical laboratories from the Canadian Association of Pathologists-Association Canadienne Des Pathologistes (CAP-ACP). Appl Immunohistochem Mol Morphol 2019;27:699–714. https://doi.org/10.1097/PAI.0000000000000800.

[13] Lozano MD, Echeveste JI, Abengozar M, Mejías LD, Idoate MA, Calvo A, et al. Cytology smears in the era of molecular biomarkers in non-small cell lung cancer: doing more with less. Arch Pathol Lab Med 2018;142:291–8. https://doi.org/10.5858/arpa.2017-0208-RA.

[14] Chang S, Park HK, Choi Y-L, Jang SJ. Interobserver reproducibility of PD-L1 biomarker in non-small cell lung cancer: a multi-institutional study by 27 pathologists. J Pathol Transl Med 2019;53:347–53. https://doi.org/10.4132/jptm.2019.09.29.

[15] Cooper WA, Russell PA, Cherian M, Duhig EE, Godbolt D, Jessup PJ, et al. Intra- and interobserver reproducibility assessment of PD-L1 biomarker in non-small cell lung cancer. Clin Cancer Res Off J Am Assoc Cancer Res 2017;23:4569–77. https://doi.org/10.1158/1078-0432.CCR-17-0151.

[16] Downes MR, Slodkowska E, Katabi N, Jungbluth AA, Xu B. Inter- and intraobserver agreement of programmed death ligand 1 scoring in head and neck squamous cell carcinoma, urothelial carcinoma and breast carcinoma. Histopathology 2020;76:191–200. https://doi.org/10.1111/his.13946.