PDL11/22: Programmed Death Ligand 1

The terminology: *We adhere to the terminology of ISO 17043 and ISO 15189 wherever possible.*
Typing conventions: *We are using commas as a decimal separator and dates in day.month.year format.*

Please visit the web page
http://www.sekk.cz/PDL1
to find complete information about PDL1 programme in one location.

Introduction

This EQA round was completed according to the document ***EQA Plan 2022***.
The scientific background of the CRC programme is under the control of the **European Society of Pathology** (ESP, www.esp-pathology.org) by means of the scientific supervisor (see bottom of this report) nominated by the ESP. Also expert laboratories (see the paragraph *Assigned values*) were selected on the basis of the recommendations of the ESP.
The task of the participant was:

1. Perform immunohistochemical staining of the physical slides using the procedure they routinely use in their laboratory.
2. Examine all primary samples (both on physical and virtual slides) – this includes calculation of the required score and final result (negative/positive).
3. Report the following information (using the web application):
   - The method used for staining.
   - Quantitative (the score) and qualitative (negative/positive) results.
   - Cut-off limits used.

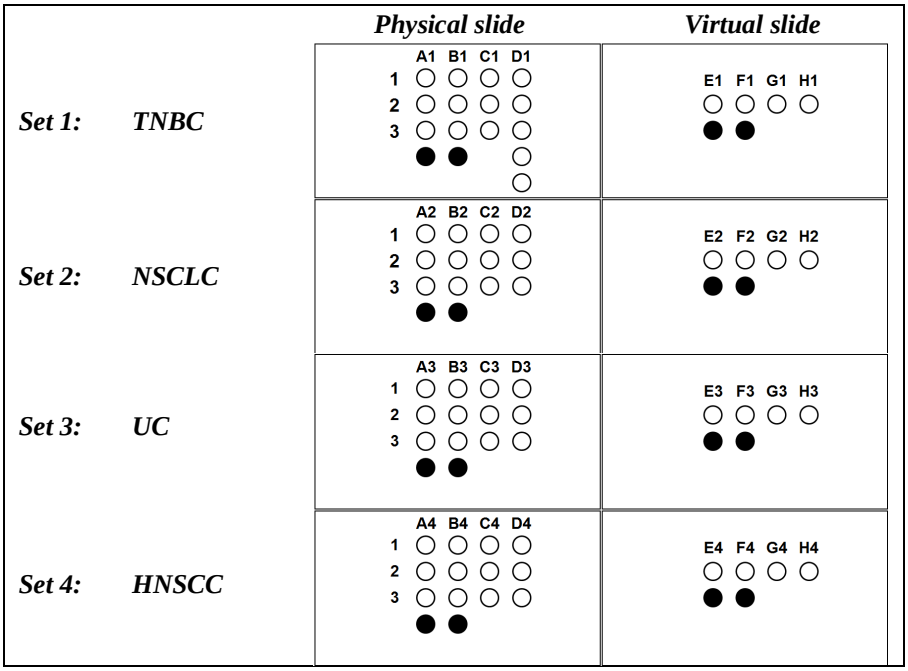
Participants

There were 56 participants in this round from 18 countries (the list of countries you can find on the website).

Samples

The samples were prepared by subcontractor. The samples were divided into 4 sets (subschemes), each set relates to one tumour type and contains one physical slide (unstained TMA section) and one virtual slide (stained TMA section).

The picture shows the map of the samples for this round:



Each column in the TMA block represents one primary sample. For each primary sample, there are a minimum of 3 cores in the TMA block of the physical slide (this "redundancy" eliminates potential problems associated with missing or damaged tissue).

Virtual PD-L1 stained slides were available on our virtual microscopy website.

Abbreviations used:

- TNBC ... triple-negative breast carcinoma
- NSCLC ... non-small cell lung cancer
- UC ... urothelial carcinoma
- HNSCC ... head-and-neck squamous cell carcinoma

PDL11/22: Programmed Death Ligand 1

The samples were shipped to the participants together with the documentation in one package via a courier service. The time of the delivery ranged from 1 to 2 days in most cases (based on the participant's country), no damage or loss of the shipments occurred, all parcels were delivered.
The participants were allowed to order spare samples in case of any sample damage in their laboratory.

Assigned values (AVs)

The AVs (expected results) for the particular **primary samples** were obtained from the consensus of the participants. In accordance with ISO 17043 classification, we have used the **CVP** (consensus value from the participants) type of AV.
There are 2 rules applied in the process of establishing the AVs:

- Consensus is reached if **80 % or more** participants agree on a result.
- Minimal size of the group assessed is **n = 5** (in case of less than 5 results the group is not assessed).

Additionally, we also received the results of 3 expert laboratories:

- University Erlangen-Nürnberg, Institut für Pathologie, Erlangen, Germany
- University Hospital Zurich, Department of Pathology and Molecular Pathology, Zurich, Switzerland
- Erasmus Universitair Medisch Centrum, Department of Pathology and Clinical Bioinformatics, Rotterdam, The Netherlands

Expert laboratories tested each tumour type (all primary samples) mentioned above.
Expert laboratories tested all samples as unknown. The task for each expert laboratory was to test the sample and report the results back to the SEKK (thus not only to confirm the results suggested by SEKK). In other words: expert laboratories tested the samples under the same conditions as regular participants.
We used the results of the expert laboratories to confirm the quality of the samples.

Evaluation of the results

As mentioned above, the participants had to calculate and report the appropriate score (quantitative result) and using the cut-off (which they had to report together with the results) decide whether the sample is negative or positive (qualitative result).
The assessment is based on the qualitative results and consists of 2 steps:

Step 1)

The results of all primary samples (A, B .... H) were sorted into these categories from the point of view of the performance assessment:

Category	***Explanation***
Expected (correct) result, marked >>> in the reports	This is the result that we expected to be found by the participants. This result is optimal for the patient's treatment. It is the result identical to the consensus of the experts.
Not assessed, marked ± in the reports	This category indicates that it would not be possible to establish the AV (the consensus among the laboratories was not reached). Without having the AV we are not able to classify the participant's result as "correct" or "incorrect".
Incorrect result	Any result which is neither "Expected" nor "Not assessed".

Step 2)

On the basis of the primary samples assessment each slide (physical, virtual) of each set (TNBC, NSCLC, UC, HNSCC) was assessed (in EQA terminology there were 8 tests assessed: TNBC physical slide, TNBC virtual slide NSCLC physical slide etc.).
The slide (one test) consists of 4 primary samples and the assessment of the slide depends on the number of the assessable primary samples on the slide this way:

- If all 4 primary samples are assessable then the slide is assessed as successful if the results of 3 or 4 primary samples are correct (that is an error in one primary sample is tolerated).
- If 3 or fewer primary samples are assessable then the slide is assessed as successful if the results of all assessed primary samples are correct (that is an error in any primary sample is not tolerated).

General problems observed in this EQA round – impulses to improve

Cut-offs used

We gave a freedom to the participants to use individual cut-offs to classify the results into negative and positive groups. We expected that different cut-offs will be used in case of the NSCLC but different cut-offs occurred also in other tumour types.
Conclusion: We will prescribe cut-offs to be used in future EQA rounds of the PDL1 programme.

Classification based on the cut-off

Suppose that cut-off is 1 %. It seems that some participants are applying the rule "negative if < 1 %" (less than) and other the rule "negative if ≤ 1 %" (less than or equal).

PDL11/22: Programmed Death Ligand 1

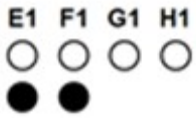
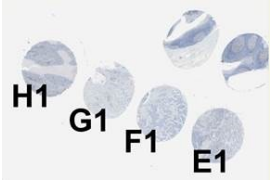
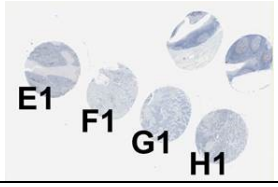
Conclusion: See the paragraph above – we'll prescribe the cut-offs in detail in the next EQA rounds.

Interpretation problems

It was expected that if the participant reports cut-off = 10 and score = 5 then their conclusion (qualitative result) will be "negative". Surprisingly in similar situations some participants classified the result as "inconclusive".
Conclusion: We recommend to use the cut-off as a strict border (regardless the fact that the qualitative result – score – has an uncertainty and this uncertainty is surely not negligible – we can try to consider the uncertainty in future).

Orientation of the virtual slides

It seems that some participants (we estimate 7 labs) reported the results of the virtual slides in wrong order.
Example for TNBC

The map of the slide was prescribed as:	
We gave this instruction on the Virtual microscopy web site	*Please follow **the map** printed in the document **General instructions** (you received it with the samples) to identify individual positions (E1, F1 etc.).*
Real virtual slides looked like:	
Expected behaviour:	Slide map aligned using tonsils and then the samples reported as: 
Wrong approach:	Some participants probably reported the results this way: 

Conclusion: We did not expect to run into this difficulty but we will improve the slide map in the next EQA round to match the virtual slide better (not requiring a rotation).

The results – each set (tumour type) and slide are discussed separately

When reading this report please view also your result sheet or summary statistics placed on the web – you can find the complete overview of the results in these documents (including the primary samples where the consensus of the experts was not reached).

PDL11/22: Programmed Death Ligand 1

Set 1: TNBC

Cut-offs reported by the participants:

Cut-off	*Count*
1 %	24
10 %	1

Physical slide (samples A1, B1, C1, D1)

Number of the participants: 25
Assigned values (AV):

Sample	*AV*	*Performance of the participants*
A1	positive	25 results were assessed and 22 of them (88 %) were successful
B1	not reached	
C1	not reached	
D1	not reached	

Assessment: Only the sample A1 was assessable (consensus reached). We assessed all results regardless of the cut-off reported. This is because the score of the sample A1 was about 20 % and thus it does not matter whether cut-off 1 % or 10 % was used.
Assessed samples - details:

Sample	*Qualitative results*		*Quantitative results (score)*	
	result	*count*		*IC [%]*
A1	Negative	2	Minimum	0
	Inconclusive	1	Robust mean	18
	Positive	22	Maximum	70

Notes on the not-assessed samples:

- B1: Robust mean of the IC score was 3 %. We received an unexpected result from one participant who reported a cut-off = 10 % and a result IC = 40 % - absolutely the highest result for this sample.
- C1: Robust mean of the IC score was 2 %. We received an unexpected result from one participant (the same as above) who reported cut-off = 10 % and a result IC = 30 % - absolutely the highest result for this sample.

In case of B1 and C1 samples 24 participants reported a cut-off = 1 % and 19 of these (79 %) reported positive results – thus the consensus was not reached by a narrow margin.

- D1: Most participants reported a score between 0 and 10 %. Surprisingly, 2 participants reported an IC = 0 % and evaluated it as "inconclusive".

Whole slide assessment: The slide assessment was based only on the sample A1 (as the samples B1, C1 and D1 were not assessable) and thus the success rate for the whole slide is identical to the success rate of the sample A1 (88 %) – this is a very good result.

Virtual slide (samples E1, F1, G1, H1)

Number of the participants: 25
Assigned values (AV):

Sample	*AV*	*Performance of the participants*
E1	positive for cut-off = 1 not reached for cut-off = 10	24 results were assessed and 24 of them (100 %) were successful
F1	negative	25 results were assessed and 22 of them (88 %) were successful
G1	not reached	
H1	positive not reached for cut-off = 10	24 results were assessed and 24 of them (100 %) were successful

Assessment: 3 samples were assessable (consensus reached). In case of E1 and H1 samples we assessed only the results of the participants who reported a cut-off = 1 %. We did not assess the results of one participant who reported a cut-off = 10 % because the average score (12 %) was near this cut-off and the group was small (only one participant).
For the sample F1 we assessed all results regardless of the cut-off reported. This is because the score of the sample F1 was near 0 % and thus it does not matter whether cut-off 1 % or 10 % was used.
Assessed samples - details:

Sample	*Qualitative results*		*Quantitative results (score)*	
	result	*count*		*IC [%]*
E1	Negative	-	Minimum	*)
	Inconclusive	-	Robust mean	12
	Positive	24	Maximum	*)

PDL11/22: Programmed Death Ligand 1

Sample	*Qualitative results*		*Quantitative results (score)*	
	result	*count*		*IC [%]*
F1	Negative	22	Minimum	*)
	Inconclusive	-	Robust mean	0
	Positive	3	Maximum	*)
H1	Negative	-	Minimum	*)
	Inconclusive	-	Robust mean	12
	Positive	24	Maximum	*)

*) Minimum/maximum not given (significantly influenced by the participants who reported the results in the wrong order).

Notes on the not-assessed samples:

- G1: Most participants reported a score between 0 and 5 %.

Whole slide assessment: The slide assessment was based on the samples E1, F1 and H1 (as the sample G1 was not assessable) and only those participants succeeded who were successful in all the samples assessed. Total success rate was 88 % which this is a good result.

Set 2: NSCLC

Cut-offs reported by the participants:

Cut-off	*Count*
1 %	37
10 %	1
50 %	14

Physical slide (samples A2, B2, C2, D2)

Number of the participants: 52

Assigned values (AV):

Sample	*AV*	*Performance of the participants*
A2	positive	52 results were assessed and 52 of them (100 %) were successful.
B2	positive	52 results were assessed and 52 of them (100 %) were successful.
C2	not reached	
D2	negative	52 results were assessed and 50 of them (96 %) were successful.

Assessment: 3 samples were assessable (consensus reached). We assessed all results regardless of the cut-off reported. This is because the score of the positive samples was above 90 % and the score of the negative sample was near 0 % and thus it does not matter whether cut-off 1 % or 10 % or 50 % was used.

Assessed samples - details:

Sample	*Qualitative results*		*Quantitative results (score)*	
	result	*count*		*TPS [%]*
A2	Negative	-	Minimum	50
	Inconclusive	-	Robust mean	94
	Positive	52	Maximum	100
B2	Negative	-	Minimum	5
	Inconclusive	-	Robust mean	97
	Positive	52	Maximum	100
D2	Negative	50	Minimum	0
	Inconclusive	-	Robust mean	0
	Positive	2	Maximum	1

Notes on the not-assessed samples:

- C2: Most participants reported the score between 0 and 5 %.

Whole slide assessment: The slide assessment was based on the samples A2, B2 and D2 (as the sample C2 was not assessable) and only those participants succeeded who were successful in all the samples assessed. Total success rate was 96 % which is an excellent result.

Virtual slide (samples E2, F2, G2, H2)

Number of the participants: 51

Assigned values (AV):

Sample	*AV*	*Performance of the participants*
E2	not reached	
F2	positive	51 results were assessed and 46 of them (90 %) were successful
G2	negative	51 results were assessed and 44 of them (84 %) were successful
H2	positive	51 results were assessed and 24 of them (100 %) were successful

PDL11/22: Programmed Death Ligand 1

Assessment: 3 samples were assessable (consensus reached). We assessed all results regardless of the cut-off reported. This is because the score of the positive samples was about 90 % and the score of the negative sample was near 0 % and thus it does not matter whether cut-off 1 % or 10 % or 50 % was used.
Assessed samples - details:

Sample	*Qualitative results*		*Quantitative results (score)*	
	result	*count*		*IC [%]*
F2	Negative	4	Minimum	*)
	Inconclusive	1	Robust mean	87
	Positive	46	Maximum	*)
G2	Negative	43	Minimum	*)
	Inconclusive	1	Robust mean	0
	Positive	7	Maximum	*)
H2	Negative	2	Minimum	*)
	Inconclusive	-	Robust mean	99
	Positive	49	Maximum	*)

*) Minimum/maximum not given (significantly influenced by the participants who reported the results in the wrong order).

Notes on the not-assessed samples:

- E2: There was a wide spread of the scores reported as almost all participants reported the score between 0 and 70 %.

Whole slide assessment: The slide assessment was based on the samples F2, G2 and H2 (as the sample E2 was not assessable) and only those participants succeeded who were successful in all the samples assessed. Total success rate was 82 % which is a good result.
We identified 4 participants who very probably reported the results in the wrong order.

Set 3: UC

Cut-offs reported by the participants:

Cut-off	*Count*
10	34
25	2

Physical slide (samples A3, B3, C3, D3)

Number of the participants: 36
Assigned values (AV):

Sample	*AV*	*Performance of the participants*
A3	positive	36 results were assessed and 36 of them (100 %) were successful
B3	positive	36 results were assessed and 36 of them (100 %) were successful
C3	negative	36 results were assessed and 33 of them (92 %) were successful
D3	negative	36 results were assessed and 29 of them (81 %) were successful

Assessment: All 4 samples were assessable (consensus reached). We assessed all results regardless of the cut-off reported. This is because the score of the positive samples was high enough (A3 about 90 and B3 about 40) and the score of the negative samples was low enough (C3 about 2 and D3 about 6) and thus it does not matter whether cut-off 10 or 25 was used.
Assessed samples - details:

Sample	*Qualitative results*		*Quantitative results (score)*	
	result	*count*		*CPS [-]*
A3	Negative	-	Minimum	12
	Inconclusive	-	Robust mean	90
	Positive	36	Maximum	100
B3	Negative	-	Minimum	10
	Inconclusive	-	Robust mean	37
	Positive	36	Maximum	100
C3	Negative	33	Minimum	0
	Inconclusive	-	Robust mean	1,6
	Positive	3	Maximum	15
D3	Negative	29	Minimum	0
	Inconclusive	1	Robust mean	5,8
	Positive	6	Maximum	100

Whole slide assessment: To assess whole slide the rule "*3 correct results of 4 must be achieved by the participant*" was used. Based on this rule 94 % of the participants succeeded. This is very good result.

PDL11/22: Programmed Death Ligand 1

Despite very good success rates we observed 1 outlier in scores reported for the assessed samples:

- D3: All participants except one reported a score below 30, one participant reported 100.

Virtual slide (samples E3, F3, G3, H3)

Number of the participants: 36
Assigned values (AV):

Sample	*AV*	*Performance of the participants*
E3	positive	36 results were assessed and 35 of them (97 %) were successful
F3	positive	36 results were assessed and 33 of them (92 %) were successful
G3	negative	36 results were assessed and 33 of them (92 %) were successful
H3	not reached	

Assessment: 3 samples were assessable (consensus reached). We assessed all results regardless of the cut-off reported. This is because the score of the positive samples was high enough (E3 above 90 and F3 about 30) and the score of the negative sample G3 was near 0 and thus it does not matter whether cut-off 10 or 25 was used.
Assessed samples - details:

Sample	*Qualitative results*		*Quantitative results (score)*	
	result	*count*		*CPS [-]*
E3	Negative	1	Minimum	*)
	Inconclusive	-	Robust mean	97
	Positive	35	Maximum	*)
F3	Negative	3	Minimum	*)
	Inconclusive	-	Robust mean	29
	Positive	33	Maximum	*)
G3	Negative	33	Minimum	*)
	Inconclusive	-	Robust mean	0
	Positive	3	Maximum	*)

*) Minimum/maximum not given (significantly influenced by the participants who reported the results in the wrong order).

Notes on the not-assessed samples:

- H3: There was a wide spread of the scores reported - almost all participants reported a score between 0 and 30.

Whole slide assessment: The slide assessment was based on the samples E3, F3 and G3 (as the sample H3 was not assessable) and only those participants succeeded who were successful in all the samples assessed. Total success rate was 92 % which is a very good result.
We identified 3 participants who very probably reported the results in the wrong order.

Set 4: HNSCC

Cut-offs reported by the participants:

Cut-off	*Count*
1	30
10	4
20	3

Physical slide (samples A4, B4, C4, D4)

Number of the participants: 37
Assigned values (AV):

Sample	*AV*	*Performance of the participants*
A4	not reached	
B4	positive	37 results were assessed and 36 of them (97 %) were successful
C4	not reached	
D4	positive	37 results were assessed and 37 of them (100 %) were successful

Assessment: 2 samples were assessable (consensus reached). We assessed all results regardless of the cut-off reported. This is because the score of the positive samples was high enough (B4 about 40 and D4 about 80) and thus it does not matter whether cut-off 1 or 10 or 20 was used.
Assessed samples - details:

Sample	*Qualitative results*		*Quantitative results (score)*	
	result	*count*		*CPS [-]*
B4	Negative	1	Minimum	5
	Inconclusive	-	Robust mean	41
	Positive	36	Maximum	85

PDL11/22: Programmed Death Ligand 1

Sample	*Qualitative results*		*Quantitative results (score)*	
	result	***count***		***CPS [-]***
D4	Negative	-	Minimum	1,2
	Inconclusive	-	Robust mean	80
	Positive	37	Maximum	100

Notes on the not-assessed samples:

- A4: There was a wide spread of the scores reported by the participants – 10 participants reported a score below 1 and 7 participants reported a score 10 or more (one reported the CPS = 100).
- C4: 17 participants reported a score below 1 and the rest (20 participants) reported a score between 1 and 10.

Whole slide assessment: The slide assessment was based on the samples B4 and D4 (as the samples A4 and C4 were not assessable) and only those participants succeeded who were successful in both the samples assessed. Total success rate was 97 % which is an excellent result.
Despite the high success rate we observed 2 outliers in scores reported for the assessed samples:

- B4: All participants except one reported a score 10 or higher, one participant reported 5.
- D4: All participants except one reported a score 20 or higher, one participant reported 1,2.

Virtual slide (samples E4, F4, G4, H4)

Number of the participants: 37
Assigned values (AV):

Sample	***AV***	***Performance of the participants***
E4	positive	37 results were assessed and 34 of them (92 %) were successful
F4	positive not reached for cut-off = 20	34 results were assessed and 30 of them (88 %) were successful
G4	not reached	
H4	not reached	

Assessment: 2 samples were assessable (consensus reached).
We assessed sample E4 regardless of the cut-off reported. This is because the score of this positive sample was about 85 and thus it does not matter whether cut-off 1 or 10 or 20 was used.
In case of the sample F4 the score was about 20 and thus we did not assess the results of 3 participants who reported cut-off = 20 (average score is very close to the cut-off and the group is too small); all other results were assessed.
Assessed samples - details:

Sample	*Qualitative results*		*Quantitative results (score)*	
	result	***count***		***CPS [-]***
E4	Negative	3	Minimum	*)
	Inconclusive	-	Robust mean	85
	Positive	34	Maximum	*)
F4	Negative	4	Minimum	*)
	Inconclusive	-	Robust mean	18
	Positive	30	Maximum	*)

Notes on the not-assessed samples:

- G4: Almost all participants reported a score between 0 and 10.
- H4: There was a wide spread of the scores reported as almost all participants reported a score between 0 and 40.

Whole slide assessment: The slide assessment was based on the samples E4 and F4 (as the samples G4 and H4 were not assessable) and only those participants succeeded who were successful in all the samples assessed. Total success rate was 86 % which is a good result.
We identified 5 participants who very probably reported the results in the wrong order.

Conclusion

This was the first round of the PDL1 programme. We have to congratulate the participants on their results because the total success in all slides was very good.
All opportunities for improvement that we identified are discussed in detail in the section "*General problems*" on page 2. It is very probable that after applying these improvements the results of the autumn round will be even better than in the current round.

Physical slides

The performance of the participants was very good.
A few participants reported (as a text comment) that some cores in a TMA column were missing – this limitation was to be expected and that is why more cores of each primary sample were placed in each column.

PDL11/22: Programmed Death Ligand 1

Virtual slides

Most mistakes arose from the wrong order of the results reported.

As a summary this round was very successful and the results achieved are very promising.

Scientific supervision: Dr. Jan von der Thüsen MA MBBS PhD
Erasmus Universitair Medisch Centrum, Pathology
Dr. Molewaterplein 40
3015 GD Rotterdam, The Netherlands
e-mail: j.vonderthusen@erasmusmc.nl

Supplements

As a supplement to this report individual participants receive:

Name of supplement	***Remark***
Confirmation of attendance	Issued only to those participants who sent us the results.
Certificate of approval	Issued only to those participants who passed successfully.
Result sheet (qualitative results)	Issued only to those participants who sent us the results.
Histograms (quantitative results)	Only for the quantitative results.

The supplements are identified by their name, EQA round identification and participant code and are intended for the needs of the participant.

Additional information

The final report, with the exception of the supplements, is public. Further information is freely available to the participants and other professionals on www.sekk.cz, in particular:

- The summary of the results of this round, including this final report.
- The document ***EQA Plan*** (contains information that applies both to this round and also the EQA in general).
- Explanation of the content of the particular supplements mentioned above.
- Contact to the EQA provider and the EQA coordinator and the list of all supervisors, including contacts.