Head-to-head comparison of composite and individual biomarkers to predict clinical benefit to PD-1 blockade in Non-Small Cell Lung Cancer

Karlijn Hummelink[1,2], Vincent van der Noort[3], Mirte Muller[2], Robert D. Schouten[2], Michel M. van den Heuvel[2,4], Daniela S. Thommen[5], Egbert F. Smit[2,6], Gerrit A. Meijer[1] and Kim Monkhorst[1]

[1]Department of Pathology, Division of Diagnostic Oncology, [2]Department of Thoracic Oncology, Division of Medical Oncology, [3]Department of Biometrics, Netherlands Cancer Institute, Amsterdam, The Netherlands, [4]Current address: Department of Pulmonary Diseases, Radboud University Medical Center, Nijmegen, The Netherlands, [5]Division of Molecular Oncology and Immunology, Netherlands Cancer Institute, Amsterdam, The Netherlands, [6]Current address: Department of Pulmonary Diseases, Leiden University Medical enter, Leiden, The Netherlands.

Corresponding authors:

Email: k.monkhorst@nki.nl (KM) and k.hummelink@nki.nl (KH)

Abstract

Background

Treatment with PD-(L)1 blocking agents has demonstrated durable efficacy in advanced NSCLC, but only in a minority of patients. Multiple biomarkers for predicting treatment benefit have been investigated, but their combined performance has not been extensively studied. Here, we assess the combined predictive performance of multiple biomarkers in a series of NSCLC patients treated with nivolumab.

Methods

Pretreatment samples from 135 patients treated with nivolumab were used to assess the predictive performance of CD8 tumor-infiltrating lymphocytes (TILs), intratumoral (IT) localization of CD8 TILs, PD-1 high expressing TILs ($PD1^T$ TILs), CD3 TILs, CD20 B-cells, tertiary lymphoid structures (TLS), PD-L1 tumor proportion score (TPS) and the Tumor Inflammation score (TIS). Patients were assigned to a training (n=55) and validation set (n=80). The primary outcome measure was Disease Control at 6 months (DC 6m) and the secondary outcome measure was DC at 12 months (DC 12m).

Results

In the validation cohort, the two best performing composite biomarkers (i.e. CD8+IT-CD8 and CD3+IT-CD8) demonstrated similar or lower sensitivity (64% and 83%) and NPV (76% and 85%) than the individual biomarkers $PD\text{-}1^T$ TILs and TIS (sensitivity: 72% and 83%, NPV: 86% and 84%) for DC 6m, respectively. Also, at 12 months, both selected composite biomarkers (CD8+IT-CD8 and CD8+TIS) showed less predictive performance

compared to PD-1^T TILs and TIS alone. PD-1^T TILs and TIS showed high sensitivity (86% and 100%) and NPV (95% and 100%) for DC 12m. PD-1^T TILs could better discriminate patients with no long-term benefit, since specificity was substantially higher as compared to TIS (74% versus 39%).

Conclusion

Composite biomarkers did not show improved predictive performance compared to PD-1^T TILs and TIS alone for both the 6- and 12-months endpoint. PD-1^T TILs and TIS identified patients with DC 12m with high sensitivity. Patients with no long-term benefit to PD-1 blockade were most accurately identified by PD-1^T TILs.

Introduction

The success of monoclonal antibodies targeting the inhibitory receptor programmed cell death protein 1 (PD-1) and its ligand programmed death-ligand 1 (PD-L1) has changed the treatment landscape of advanced stage non-small cell lung cancer (NSCLC). A subset of patients treated with these PD-(L)1 blocking agents can achieve durable responses and gain a large survival benefit(1–7). Unfortunately, the majority does not derive durable clinical benefit, highlighting the need for predictive biomarkers to support treatment decision making in clinical practice. Specifically, biomarkers to exclude patients who are unlikely to benefit from PD-1 blockade therapy can offer patients alternative treatment options.

Tumor PD-L1 expression, as detected by immunohistochemistry (IHC), has been studied as a predictive biomarker in multiple clinical trials(8). A positive correlation between PD-L1 expression and treatment outcome has been reported in advanced stage NSCLC patients(1,5–7). However, approximately 60% to 70% of patients with PD-L1 positive tumors do not respond(1,2,5). Besides this, PD-L1 assessment by IHC is hampered by intratumor heterogeneity, interassay- and interobserver variability as well as pre-analytical variation(9–14). Tumor Mutation Burden (TMB), reflecting the number of somatic mutations as a surrogate of potential tumor antigenicity, has also shown predictive potential but clinical implementation remains challenging due to the lack of a robust and predictive TMB cut-off and the technical issues that arise due to variation across platforms(15–17).

For these reasons there is an urgent need for biomarkers that can more accurately predict response to PD-(L)1 blockade in advanced NSCLC. Since PD-(L)1 blockade is

thought to reinvigorate tumor-reactive T cells(18–20), several T cell markers have been investigated. For example, the density of $CD8^+$ tumor infiltrating lymphocytes (TILs) has been correlated to response to PD-(L)1 blockade in melanoma(18), colorectal cancer(21), and NSCLC(22,23). In previous work we have shown that a distinct T cell population, termed PD-1^T TILs, can predict clinical benefit in NSCLC(24,25). Notably, these PD-1^T TILs predominantly localize in tertiary lymphoid structures (TLS)(24). B cells, which are a critical component of these TLS, have also been linked to response to PD-(L)1 blocking agents(26–28). Other studies developed predictive RNA expression signatures, such as the “tumor inflammation signature” (TIS), to characterize features of immune activity in the tumor microenvironment (TME)(29–31).

Although all of these biomarkers have shown a certain predictive potential, their accuracy is still limited which is presumably caused by multiple components that are involved in the antitumor immune response. Hence, combining biomarkers could potentially improve their predictive accuracy, as previously has been shown for the combination of TMB with PD-L1(32,33) and CD8 TILs with PD-L1(22,34). Therefore, the aim of the present study was to investigate the performance of CD8, PD-1^T TILs and CD3 TILs, $CD20^+$ B cells, TLS, PD-L1 and TIS as pairs of biomarkers, compared to single biomarkers, for prediction of clinical benefit to PD-1 blockade in NSCLC.

Methods

Patients, endpoints and samples

In this study, 162 patients with pathologically confirmed stage IV NSCLC were eligible for efficacy analysis. All patients started second or later line monotherapy

nivolumab, 3mg/kg as an IV infusion every two weeks for at least one dose, between October 2014 and August 2017 at the Netherlands Cancer Institute/Antoni van Leeuwenhoek hospital (NKI-AVL), The Netherlands. Patients with tumors harboring known sensitizing EGFR mutations or ALK translocations were excluded from treatment. Patients were randomized into a training and validation cohort. Randomization was stratified by treatment outcome at 6 months and at 12 months. Since we could only generate gene expression data in 68/162 (42%) of patients' tumors, additional stratification was done by whether mRNA expression analysis was performed or not. Stratification for missing values of other biomarkers was not performed, as the number of excluded samples per biomarker was low (range 1 to 32) (see **Supplementary Fig. S1** and later in this section).

Response was assessed per Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1. Patients with progressive disease (PD) who were not evaluable for response were determined by the treating physician as PD. The primary clinical outcome was Disease Control (DC) (complete response (CR)/partial response (PR) or stable disease (SD)) at 6 months following initiation of treatment. DC 12m (CR/PR/SD that lasted ≥12 months) was used as secondary outcome measure to predict long-term efficacy to PD-1 blockade.

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. The data was accessed for research purposes after the approval of the

Institutional Review Board (IRB) of the Netherlands Cancer Institute on January 11, 2018 (CFMPB586). After K.H., M.M., R.D.S., M.M.H., E.F.S. and K.M. retrieved archived tumor samples and response data from medical records, all patients were pseudonymized. PD-1^T TIL and PD-L1 tumor proportion score (TPS) data for 94 samples were used from our previous work as well as tertiary lymphoid structures (TLS) and $CD20^+$ B cell data for 91 samples(25). In 27 patients, none of the biomarkers could be assessed because samples did not contain tumor tissue. In one sample no tumor tissue was left for CD8 and PD-1^T TIL analysis, as well as in five samples for CD3 TIL, TLS and $CD20^+$ B cell analysis **(Supplementary Fig. S1)**. An additional number of 32 patients were excluded for PD-1^T TIL analysis based on the following criteria: samples contained less than 10,000 cells (n=12), were obtained from endobronchial lesions (n=16), contained abundant normal lymphoid tissue (n=1) and showed fixation and/or staining artefacts (n=2) **(Supplementary Fig. S1)**. As described before, we excluded bronchial biopsies because they frequently showed unspecific antibody staining due to mechanical damage, and lymph node resections due to presence of PD-1^+ T cells in normal abundant lymphoid tissue, which could potentially lead to false positive results(25). One sample was excluded for CD8 TIL, CD3 TIL, TLS, $CD20^+$ B cell and PD-L1 analysis because of fixation/staining artefacts. One sample contained less than 2,000 cells and was excluded for CD8 TIL, CD3 TIL, TLS and $CD20^+$ B cell analysis. 67 patients (41%) were excluded for mRNA expression analysis because of low RNA yield and/or low RNA quality **(Supplementary Fig. S1)**.

Immunohistochemistry

CD8 immunostaining of samples was performed on a BenchMark Ultra autostainer Instrument (Ventana Medical Systems) on 3 µm paraffin sections from FFPE blocks. 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. CD8 was detected using clone C8/144B (1/200 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).

PD-1 immunostaining was detected using clone NAT105 (Roche Diagnostics), PD-L1 using clone 22C3 (Agilent/DAKO) and CD68 using clone KP1 (Agilent/DAKO). For the double staining CD20 (yellow) followed by CD3 (purple) we used clone L26 (Agilent/DAKO) (CD20) and clone SP7 (Thermo Fisher) (CD3). All immunostainings were performed as described previously(25).

CD8, PD-1, PD-L1 and CD68 immunostainings were scanned at x20 magnification with a resolution of 0.50 per μm^2 using an Aperio slide AT2 scanner (Leica Biosystems). CD20-CD3 immunostainings were scanned at x20 magnification with a resolution of 0.24 per μm^2 using a 3Dhistech P1000 scanner. PD-L1 and CD68 data were uploaded on Slide Score, a web platform for manual scoring of digital slides using a scoring sheet (www.slidescore.com).

Digital quantification of CD8 and PD-1^T TILs

Digital image analysis was performed by a trained MD (K.H.) and supervised by an experienced pathologist (K.M.) using the Multiplex IHC v1.2 module from the HALO™ image analysis software, v2.3.2089.69 (Indica Labs). Researchers were blinded for clinical outcome. Classification of CD8 lymphocytes on single stains was performed using a computationally derived cut-off of 0.3 optical density (OD), which reflects the intensity of the staining. This cut-off was identified by manually optimizing the detection of CD8 positive stained cells in FFPE samples. An image analysis algorithm utilizing a cut-off of 0.3 OD was generated for automated analyses of CD8 lymphocytes in subsequent FFPE samples. The quantification of PD-1^T TILs was performed as described previously(25).

The number of CD8 and PD-1^T TILs per mm^2 tumor area were determined. Tumor areas were digitally annotated as described previously(25). PD-1^T TIL data of 94 samples were used from previous work(25). For regional analysis of CD8 lymphocytes, classifiers were trained to identify stromal and tumoral regions in which the CD8 lymphocytes were quantified separately. The percentage CD8 lymphocytes in tumoral regions (i.e. intra-tumoral (IT)) compared to total CD8 TILs was calculated **(Supplementary Table S1)**.

Scoring of tertiary lymphoid structures

The HALO™ image analysis software, v2.3.2089.69 (Indica Labs) was used to determine the number of TLS and the combined number of TLS and lymphoid aggregates (TLS+LA) per mm^2 tumor area on a CD20-CD3 double immunostaining as described previously(25). TLS and TLS+LA data of 91 samples were used from previous work(25) **(Supplementary Table S1)**.

CD20 and CD3 quantification by digital image analysis

The total area with CD20 expression was measured using a previously generated image analysis algorithm from the Area Quantification v1.0 module of HALO™ image analysis software (Indica Labs)(25). The same algorithm was used to measure the total area with CD3 expression. The CD20-positive and CD3-positive area were normalized per mm^2 tumor area. Cell numbers were not quantified as no reliable algorithm could be established due to dense clustering of $CD20^+$ or $CD3^+$ cells in and at the border of TLS. Tumor areas were digitally annotated as described previously(25). CD20 data of 91 samples were used from previous work(25) **(Supplementary Table S1)**.

PD-L1 scoring

PD-L1 TPS was determined using the qualitative, clinical grade LDT IHC assay (22C3 Agilent/DAKO) as described previously(25). PD-L1 TPS data of 94 samples were used from previous work(25) **(Supplementary Table S1)**. The CD68 staining was compared to the PD-L1 staining to exclude macrophages that are both $CD68^+$ and PD-$L1^+$ which can introduce false positive results.

RNA extraction and hybridization to nCounter tagset

RNA of pretreatment FFPE samples from the NKI-AVL cohorts were isolated with the AllPrep DNA/RNA FFPE isolation kit (#80234, Qiagen) according to the instructions of the manufacturer and quantified by Tapestation (Agilent). 200 to 300 ng RNA were hybridized to Nanostring PanCancer IO 360 Panel code set (Nanostring), according to the recommendations of the manufacturer. After hybridization non-bound probes were

231 washed off and the RNA-probe complex was bound to the cartridge on the Nanostring
232 Flex Prep Station according to manufacturing protocol. The cartridge was sealed and
233 transferred to the Digital Analyzer for imaging.
234
235 *Statistical analysis*
236 Patient characteristics were descriptively reported using mean ± SD, interquartile
237 range (IQR) or frequencies (percentages). The Mann-Whitney test for continuous data,
238 Fisher's exact test for categorical data and linear-by-linear association test for ordinal
239 data were used to assess differences in patient characteristics between cohorts (training
240 and validation) and between outcome groups (disease control vs PD). Differences were
241 considered statistically significant if *P<0.05, **P<0.01, ***P<0.001 or ****P<0.0001.
242 Genes in the Tumor Inflammation Signature (TIS) are normalized using a ratio of
243 the expression value to the geometric mean of the housekeeper genes used only for the
244 TIS signature and then followed by log2 transformation. The TIS score was calculated as
245 a weighted linear combination of the 18 gene expression values(29,35) **(Supplementary**
246 **Table S1)**. This analysis was performed by Nanostring as part of their intellectual
247 property.
248 In the training cohort, univariate models and bivariate logistic models for DC 6m
249 and DC 12m of treatment were constructed using CD8 TILs, IT-CD8 T cells, PD-1^{T} TILs,
250 CD3 TILs, TLS, TLS+LA, CD20^{+} B cells, PD-L1 and TIS. The bivariate models included
251 an interaction term. The bivariate logistic model produces for each patient a number
252 between 0 and 1, reflecting the probability (according to the model) of patients reaching
253 DC 6m or DC 12m. Calculation of the area under the receiver operating characteristic

(ROC) curve was used as a measure of discriminatory ability. The predictive performance of different individual and composite biomarkers 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. A point on the ROC curve matching a sensitivity of 90% for DC 6m and 90% for DC 12m was selected to calculate corresponding specificity, NPV and PPV. We further aimed for an NPV of ≥90% and a specificity of ≥50%.

Two (closely related) non-parametric approaches were considered to obtain 90% sensitivity for predicting DC 6m and DC 12m from two biomarkers: by choosing a cut-point for each of the two biomarkers and declaring the patient positive (i.e. likely to respond to PD-1 blockade) when either at least one (first method) or both (second method) biomarker values were above their respective cut-point values. The specificities obtained were either equal or worse to those obtained by the parametric method described above (i.e. via logistic regression). Therefore, these non-parametric methods were not used in this study.

Four training models were selected with a cut-off that showed the highest specificity and NPV at the prespecified sensitivities for prediction of DC 6m and DC 12m. This cut-off was used to determine sensitivity, specificity, NPV and PPV in the validation cohort.

Results

Biomarker characteristics and demographics

To assess the predictive performance of multiple biomarker combinations we first analyzed 162 pretreatment tumor samples from 162 advanced stage NSCLC patients treated with nivolumab. In total we evaluated 9 biomarkers: (1) the total number of CD8 TILs per mm^2, (2) the percentage intra-tumoral (IT) CD8 T cells of total CD8 TILs, (3) the number of PD-1^T TILs per mm^2 (4) the CD3-positive area per mm^2 to estimate the presence of CD3 TILs (5) the CD20-positive area per mm^2 to estimate the presence of B cells (6) the number of TLS and (7) the combined number of TLS and LA (referred as TLS+LA) per mm^2, (8) the PD-L1 Tumor Proportion Score (TPS) and (9) the TIS score (NanoString**) (Fig. 1).** We could successfully assess CD8 TILs and IT-CD8 T cells in 132/162 (81%), PD-1^T TILs in 103/162 (64%), CD3 TILs, $CD20^+$ B cells, TLS and TLS+LA in 128/162 (79%), PD-L1 TPS in 134/162 (83%) and TIS in 68/162 (42%) samples **(Table 1, Supplementary Fig. S1)**. Sample exclusion criteria are shown per biomarker in **Supplementary Fig. S1**.

We randomized patients with ≥2 biomarker results available (n=135) in a training (n=55) and validation (n=80) cohort. This randomization was stratified for clinical benefit to ascertain that in both cohorts 1 in 3 patients reached DC 6m and 1 in 5 patients reached DC 12m, respectively. Since a limited number of patients with TIS scores (n=68) were available, these patients were randomly distributed proportionately **(Table 1, Supplementary Fig. S1)**. For every patient the results per biomarker are shown in **Supplementary Table S1**. Demographic characteristics did not significantly differ among the training and validation cohort **(Table 2)**.

Accuracy of individual and composite biomarkers to predict DC at 6 months

Next, we determined the most optimal cut-offs for each individual and composite biomarkers in the training cohort. We aimed for a sensitivity and NPV of ≥90% to minimize undertreatment and a specificity of at least 50% to identify those patients that are unlikely to respond to PD-1 blockade therapy and can potentially benefit from alternative treatments. Since not all tumor samples were evaluable for all nine biomarkers, the number of training samples ranged from 28 to 55 **(Table 1, Supplementary Fig. S1)**. In total, 16 composite biomarkers and PD-1^{T} and TIS as individual biomarkers reached ≥90% sensitivity and ≥50% specificity **(Supplementary Table S2)**. Interestingly, these include 7/8 (88%) possible combinations with PD-1^{T} TILs and 5/8 (63%) with TIS **(Supplementary Table S2)**. However, none of these combinations did significantly improve predictive accuracy compared to PD-1^{T} TILs and TIS alone **(Supplementary Fig. S2A,B)** and were excluded from further analysis.

Next, we selected the four remaining biomarkers with the highest predictive performance for validation, being the combinations of CD8+IT-CD8 and CD3+IT-CD8, as well as PD-1^{T} TILs and TIS alone, respectively **(Supplementary Table S2)**. In the training cohort, both CD8+IT-CD8 and CD3+IT-CD8 had significantly higher probability scores in the DC 6m group (reflecting the probability of patients reaching DC 6m) compared to the PD group (CD8+IT-CD8, $P<0.0001$ and CD3+IT-CD8, $P<0.001$) **(Fig. 2A,B)**. The area under the ROC curve (AUC) was 0.83 (95% CI 0.73-0.94) for CD8+IT-CD8, and 0.78 (95% CI 0.65-0.92) for CD3+IT-CD8 **(Fig. 2C,D)**. Cut-offs of 0.167 and 0.161, respectively, correlated to a sensitivity of 94% and 94%, specificity of 62% and 54%, NPV of 96% and 95% and PPV of 50% and 47% **(Table 3)**.

Also, the PD-1^T TIL numbers and TIS scores were significantly higher in the DC 6m group than in the PD group (PD-1^T TILs, P<0.001 and TIS, P<0.01) **(Supplementary Fig. S2C,D)**. PD-1^T TILs showed an AUC of 0.82 (95% CI 0.69-0.95) and TIS an AUC of 0.81 (95% CI 0.65-0.98) **(Supplementary Fig. S2E,F)**. For PD-1^T TILs a cut-off of 90 per mm^2 was chosen, as this cut-off showed predictive value in a prior study(25). We observed a sensitivity of 92%, specificity of 67%, NPV of 95% and PPV of 52% **(Table 3)**. A score of 6.65 was chosen as optimal cut-off for TIS which correlated to a sensitivity of 100%, specificity of 55%, NPV of 100% and PPV of 47% **(Table 3)**.

Next, we evaluated the predictive performance of the four selected biomarkers in the validation cohort. The number of validation samples with successful biomarker results ranged from 40 to 79 **(Table 1, Supplementary Fig. S1)**. We observed that the predictive accuracy of CD8+IT-CD8 and CD3+IT-CD8 biomarkers was substantially lower compared to the training cohort. Specifically, probability scores in the DC 6m group did not significantly differ from scores in the PD group for CD8+IT-CD8 (P=0.08) **(Fig. 2E).** For CD3+IT-CD8 this comparison was borderline significant (P=0.01) **(Fig. 2F)**. The AUC of the ROC curve was 0.62 (95% CI 0.50-0.75) for CD8+IT-CD8 and 0.68 (95% CI 0.55-0.80) for CD3+IT-CD8 **(Fig. 2C,D)**. CD8+IT-CD8 reached a sensitivity of 64%, specificity of 56%, NPV of 76% and PPV of 41%. The predictive accuracy of CD3+IT-CD8 was higher than CD8+IT-CD8 but still lower than in the training cohort, reaching a sensitivity of 83%, specificity of 46%, NPV of 85% and PPV of 43% **(Table 3)**.

The individual biomarkers in the validation cohort showed that PD-1^T TIL numbers were significantly higher in the DC 6m group versus the PD group (P<0.01), which was not observed for TIS scores (P=0.52) **(Supplementary Fig. S2G,H)**. The discriminatory

ability of PD-1^T TILs was lower as in the training, but still reached an AUC of 0.72 (95% CI 0.57-0.87) **(Supplementary Fig. S2E)**. TIS reached an AUC of 0.57 (95% CI 0.36-0.77) **(Supplementary Fig. S2F)**. A cut-off of 90 PD-1^T TILs per mm^2 correlated to a sensitivity of 72%, specificity of 74%, NPV of 86% and PPV of 54%. A TIS score of 6.65 showed a comparable sensitivity (83%), NPV (84%) and PPV (37%) but lower specificity (39%) **(Table 3)**. In summary, these results demonstrate that a combination of CD8+IT-CD8 and CD3+IT-CD8 did not improve predictive accuracy compared to PD-1^T TILs and TIS alone. Furthermore, none of the selected biomarkers reached the prespecified performance criteria.

Accuracy of individual and composite biomarkers to predict DC at 12 months

Approximately 70-80% of patients treated in 2nd line with PD-(L)1 blockade progress within 12 months(2–4). We previously demonstrated that PD-1^T TILs could more effectively identify patients with DC 12m as compared to DC 6m, as well as a patient group without long-term benefit(25). We therefore also assessed the performance of all biomarkers to predict DC 12m. Similar to the DC 6m analysis, we determined the most optimal cut-offs for each of the composite and individual biomarkers to identify patients with DC 12m and with PD. Four patients in the training and nine patients in the validation experienced disease progression between 6 and 12 months, and were therefore included in the PD group in this analysis. 16 composite biomarkers reached ≥90% sensitivity and ≥50% specificity in the training cohort, as well as PD-1^T TILs and TIS as individual biomarkers **(Supplementary Table S3)**. We observed that 12/16 composite and 2/2 individual biomarkers (PD-1^T TILs and TIS), matched the 6-months endpoint with similar

accuracy **(Supplementary Table S2,3)**. PD-1^T TIL combinations did not significantly improve predictive accuracy compared to PD-1^T TILs alone and were excluded from further analysis **(Supplementary Fig. S3A, Supplementary Table S3)**. However, the combination of CD8 with TIS (CD8+TIS) showed an increase of 18% specificity compared to TIS alone. This combination was selected for further analysis, even though it did not reach statistical significance, possibly due to the low sample size (*P*=0.34) **(Supplementary Fig. S3B, Supplementary Table S3)**. The four biomarkers with the highest predictive performance were selected for validation. 3/4 selected biomarkers, including PD-1^T TILs, TIS and CD8+IT-CD8, matched the DC 6m selection. The fourth biomarker included the combination of CD8+TIS **(Supplementary Table S3)**.

The probability scores for DC 12m and PD are shown per sample in **Supplementary Fig. S3C** (CD8+IT-CD8, *P*<0.001) and **Supplementary Fig. S3D** (CD8+TIS, *P*<0.01). The two composite biomarkers showed a high AUC of 0.85 (95% CI: 0.73-0.96) (CD8+IT-CD8) and 0.91 (95% CI: 0.79-1.00) (CD8+TIS) in the training cohort **(Fig. 3A,B)**. A cut-off of 0.122 and 0.124, respectively, was chosen as optimal cut-off **(Table 3)**. PD-1^T TIL numbers and TIS scores are shown in **Supplementary Fig S3E** and **F**. A cut-off of 90 PD-1^T TILs per mm^2 and a TIS score of 6.65 demonstrated similar predictive accuracy as in the training for DC 6m **(Fig. 3C,D, Table 3)**.

In the validation cohort, of the two composite biomarkers, only CD8+IT-CD8 showed borderline significantly higher probability scores in the DC 12m group versus the PD group (*P*=0.03) **(Supplementary Fig. S3G,H)**. The ROCs yielded low AUCs (CD8+IT-CD8: 0.67 (95% CI: 0.53-0.81), CD8+TIS: 0.59 (95% CI 0.36-0.81)) **(Fig 3A,B)**. Furthermore, the sensitivity (68% and 29%), specificity (57% and 68%), NPV (88% and

81%) and PPV (30% and 17%) did not meet the prespecified performance criteria **(Table 3)**.

PD-1^{T} TIL numbers were significantly higher in patients with DC 12m versus PD (*P*<0.001) **(Fig. 3E)**. PD-1^{T} TILs also demonstrated a consistently high AUC (0.80, 95% CI: 0.65-0.94) and high accuracy, reaching a sensitivity of 86%, specificity of 74%, NPV of 95% and PPV of 50% **(Fig. 3C, Table 3)**. We observed an enrichment of patients with DC 12m in the ≥90 group and with PD in the <90 subgroup **(Fig. 3E)**. TIS scores did not significantly differ between the two groups (*P*=0.31) and showed a low AUC of 0.63 (95% CI 0.43-0.82) **(Fig. 3D,F)**. However, a cut-off score of 6.65 reached a sensitivity of 100%, specificity of 39%, NPV of 100% and PPV of 26% **(Table 3)**. These findings did not meet our ≥50% specificity criterium, but accurately identified all patients with DC 12m including 39% of patients with PD **(Fig. 3F)**. Taken together, PD-1^{T} TILs and TIS as individual biomarkers showed higher predictive accuracy for DC 12m compared to the combination of CD8+IT-CD8 and CD8+TIS. Notably, PD-1^{T} TILs alone was more accurate than TIS alone, as specificity and PPV were substantially higher.

Discussion

Since the introduction of PD-(L)1 blockade therapy, clinical outcome of advanced stage NSCLC has dramatically improved. Nevertheless, a subset of patients derive benefit from these treatments which consequently has led to overtreatment and unnecessarily side effects in many. In addition, health care systems deal with increasing costs. Several predictive biomarkers have been identified to support treatment decision making. Since different components in the TME can affect the tumor immune response

upon PD(L)1 blockade therapy, it is unlikely to find one single perfect biomarker. Based on the hypothesis that a predictive model should contain more than one biomarker; we here assess the predictive performance of biomarker combinations in an advanced stage NSCLC cohort treated with nivolumab. Our data showed that selected composite biomarkers did not improve predictive performance as compared to PD-1^{T} TILs and TIS alone. At 6 months, none of the selected composite and individual biomarkers reached the prespecified performance criteria in the validation cohort. At 12 months, PD-1^{T} TILs and TIS could identify patients with DC 12m with high accuracy. Patients without long-term benefit were more accurately identified by PD-1^{T} TILs than TIS.

Whereas CD8 or CD3 TILs in combination with intratumoral localization of CD8 T cells were the most accurate composite biomarkers for DC 6m in the training cohort, we observed that discriminatory ability was low in the validation cohort. The presence and localization of TILs alone might not indicate that all T cells are in a state to recognize and eliminate the tumor(36,37). In the present study, this notion is supported by the high accuracy of PD-1^{T} TILs to predict DC 12m, as these TILs have been identified as a distinct TIL subset with a high capacity of tumor recognition(24). The results are similar to our previous work because the majority of samples were re-used(25). Further refinement of this T cell population could contribute to the development of new markers or gene signatures, as recently been done by other studies(38–40). Since most of the biomarkers assessed in this study are related to antitumor immunity and are presumably correlated, PD-1^{T} combinations did not improve specificity.

Previous studies have shown the predictive potential of combining CD8+PD-L1(22,34). However, in our training cohort, CD8+PD-L1 did not meet our performance

436 criteria, and as a result, this combination was not further evaluated. Noguchi et al.
437 previously observed that PD-L1 expression on tumor cells is transient and dependent on
438 the production of IFNγ by TILs(41). Hence, variable tumor PD-L1 expression in training
439 samples might have affected the predictive accuracy of PD-L1 alone and that of PD-L1
440 combinations. Furthermore, this study is limited by the number of samples, in particularly
441 for TIS assessment. Therefore, we restricted our evaluation to two-biomarker
442 combinations instead of considering three or more. Studies involving a larger number of
443 samples are essential to further validate our findings.

444 Our results for TIS are in line with other studies that demonstrated the predictive
445 potential of this signature(29,42). Interestingly, TIS contains genes that are highly
446 expressed in PD-1^T TILs, such as LAG3 and TIGIT(24,29). A high number of PD-1^T TILs
447 or a high TIS score in pretreatment samples may serve as surrogate markers for a tumor's
448 ability to undergo durable immune reactivation upon PD-1 blockade therapy. A PD-1^T
449 TILs or TIS combination with biomarkers reflecting distinct parts of the immune response
450 could potentially improve predictive accuracy. For example, TMB can serve as a read-out
451 for immunogenic neoantigens that arise from somatic mutations(15). TMB and PD-L1
452 have previously been described as independent predictors for advanced NSCLC treated
453 with PD-1 blockade and have shown improved performance when combined(32,33).
454 Another suggestion is, in contrast, the presence of tumor-resident regulatory T cells (T_{reg})
455 in the TME. T_{reg} cells possess an immune-inhibitory function and high numbers are
456 correlated to poor patient survival(43). A combination of TMB or T_{reg} with either PD-1^T or
457 TIS could be further explored in future work.

In conclusion, this study showed that the biomarker combinations assessed here did not improve predictive performance when compared to PD-1^{T} TILs and TIS alone. PD-1^{T} TILs showed the highest predictive performance of all biomarkers, as patients with no long-term benefit were identified with high specificity and NPV.

Acknowledgements

We would like to thank the NKI-AVL Core Facility Molecular Pathology and Biobanking for supplying all IHC stainings used in this study, as well as biobank-related work and other laboratory support.

References

1. 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(19):1823–33. Available from: http://www.nejm.org/doi/10.1056/NEJMoa1606774
2. 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(17):1627–39.
3. 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(10066):255–65.
4. 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(2):123–35.

5. 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(21):2018–28.

6. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csoszi 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(7):537–46.

7. Garon EB, Hellmann MD, Rizvi NA, Carcereny E, Leighl NB, Ahn MJ, et al. Five-year overall survival for patients with advanced non-small-cell lung cancer treated with pembrolizumab: Results from the phase i KEYNOTE-001 study. J Clin Oncol. 2019;37(28):2518–27.

8. Peters S, Reck M, Smit EF, Mok T, Hellmann MD. How to make the best use of immunotherapy as first-line treatment of advanced/metastatic non-small-cell lung cancer. Ann Oncol. 2019;30(6):884–96.

9. 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(1):147–53. Available from: https://doi.org/10.1093/annonc/mdv489

10. 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(4):530–8.

11. Boothman AM, 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(8):1390–9.

12. Hong L, Negrao M V., 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(9):1449–59. Available from: https://doi.org/10.1016/j.jtho.2020.04.026

13. McLaughlin J, Han G, Schalper KA, Carvajal-Hausdorf D, Pelekanou V, Rehman J, et al. Quantitative assessment of the heterogeneity of PD-L1 expression in non-small-cell lung cancer. JAMA Oncol. 2016;2(1):46–54.

14. Butter R, A'T Hart N, Hooijer GKJ, Monkhorst K, Speel EJ, Theunissen P, et al. Multicentre study on the consistency of PD-L1 immunohistochemistry as predictive test for immunotherapy in non-small cell lung cancer. J Clin Pathol. 2020;73(7):423–30.

15. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science (80-). 2015;348(6230):124–8.

16. Ready N, Hellmann MD, Awad MM, Otterson GA, Gutierrez M, Gainor JF, et al. First-line nivolumab plus ipilimumab in advanced non-small-cell lung cancer (CheckMate 568): Outcomes by programmed death ligand 1 and tumor mutational

burden as biomarkers. J Clin Oncol. 2019;37(12):992–1000.

17. Sholl LM, Hirsch FR, Hwang D, Botling J, Lopez-Rios F, Bubendorf L, et al. The Promises and Challenges of Tumor Mutation Burden as an Immunotherapy Biomarker: A Perspective from the International Association for the Study of Lung Cancer Pathology Committee. J Thorac Oncol. 2020;15(9):1409–24.

18. Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJM, Robert L, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. Nature. 2014;515(7528):568–71.

19. 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(4):218–32. Available from: http://dx.doi.org/10.1038/s41568-019-0235-4

20. Huang AC, Postow MA, Orlowski RJ, Mick R, Bengsch B, Manne S, et al. T-cell invigoration to tumour burden ratio associated with anti-PD-1 response. Nature. 2017;545(7652):60–5.

21. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N Engl J Med. 2015;372(26):2509–20.

22. Fumet JD, 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(8):950–60. Available from: http://dx.doi.org/10.1038/s41416-018-0220-9

23. 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(16):5061–8.

24. 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(7). Available from: http://dx.doi.org/10.1038/s41591-018-0057-z

25. Hummelink K, van der Noort V, Muller M, Schouten RD, Lalezari F, Peters D, et al. PD-1^{T} TILs as a predictive biomarker for clinical benefit to PD-1 blockade in patients with advanced NSCLC. Clin Cancer Res. 2022;(10):1–14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/35852792

26. 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;(February). Available from: http://www.ncbi.nlm.nih.gov/pubmed/31942071

27. 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;(February). Available from: http://www.ncbi.nlm.nih.gov/pubmed/31942075

28. Petitprez F, de Reyniès A, 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(June 2018). Available from: http://www.ncbi.nlm.nih.gov/pubmed/31942077

29. Ayers M, Ribas A, Mcclanahan TK, Ayers M, Lunceford J, Nebozhyn M, et al. blockade IFN- γ – related mRNA profile predicts clinical response to PD-1 blockade. J Clin Invest. 2017;127(8):2930–40.

30. 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(10030):1837–46. Available from: http://dx.doi.org/10.1016/S0140-6736(16)00587-0

31. Higgs BW, Morehouse CA, Streicher K, Brohawn PZ, Pilataxi F, Gupta A, et al. Interferon gamma messenger RNA Signature in tumor biopsies predicts outcomes in patients with non–small cell lung carcinoma or urothelial cancer treated with durvalumab. Clin Cancer Res. 2018;24(16):3857–66.

32. 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(7):633–41.

33. 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(5):843-852.e4.

34. 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):1–12.

35. Danaher P, Warren S, Lu R, Samayoa J, Sullivan A, Pekker I, et al. Pan-cancer adaptive immune resistance as defined by the Tumor Inflammation Signature (TIS): Results from The Cancer Genome Atlas (TCGA). J Immunother Cancer. 2018;6(1):1–17.

596 36. Scheper W, Kelderman S, Fanchi LF, Linnemann C, Bendle G, de Rooij MAJ, et
597 al. Low and variable tumor reactivity of the intratumoral TCR repertoire in human
598 cancers. Nat Med. 2019;25(1):89–94. Available from:
599 http://dx.doi.org/10.1038/s41591-018-0266-5
600 37. Simoni Y, Becht E, Fehlings M, Loh CY, Koo SL, Teng KWW, et al. Bystander
601 CD8+ T cells are abundant and phenotypically distinct in human tumour infiltrates.
602 Nature. 2018;557(7706):575–9.
603 38. Oliveira G, Stromhaug K, Klaeger S, Kula T, Frederick DT, Le PM, et al.
604 Phenotype, specificity and avidity of antitumour CD8+ T cells in melanoma.
605 Nature. 2021;596(7870):119–25. Available from:
606 http://dx.doi.org/10.1038/s41586-021-03704-y
607 39. Caushi JX, Zhang J, Ji Z, Vaghasia A, Zhang B, Hsiue EHC, et al. Transcriptional
608 programs of neoantigen-specific TIL in anti-PD-1-treated lung cancers. Vol. 596,
609 Nature. Springer US; 2021. 126–132 p. Available from:
610 http://dx.doi.org/10.1038/s41586-021-03752-4
611 40. Lowery FJ, Krishna S, Yossef R, Parikh NB, Chatani PD, Zacharakis N, et al.
612 Molecular signatures of antitumor neoantigen-reactive T cells from metastatic
613 human cancers. Science. 2022;884(February):eabl5447. Available from:
614 http://www.ncbi.nlm.nih.gov/pubmed/35113651
615 41. Noguchi T, Ward JP, Gubin MM, Arthur CD, Lee SH, Hundal J, et al. Temporally
616 distinct PD-L1 expression by tumor and host cells contributes to immune escape.
617 Cancer Immunol Res. 2017;5(2):106–17.
618 42. Damotte D, Warren S, Arrondeau J, Boudou-Rouquette P, Mansuet-Lupo A, Biton

J, et al. The tumor inflammation signature (TIS) is associated with anti-PD-1 treatment benefit in the CERTIM pan-cancer cohort. J Transl Med. 2019;17(1):1–10. Available from: https://doi.org/10.1186/s12967-019-2100-3

43. Kim HJ, Cantor H. CD4 T-cell subsets and tumor immunity: the helpful and the not-so-helpful. Cancer Immunol Res. 2014;2(2):91–8.

Figure legends

Figure 1. **Immunohistochemical analysis of all biomarkers and digital mark-up** **(A)** The left image shows an example of a CD8 immunohistochemical staining (IHC). The black square indicates the area that is shown in the central image. The right image shows the digital markup with CD8 TILs (in brown) and all other cells (in grey). **(B)** The left image shows the same example as shown in A. The black square indicates the area that is shown in the central image. The right image shows regional analysis of only intratumoral (IT) CD8 TILs. Stromal CD8 TILs are not quantified. Red lines indicate the tumor region. Red arrows indicate IT-CD8 TILs. White arrow indicates the area with stromal CD8 TILs. **(C)** The left image shows an example of a consecutive slide stained for PD-1 IHC. The black square indicates the area that is shown in the central image. The right image shows the digital markup with PD-1^{T} TILs (in brown) and all other cells (in grey). **(D)** The left image shows an example of a consecutive slide double stained with CD20 and CD3 IHC. The black square indicates the area that is shown in the central image with $CD20^{+}$ B cells (in yellow) and $CD3^{+}$ T cells (in purple) localizing in a TLS. The right image shows the digital markup with CD20-positive areas highlighted by the intensity of the yellow staining (depicted as spectrum from yellow to red color). **(E)** Example of a consecutive slide

642 stained for PD-L1 IHC. The black square indicates the area that is shown in the right
643 image. PD-L1 IHC slides were scored manually.
644
645 Figure 2. **Performance of selected composite and individual biomarkers to predict**
646 **DC at 6 months in NSCLC patients treated with PD-1 blockade.**
647 **(A)** Probability scores of CD8+IT-CD8 in pretreatment samples from patients with disease
648 control at 6 months (DC 6m) (n=16) and progressive disease (PD) (n=39) in the training
649 cohort (n=55). Dashed line indicates a cut-off of 0.167. Medians, interquartile ranges and
650 minimum/maximum shown in boxplots, ****P<0.0001 by Mann Whitney U-test. **(B)**
651 Probability scores of CD3+IT-CD8 in pretreatment samples from patients with DC 6m
652 (n=16) and PD (n=37) in the training cohort (n=53). Dashed line indicates a cut-off of
653 0.161. Medians, interquartile ranges and minimum/maximum shown in boxplots,
654 ***P<0.001 by Mann Whitney U-test. **(C)** Receiver operating characteristic (ROC) curve
655 for predictive value of CD8+IT-CD8 for DC 6m in the training (n=55) (AUC 0.83, 95% CI:
656 0.73-0.94) and validation cohort (n=77) (AUC 0.62, 95% CI: 0.50-0.75). **(D)** ROC curve
657 for predictive value of CD3+IT-CD8 for DC 6m in the training (n=53) (AUC 0.78, 95% CI:
658 0.65-0.91) and validation cohort (n=74) (AUC 0.68, 95% CI: 0.55-0.80). **(E)** Same plot as
659 in **A** (CD8+IT-CD8) for patients with DC 6m (n=25) and PD (n=52) in the validation cohort
660 (n=77), P=0.08. **(F)** Same plot as in **B** (CD3+IT-CD8) for patients with DC 6m (n=24) and
661 PD (n=50) in the validation cohort (n=74), *P=0.02.
662
663 Figure 3. **Performance of selected composite and individual biomarkers to predict**
664 **DC at 12 months in NSCLC patients treated with PD-1 blockade.**

(A) Receiver operating characteristic (ROC) curve for predictive value of CD8+IT-CD8 for disease control at 12 months (DC 12m) in the training cohort (n=55) (AUC 0.85, 95% CI: 0.73-0.96) and validation cohort (n=77) (AUC 0.67, 95% CI: 0.53-0.81). **(B)** ROC curve for predictive value of CD8+TIS for DC 12m in the training cohort (n=28) (AUC 0.91, 95% CI: 0.79-1.00) and validation cohort (n=38) (AUC 0.59, 95% CI: 0.36-0.82). **(C)** ROC curve for predictive value of PD-1^{T} TILs for DC 12m in the training cohort (n=42) (AUC 0.82, 95% CI: 0.70-0.94) and validation cohort (n=61) (AUC 0.80, 95% CI: 0.65-0.94). **(D)** ROC curve for predictive value of TIS for DC 12m in the training cohort (n=28) (AUC 0.77, 95% CI: 0.58-0.96) and validation cohort (n=40) (AUC 0.63, 95% CI: 0.43-0.82). **(E)** PD-1^{T} TILs per mm^2 in pretreatment samples from patients with DC 12m (n=14) and PD (n=47) in the validation cohort (n=61). Dashed line indicates a cut-off of 90 PD1^{T} TILs per mm^2. Medians, interquartile ranges and minimum/maximum shown in boxplots, $^{**}P<0.01$ by Mann Whitney U-test. **(F)** TIS scores in pretreatment samples from patients with DC 12m (n=7) and PD (n=33) in the validation cohort (n=40). Dashed line indicates a cut-off score of 6.65. Medians, interquartile ranges and minimum/maximum shown in boxplots, $P=0.31$ by Mann Whitney U-test.

688 Table 1. **Total number of samples per biomarker in the training and validation cohort.**

Biomarkers	**Total samples (n)**	Training (n)					Validation (n)				
		DC 6m	PD (within 6m)	*DC 12 m*	*PD (within 12m)*	**Total**	DC 6m	PD (within 6m)	*DC 12m*	*PD (within 12m)*	**Total**
CD8 TILs IT-CD8 T cells	132	16	39	*12*	*43*	**55**	25	52	*16*	*61*	**77**
PD-1^{T} TILs	103	12	30	*9*	*33*	**42**	18	43	*14*	*47*	**61**
CD3 TILs CD20^{+} B cells TLS TLS+LA	128	16	37	*12*	*41*	**53**	24	51	*15*	*60*	**75**
PD-L1 TPS	134	16	39	*12*	*43*	**55**	25	54	*16*	*63*	**79**
TIS	68	8	20	*6*	*22*	**28**	12	28	*7*	*33*	**40**

Table 2. **Patient characteristics and treatment outcomes for training and validation cohorts.** *P*-values were calculated by Mann-Whitney, Fisher exact or linear-by-linear association tests. *S.d*, standard deviation; *IQR*, interquartile range; *PS*, Performance Score, based on the European Cooperative Oncology group (ECOG) performance status score. This is a score ranging from 0 to 5, where 0 indicates no symptoms, 1 indicates mild symptoms and above 1 indicates greater disability; *LCNEC NSCLC* type, large cell neuroendocrine carcinoma non-small cell lung cancer type; *NOS*, not otherwise specified; *KRAS*, Kirsten Rat Sarcoma viral oncogene; *PD-L1*, programmed death ligand 1; *TPS*, tumor proportion score; *CR*, complete response; *PR*, partial response; *SD*, stable disease; *PD*, progressive disease; *DC*, disease control.

		Training cohort	**Validation cohort**
	p-value	n=55	n=80
Sex	1.00		
Male, no.(%)		30 (55%)	44 (55%)
Female, no.(%)		25 (45%)	36 (45%)
Age (years), mean (SD)	0.20	62 (10.1)	65 (7.5)
Smoking (never/ex/current)	0.64	5/44/6	12/51/17
Pack years, median (IQR)	0.90	29 (20)	30 (28)
PS, no. (%)	0.46		
0		16 (29%)	16 (20%)
1		29 (53%)	50 (62%)
≥2		10 (18%)	14 (18%)
Pathology, no.(%)	0.19		
Adeno		35 (64%)	50 (62%)
Squamous		10 (18%)	20 (25%)
LCNEC, NSCLC-type		0 (0%)	3 (4%)
NSCLC, NOS		10 (18%)	7 (9%)
Mutations, no. (%)	0.86		

KRAS positive		19 (35%)	30 (38%)
PD-L1 TPS, no. (%)			
Negative <1%	1.00	30 (55%)	43 (54%)
Positive ≥1%		25 (45%)	36 (45%)
Negative <50%	0.66	43 (78%)	65 (81%)
Positive >50%		12 (22%)	14 (18%)
Unknown		0 (0%)	1 (1%)
Brain metastases, no. (%)	0.67	13 (24%)	16 (20%)
Line of treatment, no (%)	0.63		
1		0 (0%)	1 (1%)
2		42 (76%)	56 (70%)
>2		13 (24%)	23 (29%)
Best Overall Response	0.62		
CR/PR		11 (20%)	15 (19%)
SD		5 (9%)	16 (20%)
SD (PFS <6 months)		0 (0%)	6 (7%)
SD (PFS ≥6 months)		5 (9%)	10 (13%)
PD		39 (71%)	49 (61%)
DC			
at 6 months	0.85	16 (29%)	25 (31%)
at 12 months	0.83	12 (22%)	16 (20%)

Table 3. **Predictive accuracy of selected individual and composite biomarkers, summary of training and validation results**

					Training						**Validation**					
Clinical outcome	**Bio-marker**	**Predictor**	**Cut-off**	Samples (n)	***AUC (95% CI)***	**Sensitivity**	**Specificity**	**NPV**	**PPV**		Samples (n)	***AUC (95%-CI)***	**Sensitivity**	**Specificity**	**NPV**	**PPV**
DC 6 months	PD-1^{T} TILs		90	*42*	0.82 (0.69-0.95)	92%	67%	95%	52%		*61*	0.72 (0.57-0.87)	72%	74%	86%	54%
	TIS		6.65	*28*	0.81 (0.65-0.98)	100%	55%	100%	47%		*40*	0.57 (0.36-0.77)	83%	39%	84%	37%
	CD8+ IT-CD-8	probability for DC = 1 / (1 – exp (-3.5749 + 0.0031 * CD8 + 0.043 * IT-CD8)	0.167	*55*	0.83 (0.73-0.94)	94%	62%	96%	50%		*77*	0.62 (0.50-0.75)	64%	56%	76%	41%
	CD3+ IT-CD-8	probability for DC = 1 / (1 - exp (-2.3821 + 0.0806 * CD3 + 0.0175 * IT-CD8 + 0.0069 * CD3 * IT-CD8)	0.161	*53*	0.78 (0.65-0.91)	94%	54%	95%	47%		*74*	0.68 (0.55-0.80)	83%	46%	85%	43%
DC 12 months	PD-1^{T} TILs		90	*42*	0.82 (0.70-0.94)	100%	64%	100%	43%		*61*	0.80 (0.65-0.94)	86%	74%	95%	50%
	TIS		6.65	*28*	0.77 (0.58-0.96)	100%	50%	100%	35%		*40*	0.63 (0.43-0.82)	100%	39%	100%	26%
	CD8+ IT-CD-8	probability for DC = 1 / (1 - exp (-4.0644 + 0.003 * CD8 + 0.0436 * IT-CD8)	0.122	*55*	0.85 (0.73-0.96)	92%	63%	96%	41%		*77*	0.67 (0.53-0.81)	68%	57%	88%	30%
	CD8+ TIS	probability for DC = 1 / (1 - exp (-5.7952 + 0.0224 * CD8 + 0.2346 * TIS + -0.0021 * CD8 * TIS)	0.124	*28*	0.91 (0.79-1.00)	100%	68%	100%	46%		*38*	0.59 (0.36-0.82)	29%	68%	81%	17%

Figure 1

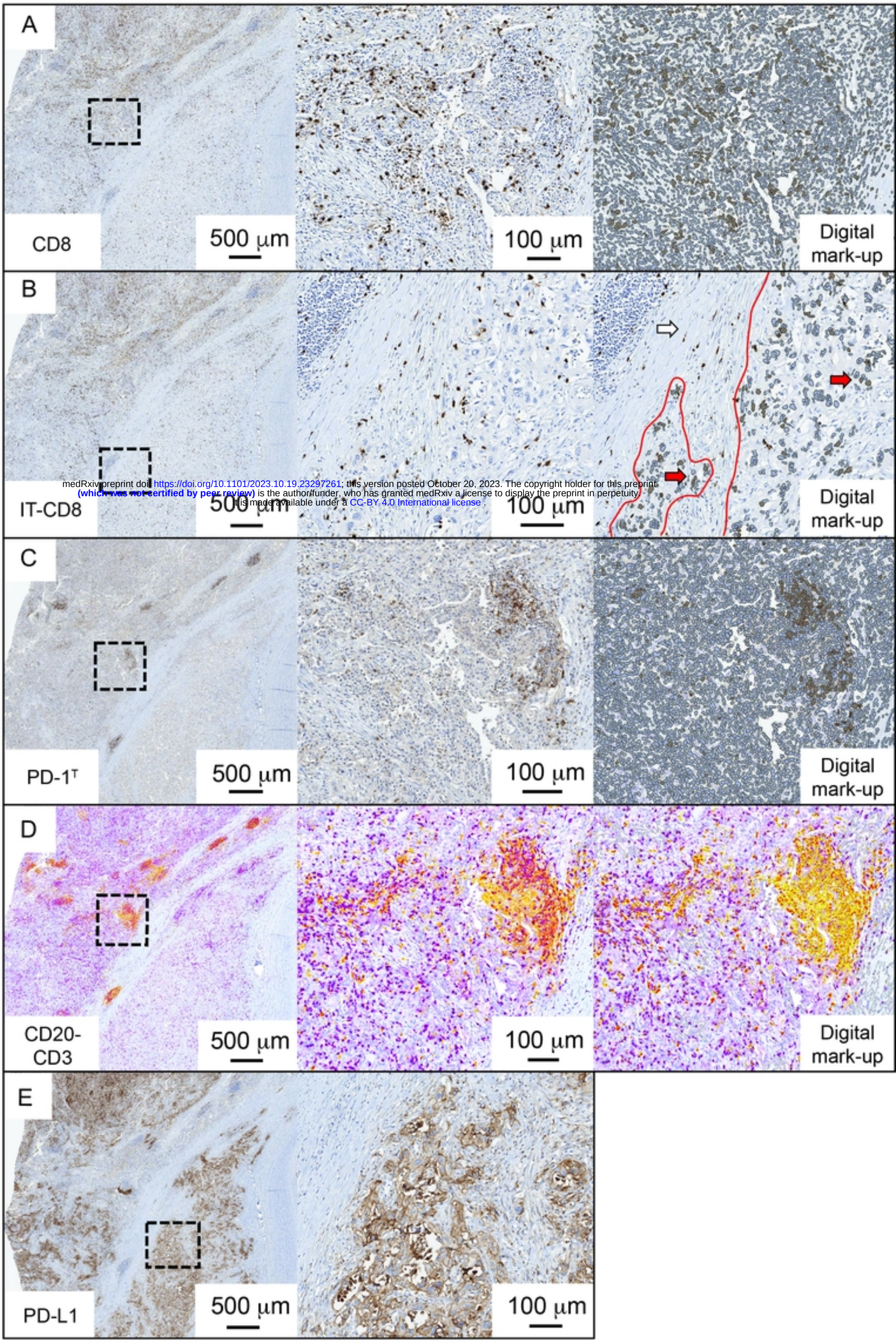
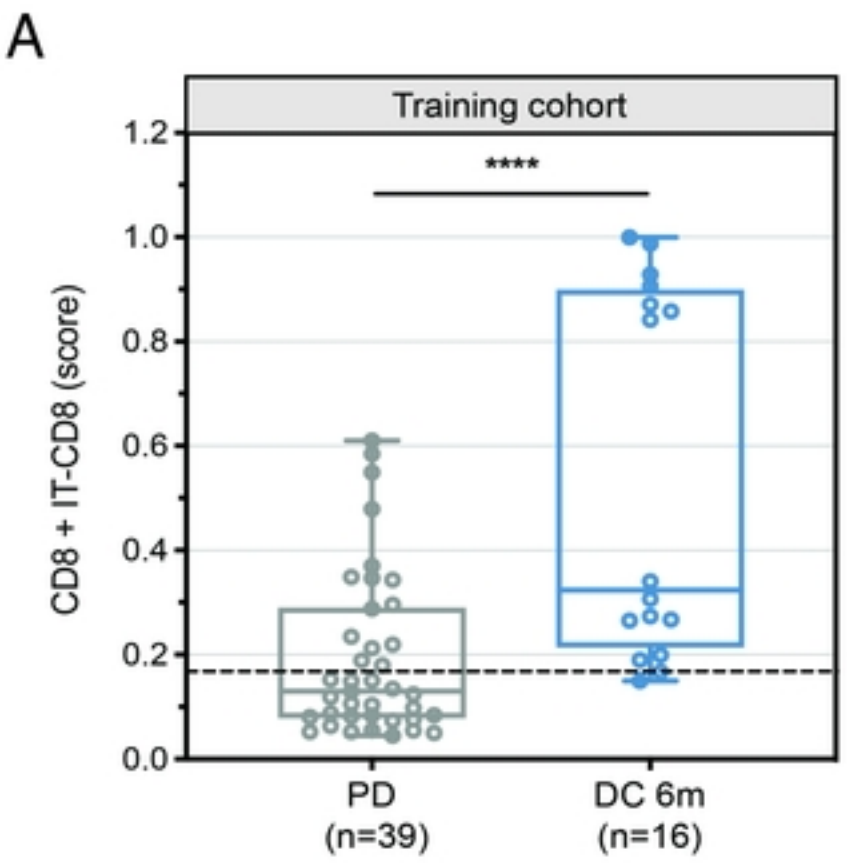
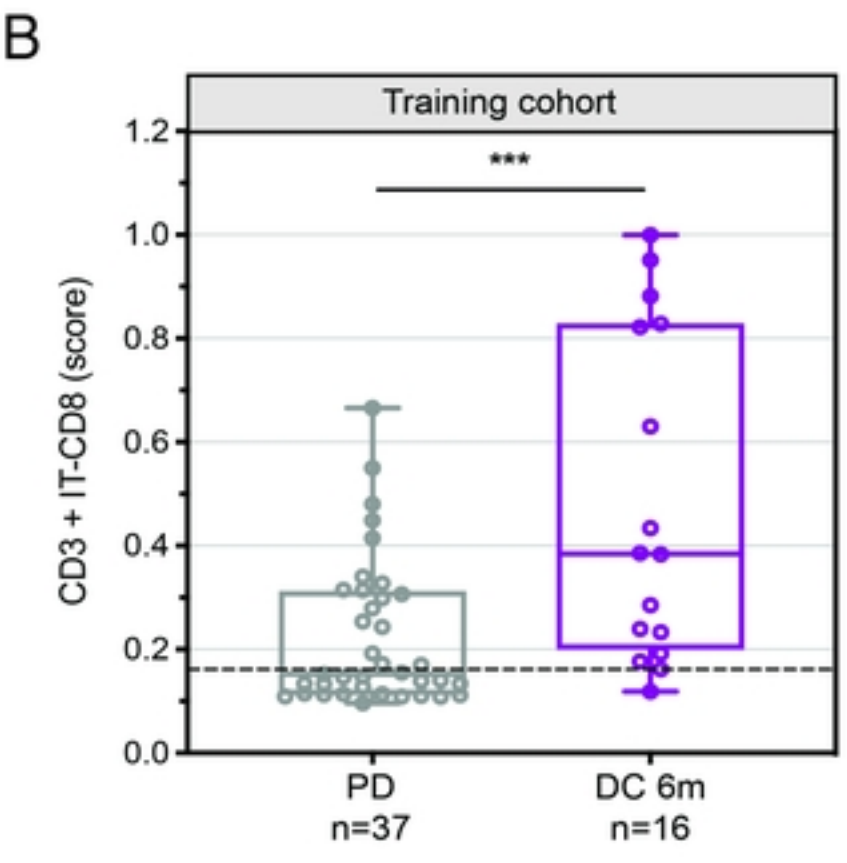
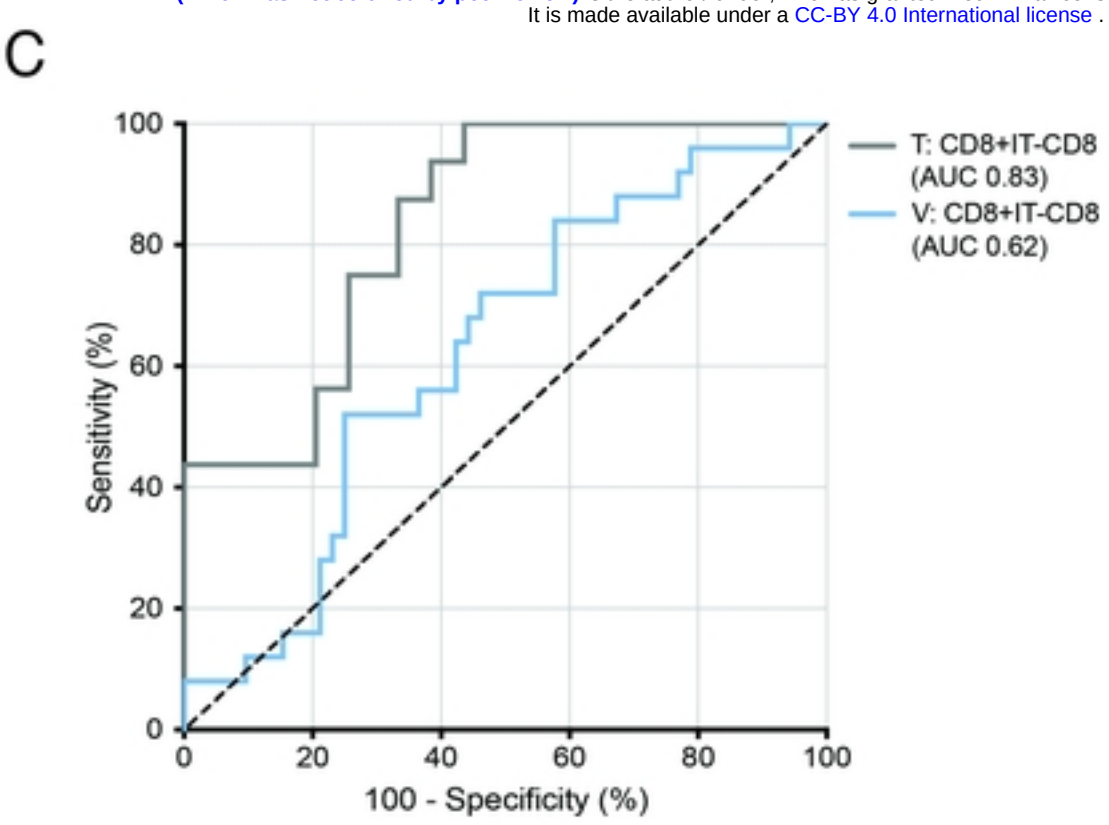


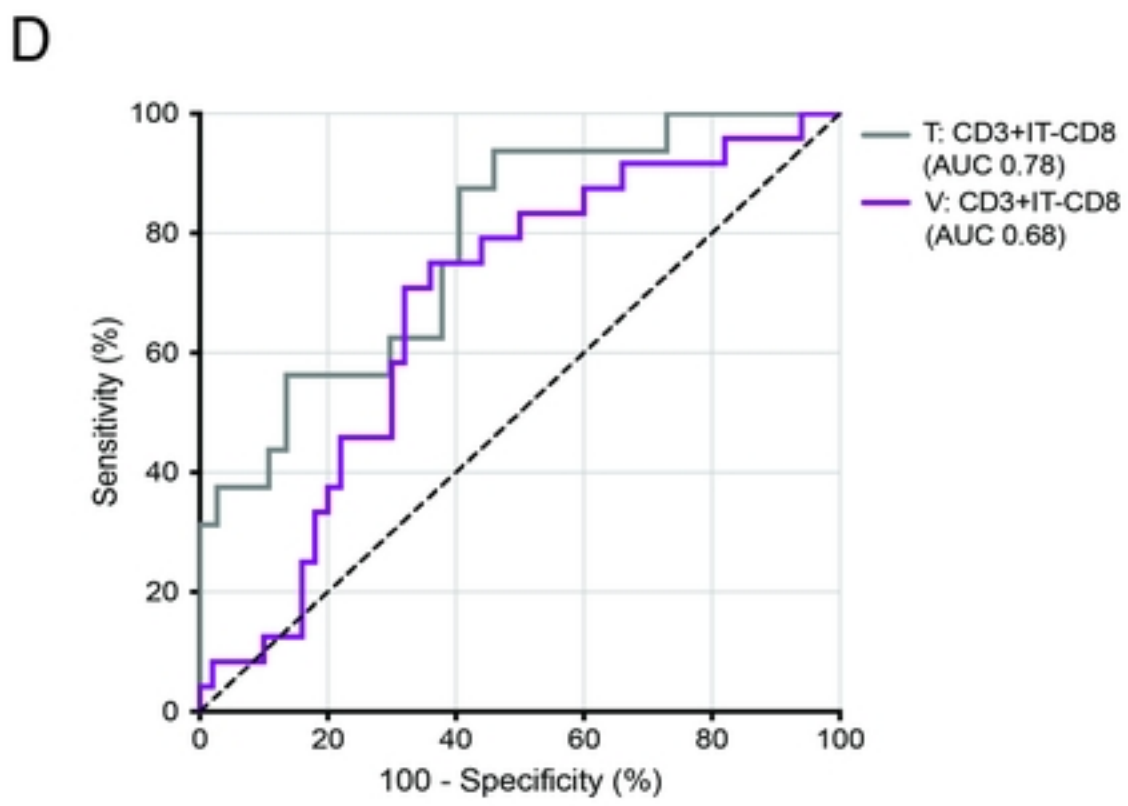
Figure 1

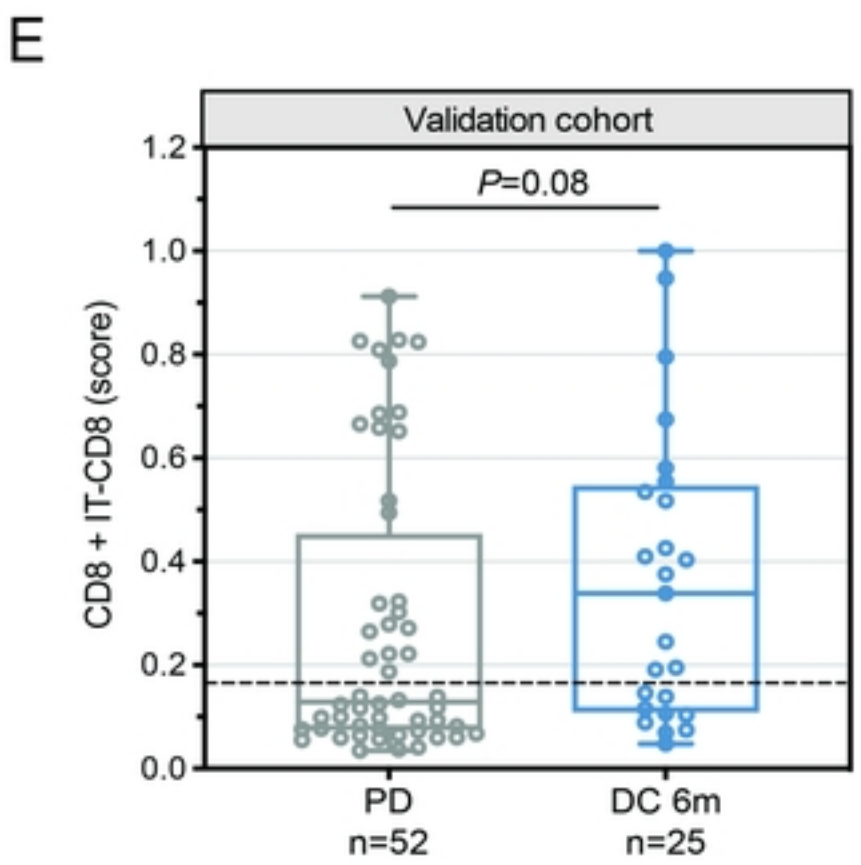
Figure 2











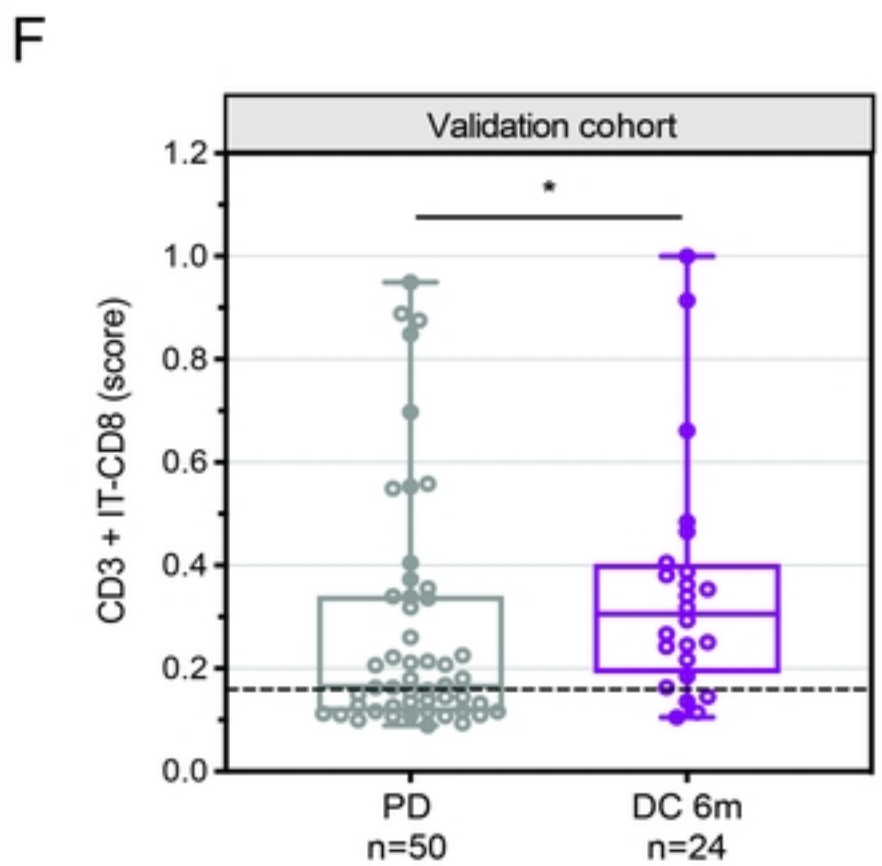


Figure 2

Figure 3

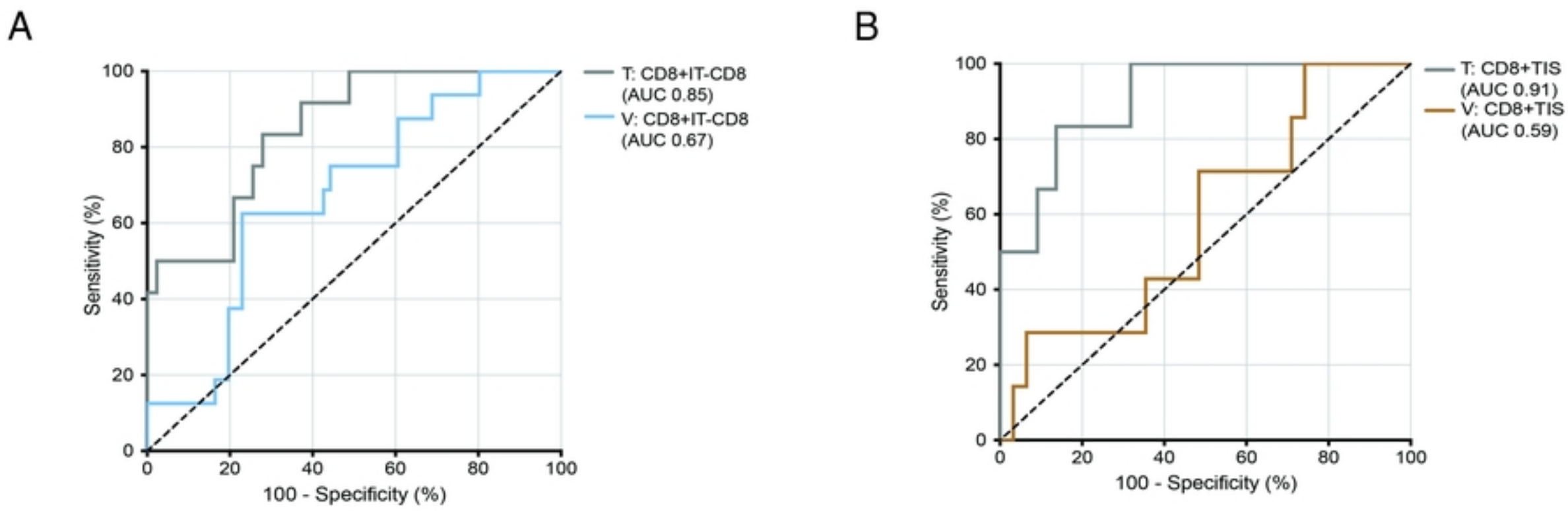


Figure 3