Peer review file

Article information: https://dx.doi.org/10.21037/jtd-21-388

Reviewer A

Comment 1:

I am happy to review for your study because of the interesting topic when immunotherapy era now. Especially, I am interested in Table 1, but other tables had nothing new compared with previous studies. So I recommend you do minor revise to focus on table 1, which should be analyzed mush in details.

1. *Can you look for which histologic subtypes would have impact on PD-L1 expression, statistically? For instance, logistic regression analysis may resolve the question by the comparison between poorly differentiated subtypes vs others, or well differentiated one vs others.*

Reply 1: We greatly appreciate for your time spent making constructive suggestions and comments, which significantly help us to improve the manuscript. In results (page 8, line 152-158), we have showed PD-L1 high expression was significantly more frequent in poorly differentiated histological variants than in well-differentiated variants (P<0.001).

Furthermore, we have added logistic regression analysis data in the results. The logistic regression analysis also showed that poorly differentiated variants had significantly more frequent high PD-L1 expression than well differentiated variants (P<0.01) (see supplementary table 2 as follow).

Supplementary Table 2 The Logistic Regression Analysis of PD-L1 expression in different subtypes of ADC (PD-L1 TPS cut-off value at 50%).

Subtype	OR	95% CI	*P* value
Lepidic vs Micropapillary	0.042	0.005-0.358	0.004
Lepidic vs Solid	0.034	0.012-0.100	<0.001
Acinar vs Micropapillary	0.066	0.022-0.195	<0.001
Acinar vs Solid	0.112	0.073-0.173	<0.001
Papillary vs Micropapillary	0.063	0.017-0.231	<0.001

Papillary vs Solid	0.132	0.080-0.218	<0.001
Micropapillary vs Solid	0.906	0.394-2.085	0.816

Changes in the text:

We have added logistic regression analysis data in results section (see page 8, line 158-161):

Furthermore, the logistic regression analysis also showed that the solid and micropapillary subtypes (poorly differentiated variants) had significantly more frequent high PD-L1 expression than lepidic, acinar and papillary subtypes (well differentiated variants) (P<0.01) (Supplementary Table 2).

We also have modified our text as advised in discussion (page11, line 237-240):

The histological subtypes of ADC may indicate PD-L1 expression status. The poorly differentiated subtypes (such as solid and micropapillary subtype) might be more likely PD-L1 high expression, while well-differentiated subtypes (lepidic, acinar and papillary subtype) might be more likely PD-L1 negative or low expression.

Comment 2:

2. Otherwise, it could be investigated whether some histologic subtypes (lepidic, papillary, or acinar) might be statistically significant negative indicator for TPS <49%.

Since the sample size was relatively large, it should be used to the maximum, I think.

Reply 2: Thank you for your helpful comment. We further analyzed the association between the histologic subtypes and PD-L1 expression status at cutoff value of 50%, and found that well-differentiated subtypes of adenocarcinoma might be more likely negative or low PD-L1expression (TPS <50%) (P<0.001, Supplementary Table 3). Therefore, well-differentiated subtypes (lepidic, acinar and papillary subtype) might indicate PD-L1 negative or low expression (TPS <50%).

Supplementary Table 3 The heterogeneous expression of PD-L1 in different subtype and differentiation of ADC (PD-L1 TPS cut-off value at 50%).

		PD-L1 status			
		TPS <50%	TPS ≥50%	Total (n=852)	*P* value
ADC	Subtype predominant				
	Lepidic	72	0	72	
	Acinar	409	9	418	
	Papillary	192	4	196	<0.001
	Micropapillary	21	7	28	
	Solid	97	41	138	
	Total	791	61	852	

Change in the text:

We have modified in discussion section (page11, Line 234-240):

Meanwhile, if using two-tiered categorization at the cut-off value of 50%, the histologic subtypes of ADC was still statistically significant associated with PD-L1 expression status (P<0.001, see Supplementary Table 3). The histological subtypes of ADC may indicate PD-L1 expression status. The poorly differentiated subtypes (such as solid and micropapillary subtype) might be more likely PD-L1 high expression, while well-differentiated subtypes (lepidic, acinar and papillary subtype) might be more likely PD-L1 negative or low expression.

Reviewer B

Major comments

Comment 1:

It seems unclear how and why 53 paired FFPE samples from over 1000 resected samples. The authors only state that they "randomly selected".

Reply 1: We greatly appreciate for your constructive suggestions and comments, which significantly help us to improve the manuscript. In this paper, the specimens for PD-L1 testing are all detected immediately after the excision or biopsy. To avoid the PD-L1 expression being influenced by the aging of FFPE tissue blocks, to avoid frozen tissue not being the most suitable for immunohistochemistry test, and the small tumor tissue which had less than 2 blocks was not enough for testing, we selected the recent FFPE blocks (<2 month, more than 3 blocks from the same resected samples) for comparison between different blocks. Therefore, only 53 paired FFPE blocks were recruited in this study.

Comment 2:

Also, how the paired blocks (A & B) were selected from the same resected samples? Were there any rules for this selection? Did they made sure to avoid paring form adjacent location? Please show the details of methodology of judge of heterogeneity within the samples.

Reply 2: Thank you for your helpful advice. All the paired blocks in this study were selected from the same resected specimen, which had no less than 3 blocks, to avoid frozen tissue not being the most suitable for PD-L1 immunohistochemistry staining or the mass was too small to be available for two blocks testing. And the 2 blocks are randomly selected. But the selection rules could not completely avoid adjacent location.

Minor comments

Comment 3:

The title of the article does not seem to fully represent the topic in this research. Please provide the title more concisely summarizing the key issues.

Reply 3: We have used other titles before, and finally we changed to the current title. The previous titles that we provide as bellow:

(1) Heterogeneity of PD-L1 Expression in Non-Small-Cell Lung Cancer

(2) Heterogenous Expression of PD-L1 in Non-Small-Cell Lung Cancer: Comparison Among Different Sampling Materials

Which title what do you think could be more representative of this research topic? Thank you for your help and giving us some advice.

Comment 4:

The authors concordance rate or kappa value among three TPS subgroups (negative, 1-49%, 50% or more). The discordance seems mainly arise on the judge of 1-49%. What if they set the cut to divide samples to two subgroups (~50% vs 50% or more, or negative vs 1-100%). Please show their concordance rate and kappa values in the discussion section, if possible.

Reply 4: According to your helpful advice, we analyzed the concordance between paired samples using two-tiered categorization at the TPS cut-off value of 1%, the concordance rate was 94.4% (kappa value: 0.886, good agreement, *P*<0.001, see table 2-1) between paired cell blocks and biopsy samples, 77.1% (kappa value: 0.5, moderate agreement, *P*=0.002, see table 3-1) between paired resected and biopsy samples, 79.2% (kappa value: 0.478, moderate agreement, *P*<0.001, see table 4-1) between different blocks of the same resected samples, and 77.6% (kappa value: 0.559, moderate agreement, *P*<0.001, see table 5-1) between paired primary and metastatic lesions, respectively. The concordance rate and kappa value were close to the data when using three TPS subgroups (negative, 1-49%, 50% or more).

Then, we further analyzed the concordance between paired samples at the TPS cut-off value of 50%, the concordance rate was 94.4% (kappa value: 0.859, good agreement, *P*<0.001, see table 2-50) between paired cell blocks and biopsy samples, 91.4% (kappa value: 0.796, good agreement, *P*<0.001, see table 3-50) between paired resected and biopsy samples, 84.9% (kappa value: 0.654, moderate agreement, *P*<0.001, see table 4-50) between different blocks of the same resected samples, and 93.9% (kappa value: 0.819, good agreement, *P*<0.001, see table 5-50) between paired primary and metastatic lesions, respectively. When using two-tiered categorization at the TPS cut-off value of 50%, the concordance rate and kappa value between paired biopsy and resected samples, paired primary and metastatic lesions were higher than that using three-tiered categorization, while the kappa value of the other two paired groups were similar when divided TPS into two or three subgroups.

Since the purpose of our study is to investigate the heterogeneity of NSCLC between different paired groups using three-tiered categorization, moreover, the concordance data of using two-tiered categorization at the TPS cut-off value of 1% or 50% is close to that of our data, so we didn't further show the data of TPS cut-off

value of 1% or 50% in the text.

Table 2-1 The discrepant expression of PD-L1 in the histology and matched cell block of the same biopsy sample.

	PD-L1 Status	Cell blocks		Total (n=54)	Kappa value	*P* value
		TPS <1%	TPS ≥1%			
Biopsy	TPS <1%	30	2	32	0.886	<0.001
	TPS ≥1%	1	21	22		
	Total	31	23	54		

Table 3-1 The heterogeneous PD-L1 expression between biopsy and matched resected specimens.

	PD-L1 Status	Biopsy		Total (n=35)	Kappa value	*P* value
		TPS <1%	TPS ≥1%			
Resection	TPS <1%	8	2	10	0.5	0.002
	TPS ≥1%	6	19	25		
	Total	14	21	35		

Table 4-1 The discrepant PD-L1 expression between different paraffin blocks of the same surgical resected specimens.

	PD-L1 Status	Block A		Total (n=53)	Kappa value	*P* value
		TPS <1%	TPS ≥1%			
Block B	TPS <1%	9	6	15	0.478	<0.001
	TPS ≥1%	5	33	38		
	Total (n=53)	14	39	53		

Table 5-1 The heterogeneity of PD-L1 expression between primary and matched lymph node metastatic lesions.

PD-L1 Status	Lymph node metastasis		Total (n=49)	Kappa value	*P* value
	TPS <1%	TPS ≥1%			

	TPS <1%	20	2	22		
Primary lesion	TPS ≥1%	9	18	27	0.559	<0.001
	Total	29	20	49		

Table 2-50 The discrepant expression of PD-L1 in the histology and matched cell block of the same biopsy sample.

	PD-L1 Status	Cell blocks		Total (n=54)	Kappa value	*P* value
		TPS <50%	TPS ≥50%			
	TPS <50%	38	2	40		
Biopsy	TPS ≥50%	1	13	14	0.859	<0.001
	Total	39	15	54		

Table 3-50 The heterogeneous PD-L1 expression between biopsy and matched resected specimens.

	PD-L1 Status	Biopsy		Total (n=19)	Kappa value	*P* value
		TPS <50%	TPS ≥50%			
	TPS <50%	23	2	25		
Resection	TPS ≥50%	1	9	10	0.796	<0.001
	Total	24	11	35		

Table 4-50 The discrepant PD-L1 expression between different paraffin blocks of the same surgical resected specimens.

	PD-L1 Status	Block A		Total (n=53)	Kappa value	*P* value
		TPS <50%	TPS ≥50%			
	TPS <50%	32	5	37		
Block B	TPS ≥50%	3	13	16	0.654	<0.001
	Total (n=53)	35	18	53		

Table 5-50 The heterogeneity of PD-L1 expression between primary and matched lymph node metastatic lesions.

	PD-L1 Status	Lymph node metastasis		Total (n=49)	Kappa value	*P* value
		TPS <50%	TPS ≥50%			
Primary lesion	TPS <50%	37	3	40	0.819	<0.001
	TPS ≥50%	0	9	9		
	Total	37	12	49		

Reviewer C

General:

Comment 1:

Although authors state that their study is "comprehensive", this is not exactly true. There were 1002 "specimens" included, but there were actually several smaller studies that were performed rather than one large and comprehensive study. The largest group was that of cases where PD-L1 expression was assessed in different histological subtypes of NSCLC. In addition, the four other smaller study sets were included from the pool of 1002 specimens.

Reply 1: Thank you for your meaningful suggestion, because of limitation of small biopsy samples number, the size of tumor, the number and the aging of FFPE blocks, only small cohort of four groups of paired samples were included in this study.

In our hospital, most of the patients suspected to lung cancer by CT or PET-CT imaging underwent surgical resection directly without preoperative diagnostic biopsy. Therefore, only about 5% of the cases (about 50 cases per year) underwent preoperative diagnostic biopsy. From the pool of 1002 specimens, only about 35 cases had more than 100 viable tumor cells in biopsy tissue for PD-L1 testing immediately.

Secondly, due to frozen tissue was not the most suitable for immunohistochemistry test, blocks of small size tumor were less than 2 not enough for testing, the PD-L1 expression could be influenced by the aging of FFPE blocks, we selected the recent FFPE tissue blocks (<2 months, >3 blocks) for PD-L1 testing.

The number of lymph node-positive rate is about 10% in 1002 surgical resected NSCLC specimens, and metastatic lesion that had less than 100 viable tumor cells was not available for PD-L1 testing. So only 53 recent paired two paraffin blocks from the same resected samples and 49 recent paired primary and lymph node metastatic lesions were selected from the 1002 resected NSCLCs pool.

Finally, some of EBUS-TBNA tissues were only used for cell blocks, and some were only used for histology paraffin blocks. There were a few samples both for cell blocks and biopsy. In addition, some cases had been tested for molecular analysis and immunohistochemical typing. Therefore, only 54 paired cases with enough remaining tumour cells (>100 tumour cells) could be used for PD-L1 IHC testing immediately at the same time in this study. Moreover, the 54 paired EBUS-TBNA cell blocks and biopsy were mainly advanced NSCLCs, which were not from the 1002 resected NSCLCs pool.

Although our study is relatively comprehensive on assessing PD-L1 expression in different histological subtypes of NSCLC. But the limitation is that relatively small cohort of 4 paired samples were included in this study to explore the heterogeneity of NSCLC between different samples. We hope that more data will be available in future studies to further confirm our results.

Comment 2:

Specific question regarding selection of two different paraffin blocks from resection specimens: How were these two blocks selected? Were they randomly added for the study or the second block was ordered because the first one was either negative or suboptimal for assessment for some reason? This is important to state because non-random selection will cause a bias in sampling and potentially different results than random sampling. The selection criteria should be clearly stated.

Reply 2: Due to frozen tissue was not the most suitable for immunohistochemistry test, blocks of small size tumor were less than 2 not enough for testing, the PD-L1 expression could be influenced by the aging of FFPE blocks, we selected the recent FFPE tissue blocks (<2 months, >3 blocks) for PD-L1 testing to investigate heterogeneous expression of PD-L1 in different paraffin blocks of non-small cell lung cancers. The 53 paired two blocks were randomly selected in recently (< 2 months) at the same time after resected. The second block was selected for PD-L1 testing not

because the first one was either negative or suboptimal for assessment for some reason, but for investigating the heterogeneity of PD-L1 expression.

Methods:

Comment 3:

What assay was used? Stating only that the 22C3 assay was used is not enough. It is not clear if that was indeed Dako PD-L1 IHC 22C3 pharmDx Assay or it was an LDT using the 22C3 clone from Dako? This is critical. It has to be precisely stated and the supplier cited as usual (e.g. Dako PD-L1 22C3 pharmDx (Dako/Agilent, city, country). If an LDT, how was the LDT validated for this purpose?

Reply 3: We are sorry about that we didn't state clear what PD-L1 22C3 assay was used in the study. We have described that Dako PD-L1 IHC 22C3 pharmDx Assay (Dako/Agilent, Santa Clara, US) was used in the study and revised in the text.

Changes in the text (see page 6, line 122-123):

using the Dako PD-L1 IHC 22C3 pharmDx assay (Dako/Agilent, Santa Clara, US).

Comment 4:

What questions were asked in this study? It appears that there were two main questions: 1. association between PD-L1 expression and histological type, and 2. concordance between paired samples of 4 different types. However, as the authors have the results for each case, I suggest that they also address accuracy of each type of sample. The gold standard is the one that is clinically relevant. Since oncologists will consider the tumor positive irrespective if they get a positive results from the biopsy or resection as positive, the "designated true value" is that of whether the case was positive with either bx or resection and each type of specimens compared to positivity rate obtained with samples combined. That would emphasize that the large sample is not always "correct" in providing the PD-L1 status of the tumor. This will allow you to calculate sensitivity of each type of specimen. Specificity cannot be addressed as any positivity is considered as correct from the oncologists' point of view (at least in our institution).

Reply 4: You have proposed a very valuable comment. All of our PD-L1 interpretation pathologists have targo training, and they all possessed considerable experiences. Therefore, the accuracy of the PD-L1 immunohistochemistry testing

results is relatively reliable. The gold standard of PD-L1 positive results is clinically relevant. Most of our resected specimens are in stage I-II and a few in stage III, at present, there is no indication for PD-L1/PD-1 inhibitors use in postoperative adjuvant therapy. We have no relevant clinical data to verify the results. It is difficult to determine which sampling type is recommended be “designated true value”. The purpose of our study is to explore the heterogeneity of PD-L1 expression in different histological type and different sampling types of non-small cell lung cancer. In practical application, oncologists will consider the tumor positive whether they get a positive result from the biopsy or resection.

Comment 5:

Table 1 Would benefit if percentages were also shown.

Reply 5: we have added the percentages in the table 1.

Changes in the text: We have modified our Table 1 as advised (see Table 1).

Comment 6:

Discussion:

In Discussion, the authors spend too much time on comparing their results to the results with SP142 assay. This is not very helpful. These two assays are used for different purposes, are using completely different scoring methods, and will be stratifying patients for different therapies.

Reply 6: You have proposed a very valuable suggestion. We have deleted some discussion content on comparing our results with SP142 assay.

Changes in the text: we have deleted the text in page 13 line 272-274:

In page 14, line 288-290:

On the other side, the authors do not really take full advantage of generated data/evidence and their meaning. Based on their results, it is feasible to consider that the current practice of selecting only one paraffin block in resection samples for PD-L1 testing has to be changed. Before reporting a negative result on a single block, perhaps additional blocks should be assessed because as much as 10% of tumors that

were negative in one block were positive in another block in this study. Actually, the discrepancy between positive and negative results was higher between the two blocks of resection specimen than that between resection and biopsy and biopsy and cytology, etc. However, the analysis of this results needs to be changed as one block results should not be compared to another block result, but to both blocks combined (e.g. block A vs. block A&B and block B vs. block A&B) because blocks are not reported independently, but should be combined to report the tumor status.

Reply: You have proposed very valuable advice. Based on the variation in our results, we recommended that if PD-L1 expression of one block was negative, adding additional block for testing may be necessary to improve the PD-L1 positive rate. The main purpose of this study is to investigate the heterogeneity of PD-L1 expression between different blocks from the same resected tissue of NSCLC. According to your comment, we further compared one block results with both blocks combined (block A vs. block A&B and block B vs. block A&B) (see tables as bellow), the concordance rate of PD-L1 expression was 81.1% (43/53) with a kappa value of 0.706 (moderate agreement, $P<0.001$) and 83% (44/53) with a kappa value of 0.734 (moderate agreement, $P<0.001$), respectively. The kappa value was close to the data of comparison of two blocks independently (A vs B) (kappa value: 0.455, moderate agreement, $P<0.001$)). If there are two blocks tested, whether the PD-L1 expression of two blocks should be average, instead of one block reported independently, it can be used for reference only. There is still not enough evidence to change the detection strategy. The shortcomings of our study: most of our resected specimens are in stage I-II and a few in stage III, there is no indication for PD-L1/PD-1 inhibitors use in postoperative adjuvant therapy at present. The final result has not been effectively verified by treatment. We hope that in the future study, it can be verified by postoperative treatment evaluation.

Changes in the text:

we have added the discussion in page14, line 294-300:

More interesting, 20.8% (11/53) of cases that were negative in one block and positive in another block in our study. These findings of the discordant PD-L1 expression between different blocks highlight the importance of taking into consideration of histological subtypes, when selecting a block for the PD-L1 testing. And if PD-L1 expression of one block is negative, adding additional block for PD-L1

IHC testing may be useful to improve the PD-L1 positive rate.

Table block A vs A+B

The discrepant PD-L1 expression between different paraffin blocks of the same surgical resected specimens.

	PD-L1 Status	Block A			Total (n=53)	Kappa value	*P* value
		TPS <1%	TPS 1-49%	TPS ≥50%			
Block A+B	TPS <1%	9	0	0	9	0.706	<0.001
	TPS 1-49%	5	20	4	29		
	TPS ≥50%	0	1	14	15		
	Total	14	21	18	53		

Table block B vs A+B

The discrepant PD-L1 expression between different paraffin blocks of the same surgical resected specimens.

	PD-L1 Status	Block B			Total (n=53)	Kappa value	*P* value
		TPS <1%	TPS 1-49%	TPS ≥50%			
Block A+B	TPS <1%	9	0	0	9	0.734	<0.001
	TPS 1-49%	6	21	2	29		
	TPS ≥50%	0	1	14	15		
	Total	15	22	16	53		

It is known that older paraffin blocks loose detectable PD-L1 at least with 22C3 clone. Authors do not state how old are their cases and if they noticed any differences between older and newer cases.

Reply: To avoid PD-L1 expression can be influenced by the aging of FFPE tissue blocks. In this paper, the specimens for PD-L1 testing are mainly detected immediately after the excision or biopsy. And the analysis of heterogeneous

expression of PD-L1 in different paraffin blocks of the same resected samples and paired primary and lymph node metastasis lesion, we selected the recent FFPE tissue blocks (<2 month) for testing immediately to avoid the influence by the aging of blocks.

Changes in the text as follow:

We have added detectable methodology in page 6, line 124-126:

To avoid PD-L1 expression maybe influenced by the aging of FFPE tissue blocks, the specimens for PD-L1 testing are all detected immediately after the excision or biopsy.

Comment 7:

References:

Authors fail to reference couple of studies that are relevant for discussion. More references are needed in discussion about association with histological type, PD-L1 testing in cytology, and overall tissue heterogeneity. Here are some examples (please see the list below), but the list is not complete and further homework should be done by authors to identify critical previously published literature and give proper credits to other good studies.

Reply 7: We are very sorry about failing to reference couple of studies that are relevant for discussion. Since we submitted this article in early March and had finished writing in February, we didn't add some reference of 2021 in the discussion. Now we have cited previously relevant references and new published literatures in the discussion.

Changes in the text:

For discussion about association with histological type and different samples, we added relevant references as follow: reference 19, reference 23.

19. Naso JR, Banyi N, Al-Hashami Z, et al. Discordance in PD-L1 scores on repeat testing of non-small cell lung carcinomas. Cancer Treat Res Commun 2021; 27:100353.

23. Bulutay P, Firat P, Zeren EH, et al. The importance of histological patterns on PD-L1 staining heterogeneity: Should we use pattern-based approach for selecting tumor samples for PD-L1 testing in lung adenocarcinomas? Turk J Med Sci 2021;51:204-13.

For discussion on cytology samples testing, we added relevant references as follow: reference 29-31.

29. Jug R, Giovacchini CX, Liu B, et al. EBUS-FNA cytologic-histologic correlation of PD-L1 immunohistochemistry in non-small cell lung cancer. J Am Soc Cytopathol 2020; 9:485-93.

30. Chauhan A, Siegel L, Freese R, et al. Performance of Ventana SP263 PD-L1 assay in endobronchial ultrasound guided-fine-needle aspiration derived non-small-cell lung carcinoma samples. Diagn Cytopathol 2021; 49:355-62.

31. Wang G, Ionescu DN, Lee CH, et al. PD-L1 testing on the EBUS-FNA cytology specimens of non-small cell lung cancer. Lung Cancer 2019; 136:1-5.

Comment 8:

Abstract:

The first sentence in the Results paragraph should be deleted and replaced by the sentence that would explicitly say how many samples were evaluated for association between PD-L1 expression and histologic type and emphasize the novel findings in their study. The second sentence should introduce three smaller sets of matched samples and show their results. It is important that authors summarize the results in this way so that it is immediately clear what was study design and results combined in this section.

Reply 8: we have revised the abstract according to your comment.

Change in the text:

The first sentence was deleted as follow:

And replaced by the sentence as follow (page 2, line 45-52):

A total of 1002 resected NSCLC specimens, including 852 adenocarcinomas (ADCs) and 150 squamous cell carcinomas (SCCs), 35 paired biopsy and resected samples, 54 paired cell block and biopsy samples, 53 paired two blocks from the same resected tissue and 49 paired primary and metastatic lesion samples were included in this study. Interesting, PD-L1 high expression is significantly more frequent in poorly differentiated subtypes than in well-differentiated subtypes in the ADC subgroup(P<0.001). In the SCC subgroup, PD-L1 high expression was significantly

more associated with the nonkeratinizing than keratinizing type (P=0.001).