# PD-L1 (lung)

The fifth assessment in the NordiQC Companion module (C5) focused on the accuracy of the PD-L1 (lung) IHC assays in order to identify patients with non-small cell lung carcinoma (NSCLC) benefitting from immune therapy with Pembrolizumab (Keytruda®), based on the tumour proportion score (TPS) either as first line treatment (TPS ≥50%) or second line treatment (TPS ≥1%). The PD-L1 expression levels in the circulated material used for the assessment were primarily characterized by the CE IVD / FDA approved companion IHC assay, 22C3 pharmDx, SK006 Dako/Agilent, but also by the CE IVD approved assays 28-8 pharmDx, SK005 Dako/Agilent, Ventana PD-L1 (SP263) assay, 790-4905 all considered to be equivalent with the 22C3 assay[1] and finally also by the newly launched 22C3 pharmDx GE006 Dako/Agilent. The PD-L1 evaluation in the individual tissue cores and levels were based on the cut-off values accordingly to the interpretation guideline for the 22C3 pharmDx assay SK006, indicating TPS cut-off levels at ≥50% and ≥1%, respectively.

1) Tsao MS et al. PD-L1 Immunohistochemistry Comparability Study in Real-Life Clinical Samples:
Results of Blueprint Phase 2 Project. J Thorac Oncol. 2018 Sep;13(9):1302-1311.

**Material**

Table 1. **Content of the TMA used for the NordiQC PD-L1 (lung) C5 assessment**

| | PD-L1 IHC TPS score* | |
|---|---|---|
| **Tissue controls** | | |
| 1. Placenta | See section for controls | |
| 2-3. Tonsil | See section for controls | |
| **NSCLC** | | |
| 4. NSCLC | No; <1% | |
| 5. NSCLC | No; <1% | |
| 6. NSCLC | Low; ≥1-49% | |
| 7. NSCLC | Low; ≥1-49% / High ≥50%** | |
| 8. NSCLC | Low; ≥1-49% | |
| 9. NSCLC | Excluded (too few tumour cells) | |
| 10. NSCLC | High; ≥50% | |
| 11. NSCLC | Low; ≥1-49% / High ≥50%** | |
| 12. NSCLC | Low; ≥1-49% / High ≥50%** | |
| 13. NSCLC | High; ≥50% | |

\* Tumour proportion score (TPS) determined by PD-L1 IHC 22C3, SK006 Dako performed in NordiQC reference lab.
\*\* These tumor cores displayed tumor heterogeneity. Depending on the blocks and reference PD-L1 levels from which the sectioned slides were circulated, the level of PD-L1 was categorized into one of the percentage ranges (TPS) highlighted in the Table 1 (for further elaboration - see description below).

All tissues were fixed in 10% neutral buffered formalin.
The participating laboratories were asked to perform the PD-L1 (lung) IHC assay, interpret the PD-L1 expression level using the TPS scoring system and submit these scores to NordiQC. This allowed assessment of the technical performance (analytical accuracy) of the PD-L1 (lung) IHC assays and provided information on the reproducibility and concordance of the PD-L1 interpretation results among the laboratories.

**PD-L1 (lung) IHC, Technical assessment**
In order to account for heterogeneity of PD-L1 expression in the individual tumour cores included in the tissue TMA blocks, reference slides were made through out the blocks. First, middle and last slides were stained for PD-L1 using the CE IVD / FDA approved 22C3 pharmDx kit SK006 (Dako), and also by the CE IVD approved assays 22C3 pharmDx GE006 (Dako), 28-8 pharmDx SK005 (Dako) and SP263 790-4905 (Ventana) in a NordiQC reference laboratory. During the assessment, TPS categories for each tissue core on the submitted slides were compared to the level in the nearest reference slide of 22C3 pharmDX SK006 (Dako). In addition, every twentieth slide throughout the blocks were staining using a validated LD assay based on 22C3 on the Omnis platform securing quality and verifying the PD-L1 status in the circulated materiel.

**Criteria for assessing a staining as <u>Optimal</u> included:**
The staining is considered perfect or close to perfect in all of the included tissues.
<u>TPS (as estimated by NordiQC assessors based on reference staining) fully concordant with the nearest NordiQC reference slide obtained in all 9 NSCLC cores.</u>

**Criteria for assessing a staining as <u>Good</u> included:**
The staining is considered acceptable in all of the included tissues. However, the protocol may be optimized to provide improved analytical accuracy, counter staining, morphology or signal-to-noise ratio.
<u>TPS category is still concordant with the nearest NordiQC reference slide in all 9 NSCLC cores.</u>

**Criteria for assessing a staining as <u>Borderline</u> included:**
The staining is considered insufficient, e.g., because of a generally too weak staining, a false negative staining or a false positive staining reaction of one of the included tissues. The protocol should be optimized.
<u>TPS is **not** found concordant with the nearest NordiQC reference slide in all 9 NSCLC cores.</u>

**Criteria for assessing a staining as <u>Poor</u> included:**
The staining is considered insufficient due to a false negative or a false positive staining reaction staining of more than one of the included tissues. An optimization of the protocol is urgently needed.
<u>TPS is **not** found concordant with the nearest NordiQC reference slide in all 9 NSCLC cores.</u>

**Participation**

| Number of laboratories registered for PD-L1 (lung) IHC C5 | 186 |
|---|---|
| Number of laboratories returning PD-L1 (lung) IHC | 176 (95%) |
| Number of laboratories returning PD-L1 scoring sheet | 155 (88%) |

**Results:** 176 laboratories participated in this assessment and 80% achieved a sufficient mark. Assessment marks for IHC PD-L1 assays and PD-L1 antibodies are summarized in Table 3 (see next page).

Table 2. **Proportion of sufficient results for PD-L1 (lung) in the five NordiQC runs performed**

|  | C1 2017 | C2 2018 | C3 2018 | C4 2018 | C5 2019 |
|---|---|---|---|---|---|
| Participants, n= | 68 | 145 | 146 | 163 | 176 |
| Sufficient results | 50% | 84% | 91% | 86% | 80% |

**Performance history**
This was the fifth NordiQC assessment of PD-L1 (lung). A slightly reduced pass rate was obtained in C5 (see Table 2) compared to the three previous assessments. The number of new participants seems to be consistently increasing.

**Conclusion**
In this fifth NordiQC run for PD-L1 (lung) in the companion module, C5, an overall pass rate of 80% was achieved. Insufficient PD-L1 IHC staining results were most frequently characterized by a reduced proportion of PD-L1 positive cells compared to the level expected as defined by the PD-L1 IHC pharmDx assay, SK006 (Dako/Agilent). This resulted in a too low TPS in one or more of the NSCLC cores.
Several companion diagnostic assays and LD assays based on clones as E1L3N, CAL10, ZR3 and BSR90 could provide optimal results. The companion diagnostic PD-L1 IHC assays, 22C3 SK006/GE006 and SK005 from Dako/Agilent and SP263 from Ventana/Roche provided a high proportion of sufficient results superior to LD assays. Within LD assays, irrespectively of Ab clone being used, meticulous calibration and validation of the assay is required.

Table 3. **Assessment marks for IHC assays and antibodies run C5, PD-L1 (lung) IHC**

| CE-IVD / FDA approved PD-L1 assays | n | Vendor | Optimal | Good | Borderline | Poor | Suff.[1] | Suff. OPS[2] |
|---|---|---|---|---|---|---|---|---|
| rmAb clone SP263, **740-4907**[3] | 16 | Ventana/Roche | 12 | 2 | 2 | - | 88% | 93% |
| rmAb clone SP263, **740-4907**[4] | 1 | Ventana/Roche | - | - | - | 1 | - | - |
| rmAb clone SP263, **741-4905**[5] | 4 | Ventana/Roche | 2 | 2 | - | - | - | - |
| rmAb clone SP263, **790-4905** | 48 | Ventana/Roche | 34 | 10 | 4 | - | 92% | 92% |
| mAb clone 22C3 pharmDX, **SK006**[6] | 24 | Dako/Agilent | 17 | 5 | - | 2 | 92% | 92% |
| mAb clone 22C3 pharmDX, **SK006**[7] | 9 | Dako/Agilent | 2 | 2 | 1 | 4 | 44% | - |
| mAb clone 22C3 pharmDX, **GE006**[8] | 3 | Dako/Agilent | 3 | - | - | - | - | - |
| rmAb clone 28-8 pharmDX, **SK005**[9] | 3 | Dako/Agilent | 2 | 1 | - | - | - | - |
| **Antibodies[10] for laboratory developed PD-L1 assays, concentrated antibodies** | n | Vendor | Optimal | Good | Borderline | Poor | Suff.[1] | Suff. OPS[2] |
| mAb clone **22C3** | 42 | Dako/Agilent | 11 | 20 | 9 | 2 | 74% | 74% |
| mAb clone **E1L3N** | 5 | Cell Signaling | 1 | 4 | - | - | 100% | 100% |
| rmAb **CAL10** | 3 3 | Biocare Zytomed Systems | 3 | 1 | - | 2 | 67% | 100% |
| rmAb clone **28-8** | 4 | Abcam | 1 | 2 | - | 1 | - | - |
| rmAb clone **ZR3** | 1 1 1 1 | Cell Marque Zeta Corporation Nordic Biosite Gene Tech | 1 | - | - | 3 | - | - |
| rmAb clone **QR1** | 1 1 | Quartett Diagomics | 1 | - | - | 1 | - | - |
| rmAb **BSR90** | 1 | Nordic Biosite | 1 | - | - | - | - | - |
| rmAb clone **SP142** | 1 | Spring Biosystems | - | 1 | - | - | - | - |
| **Ready-To-Use antibodies** | n | Vendor | Optimal | Good | Borderline | Poor | Suff.[1] | Suff. OPS[2] |
| rmAb CAL10, **API3171** | 1 | Biocare | - | - | 1 | - | - | - |
| rmAb QR1, **2-PR292-13** | 1 | Biocyc | - | 1 | - | - | - | - |
| rmAb clone MXR003, **RMA-0732** | 1 | Maixin | - | - | - | 1 | - | - |
| Total | 176 | | 91 | 51 | 17 | 17 | | |
| Proportion | | | 51% | 29% | 10% | 10% | 80% | |

1) Proportion of sufficient stains (optimal or good).
2) Proportion of sufficient stains with optimal protocol settings only, see below.
3) Approved for IMFINZI in urothelial cancers in US.
4) RTU system developed for the Ventana/Roche's automated systems (BenchMark) but used by laboratories on a different platform (Leica Bond).
5) Approved for IMFINZI, KEYTRUDA and OPDIVO in NSCLC in EU.
6) Approved for KEYTRUDA in NSCLC in EU/US.
7) RTU system developed for the Agilent/Dako's semi-automated systems (Autostainer Link48) but used by laboratories on different platforms (Ventana BenchMark and Dako Omnis).
8) Approved for KEYTRUDA in NSCLC in EU
9) Approved for OPDIVO in NSCLC in EU/US
10) mAb: mouse monoclonal antibody, rmAb: rabbit monoclonal antibody.

**Detailed Analysis**
**CE IVD / FDA approved assays**

**SP263** (740-4907, Ventana): 12 of 16 (75%) assays were assessed as optimal. Protocols with optimal results were typically based on Heat Induced Epitope Retrieval (HIER) in Cell Conditioning 1 (CC1), efficient heating time 32-72 min., 16 min. incubation of the primary Ab, OptiView as detection system and performed on BenchMark Ultra. Using these protocol settings, 13 of 14 (93%) laboratories produced a sufficient staining result (optimal or good).
One laboratory applied the same protocol settings as described above but also used an additional amplification step, and produced an optimal staining result.

**SP263** (790-4905, Ventana): 34 of 48 (71%) assays were assessed as optimal. Protocols with optimal results were typically based on HIER in CC1, efficient heating time 32-72 min., 12-44 min. incubation of the primary Ab, OptiView as detection system and performed on BenchMark Ultra. Using these protocol settings, 42 of 46 (91%) laboratories produced a sufficient staining result (optimal or good).
Two laboratories applied the same protocol settings as described above but also used an additional Amplification step. Both produced an optimal staining result.

**SP263** (741-4905, Ventana): 2 of 4 assays were assessed as optimal. Protocols with optimal results were based on HIER in CC1, efficient heating time 64 min., 16 min. incubation of the primary Ab, OptiView as detection system and performed on BenchMark Ultra. Using these protocols settings, 4 of 4 laboratories produced a sufficient staining result.

**PD-L1 IHC 22C3 pharmDx** (SK006, Dako): 17 of 24 (71%) protocols were assessed as optimal. Protocols with optimal results were based on HIER using EnVision™ Flex Target Retrieval Solution (TRS) low pH 6.1 (SK006) at 95-99°C for 20-30 min. (PT Link), 20-40 min. incubation of the primary Ab, EnVision Flex+ as the detection system and performed on Autostainer Link 48. Using these protocol settings, 22 of 24 (92%) laboratories produced a sufficient staining result.

**PD-L1 IHC 22C3 pharmDx** (GE006, Dako): 3 of 3 protocols were assessed as optimal. Protocols with optimal results were based on HIER using EnVision™ Flex Target Retrieval Solution (TRS) low pH 6.1 (GV805) at 95-99°C for 40 min. (PT Link), 40 min. incubation of the primary Ab, EnVision Flex+ as the detection system and performed on Omnis.

**PD-L1 IHC 28-8 pharmDx** (SK005, Dako): 2 of 3 protocols were assessed as optimal. Protocols with optimal results were based on HIER using EnVision™ Flex TRS low pH 6.1 at 97°C for 20 min. (PT Link), 30 min. incubation of the primary Ab, EnVison Flex+ as the detection system and performed on Autostainer Link 48. Using these protocol settings, all three laboratories produced a sufficient staining result.

Table 4 summarizes the proportion of sufficient and optimal marks for the most commonly used RTU systems. The performance was evaluated both as "true" plug-and-play systems performed strictly accordingly to the vendor recommendations and by laboratory modified systems changing basal protocol settings. Only protocols performed on the specific IHC stainer device are included.

Table 4. **Comparison of pass rates for vendor recommended and laboratory modified protocols**

| CDx assay* | Vendor recommended protocol settings** | | Laboratory modified protocol settings*** | |
|---|---|---|---|---|
| | Sufficient | Optimal | Sufficient | Optimal |
| Ventana BenchMark Ultra rmAb SP263, **740-4907** | 7/8 (88%) | 5/8 (63%) | 7/8 (88%) | 7/8 (88%) |
| Ventana BenchMark XT, GX, Ultra rmAb SP263, **741-4905** | 4/4 | 2/4 | - | - |
| Ventana BenchMark XT, GX, Ultra rmAb SP263, **790-4905** | 24/27 (89%) | 19/27 (70%) | 20/21 (95%) | 15/21 (71%) |
| Dako Autostainer Link 48+ mAb 22C3 pharmDX, **SK006** | 19/21 (91%) | 15/21 (67%) | 3/3 | 2/3 |
| Dako Omnis mAb 22C3 pharmDX, **GE006** | 3/3 | 3/3 | - | - |
| Dako Autostainer Link 48+ rmAb 28-8 pharmDX, **SK005** | 3/3 | 2/3 | - | - |

*See Table 3 for examples of approval and indications.
**Protocol settings recommended by vendor – Retrieval method and duration, Ab incubation times, detection kit, IHC stainer/equipment.
***Modifications in one or more of above mentioned parameters. Only protocols performed on the specified vendor IHC stainer are included.

**Concentrated antibodies for laboratory developed (LD) assays**
mAb **22C3**: 11 of 42 (26%) protocols were assessed as optimal of which five were stained on the Benchmark (Ventana) and five on the Omnis (Dako) platform.
On BenchMark XT/Ultra (Ventana) the protocols providing optimal results were typically based on a titre of 1:40-50, primary Ab incubation time of 44-64 min., HIER in CC1 (efficient heating time 48 min.) and OptiView as detection system. Using these protocol settings, 10 of 11 (91%) laboratories produced a sufficient staining result.
On Omnis (Dako) the protocols providing optimal results were typically based on a titre of 1:20-30 of the primary Ab, incubation time of 30-45 min., HIER in TRS low pH 6.1 (Dako) at 97°C (efficient heating time 30-50 min.) and EnVision FLEX+ as detection system. Using these protocol settings, 6 of 7 (86%) laboratories produced a sufficient staining result.
In total, 26 of 35 (74%) laboratories used optimal protocol settings and obtained a sufficient staining result.

mAb **E1L3N**: One protocol provided an optimal result. The protocol was based on HIER using a non-alkaline-buffer at 95°C for 20 min. The mAb clone E1L3N was diluted 1:150, incubated for 40 min. at room temp. and using a 3-layer detection system. Only one laboratory used these protocol settings.

rmAb **CAL10**: 3 of 6 (50%) protocols were assessed as optimal. Protocols with optimal results were typically based on a titre of 1:50-100, primary Ab incubation time of 15-45 min., HIER in an alkaline buffer for 20-30 min. using a 3-layer detection system. Using these protocol settings, 3 of 3 laboratories obtained a sufficient staining result.

Table 5. **Optimal results for PD-L1 for the most commonly used antibodies as concentrates on the four main IHC systems\***

| Concentrated antibodies | Ventana/Roche BenchMark GX/XT/Ultra | | Dako/Agilent Autostainer | | Dako/Agilent Omnis | | Leica Bond III/Max | |
|---|---|---|---|---|---|---|---|---|
| | CC1 pH 8.5 | CC2 pH 6.0 | TRS pH 9.0 | TRS pH 6.1 | TRS High pH | TRS Low pH | BERS2 pH 9.0 | BERS1 pH 6.0 |
| mAb clone **22C3** | 5/24 (21%) | - | 0/2 | 1/2 | - | 5/10 (50%) | 0/3 | - |
| mAb clone **E1L3N** | - | - | - | 1/1 | - | - | 0/4 | - |
| rmAb clone **CAL10** | 0/1 | - | - | - | - | - | 2/2 | - |

*Antibody concentration applied as listed above, HIER buffers and detection kits used as provided by the vendors of the respective platforms.
**number of optimal results/number of laboratories using this buffer

**Block construction and assessment challenges**
The tissue micro array (TMA) block constructed for this PD-L1 run consisted of 10 NSCLCs, 2 tonsils and 1 placenta. The NSCLCs were selected so the slides cut from the block would contain 2 NSCLCs being TPS negative (<1% PD-L1 positive tumour cells), 4 NSCLCs of each group: TPS low (≥1-49%) and TPS high (≥50%). Reference slides throughout the block were stained using the companion diagnostic assays 22C3 SK006, 22C3 GE006, 28-8 SK005 (all 3 Dako/Agilent) and SP263 (790-4905, Ventana) for the assessment. In total, four identical blocks was constructed but only slides from the first three blocks were sent to the participants. Tissue core no. 9 was excluded due to lack of tissue in the majority of slides. In few slides tissue core no. 12 was fragmented and not included in the assessment of these slides.

Reviewing the reference slides from the blocks, heterogenic expression of PD-L1 was seen in three of the tumor cores. In tissue core no. 7, initially scored as TPS low (≥1-49%), areas with a TPS high ≥50% was identified. For tissue core no. 11 and 12, initially scored as TPS high ≥50%, areas with TPS low ≥1-49% were observed. During the assessment, TPS categories for each tissue core on the submitted slides were compared to the level in the nearest reference slides of 22C3 pharmDX SK006 (Dako) and/or an in-house validated LD assay based on 22C3.
Heterogeneity in PD-L1 expression is well known in NSCLCs and the assessment in this sense emulated clinical settings. However, the inconsistent expression of PD-L1 in one or more of the included tissue cores was challenging for the assessment settings.

**Comments**
In this fifth NordiQC assessment for PD-L1 (lung), the prevalent feature of an insufficient staining result was a false negative staining result, being observed in 59% of the insufficient staining results (20 of 34). As shown in Table 6, a false negative staining result has been the most common reason for insufficient

staining results in all NordiQC PD-L1 (lung) assessments with an average occurrence of 74%. In this run, 18% (6 of 34) of the insufficient results were caused by a false positive staining result. Poor-signal-to-noise ratio, excessive cytoplasmic staining reaction interfering the interpretation or technical issues were observed in the remaining 23% of the insufficient results (8 of 34).

29% (51 of 176) of the participants obtained a score as Good. In 45% of these (23 of 51), this was due to a general weak staining result or a reduced TPS, but with no change in the TPS-category. In the remaining 55%, excessive background staining, impaired morphology or cytoplasmic staining reaction interfering the interpretation was observed. No obvious reason for this observation could be identified.

Table 6. **Characteristics of insufficient results in the last five NordiQC PD-L1 (lung) runs.**

|  | False Negative (TPS changes from high to low or low to negative) | False Positive (TPS changes from negative to low or low to high) | Other Cause "Technical" |
|---|---|---|---|
| C1 (50% insufficient) | 68% | 15% | 17% |
| C2 (16% insufficient) | 68% | 23% | 5% |
| C3 (9% insufficient) | 100% | - | - |
| C4 (14% insufficient) | 74% | 4% | 22% |
| C5 (20% insufficient) | 59% | 18% | 23% |
| Average (22% insufficient) | 74% | 12% | 14% |

Tissue cores no. 7, 8 and 11 were the most challenging for the laboratories in this assessment and required an accurate and carefully calibrated protocol. The majority of false negative results were especially seen in core 7, changing the TPS category from low to negative, which would change the status for immune therapy using the present guidelines.

The Ventana PD-L1 IHC assays 790-4905/740-4907/741-4905, based on the SP263 clone, was the most widely used assay for demonstration of PD-L1 and provided a pass rate of 91% (62 of 68). Applying protocol settings in compliance with the vendor recommendations, the pass rate was 91% (32 of 35). In comparison, protocols based on laboratory modified protocol settings obtained a pass rate of 90% (26 of 29).
One laboratory used the assay off label on a Leica Bond platform and obtained an insufficient result.
In this C5 run for PD-L1 (lung), 10% (7 of 68) of the laboratories using 790-4905/740-4907/741-4905 were downgraded from optimal to good or borderline due to impaired morphology, irrespective of applying optimal protocol settings. The impaired morphology was especially characterized by indistinct and diffusely contours of nuclei and tissue architecture in general. At present it is impossible to correlate if this was caused by inappropriate section handling or e.g. related to protocol/instrument issues.

The Dako/Agilent 22C3 pharmDx assay SK006 provided an overall pass rate of 92% (22 of 24). When using the recommend protocol settings from Dako, the pass rate was 91% (19 of 21). Three laboratories modified the protocols - all obtained a sufficient result.
Nine laboratories applied the SK006 RTU product on another stainer platform than the Dako Autostainer Link 48, providing a pass rate of 44% (4 of 9). All three laboratories using the Dako/Agilent 22C3 pharmDx assay GE006 for Omnis used the recommended protocol settings and all obtained optimal results.

Grouped together, and using vendor recommended protocol settings, the CE IVD approved PD-L1 IHC assays provided a pass rate of 92% (61 of 66).

Laboratory developed (LD) assays either based on a concentrated Ab, a "non-companion diagnostic approved" RTU format, or an approved companion diagnostic assay not used strictly accordingly to vendor recommendations, were used by 63% (110 of 176) of the participants. For this group a pass rate of 74% (81 of 110) was observed. The overall pass rate for LD assays is significantly lower than the approved assays used as recommended, which underlines the difficulties in validating LD assays.

The mAb clone 22C3 was the most widely used concentrated Ab within a LD assay (n=42) and the pass rate was 74% (31 of 42). This is comparable to the C4 run, where 79% of the LD assays based on this clone provided a sufficient result. In this run, 26% (11 of 42) of the LD assays based on mAb clone 22C3 as concentrate provided an optimal result compared to 75% (18 of 24) using the corresponding companion diagnostic assays SK006 and GE006 by the recommended protocol settings – see Table 4.
24 laboratories used the mAb clone 22C3 on the Ventana BenchMark stainer platform and obtained a pass rate of 71% (17 of 24). 15 protocols were based on OptiView as detection system providing a pass rate of 80% (12 of 15) and 33% optimal (5of 15) optimal. 7 protocols based on 22C3 in combination with

OptiView and Amplification provided a pass rate of 57% (4 of 7), but none optimal. When using OptiView with Amplification, an excessive aberrant granular staining pattern was seen. This pattern was accepted when it did not significantly compromise the interpretation. However, in two cases the interpretation was hampered, giving a result evaluated as insufficient. It is well known from previous assessments in NordiQC, both for other epitopes (general module) and previous PD-L1 runs, that IHC assays based on tyramide amplification can be challenging as especially low-level expressing tissue structures may be negative and if not carefully calibrated, can cause false positive staining result in structures expected to be negative.

10 laboratories used the mAb clone 22C3 as concentrate on the Dako Omnis stainer platform and obtained a pass rate of 80% (8 of 10) and 50% optimal (5 of 10). One of the insufficient results was due to an aberrant staining pattern – most likely caused by contamination. For the second insufficient result no obvious cause could be identified.

**PD-L1 interpretation and scoring consensus:**
Participants were asked to evaluate the percentage of PD-L1 positive tumour cells in each of the 9 NSCLCs included in the assessment.
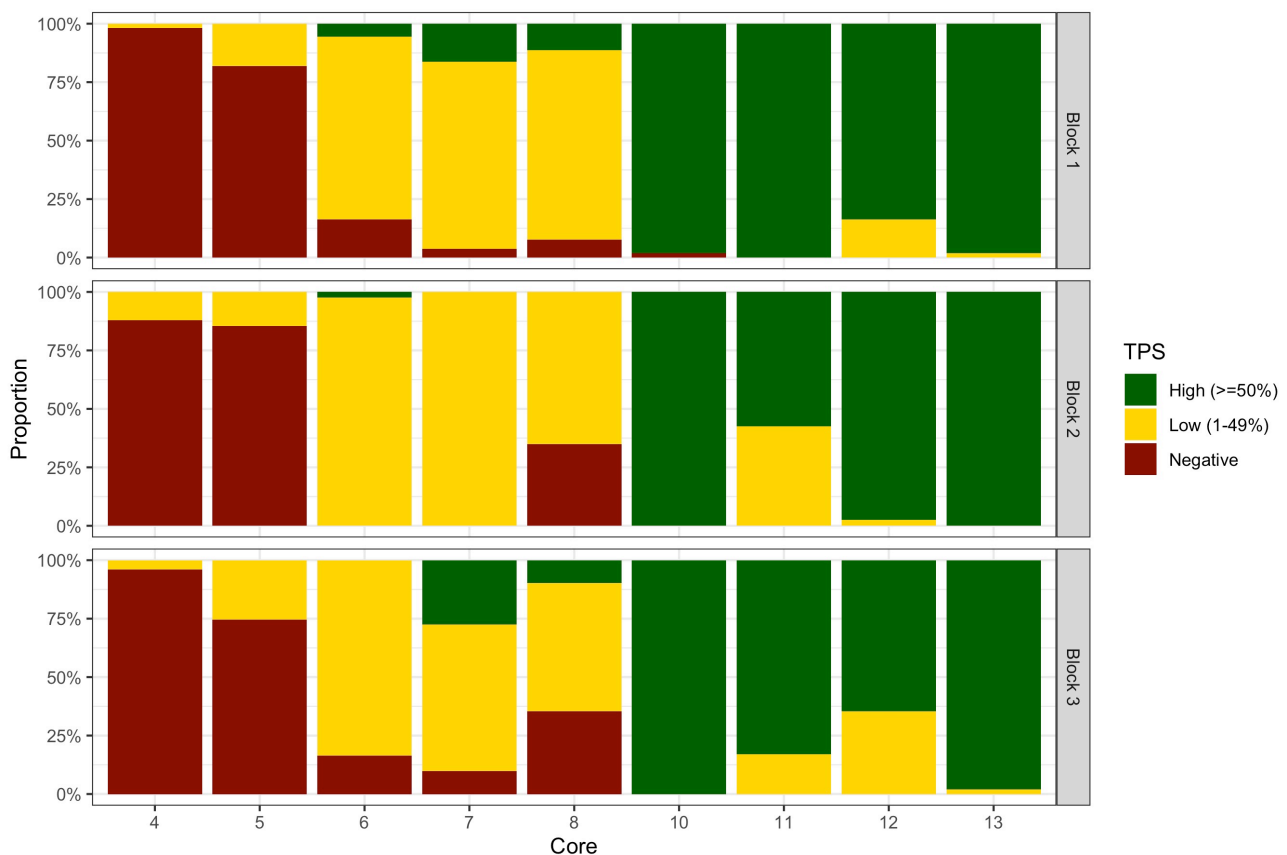


Graph 1. **NordiQC PD-L1 run C5: Tumour proportion scores across tissue core split up by different TMA blocks**

As seen in Graph 1, relatively high consensus rates were observed for the tissue cores 4, 5, 10 and 13 (in all blocks), whereas the consensus rates were significantly lower in the cores 6-8 and 11-12. Tissue core no. 9 was excluded from the assessment.

For the tissue cores 11 and 12, a large group of participants scored those as TPS low (≥1-49%). This could be explained by PD-L1 expression heterogeneity seen in cores no 11-12 across the tissue blocks. For tissue core no 11, all the participants that received slides from block 1 scored the core as TPS high (≥50%), while some slides from block 2 and 3 were scored as TPS low (≥1-49%) (see Graph 1). Core 12 was in block 2 mostly scored as TPS high (≥50%), but from block 1 and 3 large group scored this as TPS low (≥1-49%).
Core 7 was for all labs receiving slides from block 2 scored as TPS low (≥1-49%), while some labs receiving slides from block 1 and 3 scored this block TPS high (≥50%).
The cores no 6 and 8 were challenging, and no significant parameters (including specific TMA blocks/cores) could be identified unraveling the discrepancy in TPS scoring rates between participants own assessments and the results obtained in neighboring NordiQC reference slides.

When stratifying for the assessment marks, analysis revealed that scores among participants that had received an insufficient mark (borderline or poor) reported lower TPS scores than laboratories that had received a sufficient mark (see Graph 2). However, this difference was not statistically significant and a marked overlap in scores between the two groups was seen. Additionally, laboratories that received an insufficient mark tended to have more diverse TPS scores.
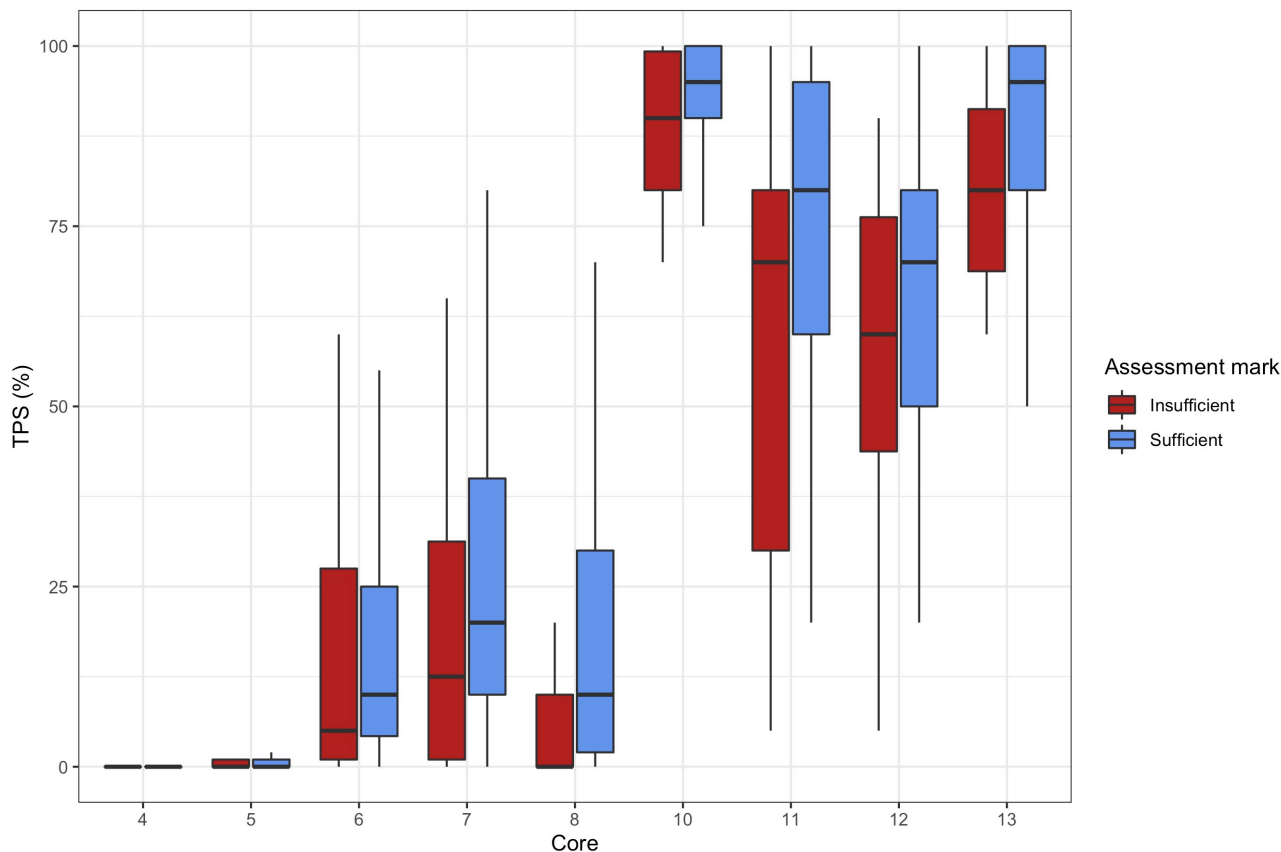


Graph 2. **NordiQC PD-L1 (lung) run C5: interpretation concordance for labs with sufficient vs. insufficient results**

## Controls

Tonsil and placenta were used as positive and negative tissue controls. In this assessment, tonsil was found to be superior to placenta, as tonsil displayed a dynamic and clinical relevant range of PD-L1 expression levels, whereas placenta virtually only contained cells (trophoblasts) with high level PD-L1 expression. Using PD-L1 IHC 28-8 (SK005, Dako/Agilent), 22C3 (GE006 and SK006, Dako/Agilent) or SP263 (790-4905/4907 and 741-4905, Ventana/Roche) and obtaining an optimal staining result, tonsil displayed the following reaction pattern: No staining reaction in the vast majority of lymphocytes including mantle zone and germinal centre B-cells, no staining reaction in superficial epithelial cells, a weak to moderate, typically punctuated membranous staining reaction of the majority of germinal centre macrophages and finally a moderate to strong staining reaction of the majority of epithelial crypt cells. It was observed that SP263 (790-4905/4907 and 741-4905, Ventana/Roche) typically provided a higher proportion of positive inter- and intra-follicular immune cells compared to the Dako/Agilent PD-L1 assays (SK005, SK006 and GE006).

However, it was observed that a fully acceptable staining pattern in tonsil could be obtained together with an insufficient result in the NSCLCs. This underlines the need to identify more reliable positive and negative (tissue) controls for PD-L1 and/or improve the interpretation criteria for a sufficient staining reaction in tonsils e.g. more accurately specify number and intensity of cells expected to be demonstrated.
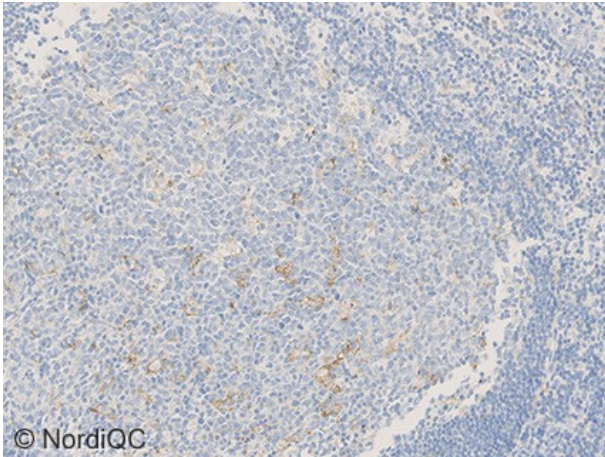
Fig. 1a
Optimal staining result of tonsil (germinal centre) using the pharmDX IHC PD-L1 assay, SK006, Dako/Agilent based on the mAb clone 22C3. Same protocol used in Figs. 2a–5a. The majority of germinal centre macrophages show an at least weak but distinct membranous staining reaction. The vast majority of lymphoid cells are negative.
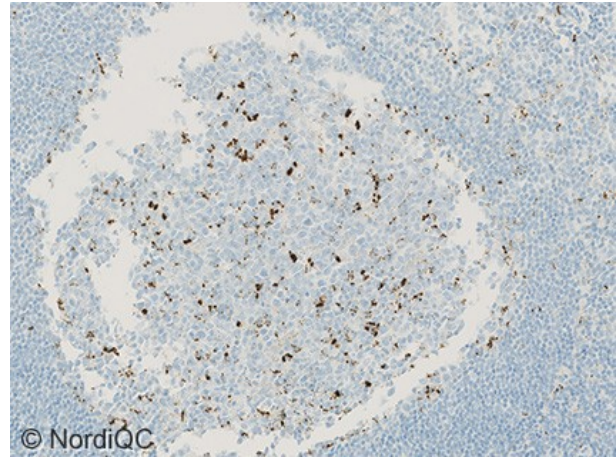


Fig. 1b
Staining result of tonsil (germinal centre) using the rmAb clone CAL10 within a laboratory developed assay. The protocol was based on HIER in an alkaline buffer and a 3-layer detection system. Same protocol used in Figs. 2b and 3b. Many germinal centre immune cells display a strong granular staining reaction, but still was unsuccessfully calibrated – see Figs. 2b and 3b. Compare with optimal staining in Fig.1a.
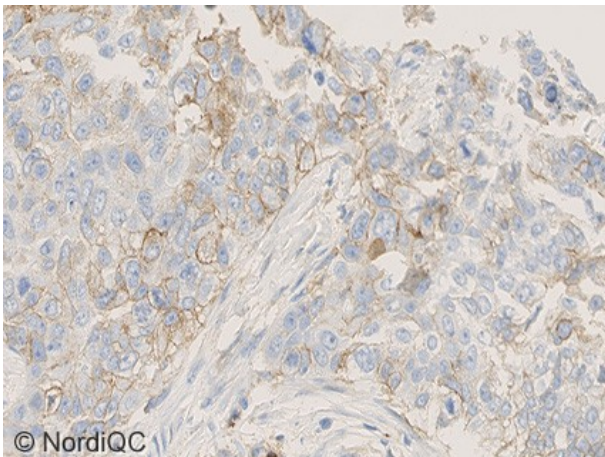


Fig. 2a
Optimal staining result of the NSCLC, tissue core no. 7, using same protocol as in Fig. 1a. Approximately, 30-40% of the neoplastic cells in the whole core show a weak but distinct membranous staining reaction. The tumour was categorized as TPS low (1-49%) and thus eligible for second line immune therapy. The same staining pattern was seen in reference slides.
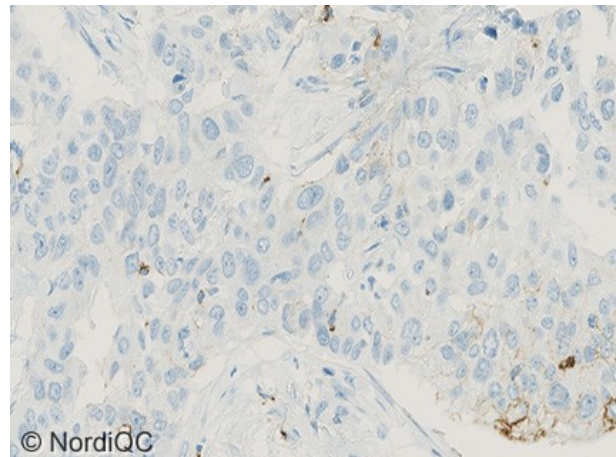


Fig. 2b
Insufficient staining result of the NSCLC, tissue core no. 7, using same protocol as in Fig. 1b.
Virtually all tumour cells are negative providing a TPS below 1%. Compare with optimal staining in Fig. 2a.
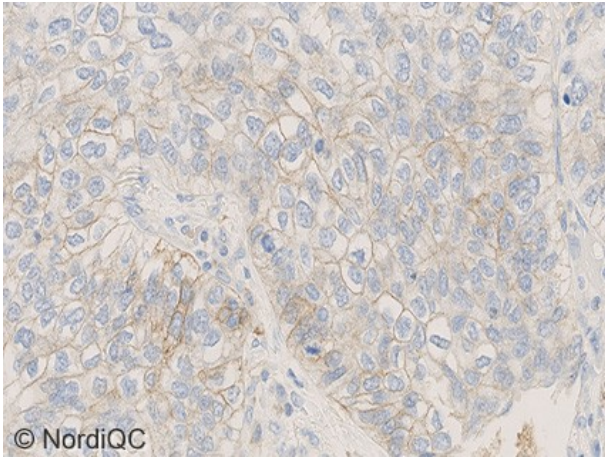
Fig. 3a
Optimal staining result of the NSCLC, tissue core no. 11, using same protocol as in Figs. 1a-2a.
The majority of the neoplastic cells are positive for PD-L1. The tumour was categorized as TPS high (≥ 50%). The same staining pattern was observed in reference slides.
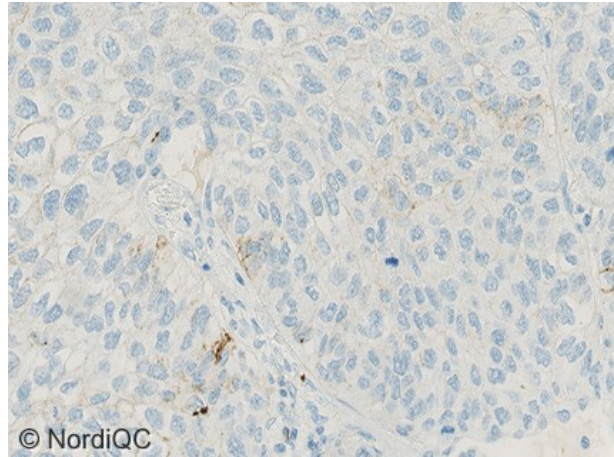


Fig. 3b
Insufficient staining result of the NSCLC, tissue core no. 11, using same protocol as in Figs. 1b-2b.
The proportion of positive tumour cells is significantly reduced compared with the optimal result in Fig. 3a, providing a result as TPS low (1-49%) and not TPS high as expected.
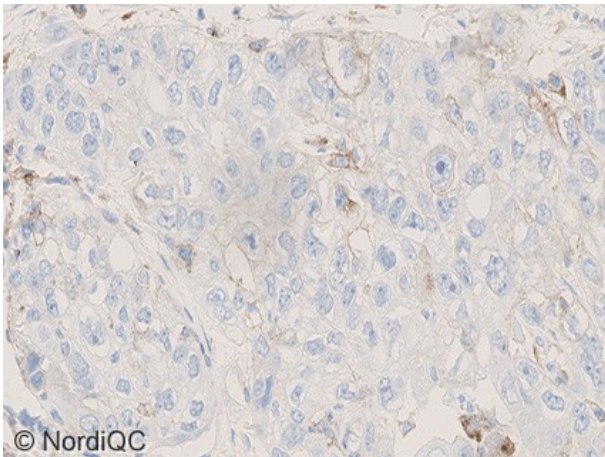


Fig. 4a
Optimal staining result of the NSCLC, tissue core no. 7, using same protocol as in Figs. 1a-3a. In this area and in the whole core, the tumour was categorized as TPS low (1-49%) and thus eligible for second line immune therapy. A weak but distinct membranous staining reaction is seen in the tumour cells. The same staining pattern was observed in reference slides.
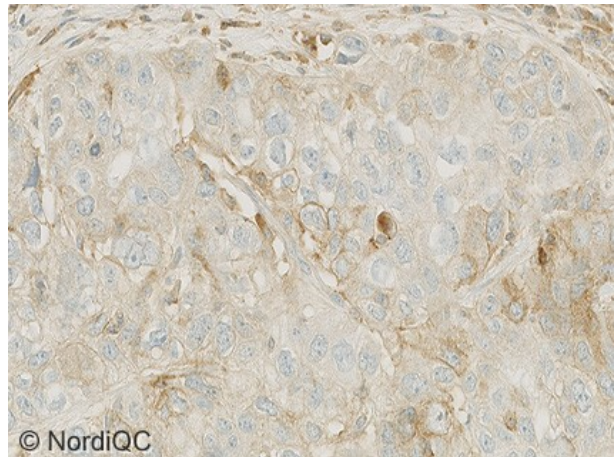


Fig. 4b
Insufficient staining result of the NSCLC, tissue core no. 7, using rmAb SP263 (740-4907, Ventana) with modified protocol settings.
A significantly increased number of positive neoplastic cells is obtained changing the TPS category from low to high. Compare with optimal result in Fig. 4a, same area.
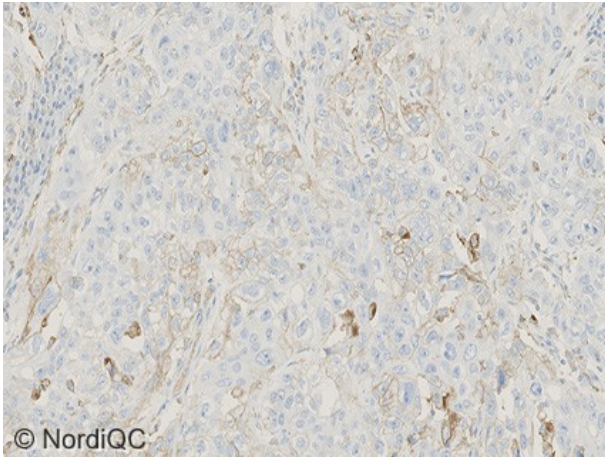
Fig. 5a
Optimal staining result of the NSCLC, tissue core no. 7, using same protocol as in Figs. 1a-4a. Approximately, 30-40% of the neoplastic cells in the whole core show a weak but distinct membranous staining reaction. The tumour was categorized as TPS low (1-49%) and thus eligible for second line immune therapy.
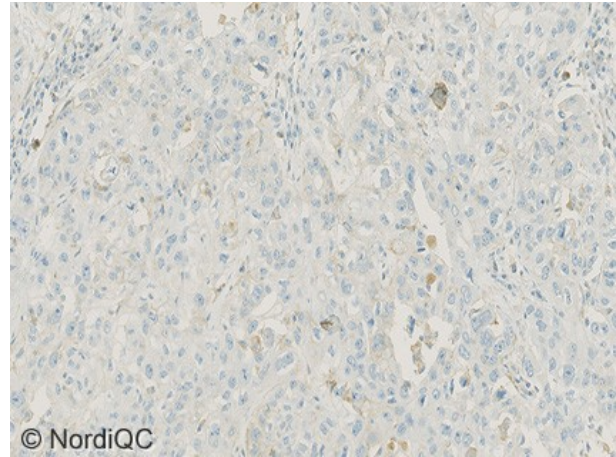


Fig. 5b
Staining result of the NSCLC, tissue core no. 7, using SK006 (Dako) on a BenchMark Ultra (Ventana) evaulated as Good. A reduced staining intensity and proportion of cells demonstrated is seen compared to the optimal and expected result shown in Fig. 5a. Despite a reduced analytical sensitivity was seen, all tumour cores were assigned in the correct TPS category.

HLK/SN/LE/RR/MV 08.07.2019