



# A comparability study of immunohistochemical assays for PD-L1 expression in hepatocellular carcinoma

Lei Shi<sup>1</sup> · Shu-Jun Zhang<sup>2</sup> · Jun Chen<sup>3</sup> · Shi-Xun Lu<sup>4</sup> · Xin-Juan Fan<sup>5</sup> · Joanna Hung-Man Tong<sup>6</sup>  · Chit Chow<sup>6</sup> · Edith Ka-Yee Tin<sup>6</sup> · Stephen Lam Chan<sup>7,8</sup> · Charing Ching-Ning Chong<sup>9</sup> · Paul Bo-San Lai<sup>8,9</sup> · Ka-Fai To<sup>6,8</sup> · Nathalie Wong<sup>6,8</sup> · Anthony Wing-Hung Chan<sup>6</sup>

Received: 26 February 2019 / Revised: 22 May 2019 / Accepted: 23 May 2019 / Published online: 23 June 2019  
© The Author(s), under exclusive licence to United States & Canadian Academy of Pathology 2019

## Abstract

Programmed death ligand 1 (PD-L1) protein expression by immunohistochemistry is a promising biomarker for PD-1/PD-L1 blockade in hepatocellular carcinoma. There are a number of commercially available PD-L1 assays. Our study aimed to compare the analytical performance of different PD-L1 assays and evaluate the reliability of pathologists in PD-L1 scoring. Consecutive sections from tumor samples from 55 patients with surgically resected primary hepatocellular carcinoma were stained with four standardized PD-L1 assays (22C3, 28–8, SP142, and SP263). We also correlated the PD-L1 protein level by immunohistochemistry with the mRNA level of those genes associated with tumor immune microenvironment by the NanoString platform. Five pathologists independently assessed PD-L1 expression on tumor cells [tumor proportion score] together with tumor-infiltrating immune cells (combined positive score). The 22C3, 28–8, and SP263 assays had comparable sensitivity in detecting PD-L1 expression, whereas the SP142 assay was the least sensitive assay. The inter-assay agreement measured by intraclass correlation coefficients for the tumor proportion score and combined positive score were 0.646 and 0.780, respectively. The inter-rater agreement was good to excellent (the overall intraclass correlation coefficient for the tumor proportion score and combined positive score was 0.946 and 0.809, respectively). Pathologists were less reliable in scoring combined positive score than tumor proportion score, particularly when using the SP142 assay. Up to 18% of samples were misclassified by individual pathologists in comparison to the consensus score at the cutoff of combined positive score  $\geq 1$ . The combined positive score by the 22C3 assay demonstrated the strongest correlation with immune-related gene mRNA signatures, closely followed by combined positive scores by the 28–8 and SP263 assays. In conclusion, the 22C3, 28–8, and SP263 assays are highly concordant in PD-L1 scoring and suggest the interchangeability of these three assays. Further improvement of the accuracy in assessing PD-L1 expression at a low cutoff is still necessary.

## Introduction

Remarkable clinical response to immune checkpoint inhibitors targeting programmed death 1 (PD-1) and programmed death ligand 1 (PD-L1) makes PD-1/PD-L1 blockade become part of the standard of treatment for

various advanced malignancies [1–4]. For advanced hepatocellular carcinoma, two recent clinical trials on anti-PD-1 antibodies (CheckMate 040 phase 1/2 trial on nivolumab and KEYNOTE 224 phase 2 trial on pembrolizumab) showed that the PD-1/PD-L1 blockade could provide a durable objective response with manageable tolerability profile [5, 6]. Nivolumab and pembrolizumab were granted accelerated approval by the U.S. Food and Drug Administration for patients with advanced hepatocellular carcinoma after the failure of sorafenib. Other phase 1 and 2 clinical trials on other anti-PD-1 (PDR001) and anti-PD-L1 (atezolizumab, avelumab, and durvalumab) agents for hepatocellular carcinoma are ongoing.

Appropriate patient selection by predictive biomarker is essential to enrich treatment efficacy. PD-L1 expression by immunohistochemistry is one of the predictive biomarkers

These authors contributed equally: Lei Shi, Shu-Jun Zhang, Jun Chen

**Supplementary information** The online version of this article (<https://doi.org/10.1038/s41379-019-0307-8>) contains supplementary material, which is available to authorized users.

✉ Anthony Wing-Hung Chan  
awh\_cha@cuhk.edu.hk

Extended author information available on the last page of the article

for PD-1/PD-L1 inhibitors [1, 3, 7]. Although its predictive role remains inconclusive in hepatocellular carcinoma, it seems to be a promising biomarker [5, 6]. The current one drug-one predictive biomarker co-development approach leads to each PD-1/PD-L1 inhibitor being associated with a unique PD-L1 immunohistochemical assay. Four standardized PD-L1 immunohistochemical assays (22C3, 28–8, SP142, and SP263) have been developed specifically for pembrolizumab, nivolumab, atezolizumab and durvalumab, respectively. Each assay employs different antibody clones, automatic staining platforms (Dako Autostainer Link48 for 22C3 and 28–8 assays, and Ventana BenchMark Ultra for SP142 and SP263 assays), staining protocols, scoring methods, and cutoff values, and produces several practical challenges. Most of the pathology laboratories do not offer all four PD-L1 assays because of the cost of each standardized assay, the availability of both automatic staining platforms and the limitation of laboratory resources. Moreover, pathological diagnosis is not mandatory for the majority of hepatocellular carcinoma patients, particularly those with advanced diseases [8, 9]. A small biopsy will be the only specimen for these patients. It may not be sufficient and cost-effective for multiple PD-L1 testing because it may be required for genomic profiling and patient-derived xenograft, which will play an increasingly important role in personalized medicine for hepatocellular carcinoma patients [10].

Several groups have conducted comparability studies to address inter-assay concordance and inter-observer agreement for PD-L1 immunohistochemistry in non-small cell lung cancers [7, 11–14]. However, such data in hepatocellular carcinoma are still limited [15]. Our study aimed to compare the analytical performance of four PD-L1 assays and evaluate the reliability of pathologists in scoring PD-L1 expression. We would also like to investigate the correlation between the PD-L1 protein level by immunohistochemistry and the mRNA level of those genes associated with tumor immune microenvironment.

## Subjects and methods

### Study material

Formalin-fixed, paraffin embedded samples from 55 patients undergoing surgical resection for primary hepatocellular carcinoma were obtained from the archives of Department of Anatomical and Cellular Pathology, Prince of Wales Hospital, Hong Kong. Written informed consent was obtained from all patients in the study. The study was approved by the institutional review board. An experienced hepatopathologist (AWHC) reviewed all samples to confirm the histological diagnosis and select a representative tumorous tissue

block for each sample. The tissue blocks were kept at room temperature with the median storage age of 59 months (interquartile range: 47–80 months). The tissue block from the largest tumor was selected if the surgical sample contained multiple tumors. All clinical and laboratory parameters were collected and reviewed from patients' records. Histological variants (steatohepatic, lymphoepithelioma-like and scirrhous hepatocellular carcinomas) were defined according to the WHO classification [16–18]. Clinicopathological characteristics of the study cohort are summarized in Supplementary Table S1.

### PD-L1 immunohistochemical assays

Consecutive 4µm thick sections were freshly cut from each tissue block for immunohistochemistry. Four standardized PD-L1 assays (22C3, 28–8, SP142 and SP263) were performed by their corresponding autostainers (Dako Autostainer Link48 for 22C3 and 28–8 assays, and Ventana BenchMark Ultra for SP142 and SP263 assays) according to the manufacturers' instructions [13]. The slides were independently scored by five gastrointestinal/liver pathologists (JC, SJZ, SL, SXL, XJF) from five different institutions. To standardize the immunohistochemical assessment, a half-day multi-head microscopy training session for these five pathologists was provided by a pathologist (AWHC) who was previously trained on the PD-L1 22C3 assay. The areas with tumor necrosis (which contributed to 0–20% of all samples) were exempted for the assessment. The tumor proportion score was determined according to the percentage of viable tumor cells with partial or complete membranous stain at any intensity. It was assigned in 1% increments over a range of 0–10 and 5% increments over a range of 10–100%. The combined positive score was calculated by dividing the number of PD-L1-stained cells (tumor cells, lymphocytes, macrophages) by the total number of viable tumor cells, multiplied by 100. The maximum combined positive score was defined as 100 [19].

### The NanoString analysis

Tumor tissue was enriched by manual macrodissection. Total mRNA was isolated from the macrodissected tumor tissues using Qiagen miRNeasy Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. The RNA sample was quantified by NanoDrop (Thermo Scientific, Wilmington, DE), and regarded as an adequate sample if it contained a 400 ng at a minimum. The sample was subsequently analyzed by the nCounter PanCancer Immune Profiling Panel (NanoString, Seattle, WA) according to the manufacturer's instructions [20]. Genes for constructing immune-related gene signatures were extracted from the literature (Supplementary Table S2) [20–22]. The gene

signature was calculated from the arithmetic mean of log base 10 transformed expression of genes.

## Statistical analyses

All statistical analyses were done by R version 3.4.4 (R Foundation for Statistical Computing, Vienna, Austria). For each sample, the consensus score for each PD-L1 assay were the median scores of all the pathologists. The PD-L1 score was dichotomized according to different specific cutoffs. However, due to the absence of clinically significant cutoff for hepatocellular carcinoma, the cutoffs were adopted from those defined in other cancers, tumor proportion score (1%, 10%, 25%, and 50%) and combined positive score (1, 10, and 20) [6, 23–25]. To evaluate the comparability of four PD-L1 assays, the intraclass correlation coefficients and Fleiss' kappa statistics were calculated among different assays for continuous scores, and dichotomized scores, respectively. Scatter plots and Bland–Altman plots were employed to compare PD-L1 assays graphically. Similarly, to assess the reliability of pathologists in PD-L1 scoring, the intraclass correlation coefficients and Fleiss' kappa statistics were calculated among all the pathologists for continuous scores and

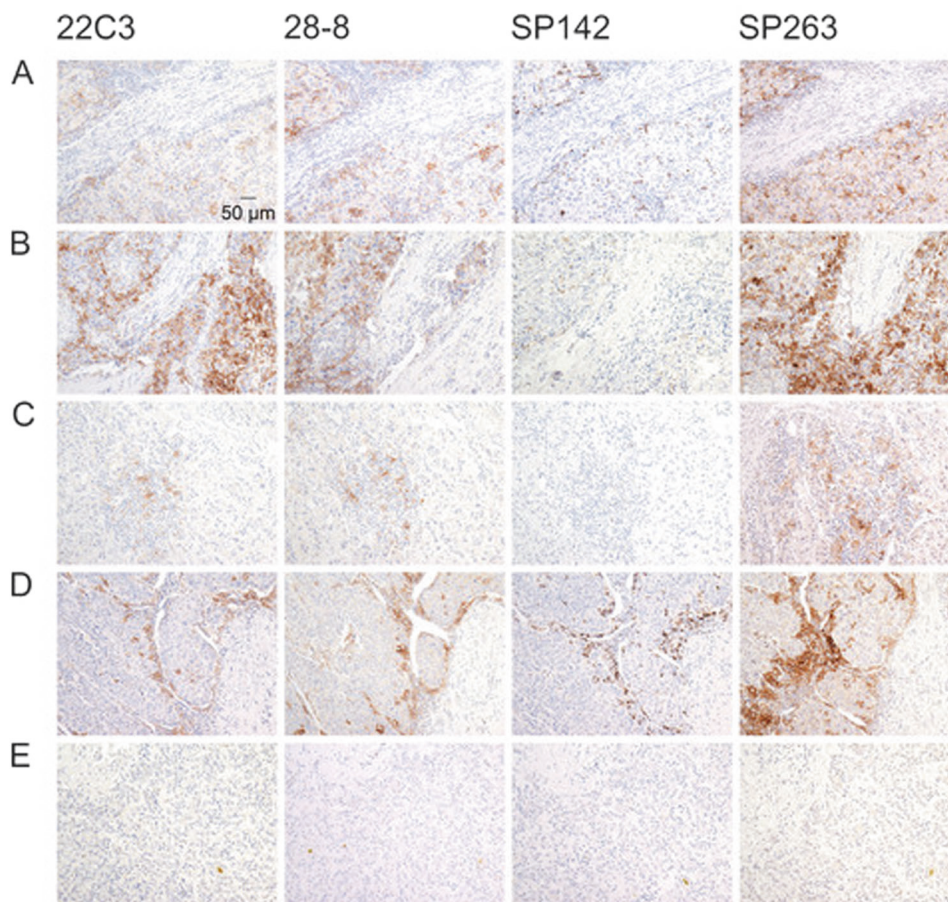
dichotomized scores, respectively. An intraclass correlation coefficient of <0.5 or less is poor reliability, 0.5–0.75 is moderate, 0.75–0.9 is good, and greater than 0.9 is excellent [26]. A Fleiss' kappa of 0.2–0.4 indicates fair agreement, 0.4–0.6 indicates moderate, 0.6–0.8 indicates substantial, and greater than 0.8 indicates almost perfect. The Wilcoxon signed rank test was used to compare the differences in PD-L1 consensus scores among four assays. Correlation between PD-L1 scores from different assays and pathologists, and between PD-L1 consensus scores and gene signatures was tested by Pearson's correlation test. Benjamini–Hochberg method was used to adjust the *P*-value for multiple comparisons. A 2-tailed *P*-value < 0.05 was considered as statistically significant.

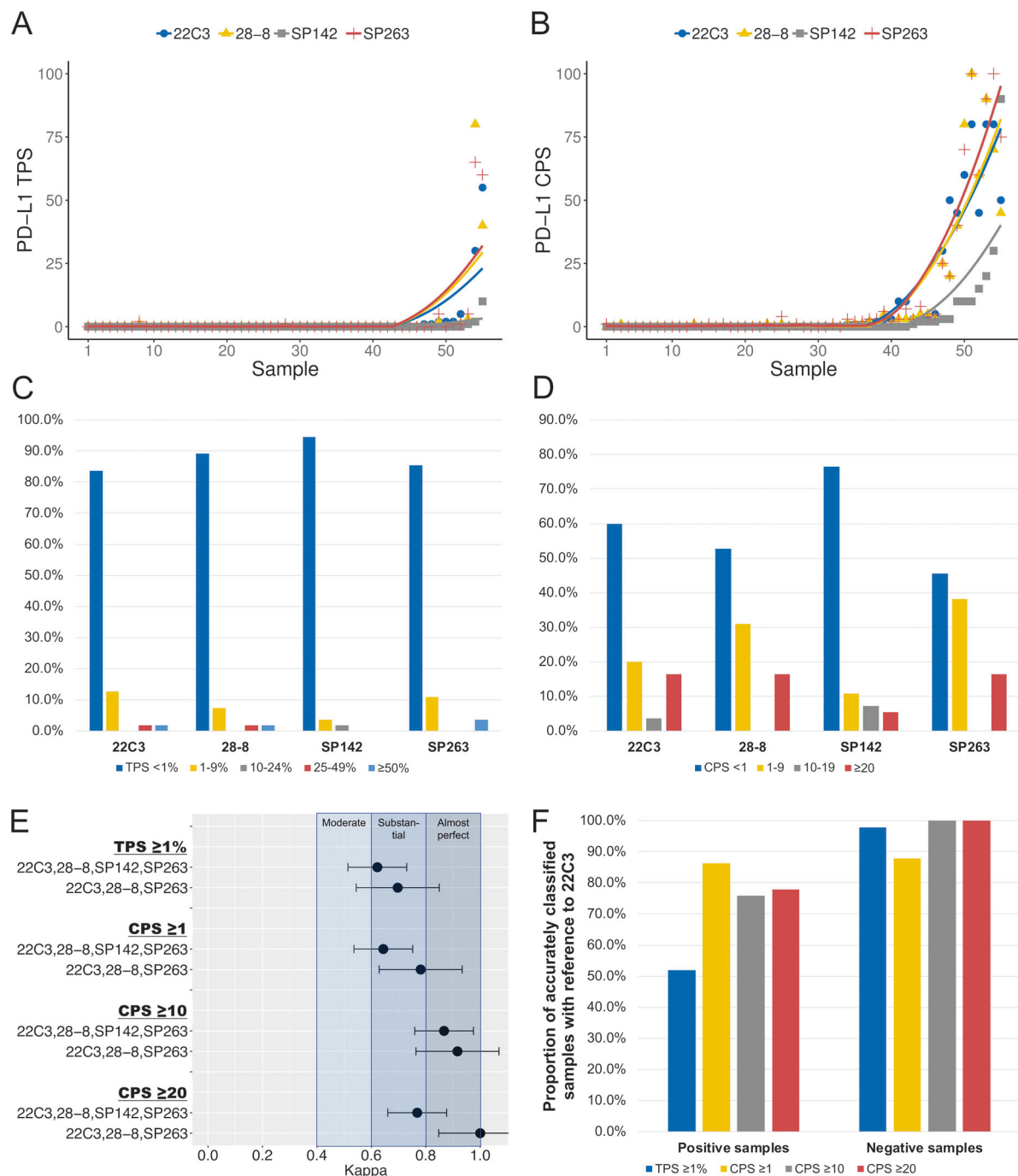
## Results

### Comparability of PD-L1 scoring between the four standardized PD-L1 assays

Figure 1 shows representative immunohistochemical images for the four PD-L1 assays. Figure 2a and b shows the best-fit curves of the consensus PD-L1 score for each

**Fig. 1** Representative samples comparing the PD-L1 protein expression by the four standardized assays for 5 different samples a–e





**Fig. 2** Comparability of PD-L1 scoring between the four standardized PD-L1 assays. Comparison of the **a** tumor proportion score (TPS) and **b** combined positive score (CPS) for the four PD-L1 assays. Frequencies distributions for the **c** tumor proportion score and **d** combined positive score at different cutoffs. **e** Inter-assay agreement among

sample among all four assays. Figure 2c, d shows the distribution of tumor proportion scores and combined positive scores categorized by pre-defined cutoffs among four assays. Figures S1–S3 shows graphically pairwise comparisons among different PD-L1 assays, whereas Table 1 summarizes the mean difference and statistical significance

different PD-L1 assays of the tumor proportion score and combined positive score at different cutoffs. **f** Proportion of correctly classified positive samples and negative samples by any one assay (28-8, SP142 and SP263) in comparison with the 22C3 assay

of pairwise comparisons of PD-L1 assays. The SP142 assay was the least sensitive in assessing the tumor proportion score and combined positive score, whereas the SP263 assay was the most sensitive in assessing the combined positive score. The 22C3, 28-8, and SP263 assays showed similar sensitivity in the tumor proportion score scoring,



**Table 1** Pairwise comparison between PD-L1 assays

Assay pair	Tumor proportion score (continuous)		Tumor proportion score (categorical) <sup>a</sup>		Combined positive score (continuous)		Combined positive score (categorical) <sup>b</sup>	
	Mean (SD)	P-value	Mean (SD)	P-value	Mean (SD)	P-value	Mean (SD)	P-value
22C3 and 28–8	−0.49 (7.12)	0.339	0.05 (0.36)	0.298	−0.02 (6.57)	0.848	−0.04 (0.33)	0.484
22C3 and SP142	1.58 (7.09)	0.009	0.18 (0.47)	0.010	6.71 (18.22)	<0.001	0.35 (0.62)	<0.001
22C3 and SP263	−0.73 (4.84)	0.347	0.00 (0.33)	1.000	−1.36 (7.23)	0.032	−0.11 (0.42)	0.065
28–8 and SP142	2.07 (11.19)	0.035	0.13 (0.23)	0.049	6.72 (20.55)	<0.001	0.38 (0.56)	<0.001
28–8 and SP263	−0.24 (3.44)	0.202	−0.05 (0.23)	0.149	−1.34 (5.87)	0.003	−0.07 (0.33)	0.129
SP142 and SP263	−2.31 (10.74)	0.014	−0.18 (0.55)	0.018	−8.07 (20.21)	<0.001	−0.45 (0.57)	<0.001

<sup>a</sup>The tumor proportion score was categorized by pre-defined cutoffs as 0: <1%; 1: 1–9%; 2: 10–24%; 3: 25–49%; 4: ≥50%

<sup>b</sup>The combined positive score was categorized by pre-defined cutoffs as 0: <1; 1: 1–9; 2: 10–19; 3: ≥20

SD standard deviation

while the 22C3 and 28–8 assays demonstrated comparable sensitivity in the combined positive score scoring. When the combined positive score was categorized by the pre-defined cutoffs, the differences between the SP263 and 22C3 assays, and the SP263 and 28–8 assays became insufficient (i.e., the 22C3, 28–8, and SP263 assays had equivalent sensitivity).

The inter-assay agreement measured by intraclass correlation coefficients for the tumor proportion score and combined positive score were 0.646 (95% confidence interval (CI): 0.528–0.753) and 0.780 (95% CI: 0.693–0.853), respectively, which significantly improved to 0.878 (95% CI: 0.817–0.922) and 0.964 (95% CI: 0.944–0.977) when the SP142 assay was excluded (Table 2). At a cutoff of tumor proportion score ≥ 1%, 6% of patients were as PD-L1 positive, and 80% of patients were classified as PD-L1 negative by all four assays. The discordant rate of PD-L1 assays using tumor proportion score ≥ 1% as the cutoff was 15% for all four assays, which was improved to 12% by excluding the SP142 assay. Similarly, at a cutoff of combined positive score ≥ 1, 24 and 44% of patients were categorized as PD-L1 positive and negative, respectively. The discordant rate of PD-L1 assays using combined positive score ≥ 1 as the cutoff was 33% for all four assays, which was reduced to 17% by excluding the SP142 assay. Figure S4 and S5 shows the heatmaps comparing four assays at different cutoffs of tumor proportion score and combined positive score. A substantial to almost perfect inter-assay agreement was shown by Fleiss' kappa statistics at a cutoff of tumor proportion score ≥ 1%, combined positive score ≥ 1, ≥ 10 and ≥ 20 (Fig. 2e). The analyses at a cutoff of tumor proportion score ≥ 10, ≥ 25 and ≥ 50 were exempted due to scanty positive samples (<5%).

As the current study did not have clinical response information for PD-1/PD-L1 blockade, we could not determine the sensitivity and specificity (as well as positive

**Table 2** Intraclass correlation coefficient for inter-assay and inter-observer agreements

	Assay	Intraclass correlation coefficient (95% confidence interval)
Inter-assay agreement		
PD-L1 tumor proportion score	22C3, 28–8, SP142, SP263	0.646 (0.528–0.753)
	22C3, 28–8, SP263	0.878 (0.817–0.922)
PD-L1 combined positive score	22C3, 28–8, SP142, SP263	0.780 (0.693–0.853)
	22C3, 28–8, SP263	0.964 (0.944–0.977)
Inter-observer agreement		
PD-L1 tumor proportion score	22C3	0.952 (0.930–0.969)
	28–8	0.957 (0.938–0.971)
	SP142	0.727 (0.637–0.810)
	SP263	0.957 (0.937–0.972)
	All assays	0.946 (0.934–0.956)
PD-L1 combined positive score	22C3	0.847 (0.788–0.898)
	28–8	0.874 (0.823–0.916)
	SP142	0.628 (0.522–0.732)
	SP263	0.851 (0.793–0.900)
	All assays	0.809 (0.744–0.842)

and negative predictive values) of the individual PD-L1 assay. In efforts to evaluate the usefulness of the PD-L1 assay in stratifying samples according to different cutoff values, we used the 22C3 assay as the reference and calculated the proportion of accurately classified positive and negative samples by other three assays as an analogue for sensitivity and specificity, respectively. At a cutoff of tumor proportion score ≥ 1%, 52% and 98% of the samples were correctly classified as a positive case and a negative case by

any one assay (28–8, SP142 and SP263) in comparison with the 22C3 assay, respectively (Fig. 2f). On the other hand, by using the combined positive score, 78–86% and 88–100% the samples were correctly classified as a positive case and a negative case, respectively.

### Reliability of pathologists in PD-L1 scoring

For the tumor proportion score scoring, the intraclass correlation coefficient among five pathologists was 0.946 (95% CI: 0.937–0.972) in overall and ranged from 0.727 to 0.957 for different assays, which indicated moderate to excellent reliability. Similarly, for the combined positive score scoring, the intraclass correlation coefficient among five pathologists was 0.809 (95% CI: 0.744–0.842) in overall and ranged from 0.629 to 0.874, which demonstrated moderate to good reliability (Table 2). Pathologists were less reliable in scoring combined positive score than tumor proportion score, particularly when using the SP142 assay (Fig. S6 and S7). A moderate to almost perfect inter-observer reliability was shown by Fleiss' kappa statistics at different cutoffs of the tumor proportion score and combined positive score (Fig. 3a). The SP142 assay in the assessment of combined positive score at a cutoff of  $\geq 20$  had the poorest inter-observer reliability. When concerning those positive samples only, the inter-observer reliability indicated by Fleiss' kappa statistics was reduced but still moderate to substantial except for the SP142 assay (Fig. 3b).

Figures S8–S11 show the heatmaps comparing the consensus score and the individual pathologists' scores at various cutoffs in different PD-L1 assays. At a cutoff of tumor proportion score  $\geq 1\%$ , 80% and 89–95% of the samples were correctly classified as a positive case and a negative case by any one pathologist in comparison with the consensus score, respectively (Fig. 3c, d). In general, over 85% of the samples could be properly stratified as a negative case at any cutoff by all four assays, whereas at least 80% of the samples could be appropriately categorized as a positive case at any cutoff by all three assays excluding the SP142 assay. Using the SP142 assay, the pathologists could only correctly classify two-third of samples as a positive case at the cutoff of combined positive score  $\geq 20$ . The overall accuracy of the individual pathologists' scoring in comparison with the consensus score ranged from 83 to 97% at various cutoffs (Fig. 3e).

### Correlation between PD-L1 scores by four assays and tumor immune microenvironment

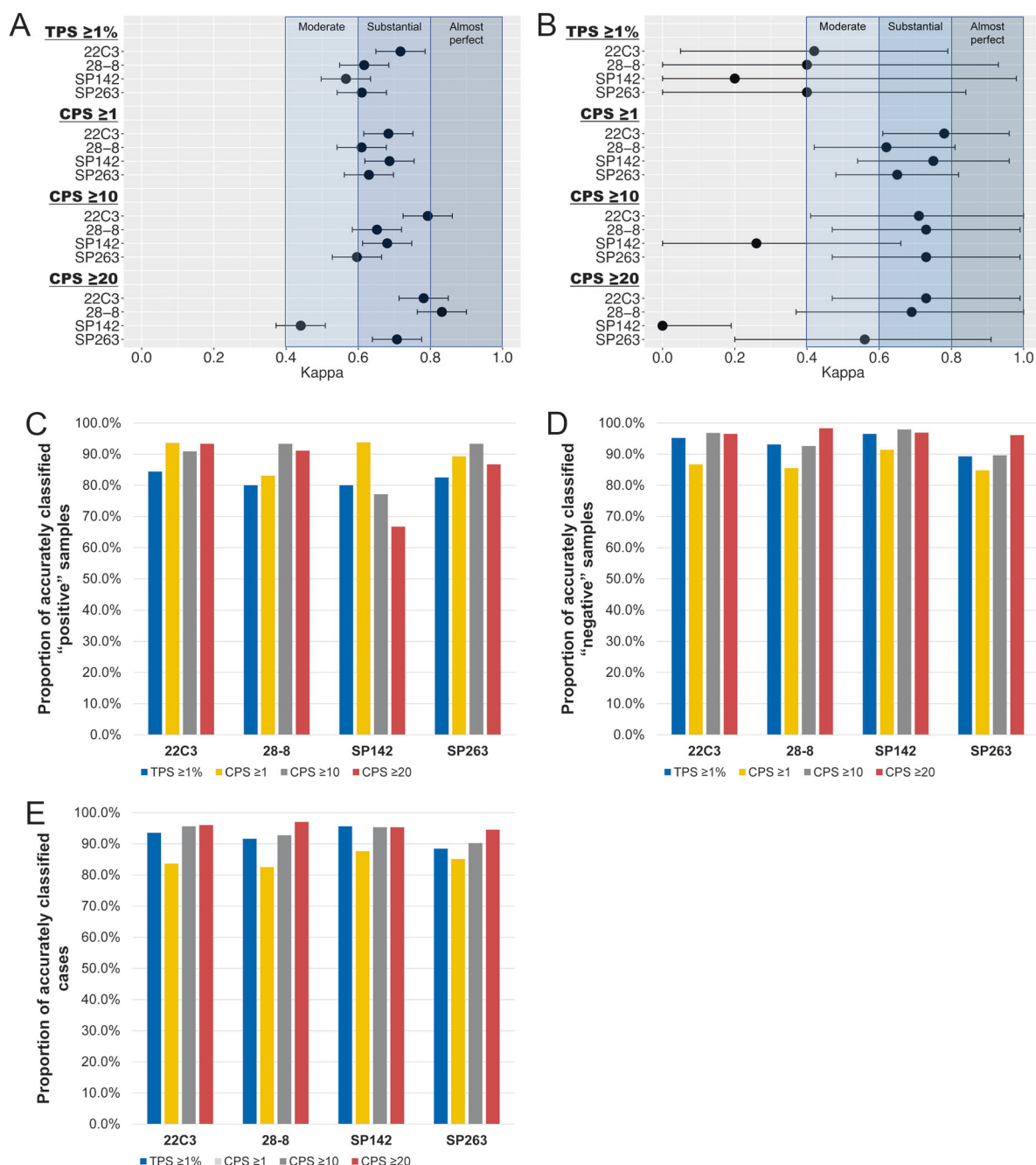
The tumor immune microenvironment was evaluated by the NanoString PanCancer Profiling Panel. Figure 4 shows the correlation between PD-L1 scores by four

immunohistochemical assays and mRNA expression levels of immune-related gene signatures. The tumor proportion score by four PD-L1 assays positively correlated with gene signatures of druggable immune checkpoints [*CD274* (encoding PD-L1), *PDCD1* (encoding PD-1), *CTLA4*, *HAVCR2* (encoding TIM3), *IDO1*, and *LAG3*], CD8-positive T-cells, T-helper cells, and interferon gamma. Compared to the tumor proportion score, the combined positive score by four PD-L1 assays showed a positive but stronger correlation with gene signatures of druggable immune checkpoints, B-cells, T-cells, CD8-positive T-cells, T-helper cells, regulatory T-cells, tertiary lymphoid structure, M1 macrophages, and interferon gamma. The combined positive score by the 22C3 assay demonstrated the strongest correlation with immune-related gene signatures, closely followed by combined positive scores by the 28–8 and SP263 assays. To evaluate the aging effect of the samples on the PD-L1 immunohistochemistry and mRNA expression, we divided our cohort into two subsets by the median storage age and compared the correlation of the 22C3 assay and immune-related gene signatures. There was no significance difference of the correlation between these two subsets (Fig. S12).

## Discussion

Our present study compared the analytical performance of the four standardized PD-L1 assays on hepatocellular carcinoma samples, and found that the 22C3, 28–8, and SP263 assays have comparable sensitivity in detecting PD-L1 expression on tumor cells (tumor proportion score) together with tumor-infiltrating immune cells (lymphocytes and macrophages) (combined positive score), whereas the SP142 assay stains a significantly lower portion of tumor cells, tumor-infiltrating lymphocytes and macrophages. We also showed that the reliability or inter-rater agreement of pathologists in scoring PD-L1 expression was good (the overall intraclass correlation coefficient for the combined positive score was 0.809) to excellent (the overall intraclass correlation coefficient for the tumor proportion score was 0.946). Finally, we demonstrated that the PD-L1 protein expression is correlated with the mRNA level of those genes associated with tumor immune microenvironment. The combined positive score has a stronger correlation with immune-related gene signatures than the tumor proportion score. Among the four different PD-L1 assays for evaluating the combined positive score, the 22C3 assay has the strongest correlation, closely followed by the 28–8 and SP263 assays.

The predictive role of the PD-L1 protein expression by immunohistochemistry for PD-1/PD-L1 blockade has not yet been well established in hepatocellular carcinoma.



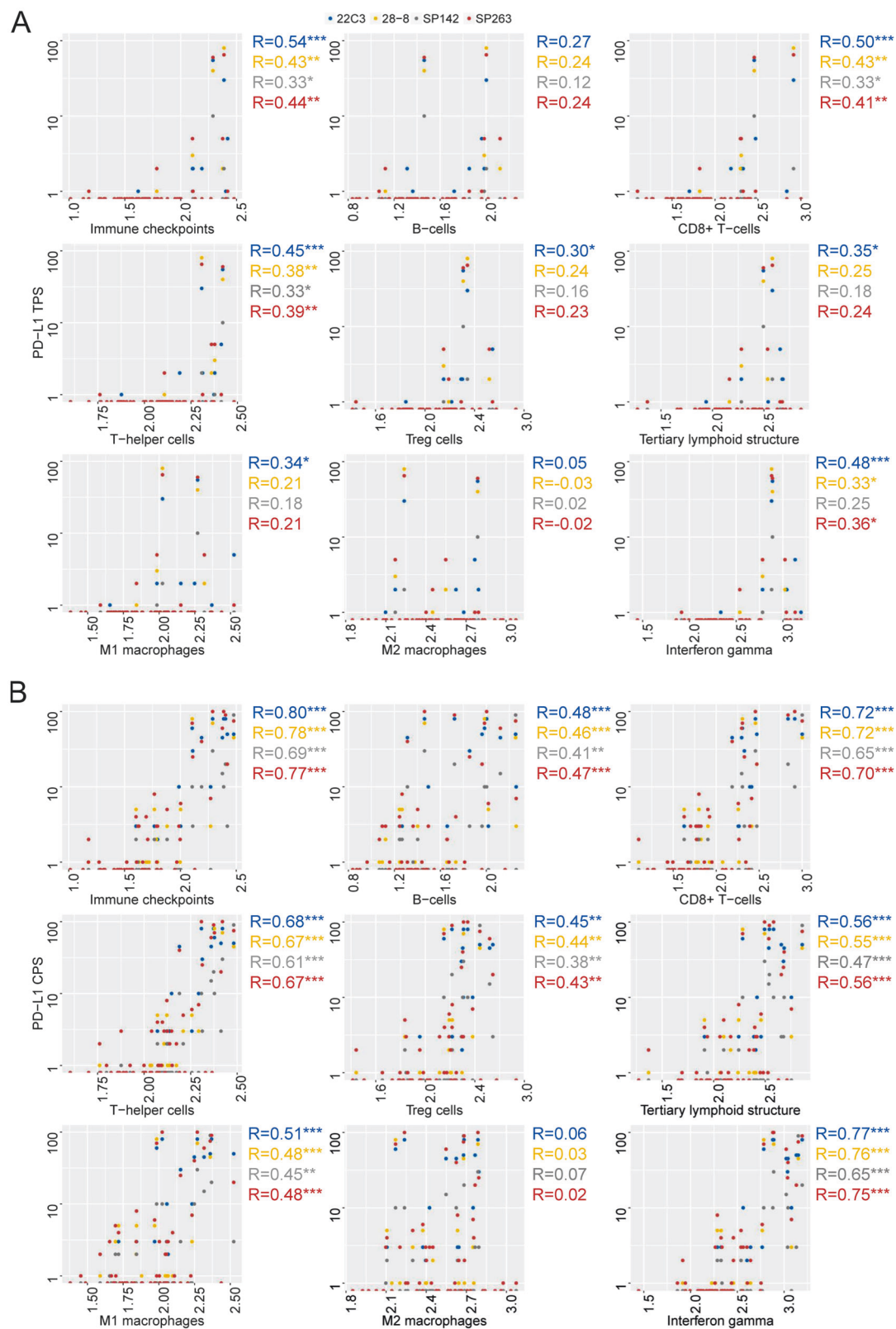
**Fig. 3** Reliability of pathologists in PD-L1 scoring. Inter-observer agreement among different pathologists in evaluating the tumor proportion score (TPS) and combined positive score (CPS) at different cutoffs for **a** all samples and **b** positive samples. Proportion of

correctly classified **c** positive samples, **d** negative samples, and **e** both positive and negative samples by any one pathologist in comparison with the consensus score

The CheckMate 040 trial showed a higher but not statistically significant objective response rate to nivolumab among hepatocellular carcinoma with positive PD-L1 expression (tumor proportion score  $\geq 1\%$ ) compared to those with negative PD-L1 expression (27% vs. 17%,  $P = 0.201$ ) [5]. However, the KEYNOTE 224 trial demonstrated a positive association between the objective response to pembrolizumab and the PD-L1 expression (in term of the

combined positive score rather than the tumor proportion score) [6]. Although ongoing phase 3 clinical trials on hepatocellular carcinoma (Checkmate 459, KEYNOTE 240, and KEYNOTE 394) are necessary to confirm the usefulness of PD-L1 expression, PD-L1 protein expression is a promising predictive biomarker.

There are a number of currently available PD-L1 assays including standardized assays and laboratory-developed



**Fig. 4** **a** Scatter plots compared PD-L1 tumor proportion score (TPS) and immune-related gene signatures. **b** Scatter plots compared PD-L1 combined positive score (CPS) and immune-related gene signatures.

The numeric values represent Pearson's  $R$  correlation coefficients.  $P$ -values are adjusted for multiple comparisons by the Benjamini–Hochberg method (\* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ )



tests [11, 12, 27]. An important practical issue for PD-L1 testing is the interchangeability between different PD-L1 assays, which is essential for better utilization of precious clinical samples, manpower and laboratory resources. We found that the 22C3, 28–8, and SP263 standardized assays are highly concordant with each other and the SP142 assay is the least sensitive assay on hepatocellular carcinoma samples, which are in agreement with those from various comparability studies of PD-L1 assays in non-small cell lung cancers [7, 11–14]. A comparability study of PD-L1 assays in hepatocellular carcinoma (the Blueprint-hepatocellular carcinoma study) showed decreasing inter-assay concordance for 22C3/SP263 assays (Pearson's  $R = 0.81$ ), 22C3/28–8 assays ( $R = 0.66$ ) and 28–8/SP263 assays ( $R = 0.51$ ) [15]. However, a direct comparison between the Blueprint-hepatocellular carcinoma study and our study is inappropriate because different scoring systems were employed. The Blueprint-hepatocellular carcinoma study used the H-score (0–300) of the PD-L1 expression in tumor cells, whereas we evaluated the PD-L1 expression by tumor proportion score and combined positive score, which are more clinically relevant scores because they are utilized in most of the clinical trials of PD-1/PD-L1 immunotherapy [1, 5, 6, 23, 24].

Another important practical issue for PD-L1 testing is the reliability of pathologists in scoring PD-L1 expression. We demonstrated moderate to excellent inter-observer agreement in assessing PD-L1 expression in hepatocellular carcinoma, which are in line with the observations in other cancers [11, 14, 19, 25, 28]. Although previous studies in other malignancies reported that pathologists are significantly less concordant in evaluating PD-L1 expression in immune cells than tumor cells [11, 14, 25, 28], we revealed high concordance among pathologists in judging the combined PD-L1 expression in tumor cells and immune cells in hepatocellular carcinoma except using the SP142 assay. Despite good reliability of pathologists in PD-L1 scoring, there were still up to 18% of our hepatocellular carcinoma samples misclassified by individual pathologists in comparison to the consensus score at the cutoff of combined positive score  $\geq 1$  (followed by 12% at tumor proportion score  $\geq 1\%$ , 9.8% at combined positive score  $\geq 10$  and 6% at combined positive score  $\geq 20$ ). The inter-observer agreement is generally poorer at a lower cutoff similarly observed in other studies [11–14, 25]. To improve the scoring accuracy, formal training program for PD-L1 assessment might be helpful, but its effect does not appear to be substantial [11, 14]. Evaluation of PD-L1 expression by automated digital image analysis is a potentially promising solution but requires proper validation in patients' sample from clinical trials [29, 30].

Compared to other three assays, the SP142 assay not only highlighted smaller amount of PD-L1 positive cells but

also provided weaker staining intensity [12, 13]. Such a staining property results in lower concordance with the other three assays and higher inter-observer variability. Nevertheless, we cannot overinterpret this finding to deny the usefulness of the SP142 assay because different groups demonstrated the clinical and biological significances of PD-L1 expression by the SP142 assay in hepatocellular carcinoma [31, 32]. Moreover, the SP142 assay showed the weakest correlation with immune-related gene signatures among the four standardized PD-L1 assays but the correlation was still significantly positive.

In the KEYNOTE 224 trial, the statistically significant association between objective response and the PD-L1 combined positive score was generated by a one-sided logistic regression test, which implies that the PD-L1 combined positive score was regarded as a continuous parameter rather than a dichotomous parameter [6]. Our current study showed that the combined positive score (in continuous form) is more strongly correlated with gene signatures associated with tumor immune microenvironment than the tumor proportion score. Hence, it is also worthy of exploring the potential predictive role of gene signatures associated with tumor immune microenvironment for PD-1/PD-L1 immunotherapy in hepatocellular carcinoma. The NanoString system is one of robust platforms for evaluation of multiplex gene expression with high sensitivity, rapid turnaround time, good reproducibility and minimal RNA requirement [33]. Compared to qRT-PCR and RNAseq, the NanoString system provides more reliable and consistent results on the formalin-fixed, paraffin embedded tissue, which contribute to most archived clinical samples [34]. More importantly, interferon gamma-related and T-cell inflamed gene expression profiles based on the NanoString system were shown to predict clinical response to pembrolizumab in multiple cancer types [21, 35].

Our present study is limited by using a retrospective cohort of hepatocellular carcinoma patients without receiving PD-1/PD-L1 immunotherapy. In the absence of clinical response data, we are unable to analyze the predictive performance of the four PD-L1 assays or elucidate the clinical significance of those samples with discordant results by different PD-L1 assays. Moreover, we did not explore novel gene expression profiles associated with PD-L1 protein expression because it may not be clinically significant to establish gene expression profiles without treatment outcome data. Furthermore, we did not evaluate PD-L1 expression in immune cells alone [immune cell proportion score] due to lacking clinical evidence supporting the predictive role of the immune cell proportion score in hepatocellular carcinoma PD-1/PD-L1 treatment [5, 6, 36].

In conclusion, the 22C3, 28–8, and SP263 assays are highly concordant in PD-L1 scoring in hepatocellular

carcinoma and suggest the interchangeability of these three assays. Pathologists are reliable in PD-L1 scoring with the high inter-observer agreement, but it is still necessary for the further improvement of the accuracy in assessing PD-L1 expression at a low cutoff. Exploration of the potential predictive role of gene signatures associated with tumor immune microenvironment for PD-1/PD-L1 immunotherapy in hepatocellular carcinoma is also warranted.

**Acknowledgements** This work was supported in part by a Collaborative Research Fund from the Hong Kong Research Grants Council (Ref. No.: C4041–17).

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

- Garon EB, Rizvi NA, Hui R, Leighl N, Balmanoukian AS, Eder JP, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. *N Engl J Med*. 2015;372:2018–28.
- Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science*. 2017;357:409–13.
- Powles T, Duran I, van der Heijden MS, Loriot Y, Vogelzang NJ, De Giorgi U, et al. Atezolizumab versus chemotherapy in patients with platinum-treated locally advanced or metastatic urothelial carcinoma (IMvigor211): a multicentre, open-label, phase 3 randomised controlled trial. *Lancet*. 2018;391:748–57.
- Schachter J, Ribas A, Long GV, Arance A, Grob JJ, Mortier L, et al. Pembrolizumab versus ipilimumab for advanced melanoma: final overall survival results of a multicentre, randomised, open-label phase 3 study (KEYNOTE-006). *Lancet*. 2017;390:1853–62.
- El-Khoueiry AB, Sangro B, Yau T, Crocenzi TS, Kudo M, Hsu C, et al. Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial. *Lancet*. 2017;389:2492–502.
- Zhu AX, Finn RS, Edeline J, Cattani S, Ogasawara S, Palmer D, et al. Pembrolizumab in patients with advanced hepatocellular carcinoma previously treated with sorafenib (KEYNOTE-224): a non-randomised, open-label phase 2 trial. *Lancet Oncol*. 2018;19:940–52.
- Hirsch FR, McElhinny A, Stanforth D, Ranger-Moore J, Jansson M, Kulangara K, et al. PD-L1 immunohistochemistry assays for lung cancer: results from phase 1 of the blueprint PD-L1 IHC assay comparison project. *J Thorac Oncol*. 2017;12:208–22.
- European Association for the Study of the Liver. EASL clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol*. 2018;69:182–236.
- Heimbach JK, Kulik LM, Finn RS, Sirlin CB, Abecassis MM, Roberts LR, et al. AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology*. 2018;67:358–80.
- Nault JC, Galle PR, Marquardt JU. The role of molecular enrichment on future therapies in hepatocellular carcinoma. *J Hepatol*. 2018;69:237–47.
- Brunstrom H, Johansson A, Westbom-Fremer S, Backman M, Djureinovic D, Patthey A, et al. PD-L1 immunohistochemistry in clinical diagnostics of lung cancer: inter-pathologist variability is higher than assay variability. *Mod Pathol*. 2017;30:1411–21.
- Adam J, Le Stang N, Rouquette I, Cazes A, Badoual C, Pinot-Roussel H, et al. Multicenter harmonization study for PD-L1 IHC testing in non-small-cell lung cancer. *Ann Oncol*. 2018;29:953–8.
- Chan AWH, Tong JHM, Kwan JSH, Chow C, Chung LY, Chau SL, et al. Assessment of programmed cell death ligand-1 expression by 4 diagnostic assays and its clinicopathological correlation in a large cohort of surgical resected non-small cell lung carcinoma. *Mod Pathol*. 2018;31:1381–90.
- Tsao MS, Kerr KM, Kockx M, Beasley MB, Borczuk AC, Botling J, et al. PD-L1 immunohistochemistry comparability study in real-life clinical samples: results of blueprint phase 2 project. *J Thorac Oncol*. 2018;13:1302–11.
- Pinato DJ, Mauri FA, Spina P, Cain O, Siddique A, Goldin RD, et al. Quantitative comparison of PD-L1 immuno-histochemical assays in hepatocellular carcinoma: the blueprint-HCC study. *J Clin Oncol*. 2018;36:91.
- Theise ND, Curado MP, Franceschi S, Htiroglou P, Kudo M, Park YN, et al. Hepatocellular carcinoma. In: Bosman FT, Carneiro F, Hruban RH, Theise ND, (eds). WHO Classification of Tumours of the Digestive System International Agency for Research on Cancer: Lyon; 2010. pp 205–16.
- Chan AW, Tong JH, Pan Y, Chan SL, Wong GL, Wong VW, et al. Lymphoepithelioma-like hepatocellular carcinoma: an uncommon variant of hepatocellular carcinoma with favorable outcome. *Am J Surg Pathol*. 2015;39:304–12.
- Chan AW, Yu S, Yu YH, Tong JH, Wang L, Tin EK, et al. Steatotic hepatocellular carcinoma: a variant associated with metabolic factors and late tumour relapse. *Histopathology*. 2016;69:971–84.
- Kulangara K, Zhang N, Corigliano E, Guerrero L, Waldroup S, Jaiswal D, et al. Clinical utility of the combined positive score for programmed death ligand-1 expression and the approval of pembrolizumab for treatment of gastric cancer. *Arch Pathol Lab Med*. 2019;143:330–7.
- Cesano A. nCounter(R) PanCancer Immune Profiling Panel (NanoString Technologies, Inc., Seattle, WA). *J Immunother Cancer*. 2015;3:42.
- Ayers M, Lunceford J, Nebozhyn M, Murphy E, Loboda A, Kaufman DR, et al. IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest*. 2017;127:2930–40.
- Cancer Genome Atlas Research Network. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell*. 2017;169:1327–41 e23.
- Bellmunt J, de Wit R, Vaughn DJ, Fradet Y, Lee JL, Fong L, et al. Pembrolizumab as second-line therapy for advanced urothelial carcinoma. *N Engl J Med*. 2017;376:1015–26.
- Burtneess B, Harrington KJ, Greil R, Soulières D, Tahara M, De Castro G, et al. LBA8\_PRKEYNOTE-048: phase III study of first-line pembrolizumab (P) for recurrent/metastatic head and neck squamous cell carcinoma (R/M HNSCC). *Ann Oncol*. 2018;29:mdy424.045.
- Rimm DL, Han G, Taube JM, Yi ES, Bridge JA, Flieder DB, et al. A prospective, multi-institutional, pathologist-based assessment of 4 immunohistochemistry assays for PD-L1 expression in non-small cell lung cancer. *JAMA Oncol*. 2017;3:1051–8.
- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155–63.
- Calderaro J, Rousseau B, Amaddeo G, Mercey M, Charpy C, Costentin C, et al. Programmed death ligand 1 expression in hepatocellular carcinoma: relationship with clinical and pathological features. *Hepatology*. 2016;64:2038–46.
- Wang C, Hahn E, Slodkowska E, Eskander A, Enepekides D, Higgins K, et al. Reproducibility of PD-L1 immunohistochemistry

- interpretation across various types of genitourinary and head/neck carcinomas, antibody clones, and tissue types. *Hum Pathol*. 2018;82:131–9.
29. Taylor CR, Jadhav AP, Gholap A, Kamble G, Huang J, Gown A, et al. A multi-institutional study to evaluate automated whole slide scoring of immunohistochemistry for assessment of programmed death-ligand 1 (PD-L1) expression in non-small cell lung cancer. *Appl Immunohistochem Mol Morphol*. 2019;27:263–9.
  30. Koelzer VH, Gisler A, Hanhart JC, Griss J, Wagner SN, Willi N, et al. Digital image analysis improves precision of PD-L1 scoring in cutaneous melanoma. *Histopathology*. 2018;73:397–406.
  31. Lu LC, Lee YH, Chang CJ, Shun CT, Fang CY, Shao YY, et al. Increased expression of programmed death-ligand 1 in infiltrating immune cells in hepatocellular carcinoma tissues after Sorafenib treatment. *Liver Cancer*. 2019;8:110–20.
  32. Yarchoan M, Xing D, Luan L, Xu H, Sharma RB, Popovic A, et al. Characterization of the immune microenvironment in hepatocellular carcinoma. *Clin Cancer Res*. 2017;23:7333–9.
  33. Veldman-Jones MH, Lai Z, Wappett M, Harbron CG, Barrett JC, Harrington EA, et al. Reproducible, quantitative, and flexible molecular subtyping of clinical DLBCL samples using the NanoString nCounter system. *Clin Cancer Res*. 2015;21:2367–78.
  34. Reis PP, Waldron L, Goswami RS, Xu W, Xuan Y, Perez-Ordóñez B, et al. mRNA transcript quantification in archival samples using multiplexed, color-coded probes. *BMC Biotechnol*. 2011;11:46.
  35. Ott PA, Bang YJ, Piha-Paul SA, Razak ARA, Bennouna J, Soria JC, et al. T-cell-inflamed gene-expression profile, programmed death ligand 1 expression, and tumor mutational burden predict efficacy in patients treated with pembrolizumab across 20 cancers: KEYNOTE-028. *J Clin Oncol*. 2019;37:318–27.
  36. Lee KH, Hsu CH, Lee MS, Ryoo BY, Verret W, He AR, et al. 1500Atezolizumab+bevacizumab in hepatocellular carcinoma (HCC): safety and clinical activity results from a phase Ib study. *Ann Oncol*. 2018;29:mdy432.002.

## Affiliations

Lei Shi<sup>1</sup> · Shu-Jun Zhang<sup>2</sup> · Jun Chen<sup>3</sup> · Shi-Xun Lu<sup>4</sup> · Xin-Juan Fan<sup>5</sup> · Joanna Hung-Man Tong<sup>6</sup>  · Chit Chow<sup>6</sup> · Edith Ka-Yee Tin<sup>6</sup> · Stephen Lam Chan<sup>7,8</sup> · Charing Ching-Ning Chong<sup>9</sup> · Paul Bo-San Lai<sup>8,9</sup> · Ka-Fai To<sup>6,8</sup> · Nathalie Wong<sup>6,8</sup> · Anthony Wing-Hung Chan<sup>6</sup>

<sup>1</sup> Department of Pathology, the Second Xiangya Hospital of Central South University, Changsha, Hunan, China

<sup>2</sup> Department of Pathology, the Fourth Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang, China

<sup>3</sup> Department of Pathology, Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University Medical School, Nanjing, Jiangsu, China

<sup>4</sup> Department of Pathology, Sun Yat-sen University Cancer Center, Guangzhou, Guangdong, China

<sup>5</sup> Department of Pathology, the Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong, China

<sup>6</sup> Department of Anatomical and Cellular Pathology, State Key Laboratory of Translational Oncology, Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong, China

<sup>7</sup> Department of Clinical Oncology, State Key Laboratory of Translational Oncology, Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong, China

<sup>8</sup> Institute of Digestive Disease, Partner State Key Laboratory of Digestive Disease, The Chinese University of Hong Kong, Hong Kong, China

<sup>9</sup> Division of Hepatobiliary and Pancreatic Surgery, Department of Surgery, The Chinese University of Hong Kong, Hong Kong, China