PD-1T TILs as a predictive biomarker for clinical benefit to PD-1 blockade in patients with advanced NSCLC

Hummelink, K.; Noort, V. van der; Muller, M.; Schouten, R.D.; Lalezari, F.; Peters, D.; ... ; Thommen, D.S.

PD-1^T TILs as a Predictive Biomarker for Clinical Benefit to PD-1 Blockade in Patients with Advanced NSCLC

Karlijn Hummelink[1,2], Vincent van der Noort[3], Mirte Muller[2], Robert D. Schouten[2], Ferry Lalezari[4], Dennis Peters[5], Willemijn S.M.E. Theelen[2], Viktor H. Koelzer[6], Kirsten D. Mertz[7], Alfred Zippelius[8], Michel M. van den Heuvel[2], Annegien Broeks[5], John B.A.G. Haanen[9], Ton N. Schumacher[10], Gerrit A. Meijer[1], Egbert F. Smit[2], Kim Monkhorst[1], and Daniela S. Thommen[9]

ABSTRACT

Purpose: Durable clinical benefit to PD-1 blockade in non–small cell lung cancer (NSCLC) is currently limited to a small fraction of patients, underlining the need for predictive biomarkers. We recently identified a tumor-reactive tumor-infiltrating T lymphocyte (TIL) pool, termed PD-1^T TILs, with predictive potential in NSCLC. Here, we examined PD-1^T TILs as biomarker in NSCLC.

Experimental Design: PD-1^T TILs were digitally quantified in 120 baseline samples from advanced NSCLC patients treated with PD-1 blockade. Primary outcome was disease control (DC) at 6 months. Secondary outcomes were DC at 12 months and survival. Exploratory analyses addressed the impact of lesion-specific responses, tissue sample properties, and combination with other biomarkers on the predictive value of PD-1^T TILs.

Results: PD-1^T TILs as a biomarker reached 77% sensitivity and 67% specificity at 6 months, and 93% and 65% at 12 months, respectively. Particularly, a patient group without clinical benefit was reliably identified, indicated by a high negative predictive value (NPV) (88% at 6 months, 98% at 12 months). High PD-1^T TILs related to significantly longer progression-free (HR 0.39, 95% CI, 0.24–0.63, $P < 0.0001$) and overall survival (HR 0.46, 95% CI, 0.28–0.76, $P < 0.01$). Predictive performance was increased when lesion-specific responses and samples obtained immediately before treatment were assessed. Notably, the predictive performance of PD-1^T TILs was superior to PD-L1 and tertiary lymphoid structures in the same cohort.

Conclusions: This study established PD-1^T TILs as predictive biomarker for clinical benefit to PD-1 blockade in patients with advanced NSCLC. Most importantly, the high NPV demonstrates an accurate identification of a patient group without benefit.

See related commentary by Anagnostou and Luke, p. 4835

Introduction

Immune checkpoint blockade (ICB) targeting the programmed cell death-1 (PD-1)/PD-ligand 1 (PD-L1) pathway has dramatically changed the treatment of patients with advanced stage non–small cell lung cancer (NSCLC). Significant improvement in survival, quality of life and a favorable safety profile compared to chemotherapy has led to the rapid and broad clinical implementation of this treatment modality (1–6). However, approximately 60% to 70% of patients progress within 6 months after treatment initiation (3, 5, 6). Hence, predictive biomarkers are needed, in particular to identify patients that are less likely to benefit to reduce overtreatment.

In analogy to molecular biomarkers that have been used for identification of patients with targetable oncogenes (7), it has been assumed that PD-L1 expression in tumors could predict benefit of anti-PD-1/PD-L1 therapy. Previous studies indeed have shown that pretreatment stratification based on high expression of PD-L1 can identify patient subgroups with improved response rates and survival (1, 2, 8), leading to the approval of PD-L1 testing for newly diagnosed advanced NSCLC. However, PD-L1 is not a perfect biomarker since multiple studies have shown conflicting results with regard to its predictive potential (3, 5, 6).

As PD-1/PD-L1 blockade is thought to reactivate dysfunctional T cells (9), an alternative strategy may be to develop biomarkers that reflect the capacity of a tumor to mount an antitumor immune response. We previously showed that the presence of a specific CD8$^+$ tumor-infiltrating lymphocyte (TIL) subpopulation, termed PD-1^T TILs, correlated with response and survival in a small cohort of patients with NSCLC treated with PD-1 blockade (10). PD-1^T TILs are a subset of PD-1+ T cells characterized by high, tumor-associated expression levels of PD-1, and are transcriptionally and functionally distinct from other TIL populations with lower or no PD-1 expression. Importantly,

[1]Department of Pathology, Division of Diagnostic Oncology, the Netherlands Cancer Institute, Amsterdam, the Netherlands. [2]Department of Thoracic Oncology, Division of Medical Oncology, the Netherlands Cancer Institute, Amsterdam, the Netherlands. [3]Department of Biometrics, the Netherlands Cancer Institute, Amsterdam, the Netherlands. [4]Department of Radiology, Division of Diagnostic Oncology, the Netherlands Cancer Institute, Amsterdam, the Netherlands. [5]Core Facility Molecular Pathology and Biobanking, Division of Molecular Pathology, the Netherlands Cancer Institute, Amsterdam, the Netherlands. [6]Department of Pathology and Molecular Pathology, University Hospital Zurich, Zurich, Switzerland. [7]Institute of Pathology, Cantonal Hospital Baselland, Liestal, Switzerland. [8]Department of Biomedicine, University Hospital Basel, Basel, Switzerland. [9]Division of Molecular Oncology and Immunology, the Netherlands Cancer Institute, Amsterdam, the Netherlands. [10]Division of Molecular Oncology and Immunology, the Netherlands Cancer Institute, Oncode Institute, Amsterdam, the Netherlands.

K. Monkhorst and D.S. Thommen contributed equally to this article.

Current address for M.M. van den Heuvel: Department of Pulmonary Diseases, Radboud University Medical Center, Nijmegen, the Netherlands; and current address for E.F. Smit, Department of Pulmonary Diseases, Leiden University Medical Center, Leiden, the Netherlands.

Corresponding Authors: Daniela S. Thommen, Division of Molecular Oncology and Immunology, Netherlands Cancer Institute, Plesmanlaan 121, Amsterdam, 1066 CX, the Netherlands. E-mail: d.thommen@nki.nl; and Kim Monkhorst, k.monkhorst@nki.nl

Clin Cancer Res 2022;28:4893–906

doi: 10.1158/1078-0432.CCR-22-0992

Translational Relevance

Despite the clinical success of anti–PD-1 treatment, robust predictive biomarkers are still lacking. As PD-1 blockade can reinvigorate dysfunctional T cells, we hypothesized that new biomarkers could be developed by assessing such direct effectors of the antitumor immune response. We previously identified a tumor-reactive T-cell population, termed PD-1^T TILs, with predictive potential in a small cohort of patients with non–small cell lung cancer (NSCLC). In this study, PD-1^T TILs were assessed as a predictive biomarker for durable clinical benefit in two NSCLC cohorts treated with PD-1 blockade, reaching high-sensitivity and high-negative predictive value. The predictive performance was superior compared with PD-L1 and tertiary lymphoid structures. Therefore, this biomarker may positively impact treatment decision making in clinical practice, as it improves patient stratification. Importantly, it specifically identifies a patient group that is unlikely to benefit from PD-1 blockade, thereby providing a tool to reduce overtreatment.

PD-1^T TILs show high tumor reactivity (10) consistent with subsequent work in other tumor types demonstrating that the capacity for tumor recognition is strongly enriched in the dysfunctional T-cell population that expresses high levels of PD-1 (11, 12). Moreover, tumor infiltration by PD-1^T lymphocytes was recently associated with immunologic response to PD-1 blockade in a number of other tumor types (13). Finally, PD-1^T TILs predominantly localize in tertiary lymphoid structures (TLS; ref. 10), which have been correlated with clinical and immunologic response to ICB in other cancer types (13–16). Collectively, these observations suggest that the presence of PD-1^T TILs in a tumor may indicate that a tumor-specific T-cell response has been mounted, and thereby represent a potential biomarker to preselect patients for treatment with PD-1 blockade. Particularly, the absence of PD-1^T TILs in a tumor may signify the lack of a tumor-reactive T-cell population and hence identify patients that are unlikely to benefit.

In this retrospective observational study, we analyzed pretreatment samples from two independent cohorts of patients with NSCLC treated with PD-1 blockade to (i) train and validate PD-1^T TILs as a predictive biomarker; (ii) explore whether certain sample characteristics such as sample type, sample location, or time of sampling influence the predictive value of this biomarker; and (iii) evaluate the potential for clinical implementation, by comparing and combining PD-1^T TILs with other predictive markers such as PD-L1 and TLS.

Materials and Methods

Patient enrollment and study endpoints

In this study, 164 patients with stage IV NSCLC were identified from two independent cohorts who started second or later line monotherapy with nivolumab ($n = 128$) or pembrolizumab ($n = 36$) between March 2015 and April 2018 at the Netherlands Cancer Institute/Antoni van Leeuwenhoek hospital (NKI-AVL), The Netherlands. All patients had pathologically confirmed stage IV NSCLC. Absence of sensitizing EGFR mutations or ALK translocations was confirmed in 145 patients, whereas in 19 patients mutation status was unknown. Patients received single agent nivolumab 3 mg/kg, administered as an intravenous infusion, every 2 weeks for at least one dose or single-agent pembrolizumab 200 mg as an intravenous infusion every 3 weeks. Nivolumab was provided within the Expanded Access Programme (EAP) from Bristol Myers Squibb or in regular care after the drug was registered. Pembrolizumab-treated patients were part of the control arm in the PEMBRO-RT study (NCT02492568; ref. 17). Patients were randomized into a training and validation set. Randomization was stratified by type of treatment (nivolumab vs. pembrolizumab) and treatment outcome at 6 months.

RECIST v1.1 was used to assess efficacy. Patients with progressive disease (PD) who were not evaluable for response by RECIST were determined by the treating physician as PD. Disease control [DC; complete response (CR)/partial response (PR) or stable disease (SD)] at 6 months following initiation of treatment was used as the primary clinical outcome measure. We assessed DC at 12 months (CR/PR/SD that lasted ≥12 months), progression-free survival (PFS) and overall survival (OS) as secondary outcome measures to predict long-term efficacy to PD-1 blockade. PFS and OS were defined as the time from the date of initiation of treatment with PD-1 blockade to the date of progression or death (for PFS) or death (for OS). Patients who had not progressed or died were censored at the date of their last follow-up.

Pretreatment formalin-fixed paraffin embedded (FFPE) tumor tissue samples were collected from all patients. Written informed consent was obtained from all patients for research usage of material not required for diagnostic use by institutionally implemented opt-out procedure. The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Research Board of NKI (CFMPB586). Forty-four patients (27%) were excluded based on the following criteria: samples contained less than 10,000 cells in the tumor area on a single cross-sectional slide ($n = 15$), were obtained more than 2 years before start of PD-1 blockade ($n = 14$), were obtained from endobronchial lesions ($n = 8$), contained normal lymphoid tissue ($n = 3$), showed fixation and/or staining artefacts ($n = 3$), or were non-NSCLC histology ($n = 1$; **Fig. 1A**; Supplementary Table S1). We excluded bronchial biopsies as they frequently showed unspecific antibody staining due to mechanical damage, and lymph node resections due to the presence of PD-1 bright T cells in normal lymphoid tissue, which could potentially lead to false-positive results. Prespecified subgroup analyses were performed to compare (i) samples derived from tumor resections and biopsies, (ii) samples from primary and metastatic sites, and (iii) samples that were obtained either directly before the start of nivolumab or pembrolizumab or before any prior line of systemic treatment.

Fresh tumor samples were collected from 16 patients with NSCLC undergoing primary surgical treatment between July 2017 and February 2019 at NKI-AVL. The study was approved by the Institutional Research Board of NKI-AVL (CFMPB484). All patients consented to research usage of material not required for diagnostic use either by opt-out procedure or via prior informed consent (after May 23, 2018). Representative tumor tissue samples were procured from surgical resection specimens by a pathologist. Half of each sample was formalin-fixed and embedded in paraffin for further histologic analysis, the other half was immediately processed into tumor fragments that were cryopreserved until further usage (see sample processing and flow cytometry analysis).

Sample processing and flow cytometry analysis

For flow cytometry analysis, cryopreserved tissue fragments were thawed and processed into single-cell suspensions by enzymatic digestion using RPMI1640 medium (Thermo Fisher Scientific) supplemented with 1% Penicillin–Streptomycin (Roche, 12.6 μg/mL Pulmozyme (Roche) and 1 mg/mL Collagenase type IV (Sigma), as described previously (14). Samples were then washed in PBS (Sigma), filtered over a 150 μm/L filter mesh, resuspended in 50 μL PBS, and

incubated with Fc receptor blocking agent (eBioscience) and with live/dead Zombie UV (Biolegend) for 20 minutes at 4°C. Cells were washed, resuspended in 50 μL of staining buffer [PBS (Sigma), 0.5% BSA (Sigma), 0.1% NaN_3 (Invitrogen)] containing the below-described antibodies, and incubated for 20 minutes at 4°C. After washing twice, cells were taken up in 200 μL IC fixation buffer (eBioscience) and incubated for 20 minutes. Subsequently, samples were washed twice before data acquisition.

For staining the following antibodies were used: anti-CD45 PerCP Cy5.5 (2D1, RRID:AB_1548697) from Invitrogen; anti-CD8 BUV563 (RPA-T8, RRID:AB_2870199), -PD-1 PE-Cy7 (EH12.1, RRID: AB_10611585), all from BD Biosciences; anti-CD3 FITC (SK7, RRID_AB2043993), -CD4 BV421 (SK3 RRID:AB_2566015), all from Biolegend. PD-1^T lymphocytes were identified by using peripheral blood T cells from healthy donors as external reference to establish the cut-off as described previously (10). Data acquisition were carried out on a BD LSR II SORP cell analyzer (BD Biosciences). Data were collected using the BD FACS Diva Software version 8.0.1, and further analyzed with FlowJo v10.6.1 (Tree Star Inc.) and GraphPad Prism v8.0e (GraphPad Software Inc.).

IHC

Separate IHC staining of consecutive FFPE tumor tissue sections were performed on a BenchMark Ultra autostainer Instrument (Ventana Medical Systems). Paraffin sections were cut at 3 μm. Sections for PD-1 staining were dried overnight at room temperature and stained within 48 hours to reduce background staining. Prior to staining, sections were initially baked at 75°C for 28 minutes and deparaffinised in the instrument with EZ prep solution (Ventana Medical Systems). Heat-induced antigen retrieval was carried out using Cell Conditioning 1 (CC1; Ventana Medical Systems) for 32 minutes (CD68 and CD20-CD3 double staining) or 48 minutes (PD-1 and PD-L1) at 95°C.

PD-1 was detected using clone NAT105 (Lot no. V0002089, Ready-to-Use, 16 minutes at RT; Roche Diagnostics; Catalog No. 7099029001). PD-L1 was detected using clone 22C3 (1/40 dilution, 1 hour at RT; Agilent/DAKO) and CD68 was detected using clone KP1 (1/10,000 dilution, 32 minutes at 37°C; Agilent/DAKO). Bound antibody was detected using the OptiView DAB Detection Kit (Ventana Medical Systems). Slides were counterstained with hematoxylin and bluing reagent (Ventana Medical Systems).

For double staining of CD20 (Yellow) and CD3 (Purple), CD20 was detected in the first sequence using clone L26 (1/800 dilution, 32 minutes at 37°C, Agilent/DAKO). CD20 bound antibody was visualized using anti-Mouse NP (Ventana Medical systems) for 12 minutes at 37°C followed by anti-NP AP (Ventana Medical systems) for 12 minutes at 37°C, followed by the Discovery Yellow Detection Kit (Ventana Medical Systems). In the second sequence of the double-staining procedure, CD3 was detected using clone SP7 (1:100 dilution, 32 minutes at 37°C; Thermo Fisher Scientific). CD3 was visualized using anti-Rabbit HQ (Ventana Medical systems) for 12 minutes at 37°C followed by anti-HQ HRP (Ventana Medical systems) for 12 minutes at 37°C, followed by the Discovery Purple Detection Kit (Ventana Medical Systems). Slides were counterstained with hematoxylin and bluing reagent (Ventana Medical Systems).

PD-1, PD-L1, and CD68 immunostainings were scanned at 20× magnification with a resolution of 0.50 per μm^2 using an Aperio slide AT2 scanner (Leica Biosystems). CD20-CD3 immunostaining was scanned at 20× magnification with a resolution of 0.24 per μm^2 using a 3DHistech P1000 scanner.

For manual scoring, PD-L1 and CD68 IHC images were uploaded on Slidescore, a digital pathology slide web platform that integrates a slide viewer with a scoring sheet (https://www.slidescore.com). PD-1^T TILs, CD20, and TLS were digitally scored as described below.

Digital quantification of PD-1^T TILs

PD-1^T TILs are a subset of PD-1+ T cells in the tumor tissue that can be identified both by flow cytometry and by immunohistochemistry (IHC). To quantify PD-1^T TILs in FFPE tissue, a digital workflow using a PD-1^T IHC scoring algorithm was established previously (10). For this study, the automated detection of PD-1^T TILs was recalibrated using the Multiplex IHC v1.2 module of the HALO software, v2.3.2089.69 (Indica Labs). To this end, an independent set of 16 NSCLC tumor samples was used to perform flow cytometry and IHC analysis in parallel. PD-1^T TILs are defined by bright, tumor-associated PD-1 expression at levels that exceed those observed on peripheral blood T cells (10). Hence, to determine the frequency of PD-1^T TILs in the NSCLC samples, PD-1 expression on intratumoral lymphocytes was assessed by flow cytometry and compared with peripheral blood T cells as external reference to establish the threshold for tumor-associated PD-1 expression (Supplementary Fig. S1A). Next, a digital IHC algorithm to quantify PD-1+ lymphocytes in matched FFPE samples was generated (Supplementary Fig. S1B). The optical density (OD) measured by this approach is reflective of staining intensity and thereby PD-1 levels. To identify the optimal OD cut-off resulting in similar frequencies of PD-1^T TILs by IHC as by flow cytometry, Pearson correlation coefficients were determined using thresholds varying from 0.2 to 0.5 OD. The percentage PD-1 bright lymphocytes obtained for each OD threshold in FFPE samples were normalized to total lymphocyte counts and compared with the flow cytometry-guided annotation of PD-1^T lymphocytes. An OD of 0.25 showed the highest Pearson correlation coefficient ($R^2 = 0.615$; $P < 0.001$; Supplementary Fig. S1C and S1D) and was selected as the threshold for further automated PD-1^T TIL quantification in FFPE tumor tissue.

For prediction of clinical benefit to PD-1 blockade, the tumor areas were measured and the number of PD-1^T TILs per mm^2 tumor area was determined (Supplementary Table S2). To this end, tumor areas were annotated with a 0.5 mm margin from the tumor border and necrotic areas were excluded with a 0.5 mm margin. Digital image analysis was carried out by a trained MD (K.H.) and supervised by an experienced pathologist (K.M.), blinded for clinical outcome. ROC curves were used in the training set to establish an optimal cut-off of 90 PD-1^T TILs per mm^2 for discriminating patients with and without clinical benefit (see Results).

PD-L1 scoring

Tumor PD-L1 expression was assessed according to the instruction manual of the qualitative, clinical grade LDT IHC assay (22C3 pharmDx; Dako) as used in routine clinical practice at NKI-AVL. As high concordance between the 22C3 and 22–8 PD-L1 antibodies has been reported (18, 19), the 22C3 clone was also used to assess the predictive value of PD-L1 for nivolumab. PD-L1 expression levels were manually scored by a trained MD (K.H.) under the supervision of an experienced pathologist (K.M.) blinded for clinical outcome. The PD-L1 tumor proportion score (TPS) was determined by calculating the percentage of PD-L1+ tumor cells of total viable tumor cells (Supplementary Table S2). PD-L1 positivity was defined as tumor cells showing circumferential and/or partial linear expression (at any intensity) of PD-L1 on the plasma cell membrane. A CD68 staining was manually evaluated and compared with PD-L1 stained slides to avoid false-positive results due to PD-L1 expressing macrophages in between tumor cells. PD-L1 IC was manually scored as the proportion of tumor area that is occupied by PD-L1+ immune cells (IC) of any

intensity (IC0: <1%; IC1: ≥1% and <5%; IC2: ≥5% and <10%; and IC3: ≥10%) as described (20, 21).

Scoring of TLS

A CD20 (yellow)/CD3 (purple) double staining was used to identify TLS. CD20-CD3 IHC images were scanned and analyzed using HALO. Lymphoid niches were manually identified based on the presence of B cell ($CD20^+$) clusters and T cell ($CD3^+$) zones as described (22, 23). Next, areas were measured in HALO and assigned as TLS (>60,000 μm^2) or lymphoid aggregate (LA; 10,000–60,000 μm^2; ref. 16). Finally, tumor areas were digitally annotated as described above and the number of TLS per mm^2 and the combined number of TLS and LA (TLS+LA) per mm^2 tumor area were determined (Supplementary Table S2).

CD20 quantification by digital image analysis

The Area Quantification v1.0 module of the HALO software was used to generate an analysis algorithm to measure the total area with CD20 expression on the CD20/CD3 images. The total $CD20^+$ area was selected because the dense clustering of $CD20^+$ cells in TLS precluded the setup of a reliable algorithm to quantify cell numbers. Tumor areas were digitally annotated as described above and the $CD20^+$ area was normalized per mm^2 tumor area (Supplementary Table S2).

Statistical analysis

Patient characteristics were descriptively reported using mean ± SD, IQR, or frequencies (percentages). Differences in patient and sample characteristics between cohorts (training and validation), between outcome groups (disease control vs. PD) and between groups created by the biomarker were assessed using the Mann–Whitney test for continuous data, Fisher exact test for categorical data, the linear-by-linear association test for ordinal variables, the unpaired *t*-test for variables with two levels and the Kruskal–Wallis test for variables with more than two levels. Differences were considered statistically significant if *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$, or ****, $P < 0.0001$.

Calculation of the area under the ROC curve (AUC) was used as a measure of discriminatory ability for the biomarkers considered. The predictive performance of different biomarkers or biomarker combinations on the same patient population was described in terms of sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) and compared using the McNemar test. The predictive accuracy of the same biomarker on different samples (e.g., resections vs. biopsies) was assessed using AUCs and compared in a one-sided permutation test. Survival curves were plotted using the Kaplan–Meier method and compared between groups identified by the various biomarkers using the log-rank test.

To assess the predictive performance of PD-1^T TILs (discretized at 90 per mm^2) and PD-L1 (discretized at either 1% or 50%) in combination, bivariate models were constructed using the validation cohort. We considered two types of models: in one case, patients were considered to have clinical benefit if both (PD-L1 and PD-1^T TILs), or one of the two markers were above their respective threshold. Patients were considered to experience disease progression if both markers were below their respective threshold. In the other case, patients were considered to have clinical benefit only if both markers (PD-L1 and PD-1^T TILs) were above their respective threshold. Patients were considered to experience disease progression if both, or one of the two markers were below their respective threshold. As the first model yielded the better predictive performance, we used this model to test the two choices for the PD-L1 threshold.

Bivariate models of PD-L1 TPS (discretized at either 1% and 50%) and PD-L1 IC (discretized at a score of 2) were constructed using all nivolumab-treated patients ($n = 94$). The same type of model was used as described for PD-1^T TILs and PD-L1 TPS above. Correlations between PD-L1 TPS and PD-L1 IC or PD-1^T TILs and PD-L1 TPS, respectively, were evaluated using linear regression analysis.

Data availability

All relevant data are presented in the paper or included as Supplementary Tables. Raw data generated in this study are available upon reasonable request from the corresponding authors.

Results

PD-1^T TILs as biomarker in NSCLC

To assess their predictive potential, we quantified PD-1^T TILs in pretreatment samples from 120 patients with advanced stage NSCLC treated with either pembrolizumab ($n = 26$) or nivolumab ($n = 94$). Because the pembrolizumab-treated cohort was substantially smaller, we randomized half of the pembrolizumab treated and one third of the nivolumab treated patients in a training set ($n = 43$). The remainder of the patients was included in a validation set ($n = 77$; **Fig. 1A**). Each sample set consisted of 30% of patients that obtained disease control (DC) at 6 months of treatment with PD-1 blockade. Clinicopathological characteristics and treatment outcomes are summarized in Supplementary Table S3. Sample characteristics are shown in Supplementary Table S4. None of these characteristics differed significantly among the training and validation set.

PD-1^T TILs are a subpopulation of PD-1+ T cells defined by a bright, tumor-specific PD-1 expression level. To quantify the PD-1^T TIL subset in FFPE tissue, we established an automated digital quantification workflow as described previously (10), allowing to reliably distinguish these cells from other PD-1+ cells (**Fig. 1B**; Supplementary Fig. S1 and Materials and Methods). Next, we determined the frequency of PD-1^T TILs per mm^2 tumor area that best discriminated patients with or without DC at 6 months (DC 6m) in the training set ($n = 43$). To minimize the risk of undertreatment due to misclassification of patients with clinical benefit, we aimed for a sensitivity and negative predictive value (NPV) of ≥90%, and a specificity of the biomarker of at least 50% to limit overtreatment. Sensitivity and specificity reflect the predictive accuracy of identifying patients with DC 6 m and with PD, respectively. The NPV reflects the probability of having no benefit to PD-1 blockade for patients with a biomarker result below threshold.

In the training set, the median number of PD-1^T TILs per mm^2 was 255 with an IQR between 86 and 356 in the DC 6 m group versus 51 (IQR, 28–84) in the PD group ($P < 0.01$; **Fig. 1C**). To select the optimal biomarker cut-off, we performed an ROC analysis. The area under the ROC curve (AUC) was 0.79 (95% CI, 0.61–0.98), demonstrating a good discriminatory ability of the biomarker (**Fig. 1D**). As cut-offs reaching the intended sensitivity and NPV ≥90% had a very low specificity (10%), we decided to select the cut-off matching the highest sensitivity as well as a specificity of at least 50% to reduce overtreatment. This resulted in a cut-off of 90 PD-1^T TILs per mm^2, reaching a sensitivity of 79% and a specificity of 83% (**Table 1**). The chosen cut-off had a high NPV of 89% as indicated by the large fraction of patients with PD in the group with less than 90 PD-1^T TILs per mm^2 (**Fig. 1E**).

To validate our findings, we next assessed the frequency of PD-1^T TILs per mm^2 in the validation set ($n = 77$). The median number of PD-1^T TILs per mm^2 was 150 (IQR, 89–231) in patients with DC 6 m versus 49 (IQR, 15–152) with PD ($P < 0.01$; **Fig. 1F**). The AUC of the ROC curve was 0.72 (95% CI, 0.60–0.84), indicating a similar

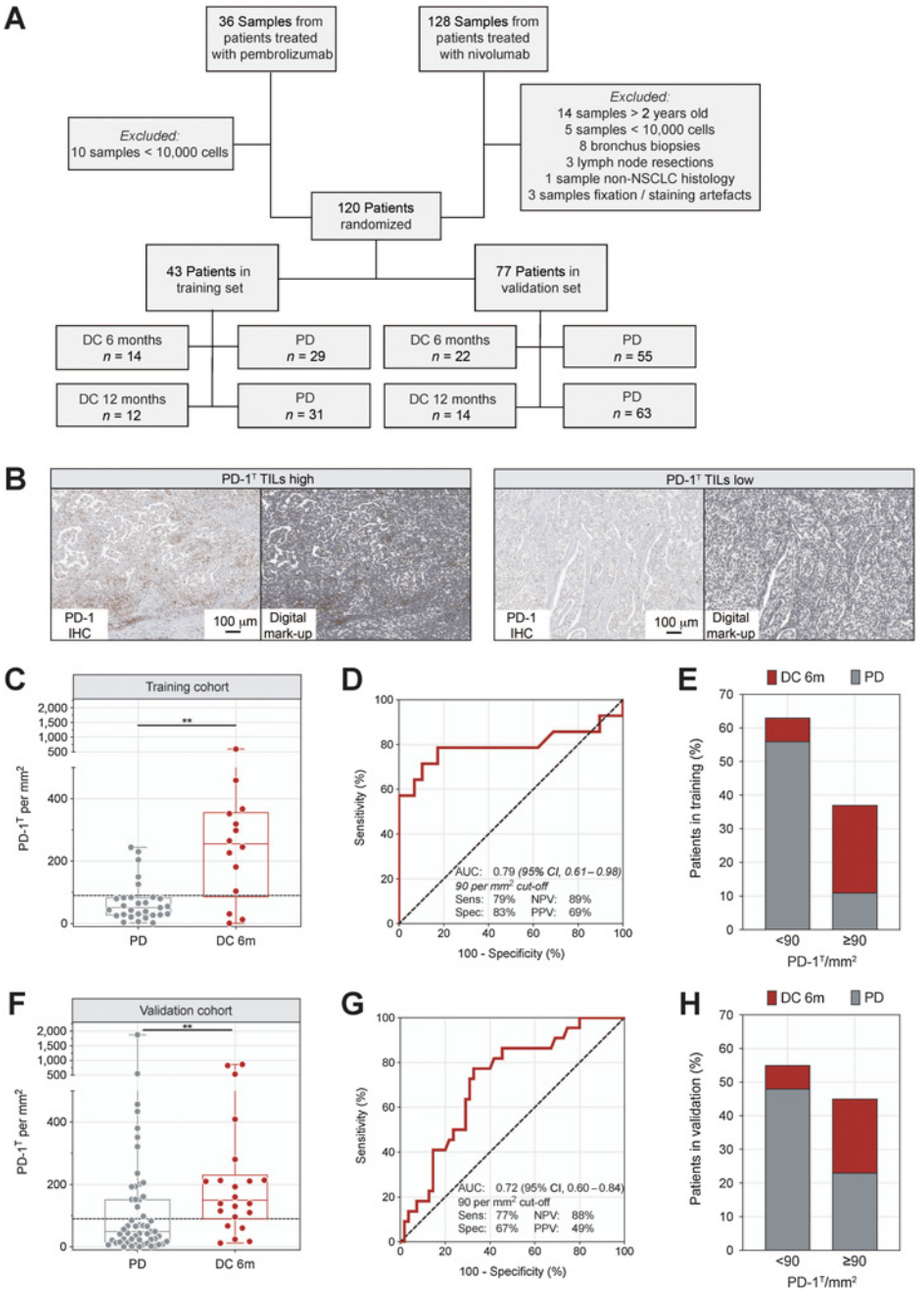

Figure 1.

PD-1^T TILs as biomarker for clinical outcome to PD-1 blockade in NSCLC. **A,** Study design for analysis of PD-1^T TILs in pretreatment samples from two retrospective patient cohorts with stage IV NSCLC treated with PD-1 blockade. The training (n = 43) and validation (n = 77) sets consisted each of 30% of patients with disease control at 6 months (DC 6 m) of treatment. Researchers were blinded for clinical outcome. **B,** Representative PD-1 IHC and digital mark-ups showing PD-1^T TILs (brown) in a PD-1^T TIL high and PD-1^T TIL low tumor sample, respectively. **C,** PD-1^T TILs per mm^2 in pretreatment samples from patients with DC 6 m (n = 14) and PD (n = 29) in the training set (n = 43). Dashed line indicates a cut-off of 90 PD-1^T TILs per mm^2. Medians, interquartile ranges and minimum/maximum shown in boxplots, **, $P < 0.01$ by Mann–Whitney U test. **D,** ROC curve for predictive value of PD-1^T TILs for DC 6 m (AUC, 0.79; 95% CI, 0.61–0.98) in the training set (n = 43). **E,** Percentage of patients with PD-1^T high (≥90 per mm^2; n = 16) and PD-1^T low (<90 per mm^2; n = 27) pretreatment samples showing DC 6 m or PD. **F** and **G,** Same plots as shown in **C** and **D** for patients with DC 6 m (n = 22) and PD (n = 55) in the validation set, **$P < 0.01$ (AUC, 0.72; 95% CI, 0.60–0.84). **H,** Same plot as shown in **E** for patients with PD-1^T high (n = 35) and PD-1^T low (n = 42) pretreatment samples in the validation set (n = 77).

performance as in the training set (**Fig. 1G**). The biomarker also reached a comparable sensitivity (77%) and NPV (88%), but somewhat lower specificity (67%; **Table 1**). Importantly, we still observed a substantial enrichment of nonresponding patients in the PD-1^T low group (**Fig. 1H**).

Assessment of secondary endpoints: DC at 12 months and survival

Since approximately 60% to 70% of patients treated in second-line with PD-(L)1 blockade progress within 6 months, and an additional 10% to 20% progress within 12 months (3, 5, 6), we also assessed the value of PD-1^T TILs to predict DC at 12 months (DC 12m). Two patients in the training and eight patients in the validation set experienced disease progression between 6 and 12 months, and were therefore included the PD group in this analysis (**Fig. 1A**). Median PD-1^T TIL numbers were comparable with the DC 6 m analysis [training set, DC 12m: 282 (IQR, 192–363), PD 44 (IQR, 27–83), $P < 0.0001$; validation set, DC 12m: 202 (IQR, 114–312), PD: 49 (IQR, 17–160), $P < 0.001$; Supplementary Fig. S2A and S2B]. Using the same cut-off of 90 PD-1^T TILs per mm^2 the ROC curve yielded a high AUC in both data sets of 0.89 (95% CI, 0.73–1.00; **Fig. 2A**, training set) and 0.78 (95% CI, 0.68–0.88; **Fig. 2B**, validation set). Importantly, in the DC 12 m analysis our predefined cut-off reached the intended criteria with a sensitivity of 92% and an NPV of 96% in the training set, and of 93% and 98%, respectively, in the validation set. In both cohorts, a specificity of >50% was maintained (84% in the training set, 65% in

Table 1. Predictive accuracy of PD-1^T TILs and PD-L1, summary of training and validation results.

	Clinical outcome	Biomarker	AUC	Cut-off	Sensitivity	Specificity	NPV	PPV
Training ($n = 43$)	DC 6 m	PD-1^T TILs per mm^2	0.7995% CI, 0.61–0.98	<90 vs. ≥90	79%	83%	89%	69%
Validation ($n = 77$)	DC 6 m	PD-1^T TILs per mm^2	0.7295% CI, 0.60–0.84	<90 vs. ≥90	77%	67%	88%	49%
		% PD-L1 TPS	0.5895% CI, 0.43–0.74	<1 vs. ≥1	41%	67%	74%	33%
				<50 vs. ≥50	23%	95%	75%	63%
Training ($n = 43$)	DC 12 m	PD-1^T TILs per mm^2	0.8995% CI, 0.73–1.00	<90 vs. ≥90	92%	84%	96%	69%
Validation ($n = 77$)	DC 12 m	PD-1^T TILs per mm^2	0.7895% CI, 0.68–0.88	<90 vs. ≥90	93%	65%	98%	37%
		% PD-L1 TPS	0.6895% CI, 0.51–0.86	<1 vs. ≥1	57%	70%	88%	30%
				<50 vs. ≥50	29%	94%	86%	50%

the validation set; **Table 1**). Notably, in both data sets only 1/43 (2%) and 1/77 (1%) samples from patients with DC 12 m showed a low frequency of PD-1^T TILs <90 per mm^2, suggesting a reliable identification of a patient group with no long-term benefit from PD-1 blockade (**Fig. 2C** and **D**).

As additional secondary endpoints, we assessed progression-free survival (PFS) and overall survival (OS) for patients with more or less than 90 PD-1^T TILs per mm^2. Because this cut-off was trained for prediction of DC at 6 months, PFS was significantly longer in PD-1^T high patients in the training set (HR, 0.30; 95% CI, 0.16–0.58; $P < 0.001$; Supplementary Fig. S2C). Notably, PFS was also significantly increased in the validation set (HR, 0.39; 95% CI, 0.24–0.63; $P < 0.0001$) in PD-1^T high patients (**Fig. 2E**). Likewise, OS was significantly longer in both the training (HR, 0.27; 95% CI, 0.14–0.53; $P < 0.0001$; Supplementary Fig. S2D) and validation set (HR, 0.46; 95% CI, 0.28–0.76; $P < 0.01$; **Fig. 2F**).

Differences between lesion-specific and overall response

Although the presence of <90 PD-1^T TILs per mm^2 was strongly associated with lack of benefit to PD-1 blockade, the PD-1^T high group was more heterogeneous with 27 of 51 patients showing progressive disease within 12 months. It is known that progression can occur heterogeneously across metastases upon PD-1 blockade (24). In addition, response assessment by RECIST criteria is based on the change in the sum of target lesion(s) and the development of new lesions (25). Thus, patients can be classified as PD based on the progression of some lesions whereas other lesions are stable or regress. To explore whether such mixed responses occur in PD-1^T high patients with PD, we assessed responses to PD-1 blockade in a lesion-specific manner. To this end, the percent increase or decrease in diameter of the biopsied lesion during treatment was determined using RECIST criteria. All patients showing PD within 12 months, and with at least two CT response assessments in which the biopsied lesion could be measured, were included in this analysis. In total, 11 PD patients in the PD-1^T high group and 14 PD patients in the PD-1^T low group could be evaluated. Interestingly, we observed that only 27% (3/11) of the biopsied lesions in the PD-1^T high group showed confirmed progression of the biopsied lesion (defined as ≥20% growth compared with smallest diameter during treatment) compared to 71% (10/14) in the PD-1^T low group. This indicates that the PD-1^T TIL biomarker correlates better with lesion-specific response than with overall radiologic response according to RECIST, and that this could account for at least part of the PD-1^T high patients with PD (**Fig. 3A**). For comparison, we also performed the same analysis in patients with DC at 12 months and found that 92% of the evaluable lesions (11/12) in the PD-1^T high group showed a durable response after a follow-up of 12 months.

Influence of patient and tissue sample characteristics on predictive potential

Several factors including tissue and patient characteristics or prior therapy can impact the predictive performance of biomarkers, as has for instance been shown for PD-L1 (26–28). We therefore explored whether clinicopathologic characteristics, intratumoral heterogeneity, sample type, sampling site, or the time of sampling influence the predictive performance of PD-1^T TILs as a biomarker. First, we examined a potential relationship of PD-1^T TILs with clinicopathologic characteristics. No significant differences were however observed between the <90 and ≥90 per mm^2 groups (Supplementary Tables S5 and S6). As heterogeneity of PD-L1 expression within lesions has been found to limit the predictive performance of this marker, we next assessed the heterogeneity of PD-1^T TILs in five resection samples of which two were PD-1^T low (<90 per mm^2) and three were PD-1^T high (≥90 per mm^2). We randomly selected 10 intratumoral areas of 1 mm^2 per sample and quantified PD-1^T TILs in each area (**Fig. 3B** and **C**). Although PD-1^T TIL frequencies varied within a sample, the vast majority of areas reflected the overall score of the sample as either PD-1^T TILs high or low. Thus, although PD-1^T TILs showed some intratumoral heterogeneity, the overall distribution could be captured by assessing a relatively small area of the tumor.

Next, we compared the potential of PD-1^T TILs to predict DC at 6 and 12 months in samples derived from either tumor resections or biopsies. Performing ROC analysis, we observed that the AUC for resected samples was higher than for biopsy samples though without reaching significance, in line with the notion that biopsies may be more prone to sampling errors (**Fig. 3D** and **E**). Next, we compared samples from primary and metastatic sites, which performed similarly with respect to prediction of treatment outcome (**Fig. 3D** and **E**). Finally, we compared samples that were taken either directly before start of anti-PD-1 treatment or before prior systemic treatment. Samples that were taken directly before anti-PD-1 treatment showed better predictive value, reaching significance in the DC 12 m subgroup [AUC 0.91 (CI, 0.82–0.99) versus 0.74 (CI, 0.61–0.88), $P = 0.04$] for samples taken prior to at least one other systemic treatment (**Fig. 3D** and **E**). In summary, these explorative analyses suggest that the predictive performance of PD-1^T TILs is even higher when assessed in a lesion-specific manner and in samples that were taken shortly before start of PD-1 blockade.

Comparison with PD-L1 as established biomarker

Pretreatment patient selection based on ≥50% or ≥1% tumor PD-L1 expression has been extensively studied, with contradictory results (1–3, 5, 6). However, improved outcomes in the KEYNOTE-024 study for patients with ≥50% PD-L1 expression have led to the implementation of PD-L1 testing in routine diagnostics (2). Therefore, we compared the predictive value of PD-1^T TILs to the PD-L1 TPS in

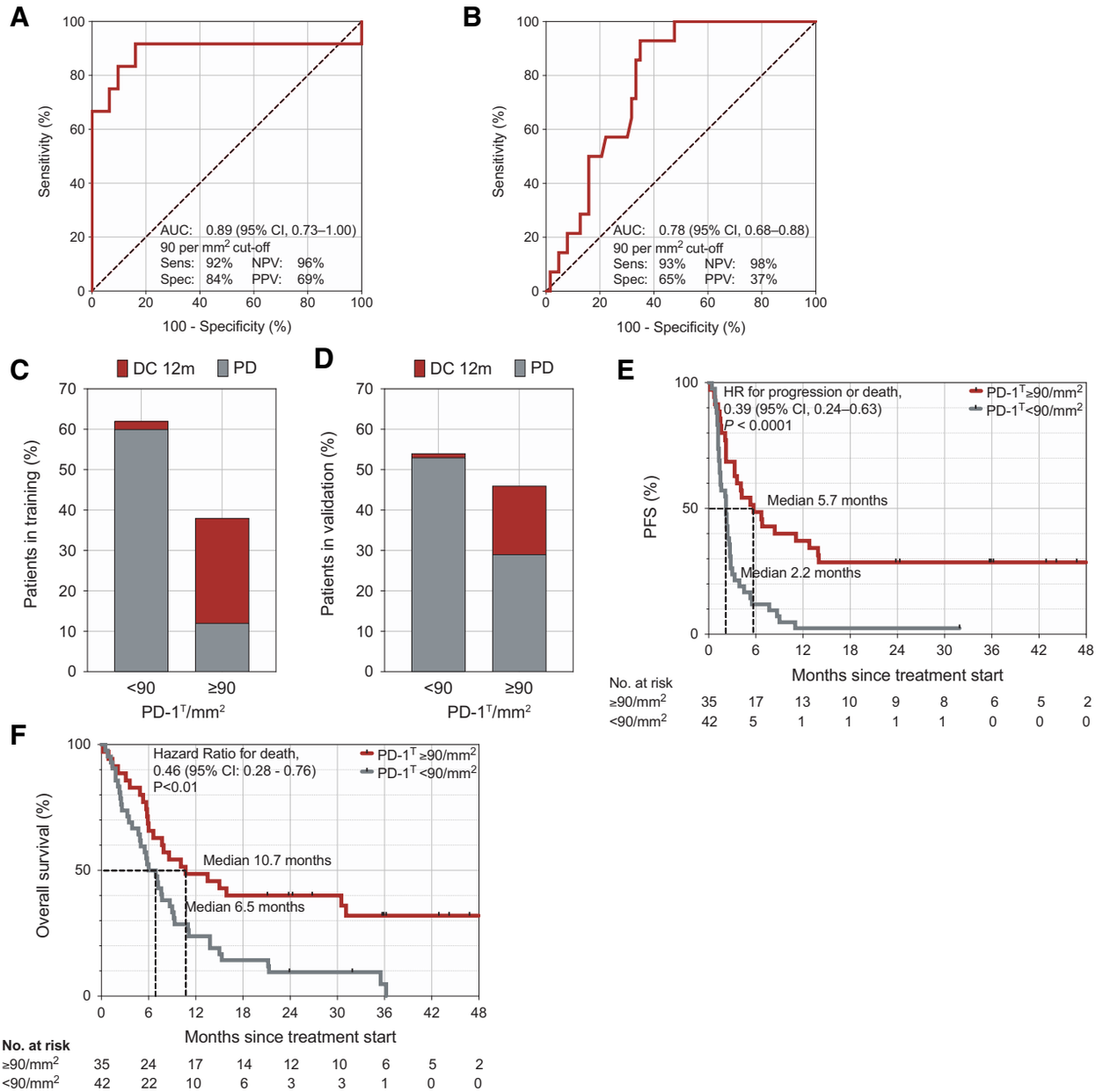

Figure 2.

PD-1^T TILs can effectively discriminate patients with long-term benefit from patients with progressive disease. **A,** ROC curve for predictive value of PD-1^T TILs for DC 12 m in the training set ($n = 43$; AUC, 0.89; 95% CI, 0.73–1.00) and (**B**) in the validation set ($n = 77$; AUC, 0.78; 95% CI, 0.68–0.88). **C,** Percentage of patients with PD-1^T high (≥90 per mm^2; $n = 16$) and PD-1^T low (<90 per mm^2; $n = 27$) pretreatment samples in the training set showing DC 12 m or PD. **D,** Same plot as in **C** for PD-1^T high ($n = 35$) and PD-1^T low ($n = 42$) pretreatment samples in the validation set ($n = 77$). **E,** PFS of patients with PD-1^T high versus PD-1^T low pretreatment samples (median, 5.7 months vs. 2.2 months; HR, 0.39; 95% CI, 0.24–0.63; ****, $P < 0.0001$) in the validation set ($n = 77$). **F,** OS (median, 10.7 months vs. 6.5 months; HR, 0.46; 95% CI, 0.28–0.76; **, $P < 0.01$). Tick marks represent data censored at the last time the patient was known to be alive and without disease progression or death. *P* value was determined by log-rank test.

the validation set (**Fig. 4A**). The fraction of disease control and PD for patients with tumors expressing ≥50%, 1% to 50%, or no PD-L1 are shown in Supplementary Fig. S3A (DC 6m) and **Fig. 4B** (DC 12m). ROC analysis of PD-L1 TPS to predict DC 6 m and DC 12 m showed a lower AUC compared with PD-1^T TILs (0.58; CI, 0.43–0.74 and 0.68; CI, 0.51–0.86; Supplementary Fig. S3B; **Fig. 4C**; **Table 1**). A cut-off using 50% PD-L1 TPS showed a substantially lower sensitivity (23%–29%) and lower NPV (75%–86%) compared with PD-1^T TILs (**Fig. 4D**; **Table 1**). Both sensitivity and NPV were slightly higher using a cut-off of 1% PD-L1 TPS (41–57% and 74–88%, respectively), but still below the values observed for PD-1^T TILs (**Fig. 4D**; **Table 1**). Also, additional cut-offs using 5% and 10% PD-L1 TPS, which have

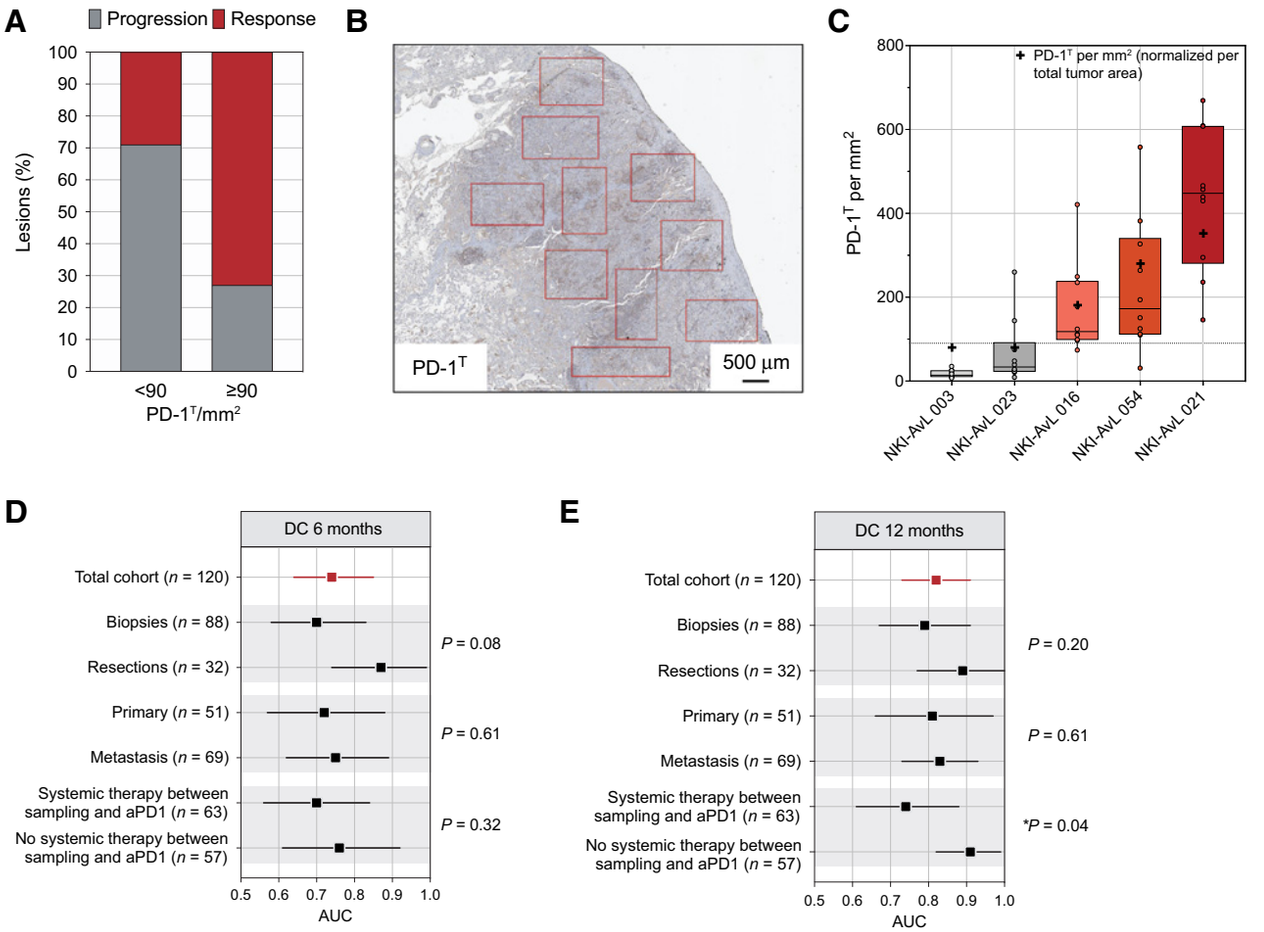

Figure 3.

Impact of lesion-specific response and tissue sample properties on the predictive accuracy of PD-1^T TILs. **A,** Percentage of responsive versus progressive lesions during treatment in the PD-1^T high (≥90 per mm^2; $n = 11$) and PD-1^T low (<90 per mm^2; $n = 14$) group of patients with PD within 12 months. A lesion was defined as progressive when ≥20% growth was seen compared with the smallest diameter during treatment. **B,** Example of a PD-1^T high IHC staining with 10 individually annotated tumor areas of 1 mm^2. **C,** Quantification of PD-1^T TILs per each mm^2 area in five resection specimens. Each dot indicates an individual measurement. Two tumors are PD-1^T low (gray shades), and three tumors are PD-1^T high (red shades). The cross indicates PD-1^T TILs per mm^2 normalized per total tumor area. **D** and **E,** The predictive value of PD-1^T TILs in the total cohort and different subgroups. Each comparison is marked in a gray square. Shown is the AUC for DC 6 m (**D**) and 12 m (**E**) with 95% CI interval. *P* value was determined by one-sided permutation test.

previously been evaluated as biomarker cut-offs for treatment with nivolumab (3, 6), did not improve prediction compared with PD-1^T TILs (Supplementary Table S7). Notably, when the predictive performance of PD-L1 TPS was assessed in the different subgroups using the full dataset as done for PD-1^T TILs, we observed a similar trend towards a higher AUC in tumor resections and in samples taken directly before start of PD-1 blockade. However, even after adjusting for these potential confounders, PD-1^T TILs remained superior to PD-L1 TPS in predicting clinical benefit (Supplementary Fig. S3C and S3D).

Next, we evaluated PFS and OS for PD-L1 TPS ≥50% and ≥1% in the validation set. Similar to reports from previous trials (1, 2, 8), PD-L1 TPS ≥50% enriched for patients demonstrating improved PFS and OS (HR, 0.36; median PFS, 30.3 months vs. 2.4 months, and HR, 0.40; median OS, 32.2 months vs. 7.2 months), of which only PFS reached significance. However, this finding was based on only 8 patients in the PD-L1 ≥50% subgroup and may therefore been prone to sample size error (**Fig. 4E** and **F**). Patients with PD-L1 TPS ≥1% showed slightly better PFS and OS in the validation set though without reaching significance which is comparable with other studies (refs. 3, 5, 6; **Fig. 4G** and **H**).

We noticed that the fraction of patients with PD-L1 TPS <1% observed here was higher than in previous studies (60% as compared with app. 30%), which could be caused by our more stringent scoring method using CD68 staining to avoid false-positive PD-L1 levels. Therefore, we assessed whether the combination of PD-L1 TPS and PD-L1 expression on immune cells (PD-L1 IC) could improve prediction. The correlation of PD-L1 TPS and PD-L1 IC was low (Supplementary Fig. S4A and S4B). Combining PD-L1 TPS at either 50% or 1% cut-off and PD-L1 IC≥2 indeed improved predictive accuracy, that is, compared with PD-L1 TPS >50%, by reaching a similar sensitivity as PD-1^T TILs but still substantially lower specificity (Supplementary Fig. S4C and S4D; Supplementary Table S8).

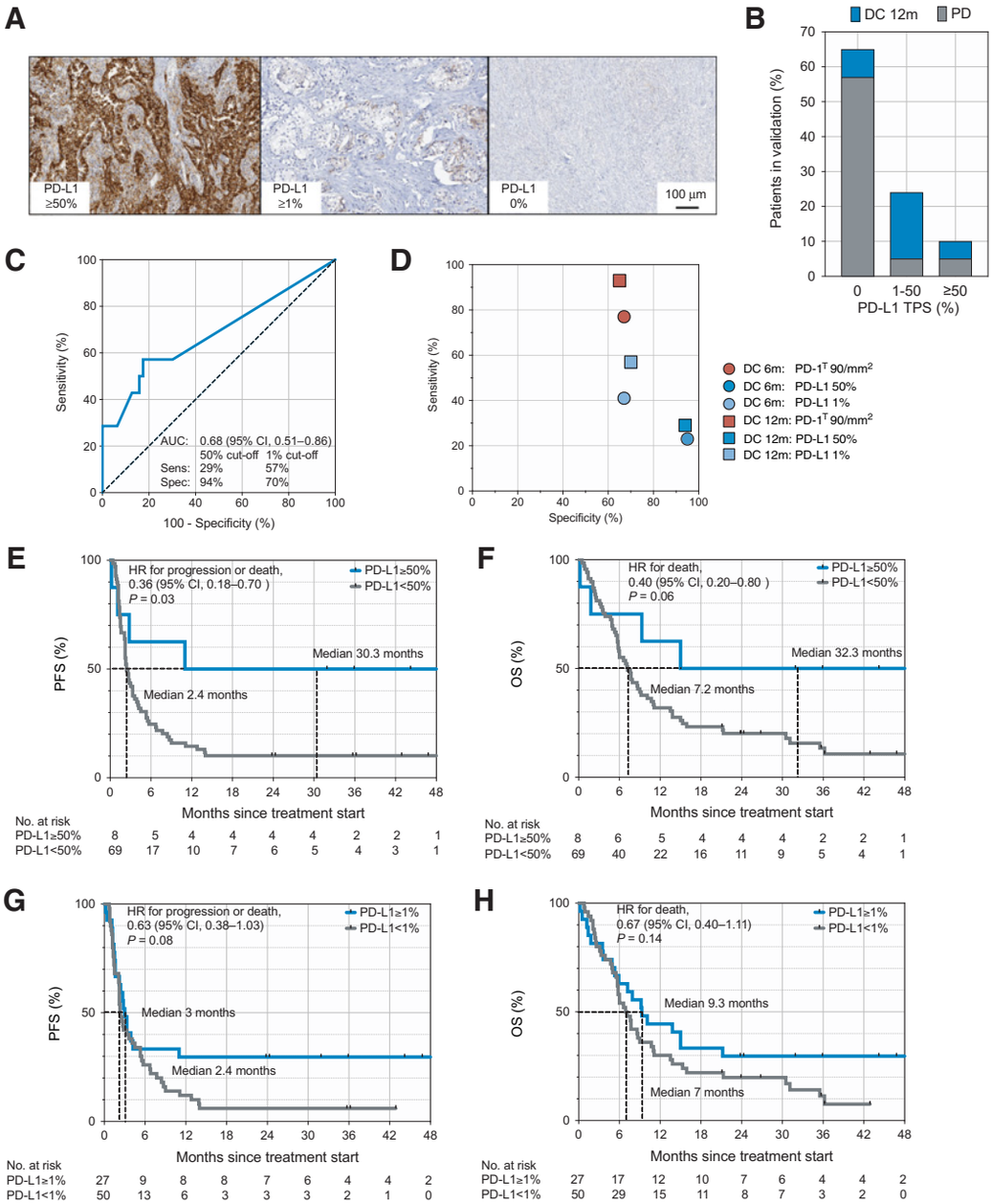

Figure 4.

Association of PD-L1 with long-term benefit and survival compared with PD-1^T TILs. **A,** IHC analysis of PD-L1. Example of NSCLC tumors with ≥50%, ≥1%, and 0% PD-L1 expression (PD-L1 TPS), respectively. **B,** Percentage of patients with ≥50% ($n = 8$), 1%–50% ($n = 19$), and 0% PD-L1 TPS ($n = 50$) in pretreatment samples showing DC 12 m or PD in the validation set ($n = 77$). **C,** ROC curve for predictive value of PD-L1 TPS for DC 12 m (AUC, 0.68; 95% CI, 0.51–0.86) in the validation set ($n = 77$). **D,** Sensitivity and specificity of PD-L1 TPS 50% and 1% for DC 6 m and 12 m in comparison with PD-1^T 90 per mm^2 in the validation set ($n = 77$). **E,** PFS (HR, 0.36; 95% CI, 0.18–0.70; *, $P = 0.03$) and (**F**) OS (HR, 0.40; 95% CI, 0.20–0.80; $P = 0.06$) of patients with ≥50% versus <50% PD-L1 TPS in pretreatment samples in the validation set ($n = 77$). **G,** PFS (HR, 0.63; 95% CI, 0.38–1.03; $P = 0.08$) and (**H**) OS (HR, 0.67; 95% CI, 0.40–1.11; $P = 0.14$) of patients with ≥1% versus <1% PD-L1 TPS in pretreatment samples in the validation set ($n = 77$). Tick marks represent data censored at the last time the patient was known to be alive and without disease progression or death. *P* value was determined by log-rank test.

Previous studies have evaluated the combination of PD-L1 TPS with other biomarkers such as TMB or CD8 and CD4 T-cell infiltration to increase predictive accuracy (29–33). Therefore, we investigated whether the combination with PD-L1 TPS could further improve the predictive value of PD-1^T TILs. The correlation between PD-1^T TILs and PD-L1 TPS was low (Supplementary Fig. S5A). Combination of the two biomarkers naturally partitioned the patient population into four groups: (ii) PD-L1 low (<50% or <1%)+PD-1^T low, (ii) PD-L1 low+PD-1^T high, (iii) PD-L1 high (≥50% or ≥1%)+PD-1^T low, and (iv) PD-L1 high+PD-1^T high (Supplementary Fig. S5B). We observed

an enrichment of PD patients in the <50% PD-L1+PD-1^T low group (35/38 when assessing DC 6 m as clinical outcome, 38/38 for DC 12m). Patients with DC 6 m or 12 m were distributed over all four groups and three out of four groups, respectively (Supplementary Fig. S5C and S5D). A similar patient distribution was found for the combination with 1% PD-L1 TPS (Supplementary Fig. S5E and S5F). The predictive value of PD-1^T+PD-L1 at 50% cut-off was comparable with PD-1^T alone with a few percent increase in sensitivity at the cost of a slightly lower specificity (Supplementary Fig. S5G; Supplementary Table S9, for details on prediction model see Materials and Methods section). The sensitivity of PD-1^T+PD-L1 at 1% cut-off was similar to PD-1^T+PD-L1 50%, but the specificity of this combination was below 50% (Supplementary Fig. S5G; Supplementary Table S9). Thus, the predictive accuracy of PD-1^T alone is not increased by parallel quantification of PD-L1 levels.

PD-1^T TILs and TLS

TLS are immune cell aggregates that form in the context of chronic inflammation and have been described in many cancer types, including NSCLC (34–36). A number of recent studies have shown that TLS and B cells as one of their main cellular components are associated with response to ICB in melanoma, renal cell carcinoma, and sarcoma (14–16). Moreover, we previously showed in a small number of NSCLC samples that PD-1^T TILs appear to predominantly localize in TLS and constitutively secrete CXCL13, a chemoattractant that is crucial for the formation of TLS (10). For 91 of our pretreatment samples for which additional FFPE material was available, we assessed whether TLS and B cells were present. To this end, CD20/CD3 double IHC staining were performed to identify TLS by the presence of B-cell clusters and T-cell zones, as described previously (10, 16, 22, 23, 35). A CD3$^+$CD20$^+$ area was defined as TLS when its size was more than 60,000 μm^2 in the annotated tumor area, and as lymphoid aggregate (LA) when between 10,000 and 60,000 μm^2. To estimate the presence of B cells, we quantified the CD20-positive area per mm^2 (Supplementary Fig. S6A). This analysis revealed that TLS and TLS and/or LA (referred as TLS+LA) were present in 30/91 (33%) and 46/91 (51%) of tumors, respectively. B cells were found in 86/91 (95%) of tumors, suggesting that the presence of these cells does not always relate to TLS and LA (**Fig. 5A**). However, in most of the 40 samples without TLS+LA, CD20$^+$ area per mm^2 was low (Supplementary Fig. S6B). Next, we wanted to assess the localization of PD-1^T TILs in relation to TLS. To improve the accuracy of this analysis we focused on tumor resections ($n = 32$), for which the annotated TLS areas based on the CD20/CD3 double staining were copied to a consecutive slide stained for PD-1 to calculate the frequency of PD-1^T TILs inside TLS (**Fig. 5B** and **C**). The frequency of PD-1^T TILs per mm^2 was significantly higher inside TLS compared with tumor areas outside of TLS ($P < 0.0001$; **Fig. 5D**). We found similar results when performing the same analysis including TLS+LA (Supplementary Fig. S6C). Of note, samples without TLS or LA showed only very low frequencies of PD-1^T TILs (**Fig. 5D**; Supplementary Fig. S6C).

Next, we investigated whether the number of TLS was associated with clinical benefit to PD-1 blockade in NSCLC as has been shown in other tumor types (14–16). The ranges of TLS and TLS+LA per mm^2 are shown in **Fig. 5E** and Supplementary Fig. S6D (DC 12m) and Supplementary Fig. S6E and S6F (DC 6m). The AUC of TLS per mm^2 was 0.62 (CI, 0.47–0.76) for DC 12 m (**Fig. 5F**) and 0.62 (CI, 0.49–0.76) for DC 6 m (Supplementary Fig. S6G), respectively, indicating a lower predictive performance than PD-1^T TILs in the same sample set. Similar results were observed for TLS+LA per mm^2 and CD20$^+$ area per mm^2 (**Fig. 5F**; Supplementary Fig. S6G). As PD-1^T TILs predominantly localize in TLS, this observation would possibly be consistent with subtypes of TLS that differ in the number of PD-1^T TILs. To investigate this, we quantified the frequency of TLS and of PD-1^T TILs inside and outside TLS in PD-1^T high ($n = 13$) and PD-1^T low resected samples ($n = 10$). Resected samples with no TLS were excluded from this analysis ($n = 9$). We found that TLS numbers did not significantly differ between both groups (**Fig. 5G**). However, tumors in the PD-1^T high group had significantly higher numbers of PD-1^T TILs inside TLS (**Fig. 5H**). In addition, in PD-1^T high tumors also significantly more PD-1^T TILs were present outside of TLS (**Fig. 5I**) compared with the PD-1^T low group. Notably, PD-1^T lymphocytes were only sparsely present in the tumor parenchyma of PD-1^T low tumors (**Fig. 5I**). Altogether, these data suggest that in tumors responding to PD-1 blockade PD-1^T TILs not only infiltrate TLS in higher numbers, but also expand in the tumor parenchyma.

Discussion

mAbs that block the PD-1–PD-L1 axis have transformed the therapeutic arsenal of advanced stage NSCLC. Nevertheless, most patients still do not benefit from PD-1 blockade, while they are exposed to the risk of treatment-related toxicity. Because of this, there is an evident clinical need for predictive biomarkers that can help reduce overtreatment. On the basis of the rationale that the presence of tumor-reactive PD-1^T T lymphocytes is indicative of an ongoing antitumor response (10), we here assess the predictive value of PD-1^T TILs using an algorithm-based quantitative PD-1 IHC assay in FFPE tissue sections. We establish PD-1^T TILs as predictive marker in two independent advanced stage NSCLC cohorts treated with PD-1 blockade. Our data show that particularly low numbers of PD-1^T TILs accurately identify a patient group with no clinical benefit. Furthermore, high PD-1^T TIL infiltration was observed in >90% of patients with DC 12m. The high sensitivity and NPV of our biomarker of more than 90% with a specificity of more than 50% should thus allow to reduce overtreatment while minimizing undertreatment.

Interobserver variability in the assessment of biomarkers often affects their predictive value. Here, we report a reliable and automated method to perform digital quantification of PD-1^T TILs, based on an approach established in our earlier work (10). Previous studies have shown the advantage of digital quantification by improving accuracy and standardization of biomarkers (37–39). Although our method allows automated quantification of PD-1^T TILs, it still requires a substantial user interaction, for instance as tumor areas need to be manually annotated. Hence, for implementation into clinical practice further studies are required to assess methods that could improve standardization across centers, for instance using artificial intelligence (AI) solutions.

Another common hurdle for biomarker development is caused by tumor heterogeneity. Therefore, we aimed to understand whether heterogeneity in PD-1^T TILs occurs within and across lesions and whether the presence of PD-1^T TILs may thus be predictive for the capacity of locally residing T cells to mediate antitumor immunity upon PD-1 blockade. To this end, we first assessed responses at lesion-level in patients defined as clinical progressors by RECIST criteria. Importantly, we observed that only a minority of the assessed lesions in the PD-1^T high group progressed as compared with the PD-1^T low group, indicating a good correlation between the biomarker and lesion-specific response. To assess the impact of intratumoral heterogeneity, we quantified PD-1^T TILs in multiple randomly selected small tumor areas. Despite some level of variation, the vast majority of individual measurements allowed to correctly classify a sample as

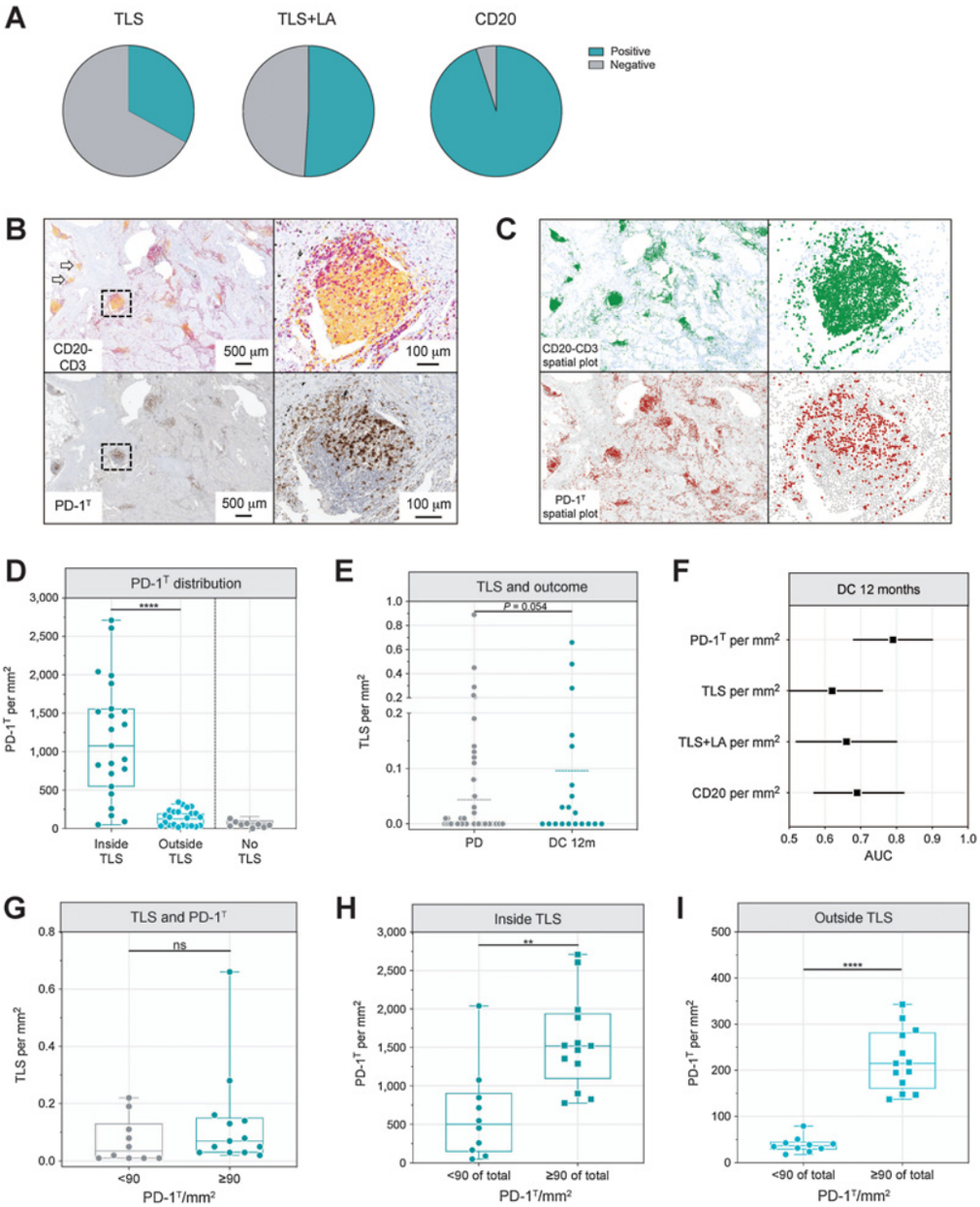

Figure 5.

PD-1^T high samples contain a higher density of PD-1^T TILs inside and outside TLS. **A,** Percentage of pretreatment samples containing TLS ($n = 30$), TLS and/or LA (referred as TLS+LA, $n = 46$), and CD20$^+$ B cells ($n = 86$) in the remaining cohort ($n = 91$). **B,** Top: Example of a CD20-CD3 IHC double staining, with the black square showing CD20$^+$ B cells (in yellow) and CD3$^+$ T cells (in purple) localizing in a TLS. Arrows indicate LA. Bottom: Example of a consecutive PD-1 IHC staining, with the black square showing PD-1^T TILs inside TLS. **C,** Digital markup showing the spatial distribution of CD20$^+$ B cells (in green) and CD3$^+$ T cells (in light blue), and digital markup of PD-1^T TILs (in red) and all other cells (in gray). **D,** PD-1^T TILs per mm^2 inside and outside TLS in resected samples ($n = 23$), and of total tumor area for resected samples with no TLS ($n = 9$). Medians, interquartile ranges, and minimum/maximum shown in boxplots (****, $P < 0.0001$ by Mann–Whitney test). **E,** TLS per mm^2 in pretreatment samples from patients with DC 12 m ($n = 20$) and PD ($n = 71$). Shown is the mean, $P = 0.054$ by Mann–Whitney test. **F,** The predictive value of PD-1^T TILs, TLS, TLS+LA, and CD20$^+$ area per mm^2 for DC 12 m ($n = 91$; note that this cohort is smaller due to the availability of FFPE material). Shown are AUCs with 95% CI interval. **G,** TLS per mm^2 in PD-1^T low (<90 per mm^2 of total tumor area; $n = 10$) and PD-1^T high (≥90 per mm^2 of total tumor area; $n = 13$) resected samples. Medians, interquartile ranges, and minimum/maximum shown in boxplots ($P = 0.18$ by Mann–Whitney test). **H,** PD-1^T TILs per mm^2 inside TLS and (**I**) outside TLS in PD-1^T low ($n = 10$) and PD-1^T high ($n = 13$) resected samples. Medians, interquartile ranges, and minimum/maximum shown in boxplots (**, $P < 0.01$; ****, $P < 0.0001$ by Mann–Whitney test).

PD-1^T high or low, suggesting that PD-1^T TIL infiltration can be reliably captured in a relatively small area of the tumor. An additional level of heterogeneity may result from, for instance, sample type, sampling site and/or the time between sampling and initiation of treatment. Thus, we explored the impact of these potential confounding factors on the predictive value of PD-1^T TILs as a biomarker. We observed a trend towards increased predictive accuracy of PD-1^T TILs when measured in tumor resections as compared with biopsies. In contrast, no difference in predictive value of the biomarker was found between primary tumors and metastases. Hong and colleagues previously observed differences in PD-L1 expression levels in distinct anatomical sites and showed that high PD-L1 was associated with better clinical outcome in lung and distant metastases, but not in lymph node biopsies (40). The low number of lymph node biopsies in our sample set precluded the separate analysis of lymph node and organ metastases, thus a possible differences in predictive value between these sample sites should be further explored in future work. Finally, we observed that samples taken immediately before start of PD-1 blockade were more accurate for prediction of clinical benefit, as shown by the higher AUC, reaching significance for prediction of DC 12m. Collectively, these data suggest that heterogeneity in PD-1^T TIL infiltration across and within lesions exist, and that PD-1^T TILs may therefore be reflective of the capacity of locally residing T cells to control tumor growth upon anti-PD-1. Hence, it will be important in future studies to address the mechanistic basis of this heterogeneity, such as differences in local antigen availability, HLA expression, or else.

A further aim of the study was to explore the association of PD-1^T TILs to other immune-related biomarkers, such as PD-L1 and TLS. PD-1^T TILs performed superior to PD-L1, because both 50% and 1% PD-L1 TPS showed substantially lower sensitivity and NPV. Predictive performance could be improved when PD-L1 TPS and PD-L1 IC were combined, but remained below that observed with PD-1^T TILs. Notably, whereas previous studies showed an additive value of PD-L1 to TMB (30, 33) and PD-L1 to CD8 (31, 32), the combination of either 50% or 1% PD-L1 TPS with PD-1^T TILs did not improve predictive accuracy. However, because the ≥50% PD-L1 group only comprised 10% of the samples in the validation set, which is different from previous reported percentages (1, 2), additional studies with subgroups that are more balanced or including PD-L1 IC should validate these findings. As another immune-related marker, TLS have recently been associated with response and survival benefit to ICB in multiple cancer types (13–16), and in previous work we observed that PD-1^T TILs predominantly localize in TLS (10). In this study, we found a lower predictive accuracy of TLS and LA compared with PD-1^T TILs. The observation that tumors with high and low PD-1^T TIL count do not show substantially different frequencies of TLS, but vary in the number of PD-1^T TILs within TLS can explain the difference in predictive value. Moreover, the increased infiltration of PD-1^T TILs in the tumor parenchyma in the PD-1^T high group suggests that not only expansion of PD-1^T TILs in TLS, but also their infiltration in the tumor may be required for an effective response upon PD-1 blockade treatment. Further studies are needed to provide a more in-depth characterization of TLS-associated and intratumoral PD-1^T TIL subsets and to investigate a potential role of TLS in the expansion of these cells.

Taken together, we here established PD-1^T TILs as a novel predictive biomarker for durable clinical benefit to PD-1 blockade in NSCLC. Importantly, the high NPV of the biomarker may allow for the reliable identification of those patients that are unlikely to benefit from PD1 blockade, thus providing a tool to reduce overtreatment.

Authors' Disclosures

W.S.M.E. Theelen reports grants from AstraZeneca and Sanofi outside the submitted work. V.H. Koelzer reports grants from Swiss National Science Foundation and other support from Indica Labs during the conduct of the study as well as grants from Roche outside the submitted work; in addition, V.H. Koelzer has a past patent application [Title: Methods for Predicting Treatment Outcome and/or for Selecting a Subject Suitable for Immune Checkpoint Therapy (NL2020422), Inventors: D.S. Thommen, V.H. Koelzer, T. Schumacher, K.D. Mertz, and A. Zippelius. Date of initial filing: February 18, 2018, by the University of Basel and the National Cancer Institute of the Netherlands (NKI).] This application was not advanced to nationalization at PCT stage. A. Zippelius reports grants from Secarna and Roche outside the submitted work. J.B.A.G. Haanen reports grants from Angen and Asher Bio; personal fees from Achilles Tx, GSK, Iotech Bio, Ipsen, Instil Bio, Molecular Partners, Neogene T, Roche, Sanofi, Scenic, and T-Knife x; and grants and personal fees from BioNTech, BMS, MSD, and Novartis outside the submitted work; in addition, J.B.A.G. Haanen reports ownership in Neogene Tx stock. T.N. Schumacher reports personal fees from Third Rock Ventures, Allogene Therapeutics, Asher Bio, Celsius, Merus, Neogene Therapeutics, and Scenic Bio outside the submitted work. G.A. Meijer reports other support from Exact Sciences, Hartwig Medical Foundation, and Sysmex and grants from CZ Health Insurance outside the submitted work; in addition, G.A. Meijer is a co-founder and board member (CSO) of CRCbioscreen BV and has research collaborations with Sysmex, Sentinel Ch. SpA, Personal Genome Diagnostics (PGDX), DELFi, and Hartwig Medical Foundation. K. Monkhorst reports grants from AstraZeneca; personal fees from MSD, Roche, AstraZeneca, Benecke, Pfizer, BMS, AbbVie, Diaceutics, Lilly, Boehringer Ingelheim, and Merck; and nonfinancial support from Roche, Takeda, Pfizer, PGDx, and Defli outside the submitted work. D.S. Thommen reports grants from Bristol Myers Squibb and Asher Bio outside the submitted work. No disclosures were reported by the other authors.

Authors' Contributions

K. Hummelink: Conceptualization, data curation, formal analysis, supervision, funding acquisition, validation, investigation, visualization, methodology, writing–original draft, project administration, writing–review and editing. **V. van der Noort:** Data curation, formal analysis, validation, methodology, writing–review and editing. **M. Muller:** Resources, data curation, formal analysis, validation, methodology. **R.D. Schouten:** Resources, data curation. **F. Lalezari:** Data curation, formal analysis, investigation, methodology. **D. Peters:** Methodology. **W.S.M.E. Theelen:** Resources, data curation, formal analysis, methodology, writing–review and editing. **V.H. Koelzer:** Validation, methodology, writing–review and editing. **K.D. Mertz:** Validation, methodology, writing-review and editing. **A. Zippelius:** Validation, methodology, writing–review and editing. **M.M. van den Heuvel:** Resources, investigation. **A. Broeks:** Resources, data curation, methodology. **J.B.A.G. Haanen:** Conceptualization, methodology, writing–review and editing. **T.N. Schumacher:** Conceptualization, methodology, writing–review and editing. **G.A. Meijer:** Conceptualization, supervision, investigation, methodology, writing–review and editing. **E.F. Smit:** Conceptualization, supervision, investigation, methodology, writing–review and editing. **K. Monkhorst:** Conceptualization, formal analysis, supervision, validation, investigation, methodology, writing–original draft, project administration, writing–review and editing. **D.S. Thommen:** Conceptualization, formal analysis, supervision, funding acquisition, validation, investigation, visualization, methodology, writing-original draft, project administration, writing–review and editing.

Acknowledgments

We thank Symbiant, Admiraal de Ruyter hospital, Medical Center Haaglanden, Pathology Laboratory Dordrecht, Alrijne hospital, Spaarne Gasthuis, Onze Lieve Vrouwe Gasthuis, Meander Medical Center, Tergooi Medical Center, Isala clinics, Gelre Hospitals, Medical Center Slotervaart, Pathan BV, Amsterdam University Medical Center, Haga hospital, Reinier de Graaf hospital, Treant Zorggroep, Rijnstate hospital, Pathology laboratory Oost-Nederland, Erasmus University Medical Center, Leiden University Medical Center, Radboud University Medical Center, Stichting PAMM, Groene Hart hospital, University Medical Center Maastricht, Pathology Friesland, University Medical Center Utrecht, Pathology laboratory Midden-Brabant, and Laurentius hospital for providing patient material included in the study. We would like to thank the NKI-AVL Core Facility Molecular Pathology and Biobanking for supplying biobank and laboratory support and the flow cytometry facility for assistance with sorting and flow cytometric analyses. We thank Marjolijn Mertz for help with digital image analysis and members of the Thommen lab for helpful

discussions. This work was supported by a KWF Young investigator grant (no. 12046) to D.S. Thommen.

The publication costs of this article were defrayed in part by the payment of publication fees. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

Note

Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Received March 28, 2022; revised May 31, 2022; accepted July 15, 2022; published first July 19, 2022.

References

1. 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.
2. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Pembrolizumab versus chemotherapy for PD-L1–positive non–small-cell lung cancer. N Engl J Med 2016;375:1823–33.
3. Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, et al. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. N Engl J Med 2015;373:1627–39.
4. Postow MA, Sidlow R, Hellmann MD. Immune-related adverse events associated with immune checkpoint blockade. N Engl J Med 2018;378:158–68.
5. Rittmeyer A, Barlesi F, Waterkamp D, Park K, Ciardiello F, Von Pawel J, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. Lancet 2017;389:255–65.
6. Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WEE, Poddubskaya E, et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. N Engl J Med 2015;373:123–35.
7. Camidge DR, Doebele RC, Kerr KM, Kerr KM. Comparing and contrasting predictive biomarkers for immunotherapy and targeted therapy of NSCLC. Nat Rev Clin Oncol 2019;16:341–55.
8. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Updated analysis of KEYNOTE-024: Pembrolizumab versus platinum-based chemotherapy for advanced non–small-cell lung cancer with PD-L1 tumor proportion score of 50% or greater. J Clin Oncol 2019;37:537–46.
9. Van Der Leun AM, Thommen DS, Schumacher TN. CD8+ T cell states in human cancer: insights from single-cell analysis. Nat Rev Cancer 2020;20:218–32.
10. Thommen DS, Koelzer VH, Herzig P, Roller A, Trefny M, Dimeloe S, et al. A transcriptionally and functionally distinct pd-1 + cd8 + t cell pool with predictive potential in non-small-cell lung cancer treated with pd-1 blockade. Nat Med 2018;24:999–1004.
11. Oliveira G, Stromhaug K, Klaeger S, Kula T, Frederick DT, Le PM, et al. Phenotype, specificity and avidity of antitumour CD8+ T cells in melanoma. Nature 2021;596:119–25.
12. Caushi JX, Zhang J, Ji Z, Vaghasia A, Zhang B, Hsiue EH-C, et al. Transcriptional programs of neoantigen-specific TIL in anti-PD-1-treated lung cancers. Nature 2021;596:126–32.
13. Voabil P, De Bruijn M, Roelofsen LM, Hendriks SH, Brokamp S, Van Den Braber M, et al. An *ex vivo* tumor fragment platform to dissect response to PD-1 blockade in cancer. Nat Med 2021;27:1250–61.
14. Cabrita R, Lauss M, Sanna A, Donia M, Skaarup Larsen M, Mitra S, et al. Tertiary lymphoid structures improve immunotherapy and survival in melanoma. Nature 2020;580:E1.
15. Helmink BA, Reddy SM, Gao J, Zhang S, Basar R, Thakur R, et al. B cells and tertiary lymphoid structures promote immunotherapy response. Nature 2020;577:549–55.
16. Petitprez F, De Reyniès AL, Keung EZ, Chen TW-W, Sun C-M, Calderaro J, et al. B cells are associated with survival and immunotherapy response in sarcoma. Nature 2020;577:556–60.
17. Theelen WSME, Peulen HMU, Lalezari F, Van Der Noort V, De Vries JF, Aerts JGJV, et al. Effect of pembrolizumab after stereotactic body radiotherapy vs pembrolizumab alone on tumor response in patients with advanced non-small cell lung cancer: results of the PEMBRO-RT Phase 2 Randomized Clinical Trial. JAMA Oncol 2019;5:1276–82.
18. Batenchuk C, Albitar M, Zerba K, Sudarsanam S, Chizhevsky V, Jin C, et al. A real-world, comparative study of FDA-approved diagnostic assays pd-l1 IHC 28-8 and 22c3 in lung cancer and other malignancies. J Clin Pathol 2018;71:1078–83.
19. Krigsfeld GS, Prince EA, Pratt J, Chizhevsky V, William Ragheb J, Novotny J Jr, et al. Analysis of real-world PD-L1 IHC 28-8 and 22C3 pharmDx assay utilisation, turnaround times and analytical concordance across multiple tumour types. J Clin Pathol 2020;73:656–64.
20. Herbst RS, Soria J-C, Kowanetz M, Fine GD, Hamid O, Gordon MS, et al. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. Nature 2014;515:563–7.
21. Fehrenbacher L, Spira A, Ballinger M, Kowanetz M, Vansteenkiste J, Mazieres J, et al. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): A multicentre, open-label, phase 2 randomised controlled trial. Lancet 2016;387:1837–46.
22. Gu-Trantien C, Garaud S, Migliori E, Solinas C, Lodewyckx JN, Willard-Gallo K. Quantifying tertiary lymphoid structure-associated genes in formalin-fixed paraffin-embedded breast cancer tissues. Methods Mol Biol 2018;1845:139–57.
23. Buisseret L, Desmedt C, Garaud S, Fornili M, Wang X, Van Den Eyden G, et al. Reliability of tumor-infiltrating lymphocyte and tertiary lymphoid structure assessment in human breast cancer. Mod Pathol 2017;30:1204–12.
24. Osorio JC, Arbour KC, Le DT, Durham JN, Plodkowski AJ, Halpenny DF, et al. Lesion-level response dynamics to programmed cell death protein (PD-1) blockade. J Clin Oncol 2019;37:3546–55.
25. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009;45:228–47.
26. Boothman A-M, Scott M, Ratcliffe M, Whiteley J, Dennis PA, Wadsworth C, et al. Impact of patient characteristics, prior therapy, and sample type on tumor cell programmed cell death ligand 1 expression in patients with advanced NSCLC Screened for the ATLANTIC Study. J Thorac Oncol 2019;14:1390–9.
27. Ilie M, Long-Mira E, Bence C, Butori C, Lassalle S, Bouhlel L, et al. Comparative study of the PD-L1 status between surgically resected specimens and matched biopsies of NSCLC patients reveal major discordances: a potential issue for anti-PD-L1 therapeutic strategies. Ann Oncol 2016;27:147–53.
28. Gniadek TJ, Li QK, Tully E, Chatterjee S, Nimmagadda S, Gabrielson E. Heterogeneous expression of PD-L1 in pulmonary squamous cell carcinoma and adenocarcinoma: Implications for assessment by small biopsy. Mod Pathol 2017;30:530–8.
29. Hu-Lieskovan S, Lisberg A, Zaretsky JM, Grogan TR, Rizvi H, Wells DK, et al. Tumor characteristics associated with benefit from pembrolizumab in advanced non–small cell lung cancer. Clin Cancer Res 2019;25:5061–8.
30. Hellmann MD, Nathanson T, Rizvi H, Creelan BC, Sanchez-Vega F, Ahuja A, et al. Genomic features of response to combination immunotherapy in patients with advanced Non-small-cell lung cancer. Cancer Cell 2018;33:843–52.
31. Fumet J-D, Richard C, Ledys F, Klopfenstein Q, Joubert P, Routy B, et al. Prognostic and predictive role of CD8 and PD-L1 determination in lung tumor tissue of patients under anti-PD-1 therapy. Br J Cancer 2018;119:950–60.
32. Althammer S, Tan TH, Spitzmüller A, Rognoni L, Wiestler T, Herz T, et al. Automated image analysis of NSCLC biopsies to predict response to anti-PD-L1 therapy. J Immunother Cancer 2019;7:1–12.
33. Rizvi H, Sanchez-Vega F, La K, Chatila W, Jonsson P, Halpenny D, et al. Molecular determinants of response to anti-programmed cell death (PD)-1 and anti-programmed death-ligand 1 (PD-L1) blockade in patients with non-small-cell lung cancer profiled with targeted next-generation sequencing. J Clin Oncol 2018;36:633–41.
34. Sautès-Fridman C, Lawand M, Giraldo NA, Kaplon H©LN, Germain C, Fridman WH, et al. Tertiary lymphoid structures in cancers: prognostic value, regulation, and manipulation for therapeutic intervention. Front Immunol 2016;7:1–11.
35. Sautès-Fridman C, Petitprez F, Calderaro J, Fridman WH. Tertiary lymphoid structures in the era of cancer immunotherapy. Nat Rev Cancer 2019;19:307–25.

36. Schumacher TN, Thommen DS. Tertiary lymphoid structures in cancer. Science 2022;375:eabf9419.
37. Amgad M, Stovgaard ES, Balslev E, Thagaard J, Chen W, Dudgeon S, et al. Report on computational assessment of tumor infiltrating lymphocytes from the International Immuno-Oncology Biomarker Working Group. NPJ Breast Cancer 2020;6:16.
38. Steele KE, Tan TH, Korn R, Dacosta K, Brown C, Kuziora M, et al. Measuring multiple parameters of CD8+ tumor-infiltrating lymphocytes in human cancers by image analysis. J Immunother Cancer 2018; 6:1–14.
39. 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.
40. Hong L, Negrao MV, Dibaj SS, Chen R, Reuben A, Bohac JM, et al. Programmed death-ligand 1 heterogeneity and its impact on benefit from immune checkpoint inhibitors in NSCLC. J Thorac Oncol 2020;15:1449–59.