Human Pathology (2019) **90**, 54–59

ELSEVIER

Human PATHOLOGY

www.elsevier.com/locate/humpath

Original contribution

PD-L1 expression in non–small cell lung cancer: evaluation of the diagnostic accuracy of a laboratory-developed test using clone E1L3N in comparison with 22C3 and SP263 assays☆,☆☆

Enrico Munari MD, PhD[a,*], Giuseppe Zamboni MD[a,b], Gianluigi Lunardi PharmD[c], Marcella Marconi BS[a], Matteo Brunelli MD, PhD[b], Guido Martignoni MD[d], George J. Netto MD[e], Linda Quatrini PhD[f], Paola Vacca PhD[f], Lorenzo Moretta MD[f], Giuseppe Bogina MD[a]

[a]*Department of Pathology, IRCCS Sacro Cuore Don Calabria, 37024 Negrar (VR), Italy*
[b]*Department of Diagnostics and Public Health, University of Verona, 37134 Verona, Italy*
[c]*Department of Oncology, IRCCS Sacro Cuore Don Calabria, 37024 Negrar (VR), Italy*
[d]*Department of Pathology, Pederzoli Hospital, 37019 Peschiera del Garda (VR), Italy*
[e]*Department of Pathology, University of Alabama at Birmingham, Birmingham, AL, 35233-7331, USA*
[f]*Immunology Area, IRCCS Bambino Gesù Pediatric Hospital, 00146 Rome, Italy*

Received 16 March 2019; revised 14 May 2019; accepted 15 May 2019

Keywords:
PD-L1;
Comparison;
E1L3N;
SP263;
22C3;
Lung cancer

Summary Different studies have evaluated the comparability of various immunohistochemical assays for PD-L1 expression evaluation, with contrasting results. Besides the important issues related to analytic performance and comparability of validated assays, not all platforms are available in all laboratories; moreover, standardized assays are very expensive, and funding for PD-L1 testing is hard to obtain, especially in the research setting. One of the most widely used and inexpensive PD-L1 clones is E1L3N (Cell Signaling Technology, Danvers, MA), which is labeled for research use only. In this work, we wanted to further study and validate in a larger cohort the analytical performance of E1L3N clone on Ventana platform (Ventana Medical Systems, Tucson, AZ) and its comparability with assays SP263 and 22C3 run onto their dedicated platforms. Serial sections of tissue microarrays built from 165 cases of resected lung cancer were stained for E1L3N onto Ventana platform following a previously reported protocol and for 22C3 and SP263 assays onto their respective platforms following manufacturer's instructions. Overall, we found very high concordance when comparing E1L3N with SP263 at both 1% and 50% cutoffs. Lower concordance was found between E1L3N and 22C3 at both cutoffs; however, 100% sensitivity was found for E1L3N compared with both SP263 and 22C3 at 50% cutoff. Given the 100% sensitivity at 50% cutoff demonstrated by E1L3N in comparison with both SP263

☆ Competing interests: The authors declare no conflicts of interest.

☆☆ Funding/Support: This study was partially supported by the Italian Association for Cancer Research (Milan, Italy) grants IG 2014 15283 and IG 2017 19920 (to Lorenzo Moretta).

* Corresponding author at: Anatomic Pathology, IRCCS Sacro Cuore Don Calabria Hospital, Via Sempreboni 5, 37024 Negrar di Valpolicella (VR), Italy.
E-mail address: enrico_munari@yahoo.it (E. Munari).

https://doi.org/10.1016/j.humpath.2019.05.003

and 22C3 and therefore the lack of false-negative cases, we propose an algorithm for PD-L1 testing in NSCLC when considering pembrolizumab as first-line therapy.

1. Introduction

Pembrolizumab is an anti-PD1 humanized monoclonal antibody that was granted Food and Drug Administration approval for the treatment of advanced lung adenocarcinoma or squamous carcinoma on the basis of PD-L1 expression on viable tumor cells, assessed with a validated assay [1-3].

Currently, pembrolizumab is the only PD1/PD-L1 inhibitor for treatment of patients with non–small cell lung cancer (NSCLC) in association with a companion diagnostic assay, namely, the 22C3 PharmDx assay (Agilent, Santa Clara, CA) using the Dako Autostainer (Dako, Carpinteria, CA). Another assay, Ventana's SP263 (Ventana Medical Systems, Tucson, AZ), has been CE-marked to guide treatment decisions in NSCLC for pembrolizumab.

Different studies have evaluated the comparability of these and other immunohistochemical assays, with contrasting results [4-7]. In this regard, we recently reported that assays 22C3 and SP263 run on appropriate platforms (Dako and Ventana, respectively) show important discrepancies in identifying PD-L1–positive cases at clinically relevant cutoffs (1% and 50%), possibly leading to underestimation of patients suitable for pembrolizumab therapy [8]. Besides the important issues related to analytic performance and comparability of validated assays, other problems regarding PD-L1 immunohistochemical evaluation exist. First of all, not all platforms are available in all laboratories; moreover, standardized assays are very expensive, whereas funding for PD-L1 testing is limited, especially in the research setting.

For these reasons, evaluation of diagnostic performance of laboratory-developed tests has been strongly advocated by the pathologists' community [7].

One of the most widely used and inexpensive PD-L1 clone is E1L3N (Cell Signaling Technology, Danvers, MA), which is labeled for research use only.

Comparability studies of E1L3N with standardized assays conducted so far have reported conflicting results [9-14]. However, a recent multicenter harmonization study reported a high weighted κ concordance coefficient (0.81) between E1L3N clone run on Ventana platform and the SP263 assay (used as the reference) [15]. This study was designed to further analyze and validate, in a larger cohort, the analytical performance of E1L3N clone on Ventana platform and its comparability with assays SP263 and 22C3 run onto their dedicated platforms to assess its usefulness as a screening tool especially in the research setting and its possible applicability in the diagnostic routine.

2. Materials and methods

2.1. Study cohort

The study cohort consisted of consecutive patients with primary NSCLC who had undergone surgical resection at the IRCCS Sacro Cuore Don Calabria Hospital of Negrar, Verona (Italy), between 2003 and 2017 with available slides and paraffin-embedded tissue blocks.

None of the patients received therapy before surgery.

Tumors were classified according to the 2015 World Health Organization classification [16,17], and staging was done using the tumor-node-metastasis staging manual (eighth edition) [18]. Investigations have been conducted according to principles expressed in the Declaration of Helsinki.

2.2. Tissue microarray construction

For every case, all hematoxylin and eosin–stained slides were reviewed for diagnosis confirmation; 1 block was then selected for tissue microarray (TMA) construction. For each block, 5 cores with a diameter of 1 mm were obtained from the diverse areas of the tumor. Overall, 8 TMAs were built.

2.3. Immunohistochemistry and scoring

From each TMA block, 5-μm sections were cut and stained with 2 validated assays according to the manufacturer's instruction: staining for SP263 was performed on the Ventana Benchmark ULTRA platform, whereas staining for 22C3 was performed on the Dako Link-48 autostainer system.

Staining with E1L3N (Cell Signaling Technology, Danvers, MA) was performed according to the protocol developed for the Ventana platform by Adam et al and reported in the article's supplementary material (Table S1, protocol for E1L3N center #5) [15].

Stained sections were scanned using Ventana iScan HT and scored based on the percentage of tumor cells showing membranous positivity, irrespective of staining intensities. A 3-tiered system was then applied using the following thresholds: $<1\%$, 1%-49% and $\geq 50\%$.

PD-L1 evaluation was performed blindly by 2 pathologists (E. M. and G. B.).

For each case, the highest scoring value across the cores was used; for discordant cases, final score was defined after

Table 1 Comparison between E1L3N and assay SP263

E1L3N *vs* SP263 Ventana								
	Agreement (%)	DP	*P* [a]	Sen (%)	Spec (%)	PPV (%)	NPV (%)	AUC
TPS ≥ 1%	94.5	1.8	.5	95.6	93.7	91.7	96.8	0.95
TPS ≥ 50%	95.8	4.2	.02	100	95	78	100	0.97

Abbreviations: DP, difference in proportion; NPV, negative predictive value; PPV, positive predictive value; Sen, sensibility; Spec, specificity; TPS, tumor proportion score; AUC, area under the ROC curve.

[a] McNemar test.

Table 2 Actual number of cases for the comparison between E1L3N and SP263 according to the thresholds <1%, 1%-49%, and ≥50%

	E1L3N			
SP263	<1%	1%-49%	≥50%	Total
<1%	90	5	1	96
1%-49%	3	35	6	44
≥50%	0	0	25	25
Total	93	40	32	165

consensus between the 2 pathologists. Part of cases stained with SP263 assay had been already evaluated by the same pathologists in a previous work [19].

Cores showing a neoplastic component ≥30% were considered as adequate; therefore, cores with lower percentages of neoplastic component were excluded.

Placenta was used as external control, whereas macrophages were used as internal control to validate the adequacy of PD-L1 staining reaction.

Overall, comparison between E1L3N and assay SP263 was possible in all cases (165); comparison between E1L3N and assay 22C3 was possible in 144 cases.

2.4. Statistical analysis

Statistical analysis was carried out using Stata version 14 (StataCorp, College Station, TX). To compare the clinical performance of the assays, difference in proportions, overall percent agreement, sensitivity, specificity, positive predictive value, negative predictive value, AUC (area under the ROC curve), and Cohen κ were calculated for all comparable cases. Wilcoxon signed-rank test was used to evaluate the differences in percent cell staining.

3. Results

The cohort consisted of 165 resected NSLC cases, the majority of which were male (76%). In terms of histology, most cases were adenocarcinomas (52%), followed by squamous cell carcinomas (36%). The remaining cases (12%) were large cell carcinomas (14 cases), adenosquamous carcinomas (4 cases), 2 large cell neuroendocrine carcinomas, and 1 sarcomatoid carcinoma. Most cases were T1 (57%) and N0 (64%).

The comparison between E1L3N and assay SP263 (Ventana platform) was possible in all cases (Table 1). At 1% cutoff, overall percent agreement was 94.5%, with a difference in proportion of 1.8. We found high sensitivity and specificity (95.6% and 93.7%, respectively), with positive and negative predictive values of 91.7% and 96.8%, respectively, and AUC of 0.95. At 50% cutoff, the overall percent agreement was slightly higher (95.8%) with a difference in proportion of 4.2. We found sensitivity to be 100% and a specificity of 95%, with positive and negative predictive values of 78% and 100%, respectively, and an AUC of 0.97. Table 2 shows the actual number of cases for the comparison between E1L3N and SP263 according to the thresholds <1%, 1%-49%, and ≥50%.

Regarding the comparison between E1L3N and assay 22C3 (Dako platform), at 1% cutoff, we found an overall percent agreement of 77.1%, with a difference in proportion of 35.4. Sensitivity was 96.7% and specificity was 71.7%, with positive and negative predictive values of 48.4% and 98.8%, respectively, and an AUC of 0.84.

At 50% cutoff, we found a higher overall percent agreement (88.2), with sensitivity of 100%; specificity of 87%; positive and negative predictive values of 39% and 100%, respectively; and an AUC of 0.94 (Table 3). The actual

Table 3 Comparison between E1L3N and assay 22C3

E1L3N *vs* 22C3 Dako								
	Agreement (%)	DP	*P* [a]	Sen (%)	Spec (%)	PPV (%)	NPV (%)	AUC
TPS ≥ 1%	77.1	35.4	<.001	96.7	71.7	48.4	98.8	0.84
TPS ≥ 50%	88.2	11.8	<.001	100	87	39	100	0.94

[a] McNemar test.

Table 4 Actual number of cases for the comparison between E1L3N and 22C3 according to the thresholds <1%, 1%-49%, and ≥50%

	E1L3N			
22C3	<1%	1%-49%	≥50%	Total
<1%	81	27	5	113
1%-49%	1	7	12	20
≥50%	0	0	11	11
Total	82	34	28	144

number of cases for the comparison between E1L3N and 22C3 according to the thresholds <1%, 1%-49%, and ≥50% is shown in Table 4.

Overall, we found E1L3N and SP263 to stain tumor cells in a similar way, with strong and intense membrane staining; 22C3 stain, on the other hand, tended to be weaker and in general more difficult to interpret (Fig. 1).

Given the 100% sensitivity at 50% cutoff of E1L3N in comparison with both SP263 and 22C3 and, therefore, the lack of false-negative cases, we propose an algorithm for PD-L1 testing in NSCLC when considering pembrolizumab as first-line therapy, as depicted in Fig. 2.

Regarding interobserver agreement for the 3 clones, at 1% cutoff, agreement was 87.2%, 93.6%, and 88.1% for E1L3N, SP263, and 22C3, respectively (Cohen κ: 0.73, 0.87, and 0.71, respectively); at 50% cutoff, agreement was 95.1%, 94.2%, and 95.4% for E1L3N, SP263, and 22C3, respectively (Cohen κ: 0.82, 0.81, and 0.76, respectively).

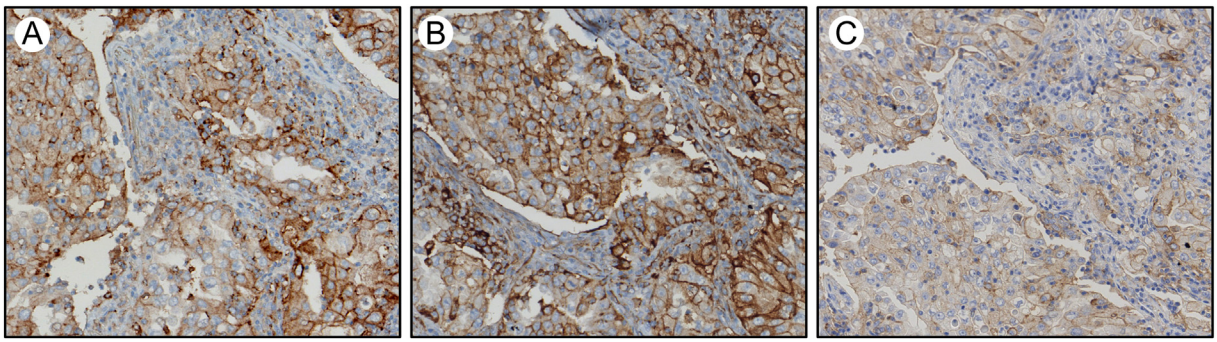

Fig. 1 Representative images of the same tissue core showing diffuse PD-L1 expression using clone E1L3N (A), SP263 (B), and 22C3 (C).

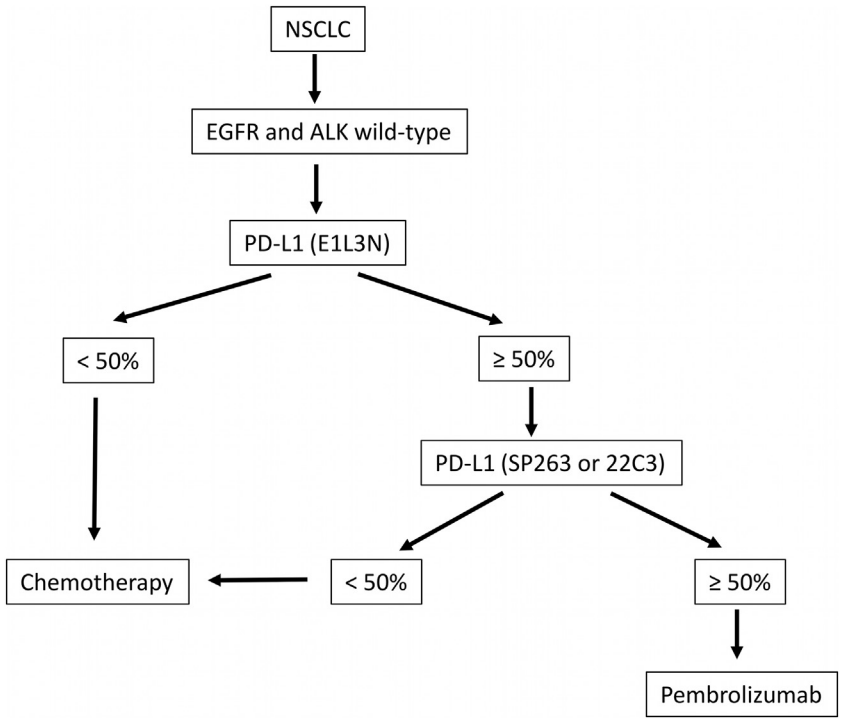

Fig. 2 Proposed algorithm for the use of PD-L1 clone E1L3N in the diagnostic setting.

4. Discussion

Higher PD-L1 expression on tumor cells generally correlates with higher rates of clinical response to treatment with PD1/PD-L1 inhibitors; however, this is not always true, as clinical responses are also observed in patients with tumors negative for PD-L1 [1,20]. It is reasonable to think that such discrepancy may be at least in part explained by the fact that a number of tumors could be misclassified, in terms of PD-L1 expression, due to different factors including expression heterogeneity in the tumors [19,21], interclone differences, and inter-/intraobserver variability. In this regard, we and others demonstrated striking discordant staining results across different cores from the same tumor, as well as important differences between validated assays (SP263 and 22C3) [6,8]. Moreover, standardized assays have been optimized to be run on appropriate platforms, but not all platforms are available in all laboratories. Finally, such assays are expensive and therefore difficult to implement, especially in the research setting.

These limitations can hamper the predictive potential of PD-L1 immunohistochemical evaluation to the point that other parameters are being explored (eg, tumor mutational burden). However, it is reasonable to believe that the predictive potential of PD-L1 immunohistochemistry may be significantly improved by reducing confounding variables, for instance, through the harmonization of PD-L1 assessment across different diagnostic materials (cytology versus diagnostic biopsies versus surgical specimens) [22], assays, and platforms to overcome existing problems related to expression heterogeneity within tumors and substantial differences existing between PD-L1 clones. In particular, for the latter issue, it is necessary to define protocols across different platforms that could improve the analytical performance of different PD-L1 clones to ameliorate their comparability. In this regard, efforts have been made to tune protocols to render different PD-L1 assays usable on nondedicated platforms.

The French harmonization study evaluated 5 anti–PD-L1 monoclonal antibodies (28-8, 22C3, E1L3N, SP142, and SP263) on 41 NSCLC surgical specimens in 7 centers using different platforms (Dako Autostainer Link 48 [Dako, Carpinteria, CA], Ventana BenchMark Ultra [Ventana Medical Systems, Tucson, AZ], and Leica Bond [Leica, Wetzlar, Germany]). Among the laboratory-developed tests that the authors developed, clone E1L3N is one of the most widely used and affordable research-use-only antibodies, largely used especially in the research setting. For this antibody, the authors reported high concordance with standardized assay SP263 (considered as the reference), with the highest weighted κ value using the Ventana platform with a specific protocol [15]. However, a limitation of this study, as reported by the authors, is that the cases were selected to equally represent the 3 most significant categories of PD-L1 expression in NSCLC (<1%, 1%-49%, and ≥50%) which might not be the case encountered in real practice, where the negative cases tend to be much more frequent.

Therefore, given the need for further validation studies, we decided to assess the concordance of E1L3N on Ventana with assays SP263 and 22C3 run onto their dedicated platforms (Ventana and Dako, respectively) and using the protocol reported by Adam et al for E1L3N [15].

Overall, we found very high concordance when comparing E1L3N with SP263 at both 1% and 50% cutoffs. Lower concordance was found between E1L3N and 22C3 at both cutoffs; however, 100% sensitivity was found for E1L3N compared with both SP263 and 22C3 at 50% cutoff.

A few previous studies evaluated the concordance of E1L3N with SP263 or 22C3 in smaller cohorts.

Among these, Conde et al evaluated E1L3N on the Ventana platform and compared results with assay SP263 in a total of 69 NSCLC cases using whole sections and found similar positivity rates at 50% cutoff but not at 1% cutoff, where difference in proportions were −12.5 and −17.3 in the discovery and validation cohorts, respectively [9].

In the work by Kim et al on 97 NSCLC cases, E1L3N and SP263 showed overall percent agreement rates of 94.8% and 95.8% of cases at 1% and 50% cutoff, respectively; slightly lower values were reported for E1L3N and 22C3, with overall percent agreement rates of 87.6% and 92.7% at 1% and 50% cutoff, respectively [10].

Smith et al compared E1L3N on Ventana platform and SP263 in 100 NSCLC cases using TMAs and found SP263 assay to be more sensitive and with a wider dynamic range than the E1L3N assay, detecting PD-L1 in 29 additional tumors [13].

Regarding interobserver variability, we found a good level of concordance between the 2 scoring pathologists for the 3 clones, in line with the results reported by our group and others [8,23]. Although we did not use whole sections for our comparisons, it should be noted that tissue cores better reflect the actual material that pathologists encounter in the daily practice because most cases for which PD-L1 evaluation is required are composed of diagnostic biopsies.

In conclusion, we could demonstrate that with an optimized protocol, the PD-L1 clone E1L3N showed high levels of overall agreement with assay SP263 (94.5% and 95.8% at 1% and 50% cutoff, respectively), whereas lower values were found with assay 22C3 (77.1% and 88% at 1% and 50% cutoff, respectively).

Importantly, at 50% cutoff, clone E1L3N showed 100% sensitivity with both SP263 and 22C3 with no false-negative cases. Thus, our study shows that clone E1L3N is a reliable and affordable tool to be used to screen large number of cases for PD-L1 expression in the research setting as well as a diagnostic test when considering pembrolizumab as first-line of therapy in NSCLC, provided that cases resulting positive are retested with an approved assay.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] Hendry S, Byrne DJ, Wright GM, et al. Comparison of four PD-L1 immunohistochemical assays in lung cancer. J Thorac Oncol 2018; 13:367-76.

[7] Buttner R, Gosney JR, Skov BG, et al. Programmed death-ligand 1 immunohistochemistry testing: a review of analytical assays and clinical implementation in non–small-cell lung cancer. J Clin Oncol 2017; 35:3867-76.

[8] Munari E, Rossi G, Zamboni G, 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.

[9] Conde E, Caminoa A, Dominguez C, et al. Aligning digital CD8(+) scoring and targeted next-generation sequencing with programmed death ligand 1 expression: a pragmatic approach in early-stage squamous cell lung carcinoma. Histopathology 2018;72:270-84.

[10] Kim H, Kwon HJ, Park SY, Park E, Chung JH. PD-L1 immunohistochemical assays for assessment of therapeutic strategies involving immune checkpoint inhibitors in non-small cell lung cancer: a comparative study. Oncotarget 2017;8:98524-32.

[11] Parra ER, Villalobos P, Mino B, Rodriguez-Canales J. Comparison of different antibody clones for immunohistochemistry detection of programmed cell death ligand 1 (PD-L1) on non–small cell lung carcinoma. Appl Immunohistochem Mol Morphol 2018;26:83-93.

[12] Scheel AH, Dietel M, Heukamp LC, et al. Harmonized PD-L1 immunohistochemistry for pulmonary squamous-cell and adenocarcinomas. Mod Pathol 2016;29:1165-72.

[13] Smith J, Robida MD, Acosta K, et al. Quantitative and qualitative characterization of two PD-L1 clones: SP263 and E1L3N. Diagn Pathol 2016;11:44.

[14] 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;3:1051-8.

[15] Adam J, Le Stang N, Rouquette I, et al. Multicenter harmonization study for PD-L1 IHC testing in non–small-cell lung cancer. Ann Oncol 2018;29:953-8.

[16] Travis WD, Brambilla E, Burke AP. WHO classification of tumors of the lung pleura thymus and heart. Lyon: IARC Press; 2015.

[17] Travis WD, Brambilla E, Nicholson AG, et al. The 2015 World Health Organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. J Thorac Oncol 2015;10:1243-60.

[18] Rami-Porta R, Bolejack V, Giroux DJ, et al. The IASLC lung cancer staging project: the new database to inform the eighth edition of the TNM classification of lung cancer. J Thorac Oncol 2014;9:1618-24.

[19] Munari E, Zamboni G, Lunardi G, et al. PD-L1 expression heterogeneity in non–small cell lung cancer: defining criteria for harmonization between biopsy specimens and whole sections. J Thorac Oncol 2018; 13:1113-20.

[20] 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.

[21] Munari E, Zamboni G, Marconi M, et al. PD-L1 expression heterogeneity in non-small cell lung cancer: evaluation of small biopsies reliability. Oncotarget 2017;8:90123-31.

[22] Munari E, Zamboni G, Sighele G, et al. Expression of programmed cell death ligand 1 in non–small cell lung cancer: comparison between cytologic smears, core biopsies, and whole sections using the SP263 assay. Cancer Cytopathol 2019;127:52-61.

[23] 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.