Biomarkers for clinical benefit to immune checkpoint blockade treatment in NSCLC

Karlijn Hummelink

Biomarkers for clinical benefit to immune checkpoint blockade treatment in NSCLC

Karlijn Hummelink

Biomarkers for clinical benefit to immune checkpoint blockade treatment in NSCLC

Biomarkers voor het voorspellen van therapeutisch voordeel van immuun checkpoint blokkade behandeling in NSCLC

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. H.R.B.M. Kummeling,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen
op dinsdag 12 november 2024 des middags te 2.15 uur

door

Karlijn Hummelink

geboren op 10 januari 1990
te Lichtenvoorde

Promotoren:
Prof. dr. G.A. Meijer
Prof. dr. E.F. Smit

Copromotoren:
Dr. K. Monkhorst
Dr. D.S. Thommen

Beoordelingscommissie
Dr. D. Cohen
Prof. dr. P.J. van Diest
Prof. dr. M.M. van den Heuvel
Prof. dr. W.L. de Laat
Prof. dr. E.E. Voest (voorzitter)

"Nae vragt, kom iej neet wies"

Voor José, mijn moeder

Scan this QR code to listen to the podcast
that Oncologie Up-to-Date recorded
with Karlijn Hummelink about this thesis.

Table of Contents

Chapter 1

General introduction
and thesis outline

General introduction and thesis outline

Epidemiology of lung cancer

Lung cancer constitutes a significant global health challenge, with an estimated 2.2 million new cancer cases and 1.8 million deaths reported worldwide in 2020. It stands as the second most frequently diagnosed cancer and is the leading cause of cancer-related mortality[1]. In the Netherlands, approximately 14.000 new cases of lung cancer were diagnosed in the year 2019, and the mortality rate exceeded 10.000 deaths attributed to lung cancer during the same period[2]. Tobacco smoking has been associated to the development of lung cancer in 80% of cases[3]. Over the decades, trends in smoking have influenced the incidence rates on different continents. For example, in the United States, the incidence rate declined from its peak of 67 per 100.000 individuals in 1992 to 43.2 per 100.000 in 2018[4], primarily due to increased smoking cessation efforts. Approximately 85% of lung cancer patients form a group of histological subtypes collectively known as non-small cell lung cancer (NSCLC)[2]. The two most common subtypes are adenocarcinoma (AD) and squamous cell carcinoma (SCC)[5]. In cases where the morphology of the tumor does not show evidence for either AD or SCC, a positive P40 (>50% of tumor cells) and negative TTF1 immunohistochemical (IHC) staining can confirm squamous cell differentiation. Otherwise, the tumor is classified as NSCLC, not otherwise specified (NOS)[5]. Almost 50% of patients are diagnosed with advanced stage disease[2] that is not curable by surgery alone, leaving systemic therapies as treatment of choice.

Treatment of advanced NSCLC

Historically, the treatment of advanced stage NSCLC has been chemotherapy consisting of a platinum doublet with either carboplatin or cisplatin with gemcitabine, pemetrexed, vinorelbine, or taxanes (paclitaxel or docetaxel). No clinically meaningful differences in outcome have been found among these cytotoxic regimens[6], with the exception of the combination pemetrexed-cisplatin which showed shorter overall survival (OS) compared to the gemcitabine-cisplatin combination in SCC[7]. The treatment landscape of NSCLC dramatically changed after the development of specific targeted therapies for the treatment of *e.g.*, *EGFR*-mutant, *ALK*-rearranged, *ROS1*-rearranged or *BRAF*[V600E]-mutant advanced-stage NSCLC. Several tyrosine kinase inhibitors (TKIs) targeting these molecular alterations have led to remarkable responses in selected patients[8].

More recently, immunotherapy, particularly immune checkpoint blockade (ICB), has introduced a new era in lung cancer care. In the tumor microenvironment (TME), activated T cells express a protein called programmed cell death 1 (PD-1). When a T

cell recognizes a specific tumor antigen presented by the major histocompatibility complex (MHC), it triggers the production of inflammatory cytokines. These cytokines can induce overexpression of programmed cell death 1 ligand 1 (PD-L1) within the tumor. The interaction between PD-L1 and the PD-1 receptor on T cells results in T cell dysfunction and subsequently immunotolerance. Consequently, tumors can protect themselves from cytotoxic (CD8[+]) T cell-mediated cell killing. Blocking the interaction between PD-1 and PD-L1 can offer an approach to restore T cell-mediated antitumor immunity[9]. The first evidence demonstrating the effectiveness of anti-PD-(L)1 antibodies in treating NSCLC came from studies involving patients with previously treated advanced NSCLC[10–12]. Since then, these treatments rapidly transitioned to the first-line treatment setting, as multiple clinical trials showed a significant improvement in survival when compared to chemotherapy alone[13–16]. Unfortunately, approximately 60-70% of patients experience disease progression within six months after initiating treatment[14–16], underscoring the need for biomarkers to support shared decision-making for therapeutic strategies. Such biomarkers are essential to improve personalized medicine, to minimize patient exposure to potential adverse effects and to reduce healthcare costs.

Biomarkers for predicting response to systemic therapy in advanced NSCLC
Although long-lasting clinical responses have been observed for patients treated with TKIs or PD-(L)1 blockade, this only accounts for a minority of patients. Therefore, biomarkers are urgently needed to estimate the probability of response to specific systemic therapeutic regimens, herein referred to as predictive biomarkers. For TKIs, most predictive biomarkers are characterized by specific genomic alterations associated with the mode of action of the involved TKIs. As a result, comprehensive molecular profiling of tumors is now routinely advised for all newly diagnosed advanced-stage NSCLC patients. This practice allows personalized therapeutic strategies tailored to patients whose tumors harbor targetable oncogenes. However, in the case of immunotherapy, which serves to restore anti-tumor immunity, potential predictive biomarkers differ fundamentally from driver oncogene biomarkers. They exhibit a continuous spectrum rather than a binary categorization, display spatial and temporal variability, and arise from multiple determinants rather than a single dominant determinant. Therefore, biomarker development is more challenging within the context of immunotherapies. The anti-tumor immune response is a complex process, requiring several steps for effective priming, activation, trafficking, and recognition of T cells to eradicate cancer cells, a process known as the cancer-immunity cycle[17]. Therefore, it is unlikely to find one single biomarker that effectively predicts response[18]. In addition to predicting which patient will respond to treatment, it is also highly relevant to predict which patients will not respond to

treatment. Biomarkers with a high negative predictive value can reliably predict this lack of therapeutic benefit, thus holding paramount significance for preventing overtreatment. As described earlier, this type of biomarker not only minimizes the risk of unnecessary side effects but also serves to reduce health care costs. Moreover, it offers the possibility of providing patients with alternative treatment options at an early stage.

The initial biomarker examined for its predictive potential in advanced NSCLC patients treated with PD-(L)1 blockade was the assessment of PD-L1 expression in tumor tissue. This was done due to its function in the inhibitory PD-L1/PD-1 pathway that is targeted by this treatment. Several studies have shown a positive correlation between high PD-L1 expression and improved response rates and survival[13,19,20]. For example, the results of the phase III KEYNOTE-024 study demonstrated that pembrolizumab led to significantly prolonged progression-free and overall survival compared to platinum-based chemotherapy in patients with PD-L1 expression levels of ≥50%[13,20]. Subsequently, PD-L1 assessment via IHC received clinical approval as a predictive biomarker test. However, different studies have published conflicting results, as some patients with PD-L1 low or PD-L1 negative tumors have also shown long-term disease control with ICB agents[14–16]. Furthermore, PD-L1 expression levels can vary due to factors such as interassay variability[21], intratumor heterogeneity[22–25] and sample characteristics including age, biopsy site, and timing[26,27]. Another common hurdle is obtaining sufficient tumor tissue samples for PD-L1 testing, as the tumor site is often difficult to reach and invasive procedures are needed.

Tumor Mutation Burden (TMB), defined as the total number of nonsynonymous mutations per sequenced coding area of a tumor genome, has subsequently been studied as predictive biomarker for PD-(L)1 blockade monotherapy. It is thought that a higher TMB increases the likelihood of tumor neoantigen production and therefore, potential immunogenicity and the killing of cancer cells[28]. While TMB has shown predictive potential, no universally applicable TMB threshold has consistently demonstrated the ability to predict overall survival. Also, technical challenges have been reported due to variation across different sequencing platforms[29].

As PD-(L)1 blockade is thought to reinvigorate dysfunctional T cells[30], the presence of tumor-infiltrating lymphocytes (TILs) has been investigated as a predictive biomarker. Although TIL density has shown predictive potential[31,32], increasing evidence suggests that not all TILs are in a state to recognize and eliminate tumor cells[33,34]. Therefore, general TIL density is not an accurate predictive biomarker for ICB response. Previous work showed that CD8$^+$ TILs with high PD-1 expression, referred

to as PD-1[T] TILs, positively correlated to treatment outcome in a small cohort of NSCLC patients treated with PD-1 blockade monotherapy[35]. These TILs display a high capacity for tumor recognition, and express and secrete CXCL13, a B cell attractant essential for the formation of tertiary lymphoid structures (TLS). Notably, PD-1[T] TILs predominantly localize within TLS[35]. TLS and B cells, a critical TLS component, have been correlated with clinical responses in other tumor types[36–39], although comprehensive data from larger NSCLC patient cohorts is currently lacking.

The clinical implementation of the aforementioned biomarkers presents challenges due to their continuous nature. Hence, reliable and automated methods are preferred for the assessment and the definition of cut-off values. For example, digital quantification methods have been described for TILs and PD-L1[40–42]. Furthermore, robust platforms such as the Nanostring nCounter platform[43] enable the development of predictive mRNA signatures capable of extracting the immune phenotype from the TME. One example of such a signature is the tumor inflammation signature (TIS)[44–47]. In addition, the combination of biomarkers could be an approach to improve predictive accuracy, as demonstrated in studies combining TMB with PD-L1[48,49] and CD8 with PD-L1[18,31].

Alternative bio-sources for biomarker development
One of the predominant challenges in biomarker testing is the availability of sufficient tumor tissue, frequently necessitating invasive procedures. Furthermore, local tissue sampling can introduce biases due to tumor heterogeneity. Notably, more and more tissue material is required to comply to the increasing number of diagnostic biomarker tests. In addition to tumor tissue, blood serves as a viable source for biomarkers. The liquid biopsy method that analyzes cell-free DNA in plasma can detect circulating tumor DNA (ctDNA). This method is minimally invasive and, in contrast to tissue biopsy, can provide a molecular profile of the tumor as well as real-time insights into tumor response dynamics during treatment. Nonetheless, sensitivity challenges persist, primarily because of the often limited fraction of plasma ctDNA available for analysis[50]. In addition, ctDNA biomarkers are less suitable for predicting responses to immunotherapy, as they do not capture characteristics of the immune infiltrate.

A variety of highly sensitive and specific technologies have been rapidly developed, primarily based on multiplex PCR (mPCR) or next-generation sequencing (NGS). These advancements enable the detection of genetic alterations in circulating nucleic acids, encompassing gene mutations, fusions, deletions, amplifications, translocations and epigenetic changes[51]. Beyond nucleic acid-based approaches,

proteomic-based profiling in liquid biopsies holds significant promise. Given that proteins represent the direct drug targets of many cancer therapies, including ICB, high dimensional proteomic data can be used for biomarker discovery.

Furthermore, other body fluids, such as pleural effusion, ascites and cerebrospinal fluid, can also serve as alternative source for biomarker identification[52]. Among these, pleural effusion stands out as an attractive bio-source for molecular profiling in NSCLC, particularly considering that approximately 30% of NSCLC patients develop malignant pleural effusion (MPE)[53]. Unfortunately, the quantity of tumor cells or the tumor cell percentage in MPE often proves insufficient for molecular analysis. Other studies have shown promising results by using cell-free DNA (cfDNA) from the supernatant of MPE[54-59].

Outline of this thesis

In spite of the impressive results observed with checkpoint inhibitor-based immunotherapies in a small subset of patients with advanced NSCLC, robust predictive biomarkers are still lacking. Notably, there are no predictive biomarkers capable of accurately identifying patients who do not derive clinical benefit to ICB treatment. In this thesis, we examined direct effectors of the anti-tumor immune response as potential biomarkers to predict response and non-response to PD-1 blockade monotherapy in advanced NSCLC. Additionally, we explored alternative bio-sources for biomarker assessment to avoid the need for invasive and complicated biopsy procedures, particularly in cases where tumor tissue is not available.

In **Chapter 2** we examine the accuracy of a tumor-reactive TIL population, known as PD-1^T TILs, as predictive biomarker in a cohort of 120 advanced stage NSCLC patients treated with PD-1 blockade. The frequency of PD-1^T TILs was quantified using digital image analysis. Additional exploratory analyses addressed the impact of lesion-specific responses, tissue sample properties, and the combination of PD-1^T TILs with other biomarkers on their predictive value.

Chapter 3 describes a study that investigates whether the predictive performance of biomarkers can be improved by combining them in pairs. The assessed biomarkers included both well-established ones, such as PD-L1, CD8/CD3 TILs and TIS, as well as recently developed biomarkers like PD-1^T TILs, CD20$^+$ B cells and TLS. All these biomarkers are known for their pivotal roles in the anti-tumor immune response upon PD-1 blockade monotherapy.

In **Chapter 4** we investigate whether a tumor's PD-1^T TIL status can be translated into an mRNA signature using the Nanostring nCounter platform. As digital quantification of PD-1^T TILs requires a substantial user interaction, a PD-1^T mRNA signature, developed on a robust clinical grade platform, will facilitates its implementation in a clinical setting. This study develops and validates a PD-1^T mRNA signature using gene expression data from 100 advanced NSCLC patients treated with PD-1 blockade from two independent cohorts.

In **Chapter 5** we highlight the use of liquid biopsies as an alternative bio-source, obviating the need for tissue biopsies. This study develops and validates a serum-derived protein signature designed to predict durable clinical benefit in 289 advanced stage NSCLC patients treated with PD-1 blockade.

1

In **Chapter 6** we describe the diagnostic yield of cell free DNA (cfDNA) extracted from pleural effusion as an alternative bio-source for molecular profiling of tumors. Currently, molecular analysis has become the mainstay in diagnostics to guide treatment selection and monitoring in advanced NSCLC. This study highlights the potential of cfDNA analysis in pleural effusion, even when tumor cell purity is low.

Last, **chapter 7** presents a general discussion of all chapters, providing a holistic view of the research findings and their implications.

References

1. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).
2. IKNL. *Integraal Kankercentrum Nederland,* https://www.iknl.nl/ (2019).
3. Alberg, A. J., Brock, M. V., Ford, J. G., Samet, J. M. & Spivack, S. D. Epidemiology of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American college of chest physicians evidence-based clinical practice guidelines. *Chest* **143**, e1S-e29S (2013).
4. Lungb @ Seer.Cancer.Gov.
5. International Agency for Research on Cancer, World health organization, I. A. of P. *Thoracic Tumours: WHO classification of Tumours.* (2021).
6. Schiller, J. H. *et al.* Comparison of Four Chemotherapy Regimens for Advanced Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **346**, 92–98 (2002).
7. Scagliotti, G. V. *et al.* Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer. *J. Clin. Oncol.* **26**, 3543–3551 (2008).
8. Tan, A. C. & Tan, D. S. W. Targeted Therapies for Lung Cancer Patients With Oncogenic Driver Molecular Alterations. *J. Clin. Oncol.* **40**, 611–625 (2022).
9. Alsaab, H. O. *et al.* PD-1 and PD-L1 checkpoint signaling inhibition for cancer immunotherapy: mechanism, combinations, and clinical outcome. *Front. Pharmacol.* **8**, 1–15 (2017).
10. Topalian, S. L. new england journal. 2443–2454 (2012).
11. Brahmer, J. R. *et al.* Phase I study of single-agent anti-programmed death-1 (MDX-1106) in refractory solid tumors: Safety, clinical activity, pharmacodynamics, and immunologic correlates. *J. Clin. Oncol.* **28**, 3167–3175 (2010).
12. Brahmer, J. R. *et al.* Safety and Activity of Anti–PD-L1 Antibody in Patients with Advanced Cancer. *N. Engl. J. Med.* **366**, 2455–2465 (2012).
13. Reck, M. *et al.* Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **375**, 1823–1833 (2016).
14. Borghaei, H. *et al.* Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N. Engl. J. Med.* **373**, 1627–1639 (2015).
15. Brahmer, J. *et al.* Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N. Engl. J. Med.* **373**, 123–135 (2015).
16. Rittmeyer, A. *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* **389**, 255–265 (2017).
17. Chen, D. S. & Mellman, I. Elements of cancer immunity and the cancer-immune set point. *Nature* **541**, 321–330 (2017).
18. Roelofsen, L. M., Kaptein, P. & Thommen, D. S. Multimodal predictors for precision immunotherapy. *Immuno-Oncology Technol.* **14**, 100071 (2022).
19. Garon, E. B. *et al.* Pembrolizumab for the treatment of non-small-cell lung cancer. *N. Engl. J. Med.* **372**, 2018–2028 (2015).
20. Reck, M. *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.* **37**, 537–546 (2019).

21. Tsao, M. S. *et al.* PD-L1 Immunohistochemistry Comparability Study in Real-Life Clinical Samples: Results of Blueprint Phase 2 Project. *J. Thorac. Oncol.* **13**, 1302–1311 (2018).

22. Ilie, M. *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.* **27**, 147–153 (2016).

23. Gniadek, T. J. *et al.* Heterogeneous expression of PD-L1 in pulmonary squamous cell carcinoma and adenocarcinoma: Implications for assessment by small biopsy. *Mod. Pathol.* **30**, 530–538 (2017).

24. Haragan, A. *et al.* Heterogeneity of PD-L1 expression in non-small cell lung cancer: Implications for specimen sampling in predicting treatment response. *Lung Cancer* **134**, 79–84 (2019).

25. McLaughlin, J. *et al.* Quantitative assessment of the heterogeneity of PD-L1 expression in non-small-cell lung cancer. *JAMA Oncol.* **2**, 46–54 (2016).

26. Boothman, A. M. *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.* **14**, 1390–1399 (2019).

27. Hong, L. *et al.* Programmed Death-Ligand 1 Heterogeneity and Its Impact on Benefit From Immune Checkpoint Inhibitors in NSCLC. *J. Thorac. Oncol.* **15**, 1449–1459 (2020).

28. Rizvi, N. A. *et al.* Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science (80-.).* **348**, 124–128 (2015).

29. Sholl, L. M. *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.* **15**, 1409–1424 (2020).

30. van der Leun, A. M., Thommen, D. S. & Schumacher, T. N. CD8+ T cell states in human cancer: insights from single-cell analysis. *Nat. Rev. Cancer* **20**, 218–232 (2020).

31. Fumet, J. D. *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* **119**, 950–960 (2018).

32. Hu-Lieskovan, S. *et al.* Tumor characteristics associated with benefit from pembrolizumab in advanced non–small cell lung cancer. *Clin. Cancer Res.* **25**, 5061–5068 (2019).

33. Simoni, Y. *et al.* Bystander CD8+ T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature* **557**, 575–579 (2018).

34. Scheper, W. *et al.* Low and variable tumor reactivity of the intratumoral TCR repertoire in human cancers. *Nat. Med.* **25**, 89–94 (2019).

35. Thommen, D. 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.* **24**, (2018).

36. Voabil, P. *et al.* An ex vivo tumor fragment platform to dissect response to PD-1 blockade in cancer. *Nat. Med.* **27**, 1250–1261 (2021).

37. Cabrita, R. *et al.* Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* (2020). doi:10.1038/s41586-019-1914-8

38. Helmink, B. A. *et al.* B cells and tertiary lymphoid structures promote immunotherapy response. *Nature* (2020). doi:10.1038/s41586-019-1922-8

39. Petitprez, F. *et al.* B cells are associated with survival and immunotherapy response in sarcoma. *Nature* **577**, (2020).

40. Amgad, M. *et al.* Report on computational assessment of Tumor Infiltrating Lymphocytes from the International Immuno-Oncology Biomarker Working Group. *npj Breast Cancer* **6**, (2020).

41. Steele, K. E. *et al.* Measuring multiple parameters of CD8+ tumor-infiltrating lymphocytes in human cancers by image analysis. *J. Immunother. Cancer* **6**, 1–14 (2018).

42. Koelzer, V. H. *et al.* Digital image analysis improves precision of PD-L1 scoring in cutaneous melanoma. *Histopathology* **73**, 397–406 (2018).

43. Veldman-Jones, M. H. *et al.* Evaluating robustness and sensitivity of the nanostring technologies ncounter platform to enable multiplexed gene expression analysis of clinical samples. *Cancer Res.* **75**, 2587–2593 (2015).

44. Ayers, M. *et al.* blockade IFN- γ – related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Invest.* **127**, 2930–2940 (2017).

45. Cristescu, R. *et al.* Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science (80-.).* **362**, (2018).

46. Damotte, D. *et al.* The tumor inflammation signature (TIS) is associated with anti-PD-1 treatment benefit in the CERTIM pan-cancer cohort. *J. Transl. Med.* **17**, 1–10 (2019).

47. Mints, M. *et al.* Tumour inflammation signature and expression of S100A12 and HLA class I improve survival in HPV-negative hypopharyngeal cancer. *Sci. Rep.* **11**, 1–11 (2021).

48. Rizvi, H. *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.* **36**, 633–641 (2018).

49. Hellmann, M. D. *et al.* Genomic Features of Response to Combination Immunotherapy in Patients with Advanced Non-Small-Cell Lung Cancer. *Cancer Cell* **33**, 843-852.e4 (2018).

50. Bettegowda, C. *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **6**, 1–12 (2014).

51. Nikanjam, M., Kato, S. & Kurzrock, R. Liquid biopsy: current technology and clinical applications. *J. Hematol. Oncol.* **15**, 1–14 (2022).

52. Siravegna, G., Marsoni, S., Siena, S. & Bardelli, A. Integrating liquid biopsies into the management of cancer. *Nat. Rev. Clin. Oncol.* **14**, 531–548 (2017).

53. Froudarakis, M. E. Pleural effusion in lung cancer: More questions than answers. *Respiration* **83**, 367–376 (2012).

54. Lin, J. *et al.* Detection of *EGFR* mutation in supernatant, cell pellets of pleural effusion and tumor tissues from non-small cell lung cancer patients by high resolution melting analysis and sequencing. *Int. J. Clin. Exp. Pathol.* **7**, 8813–8822 (2014).

55. Liu, D. *et al.* Malignant pleural effusion supernatants are substitutes for metastatic pleural tumor tissues in egfr mutation test in patients with advanced lung adenocarcinoma. *PLOS One* **9**, 1–5 (2014).

56. Lee, J. S. *et al.* Liquid biopsy using the supernatant of a pleural effusion for *EGFR* genotyping in pulmonary adenocarcinoma patients: A comparison between cell-free DNA and extracellular vesicle-derived DNA. *BMC Cancer* **18**, 1–8 (2018).

57. Asaka, S. *et al.* Rapid point-of-care testing for epidermal growth factor receptor gene mutations in patients with lung cancer using cell-free DNA from cytology specimen supernatants. *Int. J. Oncol.* **52**, 2110–2118 (2018).

58. Kawahara, A. *et al.* A Combined test using both cell sediment and supernatant cell-free DNA in pleural effusion shows increased sensitivity in detecting activating *EGFR* mutation in lung cancer patients. *Cytopathology* **29**, 150–155 (2018).

59. Shin, S., Kim, J., Kim, Y., Cho, S. M. & Lee, K. A. Assessment of real-time PCR method for detection of *EGFR* mutation using both supernatant and cell pellet of malignant pleural effusion samples from non-small-cell lung cancer patients. *Clin. Chem. Lab. Med.* **55**, 1962–1969 (2017).

PART 1

PD-1$^{\text{T}}$ TILs as precision
and clinical applicable
biomarker in NSCLC

Chapter 2

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[2,11], Annegien Broeks[5], John B. A. G. Haanen[9], Ton N. Schumacher[10], Gerrit A. Meijer[1], Egbert F. Smit[2,12], Kim Monkhorst[1]* and Daniela S. Thommen[9]*

These authors jointly supervised this work

[1]Department of Pathology, Division of Diagnostic Oncology, [2]Department of Thoracic Oncology, Division of Medical Oncology, [3]Department of Biometrics, [4]Department of Radiology, Division of Diagnostic Oncology, [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, [11]Present address: Department of Pulmonary Diseases, Radboud University Medical Center, Nijmegen, The Netherlands, [12]Present address: Department of Pulmonary Diseases, Leiden University Medical Center, Leiden, The Netherlands.

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 anti-tumor immune response. We previously identified a tumor-reactive T cell population, termed PD-1^T TILs, with predictive potential in a small cohort of non-small cell lung cancer (NSCLC) patients. 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 to PD-L1 and TLS. 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.

Abstract

Purpose
Durable clinical benefit to PD-1 blockade in 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.

Methods
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 TLS in the same cohort.

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

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 advanced stage non-small cell lung cancer (NSCLC) patients. 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 anti-tumor 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 NSCLC patients 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, 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 immunological 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)[10], which have been correlated with clinical and immunological 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

2

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 NSCLC patients treated with PD-1 blockade to (1) train and validate PD-1^T TILs as a predictive biomarker, (2) explore whether certain sample characteristics such as sample type, sample location or time of sampling influence the predictive value of this biomarker, and (3) evaluate the potential for clinical implementation, by comparing and combining PD-1^T TILs with other predictive markers such as PD-L1 and TLS.

Methods

Patient enrollment and study endpoints

In this study, 164 stage IV NSCLC patients 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, while in 19 patients the mutation status was unknown. Patients received single agent nivolumab 3 mg/kg, administered as an IV infusion, every two weeks for at least one dose or single agent pembrolizumab 200 mg as an IV 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)[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.

Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1 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). 44 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, 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 histological 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) supplemented with 1% Penicillin-Streptomycin (Roche), 12.6μg/ml Pulmozyme (Roche) and 1mg/ml Collagenase type IV (Sigma), as described previously[14]. Samples were then washed in PBS (Sigma), filtered over a 150μM filter mesh, resuspended in 50μL PBS, and incubated with Fc receptor blocking agent (eBioscience) and with live/dead Zombie UV (Biolegend) for 20 min at 4°C. Cells were washed, resuspended in 50μl of staining buffer (PBS (Sigma), 0.5% bovine serum albumin (Sigma), 0.1% NaN3 (Invitrogen)) containing the below-described antibodies, and incubated for 20 min at 4°C. After washing twice, cells were taken up in 200μl IC Fixation Buffer (eBioscience) and incubated for 20 min. 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 was carried out on a BD LSR II SORP cell analyzer (BD Biosciences). Data was 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.).

Immunohistochemistry

Separate immunohistochemistry (IHC) stainings 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 number V0002089, Ready-to-Use, 16 minutes at RT, Roche Diagnostics (Cat. # 7099029001). PD-L1 was detected using clone 22C3 (1/40 dilution, 1 hour at RT, Agilent/DAKO) and CD68 was detected using clone KP1 (1/10000 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 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[T], PD-L1 and CD68 immunostainings were scanned at x20 magnification with a resolution of 0.50 per μm² using an Aperio slide AT2 scanner (Leica Biosystems). CD20-CD3 immunostaining was scanned at x20 magnification with a resolution of

0.24 per µm² 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 previously established[10]. For the current 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 to peripheral blood T cells as external reference to establish the threshold for tumor-associated PD-1 expression **(Fig. S1A)**. Next, a digital IHC algorithm to quantify PD-1+ lymphocytes in matched FFPE samples was generated **(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 to 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) **(Fig. S1C,D)** and was selected as the threshold for further automated PD-1^T 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 mm² tumor area was determined **(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. Receiver operator characteristic (ROC) curves were used in the training set to establish an optimal cut-off of 90 PD-1^T TILs per mm² 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 **(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 (ICs) of any intensity (IC0: <1%, IC1: ≥1% and <5%, IC2: ≥5% and <10% and IC3: ≥10%) as described[20,21].

Scoring of tertiary lymphoid structures

A CD20 (yellow)/CD3 (purple) double staining was used to identify tertiary lymphoid structures (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)[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 **(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 positive 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-positive area was normalized per mm^2 tumor area **(Table S2)**.

Statistical analysis

Patient characteristics were descriptively reported using mean ±s.d., interquartile range (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's 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 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.

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). Since 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 **Table S3**. Sample characteristics are shown in **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, Fig. S1** 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 6m 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 interquartile range (IQR) between 86 and 356 in the DC 6m group versus 51 (IQR: 28-84) in the PD group (P<0.01) **(Fig. 1C)**. To select the optimal biomarker cut-off, we performed a receiver operator characteristic (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% in order 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 6m versus 49 (IQR: 15-152) with PD (p<0.01) **(Fig. 1F)**. The AUC of the ROC curve was 0.72 (CI: 0.60-0.84), indicating a similar 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 non-responding patients in the PD-1^T low group **(Fig. 1H)**.

Assessment of secondary endpoints: DC at 12 months and survival
Since approximately 60-70% of patients treated in 2nd line with PD-(L)1 blockade progress within 6 months, and an additional 10-20% 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 to the DC 6m 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, **Fig. S2A,B**). 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 (CI: 0.73-1.00) (**Fig. 2A**, training set) and 0.78 (CI: 0.68-0.88) (**Fig. 2B**, validation set). Importantly, in the DC 12m 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 the validation set) (**Table 1**). Notably, in both data sets only 1/43 (2%) and 1/77 (1%) samples from patients with DC 12m 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,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. Since 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) **(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) **(Fig. S2D)** and validation set (HR 0.46; 95% CI: 0.28-0.76, *P*< 0.01) **(Fig. 2F)**.

A

B

PD-1^T TILs high

PD-1 IHC — 100 μm — Digital mark-up

PD-1^T TILs low

PD-1 IHC — 100 μm — Digital mark-up

C Training cohort

D

AUC: 0.79 (95% CI: 0.61 - 0.98)
90 per mm² cut-off
Sens: 79%   NPV: 89%
Spec: 83%   PPV: 69%

E

F Validation cohort

G

AUC: 0.72 (95% CI: 0.60 - 0.84)
90 per mm² cut-off
Sens: 77%   NPV: 88%
Spec: 67%   PPV: 49%
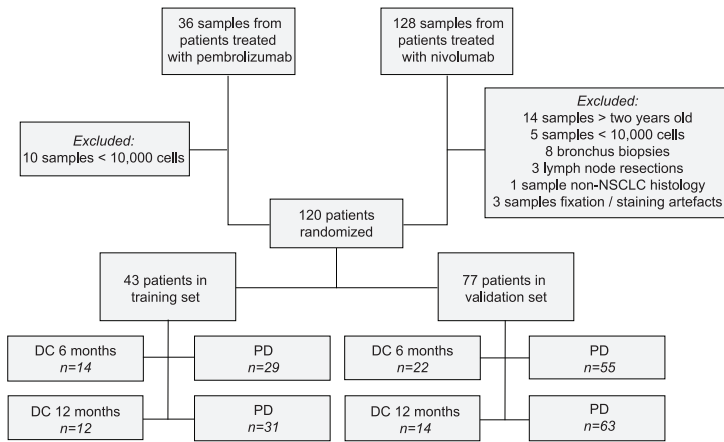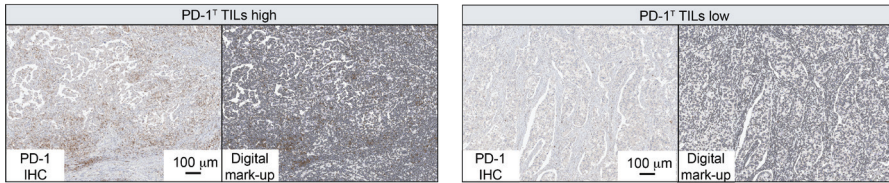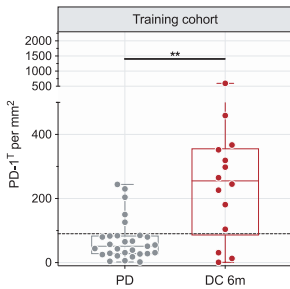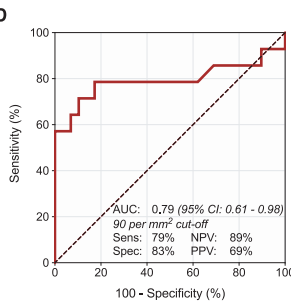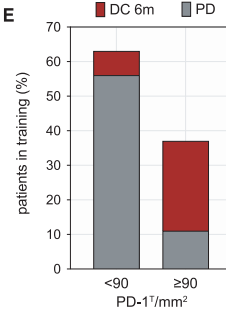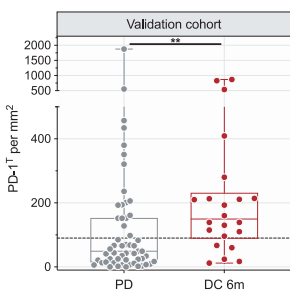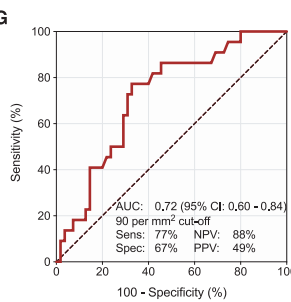
H
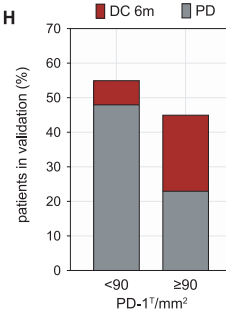
◄ **Figure 1.** PD-1^T TILs as biomarker for clinical outcome to PD-1 blockade in NSCLC.

(A) Study design for analysis of PD-1^T TILs in pretreatment samples from two retrospective stage IV NSCLC patient cohorts treated with PD-1 blockade. The training (n=43) and validation set (n=77) consisted each of 30% of patients with disease control at 6 months (DC 6m) 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 6m (n=14) and progressive disease (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)** Receiver operating characteristic (ROC) curve for predictive value of PD-1^T TILs for DC 6m (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 6m or PD. **(F-G)** Same plots as shown in **C** and **D** for patients with DC 6m (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).

Differences between lesion-specific and overall response

While 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 27/51 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 while 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 to 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 radiological 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.

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

	Clinical outcome	Biomarker	AUC
Training (n=43)	DC 6m	PD-1^T TILs per mm^2	0.79 95% CI: 0.61-0.98
Validation (n=77)	DC 6m	PD-1^T TILs per mm^2	0.72 95% CI: 0.60-0.84
		% PD-L1 TPS	0.58 95% CI: 0.43-0.74
Training (n=43)	DC 12m	PD-1^T TILs per mm^2	0.89 95% CI: 0.73-1.00
Validation (n=77)	DC 12m	PD-1^{T}TILs per mm^2	0.78 95% CI: 0.68-0.88
		% PD-L1 TPS	0.68 95% CI: 0.51-0.86

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 **(Tables S5,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,C)**. While 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, while PD-1^T TILs showed some intratumoral heterogeneity, the overall distribution could be captured by assessing relatively a small area of the tumor.

Cut-off	Sensitivity	Specificity	NPV	PPV
<90 vs ≥90	79%	83%	89%	69%
<90 vs ≥90	77%	67%	88%	49%
<1 vs ≥1	41%	67%	74%	33%
<50 vs ≥50	23%	95%	75%	63%
<90 vs ≥90	92%	84%	96%	69%
<90 vs ≥90	93%	65%	98%	37%
<1 vs ≥1	57%	70%	88%	30%
<50 vs ≥50	29%	94%	86%	50%

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,E)**. Next, we compared samples from primary and metastatic sites which performed similarly with respect to prediction of treatment outcome **(Fig. 3D,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 12m subgroup (AUC 0.91; 95% CI: 0.82-0.99 versus 0.74; 95% CI: 0.61-0.88, P=0.04) for samples taken prior to at least one other systemic treatment **(Fig. 3D,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.
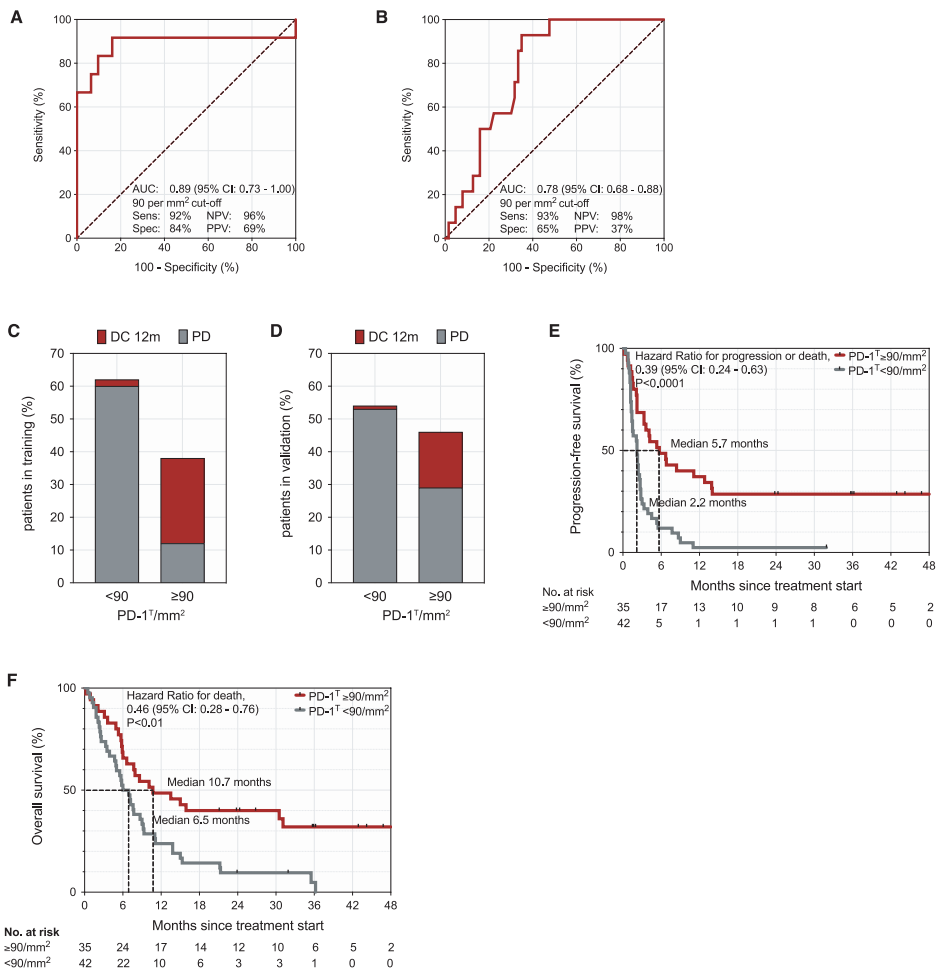
Figure 2. PD-1^T TILs can effectively discriminate patients with long-term benefit from patients with progressive disease. **(A)** ROC curve for predictive value of PD-1^T TILs for DC 12m in the training set (n=43) (AUC 0.89; 95% CI: 0.73-1.00) and **(B)** in the validation set (n=77) (AUC 0.78; 95% CI: 0.68-0.88). **(C)** Percentage of patients with PD-1^T high (≥90 per mm²) (n=16) and PD-1^T low (<90 per mm²) (n=27) pretreatment samples in the training set showing DC 12m 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)** Progression-free survival (PFS) of patients with PD-1^T high versus PD-1^T low pretreatment samples (median 5.7 versus 2.2 months, HR 0.39; 95% CI: 0.24-0.63, **** P<0.0001) in the validation set (n=77). **(F)** Overall survival (OS) (median 10.7 versus 6.5 month, 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.
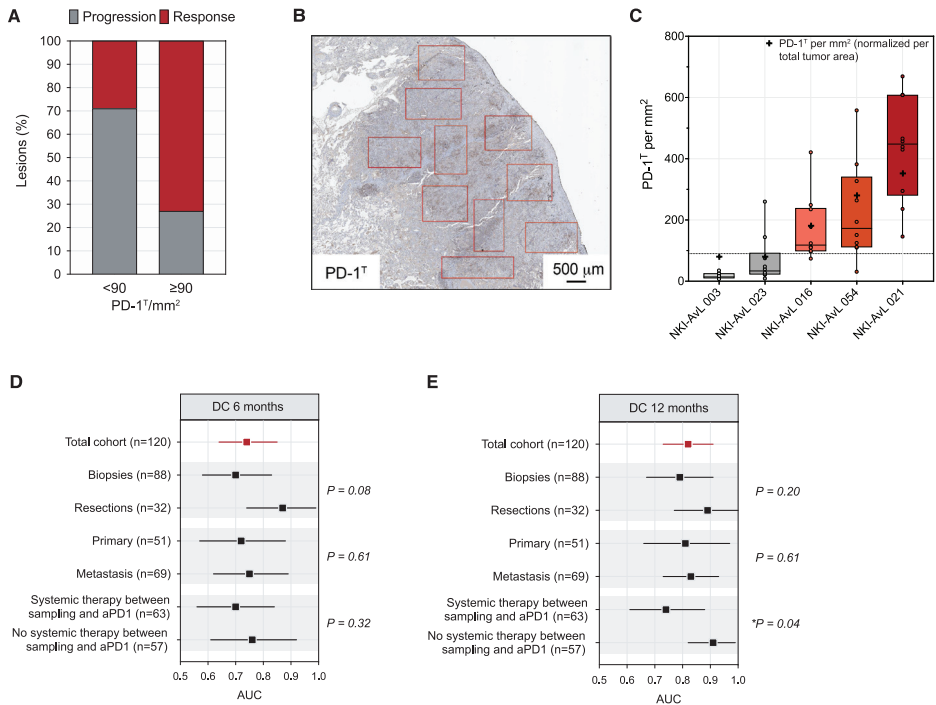
Figure 3. Impact of lesion-specific response and tissue sample properties on the predictive accuracy of PD-1^T TILs. **(A)** Percentage of responsive versus progressive lesions during treatment in the PD-1^T high (≥90 per mm²) (n=11) and PD-1^T low (<90 per mm²) (n=14) group of patients with PD within 12 months. A lesion was defined progressive when ≥20% growth was seen compared to the smallest diameter during treatment. **(B)** Example of a PD-1^T high IHC staining with 10 individually annotated tumor areas of 1 mm². **(C)** Quantification of PD-1^T TILs per each mm² area in five resection specimens. Each dot indicates an individual measurement. Two tumors are PD-1^T low (grey shades), three tumors are PD-1^T high (red shades). The cross indicates PD-1^T TILs per mm² normalized per total tumor area. **(D-E)** The predictive value of PD-1^T TILs in the total cohort and different subgroups. Each comparison is marked in a grey square. Shown is the area under the curve (AUC) for DC 6m **(D)** and 12m **(E)** with 95% CI interval. *P*-value was determined by one-sided permutation test.

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 tumor proportion score (TPS) in the validation set **(Fig. 4A)**. The fraction of disease control and PD for patients with tumors expressing ≥50%, 1-50% or no PD-L1 are shown in **Fig. S3A** (DC 6m) and

Fig. 4B (DC 12m). ROC analysis of PD-L1 TPS to predict DC 6m and DC 12m showed a lower AUC compared to PD-1^T TILs (0.58; 95% CI: 0.43-0.74, and 0.68; 95% CI: 0.51-0.86) **(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 to 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 previously been evaluated as biomarker cut-offs for treatment with nivolumab[3,6], did not improve prediction compared to PD-1^T TILs **(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 **(Fig. S3C,D)**.

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 vs 2.4 months, and HR 0.40; median OS 32.2 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,F**). Patients with PD-L1 TPS ≥1% showed slightly better PFS and OS in the validation set though without reaching significance which is comparable to other studies[3,5,6] **(Fig. 4G,H)**.

We noticed that the fraction of patients with PD-L1 TPS <1% observed here was higher than in previous studies (60% as compared to 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 **(Fig. S4A,B)**. Combining PD-L1 TPS at either 50% or 1% cut-off and PD-L1 IC≥2 indeed improved predictive accuracy, that is, compared to PD-L1 TPS≥50%, by reaching a similar sensitivity as PD-1^T TILs but still substantially lower specificity **(Fig. S4C,D, Table S8)**.

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 **(Fig. S5A)**. Combination of the two biomarkers naturally

partitioned the patient population into four groups: (1) PD-L1 low (<50% or <1%)+PD-1^T low, (2) PD-L1 low+PD-1^T high, (3) PD-L1 high (≥50% or ≥1%)+PD-1^T low and (4) PD-L1 high+PD-1^T high **(Fig. S5B)**. We observed an enrichment of PD patients in the <50% PD-L1+PD-1^T low group (35/38 when assessing DC 6m as clinical outcome, 38/38 for DC 12m). Patients with DC 6m or 12m were distributed over all 4 groups and 3 out of 4 groups, respectively **(Fig. S5C,D)**. A similar patient distribution was found for the combination with 1% PD-L1 TPS **(Fig. S5E,F).** The predictive value of PD-1^T+PD-L1 at 50% cut-off was comparable to PD-1^T alone with a few percent increase in sensitivity at the cost of a slightly lower specificity (**Fig. S5G, Table S9**, for details on prediction model see 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% (**Fig. S5G, 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 tertiary lymphoid structures
Tertiary lymphoid structures (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 μm² in the annotated tumor area, and as lymphoid aggregate (LA) when between 10,000 and 60,000 μm². To estimate the presence of B cells, we quantified the CD20-positive area per mm² **(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-positive area per mm² was low **(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-C)**. The frequency of PD-1^T TILs per mm² was significantly higher inside

TLS compared to tumor areas outside of TLS (P<0.0001) **(Fig. 5D)**. We found similar results when performing the same analysis including TLS+LA **(Fig. S6C)**. Of note, samples without TLS or LA showed only very low frequencies of PD-1^T TILs **(Fig. 5D, 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 **Fig. S6D** (DC 12m) and **Fig. S6E,F** (DC 6m). The AUC of TLS per mm^2 was 0.62 (95% CI: 0.47-0.76) for DC 12m **(Fig. 5F)** and 0.62 (95% CI: 0.49-0.76) for DC 6m **(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-positive area per mm^2 **(Fig. 5F, 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 to 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.

Figure 4. Association of PD-L1 with long-term benefit and survival compared to PD-1^T TILs. **(A)** Immunohistochemical (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 12m or PD in the validation set (n=77). **(C)** ROC curve for predictive value of PD-L1 TPS for DC 12m (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 6m and 12m in comparison to 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. ▶

A

TLS   TLS+LA   CD20

positive
negative

B

CD20-CD3   500 μm   100 μm

PD-1^T 500 μm   100 μm

C

CD20-CD3
spatial plot

PD-1^T
spatial plot

D

PD-1^T distribution

PD-1^T per mm^2

****

Inside
TLS   Outside
TLS   No
TLS

E

TLS and outcome

TLS per mm^2

p=0.054

PD   DC 12m

F

DC 12 months

PD-1^T per mm^2

TLS per mm^2

TLS+LA per mm^2

CD20 per mm^2

AUC

G

TLS and PD-1^T

TLS per mm^2

ns

<90   ≥90
PD-1^T/mm^2

H

Inside TLS

PD-1^T per mm^2

**

<90 of total   ≥90 of total
PD-1^T/mm^2

I

Outside TLS

PD-1^T per mm^2

****

<90 of total   ≥90 of total
PD-1^T/mm^2
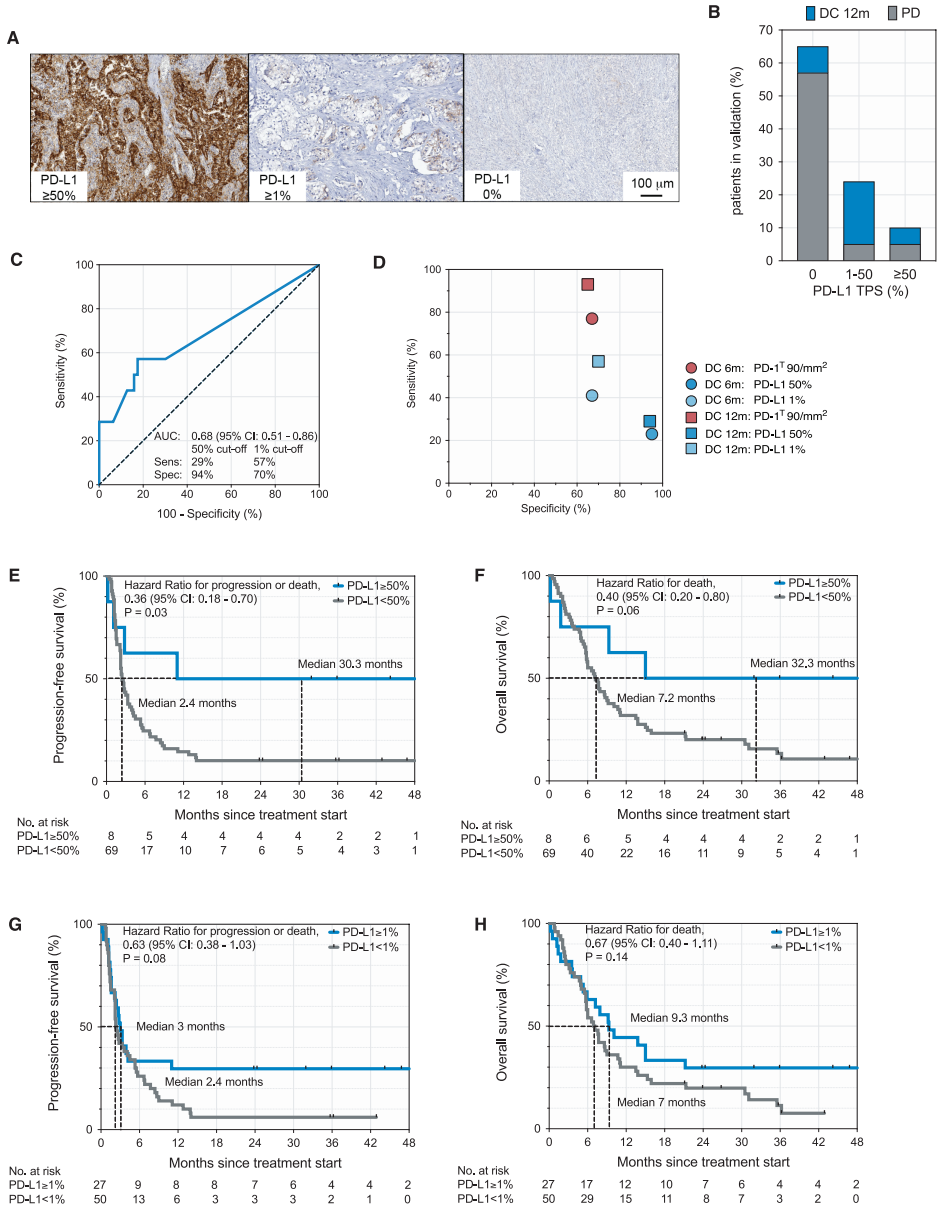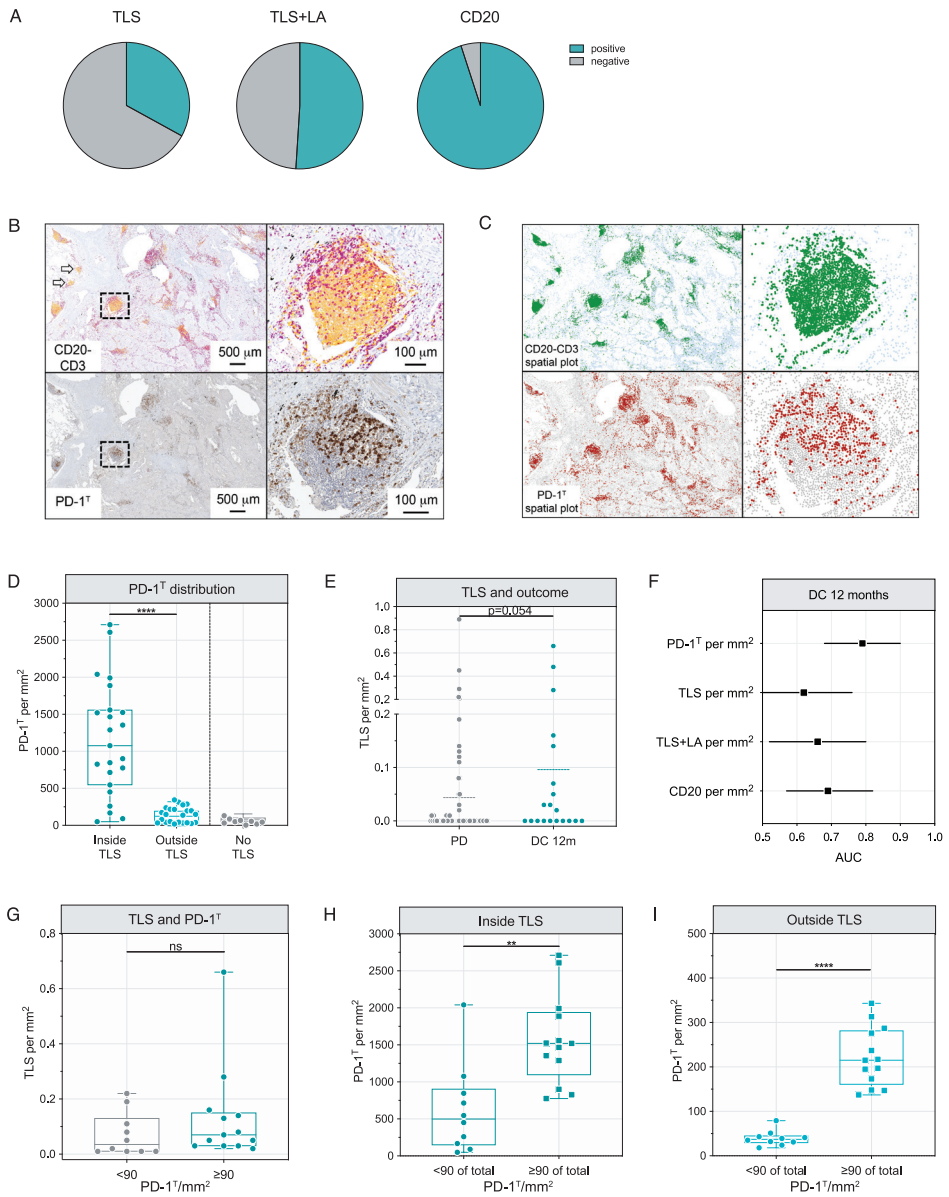
◄ **Figure 5.** PD-1^T high samples contain a higher density of PD-1^T TILs inside and outside TLS. **(A)** Percentage of pretreatment samples containing tertiary lymphoid structures (TLS) (n=30), TLS and/or lymphoid aggregates (LA) (referred as TLS+LA, n=46) and CD20+ B cells (n=86) in the remaining 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 grey). **(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 12m (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-positive area per mm^2 for DC 12m (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 and **** P<0.0001 by Mann-Whitney test.

Discussion

Monoclonal antibodies 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. Based on the rationale that the presence of tumor-reactive PD-1^T T lymphocytes is indicative of an ongoing anti-tumor 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]. While 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 anti-tumor 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 to 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 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 to 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 difference 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, since 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, since 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 the present study we found a lower predictive accuracy of TLS and LA compared to 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 PD-1 blockade, thus providing a tool to reduce overtreatment.

Acknowledgements

References

1. Garon, E. B. *et al.* Pembrolizumab for the treatment of non-small-cell lung cancer. *N. Engl. J. Med.* **372**, 2018–2028 (2015).

2. Reck, M. *et al.* Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **375**, 1823–1833 (2016).

3. Borghaei, H. *et al.* Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N. Engl. J. Med.* **373**, 1627–1639 (2015).

4. Postow, M. A., Sidlow, R. & Hellmann, M. D. Immune-related adverse events associated with immune checkpoint blockade. *N. Engl. J. Med.* **378**, 158–168 (2018).

5. Rittmeyer, A. *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* **389**, 255–265 (2017).

6. Brahmer, J. *et al.* Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N. Engl. J. Med.* **373**, 123–135 (2015).

7. Camidge, D. R., Doebele, R. C. & Kerr, K. M. Comparing and contrasting predictive biomarkers for immunotherapy and targeted therapy of NSCLC. *Nat. Rev. Clin. Oncol.* **16**, 341–355 (2019).

8. Reck, M. *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.* **37**, 537–546 (2019).

9. van der Leun, A. M., Thommen, D. S. & Schumacher, T. N. CD8+ T cell states in human cancer: insights from single-cell analysis. *Nat. Rev. Cancer* **20**, 218–232 (2020).

10. Thommen, D. 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.* **24**, (2018).

11. Oliveira, G. *et al.* Phenotype, specificity and avidity of antitumour CD8+ T cells in melanoma. *Nature* **596**, 119–125 (2021).

12. Caushi, J. X. *et al.* Transcriptional programs of neoantigen-specific TIL in anti-PD-1-treated lung cancers. *Nature* **596**, (Springer US, 2021).

13. Voabil, P. *et al.* An ex vivo tumor fragment platform to dissect response to PD-1 blockade in cancer. *Nat. Med.* **27**, 1250–1261 (2021).

14. Cabrita, R. *et al.* Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* (2020). doi:10.1038/s41586-019-1914-8

15. Helmink, B. A. *et al.* B cells and tertiary lymphoid structures promote immunotherapy response. *Nature* (2020). doi:10.1038/s41586-019-1922-8

16. Petitprez, F. *et al.* B cells are associated with survival and immunotherapy response in sarcoma. *Nature* **577**, (2020).

17. Theelen, W. S. M. E. *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.* **5**, 1276–1282 (2019).

18. Batenchuk, 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.* **71**, 1078–1083 (2018).

19. Krigsfeld, G. S. *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.* **73**, 656–664 (2020).

20. Herbst, R. S. *et al.* Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature* **515**, 563–567 (2014).

21. Fehrenbacher, L. *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* **387**, 1837–1846 (2016).

22. Gu-Trantien, C. *et al.* Quantifying Tertiary Lymphoid Structure-Associated Genes in Formalin-Fixed Paraffin-Embedded Breast Cancer Tissues. *Methods Mol. Biol.* **1845**, 139–157 (2018).

23. Buisseret, L. *et al.* Reliability of tumor-infiltrating lymphocyte and tertiary lymphoid structure assessment in human breast cancer. *Mod. Pathol.* **30**, 1204–1212 (2017).

24. Osorio, J. C. *et al.* Lesion-level response dynamics to programmed cell death protein (PD-1) blockade. *J. Clin. Oncol.* **37**, 3546–3555 (2019).

25. Eisenhauer, E. A. *et al.* New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur. J. Cancer* **45**, 228–247 (2009).

26. Boothman, A. M. *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.* **14**, 1390–1399 (2019).

27. Ilie, M. *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.* **27**, 147–153 (2016).

28. Gniadek, T. J. *et al.* Heterogeneous expression of PD-L1 in pulmonary squamous cell carcinoma and adenocarcinoma: Implications for assessment by small biopsy. *Mod. Pathol.* **30**, 530–538 (2017).

29. Hu-Lieskovan, S. *et al.* Tumor characteristics associated with benefit from pembrolizumab in advanced non–small cell lung cancer. *Clin. Cancer Res.* **25**, 5061–5068 (2019).

30. Hellmann, M. D. *et al.* Genomic Features of Response to Combination Immunotherapy in Patients with Advanced Non-Small-Cell Lung Cancer. *Cancer Cell* **33**, 843-852.e4 (2018).

31. Fumet, J. D. *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* **119**, 950–960 (2018).

32. Althammer, S. *et al.* Automated image analysis of NSCLC biopsies to predict response to anti-PD-L1 therapy. *J. Immunother. Cancer* **7**, 1–12 (2019).

33. Rizvi, H. *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.* **36**, 633–641 (2018).

34. Sautès-Fridman, C. *et al.* Tertiary lymphoid structures in cancers: Prognostic value, regulation, and manipulation for therapeutic intervention. *Front. Immunol.* **7**, 1–11 (2016).

35. Sautès-Fridman, C., Petitprez, F., Calderaro, J. & Fridman, W. H. Tertiary lymphoid structures in the era of cancer immunotherapy. *Nat. Rev. Cancer* **19**, 307–325 (2019).

36. Schumacher, T. N. & Thommen, D. S. Tertiary lymphoid structures in cancer. *Science* **375**, eabf9419 (2022).

37. Amgad, M. *et al.* Report on computational assessment of Tumor Infiltrating Lymphocytes from the International Immuno-Oncology Biomarker Working Group. *npj Breast Cancer* **6**, (2020).

38. Steele, K. E. *et al.* Measuring multiple parameters of CD8+ tumor-infiltrating lymphocytes in human cancers by image analysis. *J. Immunother. Cancer* **6**, 1–14 (2018).
39. Koelzer, V. H. *et al.* Digital image analysis improves precision of PD-L1 scoring in cutaneous melanoma. *Histopathology* **73**, 397–406 (2018).
40. Hong, L. *et al.* Programmed Death-Ligand 1 Heterogeneity and Its Impact on Benefit From Immune Checkpoint Inhibitors in NSCLC. *J. Thorac. Oncol.* **15**, 1449–1459 (2020).

Supplemental Material
Supplemental figures
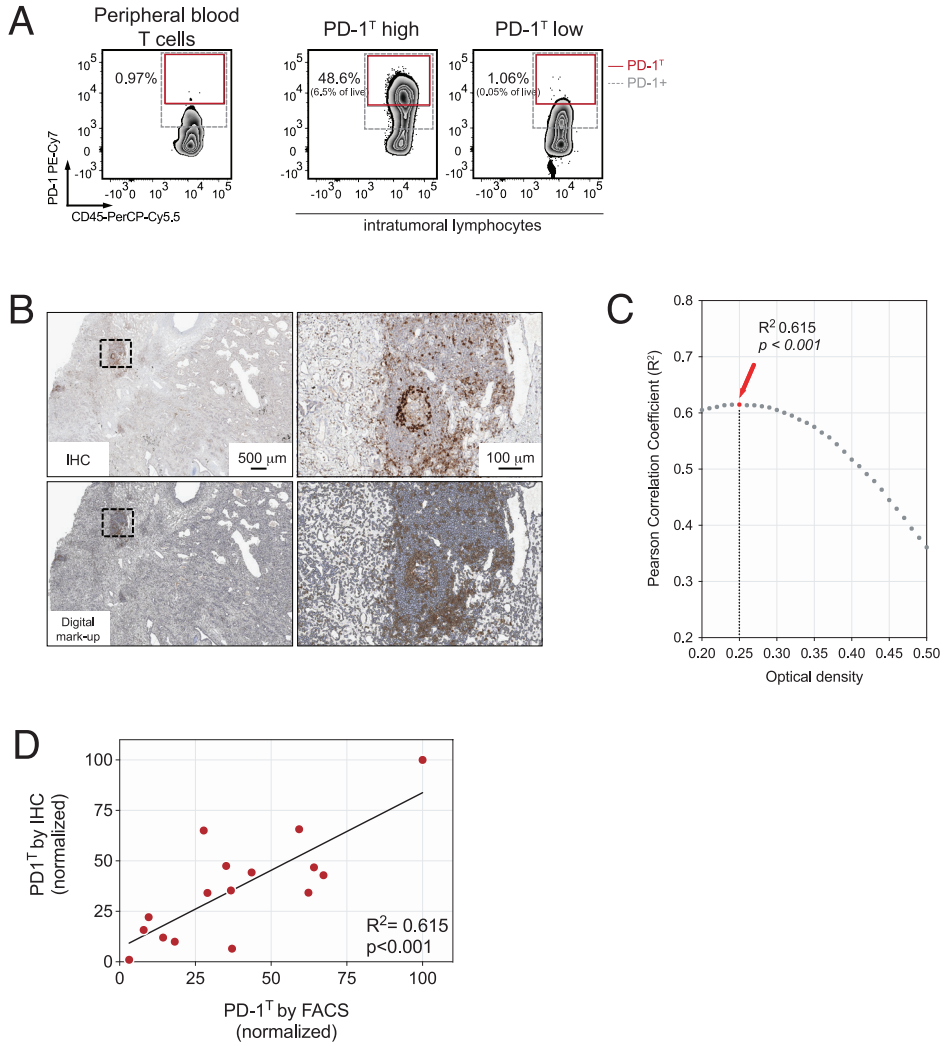
◄ **Figure S1.** Standardized identification of PD-1^T TILs by flowcytometry and digital IHC analysis. **(A)** PD-1^T expression assessed by flow cytometry in total lymphocytes. Shown is the PD-1 expression on T cells from a reference sample derived from healthy donor peripheral blood (left) and from two representative NSCLC samples with either high or low PD-1^T infiltration (right). The gate for PD-1+ T cells is shown in grey and for PD-1^T TILs in red. **(B)** Immunohistochemical (IHC) analysis of PD-1^T TILs. Example of a PD-1 IHC with digital markup; the black square shows PD-1^T TILs localized in a tertiary lymphoid structure (TLS). **(C)** Correlation of flow cytometry and digital IHC algorithm-based quantification of PD-1^T TILs in 16 NSCLC tumor samples. Shown are different OD thresholds with corresponding R^2 values. An OD of 0.25 was chosen as cut-off with the highest correlation, $R^2 = 0.615$, $P^{***} < 0.001$. R^2 and P-values were calculated using linear regression analysis. **(D)** Correlation curve showing normalized PD-1^T values for flowcytometry (FACS) and digital IHC analysis using a cut-off of 0.25 OD.
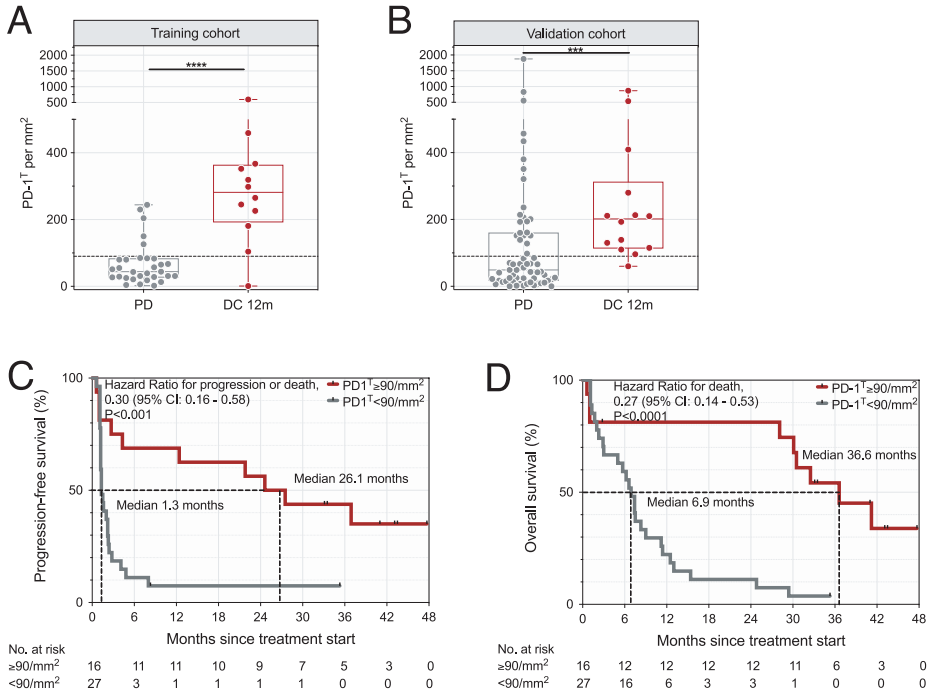
Figure S2. Association of PD-1^T TILs with long-term benefit and survival. **(A)** PD-1^T TILs per mm^2 in pretreatment samples from patients with disease control at 12 months (DC 12m) (n=12) and PD (n=31) in the training set (n=43) and **(B)** in the validation set (n=14, n=63). Dashed line indicates a cut-off of 90 PD-1^T TILs per mm^2. Medians, interquartile ranges and minimum/maximum shown in boxplots, **** $P<0.0001$ and *** $P<0.001$ by Mann Whitney U-test. **(C)** Progression-free survival (PFS) (median 26.1 versus 1.3 months, HR 0.30; 95% CI: 0.16-0.58, *** $P<0.001$) and **(D)** overall survival (OS) (median 36.6 versus 6.9 months, HR 0.27; 95% CI: 0.14-0.53, **** $P<0.0001$) 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. 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.
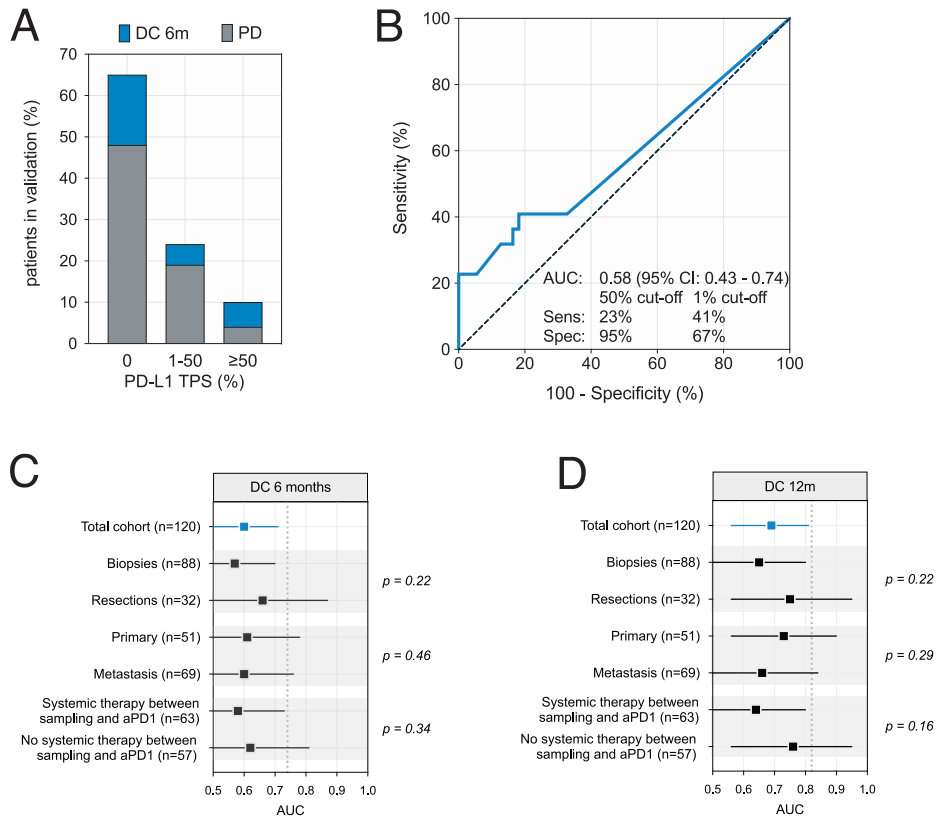
Figure S3. Association of PD-L1 TPS with DC at 6 months and impact of tissue sample properties on the predictive accuracy of PD-L1. **(A)** Percentage of patients with ≥50% (n=8), 1-50% (n=19) and 0% PD-L1 expression (n=50) in pretreatment samples showing DC 6m and PD in the validation set (n=77). **(B)** ROC curve for predictive value of PD-L1 for DC 6m (AUC 0.58; 95%-CI 0.43-0.74) in the validation set (n=77). **(C-D)** The predictive value of PD-L1 TPS in the total cohort and different subgroups. Each comparison is marked in a grey square. Shown is the area under the curve (AUC) for DC 6m **(C)** and 12m **(D)** with 95% CI interval. The dashed line indicates the AUC of PD-1[T] TILs in the total cohort. *P*-value was determined by one-sided permutation test.
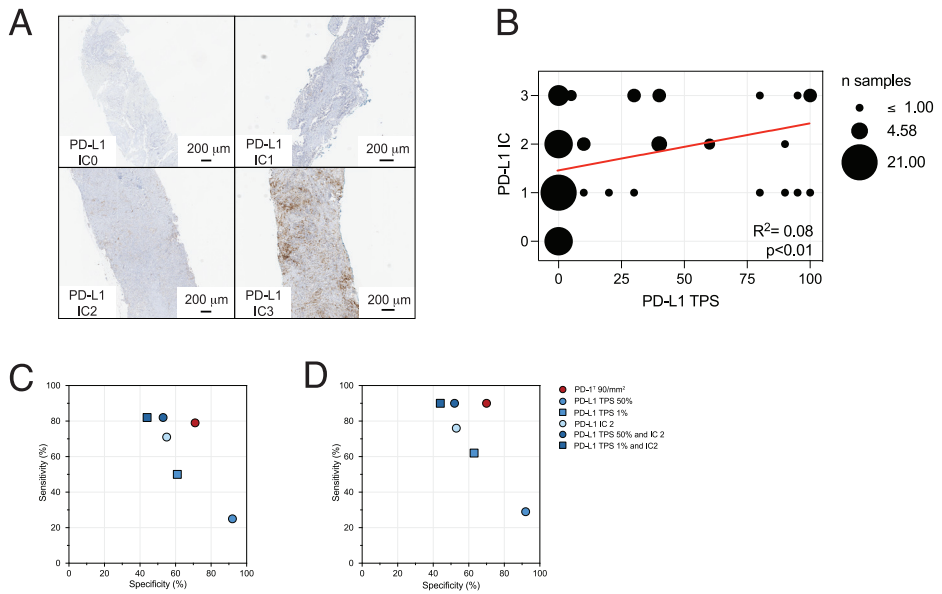
Figure S4. Combination of PD-L1 TPS and PD-L1 IC to predict clinical benefit. (**A)** Examples of IHC analysis of NSCLC tumors with IC0, IC1, IC2 and IC3 PD-L1 immune cell (IC) score. **(B)** Correlation of PD-L1 TPS scores and IC scores. The size of the circles indicates the number of tumors. R^2 and *P*-values were calculated using linear regression analysis. **(C-D)** Sensitivity and specificity of biomarkers PD-L1 TPS at either 50% or 1% cut-off, PD-L1 IC 2, and PD-L1 TPS at either 50% or 1% cut-off plus PD-L1 IC 2 for DC 6m **(C)** and 12m **(D)** in all nivolumab treated patients (n=94). Results are compared to PD-1^T 90 per mm^2 as univariate biomarker.
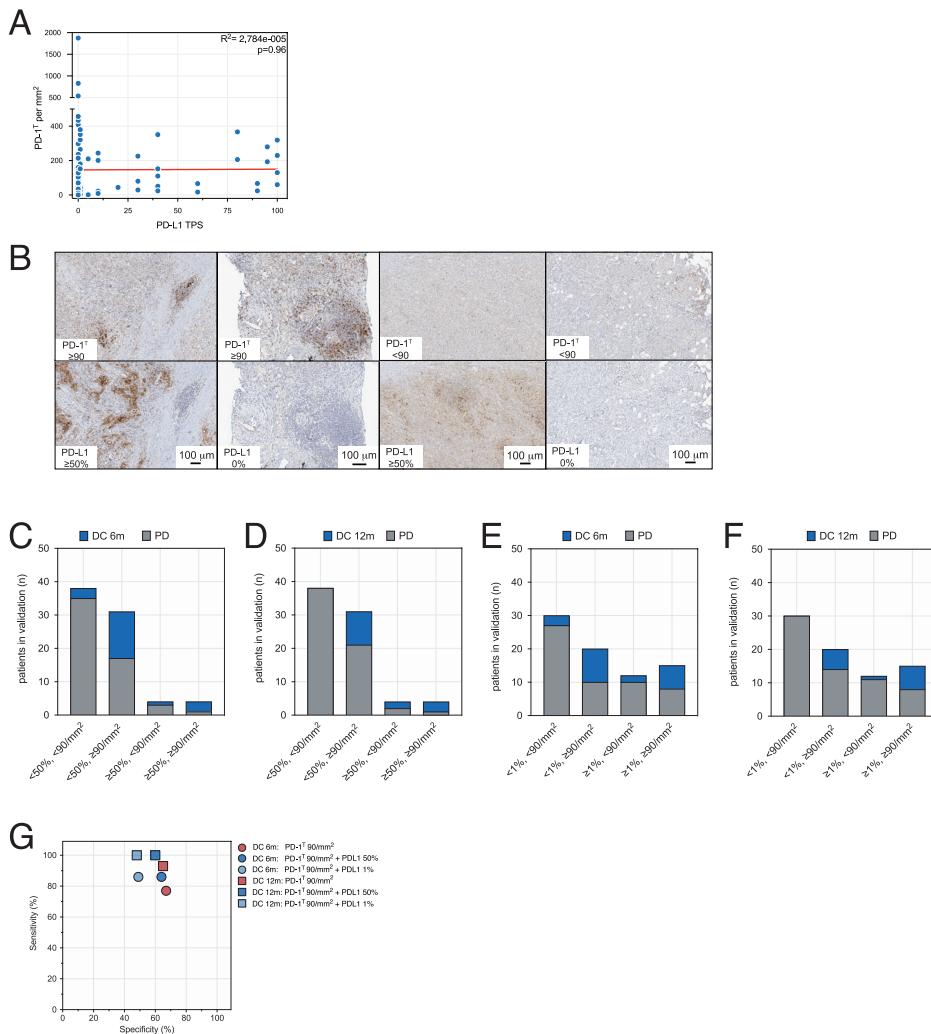
Figure S5. Combination of PD-1^T TILs and PD-L1 to predict clinical benefit. **(A)** Correlation of PD-1^T TILs and PD-L1 TPS. R^2 and *P*-values were calculated using linear regression analysis. **(B)** IHC analysis of PD-1^T TILs and PD-L1. Examples of tumors that are PD-1^T high ($\geq$90 per mm^2)+PD-L1 high ($\geq$50%), PD-1^T high+PD-L1 low (<50%), PD-1^T low (<90 per mm^2)+PD-L1 high and PD-1^T low+PD-L1 low. **(C)** Number of patients in the validation set (n=77) with either DC 6m or PD in the following subgroups: (1) PD-L1 low (<50%)+PD-1^T low (n=38), (2) PD-L1 low+PD-1^T high (n=31), (3) PD-L1 high ($\geq$50%)+PD-1^T low (n=4), and (4) PD-L1 high+PD-1^T high (n=4). **(D)** Same plot as in C for DC 12m and PD. **(E)** Number of patients in the validation set (n=77) with either DC 6m or PD in the following subgroups: (1) PD-L1 low (<1%)+PD-1^T low (n=30), (2) PD-L1 low+PD-1^T high (n=20), (3) PD-L1 high ($\geq$1%)+PD-1^T low (n=12) and (4) PD-L1 high+PD-1^T high (n=15). **(F)** Same plot as in E for DC 12m. (G) Sensitivity and specificity of composite biomarkers PD-1^T 90 per mm^2 + PD-L1 at either 50% or 1% cut-off for DC 6m and 12m using samples from the validation set (n=77). Results are compared to PD-1^T 90 per mm^2 as univariate biomarker.

2
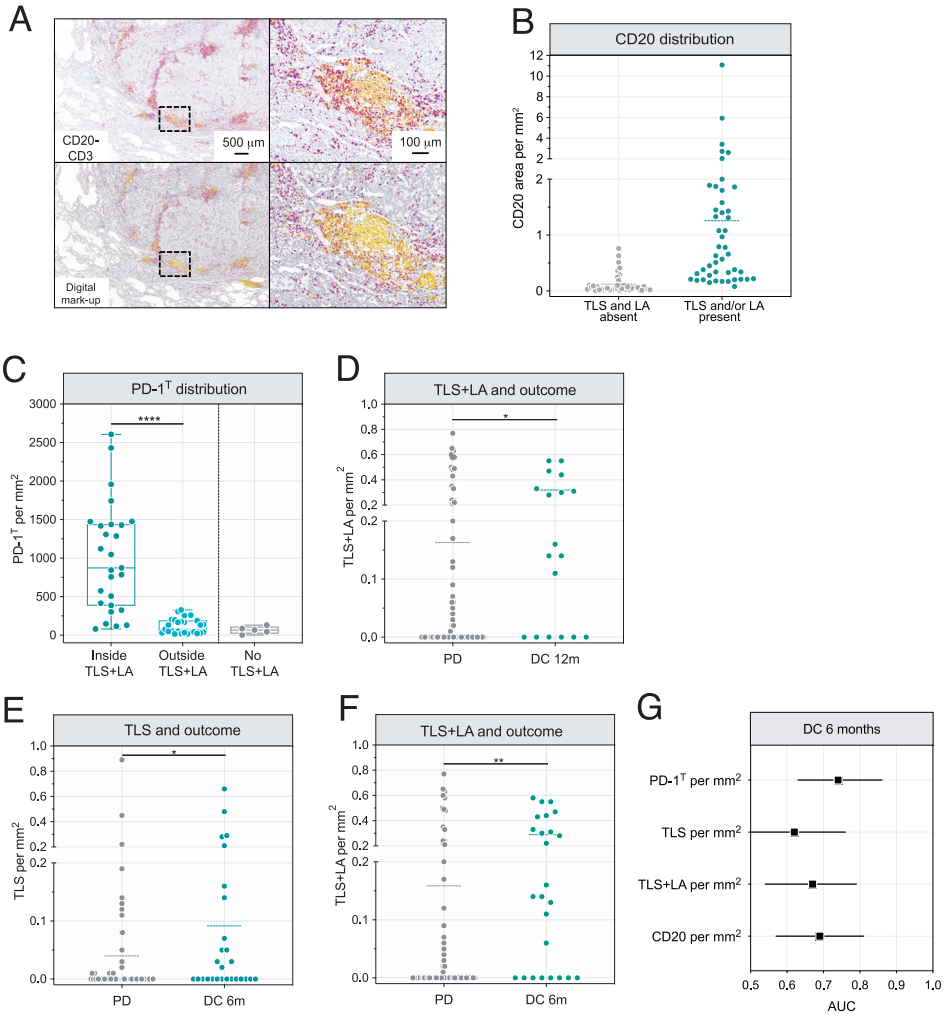
◄ **Figure S6.** Association of TLS, LA and CD20 with clinical benefit. **(A)** Top: Example of a CD20-CD3 IHC double staining with the black square showing CD20+ B cells (in yellow) and CD3+ T cells (in purple) localizing in TLS. Bottom: Digital markup to quantify CD20-positive areas. Colors indicate intensity of the staining. **(B)** CD20-positive area per mm^2 in pretreatment samples with presence (n=46) and absence (n=40) of TLS and/or LA (referred as TLS+LA). Shown is the mean. **(C)** PD-1^T TILs per mm^2 inside and outside TLS+LA in resected samples (n=27) and of total tumor area for resected samples with no TLS+LA (n=5). Medians, interquartile ranges and minimum/maximum shown in boxplots, ****$P<0.0001$ by Mann-Whitney test. **(D)** TLS+LA per mm2 in pretreatment samples from patients with DC 12m (n=20) and PD (n=71). Shown is the mean, *$P=0.02$ by Mann-Whitney test. **(E)** TLS and **(F)** TLS+LA per mm^2 in pretreatment samples from patients with DC 6m (n=27) and PD (n=64). Shown is the mean, *$P=0.03$, **$P<0.01$ by Mann-Whitney test. (G) The predictive value of PD-1^T TILs, TLS, TLS+LA and CD20-positive area per mm^2 for DC 6m (n=91, note that this cohort is smaller due to the availability of FFPE material). Shown are AUCs with 95% CI interval.

Supplemental tables

Table S1. Exclusion criteria for PD-1^T TILs assessment

List of Exclusion criteria:

- Samples containing <10,000 cells
- Samples obtained >2 years before start of ICB
- Samples from the bronchus site
- Samples containing normal lymphoid tissue

Table S2. Overview of all analyzed biomarkers per patient.

Sample ID	1 = DC 6 months, 0= PD	1 = DC 12 months, 0= PD	T=training V=validation
NKI-AvL 001	1	1	T
NKI-AvL 002	0	0	T
NKI-AvL 003	0	0	T
NKI-AvL 004	0	0	T
NKI-AvL 005	1	1	T
NKI-AvL 006	0	0	T
NKI-AvL 007	0	0	T
NKI-AvL 008	1	1	T
NKI-AvL 009	0	0	T
NKI-AvL 010	0	0	T
NKI-AvL 011	0	0	T
NKI-AvL 012	0	0	T
NKI-AvL 013	0	0	T
NKI-AvL 014	0	0	T
NKI-AvL 015	0	0	T
NKI-AvL 016	1	1	T
NKI-AvL 017	0	0	T
NKI-AvL 018	1	1	T
NKI-AvL 019	0	0	T
NKI-AvL 020	1	1	T
NKI-AvL 021	1	1	T
NKI-AvL 022	0	0	T
NKI-AvL 023	0	0	T
NKI-AvL 024	0	0	T
NKI-AvL 025	1	1	T
NKI-AvL 026	1	1	T
NKI-AvL 027	0	0	T
NKI-AvL 028	0	0	T
NKI-AvL 029	1	0	T
NKI-AvL 030	0	0	T
NKI-AvL 031	0	0	V
NKI-AvL 032	0	0	V
NKI-AvL 033	1	1	V
NKI-AvL 034	1	0	V
NKI-AvL 035	0	0	V
NKI-AvL 036	1	0	V

PD-L1 TPS	PD-1^T per mm² tumor area	TLS per mm²	TLS+LA per mm²	CD20-positive area per mm²
1	265	0.00	0.14	0.22
40	25	0.00	0.35	0.57
0	80	0.19	0.50	2.62
90	67	0.00	0.00	0.10
0	298	0.00	0.55	0.34
0	6	0.00	0.00	0.36
0	204	0.00	0.24	0.38
80	367	0.28	0.44	1.08
0	43	0.01	0.04	0.17
0	21	0.00	0.00	0.01
30	31	0.00	0.01	0.08
1	150	0.00	0.22	0.34
0	18	0.00	0.02	0.21
100	230	0.08	0.24	1.45
0	55	0.00	0.02	0.15
1	181	0.05	0.16	0.51
0	126	0.00	0.00	0.03
30	226	0.16	0.47	1.80
10	244	0.00	0.00	0.02
100	319	0.07	0.31	1.31
1	352	0.66	1.33	5.93
40	51	0.01	0.07	0.20
30	80	0.02	0.06	0.20
0	31	NA	NA	NA
0	104	NA	NA	NA
5	1	0	0	0.05
0	27	0	0	0.04
20	44	NA	NA	NA
0	31	0	0.13	0.21
30	29	0	0	0.08
40	152	0.12	0.12	0.19
0	67	0.00	0.48	2.73
0	537	0.00	0.00	0.63
90	24	0.00	0.00	0.00
0	70	0.00	0.00	0.10
0	828	0.21	0.43	1.87

Table S2. Continued

Sample ID	1 = DC 6 months, 0= PD	1 = DC 12 months, 0= PD	T=training V=validation
NKI-AvL 037	0	0	V
NKI-AvL 038	0	0	V
NKI-AvL 039	1	1	V
NKI-AvL 040	0	0	V
NKI-AvL 041	1	1	V
NKI-AvL 042	0	0	V
NKI-AvL 043	0	0	V
NKI-AvL 044	0	0	V
NKI-AvL 045	1	1	V
NKI-AvL 046	0	0	V
NKI-AvL 047	0	0	V
NKI-AvL 048	0	0	V
NKI-AvL 049	1	1	V
NKI-AvL 050	0	0	V
NKI-AvL 051	0	0	V
NKI-AvL 052	0	0	V
NKI-AvL 053	0	0	V
NKI-AvL 054	1	1	V
NKI-AvL 055	0	0	V
NKI-AvL 056	1	1	V
NKI-AvL 057	1	1	V
NKI-AvL 058	0	0	V
NKI-AvL 059	0	0	V
NKI-AvL 060	0	0	V
NKI-AvL 061	1	0	V
NKI-AvL 062	0	0	V
NKI-AvL 063	0	0	V
NKI-AvL 064	0	0	V
NKI-AvL 065	0	0	V
NKI-AvL 066	0	0	V
NKI-AvL 067	1	1	V
NKI-AvL 068	0	0	V
NKI-AvL 069	0	0	V
NKI-AvL 070	0	0	V
NKI-AvL 071	0	0	V
NKI-AvL 072	0	0	V

2

PD-L1 TPS	PD-1^T per mm² tumor area	TLS per mm²	TLS+LA per mm²	CD20-positive area per mm²
0	434	0.45	0.63	3.41
0	83	0.00	0.00	0.03
0	115	0.14	0.28	0.45
0	34	0.22	0.77	1.89
100	130	0.00	0.00	0.29
0	98	0.00	0.00	0.01
40	44	0.00	0.00	0.05
0	68	0.00	0.00	0.01
5	210	0.00	0.00	0.25
0	1875	0.89	2.98	11.07
1	25	0.00	0.65	0.79
40	351	0.00	0.00	0.07
40	110	0.00	0.11	0.33
1	43	0.00	0.03	0.17
0	6	0.00	0.00	0.02
0	457	0.00	0.00	0.51
1	321	0.00	0.00	0.06
95	280	0.03	0.33	1.40
10	201	0.00	0.21	0.78
0	213	0.03	0.30	1.08
0	409	0.48	1.28	2.03
0	41	0.11	0.49	0.97
0	21	0.05	0.17	0.66
10	9	0.00	0.00	0.05
0	17	0.00	0.00	0.76
1	3	0.00	0.00	0.01
0	26	0.08	0.58	1.43
60	17	0.00	0.00	0.19
0	13	0.00	0.00	0.00
0	128	0.00	0.00	0.02
0	211	0.00	0.55	1.58
0	85	0.00	0.00	0.06
0	152	0.03	0.20	0.63
0	13	0.00	0.00	0.04
0	7	0.00	0.00	0.00
0	236	0.14	0.60	2.00

Table S2. Continued

Sample ID	1 = DC 6 months, o= PD	1 = DC 12 months, o= PD	T=training V=validation
NKI-AvL 073	0	0	V
NKI-AvL 074	0	0	V
NKI-AvL 075	0	0	V
NKI-AvL 076	0	0	V
NKI-AvL 077	0	0	V
NKI-AvL 078	0	0	V
NKI-AvL 079	0	0	V
NKI-AvL 080	0	0	V
NKI-AvL 081	1	0	V
NKI-AvL 082	0	0	V
NKI-AvL 083	1	0	V
NKI-AvL 084	1	0	V
NKI-AvL 085	0	0	V
NKI-AvL 086	0	0	V
NKI-AvL 087	1	1	V
NKI-AvL 088	0	0	V
NKI-AvL 089	0	0	V
NKI-AvL 090	1	1	V
NKI-AvL 091	0	0	V
NKI-AvL 092	0	0	V
NKI-AvL 093	1	1	V
NKI-AvL 094	0	0	V
NKI-AvL 095	1	1	T
NKI-AvL 096	1	1	T
NKI-AvL 097	1	1	T
NKI-AvL 098	1	0	T
NKI-AvL 099	0	0	T
NKI-AvL 100	0	0	T
NKI-AvL 101	0	0	T
NKI-AvL 102	0	0	T
NKI-AvL 103	0	0	T
NKI-AvL 104	0	0	T
NKI-AvL 105	0	0	T
NKI-AvL 106	0	0	T
NKI-AvL 107	0	0	T
NKI-AvL 108	0	0	V

PD-L1 TPS	PD-1^T per mm^2 tumor area	TLS per mm^2	TLS+LA per mm^2	CD20-positive area per mm^2
0	56	0.00	0.00	0.09
0	15	0.00	0.00	0.04
1	152	0.00	0.00	0.41
80	206	0.00	0.00	0.03
0	2	0.00	0.00	0.00
0	6	0.00	0.00	0.03
0	35	0.01	0.05	0.38
10	22	0.00	0.00	0.08
0	214	0.29	0.58	1.86
0	49	0.01	0.09	0.31
0	160	0.05	0.22	0.21
0	67	0.00	0.06	0.18
60	66	0.00	0.00	0.05
0	49	0.00	0.00	0.03
95	193	0.00	0.00	0.01
40	24	0.00	0.00	0.00
0	11	0.00	0.00	0.10
100	60	0.00	0.00	0.02
1	380	0.13	0.33	1.33
0	2	0.00	0.00	0.07
0	139	0.02	0.14	0.28
0	0	0.00	0.00	0.02
60	459	NA	NA	NA
80	245	NA	NA	NA
0	595	NA	NA	NA
0	13	NA	NA	NA
0	39	NA	NA	NA
0	4	NA	NA	NA
30	85	NA	NA	NA
0	28	NA	NA	NA
0	84	NA	NA	NA
0	83	NA	NA	NA
80	65	NA	NA	NA
0	2	NA	NA	NA
0	57	NA	NA	NA
0	161	NA	NA	NA

Table S2. Continued

Sample ID	1 = DC 6 months, 0= PD	1 = DC 12 months, 0= PD	T=training V=validation
NKI-AvL 109	0	0	V
NKI-AvL 110	0	0	V
NKI-AvL 111	0	0	V
NKI-AvL 112	0	0	V
NKI-AvL 113	0	0	V
NKI-AvL 114	1	0	V
NKI-AvL 115	0	0	V
NKI-AvL 116	1	0	V
NKI-AvL 117	1	1	V
NKI-AvL 118	1	1	V
NKI-AvL 119	0	0	V
NKI-AvL 120	0	0	V

PD-L1 TPS	PD-1ᵀ per mm² tumor area	TLS per mm²	TLS+LA per mm²	CD20-positive area per mm²
0	194	NA	NA	NA
0	42	NA	NA	NA
0	552	NA	NA	NA
3	193	NA	NA	NA
0	7	NA	NA	NA
0	139	NA	NA	NA
0	25	NA	NA	NA
0	12	NA	NA	NA
2	868	NA	NA	NA
40	96	NA	NA	NA
0	66	NA	NA	NA
1	0	NA	NA	NA

Table S3. Patient characteristics and treatment outcomes for training and validation set.

		Training cohort	Validation cohort
		n=43	n=77
Sex	P=0.13		
Male, no.(%)		18 (42%)	44 (57%)
Female, no.(%)		25 (58%)	33 (43%)
Age (years), mean (s.d.)	P=0.21	64 (8)	62 (9)
Smoking (never/ex/current)	P=0.84	3/35/5	8/58/11
Pack years, median (IQR)	P=0.62	30 (21)	28 (28)
PS, no.(%)	P=0.53		
0		12 (28%)	24 (31%)
1		22 (51%)	43 (56%)
≥2		9 (21%)	10 (13%)
Pathology, no.(%)	P=0.90		
Adeno		26 (60%)	50 (65%)
Squamous		9 (21%)	12 (16%)
LCNEC, NSCLC-type		1 (2%)	2 (2%)
NSCLC, NOS		7 (16%)	13 (17%)
Mutations, no. (%)	P=0.18		
KRAS positive		22 (51%)	29 (38%)
PD-L1 TPS, no. (%)			
Negative <1%	P=0.17	22 (51%)	50 (65%)
Positive ≥1%		21 (49%)	27 (35%)
Negative <50%	P=0.39	36 (84%)	69 (90%)
Positive ≥50%		7 (16%)	8 (10%)
Brain metastases, no. (%)	P=1.00	7 (16%)	14 (18%)
Treatment, no. (%)	P=0.11		
Nivolumab		30 (70%)	64 (83%)
Pembrolizumab		13 (30%)	13 (17%)
Line of treatment, no (%)	P=0.19		
1		0 (0%)	1 (1%)
2		35 (81%)	52 (68%)
>2		8 (19%)	24 (31%)
Best Overall Response	P=0.47		
CR/PR		12 (28%)	12 (16%)
SD (PFS ≥6 months)		2 (5%)	11 (14%)
SD (PFS <6 months)		2 (5%)	4 (5%)
PD		27 (63%)	50 (65%)

Table S3. Continued

		Training cohort	Validation cohort
		n=43	n=77
DC			
6 months	P=0.68	33% (14/43)	29% (22/77)
12 months	P=0.25	28% (12/43)	18% (14/77)

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.

Table S4. Tissue sample properties of the training and validation set.

		Training cohort	Validation cohort
		n=43	n=77
Sample type	P=0.33		
Biopsy		30 (70%)	58 (75%)
Resection		13 (30%)	19 (25%)
Sample site	P=0.46		
Primary		19 (44%)	32 (41%)
Metastasis		16 (37%)	36 (47%)
Lung		1	5
Adrenal gland		3	7
Bone		0	7
Pleurae		0	1
Brain		0	2
Muscle / Subcutaneous		3	8
Liver		6	3
Mamma		2	1
Gastrointestinal		1	2
Lymph node metastasis		8 (19%)	9 (12%)

Table S5. The relationship between PD-1^T and clinicopathological variables as categories.

		PD-1^T <90 per mm^2	PD-1^T ≥90 per mm^2
		n=69	n=51
Sex	P=0.58		
Male, no. (%)		34 (49%)	28 (55%)
Female, no. (%)		35 (51%)	23 (45%)
Age (years), median (IQR)	P=0.96	63 (57 - 70)	64 (58 - 68)
Smoking (never/ex/current)	P=0.47	8/52/9	3/41/7
Pack years, median (IQR)	P=0.18	25 (13 - 40)	30 (20 - 40)
PS, no. (%)	P=0.11		
0		18 (26%)	18 (35%)
1		37 (54%)	28 (55%)
≥2		14 (20%)	5 (10%)
Pathology, no. (%)	P=0.20		
Adeno		48 (69%)	28 (55%)
Squamous		8 (12%)	13 (25%)
LCNEC, NSCLC-type		2 (3%)	1 (2%)
NSCLC, NOS		11 (16%)	9 (18%)
Mutations, no. (%)	P=1		
KRAS positive		31 (45%)	20 (39%)
KRAS negative		26 (38%)	17 (33%)
KRAS unknown		12 (17%)	14 (28%)
PD-L1 TPS, no. (%)			
Negative	*P=0.04	47 (74%)	25 (63%)
Positive ≥1%		22 (26%)	26 (37%)
Negative <50%	P=0.17	63 (91%)	42 (82%)
Positive ≥50%		6 (9%)	9 (18%)
Brain metastases, no. (%)	P=0.22	15 (22%)	6 (12%)
Treatment, no. (%)	P=0.66		
Nivolumab		53 (77%)	41 (80%)
Pembrolizumab		16 (23%)	10 (20%)
Line of treatment, no (%)	P=0.21		
1		1 (1%)	0 (0%)
2		53 (77%)	34 (67%)
>2		15 (22%)	17 (%)

Table S6. The relationship between PD-1^T and clinicopathological variables as continuous variable.

		PD-1^T TILs per mm^2 (quantiles)				
		0%	25%	50%	75%	100%
Sex	P=0.67					
Male (n=62)		0	25	67	201	1875
Female (n=58)		1	26	69	199	868
Smoking	P=0.86					
Ex (n=93)		0	24	66	194	1875
Current (n=16)		7	40	82	207	457
PS	P=0.19					
0 (n=36)		2	25	85	238	868
1 (n=65)		0	27	80	181	1875
≥2 (n=19)		0	26	66	118	457
Pathology	P= 0.60					
Adeno (n=76)		0	25	63	184	1875
Squamous (n=21)		2	31	139	213	595
LCNEC, NSCLC-type (n=3)		0	21	82	264	552
NSCLC, NOS (n=20)		7	37	67	97	126
Mutations	P= 0.42					
KRAS positive (n=51)		0	30	80	200	1875
KRAS negative (n=43)		2	20	66	212	595
KRAS unknown (n=26)		0	26	101	185	298
Brain metastases	P=0.38					
Yes (n=21)		0	29	60	152	828
No (n=99)		0	25	80	207	1875
Treatment	P=0.81					
Nivolumab (n=94)		0	25	68	206	1875
Pembrolizumab (n=26)		0	26	75	185	868
Line of treatment	P= 0.13					
1 (n=1)						
2 (n=87)		0	25	66	187	868
>2 (n=32)		0	26	104	258	1875

Table S7. Predictive accuracy of PD-L1 TPS at 1%, 5%, 10% and 50% cut-off in the validation set (n=77).

Clinical outcome	Biomarker	AUC	Cut-off
DC 6m	% PD-L1 TPS	0.58 95% CI: 0.43-0.74	<1% vs. ≥1%
			<5% vs. ≥5%
			<10% vs. ≥10%
			<50% vs. ≥50%
DC 12m	% PD-L1 TPS	0.68 95% CI: 0.51-0.86	<1% vs. ≥1%
			<5% vs. ≥5%
			<10% vs. ≥10%
			<50% vs. ≥50%

Table S8. Predictive accuracy of composite biomarkers PD-L1 TPS+PD-L1 IC in all nivolumab treated patients (n=94).

Clinical outcome	Biomarker	Cut-off
DC 6 months	% PD-L1 TPS + PD-L1 IC	<50%+<2 vs. ≥50%+<2 and <50%+≥2 and ≥50%+≥2
DC 6 months	% PD-L1 TPS + PD-L1 IC	<1%+<2 vs. ≥1%+<2 and <1%+≥2 and ≥1%+≥2
DC 12 months	% PD-L1 TPS + PD-L1 IC	<50%+<2 vs. ≥50%+<2 and <50%+≥2 and ≥50%+≥2
DC 12 months	% PD-L1 TPS + PD-L1 IC	<1%+<2 vs. ≥1%+<2 and <1%+≥2 and ≥1%+≥2

Table S9. Predictive accuracy of composite biomarkers PD-1^T TILs+PD-L1.

Clinical outcome	Biomarker	Cut-off
DC 6 months	PD-1^T TILs per mm^2 + % PD-L1 TPS	<90+<50% vs. ≥90+<50% and <90+≥50% and ≥90+≥50%
DC 6 months	PD-1^T TILs per mm^2 + % PD-L1 TPS	<90+<1% vs. ≥90+<1% and <90+≥1% and ≥90+≥1%
DC 12 months	PD-1^T TILs per mm^2 + % PD-L1 TPS	<90+<50% vs. ≥90+<50% and <90+≥50% and ≥90+≥50%
DC 12 months	PD-1^T TILs per mm^2 + % PD-L1 TPS	<90+<1% vs. ≥90+<1% and <90+≥1% and ≥90+≥1%

Sensitivity	Specificity	NPV	PPV
41%	67%	74%	33%
36%	82%	76%	44%
32%	82%	75%	41%
23%	95%	75%	63%
57%	70%	88%	30%
50%	83%	88%	39%
43%	83%	87%	35%
29%	94%	86%	50%

Sensitivity	Specificity	NPV	PPV
82	44	85	38
82	53	88	43
90	44	94	32
90	52	95	35

Sensitivity	Specificity	NPV	PPV
86%	64%	92%	49%
86%	49%	90%	40%
100%	60%	100%	36%
100%	48%	100%	30%

Chapter 3

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 Center, Leiden, The Netherlands.

Abstract

Background
The efficacy of PD-1 blocking agents in advanced NSCLC has shown prolonged effectiveness, but only in a minority of patients. Multiple biomarkers have been explored to predict treatment benefit, yet their combined performance remains inadequately examined. In this study, we assessed the combined predictive performance of multiple biomarkers in NSCLC patients treated with nivolumab.

Methods
Pretreatment samples from 135 patients receiving nivolumab were used to evaluate the predictive performance of CD8 tumor-infiltrating lymphocytes (TILs), intratumoral (IT) localization of CD8 TILs, PD-1 high expressing TILs (PD-1^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 randomly assigned to a training (n=55) and validation cohort (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%) compared to individual biomarkers PD-1^T TILs and TIS (sensitivity: 72% and 83%, NPV: 86% and 84%) for DC 6m, respectively. Additionally, at 12 months, both selected composite biomarkers (CD8+IT-CD8 and CD8+TIS) demonstrated inferior 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 more accurately discriminate patients with no long-term benefit, as specificity was substantially higher 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-month endpoints. 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 therapeutic landscape of advanced stage non-small cell lung cancer (NSCLC). A subset of patients treated with these PD-1/PD-L1 blocking agents experience durable responses, translating into a significant survival advantage[1–7]. However, the majority fails to derive durable clinical benefit, underscoring the need for predictive biomarkers to support treatment decision-making in clinical practice. Specifically, the identification of biomarkers capable of excluding patients unlikely to benefit from PD-1/PD-L1 blockade therapy can prevent unnecessary side effects and contribute to the reduction of health care costs.

The assessment of tumor PD-L1 expression through immunohistochemistry (IHC) has been a focal point in numerous clinical trials as a potential predictive biomarker[8]. Although a positive correlation between PD-L1 expression and treatment outcomes has been observed in advanced stage NSCLC patients[1,5–7], a considerable proportion (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 for potential tumor antigenicity, has also shown predictive potential. However, its clinical implementation faces challenges, including the lack of a robust and predictive TMB cut-off and technical issues related to variation across platforms[15–17].

Given these challenges, there is an urgent need for biomarkers that can more accurately predict responses to PD-1/PD-L1 blockade in advanced NSCLC. Since this treatment regimen 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 with responses to PD-1 blockade in various cancer types, including melanoma[18], colorectal cancer[21], and NSCLC[22,23]. In addition, previous work showed that a distinct T cell population, termed PD-1^T TILs, can predict clinical benefit in NSCLC[24,25]. These PD-1^T TILs predominantly localize in tertiary lymphoid structures (TLS)[24,25]. Notably, B cells, critical components of these TLS, have also been associated with response to PD-1 blocking agents[26–28]. Other studies have developed predictive RNA expression signatures, such as the "tumor inflammation signature" (TIS), characterizing features of immune activity in the tumor microenvironment (TME)[29–31].

While these individual biomarkers show predictive potential, their accuracy is limited, likely due to the multifaceted nature of the antitumor immune response. Therefore, combining biomarkers holds promise for enhancing predictive accuracy, as demonstrated previously for combinations like TMB with PD-L1[32,33] and CD8 TILs with PD-L1[22,34]. Thus, the primary objective of this study is to investigate the performance of biomarker pairs, including CD8, PD-1[T] and CD3 TILs, CD20[+] B cells, TLS, PD-L1, and TIS, and compare their efficacy against individual biomarkers in predicting 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 enrolled patients started second or later line monotherapy nivolumab, administered intravenously at a dose of 3mg/kg 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. A randomization process was employed to allocate patients into a training and validation cohort. This 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 remained relatively low, ranging from 1 to 32 (see **Fig. S1** and later in this section).

Response to treatment was evaluated according to the Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1. Patients with progressive disease (PD) who were not evaluable for response assessment were designated by the treating physician as having PD. The primary clinical outcome was Disease Control (DC), defined as achieving a complete response (CR), partial response (PR) or maintaining stable disease (SD) at the 6-month mark following the initiation of treatment. As a secondary outcome measure, DC 12m was employed, representing the persistence of CR, PR or SD for a duration of 12 months or more. This secondary endpoint aimed to serve as an indicator of the long-term efficacy to PD-1 blockade therapy.

Pretreatment formalin-fixed paraffin embedded (FFPE) tumor tissue samples were collected from all patients. Written informed consent for the research usage of material, not essential for diagnostic purposes, was obtained from each patient by an 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 by 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 as well as tertiary lymphoid structures (TLS) and CD20[+] B cell data for 91 samples were used from

previous work[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, and five samples had insufficient tumor tissue for CD3 TIL, TLS and CD20$^+$ B cell analysis **(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) **(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. A total of 67 patients (41%) were excluded for mRNA expression analysis due to low RNA yield and/or low RNA quality **(Fig. S1)**.

Immunohistochemistry
The CD8 immunostaining of samples was executed using the BenchMark Ultra autostainer Instrument (Ventana Medical Systems) on 3 μm paraffin sections from FFPE blocks. Initially, sections were baked at 75°C for 28 minutes and deparaffinized 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).

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

The immunostained slides for CD8, PD-1, PD-L1 and CD68 were subjected to scanning at a magnification of x20 with a resolution of 0.50 per μm² using an Aperio slide AT2 scanner (Leica Biosystems). Immunostained slides for CD20-CD3 were scanned at x20 magnification with a resolution of 0.24 per μm² using a

3Dhistech P1000 scanner. For manual scoring, PD-L1 and CD68 IHC images were uploaded onto Slide Score, a web platform designed for the manual scoring of digital slides using a scoring sheet (www.slidescore.com). CD8, PD-1^T, CD3 TILs, CD20$^+$ B cells and TLS were digitally scored as described below.

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, version 2.3.2089.69 (Indica Labs). Researchers were blinded for clinical outcome. For the classification of CD8 lymphocytes in single stains, a computationally derived cut-off of 0.3 optical density (OD) was used, reflecting the intensity of the staining. This cut-off was established by manually optimizing the detection of CD8 positive stained cells in FFPE samples. An image analysis algorithm utilizing a 0.3 OD cut-off was generated for automated analyses of CD8 lymphocytes in subsequent FFPE samples. The quantification of PD-1^T TILs followed a previously described methodology[25].

The frequency of CD8 and PD-1^T TILs were determined as the number per mm^2 tumor area. Tumor areas were digitally annotated as described previously[25]. PD-1^T TIL data from 94 samples were used from previous work[25] **(Table S1)**. For regional analysis of CD8 lymphocytes, classifiers were trained to distinguish stromal and tumoral regions, allowing for the separate quantification of CD8 lymphocytes in these distinct regions. The percentage of CD8 lymphocytes within tumoral regions (i.e. intra-tumoral (IT)) relative to total CD8 TILs was subsequently calculated **(Table S1)**.

Scoring of tertiary lymphoid structures
The quantification of TLS and the combined number of TLS and lymphoid aggregates (TLS+LA) per mm^2 tumor area was performed using the HALO™ image analysis software, version 2.3.2089.69 (Indica Labs). This analysis was conducted on a CD20-CD3 double immunostaining, following a previously established methodology[25]. TLS and TLS+LA data from 91 samples were used from previous work[25] **(Table S1)**.

CD20 and CD3 quantification by digital image analysis
Digital quantification of CD20 and CD3 expression involved the measurement of the total area with CD20 and CD3 expression, relatively. This was performed using a pre-established image analysis algorithm from the Area Quantification version 1.0 module of HALO™ image analysis software (Indica Labs)[25]. The resulting CD20-positive and CD3-positive areas 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 within and at the border of TLS. Tumor areas were digitally annotated as described previously[25]. CD20 data from 91 samples were used from previous work[25] **(Table S1)**.

PD-L1 scoring
PD-L1 TPS was determined using the qualitative, clinical grade, laboratory developed, IHC assay (22C3 Agilent/DAKO) as described previously[25]. PD-L1 TPS data from 94 samples were used from previous work[25] **(Table S1)**. The CD68 staining was compared to the PD-L1 staining to exclude macrophages that express both CD68 and PD-L1, as their presence could potentially introduce false-positive results.

RNA extraction and hybridization to nCounter tagset
The extraction of RNA from pretreatment FFPE samples and subsequent Nanostring analysis were performed as described previously[35].

Statistical analysis
Patient characteristics were descriptively reported using mean ± SD, interquartile range (IQR) or frequencies (percentages). The Mann-Whitney test for continuous data, Fisher's exact test for categorical data and linear-by-linear association test for ordinal data were used to assess differences in patient characteristics between cohorts (training and validation) and between outcome groups (DC vs PD). Statistical significance was considered at $*P<0.05$, $**P<0.01$, $***P<0.001$ or $****P<0.0001$.

Genes in the Tumor Inflammation Signature (TIS) are normalized using a ratio of the expression value to the geometric mean of the housekeeper genes specific to the TIS signature and then followed by log2 transformation. The TIS score, a weighted linear combination of the 18 gene expression values, was calculated as part of Nanostring's intellectual property[29,36] **(Table S1)**.

In the training cohort, univariate and bivariate logistic models were constructed for DC 6m and DC 12m using CD8 TILs, IT-CD8 T cells, PD-1[T] TILs, CD3 TILs, TLS, TLS+LA, CD20[+] B cells, PD-L1 and TIS. The bivariate models included an interaction term. The bivariate logistic model produces for each patient a number between 0 and 1, reflecting the probability (according to the model) of patients reaching DC 6m or DC 12m. Discriminatory ability was evaluated using the area under the receiver operating characteristic (ROC) curve. Predictive performance metrics (sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV)) for different individual and composite biomarkers were calculated and comparisons

were made using the McNemar test. A point on the ROC curve corresponding to 90% sensitivity for DC 6m or DC 12m was selected to determine specificity, NPV and PPV. The aim was to achieve 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. In both methods, a cut-point was chosen for each of the two biomarkers and a patient was predicted positive (i.e. likely to respond to PD-1 blockade) if at least one (first method) or both (second method) biomarker values were above their respective cut-point values. The specificities obtained with these non-parametric methods were either equal to or worse than 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 based on a cut-off that demonstrated the highest specificity and NPV at the predefined sensitivities for predicting DC 6m and DC 12m. This cut-off was then used to determine sensitivity, specificity, NPV and PPV in the validation cohort.

Results

Biomarker characteristics and demographics

To evaluate the predictive performance of various biomarker combinations, we first analyzed pretreatment tumor samples from 162 advanced stage NSCLC patients treated with nivolumab. Nine biomarkers were assessed: (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).** CD8 TILs and IT-CD8 T cells were successfully assessed 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, Fig. S1)**. The use of solely archival samples led to a subset that lacked sufficient tumor tissue, as these samples had been previously used for analyses in the standard diagnostic routine. Additional exclusion criteria for each biomarker are provided in **Fig. S1**.

Patients with results for at least two biomarkers (n=135) were randomly assigned to a training (n=55) and validation (n=80) cohort. This randomization was stratified for clinical benefit to ensure that in both cohorts, 1 in 3 patients reached disease control at 6 months (DC 6m) and 1 in 5 patients reached disease control at 12 months (DC 12m), respectively. Due to the limited availability of patients with TIS scores (n=68), these patients were proportionately distributed in the randomization process **(Table 1, Fig. S1)**. Individual results for each biomarker per patient are detailed in **Table S1**. No significant differences in demographic characteristics were observed between the training and validation cohorts **(Table 2)**.
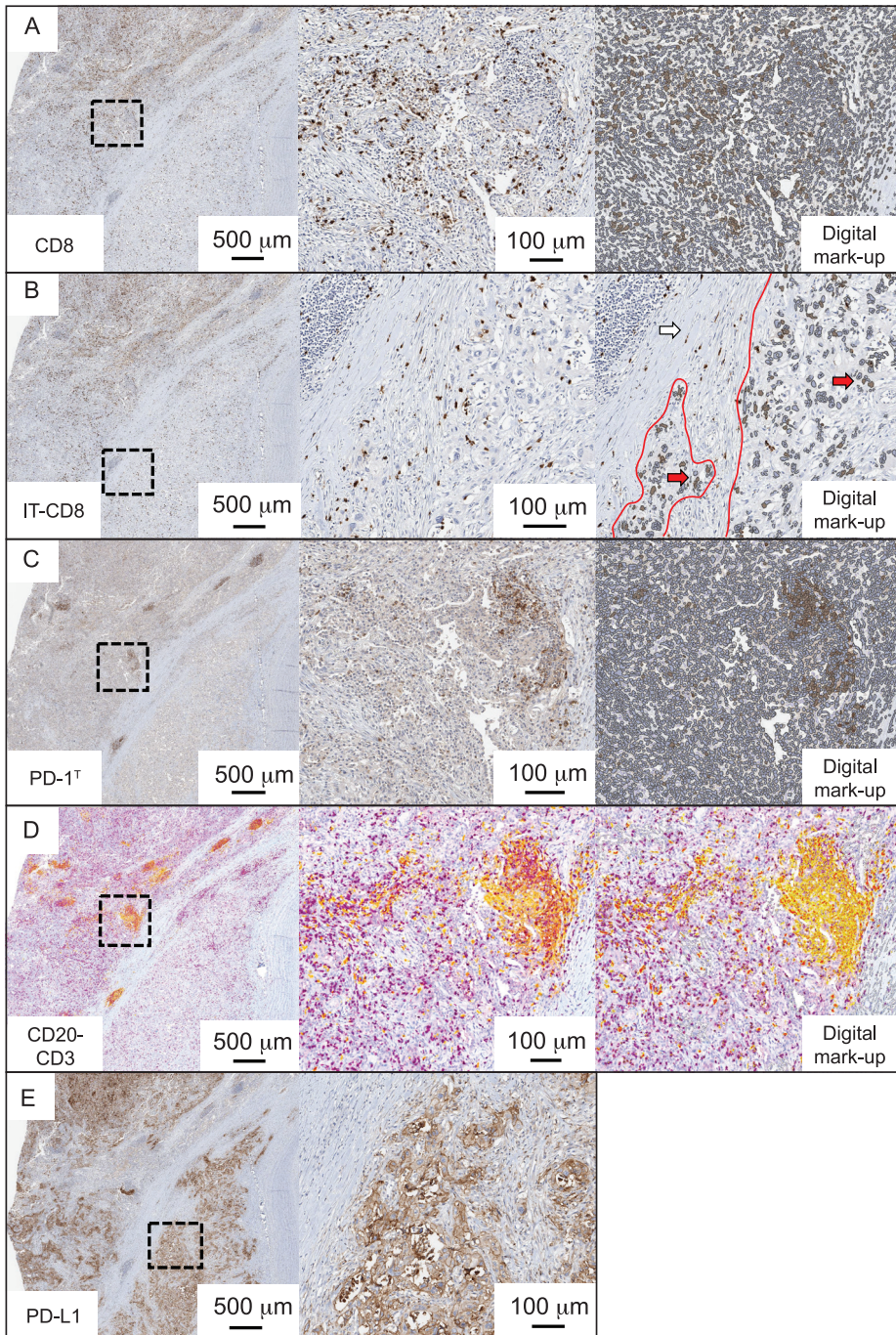
◄ **Figure 1.** Immunohistochemical analysis of all biomarkers and digital mark-up. (A) The left image shows an example of a CD8 immunohistochemical staining (IHC). The black square indicates the area that is shown in the central image. The right image shows the digital markup with CD8 TILs (in brown) 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 stained for PD-L1 IHC. The black square indicates the area that is shown in the right image. PD-L1 IHC slides were scored manually.

Accuracy of individual and composite biomarkers to predict DC at 6 months
Next, optimal cut-off values for individual and composite biomarkers in the training cohort were determined through ROC analysis. We aimed for a sensitivity and NPV of ≥90% to minimize the risk of undertreatment, while maintaining a specificity of at least 50% to identify patients unlikely to respond to PD-1 blockade therapy. The latter group can potentially benefit from alternative treatments. Since not all tumor samples were evaluable for all nine biomarkers, the number of samples in the training cohort ranged from 28 to 55 **(Table 1, Fig. S1)**. A total of 16 composite biomarkers, including PD-1^T and TIS as individual biomarkers, met the prespecified sensitivity and specificity criteria on the ROC curve **(Table S2)**. Interestingly, among these, cut-off values of 7/8 (88%) possible combinations with PD-1^T TILs and 5/8 (63%) with TIS reached these criteria **(Table S2)**. However, none significantly improved predictive accuracy compared to the standalone use of PD-1^T TILs and TIS **(Fig. S2A,B)**, leading to their exclusion from further analysis.

Subsequently, we selected the four remaining biomarkers with the highest predictive performance for validation, including the combinations of CD8+IT-CD8 and CD3+IT-CD8, as well as the individual biomarkers PD-1^T TILs and TIS **(Table S2)**. In the training cohort, both CD8+IT-CD8 and CD3+IT-CD8 demonstrated significantly higher probability scores in the DC 6m group (reflecting the probability of patients reaching DC 6m) compared to the progressive disease (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)**. Corresponding cut-off values of 0.167 and 0.161, respectively, were associated with a sensitivity of 94% for both, specificity of 62% and 54%, NPV of 96% and 95% and PPV of 50% and 47% **(Table 3)**.

Furthermore, PD-1^T TIL numbers and TIS scores were significantly higher in the DC 6m group compared to the PD group (PD-1^T TILs, $P<0.001$ and TIS, $P<0.01$) **(Fig. S2C,D)**. PD-1^T TILs showed an AUC of 0.82 (95% CI: 0.69-0.95), and TIS demonstrated an AUC of 0.81 (95% CI: 0.65-0.98) **(Fig. S2E,F)**. For PD-1^T TILs, a cut-off value of 90 per mm^2 was selected based on its predictive value in a previous study[25]. This threshold yielded a sensitivity of 92%, specificity of 67%, NPV of 95% and PPV of 52% **(Table 3)**. A score of 6.65 was identified as the optimal cut-off for TIS, resulting in 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 (CD8+IT-CD8, CD3+IT-CD8, PD-1^T TILs and TIS) in the validation cohort. The number of samples with successfully obtained biomarker results in the validation cohort ranged from 40 to 79 **(Table 1, Fig. S1)**. A decrease in the predictive accuracy of CD8+IT-CD8 and CD3+IT-CD8 biomarkers was observed when compared to the training cohort, as reflected by the AUC of the ROC curve measuring 0.62 (95% CI: 0.50-0.75) and 0.68 (95% CI: 0.55-0.80), respectively **(Fig. 2C,D)**. Moreover, the probability scores within the DC 6m group did not significantly differ from those in the PD group for CD8+IT-CD8 ($P=0.08$) **(Fig. 2E).** This comparison for CD3+IT-CD8 was borderline significant ($P=0.01$) **(Fig. 2F)**. The selected cut-off value of 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, while higher than that of CD8+IT-CD8, remained lower than in the training cohort (sensitivity: 83%, specificity: 46%, NPV: 85% and PPV: 43%) **(Table 3)**.

The individual biomarker analysis in the validation cohort showed significantly higher PD-1^T TIL numbers in the DC 6m group versus the PD group ($P<0.01$), whereas no significant difference was observed for TIS scores ($P=0.52$) **(Fig. S2G,H)**. Although the discriminatory ability of PD-1^T TILs was lower compared to the training, it still reached an AUC of 0.72 (95% CI: 0.57-0.87) **(Fig. S2E)**. A cut-off value of 90 PD-1^T TILs per mm^2 resulted in a sensitivity of 72%, specificity of 74%, NPV of 86% and PPV of 54%. Conversely, TIS obtained a low AUC of 0.57 (95% CI: 0.36-0.77) **(Fig. S2F)**. 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 the combination of CD8+IT-CD8 and CD3+IT-CD8 did not improve predictive accuracy compared to the standalone use of PD-1^T TILs and TIS. Furthermore, none of the selected biomarkers met the prespecified performance criteria.

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-value	Training cohort n=55	Validation cohort n=80
Sex	*P*=1.00		
Male, no.(%)		30 (55%)	44 (55%)
Female, no.(%)		25 (45%)	36 (45%)
Age (years), mean (SD)	*P*=0.20	62 (10.1)	65 (7.5)
Smoking (never/ex/current)	*P*=0.64	5/44/6	12/51/17
Pack years, median (IQR)	*P*=0.90	29 (20)	30 (28)
PS, no. (%)	*P*=0.46		
0		16 (29%)	16 (20%)
1		29 (53%)	50 (62%)
≥2		10 (18%)	14 (18%)
Pathology, no.(%)	*P*=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. (%)	*P*=0.86		
KRAS positive		19 (35%)	30 (38%)
PD-L1 TPS, no. (%)			
Negative <1%	*P*=1.00	30 (55%)	43 (54%)
Positive ≥1%		25 (45%)	36 (45%)
Negative <50%	*P*=0.66	43 (78%)	65 (81%)
Positive ≥50%		12 (22%)	14 (18%)
Unknown		0 (0%)	1 (1%)
Brain metastases, no. (%)	*P*=0.67	13 (24%)	16 (20%)
Line of treatment, no.(%)	*P*=0.63		
1		0 (0%)	1 (1%)
2		42 (76%)	56 (70%)
>2		13 (24%)	23 (29%)
Best Overall Response	*P*=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	*P*=0.85	16 (29%)	25 (31%)
at 12 months	*P*=0.83	12 (22%)	16 (20%)
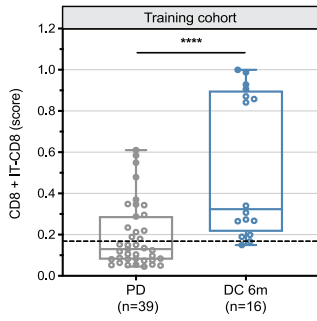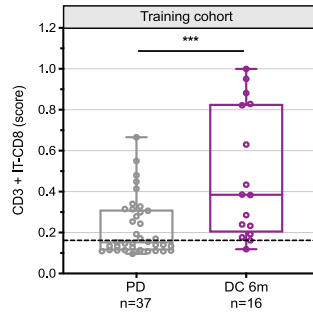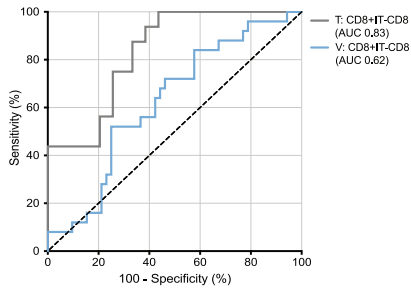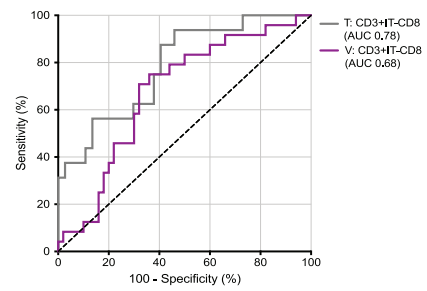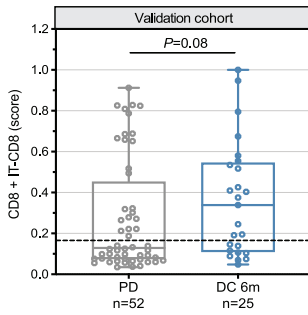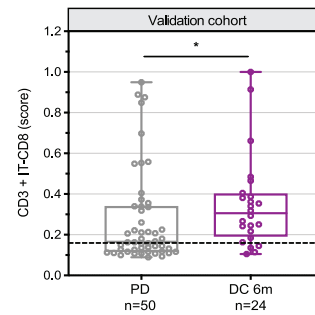
◄ **Figure 2.** Performance of selected composite and individual biomarkers to predict DC at 6 months in NSCLC patients treated with PD-1 blockade. (A) Probability scores of CD8+IT-CD8 in pretreatment samples from patients with disease control at 6 months (DC 6m) (n=16) and progressive disease (PD) (n=39) in the training cohort (n=55). Dashed line indicates a cut-off of 0.167. Medians, interquartile ranges and minimum/maximum shown in boxplots, ****P<0.0001 by Mann Whitney U-test. (B) Probability scores of CD3+IT-CD8 in pretreatment samples from patients with DC 6m (n=16) and PD (n=37) in the training cohort (n=53). Dashed line indicates a cut-off of 0.161. Medians, interquartile ranges and minimum/maximum shown in boxplots, ***P<0.001 by Mann Whitney U-test. (C) Receiver operating characteristic (ROC) curve for predictive value of CD8+IT-CD8 for DC 6m in the training (n=55) (AUC 0.83; 95% CI: 0.73-0.94) and validation cohort (n=77) (AUC 0.62; 95% CI: 0.50-0.75). (D) ROC curve for predictive value of CD3+IT-CD8 for DC 6m in the training (n=53) (AUC 0.78; 95% CI: 0.65-0.91) and validation cohort (n=74) (AUC 0.68; 95% CI: 0.55-0.80). (E) Same plot as in A (CD8+IT-CD8) for patients with DC 6m (n=25) and PD (n=52) in the validation cohort (n=77), P=0.08. (F) Same plot as in B (CD3+IT-CD8) for patients with DC 6m (n=24) and PD (n=50) in the validation cohort (n=74), *P=0.02.

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

Clinical outcome	Bio-marker	Predictor	Cut-off	Samples (n)	Training AUC (95% CI)
DC 6 months	PD-1^T TILs		90	42	0.82 (0.69-0.95)
	TIS		6.65	28	0.81 (0.65-0.98)
	CD8+ IT-CD-8	$Probability\ for\ DC = \dfrac{1}{1 - \exp(-3.5749 + 0.0031 * CD8 + 0.043 * IT - CD8)}$	0.167	55	0.83 (0.73-0.94)
	CD3+ IT-CD-8	$Probability\ for\ DC = \dfrac{1}{1 - \exp(-2.3821 + 0.0806 * CD3 + 0.0175 * IT - CD8 + 0.0069 * CD3 * IT - CD8)}$	0.161	53	0.78 (0.65-0.91)
DC 12 months	PD-1^T TILs		90	42	0.82 (0.70-0.94)
	TIS		6.65	28	0.77 (0.58-0.96)
	CD8+ IT-CD-8	$Probability\ for\ DC = \dfrac{1}{1 - \exp(-4.0644 + 0.003 * CD8 + 0.0436 * IT - CD8)}$	0.122	55	0.85 (0.73-0.96)
	CD8+ TIS	$Probability\ for\ DC = \dfrac{1}{1 - \exp(-5.7952 + 0.0224 * CD8 + 0.2346 * TIS + -0.0021 * CD8 * TIS)}$	0.124	28	0.91 (0.79-1.00)

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

Approximately 70-80% of patients undergoing second-line treatment with PD-1/PD-L1 blockade experience disease progression within 12 months[2-4]. Our previous work demonstrated the superior efficacy of PD-1^T TILs in identifying patients with DC at 12 months (DC 12m) compared to those with DC 6m, as well as identifying a subgroup without long-term benefit[25]. In light of this, we extended our analysis to evaluate the predictive performance of all biomarkers for DC 12m. Similar to the analysis performed for DC 6m, we constructed ROC curves to determine optimal cut-off values corresponding to a sensitivity of ≥90% and specificity of ≥50% for each composite and individual biomarker. Four patients in the training and nine patients in the validation cohort experienced disease progression between 6 and 12 months, allocating them into the PD group. In the training cohort, 16 composite biomarkers, along with individual biomarkers PD-1^T TILs and TIS, met the prespecified sensitivity and specificity ciritera **(Table S3)**. PD-1^T TIL and TIS combinations did not significantly improve predictive accuracy (**Fig. S3A,B** and **Table S3)**. Noteworthy, the combination of CD8 with TIS (CD8+TIS) showed an 18% increase in specificity compared to TIS

3

Training (Continued)					Validation				
Sensi-tivity	Speci-ficity	NPV	PPV	Samples (n)	AUC (95%-CI)	Sensi-tivity	Speci-ficity	NPV	PPV
92%	67%	95%	52%	61	0.72 (0.57-0.87)	72%	74%	86%	54%
100%	55%	100%	47%	40	0.57 (0.36-0.77)	83%	39%	84%	37%
94%	62%	96%	50%	77	0.62 (0.50-0.75)	64%	56%	76%	41%
94%	54%	95%	47%	74	0.68 (0.55-0.80)	83%	46%	85%	43%
100%	64%	100%	43%	61	0.80 (0.65-0.94)	86%	74%	95%	50%
100%	50%	100%	35%	40	0.63 (0.43-0.82)	100%	39%	100%	26%
92%	63%	96%	41%	77	0.67 (0.53-0.81)	68%	57%	88%	30%
100%	68%	100%	46%	38	0.59 (0.36-0.82)	29%	68%	81%	17%

alone. Despite not reaching statistical significance, possibly due to the low sample size (*P*=0.34), this combination was selected for further analysis **(Fig. S3B**, **Table S3)**. Other selected biomarkers for validation with the highest predictive performance included CD8+IT-CD8, PD-1^T TILs and TIS **(Table S3)**.

The probability scores for DC 12m and PD are depicted per sample in **Figure S3C** (CD8+IT-CD8, *P*<0.001) and **Figure 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, and optimal cut-off values of 0.122 and 0.124, respectively, were chosen **(Fig. 3A,B** and **Table 3)**. In the validation cohort, only CD8+IT-CD8 showed borderline significantly higher probability scores in the DC 12m group versus the PD group (*P*=0.03) **(Fig. S3E,F)**. The ROCs resulted in 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 and TIS scores for the training cohort are presented in **Figure S3G** and **H**. A cut-off value of 90 PD-1^T TILs per mm^2 and a TIS score of 6.65 demonstrated similar predictive accuracy as observed in the training cohort for DC 6m **(Fig. 3C,D and Table 3)**. In the validation cohort, PD-1^T TIL numbers were significantly higher in patients with DC 12m versus PD (*P*<0.001) **(Fig. 3E)**. PD-1^T TILs maintained a consistently high AUC of 0.80 (95% CI: 0.65-0.94) and demonstrated good performance (sensitivity: 86%, specificity: 74%, NPV: 95% and PPV: 50%) **(Fig. 3C** and **Table 3)**. A subgroup analysis revealed an enrichment of patients with DC 12m in the PD-1^T TILs ≥90 group and patients with PD in the PD-1^T TILs <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)**. Nevertheless, a cut-off score of 6.65 reached a sensitivity of 100%, specificity of 39%, NPV of 100% and PPV of 26% **(Table 3)**. Although these findings did not meet the prespecified ≥50% specificity criterium, they accurately identified all patients with DC 12m including 39% of patients with PD. Taken together, PD-1^T TILs and TIS, as individual biomarkers, showed superior predictive accuracy for DC 12m compared to CD8+IT-CD8 and CD8+TIS. Notably, PD-1^T TILs alone demonstrated greater performance than TIS alone, particularly in terms of specificity and PPV.

Figure 3. Performance of selected composite and individual biomarkers to predict 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 PD-1^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. ▶
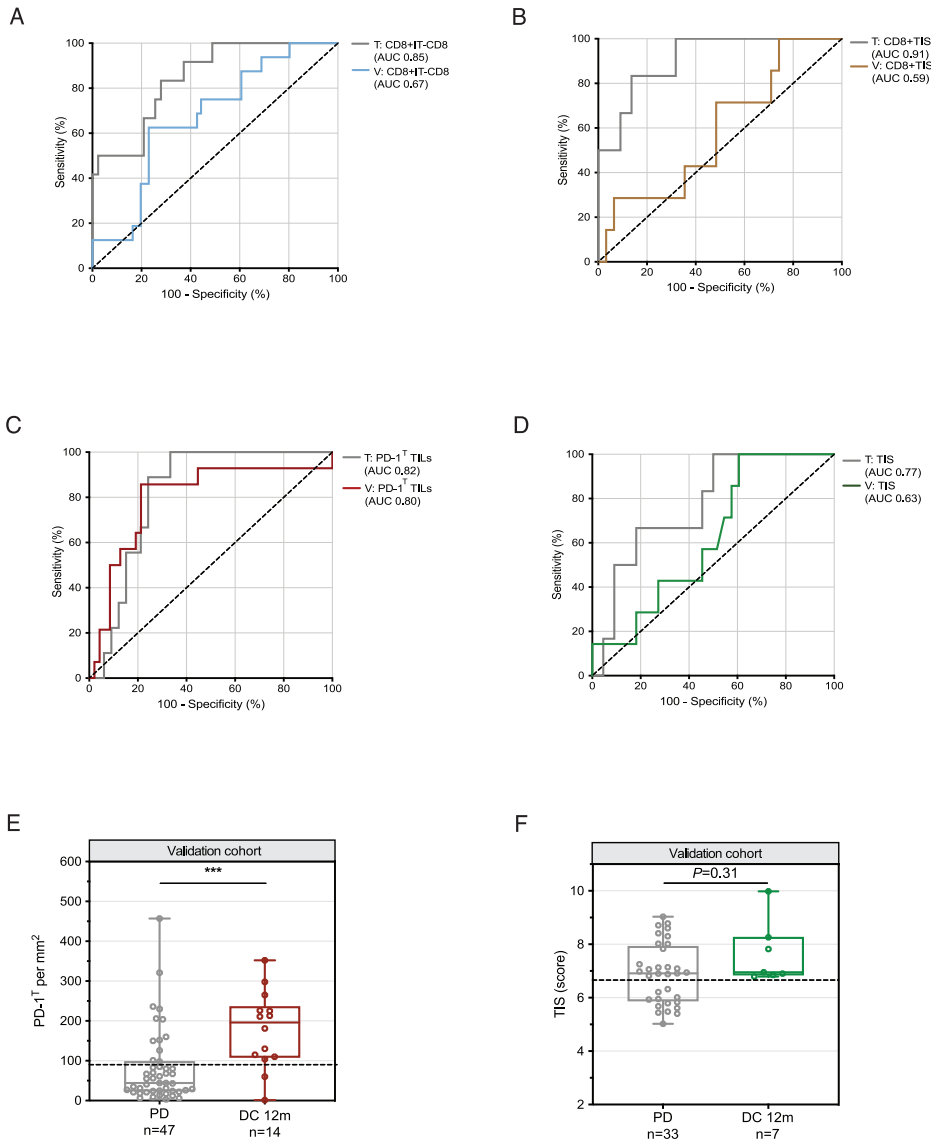
Discussion

Since the introduction of PD-1/PD-L1 blockade therapy, the clinical outcomes for advanced stage NSCLC have markedly improved. Nevertheless, a minor subset of patients derive benefit from these treatments, leading to concerns of overtreatment and unnecessary side effects. In addition, healthcare systems deal with increasing costs. Several predictive biomarkers have been identified to support treatment decision-making. Given the complex interactions within the TME influencing the immune response during PD-1/PD-L1 blockade therapy, the likelihood of identifying a single perfect biomarker is minimal. Here, we evaluate the predictive performance of combinations of biomarkers in a cohort of advanced-stage NSCLC patients treated with nivolumab. Our findings show that the selected composite biomarkers did not improve predictive performance compared to PD-1^T TILs and TIS alone. At 6 months, none of the selected biomarkers met the prespecified criteria of ≥90% sensitivity and ≥50% specificity in the validation cohort. However, at 12 months, PD-1^T TILs and TIS demonstrated a high sensitivity in identifying patients with DC 12m with. Patients without long-term benefit were more accurately identified by PD-1^T TILs than TIS.

While CD8 or CD3 TILs, in combination with intratumoral localization of CD8 T cells, emerged as the most accurate composite biomarkers for DC 6m in the training cohort, their discriminatory ability was notably low in the validation cohort. The mere presence and localization of TILs may not indicate that all T cells are in a state to recognize and eliminate the tumor[37,38]. In the present study, this notion is supported by the high predictive accuracy of PD-1^T TILs for DC 12m, given that these TILs constitute a distinct subset with an enhanced capacity for tumor recognition[24]. The consistency of these findings with our previous work can be attributed to the predominant reuse of samples[25]. Further refinement of this specific T cell population holds promise for the development of new markers or gene signatures, as shown by various studies[39–41]. Moreover, we recently developed a clinically applicable mRNA expression signature reflecting the presence of PD-1^T TILs in the TME[35]. Since most biomarkers assessed in this study are associated with 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 the training cohort, the CD8+PD-L1 combination failed to meet the sensitivity and specificity criteria, leading to its exclusion from further evaluation. PD-L1 expression on tumor cells is transient and relies on the production of IFNγ by TILs[42]. This dynamic nature of PD-L1 expression could contribute to variable tumor PD-L1 expression in the training samples, potentially impacting the predictive

accuracy of PD-L1 and of PD-L1 combinations. Furthermore, it is important to note that this study is limited by the number of samples, particularly for TIS assessment. Therefore, we restricted our evaluation to two-biomarker combinations instead of considering three or more. Studies with a larger sample size are essential to validate the robustness of our findings.

Our results for TIS align with other studies that have demonstrated the predictive potential of this signature[29,43]. Interestingly, TIS includes genes, such as *LAG3* and *TIGIT*, that are highly expressed in PD-1^T TILs[24,29]. A high number of PD-1^T TILs or a high TIS score in pretreatment samples may serve as surrogate markers for a tumor's capacity to undergo durable immune reactivation upon PD-1 blockade therapy. A PD-1^T TILs or TIS combination with biomarkers representing distinct facets of the immune response holds promise for improving predictive accuracy. For instance, TMB can serve as a read-out for immunogenic neoantigens arising from somatic mutations[15]. Previous studies have identified TMB and PD-L1 as independent predictors for advanced NSCLC treated with PD-1 blockade and have shown improved performance when combined[32,33]. Conversely, the presence of tumor-resident regulatory T cells (T_{reg}) in the TME might be considered. T_{reg} cells, known for their immune-inhibitory functions, are associated with poor patient survival when present in high numbers[44]. A combination of TMB or T_{reg} with either PD-1^T TILs or TIS warrants further investigation in future studies.

In conclusion, this study showed that the biomarker combinations assessed here did not improve predictive performance compared to PD-1^T TILs and TIS alone. PD-1^T TILs showed the highest predictive performance of all the biomarkers, accurately identifying patients without long-term benefit with high specificity and NPV.

Acknowledgements

References

1.   Reck, M. *et al.* Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer. *N Engl J Med.* 2016;375(**19**):1823–33.
2.   Borghaei, H. *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. *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. *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, E.B. *et al.* Pembrolizumab for the treatment of non-small-cell lung cancer. *N Engl J Med.* 2015;372(**21**):2018–28.
6.   Reck, M. *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, E.B. *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, E.F., Mok, T., Hellmann, M.D. 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. *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.
10. Gniadek, T.J., Li, Q.K., 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, A.M. *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. *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.
13. McLaughlin, 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. *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, N.A. *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. *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, L.M. *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, P.C. *et al.* PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature.* 2014;515(**7528**):568–71.

19. van der Leun, A.M., Thommen, D.S., Schumacher, T.N. CD8+ T cell states in human cancer: insights from single-cell analysis. *Nat Rev Cancer.* 2020;20(**4**):218–32.

20. Huang, A.C. *et al.* T-cell invigoration to tumour burden ratio associated with anti-PD-1 response. *Nature.* 2017;545(**7652**):60–5.

21. Le, D.T. *et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med.* 2015;372(**26**):2509–20.

22. Fumet, J.D. *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.

23. Hu-Lieskovan, S. *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, D.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**).

25. Hummelink, K. *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.

26. Cabrita, R. *et al.* Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature.* 2020;(February).

27. Helmink, B.A. *et al.* B cells and tertiary lymphoid structures promote immunotherapy response. *Nature.* 2020;(February).

28. Petitprez, F. *et al.* B cells are associated with survival and immunotherapy response in sarcoma. *Nature.* 2020;577(June 2018).

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

30. Fehrenbacher, L. *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.

31. Higgs, B.W. *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. *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, M.D. *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. *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. Hummelink, K. *et al.* A dysfunctional T-cell gene signature for predicting nonresponse to PD-1 blockade in non-small cell lung cancer that is suitable for routine clinical diagnostics. *Clin Cancer Res.* 2024 Feb 16;30(**4**):814-823.

36. Danaher, P. *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.

37. Scheper, W. *et al.* Low and variable tumor reactivity of the intratumoral TCR repertoire in human cancers. *Nat Med.* 2019;25(**1**):89–94.

38. Simoni, Y. *et al.* Bystander CD8+ T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature.* 2018;557(**7706**):575–9.

39. Oliveira, G. *et al.* Phenotype, specificity and avidity of antitumour CD8+ T cells in melanoma. *Nature.* 2021;596(**7870**):119–25.

40. Caushi, J.X. *et al.* Transcriptional programs of neoantigen-specific TIL in anti-PD-1-treated lung cancers. Vol. 596, *Nature.* Springer US; 2021. 126–132 p.

41. Lowery, F.J. *et al.* Molecular signatures of antitumor neoantigen-reactive T cells from metastatic human cancers. *Science.* 2022;884(February):eabl5447.

42. Noguchi, T. *et al.* Temporally distinct PD-L1 expression by tumor and host cells contributes to immune escape. *Cancer Immunol Res.* 2017;5(**2**):106–17.

43. Damotte, D. *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.

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

Supplemental material
Supplemental figures
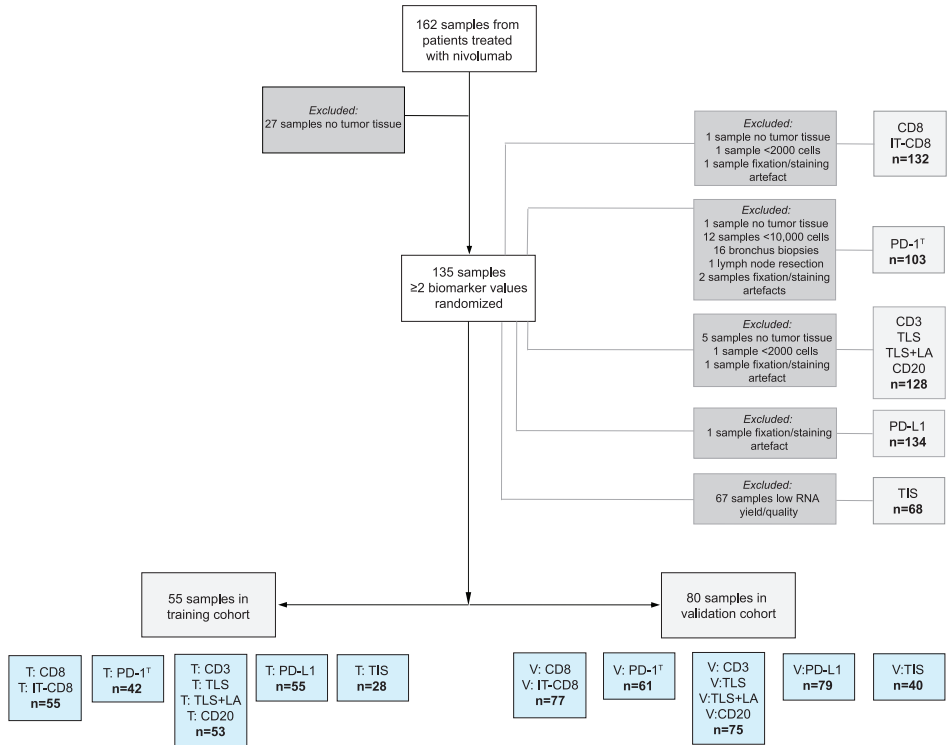
Figure S1. Flow chart of sample numbers and exclusion criteria per biomarker. In 27 samples none of the biomarkers were assessed. The right grey boxes indicate the exclusion criteria per biomarker and the right light grey boxes indicate the total number of samples that were assessed per biomarker. The remaining samples with ≥2 biomarker values were randomized in a training (n=55) and validation cohort (n=80). The blue boxes indicate the number of samples that were assessed per biomarker in the training (T) and validation (V) cohort.
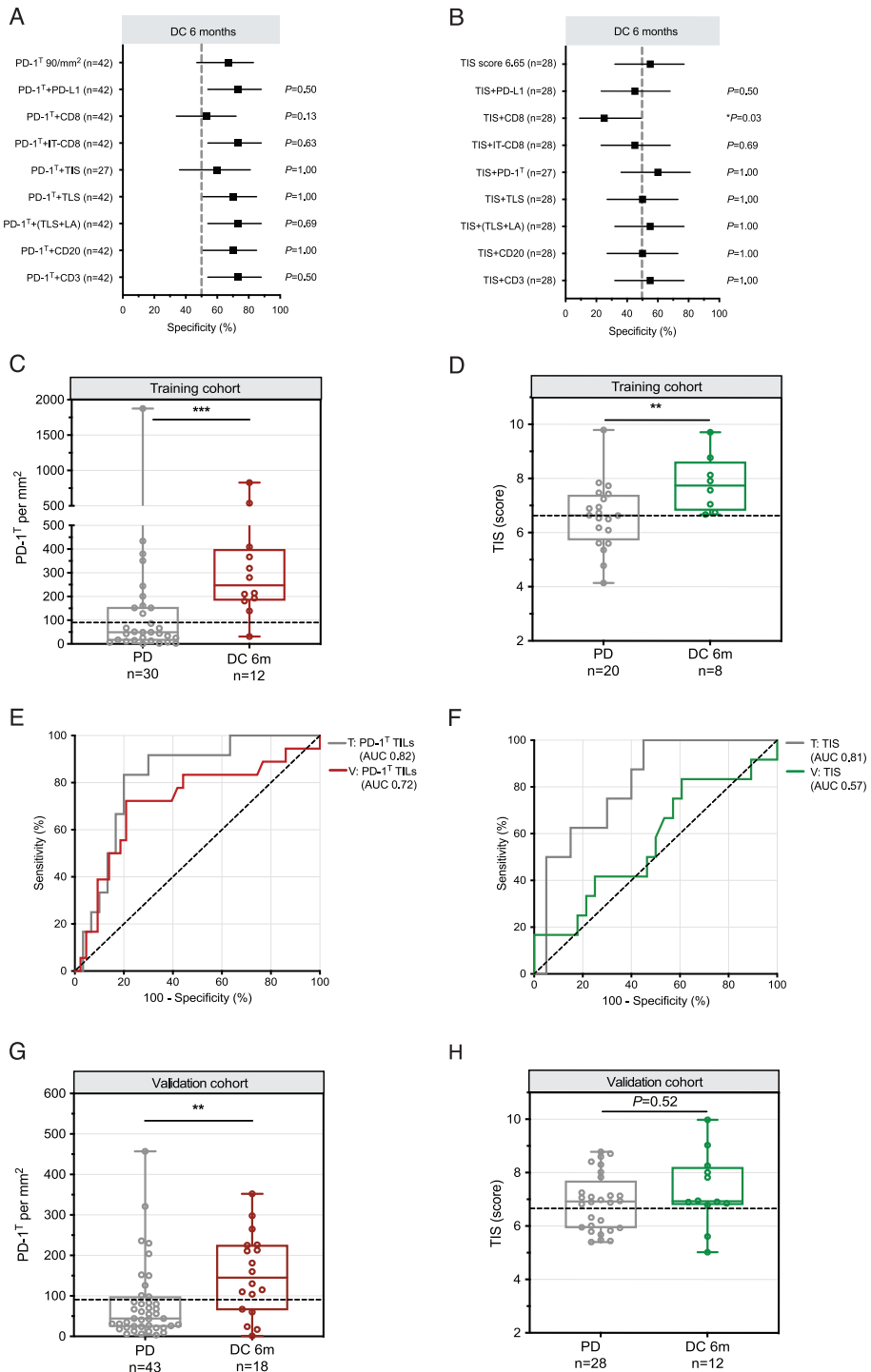
◄ **Figure S2.** Performance of PD-1^T TILs and TIS as individual and as composite biomarker to predict disease control at 6 months (DC 6m) in NSCLC patients treated with PD-1 blockade. (A) Specificity correlating to a sensitivity of ≥90% for combinations with PD-1^T TILs as predictive biomarker for DC 6m in the training cohort (n=27 or n=42). The grey dashed line indicates the prespecified specificity criterium of ≥50%. Different composite biomarkers were compared to the predictive performance of PD-1^T TILs alone. *P* values were calculated by McNemar test. (B) Same plot as in A for combinations with TIS in the training cohort (n=27 or n=28). (C) PD-1^T TILs per mm^2 in pretreatment samples from patients with DC 6m (n=12) and progressive disease (PD) (n=30) in the training cohort (n=42). 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<0.001 by Mann Whitney U-test. (D) TIS scores in pretreatment samples from patients with DC 6m (n=8) and PD (n=20) in the training cohort (n=28). Dashed line indicates a cut-off score of 6.65. Medians, interquartile ranges and minimum/maximum shown in boxplots, **P<0.01 by Mann Whitney U-test. (E) Receiver operating characteristic (ROC) curve for predictive value of PD-1^T TILs for DC 6m in the training cohort (n=42) (AUC 0.82; 95% CI: 0.69-0.95) and validation cohort (n=61) (AUC 0.72; 95% CI: 0.57-0.87). (F) ROC curve for predictive value of TIS for DC 6m in the training cohort (n=28) (AUC 0.81; 95% CI: 0.65-0.98) and validation cohort (n=40) (AUC 0.57; 95% CI: 0.36-0.77) (G) Same plot as in C (PD-1^T TILs) for patients with DC 6m (n=18) and PD (n=43) in the validation cohort (n=61), **P<0.01 by Mann Whitney U-test. (H) Same plot as in D (TIS) for patients with DC 6m (n=12) and PD (n=28), *P*=0.52 by Mann Whitney U-test.

3
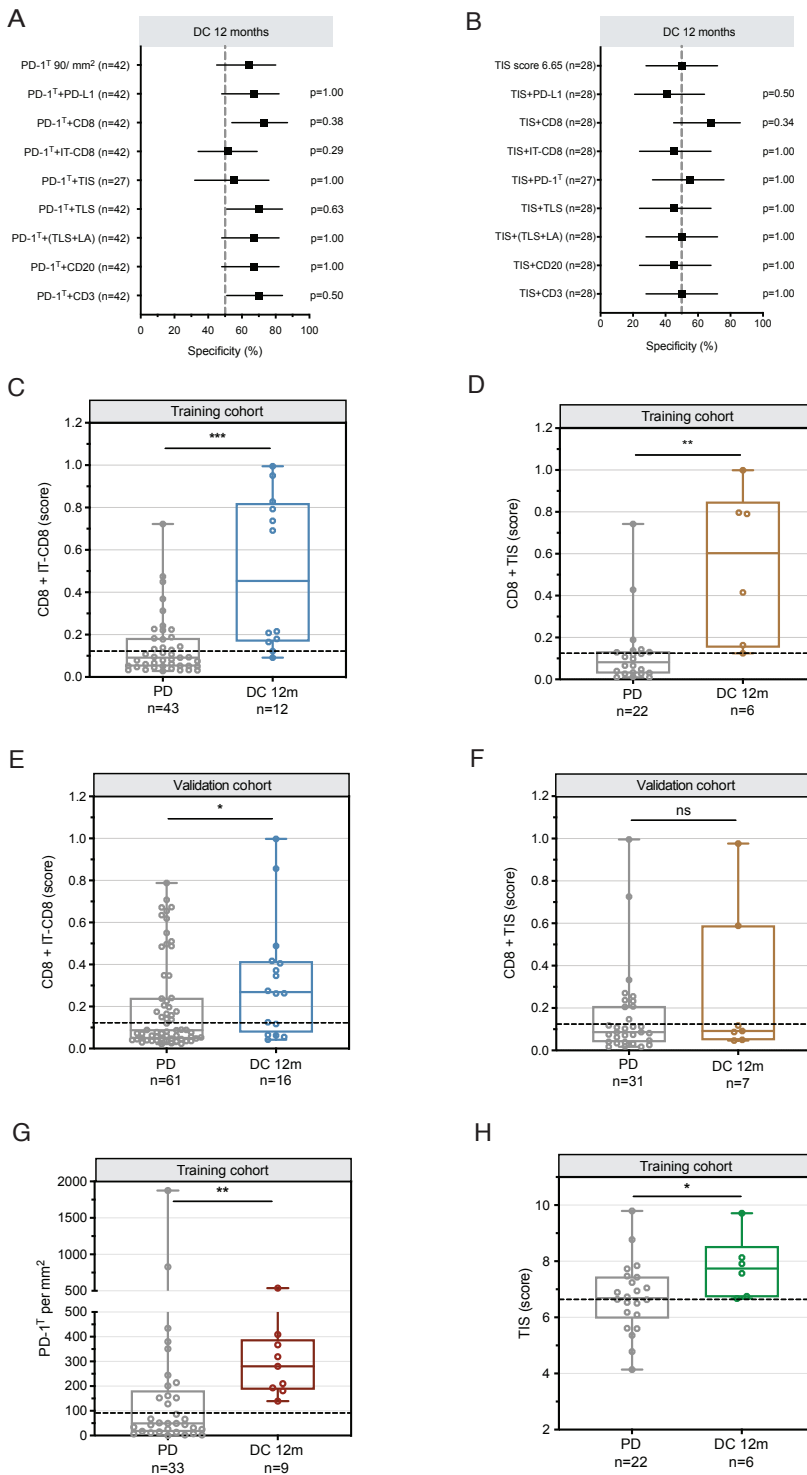
◄ **Figure S3.** Performance of PD-1^T TILs and TIS as individual and as composite biomarker to predict disease control at 12 months (DC 12m) and association of CD8+IT-CD8 and CD8+TIS with DC 12m in NSCLC patients treated with PD-1 blockade. (A) Specificity correlating to a sensitivity and NPV of ≥90% for combinations with PD-1^T TILs as predictive biomarker for DC 12m in the training cohort (n=27 or n=42). The grey dashed line indicates the prespecified specificity criterium of ≥50%. Different composite biomarkers were compared to the predictive performance of PD-1^T TILs alone. *P* values were calculated by McNemar test. (B) Same plot as in A for combinations with TIS in the training cohort (n=27 or n=28). (C) Probability scores of CD8+IT-CD8 in pretreatment samples from patients with DC 12m (n=12) and progressive disease (PD) (n=43) in the training cohort (n=55). Dashed line indicates a cut-off of 0.122. Medians, interquartile ranges and minimum/maximum shown in boxplots, ***P*<0.001 by Mann Whitney U-test. (D) Probability scores of CD8+TIS in pretreatment samples from patients with DC 12m (n=6) and PD (n=22) in the training cohort (n=28). Dashed line indicates a cut-off of 0.124. Medians, interquartile ranges and minimum/maximum shown in boxplots, ***P*<0.01 by Mann Whitney U-test. (E) Probability scores of CD8+IT-CD8 in pretreatment samples from patients with DC 12m (n=16) and PD (n=61) in the validation cohort (n=77). Dashed line indicates a cut-off of 0.122. Medians, interquartile ranges and minimum/maximum shown in boxplots, *P*=0.03 by Mann Whitney U-test. (F) Probability scores of CD8+TIS in pretreatment samples from patients with DC 12m (n=7) and PD (n=31) in the validation cohort (n=38). Dashed line indicates a cut-off of 0.124. Medians, interquartile ranges and minimum/maximum shown in boxplots, *P*=0.48 by Mann Whitney U-test. (G) PD-1^T TILs per mm^2 in pretreatment samples from patients with DC 12m (n=9) and PD (n=33) in the training cohort (n=42). 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. (H) TIS scores in pretreatment samples from patients with DC 12m (n=6) and PD (n=22) in the training cohort (n=28). Dashed line indicates a cut-off score of 6.65. Medians, interquartile ranges and minimum/maximum shown in boxplots, *P*=0.04 by Mann Whitney U-test.

Supplementary Tables

Table S1. Overview of all analyzed biomarkers per patient.

Sample ID	1 = DC 6m 0= PD	1 = DC 12m, 0= PD	T= Training, V= Validation	CD8 TILs per mm²	IT-CD8 T cells (%)	PD-1⁺TILs per mm²	CD3⁺ positive area per mm²	TLS per mm²	TLS+LA per mm²	CD20⁺ positive area per mm²
NKI-001	1	1	T	400	32.08	NA	1.28	0.00	0.00	0.15
NKI-002	0	0	T	287	49.12	152	1.79	0.12	0.12	0.19
NKI-003	0	0	T	45	23.33	44	1.07	0.00	0.00	0.09
NKI-004	0	0	T	552	24.65	NA	5.17	0.00	0.72	1.65
NKI-005	1	1	T	763	89.87	537	3.36	0.00	0.00	0.63
NKI-006	1	0	T	1273	42.31	828	9.85	0.21	0.43	1.87
NKI-007	0	0	T	295	22.62	434	3.85	0.45	0.63	3.41
NKI-008	0	0	T	336	3.22	34	2.33	0.22	0.77	1.89
NKI-009	0	0	T	164	22.06	NA	0.79	0.00	0.00	0.19
NKI-010	1	1	T	121	51.75	210	4.62	0.00	0.00	0.25
NKI-011	0	0	T	914	16.68	1875	14.26	0.89	2.98	11.07
NKI-012	0	0	T	101	20.42	67	0.46	0.00	0.00	0.10
NKI-013	0	0	T	146	32.68	NA	NA	NA	NA	NA
NKI-014	0	0	T	197	10.14	25	2.24	0.00	0.65	0.79
NKI-015	0	0	T	794	33.63	351	4.66	0.00	0.00	0.07
NKI-016	0	0	T	140	26.85	NA	0.34	0.00	0.00	0.05
NKI-017	1	1	T	1903	62.5	280	8.37	0.03	0.33	1.40
NKI-018	0	0	T	435	25.44	201	4.40	0.00	0.21	0.78
NKI-019	1	1	T	1443	21.07	409	6.88	0.48	1.28	2.03
NKI-020	1	1	T	378	24.56	NA	6.78	0.00	0.00	0.26
NKI-021	1	1	T	320	20.33	367	3.96	0.28	0.44	1.08
NKI-022	0	0	T	127	28.97	43	1.04	0.01	0.04	0.17
NKI-023	0	0	T	300	17.79	161	2.46	0.00	0.00	0.47
NKI-024	0	0	T	225	0.37	26	1.59	0.08	0.58	1.43
NKI-025	0	0	T	96	8.73	17	1.32	0.00	0.00	0.19
NKI-026	0	0	T	371	43.62	NA	1.77	0.00	0.00	0.13
NKI-027	0	0	T	14	48.35	13	0.25	0.00	0.00	0.00
NKI-028	0	0	T	928	1.09	NA	3.28	0.00	0.58	1.22
NKI-029	0	0	T	372	16.67	NA	7.04	0.48	0.48	1.28
NKI-030	0	0	T	182	79.91	128	1.89	0.00	0.00	0.02
NKI-031	0	0	T	111	17.11	7	0.16	0.00	0.00	0.00

PD-L1 TPS (%)	TIS score	CD8 TILs+ IT CD8 T cells (probability score, DC 6m)	CD3 TILs+ IT CD8 T cells (probability score, DC 6m)	CD8 TILs+ TIS (probability score, DC 12m)	CD8 TILs+ IT CD8 T cells (probability score, DC 12m)
0	NA	0.2649983	0.1919443	NA	0.1652011
40	NA	0.343007	0.3148431	NA	0.2265712
1	NA	0.0803345	0.1521966	NA	0.0509736
0	NA	0.2957622	0.3404191	NA	0.1824023
0	NA	0.9058749	0.8216733	NA	0.7926359
0	8.77	0.8708422	0.8815818	0.7421676	0.7221582
0	6.63	0.1520565	0.2535239	0.1430865	0.0932314
0	7.43	0.0839552	0.1103969	0.1380421	0.0504621
60	NA	0.1057568	0.1401703	NA	0.0657138
5	NA	0.2669763	0.6301201	NA	0.1798485
0	NA	0.4787394	0.665758	NA	0.3126318
90	7.47	0.0836776	0.1273957	0.0328116	0.052357
0	NA	0.1490646	NA	NA	0.0946666
1	6.89	0.0735271	0.1336737	0.0659117	0.0448887
40	9.79	0.5492985	0.4145797	0.0975557	0.3686506
0	NA	0.1186654	0.1390245	NA	0.0745653
95	7.91	0.9873589	0.9511964	0.9987685	0.9508837
10	7.84	0.2341446	0.3066154	0.1885522	0.1436726
0	NA	0.8416185	0.3855678	NA	0.6911857
0	8.13	0.1992826	0.4340513	0.1244547	0.1220825
80	6.75	0.149622	0.2393714	0.163071	0.0912642
0	6.73	0.1240833	0.1699447	0.0396049	0.0784285
70	NA	0.1298441	0.1717509	NA	0.0790533
0	5.60	0.054467	0.0959122	0.1070282	0.0329746
60	6.50	0.0519602	0.1146254	0.0308767	0.0320816
60	NA	0.346268	0.2791512	NA	0.2241756
0	5.36	0.1888025	0.1923131	0.012327	0.127833
0	NA	0.3486183	0.1116267	NA	0.2205889
20	NA	0.1505913	0.327581	NA	0.0911631
0	6.18	0.5844484	0.5499894	0.0653657	0.4496429
0	NA	0.0756291	0.113873	NA	0.0470361

Table S1. Continued

Sample ID	1 = DC 6m 0= PD	1 = DC 12m, 0= PD	T= Training, V= Validation	CD8 TILs per mm²	IT-CD8 T cells (%)	PD-1^{IT}TILs per mm²	CD3$^+$ positive area per mm²	TLS per mm²	TLS+LA per mm²	CD20$^+$ positive area per mm²
NKI-032	1	1	T	853	91.49	181	3.33	0.05	0.16	0.51
NKI-033	0	0	T	17	13.37	15	0.27	0.00	0.00	0.04
NKI-034	0	0	T	269	35.36	152	4.83	0.00	0.00	0.41
NKI-035	0	0	T	5	15.38	2	0.08	0.00	0.00	0.00
NKI-036	0	0	T	19	10.64	6	0.59	0.00	0.00	0.03
NKI-037	0	0	T	191	81.87	86	1.36	0.00	0.00	0.06
NKI-038	1	0	T	75	55.56	214	2.03	0.29	0.58	1.86
NKI-039	0	0	T	260	14.52	49	1.44	0.01	0.09	0.31
NKI-040	0	0	T	282	20.08	244	1.23	0.00	0.00	0.02
NKI-041	0	0	T	376	45.79	66	2.03	0.00	0.00	0.05
NKI-042	1	1	T	2731	95.76	319	11.77	0.07	0.31	1.31
NKI-043	0	0	T	513	11.19	NA	3.11	0.00	0.64	1.02
NKI-044	1	1	T	1695	1.8	NA	7.57	0.00	1.46	2.29
NKI-045	0	0	T	157	20.4	49	1.24	0.00	0.00	0.03
NKI-046	0	0	T	241	11.41	NA	NA	NA	NA	NA
NKI-047	0	0	T	330	29.98	51	2.53	0.01	0.07	0.20
NKI-048	1	1	T	152	54.29	193	1.14	0.00	0.00	0.01
NKI-049	0	0	T	52	13.3	11	0.49	0.00	0.00	0.10
NKI-050	0	0	T	507	27.05	380	3.97	0.13	0.33	1.33
NKI-051	0	0	T	85	14.47	2	0.43	0.00	0.00	0.07
NKI-052	1	1	T	473	35.65	139	1.75	0.02	0.14	0.28
NKI-053	1	0	T	595	6.53	NA	2.10	0.00	0.00	0.42
NKI-054	1	0	T	146	35.89	31	0.66	0.00	0.13	0.21
NKI-055	0	0	T	24	25.37	0	0.24	0.00	0.00	0.02
NKI-056	1	1	V	148	18.58	265	1.35	0.00	0.14	0.22
NKI-057	0	0	V	220	9.64	67	6.20	0.00	0.48	2.73
NKI-058	0	0	V	66	13.17	25	0.32	0.00	0.35	0.57
NKI-059	1	0	V	94	61.54	24	0.33	0.00	0.00	0.00
NKI-060	0	0	V	1105	67.42	NA	5.81	0.00	0.20	0.67
NKI-061	0	0	V	132	20.83	70	1.06	0.00	0.00	0.10
NKI-062	0	0	V	274	18.63	80	2.59	0.19	0.50	2.62
NKI-063	0	0	V	107	30.19	83	0.87	0.00	0.00	0.03
NKI-064	0	0	V	599	44.88	NA	2.39	0.00	0.00	0.10

3

PD-L1 TPS (%)	TIS score	CD8 TILs+ IT CD8 T cells (probability score, DC 6m)	CD3 TILs+ IT CD8 T cells (probability score, DC 6m)	CD8 TILs+ TIS (probability score, DC 12m)	CD8 TILs+ IT CD8 T cells (probability score, DC 12m)
1	7.57	0.9281668	0.8280812	0.7964289	0.8278642
0	4.14	0.0498061	0.1088811	0.0100224	0.0312862
1	6.53	0.2192833	0.4490073	0.1221511	0.138352
0	4.78	0.0522088	0.1092175	0.0098296	0.032955
0	NA	0.0448136	0.1085343	NA	0.0280398
60	NA	0.609898	0.4800232	NA	0.474431
0	7.05	0.2733109	0.3833487	0.026989	0.1878513
0	6.09	0.1035125	0.1336286	0.1290324	0.0630203
10	6.64	0.1344303	0.1463986	0.1299422	0.0822057
60	NA	0.3698593	0.3139134	NA	0.2416199
100	9.71	0.9994594	0.9996528	0.7902098	0.9947135
0	NA	0.1793685	0.1547866	NA	0.1078333
95	NA	0.8578328	0.1614933	NA	0.7372112
0	7.24	0.0974757	0.1477143	0.0476186	0.0604833
10	NA	0.0874019	NA	NA	0.0531643
40	7.73	0.2121404	0.243267	0.1180728	0.1316529
95	NA	0.3066719	0.2853273	NA	0.20799
0	NA	0.05496	0.1124674	NA	0.0343275
1	6.94	0.2877861	0.2986536	0.4280079	0.1778782
0	5.61	0.0633189	0.1138288	0.0268781	0.039405
0	6.67	0.3402331	0.2330759	0.4151931	0.2153227
0	NA	0.1896019	0.1186925	NA	0.1139911
0	NA	0.1670231	0.1765352	NA	0.1068571
0	NA	0.0821614	0.1326009	NA	0.0525204
1	6.85	0.0887281	0.1446729	0.0461784	0.0549907
0	7.06	0.0769977	0.2134284	0.0749965	0.0468721
40	NA	0.056926	0.1093362	NA	0.0354729
90	NA	0.3382343	0.2419572	NA	0.237746
80	NA	0.9118205	0.8752877	NA	0.7878656
0	7.14	0.0925615	0.1440943	0.0404469	0.0576752
0	6.94	0.1249519	0.1798511	0.1118855	0.0763194
0	6.88	0.1232627	0.1672965	0.0339259	0.0783425
100	NA	0.5171948	0.3383498	NA	0.3474269

Table S1. Continued

Sample ID	1 = DC 6m 0= PD	1 = DC 12m, 0= PD	T=Training, V=Validation	CD8 TILs per mm²	IT-CD8 T cells (%)	PD-1ᵀTILs per mm²	CD3⁺ positive area per mm²	TLS per mm²	TLS+LA per mm²	CD20⁺ positive area per mm²
NKI-065	1	1	**V**	448	44.14	115	3.20	0.14	0.28	0.45
NKI-066	1	1	**V**	3172	84.34	130	13.73	0.00	0.00	0.29
NKI-067	0	0	**V**	233	87.19	98	0.50	0.00	0.00	0.01
NKI-068	0	0	**V**	249	82.48	44	0.86	0.00	0.00	0.05
NKI-069	1	1	**V**	424	63.61	NA	2.34	0.00	0.00	1.45
NKI-070	0	0	**V**	50	62.3	68	1.24	0.00	0.00	0.01
NKI-071	0	0	**V**	66	20.29	NA	0.49	0.00	0.00	0.08
NKI-072	0	0	**V**	NA	NA	26	1.51	0.00	0.00	0.16
NKI-073	1	1	**V**	630	42.6	298	2.35	0.00	0.55	0.34
NKI-074	0	0	**V**	984	52.16	61	8.09	0.03	0.27	0.99
NKI-075	1	1	**V**	232	17.24	110	2.23	0.00	0.11	0.33
NKI-076	0	0	**V**	239	37.77	43	1.12	0.00	0.03	0.17
NKI-077	0	0	**V**	40	4.29	6	0.32	0.00	0.00	0.02
NKI-078	0	0	**V**	501	66.58	457	7.72	0.00	0.00	0.51
NKI-079	1	0	**V**	116	32.78	NA	2.34	0.00	0.00	0.03
NKI-080	0	0	**V**	732	73.97	321	2.19	0.00	0.00	0.06
NKI-081	0	0	**V**	113	17.43	6	1.70	0.00	0.00	0.36
NKI-082	1	1	**V**	279	13.04	213	3.93	0.03	0.30	1.08
NKI-083	0	0	**V**	199	4.32	NA	2.53	0.00	0.00	0.11
NKI-084	0	0	**V**	152	49.38	204	1.13	0.00	0.24	0.38
NKI-085	0	0	**V**	319	16.83	41	2.11	0.11	0.49	0.97
NKI-086	0	0	**V**	209	5.59	21	1.44	0.05	0.17	0.66
NKI-087	0	0	**V**	1442	17.82	NA	6.46	0.00	0.37	4.23
NKI-088	0	0	**V**	36	7.04	9	0.45	0.00	0.00	0.05
NKI-089	1	0	**V**	138	3.51	17	1.71	0.00	0.00	0.76
NKI-090	0	0	**V**	43	13.87	3	0.41	0.00	0.00	0.01
NKI-091	0	0	**V**	39	16.39	21	0.35	0.00	0.00	0.01
NKI-092	0	0	**V**	172	37.35	31	1.13	0.00	0.01	0.08
NKI-093	1	0	**V**	479	39.02	NA	3.49	0.00	0.00	0.43
NKI-094	0	0	**V**	240	3.75	150	2.13	0.00	0.22	0.34
NKI-095	0	0	**V**	115	10.66	18	0.49	0.00	0.02	0.21
NKI-096	1	1	**V**	582	36.78	211	4.81	0.00	0.55	1.58
NKI-097	0	0	**V**	184	10.64	101	2.40	0.10	0.60	0.91

PD-L1 TPS (%)	TIS score	CD8 TILs+ IT CD8 T cells (probability score, DC 6m)	CD3 TILs+ IT CD8 T cells (probability score, DC 6m)	CD8 TILs+ TIS (probability score, DC 12m)	CD8 TILs+ IT CD8 T cells (probability score, DC 12m)
0	NA	0.4028893	0.4048485	NA	0.2629425
100	9.98	0.9997702	0.9997072	0.5879973	0.9974879
0	7.09	0.6854384	0.3726544	0.0813578	0.5500873
40	NA	0.6514318	0.4045269	NA	0.5101457
0	NA	0.5803949	0.4843495	NA	0.4168848
0	8.02	0.3185722	0.3396215	0.0254434	0.2257876
0	NA	0.0755549	0.1278312	NA	0.0475368
NA	NA	NA	NA	NA	NA
0	NA	0.5170683	0.3180552	NA	0.3459363
60	5.83	0.8086051	0.8882897	0.9955607	0.6348037
40	6.79	0.1061145	0.1627554	0.0867333	0.0649764
1	NA	0.2214177	0.2071013	NA	0.1409171
0	5.79	0.0367493	0.093465	0.0174087	0.0227566
0	7.83	0.6581	0.9491255	0.2549388	0.4874698
70	NA	0.1387327	0.2505997	NA	0.0885215
1	8.78	0.8279293	0.5489125	0.2702141	0.6728334
0	NA	0.0770103	0.1496058	NA	0.0479004
0	6.90	0.1032161	0.184462	0.1172105	0.0626449
100	6.32	0.0589316	0.1163071	0.0738505	0.0358275
0	NA	0.2646051	0.2599037	NA	0.176134
0	NA	0.1319056	0.1577837	NA	0.0801059
0	5.95	0.0637693	0.1078106	0.0860978	0.0387583
90	NA	0.8257727	0.3183919	NA	0.6715443
10	NA	0.0406569	0.0996383	NA	0.0252655
0	NA	0.0476799	0.1050916	NA	0.0291588
1	5.44	0.0547966	0.112249	0.0170706	0.0342993
0	NA	0.0599567	0.1162457	NA	0.037682
30	5.93	0.1869745	0.205923	0.0617904	0.1196551
70	8.01	0.3752656	0.3808228	0.2075551	0.2406227
1	NA	0.0649582	0.1100494	NA	0.0393285
0	NA	0.0593131	0.1070974	NA	0.0365982
0	NA	0.4256953	0.4651745	NA	0.2746127
0	5.68	0.0722413	0.1384903	0.0714983	0.0441878

Table S1. Continued

Sample ID	1 = DC 6m 0= PD	1 = DC 12m, 0= PD	T=Training, V= Validation	CD8 TILs per mm²	IT-CD8 T cells (%)	PD-1ᵀTILs per mm²	CD3⁺ positive area per mm²	TLS per mm²	TLS+LA per mm²	CD20⁺ positive area per mm²
NKI-098	0	0	**V**	632	58.28	230	6.43	0.08	0.24	1.45
NKI-099	0	0	**V**	242	14.78	55	1.04	0.00	0.02	0.15
NKI-100	0	0	**V**	268	17.43	85	2.19	0.00	0.00	0.06
NKI-101	0	0	**V**	380	35.19	152	1.65	0.03	0.20	0.63
NKI-102	0	0	**V**	189	13.48	13	0.67	0.00	0.00	0.04
NKI-103	0	0	**V**	1012	32.4	NA	6.77	0.00	0.42	1.41
NKI-104	0	0	**V**	616	16.68	236	7.23	0.14	0.60	2.00
NKI-105	0	0	**V**	59	1.45	NA	0.33	0.00	0.00	0.03
NKI-106	0	0	**V**	124	22.94	56	1.22	0.00	0.00	0.09
NKI-107	0	0	**V**	185	23.2	126	0.83	0.00	0.00	0.03
NKI-108	1	1	**V**	175	38.34	181	3.34	0.33	1.32	7.01
NKI-109	0	0	**V**	341	99.02	206	1.96	0.00	0.00	0.03
NKI-110	1	1	**V**	381	23.05	225	3.58	0.14	0.68	2.05
NKI-111	0	0	**V**	196	27.34	35	1.46	0.01	0.05	0.38
NKI-112	0	0	**V**	48	23.99	22	0.44	0.00	0.00	0.08
NKI-113	1	1	**V**	595	34.1	226	3.76	0.16	0.47	1.80
NKI-114	0	0	**V**	52	50.51	12	0.57	0.00	0.00	0.02
NKI-115	0	0	**V**	255	3.54	NA	1.87	0.00	0.00	0.28
NKI-116	1	1	**V**	1156	18.87	NA	5.58	0.93	2.78	9.29
NKI-117	0	0	**V**	365	91.69	NA	1.40	0.00	0.00	0.04
NKI-118	1	0	**V**	260	23.93	160	2.78	0.05	0.22	0.21
NKI-119	1	0	**V**	280	15.98	67	1.32	0.00	0.06	0.18
NKI-120	1	1	**V**	1562	46.55	352	9.83	0.66	1.33	5.93
NKI-121	1	0	**V**	217	42.43	NA	1.69	0.00	0.00	0.07
NKI-122	0	0	**V**	90	25.31	24	0.21	0.00	0.00	0.00
NKI-123	0	0	**V**	396	39.23	80	3.15	0.02	0.06	0.20
NKI-124	0	0	**V**	311	62.72	31	NA	NA	NA	NA
NKI-125	1	1	**V**	245	5.09	104	NA	NA	NA	NA
NKI-126	0	0	**V**	267	11.11	NA	1.12	0.00	0.00	0.10
NKI-127	1	1	**V**	456	56.8	60	1.55	0.00	0.00	0.02
NKI-128	1	1	**V**	296	69.57	1	1.27	0.00	0	0.05
NKI-129	0	0	**V**	95	34.14	27	0.47	0.00	0	0.04
NKI-130	0	0	**V**	267	45.91	44	NA	NA	NA	NA

PD-L1 TPS (%)	TIS score	CD8 TILs+ IT CD8 T cells (probability score, DC 6m)	CD3 TILs+ IT CD8 T cells (probability score, DC 6m)	CD8 TILs+ TIS (probability score, DC 12m)	CD8 TILs+ IT CD8 T cells (probability score, DC 12m)
100	8.60	0.6655044	0.8484654	0.2370195	0.4845027
0	7.13	0.099495	0.1262077	0.08555	0.0606982
0	6.90	0.1177306	0.1625127	0.1086931	0.0718445
0	6.82	0.2783846	0.2250008	0.2328005	0.1750651
0	NA	0.0817629	0.1159974	NA	0.0501128
0	8.30	0.6881829	0.5579657	0.7256058	0.4969709
0	8.71	0.2710605	0.3359382	0.2053837	0.1653748
0	NA	0.0346187	0.0889156	NA	0.0213275
0	6.21	0.0981677	0.1555833	0.0392423	0.0614348
0	NA	0.116696	0.1445237	NA	0.072433
100	7.82	0.1947283	0.3622262	0.0497995	0.124898
80	NA	0.8241216	0.6974137	NA	0.7076549
1	NA	0.1910461	0.2449967	NA	0.1167034
0	NA	0.1396518	0.1804173	NA	0.0871661
10	6.01	0.083122	0.135228	0.019392	0.0527722
30	NA	0.4092446	0.3532502	NA	0.2617925
0	NA	0.2213779	0.221484	NA	0.1499217
0	NA	0.0673037	0.1067635	NA	0.0406925
0	NA	0.674359	0.292948	NA	0.4883256
70	8.41	0.7869513	0.5526563	0.1026149	0.6545504
0	5.61	0.1456327	0.2168293	0.1470963	0.089745
0	6.91	0.1153123	0.1356187	0.1176396	0.0701794
1	8.26	0.946989	0.9137131	0.9762271	0.8562076
1	5.02	0.2446654	0.2662741	0.1115903	0.1581614
40	NA	0.0980491	0.1315955	NA	0.0619341
30	NA	0.3226077	0.3552108	NA	0.2056915
0	NA	0.4942433	NA	NA	0.3486228
0	6.95	0.0694158	NA	0.0914363	0.0420353
2	NA	0.0929217	0.1178855	NA	0.0563683
100	NA	0.5345907	0.3402784	NA	0.3716708
5	NA	0.5542083	0.3870957	NA	0.4053088
0	6.98	0.1382523	0.1627029	0.0310978	0.088781
20	NA	0.3016551	NA	NA	0.1968249

Table S1. Continued

Sample ID	1 = DC 6m 0= PD	1 = DC 12m, 0= PD	T= Training, V= Validation	CD8 TILs per mm²	IT-CD8 T cells (%)	PD-1ᵀTILs per mm²	CD3⁺ positive area per mm²	TLS per mm²	TLS+LA per mm²	CD20⁺ positive area per mm²
NKI-131	0	0	**V**	153	42.38	29	0.87	0.00	0	0.08
NKI-132	1	0	**V**	274	4.73	NA	2.21	0.00	0.00	1.30
NKI-133	0	0	**V**	NA	NA	NA	NA	NA	NA	NA
NKI-134	1	0	**V**	927	54.17	NA	4.66	0.00	0	0.07
NKI-135	0	0	**V**	NA	NA	NA	NA	NA	NA	NA

NA, not available

Table S2. Predictive accuracy of individual and composite biomarkers at a prespecified criterium of ≥90% sensitivity for detecting disease control (DC) at 6 months in the training cohort. * Indicates an interaction term.

Model	Sensitivity ≥90%	Specificity	Specificity 95% CI lower limit	Specificity 95% CI upper limit
PD-1ᵀ * LA	0.92	73	54	88
PD-1ᵀ * PD-L1	0.92	73	54	88
PD-1ᵀ * CD3	0.92	73	54	88
PD-1ᵀ * IT-CD8	0.92	73	54	88
PD-1ᵀ * CD20	0.92	70	51	85
PD-1ᵀ	**0.92**	**70**	**51**	**85**
PD-1ᵀ * TLS	0.92	70	51	85
CD8 * IT-CD8	**0.94**	**62**	**45**	**77**
PD-1ᵀ * TIS	1.00	60	36	81
LA * TIS	1.00	55	32	77
TIS	**1.00**	**55**	**32**	**77**
CD3 * TIS	1.00	55	32	77
CD3 * IT-CD8	**0.94**	**54**	**37**	**71**
PD-1ᵀ * CD8	0.92	53	34	72
TLS * CD20	0.94	51	34	68
LA * CD20	0.94	51	34	68
TLS * TIS	1.00	50	27	73
CD20 * TIS	1.00	50	27	73
CD20	0.94	49	32	66
IT-CD8 * TIS	1.00	45	23	68

PD-L1 TPS (%)	TIS score	CD8 TILs+ IT CD8 T cells (probability score, DC 6m)	CD3 TILs+ IT CD8 T cells (probability score, DC 6m)	CD8 TILs+ TIS (probability score, DC 12m)	CD8 TILs+ IT CD8 T cells (probability score, DC 12m)
30	7.25	0.2118835	0.2109187	0.0463233	0.137822
0	NA	0.0744299	0.1140751	NA	0.0449534
0	5.48	NA	NA	NA	NA
20	9.03	0.7950019	0.6614375	0.3327138	0.6185755
0	5.4	NA	NA	NA	NA

PPV	NPV	AUC	AUC 95% CI lower limit	AUC 95% CI upper limit
58	96	85	73	97
58	96	85	74	97
58	96	86	72	99
58	96	86	70	100
55	95	85	72	98
55	**95**	**82**	**68**	**95**
55	95	84	71	97
50	**96**	**83**	**73**	**94**
47	100	83	67	99
47	100	82	67	98
47	**100**	**81**	**65**	**98**
47	100	84	69	100
47	**95**	**78**	**65**	**92**
44	94	82	68	97
45	95	73	59	88
45	95	72	58	87
44	100	86	72	100
44	100	84	68	100
44	95	71	56	86
42	100	89	74	100

Table S2. Continued

Model	Sensitivity ≥90%	Specificity	Specificity 95% CI lower limit	Specificity 95% CI upper limit
PD-L1 * TIS	1.00	45	23	68
CD8 * LA	0.94	41	25	58
PD-L1 * IT-CD8	0.94	38	23	55
TLS * CD3	0.94	38	22	55
LA * CD3	0.94	38	22	55
CD20 * CD3	0.94	38	22	55
PD-L1 * CD3	0.94	35	20	53
CD3	0.94	35	20	53
CD8 * TLS	0.94	35	20	53
CD8 * CD20	0.94	32	18	50
PD-L1 * CD20	0.94	32	18	50
CD8 * CD3	0.94	32	18	50
CD8	0.94	28	15	45
PD-L1 * CD8	0.94	26	13	42
CD8 TILs * TIS	1.00	25	9	49
PD-L1 * LA	0.94	24	12	41
TLS * IT-CD8	0.94	8	2	22
LA * IT-CD8	0.94	8	2	22
CD20 * IT-CD8	0.94	8	2	22
IT-CD8	0.94	8	2	21
TLS * LA	1.00	3	0	14
PD-L1	1.00	0	0	9
TLS	1.00	0	0	9
LA	1.00	0	0	9
PD-L1 * TLS	1.00	0	0	9

NA, not available

PPV	NPV	AUC	AUC 95% CI lower limit	AUC 95% CI upper limit
42	100	81	64	98
41	94	79	63	95
38	94	76	62	91
39	93	77	63	91
39	93	76	62	90
39	93	77	62	91
38	93	74	58	89
38	93	74	59	88
38	93	78	63	92
38	92	78	63	93
38	92	59	42	76
38	92	76	61	92
35	92	76	60	91
34	91	77	61	92
35	100	85	66	100
35	90	64	47	80
31	75	78	62	93
31	75	75	59	92
31	75	75	59	92
29	75	74	57	90
31	100	64	49	80
29	NA	54	36	72
30	NA	62	46	77
30	NA	60	44	76
30	NA	58	41	76

Table S3. Predictive accuracy of individual and composite biomarkers at a prespecified criterium of ≥90% sensitivity for detecting disease control (DC) at 12 months in the training cohort. * Indicates an interaction term.

Model	Sensitivity ≥90%	Specificity	Specificity 95% CI lower limit	Specificity 95% CI upper limit
PD-1^T * CD8	1.00	73	54	87
PD-1^T * CD3	1.00	70	51	84
PD-1T * TLS	1.00	70	51	84
CD8 * TIS	**1.00**	**68**	**45**	**86**
PD-1^T	**1.00**	**67**	**48**	**82**
PD-1^T * CD20	1.00	67	48	82
PD-1^T * LA	1.00	67	48	82
PD-1^T * PD-L1	1.00	67	48	82
CD8 * IT-CD8	**0.92**	**63**	**47**	**77**
CD3 * IT-CD8	0.92	59	42	74
LA * IT-CD8	0.92	56	40	72
PD-L1 * CD8	0.92	56	40	71
PD-1^T * TIS	1.00	55	32	76
CD8 * LA	0.92	54	37	69
PD-1^T * IT-CD8	1.00	52	34	69
CD3 * TIS	1.00	50	28	72
LA * TIS	1.00	50	28	72
TIS	**1.00**	**50**	**28**	**72**
CD-20 * IT-CD8	0.92	49	33	65
TLS * CD20	0.92	49	33	65
TLS * IT-CD8	0.92	49	33	65
CD20 * CD3	0.92	46	31	63
CD20 * TIS	1.00	45	24	68
IT-CD8 * TIS	1.00	45	24	68
TLS * TIS	1.00	45	24	68
CD20	0.92	44	28	60
IT-CD8	0.92	44	29	60
PD-L1 * IT-CD8	0.92	42	27	58
CD8 * CD3	0.92	41	26	58
CD8 * TLS	0.92	41	26	58
PD-L1 * CD3	0.92	41	26	58
PD-L1 * TIS	1.00	41	21	64
TLS * CD3	0.92	41	26	58
CD3	0.92	39	24	55

PPV	NPV	AUC	AUC 95% CI lower limit	AUC 95% CI upper limit
50	100	91	82	100
47	100	90	80	100
47	100	82	70	95
46	**100**	**91**	**79**	**100**
45	**100**	**82**	**70**	**95**
45	100	85	74	97
45	100	81	68	94
45	100	85	72	97
41	**96**	**85**	**73**	**96**
39	96	82	70	94
38	96	82	69	94
37	96	80	65	94
33	100	80	61	99
37	96	82	66	98
36	100	84	70	98
35	100	83	65	100
35	100	77	57	97
35	**100**	**77**	**57**	**97**
34	95	73	55	91
34	95	68	52	85
34	95	76	58	93
33	95	81	68	95
33	100	76	53	98
33	100	83	64	100
33	100	80	60	99
32	95	66	49	83
31	95	74	56	92
31	95	80	65	94
31	94	81	67	95
31	94	79	64	95
31	94	77	62	92
32	100	80	59	100
31	94	79	64	94
31	94	76	61	92

Table S3. Continued

Model	Sensitivity ≥90%	Specificity	Specificity 95% CI lower limit	Specificity 95% CI upper limit
LA * CD3	0.92	39	24	55
CD8	0.92	37	23	53
CD8 * CD20	0.92	37	22	53
PD-L1 * TLS	0.92	17	7	32
LA * CD20	0.92	10	3	23
PD-L1 * CD20	0.92	5	1	17
PD-L1 * LA	0.92	2	0	13
TLS * LA	1.00	2	0	13
LA	1.00	0	0	9
PD-L1	1.00	0	0	8
TLS	1.00	0	0	9

NA, not available

PPV	NPV	AUC	AUC 95% CI lower limit	AUC 95% CI upper limit
31	94	79	64	94
29	94	80	65	95
30	94	82	66	97
24	88	64	43	85
23	80	55	33	76
22	67	43	21	64
22	50	65	44	86
23	100	59	42	76
23	NA	58	40	75
22	NA	64	45	84
23	NA	59	43	76

3

Chapter 4

A dysfunctional T cell gene signature for predicting non-response to PD-1 blockade in non-small cell lung cancer that is suitable for routine clinical diagnostics

Karlijn Hummelink[1,2], Renaud Tissier[3], Linda J.W. Bosch[1], Oscar Krijgsman[4], Michel M. van den Heuvel[2,5], Willemijn S.M.E. Theelen[2], Diane Damotte[6,7,8], François Goldwasser[7,8], Karen Leroy[6,7,8,9], Egbert F. Smit[2,10], Gerrit A. Meijer[1,#], Daniela S. Thommen[4,#], and Kim Monkhorst[1,#]

#*Shared last authorship.*

[1]Department of Pathology, Division of Diagnostic Oncology, [2]Department of Thoracic Oncology, Division of Medical Oncology, [3]Biostatistics Unit, [4]Division of Molecular Oncology and Immunology, the Netherlands Cancer Institute, Amsterdam, The Netherlands, [5]Present address: Department of Pulmonary Diseases, Radboud University Medical Center, Nijmegen, The Netherlands, [6]Team Cancer, Immune Control and Escape, Cordeliers Research Center, UMRS 1138, Institut National de la Santé et de la Recherche Médicale (INSERM), Paris, France, [7]University Paris Cité, Paris, France, [8]CERTIM, Medical Oncology, Hôpital Cochin, APHP, Paris, France, [9]Department of Biochemistry, Hôpital Cochin, Européen Georges Pompidou, APHP Centre, 75015, Paris, France, [10]Present address: Department of Pulmonary Diseases, Leiden University Medical Center, Leiden, The Netherlands.

Translational relevance

PD-(L)1 blockade therapies have significantly improved the survival of a subset of patients with advanced stage non-small cell lung cancer. However, most patients don't benefit, but still are at risk of adverse effects associated. This asks for biomarkers to better predict benefit. PD-1^T TILs, a tumor- reactive T-cell population, constitute such a biomarker, and can be accurately quantified by digital image analysis in FFPE tumor tissue. Yet, this method is complex, limiting its application in routine patient care. As an alternative, a PD-1^T mRNA expression signature measured in FFPE tissue samples with industry standard technology is more suitable for implementation in routine diagnostics. The PD-1^T signature reached equally high sensitivity and negative predictive value as the digital image analysis-based IHC quantification of PD-1^T TILs, making it particularly suited for reliably identifying advanced stage non-small cell lung cancer patients who have a low chance of benefitting from PD-(L)1 blockade therapy.

Abstract:

Introduction

Since PD-1 blockade is only effective in a minority of patients with advanced stage NSCLC, biomarkers are needed to guide treatment decisions. Tumor infiltration by PD-1[T] TILs, a dysfunctional tumor infiltrating lymphocyte (TIL) pool with tumor-reactive capacity, can be detected by digital quantitative IHC and has been established as a novel predictive biomarker in NSCLC. To facilitate translation of this biomarker to the clinic, we here aimed to develop a robust RNA signature reflecting a tumor's PD-1[T] TIL status.

Methods

mRNA expression analysis using Nanostring was performed in baseline tumor samples from 41 advanced stage NSCLC patients treated with nivolumab that were selected based on PD-1[T] TIL infiltration by IHC. Samples were included as training cohort (n=41) to develop a predictive gene signature. This signature was independently validated in a second cohort (n=42). Primary outcome was disease control at 12 months (DC 12m) and secondary outcome was progression-free and overall survival.

Results

Regularized regression analysis yielded a signature using 12 out of 56 differentially expressed genes between PD-1[T] IHC high tumors from patients with DC 12m and PD-1[T] IHC low tumors from patients with progressive disease (PD). In the validation cohort 6/6 (100%) patients with DC 12m and 23/36 (64%) with PD were correctly classified with an NPV of 100% and a PPV of 32%.

Conclusion

The PD-1[T] mRNA signature showed a similar high sensitivity and high NPV as the digital IHC quantification of PD-1[T] TILs. This provides a straightforward approach allowing for easy implementation in a routine diagnostic clinical setting.

4

Introduction

Pharmacological blockade of the inhibitory immune receptor programmed cell death protein 1 (PD-1) or its ligand programmed death-ligand 1 (PD-L1) has improved the clinical outcome of many cancers, including non-small cell lung cancer (NSCLC)[1–5]. High tumor PD-L1 expression has been associated with clinical benefit in NSCLC treated with PD-1 blockade[1,2,6] and these results have led to the implementation of PD-L1 immunohistochemistry (IHC) as a biomarker in clinical practice. However, other studies have shown suboptimal correlation of PD-L1 expression to clinical outcome[3–5]. Therefore, robust biomarkers that more accurately predict who will benefit, and who not, are needed. In particular, biomarkers with a high negative predictive value (NPV) that reliably predict lack of clinical benefit are important to offer patients alternative treatments.

Previously, we reported high levels of PD-1^T tumor-infiltrating lymphocytes (TILs) as a novel predictive biomarker for long-term benefit to PD-1 blockade with a high NPV[7]. PD-1^T TILs are a subset of PD-1+ T cells with a dysfunctional phenotype, high tumor-specific expression of PD-1, and high capacity for tumor recognition[8,9]. PD-1^T TILs can be measured via algorithm-based quantitative analysis of PD-1 IHC in formalin-fixed paraffin embedded (FFPE) tumor tissue[7,8]. While digital image analysis can yield accurate quantitative PD-1 protein expression data, this method is challenging to validate across centers, impacting routine clinical care.

We therefore aimed at developing a reliable method that can detect the signal represented by PD-1^T TILs and that can easily be applied in routine clinical care. The NanoString nCounter is a robust platform that allows for measuring very low input RNA amounts isolated from FFPE tissue[10]. Already several mRNA signatures have been developed for this platform, including the Tumor Inflammation Signature (TIS) that has demonstrated predictive potential for clinical benefit to PD-1 blockade in multiple cancer types[11–13].

In the present study, we used the NanoString nCounter platform, to develop and validate an RNA expression signature that reflects a tumor's PD-1^T TIL status and predicts clinical outcome of NSCLC patients treated with PD-1 blockade.

Methods

Patient cohorts and study endpoints

In this study, 152 stage IV NSCLC patients were included that all started second or later line monotherapy with nivolumab. Patients with tumors harboring known sensitizing EGFR mutations or ALK translocations were excluded. A total of 115 patients received PD-1 blockade therapy since March 2015 at the Netherlands Cancer Institute/Antoni van Leeuwenhoek hospital (NKI-AVL), The Netherlands, and were split in a training (n=94) and validation cohort (n=21). 37 patients from the CERTIM (Immunomodulatory Therapies Multidisciplinary Study Group) treated since July 2015 at Cochin University hospital, France[13] were pooled with the validation cohort **(Fig. 1)**. All patients received nivolumab, administered per label as a single agent.

Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1 was used to determine the objective response. Patients treated at NKI-AVL, who were not evaluable for response assessment according to RECIST, were determined as progressive disease (PD) by the treating physician. Disease Control (DC) status (complete response (CR)/ partial response (PR) or stable disease (SD)) at 12 months after treatment initiation was used as the primary clinical outcome measure. Progression-free survival (PFS) and overall survival (OS) were used as secondary outcome measures. PFS and OS were defined as the time from the date of initiation of PD-1 blockade treatment to the date of progression (for PFS) or death (for OS). Patients who had not progressed or died were censored at the date of their last follow-up.

RNA or gene expression data derived from pretreatment archival formalin-fixed paraffin embedded (FFPE) tumor tissue samples were collected from the cohorts. Written informed consent was obtained from all patients treated at NKI-AVL 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 Review Board of NKI (CFMPB586). Gene expression data of the CERTIM cohort was used from Damotte et al., a study that was previously approved by the ethics committee (CPP Ile de France II, no. 2008-33, 2012 06-12, 2018 MS1) in agreement with article L.1121-1 of the French law[13]. In the NKI-AVL training cohort (n=94), 28 samples were excluded based on low RNA yield and/or low RNA quality. Eight additional samples were excluded due to quality control failure in the Nanostring nCounter profiling assay (Nanostring) **(Fig. 1)**. Eight samples in the NKI-AVL validation cohort and 6 samples in the CERTIM cohort were excluded because these were obtained 2 years or more before start of PD-1 blockade. Two additional samples in the NKI-AVL validation cohort were

excluded because these contained abundant normal lymphoid tissue (n=2) **(Fig. 1)**. We excluded lymph node resections as these contain PD-1 high-expressing T cells in normal lymphoid tissue, which could potentially lead to false positive results. Lymph node biopsies were included as these are usually targeted biopsies in the tumor region.
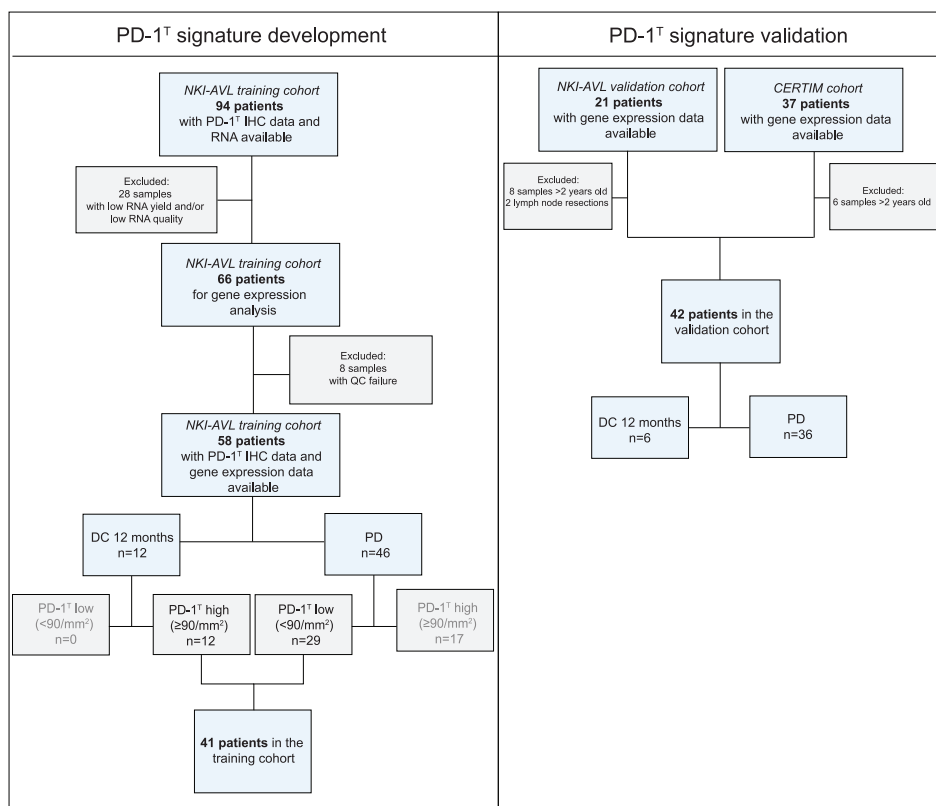

Figure 1. Study design for the development and validation of the PD-1^T signature as biomarker for non-response to PD-1 blockade in NSCLC. Overall workflow for the development of the PD-1^T signature using PD-1^T IHC high (≥90 per mm²) patients with disease control (DC) at 12 months (n=12) and PD-1^T IHC low (<90 per mm²) patients with progressive disease (n=29). An independent cohort of patients was used for validation (n=42).

Immunohistochemistry

PD-1 and PD-L1 immunostaining of samples from the NKI-AVL cohorts were performed on fresh-cut slides from FFPE blocks using an anti-PD-1 antibody (NAT105, Roche Diagnostics) and anti-PD-L1 antibody (22C3 DAKO, Agilent) on a BenchMark Ultra autostainer Instrument (Ventana Medical Systems) as described previously[7].

PD-L1 immunostaining of samples from the CERTIM cohort was performed using a different anti-PD-L1 antibody (E1L3N, Cell Signaling, catalog no. AB_2687655) on a Bond automat (Leica Biosystems) as described previously by Adam et al[14].

PD-1 IHC slides were scanned at x20 magnification with a resolution of 0.50 per µm² using an Aperio slide AT2 scanner (Leica Biosystems).

Digital quantification of PD-1[T] TILs

In 11/94 (12%) pretreatment samples automated quantification of PD-1[T] TILs was performed using an image analysis algorithm with a cut-off of 0.25 optical density (OD) of PD-1 staining as described previously[7]. In 83/94 (88%) samples PD-1[T] TIL numbers were used from a previously published cohort and quantified using the same approach[7]. PD-1[T] TIL numbers of 58 samples were used to develop the signature and are provided in **Supplementary Table S1**.

PD-L1 scoring

Tumor PD-L1 expression in pretreatment FFPE samples was assessed in the NKI-AVL training (n=58) and validation cohort (n=21) using the clinical grade LDT IHC assay with the 22C3 DAKO clone (Agilent) as described previously[7]. For 35/37 (95%) samples in the CERTIM cohort, PD-L1 tumor proportion score (TPS) data has been reported before[13]. In 2/37 (5%) the PD-L1 status was unknown. The expression levels were scored using a different anti-PD-L1 antibody (E1L3N, Cell Signaling, catalog no. AB_2687655) as previously described and validated by the PATTERN French thoracic pathologists' group[14]. PD-L1 TPS data is provided in **Supplementary Table S1**.

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 recommendations of the manufacturer and quantified by Tapestation (Agilent). RNA from the CERTIM cohort was extracted with High Pure FFPE RNA Isolation Kit (Roche Diagnostics) according to the recommendations of the manufacturer and quantified using fluorimetry with Qubit RNA XR Assay Kit (Invitrogen, Thermo Fisher). 200 to 300 ng RNA from the NKI-AVL cohorts and 30-100 ng RNA from the CERTIM cohort were hybridized to Nanostring PanCancer IO 360 Panel code set (Nanostring), according to the recommendations of the manufacturer. After hybridization non-bound probes were washed off and the RNA-probe complex was bound to the cartridge on the Nanostring Flex Prep Station (Nanostring) according to manufacturing protocol. The cartridge was sealed and transferred to the Digital Analyzer for imaging.

Statistical analysis

The Mann-Whitney, Fisher exact and linear-by-linear association tests, respectively, were used to assess differences in patient characteristics between training and validation cohorts. Differences were considered statistically significant if *P<0.05. Correlations between the PD-1^T signature and PD-1^T TILs assessed by IHC or the PD-1^T signature and the Tumor Inflammation Signature (TIS), respectively, were evaluated using linear regression analysis.

A 2-level batch effect correction on the mRNA expression data was performed on all NKI-AVL and CERTIM patients using an empirical bayes linear regression. This was performed to correct for batch effects between the NKI-AVL and the CERTIM cohort and between the different NKI-AVL cohorts. Both batch effect correction and gene expression analysis were performed with R 4.1.0 and the package limma 3.48.0. Differential gene expression analysis was performed using linear regression on the gene log-expression. Separate models were fitted for each gene, and the computation of moderated t-statistics and log-odds of differential expression was performed via empirical bayes moderation. Analysis of main biological processes involved in the gene signature was performed by gene ontology analysis using the Fisher exact test with the R package topGO 2.44.0 (SCR_014798). *P*-values were adjusted via Benjamini-Hochberg.

A prediction model was built using logistic regression combined with regularized regression for variable selection using LASSO (least absolute shrinkage and selection operator). By adding a penalization term on the coefficients of the model, the coefficients of the models are constrained to zero leading to variable selection. Due to the limited sample size and unequal distribution over the DC 12m and PD patient groups, cross validation was limited to three-fold. Thus, a three-fold cross-validation for the selection of the optimal penalization term of the regularized regression, based on the deviance, was performed. This is a goodness-of-fit statistics commonly used for generalized linear models. The results of the regression were transformed to obtain probability scores using the formula: $\frac{1}{1 + e^{-K}}$ with K being the results of the logistic regression. K was computed as $K = c_1 * STAT1 + c_2 * OAS1 + c_3 * TAP1 + \cdots + c_{12} * LAG3$. The coefficients of the prediction model are provided in **Supplementary Table S2.**

The cross-validation and prediction model building were performed with R and the package glmnet 4.1-2. Based on the NKI-AVL training cohort, a threshold was chosen from the probability scores that were provided by the prediction model to classify a patient as predicted to achieve DC upon therapy. This threshold was set at the best sensitivity (detection of DC), while keeping a satisfactory specificity. A

prediction model including only the *PDCD1* gene was built using logistic regression with samples from the NKI-AVL training cohort. All prediction models aimed to predict DC 12m and therefore the reference category used in the logistic regression was PD. Signature probability scores are provided in **Supplementary Table S1**. Receiver operating characteristic (ROC) curves were produced for each prediction model using the package pROC 1.18.0. The Area Under the ROC Curves (AUCs) were compared using the DeLong test with the pROC 1.18.0 R package.

Genes in the TIS are normalized using a ratio of the expression value to the geometric mean of the housekeeper genes that are used only for the TIS. This is then followed by a log2 transformation. The TIS score was calculated in the NKI-AVL validation cohort (n=21) as a weighted linear combination of the 18 gene expression values[11,15]. This analysis was performed by Nanostring as part of their intellectual property. For the 37 samples in the CERTIM cohort, TIS scores have been reported before[13]. TIS scores of 42 samples were used for analysis and are provided in **Supplementary Table S1**.

Results

PD-1^T signature development
To develop a predictive mRNA signature that reflects tumor infiltration by PD-1^T TILs, we first selected 94 pretreatment samples from advanced stage NSCLC patients treated with nivolumab (training cohort). 66/94 (70%) of these samples had sufficient RNA available for gene expression analysis using the Nanostring nCounter platform. Only archival samples were used, and therefore the majority with insufficient RNA were blocks that were already (partially) used for previous molecular analysis. 58/94 (62%) samples were successful in obtaining high quality mRNA expression data to be used for signature development, as 8 samples did not meet RNA quality criteria **(Fig. 1)**.

Of these 58 patients, 12 showed disease control at 12 months (DC 12m) (n=12) and 46 developed progressive disease (PD) within 12 months of treatment **(Fig. 1)**. DC12 was chosen as primary clinical outcome measure based on previous observations that patients with DC 12m were more accurately identified by the biomarker as compared to DC 6m (i.e. higher sensitivity). This was also found for a group with no long-term benefit (i.e. higher negative predictive value (NPV))[7]. Clinicopathological characteristics and treatment outcomes are summarized in **Supplementary Table S3.** In line with our previous work, the number of PD-1^T TILs per mm^2 was significantly higher in patients with DC 12m compared to patients with PD (P<0.01) **(Fig. 2A)**. PD-1^T IHC high versus low status was called using a previously established cut-off of 90 PD-1^T TILs per mm^2 **(Fig. S1A,B)**[7]. 12/12 (100%) patients with DC 12m were classified as PD-1^T IHC high. 29/46 (63%) patients with PD were classified as PD-1^T IHC low, and 17/46 (37%) as PD-1^T IHC high **(Fig. 1, 2A)**.

To obtain a distinctive mRNA expression gene set with maximum contrast, we performed differential gene expression analysis between PD-1^T IHC high patients with DC 12m (n=12) and PD-1^T IHC low patients with PD (n=29) **(Fig. 1, 2A)**. After correction for multiple testing, 54 genes were significantly higher expressed and 2 genes were significantly lower expressed in the PD-1^T IHC high DC 12m group. Some of the top ranked genes included *LAG3*, *CTLA4*, *CXCR6* and *CXCL13*, which have previously shown to be highly expressed in PD-1^T TILs[8] **(Table S4)**. In addition, various pathways related to active immune responses in the tumor microenvironment (TME), including type I interferon signaling, regulation of lymphocyte chemotaxis and natural killer cell mediated cytotoxicity were significantly upregulated in the PD-1^T IHC high DC 12m group **(Fig. 2B)**.

Regularized regression analysis (LASSO) yielded a 12-gene PD-1^T signature (*STAT1, OAS1, TAP1, HEY1, CXCL13, IFIT2, IL6, TDO2, CD6, CTLA4, CD274, LAG3*) as being most predictive **(Table 1)**. All genes in the signature were positively associated with DC 12m, except for *HEY1*, which showed a negative association **(Fig. 2C,D)**. In line with the selection strategy of the samples, we observed a high correlation between the number of PD-1^T TILs assessed by IHC and the PD-1^T signature score within samples (R^2=0.603; *P*<0.0001; **Fig. 2E**). The PD-1^T signature was able to separate the preselected PD-1^T IHC high DC 12m group from the preselected PD-1^T IHC low PD group with high significance (*P*<0.0001) **(Fig. 2D,F)**. The area under the ROC curve (AUC) was 0.97 (95% CI: 0.93-1.00) **(Fig. 2G)**. We aimed for a sensitivity of ≥90% to minimize the risk of undertreatment and a specificity of ≥50%, a strategy that was previously used for the PD-1^T TIL IHC biomarker[7]. A probability score of 0.35 was selected as optimal cut-off **(Fig. 2F)**. This cut-off resulted in a sensitivity of 92%, specificity of 93%, positive predictive value (PPV) of 85% and NPV of 96% **(Table 2)**.

PD-1^T signature validation
Next, the predictive performance of this PD-1^T signature was validated in an independent cohort of 42 patients with advanced stage NSCLC treated with nivolumab. 6/42 (14%) patients showed DC 12m and 36/42 (86%) showed PD **(Fig. 1)**. In contrast to the training cohort, tumor samples were not preselected. None of the clinicopathological characteristics differed significantly between training and validation set, except for the performance score as more patients in the validation cohort showed a higher performance score (*P*<0.01) **(Table S3)**.

In the validation cohort, in line with previously observed patterns of PD-1^T TILs quantified by IHC, PD-1^T signature scores were significantly higher in the DC 12m group than in the PD group (*P*<0.01) **(Fig. 3A)**. A high PD-1^T signature score (≥0.35) correctly identified 6/6 patients with DC 12m (sensitivity for treatment benefit 100%), and a low score (<0.35) identified 23/36 patients with PD (specificity for no treatment benefit 64%), yielding a PPV of 32%, and an NPV of 100%, with an AUC of 0.87 (95% CI: 0.74-0.99) **(Fig. 3A,B** and **Table 2)**. Similar to the training cohort, most signature genes were overexpressed in the PD-1^T signature high DC 12m group (n=6), compared to the PD-1^T signature low PD group (n=23), and the *HEY1* gene was expressed at lower levels. The PD-1^T signature high patients with PD (n=13) showed a similar gene expression profile as the PD-1^T signature high DC 12m group **(Fig. 3C)**.

In addition, in the validation cohort, progression-free survival (PFS) was significantly longer in PD-1^T signature high patients (median 8.3 months) versus PD-1^T signature low patients (median 1.8 months) with HR 0.36 (95% CI: 0.18-0.69), P<0.001. The median overall survival (OS) was 7.0 months versus 5.6 months, respectively, with HR 0.34 (95% CI: 0.17-0.68), P<0.001 **(Fig. 3D,E)**.

4
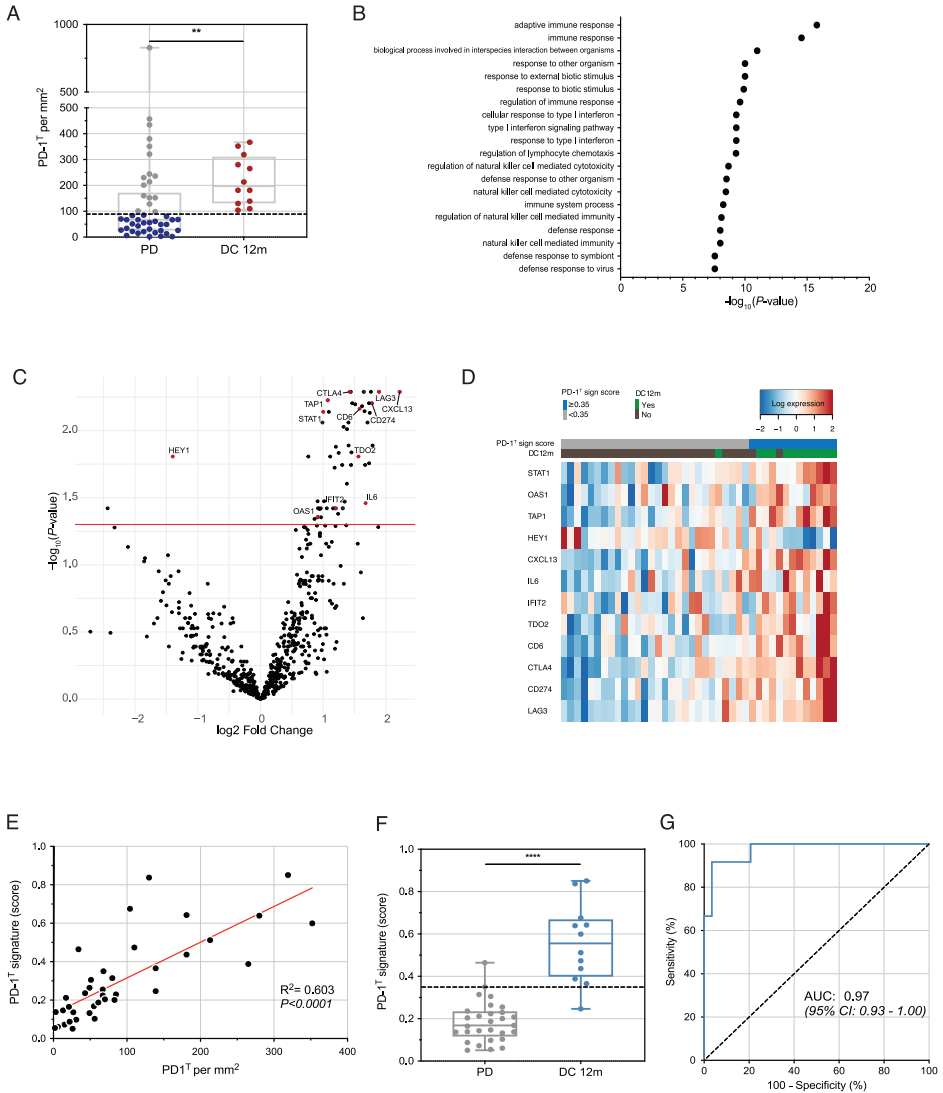
◄ **Figure 2.** PD-1^T signature development for prediction of non-response to PD-1 blockade. **(A)** PD-1^T TILs per mm², as measured by digital IHC algorithm-based quantification, in pretreatment samples from patients with disease control at 12 months (DC 12m) (n=12) and progressive disease (PD) (n=46) in the training cohort (n=58). Only patients with available PD-1^T IHC and gene expression data are included. Dashed line indicates a cut-off of 90 PD-1^T TILs per mm². Medians, interquartile ranges and minimum/maximum shown in boxplots, **P<0.01 by Mann Whitney U-test. The red colored dots indicate PD-1^T IHC high (≥90 per mm²) patients with DC 12m (n=12) and the blue colored dots indicate PD-1^T IHC low (<90 per mm²) patients with PD (n=29). **(B)** Gene set enrichment analysis displaying gene sets that are significantly enriched in PD-1^T IHC high tumors from patients with DC 12m (n=12). Pathways are ordered by P value (log transformed), P values were calculated by Fisher exact test. **(C)** Volcano plot showing the differentially expressed genes between PD-1^T IHC high tumors from patients with DC 12m and PD-1^T IHC low tumors from patients with PD (n=29) in the training cohort (n=41). 56 genes reached statistical significance. The red line indicates a P-value <0.05. P-values were computed by moderated t-statistics. **D)** Heatmap showing the expression of the 12-gene PD-1^T signature in pretreatment samples from PD-1^T IHC high patients with DC 12m (n=12) and PD-1^T IHC low patients with PD (n=29) in the training cohort (n=41) ordered by probability score. Each column represents one patient (blue: PD-1^T signature high (score of ≥0.35) and light grey: PD-1^T signature low (score of <0.35), green: PD-1^T IHC high patients with DC 12m, dark grey: PD-1^T IHC low patients with PD) and rows display genes. Positive values (red) indicate higher gene expression and negative values (blue) indicate lower gene expression. **(E)** Correlation of PD-1^T TILs and the PD-1^T signature in the training cohort (n=41), R²=0.603, ****P<0.0001. R² and P-values were calculated using linear regression analysis. **(F)** PD-1^T signature scores in pretreatment samples from PD-1^T IHC high patients with DC 12m (n=12) and PD-1^T IHC low patients with PD (n=29) in the training cohort (n=41). The dashed line indicates a cut-off score of 0.35. Medians, interquartile ranges and minimum/maximum shown in boxplots, ****P<0.0001 by Mann Whitney U-test. **(G)** Receiver operating characteristic (ROC) curve for predictive value of the PD-1^T signature for DC 12m (AUC 0.97; 95% CI: 0.93-1.00) in the training cohort (n=41).

Table 1. Overview of the up- and downregulated PD-1^T signature genes.

PD-1^T signature genes	
STAT1	UP
OAS1	UP
TAP1	UP
HEY1	DOWN
CXCL13	UP
IFIT2	UP
IL6	UP
TDO2	UP
CD6	UP
CTLA4	UP
CD274	UP
LAG3	UP

Table 2. Predictive accuracy of the PD-1^T signature and PD-L1 TPS, summary of training and validation results.

|	Biomarker	AUC
Training cohort (n=41)	PD-1^T signature	0.97 95% CI: 0.93-1.00
Validation cohort (n=42)	PD-1^T signature	0.87 95% CI: 0.74-0.99
Validation cohort (n=40)	% PD-L1 TPS	0.63 95% CI: 0.34-0.91

Comparison to PD-L1 IHC and PDCD1 gene expression

In previous work we showed that the predictive performance of PD-1^T TILs was superior to the PD-L1 tumor proportion score (TPS)[7]. Therefore, we compared the predictive performance of PD-L1 TPS to the PD-1^T signature in the validation cohort. 9/42 (21%) pretreatment samples showed a PD-L1 TPS of ≥50%, 7/42 (17%) between 1% and 50% and 24/42 (57%) showed no PD-L1 expression. In 2/42 (5%) the PD-L1 status was unknown and these patients were excluded from further analysis **(Table S1, S3)**. PD-L1 TPS was not significantly higher in patients with DC 12m compared to patients with PD (*P*=0.30) **(Fig. 4A)**. The AUC was substantially lower compared to the PD-1^T signature (AUC: 0.63; 95% CI: 0.34-0.91, *P*=0.13), indicating a lower discriminatory ability **(Fig. 4B)**. At 50% cut-off, the sensitivity was also lower (50%), as well as were specificity (82%), PPV (33%) and NPV (90%) **(Table 2)**. Furthermore, we observed that a PD-L1 TPS of ≥50% was not associated with significantly better PFS (HR 0.77; 95% CI: 0.38-1.54, median PFS 5.7 vs 2.3 months), but did show borderline significance for improved OS (HR 0.40; 95% CI: 0.20-0.81, median OS 11.4 vs 3.0 months) **(Fig 4C,D)**. Next, we assessed the predictive accuracy at 1% cut-off, as this cut-off has also been previously studied, though with contradictory results[3–5]. Here, sensitivity (50%) and NPV (88%) were similar to the performance of 50% PD-L1 TPS, with only a slightly lower specificity (62%) and PPV (19%) **(Table 2)**. PD-L1 TPS ≥1% showed similar survival outcomes as PD-L1 TPS ≥50% **(Fig. 4E,F)**.

Cut-off	Sensitivity	Specificity	PPV	NPV
<0.35 vs ≥0.35	92%	93%	85%	96%
<0.35 vs ≥0.35	100%	64%	32%	100%
<50% vs ≥50%	50%	82%	33%	90%
<1% vs ≥1%	50%	62%	19%	88%

4

Next, we explored the predictive value of *PDCD1* gene expression, encoding PD-1, and compared this to the PD-1^T signature. Using all 58 pretreatment samples with available gene expression data from the first cohort for training, we observed that *PDCD1* scores of the DC 12m group were slightly higher compared to the PD group (*P*=0.03), with an AUC of 0.71 (95% CI: 0.54-0.88) and a score of 0.20 as the optimal cut-off **(Fig. S2A,B)**. In the validation cohort, signature scores did not significantly differ between the DC 12m group and the PD group (*P*=0.06) **(Fig. S2C)**. The AUC was lower compared to the PD-1^T signature (0.74; 95% CI: 0.50-0.99, *P*=0.36) **(Fig. S2D)**, as well as the sensitivity (83%), NPV (93%) and specificity (36%) when using the predefined cut-off of 0.20 **(Table S5)**. Patients with high (≥0.20) *PDCD1* scores did not show significantly prolonged PFS and OS compared to those with low (<0.20) *PDCD1* scores in the validation cohort **(Fig. S2E,F)**.

Taken together, these findings show that the PD-1^T signature had a higher accuracy for predicting DC 12m and survival compared to PD-L1 TPS. The predictive performance of *PDCD1* alone was lower as compared to the PD-1^T signature, further highlighting that only the presence of a T cell subset with high expression of PD-1, which is reflected by the PD-1^T signature, and not total PD-1 expression is predictive for response to PD-1 blockade.

A

B

AUC: 0.87
(95% CI: 0.74 - 0.99)

C

D

E
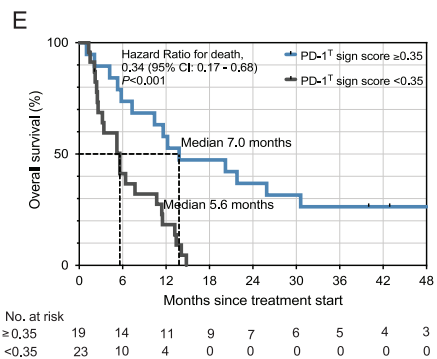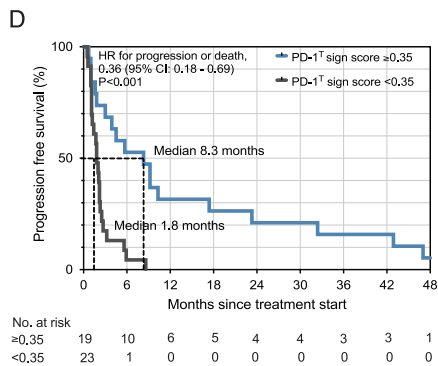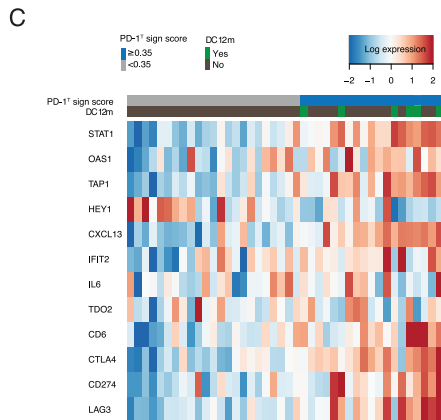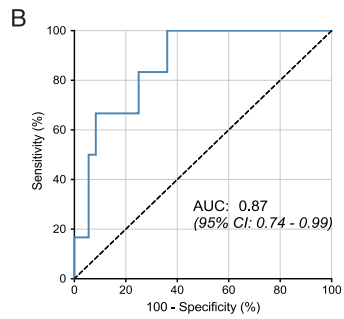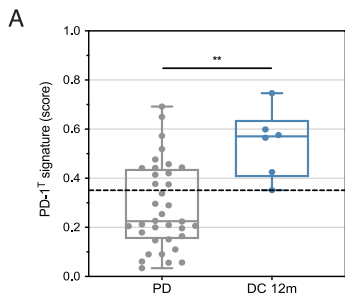
◄ **Figure 3.** PD-1^T signature validation for prediction of non-response to PD-1 blockade. **(A)** PD-1^T signature scores in pretreatment samples from patients with disease control at 12 months (DC 12m) (n=6) and patients with progressive disease (PD) (n=36) in the validation set (n=42). The dashed line indicates a cut-off score of 0.35. Medians, interquartile ranges and minimum/maximum shown in boxplots, **$P<0.01$ by Mann Whitney U-test. **(B)** Receiver operating characteristic (ROC) curve for predictive value of the PD-1^T signature for DC 12m (AUC 0.87; 95% CI: 0.74-0.99) in the validation cohort (n=42). **(C)** Heatmap showing the expression of the 12-gene PD-1^T signature in pretreatment samples from patients with DC 12m (n=6) and patients with PD (n=36) in the validation cohort (n=42) ordered by probability score. Each column represents one patient (blue: PD-1^T signature high (score of ≥0.35), light grey: PD-1^T signature low (score of <0.35), green: patients with DC 12m, dark grey: patients with PD) and rows display genes. Positive values (red) indicate higher gene expression and negative values (blue) indicate lower gene expression. **(D)** Progression-free survival (PFS) (median 8.3 months versus 1.8 months, HR 0.36; 95% CI: 0.18-0.69, ***$P<0.001$) and **(E)** overall survival (OS) (median 7.0 months versus 5.6 months, HR 0.34; 95% CI: 0.17-0.68, ***$P<0.001$) of patients with PD-1^T signature high (n=6) and PD-1^T signature low (n=36) pretreatment samples in the validation set (n=42). 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.

Figure 4. Association of PD-L1 with treatment benefit and survival. **(A)** PD-L1 tumor proportion score (TPS) in pretreatment samples from patients with DC at 12 months (DC 12m) (n=6) and patients with PD (n=34) in the validation cohort (n=40). Mean shown as dashed line, *P*=0.30 by Mann Whitney U-test. Note that for 2 patients PD-L1 TPS was unknown. **(B)** Receiver operating characteristic (ROC) curve for predictive value of PD-L1 TPS for DC 12m (AUC 0.63; 95% CI: 0.34-0.91) in the validation cohort (n=40). **(C)** Progression-free survival (PFS) (median 5.7 months versus 2.3 months, HR 0.77; 95% CI: 0.38-1.54, *P*=0.47) and **(D)** overall survival (OS) (median 11.4 months versus 3.0 months, HR 0.40; 95% CI: 0.20-0.81, *P=0.03) of patients with PD-L1 TPS ≥50% (n=9) and PD-L1 TPS <50% (n=31) pretreatment samples in the validation cohort (n=40). **(E)** PFS (median 4.2 months versus 2.1 months, HR 0.87; 95% CI: 0.46-1.63, *P*=0.65) and **(F)** OS (median 8.9 months versus 2.5 months, HR 0.56; 95% CI: 0.30-1.07, *P=0.0498) of patients with PD-L1 TPS ≥1% (n=16) and PD-L1 TPS <1% (n=24) pretreatment samples in the validation cohort (n=40). 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. ►

A

PD-L1 TPS (%)

P=0.30

PD   DC 12m

B

Sensitivity (%)

100 - Specificity (%)

AUC: 0.63
(95% CI: 0.34- 0.91)

C

Progression free survival (%)

HR for progression or death,
0.77 (95% CI: 0.38 - 1.54)
P=0.47

PD-L1≥50%
PD-L1<50%

Median 5.7 months

Median 2.3 months

Months since treatment start

No. at risk
≥50%   9   4   3   2   2   2   1   1   0
<50%   31   7   3   3   2   2   2   2   1

D

Overall survival (%)

HR for death,
0.40 (95% CI: 0.20 - 0.81)
P=0.03

PD-L1≥50%
PD-L1<50%

Median 11.4 months

Median 3.0 months

Months since treatment start

No. at risk
≥50%   9   6   4   4   4   4   3   2   2
<50%   31   9   5   3   3   2   2   2   1

E

Progression free survival (%)

HR for progression or death,
0.87 (95% CI: 0.46 - 1.63)
P=0.65

PD-L1≥1%
PD-L1<1%

Median 4.2 months

Median 2.1 months

Months since treatment start

No. at risk
≥1%   16   6   3   2   2   2   1   1   0
<1%   24   5   3   3   2   2   2   2   1

F

Overall survival (%)

HR for death,
0.56 (95% CI: 0.30 - 1.07)
P=0.0498

PD-L1≥1%
PD-L1<1%

Median 8.9 months

Median 2.5 months

Months since treatment start

No. at risk
≥1%   16   9   6   4   4   4   3   2   2
<1%   24   6   3   3   3   2   2   2   1

Discussion

In spite of the success of PD-(L)1 blockade therapies, the majority of patients do not benefit from these agents. Therefore, biomarkers that reliably allow to identify patients without clinical response are urgently needed to guide alternative treatment decisions beyond PD-1 blockade. In addition, such biomarkers should be developed using robust clinical grade platforms that can easily be implemented in a clinical setting. We previously established PD-1^T TILs, a functionally and transcriptionally distinct intratumoral T cell population with enriched tumor reactivity, as a novel predictive biomarker for long-term benefit to PD-1 blockade[7,8]. However, the measurement of PD-1^T TIL status using advanced digital image analysis-based quantification of IHC stainings is complex, and multiple – predominantly technical – sources of potential bias are challenging to cope with in a routine diagnostic setting. Therefore, we here developed a clinically applicable mRNA signature reflecting the presence of PD-1^T TILs by using the Nanostring nCounter platform. This study shows that the PD-1^T TIL IHC biomarker could successfully be translated into a gene expression signature, as the latter had a similar predictive performance to the digital IHC quantification approach[7]. Importantly, a high sensitivity and NPV (100%) was reached, which should allow to reliably identify patients without clinical benefit to PD-1 blockade alone. In this relatively small number of patients, the PD-1^T signature performed superior to PD-L1 TPS which is similar to previous work[7]. The clinical applicability of the Nanostring nCounter platform has been demonstrated previously[16] and this platform has been shown to have a high analytical sensitivity, technical reproducibility, and to generate robust data from routine FFPE samples[10]. Therefore, we expect that by using this approach the PD-1^T signature can now easily be applied in a clinical setting, and that a similar approach can be exploited for other promising biomarker candidates.

Genes in the PD-1^T signature are related to, for instance, co-inhibitory signaling (*CD274, CTLA4, LAG3*), cytokines and chemokines (*CXCL13, IL6*), interferon signaling (*IFIT2, OAS1, STAT1*), antigen presentation (*TAP1*) and angiogenesis (*HEY1*). This is in line with features of a tumor microenvironment with a pre-existing adaptive immune response as well as with ongoing immunosuppressive stimuli associated with T cell dysfunction[17,18]. Of note, *LAG3, CTLA4* and *CXCL13* correspond to the dysfunctional phenotype that characterizes PD-1^T TILs[8], indicating that the gene signature captures the presence of PD-1^T TILs in the TME. Intriguingly, the *PDCD1* gene was not among the signature genes, probably due to a partial overlap in expression levels between PD-1^T TIL high and low tumors. Notably, the predictive performance of the PD-1^T signature was better than *PDCD1* gene expression alone.

This is in line with the notion that the predictive capacity is specifically driven by the PD-1[T] TIL subset, which reflects a tumor-reactive population that is likely crucial for response to PD-1 blockade, and not simply by the presence of PD-1 positive T cells[8].

The development of predictive mRNA signatures that characterize a broader transcriptomic immune profile of the TME has gained interest. For example, the commercially available Tumor Inflammation Signature (TIS) has shown to predict clinical benefit in different cancer types, including NSCLC[11–13]. We correlated the PD-1[T] signature scores to the TIS scores in samples from the validation cohort (n=42). We observed a good correlation between the signatures (R^2=0.705, P<0.0001), in line with a partial overlap in three genes *(STAT1, CD274, LAG3)* (**Fig. S3**). While the small sample size limits conclusions that can be made from this analysis, it may further support the notion that these types of mRNA signatures can robustly detect tumor immune environments that are responsive to immune checkpoint inhibitors (ICI). In the future, it will be interesting to compare the performance of the signatures in a larger patient cohort to understand whether they report on similar features of the TME or whether there may be possible additive value for a subgroup of patients.

While the analysis of samples from two distinct expert centers on NSCLC ICI therapy strengthens the results of the current study, a number of limitations should be noted: First, the sample size was low in both the training and the validation cohorts. Thus, additional studies with higher patient numbers are needed. Second, this is a retrospective study which makes further validation in prospective studies necessary. Third, different PD-L1 IHC antibodies were used in the validation cohort which could potentially have introduced analytical differences[19,20]. Finally, an important emerging question in the clinical treatment of NSCLC is how to preselect patients for therapy with either single agent PD-1 blockade or in combination with chemotherapy. As we developed the PD-1[T] biomarker aimed at high sensitivity and NPV, the signature is currently not designed to make such predictions. Nevertheless, retraining of the biomarker for high specificity and PPV in a suitable patient cohort could be considered for this purpose. Such an approach, however, may be potentially limited by the number of false positives driven by a relatively large fraction of mixed responses in the biomarker high, PD group as shown previously[7].

Taken together, in the present study we developed a PD-1[T] gene expression signature with high sensitivity and NPV that can be used as a predictive biomarker for non-response to PD-1 blockade in NSCLC. Our data demonstrate, that our digital image-based IHC assay can reliably be replaced by a matching gene expression signature

with comparable predictive performance. This could provide a straightforward FFPE-based approach that allows for easy implementation in a routine diagnostic clinical setting. Moreover, the strategy used has the potential to bring other expression-level based biomarkers to routine clinical diagnostics where these can support shared decision making for therapeutic strategies.

4

Acknowledgements

References

1. Garon, E.B. *et al.* Pembrolizumab for the treatment of non-small-cell lung cancer. N *Engl J Med.* 2015;372(**21**):2018–28.

2. Reck, M. *et al.* Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer. *N Engl J Med.* 2016;375(**19**):1823–33.

3. Borghaei, H. *et al.* Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N Engl J Med.* 2015;373(**17**):1627–39.

4. Rittmeyer, A. *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.

5. 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.

6. Reck, M. *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. Hummelink, K. *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.

8. Thommen, D.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**).

9. Voabil, P. *et al.* An ex vivo tumor fragment platform to dissect response to PD-1 blockade in cancer. *Nat Med.* 2021;27(**7**):1250–61.

10. Veldman-Jones, M.H. *et al.* Evaluating robustness and sensitivity of the Nanostring Technologies nCounter platform to enable multiplexed gene expression analysis of clinical samples. *Cancer Res.* 2015;75(**13**):2587–93.

11. Ayers, M. *et al.* IFN- γ – related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest.* 2017;127(**8**):2930–40.

12. Cristescu, R. *et al.* Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science* (**80-**). 2018;362(6411).

13. Damotte, D. *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.

14. Adam, J. *et al.* Multicenter harmonization study for PD-L1 IHC testing in non-small-cell lung cancer. *Ann Oncol.* 2018;29(**4**):953–8.

15. Danaher, P. *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.

16. Wallden, B. *et al.* Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. *BMC Med Genomics.* 2015;8(**1**):1–14.

17. Xia, A., Zhang, Y., Xu, J., Yin, T., Lu, X.J. T Cell Dysfunction in Cancer Immunity and Immunotherapy. *Front Immunol.* 2019;10(July):1719.

18. van der Leun, A.M., Thommen, D.S., Schumacher, T.N. CD8+ T cell states in human cancer: insights from single-cell analysis. *Nat Rev Cancer.* 2020;20(**4**):218–32.

4

19. McLaughlin, 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.
20. Rimm, D.L. *et al.* A prospective, multi-institutional, pathologist-based assessment of 4 immunohistochemistry assays for PD-L1 expression in non–small cell lung cancer. *JAMA Oncol.* 2017;3(**8**):1051–8.

Supplemental material
Supplemental figures
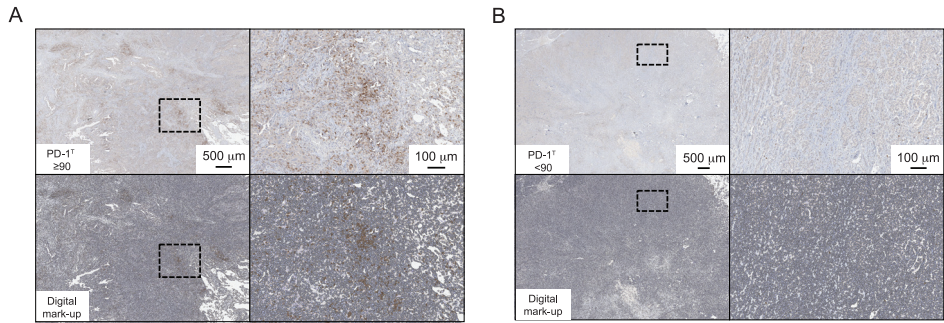
Figure S1. PD-1^T quantification by an algorithm-based immunohistochemical (IHC) imaging approach. **(A)** IHC analysis of PD-1^T TILs. Example of a PD-1^T IHC high ($\geq$90 per mm2) and **(B)** PD-1^T IHC low (<90 per mm^2) tumor with digital markup.

Figure S2. Predictive accuracy of *PDCD1* gene expression. **(A)** *PDCD1* scores in pretreatment samples from patients with disease control at 12 months (DC 12m) (n=12) and patients with progressive disease (PD) (n=46) in the training cohort (n=58). Only patients with available PD-1[T] IHC and gene expression data are included. The dashed line indicates a cut-off score of 0.20. Medians, interquartile ranges and minimum/maximum shown in boxplots, *P=0.03 by Mann Whitney U-test. **(B)** Receiver operating characteristic (ROC) curve for predictive value of *PDCD1* for DC 12m (AUC 0.71; 95% CI: 0.54-0.88) in the training cohort (n=58). **(C)** Same plot as in A for pretreatment samples from patients with DC 12m (n=6) and patients with PD (n=36) in the validation cohort (n=42). P=0.06 by Mann Whitney U-test. **(D)** ROC curve for predictive value of *PDCD1* for DC 12m (AUC 0.74; 95% CI: 0.50-0.99) in the validation cohort (n=42). **(E)** Progression-free survival (PFS) (median 2.9 months versus 2.0 months, HR 0.74; 95% CI: 0.37-1.47, P=0.35) and **(F)** overall survival (OS) (median 10.6 months versus 5.6 months, HR 0.72; 95% CI: 0.34-1.50, P=0.35) of patients with *PDCD1* high (score of ≥0.20) (n=28) and *PDCD1* low (score of <0.20) (n=14) pretreatment samples in the validation cohort (n=42).

Figure S3. Correlation of the PD-1^T signature with the Tumor Inflammation Signature (TIS). Correlation of the PD-1^T signature and the TIS in the validation cohort (n=42). R^2=0.705, P<0.0001. R^2 and P-values were calculated using linear regression analysis.

4

Supplemental tables

Table S1. Overview of all analyzed biomarkers per patient.

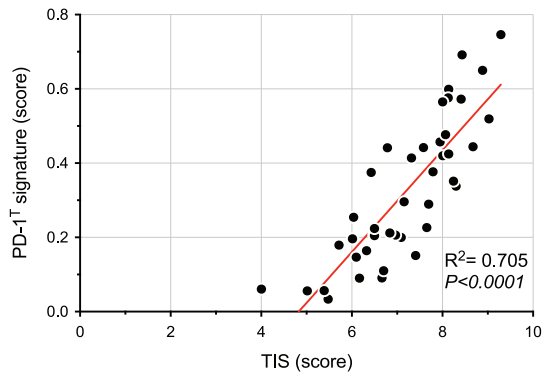| Sample ID	1= DC 12m, 0= PD	T= Training, V= validation	PD-1^T per mm^2
NKI-AVL 01	0	T	25
NKI-AVL 02	0	T	13
NKI-AVL 03	0	T	236
NKI-AVL 04	0	T	56
NKI-AVL 05	1	T	104
NKI-AVL 06	0	T	27
NKI-AVL 07	0	T	67
NKI-AVL 08	0	T	85
NKI-AVL 09	0	T	160
NKI-AVL 10	0	T	244
NKI-AVL 11	0	T	51
NKI-AVL 12	0	T	380
NKI-AVL 13	0	T	2
NKI-AVL 14	0	T	49
NKI-AVL 15	1	T	130
NKI-AVL 16	0	T	67
NKI-AVL 17	1	T	319
NKI-AVL 18	1	T	352
NKI-AVL 19	1	T	367
NKI-AVL 20	0	T	55
NKI-AVL 21	1	T	139
NKI-AVL 22	0	T	26
NKI-AVL 23	0	T	15
NKI-AVL 24	0	T	21
NKI-AVL 25	0	T	230
NKI-AVL 26	0	T	80
NKI-AVL 27	0	T	83
NKI-AVL 28	0	T	68
NKI-AVL 29	0	T	6
NKI-AVL 30	0	T	321
NKI-AVL 31	1	T	280
NKI-AVL 32	0	T	31
NKI-AVL 33	0	T	22
NKI-AVL 34	0	T	49

4

PD-L1 TPS (%)	PD-1^T signature scores	*PDCD1* signature scores	TIS scores
0	0.14313735	0.21041538	6.894
90	0.14651535	0.18958687	5.356
0	NA	0.22942241	8.714
30	0.10309239	0.21128473	6.206
1	0.6752406	0.08915112	6.951
100	0.13637567	0.20225829	6.976
0	0.25506862	0.18907535	7.473
10	0.22950462	0.23071459	6.895
0	NA	0.20943275	5.614
10	NA	0.21665969	6.643
0	0.30510214	0.21733635	7.726
1	NA	0.21748375	6.935
0	0.13819939	0.1836076	5.614
0	0.13225086	0.20029234	6.088
80	0.8372285	0.25045961	9.983
0	0.20755694	0.20377961	6.906
1	0.85051817	0.25863572	9.709
100	0.59917723	0.25248777	8.263
1	0.36533558	0.21741849	6.747
0	0.16828838	0.21972862	7.131
0	0.24642246	0.20259564	6.671
0	0.05090746	0.18342566	5.595
0	0.07232613	0.19142927	4.139
1	0.16470257	0.19862647	5.947
100	NA	0.23463825	8.602
0	0.31416743	0.20564011	6.935
0	0.20042987	0.20826304	6.875
1	0.35025771	0.21089449	8.021
10	0.06093618	0.19147168	5.785
1	NA	0.23866693	8.783
100	0.63915571	0.22100179	7.912
1	0.09786897	0.19074833	5.934
0	0.08723175	0.18909407	6.006
60	0.26389771	0.20557294	7.239

Table S1. Continued

Sample ID	1= DC 12m, 0= PD	T= Training, V= validation	PD-1^T per mm^2
NKI-AVL 35	1	T	110
NKI-AVL 36	0	T	98
NKI-AVL 37	0	T	152
NKI-AVL 38	1	T	265
NKI-AVL 39	1	T	181
NKI-AVL 40	0	T	351
NKI-AVL 41	0	T	43
NKI-AVL 42	0	T	3
NKI-AVL 43	0	T	67
NKI-AVL 44	0	T	70
NKI-AVL 45	0	T	434
NKI-AVL 46	0	T	34
NKI-AVL 47	0	T	457
NKI-AVL 48	0	T	201
NKI-AVL 49	1	T	213
NKI-AVL 50	0	T	17
NKI-AVL 51	0	T	128
NKI-AVL 52	0	T	152
NKI-AVL 53	0	T	2
NKI-AVL 54	0	T	828
NKI-AVL 55	0	T	214
NKI-AVL 56	0	T	61
NKI-AVL 57	0	T	101
NKI-AVL 58	1	T	181
NKI-AVL 59	1	V	NA
NKI-AVL 60	1	V	NA
NKI-AVL 61	0	V	NA
NKI-AVL 62	0	V	NA
NKI-AVL 63	0	V	NA
NKI-AVL 64	0	V	NA
NKI-AVL 65	0	V	NA
NKI-AVL 66	0	V	NA
NKI-AVL 67	0	V	NA
NKI-AVL 68	0	V	NA
NKI-AVL 69	0	V	NA
CERTIM 01	0	V	NA
CERTIM 02	0	V	NA

PD-L1 TPS (%)	PD-1^{T} signature scores	*PDCD1* signature scores	TIS scores
40	0.47371578	0.2086303	6.786
0	NA	0.21601595	7.093
0	NA	0.20774513	6.821
95	0.3880614	0.20910561	6.846
0	0.64223869	0.2151738	7.570
40	NA	0.23791112	9.791
0	0.23592662	0.20604298	6.730
0	0.13767804	0.19327852	5.439
0	0.22376986	0.2162488	7.057
0	0.20403531	0.19266077	7.139
0	NA	0.2198334	6.633
0	0.46417086	0.21368521	7.425
0	NA	0.1925737	7.831
10	NA	0.21595654	7.838
0	0.51176161	0.22551758	6.895
0	0.21179962	0.1968275	6.496
0	NA	0.18916553	6.182
1	NA	0.19097529	6.527
0	0.05472433	0.17966245	4.777
0	NA	0.21645211	8.767
0	NA	0.19726592	7.054
60	0.1871965	0.19130507	5.833
0	NA	0.18832904	5.680
100	0.43659426	0.20833796	7.817
0	0.59875534	0.23130271	8.136
0	0.57577079	0.23659817	8.126
0	0.33791226	0.22503081	8.299
1	0.05584476	0.1953445	5.016
70	0.57234398	0.22409197	8.412
0	0.03369218	0.19838459	5.475
20	0.51901336	0.25715829	9.026
1	0.41398173	0.20309417	7.316
100	0.16411478	0.20684874	6.324
70	0.41995247	0.21239991	8.007
0	0.05657554	0.18272455	5.389
0	0.45709704	0.19546895	7.947705372
0	0.37482841	0.19669646	6.426587946

Table S1. Continued

Sample ID	1= DC 12m, 0= PD	T= Training, V= validation	PD-1^T per mm^2
CERTIM 03	0	V	NA
CERTIM 04	0	V	NA
CERTIM 05	0	V	NA
CERTIM 06	1	V	NA
CERTIM 07	0	V	NA
CERTIM 08	0	V	NA
CERTIM 09	0	V	NA
CERTIM 10	0	V	NA
CERTIM 11	0	V	NA
CERTIM 12	0	V	NA
CERTIM 13	1	V	NA
CERTIM 14	0	V	NA
CERTIM 15	0	V	NA
CERTIM 16	1	V	NA
CERTIM 17	0	V	NA
CERTIM 18	0	V	NA
CERTIM 19	1	V	NA
CERTIM 20	0	V	NA
CERTIM 21	0	V	NA
CERTIM 22	0	V	NA
CERTIM 23	0	V	NA
CERTIM 24	0	V	NA
CERTIM 25	0	V	NA
CERTIM 26	0	V	NA
CERTIM 27	0	V	NA
CERTIM 28	0	V	NA
CERTIM 29	0	V	NA
CERTIM 30	0	V	NA
CERTIM 31	0	V	NA

NA, not available

PD-L1 TPS (%)	PD-1^T signature scores	*PDCD1* signature scores	TIS scores
0	0.09061181	0.19352412	6.663844276
50	0.19990922	0.19174377	7.089186346
0	0.37673778	0.21397528	7.790608182
100	0.42474192	0.20949471	8.13669394
0	0.20888208	0.20910316	6.48711348
0	0.28941286	0.20018483	7.695182459
0	0.29583936	0.20291918	7.152807509
0	0.2060138	0.19456563	6.963596419
5	0.20464135	0.20806045	6.498026588
0	0.21180426	0.18821267	6.839139019
100	0.74622667	0.24092749	9.289461347
0	0.47661187	0.20975658	8.067153437
0	0.11008555	0.22364154	6.701786258
60	0.56494611	0.19211549	8.003015848
0	0.22378207	0.1936509	6.498107194
90	0.09012218	0.18721871	6.168103946
0	0.35146763	0.21564433	8.243507416
0	0.44403454	0.2192682	8.673249687
5	0.15117402	0.19661584	7.40775649
0	0.1960908	0.20657123	6.012744812
NA	0.17913245	0.20169567	5.716478727
10	0.14672321	0.20961013	6.097363206
80	0.64992468	0.23851804	8.884735325
0	0.44134557	0.2070323	6.783573802
NA	0.06060794	0.17208473	4.003295911
0	0.22646001	0.20680609	7.654154777
30	0.44175787	0.22527499	7.579428401
0	0.691726	0.22474679	8.43256737
0	0.25405197	0.20390309	6.038361128

Table S2. Coefficients of the PD-1^T signature genes used in the prediction model.

Gene	Coefficient
STAT1	0.071
OAS1	0.126
TAP1	0.193
HEY1	-0.257
CXCL13	0.147
IFIT2	0.036
IL6	0.028
TDO2	0.054
CD6	0.008
CTLA4	0.219
CD274	0.006
LAG3	0.154

Table S3. 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.

	P-value	Training cohort (n=41)	Validation cohort (n=42)
Sex	*P*=0.51		
Male, no. (%)		22 (54%)	26 (62%)
Female, no. (%)		19 (46%)	16 (38%)
Age (years), mean (s.d.)	*P*=0.79	64 (9.2)	64 (9.6)
Smoking (never/ex/current)	*P*=0.39	3/32/6	6/27/9
Pack years, median (IQR)	*P*=0.27	34 (25)	30 (35)
PS, no. (%)	*P*=0.002		
0		11 (27%)	3 (7%)
1		22 (54%)	25 (60%)
≥2		8 (19%)	14 (33%)
Pathology, no. (%)	*P*=0.87		
Adeno		28 (68%)	27 (64%)
Squamous		9 (22%)	12 (29%)
LCNEC, NSCLC-type		1 (3%)	0 (0%)
NSCLC, NOS		3 (7%)	3 (7%)
PD-L1 TPS, no. (%)	*P*=0.58		
Negative		22 (54%)	24 (57%)
Positive 1-50%		10 (24%)	7 (17%)
Positive ≥50%		9 (22%)	9 (21%)
Unknown		0 (0%)	2 (5%)
Brain metastases, no. (%)	*P*=0.6028	10 (24%)	8 (19%)
Treatment, no. (%)			
Nivolumab		41 (100%)	42 (100%)
Line of treatment, no (%)	*P*=0.62		
2		32 (78%)	30 (71%)
>2		9 (22%)	12 (29%)
Best Overall Response	*P*=0.63		
CR		3 (7%)	2 (5%)
PR		9 (22%)	5 (12%)
SD		4 (10%)	5 (12%)
PD		25 (61%)	30 (71%)
DC			
≥12 months	*P*=0.12	12 (29%)	6 (14%)
≥6 months	*P*=0.63	13 (32%)	11 (26%)

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.

Table S4. Overview of significantly expressed genes between tumors from PD-1^T IHC high patients (n=12) and PD-1^T IHC low patients (n=29).

Gene	logFC	Average Expression
LAG3	1.88884548400466	-2.80041279814709
PSMB9	1.43868956579743	0.302996932117271
CTLA4	1.42200236599635	-2.47865299771225
CXCR6	1.6446782493738	-2.96961697142336
CXCL13	2.2175668262985	-1.70636645256783
SH2D1A	1.75917395751567	-3.32747140742716
TAP1	1.07179116679377	0.153764377204745
CD3E	1.46133550402829	-2.43718170364179
CD274	1.77461018365805	-2.62787569257075
CD8A	1.72830415204725	-2.03954190686031
CD3G	1.50938013286876	-3.45143656891331
CD3D	1.61835438828301	-1.69109606921838
CD6	1.58051071356482	-3.28877942082266
GZMK	1.65813054066268	-2.44753255939726
IRF1	1.08945362100745	-0.000809016866967984
STAT1	1.00072220553547	1.3077408591629
IFNG	1.74152190267805	-4.44726680139924
ZAP70	1.40194538684443	-2.43585243496581
GZMA	1.70542230583565	-2.13364043308839
IFI35	0.987717180524364	-0.841730242635249
CCR5	1.32440469417496	-2.76216599584521
CD69	1.37959073066651	-1.92613327085766
KLRK1	1.40427005952429	-2.62445715181827
CXCL10	1.78562490084712	-0.404538293424959
CD5	1.20112720197047	-2.76496364144761
TRAT1	1.44947890393937	-3.23767089629177
LY9	1.24574092018984	-2.99614570191887
HEY1	-1.39760491698939	-0.834522518489289
IFIT3	1.110231762069	-0.559615244491435
TDO2	1.56097678398257	-2.13703000505588
HLA-E	0.762834406039039	2.56854832062947
IDO1	1.73063654284901	-1.30330315884557
PRF1	1.44354677083376	-2.74740428352741
CD48	1.30252149946313	-1.62285167747091
NKG7	1.66816172872997	-1.25067906555379
IL2RG	1.18159319380198	-0.771658463957608

T-statistics	P Value	Adjusted P Value	Bayesion factor
4.83909975303927	1.78582558984076e-05	0.00513959649734931	2.84658700406427
4.68297306902122	2.95030175588419e-05	0.00513959649734931	2.3794555181907
4.58614199005501	4.01997988520925e-05	0.00513959649734931	2.09185824232994
4.56461815426475	4.30518750486993e-05	0.00513959649734931	2.02816979330291
4.55957550465733	4.374827644334e-05	0.00513959649734931	2.01326176709522
4.52323657993327	4.91044251339106e-05	0.00513959649734931	1.90597879026597
4.42871623399643	6.62333066853745e-05	0.0059420737997736	1.62819981640937
4.34655157267091	8.57844244323353e-05	0.0062493301258 9224	1.38831333947697
4.32375917934292	9.21423282844515e-05	0.0062493301258 9224	1.32204345171818
4.29919107351712	9.95116262084752e-05	0.0062493301258 9224	1.25074848681342
4.26277330674192	0.000111505813579992	0.0063659682662032	1.14533534077684
4.2243295633599	0.000125697954760364	0.0065781929657 9236	1.03441469824883
4.18473695715741	0.00014215324524755	0.00686709523195857	0.920573026443
4.14705536140784	0.00015975506400955	0.00716615572842837	0.812607525409131
4.11311177849518	0.000177416210315866	0.00725034112981547	0.715678990937824
4.10002366246689	0.000184722067001668	0.00725034112981547	0.678389223896438
4.0748260400368	0.000199620893678133	0.00737423066058042	0.606732278127318
3.98141401445179	0.00026574207772818	0.0087048654777 7766	0.342677146403622
3.97952213627703	0.000267279924727654	0.0087048654777 7766	0.337355790598474
3.96753679245733	0.000277225015215849	0.0087048654777 7766	0.303669250065301
3.92721579227505	0.000313387613303551	0.00937178195974429	0.190662987446336
3.8988727178203	0.000341502217054782	0.00974833601410923	0.111528845611906
3.77837064927973	0.000490778589434132	0.012907444760 1548	-0.222018805317814
3.77667217154252	0.000493278143700185	0.012907444760 1548	-0.226685616413969
3.75595053892489	0.000524780061140286	0.013182475135844	-0.283541712241419
3.71034624896656	0.000601092014501517	0.0145186840425751	-0.408145952388601
3.69450105663426	0.000630028960993021	0.0146540069445784	-0.451267919858972
-3.66050500135453	0.000696709420999094	0.0155998317781292	-0.543482241972573
3.63627225806104	0.000748332439827997	0.0155998317781292	-0.608956479871537
3.63533485688183	0.000750401530542069	0.0155998317781292	-0.611484888627425
3.6265525721914	0.000770055390321665	0.0155998317781292	-0.635157109489466
3.57629413697303	0.000892401931916592	0.0175133879138631	-0.7700698361775
3.54542071855368	0.000976590425975104	0.0180202149542676	-0.85246821215889
3.54275584938563	0.000984204057044527	0.0180202149542676	-0.859563203214019
3.53581193874381	0.00100431134299262	0.0180202149542676	-0.878037769613107
3.50991373160624	0.0010828328324166	0.018889417 1877119	-0.946774068063469

Table S4. Continued

Gene	logFC	Average Expression
CCL4	1.37652182502365	-1.15162060939353
CD244	1.0140109917243	-4.13601859978355
GIMAP4	0.910008059694032	-0.110032084999837
CCL5	1.34205185377604	-0.41616328442252
IL6	1.6751565235166	-1.85138061368234
CXCL14	-2.43479633694284	1.19030995683133
TNFRSF1B	0.931778521160674	-0.633533347850366
IL2RB	1.20344276687788	-1.41007205577328
CD40	0.917900814704211	-0.693642929932606
IFIT2	1.18990439064898	-1.60118350410009
SLAMF7	1.310766818109	-2.38257048577135
HLA-F	1.05618112032267	-0.105931695806803
NCR1	0.960005951631412	-4.16203875925747
CD7	1.15797115824849	-2.92732873037538
HLA-C	0.943096476582807	3.12629587925205
HSD11B1	1.23825740268296	-2.96366472328225
OAS1	0.916330131374769	0.7835427792938
OAS2	0.964878990316807	0.159982062948577
PIK3CD	0.910004000640195	-1.47076303939425
RRM2	0.861345905307591	-0.16258390339206

Table S5. Predictive accuracy of *PDCD1* gene expression signature, summary of training and validation results.

	Biomarker	AUC	Cut-off
Training cohort (n=58)	*PDCD1*	0.71 95% CI: 0.54-0.88	<0.20 vs ≥0.20
Validation cohort (n=42)	*PDCD1*	0.74 95% CI: 0.50-0.99	<0.20 vs ≥0.20

T-statistics	P Value	Adjusted P Value	Bayesion factor
3.40516640779077	0.00146460040912612	0.0248586231603028	–1.22202988086957
3.29001171724326	0.00203162237460295	0.0335752329276488	–1.5193151090479
3.28044048418724	0.0020871554928o246	0.0336085551148703	–1.54376365326564
3.26944024694903	0.00215276026558408	0.0337983361696701	–1.57181186993599
3.25141577878206	0.00226450748828835	0.0346856268937826	–1.61765275641612
–3.20732072434023	0.00256153449492565	0.0379144834890493	–1.72917479474905
3.18603193642275	0.00271780639232291	0.0379144834890493	–1.78269593555762
3.17866032782817	0.00277399012810882	0.0379144834890493	–1.80117926832359
3.17042387987491	0.00283806429807378	0.0379144834890493	–1.8218009080922
3.17034219468396	0.00283870670848873	0.0379144834890493	–1.82200526392905
3.16800927938585	0.00285711198423105	0.0379144834890493	–1.82784030499371
3.15810405319661	0.00293652191218481	0.0379144834890493	–1.85258644150503
3.15005089263706	0.00300261649391027	0.0379144834890493	–1.8726713287389
3.14812028847099	0.00301866906759946	0.0379144834890493	–1.877481749504
3.13625194586673	0.00311914668878484	0.0384083160893506	–1.90701465469001
3.09942853822148	0.00345136315249579	0.0416818473032184	–1.99821434057035
3.07351166731402	0.00370488487846314	0.0438993906353746	–2.06200681564335
3.05982631055085	0.0038457314590148	0.0442309269563468	–2.0955590246727
3.05716235556109	0.00387372767929789	0.0442309269563468	–2.10207942725563
3.0401575234094	0.00405700347078921	0.045496396065279	–2.14361792408683

Sensitivity	Specificity	PPV	NPV
92%	41%	29%	95%
83%	36%	18%	93%

PART 2

Serum and pleural effusion
as bio-sources for diagnostic
biomarker tests

Chapter 5

A serum protein classifier identifying patients with advanced non-small cell lung cancer who derive clinical benefit from treatment with immune checkpoint inhibitors

Mirte Muller[1]*, Karlijn Hummelink[2]*, Daniel Hurkmans[3] , Anna-Larissa Niemeijer[4], Kim Monkhorst[2], Joanna Roder[5] , Carlos Oliveira[5] , Heinrich Roder[5], Joachim G. Aerts[3]#, Egbert F. Smit[1]#.

* These authors contributed equally as first-author
#These authors jointly supervised this work

[1]Department of Thoracic Oncology, [2]Department of Pathology, Netherlands Cancer Institute, Amsterdam, The Netherlands, [3]Department of Pulmonary Diseases, Erasmus Medical Center, Rotterdam, The Netherlands, [4]Department of Pulmonary Diseases, Vrije Universiteit VU Medical Center, Amsterdam, The Netherlands, [5]Biodesix, Boulder, Colorado.

Translational Relevance

Predictive biomarkers for the efficacy of PD-(L)1 inhibition in non-small cell lung cancer (NSCLC) beyond PD-L1 are lacking. We retrospectively developed a pretreatment proteomic signature derived from peripheral blood that was able to stratify patients for benefit of nivolumab in treatment of relapsed NSCLC. A signature consisting of 274 mass spectral features derived from a development set of 116 patients was associated with progression free survival (PFS) and overall survival (OS) across 2 validation cohorts (n=98 and n=75). In pooled analysis, a significantly better OS was demonstrated for "sensitive" relative to "not sensitive" patients, hazard ratio (HR) 0.58 (95% CI: 0.38-0-87, *P*=0.009). There was no significant association with clinical factors including PD-L1 immunohistochemistry. Further prospective exploration of the predictive capabilities of this assay is underway.

Abstract

Purpose
Pretreatment selection of non-small-cell lung cancer (NSCLC) patients who derive clinical benefit from treatment with immune checkpoint inhibitors would fulfill an unmet clinical need by reducing unnecessary toxicities from treatment and result in substantial health care savings.

Patients and Methods
In a retrospective study, mass spectrometry (MS) based proteomic analysis was performed on pretreatment sera derived from advanced NSCLC patients treated with nivolumab as part of routine clinical care (n=289). Machine learning combined spectral and clinical data to stratify patients into three groups with good ("sensitive"), intermediate and poor ("resistant") outcomes following treatment in the second-line setting. The test was applied to three independent patient cohorts and its biology investigated using protein set enrichment analyses (PSEA).

Results
A signature consisting of 274 MS features derived from a development set of 116 patients was associated with progression free survival (PFS) and overall survival (OS) across 2 validation cohorts (n=98 and n=75). In pooled analysis, significantly better OS was demonstrated for "sensitive" relative to "not sensitive" patients treated with nivolumab, HR 0.58 (95% CI: 0.38-0-87, *P*=0.009). There was no significant association with clinical factors including PD-L1 expression, available from 133/289 patients. The test demonstrated no significant association with PFS or OS in a historical cohort (n=68) of second-line NSCLC patients treated with docetaxel. PSEA revealed proteomic classification to be significantly associated with complement and wound healing cascades.

Conclusions
This serum-derived protein signature successfully stratified outcomes in cohorts of advanced NSCLC patients treated with second line PD-1 checkpoint inhibitors and deserves further prospective study.

Introduction

The addition of immune checkpoint inhibitors (CPIs) to the armamentarium of medical treatment of advanced non-small cell lung cancer (NSCLC) has increased survival for a minority of patients. Historically, in patients with metastatic disease, 2-year survival rates following platinum-based chemotherapy were 10-20%[1]. In recent phase III studies, either comparing CPIs alone or CPI-chemotherapy to chemotherapy[2], 2-year survival rates in the CPI arms range from 32% to 67%. In addition, long term follow-up of patients treated in early single agent CPI studies indicates that 5-year survival of 15-20% may be expected, even in heavily pretreated patients[3,4].

At the same time, it is clear that not all patients benefit from treatment with CPIs. Indeed, response rates and survival times can be augmented by pretreatment selection based on tumor characteristics such as PD-(L)1 expression[5], staining of CD8 positive cells[6], tumor mutational burden (TMB)[7] and other genomic markers[8,9]. The predictive power of the best studied of these, PD-L1 immunohistochemistry, is far from perfect. For example, in previously treated NSCLC patients with PD-L1 staining of at least 50%, the objective response rate (ORR) to pembrolizumab is 44%[5]. Thus, alternative predictive biomarkers for response and clinical benefit are needed. We sought to develop a serum-based, pretreatment protein test to avoid the need for tissue biopsies, which are typically required to analyze tumor-related biomarkers. Here, we report on the development of such a test in advanced NSCLC treated with single agent CPI in the second-line setting.

Methods

Patient cohorts and sample sets

Pretreatment serum samples, collected within one month of immunotherapy initiation, were available from four cohorts of patients. The development set consisted of 116 samples from patients treated at the Netherlands Cancer Institute between May 2015 and March 2017. Validation set 1 consisted of 98 samples from patients treated at Vrije Universiteit Medical Center or Netherlands Cancer Institute between June 2015 and July 2018. Validation set 2 comprised samples from 75 patients treated at Erasmus University Medical Centre in Rotterdam between April 2016 and July 2017. Patients, identified according to criteria established in the phase 3 trials demonstrating benefit for nivolumab over docetaxel[10,11], received nivolumab 3mg/kg, administered as an intravenous infusion, every 2 weeks, for advanced NSCLC after platinum containing chemotherapy as part of routine clinical care. Patients in the development cohort and validation set 2 were treated in second line. Validation set 1 contained 58 patients treated in second line and 40 patients treated in higher lines. The cohorts comprised all patients in the respective institutions who provided pretreatment serum samples available for analysis, were eligible for immunotherapy as routine care, and who received at least one dose of nivolumab. Response to treatment was evaluated according to RECIST v1.1 every 6 weeks for the first 12 weeks and every three months thereafter. In addition, a fourth cohort of pretreatment serum samples (chemotherapy cohort) had been collected from patients with advanced NSCLC treated in second line with chemotherapy while enrolled in a clinical trial (NCT00989690)[12]. Samples were available for 68 of the 74 patients who received docetaxel ($75mg/m^2$ every 21 days) in this study. Trial inclusion and exclusion criteria have been published elsewhere[12]. All samples were obtained in the context of biobanking protocols or a clinical trial for which institutional review board approval was sought and obtained. All patients provided written informed consent according to local ethical standards and adhered to standards set out in the Declaration of Helsinki. Progression free survival (PFS) was measured from start of treatment until progression of disease, death or loss to follow up. Overall survival (OS) was defined as time from start of therapy until death or loss to follow up.

PD-L1 immunohistochemistry

Tumor PD-L1 expression scoring was performed according to the instruction manual of the qualitative immunohistochemical assay developed as a complementary diagnostic tool for nivolumab (PD-L1 IHC 22C3 pharmDx, Dako, Carpinteria, CA). PD-L1 expression levels were determined by observing complete circumferential or partial linear expression (at any intensity) of PD-L1 on the plasma cell membrane of

viable tumor-cells. In parallel, the pattern of staining in CD4 stained slides, which also stain CD4[+] lymphocytes and macrophages, was evaluated and compared to PD-L1 stained slides in order to avoid false positive assessment due to PD-L1 expressing macrophages in between tumor cells. Assessment of expression levels was performed in sections that included at least 100 tumor cells that could be evaluated.

Spectral acquisition and processing
Samples were processed using standardized operating procedures. We used the Deep MALDI® method of mass spectrometry on a matrix-assisted laser desorption/ionization (MALDI) time-of-flight mass spectrometer (SimulTof Systems, Marlborough, MA, USA) to generate reproducible mass spectra from small amounts of serum (3 μL)[13]. This approach reveals mass spectral (MS) peaks with a greater dynamic range than previously possible by exposing the samples to 400,000 MALDI laser "shots", rather than the several thousand used in standard applications. The spectra were processed to render them comparable between patients and 274 MS features (peaks) were selected for further analysis for their known reproducibility and stability (listed in supplement). Sample processing and MS analysis followed methods previously presented[14,15] and are outlined in the supplementary materials. Parameters for these procedures were established using only the 116-sample development set, and this fixed procedure was applied to all other sample sets without modification.

Test Development
Test development was carried out using the Diagnostic Cortex® platform[16], which has been used previously to design tests able to stratify patients by outcome in various settings, for example, to identify patients with advanced melanoma likely to be sensitive to checkpoint inhibitors[14,15]. The approach incorporates machine learning concepts and elements of deep learning[17] to facilitate test development in cases where there are more measured attributes than samples. The potential for overfitting is minimized, thus allowing the creation of tests that can generalize to unseen datasets. Tests are created averaging over many splits of the development set into training and test sets, and reliable test performance estimates can be obtained from the development set by restricting averages to the test set evaluations ('out-of-bag estimates')[18].

For successful supervised learning, suitable training class labels are required. We used a semi-supervised approach[19] that does not require accurate pre-specification of patients into better or worse outcome training classes and allows us to be guided by the gold standard time-to-event outcomes of OS and PFS. An approximation is

made for training classes, with patients with the lowest time-to-event outcome times assigned to the 'negative' class and those with the highest time-to-event outcome times assigned to the 'positive' class. A classifier is constructed using these training classes and used to generate classifications for all samples in the development set using out-of-bag evaluations. These resulting classifications are then used as improved training class labels for a second iteration of classifier construction. This simultaneous iterative refinement of the classifier and the classes used in training generally converges quickly and reveals the underlying structure of the MS data and its association with clinical outcomes[19]. Full details of the application of the method in this setting are provided in the supplementary materials.

One classifier previously developed with the Diagnostic Cortex platform was used as part of the developed test. BDX008 was created to stratify patients with advanced melanoma into groups with better and worse outcomes when treated with nivolumab[15]. It has been validated in multiple independent cohorts of melanoma patients treated with CPIs[15,20]. Also, it has demonstrated some ability to stratify overall survival of patients with advanced NSCLC treated with nivolumab[21]. A version of BDX008, adapted for the spectral preprocessing parameters and feature definitions in this project, was created (See Supplemental Data: Methods for details).

Preliminary statistical considerations showed a binary split of the development set into equal-sized groups would allow detection of a hazard ratio between the groups of 0.55 with 90% power, assuming fully mature clinical data and a significance level of 95%. Similar considerations for a ternary split into equal size subgroups would allow detection of a hazard ratio of 0.48 under the same specifications.

All reference data and test parameters were generated solely using the development set. Validation sets and the chemotherapy cohort were never used in the creation of any components of the test. All elements of the classification algorithms were locked prior to running the test on the validation sets and chemotherapy cohort.

Protein set enrichment analysis (PSEA)
This analysis applies the gene set enrichment analysis (GSEA) method[22] to protein expression data. The method identifies expression differences that are consistent across prespecified groups or sets of attributes, in this case, sets of proteins that are associated with particular biological processes. Two additional independent reference sets of serum samples with matched MS data and protein expression data were used for this set enrichment analysis. One sample set was composed of 49 samples with protein expression data from a panel of 1,129 proteins; the second

set had 100 samples with protein expression data from a panel of 1,305 proteins. (Protein expression measurements were generated by SomaLogic, Boulder, CO, USA.) Specific protein sets were created as the intersection of the list of the panel targets and results of queries for biological functions from GeneOntology, using AmiGO2 tools (http://amigo.geneontology.org/amigo) and UniProt databases (https://www.uniprot.org/). The PSEA method associated test classification with these biological functions via a rank-based correlation of the measured protein expressions with the test classifications of the reference samples[23]. The mass spectral features associated with biological processes (in particular Immune Response Type 2) were determined using Spearman correlation of the measured protein expressions with the mass spectral features[23] using the 49-sample reference set only. While the implementation closely follows the GSEA approach, we employed an extension of the standard method that increases the statistical power to detect associations between phenotype (test classification subgroup) and biological process[24]. The PSEA was carried out using a C# implementation and Matlab (MathWorks, Natick, MA, USA). PSEA *P*-values were defined as described by Subramanian and colleagues[20]. False discovery rates (FDRs) for the PSEA calculations were assessed using the method of Benjamini and Hochberg[25].

Other Statistical analysis
All analyses, except the PSEA, were carried out using SAS9.3 (SAS Institute, Cary, NC, USA) or PRISM (GraphPad, La Jolla, CA, USA). Survival/progression-free survival plots and medians were generated using the Kaplan–Meier method. Association between test classification and categorical or continuous variables was assessed using Fisher's exact test and Mann-Whitney test, respectively. All *P*-values are two-sided.

Results

Patient characteristics and overall outcomes for all four cohorts are summarized in **Table 1** and were typical of patients with advanced NSCLC treated predominantly in the second-line setting. Clinicopathological characteristics were generally similar between the four cohorts, although the proportion of patients with performance status 2 or higher was larger in the development cohort and validation set 1, and the proportion of patients with performance status 0 was higher in the chemotherapy cohort. PD-L1 status was not available for the chemotherapy cohort and was missing for at least one-third of patients in the other three cohorts.

Development of the test
A ternary test was developed that was able to stratify the development set of 116 samples into three groups with different outcomes after anti-PD-1 treatment, i.e., the resistant group (with poor outcomes), the intermediate group (with intermediate outcomes) and the sensitive group (with good outcomes). The ternary test result was generated by combining the results of three binary classification algorithms (classifiers). Each of the three classifiers stratified patients into two groups: "positive", with better outcomes and "negative" with worse outcomes. The binary results were integrated, as shown in **figure 1**, to yield the final test result. First, classifications were generated for all samples by Classifier A, the version of the pre-existing BDX008 test adapted to the spectral processing used in this project. To identify a group of patients least likely to have good outcomes, the patients classified as negative by Classifier A were subsequently classified by Classifier C. This classifier was developed using the subset of MS features found to be associated with Immune Response type 2 by PSEA and a subset of the development cohort enriched for inferior outcomes, by excluding patients designated as BDX008+ and having performance status 0. (The MS features in this subset are listed in the supplementary materials.) Samples designated as negative by both Classifier A and Classifier C were classified as "resistant". To identify a group of patients likely to have the best outcomes, the patients classified as positive by Classifier A were further classified by Classifier B. This classifier was developed using all 274 mass spectral features on a subset of the development set enriched for better outcomes, by excluding patients who were classified both as BDX008- and negative by Classifier C. Samples designated positive by both Classifier A and Classifier B were classified as "sensitive". All samples not classified as "sensitive" or "resistant" were classified as "intermediate". More details of the test development process and parameters are provided in the supplement. Reproducibility was assessed by running the test on the 98 serum samples of validation set 1 twice, 13 months apart. Concordance between classifications was 85%.

For identification of patients with resistant outcomes (resistant vs not resistant (i.e., sensitive and intermediate)), concordance was 91% and for identification of patients with sensitive outcomes (sensitive vs not sensitive (i.e., resistant and intermediate)), concordance was 93%.

Table 1. Patient characteristics and outcomes for all cohorts: development, validation set 1, validation set 2, and chemotherapy cohort.

		Development (n = 116)	Validation 1 (n = 98)	Validation 2 (n = 75)	Chemotherapy (n = 68)
Age	median (range)	65 (43-83)	64 (29-77)	65 (35-78)	64 (39-77)
Gender, n (%)	Male	66 (57)	51 (52)	48 (64)	52 (76)
	Female	50 (43)	47 (48)	27 (36)	16 (24)
PS, n (%)	0	36 (32)	20 (20)	18 (32)	35 (51)
	1	60 (54)	65 (66)	37 (66)	29 (43)
	2+	15 (14)	13 (13)	1 (2)	4 (6)
Smoking status, n (%)	Ever	104 (91)	88 (92)	61 (92)	64 (94)
	Never	10 (9)	8 (8)	5 (8)	4 (6)
Histology, n (%)	Adeno	77 (66)	42 (74)	49 (65)	47 (75)
	Squamous	26 (22)	10 (18)	17 (23)	12 (19)
	Other	13 (11)	5 (9)	9 (12)	4 (6)
PD-L1, n (%)	≥1%	33 (28)	12 (14)	16 (21)	0 (0)
	<1%	43 (37)	30 (29)	9 (12)	0 (0)
	NA	40 (34)	56 (57)	50 (67)	68 (100)
Response, n (%)	CR	1 (1)	1 (0)	0 (0)	0 (0)
	PR	16 (14)	28 (28)	15 (20)	7 (10)
	SD	19 (16)	26 (33)	25 (33)	23 (34)
	PD	65 (56)	37 (33)	31 (41)	22 (32)
	NA/NE	15 (13)	6 (7)	4 (5)	16 (24)
PFS (months)	Median	2.6	4.1	4.3	3.5
OS (months)	Median	8.5	8.4	12.0	8.0

Abbreviations: %, % of patients with available data; adeno, adenocarcinoma; NA, not available; NE, not evaluable; PD, progressive disease; PD-L1, PD-L1 expression; PS, performance status; SD, stable disease.
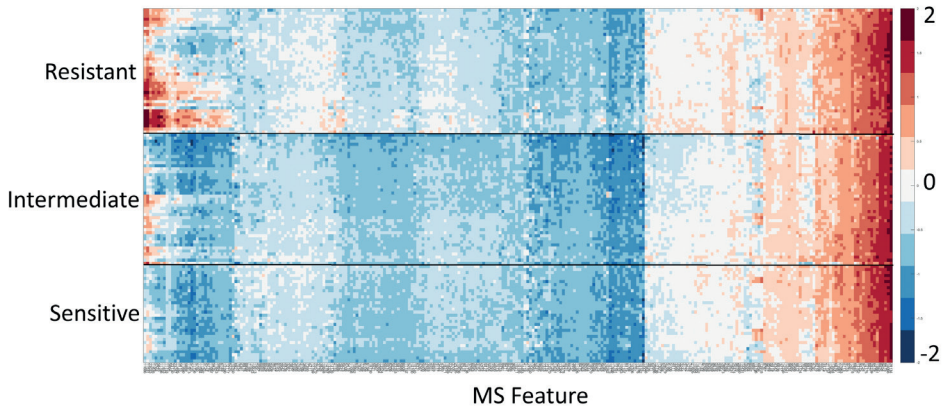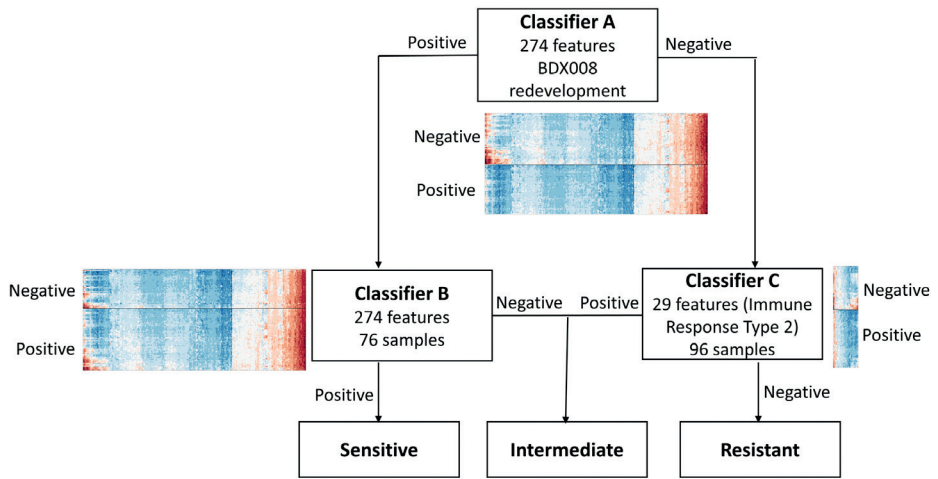
Figure 1. Schema showing how the final test result is produced from the three classifiers A, B and C. Heatmaps within the scheme show $\log_{10}$ values of features used in each classifier (x-axis) for the development cohort of 116 samples, grouped by individual classifier results, negative or positive. The heatmap below the schema shows the $\log_{10}$ values of all 274 features used within the test for all samples in the development cohort, grouped by test classification (resistant, intermediate or sensitive). Larger versions of the heatmaps are in the Supplementary Data.

The test was able to stratify patients into three groups (sensitive, intermediate and resistant) with different OS and PFS. Of the 116 samples in the development set, 41 (35%) were classified as resistant, 43 (37%) as intermediate and 32 (28%) as sensitive. Kaplan-Meier plots of OS and PFS by classification groups are shown in **Figures 2A** and **2B**. PFS for the resistant subgroup was significantly shorter than for the other groups (resistant vs sensitive: HR=0.33 (95% CI:0.19-0.58), *P*<0.001; resistant vs intermediate: HR=0.59 (95% CI: 0.37-0.96), *P*=0.035). Median PFS was 1.4 (95% CI: 1.3-2.3) months for the resistant group,

4.3 (95% CI: 1.4-5.7) months for the intermediate group and 9.1 (95% CI: 2.5-undefined) months for the sensitive group. OS for the resistant subgroup was significantly shorter than for the sensitive subgroup and numerically shorter than for the intermediate group (resistant vs sensitive: HR=0.34 (95% CI: 0.19-0.64), $P<0.001$; resistant vs intermediate: HR=0.63 (95% CI: 0.38-1.06), $P=0.083$. Median OS was 4.3 (95% CI: 2.0-7.9) months for the resistant subgroup, 10.4 (95% CI: 5.9-11.4) months for the intermediate group and 17.3 (95% CI: 8.5-undefined) months for the sensitive group. Test classification was also associated with response ($P<0.001$, see Supplemental Data: Results Supplementary Table 12). Eighty-five percent of patients classified as resistant experienced progressive disease as best response and only ten percent had a response (all partial). In the sensitive group, only 28% of patients had progressive disease as best response and 28% achieved a response (1 CR and 8 PRs as best response out of 32 patients).

For differentiating patients with the worst outcome from the remainder of the cohort, we compared the resistant subgroup with the "not resistant" group, i.e., the combination of intermediate and sensitive subgroups, see **Figures 2C** and **2D**. The resistant subgroup had significantly inferior OS and PFS than the other patients (HR=0.48 (95% CI:0.30-0.77), $P=0.002$ for OS and HR=0.46 (95% CI: 0.30-0.71), $P<0.001$ for PFS). These differences remained significant for PFS ($P=0.015$) and trended to significance for OS ($P=0.062$) in multivariate analysis when adjusted for other baseline characteristics, including performance status and PD-L1 expression.

The patients with the best outcomes (sensitive subgroup) were compared with the "not sensitive" group, i.e., the remainder of the cohort (resistant+intermediate subgroups), **Figures 2E** and **2F**. Patients classified as sensitive had significantly better OS and PFS than patients classified as not sensitive (HR=0.45; 95% CI: 0.25-0.79, $P=0.006$ for OS and HR=0.45; 95% CI: 0.27-0.76, $P=0.003$ for PFS). Median OS was 17.3 (95% CI: 8.5-undefined) months for the sensitive group, compared with 6.0 (95% CI: 4.3-9.2) months for the not sensitive group; median PFS was 9.1 (95% CI: 2.5-undefined) months for the sensitive group, compared with only 1.8 (95% CI: 1.4-2.7) months for the not sensitive group. In multivariate analyses, while the effect sizes for OS and PFS remained substantial (HR=0.60 and 0.63, respectively), classification sensitive vs not sensitive did not retain its independent significance as a predictive factor (Supplemental Data: Results Supplementary Tables 13 and 14).

Baseline patient characteristics showed no association with test classification for $P<0.05$ (Supplemental Data: Results Supplementary Table 15). In particular, PD-L1 expression was not significantly correlated with test classification ($P=0.387$ for ternary classification vs. PD-L1[+]/PD-L1[-]/NA).
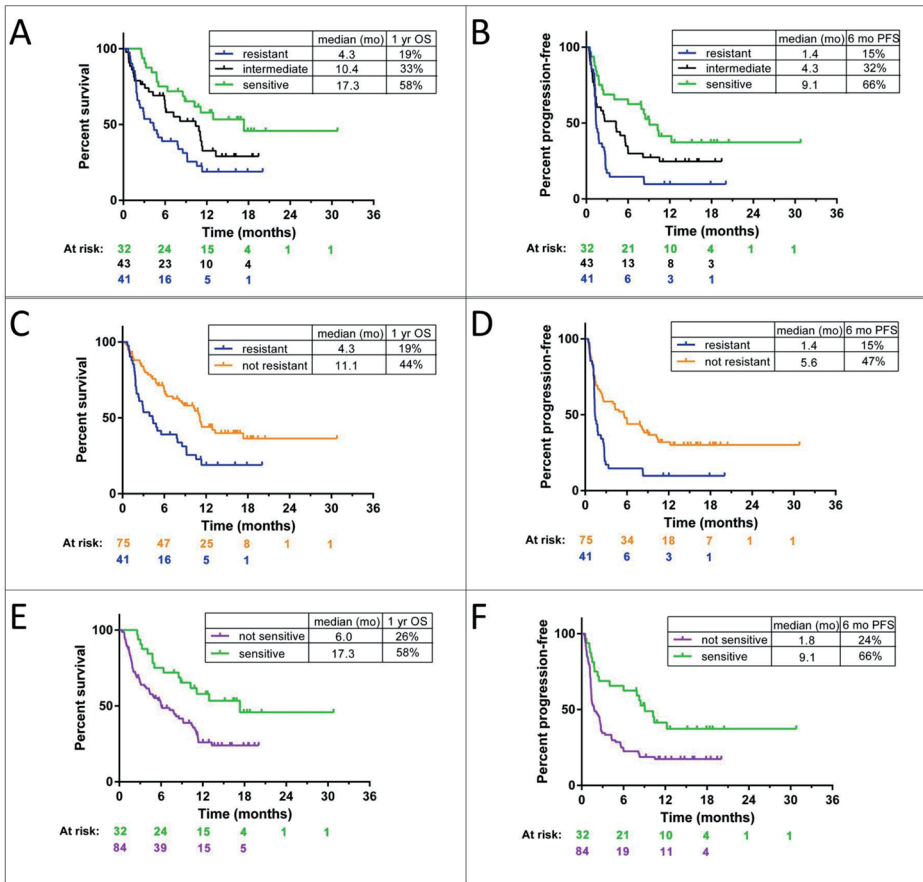
Figure 2. A–F, Outcomes by test classification for the development cohort.

Validation

The locked test was applied to samples from validation sets 1 and 2 and the chemotherapy cohort. Validation set 1 had been used in a previous investigation[26] and therefore, while it was not used in test development, validation set 1 could not be run blinded to clinical data. The chemotherapy cohort was a subset of a previously analyzed clinical trial comparing chemotherapy and targeted therapy and hence could also not be tested blinded to clinical data. Testing of validation set 2 was completely blinded to all clinical data. Statistical consideration of power to detect the effect sizes observed in the development cohort for each validation set and the chemotherapy cohort is outlined in the supplement.

Within the validation sets, the number and proportions of patients assigned to each classification group were 37 (38%)/32 (43%) resistant, 30 (31%)/19 (25%) intermediate and 31 (32%)/24 (32%) sensitive for set 1/set 2 respectively. Kaplan-Meier plots of OS by test classification resistant vs. not resistant and sensitive vs. not sensitive are shown for the validation sets in **Figure 3A-D**. In validation set 1, **Figures 3A** and **3B**, patients classified as resistant had significantly worse OS than not resistant patients (HR=0.60; 95% CI: 0.37- 0.97, *P*=0.037) and patients classified as sensitive had significantly better OS than not sensitive patients (HR=0.56; 95% CI: 0.33-0.97, *P*=0.038). One year survival for the sensitive group was 65% and the corresponding median was 15.3 (95% CI: 8.8-undefined) months. In contrast, median OS was only 4.8 (95% CI: 2.9-9.3) months in the resistant group, with 29% OS at one year. PFS was numerically superior in the sensitive group and inferior in the resistant group, but the differences in outcome were smaller and did not reach statistical significance, see Supplemental Data: Results Supplementary Figures 1 and 2. Analysis of the subgroup of patients treated with nivolumab in third or higher line (n=40), showed similar behavior in OS and PFS, with resistant patients showing a trend to shorter outcomes (HR=0.49; 95% CI: 0.23-1.04, *P*=0.062 for OS and HR=0.50; 95% CI: 0.25-1.02, *P*=0.057 for PFS) and sensitive patients showing numerically longer survival (HR=0.48; 95%CI: 0.21-1.10, *P*=0.082 for OS and HR=0.62; 95% CI: 0.31-1.23, *P*=0.172 for PFS). Kaplan-Meier plots for this subgroup are shown in the Supplementary data.

Results for validation set 2 are shown in **Figures 3C** and **3D**. Patients classified as resistant had worse OS than not resistant patients (HR=0.39; 95% CI: 0.19-0.77, *P*=0.007). The comparison of OS between the sensitive group and the not sensitive patients yielded a HR of 0.58, but did not show a significant difference (*P*=0.179). However, for ternary test classifications, the sensitive group had longer OS than the resistant group (HR=0.41; 95% CI: 0.18-0.94, *P*=0.036). Full analysis for the sensitive/intermediate/ resistant classifications can be found in Supplemental Data: Results. Analysis of PFS showed only numerical differences between classification groups.

As results were consistent across cohorts, within the limits of relatively small subgroup sizes, a pooled analysis of all patients treated in second line with nivolumab was carried out stratified by cohort (n=249). There was no indication of any correlation of PD-L1 expression with test classification (*P*=0.292, 0.810, 0.337 for ternary, resistant vs. not resistant, sensitive vs. not sensitive test classifications), although positive PD-L1 expression was a predictor of improved OS and PFS in the pooled analysis (HR=1.60; 1.01-2.54, *P*=0.046 for OS and HR=1.61; 1.07-2.44, *P*=0.023 for PFS). Indeed, analysis including test classification and PD-L1 expression demonstrated both to be independent predictors of PFS (see Supplementary data). Within the

pooled second-line population, multivariate analysis showed that the resistant vs not resistant stratification was a significant independent predictor of OS ($P<0.001$) and PFS ($P=0.006$) when adjusted for multiple baseline factors, **Table 2**. The sensitive vs not sensitive stratification was a significant independent predictor of OS ($P=0.009$) and showed a trend to prediction of PFS ($P=0.079$).

Figures 3E and **3F** show OS for classification groups obtained by applying the test to pretreatment samples of the chemotherapy cohort, in which patients received docetaxel as second-line therapy. There was no indication that the test was able to stratify patients by outcome following this single agent chemotherapy ($P=0.471$ and $P=0.165$ for OS comparison of resistant vs not resistant and sensitive vs not sensitive, respectively).

Protein set enrichment
To examine the potential biological mechanisms underlying the test, the association of test classification with various biological processes was assessed using PSEA methods[22-24]. The results are summarized in **Table 3**. Acute phase response, acute inflammatory response, wound healing, and complement activation were identified as associated with test classification with $P<0.001$. Additionally, innate immune response and chronic inflammatory response were identified as associated with $P<0.01$. Similar analysis was performed comparing the sensitive subgroup with the remaining patients. Only immune tolerance and suppression were identified as associated with test classification with $P<0.01$ (FDR<0.1). Full results for sensitive vs not sensitive phenotype are contained in the Supplemental Data: Results Supplementary Table 21.
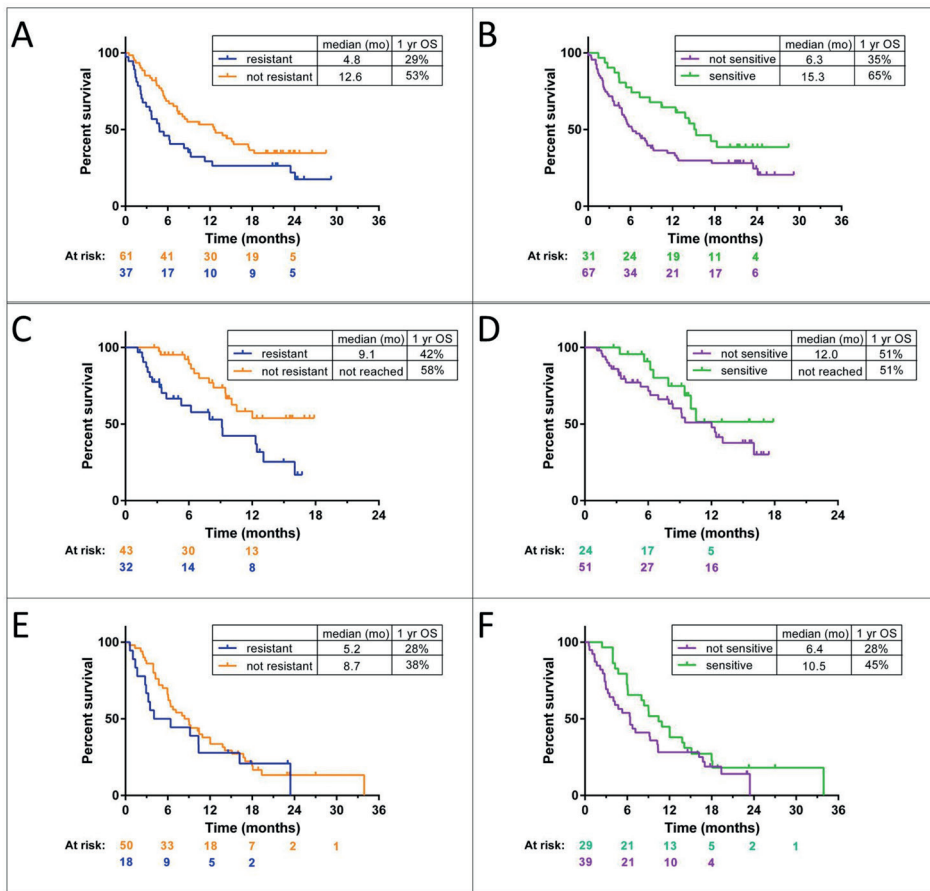
Figure 3. Kaplan–Meier plots of OS for the validation sets and the chemotherapy cohort. **A**, Validation set 1 resistant versus not resistant. **B**, Validation set 1 sensitive versus not sensitive. **C**, Validation set 2 resistant versus not resistant. **D**, Validation set 2 sensitive versus not sensitive. **E**, Chemotherapy cohort sensitive versus not sensitive. **F**, Chemotherapy cohort resistant versus not resistant.

Table 2. Multivariate analysis of OS and PFS stratified by cohort for the pooled second-line population for test classification resistant versus not resistant (analysis 1) and test classification sensitive versus not sensitive (analysis 2).

		OS		PFS	
Analysis 1		**HR (95% CI)**	***P***	**HR (95% CI)**	***P***
Test classification (vs. resistant)	Not resistant	0.52 (0.37–0.74)	<0.001	0.64 (0.47–0.88)	0.006
Histology (vs. adeno)	Squamous	0.83 (0.54–1.28)	0.403	1.10 (0.75–1.60)	0.639
	NA/other	1.10 (0.66–1.85)	0.711	1.09 (0.69–1.70)	0.718
Age (vs. ≥65)	<65	1.14 (0.80–1.63)	0.455	1.27 (0.93–1.73)	0.130
Gender (vs. male)	Female	0.52 (0.35–0.76)	0.001	0.69 (0.50–0.96)	0.027
PS (vs. 0)	1	1.56 (1.02–2.39)	0.040	1.37 (0.96–1.97)	0.084
	2+	3.66 (2.00–6.67)	<0.001	2.30 (1.31–4.06)	0.004
	NA	2.29 (1.15–4.54)	0.018	1.90 (1.05–3.46)	0.035
Smoking (vs. ever)	Never	1.87 (0.96–3.64)	0.064	1.47 (0.81–2.67)	0.209
	NA	0.76 (0.30–1.92)	0.559	0.76 (0.33–1.76)	0.521
PD-L1 (vs. positive)	Negative	1.20 (0.74–1.94)	0.461	1.31 (0.85–2.03)	0.218
	NA	0.84 (0.52–1.36)	0.474	0.86 (0.57–1.30)	0.476
Analysis 2		**HR (95% CI)**	***P***	**HR (95% CI)**	***P***
Test classification (vs. not sensitive)	Sensitive	0.58 (0.38–0.87)	0.009	0.73 (0.51–1.04)	0.079
Histology (vs. adeno)	Squamous	0.84 (0.54–1.30)	0.428	1.12 (0.76–1.63)	0.573
	NA/other	1.13 (0.68–1.87)	0.648	1.10 (0.70–1.71)	0.683
Age (vs. ≥65)	<65	1.06 (0.75–1.51)	0.750	1.21 (0.89–1.65)	0.227
Gender (vs. male)	Female	0.49 (0.33–0.72)	<0.001	0.66 (0.47–0.91)	0.011
PS (vs. 0)	1	1.41 (0.92–2.17)	0.116	1.32 (0.92–1.90)	0.136
	2+	3.31 (1.78–6.13)	<0.001	2.19 (1.22–3.91)	0.008
	NA	2.25 (1.14–4.45)	0.020	1.95 (1.07–3.55)	0.028
Smoking (vs. ever)	Never	1.82 (0.93–3.57)	0.082	1.48 (0.81–2.71)	0.205
	NA	0.83 (0.33–2.11)	0.693	0.84 (0.36–1.95)	0.676
PD-L1 (vs. positive)	Negative	1.22 (0.75–1.98)	0.417	1.34 (0.87–2.07)	0.189
	NA	0.97 (0.60–1.55)	0.882	0.94 (0.62–1.41)	0.755

Abbreviations: Adeno, adenocarcinoma; NA, not available; PS, performance status.

5

Table 3. Results of PSEA of test classifications resistant versus not resistant.

Biological process	P value of association	FDR
Acute phase response	<0.0001	<0.002
Acute inflammatory response	0.0001	<0.002
Wound healing	0.0002	<0.002
Complement activation	0.0005	<0.003
Innate immune response	0.0014	<0.01
Chronic inflammatory response	0.0044	<0.02
Extra cellular matrix	0.0231	<0.08
IFN type 1	0.0315	<0.1
Cellular component morphogenesis	0.0317	<0.1
Immune tolerance and suppression	0.0526	<0.2
B-cell-mediated immunity	0.0526	<0.2
Angiogenesis	0.0753	<0.2
NK-cell-mediated immunity	0.1222	<0.3
Behavior	0.1270	<0.3
Cytokine production involved in immune response	0.3198	<0.5
Glycolysis and positive regulators	0.3560	<0.6
Epithelial-mesenchymal transition	0.4548	<0.6
Type 17 immune response	0.4668	<0.6
Type 1 immune response	0.5102	<0.7
Type 2 immune response	0.7791	<0.9
Response to hypoxia	0.9287	<1
T-cell-mediated immunity	0.9861	<1
IFNγ	0.9884	<1

Abbreviation: NK, natural killer.

Discussion

Here, we report the establishment of a pretreatment serum proteomic classifier that separates those patients who obtain little from those that obtain durable clinical benefit from treatment with the PD-1 inhibitor nivolumab as second-line treatment for advanced NSCLC. Based on 274 MS features, patients could be classified as being resistant, intermediate or sensitive. The difference in OS between resistant and not resistant patients was highly significant: the HR was 0.48, and median survival times were 4.3 months vs 11.1 months, respectively. The test was validated while blinded to clinical outcome data with an independent set of advanced NSCLC patients, treated at a different institution. The classifier failed to stratify outcomes within a historical cohort of advanced NSCLC patients treated with docetaxel as second-line therapy. Moreover, test classification, as expected, was independent of well-established clinical factors and notably showed no evidence of association with PD-L1 expression.

A serum test would have obvious advantages, such as ease of detection using one blood draw. Also, the test may avoid the issue of intra patient tumor heterogeneity and could assess host factors that are not captured by examination of the tumor microenvironment in histological samples. Further characterization of the classifier revealed that the classification phenotypes identified are associated with biological processes known to confer a poor prognosis in lung cancer. Several lines of research indicate that complement, as a member of a diverse family of innate immune proteins, is involved in dysregulation of mitogenic signaling and escape from immune surveillance[27,28]. Complement activation, as measured by Cd4, a stable complement degradation product, in serum of early-stage NSCLC patients was significantly associated with poor prognosis[29]. A number of authors have identified the ratio of the acute phase protein, serum C-Reactive Protein, to albumin as a negative prognostic factor in both early and advanced NSCLC[30]. Intratumoral wound healing signatures, as measured by mRNA expression arrays, are considered to be T-cell suppressive and have been observed in several tumor types, amongst them NSCLC[31]. Interestingly, sera derived from patients with tumors exhibiting wound healing signatures elicited identical signatures from non-tumor associated fibroblasts, which were found to be a powerful predictor of an unfavorable clinical course[32]. These observations may provide the biological basis of our findings, although a direct link between the abundance of these circulating proteins and absence of a response to PD-1 inhibitors remains to be established.

5

The results obtained in this study do not stand alone. Weber and colleagues identified a protein classifier from sera of melanoma patients treated with PD-1 inhibitors, employing the same technology that was used in our study. This was validated in multiple patient cohorts treated with PD-1 inhibitors and CTLA4 antagonists[14]. As here, they were able to identify, prior to initiation of treatment, patients who had a favorable outcome following treatment. Biological processes associated with that classifier included complement, wound healing and acute phase pathways, all upregulated in the poor prognosis group, corroborating our results. Further evidence that the pretreatment circulating proteome provides important information on checkpoint efficacy was provided in the context of a phase II study where atezolizumab was compared with docetaxel as second-line treatment in 272 advanced NSCLC patients[33]. Similar to our results, a serum protein classifier was established that identified patients with poor (median OS 7.3 months, n=60 (45%)) and good (median OS not reached, n= 73 (55%)) outcomes. This classifier was shown in blinded validation to be predictive for atezolizumab vs. docetaxel for OS and PFS (interaction $P<$0.01). In that study, as in our own, there was no association between test classification and tumor PD-L1 expression; there was also no association with TMB. Also, among the biological processes that were most significantly associated with classification by PSEA, acute inflammation and complement activation ranked in the top three.

There are some limitations to our results. Obviously, the number of patients is low and all three immunotherapy-treated cohorts come from one geographic area and were investigated retrospectively. Also, for historical reasons, validation blinded to all clinical data was only possible for validation set 2. Although we made strong efforts to obtain sufficient tumor tissue samples, we were not able to obtain PD-L1 expression data on all patients. Several factors contributed to this: many patients are diagnosed on the basis of cytology alone and so have no tissue available for PD-L1 analysis; at the time of treatment initiation for these patients, use of PD-L1 expression was still somewhat investigational; and positive PD-L1 expression status is not mandatory for administration of nivolumab in the second and higher line setting. Unfortunately, TMB data was not collected. Investigation of larger cohorts with more complete information on TMB and PD-L1 expression would be useful to examine with more precision the level of association of these markers and how much complementary information each can provide to predict outcome. The non-immunotherapy-treated control set is small and restricted to one therapy. It would be of interest to study the performance of the test in larger control cohorts in other standard-of-care non-immunotherapy regimens to be able to explore the test's predictive potential.

Additional validation of the test in other larger cohorts of patients treated with CPIs is necessary. So far, we have investigated the ability of the test to stratify outcome for patients receiving checkpoint blockade monotherapy in the second- and higher-line setting, after platinum-based chemotherapy. However, now immunotherapy is moving into the first-line setting, either as monotherapy for patients with PD-L1 expression greater than 50%, or in combination with chemotherapy. It is of interest to evaluate the performance of the test in these first-line settings. A prospective trial, comparing outcomes between mono- immunotherapy and the chemo-immunotherapy combination in front-line patients with high PD-L1 expression is in the final stages of design. Studies in earlier stage patients receiving durvalumab with chemoradiation would also be informative. Evaluation of the test with appropriate comparator non-immunotherapy regimens in a prospective, randomized setting would be required to unequivocally determine its predictive power and clinical utility.

5

References

1. Goldstraw, P. *et al.* The IASLC Lung Cancer Staging Project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. *J Thor Oncol.* 2016;**11**:39-51.
2. Peters, S., Reck, M., Smit, E.F., Mok, T.S., Hellman, M.D. How to make best use of immunotherapy as first-line treatment for advanced/metastatic non-small cell lung cancer? *Ann Oncol.* 2019; March 26.
3. Gettinger, S, *et al.* Five year follow up of nivolumab in previously treated advanced non-small-cell lung cancer: results from the CA209-003 study. *J Clin Oncol.* 2018;**36**:1675-1684.
4. Garon, E.B. *et al.* Five-year overall survival for patients with advanced non-small-cell lung cancer treated with pembrolizumab: results from the Keynote 001 study. *J Clin Oncol.* 2019; 37(**28**):2518-2527.
5. Garon, E.B. *et al.* Pembrolizumab for the treatment of non-small-cell lung cancer. *New Eng J Med.* 2015;**372**:2018-2028.
6. Sun, R. *et al.* A radiomics approach to assess tumor-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol.* 2018;**19**;1180-1191.
7. Hellmann, M. *et al.* Genomic features of response to combination immunotherapy in patients with advanced non-small-cell lung cancer. *Cancer Cell* 2018;**33**:843-852.
8. Duruisseaux, M. *et al.* Epigenetic prediction of response to anti-PD-1 treatment in non-small-cell lung cancer: a multicenter, retrospective analysis. *Lancet Respir Med.* 2018;**6**:771-781.
9. Cristescu, R. *et al.* Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science.* 2018;**362**:197.
10. Brahmer, J. *et al.* Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N Engl J Med.* 2015; 373(**20**):123-35.
11. Borghaei, H. *et al.* Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N Engl J Med.* 2015;373(**17**):1627-39.
12. Gregorc, V. *et al.* Predictive value of a proteomic signature in patients with non-small-cell lung cancer treated with second-line erlotinib or chemotherapy (PROSE): a biomarker-stratified, randomized phase 3 trial. *Lancet Oncology.* 2014;**15**:713-721.
13. Tsypin, M., Asmellash, S., Meyer, K., Touchet, B., Roder, H. Extending the information content of the MALDI analysis of biological fluids via multi-million shot analysis. *PLoS ONE* 14(**12**): e0226012.
14. Weber, J.S. *et al.* A Serum Protein Signature Associated with Outcome after Anti-PD-1 Therapy in Metastatic Melanoma. *Cancer Immunol Res.* 2018;**6**:79-86.
15. Ascierto, P.A. *et al.* Proteomic test for anti-PD-1 checkpoint blockade treatment of metastatic melanoma with and without BRAF mutations. *J Immunother Cancer.* 2019;**7**:91.
16. Roder, J. *et al.* A dropout-regularized classifier development approach optimized for precision medicine test discovery from omics data. *BMC Bioinformatics.* 2019;**20**:325.
17. Goodfellow, I., Bengio, Y., Courville, A. Deep Learning. *Cambridge: MIT Press;* 2016. p.5-11.
18. Breiman, L. Out-of-bag estimation. *Technical Report.* Department of Statistics, University of California, 1996.

19. Roder, H., Oliveira, C., Net, L., Linstid, B., Tsypin, M., Roder, J. Robust Identification of Molecular Phenotypes using Semi-Supervised Learning. *BMC Bioinformatics*. 2019;**20**:273.

20. Weber, J. *et al.* Pre-treatment patient selection for nivolumab benefit based on serum mass spectra. *J Immunother Cancer*. 2015:**3**(Suppl 2):P103.

21. Grossi, F. *et al.* Evaluation of Pretreatment Serum Tests for Nivolumab Benefit in Patients with Non-Small Cell Lung Cancer. *J Thorac Oncol*. 2017;12(**1**):S1322.

22. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;**102**:15545-50.

23. Grigorieva, J., Asmellash, S., Oliveira, C., Roder, H., Net, L., Roder, J. Application of protein set enrichment analysis to correlation of protein functional sets with mass spectral features and multivariate proteomic tests. *Clin Mass Spectrometry*. 2020; **15**:44-53.

24. Roder, J., Linstid, B., Oliveira, C. Improving the power of gene set enrichment analyses. *BMC Bioinformatics* 2019; **20**:257.

25. Benjamini, Y., Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J Royal Stat Soc. Series B* (Methods). 1995;**57**:289-300.

26. Goldberg, S.B. *et al.* Mass spectrometry-based test predicts outcome on anti-PD-1 therapy for patients with advanced non-small cell lung cancer with brain metastases. *J Immunother Cancer*. 2017;**5**(Suppl 2):86.

27. Rutkowski, M.J., Sughrue, M.E., Kane, A.J., Mills, S.A., Parsa, A.T. Cancer and the complement cascade. *Mol Cancer Res*. 2010;8(**11**):1453-65.

28. Wang, Y. *et al.* Autocrine complement inhibits IL10 dependent T-cell-mediated antitumor immunity to promote tumor progression. *Cancer Disc*. 2016:**6**;1022-1035.

29. Ajona, D. *et al.* Investigation of complement activation product C4D as a diagnostic and prognostic biomarker for lung cancer. *J Natl Cancer Inst*. 2013;105(**18**):1385-93.

30. McMillan, D.C. The systemic inflammation-based Glasgow score: a decade of experience in patients with cancer. *Cancer Treat Rev*. 2013;**39**:534-40.

31. Hugo, W. *et al.* Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell*. 2016;**165**:35-44.

32. Chang, H.Y. *et al.* Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biology*. 2004;**2**:0206.

33. Kowanetz, M. *et al.* Evaluation of immune-related markers in the circulating proteome and their assocation with atezolizumab efficacy in patients with 2L+ NSCLC. *J Immunother Cancer* 2018;**6**(Suppl 1):114.

5

Supplemental material
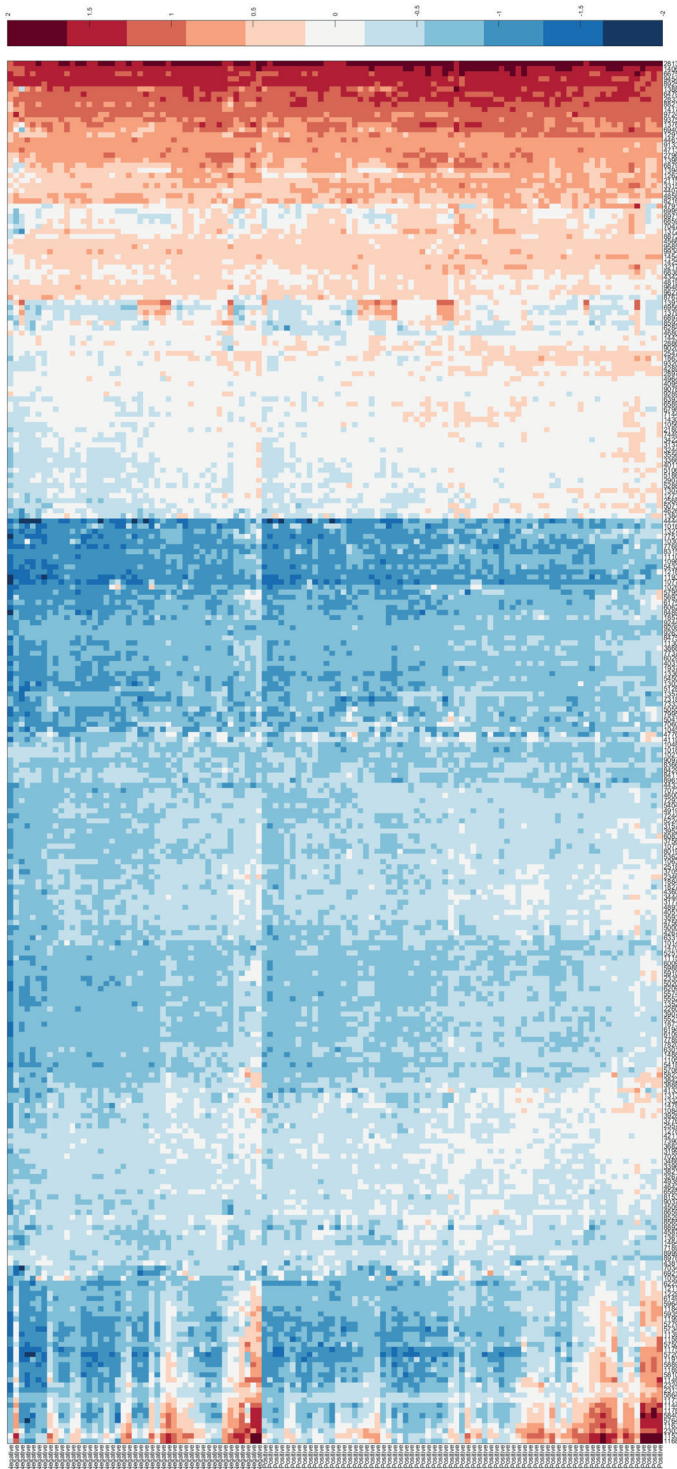## Supplementary results

1. Development of the test
   The following pages provide heatmaps for samples in the development cohort of the features used for each of classifier A, classifier B, classifier C, and for the test. Samples are sorted into the classification groups of each classifier or the test. The mass spectral features are labelled on the *x*-axis. The heatmap plots $\log_{10}$ of the feature values, with the red/brown for the largest features and dark blue for the smallest features.

2. Reproducibility of Test Classifications
   To assess the reproducibility of the test, the test was run from scratch on samples from validation set 1 on two occasions more than one year apart. The classifications obtained for run1 and run2 are compared in **Supplementary Table S9** for classification sensitive vs. intermediate vs. resistant, in **Supplementary Table S10** for the binary combination resistant vs. not resistant (intermediate and sensitive), and in **Supplementary Table S11** for the binary combination not sensitive (intermediate and resistant) vs. sensitive. Classification concordance is 85% for the three-way classifications, 91% for the resistant / not resistant combination and 93% for the not sensitive / sensitive combination.
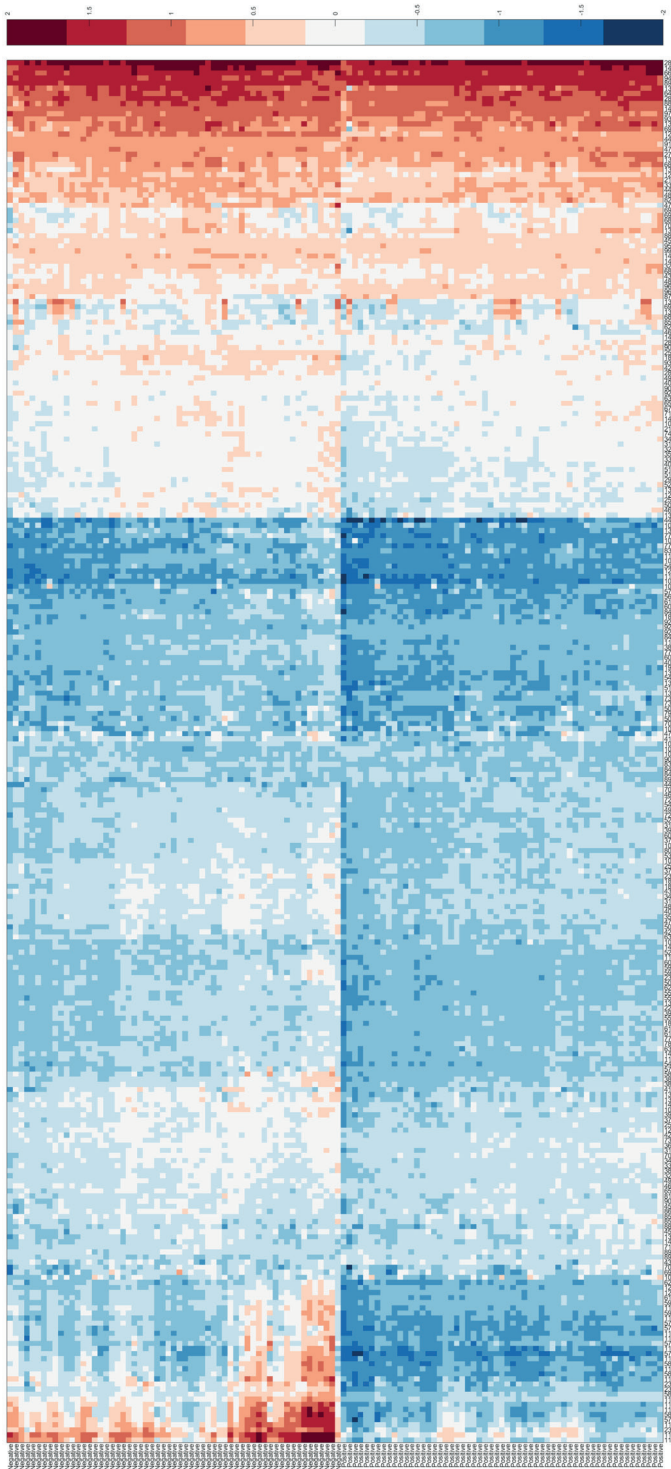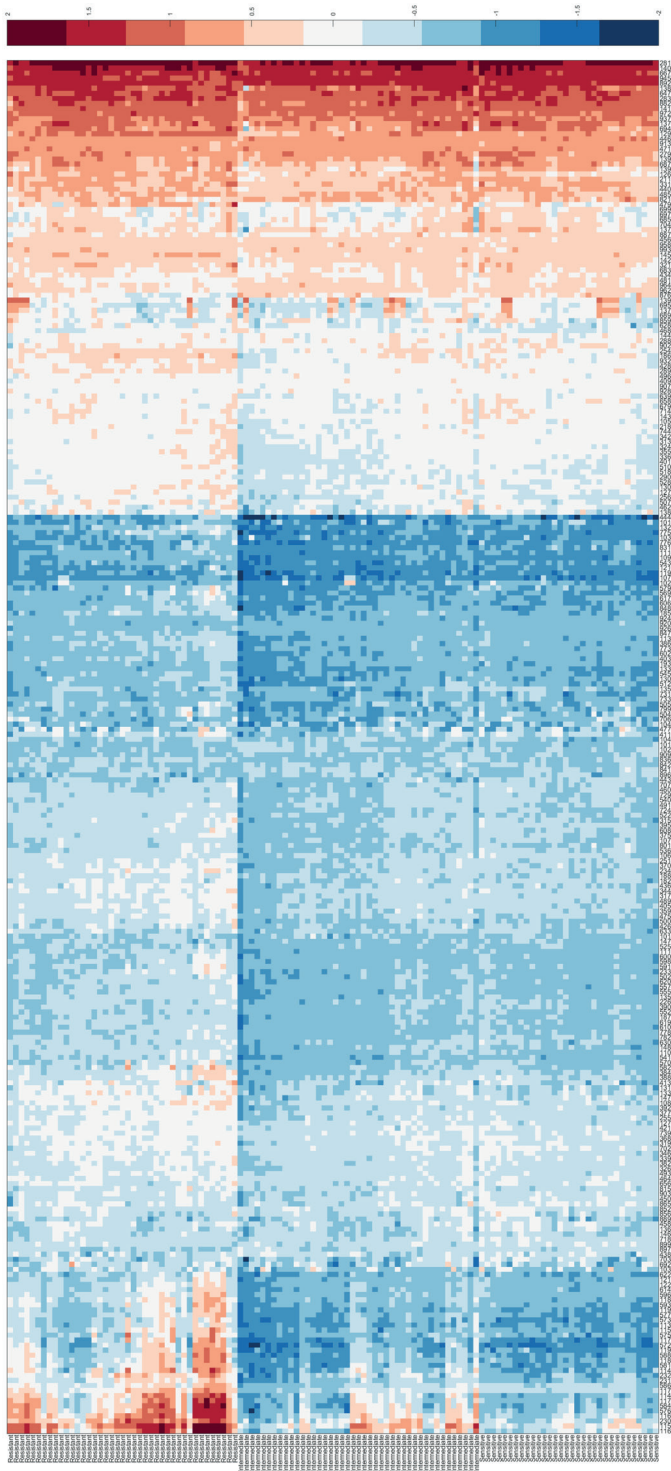
Supplementary Figure S1: Heatmap for the samples in the development cohort of 274 mass spectral features used for classifier A (BDX008 redevelopment).

Supplementary Figure S2: Heatmap for the samples in the development cohort of 274 mass spectral features used for classifier B.
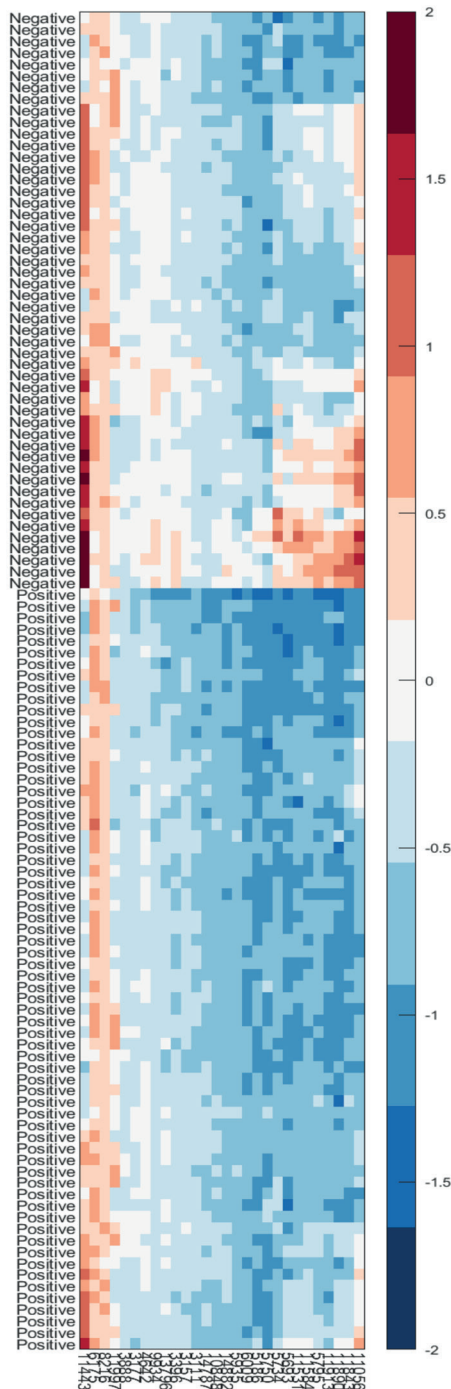
Supplementary Figure S3: Heatmap for the samples in the development cohort of the 29 mass spectral features (associated with Immune Response type 2) used for classifier C.

Supplementary Figure S4: Heatmap for the samples in the development cohort of 274 mass spectral features used for the test.

Supplementary Table S9: Classification concordance (sensitive vs. intermediate vs. resistant).

		Run2		
		Resistant (n=37)	Intermediate (n=30)	Sensitive (n=31)
Run1	Resistant (n=40)	34	6	0
	Intermediate (n=22)	2	19	1
	Sensitive (n=36)	1	5	30

Supplementary Table S10: Classification concordance (resistant vs. not resistant).

		Run2	
		Resistant (n=37)	Not resistant (n=61)
Run1	Resistant (n=40)	34	6
	Not resistant (n=58)	3	55

Supplementary Table S11: Classification concordance (sensitive vs. not sensitive).

		Run2	
		Not sensitive (n=67)	Sensitive (n=31)
Run1	Not sensitive (n=62)	61	1
	Sensitive (n=36)	6	30

3. Association of response with test classification

Supplementary Table S12: Response by test classification in the development cohort.

		Resistant (*n=41*)	Intermediate (*n=43*)	Sensitive (*n=32*)
		n (%)	n (%)	n (%)
Response	CR	0 (0)	0 (0)	1 (3)
	PR	4 (10)	4 (9)	8 (25)
	SD	1 (2)	11 (26)	7 (22)
	PD	35 (85)	21 (49)	9 (28)
	NA	1 (2)	7 (16)	7 (22)

4. Multivariate analysis of the development cohort

Supplementary Table S13: Multivariate analysis of OS and PFS for the development cohort by test classification resistant vs. not resistant.

	OS		PFS	
	HR (95% CI)	*P* value	HR (95% CI)	*P* value
Test classification (not resistant vs. resistant)	0.59 (0.34-1.03)	0.062	0.53 (0.32-0.89)	0.015
ECOG PS (1 vs. 0)	1.71 (0.90-3.22)	0.100	1.36 (0.78-2.35)	0.277
ECOG PS (≥2 vs. 0)	4.67 (2.05-10.66)	<0.001	2.50 (1.19-5.25)	0.016
Never vs. ever smoker	1.88 (0.84-4.23)	0.126	1.20 (0.54-2.65)	0.657
Squamous vs. non-squamous	1.02 (0.56-1.84)	0.960	1.04 (0.62-1.76)	0.876
PD-L1 (<1% vs. ≥1%)	1.53 (0.79-2.95)	0.205	1.40 (0.76-2.58)	0.285
PD-L1 (NA vs. ≥1%)	0.85 (0.41-1.77)	0.669	0.86 (0.45-1.65)	0.655

Supplementary Table S14: Multivariate Analysis of OS and PFS for the development cohort by test classification sensitive vs. not sensitive.

	OS		PFS	
	HR (95% CI)	*P* value	HR (95% CI)	*P* value
Test classification (sensitive vs. not sensitive)	0.60 (0.30-1.20)	0.150	0.63 (0.35-1.15)	0.132
ECOG PS (1 vs. 0)	1.64 (0.86-3.13)	0.133	1.39 (0.80-2.40)	0.245
ECOG PS (≥2 vs. 0)	4.80 (2.11-10.91)	<0.001	2.68 (1.28-5.57)	0.009
Never vs. ever smoker	2.14 (0.92-4.95)	0.077	1.31 (0.58-2.95)	0.512
Squamous vs. non-squamous	0.99 (0.55-1.81)	0.988	1.05 (0.62-1.78)	0.856
PD-L1 (<1% vs. ≥1%)	1.65 (0.86-3.14)	0.130	1.66 (0.92-2.99)	0.091
PD-L1 (NA vs. ≥1%)	1.02 (0.50-2.07)	0.958	1.08 (0.57-2.04)	0.811

5. Patient characteristics by test classification

Supplementary Table S15: Patient characteristics by test classification (development cohort).

		Resistant (n=41)	Intermediate (n=43)	Sensitive (n=32)	*P* value
		n (%)	n (%)	n (%)	
Gender	Male	25 (61)	22 (51)	19 (59)	0.649
	Female	16 (39)	21 (49)	13 (41)	
Age	Median (range)	68 (52-83)	63 (43-80)	64 (49-82)	--
Histology	Two primary tumors	1 (2)	1 (2)	0 (0)	0.474
	Adenocarcinoma	25 (61)	27 (63)	25 (78)	
	NSCLC-NEC	0 (0)	2 (5)	1 (3)	
	NSCLC-NOS	2 (5)	4 (9)	2 (6)	
	Squamous	13 (32)	9 (21)	4 (13)	
Performance Status (WHO)	0	8 (20)	13 (30)	15 (47)	0.112
	1	22 (54)	24 (56)	14 (44)	
	2	8 (20)	3 (7)	1 (3)	
	3	2 (5)	1 (2)	0 (0)	
	NA	1 (2)	2 (5)	2 (6)	
Smoking Status	Current	10 (24)	6 (14)	7 (22)	0.431
	Former	26 (63)	35 (81)	20 (63)	
	Never	4 (10)	2 (5)	4 (13)	
	NA	1 (2)	0 (0)	1 (3)	
Brain metastases at start of therapy	No	31 (76)	32 (74)	24 (75)	>0.999
	Yes	10 (24)	11 (26)	8 (25)	
Previous radiotherapy	No	13 (32)	15 (35)	15 (47)	0.389
	Yes	28 (68)	28 (65)	17 (53)	
Previous thoracic radiotherapy	No	30 (73)	24 (56)	25 (78)	0.087
	Yes	11 (27)	19 (44)	7 (22)	
PD-L1 expression	Positive (≥1%)	10 (24)	13 (30)	10 (31)	0.387
	Negative (<1%)	16 (39)	19 (44)	8 (25)	
	NA	15 (37)	11 (26)	14 (44)	

6. Power considerations for validation of the test on validation set 1, validation set 2 and the chemotherapy cohort

For each sample set, the power to detect the effect sizes between test classification groups observed in the development cohort was estimated using the proportions of each classification observed in development and the proportions of each classification observed in the sample set upon testing. All calculations assume fully mature clinical data and $\alpha=0.05$, two-sided.

i. Validation set 1 (*n*=98)

The expected number of patients of each classification group given development set data are 35 resistant, 36 intermediate and 27 sensitive. On running the test on the samples of validation set 1, 37 samples were classified as resistant, 30 as intermediate and 31 as sensitive.

Supplementary Table S16: Power to detect effect size observed in the development set in validation set 1.

| Comparison	PFS		OS	
	Classifications predicted	Classifications assigned	Classifications predicted	Classifications assigned
Resistant vs. intermediate	60%	57%	49%	47%
Intermediate vs. sensitive	57%	57%	62%	62%
Resistant vs. sensitive	99%	>99%	99%	99%
Resistant vs. not resistant	96%	96%	94%	94%
Sensitive vs. not sensitive	94%	96%	94%	96%

ii. Validation Set 2 (*n*=75)

The expected number of patients of each classification group given the development set data are 26 resistant, 28 intermediate and 21 sensitive. On running the test on the samples of validation set 2, 32 samples were classified as resistant, 19 as intermediate and 24 as sensitive.

Supplementary Table S17: Power to detect effect size observed in the development set in validation set 2.

Comparison	PFS		OS	
	Classifications predicted	Classifications assigned	Classifications predicted	Classifications assigned
Resistant vs. intermediate	49%	45%	40%	36%
Intermediate vs. sensitive	47%	43%	52%	47%
Resistant vs. sensitive	97%	98%	96%	98%
Resistant vs. not resistant	89%	91%	86%	88%
Sensitive vs. not sensitive	87%	87%	87%	87%

iii. Chemotherapy Cohort (n=68)

The expected number of patients of each classification group given development set data are 24 resistant, 25 intermediate and 19 sensitive. On running the test on the samples of the chemotherapy cohort, 18 samples were classified as resistant, 21 as intermediate and 29 as sensitive.

Supplementary Table S18: Power to detect effect size observed in the development set in the chemotherapy cohort.

Comparison	PFS		OS	
	Classifications predicted	Classifications assigned	Classifications predicted	Classifications assigned
Resistant vs. intermediate	46%	38%	37%	30%
Intermediate vs. sensitive	43%	48%	48%	53%
Resistant vs. sensitive	95%	96%	94%	95%
Resistant vs. not resistant	86%	81%	82%	76%
Sensitive vs. not sensitive	84%	90%	84%	90%

7. Kaplan-Meier plots of OS by test classification sensitive vs. intermediate vs. resistant in the validation sets
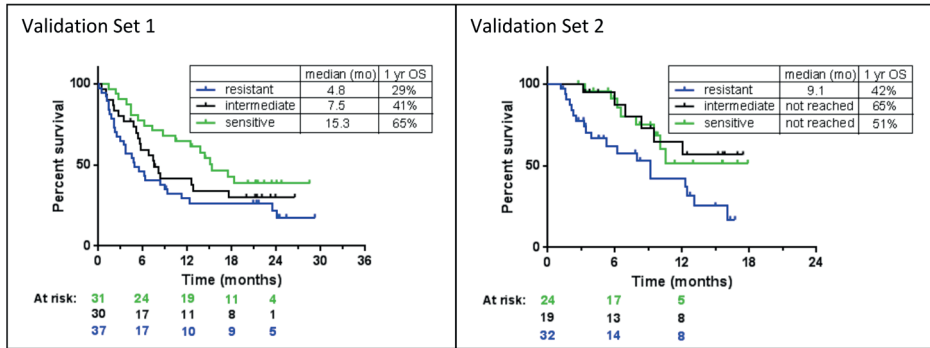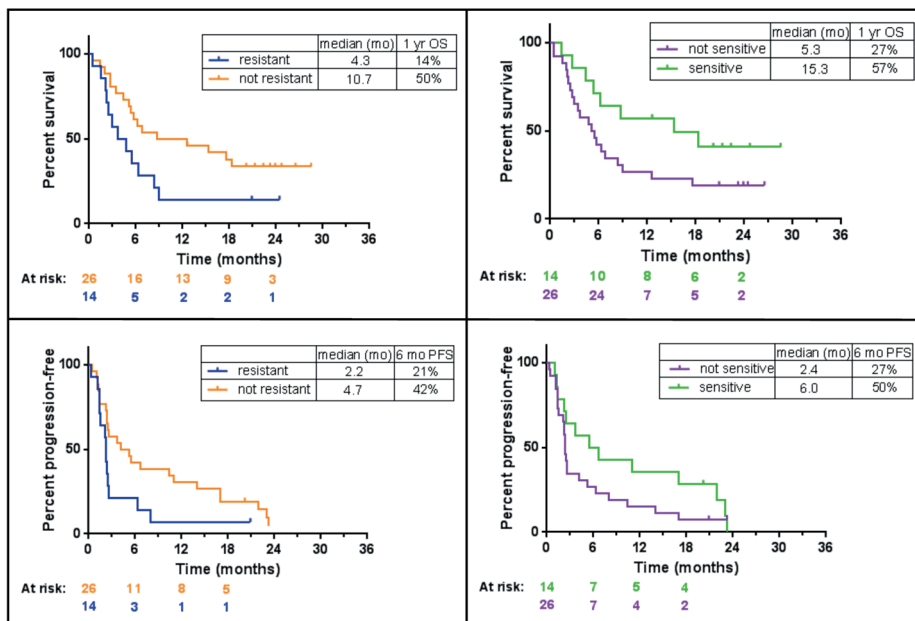

Supplementary Figure 5: Kaplan-Meier plots of OS by test classification sensitive vs. intermediate vs. resistant in the validation sets.

8. Kaplan-Meier plots of PFS for the validation sets
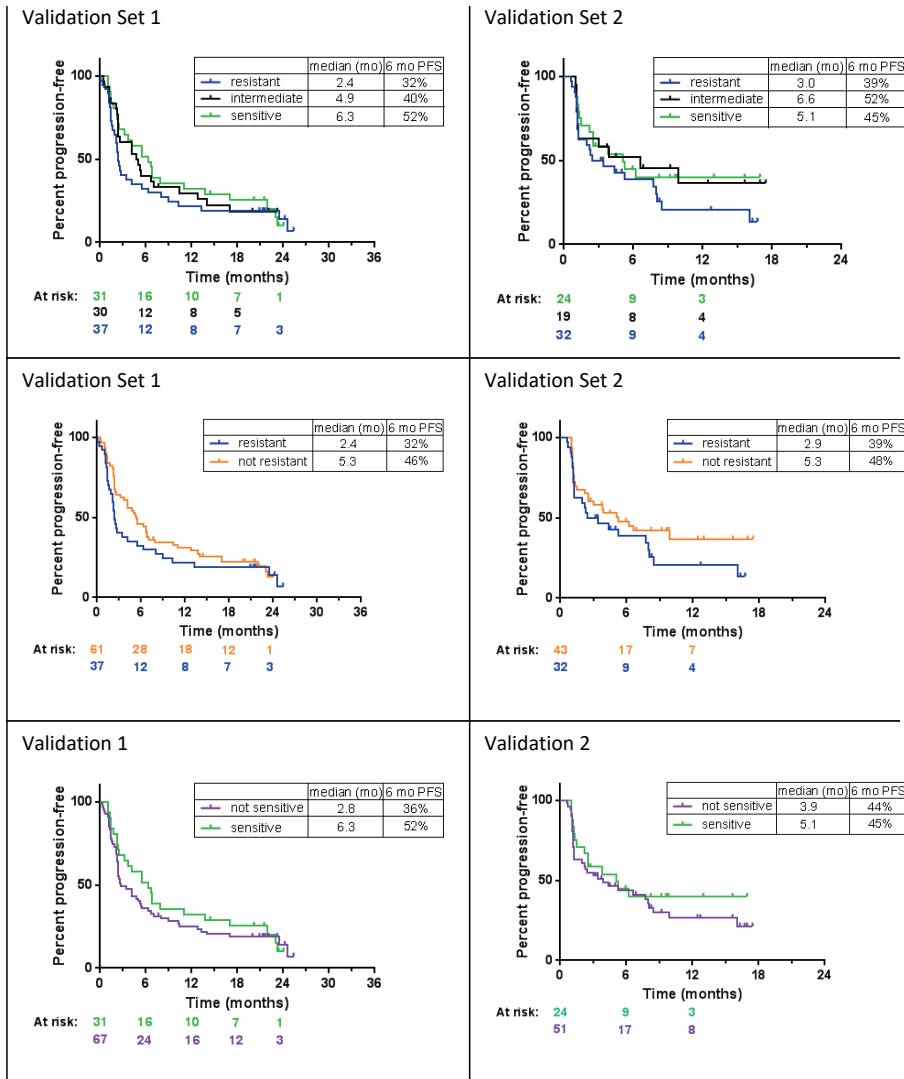
Supplementary Figure 6: Kaplan-Meier plots of PFS by test classification in the validation sets.

9. Kaplan-Meier plots of OS and PFS by test classification for the subgroup of validation set 1 treated with CPIs in third or higher line


Supplementary Figure 7: Kaplan-Meier plots of OS and PFS by test classification for patients in validation set 1 treated with CPIs in third or higher lines.

10. Association of PD-L1 status with test classification in the pooled analysis of second line patients

Supplementary Table 19: PD-L1 status by test classification (pooled second line patients)

|	Resistant (n=96)	Intermediate (n=80)	Sensitive (n=73)
PD-L1 Positive (≥1%)	21 (22)	20 (25)	16 (22)
PD-L1 Negative (< 1%)	25 (26)	28 (35)	16 (22)
NA	50 (52)	32 (40)	41 (56)

NA, not available
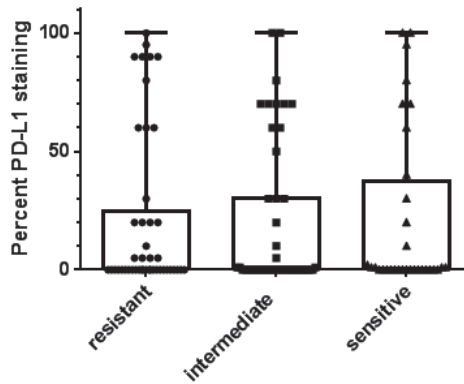
Supplementary Figure 8: Dot plot of PD-L1 staining by test classification in the pooled analysis of second line patients with known staining. Less than 1% is shown as 0. Whiskers show minimum and maximum. Boxes show the median and quartiles. Median and first quartile are both 0% for resistant and intermediate. Median and first quartile are 0% and 0.5% for sensitive.

5

11. Additional Univariate and Multivariate Analyses of the Pooled Second-Line Cohort

Supplementary Table 20: Cox proportional hazards analyses stratified by cohort with test classification and PD-L1 status as explanatory variables.

		OS		PFS	
Analysis 1		HR (95% CI)	*P* value	HR (95% CI)	*P* value
Test Classification (vs. resistant)	not resistant	0.50 (0.36-0.70)	<0.001	0.60 (0.44-0.81)	0.001
PD-L1 (vs. positive)	negative	1.51 (0.95-2.40)	0.081	1.52 (1.00-2.29)	0.049
	NA	0.96 (0.62-1.50)	0.861	0.92 (0.62-1.37)	0.681
Analysis 2		HR (95% CI)	*P* value	HR (95% CI)	*P* value
Test Classification (vs. not sensitive)	sensitive	0.54 (0.36-0.80)	0.002	0.66 (0.47-0.92)	0.015
PD-L1 (vs positive)	negative	1.53 (0.96-2.43)	0.072	1.55 (1.03-2.35)	0.037
	NA	1.09 (0.70-1.69)	0.717	1.01 (0.69-1.50)	0.951

12. Kaplan-Meier plots of OS by test classification resistant vs. intermediate vs. sensitive in the chemotherapy set
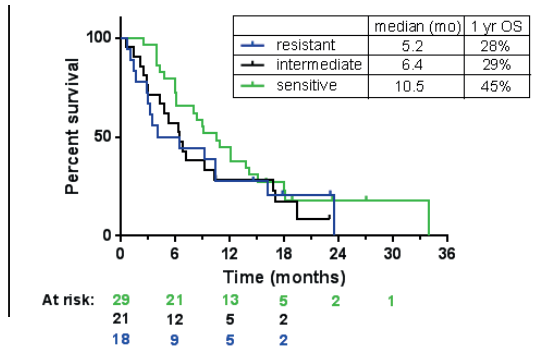
Supplementary Figure 9: Kaplan-Meier plots of OS by test classification sensitive vs. intermediate vs. resistant in the chemotherapy set.

13. Kaplan-Meier plots of PFS for the chemotherapy set
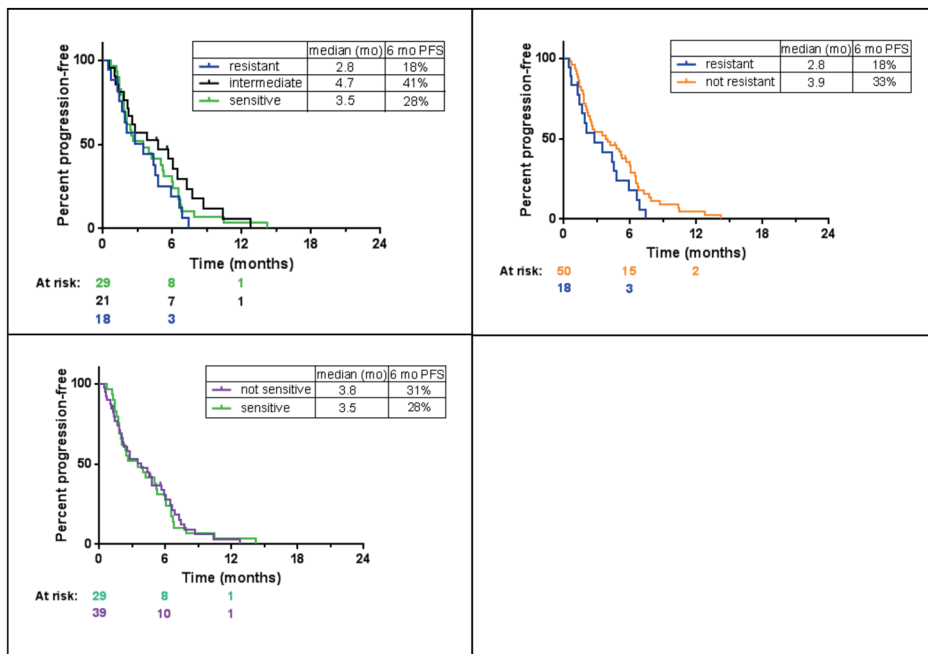
Supplementary Figure 10: Kaplan-Meier plots of OS by test classification sensitive vs. intermediate vs. resistant in the chemotherapy cohort.

14. Protein set enrichment analysis of association of biological processes with test classification sensitive vs. not sensitive

Supplementary Table 21: PSEA for association of biological processes with test classification sensitive vs. not sensitive

Biological Process	P value of association	FDR
Immune tolerance and suppression	0.0035	<0.1
Acute inflammatory response	0.0154	<0.2
Acute phase response	0.0170	<0.2
Cytokine production involved in immune response	0.0665	<0.4
Complement activation	0.1372	<0.5
Innate immune response	0.1523	<0.5
Angiogenesis	0.1532	<0.5
NK cell mediated immunity	0.1717	<0.5
B cell mediated immunity	0.2683	<0.7
Wound healing	0.2760	<0.7
Type 2 immune response	0.3887	<0.8
Extra cellular matrix	0.4302	<0.8
Epithelial-Mesenchymal Transition	0.4332	<0.8
Chronic Inflammatory response	0.5087	<0.8
IFN type 1	0.5488	<0.8
IFN-Gamma	0.5558	<0.8
Type 17 immune response	0.5576	<0.8
Response to hypoxia	0.5601	<0.8
Cellular component morphogenesis	0.6322	<0.8
T cell mediated immunity	0.6769	<0.8
Type 1 immune response	0.7802	<0.9
Glycolysis and positive regulators	0.8013	<0.9
Behavior	0.8487	<0.9

5

Chapter 6

Cell free DNA in the supernatant of pleural effusion can detect driver and resistance mutations and can guide TKI treatment decisions

Karlijn Hummelink[1,2], Mirte Muller[2], Theodora Linders[3], Vincent van der Noort[4], Petra M. Nederlof[1], Paul Baas[2], Sjaak Burgers[2], Egbert F. Smit[2], Gerrit A. Meijer[1], Michel M. van den Heuvel[2,5], Daan van den Broek[3], Kim Monkhorst[1]

[1]Department of Pathology, [2]Department of Thoracic Oncology, [3]Department of Laboratory Medicine, The Netherlands Cancer Institute – Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands, [4]Department of Biometrics, The Netherlands Cancer Institute – Antoni van Leeuwenhoek Hospital, The Netherlands, [5]Department of Pulmonary Diseases, Radboud University Medical Center, Nijmegen, The Netherlands.

Abstract

Objectives:
Molecular profiling of tumors has become the mainstay of diagnostics for metastasized solid malignancies and guides personalized treatment, especially in non-small cell lung cancer (NSCLC). In current practice, it is often challenging to obtain sufficient tumor material for reliable molecular analysis. Cell-free (cfDNA) in blood or other bio-sources could present an alternative approach to obtain genetic information from the tumor. In a retrospective cohort we analyzed the added value of cfDNA analysis in pleural effusions for molecular profiling.

Methods:
We retrospectively analyzed both the supernatant and the cell pellet of 44 pleural effusions sampled from 39 stage IV patients with *KRAS* (23) or *EGFR* (16) mutated tumors to detect the original driver mutation as well as for *EGFR* T790M resistance mutations. Patients were diagnosed with either NSCLC (n=32), colon carcinoma (n=4), appendiceal carcinoma (n=2) or adenocarcinoma of unknown primary (n=1). Samples collected in the context of routine clinical care were stored at the Netherlands Cancer Institute biobank. We used droplet digital PCR for analysis.

Results:
The driver mutation could be detected in 36 of the 44 pleural effusions by analysis of both supernatant (35 out of 44 positive) and cell pellet (31 out of 44 positive). In 7 out of 20 pleural effusions from patients with *EGFR* mutation-positive tumors, a T790M mutation was detected. All 7 supernatants and cell pellets were positive.

Conclusions:
CfDNA in pleural effusion can be used to detect driver mutations as well as resistance mechanisms like *EGFR* T790M in pleural effusion with high accuracy and is therefore a valuable bio-source.

Introduction

In current practice, molecular profiling of tumors has become essential to offer targeted therapy for several types of metastasized malignancies. Especially in non-small cell lung cancer (NSCLC), targeted therapy has been shown to be highly effective, and multiple tyrosine kinase inhibitors (TKI) for activating mutations or rearrangements in genes like *EGFR* (epidermal growth factor receptor), *BRAF* (B-Raf proto-oncogene serine/threonine kinase), *ALK* (anaplastic lymphoma kinase) and *ROS1* (Ros oncogene 1 receptor tyrosine kinase) have become available[1]. Also, early detection of resistance mutations like *EGFR* T790M is important as they can guide the next line of therapy[2].

Unfortunately, obtaining tumor tissue for molecular analysis can be challenging. The site of the tumor can be difficult to reach and it often requires invasive procedures to obtain adequate amounts of vital tumor. In this respect, pleural effusion could be an attractive alternative bio-source for molecular analysis, especially as approximately ~30% of NSCLC patients develop pleural effusion[3]. Usually, DNA is isolated from a cell block or a Giemsa slide, but analysis of cell-free (cfDNA) in the supernatant has shown promising results[4-9]. Moreover, since the amount of tumor cells or the tumor cell percentage is often insufficient for analysis, this cell-free compartment is highly interesting.

Here we explore the diagnostic yield of cfDNA analysis in pleural effusion samples from patients with NSCLC and other cancer types for *EGFR* and *KRAS* (KRAS proto-oncogene GTPase) driver mutation detection, as well as for *EGFR* resistance mutations. Analyses of cfDNA from the supernatant were compared head-to-head with analyses of the cell pellets **(figure 1)**.
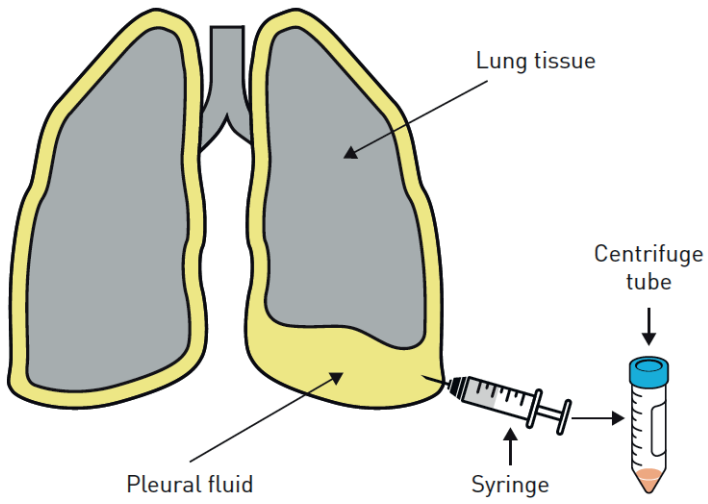
6

Figure 1. A schematic overview of the collection of pleural effusion. After pleurocentesis the fluid was centrifuged in order to separate the supernatant from the cell pellet.

Methods

Supernatant and corresponding cell pellets from 39 stage IV patients, presented at the Netherlands Cancer Institute (Amsterdam, The Netherlands) between 2009 and 2016, with *EGFR* (n=16) or *KRAS* (n=23) mutation-positive tumors were obtained from the Netherlands Cancer Institute biobank. In total, 44 paired (supernatant and cell pellet) samples were available from 39 patients diagnosed with NSCLC (n=32), colon carcinoma (n=4), appendiceal carcinoma (n=2), and adenocarcinoma of unknown primary (n=1) **(Table 1)**. The reason for pleurocentesis was either diagnostic (n=16) or therapeutic (n=28). These samples were leftover material that had been stored as routine laboratory assessment after diagnostics.

This study was approved by the Institutional Research Board of the Netherlands Cancer Institute (CFMPB497).

All driver gene and *EGFR* T790M mutations were detected in tumor tissue or cytology samples with clinically validated assays using high resolution melting, fragment analysis, Sanger sequence analysis, MassARRAY technology or next-generation sequencing (NGS) (data not shown). These analyses were done in multiple hospitals in the Netherlands.

The supernatant was separated from the cell pellet after centrifugation (1700xg for 10 min). The pellet was resuspended in 0.5 ml of 0.9% NaCl. Both samples were stored at -30°C. In total, 400µl from the cell pellet was isolated using the QIAsymphony DSP DNA Midi Kit (Qiagen, Hilden, Germany). At least 10% from the sample was analyzed, from a median eluted volume of 200µl. CfDNA was isolated from a median (range) of 1 (0.75-4) ml pleural effusion using the QIAsymphony Circulating DNA kit (Qiagen). At least 20% of the sample was analyzed, from a median eluted volume of 90µl. The Bio-Rad (Hercules, CA, USA) QX200 droplet digital PCR (ddPCR) was used for mutation detection using Bio-Rad PrimePCR ddPCR mutation assays for *KRAS* (*KRAS* Screening Multiplex 186-3506), *EGFR* T790M (dHsaCP2000019 and dHsaCP2000020), *EGFR* exon 19 deletion screening assay[10], *EGFR* L858R (dHsaCP2000021 and dHsaCP2000022), *EGFR* G719X (validated laboratory-developed method using IDT (Coralville, IA, USA), *EGFR* wild-type for G719 (HEX), *EGFR* G719S and *EGFR* G719A (both FAM). The limit of blank and the limit of detection were determined for each individual assay using a Clinical and Laboratory Standards Institute (CLSI) EP17 protocol[11]. Results were analyzed using Quantasoft software version 1.6.6 (www.quantasoft.com). Supernatant results were normalized to the amount of cfDNA in 1 ml of fluid.

6

Table 1. Patient and sample characteristics

	All patients	All paired samples	EGFR exon 19 deletion	EGFR L858R	EGFR G719X	KRAS
			Paired samples per tumor driver mutation			
Patients/samples	39	44	15	4	1	24
Tumor type						
Non-small cell lung cancer	32	35	15	4	1	17
Colon carcinoma	4	4				4
Appendiceal carcinoma	2	2				2
Adenocarcinoma of unknown primary	1	1				1

Data are presented as n.

Results

The majority of the 44 paired samples were either both positive (30 out of 44) or both negative (8 out of 44) for the original driver mutation. In five cases the driver mutation was only detected in the supernatant and in one case only in the cell pellet **(Table 2)**. Thus, in 36 of the 44 paired samples a driver mutation was detected using both the supernatant and the cell pellet. Testing only the supernatant would detect 35 of the 36 drivers (97%) and analysis of only the cell pellet would detect 31 of the 36 drivers (86%). These results indicated that the supernatant was an excellent bio-source for ddPCR driver detection. Optimal sensitivity was reached when both the cell pellet and supernatant were analyzed.

In five paired samples the driver mutation was only detected in the supernatant. Reviewing the cytology reports for these cases showed that in two of these samples no tumor cells were seen, in two cases no cytologically analysis was performed, and in one case the tumor cell percentage was only 1%. In four of these cell pellets no mutant copies were found and in one the amount was below the limit of detection. In one sample only the cell pellet was positive for the primary driver mutation. This cell pellet had a very low estimated tumor cell percentage and a borderline result of only 1 mutant copy/µl in the cell pellet (**Table S1**). No mutant copies were detected in the supernatant by ddPCR. The cfDNA was isolated from 0.75 ml of supernatant and showed a very low concentration of 0.005 ng/µl measured by Qubit (Invitrogen, Carlsbad, CA, USA). Therefore, the mutation could easily be missed in our analysis.

To evaluate whether pleural effusion could be used as a bio-source for resistance analysis after progression on *EGFR* TKIs we analyzed 20 paired pleural effusions sampled from the 16 patients with *EGFR*-positive tumors for the presence of *EGFR* T790M. *EGFR* T790M was detected in seven out of the 20 paired samples. All seven supernatants and cell pellets were positive **(Table 3)**. Four out of the seven *EGFR* T790M mutation-positive pleural effusion samples had very low estimated tumor cell percentages and in one case no tumor cells were seen by the pathologist **(Table S1)**. Five of the seven pleural effusions were sampled from patients progressing on first-generation *EGFR* TKIs (erlotinib, gefitinib). Furthermore, two cases with a positive supernatant, but with a cell pellet showing a borderline result of 1 copy/µl, identified two patients with a durable response to osimertinib. The cytology reports of these pleural effusion samples showed that in one sample no tumor cells were seen and in one sample the tumor cell percentage was very low **(Table S1)**. These results indicated that the supernatant is a good bio-source to detect *EGFR* T790M. In four patients *EGFR* T790M had already been detected in tumor tissue samples by clinically validated

diagnostic assays. *EGFR* T790M was confirmed in all four matched supernatants and cell pellets. Two out of the seven samples were initially taken specifically for molecular resistance analysis by NGS/TSACP V1.0 (MiSeq; Illumina, San Diego, CA, USA) using the cell pellet: in one of these samples NGS could be successfully performed but no T790M mutation was found, while for the other sample in the end no NGS was done due to a low estimated tumor cell percentage. Both samples had tumor cell percentages ≥5% **(Table S1)**.

Table 2. Detection of original driver mutations by droplet digital PCR in paired samples (supernatant and cell pellet available) of pleural effusions.

Supernatant	Cell pellet		
	Driver positive	**Driver negative**	**Total**
Driver positive	30	5	35
Driver negative	1	8	9
Total	31	13	44

Data are presented as n.

Table 3. Detection of *EGFR* T790M by droplet digital PCR in paired samples (supernatant and cell pellet available) of pleural effusions

Supernatant	Cell pellet		
	T790M positive	**T790M negative**	**Total**
T790M positive	7	0	7
T790M negative	0	13	13
Total	7	13	20

Data are presented as n.

Discussion

This study investigated whether cfDNA from the supernatant of pleural effusions could be used for detecting driver mutations and *EGFR* resistance mutations. Our results clearly show that the cell-free fraction of pleural effusions is an excellent source for cfDNA and that it can be used to detect driver as well as resistance mutations effectively with ddPCR. We showed a high concordance rate between supernatant and cell pellet, which is in line with a previous study[6], but a slightly lower rate has been reported elsewhere[7]. Furthermore, our results suggest that by using cfDNA in pleural effusion as a bio-source, the molecular testing has gained in sensitivity and that optimal sensitivity can be reached by analyzing both the supernatant and the cell pellet. In addition to analysis of cfDNA in pleural effusion from NSCLC patients, we could also detect *KRAS* mutations in the supernatant samples from patients with colon carcinoma, appendiceal carcinoma and adenocarcinoma of unknown primary, which has not been reported before.

In five pleural effusion samples driver mutations were detected only in the supernatant even though no or very few tumor cells were seen in the cytopathological results. In our study, mutations were only called when the amount of copies found by ddPCR was above the limit of detection, determined according to the CLSI EP17 protocol[11]. The limit of detection was set with a confidence level of 99% to prevent false-positive results. Tumor DNA in the five corresponding cell pellets was not present or too low to detect the driver mutation. Importantly, since the detection of a driver mutation in the supernatant provides no proof of actual presence of tumor cells it can at present not be used for staging purposes. The origin of cfDNA in the supernatant has not been well studied, although several theories have been described[5,8,9]. Most likely this cfDNA is released from necrotic cells in the pleural cavity, but may also have been leaked from the circulation as a transudate. Alternatively, the cfDNA could have been derived from tumor cells damaged in the pre-analytical phase, *e.g.* by centrifugation[5], or from exosomes secreted by the tumor[8,9]. Therefore, a high detection rate in the supernatant is expected and also observed in our study. In one pleural effusion sample only the cell pellet was positive for the driver mutation. Most likely too little tumor DNA was present in the supernatant.

Resistance mutations are subclonal events in the tumor that often occur during TKI treatment. The *EGFR* T790M mutation, for example, is common in tumors of patients progressing on first-generation *EGFR* TKI (erlotinib, gefitinib)[2]. Osimertinib is frequently given as alternative treatment after this mutation is detected in a biopsy of a progressive lesion[2]. As obtaining tumor tissue is challenging and *EGFR*

T790M is a subclonal event, sensitive techniques are needed for molecular testing. Plasma genotyping is an alternative approach, but the sensitivity of detecting *EGFR* T790M in plasma is only 70%[11]. Therefore, other bio-sources for molecular testing could be useful. In our study, seven paired samples were positive for *EGFR* T790M, of which all were concordant between the supernatant and the cell pellet. Based on these results, both the supernatant and the cell pellet performed equally well as bio-sources for detecting *EGFR* T790M. The pleural effusion samples from four of the five patients progressing on first-generation *EGFR* TKIs showed a higher amount of T790M copies in both the supernatant and the cell pellet compared with the two patients showing a durable response to osimertinib. In prospective studies it will be interesting to see if the amount of *EGFR* T790M copies, measured by ddPCR in the supernatant and the cell pellet of pleural effusions, correlates to outcome in patients with *EGFR*-positive tumors treated with osimertinib.

This study is limited by the small sample size. The supernatant performed better as a bio-source than the cell pellet for detecting *EGFR* or *KRAS* primary driver mutations, but not for detecting *EGFR* T790M. Studies with a larger sample size are needed to confirm our results. Besides this, we could only compare a subset of the paired samples with results obtained in molecular diagnostics with NGS/TSACP version 1.0 using the cell pellet. A direct comparison with a larger set of paired samples could confirm the additive value of using cfDNA in the supernatant as a bio-source for molecular profiling in a clinical setting.

In conclusion, we showed that analysis of cfDNA in pleural effusion can robustly detect *EGFR* and *KRAS* driver mutations and *EGFR* resistance mutations. Therefore, cfDNA is a valuable bio-source for molecular testing, even when tumor cell purity is low.

Acknowledgements

References

1. Li, T., Kung, H.J., Mack, P.C., Gandara, D.R. Genotyping and genomic profiling of non-small-cell lung cancer: implications for current and future therapies. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2013;31(**8**):1039-49.

2. Cross, D.A. *et al.* AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer. *Cancer discovery.* 2014;4(**9**):1046-61.

3. Froudarakis, M.E. Pleural effusion in lung cancer: more questions than answers. *Respiration.* 2012;83(**5**):367-76.

4. Liu, D. *et al.* Malignant pleural effusion supernatants are substitutes for metastatic pleural tumor tissues in EGFR mutation test in patients with advanced lung adenocarcinoma. *PloS one.* 2014;9(**2**):e89946.

5. Lin, J., Gu, Y., Du, R., Deng, M., Lu, Y., Ding, Y. Detection of EGFR mutation in supernatant, cell pellets of pleural effusion and tumor tissues from non-small cell lung cancer patients by high resolution melting analysis and sequencing. *International journal of clinical and experimental pathology.* 2014;7(**12**):8813-22.

6. Asaka, S. *et al.* Rapid point-of-care testing for epidermal growth factor receptor gene mutations in patients with lung cancer using cell-free DNA from cytology specimen supernatants. *International Journal of Oncology.* 2018;52(**6**):2110-2118

7. Kawahara, A. *et al.* A combined test using both cell sediment and supernatant cell-free DNA in pleural effusion shows increased sensitivity in detecting activating EGFR mutation in lung cancer patients. *Cytopathology.* 2018;29(**2**):150-155

8. Shin, S., Kim,J ., Kim, Y., Cho, S.M., Lee, K.A. Assessment of real-time PCR method for detection of EGFR mutation using both supernatant and cell pellet of malignant pleural effusion samples from non-small-cell lung cancer patients. *Clin Chem Lab Med.* 2017;55(**12**):1962-1969

9. Lee, J.S., *et al.* Liquid biopsy using the supernatant of a pleural effusion for EGFR genotyping in pulmonary adenocarcinoma patients: a comparison between cell-free DNA and extracellular vesicle-derived DNA. *BMC Cancer.* 2018;18(**1**):1236.

10. Yung, T.K., Chan, K.C., Mok, T.S., Tong, J., To, K.F., Lo, Y.M. Single-molecule detection of epidermal growth factor receptor mutations in plasma by microfluidics digital PCR in non-small cell lung cancer patients. *Clin Cancer Res.* 2009;15(**6**):2076-84.

11. Oxnard, G.R, *et al.* Association between plasma genotyping and outcomes of treatment with Osimertinib (AZD9291) in advanced Non-Small-Cell Lunc Cancer. *Journal of Clinical Oncology.* 2016;334(**28**):3375-82.

6

Supplemental Material
Supplementary Tables

Table S1 *(part 1)*. Sample characteristics with initial clinical application in molecular diagnostics

	Diagnosis	Driver mutation (TISSUE)	*EGFR* T790M detected in tumor tissue	Reason for pleurocentesis
NKI-1	NSCLC	*EGFR* L858R		diagnostic
NKI-2	NSCLC	*KRAS*		diagnostic
NKI-3	NSCLC	*KRAS*		therapeutic
NKI-4	NSCLC	*KRAS*		therapeutic
NKI-5	NSCLC	*EGFR* L858R		therapeutic
NKI-6	NSCLC	*KRAS*		diagnostic
NKI-7	NSCLC	*KRAS*		therapeutic
NKI-8	appendiceal carcinoma	*KRAS*		therapeutic
NKI-9	NSCLC	*KRAS*		diagnostic
NKI-10	NSCLC	*EGFR* exon 19 del		therapeutic
NKI-11	NSCLC	*EGFR* exon 19 del	*EGFR* T790M	diagnostic
NKI-12	NSCLC	*EGFR* L858R		therapeutic
NKI-13	NSCLC	*EGFR* exon 19 del		therapeutic
NKI-14	NSCLC	*EGFR* exon 19 del		diagnostic
NKI-15	NSCLC	*KRAS*		therapeutic
NKI-16	NSCLC	*EGFR* exon 19 del		diagnostic

Molecular diagnostics performed in clinical practice				
Time of sampling	Clinical application	Assay	Results	Tumor cell percentage (estimated by pathologist on Giemsa slide)
progressive disease during erlotinib treatment	molecular profiling + T790M	NGS	*EGFR* L858R NO T790M	5%
	unknown	sanger seq	*KRAS*	10%
				20%
				20%
progressive disease during erlotinib treatment				30%
				50%
				60%
				60%
				60%
				unknown
progressive disease during gefitinib treatment	T790M	not possible		<5%
				<5%
				10–20%
	molecular profiling + T790M	NGS	*EGFR* exon 19 del NO T790M	20–80%
				40–50%
				high

6

Table S1 *(part 1)*. Continued

	Diagnosis	Driver mutation (TISSUE)	*EGFR* T790M detected in tumor tissue	Reason for pleurocentesis
NKI-17	NSCLC	*KRAS*		therapeutic
NKI-18	NSCLC	*EGFR* exon 19 del	*EGFR* T790M	therapeutic
NKI-19	NSCLC	*EGFR* exon 19 del		therapeutic
NKI-20	NSCLC	*EGFR* exon 19 del		therapeutic
NKI-21	NSCLC	*EGFR* L858R	*EGFR* T790M	therapeutic
NKI-22	NSCLC	*KRAS*		therapeutic
NKI-23	appendiceal carcinoma	*KRAS*		therapeutic
NKI-24	NSCLC	*KRAS*		therapeutic
NKI-25	NSCLC	*KRAS*		diagnostic
NKI-26	NSCLC	*EGFR* G719X		diagnostic
NKI-27	NSCLC	*EGFR* exon 19 del		diagnostic
NKI-28	colon carcinoma	*KRAS*		therapeutic
NKI-29	NSCLC	*EGFR* exon 19 del		diagnostic
NKI-30	NSCLC	*EGFR* exon 19 del		therapeutic
NKI-31	NSCLC	*EGFR* exon 19 del	*EGFR* T790M	therapeutic
NKI-32	NSCLC	*KRAS*		therapeutic
NKI-33	NSCLC	*KRAS*		diagnostic
NKI-34	NSCLC	*EGFR* exon 19 del		therapeutic
NKI-35	colon carcinoma	*KRAS*		therapeutic

| | Molecular diagnostics performed in clinical practice | | | |
Time of sampling	Clinical application	Assay	Results	Tumor cell percentage (estimated by pathologist on Giemsa slide)
				low
durable response to osimertinib treatment				very low
progressive disease during erlotinib treatment				very low
				very low
progressive disease during erlotinib treatment				unknown
				unknown
				unknown
				unknown
				50%
				5-10%
	molecular profiling	fragment analysis	*EGFR* exon 19 del	30%
				1%
	T790M	not possible		no tumor
				no tumor
durable response to osimertinib treatment				no tumor
				no tumor
				no tumor
				unknown
				unknown

6

Table S1 *(part 1).* Continued

	Diagnosis	Driver mutation (TISSUE)	*EGFR* T790M detected in tumor tissue	Reason for pleurocentesis
NKI-48	unknown origin	*KRAS*		therapeutic
NKI-49	NSCLC	*EGFR* exon 19 del		therapeutic
NKI-50	NSCLC	*KRAS*		therapeutic
NKI-51	NSCLC	*EGFR* exon 19 del		therapeutic
NKI-52	colon carcinoma	*KRAS*		therapeutic
NKI-53	NSCLC	*KRAS*		diagnostic
NKI-54	colon carcinoma	*KRAS*		diagnostic
NKI-55	NSCLC	*KRAS*		diagnostic
NKI-56	NSCLC	*KRAS*		diagnostic

Table S1 *(part 2).* Concentrations of the original driver mutation and *EGFR* T790M in the supernatant and/or the cell pellet measured by ddPCR

	Supernatant			
	Original driver copies /ml	WT original driver copies /ml	*EGFR* T790M copies /ml	WT *EGFR* T790M copies /ml
NKI-1	369	781	72	1298
NKI-2	28820	80300		
NKI-3	2110	7150		
NKI-4	103	418		
NKI-5	13662	27588	12474	26136
NKI-6	594	1474		
NKI-7	3018	11242		
NKI-8	8716	4906		
NKI-9	881	4077		
NKI-10	2160	13772	0	13200
NKI-11	2231240	2063600	1196800	3080000
NKI-12	90	2948	0	3322
NKI-13	23835	123464	0	155100
NKI-14	99568	17248	0	112068

	Molecular diagnostics performed in clinical practice			
Time of sampling	Clinical application	Assay	Results	Tumor cell percentage (estimated by pathologist on Giemsa slide)
				very low
				very low
				unknown
				no tumor
				no tumor
	molecular profiling	not possible		no tumor
				no tumor
	molecular profiling	not possible		no tumor
				no tumor

Cell pellet			
Original driver copies /μl	WT original driver copies /μl	*EGFR* T790M copies /μl	WT *EGFR* T790M copies /μl
22	1530	4	1613
300	8820		
3778	11780		
94	565		
341	4500	259	4670
955	1160		
325	13450		
41	22		
7265	12150		
165	4755	0	5035
865	5765	429	6045
210	22500	0	21840
483	1690	0	1950
7987	4195	0	11725

Table S1 *(part 2)*. Continued

| | Supernatant | | | |
| | Original driver | WT original driver | *EGFR* T790M | WT *EGFR* T790M |
	copies /ml	copies /ml	copies /ml	copies /ml
NKI-15	185790	40700		
NKI-16	25628	42680	0	29260
NKI-17	557	35596		
NKI-18	1329	3614	125	6226
NKI-19	4463800	9042000	45100	11550000
NKI-20	7121	16500	0	6710
NKI-21	2596	28622	1364	25872
NKI-22	25740	27940		
NKI-23	317900	1337600		
NKI-24	1650	1650		
NKI-25	113850	135300		
NKI-26	1102200	839300	0	1958000
NKI-27	18	1276	0	1364
NKI-28	167	2332		
NKI-29	84	6248	0	6578
NKI-30	13	968	0	1276
NKI-31	88	7193	44	6795
NKI-32	3718	29304		
NKI-33	2420	8778		
NKI-34	35	1518	0	1540
NKI-35	242	3322		
NKI-48	0	645		
NKI-49	0	1056	0	1034
NKI-50	0	9438		
NKI-51	0	1320	0	1012
NKI-52	0	15972000		
NKI-53	0	814		
NKI-54	0	42724		
NKI-55	0	566720		
NKI-56	0	7612000		

WT, wildtype

Cell pellet Original driver copies /µl	WT original driver copies /µl	*EGFR* T790M copies /µl	WT *EGFR* T790M copies /µl
25570	5340		
28568	34640	0	61840
8	980		
257	5	1	76
2193	7570	38	9200
531	8100	0	8850
334	1354	217	1542
622	3315		
468	2308		
124	1194		
810	2640		
13160	1080	0	13960
807	9330	0	10550
0	484		
1	138	0	7365
0	2820	0	2800
1	802	1	579
6	4870		
0	1300		
0	2625	0	2800
0	110		
1	317		
0	9060	0	9420
0	5990		
0	4770	0	5005
0	23980		
0	426		
0	27560		
0	9170		
0	6560		

6

PART 3

Summary and
future perspectives

Chapter 7

Summary and
future perspectives

Summary and future perspectives

The primary objective of this thesis was to identify novel biomarkers that can improve the pretreatment selection for PD-1 blocking agents in advanced non-small cell lung cancer (NSCLC), specifically for patients who are less likely to benefit from this treatment.

Chapter 2 of this thesis focused on the assessment of biomarkers that are direct effectors of the anti-tumor immune response. One of the direct effectors, PD-1^T tumor infiltrating lymphocytes (TILs), emerged as a key biomarker capable of predicting clinical benefit to PD-1 blockade in NSCLC with a high negative predictive value (NPV). **Chapter 3** assessed whether combining PD-1^T TILs with other established immune-related biomarkers could improve the predictive accuracy of PD-1^T TILs. To facilitate the clinical application of PD-1^T TILs as biomarker, **chapter 4** of this thesis focused on the development of a gene signature closely associated with PD-1^T TILs. This signature was constructed using NanoString technology, a robust clinical-grade platform known for its proficiency in mRNA profiling, particularly from formalin-fixed paraffin-embedded (FFPE) tissue samples. **Chapter 5** and **6** of this thesis highlight alternative bio-sources for biomarker testing, aimed to mitigate the need for invasive tumor biopsy procedures. All chapters will be discussed in more detail in the following sections.

Chapter 2; PD-1^T TILs as key effectors in the anti-tumor immune response, emerging as a promising biomarker beyond PD-L1

Pharmacological blockade of the inhibitory immune receptor PD-1 and its ligand PD-L1 has transformed the treatment landscape of advanced stage NSCLC. These immune checkpoint blocking (ICB) agents have demonstrated the capacity to induce durable responses, with estimated 5-year overall survival (OS) rates ranging from 16% to 23%[1]. Nevertheless, a substantial proportion of patients, approximately 60% to 70%, experience disease progression within the first six months after treatment initiation[2-4]. This raises concerns about the unnecessary exposure of patients to adverse effects, the financial burden, and the potential delay in exploring alternative (experimental) therapeutic options. Hence, the identification of biomarkers capable of selecting patients that will not derive benefit from PD-(L)1 blockade monotherapy has become an urgent necessity. Crucially, false negative test results should be minimized, as the impressive long-term responses observed with PD-(L)1 blockade have made clinicians hesitant to withhold this treatment from their patients. To achieve this goal, a sensitivity and NPV of ≥90% are considered necessary to avoid undertreatment. Concurrently, a specificity of ≥50% is required to effectively reduce overtreatment.

Tumor specific CD8[+] T cells possess the capacity to recognize and eliminate cancer cells. However, they are often functionally impaired in the tumor microenvironment (TME), for instance via the PD-1/PD-L1 signaling pathway. Therapeutic blockade of this pathway has the potential to reinvigorate dysfunctional T cells, which are characterized by high expression of inhibitory receptors such as PD-1, CTLA-4 and TIM-3, and secrete CXCL13. Furthermore, they show an increased capacity for tumor recognition[5–7]. We hypothesized that a new biomarker could be developed by assessing such direct effectors of the anti-tumor immune response. In **chapter 2** we validated PD-1[T] tumor infiltrating lymphocytes (TILs), a dysfunctional CD8[+] TIL subpopulation[7], as a novel biomarker for prediction of treatment benefit to PD-1 blockade in NSCLC. Importantly, the high NPV of this biomarker may allow for reliable identification of patients unlikely to benefit from PD-1 blockade. Furthermore, PD-1[T] TILs have been found across cancer types and have shown similar predictive potential[8], making it interesting to test this biomarker in other cancer types for which PD-1 blockade monotherapy is available.

Unlike PD-1[T] TILs, which serve as a direct indicator of an effective tumor-specific T cell response, the commonly used biomarker in routine clinical practice, PD-L1, is primarily considered a surrogate marker. In essence, PD-L1, expressed on the surface of tumor cells, binds to the PD-1 receptor on activated T cells, leading to an inhibitory effect. PD-(L)1 blocking agents target this interaction, thereby enhancing the 'pre-existing' anti-tumor immune activity. Nevertheless, a subset of patients with low PD-L1 expression or PD-L1 negative tumors can still show long-term disease control when treated with agents targeting the PD-1/PD-L1 axis[1,9]. Importantly, in **chapter 2**, we demonstrated that PD-1[T] TILs outperformed the PD-L1 tumor proportion score (TPS), and the combination of both did not improve predictive accuracy. This suggests that PD-1[T] TIL scoring has the potential to serve as a biomarker for response to PD-1 blockade monotherapy, irrespective of PD-L1 status.

Further, in **chapter 2,** we describe that PD-1[T] TILs predominantly reside in tertiary lymphoid structures (TLS)[7]. TLS are characterized by the presence of a central cluster of B cells surrounded by a border zone of T cells. They tend to form in chronic inflammation and in various cancer types, including NSCLC[10–12]. TLS have been associated with favorable responses and improved survival outcomes following ICB treatment[8,13–16]. We observed that using TLS as a single biomarker resulted in lower predictive accuracy compared to PD-1[T] TILs. It remains to be investigated whether factors such as spatial heterogeneity of TLS in peri- and intratumoral regions may account for this outcome. Our findings revealed that the frequency of TLS did not significantly differ between PD-1[T] high (≥90 per mm^2) and PD-1[T] low (<90 per

mm^2) tumors. In contrast, PD-1^T high tumors showed a higher number of PD-1^T TILs both within and outside TLS, potentially explaining the difference in predictive performance. These results suggest that the expansion of PD-1^T TILs, both within TLS and within the tumor parenchyma, is essential for an effective response to PD-1 blockade therapy. Digitally quantifying PD-1^T TILs in both tumor regions separately and across larger study cohorts is necessary to confirm these results. Additionally, spatial profiling of TLS-associated and intratumoral PD-1^T TIL subsets would be valuable to investigate whether TLS actively participate in the anti-tumor immune response during ICB treatment or if they are solely a characteristic of an inflamed TME. Therefore, gaining insight into how TLS shape the state and reactivation of tumor-specific T cells is imperative. This understanding could potentially improve biomarker development and, for instance, by inducing TLS formation, create opportunities for novel therapeutic strategies.

Chapter 2; confounding factors on the predictive potential of PD-1^T TILs

In clinical practice, biopsy sampling for biomarker testing is typically guided by factors such as the size of the lesion, accessibility, and the likelihood of obtaining accurate information. Usually, only one lesion is sampled, even in cases of advanced disease with multiple metastases. In stage IV disease, this commonly involves core needle biopsies or, in cases of oligoprogression, surgical resection of a metastasic lesion. Previous clinical trials of PD-(L)1 blockade monotherapy, which included PD-L1 immunohistochemical (IHC) testing, did not impose specific requirements regarding the type or location of the biopsy samples (i.e. primary versus metastases), resulting in a heterogenous mix of samples. PD-L1 IHC results can be influenced by several factors such as heterogeneity within the same tumor (spatial heterogeneity), sampling time between two different treatments (temporal heterogeneity), as well as differences in sampling between primary and metastatic tumor sites[18–21]. This leads to an important question: 'How should clinicians define the optimal sample that is most representative?'

To this end, we examined several potential confounding factors at the sample level for PD-1^T TILs. We observed a trend toward increased predictive accuracy of PD-1^T TILs when assessing tumor resections compared to biopsies. This aligns with expectations, as regional differences within the same tumor can be more precisely delineated in resected specimens. To investigate spatial heterogeneity, we quantified PD-1^T TILs in randomly selected small areas of resected specimens. While we did observe a certain degree of variability, the samples were classified as either PD-1^T high or PD-1^T low. We established this biomarker cut-off as the most predictive for discriminating treatment benefit from disease progression. Our findings indicate that it is possible to detect PD-1^T

TIL infiltration in relatively small biopsy samples, which is a promising development considering the challenges associated with biopsy sampling in NSCLC. In our study, we identified temporal heterogeneity, a phenomenon previously noted in PD-L1 IHC assessments[18,19]. Our results suggest that samples collected immediately before the initiation of PD-1 blockade monotherapy provided more accurate predictions of clinical benefit. It is worth noting that we excluded substantially older samples from our analysis, as archival samples might no longer accurately represent the tumor's immunological status. No difference in predictive accuracy was observed between samples collected from primary or metastatic sites.

In **chapter 2** we further highlight that lesion-specific response better correlated with the frequency of PD-1^T TILs compared to the overall response assessment by RECIST 1.1 criteria. Specifically, only a minority of the PD-1^T high lesions progressed compared to the PD-1^T low lesions. This observation may contribute to the reduced specificity and positive predictive value (PPV) of our biomarker, given that a significant number of tumors from patients with PD were PD-1^T high. Previous work by Osorio and colleagues examined the dynamics of individual metastases following PD-1 blockade monotherapy and demonstrated that progression tended to occur heterogeneously across different metastatic sites[22]. Therefore, the PD-1^T TIL status may reflect the local, and less likely the systemic anti-tumor immunity. Hence, we expect that the frequency of these tumor-reactive TILs may vary among different metastatic lesions.

While it may not always be clinically feasible to obtain fresh tumor tissue or biopsy multiple lesions within a patient, clinicians should recognize the limitations of biomarkers. Biomarkers can serve as valuable diagnostic tools in shared decision-making for therapeutic strategies. There is potential for further refinement of PD-1^T TILs as a biomarker by integrating it into multimodal predictors that comprehensively capture the tumor-immune microenvironment and other tumor-related factors. This aspect is explored in **chapter 3** of this thesis. Additionally, a more in-depth characterization of the tumor reactive TIL subset holds promise for the development of novel gene signature biomarkers, a subject that is investigated in **chapter 4** of this thesis.

Chapter 3; The additive value of biomarker combinations

The interaction between tumors and the immune system constitutes a complex spatiotemporal process, involving various stimulating and inhibitory factors acting at different stages of the cancer immunity cycle[23]. PD-(L)1 blockade therapy relies on promoting a dynamic anti-tumor immune response, and is not limited to targeting

a single oncogenic alteration or other autonomous features of cancer cells for which tyrosine kinase inhibitor therapy or chemotherapy is available. Therefore, we hypothesized that a predictive model should contain more than one biomarker to capture the complex interplay of different components within the TME. Previous studies have already shown the additive value of combining PD-L1 with tumor mutational burden (TMB)[24,25] and PD-L1 with CD8 TILs[26,27]. As demonstrated in **chapter 2**, combining PD-1[T] TILs with either 50% or 1% PD-L1 TPS did not improve predictive accuracy. A more comprehensive analysis was performed in **chapter 3**, where we evaluated individual and pairs of biomarkers as continuous variables. The tested biomarkers included various T cell markers (CD8 TILs, PD-1[T] TILs, CD3 TILs), a B-cell marker (CD20), TLS, PD-L1 TPS and the tumor inflammation signature (TIS). The TIS is an mRNA signature that has demonstrated predictive potential in different cancer types treated with PD-1 blockade monotherapy[28–30]. Our findings in **chapter 3** indicated that composite biomarkers did not provide improved predictive performance compared to the use of PD-1[T] TILs or TIS alone.

In contrast to PD-1[T] TILs, a distinct tumor-reactive TIL subset[7], we assessed general TIL density as this biomarker has been correlated to response to PD-1 blockade monotherapy in different cancer types[26,31–34]. In addition, the specific localization of CD8 TILs within the tumor and the peritumoral compartment (i.e. stroma) has been proposed as a biomarker for response to PD-blockade monotherapy[31,35]. In our study, combining CD8 or CD3 density with intratumoral localization of CD8 showed limited discriminatory ability in the validation cohort. Notably, PD-1[T] TILs showed the highest predictive performance among all the biomarkers, particularly in identifying patients with no long-term benefit. These findings align with other studies that have observed that not all TILs are in a state to recognize and eliminate tumor cells[7,36,37].

Chapter 3 also demonstrates that combining PD-1[T] TILs with any of the other biomarkers did not improve predictive capacity. This observation can be attributed to the interdependence between biomarkers, as they all reflect aspects of the anti-tumor immune response. Consequently, we propose combinations of PD-1[T] TILs with alternative biomarkers that capture distinct facets of the anti-tumor immune response. For example, TMB can serve as an indicator of immunogenic neoantigens resulting from somatic mutations[38]. Additionally, mutations in *STK11/LKB1* and *KEAP1* are associated with a TME characterized by T-cell exclusion, low or absent PD-L1 expression, downregulation of MHC class II in *STK11* mutated tumors, and down-regulation of type I interferon and other cytokines in *KEAP1* mutated tumors. These subgroups have been linked with primary resistance to PD-1 blockade monotherapy

7

in *KRAS* mutated NSCLC tumors[39]. Furthermore, high levels of *Beta-2-microglobulin* (*B2M*) mRNA, a component of MHC class I molecules, in baseline tumor samples have been correlated with an enhanced response to PD-1 blockade monotherapy in melanoma[40]. Lastly, the expansion of immune-inhibitory populations within the TME, such as myeloid-derived suppressor cells (MDSCs) and regulatory T cells (T$_{regs}$), has been associated with poor patient survival[41–43].

Chapter 4; A dysfunctional T cell gene signature that is more suitable for routine clinical diagnostics

In **chapter 2,** we reported a reliable automated method for digital quantification of PD-1^T TILs using IHC, based on an approach from earlier work[7]. However, IHC staining results may vary, primarily due to preanalytical factors such as tissue and slide preparation, and analytical factors such as antibody validation and the use of different staining platforms. Furthermore, a substantial degree of user involvement is necessary for the digital quantification of PD-1^T TILs, particularly as the delineation of tumor regions requires manual annotation. Computational pathology, using artificial intelligence (AI) techniques, holds the potential to enhance standardization across different medical centers. AI can perform complex visual pattern recognition by compiling subtle visual cues that are related to cell counts, cellular morphology, textures, and spatial patterns. For instance, the digital quantification of PD-L1 expression levels can be accomplished through IHC using machine learning techniques[44,45]. However, AI-based biomarkers come with several limitations. For example, it requires high-quality training data because the presence of artefactual images can introduce noise, necessitating a substantial number of images for the model to attain satisfactory performance. Furthermore, multiple datasets are needed to accurately represent diverse real-world patient populations, which can vary significantly between different hospitals. Therefore, data standardization and quality control systems are essential before application in routine clinical diagnostics[46].

An ideal platform for biomarker assessment in routine diagnostic applications should deliver sensitive, specific, and reproducible results while requiring minimal complexity for hands-on laboratory work and data analysis. The Nanostring nCounter platform has demonstrated its ability to meet this criteria, even when dealing with low-quality RNA samples obtained from FFPE samples[47]. This technique hybridizes fluorescent barcodes directly to specific nucleic acid sequences, enabling individual counting without the need of amplification, and it offers enhanced precision compared to microarray and RNA sequencing[47,48]. Importantly, the Nanostring nCounter platform has previously demonstrated its clinical applicability[49].

To facilitate the translation of the PD-1^T TILs biomarker into clinical practice, **chapter 4** aimed to develop a robust method suitable for use in routine clinical diagnostics. Using the Nanostring nCounter platform, we developed an mRNA signature that reflects a tumor's PD-1^T TIL status and predicts the clinical outcome of NSCLC patients treated with PD-1 blockade monotherapy. The PD-1^T gene expression signature showed equally high sensitivity and NPV as the digital image analysis-based IHC quantification of PD-1^T TILs. This mRNA signature enables accurate identification of patients with long-term benefit (DC 12m) from PD-1 blockade monotherapy as well as those with a low chance of benefitting from it.

The PD-1^T signature successfully matches the dysfunctional expression profile of PD-1^T TILs, as it includes the co-inhibitory signaling molecules *LAG3* and *CTLA4*. High expression of these genes, alongside *PD-1*, characterizes T cells in a late stage of differentiation marked by a loss of proliferative capacity and impaired cytotoxic function. This impairment is largely attributable to chronic tumor antigen stimulation within the TME[7,50]. Importantly, it is noteworthy that all of these markers have been correlated to tumor reactivity[7,36]. Interestingly, another gene in the PD-1^T signature is *CXCL13*, a chemokine constitutively expressed and secreted by PD-1^T TILs. *CXCL13* serves as a B cell attractant, contributing to the formation and maintenance of TLS. PD-1^{high}CXCL13$^+$ T cells predominantly localize in TLS[7]. A recent study has established a correlation between CXCL13$^+$CD8$^+$ T cells and favorable responses to ICB[51], underscoring the robustness of CXCL13 as a marker for tumor-reactive T cells and as a predictive biomarker. The PD-1^T signature also includes genes related to interferon signaling *(IFIT2, OAS1, STAT1)*, antigen presentation *(TAP1)* and angiogenesis *(HEY1)* as well as the cytokine *IL6* and the well known co-inhibitory molecule *CD274* (i.e. PD-L1). Collectively, these PD-1^T signature genes characterize both a pre-existing adaptive immune response and the immunosuppressive stimuli associated with T cell dysfunction[52,53]. As mentioned earlier, it is thought that PD-1 blockade therapy can reinvigorate these dysfunctional T cells.

We can conclude that the digital-image based IHC method can reliably be replaced by a matching gene expression signature. Since we developed the PD-1^T signature on the Nanostring nCounter platform, a well-established industry-standard technology, the biomarker is now more suitable for implementation. Importantly, **chapter 4** illustrates that the approach used has the potential to pave the way for the integration of other expression-level-based biomarkers into routine clinical diagnostics. These biomarkers can play an important role in supporting shared decision-making for therapeutic strategies.

Integration of the PD-1^T signature into prospective trials

Biomarkers hold a pivotal role in guiding treatment decisions, but how can we integrate them into clinical trial designs? The treatment landscape for advanced NSCLC has evolved, with ongoing investigations into the synergistic potential of immunotherapy combination therapies. A noteworthy development is the adoption of the combination of chemotherapy and pembrolizumab as the standard of care for newly diagnosed advanced NSCLC patients without targetable oncogenes. This is a result of the improved OS observed within the cohort receiving pembrolizumab plus chemotherapy, in contrast to those receiving chemotherapy alone. Importantly, this effect was observed irrespective of the patients' tumor PD-L1 status[54,55]. A debate persists concerning the methodological and essential aspects of PD-L1 IHC testing to determine the ideal candidates for combined chemotherapy and pembrolizumab versus those who should receive pembrolizumab monotherapy. Many physicians tend to favor pembrolizumab monotherapy for patients with a PD-L1 TPS of 50% or higher. This preference is based on the observation that pembrolizumab monotherapy offers improved tolerability and improved health-related quality of life compared to chemotherapy[54,57,58]. However, it is important to note that no trial has conducted a direct comparison between pembrolizumab plus chemotherapy and pembrolizumab monotherapy[59]. Consequently, the majority of patients in clinical practice are prescribed the dual therapy. This current approach underscores the need for more accurate and robust biomarkers in the field.

How can the PD-1^T signature be incorporated in a clinical diagnostic routine where PD-L1 IHC testing is already widely implemented? As previously mentioned, the predictive utility of PD-L1 remains limited to the subgroup characterized by a TPS ≥50%. The PD-1^T signature biomarker may be strategically used to improve patient stratification within this PD-L1 above 50% group (PD-L1^{high}) or in the PD-L1 below 50% group (PD-L1^{low}). To address these considerations, a prospective observational study could be devised to investigate the following clinical questions: 1) Is the omission of chemotherapy in the PD-1^T signature high (PD-1^{T-high}) + PD-L1^{low} group feasible? 2) Should chemotherapy be added in the PD-1^T signature low (PD-1^{T-low}) + PD-L1^{high} patients? The concurrent assessment of PD-L1 TPS and PD-1^T signature scores can be determined in parallel with minimal tissue sample requirements. A third clinical question of interest could be: Is leaving out pembrolizumab (so patients will only receive chemotherapy) in the PD-1^{T-low} + PD-L1^{low} group non inferior to the combination of chemotherapy plus pembrolizumab? This question is complicated to assess since the effect of chemotherapy on the tumor-reactive capacity of PD-1^T TILs is currently unknown. There is a possibility that chemotherapy creates an immunogenic environment, potentially priming new tumor-reactive T cells and

thus enhancing susceptibility to PD-1 blockade therapy. Historical outcome- and survival data of chemotherapy-only treated patients are needed to differentiate the distinct effects of chemotherapy and PD-1 blockade therapy. It is noteworthy that conducting a future phase three clinical trial excluding PD-1 blockade therapy in PD-1$^{\text{T-low}}$ + PD-L1$^{\text{low}}$ patients may face challenges. A subset of patients may not participate in the trial to avoid the risk of being randomized into the chemotherapy treated arm. The reluctance comes from the high demand for ICB treatment, and de-escalation studies encounter a lack of popularity. This sentiment refers to the principal that 'favoring more extensive treatment is a prevailing preference over opting for less intervention'. However, patients allocated to the chemotherapy-only arm should retain the option to receive PD-1 blockade monotherapy as a second-line intervention upon disease progression. Furthermore, PD-1 blockade overtreatment leads to unnecessary side effects in patients and health care systems deal with a substantial increase in costs. Consequently, it is crucial to improve personalized medicine to guide treatment decisions, thereby reducing costs. The resources saved could, for instance, be allocated to studies evaluating alternative treatment modalities tailored for patients who show primary or secondary resistance to PD-1 blockade therapy.

Chapter 5 and 6; Serum and pleural effusion as bio-sources for diagnostic biomarker tests

In **chapter 5** and **6** of this thesis we investigated alternative bio-sources for biomarker assessment. In current diagnostics, a tissue biopsy is a requisite in the majority of cases and, as such, continues to serve as a golden standard. However, it is an invasive, cumbersome procedure and small tumors may require multiple attempts to secure an adequate tumor tissue sample. In contrast, liquid biopsies offer a minimally invasive approach, enabling the retrieval of information from all tumor sites, thereby circumventing the loss of critical information attributed to tumor heterogeneity.

Mass spectrometry (MS) is a powerful tool for analyzing the proteomic landscape within blood samples. In combination with machine learning algorithms, this approach can be used to develop proteomic signatures with predictive capacity. In **chapter 5** of this thesis, we performed a MS-based proteomic analysis using pretreatment sera derived from 289 advanced-stage NSCLC patients who received second-line nivolumab treatment, with the aim of constructing a predictive protein signature. By using machine learning, which incorporated spectral data alongside clinical information, we stratified patients into three distinct groups with good ("sensitive"), intermediate, and poor ("resistant") outcomes following treatment. Our results demonstrated a strong association between the signature and PFS as well

as OS across two validation cohorts. In pooled analysis, a trend in better PFS was observed, and a significantly improved OS was demonstrated for patients categorized as "sensitive" compared to those categorized as "not sensitive". For "resistant" versus "not resistant" patients, both PFS and OS were significant independent predictors. In contrast, our test did not show the ability to stratify patients based on treatment outcome within a historical cohort of patients treated with chemotherapy.

Recently, Roder and colleagues conducted a blind and retrospective validation of our protein signature, now referred as the 'Primary Immune Response' (PIR) test, in the POPLAR and OAK trials[60]. This study also showed prolonged PFS and OS in the "not resistant" group compared to the "resistant" group in both cohorts. When comparing the "sensitive" and "not sensitive" groups, both PFS and OS demonstrated superiority in the POPLAR cohort. In the OAK cohort, only OS showed improvement, and these findings remained consistent even after adjusting for PD-L1 expression[60]. As previously described, it is of interest to assess the performance of the test in the present clinical setting, particularly considering that the standard first-line treatment for advanced NSCLC patients lacking targetable driver mutations now involves dual therapy comprising chemotherapy and pembrolizumab.

Advanced-stage NSCLC patients harboring tumors with molecular alterations such as *EGFR* mutations, *ALK* and *ROS1* rearrangements, or the *BRAF*^{V600E} mutation are eligible for specific targeted therapies. Therefore, it is imperative that all newly diagnosed patients receive comprehensive molecular tumor profiling. Also, the detection of resistance mechanisms during treatment with tyrosine kinase inhibitors (TKI) can guide the next line of therapy. In **chapter 6**, we emphasize the utility of cell-free DNA (cfDNA) in the supernatant of pleural effusion as an alternative bio-source, rather than relying on the cell pellet. In this study, driver and resistance mutations were detected with high accuracy in patients with NSCLC, colon and appendiceal cancer using droplet digital PCR. Notably, the supernatant showed a consistently high level of concordance with the cell pellet. Interestingly, its sensitivity surpassed that of the cell pellet, as mutations could still be detected in the supernatant even tumor purity was very low. As a result, the analysis of cfDNA in the supernatant has been institutionally implemented at the Netherlands Cancer Institute.

Conclusion

This thesis has made several contributions to the field of biomarkers and its diagnostic approaches related to predicting clinical outcomes in advanced-stage NSCLC patients treated with PD-1 blockade therapy. The key findings can be summarized as follows:

1) We have established the presence of a distinct population of tumor-reactive T cells, known as PD-1^T TILs, as a novel biomarker for long-term benefit to PD-1 blockade in NSCLC with high NPV.
2) We have observed that selected biomarker combinations did not improve predictive accuracy when compared to PD-1^T TILs as a single biomarker.
3) We have developed a robust PD-1^T gene signature, reflecting a tumor's PD-1^T TIL status, offering a more practical and readily implementable option for routine clinical diagnostics.
4) A serum-based protein signature has been constructed, allowing for outcome stratification without the need for invasive tissue biopsies.
5) Finally, we have demonstrated the efficacy of cfDNA extracted from the supernatant of pleural effusion for the accurate detection of targetable oncogenes and resistance mechanisms.

Together, these findings advance our capacity to tailor treatment strategies and diagnostic approaches for advanced-stage NSCLC patients undergoing PD-1 blockade therapy.

7

References

1. Garon, E. B. *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.* **37**, 2518–2527 (2019).

2. Brahmer, J. *et al.* Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N. Engl. J. Med.* **373**, 123–135 (2015).

3. Borghaei, H. *et al.* Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N. Engl. J. Med.* **373**, 1627–1639 (2015).

4. Rittmeyer, A. *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* **389**, 255–265 (2017).

5. Oliveira, G. *et al.* Phenotype, specificity and avidity of antitumour CD8+ T cells in melanoma. *Nature* **596**, 119–125 (2021).

6. Caushi, J. X. *et al.* Transcriptional programs of neoantigen-specific TIL in anti-PD-1-treated lung cancers. *Nature* **596**, (Springer US, 2021).

7. Thommen, D. 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.* **24**, (2018).

8. Voabil, P. *et al.* An ex vivo tumor fragment platform to dissect response to PD-1 blockade in cancer. *Nat. Med.* **27**, 1250–1261 (2021).

9. Herbst, R. S. *et al.* Long-term outcomes and retreatment among patients with previously treated, programmed death-ligand 1-positive, advanced non-small-cell lung cancer in the KEYNOTE-010 study. *J. Clin. Oncol.* **38**, 1580–1591 (2020).

10. Sautès-Fridman, C. *et al.* Tertiary lymphoid structures in cancers: Prognostic value, regulation, and manipulation for therapeutic intervention. *Front. Immunol.* **7**, 1–11 (2016).

11. Sautès-Fridman, C., Petitprez, F., Calderaro, J. & Fridman, W. H. Tertiary lymphoid structures in the era of cancer immunotherapy. *Nat. Rev. Cancer* **19**, 307–325 (2019).

12. Schumacher, T. N. & Thommen, D. S. Tertiary lymphoid structures in cancer. *Science* **375**, eabf9419 (2022).

13. Cabrita, R. *et al.* Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* (2020).

14. Helmink, B. A. *et al.* B cells and tertiary lymphoid structures promote immunotherapy response. *Nature* (2020).

15. Petitprez, F. *et al.* B cells are associated with survival and immunotherapy response in sarcoma. *Nature* **577**, (2020).

16. Vanhersecke, L. *et al.* Mature tertiary lymphoid structures predict immune checkpoint inhibitor efficacy in solid tumors independently of PD-L1 expression. *Nat. Cancer* **2**, 794–802 (2021).

17. Roelofsen, L. M., Kaptein, P. & Thommen, D. S. Multimodal predictors for precision immunotherapy. *Immuno-Oncology Technol.* **14**, 100071 (2022).

18. Boothman, A. M. *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.* **14**, 1390–1399 (2019).

19. Hong, L. *et al.* Programmed Death-Ligand 1 Heterogeneity and Its Impact on Benefit From Immune Checkpoint Inhibitors in NSCLC. *J. Thorac. Oncol.* **15**, 1449–1459 (2020).

20. Ilie, M. *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.* **27**, 147–153 (2016).

21. Gniadek, T. J. *et al.* Heterogeneous expression of PD-L1 in pulmonary squamous cell carcinoma and adenocarcinoma: Implications for assessment by small biopsy. *Mod. Pathol.* **30**, 530–538 (2017).

22. Osorio, J. C. *et al.* Lesion-level response dynamics to programmed cell death protein (PD-1) blockade. *J. Clin. Oncol.* **37**, 3546–3555 (2019).

23. Chen, D. S. & Mellman, I. Elements of cancer immunity and the cancer-immune set point. *Nature* **541**, 321–330 (2017).

24. Hellmann, M. D. *et al.* Genomic Features of Response to Combination Immunotherapy in Patients with Advanced Non-Small-Cell Lung Cancer. *Cancer Cell* **33**, 843-852.e4 (2018).

25. Rizvi, H. *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.* **36**, 633–641 (2018).

26. Fumet, J. D. *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* **119**, 950–960 (2018).

27. Althammer, S. *et al.* Automated image analysis of NSCLC biopsies to predict response to anti-PD-L1 therapy. *J. Immunother. Cancer* **7**, 1–12 (2019).

28. Ayers, M. *et al.* blockade IFN- γ – related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Invest.* **127**, 2930–2940 (2017).

29. Cristescu, R. *et al.* Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science (80-.).* **362**, (2018).

30. Damotte, D. *et al.* The tumor inflammation signature (TIS) is associated with anti-PD-1 treatment benefit in the CERTIM pan-cancer cohort. *J. Transl. Med.* **17**, 1–10 (2019).

31. Tumeh, P. C. *et al.* PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515**, 568–571 (2014).

32. Le, D. T. *et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N. Engl. J. Med.* **372**, 2509–2520 (2015).

33. Hu-Lieskovan, S. *et al.* Tumor characteristics associated with benefit from pembrolizumab in advanced non–small cell lung cancer. *Clin. Cancer Res.* **25**, 5061–5068 (2019).

34. Gataa, I. *et al.* Tumour-infiltrating lymphocyte density is associated with favourable outcome in patients with advanced non–small cell lung cancer treated with immunotherapy. *Eur. J. Cancer* **145**, 221–229 (2021).

35. Hashemi, S. *et al.* Surprising impact of stromal TIL's on immunotherapy efficacy in a real-world lung cancer study. *Lung Cancer* **153**, 81–89 (2021).

36. Simoni, Y. *et al.* Bystander CD8+ T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature* **557**, 575–579 (2018).

37. Scheper, W. *et al.* Low and variable tumor reactivity of the intratumoral TCR repertoire in human cancers. *Nat. Med.* **25**, 89–94 (2019).

38. Rizvi, N. A. *et al.* Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science (80-.).* **348**, 124–128 (2015).

7

39. Ricciuti, B. *et al.* Diminished Efficacy of Programmed Death-(Ligand)1 Inhibition in STK11- and KEAP1-Mutant Lung Adenocarcinoma Is Affected by KRAS Mutation Status. *J. Thorac. Oncol.* **17**, 399–410 (2022).

40. Zhao, Y. *et al.* B2M gene expression shapes the immune landscape of lung adenocarcinoma and determines the response to immunotherapy. *Immunology* **164**, 507–523 (2021).

41. Weide, B. *et al.* Myeloid-derived suppressor cells predict survival of patients with advanced melanoma: Comparison with regulatory T cells and NY-ESO-1- or melan-A-specific T cells. *Clin. Cancer Res.* **20**, 1601–1609 (2014).

42. Kim, H. J. & Cantor, H. CD4 T-cell subsets and tumor immunity: the helpful and the not-so-helpful. *Cancer Immunol. Res.* **2**, 91–98 (2014).

43. Hou, A., Hou, K., Huang, Q., Lei, Y. & Chen, W. Targeting Myeloid-Derived Suppressor Cell, a Promising Strategy to Overcome Resistance to Immune Checkpoint Inhibitors. *Front. Immunol.* **11**, 1–19 (2020).

44. Koelzer, V. H. *et al.* Digital image analysis improves precision of PD-L1 scoring in cutaneous melanoma. *Histopathology* **73**, 397–406 (2018).

45. Ebert, M. P. *et al.* Second-line therapy with nivolumab plus ipilimumab for older patients with oesophageal squamous cell cancer (RAMONA): a multicentre, open-label phase 2 trial. *Lancet Heal. Longev.* **3**, e417–e427 (2022).

46. Ghaffari Laleh, N., Ligero, M., Perez-Lopez, R. & Kather, J. N. Facts and Hopes on the Use of Artificial Intelligence for Predictive Immunotherapy Biomarkers in Cancer. *Clin. Cancer Res.* **29**, 316–323 (2023).

47. Veldman-Jones, M. H. *et al.* Evaluating robustness and sensitivity of the nanostring technologies ncounter platform to enable multiplexed gene expression analysis of clinical samples. *Cancer Res.* **75**, 2587–2593 (2015).

48. Geiss, G. K. *et al.* Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* **26**, 317–325 (2008).

49. Wallden, B. *et al.* Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. *BMC Med. Genomics* **8**, 1–14 (2015).

50. Wherry, E. J. & Kurachi, M. Molecular and cellular insights into T cell exhaustion. *Nat. Rev. Immunol.* **15**, 486–499 (2015).

51. Liu, B., Zhang, Y., Wang, D., Hu, X. & Zhang, Z. Single-cell meta-analyses reveal responses of tumor-reactive CXCL13 + T cells to immune-checkpoint blockade. *Nat. Cancer* **3**, 1123–1136 (2022).

52. Xia, A., Zhang, Y., Xu, J., Yin, T. & Lu, X. J. T Cell Dysfunction in Cancer Immunity and Immunotherapy. *Front. Immunol.* **10**, 1719 (2019).

53. van der Leun, A. M., Thommen, D. S. & Schumacher, T. N. CD8+ T cell states in human cancer: insights from single-cell analysis. *Nat. Rev. Cancer* **20**, 218–232 (2020).

54. Gandhi, L. *et al.* Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. *N. Engl. J. Med.* **378**, 2078–2092 (2018).

55. Paz-Ares, L. *et al.* Pembrolizumab plus Chemotherapy for Squamous Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **379**, 2040–2051 (2018).

56. Reck, M. *et al.* Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **375**, 1823–1833 (2016).

57. Wang, D. Y. *et al.* Fatal Toxic Effects Associated With Immune Checkpoint Inhibitors: A Systematic Review and Meta-analysis. *JAMA Oncol.* **4**, 1721–1728 (2018).

58. Reck, M. *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.* **37**, 537–546 (2019).

59. Peters, S., Reck, M., Smit, E. F., Mok, T. & Hellmann, M. D. How to make the best use of immunotherapy as first-line treatment of advanced/metastatic non-small-cell lung cancer. *Ann. Oncol.* **30**, 884–896 (2019).

60. Roder, H. et al. Abstract 26: Validation of the Primary Immune Response (PIR) test in advanced non-small cell lung cancer (NSCLC): blinded retrospective analyses from the POPLAR and OAK trials. *J. Immunother. Cancer* (2021).

7

Chapter 8

Nederlandse samenvatting

Achtergrond

Longkanker vormt wereldwijd een groot gezondheidsprobleem en staat op de tweede plaats van veelvoorkomende kankers. Daarnaast is longkanker de belangrijkste oorzaak van kanker gerelateerde sterfte. In 2019 werden in Nederland 14.000 patiënten met longkanker gediagnosticeerd en 10.000 patiënten kwamen te overlijden aan de gevolgen van longkanker. In gemiddeld 80% van de patiënten met longkanker is roken de oorzaak. Het meest voorkomende tumortype (85% van de patiënten) is niet-kleincellige longkanker (NSCLC, i.e. non-small cell lung cancer). Ongeveer de helft van de NSCLC-patiënten krijgt de diagnose longkanker in een gevorderd stadium (stadium IV), wat betekent dat er metastasen op afstand van de primaire tumor te vinden zijn. In dit stadium kan de ziekte niet meer chirurgisch worden behandeld en spelen systemische therapieën een centrale rol in het behandeltraject.

Historisch gezien bestond de behandeling van stadium IV NSCLC voornamelijk uit chemotherapie. Deze aanpak veranderde drastisch na de introductie van doelgerichte therapieën, welke gericht ontworpen zijn voor tumoren met een specifiek moleculair profiel. Deze therapeutische middelen, bekend als tyrosinekinase-remmers (TKI, i.e. tyrosine kinase inhibitors), lieten een significant verbeterde overleving zien in ongeveer 30% van de stadium IV NSCLC patiënten. Om patiënten te selecteren die mogelijk in aanmerking komen voor TKI-behandeling wordt nu in de Nederlandse richtlijn aanbevolen om een uitgebreide moleculaire analyse uit te voeren bij alle nieuw gediagnosticeerde patiënten met stadium IV NSCLC.

Voor de patiënten zonder oncogene driver mutatie heeft de ontwikkeling van immuuntherapie, met name de immuun checkpointremmers (ICI, i.e. immune checkpoint inhibitors), geleid tot een paradigmaverschuiving in de behandeling. Een voorbeeld van dit type behandeling is anti-PD-1 therapie, dat gericht is op het blokkeren van de interactie tussen PD-L1 (programmed death-ligand 1), een receptoreiwit dat tot expressie komt op tumorcellen, en PD-1 (programmed cell death protein 1), dat tot expressie komt op geactiveerde T- cellen. Hierdoor kunnen tumorcellen zichzelf minder goed beschermen tegen cytotoxische (CD8$^+$) T-cel-gemedieerde celdood. Hoewel anti-PD-1 therapie nu onderdeel uitmaakt van de eerstelijnsbehandeling voor patiënten zonder oncogene driver mutatie, ondervindt ongeveer 60-70% ziekteprogressie binnen zes maanden na de start van de behandeling. Dit gegeven benadrukt de noodzaak voor zogenaamde 'voorspellende biomarkers' die de respons op anti-PD-1 therapie kunnen voorspellen. Daarnaast is het essentieel om patiënten te onderscheiden die geen baat hebben bij anti-PD-1 therapie. Biomarkers met een hoge negatief voorspellende waarde (NPV, i.e. negative predictive value) kunnen

8

betrouwbaar aangeven welke patiënten geen therapeutisch voordeel zullen behalen. Een sensitiviteit en NPV van ≥90% wordt noodzakelijk geacht om onderbehandeling te voorkomen. Tegelijkertijd is een specificiteit van ≥50% vereist om overbehandeling substantieel te verminderen. Dit type biomarker biedt patiënten de mogelijkheid om in een vroege fase alternatieve behandelingsopties te verkennen. Daarnaast wordt het risico op bijwerkingen geminimaliseerd en draagt het bij aan het verminderen van de kosten van de gezondheidszorg.

De expressie van PD-L1 op tumorcellen is als een van de eerste biomarkers onderzocht. Hierbij wordt het percentage PD-L1 positieve tumorcellen geschat met behulp van een immuunhistochemische (IHC) kleuring voorafgaand aan behandeling. Verschillende onderzoeken hebben een positieve correlatie aangetoond tussen een hoge PD-L1 score (TPS, i.e. tumor proportion score) en verbeterde overleving na behandeling met anti-PD-1 therapie. Sindsdien wordt deze biomarkertest routinematig uitgevoerd in de diagnostiek. Echter zijn er tegenstrijdige resultaten die aangeven dat tumoren met een lage of negatieve PD-L1 TPS ook goed kunnen responderen op anti-PD-1 therapie. Daarom zijn er betere voorspellende biomarkers nodig.

Het primaire doel van dit proefschrift was het identificeren van biomarkers die beter kunnen voorspellen welke patiënten met stadium IV NSCLC kunnen responderen op anti-PD-1 therapie en welke patiënten niet.

Hoofdstuk 2 en 3; PD-1^T TILs als veelbelovende biomarker en de toegevoegde waarde van biomarkercombinaties

Tumor-specifieke cytotoxische CD8$^+$ T-cellen hebben de capaciteit om tumorcellen te herkennen en te elimineren. Echter kunnen ze functioneel beperkt zijn in de omgeving van de tumor (TME, i.e. tumor micro-environment), bijvoorbeeld via de PD-1-PD-L1 interactie. In de literatuur wordt verondersteld dat anti-PD-1 therapie disfunctionele T-cellen weer kan reactiveren. De frequentie van tumorinfiltrerende lymfocyten (TILs) in de TME heeft voorspellende waarde getoond als biomarker voor respons op anti-PD-1 therapie. Echter is er steeds meer wetenschappelijk bewijs dat niet alle TILs een actieve rol spelen in de immuunrespons tegen tumorcellen. Uit een eerdere studie bleek dat cytotoxische CD8$^+$ T-cellen met een hoge expressie van PD-1, bekend als PD-1^T TILs, gericht tumorcellen kunnen herkennen en elimineren. Bovendien is er een sterke correlatie aangetoond tussen de aanwezigheid van PD-1^T TILs en de respons op anti-PD-1 therapie in een klein cohort van patiënten met stadium IV NSCLC.

Hoofdstuk 2 van dit proefschrift beschrijft de voorspellende waarde van PD-1^T TILs in een cohort van 120 stadium IV NSCLC-patiënten behandeld met anti-PD-1 therapie. Hierbij werd de frequentie van PD-1^T TILs in tumorweefsel, afgenomen voor start van de therapie, digitaal gescoord met een PD-1 IHC-algoritme. PD-1^T TILs als biomarker behaalde een sensitiviteit van 77% en een specificiteit van 67% na 6 maanden behandeling zonder progressie van ziekte, en respectievelijk 93% en 65% na 12 maanden. Een ander belangrijk gegeven is dat de hoge NPV (88% op 6 maanden en 98% op 12 maanden) zorgde voor een nauwkeurige identificatie van een kleine groep patiënten die geen baat had bij anti-PD-1 therapie. Een hoog aantal PD-1^T TILs was bovendien gecorreleerd aan een langere progressievrije overleving (HR 0.39, 95% BI: 0.24-0.63, *P*<0.0001) en algehele overleving (HR 0.46, 95% BI: 0.28-0.76, *P*<0.01). De voorspellende waarde van PD-1^T TILs was ook nauwkeuriger wanneer alleen naar de respons van de gebioopteerde laesie werd gekeken en als tumorweefsel werd gebruikt dat kort voor de start van de behandeling was afgenomen. Daarnaast werd aangetoond dat PD-1^T TILs een nauwkeurigere biomarker was dan PD-L1. Een combinatie van beide biomarkers toonde geen verbetering in voorspellende waarde.

Uit een eerdere studie is bekend dat PD-1^T TILs voornamelijk voorkomen in tertiaire lymfoide structuren (TLS). TLS worden gekenmerkt door de aanwezigheid van een centraal cluster van B-cellen, omgeven door een zone van T-cellen. TLS kunnen worden gevormd ten gevolge van een chronische ontsteking, evenals door verschillende soorten kanker zoals NSCLC. **Hoofdstuk 2** beschrijft dat TLS een lagere voorspellende waarde heeft in vergelijking met PD-1^T TILs. Mogelijk hebben factoren zoals ruimtelijke heterogeniteit van TLS in peri- en intratumorale gebieden invloed op dit resultaat, echter moet dit nog worden onderzocht.

Verschillende factoren in de TME kunnen de immuunrespons tijdens anti-PD-1 therapie beïnvloeden. Daarom is het onwaarschijnlijk om één perfecte biomarker te vinden. In **hoofdstuk 3** is onderzocht of de voorspellende waarde verbeterd kan worden door biomarkers te combineren. De volgende biomarkers werden, zowel individueel als in combinatie van 2 biomarkers, onderzocht in een cohort van 135 stadium IV NSCLC patiënten behandeld met anti-PD-1 therapie: CD8$^+$ TILs, intratumorale (IT) lokalisatie van CD8$^+$ TILs, PD-1^T TILs, CD3$^+$ TILs, CD20$^+$ B cellen, TLS en de tumor inflammation signature (TIS). De TIS is een genexpressie biomarker opgebouwd uit een set van 18 genen. Eerdere studies hebben de voorspellende waarde van deze biomarker aangetoond in verschillende soorten kanker behandeld met anti-PD-1 therapie.

8

Hoofdstuk 3 concludeert dat samengestelde biomarkers geen toegevoegde voorspellende waarde hebben in vergelijking met PD-1^T TILs en TIS als individuele biomarkers. Dit was zowel voor het primaire eindpunt op 6 maanden, als voor het secundaire eindpunt op 12 maanden zonder progressie van ziekte. PD-1^T TILs en TIS identificeerde patiënten met een langdurige respons van meer dan 12 maanden met een hoge sensitiviteit (86% en 100%) en hoge NPV (95% en 100%). Patiënten die geen of weinig baat hadden bij behandeling met anti-PD-1 therapie werden het meest nauwkeurig geïdentificeerd door PD-1^T TILs (specificiteit PD-1^T TILs 74% versus TIS 39% op 12 maanden).

Hoofdstuk 4; Een PD-1^T signatuur dat geschikter is voor de dagelijkse diagnostiek
Hoofdstuk 2 beschrijft een digitaal PD-1 IHC algoritme dat het aantal PD-1^T TILs kan kwantificeren in tumorweefsel. Echter is deze methode complex en wordt beperkt door verschillende – voornamelijk technische – vormen van bias, waardoor de implementatie van deze biomarker in de diagnostiek lastig is. Voorbeelden van bias zijn pre-analytische factoren, zoals het voorbereidingsproces van weefsel tot coupe en analytische factoren, zoals de validatie van antilichamen en het gebruik van verschillende IHC-kleuring apparatuur. Bovendien vereist het digitaal kwantificeren van PD-1^T TILs handmatige annotatie van de tumorgebieden. **Hoofdstuk 4** onderzocht een betrouwbare methode om het signaal van PD-1^T TILs in tumoren te kunnen weergeven. Een voorwaarde is dat deze methode ook eenvoudig moet kunnen worden toegepast in de dagelijkse praktijk.

Een optimaal platform voor de evaluatie van biomarkers in diagnostische toepassingen moet gevoelige, specifieke en reproduceerbare resultaten opleveren. Tegelijkertijd is er voor laboratoriumwerk en data-analyse een minimale complexiteit vereist. Het Nanostring nCounter-platform heeft aangetoond te voldoen aan deze criteria, zelfs bij het verwerken van lage kwaliteit RNA-monsters verkregen uit FFPE- (formalin-fixed, paraffin-embedded) materiaal. Deze techniek hybridiseert fluorescerende barcodes rechtstreeks aan nucleïnezuursequenties, waardoor individuele telling mogelijk is zonder amplificatiestap. Bovendien zijn Nanostring analyses gevoeliger in vergelijking met microarrays en RNA-sequencing en heeft het platform eerder zijn klinische toepasbaarheid aangetoond.

Hoofdstuk 4 beschrijft de ontwikkeling en de validatie van een mRNA genexpressie signatuur dat de PD-1^T TIL status in tumoren detecteert en een voorspelling maakt van de klinische uitkomst van stadium IV NSCLC patiënten behandeld met anti-PD-1 therapie. Hierbij werd gebruik gemaakt van het Nanostring nCounter-platform. De resultaten van deze studie toonden aan dat de PD-1^T TIL IHC-biomarker succesvol kon worden omgezet in een mRNA genexpressie signatuur. De PD-1^T genexpressie

signatuur toonde namelijk een vergelijkbare voorspellende waarde als de digitale IHC-kwantificeringsbenadering. Een hoge NPV van 100% toonde aan dat patiënten met weinig of geen respons na behandeling met anti-PD-1 therapie betrouwbaar konden worden geïdentificeerd. Op basis van deze resultaten en het gebruikte platform wordt verwacht dat de PD-1^T signatuur gemakkelijker kan worden toegepast in de dagelijkse praktijk.

Hoofdstuk 5 en 6; Bloedserum en pleuravocht als biologische bronnen voor diagnostische biomarker-testen
Hoofdstuk 5 en **6** van dit proefschrift onderzochten alternatieve biologische bronnen voor het testen van biomarkers. In de huidige diagnostiek is een weefselbiopsie in de meeste gevallen vereist. Deze procedure is echter invasief, tijdrovend en bij kleine tumoren zijn er mogelijk meerdere pogingen nodig om voldoende tumorweefsel te verkrijgen. Daarentegen bieden vloeibare biopsieën, ook wel liquid biopsies genoemd, een minimaal invasief alternatief. Via het bloed kan bijvoorbeeld informatie vanuit alle tumorgebieden worden verzameld en is het verlies van cruciale informatie als gevolg van lokale tumorheterogeniteit beperkt.

Massaspectometrie (MS) is een gevoelige methode voor het analyseren van eiwitten in bloed. **Hoofdstuk 5** onderzocht bloedsera, afgenomen voorafgaande aan anti-PD-1 therapie van 289 stadium IV NSCLC patiënten met behulp van MS. Het doel was om een voorspellende eiwit-signatuur te ontwikkelen met behulp van machine learning, waarbij spectrale gegevens samen met klinische informatie werden geïntegreerd. Patiënten werden geclassificeerd in drie verschillende groepen: voorspelde goede ("sensitieve"), intermediaire en slechte ("resistente") resultaten na de behandeling. Er werd in twee validatiecohorten een sterke associatie gevonden tussen de eiwit-signatuur en zowel de progressie-vrije, als de algehele overleving.

Na het samenvoegen van de twee validatiecohorten werd een significant betere algehele overleving aangetoond in patiënten die als "sensitief" waren geclassificeerd in vergelijking met degenen die als "niet sensitief" waren geclassificeerd. Voor "resistente" patiënten ten opzichte van "niet resistente" patiënten waren zowel de progressie-vrije, als de algehele overleving significante onafhankelijke voorspellers. Daarentegen was de eiwit-signatuur niet voorspellend voor respons in een historisch cohort van patiënten behandeld met chemotherapie.

Stadium IV NSCLC patiënten met tumoren die moleculaire veranderingen tonen zoals *EGFR* of *BRAF* V600E mutaties en *ALK* of *ROS1* translocaties komen in aanmerking voor behandeling met TKI's. Daarom krijgen alle nieuw gediagnosticeerde stadium

8

IV NSCLC patiënten een uitgebreide moleculaire analyse van de tumor. Tevens kan het detecteren van resistentiemechanismen tijdens de behandeling met TKI's een eventuele volgende stap in de behandeling bepalen. **Hoofdstuk 6** beschrijft het nut van cel-vrij DNA (cfDNA, i.e. cell free DNA) in het supernatant van pleuravocht als alternatieve biologische bron voor moleculaire biomarker testen, in plaats van het cel sediment. De resultaten van deze studie laten zien dat mutaties in cfDNA betrouwbaar konden worden gedetecteerd in patiënten met NSCLC, colon- en appendixcarcinoom bij gebruik van droplet digital PCR. De resultaten van het supernatant kwamen overeen met de resultaten van het cel sediment. Echter was de gevoeligheid voor het detecteren van mutaties in het supernatant hoger dan in het cel sediment, omdat mutaties vaker konden worden gedetecteerd in pleuravochten met een zeer laag percentage tumorcellen. Naar aanleiding van de studieresultaten is de analyse van cfDNA in het supernatant institutioneel geïmplementeerd in het Nederlands Kanker Instituut.

Conclusie

Dit proefschrift heeft bijgedragen aan de kennis over voorspellende biomarkers en de bijbehorende diagnostische toepassingen bij patiënten met stadium IV NSCLC die zijn behandeld met anti-PD-1 therapie. De belangrijkste bevindingen kunnen als volgt worden samengevat:

1) We hebben aangetoond dat de aanwezigheid van een tumorreactieve T- cel populatie, bekend als PD-1^T TILs, kan fungeren als een voorspellende biomarker voor zowel langdurige respons als geen of nauwelijks respons op behandeling met anti-PD-1 therapie bij stadium IV NSCLC.
2) We hebben vastgesteld dat een combinatie van biomarkers de voorspellende waarde niet verbeterde in vergelijking met PD-1^T TILs als individuele biomarker.
3) We hebben een PD-1^T genexpressie signatuur ontwikkeld die de aanwezigheid van PD-1^T TILs in een tumor weergeeft. Deze biomarker biedt een praktische en een direct toepasbare optie voor de routinematige diagnostiek.
4) We hebben een op bloedserum gebaseerde eiwit-signatuur ontwikkeld waarmee klinische uitkomststratificatie mogelijk is zonder de noodzaak van een biopsie.
5) Ten slotte hebben we aangetoond dat cfDNA, geëxtraheerd uit het supernatant van pleuravocht, kan worden gebruikt voor het nauwkeurig detecteren van zowel moleculaire veranderingen in de tumor als resistentiemechanismen.

Concluderend bevorderen de bevindingen in dit proefschrift onze mogelijkheid om behandelingsstrategieën en diagnostische toepassingen op maat te maken voor patiënten met stadium IV NSCLC die worden behandeld met anti-PD-1 therapie.

8

APPENDICES

Appendix

Chapter 5 –
Supplemental material: methods

Methods

1. Acquisition of Mass Spectra from Serum Samples
    a. Processing of Serum Samples

Samples were thawed and 3 μl aliquots of each test sample and quality control/reference serum (a pooled sample obtained from serum of five healthy patients, "SerumP3") spotted onto serum cards (Therapak). The cards were allowed to dry for 1 hour at ambient temperature after which the whole serum spot was punched out with a 6mm skin biopsy punch (Acuderm). Each punch was placed in a centrifugal filter with 0.45 μm nylon membrane (VWR). One hundred μl of HPLC grade water (JT Baker) was added to the centrifugal filter containing the punch. The punches were vortexed gently for 10 minutes then spun down at 14,000 rcf for two minutes. The flow-through was removed and transferred back on to the punch for a second round of extraction. For the second round of extraction, the punches were vortexed gently for three minutes then spun down at 14,000 rcf for two minutes. Twenty μl of the filtrate from each sample was then transferred to a 0.5 ml eppendorf tube for mass spectral analysis.

All subsequent sample preparation steps were carried out in a custom designed humidity and temperature control chamber (Coy Laboratory). The temperature was set to 30 °C and the relative humidity at 10%.

An equal volume of freshly prepared matrix (25 mg of sinapinic acid per 1 ml of 50% acetonitrile: 50% water plus 0.1% TFA) was added to each 20 μl serum extract and the mix vortexed for 30 sec. The first three aliquots (3 x 2 μl) of sample: matrix mix were discarded into the tube cap. Eight aliquots of 2 μl sample: matrix mix were then spotted onto a stainless steel MALDI target plate (SimulTOF). The MALDI target was allowed to dry in the chamber before placement in the MALDI mass spectrometer.

    b. Qualification of Mass Spectrometer

The instrument qualification procedure is performed periodically to assess the reproducibility and quality of spectra compared to a defined 'Gold Standard' run. The machine qualification sample set used for this analysis contains 40 serum samples representative of expected diversity in human serum. The serum samples are prepared as a batch and spectra acquired as defined for the experimental samples. Spectral processing is performed using parameters defined in the machine qualification standard operating procedures, with methods similar to those

A

defined here for the experimental samples. A visual inspection of the spectra is performed. A concordance analysis compares a set of 85 feature values from the machine qualification batch with those of the Gold Standard run. Concordance plots for all 85 features, plotting the feature values for the run against those for the 'Gold Standard' run, are generated and examined. A summary statistic is computed, which assesses how close the slopes of the 85 concordance plots are to the ideal slope of 1. Qualification metrics assessing spectral quality and concordance of the spectra with the Gold Standard must be met for the mass spectrometer to be qualified for spectral acquisition from experimental samples.

c. Spectral Acquisition

Spectra were obtained using a qualified MALDI-TOF mass spectrometer (SimulTOF 100 s/n: LinearBipolar 11.1024.01 from Virgin Instruments, Sudbury, MA, USA). The instrument was set to operate in positive ion mode, with ions generated using a 349 nm, diode-pumped, frequency-tripled Nd:YLF laser operated at a laser repetition rate of 0.5 kHz. External calibration was performed using a mixture of standard proteins (Bruker Daltonics, Germany) consisting of insulin (m/z 5734.51 Da), ubiquitin (m/z, 8565.76 Da), cytochrome C (m/z 12360.97 Da), and myoglobin (m/z 16952.30 Da).

Spectra from each MALDI spot were collected as 800-shot spectra that were 'hardware averaged' as the laser fires continuously across the spot while the stage is moving at a speed of 0.25 mm/sec. A minimum intensity threshold of 0.01 V was used to discard any 'flat line' spectra. All 800-shot spectra with intensity above this threshold were acquired without any further processing.

2. Mass Spectral Processing

Spectral processing is necessary for two main reasons. First, it is used to average together many of the 800-shot 'raster' spectra that were collected on the mass spectrometer to create the Deep MALDI[1] averages. This allows a deeper and less noisy probing of the serum proteome. Second, spectral processing is essential to render the deep MALDI average spectra reproducible and comparable across samples.

a. Processing of Raster Spectra to Deep MALDI averages

Raster spectra were rescaled in the mass/charge (m/Z) axis relative to a standard reference spectrum to correct for any overall alignment issues in m/Z. To improve the signal to noise ratio of the spectra, a ripple filter was then applied. For a finer

adjustment of alignment in the m/Z axis, background was subtracted from the raster spectra and peaks are located. This peak list was then used to align raster spectra prior to background subtraction, using a set of 41 fixed alignment points, **Supplementary Table 1**.

Supplementary Table 1: Points in m/Z used to align the raster spectra

3168	7202	8919	12173	15127	28298
4153	7563	8994	12572	15263	
4183	7614	9133	12864	15869	
4792	7934	9310	13555	17253	
5773	8034	9427	13763	18630	
5802	8206	10739	13882	21066	
6433	8684	10938	14040	23024	
6631	8812	11527	14405	28090	

The raster spectra were then filtered to keep only spectra where at least 20 peaks were found and at least 5 of the alignment points in Supplementary Table 2 were used in their alignment. Deep MALDI, 400,000 laser-shot averages were then created by averaging together 500 randomly selected from the pool of filtered raster spectra.

b. Processing of Deep MALDI Average Spectra

"Spectrum" refers to a Deep MALDI average spectrum. The background in each spectrum was estimated and subtracted. Spectra were then normalized in several stages. First, a coarse normalization was performed using a partial ion current (PIC) approach using regions of the spectra that showed low variability across the population of interest and that showed no sign of association with outcomes. Spectral regions known to be intrinsically unstable were excluded from the regions used for normalization. The regions used in this normalization step were defined using the development set of spectra and then held fixed as fully specified parameters of the test. For each spectrum the area under the spectra within each m/Z region used for normalization was calculated and summed to produce a normalization coefficient. Each spectrum was then divided by its own normalization coefficient to yield the normalized spectrum. **Supplementary Table 2** gives the upper and lower limits in m/Z of the regions used in normalization.

Supplementary Table 2: m/Z regions used for coarse normalization

Lower limit m/Z	Upper limit m/Z
6100	7500
8500	10700
13100	15000

Spectra then underwent an additional baseline-smoothing background subtraction using Eilers' method (Eilers PHC, Boelens HFM. Baseline Correction with Asymmetric Least Squares Smoothing. 2005 Leiden Univ. Medical Centre Report.

https://zanran_storage.s3.amazonaws.com/www.science.uva.nl/ContentPages/443199618.pdf) aimed at increasing spectral reproducibility and limiting batch-to-batch variation and a second partial ion current normalization was carried using the m/Z regions in **Supplementary Table 3**.

Supplementary Table 3: m/Z regions used for second normalization

Lower limit m/Z	Upper limit m/Z
6931.26	6963.03
6963.58	6978.64
6979.01	7011.52
7012.07	7030.07
7030.26	7039.99
7066.99	7084.26
13864.91	13924.11
13925.45	13959.08
13959.98	14002.58
14008.41	14076.57
14077.02	14122.31
14124.55	14179.70
14180.60	14228.58
14229.93	14279.70
14280.60	14323.20
14412.88	14457.73
14464.45	14514.23
14516.47	14570.28
14571.18	14618.26
21006.87	21124.88
21125.85	21221.13
21221.61	21322.70

Although spectral alignment is typically very good, due to the alignment carried out in the raster processing, the averaged spectra were aligned to address minor differences in peak positions that might still be present. This was carried out using a fixed set of alignment points, which are listed in **Supplementary Table 4**.

Supplementary Table 4: Points in m/Z used to align the spectra

3315	8916	15872	28293
4153	9423	16078	
4457	9714	17256	
4710	12868	17383	
5066	13766	18631	
6433	14045	21069	
6631	14093	21168	
7934	15131	28084	

Features were defined on the mass spectra. These were defined by visual inspection of an indication representative set of spectra. A feature is an m/Z region in the mass spectrum, specified by its lower m/Z limit and its upper m/Z limit. While features were defined based on the location of mass spectral peaks in typical spectra, for any individual spectrum the feature may or may not contain a well-defined mass spectral peak. Once the features were defined, their definitions became parameters in the fully-specified test. For each feature and spectrum, a feature value was defined as the integrated area under the spectrum within the feature. For this test 282 features were defined, of which 274 were used during classifier development. The mass spectral ranges defining the 274 features used in classifier development are listed in **Supplementary Table 5.**

A

Supplementary Table 5: Mass spectral ranges (mass/charge regions) defining the 274 features used in classifier development

Lower m/Z	Center m/Z	Upper m/Z	Lower m/Z	Center m/Z	Upper m/Z	Lower m/Z	Center m/Z	Upper m/Z
3071.22	3085.19	3099.16	4397.29	4407.28	4417.26	5685.16	5693.39	5701.62
3099.64	3111.21	3122.77	4427.4	4433.22	4439.04	5701.82	5708.3	5714.77
3125.22	3137	3148.78	4439.75	4443.96	4448.18	5714.97	5720.49	5726.01
3149.02	3156.94	3164.86	4449.38	4461.23	4473.07	5726.03	5734.42	5742.81
3165.7	3177.13	3188.57	4502.53	4508.75	4514.98	5743.56	5750.24	5756.93
3189.67	3198.85	3208.03	4553.3	4565.57	4577.84	5757.29	5764.16	5771.03
3208.33	3216.82	3225.3	4580.87	4586.84	4592.81	5771.12	5778.62	5786.11
3231	3243.53	3256.07	4593.22	4599.59	4605.96	5786.29	5794.96	5803.62
3256.9	3267	3277.1	4618.51	4625.99	4633.46	5803.89	5810.08	5816.27
3305.44	3314.98	3324.51	4633.75	4642.42	4651.09	5816.42	5822.76	5829.11
3353.79	3366.37	3378.94	4667.54	4679.99	4692.43	5832.02	5840.46	5848.89
3384.77	3396.02	3407.27	4698.76	4713.31	4727.86	5850.08	5863.91	5877.73
3410.04	3422.21	3434.37	4747.49	4755.82	4764.15	5879.59	5888.74	5897.9
3434.51	3443.9	3453.3	4770.57	4776.34	4782.12	5898.07	5909.77	5921.47
3454.74	3466.38	3478.02	4782.62	4790.85	4799.08	5922.62	5934.73	5946.84
3540.72	3555.53	3570.35	4807.16	4819.05	4830.95	5949.41	5963.9	5978.4
3583.14	3593.09	3603.05	4845.9	4857.7	4869.49	5978.83	5987.76	5996.69
3667.06	3681.88	3696.7	4885.05	4893.33	4901.61	5998.01	6008.58	6019.14
3697.35	3705.33	3713.31	4910.19	4919.12	4928.06	6020.13	6028.93	6037.72
3747.09	3755.81	3764.52	4928.26	4938.24	4948.23	6054.61	6061.94	6069.27
3766.63	3776.4	3786.17	4949.44	4964.3	4979.15	6069.47	6082.86	6096.26
3811.87	3821.31	3830.74	4989.38	5000.07	5010.76	6099.57	6109.12	6118.68
3832	3841.64	3851.28	5012.17	5020.4	5028.63	6134.48	6148.68	6162.88
3860.09	3867.51	3874.93	5033.64	5041.17	5048.71	6165.75	6175.04	6184.34
3877.78	3888.14	3898.49	5048.95	5054.98	5061	6186.65	6194.45	6202.25
3899.28	3907.43	3915.58	5061.1	5070.88	5080.67	6202.33	6209.35	6216.38
3915.7	3927.75	3939.8	5093.87	5106.47	5119.06	6216.68	6224.86	6233.04
3943.3	3952.26	3961.21	5120.38	5127.97	5135.56	6275.16	6284.15	6293.14
3999.11	4011.41	4023.71	5162.99	5185.8	5208.61	6293.16	6301.49	6309.82
4023.91	4031.43	4038.96	5209.58	5224.44	5239.3	6322.27	6331.46	6340.64
4039.25	4051.4	4063.54	5240.4	5251.05	5261.69	6378.77	6393.09	6407.42
4080.14	4094.83	4109.53	5274.04	5288.09	5302.14	6409.41	6479.04	6548.68
4112.17	4119.37	4126.57	5351.59	5362.36	5373.12	6553.89	6564.68	6575.47
4127.25	4133.39	4139.52	5396.97	5404.07	5411.16	6575.85	6589.26	6602.67
4198.95	4210.81	4222.68	5411.52	5418.07	5424.63	6604.74	6675.06	6745.39

Supplementary Table 5: Continued

Lower m/Z	Center m/Z	Upper m/Z	Lower m/Z	Center m/Z	Upper m/Z	Lower m/Z	Center m/Z	Upper m/Z
4258.83	4266.5	4274.18	5424.89	5431.54	5438.18	6779.07	6798.12	6817.17
4276.79	4289.24	4301.69	5442.72	5449.54	5456.36	6825.83	6837.67	6849.52
4332.05	4341.63	4351.22	5512.62	5520.5	5528.38	6849.89	6859.44	6868.99
4351.4	4359.83	4368.27	5540.44	5552.25	5564.06	6869.08	6878.92	6888.75
4372.4	4381.03	4389.66	5564.15	5573.62	5583.09	6889.03	6896.99	6904.95

Lower m/Z	Center m/Z	Upper m/Z	Lower m/Z	Center m/Z	Upper m/Z	Lower m/Z	Center m/Z	Upper m/Z
6911.6	6920.97	6930.34	9066.55	9077.91	9089.26	11927.82	11938.26	11948.69
6930.88	6939.55	6948.22	9089.32	9096.91	9104.51	11974.12	11997.34	12020.56
6948.87	6956.18	6963.49	9112.47	9133.46	9154.45	12084.48	12116.9	12149.32
6963.58	6971.11	6978.64	9196.31	9207.88	9219.45	12151.24	12160.63	12170.03
6979.01	6995.27	7011.52	9234.27	9243.94	9253.6	12266.86	12290.16	12313.47
7011.77	7019.83	7027.88	9254.17	9263.3	9272.44	12552.61	12629.48	12706.34
7029.37	7033.6	7037.84	9272.68	9289.41	9306.14	12723.06	12738.33	12753.59
7037.91	7046.82	7055.73	9308.35	9319.83	9331.31	12769.89	12789.06	12808.24
7055.81	7060.15	7064.5	9341.1	9374.82	9408.53	12834.49	12917.52	13000.55
7065.49	7072.9	7080.31	9411.21	9454.03	9496.84	13018.32	13031.4	13044.48
7118.24	7143.95	7169.66	9560.23	9585.25	9610.26	13049.54	13076.86	13104.18
7178.66	7189.32	7199.97	9613.48	9626.56	9639.65	13119.56	13135.29	13151.02
7234.04	7243.67	7253.3	9639.94	9647.57	9655.2	13265.3	13276.12	13286.94
7279.59	7292.85	7306.11	9688.55	9723.57	9758.58	13304.84	13325.96	13347.09
7309.51	7318.12	7326.73	9903.45	9934.33	9965.21	13351.99	13364.15	13376.31
7327.41	7332.74	7338.06	10128.04	10139.87	10151.71	13501.19	13524.33	13547.48
7375.19	7390.07	7404.95	10152.46	10161.84	10171.22	13554.22	13569.52	13584.82
7406.19	7448.51	7490.84	10171.98	10184.57	10197.16	13602.38	13612.58	13622.78
7731.02	7736.79	7742.56	10197.54	10211.07	10224.6	13708.2	13723.6	13739
7742.75	7751.34	7759.93	10249.52	10262.23	10274.94	13740.4	13762.02	13783.64
7760.24	7767.77	7775.3	10295.62	10305.69	10315.75	13783.92	13795.98	13808.04
7776.52	7788.92	7801.31	10328.34	10350.14	10371.93	13832.96	13846	13859.04
7803.18	7820.32	7837.46	10435.64	10450.45	10465.26	13860.73	13881.13	13901.52
7984.8	7994.91	8005.01	10465.61	10482.62	10499.63	13905.76	13917.74	13929.71
8006.66	8018.69	8030.72	10518.75	10564.38	10610.01	13929.96	13944.37	13958.78
8131.01	8153.05	8175.09	10615.18	10638.37	10661.56	13959.98	13981.28	14002.58
8192.54	8215.68	8238.82	10711.79	10737.82	10763.85	14014.11	14067.59	14121.06
8306.66	8314.7	8322.74	10764.79	10775.15	10785.51	14122.86	14174.53	14226.2
8353.19	8366.02	8378.85	10828.47	10847.99	10867.5	14229.93	14254.82	14279.7

A

Supplementary Table 5: Continued

Lower m/Z	Center m/Z	Upper m/Z	Lower m/Z	Center m/Z	Upper m/Z	Lower m/Z	Center m/Z	Upper m/Z
8401.71	8411.17	8420.63	10951.44	10963.37	10975.3	14280.6	14301.9	14323.2
8420.71	8428.79	8436.87	11028.77	11056.4	11084.03	14401.51	14431.22	14460.94
8466.84	8474.84	8482.84	11090.89	11107.43	11123.96	14462.27	14541.41	14620.56
8483.32	8489.05	8494.77	11132.45	11152.43	11172.4	14623.06	14642.87	14662.69
8516.01	8528.96	8541.91	11285.82	11305.1	11324.39	14684.56	14699.66	14714.76
8555.29	8565.12	8574.94	11378.42	11392.26	11406.11	14764.89	14786.87	14808.84
8575.31	8592.03	8608.74	11428.16	11442.74	11457.32	14859.96	14882.15	14904.35
8650.35	8659.11	8667.86	11468.24	11485.3	11502.35	18248.47	18271.03	18293.59
8754.04	8766.76	8779.48	11513.71	11530.99	11548.26	18548.49	18570.16	18591.84
8799.09	8820.53	8841.97	11567.26	11584.42	11601.59	18603.02	18630.68	18658.34
8860.56	8871.76	8882.96	11611.34	11634.82	11658.3	18708.84	18730.03	18751.21
8882.98	8891.91	8900.84	11670.69	11686.46	11702.22	18811.43	18848.65	18885.87
8904.09	8925.16	8946.24	11719.74	11732.72	11745.69	19343	19378.06	19413.11
8954.36	8961.34	8968.33	11746.38	11756.13	11765.89	20886.02	21156.71	21427.39
8968.81	8978.23	8987.65	11769.8	11786.1	11802.4	21669.84	21804.69	21939.55
8988.02	8998.68	9009.33	11826.75	11843.48	11860.2	22566.7	22604.25	22641.79
9010.43	9019.53	9028.62	11876.81	11889.88	11902.95	22999.81	23033.14	23066.47
9028.78	9037.31	9045.84	11903.39	11913.25	11923.11	23097.51	23130.26	23163.02

Lower m/Z	Center m/Z	Upper m/Z
23213.01	23246.92	23280.82
23305.86	23353.04	23400.22
23429.2	23467.05	23504.91
25144.7	25185.27	25225.84
25429.35	25473.61	25517.87
25519.61	25570.37	25621.14
25624.4	25686.01	25747.63
27915.48	27962.78	28010.08
28037.85	28133.53	28229.2
28237.01	28338.55	28440.09
28800.67	28859.9	28919.13
28924.34	28972.72	29021.1
29030.65	29078.6	29126.55

To ensure that spectral data can be reproducibly generated, the QC/reference samples that were included at the beginning and end of each batch of samples run were used to batch correct the feature values of each batch of samples. Batch correction parameters were determined by comparing the feature values of the QC/reference samples within the batch to gold standard values for the QC/reference sample defined from those contained within the first batch of the development set samples that were run. Note that development set samples and test samples were not used in determining batch correction parameters. Once the batch correction parameters were obtained for a particular batch, all feature values for each spectrum in the batch were corrected. This process adjusted for small m/Z dependent changes in mass spectral performance of the mass spectrometer or between mass spectrometers.

The final step in processing of the spectra was another partial ion current normalization step. Using the development set of spectra, a subset of the 274 identified features was defined for final spectral normalization comprising features that had low coefficients of variation and were not associated with the clinical outcome variables in the development set. The m/Z regions (also features) used and fixed as parameters in the fully-specified test are listed in **Supplementary Table 6**.

Supplementary Table 6: m/Z regions used for final normalization

Lower limit m/Z	Upper limit m/Z
4351.40	4368.27
4372.40	4389.66
4553.30	4577.84
5396.97	5411.16
8353.19	8378.85
8401.71	8420.63
8954.36	8968.33
8968.81	8987.65
8988.02	9009.33
9028.78	9045.84
9089.32	9104.51
9196.31	9219.45
9254.17	9272.44
9639.94	9655.20
10128.04	10151.71
10171.98	10197.16
10197.54	10224.60
14401.51	14460.94

A

The final normalization coefficient was calculated for each spectrum and all feature values were divided by the normalization coefficient for the spectrum/sample to produce the final processed feature values. These were the feature values that were used for the development set samples for the creation of the test and these would be the feature values that would be input into the test classification algorithm when performing the fully specified test on a new sample.

3. Machine Learning: Development of the Classification Algorithm

A test classification is generated using a combination of 3 binary classifiers. Each classifier was created using the classifier development approach outlined in the next section.

a. A hierarchical classifier development platform designed for problems where the number of available instances is smaller than the number of measured attributes

Each classifier was created using a hierarchical classifier development platform designed specifically to work well in settings where the number of attributes (features) measured for each instance (sample) exceeds the number of instances available for classifier training. It incorporates aspects of traditional and modern machine learning, including bagging, boosting, and regularization using dropout, with the aim of producing classifiers with reliable performance estimates from relatively small sample sets while minimizing chances of overfitting to peculiarities in the development set data. This platform has been used in several other personalized medicine projects. Details of the approach can be found in Roder J, Roder H. Classification generation method using combination of mini-classifiers with regularization and uses thereof. United States patent US 2016; 9,477,906. (ed. Office, U.S.P.a.T.) (Biodesix, USA, 2016), and references [14] and [17].

b. Training Class Definition and a Semi-Supervised Approach to Simultaneous Refinement of Training Class Labels and Classifier

This approach used supervised learning, i.e., training class labels were needed for each instance (sample). To solve this particular classification problem using supervised learning one would need to know which patients received durable benefit from immune therapy and which did not. It was not *a priori* clear how to unambiguously define durable benefit from time-to-event data in a way that revealed underlying information in the molecular data. We employed an approach

that simultaneously refined training class labels for classifier development at the same time as the classifier itself. This approach is explained in detail in [17].

c. Classifier A: Redevelopment of BDX008 to use current mass spectral preprocessing and feature definitions

The development of BDX008 has been outlined in detail in Ascierto et al[13]. This used different spectral processing and feature definitions than we use here. Hence, an alternative version of this test was developed to allow standardization of spectral processing and mass spectral features across all three classifiers. The redevelopment used spectra from 113 of the 119 serum samples obtained from patients with melanoma prior to treatment with nivolumab originally used to develop BDX008. All 274 mass spectral features defined herein were used in the approach. All other parameters were kept the same as in BDX008 development.

d. Sample subsets and parameters used to define classifiers B and C

Supplementary Table 7: Parameters used in classifier development of classifiers B and C

	Classifier C	Classifier B
Samples used for development	96 development set samples that were not both BDX008+ and from a patients with performance status =0	76 development set samples that were not both BDX008- and in the poor prognosis group of Classifier C
k (for kNN)	11	11
# training/test split realizations	625	625
# dropout iterations	20,000	200,000
Averaging over dropout iterations	Weight-based	Weight-based
Filtering metric	PFS Hazard ratio between classes	PFS Hazard ratio between classes
Initial assignment of training classes	PFS time < median vs PFS time > median	PFS time < median vs PFS time > median
MS features used	29 features associated with Immune Response Type 2 (See Supplementary Table 8)	All 274 features

A

The mass spectral features associated with Immune Response Type 2 were determined by protein set enrichment analysis using the 49-sample reference set with both mass spectral data and protein expression data using the methods described in Grigorieva et al[21]. They are listed in **Supplementary Table 8**.

Supplementary Table 8: Subset of mass spectral features identified as being associated with immune response type 2 and used in development of Classifier C. The center m/Z rounded to whole Daltons is provided.

Rounded center m/Z	Rounded center m/Z	Rounded center m/Z
3111	5795	11584
3157	5935	11843
3177	6009	11890
3396	6175	11913
3888	8216	11997
4642	9934	13796
5186	10848	14787
5693	11056	14882
5734	11443	23353
5750	11531	

Appendix

Chapter 2 –

Commentary by Anagnostou V and Luke JJ in Clinical Cancer Research Quantitative spatial profiling of TILs as the next step beyond PD-L1 testing for immune checkpoint blockade

Commentary

Summary

Analysis of tumor-infiltrating lymphocyte (TIL) functional states, particularly tumor-reactive PD-1[T] TILs, within specific spatial context, can serve as a biologically informed predictive marker of immunotherapy that may be superior to standard clinical biomarkers. High-plex quantitative immune cell phenotyping within their spatial context has tremendous potential in immuno-oncology.

In this issue of *Clinical Cancer Research*, Hummelink and colleagues report on programmed cell death protein 1 (PD-1[T]) TILs, a tumor-reactive tumor-infiltrating T lymphocyte (TIL) pool, as a predictive biomarker for immunotherapy in non–small cell lung cancer[1] (NSCLC). PD-1[T] TILs represent an intratumoral CD8[+] T-cell population with high PD-1 expression, distinct transcriptional profiles, and increased tumor recognition capacity[2]. This subset of PD-1+ tumor-infiltrating T cells is preferentially recruited in tertiary lymphoid structures (TLS) and can be identified by bright PD-1 expression that can be digitally quantified and distinguished from other PD-1+ cells[1,2]. Hummelink and colleagues report their findings on the predictive accuracy of PD-1[T] TILs in the context of immune checkpoint inhibitor (ICI) therapy for patients with NSCLC receiving nivolumab or pembrolizumab. Following a digital workflow for PD-1[T] TIL quantification in formalin-fixed paraffin-embedded tissue, the authors evaluated the association of PD-1[T] TIL density with clinical outcomes, focusing on disease control at 6 months (a surrogate endpoint also known as durable clinical benefit[3]) as their primary endpoint. The predictive accuracy of PD-1[T] TIL density (AUC ROC, 0.72–0.79) was superior to that of programmed death-ligand 1 (PD-L1) TPS score, commonly used in NSCLC to identify tumors more likely to regress with ICIs (AUC ROC, 0.58). Notably, the predictive nature of PD-1[T] TILs may be enhanced for determining long-term clinical outcome and sustained clinical response past 6 months (ROC AUC, 0.79–0.89 for prediction of disease control at 12 months). As PD-1[T] TILs were predominantly found in TLSs[1,2] and the role of mature TLS in antitumor immune responses in the context of immune checkpoint blockade[4], the authors investigated the incremental value of assessing PD-1[T] TILs over the number of TLS within the analyzed tumors; these analyses showed that the predictive value of PD-1[T] TILs was not driven by TLS density alone (ROC AUC for the latter 0.62). Taken together, these findings build on the previously reported role of this functionally distinct subset of CD8[+] intratumoral T cells[2] and support PD-1[T] TILs as a putative determinant of response to immune checkpoint blockade and suggest that prospective validation in larger cohorts should be prioritized.

The study of Hummelink and colleagues, emphasizes that a nuanced spatially informed quantitative analysis, that captures T-cell populations with unique functional properties and tumor recognition capacities, may more accurately identify individuals more likely to respond to immune checkpoint blockade compared with conventionally used biomarkers. Currently established predictive biomarkers of ICI response include microsatellite instability[5] (MSI), which is detected in <5% of human cancers, as well as PD-L1 expression and tumor mutation burden (TMB), that both suffer from technical and biological limitations. The clinical utility of PD-L1 testing varies on the basis of the cancer type evaluated and the ICI therapy considered[6], with several phase III trials failing to reproduce the association between PD-L1 expression and ICI response[7,8]. Similarly, with the exception of MSI-high tumors, the predictive value of TMB is cancer-lineage dependent[9] and not consistently predictive of ICI response[10,11]. In contrast to PD-L1 expression or TMB that serve as surrogates of an antitumor immune response, PD-1^T TILs are an indicator that an effective tumor-specific T-cell response has occurred and can therefore serve as a biologically relevant measure of clinical outcomes. Furthermore, as PD-1^T TIL density was largely independent from PD-L1 TPS in the study by Hummelink and colleagues; it is conceivable that PD-1^T TIL density may be informative for PD-L1 negative tumors as well as tumors with PD-L1 TPS in the gray zone of 1% to 50% (**Figure 1**). Conceptually, PD-1^T TILs can be used as a footprint for active tumor-specific adaptive immune responses and therefore might enable patient selection for ICIs in cancers with marginal anti–PD-1 response rates, for example ovarian and breast cancer.

The value of TILs in reflecting adaptive antitumor immune responses and ultimately clinical responses with ICI therapy has been previously demonstrated[12], with emerging studies supporting the additive benefit of considering TIL functional profiles and their spatial localization within the tumor microenvironment (TME). To this end, spatially resolved multiplex immunofluorescence analyses have uniquely enabled spatial mapping of immune cells and assessment of their heterogeneity in the TME[13-15], revealing relationships among TIL subpopulations that are linked with differential ICI clinical outcomes[16]. Furthermore, evaluation of PD-1/PD-L1 proximity rather than PD-L1 expression alone may more optimally distinguish tumors more likely to regress with ICI therapy[17]. In addition to evaluation of the PD-1/PD-L1 axis, spatially resolved quantitative immunofluorescence approaches have the potential to interrogate interactions and localization of immunoregulatory molecules such as IDO-1, LAG-3, TIGIT, TIM-3, and VISTA, providing a unique opportunity to understand mechanisms of response and resistance to novel checkpoint inhibitors currently tested in clinical trials. Overall, these approaches have been shown to more accurately predict ICI response compared with PD-L1 expression and TMB[18].
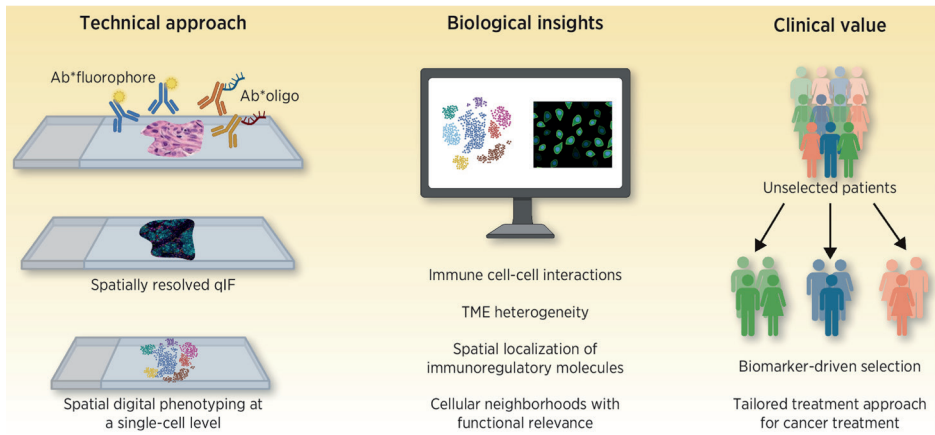
Figure 1. Impact and roles for spatially resolved high-plex assays in immunotherapy treatment and biomarker discovery. Spatially resolved high-plex methods, including quantitative immunofluorescence and digital spatial profiling, are high-throughput approaches that allow for simultaneous identification of multiple biomarkers in their spatial context. These methods have the unique potential to provide insights in the phenotype and spatial localization of immune cell subsets and thus serve as biology-informed biomarkers reflecting the quality and architecture of antitumor immune responses. As such, they can be incorporated in patient selection strategies for cancer immunotherapy and be used as a platform for novel biomarker discovery. (Adapted from an image created with BioRender. com.)

Similarly, spatial phenotyping by reconstruction of cellular neighborhoods has pointed towards local enrichment in immune cell subpopulations and differential organization of the TME that is reflective of distinct antitumor immunity states[19]. Implementation of photo-cleavable oligonucleotide tags attached to antibodies or RNA probes has further increased the multiplexing capacity, dynamic range and level of detection of digital spatial profiling approaches. Spatial transcriptomics represent another avenue of interrogation of immune cell spatial heterogeneity, with neoantigen-reactive T-cell clones shown to harbor unique transcriptomic profiles that are further differentiated in the TME of ICI responsive tumors[20]. While spatially resolved and high-plex assays may uniquely assess the immune contexture of tumors at a single-cell resolution, further standardization is required to generate analytical platforms that allow for measurement of complex spatial associations. Notably, these approaches are more likely to succeed when representative of spatial and functional interactions, following the paradigm of the study by Hummelink and colleagues that relied on interrogation of a TIL subset previously functionally characterized and found to be tumor-reactive[2].

Collectively, high-plex quantitative evaluation of immune cell subpopulation phenotypes, in their spatial context, holds unique promise as a near-term improved biomarker of treatment response and has tremendous potential for ICI biomarker discovery, especially for the subset of tumors with low PD-L1 expression and/or low TMB.

1. Hummelink, K. *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;**28**:4893–906.
2. Thommen, D.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**:994–1004.
3. Anagnostou, V. *et al.* Immuno-oncology trial endpoints: capturing clinically meaningful activity. *Clin Cancer Res* 2017;**23**:4959–69.
4. Fridman, W.H., Meylan, M., Petitprez, F., Sun, C.M., Italiano, A., Sautes-Fridman, C. B cells and tertiary lymphoid structures as determinants of tumor immune contexture and clinical outcome. *Nat Rev Clin Oncol* 2022;**19**:441–57.
5. Cercek, A. *et al.* PD-1 blockade in mismatch repair-deficient, locally advanced rectal cancer. *N Engl J Med* 2022;**386**:2363–76.
6. Doroshow, D.B. *et al.* PD-L1 as a biomarker of response to immune checkpoint inhibitors. *Nat Rev Clin Oncol* 2021;**18**:345–62.
7. Brahmer, J. *et al.* Nivolumab versus docetaxel in advanced squamous cell non-small cell lung cancer. *N Engl J Med* 2015;**373**:123–35.
8. Rittmeyer, A. *et al.* Atezolizumab versus docetaxel in patients with previously treated non-small cell lung cancer (OAK): a phase III, open-label, multicenter, randomized controlled trial. *Lancet* 2017;**389**:255–65.
9. Litchfield, K. *et al.* Meta-analysis of tumor- and T-cell–intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* 2021;**184**:596–614.
10. McGrail, D.J. *et al.* High tumor mutation burden fails to predict immune checkpoint blockade response across all cancer types. *Ann Oncol* 2021;**32**:661–72.
11. Anagnostou, V., Bardelli, A., Chan, T.A., Turajlic, S. The status of tumor mutational burden and immunotherapy. *Nature cancer* 2022;**3**:652–6.
12. Tumeh, P.C. *et al.* PD-1 blockade induces responses by inhibiting adaptive immune resistance. Nature 2014;**515**:568–71.
13. Berry, S. *et al.* Analysis of multispectral imaging with the AstroPath platform informs efficacy of PD-1 blockade. *Science* 2021;372.
14. Carstens, J.L. *et al.* Spatial computation of intratumoral T cells correlates with survival of patients with pancreatic cancer. *Nat Commun* 2017;**8**:15095.
15. Giraldo, N.A. *et al.* Spatial UMAP and image cytometry for topographic immuno-oncology biomarker discovery. *Cancer Immunol Res* 2021;**9**:1262–9.
16. Lopez de Rodas, M. *et al.* Role of tumor- infiltrating lymphocytes and spatial immune heterogeneity in sensitivity to PD-1 axis blockers in non-- small cell lung cancer. *J Immunother Cancer* 2022;10.
17. Gavrielatou, N. *et al.* Association of PD-1/PD-L1 co-location with immunotherapy outcomes in non-- small cell lung cancer. *Clin Cancer Res* 2022;**28**:360–7.
18. Lu, S. *et al.* Comparison of biomarker modalities for predicting response to PD-1/PD-L1 checkpoint blockade: a systematic review and meta-analysis. *JAMA Oncol* 2019;**5**:1195–204.
19. Schurch, C.M. *et al.* Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. *Cell* 2020;**183**:838.
20. Caushi, J.X. *et al.* Transcriptional programs of neoantigen-specific TIL in anti–PD-1—treated lung cancers. *Nature* 2021;**596**:126–32.

A

Appendix

List of publications

List of publications

Hummelink K, van der Noort V, Muller M, Schouten RD, van den Heuvel MM, Thommen DS, Smit EF, Meijer GA, Monkhorst K. Head-to-head comparison of composite and individual biomarkers to predict clinical benefit to PD-1 blockade in non-small cell lung cancer. *PloS One.* 2024 Jul 31;19(7):e0293707. Doi: 10.1371/journal. pone.0293707. eCollection 2024.

Van de Haar J*, Mankor JM*, **Hummelink K**, Monkhorst K, Smit EF, Wessels LFA, Cuppen E#, Aerts JGJV#, Voest EE#. Combining genomic biomarkers to guide immunotherapy in non-small cell lung cancer. *Clin Cancer Res.* 2024 Apr 1;30(7):1307-1318. Doi: 10.1158/1078-0432.CCR-23-4027. *Co-first author. #These authors jointly supervised this work.*

Hummelink K, Tissier R, Bosch LJW, Krijgsman O, van den Heuvel MM, Theelen WSME, Damotte D, Goldwasser F, Leroy K, Smit EF, Meijer GA#, Thommen DS#, Monkhorst K#. A dysfunctional T cell gene signature for predicting non-response to PD-1 blockade in non-small cell lung cancer that is suitable for routine clinical diagnostics. *Clin Cancer Res.* 2024 Feb 16;30(4):814-823. Doi: 10.1158/1078-0432.CCR-23-1061. #*These authors jointly supervised this work.*

Murray JC*, Sivapalan L*, **Hummelink K***, Balan A, White JR, Niknafs N, Rhymee L, Pereira G, Rao N, Weksler B, Bahary N, Phallen J, Leal A, Bartlett DL, Marrone KA, Naidoo J, Goel A, Levy B, Rosner S, Hann CL, Scott SC, Feliciano J, Lam VK, Ettinger DS, Kay Li Q, Illei PB, Monkhorst K, Scharpf RB, Brahmer JR, Velculescu VE, Zaidi AH#, Forde PM#, Anagnostou V#. Elucidating the heterogeneity of immunotherapy response and immune-related toxicities by longitudinal ctDNA and immune cell compartment tracking in lung cancer. *Clin Cancer Res.* 2024 Jan 17;30(2):389-403. Doi: 10.1158/1078-0432.CCR-23-1469. *Co-first author. #These authors jointly supervised this work.*

Muller M*, Best MG*, van der Noort V, Hiltermann TJN, Niemeijer AN, Post E, Sol N, in 't Veld SGJG, Nogarede T, Visser L, Schouten RD, van den Broek D, **Hummelink K**, Monkhorst K, de Langen AJ, Schuuring E, Smit EF, Groen HJM, Wurdinger T, van den Heuvel MM. Blood platelet RNA profiles do not enable for nivolumab response prediction at baseline in patients with non-small cell lung cancer. *Tumour Biol.* 2024;46(s1):S327-S340. Doi: 10.3233/TUB-220037. *Co-first author.*

A

Niknafs N, Balan A, Cherry C, **Hummelink K,** Monkhorst K, Shao XM, Belcaid Z, Marrone KA, Murray J, Smith KN, Levy B, Feliciano J, Hann CL, Lam V, Pardoll DM, Karchin R, Seiwert TY, Brahmer JR, Forde PM, Velculescu VE, Anagnostou V. Persistent mutation burden drives sustained anti-tumor immune responses. *Nat Med.* 2023 Feb;29(2):440-449. Doi: 10.1038/s41591-022-02163-w. Epub 2023 Jan 26.

Ibanez-Molero S, van Vliet A, Pozniak J, **Hummelink K**, Terry AM, Monkhorst K, Sanders J, Hofland I, Landeloos E, van Herck Y, Bechter O, Kuilman T, Zhong W, Marine JC, Wessels L, Peeper DS. SERPIN B9 is commonly amplified and high expression in cancer cells correlates with poor immune checkpoint blockade response. *Oncoimmunology.* 2022 Nov 29;11(1):2139074. Doi: 10.1080/2162402X.2022.2139074. eCollection 2022.

Hummelink K, van der Noort V, Muller M, Schouten RD, Lalezari F, Peters D, Theelen WSME, Koelzer VH, Mertz KD, Zippelius A, van den Heuvel MM, Broeks A, Haanen JBAG, Schumacher TN, Meijer GA, Smit EF, Monkhorst K[#], Thommen DS[#]. PD-1^T TILs as a predictive biomarker for clinical benefit to PD-1 blockade in patients with advanced NSCLC. *Clin Cancer Res.* 2022 Nov. 14;28(22):4893-4906. Doi: 10.1158/1078-0432.CCR-22-0992. [#]*These authors jointly supervised this work.*

Trebeschi S, Bodalal Z, Boellaard TN, Tareco Bucho TM, Drago SG, Kurilova I, Calin-Vainak AM, Delli Pizzi A, Muller M, **Hummelink K**, Hartemink KJ, Nguyen-Kim TDL, Smit EF, Aerts HJWL, Beets-Tan RGH. Prognostic value of deep learning-mediated treatment monitoring in lung cancer patients receiving immunotherapy. *Front Oncol.* 2021 Mar 2;11:609054. Doi: 10.3389/fonc.2021.609054. eCollection 2021.

Muller M*, **Hummelink K***, Hurkmans DP, Niemeijer AN, Monkhorst K, Roder J, Oliveira C, Roder H, Aerts JG[#], Smit EF[#]. A serum protein classifier identifying patients with advanced non-small cell lung cancer who derive clinical benefit from treatment with immune checkpoint inhibitors. *Clin Cancer Res.* 2020 Oct 1;26(19):5188-5197. Doi: 10.1158/1078-0432.CCR-20-0538. Epub 2020 Jul 6. *Co-first author. [#]These authors jointly supervised this work.*

Anagnostou V*, Niknafs N*, Marrone K, Bruhm DC, White JR, Naidoo J, **Hummelink K**, Monkhorst K, Lalezari F, Lanis M, Rosner S, Reuss JE, Smith KN, Adleff V, Rodgers K, Belcaid Z, Rhymee L, Levy B, Feliciano J, Hann CL, Ettinger DS, Georgiades C, Verde F, Illei P, Kay Li Q, Baras AS, Gabrielson E, Brock MV, Karchin R, Pardoll DM, Baylin SB, Brahmer JR, Scharph RB, Forde PM, Velculescu VE. Multimodal genomic features predict outcome of immune checkpoint blockade in non-small-cell lung cancer. *Nature Cancer.* 2020 Jan 13;(1):99-111. Doi: 10.1038/s43018-019-0008-8. Epub 2020 Jan 13. *Co-first author.*

De Vries R*, Muller M*, van der Noort V, Theelen WSME, Schouten RD, **Hummelink K**, Muller SH, Wolf-Lansdorf M, Dagelet JWF, Monkhorst K, Maitland-van der Zee AH, Baas P, Sterk PJ, van den Heuvel MM. Prediction of response to anti-PD-1 therapy in patients with non-small-cell lung cancer by electronic nose analysis of exhaled breath. *Annals of Oncology.* 2019 Oct 1;30(10):1660-1666. Doi: 10.1093/annonc/mdz279. *Co-first author.*

Hummelink K, Muller M, Linders TC, van der Noort V, Nederlof PM, Baas P, Burgers S, Smit EF, Meijer GA, van den Heuvel MM, van den Broek D, Monkhorst K. Cell-free DNA in the supernatant of pleural effusion can be used to detect driver and resistance, and can guide tyrosine kinase inhibitor treatment decisions. *ERJ Open Research.* 2019 Mar 25;5(1):00016-2019. Doi: 10.1183/23120541.00016-2019. eCollection 2019 Feb.

A

Appendix

Curriculum Vitae

Curriculum Vitae

Karlijn Hummelink was born on the 10th of January 1990 in Lichtenvoorde, the Netherlands. After graduating from Atheneum at the Marianum College in Groenlo in 2008, she started medical school at the Radboud University in Nijmegen. Her interest in pathology was sparked in 2012 during a scientific internship at the molecular neuropathology laboratory led by Professor Kenneth Aldape at the MD Anderson Cancer Center in Houston, Texas, USA. Throughout this internship, she conducted fundamental laboratory research focusing on the proneural to mesenchymal transition in glioblastoma. After obtaining her medical degree in October 2015, she started working as a pathology resident not in training at the Netherlands Cancer Institute in Amsterdam. In January 2017, she started on her PhD trajectory, aiming to identify biomarkers for clinical benefit to immune checkpoint blockade in advanced-stage non-small cell lung cancer. Under the supervision of Professor Gerrit Meijer (Pathology department, NKI) and Professor Egbert Smit (Thoracic Oncology, NKI), along with co-promotors Dr. Kim Monkhorst (Pathology department, NKI) and Dr. Daniela Thommen (Division of Molecular Oncology and Immunology, NKI), this endeavor resulted in the research presented in this thesis. During her PhD, she received a travel grant from the Nijbakker Morra stichting, enabling her to visit the genomics laboratory of Professor Victor Velculescu at the Johns Hopkins University in Baltimore, Maryland, USA in 2018.

Currently, Karlijn resides in Amsterdam and has been working as a resident in pathology at Amsterdam UMC since July 2021

A

Appendix

Dankwoord

Dankwoord

Toen ik in 2017 aan mijn PhD-traject begon, kon ik alleen maar dromen van dit eindresultaat. In de afgelopen zeven jaar heb ik ontzettend veel geleerd. Hoewel er soms pittige periodes waren, heb ik vooral veel plezier beleefd aan het tot stand komen van dit proefschrift. Graag neem ik nu de tijd om iedereen te bedanken die heeft bijgedragen aan dit werk.

Promotieteam

Kim, als eerste wil ik jou bedanken. Wij leerden elkaar kennen toen ik als ANIOS pathologie op de afdeling begon. Al snel werkten we samen aan verschillende biomarkerstudies, waarbij jouw enthousiasme en ambities mij overtuigden om aan een PhD-traject te beginnen. Ik denk met plezier terug aan onze lange gesprekken en discussies in jouw groene, met planten gevulde kantoor, waar altijd ruimte was voor een koffietje en een lolletje. Ik heb veel geleerd van jouw kritische, scherpe en pragmatische blik op het uitvoeren van translationeel biomarkeronderzoek. Bijvoorbeeld hoe je onderzoeksvragen zorgvuldig formuleert, de aannames die je maakt kritisch afweegt en deze vertaalt naar het klinische probleem dat je probeert op te lossen. Daarnaast heb je me geleerd om in de diagnostiek niet te verzanden in details, maar te focussen op de consequenties van je handelen voor de patiënt. Ook heb ik door jou meer inzicht gekregen in de moleculaire biologie van longkanker. Deze kennis zal ik zeker meenemen in mijn eigen verdere carrière. Dank ook voor je enorme toewijding tijdens dit lange traject. Het verliep niet altijd even snel, maar ik ben trots op de resultaten die we samen hebben bereikt!

Gerrit, dank voor je vertrouwen in mij. Ik herinner me nog goed hoe ik, onervaren en net afgestudeerd in de geneeskunde, bij jou binnenkwam. Je hebt me de vele facetten van translationeel onderzoek bijgebracht en wist me steeds bij te sturen als ik vastliep. Ook in moeilijke momenten kon ik bij je aankloppen voor goede raad. Je kon me wijzen op de hiaten in mijn kennis en me zo prikkelen om me meer te verdiepen. Ik bewonder je kritische blik en de enorme hoeveelheid kennis die je bezit, zowel als onderzoeker en als patholoog. Je wist altijd op het juiste moment de scherpste vragen te stellen. Ook zal ik jouw advies om in de pathologie niet te vervallen in picture matching, zeker meenemen in mijn verdere opleiding tot patholoog.

Daniela, since our first meeting during my second year of my PhD trajectory, I have witnessed the start of what has become a truly fruitful collaboration. At that time, you had just published your groundbreaking research on PD-1^T TILs in *Nature Medicine*, which inspired the promotion team to validate this promising biomarker

in larger NSCLC cohorts. I am incredibly grateful that you became my second co-promotor and that this topic has become a central part of my thesis. Thank you for sharing your exceptional knowledge about cancer and the immune system. I have also learned a great deal from you about structuring and writing a paper, as well as the importance of figures in illustrating findings. I admire your dedication and relentless pursuit of excellence in research. I wish you all the best in your future endeavors and will continue to follow your work with great interest.

Egbert, als vooraanstaande en ervaren longarts was jouw bijdrage aan mijn promotieteam van grote waarde. De resultaten in dit proefschrift waren niet mogelijk geweest zonder jouw waardevolle inbreng. We hebben interessante discussies gevoerd over de eisen waaraan biomarkers moeten voldoen in de huidige praktijk van (over)behandeling met immuun checkpoint remmers. Ik bewonder jouw kennis en tomeloze inzet om NSCLC-patiënten in Nederland de best mogelijke behandeling te kunnen bieden. Dank voor al je opbouwende feedback en het altijd vrijmaken van tijd om met mij te brainstormen.

Lees- en promotiecommissie
Prof. dr. **P. J. Van Diest**, prof. dr. **E. E. Voest**, prof. dr. **W. L. de Laat**, prof. dr. **M. M. van den Heuvel** en dr. **D. Cohen**, hartelijk dank voor de tijd die jullie hebben genomen om mijn proefschrift te lezen en voor jullie deelname als opponent tijdens mijn verdediging.

Co-auteurs
Heel veel dank aan alle **co-auteurs** die aan dit proefschrift hebben meegewerkt; zonder jullie zou dit eindresultaat niet mogelijk zijn geweest. Graag wil ik een aantal personen in het bijzonder uitlichten.

Vincent, aan het begin van mijn PhD zag ik behoorlijk op tegen de statistiek die bij translationeel biomarkeronderzoek komt kijken. Dankzij jouw enthousiasme heb ik onze meetings echter altijd als zeer plezierig ervaren en kijk ik terug op een fijne samenwerking. Ik heb veel geleerd en kan nu artikelen veel beter op waarde schatten. De kennis die ik heb opgedaan zal zeker van pas komen in mijn verdere carrière. Hartelijk dank daarvoor!

Mirte, dank voor de fijne samenwerking, de gezelligheid en de vele leuke uitstapjes, zowel in het binnen- als buitenland. Onze hoogtepunten waren ongetwijfeld de overnachting in het Amstel hotel en de trip naar Barcelona! Samen met **Robert** en **Michel** heb je een uitgebreide klinisch geannoteerde patiëntendatabase opgebouwd. Hierbij mijn dank aan jullie drieën voor het delen van deze database met mij.

Annegien, dank voor je betrokkenheid en hulp bij onder andere mijn PALGA-aanvragen voor het verzamelen van tumorsamples, de AVG-kwesties en alle laboratorium ondersteuning die ik heb ontvangen om mijn analyses te kunnen doen. **Dennis**, ook jij bedankt voor je laboratorium ondersteuning, in het bijzonder voor het opzetten van de PD-1 immunohistochemische kleuring.

Daan, dank voor de fijne samenwerking en het delen van je waardevolle kennis over liquid biopsies.

Renaud, thank you for your significant contribution to the development of the PD-1^{T} signature. I am very proud that we were able to publish this work in a reputable journal like *Clinical Cancer Research*.

TGO-groep, NKI-AvL
In het bijzonder wil ik alle collega's van de **TGO-groep** bedanken die mij in 2017 volledig opnamen in de groep, ondanks dat we niet direct samenwerkten aan een onderzoeksproject. Na een week hard werken heb ik enorm genoten van de gezellige vrijdagmiddagborrels in Radion, waar de beste verhalen op tafel kwamen na enkele (of meerdere) speciaal biertjes. De gezelligheid werd verder gekenmerkt door onze lunches om 11:37u stipt, de congressen, de boottochtjes door de Amsterdamse grachten, het squashen en het samen sushi eten. Jullie zijn een toegewijde en hechte groep die altijd klaarstond om mij te helpen wanneer ik ergens niet uitkwam. Ik bewonder de mooie wetenschappelijke successen die jullie hebben behaald en wens jullie het allerbeste voor de toekomst. Ik zal jullie werk zeker blijven volgen!

CFMPB, NKI-AvL
Ik ben alle collega's van de **CFMPB**, van administratieve medewerkers tot de analisten van het lab, enorm dankbaar voor de ondersteuning die zij mij hebben geboden bij al mijn projecten. Er moesten veel, heel veel samples worden ingeschreven, gekleurd en gesequenced worden. Zonder jullie hulp had ik dit niet kunnen realiseren!

Stafleden Thorax-oncologie, NKI-AvL
Dank aan alle **longartsen** van de thorax-oncologie voor de leerzame momenten tijdens de research meetings.

Opleiders Pathologie, AmsterdamUMC
Ellis, **Lianne** en **Elisabeth**, dank dat jullie mij de kans hebben gegeven om zowel voor als tijdens mijn opleiding extra tijd aan mijn proefschrift te besteden.

A

Buiten mijn werk om heb ik een hele fijne groep mensen om me heen die zeer dierbaar voor mij zijn en die ik voor geen goud zou willen missen! Dank dat jullie deel uitmaken van mijn leven. Nu wil ik graag het woord tot jullie richten.

Paranimfen
Lieve **Dorothée**, koffiemaatje, toen ik jou leerde kennen was mijn promotie al een paar jaar onderweg. Koffie heeft ons verbonden, en ik heb altijd genoten van onze koffiemomenten op het werk. Vooral in de ochtend als we nog even wakker moesten worden! Zelfs als ik het te druk had om met je mee te gaan, zorgde je ervoor dat ik koffie kreeg zonder dat ik erom vroeg. Eeuwige dank! Al snel leerden we elkaar beter kennen en groeide onze relatie van collega's naar vriendinnen die elkaar ook buiten het werk zien. We delen dezelfde humor en kunnen enorm lachen samen. Festivals bezoeken is altijd een feest met jou! Daarnaast konden we alles met elkaar bespreken, en was je mijn luisterend oor tijdens proefschriftperikelen. Dank dat je aan mijn zijde wilt staan als paranimf.

Lieve **Jeanine**, we go 'way back', al vanaf ons eerste jaar als studenten in Nijmegen tot nu; jij als gepromoveerde en ik als bijna gepromoveerde! Nadat we ons geneeskunde diploma hadden behaald, was jij degene die mij kennis liet maken met Amsterdam. Samen door de straten struinen, leuke eettentjes ontdekken en natuurlijk feesten. Je gaf me het gevoel dat ik me hier echt thuis zou kunnen voelen. Als mede-promovendus konden we alles met elkaar delen. Het was altijd een opluchting om mijn frustraties met je te kunnen delen of samen successen te vieren. Wij samen betekent gezelligheid, in een deuk liggen, dansjes wagen, lekker eten (en drinken), en alles met elkaar delen. Hoewel we inmiddels geen buren meer zijn, weten we allebei dat deze vriendschap voor altijd is! Ik ben heel blij dat je naast me wil staan als paranimf op deze bijzondere dag.

Vriendinnen
Lieve **Caroline**, **Angelique**, **Renske**, **Ilse** en **Imke** (Team Tequila), dank voor alle interesse en steun die ik gedurende mijn PhD-jaren van jullie heb ontvangen. We kennen elkaar al heel lang, en dat maakt deze groep zo bijzonder. Vanaf het eerste jaar geneeskunde hebben we samen vele onvergetelijke momenten, avonturen en mijlpalen beleefd. We wonen niet meer bij elkaar in de buurt maar we zien elkaar nog steeds regelmatig en dat waardeer ik enorm!

Lieve **Lies**, hoewel we elkaar pas als coassistenten hebben leren kennen, was dat het begin van een bijzondere en hechte vriendschap. We begonnen als sportmaatjes, maar al snel werd onze tijd vooral gevuld met veel geklets, samen eten, wandelingen

maken, shoppen, vakanties en vooral veel feesten. We klikten meteen omdat we zoveel gemeen hebben; we houden van dezelfde dingen en hebben zelfs dezelfde kledingsmaak (hoe vaak hebben we niet hetzelfde gekocht?)! Onze wilde tijden liggen inmiddels een tijdje achter ons en er is veel veranderd in ons werkende en gewone leven. Toch kan ik me een leven zonder onze vriendschap niet meer voorstellen. Het voelt altijd vertrouwd om weer bij jou te zijn.

Familie
Lieve familie Verhorst; **Rina**, **Leendert**, **Remco**, Leonie, Olivia, **Quinn**, **Lisette**, **Olger**, **Sofieke** en **Jona**, een warme familie met veel gezelligheid. Ik ben erg blij dat ik jullie dankzij Erik heb leren kennen.

Mijn roots liggen een flink stuk van Amsterdam vandaan, in de Achterhoek, in *Zöwent* zoals we daar zeggen. Daar is het echt *thuuskomm'n*. Ik ben ontzettend blij en trots om eindelijk tegen jullie te kunnen zeggen: het boekje is klaar! Lieve **Monique**, **Jeannette**, **Luuk** en **Koen** dank voor jullie onvoorwaardelijke steun gedurende dit lange traject. Ik geniet altijd van het samenzijn met de hele familie aan de lange tafel. De tafel die net groot genoeg is voor ons allemaal. Lieve **Owen** en **Elin**, jullie zijn nog klein maar ooit zal ik jullie vertellen over dit PhD-avontuur. Ik geniet altijd van jullie aanwezigheid en jullie zijn mij zeer dierbaar.

Lieve **Coen** en **papa**, we hebben al zoveel samen meegemaakt sinds mijn jonge jaren. Dank voor het vertrouwen, de steun en de liefde die ik altijd van jullie heb ontvangen, zowel in de mooie als in de moeilijke periodes van mijn leven. Jullie weten als geen ander hoe groot mijn wens was om dit PhD-traject te volgen en hoeveel werk het mij heeft gekost. Dit is waarschijnlijk mijn eerste en laatste boek, maar het betekent veel voor mij om deze mijlpaal samen met jullie te vieren.

Lieve **mama**, je bent er niet meer maar toch draag ik je altijd bij me. Mijn doel was om de wetenschap een stukje verder te brengen als eerbetoon aan jou. Ik ben er trots op dat dit gelukt is. Ik draag daarom dit proefschrift op aan jou.

Het slot van dit dankwoord is voor mijn rots in de branding, mijn grote liefde en beste vriend tegelijkertijd, lieve **Erik**. Jij hebt een groot deel van het proces naar dit eindresultaat meegemaakt. We hebben samen mijn publicaties gevierd, en je was er altijd om mij te steunen tijdens moeilijke momenten. Jij geeft mij rust en een echt thuisgevoel, een plek waar ik mijn batterij kan opladen en waar je me altijd weer kunt laten lachen. Geen woorden kunnen volledig beschrijven hoe dankbaar ik ben dat jij in mijn leven bent. Op naar nog vele jaren samen!

A