Association of digital and manual quantification of tumor PD-L1 expression with outcomes in nivolumab-treated patients

Chunzhe Duan,[1,a] Michael Montalto,[1*,2,a] George Lee,[1] Dimple Pandya,[1] Daniel Cohen,[1] Han Chang,[1] Hao Tang,[1] Nishant Agrawal,[2] Hunter Elliott,[2] Benjamin Glass,[2] Ilan Wapinski,[2] Robin Edwards,[1] Andrew H. Beck,[2,b] Vipul Baxi[1,b]

[1]Bristol Myers Squibb, Princeton, NJ, USA; [2]PathAI, Boston, MA, USA

[a]Co-first authors; [b]Co-senior authors; *At the time the analysis was conducted

2017

Background

- Programmed death ligand 1 (PD-L1) is a biomarker associated with response to programmed death-1 (PD-1) and PD-L1 immune checkpoint inhibitors[1]
- Four immunohistochemistry (IHC) assays for PD-L1 expression have been approved across a number of tumor types[1]
- Manual assessment of PD-L1 expression is limited by variability in assay and scoring methodology, heterogeneous tumor PD-L1 expression, and interobserver variability[1-3]
  - Interpathologist discordance may increase for tumor samples with low expression of PD-L1[2]
- Digital quantification may improve the precision of PD-L1 assessment by overcoming certain limitations observed with manual scoring, as suggested in various tumor types, including non-small cell lung cancer, melanoma (MEL), and urothelial carcinoma (UC)[3-6]

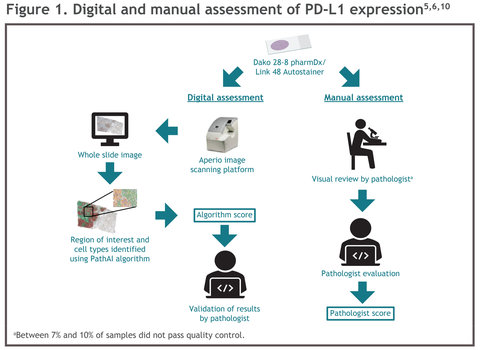
Study objectives

- We compared a digital quantification method for PD-L1 assessment with manual scoring in MEL and UC samples to assess:
  - The correlation of PD-L1 expression scores obtained from digital and manual assessment
  - The prevalence of samples with PD-L1–positive (PD-L1+) tumor cells (TCs) ≥ 1% and ≥ 5% by digital and manual quantification
  - The association of PD-L1 expression scores derived by digital and manual assessment with clinical outcomes

Methods

- PD-L1 expression was retrospectively determined in pretreatment samples from patients with MEL treated with first-line nivolumab (NIVO) as part of exploratory analyses of CheckMate 067 (NCT01844505)[7] and CheckMate 238 (NCT02388906),[8] and patients with UC treated with second-line NIVO as part of CheckMate 275 (NCT02387996)[9]
- For manual and digital assessment of PD-L1 expression, samples were stained using the Dako PD-L1 IHC 28-8 pharmDx assay (Agilent, Santa Clara, CA) (**Figure 1**)
  - Stained slides were manually scored for PD-L1+ TCs by certified pathologists from LabCorp (Burlington, NC) (pathologist score)[9]
    - Score derived from percentage of TCs with PD-L1 membrane staining at any level
  - The same slides were scanned with the Aperio AT2 (Leica Biosystems, Vista, CA) and analyzed using the PathAI artificial intelligence (AI)-powered image analysis algorithm (PathAI research platform, Boston, MA) for PD-L1+ expression in the membrane of TCs (algorithm score)[5,6]
    - Deep-learning models were trained to recognize PD-L1+ TCs using cellular and tissue region annotations collected via the PathAI pathologist network, while automatically excluding regions of background staining or sample pigmentation. Outputs consisting of quantitative features summarizing slide-level PD-L1 positivity for TCs were generated for each sample
    - Quality control was performed on all PD-L1 scoring results derived from digital analysis by a board-certified pathologist
    - Samples were excluded from analysis if the algorithm failed to identify the predefined region of interest
      - These included 21 samples from CheckMate 067, 28 from CheckMate 238, and 25 from CheckMate 275
- Best overall response was evaluated according to Response Evaluation Criteria in Solid Tumors (RECIST) v1.1 and categorized as complete response (CR), partial response (PR), stable disease (SD), progressive disease (PD), or response not evaluable (NE)

Figure 1. Digital and manual assessment of PD-L1 expression[5,6,10]



[a]Between 7% and 10% of samples did not pass quality control.

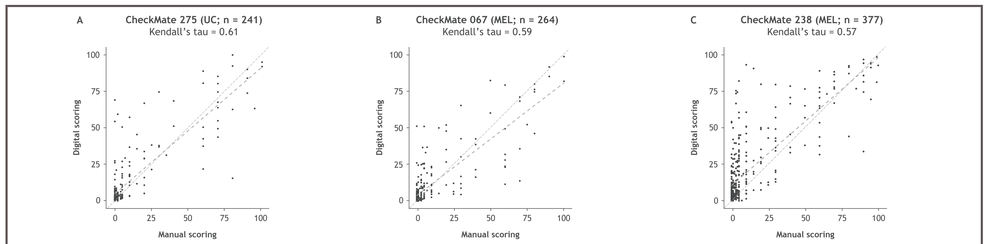
Statistical analysis

- Association of clinical efficacy with PD-L1+ TCs was assessed using cutoffs of 1% and 5%, as evaluated by digital and manual assessment
  - Kendall's tau coefficient was used to evaluate the correlation between digital and manual PD-L1+ TC scores within each trial
  - Objective response rate (ORR) was calculated using the sum of patients who achieved a CR or PR vs those who achieved a response of SD, PD, or were NE
  - Odds ratios were calculated using logistic regression to examine associations with objective response
  - Hazard ratios (HRs) were calculated using Cox proportional hazards models to examine associations with overall survival (OS) in CheckMate 275 and CheckMate 067 and recurrence-free survival (RFS) in CheckMate 238
  - Kaplan-Meier curves were drawn to illustrate comparisons of OS and RFS in patient groups

Results

- The correlation between digital and manual quantification of PD-L1+ TC expression was moderate across the trials (**Figure 2**)
  - For CheckMate 275 (n = 241), tau = 0.61; for CheckMate 067 (n = 264), tau = 0.59; and for CheckMate 238 (n = 377), tau = 0.57
- Digital assessment of PD-L1 expression demonstrated higher prevalence and identified more samples with ≥ 1% and ≥ 5% PD-L1+ TCs than manual scoring (**Table 1**)
  - When digital analysis was compared with manual assessment:
    - At a cutoff of ≥ 1%, PD-L1+ TC prevalence was increased by 22% in samples from patients with UC from CheckMate 275, 25% in samples from patients with MEL from CheckMate 067, and 13% in samples from patients with MEL from CheckMate 238
    - At a cutoff of ≥ 5%, PD-L1+ TC prevalence was increased by 7% in samples from patients with UC from CheckMate 275, 10% in samples from patients with MEL from CheckMate 067, and 25% in samples from patients with MEL from CheckMate 238
- At cutoffs of ≥ 1% and ≥ 5%, associations between PD-L1+ TCs and objective response were similar for digital and manual assessment in samples from patients with UC from CheckMate 275 and samples from patients with MEL from CheckMate 067 (**Table 2**), while assessments of PD-L1+ TCs by digital and manual scoring demonstrated better or similar associations with survival compared with manual scoring only (**Figure 3**) or digital scoring only (**Figures 3 and 4**) across all trials at those same cutoffs
  - A trend for higher ORR was observed when PD-L1 expression was assessed by digital analysis compared with manual scoring (**Table 2**)

Figure 2. Correlation between digital and manual assessment of PD-L1 expression[a]



[a]Line with longer dashes represents actual correlation (with 95% CI). CI, confidence interval.

Table 1. Prevalence of PD-L1 expression by digital and manual scoring

	Evaluable samples, n	Prevalence, n (%)		Additional samples identified by digital only, n (%)[a]	Additional samples identified by manual only, n (%)
		Digital	Manual		
≥ 1% PD-L1+ TCs					
CheckMate 275 (UC)	241	166 (69)	113 (47)	58 (24)	5 (2)
CheckMate 067 (MEL)	264	173 (66)	160 (61)	36 (14)	23 (9)
CheckMate 238 (MEL)	377	307 (81)	259 (69)	66 (18)	18 (5)
≥ 5% PD-L1+ TCs					
CheckMate 275 (UC)	241	90 (37)	74 (31)	28 (12)	12 (5)
CheckMate 067 (MEL)	264	103 (39)	76 (29)	36 (14)	9 (3)
CheckMate 238 (MEL)	377	234 (62)	139 (37)	104 (28)	9 (2)

[a]Samples that were identified by digital as ≥ 1% or ≥ 5% but < 1% or < 5% by manual.

Table 2. Association of objective response with PD-L1 expression scored by digital and manual assessment

	Method	Overall ORR, %		Odds ratio (95% CI)	Method	Association with ORR	ORR, %
		PD-L1 TC cutoff					
		< 1%	≥ 1%				
CheckMate 275 (UC)[a]	Digital	13.3	22.9	2.05 (0.95-4.43)	Samples not identified by digital or manual		12.9
					Samples identified by manual and digital		25.0
	Manual	15.6	24.8	1.70 (0.89-3.25)	Additional samples identified by digital only		19.0
					Additional samples identified by manual only		20.0
CheckMate 067 (MEL)[b]	Digital	37.4	54.0	1.82 (1.07-3.11)	Samples not identified by digital or manual		35.3
					Samples identified by manual and digital		57.2
	Manual	37.5	55.3	1.89 (1.12-3.18)	Additional samples identified by digital only		41.7
					Additional samples identified by manual only		43.5
		< 5%	≥ 5%				
CheckMate 275 (UC)[a]	Digital	13.2	31.1	2.99 (1.55-5.76)	Samples not identified by digital or manual		13.7
					Samples identified by manual and digital		33.9
	Manual	15.6	29.7	2.27 (1.18-4.39)	Additional samples identified by digital only		25.0
					Additional samples identified by manual only		8.3
CheckMate 067 (MEL)[b]	Digital	40.1	61.2	2.28 (1.35-3.85)	Samples not identified by digital or manual		41.2
					Samples identified by digital and manual		64.2
	Manual	43.9	59.2	1.77 (1.01-3.09)	Additional samples identified by digital only		55.6
					Additional samples identified by manual only		22.2

Database lock in 2019: CheckMate 275, June 14; CheckMate 067, January 18; CheckMate 238, April 3.
[a]Odds ratios are adjusted for ECOG PS, liver metastatic status, and hemoglobin; [b]Odds ratios are adjusted for ECOG PS, liver metastatic status, lactate dehydrogenase, and *BRAF* mutation. *BRAF*, B-Raf proto-oncogene, serine/threonine kinase; ECOG PS, Eastern Cooperative Oncology Group performance status.

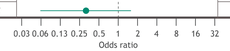
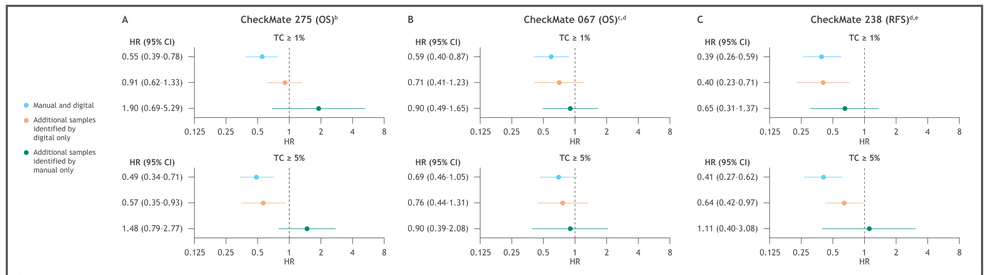
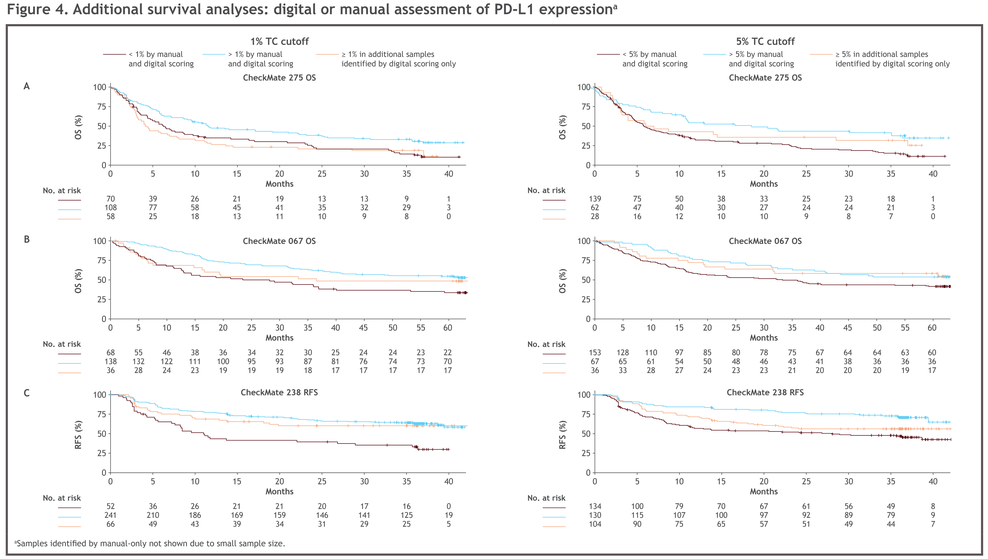


Figure 3. Survival based on digital or manual assessment of PD-L1 expression[a]



[a]Compared with digital and manual assessment ≤ 1% or ≤ 5%; [b]Adjusted for ECOG PS, liver metastatic status, and hemoglobin; [c]Adjusted for ECOG PS, liver metastatic status, lactate dehydrogenase, and *BRAF* mutation; [d]NIVO only; [e]Adjusted for ECOG PS, American Joint Committee on Cancer stage, lactate dehydrogenase, and *BRAF* mutation.

Figure 4. Additional survival analyses: digital or manual assessment of PD-L1 expression[a]



[a]Samples identified by manual-only not shown due to small sample size.

Conclusions

- In post-hoc exploratory analyses of NIVO clinical trials, digital assessment identified higher prevalence of samples with PD-L1+ TC expression at cutoffs of ≥ 1% and ≥ 5% from patients with UC and MEL compared with manual scoring
- While we identified moderate correlations between digital and manual quantification of PD-L1+ TCs across all trials, this is limited primarily by interpathologist variability based on previously published data[1,11,12]
- Similar associations between PD-L1 expression and overall response were demonstrated between manual and digital scoring
- Additional patients identified as PD-L1+ across all cutoffs by both manual and digital scoring and digital scoring only demonstrated an improved or maintained response and survival compared with patients identified as PD-L1 negative by both manual and digital assessment
  - Results from this study suggest digital quantification may identify patients who may show clinical benefit from NIVO treatment with greater sensitivity
- Comparison of manual and digital scoring of PD-L1 expression is warranted in further analyses in other tumor types and cell populations
  - For example, digital quantification of PD-L1 expression on immune cells and identification of key geographic features may be informative

References

1. Rimm DL, et al. *JAMA Oncol* 2017;3:1051-1058.
2. Büttner R, et al. *J Clin Oncol* 2017;35:3867-3876.
3. Kearney S, et al. *Cancer Res* 2017;77(suppl 13). Abstract 4582.
4. Koelzer VH, et al. *Histopathology* 2018;73:397-406.
5. Baxi V, et al. Oral presentation at the Society for Immunotherapy of Cancer (SITC) Annual Meeting; November 6-10, 2019; National Harbor, MD, USA. Abstract O65.
6. Beck A, et al. Poster presentation at the Society for Immunotherapy of Cancer (SITC) Annual Meeting; November 6-10, 2019; National Harbor, MD, USA. Abstract P730.
7. Hodi FS, et al. *Lancet Oncol* 2018;19:1480-1492.
8. Weber JS, et al. *Ann Oncol* 2019;30(suppl 5). Abstract 1310O.
9. Sharma P, et al. *Lancet Oncol* 2017;18:312-322.
10. Bera K, et al. *Nat Rev Clin Oncol* 2019;16:703-715.
11. Brunnstrom H, et al. *Mod Pathol* 2017;30:1411-1421.
12. Rehman J, et al. *Mod Pathol* 2017;30:340-349.

Acknowledgments

- The patients and families who have made the trial possible
- The clinical study teams for CheckMate 275 and protocol manager Toshiki Maeda; CheckMate 067 and protocol manager Arancha Campos; and CheckMate 238 and protocol manager Jennifer Schick
- Dako, an Agilent Technologies, Inc. company, for collaborative development of the PD-L1 IHC 28-8 pharmDx assay (Santa Clara, CA)
- Bristol-Myers Squibb Company (Princeton, NJ) and ONO Pharmaceutical Company Ltd. (Osaka, Japan)
- The study was supported by Bristol-Myers Squibb Company
- All authors contributed to and approved the presentation; writing and editorial assistance were provided by Emily Motola, PharmD, and Jay Rathi, MA, of Spark Medica Inc, funded by Bristol-Myers Squibb Company

Email: Emily.Duan@bms.com