Implementation of PD-L1 22C3 IHC pharmDx™ in Cell Block Preparations of Lung Cancer: Concordance with Surgical Resections and Technical Validation of CytoLyt® Prefixation

Si Kei Lou^{a, b} Hyang Mi Ko^{a, b} Tomonari Kinoshita^c Scott MacDonald^a
Jessica Weiss^d Katarzyna Czarnecka-Kujawa^{c, e} Scott L. Boerner^{a, b}
Kazuhiro Yasufuku^{c, f} Ming-Sound Tsao^{a, b} Joerg Schwock^{a, b}

^aDivision of Pathology, University Health Network, Toronto, ON, Canada; ^bDepartment of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada; ^cDivision of Thoracic Surgery, University Health Network, Toronto, ON, Canada; ^dDivision of Biostatistics, University Health Network, Toronto, ON, Canada; ^eDivision of Respiriology, University Health Network, Toronto, ON, Canada; ^fDivision of Surgical Oncology, University Health Network, Toronto, ON, Canada

Keywords

Biomarker · Cell blocks · Concordance · Immunohistochemistry · Lung cytopathology · Programmed death ligand-1 · Pulmonary · Validation

Abstract

Background: Programmed death ligand-1 (PD-L1) assessed by immunohistochemistry (IHC) is used as biomarker for pembrolizumab therapy in advanced stage lung cancer patients. However, data permitting direct performance comparison between cytology and surgical specimen types are limited since both specimens from a single tumor site are infrequently available. In addition, alcohol fixation used with cytology specimens requires technical validation of the PD-L1 IHC assay before clinical use. We here report our experience with implementation of the PD-L1 22C3 IHC pharmDx™ assay for cytologic samples at a large tertiary cancer

center. **Study Design:** Archival formalin-fixed (FF), paraffin-embedded cell blocks (CBs) and subsequent lung tumor resections (LTRs) from the same anatomical site were used for a direct comparison of PD-L1 tumor proportion scores (TPSs). TPS values were independently determined by one surgical lung pathologist and two cytopathologists blinded to the specimen pairs. An interim analysis was performed to facilitate the pooling of expertise among observers. After PD-L1 22C3 IHC pharmDx™ implementation for FF cytology specimens, dual-processed samples were used for a prospective

Current affiliation of T. Kinoshita: Division of General Thoracic Surgery, Department of Surgery, Keio University School of Medicine, Tokyo, Japan. Preliminary results of this study were presented at the World Conference on Lung Cancer 2018 of the International Association for the Study of Lung Cancer (Toronto, ON, Canada) and the International Congress of Cytology 2019 of the International Academy of Cytology (Sydney, NSW, Australia).

technical validation of CytoLyt® prefixation (CF). Digital image analysis was performed for a subset of dual-processed specimens. **Results:** Eighty-one CBs and LTRs were included for comparison of the specimen types. PD-L1 assessment in CBs had an accuracy, sensitivity, specificity, positive predictive value, and negative predictive value of 88.9/72.8, 66.7/73.5, 95.2/72.3, 80.0/65.8, and 90.9/79.1% for the $\geq 50/\geq 1\%$ cutoff, respectively. The intraclass correlation coefficient was 0.84 (95% confidence interval [CI]: 0.76, 0.90), and it improved after interim analysis (before: 0.79 and after: 0.92). The overall concordance between CF and FF for the categories defined by the $\geq 50/\geq 1\%$ cutoff values was 90.4% (95% CI: 79.0, 96.8). Similar assay performance was confirmed by digital analysis. **Conclusions:** PD-L1 22C3 IHC pharmDx™ shows good reliability if used with CB preparations. CF does not impact assay results significantly. Clinical validation with outcome data is needed, and digital methods of assessment should be further investigated.

© 2020 The Author(s)
Published by S. Karger AG, Basel

Introduction

Lung cancer is the second most common cancer diagnosed in the USA and Canada and is the leading cause of cancer-related mortality [1, 2]. Non-small cell lung cancer (NSCLC), mainly adenocarcinoma and squamous cell carcinoma, is the most common form of lung cancer [3]. Immune checkpoint inhibitors, such as nivolumab and pembrolizumab, interfere with the interaction of programmed death receptor-1 (PD-1) and its ligand, programmed death ligand-1 (PD-L1), to enhance the host's natural antitumor immune response. Therapeutic PD-1/PD-L1 blocking antibodies have shown remarkable response rates and reduced risk of death in NSCLC patients whose tumors were determined to be PD-L1 positive by immunohistochemistry (IHC) in histologic tissue samples [4–7]. Specifically, patient eligibility for pembrolizumab therapy is determined using the 22C3 IHC pharmDx™ assay (Agilent Technologies/Dako, Carpinteria, CA, USA) as a companion diagnostic test, which was developed using formalin-fixed (FF), paraffin-embedded histologic specimens. However, the majority of NSCLC patients present with advanced stage disease and a significant proportion is diagnosed by cytology alone, making cytologic specimens (e.g., endobronchial ultrasound-guided transbronchial needle aspiration [EBUS-TBNA], effusion fluids) pivotal in diagnostic, prognostic, and predictive testing [8, 9]. Concerns related to the validity of PD-L1 testing of cytologic samples and the com-

mon use of non-formalin fixatives in cytology processing have, in the past, frequently necessitated additional sampling with associated procedure-related risks and costs.

Currently, only a limited number of studies have reported the use of the PD-L1 22C3 IHC pharmDx™ assay in specimen types other than FF paraffin-embedded histologic material (i.e., surgical resection specimens or core needle biopsies). These studies focused mainly on cell block (CB) and/or smear preparations with or without inclusion of histologic samples. Matched cytologic and histologic specimens from the same anatomic site of tumor manifestation (primary, metastatic) make up varying, often small subsets of the examined materials [10–22]. Even fewer studies specifically address preanalytical variables in clinical samples such as non-formalin fixatives or cold ischemia time. Although loss of PD-L1 immunoreactivity was noted after CytoLyt® fixation in a small study [23], others have reported no difference after exposure to alcohol fixatives [24, 25]. Notably, PD-L1 IHC validation may be complicated by the biological complexity of the marker with known heterogeneous expression, choice of multiple PD-L1 antibodies and assays with unique scoring methods/cutoff values, and a variety of potentially relevant preanalytical variables (e.g., fixation agent and duration and CB processing method).

The purpose of this study was to evaluate the performance of CB cytology samples for the PD-L1 22C3 IHC pharmDx™ assay in our laboratory. To this end, we performed a cyto-histologic comparison of FF paraffin-embedded CB and lung tumor resection (LTR) specimens matched for origin from the same primary lung tumor site. After the performance characteristics for CB specimens were established, we examined the effect of CytoLyt® prefixation (CF; Cytyc/Hologic, Marlborough, MA, USA) using dual-processed clinical samples and control materials in order to fully implement the assay for cytology in our clinical practice.

Materials and Methods

Case Selection for Cytology-Histology Correlation

CT-guided transthoracic fine-needle aspiration (FNA)-derived CBs for NSCLC with subsequent LTR specimen were identified through a search of the laboratory information system at the UHN (January 2012 to March 2018). The inclusion criteria were as follows: (1) sample collected with on-site assessment (to ensure fixation with 10% neutral buffered formalin [NBF] as per the standard operating procedure), (2) preliminary evidence of adequate CB tumor cellularity (CB-based IHC studies included in the diagnostic report), and (3) primary lung origin without neuroendocrine (small cell/large cell) differentiation. Of 114 cases, 33 (23%) were subse-

quently excluded due to (1) less than 100 tumor cells in the CB section ($n = 26$) and (2) >1 tumor in the LTR or marked difference in morphology between CB and tested portion of the LTR ($n = 7$), resulting in 81 cases for final analysis (71%). The time interval between specimens was recorded based on accessioning dates.

CB Preparation and IHC

CBs were generated from normal saline (NS) needle rinse fluid after formalin fixation and processing using the HistoGel™ (Thermo Fisher Scientific, Waltham, MA, USA) method. Cell pellets were formed by centrifugation (1 min, 600 g), resuspended in HistoGel™, wrapped in lens paper, and submitted for standard paraffin processing overnight. Automated processing was done for CBs and LTR materials using the program settings for small and large histologic specimens, as appropriate. Evaluation and diagnosis followed the 2015 World Health Organization classification of lung tumors [3]. IHC analysis of protein markers, including but not limited to, thyroid transcription factor 1, p40, chromogranin, and synaptophysin, when necessary, was used to establish the diagnosis. CB section and a single representative LTR section were freshly cut at 4-micron thickness and mounted on coated slides. The 22C3 IHC pharmDx™ assay (Agilent Technologies/Dako, Carpinteria, CA, USA) was performed on the Dako Autostainer Link 48 according to the manufacturer's protocol. PD-L1 tumor proportion score (TPS) values were determined independently by 3 observers including an expert pulmonary pathologist and two cytopathologists (M.S.T., J.S., and H.M.K.) in a blinded fashion. The cytopathologists underwent formal PD-L1 interpretation training (sponsored by Merck Canada Inc., Kirkland, QC, Canada) prior to study commencement. Selected cases (e.g., TPS values close to cutoff points) underwent multi-header consensus review after scoring of 43 CB-LTR pairs was completed ($n = 5/43$, 12%), and an interim statistical analysis was performed at this point before completion of the entire 81 case cohort. No consensus review was performed upon completion of the cohort. A "final" TPS for subsequent comparisons between CB and LTR specimens was generated (final TPS = expert pulmonary pathologist TPS + closest cytopathologist TPS/2). The CB tumor cellularity was assessed independently by the two cytopathologists and categorized into three groups: 100–200, 200–500, and >500 cells.

Preparation of Control Tissues and Dual-Processed Clinical Specimens

Fresh tonsillar and placental tissues were minced (approx. 3–5 mm fragments) using a scalpel blade, subjected to different pre-analytical conditions (NBF fixation, 24 h CF, and delayed NBF fixation after storage in NS), and processed as CBs. Clinical NSCLC specimens were prospectively collected (June 2018 to September 2019) and consisted of 2 specimen types: (1) pleural fluids (PLFL) submitted for CF (for 1–7 h followed by at least 4 h NBF) and immediate NBF fixation prior to CB preparation and (2) EBUS-TBNA specimens with separate needle passes from one single anatomical site placed in CytoLyt® and NBF at the time of the procedure. The EBUS-TBNA specimens remained in the respective fixative until arrival at the laboratory (same day or up to multiple days on a weekend). Specimens with less than 100 morphologically identifiable tumor cells per section were excluded. Specimens clinically reported as positive (TPS $\geq 1\%$), deemed borderline at the $\geq 1\%$ cutoff (any fixative) upon review, and a subset of negative cases (38/52; 73%) were rescored (J.S. and H.M.K.; blinded to both sample pairs and fixative used). The remaining cases were re-

Table 1. Comparison of PD-L1 TPS categories from CBs and LTRs

	LTR specimens			<i>n</i> (%)
	<1%	1–49%	≥50%	
<i>CB specimens</i>				
<1%	34	8	1	43 (53.1)
1–49%	13	5	5	23 (28.4)
≥50%	0	3	12	15 (18.5)
<i>n</i> (%)	47 (58.0)	16 (19.8)	18 (22.2)	81

PD-L1, programmed death ligand-1; TPS, tumor proportion score; CB, cell block; LTR, lung tumor resection.

viewed to ensure the absence of PD-L1 tumor cell reactivity. Cases with discrepant TPS results (at the clinically relevant cutoff points ≥ 50 and $\geq 1\%$) after rescoring underwent multi-header consensus review by the two raters including formal cell count (if required) to resolve the discrepancy. The mean TPS of the two raters was used as "final" TPS for all other cases.

Computational Image Analysis

Digital image analysis was performed with control tissues and a subset of dual-processed PLFL specimens ($n = 12$: 4 cases with TPS <1%, 4 cases with TPS 1–49%, and 4 cases with TPS $\geq 50\%$). Image analysis was limited to PLFL in order to avoid potential sampling bias, which may affect results in EBUS-TBNA specimens. Slides were scanned at $\times 200$ (Aperio Scanscope AT2 Whole Slide Scanner; Leica Biosystems Inc., Nussloch, Germany). The region of interest (ROI) (i.e., deep crypt reticulated epithelium in tonsillar tissue, syncytiotrophoblast in placental tissue, and neoplastic cells in the PLFL) was digitally annotated using HALO digital image analysis software version 2.0 (Indica Labs, Corrales, NM, USA). A supervised machine learning algorithm (random forest classifier), cell recognition, and nuclear segmentation were trained to recognize and optimized on the cells within the ROI. Color deconvolution for the nuclear counterstain (tonsils: RGB 0.293, 0.313, 0.166; placenta: RGB 0.369, 0.385, 0.251; and PLFL: 0.503, 0.487, 0.279) and the diaminobenzidine reaction product (tonsils: RGB 0.600, 0.915, 1.064; placenta: RGB 0.587, 0.939, 1.064; and PLFL: 0.346, 0.543, 0.699) was set to optimize for ROI cell recognition. A combination of exclusion annotation and random forest classifier was used to exclude surface epithelium of the tonsillar slides and inflammatory and stromal cells. PD-L1 IHC was quantified using the HALO image analysis Membrane v1.4 algorithm using the following metrics: percentage of positive staining (PP: strong, moderate, and weak) and H-score (HS). HS evaluates the proportion of strongly, moderately, and weakly staining cells with a range of 0–300 ($HS = [3 \times \% \text{ strongly staining cells}] + [2 \times \% \text{ moderately staining cells}] + [1 \times \% \text{ weakly staining cells}]$). All annotations and image analysis markups were assessed visually to verify the performance and accuracy. A mean of 4,264 crypt epithelial cells (median: 3,676; range: 1,230–11,037) and a mean of 139,536 syncytiotrophoblast nuclei (median: 133,721; range: 82,372–224,585) were scored per each tonsillar and placental slide, respectively. For PLFL specimens, a mean of 9,873 malignant cells (median: 6,055; range: 169–55,962) was scored per slide.

Table 2. Test parameters for PD-L1 22C3 IHC pharmDx™ in CBs compared to LTRs

TPS cutoff	Total cohort (n = 81)		Before interim analysis (n = 43)		After interim analysis (n = 38)	
	≥50%	≥1%	≥50%	≥1%	≥50%	≥1%
Sensitivity	66.7 (41.0, 86.7)	73.5 (55.6, 87.1)	77.8 (40.0, 97.2)	73.7 (48.8, 90.0)	55.6 (21.2, 86.3)	73.7 (44.9, 92.2)
Specificity	95.2 (86.7, 99.0)	72.3 (57.4, 84.4)	91.2 (76.3, 98.1)	70.8 (48.9, 87.4)	100.0 (88.1, 100)	73.9 (51.6, 89.8)
Accuracy	88.9 (80.0, 94.8)	72.8 (61.8, 82.1)	88.4 (74.9, 96.1)	72.1 (56.3, 84.7)	89.5 (75.2, 97.1)	73.7 (56.9, 86.6)
PPV	80.0 (51.9, 95.7)	65.8 (48.6, 80.4)	70.0 (34.8, 93.3)	66.7 (43.0, 85.4)	100.0 (47.8, 100)	64.7 (38.3, 85.8)
NPV	90.9 (81.3, 96.6)	79.1 (64.0, 90.0)	93.9 (79.8, 99.3)	77.3 (54.6, 92.2)	87.9 (71.8, 96.6)	81.0 (58.1, 94.6)
FPR	4.8 (1.0, 13.3)	27.7 (15.6, 42.6)	8.8 (1.9, 23.7)	29.2 (12.6, 51.1)	0.0 (0, 11.9)	26.1 (10.2, 48.4)
FNR	33.3 (13.3, 59.0)	26.5 (12.9, 44.4)	22.2 (2.8, 60.0)	26.3 (9.1, 51.2)	44.4 (13.7, 78.8)	26.7 (7.8, 55.1)

All values are in percentage (%) and 95% confidence intervals are within parentheses. PD-L1, programmed death ligand-1; IHC, immunohistochemistry; CB, cell block; LTR, lung tumor resection; FNR, false-negative rate; FPR, false-positive rate; NPV, negative predictive value; PPV, positive predictive value; TPS, tumor proportion score.

Table 3. Comparing continuous PD-L1 TPS between LTRs and CBs

	Total cohort	Before interim analysis	After interim analysis
Cases, n	81	43	38
ICC	0.84 (0.76, 0.90)	0.79 (0.64, 0.88)	0.92 (0.85, 0.96)
Pearson’s correlation coefficient	0.84	0.79	0.92
Kappa square weights	0.68	0.66	0.70

95% confidence intervals are within parentheses. PD-L1, programmed death ligand-1; TPS, tumor proportion score; CB, cell block; LTR, lung tumor resection; ICC, intraclass correlation coefficient.

Statistical Analysis
Software used for statistical calculations was R version 3.5.3. Sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), false-positive rate (FPR), and false-negative rate (FNR) were calculated for the TPS cutoff values ≥50% and ≥1%. Intraclass correlation coefficient (ICC), Pearson’s correlation coefficient (R), and kappa (κ) values were calculated for TPS raw value comparisons using standard methods.

Results

CB-LTR Concordance of PD-L1 22C3 IHC pharmDx™
Eighty-one CB-LTR pairs from 45 female and 36 male patients were available for final analysis. Seventy (86%) cases were diagnosed as adenocarcinoma, 10 (12%) as squamous cell carcinoma, and 1 (1%) as adenosquamous carcinoma. The median patient age was 71 years (range: 42–89 years). The median interval between FNA and tumor resection was 1 month (range:

0–4 months). None of the patients had interval treatment. At the ≥50/≥1% TPS cutoff, 15/38 CBs (18.5/46.9%) and 18/34 LTRs (22.2/42.0%) had a “positive” final TPS (Table 1 and see online suppl. Table 1; see www.karger.com/doi/10.1159/000508628 for all online suppl. material). There was agreement in 51/81 (63%) cases. Accuracy, sensitivity, specificity, PPV, NPV, FPR, and FNR were 88.9/72.8, 66.7/73.5, 95.2/72.3, 80.0/65.8, 90.9/79.1, 4.8/27.7, and 33.3/26.5%, respectively, for the clinically relevant ≥50/≥1% cutoff values (Table 2). Thirteen of 81 cases were assessed as TPS 1–49% in the CB and as TPS <1% in the LTR. An example of a false-positive result at the ≥50% cutoff is illustrated in Figure 1a, b. The only 2-step discordance (case 43) was likely caused by intratumoral heterogeneity in PD-L1 expression (Fig. 1c, d). The ICC between CB- and LTR-derived TPS values was 0.84 (95% confidence interval [CI]: 0.76, 0.90). The square weighted κ was 0.68 (moderate agreement).

Fig. 1. a–d Examples of discordant PD-L1 TPS in LTRs and CBs. False-positive case (case 39): LTR with TPS of 10% ($\times 100$) (**a**) and corresponding CB with TPS of 85% ($\times 630$) (**b**). False-negative case (case 43): LTR with TPS of 55% ($\times 100$) (**c**) and corresponding CB with TPS of 0% ($\times 630$) (**d**). Asterisks indicate tumor areas of PD-L1 expression. Arrows indicate tumor areas without PD-L1. PD-L1, programmed death ligand-1; TPS, tumor proportion score; LTR, lung tumor resection; CB, cell block.

Impact of Interim Analysis, Observer Specialty, Tumor Cellularity, and Specimen Interval

Reliability improved after interim analysis performed with completion of the first 43 pairs (ICC: 0.79 before vs. 0.92 after; squared weighted κ : 0.66 before vs. 0.70 after) (Table 3). Correlation between cytopathologists and the pulmonary pathologist was high for both the TPS values derived from CB and LTR (Pearson R : 0.93–0.97). Individual agreement improved somewhat with interim analysis (0.91–0.97 before vs. 0.96–0.98 after). The weighted κ for CB tumor cellularity estimates showed

only moderate agreement between two raters ($\kappa = 0.67$) perhaps due to difficulty in distinction between tumor cells and admixed non-neoplastic cells. Cellularity appeared to impact PD-L1 agreement between raters at least at either end of the spectrum, with low cellularity linked to diminished agreement for continuous TPS values and high cellularity linked to improved agreement (online suppl. Table 2). No clear impact on the TPS was seen for the time interval between CB and LTR (online suppl. Fig. 1).

Fig. 2. a–d PD-L1 IHC in FF versus CF tonsillar and placental tissue. Tonsillar (**a, b**) and placental control tissues (**c, d**) after direct formalin fixation (**a, c**) and CF (**b, d**) show similar PD-L1 immunohistochemical reactivity with minimal differences in staining quality ($\times 200$). PD-L1, programmed death ligand-1; IHC, immunohistochemistry; FF, formalin-fixed; CF, CytoLyt®-prefixed.

Assessment of Preanalytical Variables with Control Tissues

CF prior to CB preparation and short delays in formalin fixation due to transient storage of needle rinse material in NS solution (usually within the range of minutes) represent preanalytical variables of potential significance in our practice environment. The impact of these two variables was first examined in control tissues (tonsil and placenta), which revealed minimal variances in staining intensity and distribution of uncertain clinical impact (Fig. 2). Image analysis revealed a slight decrease in PP (51.66 vs. 40.41) and HS (144.23 vs. 114.31) with CF compared to formalin fixation in placental tissue. Although

cold ischemia time could not be controlled due to logistical reasons, there were no observable or measurable differences with delayed formalin fixation after prolonged storage in NS at room temperature (online suppl. Table 3).

Technical Assay Validation with Dual-Processed Clinical CB Specimens

Fifty-two dual-processed clinical samples were identified, including 28 (54%) pleural effusions and 24 (46%) EBUS-TBNA specimens. The clinicopathologic features of the patients were as follows: 30 patients were men (58%) and 22 were women (42%), with a median age of 73 years (range: 49–90 years). Forty-four (85%) cases

Table 4. Test parameters for CF CB specimens compared to direct formalin fixation

	TPS cutoff $\geq 50\%$	TPS cutoff $\geq 1\%$
Sensitivity	95.2 (76.2, 99.9)	96.8 (83.3, 99.9)
Specificity	96.8 (83.3, 99.9)	90.5 (69.6, 98.8)
Accuracy	96.2 (86.8, 99.5)	94.2 (84.1, 98.8)
PPV	95.2 (76.2, 99.9)	93.8 (79.2, 99.2)
NPV	96.8 (83.3, 99.9)	95.0 (75.1, 99.9)
FPR	3.2 (0.1, 16.7)	9.5 (1.2, 30.4)
FNR	4.8 (0.1, 23.8)	3.2 (0.1, 16.7)

All values are in percentage (%) and 95% confidence intervals are within parentheses. CF, CytoLyt[®]-prefixed; CB, cell block; TPS, tumor proportion score; FNR, false-negative rate; FPR, false-positive rate; NPV, negative predictive value; PPV, positive predictive value.

were diagnosed as adenocarcinoma, 7 (13%) as squamous cell carcinoma, and 1 (2%) as large cell neuroendocrine carcinoma. TPS values obtained for CF and FF CBs showed a strong positive correlation (Pearson's correlation coefficient $R = 0.97$). Accuracy, sensitivity, specificity, PPV, NPV, FPR, and FNR were 96.2/94.2, 95.2/96.8, 96.8/90.5, 95.2/93.8, 96.8/95.0, 3.2/9.5, and 4.8/3.2%, respectively, for the TPS cutoff values of $\geq 50/\geq 1\%$ (Table 4). Five of 52 CF-FF pairs were discordant. Four discordances were found with EBUS-TBNA specimens and 3 were seen with TPS values close to the $\geq 1\%$ cutoff. No two-step discordances (crossing of both cutoff values) were seen (online suppl. Table 4). A subset of PLFL cases ($n = 12$; 23%) including 4 CF-FF pairs for each TPS category ($<1\%$, 1–49%, and $\geq 50\%$) were submitted to digital image analysis. The ICCs between CF and FF for the parameters PP and HS were both 0.97 (95% confidence interval: 0.91, 0.99), supporting a negligible impact of specimen prefixation with CytoLyt[®] on PD-L1 22C3 IHC pharmDx[™] assay performance at least in comparison to other confounding factors such as intra-/interobserver variability and spatial heterogeneity of marker expression (online suppl. Table 5).

Discussion

Immune checkpoint inhibitors have become part of the current standard of care in the clinical management of advanced NSCLC. The PD-L1 TPS cutoff values $\geq 50\%$ and $\geq 1\%$, as assessed by the 22C3 IHC pharmDx[™] assay, are used to select patients for first- and second-line ther-

apy with pembrolizumab, respectively. The ability to use cytology specimens for the purpose of patient allocation into the appropriate treatment category allows patients who are not surgical candidates to qualify for immune therapy while avoiding additional invasive testing. Although prior studies have reported mainly a good concordance between cytology and histology for PD-L1 expression, comparability between studies and generalizability of those results have been hampered by significant differences in study design with respect to cohort sizes, inclusion criteria, technical preparation methods, and assay type/antibody clone. Rather than performing a bulk comparison of diagnostic rates, we choose to examine a narrowly defined, homogeneous cohort of cases matched for the anatomical site of specimen origin as starting point for our assay implementation. This first step also served to facilitate an exchange of expertise between observers with different subspecialties (pulmonary pathology and cytopathology) and to establish basic performance characteristics as underpinning for clinical utilization. Next, we performed a technical validation in order to be able to offer assay results on samples submitted in CytoLyt[®], which represents the most common non-formalin fixative in our practice. Digital image analysis was employed in a subset of cases as additional means of PD-L1 IHC evaluation to alleviate concerns with respect to the subjectivity of the assay. As a result, the PD-L1 22C3 IHC pharmDx[™] was implemented in our laboratory on the basis of a good reliability of the CB-derived assay results and a negligible impact of CytoLyt[®] on the assay performance.

A number of recent studies examined the 22C3 IHC pharmDx[™] assay with CB preparations (online suppl. Table 6). The study by Skov and Skov [10] is similar to our CB-LTR comparison in terms of design; however, important differences include (1) fewer resections ($n = 60$); (2) inclusion of neuroendocrine tumors, mesotheliomas, and metastatic lesions; (3) inclusion of CBs with less than 100 tumor cells; and (4) IHC scoring performed by a single pathologist [10]. The concordance rates reported by Skov and Skov [10] were 94 and 85% at the TPS cutoffs of 50 and 1%, respectively (89 and 73% in our study). Heymann et al. [11] reported a study of 214 specimens from 188 patients, which included 40 cytology samples. Interestingly, cytology specimens were more likely to be PD-L1 positive compared to histology; however, only 23 patients had more than one specimen and only 8 of those included a cytology specimen [11]. Noll et al. [12] found good concordance between specimen types in a study of 41 cases with 22C3 IHC pharmDx[™] performed on Papa-

nicolaou-stained direct smears, CBs, and core needle biopsies. Notably, a better correlation was found between smears and core needle biopsies than CBs and core needle biopsies [12]. Capizzi et al. [13] reported a comparison between smears and histologic material. Three different antibody clones including 22C3 were examined. The cytology-histology correlation at the $\geq 1\%$ cutoff was found to be weak [13]. A similar study focused on smears as substrate was conducted by Lozano et al. [20] who used mainly CBs as reference (12/62 patients with surgical specimens only) and reported high concordance. Torous et al. [14] performed a bulk comparison with 94 CB and 138 surgical pathology specimens. The CB processing method reported in their study is similar to ours (i.e., CytoLyt[®] fixation prior to Histogel CB preparation). No statistically significant difference in the proportions of the clinically relevant groups for CBs compared to surgical pathology specimens was seen. However, the authors note that no direct conclusions on specimen comparability was possible due to a lack of paired specimens in their cohort. Of importance is their preliminary finding of similar response and disease control rates in a small subset of patients after treatment with pembrolizumab regardless of specimen type [14]. Of note, benefit from immune checkpoint inhibitor treatment regardless of specimen type has also been reported recently by other groups in abstract form [26, 27]. Based on data from a large study with 1,419 consecutive cases including 371 CB specimens, Wang et al. [15] concluded that PD-L1 IHC performs well with CBs as substrate when TPS $\geq 50\%$ was used as endpoint. Fixation with alcohol did not appear to affect the high PD-L1 expression rate (42% formalin only vs. 40% combined methanol/alcohol and formalin) [15]. Ilie et al. [16] compared bronchial biopsies with bronchial wash- and pleural effusion-derived CBs and reported a high concordance regardless of TPS cutoff point ($\geq 1\%$ and $\geq 50\%$). Xu et al. [17] found a moderate agreement for a cohort of 52 paired specimens from the same patients. However, agreement was only fair for the subset of squamous cell carcinoma ($n = 29$) and categories of positivity (1–49%, $\geq 50\%$) were not separately reported [17]. Sakata et al. [18] demonstrated high concordance rates of 82 and 87% at $\geq 50\%$ and $\geq 1\%$ comparing EBUS-TBNA with resected tumor specimens but noted a decrease in sensitivity and PPV at the $\geq 50\%$ cutoff. Wang et al. [22] examined a large EBUS-TBNA cohort including a subset of paired surgical specimens (resections and biopsies), which resulted in 91.3% concordance. The authors of this study noted that a large sample size of matched cytology and surgical specimens is needed for

more solid conclusions [22]. Hernandez et al. [19] demonstrated 67% (35/52) overall agreement in a cohort of paired cases. Agreement was higher with paired excisional samples (12/14 [86%]) than small biopsies (23/38 [61%]), and low CB cellularity (< 100 cells) was linked to a diminished agreement with surgical specimens ($\kappa = 0.19$ vs. 0.63) [19]. Mei et al. [21], in a study focused on clinicopathologic and molecular features, reported similar PD-L1 positive rates (56.0 vs. 52.1%) and high PD-L1 expression rates (31.0 vs. 28.5%) in cytology and surgical specimens.

Our study represents the largest and most homogeneous cohort of anatomical site-matched cytology-histology specimen pairs to date and adds to the growing body of evidence supporting the clinical value of cytology in PD-L1 IHC evaluation. The main sources of disagreement between specimen types appear to be (1) intratumoral heterogeneity, (2) difficulty in distinction of rare positive tumor cells from mimics, and (3) more broadly defined observer-related variability/subjectivity of IHC assessment (e.g., absence or presence of weak or partial membrane staining, differentiation of membrane vs. cytoplasmic staining, and PD-L1 expression close to the cutoff points). Intratumoral heterogeneity is expectedly of greater concern if CB cellularity is low as demonstrated by Hernandez et al. [19] and confirmed by our own data. Munari et al. [28] suggested a minimum of 4 biopsies to minimize the risk of misclassification due to intratumoral heterogeneity. Disagreement found with a substantial proportion of CB cases from the 1–49% group in our cohort, which were negative (TPS $< 1\%$) upon evaluation of the LTR specimen, points to the difficulty of differentiating rare genuinely PD-L1-reactive tumor cells from mimics (e.g., macrophages or degenerated/necrotic cells) in a context where architectural information is diminished. Practical day-to-day experience and prior literature support that, especially in cases where tumor cell PD-L1 expression is low and strong membranous reactivity is present in tumor-infiltrating immune cells, false-positive results may occur [29]. The evaluation of additional markers on adjacent sections (e.g., p40 and TTF1) or as double stain on the same slide, while not a universal remedy and not tested as part of this study, can serve to mitigate the problem in these instances. Importantly, small histologic specimens such as core needle biopsies, although not included in our study, carry a disagreement rate, which is likely similar to CBs. Although differences in study design prevent direct comparisons, data from our own institution with a concordance rate for biopsies close to that reported here for CB specimens appear to

support this conclusion [30]. Interestingly, biopsies tended to underestimate the PD-L1 expression in this study. With respect to observer-related variability/subjectivity of PD-L1 evaluation, digital image analysis may be a potential solution. Based on our preliminary evaluation, there appears to be a high correlation between manual and digital assessment. On a practical level, independent assessment of cases by a second observer may provide a degree of quality control especially if PD-L1 scoring is perceived as challenging from the morphological perspective or whenever the initial TPS value falls close to a clinically relevant cutoff point. An opportunistic or more systematic form of intradepartmental case review may be helpful in these situations. In addition, flow cytometry may be a potential method of automated, objective evaluation for cell suspensions (e.g., pleural effusions) [31].

Aside from analytical problems, preanalytical factors such as the frequent use of non-formalin fixatives with cytology samples pose challenges with IHC testing [32]. Technical validation of a satisfactory test performance is therefore required at the level of each laboratory [33]. A study reported as abstract cautioned against the use of CytoLyt® [23]. In contrast, a study using PD-L1 expressing cell lines and E1L3N antibody showed no difference with CytoRich Red [25]. The largest study to date was conducted by Gosney et al. [24] who examined 50 pairs (12 CytoLyt® and 38 CytoRich Red) of dual-processed specimens in a design similar to ours and, importantly, used the 22C3 IHC pharmDx™ assay. No difference was seen between alcohol-based fixatives and formalin; however, slides were assessed in pairs. In our study, we found high concordance (>90%) using a clinically representative cohort of dual-processed (CytoLyt® only) clinical samples from the same source/anatomical site after blinded rescoring. Interestingly, 4/5 of the discrepant cases were EBUS-FNA specimens, which raises the possibility of a sampling effect (due to distinct areas targeted with successive FNA passes), whereas PLFL tend to be homogeneous. Tonsillar and placental tissues, known to express PD-L1 [34, 35], were included in our study to assess CF and potential effects of delayed fixation. Although we were unable to control the cold ischemia time, storage in NS did not appear to have any effect on PD-L1 IHC.

Our study differs from other reports with respect to a relatively low proportion of positive cases in the cytology-histology comparison set. Wang et al. [22] reported a low positivity rate in primary lung tumors (25.8%, TPS ≥ 50%), which is close to the one observed by Garon et al. [4] (23.2%) for the KEYNOTE-001 trial and ours (22.2%). Similarly, a 2.3-fold higher PD-L1 expression at the TPS

≥ 50% was reported in metastatic than in early/locally advanced stage cancer, which may explain our findings [15]. A different potential cause for the low rate of positivity in our cohort includes the use of samples after prolonged storage, for which contradictory results have been published [15, 36]. Regardless of cause, these findings support the role of repeat testing in the setting of recurrent lung cancer. Another limitation of our study is the lack of clinical data with respect to PD-L1/PD-1 inhibition therapy. To date, only few retrospective studies have begun to explore this area [14, 26, 27]. Rather than retrospective data collection, the up-front inclusion of cytology specimens in biomarker studies requires further advocacy efforts. Also, we only examined the 22C3 IHC pharmDx™ assay; however, harmonization studies may pave the way for a broader application of individual assays and may ease the burden put on single laboratories trying to respond to clinical demands [37–39].

In summary, we show good concordance of PD-L1 22C3 IHC pharmDx™ results between cytology and site-matched resection specimens, and a negligible impact of CF on the performance of this assay compared to direct formalin fixation. We used digital image analysis as previously advocated for by others [29] to substantiate our results. Future research requires the assessment of clinical benefit derived from PD-L1/PD-1 inhibition after PD-L1 IHC performed with cytology specimens (i.e., clinical validation). In addition, more objective and automated means of PD-L1 testing should be explored.

Acknowledgements

The authors would like to express their gratitude to the staff of the cytopathology and immunohistochemistry laboratories for their support.

Statement of Ethics

The study was reviewed and approved by the University Health Network Research Ethics Board (#18-5680).

Disclosure Statement

J.S. has accepted honoraria from Olympus and Merck within the past 3 years. M.S.T. has received research grant and/or honoraria from Merck, AstraZeneca, Bristol-Myers Squibb, and Hoffmann La Roche/Ventana unrelated to this study. No conflicts of interest were reported by the other authors.

Funding Sources

The study received intramural funding from the Division of Pathology, University Health Network, Toronto, ON, Canada.

Author Contributions

All authors made substantial contributions to the conception or design of the work, or to the acquisition, analysis, or interpretation of data for the work; participated in drafting the work or revising it critically for important intellectual content; approved the final version to be published; and agreed to be accountable for all aspects of the work.

References

- 1 Siegel RL, Fedewa SA, Miller KD. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;65(6):457–80.
- 2 Canadian Cancer Statistics Advisory Committee. *Canadian Cancer Statistics 2019*. Toronto, ON: Canadian Cancer Society; 2019 [accessed 2020 Feb 3]. Available from: cancer.ca/Canadian-Cancer-Statistics-2019-EN.
- 3 Travis D, Brambilla E, Burke AP, Marx A, Nicholson AG. *WHO classification of tumours of the lung, pleura, thymus and heart*. 4th ed. Lyon: IARC; 2015.
- 4 Garon EB, Rizvi NA, Hui R, Leigh N, Balmanoukian AS, Eder JP, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. *N Engl J Med*. 2015;372(21):2018–28.
- 5 Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, et al. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N Engl J Med*. 2015;373(17):1627–39.
- 6 Herbst RS, Baas P, Kim DW, Felip E, Pérez-Gracia JL, Han JY, 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(10027):1540–50.
- 7 Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csósz T, Fülöp A, et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N Engl J Med*. 2016;375(19):1823–33.
- 8 Zhou F, Moreira AL. Lung carcinoma predictive biomarker testing by immunoperoxidase stains in cytology and small biopsy specimens: advantages and limitations. *Arch Pathol Lab Med*. 2016;140(12):1331–7.
- 9 Mino-Kenudson M. Programmed death-ligand 1 immunohistochemistry testing for non-small cell lung cancer in practice. *Cancer*. 2017;125(7):521–8.
- 10 Skov BG, Skov T. Paired comparison of PD-L1 expression on cytologic and histologic specimens From malignancies in the lung assessed with PD-L1 IHC 28-8pharmDx and PD-L1 IHC 22C3pharmDx. *Appl Immunohistochem Mol Morphol*. 2017;25(7):453–9.
- 11 Heymann JJ, Bulman WA, Swinarski D, Pagan CA, Crapanzano JP, Haghighi M, et al. Programmed death-ligand 1 expression in non-small cell lung carcinoma: comparison among cytology, small biopsy, and surgical resection specimens. *Cancer*. 2017;125(12):896–907.
- 12 Noll B, Wang WL, Gong Y, Zhao J, Kalhor N, Prieto V, et al. Programmed death ligand 1 testing in non-small cell lung carcinoma cytology cell block and aspirate smear preparations. *Cancer Cytopathol*. 2018;126(5):342–52.
- 13 Capizzi E, Ricci C, Giunchi F, Zagnoni S, Ceccarelli C, Gómez BUÁ, et al. Validation of the immunohistochemical expression of programmed death ligand 1 (PD-L1) on cytological smears in advanced non small cell lung cancer. *Lung Cancer*. 2018;126:9–14.
- 14 Torous VF, Rangachari D, Gallant BP, Shea M, Costa DB, VanderLaan PA. PD-L1 testing using the clone 22C3 pharmDx kit for selection of patients with non-small cell lung cancer to receive immune checkpoint inhibitor therapy: are cytology cell blocks a viable option? *J Am Soc Cytopathol*. 2018;7(3):133–41.
- 15 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(6):1417–22.
- 16 Ilie M, Juco J, Huang L, Hofman V, Khambata-Ford S, Hofman P. Use of the 22C3 anti-programmed death-ligand 1 antibody to determine programmed death-ligand 1 expression in cytology samples obtained from non-small cell lung cancer patients. *Cancer Cytopathol*. 2018;126(4):264–74.
- 17 Xu H, Bratton L, Nead M, Russell D, Zhou Z. Comparison of programmed death-ligand 1 (PD-L1) immunostain for nonsmall cell lung carcinoma between paired cytological and surgical specimens. *Cytojournal*. 2018;15:29.
- 18 Sakata KK, Midthun DE, Mullan JJ, Kern RM, Nelson DR, Edell ES, et al. Comparison of programmed death ligand-1 immunohistochemical staining between endobronchial ultrasound transbronchial needle aspiration and resected lung cancer specimens. *Chest*. 2018;154(4):827–37.
- 19 Hernandez A, Brandler TC, Zhou F, Moreira AL, Schatz-Siemers N, Simsir A. Assessment of programmed death-ligand 1 (PD-L1) immunohistochemical expression on cytology specimens in non-small cell lung carcinoma. *Am J Clin Pathol*. 2019;151(4):403–15.
- 20 Lozano MD, Abengozar-Muela M, Echeveste JJ, Subtil JC, Bertó J, Gúrpide A, et al. Programmed death-ligand 1 expression on direct Pap-stained cytology smears from non-small cell lung cancer: comparison with cell blocks and surgical resection specimens. *Cancer Cytopathol*. 2019;127(7):470–80.
- 21 Mei P, Shilo K, Wei L, Shen R, Tonkovich D, Li Z. Programmed cell death ligand 1 expression in cytologic and surgical non-small cell lung carcinoma specimens from a single institution: association with clinicopathologic features and molecular alterations. *Cancer Cytopathol*. 2019;127(7):447–57.
- 22 Wang G, Ionescu DN, Lee CH, Hiruki T, Myers R, Shaipanich T, et al. PD-L1 testing on the EBUS-FNA cytology specimens of non-small cell lung cancer. *Lung Cancer*. 2019;136:1–5.
- 23 Alex D, Buonocore D, Rekhman N, Jungbluth A, Frosina D, Hellmann M, et al. Loss of PD-L1 immunoreactivity in CytolTyt-fixed cell blocks. *J Am Soc Cytopathology*. 2017;6(5):S44.
- 24 Gosney JR, Haragan A, Chadwick C, Giles TE, Grundy S, Tippet V, et al. Programmed death ligand 1 expression in EBUS aspirates of non-small cell lung cancer: is interpretation affected by type of fixation? *Cancer Cytopathol*. 2020;128(2):100–6.
- 25 Pejchal M, Hutchinson L, Dresser K, Woda B, Fischer A. Alcohol and formalin fixation give identical PD-L1 testing results: a study using cell lines. *Lab Invest*. 2018;98:167.
- 26 Wang Y, Butcher M, Naqvi A, Cutz J, Juegens R. P2.09-08 clinical outcomes of histology versus cytology PD-L1 22C3 antibody testing in advanced non-small cell lung cancer. *J Thorac Oncol*. 2018;13(10):S764.
- 27 Guo K, Kasymjanova G, Wang H, Sakr L, Small D, Cohen V, et al. P1.04-16 comparison of clinical response to checkpoint inhibitors in advanced NSCLC with high PD-L1 expression tested on cytology versus biopsy samples. *J Thorac Oncol*. 2018;13(10):S531.
- 28 Munari E, Zamboni G, Marconi M, Sommaggio M, Brunelli M, Martignoni G, et al. PD-L1 expression heterogeneity in non-small cell lung cancer: evaluation of small biopsies reliability. *Oncotarget*. 2017;8(52):90123–31.
- 29 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–S4.
- 30 Albaqer T, Santiago R, Leung Y, Pal P, Khan Z, Torlakovic E, et al. P1.07-022 routine PD-L1 immunohistochemistry testing by 22C3 in a Canadian reference testing center. *J Thorac Oncol*. 2017;12(11):S2004.

- 31 Yoon J-Y, Schwock J, Nayyar R, Pal P, Quest G, Tsao MS, et al. PD-L1 lineage-specific quantification in malignant pleural effusions of lung adenocarcinoma by flow cytometry [1892]. *Mod Pathol*. 2019;32(Suppl 2):71.
- 32 Pillappa R, Kraft AO. Immunohistochemical validation studies in effusion cytology: a cautionary tale. *Cancer Cytopathol*. 2019;127(11):680–3.
- 33 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(10):699–714.
- 34 Lyford-Pike S, Peng S, Young GD, Taube JM, Westra WH, Akpeng B, et al. Evidence for a role of the PD-1:PD-L1 pathway in immune resistance of HPV-associated head and neck squamous cell carcinoma. *Cancer Res*. 2013;73(6):1733–41.
- 35 Veras E, Kurman RJ, Wang TL, Shih IM. PD-L1 expression in human placentas and gestational trophoblastic diseases. *Int J Gynecol Pathol*. 2017;36(2):146–53.
- 36 Midha A, Sharpe A, Scott M, Walker J, Shi K, Ballas M, et al. PD-L1 expression in advanced NSCLC: primary lesions versus metastatic sites and impact of sample age. *J Clin Oncol*. 2016;34(15_Suppl 1):3025.
- 37 Scheel AH, Dietel M, Heukamp LC, Jöhrens K, Kirchner T, Reu S, et al. Harmonized PD-L1 immunohistochemistry for pulmonary squamous-cell and adenocarcinomas. *Mod Pathol*. 2016;29(10):1165–72.
- 38 Hirsch FR, McElhinny A, Stanforth D, Ranger-Moore J, Jansson M, Kulangara K, 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(2):208–22.
- 39 Rimm DL, Han G, Taube JM, Yi ES, Bridge JA, Flieder DB, 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):1051–8.