Inter-reader Reliability of Programmed Death Ligand-1 (PD-L1) Scoring Using the VENTANA PD-L1 (SP263) Assay in Non-Small Cell Lung Cancer (NSCLC)

Gareth H. Williams[1], Andrew G. Nicholson[2], David R. J. Snead[3], Erik Thunnissen[4], Sylvie Lantuejoul[5], Paul Cane[6], Keith M. Kerr[7], Marco Loddo[1], Marietta L. Scott[8], Paul W. Scorer[8], Craig Barker[8]

[1]Oncologica UK Ltd., Cambridge, UK; [2]Royal Brompton and Harefield NHS Foundation Trust and National Heart and Lung Institute, Imperial College, London, UK; [3]University Hospitals Coventry and Warwickshire NHS Trust, Coventry, UK; [4]VU University Medical Center, Amsterdam, The Netherlands; [5]Centre Léon Bérard UNICANCER and Grenoble Alpes University, Lyon, France; [6]Guy's and St Thomas' NHS Foundation Trust, London, UK; [7]Aberdeen University Medical School and Aberdeen Royal Infirmary, Aberdeen, UK; [8]Precision Medicine Laboratories, Precision Medicine and Genomics, IMED Biotech Unit, AstraZeneca, Cambridge, UK

155P

Background

- Multiple immunotherapies targeting the immune checkpoint programmed cell death-1/programmed cell death ligand-1 (PD-1/PD-L1) pathway are approved for the treatment of patients with various cancers, including non-small cell lung cancer (NSCLC) and urothelial carcinoma (UC)[1]
- Clinical data from multiple studies across several indications have demonstrated that PD-L1 expression level can be used to enrich for responders to PD-1/PD-L1 targeted monotherapy[2-4]
- The VENTANA PD-L1 (SP263) Assay is an immunohistochemical diagnostic assay for anti-PD-L1 immune checkpoint inhibitors[5]
  - It is approved by the US FDA as a complementary diagnostic for use with durvalumab in UC[6]
  - In Europe, it is CE marked for use with durvalumab in NSCLC and UC, and with pembrolizumab and nivolumab in NSCLC[7]
- There is a need to assess the inter-observer reproducibility between pathologists for the VENTANA PD-L1 (SP263) Assay, to determine PD-L1 expression level at multiple clinically relevant PD-L1 cut-offs and scoring algorithms used with anti-PD-1/PD-L1 therapies in NSCLC

Study objectives

- The primary objective is to investigate inter-observer reliability, applied to PD-L1 scoring of tumour cells (TCs) in NSCLC tumour samples, using the VENTANA PD-L1 (SP263) Assay
- The exploratory objective is the inter-observer reliability of PD-L1 scoring in immune cell (IC) infiltrates in NSCLC tumour samples using the VENTANA PD-L1 (SP263) Assay

Methods

- Six expert pulmonary pathologists from different European sites independently scored 200 NSCLC tumour samples stained using the VENTANA PD-L1 (SP263) Assay (Ventana Medical Systems, Inc.)
  - Archival, commercially sourced, formalin-fixed, paraffin-embedded resections were selected to represent the dynamic range of PD-L1 expression
- The proportion of TCs with PD-L1 membrane staining (TC score) was scored based on pre-defined intervals (<1%, 1–4%, 5–9%, 10–15%, etc.). The proportion of ICs present as a proportion of the tumour environment (PIC value) and the percentage of ICs staining positive for PD-L1 (IC score) were recorded as continuous variables
  - PD-L1 expression level was evaluated at cut-offs of ≥1%, ≥25% and ≥50% for TC score and at a cut-off of ≥25% for IC score
- PD-L1 staining in alveolar macrophages was excluded from assessment by all pathologists

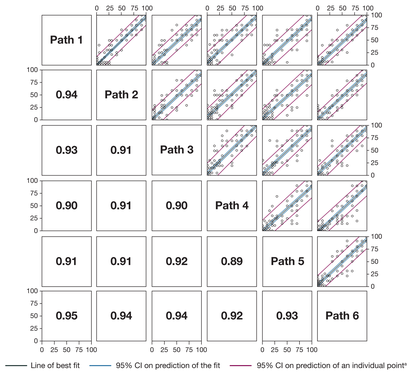
Statistical analyses

- Pathologist concordance for PD-L1 classification across TC ≥1%, ≥25% and ≥50% was determined by generating tables of agreement in PD-L1 expression level between pathologists and calculating the overall percent agreement (OPA), average positive agreement (APA) and average negative agreement (ANA) and the lower bound of the 95% confidence interval (CI)
  - Percent agreement data for IC ≥25% cut-off were not included as very few samples were above the cut-off, rendering the analysis less informative
- Reliability across pathologists for the TC score, PIC value and IC score was determined using the lower bound of the 95% CI on the intra-class correlation coefficient[8] (ICC; single-observer, absolute-agreement, 2-way random effects model)
  - Reliability was defined as poor (<0.5), moderate (≥0.5 to <0.75), good (≥0.75 to <0.9) or excellent (≥0.9)[8]
- Pathology reliability in classifying patient samples as PD-L1 high or low was assessed using Fleiss' kappa
  - Reliability was interpreted using kappa as poor (<0), slight (0.01–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80) or almost perfect (0.81–1.00)[9]
- All statistical analyses were performed using R version 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria)

Results

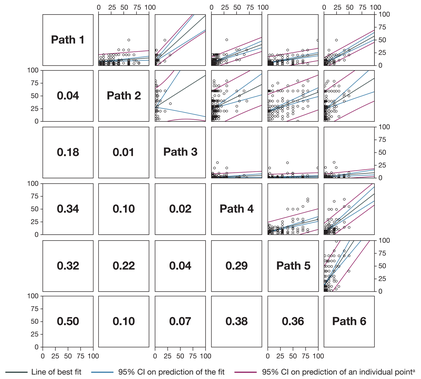
- TC scoring showed strong concordance between pathologists across cut-offs (**Table 1**)
  - OPA was ≥90% for TC ≥1% and >94% for both TC ≥25% and ≥50%
- TC scoring showed strong 1:1 pair-wise correlations (R^2 >0.89), with no systematic bias between pathologists (**Figure 1**)
- Systematic and substantial differences in PIC values (R^2 range: 0.01–0.50) and IC scores (R^2 range: 0.26–0.46) were noted between pathologists with poor pair-wise correlations (**Figures 2** and **3**)
- ICC indicated 'excellent' reliability for TC scoring between pathologists (ICC = 0.96) (**Table 2**)
  - Reliability was 'poor' between pathologists for both PIC values (ICC = 0.21) and IC scores (ICC = 0.48)
- Fleiss' kappa indicated 'substantial' agreement between pathologists for TC ≥1% and 'almost perfect' agreement for TC ≥25% and ≥50% cut-offs in classifying patients by PD-L1 expression level as either high or low (**Table 3**)

Figure 1. Pair-wise scatterplots for PD-L1 TC scoring by pathologists



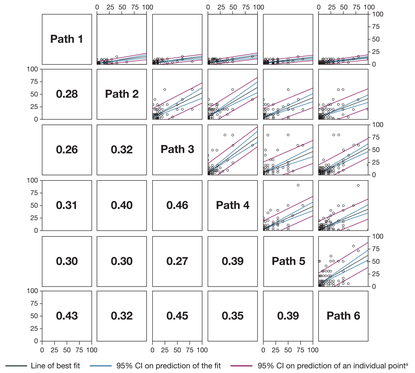
[a]95% of the data are expected to lie within these boundaries
CI, confidence interval; Path, pathologist; PD-L1, programmed cell death ligand-1; TC, tumour cell; TC score, proportion of TCs with PD-L1 membrane staining

Figure 2. Pair-wise scatterplots for PD-L1 PIC scoring by pathologists



[a]95% of the data are expected to lie within these boundaries
CI, confidence interval; Path, pathologist; PD-L1, programmed cell death ligand-1; PIC value, proportion of immune cells present as a proportion of the tumour environment

Figure 3. Pair-wise scatterplots for PD-L1 IC scoring by pathologists



[a]95% of the data are expected to lie within these boundaries
CI, confidence interval; IC score, percentage of immune cells staining positive for PD-L1; Path, pathologist; PD-L1, programmed cell death ligand-1

Table 1. Overall agreement between pathologists for PD-L1 TC scores across cut-offs

PD-L1 cut-off	OPA (95% CI)	APA (95% CI)	ANA (95% CI)
TC ≥1%	90.6% (89.6%)	92.1% (91.2%)	88.5% (87.3%)
TC ≥25%	95.0% (94.3%)	92.8% (91.7%)	96.2% (95.6%)
TC ≥50%	94.4% (93.6%)	90.2% (88.8%)	96.1% (95.5%)

Lower bound of the 95% CI for all calculations is >85%, the value typically required to be exceeded for ILR studies
ANA, average negative agreement; APA, average positive agreement; CI, confidence interval; ILR, inter-laboratory reproducibility; OPA, overall percent agreement; PD-L1, programmed cell death ligand-1; TC, tumour cell; TC score, proportion of TCs with PD-L1 membrane staining

Table 2. ICC for reliability of scores between pathologists

Class	ICC (95% CI)	Interpretation of ICC lower bound
TC scores	0.96 (0.95–0.97)	Excellent
PIC values	0.21 (0.11–0.31)	Poor
IC scores	0.48 (0.40–0.56)	Poor

CI, confidence interval; IC score, percentage of immune cells staining positive for PD-L1; ICC, intra-class correlation coefficient; PD-L1, programmed cell death ligand-1; PIC value, proportion of immune cells present as a proportion of the tumour environment; TC, tumour cell; TC score, proportion of TCs with PD-L1 membrane staining

Table 3. Fleiss' kappa for reliability in classification of patients based on TC scores between pathologists

PD-L1 cut-off	Kappa	Interpretation of reliability
TC ≥1%	0.806	Substantial
TC ≥25%	0.890	Almost perfect
TC ≥50%	0.863	Almost perfect

Kappa values are for single cut-offs (i.e. the pathologists score as high or low)
PD-L1, programmed cell death ligand-1; TC, tumour cell; TC score, proportion of TCs with PD-L1 membrane staining

Conclusions

- Assessment of TC scores by expert pathologists was highly reproducible in NSCLC tumour samples using the VENTANA PD-L1 (SP263) Assay, building confidence in the performance of this assay in patient selection for anti-PD-L1 therapy
- PD-L1 staining on ICs is not included in the assays that aim to identify patients eligible for immuno-oncology monotherapy treatment in NSCLC.[10,11] PIC and IC scoring were not reproducible in NSCLC tumour samples, suggesting that the assessment methodology is unreliable for this tumour type and assay
- The results for NSCLC are in contrast with those for UC, where inter-laboratory reproducibility studies have shown strong agreement in both TC and IC scoring[12]
- Further studies are required to investigate this difference in the pathology of tumour types

References

1. Chae YK et al. *J Immunother Cancer* 2018;6:39
2. Massard C et al. *J Clin Oncol* 2016;34:3119-25
3. Herbst RS et al. *Lancet* 2016;387:1540-50
4. Borghaei H et al. *N Engl J Med* 2015;373:1627-39
5. Rebelatto M et al. *J Clin Oncol* 2015;33:8033-3
6. Ventana Medical Systems. VENTANA PD-L1 (SP263) Assay. Available at: https://www.accessdata.fda.gov/cdrh_docs/pdf16/P160046C.pdf. Last accessed 8 Aug 2018
7. Ventana Medical Systems. VENTANA PD-L1 (SP263) Assay. Available at: https://diagnostics.roche.com/global/en/products/tests/ventana-pd-l1-_sp263-assay2.html. Last accessed 10 Aug 2018
8. Koo TK et al. *J Chiropr Med* 2016;15:155-63
9. Landis JR et al. *Biometrics* 1977;33:159-74
10. Rizvi N et al. *J Immunother Cancer* 2015;3:P171
11. Reck M et al. *N Engl J Med* 2016;375:1823-33
12. Zajac M et al. *Arch Pathol Lab Med* 2018 (in press)

Acknowledgements

This study is sponsored by AstraZeneca. The authors would like to thank Bridgette Higgins, Oncologica UK Ltd., Cambridge, for her valuable contribution to the study. Commercial tissue samples were supplied by Avaden Biosciences Inc., Seattle, WA, USA. Medical writing support, which was in accordance with Good Publication Practice (GPP3) guidelines, was provided by Kaveri Sidhu, PhD, of Cactus Communications, and was funded by AstraZeneca.
Copies of this poster obtained through QR (Quick Response) and/or text key codes are for personal use only and may not be reproduced without written permission of the authors.
Author email: gareth.williams@oncologica.com

Presented at the European Society for Medical Oncology (ESMO) 2018 Congress, Munich, Germany, 19-23 October